

# Katerino Mome

---



**Catherine Rudin**

# **Katerino Mome**

**Studies in Bulgarian Morphosyntax  
in Honor of Catherine Rudin**

**Edited by**

**Steven L. Franks  
Vrinda Chidambaram  
Brian D. Joseph  
Iliyana Krapova**

**Bloomington, Indiana, 2018**



© 2018 by the authors. All rights reserved.

This volume was published in September 2018.

ISBN 978-0-89357-490-2

Technical Editor: Renata Uzzell and Phillip Weirich

Cover Design by Tracey Theriault

Images: Front cover photograph of Balchik Botanical Gardens by Steven L. Franks. Back cover photograph of Catherine Rudin by Kristina Mihajlovic. Frontispiece photograph of Catherine Rudin courtesy of Ali Eminov.

#### Library of Congress Cataloging-in-Publication Data

Names: Franks, Steven, editor. | Chidambaram, Vrinda, 1980- editor. | Joseph, Brian D., editor. | Krapova, Iliana, editor. | Rudin, Catherine, 1954- honoree.

Title: Katerino Mome : studies in Bulgarian morphosyntax in honor of Catherine Rudin / edited by Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, Iliyana Krapova.

Description: Bloomington : Slavica Publishers, 2018. | Includes bibliographical references.

Identifiers: LCCN 2018033511 | ISBN 0893574902

Subjects: LCSH: Bulgarian language--Morphosyntax.

Classification: LCC PG870 .K38 2018 | DDC 491.8/159--dc23

LC record available at <https://lcn.loc.gov/2018033511>

Slavica Publishers  
Indiana University  
1430 N. Willis Drive  
Bloomington, IN 47404-2146  
USA

[Tel.] 1-812-856-4186  
[Toll-free] 1-877-SLAVICA  
[Fax] 1-812-856-4187  
[Email] [slavica@indiana.edu](mailto:slavica@indiana.edu)  
[www] <http://www.slavica.com/>

## Contents

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova	
Introduction .....	vii
Publications by Catherine Rudin .....	xv
Olga Arnaudova	
Clitic Doubling in Bulgarian .....	1
Loren Billings	
Ordering Restrictions if Both Internal Arguments are <i>Wh</i> -Phrases.....	21
Željko Bošković	
On Affixal Articles: An Argument from Bantu for Bulgarian, Romanian, and Icelandic .....	37
Wayles Browne	
Bulgarian Pronouns: What They Don't Distinguish that Most of Slavic Does .....	53
Vrinda Chidambaram	
Clitic-Doubles of Conjoined DP Objects in Macedonian and Bulgarian .....	65
Elena Dimova and Christine Tellier	
Bulgarian Multiple <i>Wh</i> Relatives Revisited .....	79
Steven L. Franks	
A Bulgarian Solution to the Slavic Q Question? .....	93
Victor A. Friedman	
The Pluperfect in Bulgarian and Macedonian: From <i>Bai Ganyo</i> to the <i>Bombi</i> .....	121
Iliyana Krapova and Guglielmo Cinque	
Genitive/Dative Case Puzzles in the Bulgarian DP .....	131
John Leafgren	
The Use of Long and Short Definite Articles in Bulgarian .....	165

Petya Osenova	
Heads and Dependents in Bulgarian .....	183
Roumyana Pancheva	
How Many Flowers! So Many Colors!	
Number Marking in Cardinality Exclamatives in Bulgarian .....	197
Teodora Radeva-Bork	
A Glimpse into the Acquisition of Bulgarian Morphosyntax:	
Pronominal Clitics .....	235
Lilia Schürcks	
Binding and Labeling in Bulgarian .....	251
Vesela Simeonova	
Two Declarative Complementizers in Bulgarian .....	271
Mila Tasseva-Kurktchieva and Stanley Dubinsky	
On the NP/DP Frontier: Bulgarian as a Transitional Case .....	287

## Introduction

We know Catherine Rudin as a linguist who specializes in the syntax of Bulgarian, both in the context of grammatical theory and in comparison with other languages of the Balkan region, especially Macedonian. But she is a person of many lives, with diverse interests and seemingly boundless energy. Catherine rarely misses “our” conferences—such as the Slavic Linguistics Society (SLS), Formal Approaches to Slavic Linguistics (FASL), the American Association of Teachers of Slavic and East European Languages (AATSEEL), or the Biennial Conference on Balkan and South Slavic Linguistics, Literature and Folklore—but she also takes part in meetings of many other kinds. These include not just familiar national linguistic gatherings such as the Linguistic Society of America, and sundry meetings of specialists in Native American languages (where she has an entire other life!), but also gatherings of dancers, singers, and folklore enthusiasts of various ilks. (“Dance and music” constitutes an impressive category on her CV.)

To us, her friends and colleagues, Catherine has been a tireless fixture in the field, writing and presenting, organizing meetings, and working tirelessly to advance the Slavic Linguistics Society: as Treasurer, Chair of the Board, and long-time Associate Editor for the *Journal of Slavic Linguistics (JSL)*. Thankfully, even after official “retirement,” she shows every sign of remaining so. She continues to be a popular invited speaker and her recent CV lists no fewer than six publications in press, in preparation, or to appear. And of course she continues to serve SLS and *JSL*. It is thus with great pleasure, pride, and esteem that we present this volume to Catherine in honor of her intellectual work, her professional engagement, and her unstinting friendship. We hope that this homage will inspire and delight.

Although we aimed to limit their content to Bulgarian morphosyntax, the papers included in these pages ended up reflecting the diversity of Catherine’s interests and, more importantly, the diversity of her friendships, even within our small field. What follows are short summaries of each of these contributions.

**Olga Arnaudova’s** paper gives an overview of clitic doubling in Bulgarian and argues that this phenomenon corresponds to two types of constructions in view of properties like degree of obligatoriness, distinctive features, and distribution. The author proposes assimilating each of them to a different applicative structure: a low applicative with an Experiencer argument projected in a clause-internal position and a high applicative base-generated above the

clause. Clitic Left Dislocation and Clitic Doubling structures are compared to structures involving focus-topic chains, where typically no doubling occurs.

**Loren Billings** explores the topic of multiple *wh*-questions and in particular sheds light on the ordering of *wh*-phrases in ternary *wh*-questions. He concludes that the ordering we find results not from inherent argument ordering preferences or arboreal asymmetry but rather from restrictions based on animacy and consecutive homophony.

The paper by **Željko Bošković** differs somewhat in orientation from the other contributions, in that rather than concentrating on Bulgarian (or even South Slavic), it uses well-motivated analyses of other, superficially very different languages as a foil to show that an otherwise credible account of article placement in Bulgarian would not in fact be typologically consistent. As such, the paper serves a useful methodological point, one that Catherine's more recent work also reflects: syntax is moving away from simple investigations of individual languages to comparisons of unrelated languages (as has indeed long been the norm for typology). This paper reveals how one can tease apart different analyses of Bulgarian by looking at Bantu, thereby illustrating a very promising (if surprising) research strategy.

In his paper **Wayles Browne** carefully examines Bulgarian personal and *wh*-pronouns. He shows that Bulgarian pronouns are not sensitive to certain contrasts, and therefore, obscure distinctions that would be clear in other related languages. For example, in Bosnian, Croatian, or Serbian, he observes, a personal pronoun can have a DP antecedent but not a clausal CP antecedent, whereas in Bulgarian, a pronoun can take either a DP or a CP as its antecedent. The paper considers five different ways in which Bulgarian pronouns can be ambiguous.

**Vrinda Chidambaram** addresses in her paper an intriguing topic that not only adds to our descriptive knowledge of Bulgarian and Macedonian but also has consequences of a theoretical nature. She deals with clitic doubling, focusing her attention on "what happens when the clitic-double corresponds to an object consisting of conjoined DPs differing in definiteness". The two languages differ in their treatment of such sentences, making for an important descriptive contribution within her comparative syntax approach. Moreover, this allows for a contribution to the theory of First Conjunct Agreement, insofar as that is the basis for the divergence between the two languages.

Building on Rudin's work and adopting her conclusion that multiple *wh*-relatives in Bulgarian are different from correlatives, **Elena Dimova** and **Christine Tellier** argue that the first *wh*-phrase in multiple *wh*-relatives is not in Spec, CP. The authors argue that it occupies a Topic projection given that it has the properties of a topicalized constituent, while the second *wh*-phrase is a free relative selected by both the main and the embedded verb. The proposed analysis offers a derivation of multiple relatives in terms of labeling, which



also explains the transparency of Topics for the purpose of selection in such contexts.

The paper by **Steven Franks** addresses the question of whether Bulgarian instantiates a QP category analogous to that of other Slavic languages in spite of its morphological impoverishment. To establish the depth of nominal structure, the author discusses the so-called *brojna forma* as a possible (surface) instantiation of Q, and on this bases his conclusion that this may indeed be the case for combinations of the *brojna forma* with numerals, classifiers, and quantifiers. In order to show that projecting a QP counts for the purposes of binding and c-command, the author examines Condition B and Condition C effects in nominal phrases. While the absence of the former is expected by the DP-analysis of Bulgarian, new data reported in the paper give a clue that it is precisely Q that is responsible for suppressing expected Condition C effects.

**Victor Friedman** engages in a comparative study of the functioning of the pluperfect tense in Bulgarian and in Macedonian. Working with translations of Aleko Konstantinov's 1895 novel *Bai Ganyo* (an appropriate corpus because Catherine Rudin worked with Friedman on an annotated English translation) and a present-day Macedonian corpus of wiretapped material, the *Bombi*, Friedman argues that the pluperfect is becoming obsolete in Macedonian while it remains alive and functional in Bulgarian. This then represents another way in which the two languages are diverging.

**Iliyana Krapova** and **Guglielmo Cinque** examine two puzzles in Bulgarian syntax and argue that these can be understood in terms of how DP-internal arguments receive abstract Case. The first concerns the differential ability of lexical and pronominal arguments (in event/process nominal DPs) to act as DP-internal subjects; the second concerns clitic doubling in ordinary object nominals. Their case-conflict account supports the universalist idea that all languages have the same abstract Cases but differ in terms of the morphological cases which ultimately spell them out.

**John Leafgren** tackles a long-standing issue in Bulgarian morphosyntax, namely, the ways in which the distinction between long and short forms of the definite articles is manifested in masculine singular nouns. He approaches the topic through a corpus-based study of contemporary formal and informal oral and written texts. He finds that a variety of factors, ranging from phonological to syntactic to discourse-controlled, play a role in the distribution of these variants

**Petya Osenova** considers the concept of syntactic head from various perspectives. She describes how different theoretical models approach the concept of head, asking, in particular, how heads relate to the phrases that contain them and how they relate to external phrases (in generative terms, complement and specifier phrases).

**Roumyana Pancheva's** contribution explores reasons behind the choice between count (*brojna*) and regular plural marking in Bulgarian, in particular

from the perspective of why the former fails to be used with the “cardinality” exclamative *wh*-expressions *kolkova*, *tolkova*, even though these same lexical items take count forms otherwise. She also treats differences between normative and colloquial variants. The account relies on the unusual assumption that count NPs are semantically singular, as well as on the proposal that the resistance of exclamatives to count-marked NPs follows from the idea that these are formed using a cardinality measure based on estimation (rather than counting *per se*).

**Teodora Radeva-Bork**’s paper is a study of children’s acquisition of clitics. She looks specifically at the acquisition of direct object clitics and clitic doubling. She outlines three experimental studies and their results, which show that direct object clitics emerge early and are used appropriately (i.e., used in place of full DPs, occur in the correct syntactic position, etc.) from an early age ( $\approx 2;3$ ), while CD emerges much later and with far less consistency with adult usage. The paper then turns to the theoretical implications of these results.

**Lilia Schürcks**, in her paper, reassesses the mechanisms governing the distribution in Bulgarian of the [+refl,  $-\phi$ ] forms *sebe si*, *svoj si*, *se vis-à-vis* the [+refl,  $+\phi$ ] forms *nego si*, *negov si*. She specifically examines c-command and the locality requirement within Binding Theory and the Minimalist Program, with the goal of reformulating her previously enunciated Degree of Markedness Spell-Out Principle.

In her paper, **Vesela Simeonova** considers the semantics of the complementizers *deto* and *če*, which largely occur in complementary distribution. She explores the contexts in which these complementizers are used and concludes that each one serves a distinct semantic function: while *deto* introduces content, *če* is an exemplifying complementizer.

Finally, **Mila Tasseva-Kurktchieva** and **Stanley Dubinsky** argue that even though Bulgarian may in some nominal contexts project a DP structure, this language is a weak DP language in that DP is not always required to project. To support their view, the authors discuss Neg raising and subject expletives and argue that these phenomena, in Bošković’s (2012) typology, characterize Bulgarian as an NP-language, while clitic doubling and obligatory number morphology characterize it as a DP-language. The proposal that Bulgarian is a weak DP-language brings out the possibility of reconsidering the NP/DP dichotomy into a tripartite typology: strong DP languages (English), weak DP-languages (Bulgarian) and strong NP-languages (Chinese).

In the remainder of this brief introduction, each of us offers a few personal comments and/or reflections on Catherine, thinking about first meetings, time spent together, Catherine’s contributions, and the like.

**Steven Franks** cannot recall when he first met Catherine, although it was surely at an LSA meeting in the early 1980s (probably at the 1981 meeting in New York, where she talked on “Bulgarian Free Relatives and the Matching Effect,” or the 1982 meeting in San Diego, where she talked on “Movement, Bind-

ing, and Island Conditions in Bulgarian Relative Clauses.” Despite a paucity of records from that period, in one of his recent office moves he came across a postcard from Catherine from April 1983 confirming his AATSEEL presentation and—in perfect Catherine form—asking about some Macedonian data in light of Bulgarian. Since that time their paths have crossed regularly; they co-authored a number of papers, and he even flew out to Nebraska to work with her on one of them. Catherine came to Bloomington many times, for conferences and workshops (where she was invariably a featured speaker), and she taught Bulgarian at the IU Summer Language Workshop (then SWSEEL) in 1993 and 1994. From those summers, Steven has fond memories of hikes with Catherine and her family at places like McCormick’s Creek State Park. She and her husband Ali Eminov both received their PhDs from IU–Bloomington and always enjoyed return visits. Indeed, she is one of their most prominent Linguistics PhDs, and is being honored with the department’s Distinguished Alumni Award in October 2018, joining the ranks of such notables as Ken Hale, Dell Hymes, and George Lakoff.

Steven emphasizes that Catherine has been instrumental from the start in making the Slavic Linguistics Society such a well-functioning and successful organization. Her constant dedication, discernment, and concern have long played an essential role in virtually all decision processes. Nothing is done without including Catherine, who sometimes points out potential difficulties and invariably adds vital feedback. She remembers everything we said at every meeting, recalls everything we promised in every e-mail, and follows through on everything quickly and conscientiously. She is, in a word, the driving force behind SLS. She has also been the longest serving Associate Editor for *JSL*, and even agreed to continue after Steven stepped down as Editor-in-Chief at the beginning of the year. In short, they have been a team, and he has relied on her insights over many, many years. (He says his name appears 13 times on her CV, so when repeatedly asked whether there was going to be a Festschrift for Catherine, he finally realized he needed to make it happen.) Steven also recalls how their joint research typically began: by being “stuck” together after a conference. Their work on Bulgarian clitics came from waiting for hours in the airport in Ottawa after the 2003 FASL meeting, and their work on universal concessive constructions came from hanging around at the end of the 2011 SLS meeting in Aix-en-Provence (which no one was in a rush to leave). But they also had a lot of fun together, touring ruined castles in the countryside near Heidelberg (SLS in 2015), or just walking until exhausted (Steven, not Catherine) around places like Szczecin (SLS in 2013) and Ljubljana (SLS in 2017).

Although **Vrinda Chidambaram** never took a class with Catherine, she can speak a bit about her tremendous patience as a teacher. One thing that Catherine does not always reveal about herself is that she is a phenomenal singer and a seasoned performer of Bulgarian folk song and dance. And

Vrinda, as an enthusiastic student of folk songs, was eager to learn from her. At the meeting of the Slavic Linguistics Society in Szczecin, Vrinda recalls how she and Catherine sneaked away from an afternoon session to sit on the steps of a neighboring building and sing. Vrinda had expressed nervousness about her upcoming talk, and Catherine suggested there was nothing better to strengthen her constitution than to take a break and sing. So they retreated to a quiet corner, and Catherine taught her to sing the melodies and harmonies of several Bulgarian folksongs. Among these was the popular “Katerino Mome,” inspiring Vrinda to choose it as part of the title for the present volume. Vrinda recalls that Catherine patiently repeated each phrase as many times as it took to sink in, until they could sing several songs in rich harmony at the final banquet of SLS. This became a new tradition: Vrinda sang with Catherine again at SLS in Ljubljana and she is looking forward to doing it in Eugene after this volume has been officially presented. Catherine is a person who communicates her love for and knowledge of Bulgarian language, music, and dance in a way that few others can. She is an expert who refuses to intimidate but rather insists on encouraging all those around her. This, just as much as genius in the study of language, is her gift.

**Brian Joseph** recalls first meeting Catherine in the early 1980s, during a summer when she was working in the Ohio State University Department of Slavic and East European Languages and Literatures (as it was then known) on a project to develop pedagogical materials for Bulgarian. (This project resulted in her co-authored 1984 book *Manual for Individualized Studies: Bulgarian Intermediate 2*.) He remarks what a pleasure it was to get to know her—a kindred spirit interested in Balkan syntax—and he notes how that pleasure has continued cyclically over the years at numerous Balkan, South Slavic, and Linguistics conferences in various venues. For him, she was then, just as she remains still, a refreshing voice examining aspects of Bulgarian and Balkan syntax from a formal perspective but always in a way that was suitably rigorous, empirically sound, yet theoretically nondoctrinaire. These are the traits which have given the scholarly community great confidence in her work and which have helped to give the work the visibility it so richly deserves.

But although she is best known for her early—and continuing—work on Bulgarian syntax, Brian recognizes that Catherine is so much more than that. As a linguist, her interests range over diverse languages such as Omaha-Ponca (Siouan) as well as various Balkan languages, including Turkish and diverse fields, especially the sociology of language with regard to the status of Turkish in Bulgaria and language documentation with regard to Omaha-Ponca. But as a person, she really shines, as she is an exceptionally talented singer and dancer who has thrilled those in attendance at conferences with her sometimes impromptu performances. Moreover, on a personal level, she is friendly and open and—as Brian can confirm from personal experience—a fine house guest. He is particularly pleased to be able to contribute to this recognition of

all that she has accomplished and the gracious—and graceful—way in which she has accomplished it.

**Iliyana Krapova** met Catherine in the fall of 1995 when she invited her to speak at the First Conference of Balkan and South Slavic languages, held in Iliyana's home town of Plovdiv. This was an emblematic conference since it took place just several years after the democratic changes of 1989, and the organizers wanted to celebrate the new era of opening up to the Western world. Catherine was also interviewed by the local radio station, and Iliyana remembers being struck by her revelation that her love for Bulgarian syntax was a follow-up on her passion for Bulgarian folk dancing. Later on, she had several occasions to see and talk to Catherine at conferences in Bulgaria and the U.S. With the years, she came to realize that Catherine is a very gentle, earnest, and down-to-earth person, a person of peace in heart and mind. These human qualities complement her work and are a source of inspiration for many.

Those who have followed in her footsteps (as well as those who have not) recognize Catherine Rudin as a pioneer of Bulgarian formal linguistics and one of the founding scholars of the study of Bulgarian syntax. Iliyana explains that, until the 1980s, syntax used to be a marginal theoretical subject for Bulgarian academics compared to traditional areas like phonology and morphology, so Catherine's contributions helped a lot in modernizing the field and changing fundamental aspects of the way younger generations now think about the importance of doing syntax. The periods of research Catherine spent in Bulgaria in the early 1980s, in particular at the Institute for Bulgarian Language, as well as her discussions with the late Jordan Penchev, led to interesting discoveries. Unexpected facts and properties of the language (especially with respect to *wh*-words, focus, topic, and complementizers) were brought to light and received their transformational-generative labels as well as a thorough analysis in her 1982 Indiana University dissertation, published by Slavica in 1986 as *Aspects of Bulgarian Syntax*. As happens with ground-breaking work in general, this book has not lost its value over time. (Slavica released a second revised edition in 2013.) Many researchers still refer to it not only for its analyses, but also for Catherine's solid descriptions of the many phenomena relevant to the organization of the left periphery and the discourse properties of Bulgarian syntax. Catherine has a keen eye for data, which is why her formal descriptions have long constituted a suitable basis for further work. Her contributions on issues as diverse as clitics and clitic doubling, multiple *wh*-fronting in questions and relatives, *da*-clauses, and *li*-questions are all well known to the international community of Slavic and general linguists. Special mention should be made of her 1988 *NLLT* paper "On Multiple Questions and Multiple *Wh*-Fronting," which opened up an extremely productive line of research into many issues of syntactic variation in Slavic and beyond.

Besides the four of us, many other people worked hard to make this volume possible. The editors wish to thank Slavica Publishers for seeing this

project through, Catherine's sons Deniz and Adem and her husband Ali for their assistance in various matters, and also Dean Yasuko Taoka (Wayne State College, School of Arts and Humanities) for securing financial support. Finally, we acknowledge the help of the following additional individuals for providing valuable feedback on one or more of the papers in this volume:

Loren Billings  
 Barbara Citko  
 Miloje Despić  
 Mila Dimitrova-Vulchanova  
 Elena Dimova  
 Lydia Grebenyova  
 Boris Harizanov  
 Christina Kramer

John Leafgren  
 Olga Mladenova  
 Roumyana Pancheva  
 Asya Pereltsvaig  
 Vesela Simeonova  
 Anastasia Smirnova  
 Mila Tasseva-Kurktchieva  
 Francis Tyers

**Steven L. Franks**  
 Bloomington, Indiana

**Vrinda Chidambaram**  
 Riverside, California

**Brian D. Joseph**  
 Columbus, Ohio

**Iliyana Krapova**  
 Venice, Italy

July 2018

## Publications of Catherine Rudin

### Books

- 2013 *Aspects of Bulgarian syntax: Complementizers and Wh constructions.* 2nd revised edition, with new introduction. Bloomington, IN: Slavica Publishers.
- 1986 *Aspects of Bulgarian syntax: Complementizers and Wh constructions.* Columbus, OH: Slavica Publishers.
- 1984 *Manual for individualized studies: Bulgarian intermediate 2.* With Instructor's Manual. Columbus, OH: Ohio State University Slavic Papers #12 and #12A. (with Ljubomira P. Gribble)

### Articles

- In prep. "Remembering BSA/BAN I: Bulgarian studies at the University of Wisconsin in the 1970s". Paper presented at the 10th joint Bulgarian-American conference; Bulgarian Studies Association/Bulgarian Academy of Sciences. Sofia, Bulgaria, 26–29 June 2016. To appear in an edited conference volume.
- In press "Balkan Slavic divergences: A meditation on discovering micro-variation with some help from my friends". Donald L. Dyer and Jane Hacking, eds. *Celebration of the retirement of Christina Kramer.* (= *Balkanistica* 32.2) To appear Spring 2019.
- In press "Language contact continues: Bulgarian-Turkish code switching and mixing in the 21st century". *Celebrating twenty years of the Kenneth E. Naylor Lectures in South Slavic and Balkan Linguistics* (= *Balkanistica* 32.1). To appear Spring 2019.
- In press "Balkan Slavic and Balkan Turkish". Marc L. Greenberg, ed. *Encyclopedia of Slavic Languages and Linguistics.* Leiden: Brill. With coauthor Julian Rentzsch.
- In press "The new Bulgarian: Turkisms and Europeanisms in the language of Bai Ganyo and Nov Zhivot". Short monograph forthcoming from *Balkanistica*/SEESA publications.
- In press "Balkan Slavic comparatives". To appear in *Formal approaches to Slavic linguistics: The Cornell meeting, 2016.*

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin.* Bloomington, IN: Slavica Publishers, 2018, xv–xxii.

- 2018 “Multiple determination in Bulgarian and Macedonian: An exploration of structure, usage, and meaning”. Stephen M. Dickey and Mark Richard Lauersdorf, eds. *V zeleni drželi zeleni breg: Studies in Honor of Marc L. Greenberg*. Bloomington, IN: Slavica Publishers, 263–86.
- 2017 “Indefinite and universal *wh* clauses in Macedonian”. *Balkanistica* 30(2): 257–78.
- 2016 “Coordination in Omaha Ponca and in Siouan”. Catherine Rudin and Bryan Gordon, eds. *Advances in the study of Siouan languages and linguistics*. Berlin: Language Science Press, 369–95. [*Studies in diversity linguistics*, 10.]
- 2015 “Whatever: *Wh*-universal constructions in Macedonian and Bulgarian”. *Balkanistica* 28: 409–32.
- 2015 “Invariant *-to* in Bulgarian” Miriam Shrager, Edna Andrews, George Fowler, and Steven Franks, eds. *Studies in accentology and Slavic linguistics in honor of Ronald F. Feldstein*. Bloomington, IN: Slavica Publishers, 99–136.
- 2015 “Syntactic and typological aspects of universal concessive conditionals in Bulgarian”. Anisava Miltenova, Cynthia Vakareliyska, and Christine Holden, eds. *Rethinking the past—looking to the future: Proceedings of the Ninth Joint Meeting of Bulgarian and North American Scholars, Eugene, Oregon, May 31–June 1, 2012*. Sofia: Institute of Literature BAS, 109–22. (Coauthored with Steven Franks.)
- 2014 “Sorting out *-to* and *što*: Bulgarian and Macedonian relative markers”. Jacek Witkoś and Sylwester Jaworski, eds. *New insights into Slavic linguistics*. Berlin: Peter Lang, 317–28.
- 2014 “Focusing on irrealis concessions”. Cassandra Chapman, Olena Kit, and Ivona Kučerová, eds. *Formal approaches to Slavic linguistics: The McMaster meeting, 2013*. Ann Arbor, MI: Michigan Slavic Publications, 266–86. (Coauthored with Steven Franks.)
- 2013 “Mnogokratni svobodni *k*-relativi i korelativi” (“Multiple *Wh*-free relatives and correlatives”). Iliyana Krapova, Jovka Tisheva, and Svetla Koeva, eds. *Novi izsledvanija po generativen sintaksis na Bălgarskija ezik*. Sofia: BAN Marin Drinov, 137–64.
- 2013 “Uvodni dumi” (“Introductory remarks”). Iliyana Krapova, Jovka Tisheva, and Svetla Koeva, eds. *Novi izsledvanija po generativen sintaksis na Bălgarskija ezik*. Sofia: BAN Marin Drinov, 15–16.
- 2012 “Bai Ganyo’s revenge: The persistence of Turkisms in Modern Bulgarian”. *Balkanistica* 25(2): 275–90.
- 2010 “Preface”. Anastasia Smirnova, Vedrana Mihaliček, and Lauren Ressue, eds. *Formal studies in Slavic linguistics*. Cambridge: Cambridge Scholars Publishing.



- 2009 "The Bulgarian relative marker *-to*". Vrinda Chidambaram, Steven Franks, and Brian Joseph, eds. *A linguist's linguist: Studies in South Slavic linguistics in honor of E. Wayles Browne*. Bloomington, IN: Slavica Publishers, 403–22.
- 2008 "Pair-list vs. single pair readings in multiple *Wh*-free relatives and correlatives". Emily Tummons and Stephanie Lux, eds. *Proceedings of the 2007 Mid-America Linguistics Conference*. Lawrence, KS: University of Kansas, Department of Linguistics, 257–67. [*Kansas Working Papers in Linguistics* 30.] Online version available at <https://kuscholarworks.ku.edu/dspace/handle/1808/276>.
- 2007 "Multiple *Wh*-relatives in Slavic". Richard Compton, Magda Goledzinowska, and Ulyana Savchenko, eds. *Formal approaches to Slavic linguistics: The Toronto meeting, 2006*. Ann Arbor, MI: Michigan Slavic Publications, 282–307.
- 2006 "What makes clitic doubling obligatory?" Ivan Kostadinov Kucarov, Vasil Gospodinov Rajnov, Meri Petrova Lakova, Vladko Danailov Murdarov, and Svetla Peneva Koeva, eds. *Ezikovedski izsledvanija v chest na 75-godishninata ot rozhdenieto na St.N.S. i St.D.F.N. Jordan Penčev Penčev*. Sofia: Bulgarian Academy of Sciences, 267–78. (Coauthored with Steven Franks.)
- 2005 "Omaha-Ponca". Keith Brown, ed. *Encyclopedia of language and linguistics*, 2nd edition. Edinburgh: Elsevier. [Vol. 9, 25–29, article id. #4960.] (Coauthored with Kathleen Shea.)
- 2005 "In Memoriam: Jordan Penčev". *Journal of Slavic linguistics* 13(1): 3–4.
- 2005 "Bulgarian clitics as  $K^0$  heads". Steven Franks, Frank Y. Gladney, and Mila Tasseva-Kurktchieva, eds. *Formal approaches to Slavic linguistics: The South Carolina meeting, 2004*. Ann Arbor, MI: Michigan Slavic Publications, 104–16.
- 2002 "From the co-editors: Why formal linguistics?" *Balkanistica* 15: ix–xii.
- 2002 "Clitic pronoun ordering in the Balkan languages". Victor A. Friedman and Donald L. Dyer, eds. *Of all the Slavs my favorites: Studies in South Slavic and Balkan linguistics in honor of Howard I. Aronson on the occasion of his 66th birthday*. Bloomington, IN: Slavica Publishers, 341–57.
- 1999 "Macedonian and Bulgarian *li* questions: Beyond syntax". *Natural language and linguistic theory* 17(3): 541–86. (Coauthored with Christina Kramer, Loren Billings, and Matthew Baerman.)
- 1998 "Focus in Russian and Bulgarian yes-no questions". Elena Benedicto, Maribel Romero, and Satoshi Tomioka, eds. *Proceedings of the Workshop on Focus*. University of Massachusetts Occasional

- Papers in Linguistics* 21: 209–26. (Coauthored with Tracy H. King and Roumyana Izvorski.)
- 1997 “The your guys’s possessive and the structure of DP”. Clifton Pye, ed. *1996 MidAmerica Linguistics Conference Papers*. Lawrence, KS: University of Kansas, 388–95.
- 1997 “Against *li* lowering in Bulgarian”. *Lingua* 102: 187–94. (Coauthored with Roumyana Izvorski and Tracy Holloway King.)
- 1997 “AgrO and Bulgarian pronominal clitics”. Martina Lindseth and Steven Franks, eds. *Formal approaches to Slavic linguistics: The Indiana meeting, 1996*. Ann Arbor, MI: Michigan Slavic Publications, 224–52.
- 1997 “Kakvo li e *Li*: Interrogation and focusing in Bulgarian”. *Balkanistica* 10: 335–46.
- 1996 “Syntax and prosody: *Li* questions in Macedonian and Bulgarian”. Artemis Alexiadou, Nanna Fuhrhop, Paul Law, and Sylvia Löhken, eds. *Papers in phonetics and phonology*. Berlin: ZAS Papers in linguistics, 1–26. [*Zentrum für Allgemeine Sprachwissenschaft Papers in Linguistics*, 7] (Coauthored with Loren Billings and Christina Kramer.)
- 1996 “Optimality and superiority: A new approach to *wh*-ordering”. Jindřich Toman, ed. *Formal approaches to Slavic linguistics: The College Park meeting, 1994*. Ann Arbor, MI: Michigan Slavic Papers, 35–60. (Coauthored with Loren Billings.)
- 1996 “On pronominal clitics”. Mila Dimitrova-Vulchanova and Lars Hellan, eds. *Papers from First Conference on Formal Approaches to South Slavic Languages*. University of Trondheim Working Papers in Linguistics 28: 229–46.
- 1996 “Multiple questions south, west, and east: A government-binding approach to the typology of *wh*-movement in Slavic languages”. *International journal of Slavic linguistics and poetics* XXXIX/XL: 103–22. [V. A. Zvegintsev.]
- 1996 “Grammar as a mental organ: A survey of the GB perspective”. *International journal of Slavic linguistics and poetics* XXXIX/XL: 44–65. [V. A. Zvegintsev.] (Coauthored with Steven Franks.)
- 1995 “Change in progress: An update on Bulgarian Turkish”. *Indiana Slavic studies* 7: 168–73.
- 1994 “Bulgarian relativization strategies”. Charles E. Gribble, ed. *James Daniel Armstrong: In Memoriam*. Columbus, OH: Slavica Publishers, 164–174. (Written 1981)
- 1994 “On focus position and focus marking in Bulgarian questions”. Alice Davison, Nicole Maier, Glaucia Silva, and Wan Su Yan, eds. *Papers from the Fourth Annual Meeting of the Formal Linguistics*

- Society of MidAmerica*. Iowa City, IA: University of Iowa, Department of Linguistics, 252–65.
- 1994 “Topic and focus in Bulgarian”. *Acta linguistica academiae scientiarum hungaricae* 40(3–4): 1–19. (Volume dated 1990–91 but appeared in 1994.)
- 1993 “Bulgarian nationalism and Turkish language in Bulgaria”. Eran Fraenkel and Christina Kramer, eds. *Language contact—Language conflict*. Berlin: Peter Lang, 43–71. (Coauthored with Ali Eminov.)
- 1993 “Articles and the structure of NP in Omaha”. Evan Smith and Flore Zephir, eds. *1992 Mid-America Linguistics Conference/Conference on Siouan /Caddoan Languages*. Columbia, MO: University of Missouri, 361–71.
- 1993 “Omaha language preservation: A status report”. *Proceedings of the XVth International Congress of Linguists*. Québec: Les presses de l’université Laval.
- 1992 “The complementizer system of Modern Bulgarian”. *Indiana Slavic studies* 6/*Balkanistica* 8: 123–30.
- 1991 “Observations on *wh*-questions in Omaha”. Frances Ingemann, ed. *Proceedings of the 1990 MidAmerica Linguistics Conference*. Lawrence, KS: University of Kansas, 573–83.
- 1991 “Omaha language preservation in the Macy, Nebraska public school”. Cheryl A. Roberts, Timothy J. Riney, and Stephen J. Gaies, eds., *Proceedings of 1989 MidAmerica Linguistic Conference*. [no page numbers. Also appeared as ERIC microfiche no. ED 347 785.]
- 1990 “Bulgarian Turkish: The linguistic effects of recent nationality policy”. *Anthropological linguistics* 32(1–2): 149–62. (Coauthored with Ali Eminov. Issue is dated 1990, but appeared in 1992.)
- 1988 “On multiple questions and multiple *wh*-fronting”. *Natural language and linguistic theory* 6: 445–501. (Slightly earlier version appeared as ERIC microfiche no. ED 285 403.) [Reprinted in: *Comparative grammar: Critical concepts*. Vol. 4. Ian Roberts, ed. New York: Routledge, 2007.]
- 1988 “Finiteness and opacity: Evidence from the Balkans”. Michael Hammond, Edith Moravcsik, and Jessica Wirth, eds. *Studies in syntactic typology*. Amsterdam: John Benjamins, 37–52.
- 1988 “Multiple questions in South Slavic, West Slavic, and Romanian”. *Slavic and East European journal* 32(1): 1–24.
- 1986 “Changing the rules: Extra prepositions in relative clauses”. Anne M. Farley, Peter T. Farley, and Karl-Erik McCullough, eds. *CLS* 22. Chicago, IL: Chicago Linguistic Society, 277–85.

- 1986 "Multiple *wh*-fronting constructions". Paper presented at the annual meeting of the Linguistics Society of America/American Association for Applied Linguistics, New York, New York, 27–30 December 1986. Also appeared as [ERIC Document Reproduction Service no. ED 284 435.]
- 1985 "*DA* clauses, finiteness, and opacity". *Folia Slavica* 7(3): 435–51.
- 1984 "Comparing comparatives". Joseph Drogo, Veena Mishra, and David Testen, eds. *CLS 20*. Chicago, IL: Chicago Linguistic Society, 334–43.
- 1984 "Comparatives and equatives in Bulgarian and the Balkan languages". Kot K. Shangriladze and Erica W. Townsend, eds. *Papers for the V Congress of Southeast European Studies*. Columbus, OH: Slavica Publishers, 328–37.
- 1983 "*DA* and the category AUX in Bulgarian". Frank Heny and Barry Richards, eds. *Linguistic categories: Auxiliaries and related puzzles*. Dordrecht: D. Reidel, 3–20.
- 1983 "Turkish velars and the psychological reality of competing grammars". Frances Ingemann, ed. *1982 MidAmerica Linguistics Papers*. Lawrence, KS: University of Kansas, 159–78.
- 1981 "Who what to whom said?: An argument from Bulgarian against cyclic *wh*-movement". Roberta A. Hendrick, Carrie S. Masek, and Mary Frances Miller, eds. *CLS 17*. Chicago, IL: Chicago Linguistic Society, 353–60.
- 1980 "Phonetic evidence for a phonological rule: G-deletion in Turkish". *Research in phonetics* 1: 217–32.
- 1977 "Towards a unified treatment of *WILL*". *Language sciences* 48: 14–17.

### Online Project

- In prep. Omaha and Ponca digital dictionary project. Funded by NSF/NEH Documenting Endangered Languages grant 2008–2011; work has continued sporadically since. (Co-Principal Investigator with Mark Awakuni-Swetland, UNL). Public site: <http://omahaponca.unl.edu>.

### Edited Volumes

- In prep. *Papers from the 10th North American-Bulgarian Meeting*. Coeditor with Vessela Warner, Anisava Miltenova, and Ana Luleva. Sofia: Bulgarian Academy of Sciences

- 2018 *Journal of Slavic linguistics* 25.2. Special 25th anniversary “State of the field” issue. Coeditor with Steven Dickey, Laura Janda, and Keith Langston. Bloomington, IN: Slavica Publishers.
- 2016 *Advances in the study of Siouan languages and linguistics* Coeditor with Bryan Gordon. Berlin: Language Science Press [*Studies in diversity linguistics*, 10.]
- 2002 *Balkanistica* 15. *Papers from the Third Conference on formal approaches to South Slavic and Balkan Linguistics*. Coeditor with Mila Dimitrova-Vulchanova, Donald Dyer, and Iliyana Krapova.
- 2002 *Dictionary of Turkisms in Bulgarian*. Coeditor with Alf Grannes, Kjetil Rå Hauge and Hayriye Süleymanoğlu. Oslo: University of Oslo Press. (Copy-editor and checker.)

## Reviews

- 2005 Review of Boeckx, Cedric and Kleanthes K. Grohmann, eds. *Multiple wh movement*. *Journal of Linguistics* 41(2): 437–41.
- 2004 Review of Hauge, Kjetil Rå. *A short grammar of contemporary Bulgarian*. *Journal of Slavic linguistics* 9(2): 325–29. [cover date 2001, appeared spring 2004.]
- 2004 Review of Moskovsky, Christo. *Aspects of binding in Bulgarian*. *Slavic and East European journal* 48(3), 526–27.
- 2003 Review of Henk van Riemsdijk, ed. *Clitics in the languages of Europe*. *Journal of linguistics* 39(3): 683–88. [Nov. 2003]
- 2003 Review of Leanne Hinton and Kenneth L. Hale, eds. *The green book of language revitalization in practice*. *Modern language journal* 87(4): 630–31.
- 2003 Review: María Luisa Rivero and Angela Ralli, eds. *Comparative syntax of Balkan languages*. *Language* 79(2): 423–25.
- 2002 Book Notice: Mila Dimitrova-Vulchanova and Lars Hellan, eds. *Topics in South Slavic syntax and semantics*. *Language* 78(1): 197–98.
- 2002 Book Notice: István Kenesei, ed. *Crossing boundaries: Advances in the theory of Central and Eastern European languages*. *Language* 78(1): 198–99.
- 2001 Review of Christina Bratt Paulston and Donald Peckham, eds. *Linguistic minorities in Central and Eastern Europe*. *Anthropological linguistics* 42(2): 587–88.
- 2000 Book Notice: William O’Grady, Michael Dobrovolsky, and Mark Aronoff, eds. *Contemporary linguistics: An introduction*. *Language* 76(2): 470.
- 2000 Book Notice: Albert White Hat. *Reading and writing the Lakota language*. *Language* 76(2): 470–71.

- 1999 Review of Shirley Silver and Wick R. Miller. *American Indian languages: Cultural and social contexts*. *Great Plains review* 9(1): 178–79.
- 1998 Book Notice: Lesley Milroy and Pieter Muysken, eds. *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. *Language* 74(2): 409–10.
- 1998 Book Notice: Ronald Wardhaugh. *Understanding English grammar: A linguistic approach*. *Language* 74(1): 202–03.
- 1997 Book Notice: Jack Feuillet. *Bulgare*. *Language* 73(2): 438–39.
- 1995 Book Notice: Anthony Mattina and Timothy Montler, eds. *American Indian linguistics and ethnography in honor of Lawrence C. Thompson*. *Language* 71(1): 175–76.
- 1994 Book Notice: Willem Fase, Koen Jaspaert, Sjaak Kroon, eds. *Maintenance and loss of minority languages*. *Language* 70(1): 202.
- 1991 Review of Julia Kristeva. *Language: The unknown: An initiation into linguistics*. *European studies journal* 7(1): 80–82
- 1989 Review of Charles Gribble. *Reading Bulgarian through Russian*. *Modern language journal* 73(3): 351.
- 1987 Review of Dimitar Tilkov, Stojan Stojanov, and Kostadin Popov, eds. *Gramatika na sâvremenija bâlgarski knižoven ezik v tri toma (Grammar of the Contemporary Bulgarian Literary Language in Three Volumes)*. *Slavic and East European journal* 31(4): 646–48.
- 1985 Review of Catherine V. Chvany and Richard D. Brecht, eds. *Morphosyntax in Slavic*. *Slavic and Eastern European journal* 29(3): 353–55

### Translations

- In prep. *Folk dance ensembles as a cultural phenomenon in Bulgaria*. (Naroden Tancov Sâstav kato Kulturno Javlenie v Bulgarija) by Daniela Ivanova Nyberg.
- 2010 *Bai Ganyo: Incredible tales of a modern Bulgarian*. by Aleko Konstantinov. Victor A. Friedman, ed. Translated by Victor A. Friedman, Christina E. Kramer, Grace E. Fielder, and Catherine Rudin. (Rudin translated the following three stories: “Baj Ganjo returns from Europe,” “Baj Ganjo and the elections,” “Baj Ganjo the journalist”.)

# Clitic Doubling in Bulgarian\*

Olga Arnaudova

*Abstract:* This paper gives an overview of clitic doubling in Bulgarian and argues that it displays a degree of obligatoriness, distinctive features, and distribution, which should be reflected in the syntactic representation. *Type 1* includes absolute obligatoriness of clitic doubling with Experiencer predicates, while *Type 2* involves “structure-induced obligatoriness” with clitic left dislocation. Clitic doubling structures are compared to structures involving focus-topic chains, where typically no doubling occurs. The two types related to doubling display two different ways of applying arguments to the clause—one clause-internally with applicative arguments (Experiencers), and the other with base-generated arguments above the clause (Clitic Left Dislocation). While displaying an array of features which distinguishes them, Type 1 and Type 2 structures can be unified from a semantic point of view as they both display a discourse subject.

## 1. Introduction

Clitic reduplication or doubling is one of the distinguishing characteristics of Bulgarian (see Rudin 1997 and Franks and King 2000). It is attested in structures where an accusative or dative clitic pronoun is doubled by a noun or full pronoun. These structures appear to have redundant thematic structures, as the verb is simultaneously assigning two identical thematic roles: one to the clitic and one to the noun. In (1), a left-dislocated element, the proper name *Ivan* (referred to as “associate”) is linked to a coindexed resumptive clitic pronoun *go*.

---

\* I am grateful to Iliyana Krapova for discussions on many of the issues raised in this work and to an anonymous reviewer for comments that improved the quality of this paper. I will use in this paper the term “clitic doubling” to refer to any resumption of clitics with full arguments but will refer to specific clitic-doubling manifestations such as Clitic Left Dislocation or Clitic Doubling Proper, where necessary.

- (1) Ivan Marija go vidja.  
 Ivan Mary him<sub>CL.ACC</sub> saw<sub>3SG</sub>  
 ‘Mary saw Ivan./’Ivan was seen by Mary.’

Starting with Miklošič (1861), clitic reduplication is viewed as pan-Balkan, and numerous papers were devoted to the distribution of reduplication in Balkan languages (see Asenova 2002). Bulgarian and Macedonian are the only Slavic languages displaying this phenomenon. In the last few decades, significant attempts have been made to systemize doubling and study the syntactic properties associated with it (Rudin 1986, Rudin 1997, Tomić 2000, Arnaudova 2002, Krapova 2002, Franks & Rudin 2005, Bošković 2008, Harizanov 2014, among others). One of the important questions that researchers have asked is whether clitic reduplication should be presented in a uniform way or rather consists of a number of (optional) phenomena. Another question is how the doubled element should be viewed—as an adjunct or as an argument of the verb. The discussion is complicated by differences between written and spoken variants of the language and varying speakers’ judgments.<sup>1</sup>

Before arriving at a uniform construction or analysis (if ever possible), we need to pose the question whether clitic-doubling cases should not be separated into several sub-types in order to be better understood structurally and semantically. In this paper, I investigate this question in more detail and suggest that a unified approach cannot account for all the cases at hand. Furthermore, I argue that at least in some cases the optionality noticed in previous generative and non-generative research might be linked to different varieties of the language, with the spoken variety requiring a clitic and the written variety not tolerating the (overt) use of clitics whenever possible. In what follows I present a review of the known cases of clitic reduplication in Bulgarian and first discuss their optionality. Then I turn to distinguish between two main types of clitic doubling and discuss environments where doubling is not attested (focus-topic chains).<sup>2</sup> Finally, I examine the two types in more detail and propose an analysis.

<sup>1</sup> In representative grammars of Bulgarian, it is acknowledged that we may have to deal with two different varieties of Bulgarian and with some kind of avoidance of certain structures in literary Bulgarian due to influence from Russian (see Andrejčin et al. 1977: §522, p. 376).

<sup>2</sup> In generative grammar true clitic doubling is a case where the clitic and the noun, in situ, are both arguments and are obligatory regardless of language varieties, as in the Spanish *Lo vimos a Juan* ‘We saw CL Juan’. I will use in this paper the term “clitic doubling” in a broader sense to refer to any resumption of clitics with full arguments.



## 2. Obligatory or Optional?

One case is when we have an associate (noun or full pronoun which is a direct or indirect object) appearing in the left periphery of the clause,<sup>3</sup> as shown in (2). It has to be noted that structures like (2), while fully grammatical, are rarely found in the literary variety of the language and therefore are not discussed in traditional grammars of Bulgarian that normally use examples from written texts. In the spoken variety, the clitic is always present with the exception of the focus-topic chain discussed later in this paper.

### Associate-Clitic

- (2) a. Tetradkata /edna tetradka sâm ja zabravit.  
 notebook<sub>DEF</sub>/one notebook AUX it<sub>CL.ACC</sub> forgotten  
 'I forgot the notebook/a notebook.'
- b. Na men, pensionera, mi dadoxa! (personal  
 to me, pensioner<sub>DEF</sub> me<sub>CL.DAT</sub> gave<sub>3PL</sub> conversation)  
 'They gave (it) to me.'
- c. A Daniel savsem sâm go zarjazala. (from an  
 Daniel completely AUX him<sub>CL.ACC</sub> abandoned email)  
 'I completely left Daniel.'

In (3) below, multiple nouns appear as associates, each doubled by a clitic (with the exception of dislocated subjects, such as *az* 'I', which do not have a corresponding (overt) clitic). Note that the order of the associates can be *učebnika, na Stoyan, az, na Stoyan, az, učebnika, az, na Stoyan, učebnika* etc. which clearly indicates free order among them.

- (3) Az učebnika na Stojan mu go dadox.  
 I textbook<sub>DEF</sub> to Stojan him<sub>CL.DAT</sub> it<sub>CL.ACC</sub> gave<sub>1SG</sub>  
 Lit. As for me, the textbook, and Stoyan, I gave it to him.'  
 (Arnaudova, 2002)

In (4), the associate follows the clitic. We can tentatively assume that the division between spoken and written varieties of the language is also one of the determining factors in deciding whether the clitic is used or not with *anticipatio* (along with other important factors discussed later in this paper).

<sup>3</sup> The argument in this case can be also indefinite but needs to be specific. The associate has been in the past shown to be a specific, definite, direct or indirect object (cf. Leafgren 1997).

*Clitic-Associate (anticipatio)*

- (4) a. Decata (ja) obiĉat neja.  
 kids<sub>DEF</sub> her<sub>CL.ACC</sub> love her  
 ‘The kids love her.’
- b. Marija (mu) izprati pismo na rabotnika.  
 Maria him<sub>CL.DAT</sub> sent<sub>3SG</sub> letter to worker<sub>DEF</sub>  
 ‘Maria sent a letter to the worker.’

(Harizanov 2014)

See also the clitic doubled ditransitive construction in (5) where clitic doubling is also attested:

- (5) Ivan (i) (ja) prati knigata na Gabriela.  
 Ivan it<sub>CL.DAT</sub> her<sub>CL.ACC</sub> sent<sub>3SG</sub> book<sub>DEF</sub> to Gabriela  
 ‘Ivan sent Gabriela the book.’

Bulgarian also possesses genuine clitic doubling (CD) constructions that exhibit many of the properties of their Romance counterparts, with the difference that in Bulgarian this type is found predominately with Experiencer predicates, as illustrated in (6).<sup>4</sup> In all of the cases below, doubling is absolutely obligatory, both in written and colloquial speech and the absence of the clitic leads to an ungrammatical sentence.

- (6) a. Ivan \*(go) boli gârloto.  
 Ivan him<sub>CL.ACC</sub> aches throat<sub>DEF</sub>  
 ‘Ivan’s throat is sore.’
- b. Na Ivan \*(mu) xaresa filmât.  
 to Ivan him<sub>CL.DAT</sub> liked film<sub>DEF</sub>  
 ‘Ivan liked the film.’  
 (lit. ‘The movie appealed to Ivan.’)

<sup>4</sup> It is the choice of predicate (e.g., psych and physical perception predicates, modal predicates, predicates with possessor datives, etc.) and the obligatoriness of the doubling clitic that distinguish “proper” CD constructions (see Manolova 1979 and Krapova and Cinque 2008 for discussion).

See also an example with the *feel-like* construction in (7). In (7b) an overt pronoun coexists with the clitic:<sup>5</sup>

- (7) a. Spi mu se.  
 sleep him<sub>CL.DAT</sub> REFL  
 ‘Peter feels like sleeping.’
- b. Na mene mi se spi.  
 to me me<sub>CL.DAT</sub> REFL sleep  
 ‘I feel like sleeping./I am sleepy.’

Clitic doubling is found also with affected arguments/possessors (the so-called **malefactives** and **benefactives**), as pointed out by Iliyana Krapova (p.c.). Please note that unlike the Experiencer constructions, mentioned above, the clitic with benefactives/malefactives can be omitted in the written variety of the language.

- (8) Az (mu) šcupix/opravix na Ivan koleloto.  
 I him<sub>CL.DAT</sub> broke<sub>1SG</sub>/fixed<sub>1SG</sub> to Ivan bicycle<sub>DEF</sub>  
 ‘I broke/fixed Ivan’s bicycle.’

As in Italian, clitic left-dislocated elements are not limited to DPs and are also found with full clauses. An example with a CLLDed CP is given in (9):

- (9) [CP Če Plamen ne e dobar specialist] go  
 that Plamen NEG AUX good specialist it<sub>CL.ACC.SG</sub>  
 razbraxa veče vsički.  
 found.out<sub>3PL</sub> already all  
 ‘Everybody understood that Plamen is not a good specialist.’

There is also doubling of *wh*-words, as in (10a). The hierarchy of the *wh*-words exemplified in (10b) where no doubling is attested, is violated in (10a). This type of doubling is found predominately in colloquial Bulgarian (see Krapova and Cinque 2008 for detailed discussion):

<sup>5</sup> This construction is productive throughout Slavic with transitive and intransitive agentive verbs, and is used to express the fact that someone (the dative Experiencer) ‘feels like V-ing’, e.g., *spi mi se* (‘I feel like sleeping’ lit.: it sleeps to me). For a discussion of this construction in Slavic see Rivero and Sheppard (2003) and Marušič and Žaucer (2006). Rivero & Sheppard also observe that in West Slavic languages (Polish and Czech) the ‘feel-like’ interpretation is not available.

- (10) a. specific set, no hierarchy of *wh*-words

Na kogo koj mu podari kniga?  
 who<sub>DAT</sub> who<sub>NOM</sub> him<sub>CL.DAT</sub> donated<sub>3SG</sub> book?

‘Who gave a book as a present to whom?’

- b. multiple sets; fixed hierarchy of *wh*-words

Koj na kogo podari kniga?  
 who<sub>NOM</sub> who<sub>DAT</sub> donated<sub>3SG</sub> book?

‘Who gave a book as a present to whom?’

To summarize the discussion above, clitics appear with overt associates in the following main cases:

- (11) **Type 1:** Predicate-induced obligatoriness

- The associate is an Experiencer, or
- An affected argument (benefactive or malefactive) where an action of an individual affects another individual.

**Type 2:** Structure-induced obligatoriness (obligatory in colloquial speech; suppressed in written Bulgarian at least for some of the constructions)

- The associate is a definite or indefinite specific dislocated dative or accusative (usually to the left but right “dislocations” are also attested)
- When *wh*-words in the left periphery appear to violate Superiority

In what follows, I present some interesting characteristics pertaining to Type 1 and Type 2 structures and differentiate them on the basis of syntactic tests. I argue that Type 1 and Type 2 correspond to two different ways of applying arguments to the clause but might be united from a semantic point of view, as they both display categorical statements in the sense of Kuroda 1972. In the next section, I differentiate them from topic-focus chains where no doubling occurs.

### 3. Topic-Focus Chains (No Doubling)

Consider now the case of the Topic-Focus chains, discussed in Büring (2003), where no doubling is found regardless of the nature of the associate (specific or not) or its position:

- (12) CT.....F; sets of pairs (Büring 2003)

In (13a) and (13b) below we have a set of ordered pairs and contexts where the subject and the object are both questioned (new information). In (13a), a set of pairs is evoked {(children, ice cream), (mothers, coffee)} and the persons are contrastive topics in the left periphery of the clause. In (13b), the ordered pairs are {(ice cream, children), (coffee, mothers)} with the items purchased appearing as contrastive topics in the left periphery. Notice the absence of clitics in these constructions.

- (13). a. {(children, ice cream), (mothers, coffee)}  
 [Who bought what? Who bought ice cream and who bought coffee?]  
 S-V-O  
 Decata-CT kupixa sladoled-F, a majkite-CT (kupixa)  
 children<sub>DEF</sub> bought<sub>3PL</sub> ice cream while mothers<sub>DEF</sub> (bought<sub>3PL</sub>)  
 kafe-F.  
 coffee
- b. {(ice cream, children), (coffee, mothers)}-“reversed” pairs  
 [Who bought what? What did the children and the mothers buy?]  
 O-V-S  
 Sladoled-CT kupixa deca/decata-F, a kafe (kupixa)  
 ice cream bought<sub>3PL</sub> children<sub>DEF</sub> while coffee (bought<sub>3PL</sub>)  
 majkite-F.  
 mothers<sub>DEF</sub>
- c. \*Sladoled decata kupixa, a kafe majkite kupixa. \*O-S-V  
 d. \*Decata sladoled kupixa, a majkite kafe kupixa. \*S-O-V

In this context clitic doubling is not felicitous. On the other hand, left dislocated topics are felicitous, where the context is not linked to ordered pairs but to the event as a whole as in (14):<sup>6</sup>

<sup>6</sup> It is true that Topic-Focus pairs are also possible with CLLD constructions when one of the elements is not dislocated, as also noted for Spanish by Arregi (2003). This is achieved when a focused element is found in the lower clausal domain. In my view, however, these cases only superficially resemble their counterparts in the constructions with contrastive topics and involve accidental pairings (*knigite, Ivan*)/(*Ivan, knigite*) of the salient topic and the focused element, since the “real” pairing in this case is between the salient element and the whole predicate (*gi varna Ivan*) which is relevant to the interpretation of the topic (see section 7.1). Note also that the intonational properties of these constructions are quite distinct from those of their CT-F counterparts.

- (14) \*Who returned what yesterday? \*What did Ivan return? \*Who returned the books?

√What happened to Ivan, Mary, and the books (salient items)? The event is not known initially.

T1/T2 ... CL1 CL2 V (the order of the fronted elements is free)

- (15) a. Ivan na Marija knigite včera i gi  
Ivan to Mary books<sub>DEF</sub> yesterday her<sub>CL.DAT</sub> them<sub>CL.ACC</sub>  
vârna.  
gave back<sub>3SG</sub>

Additionally, no clitic doubling is found with contractive (exhaustive) foci (cf. Kiss 1998 for Hungarian; Arnaudova 2003 for Bulgarian) as shown in (16) and (17):

- (16) Contrastive (exhaustive) F..... (non contrastive) T

\*Who killed what? \*Who did kill the deer?

√What did the hunter kill? (a set of alternatives in a single pair is evoked, restricted or unrestricted, and one member is exhaustively selected)

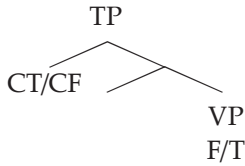
√Is it a rabbit that the hunter killed? (from a set).

Ó-V-S

- (17) VÂLKA (\*go) ubi včera lovecât.  
wolf<sub>DEF</sub> \*him<sub>CL.ACC</sub> killed<sub>3SG</sub> yesterday hunter<sub>DEF</sub>  
'It was the wolf that the hunter killed yesterday.'

The answer is true if and only if a wolf was killed by the hunter and nothing else. Contrastive topic/focus constructions in Bulgarian can be united on the view that they involve (sets of) ordered pairs where the higher element is valuing an EPP feature (Chomsky 1995) while the other element in the VP is a non-contrastive topic or focus. On this view, foci and topics can be united under a split chain hypothesis having a non-contrastive member in the *vP*-domain as the other member of the pair. This is shown in (18):

- (18) Contrastive topic or Focus—F/T (non-contrastive)



#### 4. Experiencer Constructions and Clitic Left Dislocated Structures<sup>7</sup>

Arnaudova and Krapova (2007) show that the Experiencer argument displays subject-like properties and hence qualifies as a quirky subject, i.e., a non-Nominative subject. The Experiencer can be a Dative (Prepositional), Genitive, or Accusative object, while the Theme is invariably Nominative. Evidence for treating Experiencer Datives as filling a position distinct from that of topicalized datives comes from the fact that in Bulgarian there are contexts where Experiencer fronting is perfectly fine, while fronting of a dative object of a transitive verb, which is an instance of left dislocation, is quite marginal. Consider now two such cases illustrated by the adverbial and the relative clauses below:

- (19) a. <sup>?</sup>Vsički se pritesnixa, zaštoto na Ivan sâm (mu)  
 all REFL got worried<sub>3PL</sub> because to Ivan am him<sub>CL.DAT</sub>  
 pomognal.  
 helped  
 ‘Everybody got worried because I have helped Ivan.’
- b. Vsički se pritesnixa, zaštoto sam (mu) pomognal  
 all REFL got worried<sub>3PL</sub> because am him<sub>CL.DAT</sub> helped  
 na Ivan.  
 to Ivan  
 ‘Everybody got worried because I have helped Ivan.’
- (20) a. Vsički se pritesnixa, zaštoto na Ivan mu  
 all REFL got worried<sub>3PL</sub> because to Ivan him<sub>CL.DAT</sub>  
 dopada bâlgaristikata.  
 appeal<sub>3SG</sub> Bulgarian studies<sub>DEF</sub>  
 ‘Everybody got worried because Ivan likes Bulgarian studies.’

<sup>7</sup> This section is based on joint work with Iliyana Krapova, presented at Formal Approaches to Slavic Languages 15 (Arnaudova and Krapova 2007).

- (20) b. Vsički se pritesnixa,      zaštoto Ivan predpočita  
 all      REFL got worried<sub>3PL</sub> because Ivan prefer<sub>3SG</sub>  
 bâlgaristikata.  
 Bulgarian studies<sub>DEF</sub>  
 ‘Everybody got worried because Ivan prefers Bulgarian studies.’

In (19a) we see that in a sentence that does not contain an Experiencer predicate, the attempt to front (i.e., to topicalize) a regular prepositional (indirect) object (*na Ivan*) in an adverbial (*because*) clause produces a marginal result. The reason for this marginality lies in the fact that the discourse factors which would motivate this more marked order are weak, hence it becomes difficult to topicalize the object. The most natural order will be the one in which the object *na Ivan* appears in its regular postverbal position, as in (19b). Now look at (20) above with the Experiencer predicate *dopada*. (20a) shows that when we have an Experiencer (appearing in the preverbal position), the sentence is perfectly all right. In (20b) we have a regular Nominative subject which is comparable to (20a). This we take as direct evidence that non-Nominative Experiencer objects have subject-like properties.

Next, consider the co-occurrence of bare quantifiers and indefinites with clitic structures. There is a sharp difference in grammaticality between left-dislocating and Experiencer fronting of a negative quantifier, as the contrast in (21) shows. The same holds for other bare quantifiers like the indefinite *njakoj* ‘someone’, *edin* ‘one’ (not shown here).

- (21) a. \*Na nikogo ne sâm mu pisal.      CLLD  
 to nobody NEG am him<sub>CL.DAT</sub> written  
 ‘To nobody have I written’  
 b. Na nikogo ne mu xaresa pismoto mi.      Exp  
 to nobody NEG him<sub>CL.DAT</sub> appealed<sub>3SG</sub> letter<sub>DEF</sub> my  
 ‘Nobody liked my letter.’

Next consider anaphor binding and pronominal binding facts (see also Harizanov 2014). The contrast below in (22) shows that similarly to other languages (e.g., Russian, as discussed in Franks 1995: 253) the Dative Experiencer, on par with subjects, shows the potential to bind an anaphor. Failure to front the appropriate kind of constituent affects binding relations and produces ungrammaticality as a Principle A violation, cf. (22a) and (22b). This property will be considered later when the status of Experiencers in Bulgarian is analysed and an applicative approach is proposed.



- (22) a. Ivan go dojadja na sebe si.  
 Ivan him<sub>CL.ACC</sub> got-angry<sub>3SG</sub> at himself  
 'Ivan got angry with himself.'
- b. \*Na sebe si go dojadja Ivan.  
 to himself him<sub>CL.ACC</sub> got-angry<sub>3SG</sub> Ivan.

Pronominal binding facts illustrate lack of WCO effects in (23a), as opposed to the (23b) example, which once again indicates that the fronted Experiencer occupies a clause-internal A-position:

- (23) a. ?Na vsjaka krasiva žena ì xaresva nejnoto  
 to every beautiful woman her<sub>CL.DAT</sub> appeal<sub>3SG</sub> her<sub>DEF</sub>  
 sobstveno kuče.  
 own dog  
 'Every beautiful woman likes her own dog.'
- b. \*Nejnoto sobstveno kuče ì xaresva na vsjaka  
 her<sub>DEF</sub> own dog her<sub>CL.DAT</sub> appeal<sub>3SG</sub> to every  
 krasiva žena.  
 beautiful woman

The Experiencer Subject appears as an answer to a *wh*-question and is compatible with focusing adverbs such as *only*, *even*, and *also* (not shown).

If we apply the tests to Accusative Experiencers (24–28), i.e., to Experiencers in psych constructions with accusative clitics, we get practically the same results as with Dative Experiencers. The cluster of properties illustrated below lead us to consider that Accusative Experiencers are *quirky* subjects in A-positions.

Word order: AccExp -V- Theme Nom and Theme Nom - V - AccExp

- (24) a. Petârčo go boli gârloto /gârlò.  
 Peter<sub>DIMIN</sub> him<sub>CL.DAT</sub> aches<sub>3SG</sub> throat<sub>DEF</sub>/throat
- b. Gârloto /gârlò go boli Petârčo.  
 throat<sub>DEF</sub>/throat him<sub>CL.DAT</sub> aches Peter<sub>DIMIN</sub>  
 'Little Peter has a sore throat.'

## Accusative Experiencers vs CLLD

- (25) a. <sup>?</sup>Vsički se pritesnili, zaštoto Marija ja bil  
 all REFL got-worried<sub>3PL</sub> because Mary her<sub>CL.ACC</sub> was  
 celunal Ivan.  
 kissed Ivan.
- b. Vsički se pritesnili, zaštoto Marija ja  
 All REFL got-worried<sub>3PL</sub> because Mary her<sub>CL.ACC</sub>  
 zaboljal koremât.  
 started-to-ache<sub>3SG</sub> stomach<sub>DEF</sub>
- (26) a. <sup>?</sup>Onezi, deto Ivan (go) čakati, sa negovite studenti.  
 those that Ivan him<sub>CL.ACC</sub> wait<sub>3PL</sub> are his<sub>DEF</sub> students
- b. Onova, deto Ivan go boli naj-mnogo, e dušata.  
 that that Ivan him<sub>CL.ACC</sub> aches<sub>3SG</sub> most is soul<sub>DEF</sub>  
 ‘What hurts Ivan the most, is his soul.’

Accusative Experiencers and bare quantifiers/indefinites are fine:

- (27) a. \*Nikogo ne go sreštnax po pâtja nasam. CLLD  
 nobody<sub>ACC</sub> NEG him<sub>CL.ACC</sub> met<sub>1SG</sub> on way<sub>DEF</sub> to-here  
 ‘I met no one on my way here’
- b. <sup>?</sup>Njakogo go sreštnax po pâtja nasam.  
 someone him<sub>CL.ACC</sub> met on way<sub>DEF</sub> to-here  
 ‘I met someone on my way here’
- (28) a. Nikogo ne go boli glavata. Exp  
 nobody<sub>ACC</sub> NEG him<sub>CL.ACC</sub> aches head<sub>DEF</sub>  
 ‘Nobody has a headache’
- b. Njakogo maj go boli glavata.  
 somebody perhaps him<sub>CL.ACC</sub> aches head<sub>DEF</sub>  
 ‘Perhaps someone has a headache.’

Considering the evidence presented so far, it is more likely that the Experiencer “subject” occupies a position that is internal to the clause, unlike the clitic reduplicated arguments in CLLD constructions, where, the reduplicated argument is Topicalized or dislocated and hence occupies a non-argumental position in the Left Periphery of the clause (in the CP domain). The surface

structure position of the Experiencer is therefore lower than the position of the dislocated phrase:

- (29) a. Type 1  
 ExperiencerSubject ... VP  
 |\_\_\_\_\_C domain\_\_\_\_\_|
- b. Type 2  
 CLLD  
 |\_\_\_\_Infl domain\_\_\_\_\_|

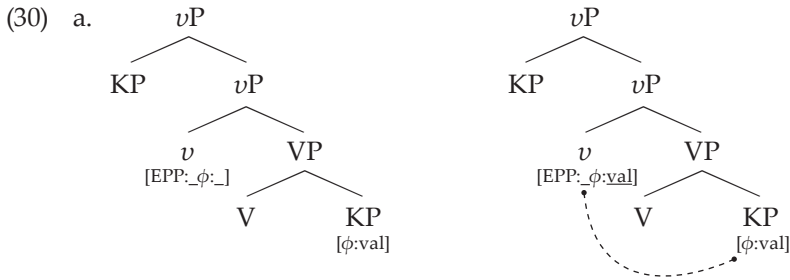
## 5. Previous Accounts of Clitic Doubling

### 5.1. Early Generative Accounts of Bulgarian

For Bulgarian, Rudin 1997 and Franks and King 2000 adopt the view that clitics are functional agreement heads which may optionally associate with full DPs located in VP. According to this view, the associated arguments appear in the usual VP-internal positions (overt or null) and may raise overtly to a higher agreement projection to (optionally) check features in a spec-head configuration, and then even higher, undergoing a movement with A and A-bar properties (see Franks and King 2000 for discussion). The NP raises (covertly) to the specifier of this clitic projection to check a feature associated with definiteness. This analysis is based on a view that relies on the optionality of clitics in all clitic doubling constructions (see also Rudin 1997: 24–25) and raises the question of how this optionality should be accounted for. Unlike genuine object agreement markers, object clitics in Bulgarian are not always attested, as predicted by this analysis.

### 5.2. Recent Proposals

Franks and Rudin (2005, 2006) and Harizanov (2014) argue that regardless of the nature of the predicate, clitic doubling is the outcome of syntactic movement. According to Franks and Rudin 2006, pronominal clitics are heads of KPs (CaseP), which take a possibly null DP as a complement (the associate). The associate (a topic or an oblique object) moves through SpecKP to a higher position in the clause. In the analysis presented by Harizanov (2014), the nominal argument first merges in its base position and then merges again as the specifier of *v* if attracted by an EPP-feature. Then, the higher occurrence of the argument is reduced to its K head (i.e., Case- and  $\varphi$ -features) by the application of m-merger, while the lower occurrence is pronounced in full. As a result, clitic doubling of the kind found in Bulgarian involves, descriptively speaking, spelling out both the head and the foot of a movement chain as shown in (30):



While attractive, these approaches do not capture, in my view, the many differences observed between Type 1 and Type 2 constructions and their semantic nature.

### 5.3. Proposal

I propose that Type 1 and Type 2 cases are linked to the pronoun they double in the following way: 1) the CLLD antecedent is linked to a clitic-variable and appears in a domain higher than TP, and (2) the Experiencer is a high applicative in the sense of Pylkkänen 2002 and Cuervo 2003. The analysis is discussed in more detail below.

#### 5.3.1. CLLD

First, let's consider the position of the CLLD element, which is distinct from the one of the Experiencer, as show in the previous section. As proposed in Arnaudova 2003, I will argue that the left-hand noun is base-generated (see Cinque 1990). It is the subject of predication and takes a predicate containing a variable (the clitic):

(31) [XP DP [IP cl ... t.]]

CLLD has no syntactic analogue in English or in any other Slavic language.<sup>8</sup>

Consider now (32), where the NP *Ivan* can be either a dislocated object, as in (a), or a subject, as in (b):

<sup>8</sup> The minimality constraints presented in Arnaudova 2003 can be attributed to the functioning of an operator associated with a dislocated element and binding a clitic (or pro):

- (i) OpP wh/foc \*OpP clitic/pro verb EC

Another piece of evidence that dislocated elements are in a higher domain comes from the intonational properties of sentences.

- (32) a. Ivan go vidjaxa.  
 Ivan <sub>CL<sub>ACC</sub></sub> (they) saw  
 ‘As for Ivan, they SAW him.’  
 For some x (x = Ivan) they saw x.
- b. Ivan dojde.  
 Ivan came<sub>3SG</sub>  
 ‘Ivan came.’

In (33), the referent is picked up again by the description, similar to what we find with a so-called E-type pronoun and its antecedent (see Evans 1980):

- (33) there is an x (x = Ivan)  
 the x (such that x = Ivan) came/was seen etc.

External restrictor domains define a separate background existential presupposition related to discourse and identify an entity (which can be viewed as an inherent topic or event). This external domain has been equated semantically with “subject of predication” (Reinhart 1981), “higher predication domain”, and “argument externalization” (Zubizarreta 1999; Arnaudova 2003), and has been described in syntactic terms as realized by adjuncts or, more recently, by elements occurring in the specifier positions of topic operators (Zubizarreta 1999). As a result, the dislocated element is an argument but is felt to be “removed” from the domain of the predication, providing an independent description of the referent. The presence or absence of the dislocated element does not alter the focus-topic structure of the lower predication domain, which is on the event or on an internal argument inside it.

### 5.3.2. Experiencers and Other Clause-Internal Associates

Alexopoulou et al. 2003 examine a class of Broad Subjects that have the properties normally associated with subjects in A-positions. They differentiate them from left-dislocated constituents in A'-positions and argue that Broad Subjects are “arguments of the sentential predicate” without being assigned a thematic role by the lexical predicate. Furthermore, they provide evidence that these subjects are not dislocated but are merged (and not moved) to [specTP]. The notion of Broad Subject is further developed by the study on applicatives across languages. Pylkkänen (2002) proposes that cross-linguistically there are two kinds of applied arguments (applicatives): high applicatives where the applicative head denotes a thematic relation between an individual and an event and low applicatives, where the applicative head denotes a transfer of possession relation. With high applicatives, a DP which is external to

$vP$  appears in the Specifier position of an Applicative Phrase and is related to an event expressed by the  $vP$ . With low applicatives, on the other hand, Pylkkänen argues that two individuals are related to each other and there is a transfer of possession. A low applicative head takes an object DP (theme) as its complement and relates it to the DP licensed in its specifier (applied argument). The applicative phrase then combines with the verb. Pylkkänen considers the double object construction (DOC) as an instance of low applicative construction in which there is a transfer of possession. An example is given in (34) where indirect object is an intended *recipient* of the direct object.

(34) Low recipient applicative: English

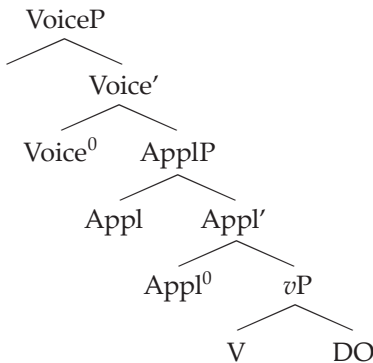
I wrote John a letter.

*I wrote a letter and the letter becomes the possession of John.*

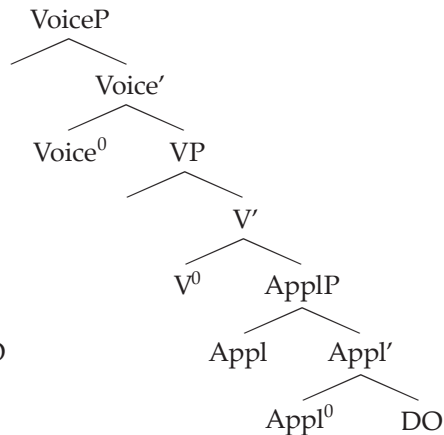
To sum up, low applicatives relate a recipient or a source to an individual, which is the internal argument of a verb being involved in a dynamic relation of transfer possession, while high applicatives relate an individual to an event. The respective structures for high and low applicatives are given in (35). The structure in (35) contains a VoiceP and the subject is located in its Specifier.

(35) (Pylkkänen 2002)

a. High Applicatives



b. Low Applicatives



For constructions such as the double-object construction in (5) above, we can consider the low applicative approach.<sup>9</sup> Following Cuervo 2003, we can con-

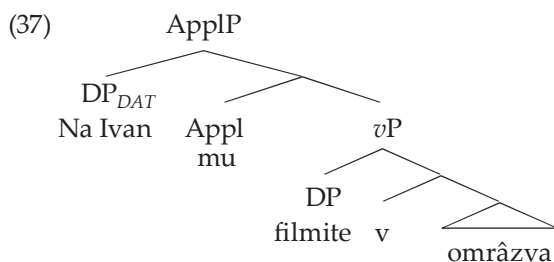
<sup>9</sup> See Slavkov 2008 for a low-applicative analysis of the double-object construction in Bulgarian.

sider these constructions as unaccusative predicates that express transfer of a theme (pleasure/hunger/mental/physical state) to a goal/location/affectee. Here the applied argument appears in dative form expressing the possession/affectedness relation to the theme. The low-applicative analysis could be also applicable to the benefactives/malefactives discussed above. I will not examine it here.

Let us now turn to Type 1 constructions with Experiencers. As in Cuervo 2003 we can argue that the clitic doubled dative argument with Experiencers is an applicative merged higher than the nominative DP theme. Consider now the following example:

- (36) Na Ivan mu omrâznaxa filmite  
 To Ivan him<sub>CL.DAT</sub> became.boring<sub>3PL</sub> movies<sub>DEF</sub>  
 "Ivan got tired of the movies"

The proposed structure is given in (37):<sup>10</sup>



The optional Experiencer in sentence-initial position is an additional argument which is being applied in the sense of Pylkänen 2002 and Cuervo 2003. In (37) the dative argument is licensed by a high applicative that takes the *vP* as its complement. The light verb structure contains all the arguments. The Experiencer DP is in sentence-initial position (*Na Ivan*) and is coreferential with the dative clitic pronoun. The doubled element in this case is either a discourse subject or a quirky subject and is *categorical* in nature (following Kuroda 1972) with the clitic being a marker of a categorical statement similarly to the *wa* particle in Japanese. Semantically, a categorical statement consists of two parts (double statement), one of which is the recognition of existence of an entity and the second—the predication itself, or the property attributed to that entity. We leave for future research to determine if the restrictor domains in structures with clitic left-dislocated elements discussed above are of a similar type.

<sup>10</sup> We can possibly extend the applicative analysis to accusative Experiencers discussed in (24–28).

## 6. Conclusion

In this paper I show that the presence of clitics in clitic doubling constructions in Bulgarian is not optional. It is also not necessarily linked to topicality, as seen from the evidence from focus-topic chains presented in this paper. The presence of the clitics together with an associate signals the presence of an applied argument, either clause-internally as an applicative (Type 1 with Experiencer predicates and ditransitives), or clause-externally in a higher predication domain. (Type 2 with clitic left-dislocated structures).

The proposal for Bulgarian is then that full-fledged arguments can be base-generated at two different levels of predication: as applicatives within the *vP* or in a higher predication domain above IP (TP). While displaying an array of features which distinguish them, Type 1 structures with left clitic dislocations and Type 2 structures with Experiencers can tentatively be unified from a semantic point of view as they both display a discourse subject.

Conseil des écoles catholiques du Centre-Est Ottawa (Canada)  
Education Permanente  
oarnaudova@gmail.com

## References

- Alexopoulou, Theodora, Edit Doron, and Caroline Heycock. (2003) "Broad subjects and clitic left dislocation". David Adger, Cécile de Cat, and George Tsoulas, eds. *Peripheries: Syntactic edges and their effects*. Dordrecht: Kluwer, 329–58.
- Andrejčin, Ljubomir, Konstantin Popov, Stojan Stojanov. (1977) *Gramatika na bŕlgarskija ezik*. Sofia: Nauka i izkustvo.
- Arnaudova, Olga (2002) "Clitic left dislocation and argument structure in Bulgarian". Jindřich Toman, ed. *Formal approaches to Slavic linguistics: The second Ann Arbor meeting 2001*. Ann Arbor, MI: Michigan Slavic Publications, 23–46.
- . (2003) *Focus and Bulgarian clause structure*. Ph.D dissertation, University of Ottawa.
- . (2010) *Focus and Bulgarian clause structure: Word order variation and prosody*. Saarbrücken: VDM Verlag.
- Arnaudova, Olga and Iliyana Krapova. (2007) "Clitic reduplication in Bulgarian: Towards a unified account". Richard Compton, Magdalena Golezdzinowska, and Ulyana Savchenko, eds. *Formal approaches to Slavic linguistics: The Toronto meeting, 2006*. Ann Arbor, MI: Michigan Slavic Publications, 1–24.



- Arregi, Karlos. (2003) "Clitic left dislocation is contrastive topicalization". *Proceedings of the 26th Annual Penn Linguistics Colloquium. Penn working papers in linguistics* 9(1): 31–44.
- Asenova, Petja. (2002) *Balkansko ezikoznanie: Osnovni problemi na balkanskija ezikov sajuz*. London: Faber.
- Bošković, Željko. (2008) "What will you have, DP or NP?" Emily Elfner and Martin Walkow eds. *NELS 37* Amherst: University of Massachusetts, Graduate Linguistic Student Association, 101–14.
- Büring, Daniel. (2003) "On D-trees, beans, and B-accents". *Linguistics and philosophy* 26(5): 511–45.
- Chomsky, Noam. (1995) *The minimalist program*. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. (1990) *Types of A-dependencies*. Cambridge, MA: MIT Press.
- Cuervo, María Cristina. (2003) *Datives at large*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Evans, Gareth. (1980) "Pronouns". *Linguistic inquiry* 11: 337–62.
- Franks, Steven L. (1995) *Parameters of Slavic morphosyntax*. New York: Oxford University Press.
- Franks, Steven L. and Tracy H. King. (2000) *A handbook of Slavic clitics*. New York: Oxford University Press.
- Franks, Steven L. and Catherine Rudin. (2005) "Bulgarian Clitics as K<sup>0</sup> heads". Steven L. Franks, Frank. Y. Gladney, and Mila. Tasseva-Kurktchieva, eds. *Formal approaches to Slavic linguistics: The South Carolina meeting, 2004*. Ann Arbor, MI: Michigan Slavic Publications, 104–16.
- Franks, Steven L. and Catherine Rudin. (2006) "What makes clitic doubling obligatory?" Ivan Kucarov et al., eds. *Ezikovedski izsledvanija c čest na 75-godišnjinata na st.n.s i st.d.f.n Jordan Penčev Penčev*, 267–78.
- Harizanov, Boris. (2014) "Clitic doubling at the syntax-morphophonology interface: A-movement and morphological merger in Bulgarian". *Natural language and linguistic theory* 32(4): 1033–88.
- Krapova, Iliyana. (2002) "On the left periphery of the Bulgarian sentence". *University of Venice working papers in linguistics* 12: 107–28.
- Krapova, Iliyana and Guglielmo Cinque. (2008) "Clitic reduplication constructions in Bulgarian". Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 257–87.
- Kuroda, S.-Y. (1972) "The categorical and the thetic judgment". *Foundations of language* 9: 153–85.
- Leafgren, John R. (1997) "Bulgarian clitic doubling: Overt topicality". *Journal of Slavic Linguistics* 5(1): 117–43.
- Manolova, Lilija. (1979) "Upotreba na udvoenoto dopâlnenie v bâlgarskija knižoven ezik". *Izvestija na Instituta za bâlgarski ezik*, 13. Sofia: BAN
- Marušič, Franc and Rok Žaucer. (2006) "On the intensional FEEL-LIKE construction in Slovenian: A case of a phonologically null verb". *Natural language and linguistic theory* 24(4): 1093–1159.

- Miklošič, Franz. (1861) "Die slavischen Elemente in Rumunischen". *Denkschriften der Kaiserlichen Akademie der Wissenschaften: Philosophisch-historische Klasse* 12: 1–70
- Penčev, Jordan. (1993) *Sintaksis na sâvremennija bâlgarski knizhoven ezik*. Plovdivsko univerzitetsko izdatelstvo.
- Pylkkänen, Liina. (2002) *Introducing arguments*. Ph.D. Dissertation, Massachusetts Institute of Technology.
- Reinhart, Tanya. (1981) "Pragmatics and linguistics: An analysis of sentence topics". *Philosophica* 27: 53–93.
- Rivero, María Luisa and Milena Sheppard. (2003) "Indefinite reflexive clitics in Slavic: Polish and Slovenian". *Natural language and linguistic theory* 21(1): 89–155.
- Rudin, Catherine. (1986) *Aspects of Bulgarian syntax: Complementizers and wh-constructions*. Columbus, OH: Slavica.
- . (1997) "Agr-O and Bulgarian pronominal clitics". M. Lindseth and S. Franks, eds. *Formal approaches to Slavic linguistics: The Indiana meeting 1996*. Ann Arbor, MI: Michigan Slavic Publications, 224–52.
- Slavkov, Nikolay. (2008) "Formal consequences of dative clitic doubling in Bulgarian ditransitives: An applicative analysis". *Journal of Slavic linguistics* 16: 139–66.
- Tomić, Olga. (2000) "On clitic sites". Frits H. Beukema and Marcel den Dikken, eds. *Clitic Phenomena in European Languages*. Amsterdam: John Benjamins, 293–316.
- Zubizarreta, Maria Luisa. (1999) "The clitic projection in questions". *Catalan working papers in linguistics* 7: 253–77.

# Ordering Restrictions if Both Internal Arguments are *Wh*-Phrases\*

Loren A. Billings

*Abstract:* Bulgarian ditransitive data in which both internal arguments are interrogative, as well as those with all three arguments as *wh*-phrases, are amassed and examined. The need to consider these structures is justified. Evidence of an arboreal asymmetry between the objects has not been found.

## 0. Introduction

After being invited to contribute to this volume, I consulted the recent literature on the morphosyntax of Bulgarian. One passage struck me in particular: “there is no consensus on the ordering restrictions (or lack thereof)” in questions with three *wh*-phrases (Bailyn 2017: 29, n. 2). Having worked on multiple questions in Bulgarian quite a while back (Billings and Rudin 1996), I set out to ascertain this observation’s accuracy. This paper sums up my findings on the overall issue.<sup>1</sup>

---

\* Comments from anonymous referees improved the paper considerably. Thanks also go to J. Allen, V. Chidambaram, B. Citko, R. Compton, D. Dyer, S. Fischer, G. Fowler, S. Franks, L. Grebenyova, S. Harves, F. Jaeger, I. Krapova, J. Leafgren, M. Sherkina-Lieber, V. Simeonova, J. Toman, and P. Weirich for various help along the way. They, and especially my erstwhile co-author cited below, aren’t to blame for errors herein.

<sup>1</sup> Three other studies, like Bailyn’s, steer clear of such questions. To begin, Citko eschews this issue altogether (at least in Bulgarian) by considering only (i) an external *wh*-argument with one internal *wh*-argument, (ii) an internal *wh*-argument with a *wh*-adjunct, or even (iii) a combination of all three (respectively: 1998: 98, 107–108; 97, 109; 110). Citko and Gračanin-Yuksek (2013: 14–15) likewise address the same three categories of *wh*-phrase but with no more than one internal *wh*-argument in the same example. Finally, Krapova and Cinque do not consider questions in which there are “more than two *wh*-phrases, which are said to allow free ordering of all but the first *wh*-phrase” (2005: 173, n. 2/2008: 319, n. 1). As examples (14a–b) below show clearly, this is true only if all three of the argument *wh*-phrases denote humans. I find it interesting that the very first studies of multiple *wh*-fronting in Slavic included Polish data with both internal arguments as *wh*-phrases (Wachowicz 1974a: 161–62, 1974b: 71, cited in part by Cheng 1991: 77/1997: 64). See also Cheng (1991: 94/1997: 77).

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin*. Bloomington, IN: Slavica Publishers, 2018, 21–36.

There certainly have been claims about the relative order of the fronted DO and IO if both are *wh*-phrases, though—surprisingly to me—not many.<sup>2</sup> Such claims are usually linked to single examples. I therefore consider these claims as the data are presented. In each case, I assess the relevant evidence mainly in terms of three factors: arboreal superiority, animacy, and whether consecutive *wh*-homophones occur.

All of the data presented in this paper involve verbs with three arguments, with at least both of the internal arguments being *wh*-phrases. Section 1 examines questions with just one unique verb, *pitam* ‘ask’, in which *na* ‘to’ (which is historically a preposition) doesn’t precede either internal argument.<sup>3</sup> The two objects in that section’s data are *kakvo* ‘what’ and *kogo* ‘whom’. The remainder of the paper then considers the relevant data with all other ditransitive verbs; these each require IOs preceded by *na*. Section 2 looks at such questions where the two internal arguments differ in animacy: *kakvo* ‘what’ and *na kogo* ‘to whom’. Section 3 then assesses questions in which both *wh*-phrases are human. Here the internal arguments are *kogo* ‘whom’ and *na kogo* ‘to whom’. In addition, after starting with questions where only the internal arguments are *wh*-phrases, each section moves to data where there is also a human external *wh*-argument, encoded by *koj* ‘who’. For expository convenience, modifying Błaszczak and Fischer’s distinction (2001: 60), I call these binary and ternary questions, respectively.<sup>4</sup>

## 1. Questions with *kakvo* ‘what’ plus *kogo* ‘whom’

This section’s data are limited to two combinations of *wh*-phrases: the entirety of the data reported in the literature. Still, a number of issues relevant to the rest of this paper are introduced and elucidated here.

<sup>2</sup> These abbreviations are used: DO direct object, FUT future (tense), IO indirect object, M masculine (gender), PRS present (tense), PST past (tense)/aorist, Ptcp (past) participle, Q interrogative, S subject, SG singular (number), 2 second (person), 3 third (person).

<sup>3</sup> Following Bošković (1997: 238–39, n. 17), I assume *na* is no longer a syntactic head. In the translations, I use ‘(to) whom’ for (*na*) *kogo* regardless how stilted it sounds in the English. As such, I generally don’t gloss the morphological case of each *wh*-phrase.

<sup>4</sup> This study does not discuss either non-argument or D[iscourse]-linked *wh*-phrases. The latter are doubled by object-agreement clitics. The only data I am aware of where there are two internal *wh*-arguments but no external argument require such doubling (Billings and Rudin 1996: 40, 44; Błaszczak and Fischer 2001: 59; Dimova 2010: 67; Fanselow 2004: 108; Grohmann 2006: 282, n. 26; Ishioka 2005: 153; Jaeger 2004: 212; Krapova and Cinque 2005: 172/2008: 318; Lahne 2012: 291; Rudin 1986: 161/2013: 171).

In Bulgarian perhaps the most startling property of multiple questions is that the contrast in binary (1) can disappear in (2) if *koj* ‘who’ replaces the external argument in (1): *Ivan*, a personal name.<sup>5</sup>

- (1) a. <sup>(?)</sup>\***Kakvo kogo** e pital Ivan?  
 b. **Kogo kakvo** e pital Ivan?  
 whom what *S<sub>PRS.3SG</sub>* ask<sub>PTCP.M.SG</sub> Ivan  
 ‘Whom has Ivan asked what?’ [Bošković 1997: 239]
- (2) a. <sup>(\*)</sup>**Koj kakvo kogo** e pital?  
 b. **Koj kogo kakvo** e pital?  
 who whom what *S<sub>PRS.3SG</sub>* ask<sub>PTCP.M.SG</sub>  
 ‘Who has asked whom what?’  
 c. <sup>\*</sup>**Kakvo koj kogo** e pital? e. <sup>\*</sup>**Kogo koj kakvo** e pital?  
 d. <sup>\*</sup>**Kakvo kogo koj** e pital? f. <sup>\*</sup>**Kogo kakvo koj** e pital?

Rudin (1986: 115–16/2013: 124) lists each of (2a, c–f) with a preceding asterisk; nearly every subsequent work shows (2a) as fully acceptable.<sup>6</sup> Bošković (1997: 239–40, including n. 18), the first author to show (2a) as grammatical, doesn’t elucidate how this example, unacceptable in Rudin 1986/[2013], came to be completely acceptable in his own study.

To account for his ordering preference in (1b), Bošković suggests that *kakvo* ‘what’ be “Case-checked structurally in the Spec[ifier] of an Agr[reement]P[hrase] that is lower than the SpecAgrP in which *kogo* [‘whom’] is Case-checked” (1997: 239–40), adding that “there would be an AgrP for each object NP” (1997: 240). Most though not all of the publications that list (1b) also repeat such an argument by Bošković.<sup>7</sup>

<sup>5</sup> This loss of contrast in ternary questions occurs again in §2. See also only (16) in §3.

<sup>6</sup> Just one author (who does not cite Bošković 1997) lists Rudin’s judgment about the unacceptability of (2a): Mulders (1997: 138) reports (2b) as the only acceptable order. Later, the opposite (good and bad) *wh*-orders in only (2b, d) are contrasted (1997: 139).

<sup>7</sup> These studies have cited both (1a–b) and (2a–b): Błaszczak and Fischer 2001: 59–60; Bošković 2002: 366, 2010: 23; *Grebenyova 2004: 173; Lambova 2000: 248* [but without final *Ivan* in (1a–b)], 2001: 331–32; *Reglero 2003: 190–91*; and Richards 1997: 277, 299, 332, 2001: 213–14, 248–49, 282. Only (1a–b) also appear in *Boeckx and Grohmann 2003: 3*; *Dukova-Zheleva 2010: 34*; *Jeong 2004: 111, n. 5*; *Krapova 2006: 248*; and *Krapova and Cinque 2005: 179/2008: 323*. Incidentally, the italicized sources so far in this footnote actually show (1a) as totally unacceptable: preceded by an asterisk (rather than by <sup>?</sup> as reported by Bošković). Next, Moro (2011: 409) and Richards (1997: 12, 2001: 1) each list just (1b). Finally, each of *Boeckx (2003: 17)*, *Boeckx and Grohmann (2003: 5)*, *Boeckx and Hornstein (2008: 198)*, and *Jeong (2004: 103)* list all of (2a–f). Only (2a–b) also ap-

Another approach to the contrast in (1a–b) relies on the differing animacy of these *wh*-phrases: *kakvo* ‘what’ vs. *kogo* ‘whom’. Observe that all three arguably good orders begin with human *wh*-phrases: *kogo* ‘whom’ in (1b) or *koj* ‘who’ in (2a–b). To be sure, animacy alone would not rule (2e–f) out. Whereas both superiority and animacy (of only the first *wh*-phrase) hold for ternary (2a–b), perhaps animacy is what differentiates binary (1a–b) from each other. No publication to date has taken up this issue. Though Billings and Rudin 1996 argues for animacy as a factor in ordering *wh*-phrases, it does not consider data like (1) and (2), and Bošković (1997) does not accept animacy as a factor.<sup>8</sup> Moreover, the wording of the animacy proposal in Billings and Rudin (1996: 46) would not affect (1a–b): “If the external argument is not a *wh* phrase, [...] then there can be any ordering in the *wh* cluster.” Indeed, Dukova-Zheleva (2010: 161) points out that her experimental results, similar to the contrast in (1a–b) above, corroborate the more general observation in Rudin 1986/2013, discussed below, that animate *wh*-phrases are allowed to precede inanimate ones, “but also revealed that this is a general property (rather than a property observed only in external arguments as suggested by Billings and Rudin [...]).”

The contrast between the only fully acceptable order in binary (1b) and the variation, to at least some speakers, in ternary (2a–b) requires an account in any theory. An approach rejecting animacy is forced to propose that the internal arguments are not equidistant to their *wh*-destination because one

---

pear in Bošković 1997: 239; Dimova 2010: 27; Fanselow 2004: 109; Gärtner and Michaelis 2007: 189; Liakin 2007: 284–85; McKinney-Bock 2013: 44; Mizuguchi 2014: 25; Pesetsky 2000: 24; Pesetsky and Torrego 2001: 368; and Sabel 2001: 538, n. 4—whereas Grewendorf (2001: 98) lists only (2b), and Richards (1997: 332, 2001: 282 [both of his works also citing Roumyana Izvorski, personal communication]) reports only (2e).

<sup>8</sup> In a subsequent publication, Bošković (2010: 8, n. 7) flatly rejects any such animacy effect: “It should be noted here that Billings and Rudin [...] report a possible exception to Superiority in Bulgarian concerning inanimate *wh*-phrases. My informants do not share the crucial judgments reported by Billings and Rudin.” Bošković then adds that “in Pesetsky (2000: 57) [...] the same holds for his and Norvin Richards’s informants.” The page number cited there is incorrect. In fact, Pesetsky (2000: 98, n. 24) writes that Billings and Rudin “identify several factors that license exceptions to the Superiority effect among their Bulgarian consultants. These include animacy and avoidance of phonetic identity. My informants and Richards’s have not assented to the judgments reported by Billings and Rudin.” Incidentally, Richards (to my knowledge) does not weigh in on either of the factors that Pesetsky mentions regarding Bulgarian in any publication. However, Richards (2010: 51) acknowledges the avoidance of consecutive *wh*-homophones in Serbian. Moreover, the excerpt from Bošković (2010) is somewhat inaccurate in the sense that Pesetsky rejects Billings and Rudin’s observations about both animacy and consecutive homophones, yet Bošković has routinely accepted the observation in Billings and Rudin 1996 about consecutive homophony (e.g., Bošković 1997: 238, n. 17, 2002: 365, 2010: 11, 19–20; Bošković and Franks 2002: 51–52, 69, n. 6).

of these is merged arboreally higher than the other. By contrast, any model sensitive to animacy need not resort to such an assumption.

To summarize the overall discussion so far, (2a–b) show how in Bulgarian there is variation for some speakers in a ternary question. This difference is exactly the kind described by Bailyn. I will not attempt to sort out this empirical discord. I will assume only that there are two groups of judgments: good and bad (2a). However, using either superiority or animacy can select (1b) over (1a), though Billings and Rudin's proposal for harnessing animacy as a factor would still need to be modified to account for the data in (1) and (2). A third factor, consecutive homophony, is not relevant to any of the data thus far.<sup>9</sup>

An analysis of the syntax and semantics of *pitam* 'ask'—taking two objects, neither of which is preceded by *na*—lies beyond the scope of this paper. I have presented the facts, along with the disagreements in judgments, in hopes that a later study might make sense of *pitam* and how it differs from the other ditransitive verbs in its *wh*-ordering.

## 2. Questions with *kakvo* 'what' plus *na kogo* 'to whom'

As in the preceding section, animacy is relevant here. Once more, authors disagree about the acceptability of various *wh*-phrase orders.

Starting with data involving just internal arguments as *wh*-phrases, the literature reports three distinct patterns of acceptability judgments, each inconsistent with the other two.<sup>10</sup> The first set, shown in (3) and (4), requires the IO to appear first in the *wh*-cluster, as in (1b) above.

- (3) a. \***Kakvo na kogo** e kazal Petâr?  
 b. Na **kogo kakvo** e kazal Petâr?  
 to whom what S<sub>PRS.3SG</sub> say<sub>PTCP.M.SG</sub> Petâr  
 'What has Petâr said to whom?' [Rudin 1986: 119/2013: 127]

<sup>9</sup> It would be worthwhile to ascertain the acceptability of the colloquial counterpart of (2b), as reported in Billings and Rudin 1996: 44, where the DO *wh*-phrase is *koj* 'who' (as is the external-argument *wh*-phrase). Alas, I haven't undertaken to elicit such data.

<sup>10</sup> Comorovski (1996: 127) also lists only (3a). Lambova (2004: 15, 96) lists data identical to (5a–b) except for having *e dal* 'has given' and *si dal* 'have<sub>2SG</sub> given<sub>PTCP.M.SG</sub>' (both in place of *e kazal* 'has said'). Lambova (2002: 122, 2008: 338) also shows (6a–b) without *Ivan* (and translated, respectively, as 'What has he given to whom as a gift?' or 'What did he give to whom as a present?'). Similarly, both Bošković (2002: 352) and Krapova (2006: 252) use *Ivan* at the end of (7b); in the translation, 'Ivan' replaces 'Marta'. Dukova-Zheleva (2010: 28, 32, 35) lists data similar to (7a–b) but either ending in *kupi Marija?* '... did Marija buy ...?' or without *Marta* at the end (her translation being 'What did you give to whom?'). Dimova (2010: 81) lists (7a) without *Marta* and preceded by a question mark, consistent with the first judgment set. Finally, Ishioka (2005: 166) cites (8a–b).

- (4) a. <sup>?</sup>**Kakvo** na **kogo** e pokazal Ivan?  
 b. Na **kogo** **kakvo** e pokazal Ivan?  
 to whom what S<sub>PRS.3SG</sub> show<sub>PTCP.M.SG</sub> Ivan  
 ‘What has Ivan shown to whom?’  
 [Krapova and Cinque 2005: 179/2008: 323; parentheses in (4a) *sic*]

The next set of judgments, in (5) and (6), are the opposite of those in (3) and (4) immediately above.<sup>11</sup> Here only the (a) examples are good:

- (5) a. **Kakvo** na **kogo** e казал?  
 what to whom S<sub>PRS.3SG</sub> say<sub>PTCP.M.SG</sub>  
 ‘What has he said to whom?’  
 b. \*Na **kogo kakvo** e казал? [Lambova 2003: 317]
- (6) a. **Kakvo** na **kogo** e podaril Ivan?  
 what to whom S<sub>PRS.3SG</sub> gift<sub>PTCP.M.SG</sub> Ivan  
 ‘What has Ivan given to whom as a gift?’  
 b. \*Na **kogo kakvo** e podaril Ivan? [Lambova 2004: 80, n. 22]

The third distinct set of judgments is shown in (7) and (8), where multiple authors report that either order of the *wh*-phrases is accepted.

- (7) a. **Kakvo** na **kogo** dade Marta?  
 what to whom give<sub>PST.2/3SG</sub> Marta  
 b. Na **kogo kakvo** dade Marta?  
 ‘What did Marta give to whom?’ [Dukova-Zheleva 2010: 262–63]
- (8) a. **Kakvo** na **kogo** e dal Ivan?  
 what to whom S<sub>PRS.3SG</sub> give<sub>PTCP.M.SG</sub> Ivan  
 b. Na **kogo kakvo** e dal Ivan?  
 ‘What has Ivan given to whom?’ [Grewendorf 2001: 97, n. 19]

Krapova and Cinque contrast the first and third sets: “Some speakers seem to fully accept such examples [i.e., in (8a–b)] while others seem to fully exclude them[, as in (3a–b)]” (2008: 323, n. 7/cf. 2005: 179, n. 7). Contrasting the last two

<sup>11</sup> Richards (2014: 174) shows the pair of examples in (8) but lists (8b) as bad, implying that these are as reported in Rudin (1988b) but not specifically attributing the example to any previous study. Though this pattern is as in (5) and (6)—otherwise reported only in Lambova 2002, 2003, 2004, 2008—Richards doesn’t cite any work by Lambova.



sets of judgments, Jaeger (2004: 210–11) cites examples (5a–b) and (8a–b). No publication until now has contrasted either just the first two sets or all three sets of conflicting judgments.

Though the variation represented by the three judgment sets in (3) and (4), (5) and (6), and (7) and (8) is not what Bailyn is referring to, because these include only two *wh*-phrases each, this constitutes by far the greatest variation reported in this paper. Rather than trying to account for these three judgment sets, I compare the consensus in (1), where there is no historically prepositional *na*, to the extreme variation in (3) through (8), where *na* is obligatory. To begin, for the most part, Bošković has remained silent about the data here in section 2, listing only in passing (2002: 352) an example similar to (7b), which implies only that that any speakers he consulted disfavor the second judgment set, in (5) and (6). Next, Dukova-Zheleva accepts that the contrast in (1a–b) above reflects the syntactic superiority of the IO prior to *wh*-movement, whereas she assumes that in (7) it is the DO that is arboreally superior (2010: 34–35, 262–63), and the animacy of only the IO in (7) allows superiority to be overridden optionally. Similarly, Lambova's earlier work (2000: 248, 2001: 331–32) assumes that (1b) reflects the superiority relationship between the internal arguments, with the IO being higher. However, Lambova's later work (2002, 2003, 2004, 2008) does not cite (1a–b) at all, and consistently uses only data like (5) and (6) instead to illustrate that the DO is the arboreally higher internal argument. (Bošković 1997 is not even cited in Lambova 2002, 2003, 2008.) Indeed, the very same clause is repeated verbatim to establish seemingly conflicting views: "The structurally highest *wh*-phrase surfaces first" (2000: 248, 2001: 331, 2004: 15). Lambova adds, "I assume that in a double object construction the direct object surfaces before the indirect one as indicated by the Superiority effect in [the contrast between (6a) and (6b)]" (2004: 80, n. 22). If Lambova's first two and last four works are to be reconciled, then they resemble Dukova-Zheleva's approach—except that there is no optionality to contend with in (5) and (6), where only the (a) examples are reported to be acceptable. As it were, the sum of Dukova-Zheleva's statements is Lambova's combined proposals but with the animacy effect added. In my view, the easiest of the three sets of judgments in this section to justify, assuming (1b) above, is the first one. Arboreal asymmetry is not necessary to differentiate the internal arguments from each other; animacy simply determines the preferred outcome in each instance.<sup>12</sup>

<sup>12</sup> Dukova-Zheleva (2010: 67, 81–84); Pesetsky (2000: 21); and Rudin (1988a: 7, 1988b: 451) each report *wh*-extraction data from ditransitive clauses, with only the two internal arguments as *wh*-phrases. All three seem consistent with the first set of judgments; Pesetsky's and Rudin's data, also with the third judgment set. Other data seem to be like the second judgment set (Richards 1997: 295, 2001: 245). Yet other data appear to resemble the third judgment set (Richards 1997: 298, 2001: 248, 288).

Moving, then, to ternary questions—with *koj* ‘who’, *kakvo* ‘what’, and *na kogo* ‘to whom’—we see near consensus.<sup>13</sup> As long as *koj* ‘who’ is initial, then the internal *wh*-arguments can appear in either order:

- (9) a. *Koj kakvo na kogo e kazal?*  
 who what to whom  $S_{PRS.3SG}$  say $_{PTCP.M.SG}$   
 [Lambova 2003: 317, 2004: 2; Rudin 1986: 116, 119/2013: 124, 127]
- b. *Koj na kogo kakvo e kazal?*  
 ‘Who has said what to whom?’  
 [Lambova 2003: 317; Rudin 1986: 116/2013: 124]
- (10) a. *Koj kakvo na kogo dade?*  
 who what to whom give $_{PST.2/3SG}$
- b. *Koj na kogo kakvo dade?*  
 ‘Who gave what to whom?’ [Dimova 2010: 23]

As in the view in which (2a) is allowed, the only requirement relevant to ordering the *wh*-phrases in (9) and (10) is that *koj* ‘who’ be first.

Since the *wh*-phrases in this section are each distinct as to their form, consecutive homophony does not arise—even in the register where *na koj* ‘to who’ is the *wh*-IO (Billings and Rudin 1996: 43–46). Nonetheless, the precise analogue of (9) and (10) in a non-Indo-European language demonstrates that homophony can be a factor:

- (11) Inuktitut (South Baffin dialect)
- a. *Kina sunamit kinamut tunisiva:?*  
 who what $_{ACCUSATIVE}$  who $_{ALLATIVE}$  give $_{3SG.Q}$   
 ‘Who gave what to whom?’
- b. *\*Kina kinamut sunamit tunisiva:?* [Sherkina-Lieber 2004: 125]

As in Bulgarian, the only syntactic ordering requirement is that *kina* ‘who’ (unsuffixed) be first. Because the IO in colloquial Bulgarian is marked by preceding *na*, if *koj* ‘who’ immediately precedes the IO *na koj* ‘to who’, then

<sup>13</sup> The data in (9) and (10) are representative. Billings and Rudin (1996: 37) also list (9b). Pesetsky (2000: 6, 19) lists (10a). There are several more pairs like these in the literature; see, e.g., Lambova 2002: 122, 2004: 15, 95, 2008: 338 and Richards 1997: 98. However, Pesetsky (2000: 24, 48, repeated and cited in Lotfi 2003: 170–71) also lists (10a–b) with one question mark before only (10a); this is the only such deviating judgment I am aware of. I don’t know what to make of (10a) being good in Pesetsky 2000: 6, 19 but only slightly degraded elsewhere in the same publication (2000: 24, 48).

consecutive homophony would still not occur. By contrast, because case is marked suffixally, *kina* ‘who’ followed by *kinamut* ‘who<sub>ALLATIVE</sub>’ allows for *kina* to occur twice in succession. Thus, consecutive homophony is also relevant here, ruling out (11b).

To conclude section 2, there are three distinct sets of judgments in the binary questions (with *kakvo* ‘what’ plus *na kogo* ‘to whom’). By contrast, there is near consensus about the corresponding ternary data.

### 3. Questions with *kogo* ‘whom’ plus *na kogo* ‘to whom’

The last type of data discussed in this paper is where both of the internal *wh*-arguments encode human beings. Unlike in the previous sections, here there is consensus about the data but alas a distinct lack of concord about how this evidence is to be explained. Fortunately, this problem can be solved with more careful elucidation of the facts.

Just two pairs of such binary questions exist in the literature:

- (12) a. **Kogo** na **kogo** e pokazal Ivan?  
whom to whom *S<sub>PRS.3SG</sub>* show<sub>*PTCP.M.SG*</sub> Ivan  
‘Whom has Ivan pointed out to whom?’
- b. \*Na **kogo kogo** e pokazal Ivan? [Rudin 1986: 116/2013: 124]
- (13) a. **Kogo** na **kogo** šte predstavíš?  
whom to whom *FUT* introduce<sub>*PRS.2SG*</sub>  
‘Whom will you introduce to whom?’
- b. \*Na **kogo kogo** šte predstavíš?  
[Krapova and Cinque 2005: 177/2008: 322]

And there is only one corresponding ternary question, as follows.<sup>14</sup>

<sup>14</sup> Billings and Rudin (1996: 41, 45); Bošković (1997: 238, n. 17, 2010: 11); and Ishioka (2005: 152) each list (12a–b). Błaszczak and Fischer (2001: 27) and Comorovski (1996: 127) also list only (12a). Only Mišmaš (2015: 91) also lists (13a–b). Arnaudova (2003: 155, n. 20); Billings and Rudin (1996: 43, 45); Błaszczak and Fischer (2001: 62); Bošković (1997: 238, n. 17, 2002: 365); Bošković and Franks (2002: 52–53); Bošković and Nunes (2007: 46); and Fanselow (2004: 110) list (14a–b). Rudin lists only (14a), as does Lahne (2012: 289); both add that all other orders are illicit. Błaszczak and Fischer (2001: 27) also list only (14a), whereas Franks (2017: 135) lists only (14b). Ishioka (2005: 166) lists (15a–b). Dukova-Zheleva (2010: 67, 81, 83–84) reports an experiment using *wh*-phrases identical to (15a–b) but extracted to a higher clause. Participants favored, significantly,  $p < .001$ , (15a). Błaszczak and Fischer (2001: 62) and Fanselow (2004: 110) list (16a–b). Bošković and Franks (2002: 69, n. 6) allude to, and Franks (2017: 135) lists, only (16b).

- (14) a. *Koj kogo na kogo e pokazal?*  
 who whom to whom  $S_{PRS,3SG}$  show $_{PTCP.M.SG}$
- b. \**Koj na kogo kogo e pokazal?*  
 ‘Who has pointed out whom to whom?’ [Rudin 1988b: 473]

Numerous studies, including the original sources of each of (12) through (14), have used these data to argue that if there are two internal *wh*-arguments, the DO must go first—though Krapova and Cinque (2005: 176/2008: 321) and Rudin (1986: 116/2013: 124) are both careful to observe that this statement is true only if both *wh*-phrases specifically denote humans. However, Rudin’s wording has not been as precise in a later paper: “In Bulgarian, a nominative must precede an accusative Wh-word, and when a Wh-word indirect object is also present, the order of the three Wh-words must be subject, direct object, indirect object” (1988b: 472), adding, in connection with (14a), that “no other word order possible” (1988b: 473). Note no mention of animacy in these excerpts. Lahne (2012: 289) apparently takes this to mean that questions like (9b) and (10b) above are similarly ruled out, proposing (2012: 290) an account of (14a) based on that erroneous assumption.<sup>15</sup> Similarly to Rudin (1988b), Krapova and Cinque (2005: 176/2008: 321), write, “Multiple questions containing two [+human] *wh*-objects [...] show a strict ordering. As noted by Billings and Rudin (1996, 41), [...] the direct *wh*-object must always precede the indirect *wh*-object.” The mention of animacy in the first quoted sentence does not seem to affect “always” in the second sentence. A diagram (2005: 177/2008: 322, cited in Rojina 2011: 44) and table (2005: 184/2008: 327, cited in both Mišmaš 2015: 91 and Rojina 2011: 49) also formalize this overgeneralization about the precedence.<sup>16</sup> It would be instructive to peruse the rest of the page that Krapova and Cinque cite (i.e., Billings and Rudin 1996: 41): “It looks as though the DO must be superior to (or higher in the syntactic tree than) the IO. We return to this issue (in §[...]) using data from colloquial Bulgarian, showing that **there is no syntactic *wh*-ordering requirement** [...] and this factor can be conveniently controlled for” (emphasis added/L.A.B.). The colloquial data in the later section alluded to in this excerpt are now listed in (15) and (16), both from Billings and Rudin 1996: 45, corresponding to (12) and (14) above in the standard register of Bulgarian, respectively.<sup>17</sup>

<sup>15</sup> Lahne (2012: 291) cites Billings and Rudin’s paper but cites neither (9b) above nor the pages that discuss avoidance of consecutive *wh*-homophony (1996: 37, 41, 44–46).

<sup>16</sup> See a similar claim by Krapova and Cinque (2005: 190, n. 18/2008: 331, n. 18).

<sup>17</sup> Krapova also writes, “*wh*-Direct objects (*kogo* [‘whom’]) must precede *wh*-Indirect objects (*na kogo* [‘to whom’])” (2006: 244). This predicts that if the DO is *kakvo* ‘what’ (rather than *kogo* ‘whom’), then the order of *wh*-phrases would be as in (5) and (6): a *wh*-phrase ordering that is not recognized by Krapova and Cinque (2005/2008).

- (15) a. **Kogo na koj e pokazal Ivan?**  
 whom to who  $S_{PRS,3SG}$  show $_{PTCP,M,SG}$  Ivan
- b. **Na koj kogo e pokazal Ivan?**  
 ‘Whom has Ivan pointed out to who?’
- (16) a. **Koj kogo na koj e pokazal?**  
 who whom to who  $S_{PRS,3SG}$  show $_{PTCP,M,SG}$
- b. **Koj na koj kogo e pokazal?**  
 ‘Who has pointed out whom to who?’

In this (moderately) colloquial register, the *wh*-IO is *na koj* ‘to who’ (rather than *na kogo* ‘to whom’ in the literary language) but the *wh*-DO remains *kogo* ‘whom’ (as in standard Bulgarian). The result is that no instance of *koj* ‘who’ (either by itself or as part of *na koj* ‘to who’) can result in consecutive *wh*-homophony, thus predicting that either order is possible, which is indeed what we find.<sup>18</sup> In (16b) this is true as long as *koj* ‘who’ is cluster-initial (owing to external-argument superiority).

To summarize section 3, then, consecutive homophony is avoided in both binary questions, in (12a) and (13a), and ternary ones, in (14a). Thus, unlike both superiority and animacy (which are relevant to the selection of only the first of two or more *wh*-phrases), this constraint affects any overt sequence of *wh*-forms that are pronounced alike.

#### 4. Conclusion

Stitching together the three body sections of this paper, I start with the ternary data. We see that if *koj* ‘who’ is the external argument, then it must be initial, and the order of two internal *wh*-arguments is free of syntactic or pragmatic constraints, as in (2)—for at least some speakers—as well as (9), (10), and (16). Only if the two internal *wh*-arguments result in consecutive homophony is one order eliminated, in (14b). The more complicated data are the binary questions. The data from sections 1 and 3—in (1) vs. (12) or (13)—from a syntactic point of view appear contradictory. However, once animacy in (1) and the avoidance of consecutive homophony in (12) and (13) are considered, then there is no need to propose that the two internal arguments are merged in an asymmetric arboreal relation. (I’ll leave it to others to connect the pipes.) We

<sup>18</sup> Krapova and Cinque seem to have independently observed that consecutive *wh*-homophones are disallowed, though without any examples or specifics: “the apparent ban on combining a *kakvo* subject with a *kakvo* object, [...] regardless of order” (2005: 181, n. 8/2008: 325, n. 8). For representative examples, see Franks (2017: 132). I am preparing a separate paper on that related issue (currently entitled “Failure to undergo *wh*-fronting in order to avoid overt ‘who who’ or ‘what what’ sequences”).

are thus left with the thorniest of the binary-question data, in (3) through (8). I'm inclined to go with the first set of judgments, in (3) and (4), where animacy decides the relative order of the two *wh*-objects: *na kogo* 'to whom' precedes *kakvo* 'what'. Still, there is some credibility with the third set of judgments, in (7) and (8), both in the array of authors that have reported these data and the internal consistency of individual authors' accounts. If there exist speakers for whom animacy is not a factor, in (7) and (8), I wonder why no author has reported that (1a) is fully acceptable for even one of the speakers consulted. The only clue in this regard is that (1a) is not deemed fully unacceptable by Bošković and several other authors—though not all of the ones that cite (1a). Alas, I cannot pursue the issue further here.

Returning to my opening discussion, Bailyn's avoidance of ternary questions as such is valid, but only with regard to (2a). There is consensus about the facts in all other such data: (9) and (10), each with free variation, and (14), where consecutive homophony causes only one order to result. Still, why (14) remains elusive in much of the literature also might be what Bailyn is alluding to. The real challenge turns out to be binary questions, with the only *wh*-phrases being the internal arguments. Though the solutions have been few, perhaps the most satisfying result of this investigation for me has been that both categories of question—binary and ternary—still require investigation. Looking at both types has crystallized my view that both superiority and animacy are relevant to selecting only the first of two or more *wh*-phrases. By contrast, consecutive homophony is relevant to the entire *wh*-cluster. Finally, I remain unconvinced that, prior to fronting, either of the internal *wh*-arguments merges arboreally higher than the other.

Palacký University  
sgnillib@gmail.com

## References

- Arnaudova, Olga. (2003) *Focus and Bulgarian clause structure*. Ph.D. dissertation, University of Ottawa.
- Bailyn, John Frederick. (2017) "Bulgarian superiority and minimalist movement theory". Yohei Oseki, Masha Esipova, and Stephanie Harves, eds. *Formal approaches to Slavic linguistics: The NYU meeting, 2015*. Ann Arbor, MI: Michigan Slavic Publications, 27–49. [*Michigan Slavic materials*, 62.]
- Billings, Loren and Catherine Rudin. (1996) "Optimality and superiority: A new approach to multiple-*wh* movement". Jindřich Toman, ed. *Formal approaches to Slavic linguistics: The College Park meeting, 1994*. Ann Arbor, MI: Michigan Slavic Publications, 35–60. [*Michigan Slavic materials*, 38.]

- Błaszczak, Joanna and Susann Fischer. (2001) *Multiple Wh-Konstruktionen im Slavischen: Ein „state of the art report“*. Potsdam: Publikationsstelle, Universitätsbibliothek. [*Linguistics in Potsdam*, 14.]
- Boeckx, Cedric. (2003) “Symmetries and asymmetries in multiple checking”. Cedric Boeckx and Kleanthes K. Grohmann, eds. *Multiple wh-fronting*. Amsterdam: Benjamins, 17–26. [*Linguistik aktuell*, 64.]
- Boeckx, Cedric and Kleanthes K. Grohmann. (2003) “Introduction”. Cedric Boeckx and Kleanthes K. Grohmann, eds. *Multiple wh-fronting*. Amsterdam: Benjamins, 1–15. [*Linguistik aktuell*, 64.]
- Boeckx, Cedric and Norbert Hornstein. (2008) “Superiority, reconstruction, and islands”. Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarretta, eds. *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*. Cambridge, MA: MIT Press, 197–225. [*Current studies in linguistics*, 45.]
- Bošković, Željko. (1997) “On certain violations of the superiority condition, AgrO, and economy of derivation”. *Journal of linguistics* 33(2): 227–54.
- . (2002) “On multiple *wh*-fronting”. *Linguistic inquiry* 33(3): 351–83.
- . (2010) “Wh-phrases and *wh*-movement in Slavic”. Steven L. Franks, ed. *Contemporary issues in Slavic linguistics*. Durham, NC: Slavic and East European Language Resource Center, 1–45. [*Glossos*, 10.]
- Bošković, Željko and Steven Franks. (2002) “Phonology-syntax interactions in South Slavic”. *Balkanistica* 15: 49–74.
- Bošković, Željko and Jairo Nunes. (2007) “The copy theory of movement: A view from PF”. Norbert Corver and Jairo Nunes, eds. *The copy theory of movement*. Amsterdam: Benjamins, 13–74. [*Linguistik aktuell*, 107.]
- Cheng, Lisa Lai Shen. (1991) *On the typology of Wh-questions*. Ph.D. dissertation, Massachusetts Institute of Technology [published (1997) New York: Garland].
- Citko, Barbara. (1998) “On multiple WH movement in Slavic”. Željko Bošković, Steven Franks, and William Snyder, eds. *Formal approaches to Slavic linguistics: The Connecticut meeting, 1997*. Ann Arbor, MI: Michigan Slavic Publications, 97–113. [*Michigan Slavic materials*, 43.]
- Citko, Barbara and Martina Gračanin-Yuksek. (2013) “Towards a new typology of coordinated *wh*-questions”. *Journal of linguistics* 49(1): 1–32.
- Comorovski, Ileana. (1996) *Interrogative phrases and the syntax-semantics interface*. Dordrecht: Kluwer. [*Studies in linguistics and philosophy*, 59.]
- Dimova, Elena. (2010) *Questions à mouvement multiple en bulgare*. M.A. thesis, Université de Montréal.
- Dukova-Zheleva, Galina. (2010) *Questions and focus in Bulgarian*. Ph.D. dissertation, University of Ottawa.
- Fanselow, Gisbert. (2004) “The MLC and derivational economy”. Arthur Stepanov, Gisbert Fanselow, and Ralf Vogel, eds. *Minimality effects in syn-*

- tax. Berlin: Mouton de Gruyter, 73–123. [*Studies in generative grammar*, 70.]
- Franks, Steven L. (2017) *Syntax and spell-out in Slavic*. Bloomington, IN: Slavica Publishers.
- Gärtner, Hans-Martin and Jens Michaelis. (2007) "Some remarks on locality conditions and minimalist grammars". Uli Sauerland and Hans-Martin Gärtner, eds. *Interfaces + Recursion = Language?* Berlin: Mouton de Gruyter, 161–95. [*Studies in generative grammar*, 89.]
- Grebenyova, Lydia. (2004) "Interpretation of Slavic multiple *wh*-questions". Olga Arnaudova, Wayles Browne, María Luisa Rivero, and Danijela Stojanović, eds. *Formal approaches to Slavic linguistics: The Ottawa meeting, 2003*. Ann Arbor, MI: Michigan Slavic Publications, 169–85. [*Michigan Slavic materials*, 49.]
- Grewendorf, Günther. (2001) "Multiple *wh*-fronting". *Linguistic inquiry* 32(1): 87–122.
- Grohmann, Kleanthes K. (2006) "Top issues in questions: Topics—topicalization—topicalizability". Lisa Lai-Shen Cheng and Norbert Corver, eds. *Wh-movement: Moving on*. Cambridge, MA: MIT Press, 249–88. [*Current studies in linguistics*, 42.]
- Ishioka, Seizo. (2005) "Mitigation of the *wh*-island constraint". *Nagaoka gijutsu kagaku daigaku gijutsu kagaku ronshu (University of Nagaoka technology and science bulletin of language science and humanities)* 19: 135–70.
- Jaeger, T. Florian. (2004) "Topicality and superiority in Bulgarian *wh*-questions". Olga Arnaudova, Wayles Browne, María Luisa Rivero, and Danijela Stojanović, eds. *Formal approaches to Slavic linguistics: The Ottawa meeting, 2003*. Ann Arbor, MI: Michigan Slavic Publications, 207–27. [*Michigan Slavic materials*, 49.]
- Jeong, Youngmi. (2004) "Multiple superiority effects in Basque". *University of Pennsylvania working papers in linguistics* 10(1): 103–14.
- Krapova, Iliyana. (2006) "On D-linking in Bulgarian multiple *wh*-question". Ivan Kostadinov Kucarov, Vasil Gospodinov Rajnov, Meri Petrova Lakova, Vladko Danailov Murdarov, and Svetla Peneva Koeva, eds. *Ezikovedski izsledvanija v čest na 75-godišnjinata ot roždenieto na st.n.s. i st. d.f.n. Jordan Penčev Penčev*. Plovdiv: Artgraf, 244–54.
- Krapova, Iliyana and Guglielmo Cinque. (2005) "On the order of *wh*-phrases in Bulgarian multiple *wh*-fronting". *University of Venice working papers in linguistics* 15: 171–97.
- . (2008) "On the order of *wh*-phrases in Bulgarian multiple *wh*-fronting". Gerhild Zybatow, Luka Szucsich, Uwe Junghanns, and Roland Meyer, eds. *Formal description of Slavic languages: The fifth conference, Leipzig 2003*. Frankfurt am Main: Lang, 318–36. [*Linguistik international*, 20.]
- Lahne, Antje. (2012) "Specificity-driven syntactic derivation". Myriam Uribe-Etxebarria and Vidal Valmala, eds. *Ways of structure building*. Ox-



- ford: Oxford University Press, 271–96. [*Oxford studies in theoretical linguistics*, 40.]
- Lambova, Mariana. (2000) “The typology of multiple Wh-fronting in Slavic revisited”. Tracy Holloway King and Irina A. Sekerina, eds. *Formal approaches to Slavic linguistics: The Philadelphia meeting, 1999*. Ann Arbor, MI: Michigan Slavic Publications, 238–58. [*Michigan Slavic materials*, 45.]
- . (2001) “On A-bar movements in Bulgarian and their interaction”. *The linguistic review* 18(4): 327–74.
- . (2002) “On the representation of topic and focus in Bulgarian”. Jindřich Toman, ed. *Formal approaches to Slavic linguistics: The second Ann Arbor meeting, 2001*. Ann Arbor, MI: Michigan Slavic Publications, 119–42. [*Michigan Slavic materials*, 47.]
- . (2003) “When is splitting the *wh*-cluster in Bulgarian possible?”. Wayles Browne, Ji-Yung Kim, Barbara H. Partee, and Robert A. Rothstein, eds. *Formal approaches to Slavic linguistics: The Amherst meeting, 2002*. Ann Arbor, MI: Michigan Slavic Publications, 315–34. [*Michigan Slavic materials*, 48.]
- . (2004) *On information structure and clausal architecture: Evidence from Bulgarian*. Ph.D. dissertation, University of Connecticut.
- . (2008) “Multiple fronting in Bulgarian: Clustering and separability”. Gerhild Zybatow, Luka Szucsich, Uwe Junghanns, and Roland Meyer, eds. *Formal description of Slavic languages: The fifth conference, Leipzig 2003*. Frankfurt am Main: Lang, 337–52. [*Linguistik international*, 20.]
- Liakin, Denis. (2007) “Les questions multiples : Le débat continue”. *La revue canadienne de linguistique* 52(3): 279–312.
- Lotfi, Ahmad R. (2003) “Persian *wh*-riddles”. Cedric Boeckx and Kleantes K. Grohmann, eds. *Multiple wh-fronting*. Amsterdam: Benjamins, 161–86. [*Linguistik aktuell*, 64.]
- McKinney-Bock, Katherine S. (2013) *Building phrase structure from items and contexts*. Ph.D. dissertation, University of Southern California.
- Mišmaš, Petra. (2015) *On the optionality of wh-fronting in a multiple wh-fronting language*. Ph.D. dissertation, University of Nova Gorica.
- Mizuguchi, Manabu. (2014) “Free merge and superiority effects on *wh*-movement”. Sarah Lundquist, ed. *Proceedings of WIGL 2014*. Madison: Department of Linguistics, University of Wisconsin, 16–30. [*LSO working papers in linguistics*, 9.]
- Moro, Andrea. (2011) “Clause structure folding and the ‘Wh-in-situ effect’”. *Linguistic inquiry* 42(3): 389–411.
- Mulders, Iris. (1997) “Mirrored specifiers”. Jane A. Coerts and Helen de Hoop, eds. *Linguistics in the Netherlands 1997*. Amsterdam: Benjamins, 135–46. [*Linguistics in the Netherlands*, 14.]
- Pesetsky, David. (2000) *Phrasal movement and its kin*. Cambridge, MA: MIT Press. [*Linguistic inquiry monographs*, 37.]

- Pesetsky, David and Esther Torrego. (2001) "T-to-C movement: Causes and consequences". Michael Kenstowicz, ed. *Ken Hale: A life in language*. Cambridge, MA: MIT Press, 355–426. [*Current studies in linguistics*, 36.]
- Reglero, Lara. (2003) "Non-*wh*-fronting in Basque". Cedric Boeckx and Kleantes K. Grohmann, eds. *Multiple wh-fronting*. Amsterdam: Benjamins, 187–227. [*Linguistik aktuell*, 64.]
- Richards, Norvin W. (1997) *What moves where when in which language?* Ph.D. dissertation, Massachusetts Institute of Technology.
- . (2001) *Movement in language: Interaction and architectures*. Oxford: Oxford University Press.
- . (2010) *Uttering trees*. Cambridge, MA: MIT Press. [*Linguistic inquiry monographs*, 56.]
- . (2014) "A-bar movement". Andrew Carnie, Yosuke Sato, and Daniel Siddiqi, eds. *The Routledge handbook of syntax*. London: Routledge, 167–91.
- Rojina, Nina. (2011) *The syntactic structures of Russian wh-questions*. Ph.D. dissertation, Université de Genève.
- Rudin, Catherine. (1986/2013) *Aspects of Bulgarian syntax: Complementizers and WH constructions*. 2nd ed. Bloomington, IN: Slavica Publishers.
- . (1988a) "Multiple questions in South Slavic, West Slavic, and Romanian". *The Slavic and East European journal* 32(1): 1–24.
- . (1988b) "On multiple questions and multiple Wh fronting". *Natural language and linguistic theory* 6(4): 445–501.
- Sabel, Joachim. (2001) "Deriving multiple head and phrasal movement: The cluster hypothesis". *Linguistic inquiry* 32(3): 532–47.
- Sherkina-Lieber, Marina. (2004) "Focus fronting in *wh*-questions in Inuktitut". *Toronto working papers in linguistics* 23(1): 119–32.
- Wachowicz, Krystyna A. (1974a) "Against the universality of a single Wh-question movement". *Foundations of language* 11(2): 155–66.
- . (1974b) *On the syntax and semantics of multiple questions*. Ph.D. dissertation, University of Texas, Austin.

# On Affixal Articles: An Argument from Bantu for Bulgarian, Romanian, and Icelandic\*

Željko Bošković

*Abstract:* The paper argues against the N-to-D movement analysis of article affixation in the N-D word order in Bulgarian, Icelandic, and Romanian based on these languages not displaying a locality effect that is attributed to N-to-D movement in Bantu languages.

## 1. Introduction

Affixal articles in languages like Bulgarian, Romanian, and Icelandic have attracted a considerable amount of attention. While the elements in question show a number of rather interesting and hotly debated properties, the main debate has focused on the issue of how article placement is accomplished in constructions like (1).<sup>1</sup>

---

\* It is a great pleasure and a privilege to be able to dedicate this paper to Catherine Rudin for her many invaluable and lasting contributions to the field of Slavic linguistics. For helpful comments on the paper, I thank Steven Franks and an anonymous reviewer.

<sup>1</sup> To mention just some of those (less known) additional issues here (for a more general discussion and references regarding Bulgarian, see Franks and King 2000), Bošković (2008b) shows that affixal article languages do not show *wh*-island effects. Rudin (1988) actually notes that Bulgarian and Romanian do not show *wh*-island effects, attributing this to the availability of multiple *wh*-fronting in these languages. However, Bošković (2008b) shows that the *wh*-island effect is voided in a number of affixal article languages which do not have multiple *wh*-fronting; in particular, the *wh*-island effect is voided in Bulgarian, Romanian, Icelandic, Swedish, Albanian, and Hebrew (there are actually contexts where these languages do show *wh*-island effects; importantly, these contexts are the same for all the languages in question). All these languages have affixal articles, but only the first two have multiple *wh*-fronting.

Despić (2011, 2015) and Marelj (2008, 2011) examine the possibility of reflexive possessive anaphors in the nominal domain being bound outside of the nominal domain (i.e., constructions like \**John sold himself's book*). They show that such anaphors are possible in languages without articles, i.e., NP languages in Bošković's (2008a, 2012) typology, but not in languages with articles, i.e., DP languages. Furthermore, Despić

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin*. Bloomington, IN: Slavica Publishers, 2018, 37–52.

- (1) topka-ta  
ball<sub>DEF</sub>

There are two main lines of research here: (i) the postnominal article placement in (1) is accomplished in the syntax through N-to-D movement; (ii) the postnominal article placement in (1) is accomplished in PF through a process akin to Chomsky's (1957) affix hopping (Morphological Merger and Prosodic Inversion fall within this line of research, for ease of exposition I use the term Morphological Merger for this type of analysis): in the syntax, the article is in D and *topka* is in a lower position, with the article placed following the noun in PF so that its prosodic property, namely the suffix requirement, can be satisfied.

There are many works on the issue in question. While this paper will address the issue, its scope will be rather limited. I will not address already existing analyses and arguments, or even discuss the full paradigm pertaining to article placement in languages like Bulgarian;<sup>2</sup> rather, I will simply point out that a property of Bantu languages has relevance for the N-to-D movement vs. Morphological Merger debate regarding structures like (1) (though it

---

shows that affixal article languages behave like NP languages in this respect, proposing a phase-based analysis where the affixal status of D affects the phasehood of DP. It should be noted here that LaTerza 2016 claims that affixal article languages behave like NP languages regarding the ability of possessors to bind out of their nominal domain; however, Franks (this volume) shows that the claim is actually factually incorrect—affixal article languages do not differ from other article languages in this particular respect. Also worth noting is Talić (2017), who argues that affixal article languages actually represent a distinct type, different from both languages with articles and languages without articles (from Bošković's typology).

The point of the discussion in this footnote is to note that affixal articles are a rather complex phenomenon, which affects many properties (this is not that surprising in light of Bošković 2008a, 2012, where it is shown that articles (i.e., the presence vs lack of articles) have wide ranging effects both syntactically and semantically even with respect to phenomena that at least superficially seem to have nothing to do with the nominal domain). This paper will not attempt to address the complexity of the phenomenon in question: as noted below, I confine the discussion here to one particular point, namely whether constructions like (1) involve N-to-D movement.

<sup>2</sup> For example, I do not discuss the possibility of A-D order, which (where allowed) is handled in the same way as the N-D order under Morphological Merger analyses, and in terms of A/AP-movement under movement analyses. At any rate, there are numerous works on the N-D order in the languages under consideration, see for example Halpern 1995, Tomić 1996, Dimitrova-Vulchanova and Giusti 1999, Franks and King 2000, Embick and Noyer 2001, Franks 2001, Julien 2005, Dost and Gribanova 2006, Koev 2011, Talić 2017, and, for N-to-D movement analyses, Dobrovie-Sorin 1987, Cornilescu 1992, Delsing 1993, Sigurðsson 1993, Fowler and Franks 1994, Grosu 1994, Giusti 1995, Arnaudova 1996, Longobardi 1996, Ungureanu 2006, Lohrmann 2010, 2011, and Harðarson 2017, among many others.

should be noted that the focus of the discussion will actually be on whether (1) involves N-to-D movement).<sup>3</sup>

What is relevant here is that Bantu languages quite clearly have N-to-D movement, in fact in all constructions: Bantu traditional Noun Phrases (TNP) are N-initial, which is typically attributed to N-to-D movement (see for example Carstens 2010; note that the term TNP is used neutrally here, simply to refer to the nominal domain, more precisely, the highest projection in the nominal domain, without commitment to its categorial status). Taking for granted that Bantu has N-to-D movement, I will point out that the lack of a parallelism between Bantu and affixal article languages like Bulgarian, Romanian, and Icelandic with respect to a particular phenomenon where N-to-D movement has been claimed to be crucially involved argues against the N-to-D movement analysis of affixal article languages (from now on, I will refer to Bulgarian as representative of this language group).

## 2. On the Complex NP Constraint and Article Affixation

Of interest to us here is the Complex NP Constraint, given in (2), where a complex NP is a noun modified by a clause.<sup>4</sup>

- (2) The Complex NP Constraint (CNPC): Extraction from complex NPs is disallowed.

The effect of (2) is illustrated by (3).<sup>5</sup>

- (3) \*How<sub>i</sub> did you hear [<sub>NP</sub> rumors [<sub>CP</sub> that [<sub>IP</sub> John bought a house *t<sub>i</sub>*]]]?

Bošković (2015) shows that the effect in question is much more general. Extraction is banned not only from clausal complements of nouns but, in fact, from all complements of nouns (i.e., it is banned from PP, DP, and NP, as well

<sup>3</sup> Additional possibilities that are consistent with the relevant Bantu data will also be briefly noted below.

<sup>4</sup> My focus here will be on traditional clausal complements, not relative clauses. Extraction from relative clauses is banned independently of (1) because these are adjuncts, extraction from adjuncts being disallowed.

<sup>5</sup> The effect also arises with argument extraction, though it is slightly weaker in this case reflecting the well-known (but ill-understood) argument/adjunct asymmetry in the strength of the violation with extraction out of islands:

- (i) ?\*What<sub>i</sub> did you hear [<sub>NP</sub> rumors [<sub>CP</sub> that [<sub>IP</sub> John bought *t<sub>i</sub>*]]]?

Following the standard practice for works that do not specifically deal with the argument/adjunct difference in question, in what follows I will abstract away from this difference and simply mark all degraded extractions out of islands with \*.

as clausal complements of nouns).<sup>6</sup> Furthermore, Bošković (2015) shows that this effect actually also holds for AP, PP, and ergative VP (all of which are projections of lexical heads): extraction is also banned from the complements of adjectives, prepositions, and ergative verbs. The only exception to the general ban on extraction out of complements of lexical heads (the Complex XP Constraint) concerns transitive, non-ergative VP.<sup>7</sup> Bošković (2015) also proposes a deduction of the Complex XP Constraint based on an approach to successive-cyclic movement, which quite generally makes successive-cyclic movement more difficult, while still allowing it to take place in the case where it is allowed, namely, with complements of non-ergative verbs, as in (4).

- (4) How<sub>i</sub> did you [<sub>VP</sub> think [<sub>CP</sub> that [<sub>IP</sub> John bought a house *t<sub>i</sub>*]]]?

An alternative account of the Complex XP Constraint is presented in Bošković (2016), the gist of both accounts being that extraction is banned from a double phase configuration, where a phasal head takes a phase as its complement. This is stated in (5).

<sup>6</sup> One relevant case from Greek regarding DP complements of nouns is given in (i) and (ii). Example (i) involves extraction of a genitive DP complement of a noun, which is acceptable, while (ii), which is unacceptable, involves extraction out of the genitive DP complement of the noun.

- (i) Tu vivliu<sub>i</sub> mu ipes pos dhiavases tin [kritiki *t<sub>i</sub>*]  
 the<sub>GEN</sub> book<sub>GEN</sub> me said<sub>2S</sub> that read<sub>2S</sub> the review  
 ‘You told me you read the review of the book.’ (Horrocks and Stavrou 1987)
- (ii) \*Tu vivliu<sub>i</sub> mu ipes pos dhiavases tin [<sub>NP</sub> enstasi [tis kritikis *t<sub>i</sub>*]]  
 the<sub>GEN</sub> book<sub>GEN</sub> me said<sub>2S</sub> that read<sub>2S</sub> the objection the<sub>GEN</sub> review<sub>GEN</sub>  
 ‘You told me you read the objection to the review of the book.’ (Bošković 2015)

<sup>7</sup> The impossibility of extraction out of complements of adjectives, prepositions, and ergative verbs is illustrated in (i), (ii), and (iii), respectively, for CP complements of these heads (the preposition case is illustrated with Spanish, since prepositions in English do not take finite CP complements).

- (i) \*How<sub>i</sub>/Why<sub>i</sub> are you [<sub>AP</sub> proud [<sub>CP</sub> that John hired Mary *t<sub>i</sub>*]]?
- (ii) \*¿Cómo<sub>i</sub> se acordó [<sub>PP</sub> de [<sub>CP</sub> que [Pedro preparaba  
 how clitic remembered<sub>3SG</sub> of that Pedro prepared<sub>IMPERFECT</sub>  
 la comida *t<sub>i</sub>*]]]?  
 the food  
 [Intended] ‘How did she remember that Pedro used to cook food?’
- (iii) a. \*How<sub>i</sub> did it depress Mary [that John was fired *t<sub>i</sub>*]?  
 b. \*How<sub>i</sub> does it bother Bill [that John fixed the car *t<sub>i</sub>*]?

- (5) The Phase-over-Phase Constraint: Extraction is banned from phases that function as complements of phasal heads (i.e., the double-phase configuration in (6)).
- (6) [XP=Phase [YP=Phase]]

Phases are taken to define locality domains for syntactic movement, the crucial mechanism here being the Phase-Impenetrability Condition (PIC), which requires movement to proceed via phasal edges. While Chomsky (2000) assumes that a particular phrase is a phase or not regardless of its syntactic context (e.g., CP is always a phase and IP is never a phase), many authors have argued for various contextual approaches to phasehood. In these, the phasal status of  $\alpha$  depends on the syntactic context in which it occurs (this follows the spirit of Chomsky's 1986 *Barriers*, where we cannot determine whether CP is a barrier or not without knowing its syntactic context—CP can be a barrier or not, depending on its structural position). In particular, Bošković (2015, 2016) argues for a contextual approach to phasehood in which structure is divided into two domains, thematic and non-thematic (i.e., functional), where the highest phrase in each of these domains functions as a phase. In other words, the highest phrase in the thematic domain and the highest phrase in the functional domain count as phases. As a result, the NP, as the highest phrase in the thematic domain of the Noun, and the CP of its complement, as the highest phrase in the functional domain, count as phases in (3). This means that (3) involves a double-phase configuration, as shown in (7), where phases are given in bold.

- (7) \*How<sub>i</sub> did you hear [**NP** rumors [**CP** that [**IP** John bought a house *t<sub>i</sub>*]]]?

Given the PIC, which requires movement to proceed via phasal edges, movement has to proceed successive-cyclically through the edge of the CP and the NP in (7), which Bošković (2015, 2016) shows results in a violation. I focus here on the account presented in Bošković (2015). This account adopts antilocality, which bans movement steps that are too short (see Bošković 1994, 1997, Saito and Murasugi 1999, Abels 2003, Grohmann 2003, among many others), defining antilocality within the labeling system of Chomsky (2013).

In this system, labeling is not forced as part of the Merge operation; hence unlabeled objects are allowed during the derivation, with labels provided at the point when a phasal level is reached through a labeling algorithm (LA). According to the LA, when a head and a phrase merge, the head projects (i.e., it provides the label for the resulting object). There are two ways to label when two phrases merge, via feature sharing or traces, traces being ignored for the purpose of labeling. To illustrate the former case (which is similar to traditional Spec-head agreement), when *what* merges with the *wh*-CP in *I wonder*

[*what*<sub>i</sub> [C [*he said t*]]] (the sister of *what* is a CP at the point of this merger), both *what* and the CP have the interrogative Q-feature, which determines the label.

It is the latter case, however, which is important for our purposes. Chomsky assumes that successive-cyclic movement does not involve feature sharing, essentially following Bošković (1997, 2002, 2007, 2008c). There is then no feature sharing between *that* and the *wh*-phrase which passes through its edge in (8). As a result, the embedded clause cannot be labeled when *what* moves to its edge (indicated with ? in (9)). When *v* is merged into the structure, *what* undergoes movement. Since the element merged with the *that*-CP is now a trace, ? is labeled as CP after the movement of *what*.

(8) *What*<sub>i</sub> do you think [*t*'<sub>i</sub> *that* [*he bought t*<sub>i</sub>]]

(9) *v* [<sub>VP</sub> think [<sub>?</sub> *what* [<sub>CP</sub> *that* [*he bought t*<sub>i</sub>]]]]

Bošković (2015) shows that given this approach to labeling, the Complex NP Constraint (and the Complex XP Constraint and (5) more generally) follows from antilocality, which Bošković (2015) states as a requirement that movement must cross a labeled projection. As noted above, movement in (7) has to proceed successive-cyclically through the edges of the CP and the NP, the CP and the NP being phases. As is always the case with successive-cyclic movement in the labeling framework, movement to the edge of the CP does not involve agreement, which means that the object created by the merger of *how* and the CP in question is not labeled at that point. This is shown in (10a). *N* is then merged into the structure and the *wh*-phrase moves to the edge of the NP. Notice now that the movement in question does not cross a labeled category, hence it violates antilocality. (It only crosses ? in (10b); it does not cross NP, since the movement involves merger with this NP).<sup>8</sup>

<sup>8</sup> Given the LA, where traces are ignored for the purpose of labeling, the complement of the noun will be later labeled as CP, when the next phasal level, NP, is completed, but this is too late for our purposes: at the point of movement the element in question is unlabeled.

Notice that the antilocality problem does not arise in (4) due to the presence of an additional projection in the thematic domain of the verb, namely *vP*, which is the projection where the external theta-role is assigned. The relevant structure of (4) is given in (i).

(i) ... *how*<sub>i</sub> [<sub>vP</sub> [<sub>VP</sub> think [<sub>?</sub> *how* [<sub>CP</sub> *that* [<sub>IP</sub> *John bought a house t*<sub>i</sub>]]]]?

As a result, there are two thematic projections in the thematic domain of the verb: VP and *vP*. Since the highest projection in the thematic domain is a phase, *vP* is a phase, but VP is not. *Wh*-movement here then proceeds from the edge of the CP to the edge of the *vP*. This movement crosses a labeled projection, namely VP, so that antilocality is not violated. Notice also that in the ergative VP example (iii) from footnote 7, *vP* is likely present, as indicated by V-movement (the verb precedes both complements so



- (10) a. [<sub>?</sub>how<sub>i</sub> [<sub>CP</sub> that [<sub>IP</sub> John bought a house  $t_i$ ]]]  
 b. how<sub>i</sub> [<sub>NP</sub> rumors [<sub>?</sub>how [<sub>CP</sub> that [<sub>IP</sub> John bought a house  $t_i$ ]]]]

The details of the account are actually not important for our purposes, the reader should simply bear in mind that we are dealing with a phasehood effect. What is important is a particular proposal concerning voiding of phasehood, which voids phasal locality effects in a number of configurations, including the one discussed here.

Consider the configuration in (11), where X and Y are phasal heads:

- (11) [<sub>XP</sub> Y<sub>i</sub>+X [<sub>YP</sub>  $t_i$ ]]

Bošković (2015) presents a number of constructions where in the case of a complex phase, i.e., a phasal projection that is headed by two phasal heads due to head-movement of the lower phasal head to the higher phasal head, the two phases are collapsed into one, with the lower phase losing its phasehood. This situation is abstractly represented in (11). Since the head of phase YP moves to a phasal head, X, YP ceases to be a phase, which means that phrasal movement out of YP need not proceed through the edge of YP.

Bošković (2015) gives a number of cases from a wide variety of languages that instantiate this phase collapsing effect.<sup>9</sup> Importantly, one such case involves the Complex NP Constraint. Although islandhood in general displays a good amount of crosslinguistic variation, the Complex NP Constraint is one island that is rather resistant to crosslinguistic variation. Bošković (2015)

the examples should involve V-movement). However, since *vP* here is not a thematic projection (no theta-role is assigned in *SpecvP*), VP rather than *vP* is a phase in (ii), in contrast to (i). This leads to an antilocality violation, since movement from the CP edge to the VP edge does not cross a labeled projection, as can be seen in (ii) (with V-movement and irrelevant structural details ignored):

- (ii) ... how<sub>i</sub> [<sub>VP</sub> depress Mary [<sub>?</sub>how [<sub>CP</sub> that [<sub>IP</sub> John was fired  $t_i$ ]]]]?

<sup>9</sup> One relevant case involves article incorporation in Galician, which involves movement of the definite article to the *v+V* complex (see Uriagereka 1988, 1996; Bošković 2013, 2015, 2017). Article incorporation in Galician quite generally voids the islandhood of the DP from which the article incorporates (including subject, adjunct, and conjunct islands), which Bošković (2015, 2017) analyzes in terms of phase collapsing, with article incorporation voiding the phasehood of the relevant DPs (see footnote 13 for illustration). It should be noted that phase collapsing crucially differs from phase sliding/extension, proposed in Gallego and Uriagereka (2007) and den Dikken (2007), where any movement of phasal head Y voids the phasehood of YP. Bošković (2015) argues that the effect under consideration occurs only if Y moves to another phasal head. This will be important below, since under phase collapsing, phasehood of NP will be voided *only* if N moves all the way to D (DP being a phase as the highest functional projection in the nominal domain).

shows, however, that there is a group of languages that resist islandhood in this case, i.e., which do not show the Complex NP Constraint effect. These are Bantu languages, as illustrated by the Setswana example in (12).<sup>10</sup>

- (12) Ke m-ang yo o utlw-ile-ng ma-gatwe a gore  
 it C1-who C1Rel 2sgSM hear-Perf-Rel C6-rumor C6SM that  
 ntša e lom-ile?  
 C9-dog C9SM bite-Perf  
 ‘Who did you hear rumors that a dog bit?’

As in other Bantu languages, in Setswana the noun always precedes *all* other NP-elements, a fact which is analyzed in terms of N-to-D movement (see Carstens 2010 on the N-to-D analysis of the N-initial word order in Bantu). Bošković (2015) argues that this is exactly what is responsible for the lack of the Complex NP Constraint effect in Setswana. The exceptional behavior of Setswana with respect to the Complex NP Constraint in fact follows rather straightforwardly under phase collapsing, given that Setswana has N-to-D movement, as indicated by the N-initial nature of DPs in Setswana. As a result of N-to-D movement, the object DP in (12) is a complex phasal domain, headed by two phasal heads, D and N. Since we are dealing with one phase, the NP is not a phase here. In other words, N-to-D movement voids the phasehood of NP. This means that movement need not proceed through the edge of the NP (see the structure in (13), where the relevant traces are given as copies), which makes movement out of the CP complement of N possible. Since the first phasal head above the embedded CP in (12) is D, the *wh*-phrase will move from the edge of CP to the edge of DP. The movement in question does cross a labeled projection, namely NP, so there is no antilocality violation here, in contrast to (7)/(10).<sup>11</sup>

<sup>10</sup> Bulu and Swahili, also Bantu languages, also do not show CNPC effects. Note that (12) involves argument extraction because adjuncts do not undergo *wh*-movement in Setswana (*wh*-movement actually involves clefting in Setswana).

<sup>11</sup> Bošković (2015) also treats traditional reanalysis cases like (i), from (Ross 1967), where the CNPC effect is voided, in terms of phase collapsing:

- (i) the money, which I am making the claim that the company squandered,  
 amounts to \$400,000

These are lexically conditioned and analyzed in terms of reanalysis/complex predicate formation (for *make-the-claim*); see Chomsky (1980), Kayne (1981), Cinque (1990), and Davies and Dubinsky (2003). Bošković (2015) suggests a phase collapsing analysis, involving covert N-to-D-to-*v*+V movement, which creates a complex predicate *make-the-claim* and which voids the phasehood of NP (as in Bantu) and DP (see footnote 13 for the voiding of DP phasehood here).

- (13) ... who<sub>i</sub> [<sub>DP</sub> rumors<sub>j</sub>+D [<sub>NP</sub> rumors<sub>j</sub> [<sub>? who<sub>i</sub> [<sub>CP</sub> that [<sub>IP</sub> a dog bit *t<sub>i</sub>*]]]]]]?</sub>

Again, the details of the account are not important for our purposes. What is important is that N-to-D movement voids the Complex NP Constraint effect, as we can see in Bantu languages like Setswana, which clearly have N-to-D movement in all contexts, as shown by the N-initial status of TNPs.

The above discussion provides us with a tool to test the proposed analyses of (1) in affixal article languages like Bulgarian, Romanian, and Icelandic. If in such languages the N-D order arises as a result of N-to-D movement, the Complex NP Constraint effect should get voided in such languages, just as it does in Bantu. But the data below show that Bulgarian, Romanian, and Icelandic do exhibit the Complex NP Constraint effect, just like English, and in contrast to Bantu.<sup>12</sup>

- (14) \*Kakvo<sub>i</sub> ču slux-a [če Ivan e kupil *t<sub>i</sub>*?  
 what hear<sub>AOR.2SG</sub> rumor<sub>DEF</sub> that Ivan is bought  
 ‘What did you hear the rumor that Ivan bought?’ (Bulgarian)

- (15) \*Ce<sub>i</sub> ai auzit zvon-ul [că Ion a cumpărat *t<sub>i</sub>*?  
 what have<sub>2SG</sub> heard rumor<sub>DEF</sub> that Ion has bought  
 ‘What did you hear the rumor that Ion bought?’ (Romanian)

- (16) \*Hvað<sub>i</sub> heyrðir þú orðróm-inn um [að Jón hefði  
 what heard you rumor<sub>DEF</sub> about that Jón have<sub>SUBJ.PAST</sub>  
 keypt *t<sub>i</sub>*?  
 bought  
 ‘What did you hear the rumor that Jón bought?’ (Icelandic)

This then argues against the N-to-D account of the post-nominal placement of the article in the languages in question. On the other hand, the sensitivity of Bulgarian, Romanian, and Icelandic to the CNPC is fully consistent with the Morphological Merger account of article placement.

It should, however, be noted that the CNPC data do not necessarily argue for this account, they merely argue against the N-to-D account, the point being the lack of an expected parallelism with Bantu, which would be expected if Bulgarian, Romanian, and Icelandic were to involve N-to-D movement in the derivation of the N-D word order. One can in fact think of different accounts of the order in question that would not involve Morphological Merger. For example, it is possible that affixal articles are located in a projection lower

<sup>12</sup> The Bulgarian, Romanian, and Icelandic data below were provided by Vesela Simonova, Vanessa Petroj, and Gísli Harðarson, respectively.

than DP (see for example Julien 2005), which might not be surprising in light of the mixed behavior of affixal article languages discussed by Talić (2017) and noted in footnote 1. N could still move to the article, but the phasehood of the NP projection would not be voided, since N then would not move to the highest head in the nominal domain. (Recall that only movement to a phasal head, which is the highest head in the nominal domain, voids phasehood.) Another possibility, explored in Koev (2011) and Petroj (2014), is that N undergoes agreement with D, with the article being the morphological realization of this agreement (under this analysis, *-a* on *sluxa* ‘the rumor’ in (14) is not an article in the first place, so the construction does not involve N-to-D). In other words, the lack of a parallelism with Bantu discussed above does not uniquely pick the Morphological Merger analysis, it merely argues against the N-to-D movement analysis of (1).

In summary, languages with N-to-D movement allow extraction out of complex NPs. Affixal article languages like Bulgarian, Romanian, and Icelandic do not allow such extraction, which argues against the N-to-D analysis of the N-D order in these languages.

### 3. The Definiteness Effect

It should, however, be noted that there is an interfering factor not yet considered. It is conceivable that (14–16) are ruled out independently of N-to-D movement due to a definiteness effect (i.e., the ban on extraction from definite NPs; cf., e.g., Chomsky 1986 or Fiengo and Higginbotham 1980), since (14–16) also involve extraction out of definite NPs (a necessity since the article in question is definite). This is a factor that is difficult to control for, since the definiteness effect is often relaxed, and its relaxation is subject to crosslinguistic variation (in addition to ill-understood contextual factors). Thus, Spanish is more permissive regarding extraction out of definite DPs than English; see Ticio (2003, 2005). One relevant example from Spanish is given in (17); see also Greek (i) in footnote 6:<sup>13</sup>

<sup>13</sup> However, even (17) becomes unacceptable if the article is replaced by a demonstrative or if the PP that is extracted is interpreted as an agent or a possessor (see Ticio 2003, 2005 and references therein). It should be noted that the definiteness effect can be voided under head movement/phase collapsing. However, what is needed for this is for D to undergo movement, as can be seen in Galician (i). As noted in footnote 9, Galician has article-to-*v*+V incorporation, which voids phasehood/locality violations under phase collapsing. Thus, like English, Galician disallows extraction from definite DPs, as shown by (ia). However, the violation is voided when the head of the DP incorporates into the verb, as (ib) shows.

- (i) a. \*E de quén<sub>i</sub> viche [DP o retrato t<sub>i</sub>]  
and of who saw<sub>2SG</sub> the portrait

- (17) ¿De qué cantante<sub>i</sub> salieron publicadas [las fotos t<sub>i</sub>]?  
 of which singer were published the photos  
 (Ticio 2005: 238)

More relevant for our purposes is that Bulgarian allows extraction out of definite NPs in examples like (18), which makes it less likely that (14) is ruled out due to a definiteness effect.<sup>14</sup>

- (18) a. Na koja disertacija<sub>i</sub> pročete [komentari-te t<sub>i</sub>]?  
 on which dissertation read<sub>AOR.2SG</sub> comments<sub>DEF</sub>  
 ‘On which dissertation did you read the comments?’  
 b. Na kogo<sub>i</sub> vidja [sestra-ta t<sub>i</sub>]?  
 of who saw<sub>AOR.2SG</sub> sister<sub>DEF</sub>  
 ‘Whose sister did you see?’

For discussion of the definiteness effect in Icelandic, the reader is referred to Harðarson (2017). What is important for our purposes is that examples involving a definiteness effect violation, such as (19a), are quite clearly worse than CNPC violations like (16) in Icelandic, as the following data show.<sup>15</sup>

- (19) a. \*Um hvaða bók<sub>i</sub> last þú [gagnrýni-na t<sub>i</sub>]?  
 about what book read you review<sub>DEF</sub>  
 ‘Of which book did you read the review?’

- (i) b. E de quén<sub>j</sub> viche-lo<sub>i</sub> [DP [D' t<sub>i</sub> [NP retrato t<sub>j</sub>]]]?  
 and of whom saw<sub>2SG</sub>-the portrait  
 ‘So, who have you seen the portrait of?’ (Uriagereka 1988)

Bošković (2015) implements the definiteness effect by assuming that a definite D cannot work as an attractor. As a result, movement of the *wh*-phrase to SpecDP is not possible. Since DP is a phase, which requires movement through SpecDP, (ia) is ruled out. Regarding (ib), D moves to the complex *v+V* head, which is a phasal head. This voids the phasehood of DP, rendering movement through the edge of DP unnecessary in (ib). It is worth noting here that, as observed in Bošković (2015), Galician still shows the Complex NP Constraint effect. As discussed in the text, to void the Complex NP Constraint effect, N-to-D movement is needed, D-to-*v* does not suffice here since such movement does not affect the phasehood of NP.

<sup>14</sup> Romanian disallows this kind of extraction regardless of the definiteness effect:

- (i) \*De care carte<sub>i</sub> ai citit recenzia/ o recenzie t<sub>i</sub>?  
 of which book have<sub>2SG</sub> read review<sub>DEF</sub>/ a review  
 ‘Of which book did you read the/a review?’

<sup>15</sup> The judgment for (16) is adjusted in (19c) to allow for a comparison of the constructions in question, which was not at issue in the above discussion.

- (19) b. <sup>?</sup>Um hvaða bók<sub>i</sub> last þú [gagnrýni t<sub>i</sub>]  
 about what book read you review  
 ‘Of which book did you read a review?’
- c. <sup>??/\*?</sup>Hvað<sub>i</sub> heyrðir þú orðróm-inn um [að Jón  
 what heard you rumor<sub>DEF</sub> about that Jón  
 hefði keypt t<sub>i</sub>]  
 have<sub>SUBJ.PAST</sub> bought  
 ‘What did you hear the rumor that Jón bought?’

The difference in the grammaticality status of (19a) and (19c) indicates that they do not involve the same violation, i.e., that (19c) is not ruled out due to a definiteness effect; if it were (19c) would be expected to have the same status as (19a).

It should also be noted that Ormazabal (1991) observes that even where the definiteness effect is observed, it is confined to cases where the extracted element and the definite article “modify” the same NP, which would make it irrelevant to CNPC examples like (14–16), since the extracted element is not base-generated within the NP complement of the article.

#### 4. Conclusion

In conclusion, this paper has shown that a locality effect that is attributed to N-to-D movement in Bantu languages, which clearly have N-to-D movement, does not arise in affixal article languages like Bulgarian, Icelandic, and Romanian. This raises a problem for the N-to-D movement analysis of the N-D order (i.e., article affixation) in Bulgarian, Icelandic, and Romanian.

University of Connecticut  
 zeljko.boskovic@uconn.edu

#### References

- Abels, Klaus. (2003) *Successive cyclicity, antilocality, and adposition stranding*. Ph.D. dissertation, University of Connecticut, Storrs.
- Arnaudova, Olga. (1996) “The Bulgarian noun (adjective) movement to D”. *University of Trondheim Working Papers in Linguistics* 28: 1–28. Dragvoll, Norway: NTNU Linguistics Department.
- Bošković, Željko. (1994) “D-structure, theta-criterion, and movement into the-ta-positions”. *Linguistic analysis* 24: 247–286.
- . (1997) *The syntax of nonfinite complementation: An economy approach*. Cambridge, MA: MIT Press.

- Bošković, Željko. (2002) "A-movement and the EPP". *Syntax* 5: 167–218.
- . (2007) "On the locality and motivation of Move and Agree: An even more minimal theory". *Linguistic inquiry* 38: 589–644.
- . (2008a) "What will you have, DP or NP?" Emily Elfner and Martin Walkow eds. *NELS 37* Amherst: University of Massachusetts, Graduate Linguistic Student Association, 101–14.
- . (2008b) "On the operator freezing effect". *Natural language and linguistic theory* 26: 249–87.
- . (2008c) "On successive cyclic movement and the freezing effect of feature checking". Jutta M. Hartmann, Veronika Hegedűs, and Henk van Riemsdijk, eds. *Sounds of silence: Empty elements in syntax and phonology*. Amsterdam: Elsevier North Holland, 195–233.
- . (2012) "On NPs and clauses". Günther Grewendorf and Thomas Ede Zimmermann, eds. *Discourse and grammar: From sentence types to lexical categories*. Berlin: Mouton de Gruyter, 179–242..
- . (2013) "Traces do not head islands: What can PF deletion rescue?". Yoichi Miyamoto, Daiko Takahashi, Hideki Maki, Masao Ochi, Koji Sugisaki, and Asako Uchibori, eds. *Deep insights, broad perspectives: Essays in honor of Mamoru Saito*. Tokyo: Kaitakusha, 56–93.
- . (2015) "From the complex NP constraint to everything: On deep extractions across categories". *The linguistic review* 32: 603–69.
- . (2016) "What is sent to spell-out is phases, not phasal complements". *Linguistica* 56: 25–56.
- . (2017) "On the coordinate structure constraint and the adjunct condition". Unpublished ms, University of Connecticut.
- Carstens, Vicki. (2010) "Head-movement in Bantu DPs". Paper presented at the North East Linguistic Society 41, University of Pennsylvania, October 2010.
- Chomsky, Noam. (1957) *Syntactic structures*. The Hague: Mouton.
- . (1980) "On binding". *Linguistic inquiry* 11: 1–46.
- . (1986) *Barriers*. Cambridge, MA: MIT Press.
- . 2000. "Minimalist inquiries". Roger Martin, David Michaels, and Juan Uriagereka, eds. *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press, 89–155.
- . 2013. "Problems of projection". *Lingua* 130:33–49.
- Cinque, Guglielmo. (1990) *Types of A'-dependencies*. Cambridge, MA: MIT Press.
- Cornilescu, Alexandra. (1992) "Remarks on the determiner system of Rumanian: The demonstratives *al* and *cel*". *Probus* 4: 189–260.
- Davies, William D. and Stanley William Dubinsky. (2003) "On extraction from NPs". *Natural language and linguistic theory* 21: 1–37.
- Delsing, Lars-Olof. (1993) *The internal structure of Noun Phrases in the Scandinavian languages*. Ph.D. dissertation, University of Lund, Lund.

- Despić, Miloje. (2011) *Syntax in the absence of determiner phrase*. Ph.D. dissertation, University of Connecticut, Storrs.
- . (2015) "Phases, reflexives and definiteness". *Syntax* 18: 201–34.
- den Dikken, Marcel. (2007) "Phase extension: Contours of a theory of the role of head movement in phrasal extraction". *Theoretical linguistics* 33: 1–41.
- Dimitrova-Vulchanova, Mila and Giuliana Giusti. (1999) "Possessors in the Bulgarian DP". Mila Dimitrova-Vulchanova and Lars Hellan, eds. *Topics in South Slavic syntax and semantics*. Amsterdam: John Benjamins, 163–92.
- Dobrovie-Sorin, Carmen. (1987) "A propos de la structure du groupe nominal en Roumain". *Rivista di grammatica generativa* 12: 126–51.
- Dost, Ascander, and Vera Gribova. (2006) "Definiteness marking in the Bulgarian". Donald Baumer, David Montero, and Michael Scanlon, eds. *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadia Proceedings Project, 132–40.
- Embick, David, and Rolf Noyer. (2001) "Movement operations after syntax". *Linguistic inquiry* 32: 555–95.
- Fiengo, Robert, and James Higginbotham. (1980) "Opacity in NP". *Linguistic analysis* 7: 395–421.
- Fowler, George and Steven L. Franks. (1994) "DP in Bulgarian". George Fowler, Henry Cooper, and Jonathan Ludwig, eds. *Indiana Slavic Studies, Proceedings of the 9th Biennial Conference on Balkan and South Slavic Linguistics*. Bloomington, IN: Indiana University Linguistic Club, 69–78.
- Franks, Steven L. (2001) "The internal structure of Slavic Noun Phrases, with special reference to Bulgarian". Adam Przepiórkowski and Piotr Bański, eds. *Generative linguistics in Poland: Syntax and morphosyntax*. Warszawa: Instytut Podstaw Informatyki PAN, 53–69.
- . (this volume) "A Bulgarian solution to the Slavic Q question?" 93–120.
- Franks, Steven L. and Tracy H. King. (2000) *A handbook of Slavic clitics*. Oxford: Oxford University Press.
- Gallego, Ángel J., and Juan Uriagereka. (2007) "Sub-extraction from subjects: A phase theory account". José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez, and María José Cabrera, eds. *Romance linguistics 2006*. Amsterdam: John Benjamins, 149–62.
- Giusti, Giuliana. (1995) "Heads and modifiers among determiners: Evidence from Romanian". Guglielmo Cinque and Giuliana Giusti, eds. *Advances in Roumanian linguistics*. Amsterdam: John Benjamins, 103–25.
- Grohmann, Kleanthes. (2003) *Prolific domains: On the antilocality of movement dependencies*. Amsterdam: John Benjamins.
- Grosu, Alexander. (1994) *Three studies in locality and case*. New York: Routledge.
- Halpern, Aaron Lars. (1995) *On the placement and morphology of clitics*. Stanford, CA: CSLI.



- Harðarson, Gísli Rúnar. (2017) *Cycling through grammar: On compounds, noun phrases and domains*. Ph.D. dissertation, University of Connecticut, Storrs.
- Horrocks, Geoffrey, and Melita Stavrou. (1987) "Bounding theory and Greek syntax". *Journal of linguistics* 23: 79–108.
- Julien, Marit. (2005) *Nominal Phrases from a Scandinavian perspective*. Amsterdam: John Benjamins.
- Kayne, Richard. (1981) "ECP extensions". *Linguistic inquiry* 12: 93–133.
- Koev, Todor. (2011) "Definiteness as agreement: Evidence from Bulgarian". Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer, and Barbara Tomaszewicz, eds. *Proceedings of the 28th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 133–41.
- LaTerza, Ivana. (2016) "Binding in English and South Slavic and the parameterized DP hypothesis". *Linguistic inquiry* 47: 741–53.
- Longobardi, Giuseppe. (1996) "The syntax of N-raising: A minimalist theory". OTS Working Papers, Utrecht University.
- Lohrmann, Susanne. (2010) *The structure of the DP and its reflex in Scandinavian*. Ph.D. dissertation, University of Stuttgart.
- . (2011) "A unified structure for Scandinavian DPs". Petra Sleeman and Harry Perridon, eds. *The noun phrase in Romance and Germanic*. Amsterdam: John Benjamins, 111–26.
- Marelj, Marijana. (2008) "Probing the relation between binding and movement: A left branch extraction and pronoun insertion strategy". Emily Elfner and Martin Walkow, eds. *NELS 37*. Amherst: University of Massachusetts, Graduate Linguistic Student Association, 73–86.
- . (2011) "Bound-variable anaphora and Left Branch Condition". *Syntax* 14: 205–29.
- Ormazabal, Javier. (1991) "Asymmetries on wh-movement and some theoretical consequences". Unpublished ms., University of Connecticut, Storrs.
- Petroj, Vanessa. (2014) "Code-switching within the TNP: Romanian and Serbo-Croatian in interaction". Unpublished ms. University of Connecticut, Storrs.
- Ross, John Robert. (1967) *Constraints on variables in syntax*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Rudin, Catherine. (1988) "On multiple questions and multiple wh-fronting". *Natural language and linguistic theory* 6: 445–501.
- Saito, Mamoru, and Keiko Murasugi. (1999) "Subject predication within IP and DP". Kyle Johnson and Ian Roberts, eds. *Beyond principles and parameters*. Dordrecht: Kluwer, 167–88.
- Sigurðsson, Halldór Ármann. (1993) "The structure of the Icelandic DP". *Studia linguistica* 47: 177–97.
- Talić, Aida. (2017) *From A to N and back: Functional and bare projections in the domain of N and A*. Ph.D. dissertation, University of Connecticut, Storrs.

- Ticio, Emma. (2003) *On the structure of DPs*. Ph.D. dissertation, University of Connecticut, Storrs.
- . (2005) "Locality and anti-locality in Spanish DPs". *Syntax* 8: 229–286.
- Tomić, Olga. (1996) "The Balkan Slavic nominal clitics". Aaron Halpern and Arnold Zwicky, eds. *Approaching second: Second position clitics and related phenomena*. Stanford: CSLI Publications, 511–35.
- Ungureanu, Mona-Luiza. (2006) *DP-internal structure and movement in Romanian*. Ph.D. dissertation, McGill University, Montreal.
- Uriagereka, Juan. (1988) *On government*. Ph.D. dissertation, University of Connecticut, Storrs.
- . (1996) "Determiner clitic placement". Robert Freidin, ed. *Current issues in comparative grammar*. Dordrecht: Kluwer, 257–95.

# Bulgarian Pronouns: What They Don't Distinguish that Most of Slavic Does

Wayles Browne

*Abstract:* The evolution of pronouns in Bulgarian has obscured several distinctions that are still clearly made in most other Slavic languages. This paper points out four such instances, or even five, supporting them with recent examples of usage. One is the difference between interrogative 'who' and 'which', another is the difference between interrogative 'what' and 'what kind of', a third shows that personal pronouns like 'it' can stand for neuter indefinite or neuter demonstrative antecedents where other Slavic languages would need to use 'this' or 'that', a related fourth shows that neuter indefinite or neuter demonstrative antecedents can be followed by the same sort of relative pronouns that other antecedents get, and the fifth is the use of the indefinite pronoun *nešto* as a noun, singular *nešto*, plural *nešta*, meaning 'thing, things'.

## 0. Introduction

In her even now fundamental study *Aspects of Bulgarian syntax: Complementizers and WH constructions* (Rudin 1986), Catherine Rudin mentions in a footnote (p. 124) that, as compared with Serbo-Croatian [and Bosnian] *što* used in two types of relative constructions, "the lack of case marking obscures the distinction between 'declinable' and 'indeclinable' (WH word and complementizer) *što*. Since *što* relatives are marginal in modern standard Bulgarian, I do not consider them any further here." In fact, obscuring of distinctions is a motif that recurs in multiple places in the Bulgarian pronoun system. This paper will point out four such instances, or even five. Each of them is more or less well known, but I am not aware that they have been presented side by side until now.

## 1. Interrogative 'who' and 'which'

Bulgarian has largely ceased to distinguish the interrogative pronoun *koj* 'who' as a pro-noun from the interrogative masc. *koj*, fem. *koja*, neut. *koje*, plural *koi* 'which; which (one)' as a pro-adjective (and likewise the relative pronoun *kojto* from the relative pro-adjective *kojto*, *kojato*, *koeto*, *koito*). The nominative

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin*. Bloomington, IN: Slavica Publishers, 2018, 53–63.

form *koj* ‘who’ was taken over from *koj/koja/koe/koi* centuries ago, replacing the reflex of Common Slavic *\*kъto* seen in almost all other Slavic languages,<sup>1</sup> but the distinction between ‘who’ and ‘which one’ was still assured because *koj* ‘who’ was masculine singular only, and still had the dative form *komu* and the accusative *kogo* going back to the paradigm of *\*kъto*. Both of these inflected forms were kept in the early 20th century (Beaulieux 1933/1950: 87), but *komu* was felt to be archaic (and replaced by *na kogo* ‘to whom’) later in the century (Rudin 1986: 9; Nicolova 1986: 153). More recently *kogo* has been giving way to *koj* (Rudin 1986: 9: interrogative *kogo* and relative *kogoto* “is sometimes replaced by *koj(to)* in colloquial speech”; Nicolova 2017: 272).

Interrogative *koj* ‘who’ was earlier a masculine singular word no matter what gender and number the answer to it might turn out to be (Beaulieux 1933/1950: 88: “En cette valeur pronominale, la forme du masculin singulier *koj?* s’emploie, comme le français *qui ?*, sans acception de genre ni de nombre : *kój me víka? qui m’appelle?*”, that is, “In this pronominal value, the masculine singular form *koj?* is used, like French *qui?*, without regard for gender or number: *kój me víka? ‘who is calling me?’”).*

- (1) *Koj e bil tuk?*  
 ‘Who has been [masc.sg.] here?’

This was seen in its agreement with verbs or other predicates even if many people had been here before us. It has recently come to have the plural form *koi* and all genders of the singular (fem. *koja* and neut. *koe*), thus becoming entirely homophonous with the ‘which one’ interrogative. Nicolova (1986: 154) records this usage but says (Nicolova 2017: 273) “In everyday speech a person can occasionally be denoted by the independent forms *kojā, koè, koi*, when the speaker wants to indicate explicitly that the question refers to a female person, to a child, or to more than one person.” In fact, such usage is highly frequent, as the internet attestations below will show.<sup>2</sup>

<sup>1</sup> The same *koj* for ‘who’ is seen in Macedonian, but this paper will not treat Macedonian due to lack of information about the most recent changes in it.

<sup>2</sup> In this respect modern Bulgarian contrasts with other Slavic languages. As Vrinda Chidambaram and Jozef Müller kindly confirm, “any  $\phi$ -feature agreement besides [masc.sg.] is strictly prohibited in interrogatives in Slovak, even when you in fact know the answer” as in (\*i) vs. (ii) when addressing a rhetorical question to their daughter.

- (i) *\*Kto zjed-la všetku čokoládu?*  
 Who ate [fem.sg.past] all chocolate  
 ‘Who ate all the chocolate?’ (when you know who ate it and it is someone who identifies as female)

When a set of items is presupposed and the items are feminine singular, whether human or not, *koja* might be glossed 'which (one)', as in the titles of two newspaper stories:

- (2) *Koja e naj-težkata tečnost?*  
 'Which [fem.sg.] is the heaviest liquid [fem.sg.]?'  
 (<http://nakratko.bg/category/51/146296/>)<sup>3</sup>
- (3) *Ivana i Melanija v spor koja e pârвата dama.*  
 'Ivana and Melania in disagreement which one [fem.sg.] is the First Lady.'  
 (<http://glasove.com/categories/izgubenata-bylgariya/news/ivana-i-melaniya-v-spor-koya-e-pyrvata-dama-na-sash/>)

But when there is no presupposed set, *koja* is 'who' when one seeks to identify a female person:

- (4) *Koja e Angela Merkel?*  
 'Who is Angela Merkel?'  
 (<https://www.tvevropa.com/2017/09/koya-e-angela-merkel/>)
- (5) *Koj e Kont Androvanti i koja e Mara Gidik?*  
 'Who [masc.sg.] is Kont Androvanti and who [fem.sg.] is Mara Gidik?'  
 (A newspaper story about why one street in Burgas is named Kont Androvanti Street and another is Mara Gidik Street, see  
<http://faragency.bg/bg/koy-e-kont-androvanti-i-koya-e-mara-gidik/>)

Neuter *koe* is also 'which one' for neuter things, as well as 'which one' for things of mixed genders and for ideas (sentences or phrases):

- 
- (ii) *Kto zjed-ol všetku čokoládu?*  
 Who ate [masc.sg.past] all chocolate  
 'Who ate all the chocolate?' (when you know who ate it and it is someone who identifies as female)

In Bosnian-Croatian-Serbian, too, *tko/ko* 'who' demands masculine singular agreement. In the Bosnian poet Saša Skenderija's *Sretniji dani Černobila: kratki porodični roman* (unpublished manuscript, Prague 2018, p. 10) the father figure criticizes the mother by asking her (iii), though both know well that it was she who didn't want to.

- (iii) *Ko nije htio da mu kupi nove adidaske?*  
 'Who didn't want [masc.sg.] to buy him new Adidas tennis shoes?'

<sup>3</sup> Unless otherwise specified, all websites were last accessed 25 March 2018.

- (6) Koe e naj-lošoto vreme za započvane na dieta?  
 ‘Which [neut.sg.] is the worst time [neut.sg.] for beginning a diet?’  
 (<https://dariknews.bg/novini/liubopitno/koe-e-naj-loshoto-vreme-za-zapochvane-na-dieta-video-2063676>)
- (7) Koe e po-važno, kakvo kazvaš ili kak go kazvaš?  
 ‘Which [neut.sg.] is more important, what you say or how you say it?’  
 (<http://www.gnezdoto.net/mydrost/870-koe-e-po-vajno-kakvo-kazvash-ili-kak-go-kazvash>)
- (8) Ajrjan, rakija, čaj, kafe, koe e naj-tursko?  
 ‘Ayran [masc.sg.], raki [fem.sg.], tea [masc.sg.], coffee [neut.sg.], which one [neut.sg.] is the most Turkish?’  
 (<http://www.dw.com/bg/айрян-ракия-чай-кафе-кое-е-най-турско/a-40463340>)

But *koe* is also ‘who’ for persons being referred to by a neuter-gender word, whatever their real gender may be:

- (9) Koe ot tezi bebeta e momiče?  
 ‘Which [neut.sg.] of these babies [shown on a picture] is a girl?’ (both *bebe* and *momiče* are neuter)  
 (<http://www.zajenata.bg/кое-е-от-бебета-е-момиче!-тестът-който-разкривавсичко-за-личността-ви---точен-е-100!-news86762.html>)
- (10) Koe e deteto ot šokoladite Kinder?  
 ‘Who [neut.sg.] is the child [neut.sg.] on the Kinder chocolates?’ (i.e., on their wrapping and ads)  
 (<http://www.danybon.com/komunikacii/pr-i-reklama/dete-shokoladi-kinder/>)
- (11) New girl, koe e tova novo momiče?  
 ‘New Girl [TV series], who [neut.sg.] is that new girl?’ (where *momiče* ‘girl’ is a neuter noun)  
 (<http://serial.spisanie.to/комедия/568-new-girl-кое-е-това-ново-момиче>)
- (12) Razkrixa koe e misterioznoto gadže na Riana.  
 ‘They revealed who [neut.sg.] Rihanna’s mysterious boyfriend is.’  
 (where *gadže* despite his masculine qualities is neuter)  
<https://www.actualno.com> › Лайфстайл › Шоубизнес)

An example for plural *koi* in the sense ‘who, who all’:

- (13) Beše mi interesno da uznaja koi ne sa xaresali moja komentar...  
 'It was interesting for me to learn who [pl.] did not like my  
 comment...'

(<http://www.divino.bg/statii/еносиазъм>)

Finally we give an example explicitly contrasting *koi* 'who [pl.]' and *koj* 'who [sing.]':

- (14) Mnogo xora me pitat koe e normalno i koe ne e, koi sa normalni i koi  
 ne sa i koj opredelja normalnostta kato cjalo.  
 'Many people ask me what is normal and what is not, who [pl.] are  
 normal and who [pl.] are not, and who [sing.] defines normality as a  
 whole.'

(<http://www.forumat-bg.com/politika/2435-meglana-kuneva-razlichnite-ot-men-ne-sa-normalni-da-kazhem-da-na-normalnostta>)

## 2. Interrogative 'what'

Most of the Slavic languages have a reflex of Common Slavic \**čvto* for interrogative 'what'. As mentioned above, *što* is marginal as a relativizer in modern Bulgarian; it is equally marginal as an interrogative. It has almost entirely (Nicolova 1986: 155, and 2017: 274) been replaced by *kakvo*, which in form is the neuter singular of the inherited pro-adjective for qualities, masc. *kakâv*, fem. *kakva*, neut. *kakvo*, plural *kakvi* 'what kind of, like what'. This is already an instance of obscured distinctions. We can use *kakvo* unambiguously to ask about the identity of a masculine, feminine, or plural noun without confusion, e.g.,

- (15) a. Kakvo e životât?  
 'What is life [masc.sg.]?'  
 b. Kakvo e ljubovta?  
 'What is love [fem.sg.]?'  
 c. Kakvo sa oblacite?  
 'What are clouds [plural]?'

But if we are asking about a neuter noun like *ime* 'name', a question with *kakvo* can be ambiguous.

- (16) Kakvo e nejnoto ime?  
 'wh... is her name?'

could be 'What is her name?' or 'What kind is her name (a French name, a Spanish name, etc.)?' A real-life example is the discussion of giving a name to

the dwarf planet now called Eris (Erida, in Bulgarian) in the Wikipedia article “Ерида (планета джудже)”. The question is raised

- (17) *kakvo bi moglo da bâde imeto*  
 ‘wh... could be the name’

and one has to read the entire paragraph to see whether it is ‘what’ or ‘what kind of name’.

Similar to what we saw in section 1, where originally masculine-singular-only *koj* ‘who’ is taking on the other gender and number forms, *kakvo* ‘what’ used as a predicate can agree with its subject in gender and number. This is most frequently seen with copulas (Browne 1999: 203):

- (18) a. *Kakâv ste po profesija?*  
 ‘What [masc.sg.] are you by profession?’  
 for a masculine (in formal address),  
 b. *Kakva ste po narodnost?*  
 ‘What [fem.sg.] are you by nationality?’  
 for a feminine (formal address),  
 c. *Kakvi ste po profesija?*  
 ‘What [pl.] are you by profession?’  
 for a plural subject

But it applies to secondary predicates on verbs as well (Nicolova 1986: 157):

- (19) *Kakâv go naznačixa?*  
 ‘As what [masc.sg.] did they appoint him?’

That is to say, the homonymy between the four gender-number forms of ‘what kind’ and the four gender-number forms of ‘what’ has become complete.

### 3. Antecedents for Personal Pronouns

In most Slavic languages, neuter personal pronouns are choosy about their antecedents. Thus in BCS, the pronoun *ono* nom., *ga/njega* acc., can have an antecedent that is a neuter noun, such as *pismo* ‘a letter’:<sup>4</sup>

<sup>4</sup> Here and below, wherever necessary, antecedents and their anaphoric or relative pronouns are given in **boldface**.



- (20) Daj mi **pismo**. Želim **ga** čitati.  
 'Give me **the letter**. I want to read **it**.'

But it cannot have an antecedent that is a neuter demonstrative pronoun or neuter indefinite pronoun:

- (21) a. Daj mi **ovo**. Želim \***ga** čitati.  
 'Give me **this**. I want to read **it**.'  
 b. Daj mi **nešto**. Želim \***ga** čitati.  
 'Give me **something**. I want to read **it**.'

Rather, one refers to such an antecedent with *to*, the unmarked demonstrative in BCS, and says:

- (22) Daj mi **ovo**. Želim **to** čitati.  
 'Give me **this**. I want to read **this/that** [neut.sg].'

(Browne 1981/1986, chapters 2 and 7). The same constraint holds for antecedents that are infinitives, clauses, or unexpressed ideas:

- (23) a. **Tvoja je sestra došla** iako \***ga** nisam očekivao.  
 'Your sister came although I didn't expect **it**.'  
 b. ... iako **to** nisam očekivao.  
 'although I didn't expect **this/that**.'

As shown in (23b) the pronoun again has to be the demonstrative *to*.

Modern Bulgarian makes no such distinction between types of antecedent. The neuter singular pronoun [no longer a demonstrative] *to*, accusative *nego/go* can refer to all the abovementioned kinds of antecedents. Here is a neuter demonstrative antecedent:

- (24) Daj mi **tova**. Iskam da **go** četa.  
 'Give me **this**. I want to read **it**.'

A neuter indefinite antecedent:

- (25) Daj mi **nešto**. Iskam da **go** četa.  
 'Give me **something**. I want to read **it**.'

And a phrasal or sentential one:

- (26) **Kato cjalo igraxme dosta dobre**, no az lično **go** očkavax.  
 ‘**In general we played pretty well**, but I personally expected it.’  
 (<https://leedsunitedbg.com/forums/?view=thread&id=148&part=3>)

See also the *go* in the last clause of (7) above (referring back to *kakvo* ‘what’ or perhaps to *kakvo kazvaš* ‘what you say’).

We conjecture that the loss of antecedent-sensitivity in Bulgarian is connected with the rise of clitic doubling. At a certain point in its history, Bulgarian began to use clitic doubling as a way to reinforce direct and indirect objects having certain functions in a clause (definite or specific topics and the like). As soon as that happened, every kind of object needed a clitic that would be able to double it. Noun objects and personal-pronoun objects already could be referred to by the clitic forms of personal pronouns, as in BCS (20), so these same clitics could serve to double them, including the 3rd person singular neuter *go*. But the abovementioned demonstratives, indefinites, and phrasal/sentential objects at first did not have a clitic to refer to them. *Go* was pressed into service as a doubling clitic, and then there no longer was any rationale to prohibit it as an anaphoric pronoun referring back to the same group of items. Another factor may well have been the loss of the nominative neuter personal pronoun *ono* and its replacement with *to*, built on the same former demonstrative stem as *toj* ‘he’ and *tja* ‘she’.

#### 4. Antecedents for Relative Pronouns

The same group of items that could not be antecedents for personal pronouns in section 3 also cannot be antecedents for the relative pronoun corresponding to ‘which’ in most of the Slavic languages. Using BCS once again as a point of comparison, *pismo* ‘a letter’ can be the antecedent for a relative clause made with *koji* (neuter *koje*), as in (27):

- (27) **pismo koje** čitam  
 ‘**the letter** [neut.sg.] **which** [neut.sg.] I am reading’

But a demonstrative could not be:

- (28) \***ovo koje** čitam  
 ‘**this** [neut.sg.] **which** [neut.sg.] I am reading’

An indefinite also cannot antecede a *koji/koje* relative:

- (29) \***nešto koje** čitam  
 ‘**something** [neut.sg.] **which** [neut.sg.] I am reading’

Nor can a phrase or clause:

- (30) **Tvoja je sestra došla, \*koje** nisam očekivao.  
'Your sister came, **which** [neut.sg.] I didn't expect.'

Instead one must use the declinable *što* as a relativizer (Browne 1981/1986, chapter 7):

- (31) **Tvoja je sestra došla, što** nisam očekivao.

The same dependency is found, e.g., in Russian and Polish. Modern Bulgarian has no such constraint on choice of relativizers. Neuter *koeto* (= interrogative *koje* plus the relative marker *-to*) can relativize demonstrative antecedents, indefinite antecedents, and phrasal/clausal antecedents:

- (32) a. **tova, koeto** ne očekvax  
'that [neut. sg.] **which** I didn't expect'  
b. **nešto, koeto** ne očekvax  
'something [neut. sg.] **which** I didn't expect'  
c. **ima problemi s filma..., koeto** ne očekvax  
'there are problems [pl.] **with the film** [masc. sg.]..., **which** [neut. sg.] I didn't expect'

(www.segabg.com/index.php?iid=11430&sid=7).

Or, to cite an example from classical literature:

- (33) Baj Gan'ó, predi da počne objada, načeva da se krâsti, ama **xem se krâsti, xem se usmixva**, s **koeto** iska da pokaže na stopanite, če toj ne e ot onezi prostite xorica  
'Bai Ganyo, before starting his lunch, **crosses himself, smiling at the same time**, with **which** [neut. sg.] he means to show to his hosts that he is not one of those simple folk'  
(A. Konstantinov, quoted in Nicolova 2017: 270)

If, as argued in Browne (1981/1986), relative pronouns are derived via personal pronouns, exactly this difference between BCS and Bulgarian is to be expected. Just as Bulgarian has obscured the distinction between possible types of antecedents for personal pronouns, it has also obscured the distinction between possible types of antecedents for relative clauses.

## 5. A Pronoun Can Be a Noun

A curious further instance of a Bulgarian pronoun losing a distinction is the word *nešto*. It began life as an indefinite neuter singular pronoun ‘something’, which it still frequently is, but it also now functions as a neuter noun meaning ‘thing’. We can see its noun-like behavior in the fact that it can take a definite article, which pronouns do not normally do: e.g., *neštoto* ‘the thing’ but not *\*tjata* ‘the she’, *\*kojat* ‘the who’,<sup>5</sup> and particularly in the fact that it forms a plural *neštá* with the shift of stress typical for neuter nouns in *-o* (cf. *sélo* ‘village’, plural *selá*). Further, as a pronoun it takes postmodifiers, e.g., *nešto normalno* ‘something normal’, but as a noun, particularly in the plural, it takes a premodifying adjective: *normalni nešta* ‘normal things’ (though also in the singular: *normalno nešto* ‘a normal thing’). Unlike the English-language occasionalism ‘somethings’, the plural form can be part of a terminological system, as in *Internet na neštata* ‘the Internet of Things’, *položenie na neštata* ‘state of affairs’. So we can say this word has obscured the distinction between indefinite pronoun and lexical noun.

Cornell University  
ewb2@cornell.edu

## References

- Beaulieux, Léon. (1950) *Grammaire de la langue bulgare*. 2nd ed. Paris: Institut d'études slaves.
- Browne, Wayles. (1981) “Relativna rečenica u hrvatskom ili srpskom jeziku u poređenju s engleskom situacijom.” Ph.D. dissertation, University of Zagreb.
- . (1986) *Relative clauses in Serbo-Croatian in comparison with English*. Zagreb: Institute of Linguistics [*Yugoslav Serbo-Croatian-English contrastive project. New studies*, 4.]
- . (1999) “Vsjako načalo e trudno: Traps in lesson I in the South Slavic languages.” Mila Dimitrova-Vulchanova and Lars Hellan, eds. *Topics in South Slavic syntax and semantics*. Amsterdam: John Benjamins.
- “Ерида (планета джудже)”. In the Bulgarian wikipedia: [https://bg.wikipedia.org/wiki/Ерида\\_\(планета\\_джудже\)](https://bg.wikipedia.org/wiki/Ерида_(планета_джудже)) (last accessed 6 September 2018).
- Nicolova, Ruselina. (1986) *Bálgarskite mestoimenija*. Sofija: Nauka i izkustvo.

<sup>5</sup> The musical group that bears this name is called neither *\*Kojat* nor *\*Xuto* in Bulgarian. The Bulgarian Wikipedia lists it as follows: “Ху” (на английски: *The Who*, Дъ Ху) е британска рокгрупа, създадена през 1964 година в Лондон... “**‘Hu’** (in English: *The Who*, **Dâ Hu**) is a British rock group, formed in 1964 in London...” (accessed 6 Sept. 2018).

- Nicolova, Ruselina. (2017) *Bulgarian grammar*. Berlin: Frank & Timme.
- Rudin, Catherine. (1986) *Aspects of Bulgarian syntax: Complementizers and wh constructions*. Columbus, OH: Slavica Publishers.
- Skenderija, Saša. (2018) "Sretniji dani Černobila: Kratki porodični roman". Unpublished ms, Prague.
- "Xy". In the Bulgarian wikipedia: <https://bg.wikipedia.org/wiki/Xy> (last accessed 6 September 2018).

# Clitic-Doubles of Conjoined DP Objects in Macedonian and Bulgarian\*

Vrinda Chidambaram

*Abstract:* It is well known that Bulgarian and Macedonian differ with respect to the constraints and requirements on clitic doubling (CD). In Macedonian, CD of direct objects is strictly limited to definite DPs, whereas the parameters of CD in Bulgarian are more subtle. In this paper, I explore the consequences of pairing two objects, differing in definiteness, within a single conjoined phrase. The result in Macedonian appears to be First Conjunct Agreement (FCA), whereas Bulgarian allows Full Agreement. I weigh competing theories of FCA using the data from Macedonian and Bulgarian.

## 0. Introduction

There is a wealth of literature on the phenomenon of clitic doubling (CD) across languages—see Anagnostopoulou 2006 for a detailed survey—and the broad extent of the work done on the topic accurately reflects its cross-linguistic prevalence. However, while CD surfaces in a wide variety of languages, each language has its own special rules of how and when it occurs. Even languages that are as closely related as Bulgarian (Bg) and Macedonian (Mac) diverge with respect to the contexts giving rise to CD.

In this paper, I consider one particular disparity between Mac and Bg; namely what happens when the clitic double corresponds to an object consisting of conjoined DPs differing in definiteness. Specifically, Mac prohibits a plural clitic double of a conjoined DP if either of the conjuncts is indefinite.

- (1) \*Gi vidov profesor-ot i eden student.  
them saw<sub>1SG</sub> professor<sub>DEF</sub> and a/one student  
'I saw the professor and a student.'

---

\* I offer my deepest gratitude to my infinitely patient consultants: Boban Arsenijević, Biljana Garvanlieva, Anton Koychev, Marko Kostovski, Iliyana Krapova, Roumyana Pancheva, Teodora Radeva-Bork, Vesela Simeonova, Jelena Stojković, Biljana Temova, Žane Temova, Georgia Tsovev, and Ognen Vangelov. I am also immensely grateful to the anonymous reviewers for their suggestions and remarks on a previous version of this paper.

In contrast, some speakers consider the analogous Bulgarian sentence to be grammatical:<sup>1</sup>

- (2) Vidjax gi profesora i edin student.  
 saw<sub>1SG</sub> them the-professor and a/some student  
 'I saw the professor and a/some student.'

In addition to (1) and (2), we find another closely related data pair, in which the Mac sentence (3) is grammatical while the Bg sentence (4) is not.

- (3) Go vidov profesor-ot i eden student.  
 him saw<sub>1SG</sub> professor<sub>DEF</sub> and a/one student  
 'I saw the professor and a student.'
- (4) \*Vidjax go profesora i edin student.  
 saw<sub>1SG</sub> him professor<sub>DEF</sub> and a/one student  
 [Intended] 'I saw the professor and a student.'

The Mac sentence (3) includes a singular CD, which suggests that it is doubling only the first, definite conjunct. This looks quite similar to what we see in cases of First Conjunct Agreement (FCA). FCA occurs in many languages and there have been a number of theoretical proposals to account for it, however, most (if not all) the work done on FCA has centered on verbal, adjectival, and participial (which could be seen as a verb-adjective hybrid category) agreement. The theories largely fall into two categories: structure-based and Agree-based. The structure-based theory suggests that Full Agreement (FA), in which the agreement is with the whole conjoined object, and FCA are the result of differing syntactic structures. The Agree-based theory suggest that FA and FCA are distinguished by differing applications of the Agree operation.

Thus far, no work has been done regarding instances of FCA in CD. So, the question of how the theories of FCA account for first conjunct agreeing clitic doubles has yet to be answered. In this paper, I intend to show how each of the two dominant competing theories of FCA handles the data in (1–4) and

<sup>1</sup> While the judgments for Macedonian are fairly robust (my consultants largely agreed with one another), the judgments for Bulgarian vary widely. Interestingly, my Bulgarian consultants felt very strongly about their own judgments (there was no hesitation or any other indication of uncertainty), but the inter-speaker variation among them was remarkable. The variation does not seem geographically dependent, as speakers from the same areas disagreed, nor does it seem to depend on age, socio-economic, or educational factors. This phenomenon, in and of itself, seems worthy of investigation, but that is best left for another paper.

suggest that the data may be better supported by a structure-based account. The paper is organized as follows: section 1 provides some essential background on the distribution of CD in Mac and Bg and the prevailing theories accounting for it. Section 2 is an outline of the competing theories of FCA and illustrates how they apply to the Mac data. In section 3 I turn to the Bg data and suggest how particular interpretive constraints therein may be more readily explained if we adopt a structure-based account of FA/FCA. Section 4 concludes the paper.

## 1. Background

Clitic-doubling is a superficial violation of the Theta Criterion: it involves multiple instantiations of a single argument (Harizanov 2016). That is, an R-expression or long-form pronominal object of a verb is reinforced by a clitic pronoun; this much is true of CD across languages. However, in the particulars of how and when doubling occurs, there is abundant variation.

Mac and Bg are somewhat similar in terms of the general principles guiding the use of clitic doubles. Some have suggested that their presence is dependent upon the definiteness, specificity, and/or topicality of the object (Leafgren 1997, Rudin 1997, Jaeger 2004, Guentchéva 2008; see also Dimitrova-Vulchanova 1999 and Dimitrova-Vulchanova and Hellan 1999). The Macedonian examples below illustrate the apparent dependence of clitic doubling on the definiteness/specificity of the direct object.

- (5) \*(Go) vidov profesor-ot.  
 him saw<sub>1SG</sub> professor<sub>DEF</sub>  
 'I saw the professor.'
- (6) (\*Go) vidov (eden) profesor.  
 him saw<sub>1SG</sub> (a) professor  
 'I saw a professor.'

Others have suggested that these features of the object, in fact, play no role in licensing a clitic double (Cinque and Krapova 2008), offering a number of examples in which a variety of distinct licensing mechanisms operate but each giving rise to syntactic configurations that superficially look like simple object clitic doubling. For example, they distinguish between Clitic Right Dislocation (CLRD), as discussed also in Arnaudova (2002), and true Clitic Doubling. For CLRD, the Topicality of the associate is the crucial trigger for the clitic, whereas true Clitic Doubling occurs as a reflex of certain predicates. Cinque and Krapova (2008) do, however, suggest the possibility that some of these mechanisms may actually be, on some structural level, the same, al-



though motivated by distinct syntactic properties. Although clitic doubling phenomena can be witnessed in a number of contexts and for a variety of reasons, I focus here on the CLRD type of reduplication, and purely for the sake of simplicity refer to this as Clitic Doubling (CD).

### 1.1. Definiteness and CD in Macedonian and Bulgarian

Unlike Bulgarian, Macedonian generally requires the presence of a clitic to double definite direct and definite or indefinite indirect objects.

- (7) Biljana \*(mu) \*(go) dava podarok-ot na Ognen.  
 Biljana him<sub>DAT</sub> it<sub>ACC</sub> gives gift<sub>DEF</sub> to Ognen  
 ‘Biljana gives the gift to Ognen.’

- (8) Biljana mu (\*go)<sup>2</sup> dava eden podarok na edno momče.  
 Biljana him<sub>DAT</sub> it<sub>ACC</sub> gives a gift to a boy  
 ‘Biljana gives a gift to a boy.’

In Bulgarian, such doubling (namely, with definite objects) is often optional;<sup>3</sup> while it is certainly possible to get CD in the same context as (7), it is not strictly necessary as it is in Macedonian.

## 2. Macedonian

Because the occurrence of CD is more easily predicted by definiteness in Mac, I use it as a point of departure. If we have a conjoined direct object with conjuncts differing in their definiteness feature, what kind of CD, if any, should we expect? There are two distinct possibilities that arise in Mac, examined in section 2.1 and section 2.2.

### 2.1. First Conjunct Agreement

In Mac, the direct object clitic may double only the first, definite conjunct of a coordinated DP.

<sup>2</sup> This can actually be grammatical under the specific-indefinite reading of the direct object (i.e., some specific gift), an issue I address in section 3.

<sup>3</sup> As pointed out in Cinque and Krapova 2008, this stands in contrast to what they consider true clitic doubling, in which the nature of the predicate not only licenses, but in fact necessitates the presence of a clitic double.

- (9) Im            **ja**            dade [kniga-ta            i    edna igračka]  
 them<sub>DAT.PL</sub>    **it**<sub>FEM.SG</sub>    gave    **book**<sub>DEF.FEM.SG</sub>    and    a            toy  
 na deca-ta.  
 to    kids<sub>DEF</sub>  
 ‘She gave the kids the book and a toy.’

The bracketed phrase is the coordinated DP direct object of the verb *dade* (‘gave’), and one might reasonably expect that the multiple objects denoted by this argument would give rise to a plural CD. Indeed, when both objects are definite, that is precisely what we find.

- (10) **Gi**            vidov [professor-ot            i    student-ot].  
**them**<sub>ACC.PL</sub>    saw<sub>1SG</sub>    professor<sub>DEF.MASC.SG</sub>    and    student<sub>DEF.MASC.SG</sub>  
 ‘I saw the professor and the student.’

However, as we saw in example (1), a plural CD is disallowed when one conjunct is indefinite. Moreover, when the conjuncts differ in definiteness, their ordering is fixed:<sup>4</sup> the definite object must precede the indefinite. The CD, then, agrees only with the first conjunct.<sup>5</sup>

- (11) **Go**            vidov [professor-ot            i    eden student].  
**him**<sub>ACC.SG.MASC</sub>    saw<sub>1SG</sub>    professor<sub>DEF.MASC.SG</sub>    and    a/one student  
 ‘I saw the professor and the student.’

### 2.1.1. A Structure-Based Derivation of FCA for CD

Although first conjunct agreement of clitic doubles will be derived differently from FCA of verbs or adjectives, the underlying mechanisms that result in first conjunct agreeing CD in Mac may be the same.

Benmamoun (1992) proposes one theory of FCA, suggesting that coordination in all contexts is phrasal but that there exist multiple mechanisms (Spec-head agreement and government) that can be applied to derive agreement, and that using different mechanisms results in different surface  $\phi$ -feature agreement. This theory was used to explain data in which we find two

<sup>4</sup> The reason for this particular ordering constraint is unclear, but it is certainly worthy of attention and deserving of more extensive research.

<sup>5</sup> An anonymous reviewer suggested that sentence (9) might involve an intonational break prior to the final conjunct, indicating that it is some kind of afterthought (thus explaining the singular agreement on the CD). I returned to my Macedonian consultants to determine whether this was the case, and they responded that there is no break and that the prosody is that of a normal declarative sentence.

distinct agreement patterns for coordinated NPs, FCA and Full Agreement (FA):

- (12) a. *ža*                      *ʃomar w kariim*                      (Moroccan Arabic)  
           *came<sub>MASC.3SG</sub>*        Omar and Karim
- b. *žaw*                      *ʃomar w kariim*  
           *came<sub>MASC.3PL</sub>*        Omar and Karim
- ‘Omar and Karim came.’

This theory (particularly its reliance on government) became conceptually problematic following the advent of the Minimalist Program.

Aoun, Benmamoun, and Sportiche (1994, 1999) offer an alternative theory, suggesting that FCA results from clausal coordination with ellipsis; i.e., the agreeing head in a Spec-head relation to the first of the two conjuncts is the only one to surface at PF, concretely, that (12a) is derived as shown in (13).

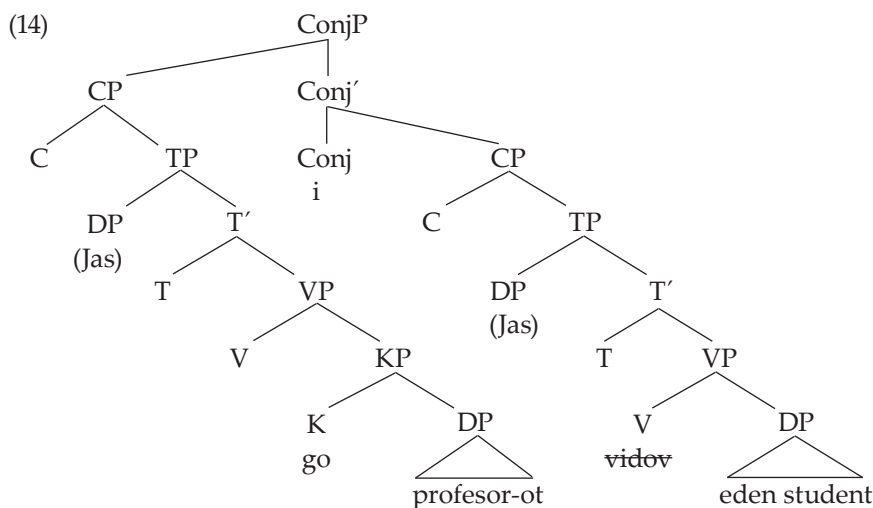
- (13) [ConjP [*ža ʃomar*] w [*ža kariim*]]  
       [ConjP [*came<sub>MASC.3SG</sub> Omar*] and [*came<sub>MASC.3SG</sub> Karim*]]

In summary, Aoun et al. (1994, 1999) argue that the variation in agreement stems from two distinct coordination structures: phrasal (i.e., coordinated NPs resulting in FA) and clausal (e.g., as in (13), resulting in FCA).

Although the precise mechanics are certainly different from FCA of verbs, it is possible to extrapolate the Aoun et al. (1994, 1999) proposal and apply it to CD in Macedonian. The tree in (14) on the facing page shows the derivation of (11) according to the Aoun et al. proposal and also following Franks and Rudin 2005, in which clitic doubles are  $K^0$  that become dislocated from their DP associates.

## 2.2. Adverbial Modification and Single-Event Reading

Although framework-specific considerations were one reason for the departure from previous analyses, the crucial factor motivating the Aoun et al. (1994, 1999) proposal concerned “number-sensitive items” (NSI). NSIs are elements that naturally require agreement with a particular number. For example, the adverbial modifier ‘together’ requires a non-singular antecedent, as illustrated in the following examples from Aoun et al. (1994, 1999).



- (15) \*ža                    ʃomar w kariim bžužhum    (Moroccan Arabic)  
 came<sub>MASC.3SG</sub> Omar and Karim together  
 'Omar and Karim came together.'

The singular number feature on the verb conflicts with the requirement of the NSI (namely, that it needs a non-singular subject), yielding an ungrammatical sentence. Aoun et al. offer this as evidence that FCA constructions involve clausal coordination while FA constructions (like those forced by the presence of an NSI) involve phrasal coordination. We find a phenomenon analogous to (15) in Macedonian.<sup>6</sup>

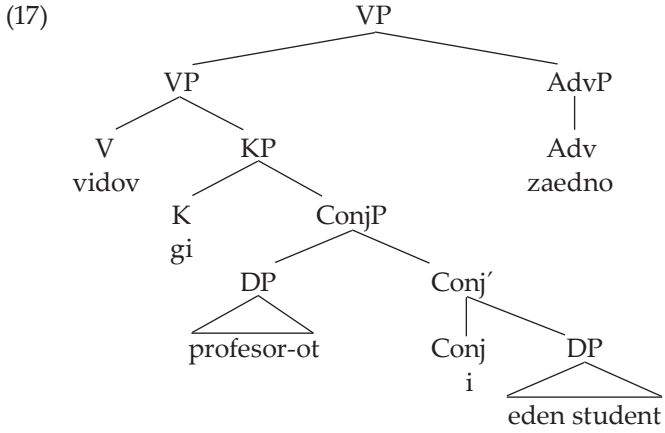
- (16) Gi    vidov   profesor-ot   i    eden student zaedno.  
 them saw<sub>1SG</sub> professor<sub>DEF</sub> and a/one student together  
 'I saw the professor and a/some student together.'

The presence of the NSI *zaedno* forces a single event reading; i.e., the professor and the student were seen simultaneously as part of a single-event. This

<sup>6</sup> Interestingly, this construction works even if neither of the conjuncts is definite, but in this case there is a clear preference for the specific indefinite reading on one or both conjuncts; see (i):

- (i) Gi    vidov   eden profesor   i    eden student zaedno.  
 them saw<sub>1SG</sub> a/one professor and a/one student together  
 'I saw a certain professor and a (certain) student together.'

suggests that there is only one VP whose head takes the KP headed by *gi* as a complement, as in (17).



**2.3. The Mechanics of CD Agreement**

One still unresolved question concerns the nature of the agreement operation that results in the CD being singular in (14) and plural in (17). The Aoun et al. account is compelling precisely because both FA and FCA are achieved via the Spec-head agreement of the clausal subject DP and the V-T complex head. Achieving the agreement properties of CD, however, is not as straightforward, since the clitic is analyzed as a head.

There are two possibilities: feature spreading and phrasal movement, both of which are compatible with the data presented. Under a feature spreading account, one would say that the number features of the XP sister to K<sup>0</sup> percolate up to KP and are then inherited by K<sup>0</sup>. Alternatively, we could posit movement and subsequent Spec-head agreement. According to this analysis, K<sup>0</sup> requires a Specifier and its sister is extracted and raised to Spec-KP. The data do not obviously favor one these analyses over the other (either one predicts the correct agreement pattern), so the question of feature movement vs. phrasal movement remains open.

**2.4. Issues with the Phrasal vs. Clausal Coordination Account**

Although it works fairly well for Macedonian, the Aoun et al. account of FCA encounters problems as we look at a broader set of data. For example, Marušič, Nevins, and Saksida (2007) note that NSIs, like ‘together’, do not necessarily prohibit FCA in Slovene, which suggests that the clausal coordination analysis cannot be extended to all instances of FCA/FA variation.

Subsequent proposals (e.g., Bošković 2009, Benmamoun et al. 2009) have relied upon Agree (in the case of Benmamoun et al., Agree as well as a PF rule) in order to derive FCA. These proposals account for a much wider array of data and are in some ways simpler.

### 2.4.1. An Agree-Based Account of FCA in CD

Following Benmamoun et al. (2009), we could suggest that there is no structural difference distinguishing FCA from FA, and that both are derived from the structure we see in (17). Under this analysis, the variation would result from *how* agreement arises; i.e., whether Agree causes the features of the  $K^0$  probe to be valued within the narrow syntax (in which case, there would be FA) or at PF, in which case a linear order rule applies, yielding FCA. Given the Mac data, the latter would be triggered specifically when an NSI is present.

## 3. Bulgarian

One difference between Mac and Bg is that the plural CD in Bg can occur with a conjoined DP in the absence of an NSI. Moreover, Bg allows a multiple-event reading in such sentences.

- (18) a. Vidjax gi profesora i edin student.  
           saw<sub>1SG</sub> them professor<sub>DEF</sub> and a/some student
- b. Profesora i edin student gi vidjax.  
           professor<sub>DEF</sub> and a/some student them saw<sub>1SG</sub>  
           ‘I saw the professor and a/some student.’

It should be noted that these data are not accepted by all speakers (see footnote 1). However those who do find (18) grammatical are consistent with respect to the interpretive requirements: if the NSI *zaedno* (‘together’) is omitted, as in (18), the sentences can be read either as single-event (i.e., I saw the the professor and a student together) or as multiple-event (i.e., I saw the professor and a student on distinct occasions).

The single-event interpretation is most natural in (18b), in which the conjoined object appears sentence-initially. This could be an indication (as some native speakers have confirmed) that *profesora i edin student* represents a topicalized phrase. In Bg, topicalization automatically triggers CD, thus giving the plural clitic *gi*.

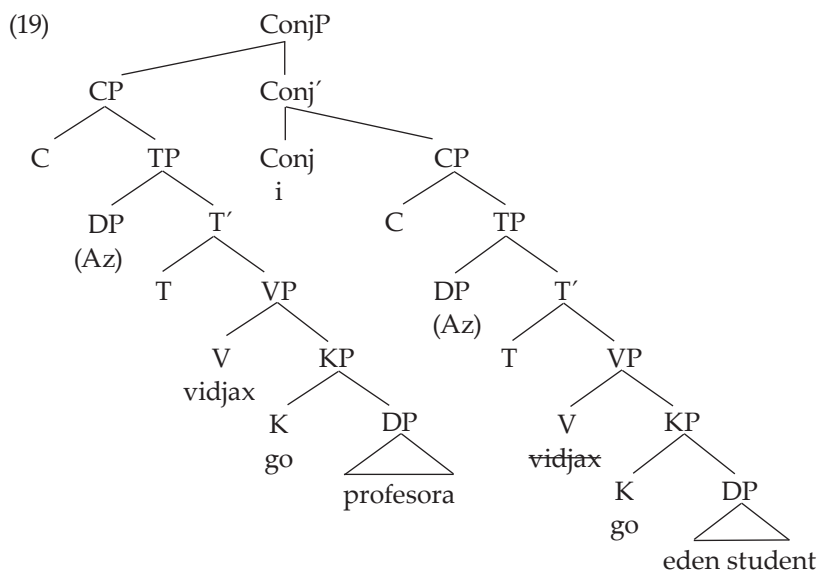
It is the multiple-event interpretation, however, which yields a curious side effect: the indefinite DP *edin student* is obligatorily interpreted as a specific indefinite, ‘a certain student’. As it turns out, this relationship hinges entirely on the model of agreement one adopts (Agree-based vs. structure-based).

If we adopt an Agree-based account, the syntax of (18) is analogous to what we find in derivation (17). Because this example involves FA, there is no need to invoke any of the mechanisms resulting in FCA; it simply requires a single Agree operation in which the probe,  $K^0$ , receives its plural number feature from  $\text{Conj}^0$ . But in no version of this analysis is there a way to explain the different interpretive possibilities as a reflex of the narrow syntax. This is not necessarily a problem; one might argue that the interpretation of the indefinite DP is determined at LF. However, we might see these Bulgarian data and their interpretive possibilities as supporting the structure-based account proposed by Aoun et al.

The similarity between Bg sentence (18) and Mac sentence (10) is compelling. Both have two potential interpretations: single-event and multiple-event. Given the Aoun et al. proposal, we could attribute this to a structural difference, i.e., clausal vs. phrasal coordination. Clausal coordination would involve two VPs, resulting in a multiple-event reading, while phrasal coordination would involve only one VP, yielding a single-event reading. However, in that case, we need to reflect on how plural agreement arises under clausal coordination. The Bg data provide crucial insight into this question.

In Bulgarian, the difference between a single-event interpretation and a multiple-event one coincides with different interpretations of the indefinite *edin student*; i.e., when it is specific indefinite, the only available reading of the sentence involves multiple events. If we adopt the Aoun et al. hypothesis, this suggests that the specific indefinite must be part of a biclausal structure, thus yielding the multiple-event reading. So we expect a derivation similar to (14) with one significant contrast: the specificity of the indefinite triggers CD. This means that the DP *edin student* is embedded within KP headed by a clitic double, *go*.

In structure (19), we find two clauses, each with its own direct object argument and each of these arguments, in turn, with its own dedicated CD. But in the surface word order, we do not find \**go go*, rather we find the plural clitic *gi*.



Across languages we find restrictions against consecutive phonologically identical segments, which frequently result in haplology (e.g., Amharic (Kramer 2014)), and Bulgarian exhibits this restriction with respect to pronominal clitics (Mišeska-Tomić 2006: 105). One might speculate, however, that if the features of the two clitics fully overlap, they might be reanalyzed into a single clitic that unifies the feature complexes of each individual term. In the case of (19), this would yield a plural clitic at linearization.

#### 4. Conclusion

This paper explores a complex tripartite syntactic relationship between coordination, agreement, and interpretation. In Macedonian we find an instance of First Conjunct Agreement which does not occur in Bulgarian, and conversely, in Bulgarian we find an instance of Full Agreement which does not occur (except under special circumstances, i.e., modification by an NSI) in Macedonian. Most (if not all) accounts of FCA vs. FA focus on verbal agreement, adjectival agreement, or participial agreement (which could be seen as a hybrid of the other two), but in this paper, I explore agreement with clitic doubles. I outline two competing theories of FCA, namely, structure-based theories and Agree-based theories, and illustrate how each applies to the clitic-doubling data. Finally, I show that both models can account for the data, but that a structure-based model may be more effective at explaining the interpretive constraints found in Bulgarian.



University of California-Riverside  
vschidambaram@gmail.com

## References

- Anagnostopoulou, Elene. (2006) "Conditions on clitic doubling in Greek". Henk van Riemsdijk, ed. *Clitics in the languages of Europe*. Berlin: Mouton de Gruyter, 761–98.
- Aoun, Joseph and Elabbas Benmamoun. (1999) "Gapping, PF Merger, and patterns of partial agreement". Shalom Lappin and Elabbas Benmamoun, eds. *Fragments: Studies in ellipsis and gapping*. Oxford: Oxford University Press, 175–87.
- Aoun, Joseph, Elabbas Benmamoun, and Dominique Sportiche. (1994) "Agreement and conjunction in some varieties of Arabic". *Linguistic inquiry* 25: 195–220.
- . (1999) "Further remarks on first conjunct agreement". *Linguistic inquiry* 30(4): 669–81.
- Arnaudova, Olga. (2002) "Clitic left dislocation and argument structure in Bulgarian". Jindřich Toman, ed. *Formal Approaches to Slavic Linguistics 10: The second Ann Arbor meeting 2001*. Ann Arbor, MI: Michigan Slavic Publications, 23–46.
- Benmamoun, Elabbas. (1992) *Functional and inflectional morphology: Problems of projection, representation, and derivation*. Ph.D. dissertation, University of Southern California.
- Benmamoun, Elabbas, Archana Bhatia, and Maria Polinsky. (2009) "Closest conjunct agreement in head-final languages". *Linguistic variation yearbook* 9(1): 67–88.
- Bošković, Željko. (2009) "Unifying first and last conjunct agreement". *Natural language and linguistic theory* 27(3): 455–96.
- Cinque, Guglielmo and Krapova, Iliyana. (2008) "Clitic reduplication constructions in Bulgarian". Dalina Kalluli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 257–88.
- Dimitrova-Vulchanova, Mila. (1999) "Clitics in the Slavic languages". Henk van Riemsdijk, ed. *Clitics in the languages of Europe*. Berlin: Mouton de Gruyter, 83–122.
- Dimitrova-Vulchanova, Mila. and Lars Hellan. (1999) "Clitics and Bulgarian clause structure". Henk van Riemsdijk, ed. *Clitics in the languages of Europe*. Berlin: Mouton de Gruyter, 469–514.
- Franks, Steven L. and Catherine Rudin. (2005) "Bulgarian clitics as K<sup>0</sup> heads". Steven L. Franks, Frank Y. Gladney, and Mila Tasseva-Kurktchieva, eds. *Annual workshop on formal approaches to Slavic linguistics 13: The Columbia meeting 2004*. Ann Arbor, MI: Michigan Slavic Publications, 104–16.

- Guentchéva, Zlatka. (2008) "Object clitic doubling constructions and topicality in Bulgarian". Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 203–26.
- Harizanov, Boris. (2016) "Clitic doubling at the syntax-morphophonology interface: A-movement and morphological merger in Bulgarian". *Natural language and linguistic theory* 32(4): 1033–88.
- Jaeger, T. Florian. (2004) "Topicality and superiority in Bulgarian wh-questions". Olga Arnaudova, E. Wayles Browne, María Luisa Rivero, and Danijela Stojanović, eds. *Formal approaches to Slavic linguistics: The Ottawa meeting, 2003*. Ann Arbor, MI: Michigan Slavic Publications, 207–28.
- Kramer, Ruth. (2014) "Clitic doubling or object agreement: The view from Amharic". *Natural language and linguistic theory* 32: 593–634.
- Leafgren, John R. (1997) "Bulgarian clitic doubling: overt topicality". *Journal of Slavic linguistics* 5: 117–44.
- Marušič, Franc, Andrew Nevins, and Amanda Saksida. (2007) "Last-Conjunct agreement in Slovenian". Richard Compton, Magdalena Golezinska, and Ulyana Savchenko, eds. *Formal approaches to Slavic linguistics: The Toronto meeting, 2006*. Ann Arbor, MI: Michigan Slavic Publications, 210–27.
- Mišeska-Tomić, Olga. (2006) *Balkan sprachbund morpho-syntactic features*. Dordrecht: Springer. [*Studies in natural language and linguistic theory* 68]
- Rudin, Catherine. (1997) "AgrO and Bulgarian pronominal clitics". Martina Lindseth and Steven L. Franks, eds. *Formal approaches to Slavic linguistics: The Indiana meeting, 1996*. Ann Arbor, MI: Michigan Slavic Publications, 224–52.

# A Bulgarian Solution to the Slavic Q Question?\*

Steven L. Franks

*Abstract:* The robust case and agreement evidence for positing a special Q(uantifier)P(hrase) in Russian is missing in Bulgarian. This paper asks whether such a Q category is nonetheless also warranted in that language. One relevant morphological fact is the *brojna* ('count') *forma*, another is the existence of numeral classifiers in Bulgarian, a third is the interaction between definiteness inflection and numerals. Also treated are binding and other c-command data that necessitate a QP in BCMS, where the extent to which these might carry over to Bulgarian is explored. It is concluded that, just as in other Slavic languages, on the whole the evidence supports the postulation of a distinct QP projection in Bulgarian as well.

## 0. The Question

For many Slavic languages, there is good reason to posit a special Quantifier category distinct from other, more familiar parts of speech. This paper addresses the question of whether Bulgarian (Bg) warrants this as well, despite the fact that, given its morphological impoverishment, the traditional arguments in the generative literature based on languages like Russian do not readily carry over to Bg.<sup>1</sup> The evidence for a Q(uantifier)P(hrase) in Bg, it will be concluded, is suggestive but far from definitive.

## 1. A Little Background: Q in Russian

The mixed properties of Slavic numerals have long puzzled linguists. Classic studies of Russian by Soviet scholars such as Suprun (1959) led to more theoretically oriented treatments by (at the time) young Western linguists such as

---

\* I am grateful to many individuals, including Loren Billings, Željko Bošković, Guglielmo Cinque, Miloje Despić, Tom Grano, Boris Harizanov, Boban Karapejovski, Pasha Koval, Iliyana Krapova, Ivana LaTerza, Petya Osenova, Roumyana Pancheva, Anita Peti-Stantić, Catherine Rudin, Aida Talić, Vesela Simeonova, Arthur Stepanov, Ksenia Zanon, and an anonymous reviewer for their advice on various aspects of this paper.

<sup>1</sup> While most of what I will have to say applies to Macedonian as well, I leave that language for future consideration.

Corbett (1978), Neidle (1982/1988), and Pesetsky (1982). Corbett demonstrated that numerals universally fall along a categorial continuum (in Ross's 1972 terms, a "squish") from adjectival to nominal. Thus, 'one' is always the most adjectival numeral and, from there, numbers become increasingly nounlike the higher their cardinality. Neidle, working within the Lexical-Functional Grammar framework of Bresnan (2001), and Pesetsky, working within the Government-Binding (GB) framework of Chomsky (1981), both attacked the problem in more formal terms. They argued for discrete solutions, assigning numerals with different properties to different categories such as A(djective), N(oun), and Q(uantifier). The thrust of Neidle's account was that numerals could have mixed paradigms, Russian *pjat'* 'five' for example being an N in the direct cases—NOM(inative) and ACC(usative)—but an A in the oblique cases. Compare (1a) and (1b):

- (1) a. *pjat'* butylok b. s *pjat'ju* butylkami  
 five<sub>NOM/ACC</sub> bottles<sub>GEN</sub> with five<sub>INST</sub> bottles<sub>INST</sub>

In (1a) *pjat'* seems to govern GEN(itive) on the N, whereas in (1b) *pjat'ju* seems to agree in INST(rumental) case with the N.<sup>2</sup> The thrust of Pesetsky's account was that, although numerals were Qs that combine with the nouns over which they quantify, the resulting phrase could be either a QP or an NP, with distinct properties depending on whether or not the Q was the head. Thus, (1a) is actually ambiguous, as demonstrated by the alternatives in (2):

- (2) a. *pjat'* butylok stojalo b. *pjat'* butylok stojali  
 five bottles<sub>GEN</sub> stood<sub>NEUT</sub> five bottles<sub>GEN</sub> stood<sub>PL</sub>

In (2a) the Q *pjat'* seems to cause the neuter verb *stojalo* 'stood' not to agree, whereas in (2b) the N *butylok* 'bottles' seems to cause the plural verb *stojali* 'stood' to agree. For Pesetsky, this meant *pjat' butylok* 'five bottles' was a QP in (2a) but an NP in (2b).

One reason it makes sense to think of numerals as assigning GEN is that certain non-numeric quantifiers also have this property, such as Russian *neskol'ko* 'several' and *mnogo* 'many':

- (3) a. *neskol'ko* butylok b. *mnogo* butylok  
 several bottles<sub>GEN</sub> many bottles<sub>GEN</sub>

<sup>2</sup> Following Babby's generally adopted 1987 terminology, (1a) instantiates a "heterogeneous" case pattern whereas (1b) is "homogeneous."

Additionally, *mnogo* (but not *neskol'ko*) has an unequivocally adjectival form, *mnogie*, which fails to assign GEN and, when used as a subject, requires that the verb agree:

- (4) *mnogie*      *butylki*      *stojali*/\**stojalo*  
       *many*<sub>NOM</sub>    *bottles*<sub>NOM</sub>    *stood*<sub>PL</sub>/*stood*<sub>NEUT</sub>

Adjectival *mnogie* has more of an individuating sense than the quantifier *mnogo* which, as a Q assigns GEN.

At least since Babby (1987), considerable attention has been focused by generative grammarians on the rise of a specialized Q category in the history of Slavic. That is, given an expression such as *pjat' butylok*, Babby pointed out that in Old Russian *pjat'* was a (NOM or ACC) noun and *butylok* was a simple adnominal GEN. Numerals such as 'one, two, three, four' were adjectival, except that 'one' appeared with a singular noun, 'two' with a dual, and 'three, four' with a plural. According to Babby (1987: 104), "*pjat'* has been reanalyzed as a quantifier, a new grammatical category with morphosyntactic properties which differ from those of both nouns and adjectives," while "*butylok*, the head of the adnominal complement in O[ld] R[ussian], has been reanalyzed as the head of the entire NP in M[odern] R[ussian]." Consequently, "the GEN marking on *butylok*, which was the adnominal GEN in OR, has been reanalyzed in MR as the 'quantitative' GEN," which, for Babby, was a special "GEN assigned to a non-maximal projection of N<sup>0</sup> in the scope of the quantifier."

In a series of works, most notably Franks (1994, 1995), I too argued that Russian had developed a special quantificational genitive, dubbed GEN(Q), that differed from the regular genitive in being a "structural" case.<sup>3</sup> This explained the alternation in (1), since the structural GEN(Q) could be overridden by inherent INST in (1b) but not by structural NOM or ACC in (1a). For (2a) versus (2b), I then argued that the Q *pjat'* assigned GEN to the NP *butylok* in both, but that in the latter the QP dominating *pjat'* was itself contained within a DP:<sup>4</sup>

<sup>3</sup> In the Chomskyan tradition "structural" means dissociated from theta-roles, and in the GB framework it also meant assigned at S-structure rather than D-structure. While I exploited the latter property to handle GEN(Q), in my system "structural" was simply the unmarked value for the feature opposition [±oblique]. More precisely, I argued that the structural GEN(Q) was [-oblique] and the inherent GEN was [+oblique], also extending the opposition to the special DAT(ive) (or locative) assigned by the distributive preposition *po* in the various languages. For evidence of DAT(Q) in Russian, see example (25) below. For West Slavic, I showed that there is variation between these alternatives.

<sup>4</sup> This analysis has several advantages over those of Pesetsky and Babby. By respecting X-bar syntax, we need not claim GEN(Q) is assigned to a non-maximal projec-

- (5) a. [QP Q [NP<sub>GEN</sub>]]                      b. [DP D [QP Q [NP<sub>GEN</sub>]]]

QP subjects do not induce agreement, but DP subjects do. Since that time, two innovations require a small adjustment to the original account. The first is that a growing body of compelling research, initiated by Bošković (2005), now demonstrates a clear dichotomy between NP-languages such as Russian or BCMS,<sup>5</sup> on the one hand, and DP-languages such as English or Bg, on the other. The second is the realization that Slavic clitics such as BCMS ACC *ga* or DAT *mu* should be analyzed as K<sup>0</sup> heads of KPs, so that when one needs to posit a projection above NP or QP in NP-languages, that projection is typically going to be a KP (rather than a DP).<sup>6</sup> Updating the earlier account, then, QPs in Russian are embedded in KPs, albeit optionally when NOM (or ACC), giving rise to the contrast between no subject-verb agreement (2a)/(5a) and subject-verb agreement (2b)/(5b).<sup>7</sup> In sum, when the quantified phrase is a bare QP it lacks case and the predicate cannot agree with it. However, as a DP/KP it does have case, hence as a NOM subject it causes the verb to show (plural) agreement. On the other hand, since oblique/inherent case must be assigned, presumably for theta-theoretic or lexical requirement reasons, only the DP/KP option is possible in necessarily instrumental (1b).

## 2. Extending the Analysis: Does BCMS Really Need GEN(Q)?

Unlike earlier studies, which concentrated on Russian, the importance of Franks (1994) was its attempt to accommodate the different properties of numeral phrases in a range of Slavic languages by tinkering with accounts developed initially for Russian. In particular, I argued that (i) the option of projecting only up to QP (hence non-agreement) was only true of Russian/East

---

tion of N and we achieve a consistent structural relationship between the Q and its GEN(Q) NP complement.

<sup>5</sup> I use this term as a catch-all for the linguistic systems of Bosnia, Croatia, Montenegro, and Serbia, putting aside any differences as not relevant to the discussion at hand.

<sup>6</sup> While true of BCMS, for Bg, which has definite articles, Franks and Rudin (2005) argue that this language has both DPs and KPs. Hence, pronominal clitics such as ACC *go* occupy K<sup>0</sup> but KP dominates DP, roughly:

- (i) [KP K [DP D [NP ... ]]]

Clitic doubling in Bg is also a consequence of the possibility of simultaneously realizing both KP and DP, an idea in keeping with Bošković's (2008) observation that "clitic doubling is possible only in languages with articles." See also Franks (2017) for related discussion.

<sup>7</sup> Since no such optionality is available for regular (i.e., non-QP) NPs, which always trigger agreement when NOM, under this account we must assume that these are always embedded in KPs.

Slavic and (ii) whether or not GEN(Q) was a structural case varied across the languages. Specifics of how (various instantiations of) the approach handled the facts of Slavic languages such as BCMS, Polish, Czech, and Upper/Lower Sorbian can be found in much of my work, including Franks (1994, 1995, 1998, 2000, 2002, 2003, 2009), so I will not attempt to reproduce those ideas here.

Nonetheless, since the present paper looks at Bg, let us briefly consider its Balkan neighbor BCMS. In previous studies, especially Franks (1995: ch. 4) and Franks (2002: sec. 4.5), I did argue that BCMS warrants a dedicated Q category distinct from regular N and A, this despite the fact that BCMS neither countenances anything like (1b) nor exhibits the same contrast as in (2).<sup>8</sup> That is, given the absence of INST in the BCMS version of Russian (1b), as seen in (6), I argued that GEN(Q) is inherent in that language:

- (6) sa pet boca  
with five bottles<sub>GEN</sub>

This of course makes it no different from the canonical GEN assigned by verbs, prepositions, and the adnominal genitive. Despite this, in BCMS, numerals such as *pet* are clearly not nouns. For one thing, they do not decline: any N replacing *pet* in (6) would be morphologically INST. And predicates do not ordinarily show plural agreement:

- (7) pet boca je stajalo  
five bottles<sub>GEN</sub> AUX<sub>3SG</sub> stood<sub>NEUT</sub>

<sup>8</sup> To be sure, as the reviewer reminds me, while this is true for the higher numerals, the situation is very different in BCMS for the paucal numerals *dva/dve*, *tri*, *četiri* ‘two, three, four’. Despić (2013b) argues that, whereas in Russian these are usually treated as assigning genitive singular, in BCMS, as subjects they actually modify nominative expressions with paucal number (which can look like neuter plural):

- (i) Dva mlada/\*mladog(a) čoveka su došla.  
two young<sub>PAUC/\*GEN</sub> men<sub>PAUC/GEN</sub> AUX<sub>3PL</sub> arrived<sub>PAUC/NEUT.PL</sub>  
‘Two young men arrived.’

Even in Russian there are complexities that suggest these do not really take genitive singular, since sometimes the stress of the paucal noun is distinct from that of the genitive and there exist both genitive and nominative options for modifying adjectives. And of course in Slavic languages such as Polish these occur with unambiguously nominative plural nouns, plural modifiers, and plural verbs. That the paucal numeral is agreeing in BCMS also makes sense because, as Despić notes, these can also decline, e.g., for the INST assigned by prepositions we have two options (although the first is not viable with verbs that take INST):

- (ii) a. sa dve žene b. sa dvema ženama  
with two<sub>FEM</sub> women<sub>PAUC/PL</sub> with two<sub>FEM.INST</sub> women<sub>INST</sub>

I argued that plural cannot percolate up from *boca* ‘bottles’ to the containing phrase (characterized in Franks 1995 as an NP and in Franks 2002 as a DP, but now arguably best regarded as a KP) because it is blocked by the case mismatch, even though as a subject the entire phrase in (7) must be NOM, which is eventually spelled out as neuter by default. Franks (2002) further exploited the existence of QPs in BCMS (and Polish) by showing that the most insightful way to understand the otherwise mystifying limitations on the contexts in which numerically quantified expressions can occur is by restricting the distribution of QPs in terms of the case markedness of their environments.<sup>9</sup> In short, although the argumentation is more subtle for BCMS than for Russian, this language also needs a Q category different from N and A.<sup>10</sup>

Be that as it may, there is one much newer proposal which deserves some scrutiny and which further demonstrates that Q, even in BCMS, must be a distinct category. This is Bošković’s (2014, 2016) argument that, rather than positing fixed phase heads, it is instead only the highest phrase in the extended projection of a lexical category which counts as a phase. This means that processes which target phases and their complements or which require the exploitation of phase edges, such as ellipsis, movement, and binding, are sensitive to what phrases project in any given nominal domain. The important contrast for present purposes is between adjectives, which, Bošković argues, are adjoined to NP, versus quantifiers, which project a new phrase above NP. Let us review this idea before asking what evidence there is for a dedicated Q in Bg.

The relevant argument can be constructed on the basis of the binding facts in BCMS. First, note that because the possessive adjective is adjoined to NP it also c-commands out of that NP. This can be seen in the following examples (8a, b) from Despić (2011, 2013a), where Binding Conditions B and C force disjoint reference. Throughout this paper, clitic pronouns are represented using small caps.

<sup>9</sup> Specifically, I concluded the following:

- (i) Polish: QPs cannot occur in a DP with any marked case feature.
- (ii) Russian: QPs cannot occur in a DP with more than one marked case feature.
- (iii) BCMS: QPs cannot occur in a DP with more than one marked case feature (or, for some speakers, two).

Note that these generalizations crucially rely on positing QPs in these languages in the first place. For motivating data the reader is referred to Franks 2002.

<sup>10</sup> As noted in footnote 8 above, the paucal numerals have a number of properties which distinguish them from numerals such as *pet* ‘five’. It is likely that the latter share part-of-speech features with Ns and the former share part-of-speech features with As, but that both have an additional categorial feature identifying them as Quantifiers.



- (8) a. \*<sub>[NP</sub> **Kusturicin**<sub>i</sub> [<sub>NP</sub> najnoviji film]] **ga**<sub>i</sub> je  
 Kusturica's latest movie HIM AUX<sub>3SG</sub>  
 zaista razočarao.  
 really disappointed  
 [Intended] 'Kusturica<sub>i</sub>'s latest movie really disappointed him<sub>i</sub>.'
- b. \*<sub>[NP</sub> **Njegov**<sub>i</sub> [<sub>NP</sub> najnoviji film]] je zaista  
 his latest movie AUX<sub>3SG</sub> really  
 razočarao **Kusturicu**<sub>i</sub>.  
 disappointed Kusturica  
 [Intended] 'His<sub>i</sub> latest movie really disappointed Kusturica<sub>i</sub>.'

Despić argues that, assuming an NP-adjunction structure, not only will Condition B and C effects arise when the pronoun or R-expression is at the left-edge as in (8), but they will also arise even when ostensibly protected by another modifier, such as the demonstrative *ovaj* 'this':

- (9) a. \*<sub>[NP</sub> **Ovaj** [<sub>NP</sub> **Kusturicin**<sub>i</sub> [<sub>NP</sub> najnoviji film]]] **ga**<sub>i</sub> je  
 this Kusturica's latest movie HIM AUX<sub>3SG</sub>  
 zaista razočarao.  
 really disappointed  
 [Intended] 'This latest movie of Kusturica<sub>i</sub>'s really disappointed him<sub>i</sub>.'
- b. \*<sub>[NP</sub> **Ovaj** [<sub>NP</sub> **njegov**<sub>i</sub> [<sub>NP</sub> najnoviji film]]] je zaista  
 this his latest movie AUX<sub>3SG</sub> really  
 razočarao **Kusturicu**<sub>i</sub>.  
 disappointed Kusturica  
 [Intended] 'This latest movie of his<sub>i</sub> really disappointed Kusturica<sub>i</sub>.'

These judgments are particularly striking for an English speaker, since in English all the intended readings are perfectly natural. The difference, it is claimed, has to do with the fact that possessives and demonstratives in English entail an additional DP projection above NP.<sup>11</sup>

<sup>11</sup> Russian, which is also an NP-language, behaves similarly to BCMS:

- (i) \*Èti ego<sub>i</sub> prijatelji kritikovali Ivana<sub>i</sub>.  
 these his friends criticized Ivan  
 [Intended] 'These friends of his<sub>i</sub> criticized Ivan<sub>i</sub>.'  
 [similarly: \*Ego<sub>i</sub> (èti) prijatelji kritikovali Ivana<sub>i</sub>.]

Now for the interesting part if the element to the left of the offending expression is a QP-projecting quantifier, such as *pet* ‘five’ or *mnogo* ‘many’, then the disjoint-reference effect disappears. This discovery was first reported by Despić (2011: 70–71), who observes: “It has been argued by a variety of authors (e.g., Franks, 1994, Bošković, 2006) that certain numerals and quantifiers in SC project QP, taking the whole NP as its complement, e.g., [QP [Q' Q NP]] ... When a quantifier of this type [is added], Condition B effects disappear, as expected.” Here is one of Despić’s original examples:<sup>12</sup>

- (10) [QP [Q' Mnogo [NP **Kusturicinih**<sub>i</sub> [NP prijatelja ]]]] je  
           many          Kusturica<sub>i</sub>'s<sub>GEN</sub>      friends<sub>GEN</sub>      AUX<sub>3SG</sub>  
 kritikovalo **njega**<sub>i</sub>.  
 criticized      him<sub>i</sub>  
 ‘Many of Kusturica<sub>i</sub>’s friends criticized him<sub>i</sub>.’

He comments that because “*mnogo* projects a QP immediately dominating the subject NP and the possessor *Kusturicinih* ‘Kusturica’s’ ... the possessor does not c-command the object pronoun, and consequently Condition B is not violated.” Despić (pp. 71–72) also offers the following minimal pair:

- (11) a. \***Njegov**<sub>i</sub> prijatelj je      kritikovao **Kusturicu**<sub>i</sub>.  
           his      friend      AUX<sub>3SG</sub>      criticized      Kusturica  
 [Intended] ‘His<sub>i</sub> friend criticized Kusturica<sub>i</sub>.’  
 b. [QP [Q' Mnogo [NP **njegovih**<sub>i</sub> [NP prijatelja ]]]] je  
           many          his<sub>GEN</sub>          friends<sub>GEN</sub>      AUX<sub>3SG</sub>  
 kritikovalo **Kusturicu**<sub>i</sub>.  
 criticized      Kusturica<sub>i</sub>  
 ‘Many of his<sub>i</sub> friends criticized Kusturica<sub>i</sub>.’

The sentence in (11a) violates Condition C since the pronominal possessor c-commands the object R-expression. In (11b), on the other hand, no Con-

See Zanon (2015) for additional applications of Bošković’s relativized phase system to Russian. Thanks also to Ksenia Zanon (p.c.) and Pasha Koval (p.c.) for providing Russian judgments.

<sup>12</sup> And from Bošković (2014), who builds on Despić’s insights:

- (i) [QP Pet/Mnogo [NP **njegovih**<sub>i</sub> [NP filmova]]] je  
           five/many          his<sub>GEN</sub>          movies<sub>GEN</sub>      AUX<sub>3SG</sub>  
 proslavilo      **Kusturicu**<sub>i</sub>.  
 made-famous      Kusturica  
 ‘Five/Many of his<sub>i</sub> movies made Kusturica<sub>i</sub> famous.’

dition C effect arises because the QP blocks the possessor *njegovih* 'his' from c-commanding *Kusturicu* 'Kusturica'. Finally, in a footnote Despić observes that the agreeing quantifier *mnogi* 'many', equivalent to Russian *mnogie* in (4), behaves not like a Q but like an adjective. Since it is adjoined to NP it does not project a QP and accordingly does not block violations of Conditions B and C.<sup>13</sup>

- (12) \*<sub>2</sub>[<sub>NP</sub> Mnogi [<sub>NP</sub> **njegovi**<sub>i</sub> [<sub>NP</sub> [<sub>NP</sub> prijatelj<sub>i</sub>]]]] su kritikovali  
           many<sub>NOM</sub> his<sub>NOM</sub> friends<sub>NOM</sub> aux<sub>3PL</sub> criticized  
**Kusturicu**<sub>i</sub>.  
 Kusturica  
 [Intended] 'Many of his<sub>i</sub> friends criticized Kusturica<sub>i</sub>'

The upshot is that, once again, we see evidence that certain lexical items with quantificational semantics need to be kept categorially distinct, as Qs, from Ns and As. Presumably they are characterized by an additional [+Q] feature, on top of those that define the major parts-of-speech (traditionally, [ $\pm$ N,  $\pm$ V]).<sup>14</sup>

### 3. What about Bulgarian?

The remainder of this paper is concerned with Bg, where we examine the extent to which the kinds of evidence adduced for Q in Russian and BCMS pertain to that language. While much of the particular arguments do not carry over, it will nonetheless be concluded that there is reason to separate out quantifiers in Bg as well.

#### 3.1. The *brojna forma*

As a point of departure, recall that many of the facts motivating a special Q category distinct from N are intimately related to case. Since Bg has a highly

<sup>13</sup> As expected, Ksenia Zanon (p.c.) confirms that Despić's BCMS judgments hold of Russian as well:

- (i) Mnogo Mašiny<sub>x</sub><sub>i</sub> prijatelej ee<sub>i</sub> kritikovalo.  
       many Masha's friends her criticized  
       'Many of Masha<sub>i</sub>'s friends<sub>i</sub> criticized her<sub>i</sub>.'
- (ii) ?\*Mnogie Mašiny<sub>i</sub> prijateli ee<sub>i</sub> kritikovali.  
       many Masha's friends her criticized  
       [Intended] 'Many of Masha<sub>i</sub>'s friends<sub>i</sub> criticized her<sub>i</sub>.'

<sup>14</sup> As noted in footnote 3, this is also true of certain quantificational Ps, such as distributive *po*, and may even extend to Vs such as those in Russian (and elsewhere) prefixed with quantificational *na-*.

impoverished case system, with no real case on nominals to speak of, the kind of pattern displayed by Russian (1) will of course not carry over. There is, however, a special *brojna forma* (or count form) in *-a*. This occurs on (in most dialects, only non-human) maculine nouns after numerals and is morphologically distinct from the regular *-i* or *-ove* plurals.<sup>15</sup> Here are some examples from Andrejčin et al. (1977: 122–23) and Hauge (1999: 24):

- (13) a. dva **brjaga**    b. pet **konja**    c. sto **kilometra**  
 two shores<sub>BF</sub>    five horses<sub>BF</sub>    hundred kilometers<sub>BF</sub>

We might therefore want to say that, just as in Russian, Qs in Bg select a particular form on their complement NPs. And, just as with Russian, there is some idiosyncrasy mentioned in the literature both with respect to which elements count as Q and how particular Ns behave.<sup>16</sup> For example, the *brojna forma* almost always preserves the stem,<sup>17</sup> including stem vowels (*pet orela* ‘five eagles’ vs. regular plural *orli*), which also means that the ending is never stressed (compare *dva brjága* with articulated *bregá* ‘the shore’). To my mind, this suggests that the *-a* suffix is added early in the derivation, blocking the regular plural and its effects.

There is considerable variation in its use and in particular its competition with the regular plural. Two recent papers that examine this competition are Kristozova (2012) and Stateva and Stepanov (2016). The latter is an interesting study of the factors which affect failure to use the *brojna forma* in contexts that call for it. They observe that distance between the numeral and the noun has a direct impact and that definiteness also matters. So although the plural *prozorci* ‘windows’ is technically ungrammatical in (14), the more intervening material, the more likely a Bg speaker is to produce and/or accept it.<sup>18</sup> Stateva and Stepanov (2016) argue that this is an agreement error, i.e., a matter of performance.

<sup>15</sup> Although most reference grammars just repeat the standard restrictions or mention the form in passing, Nicolova (2017) provides somewhat more detail. One excellent early source entirely dedicated to its use is Bernard (1954). I thank the many friends and colleagues who responded to my electronic request for relevant references.

<sup>16</sup> For some discussion of norms and variation, see Pancheva (this volume).

<sup>17</sup> Two exceptions are (*kilo*)*metâr*, as in (13c), and *litâr*, possibly because the *â* is epenthetic rather than underlying.

<sup>18</sup> Because they found a cumulative effect for adjectives but only a minimal (not statistically significant) effect for adverbs, they argue that structural rather than linear distance is at play here. The presence of an article also increases the likelihood of the regular plural.

- (14) pet(te) (stari) ((mnogo) prašasali) (dârveni) **prozoreca**/\*prozorci  
 five<sub>(DEF)</sub> (old<sub>PL</sub>) ((very)) dusty<sub>PL</sub> (wooden<sub>PL</sub>) windows<sub>BF/\*PL</sub>

The choice is also sensitive to the distinction between true quantifiers and cardinal numerals, in that only the latter take the *brojna forma*:

- (15) a. mnogo vestnici/\***vestnika** b. malko stolove/\***stola**  
 many newspapers<sub>PL/\*BF</sub> few chairs<sub>PL/\*BF</sub>

Interestingly, *kolko* does when it has an interrogative meaning, because there it asks for a numeric response, but not when it is exclamatory. Consider the following pair based on Andrejčin et al. (1977: 122):<sup>19</sup>

- (16) a. Kolko **romana** e napisal Vazov?  
 how-many novels<sub>BF</sub> AUX<sub>3SG</sub> wrote Vazov  
 ‘How many novels did Vazov write?’  
 b. Kolko romani e pročel tozi čovek!  
 how-many novels<sub>PL</sub> AUX<sub>3SG</sub> read that person  
 ‘How many novels that person has read!’

The only formal treatment of this of which I am aware can be found in the present volume, but goes well beyond the purview of this paper.<sup>20</sup> For our purposes, the crucial point is simply that such a special form exists and that it is sensitive to numeric quantifiers, implying the need for some kind of quantifier functional projection in Bg as well.

### 3.2. Numeral Classifiers

Cinque and Krapova (2007) provide additional evidence for such a category. They discuss classifiers in Bg such as *dúši* ‘people (lit. souls)’, as in (17):

<sup>19</sup> *Tolkova* ‘so many’ is similar. Bernard (1954: 36) contrasts (i) with (ii):

- (i) Ihave *tolkova stola*.  
 have<sub>1PL</sub> so-many chairs<sub>BF</sub>  
 ‘We have this many (a given number of) tables.’  
 (ii) *Izpočupixa tolkova stolove!*  
 broke<sub>3PL</sub> so-many chairs<sub>PL</sub>  
 ‘So many tables broke!’

<sup>20</sup> This is Pancheva’s contribution to the volume. Other discussions in the literature include Ouwayda (2014: 94–98) and Stepanov and Stateva (2018: section 6).

- (17) a. trima **dúši** studenti/rabotnici/bâlgari  
 three persons students/workers/Bulgarians  
 ‘three students/workers/Bulgarians’
- b. Kolko **dúši** imaše tam?  
 how-many persons were there  
 ‘How many people were there?’

Such classifiers can only occur in numeral contexts, as above.<sup>21</sup> Cinque and Krapova point out that this form, which they call a “sortal” classifier, is accentually distinct from *duši*, the regular plural of *duša* ‘soul’. Of particular interest in the present context is the fact that *dúši* tracks exactly the same conditions as the *brojna forma* does, i.e., it is only possible after numbers and also shows the pattern in (15) and (16).

One curious difference, however, is that unlike the *brojna forma* the sortal classifier cannot be modified: \**trima dobri dúši studenti* ‘three good persons students’. This makes sense if *dúši* is actually the head of some functional projection above NP, as in (18a). Numerals which induce the *brojna forma* could then be analyzed as its specifier instead, with a  $\emptyset$  head, as in (18b).<sup>22</sup>

- (18) a. [QP NumP [Q' dúši [NP<sub>PL</sub>]]]      b. [QP NumP [Q'  $\emptyset$  [NP<sub>BF</sub>]]]

An interesting complication, drawn to my attention by Iliyana Krapova, is that the sortal classifier itself can appear in the *brojna forma* if it refers to an inanimate object. Her example is given in (19a), for which she suggests the structure in (19b):

- (19) a. pet broja prozorci/\*prozoreca  
 five number<sub>BF</sub> windows<sub>PL/\*BF</sub>  
 ‘five windows’
- b. [QP [NumP pet] [Q' broja<sub>BF</sub> [NP prozorci<sub>PL</sub>]]]

It is however unclear to me that simply combining (18b) with (18a) will suffice. The problem is that the claim implicit in (18) is that the *brojna forma* results specifically from having a silent Q.<sup>23</sup> That is, according to this analysis, for QP

<sup>21</sup> There is a large literature on classifiers and their relationship to numerals (cf. especially Simpson 2005 for Southeast Asian languages) and to quantifiers (cf., e.g., Löbel 1990 for German).

<sup>22</sup> Although not shown here, this phrase can itself be inside a DP, KP, or both (cf. footnote 6).

<sup>23</sup> This proposal resembles Bailyn’s (2012) analysis of Russian predicators, where Pred assigns INST only when itself silent. This idea also extends to his account of Russian

to exist, either SpecQP or Q<sup>0</sup> has to be occupied. Krapova suggests we might then conclude that whenever Q appears empty, it must really contain a silent sortal classifier. While this makes sense, unifying overt elements such as *dúši* and *broj* with  $\emptyset$ , where only the latter selects the *brojna forma* (cf. footnote 23), the possibility of (19a) implies additional structure in order to license the *brojna forma* on the sortal classifier *broj*. One might contend that *broja* enters the structure as a sort of light noun complement to null Q, receives the *brojna forma* as such, and then raises to Q:

(20) [QP pet [Q' **broja**<sub>BF</sub> [<sub>NP</sub> ~~broja~~ [NP prozorci<sub>PL</sub>]]]]

If correct, it is worth noting that only sortal classifiers can do this. Cinque and Krapova oppose these to “mensural” classifiers such as *butilki* ‘bottles’ in (21):

(21) dve butilki vino  
two bottles wine

Mensural classifiers are an open-ended class, whereas Cinque and Krapova only acknowledge three sortal classifiers (or four if one counts archaic *glavi* ‘heads’ as in *deset glavi ovce* ‘ten head (of) sheep’). They also differ in being able to occur without a numeral as in (22a) and in allowing modification as in (22b):

(22) a. butilkite vino                      b. dve pâlni lâžici sirup  
bottles<sub>DEF</sub> wine                              two full spoons syrup

Mensural classifiers, although they too originate in *n* and therefore (if masculine) also appear in the *brojna forma* in the context of numerals, remain in situ. We see this on *paket* in (23):

(23) a. dva paketa/\*paketi molivi/\*moliva  
two packs<sub>BF/\*PL</sub> pencils<sub>PL/\*BF</sub>  
‘two packs of pencils’  
b. [QP dva [Q'  $\emptyset$  [<sub>NP</sub> paketa<sub>BF</sub> [NP molivi<sub>PL</sub>]]]]

Note that as (19) and (23) show it is only the classifier which appears in the *brojna forma*. This follows from the proposed structure, since the classifier protects NP from being selected by Q.

---

numerals, in that, as mentioned in the following footnote, only a null Q assigns GEN to its complement NP.

This account has several other advantages. One is that it may explain why mensural classifiers are more nominal, with structures for (22) as follows:

- (24) a. [DP [<sub>NP</sub> butilkite [<sub>NP</sub> vino]]]  
 b. [QP dve [Q'  $\emptyset$  [<sub>NP</sub> pâlni lâžici [<sub>NP</sub> sirop]]]]

Unlike sortal classifiers, which are selected by and raise to Q, mensural classifiers are simply light *n* and, as such, they admit modifiers compatible with their bleached content. Another is that the assumed structure comports well with the analysis in Franks 1995: 149 and ff. There I argued that numerals such as *pjat* 'five' are in SpecQP on the basis of the fact that the distributive preposition *po* can assign it dative in Russian.<sup>24</sup> I further claimed that this dative must be a structural DAT(Q), since it can be assigned to the specifier of its complement under GB's Exceptional Case Marking:<sup>25</sup>

- (25) [PP *po* [QP *pjati*<sub>DAT</sub> [Q'  $\emptyset$  [<sub>NP</sub> *rublej*<sub>GEN</sub>]]]]  
 'five rubles each'

The Russian null Q assigns GEN(Q) just as the Bg null Q selects for the *brojna forma* on its complement NP. Interestingly, the so-called paucal numerals ('two', 'three', and 'four') cannot be directly assigned case. Thus, 'two rubles each' can only be expressed as in (26b), not (26a):

- (26) a. \*[PP *po* [QP *dvum*<sub>DAT</sub> [Q'  $\emptyset$  [<sub>NP</sub> *rublej*<sub>GEN</sub> / *rubljam*<sub>DAT</sub> ]]]]  
 b. [PP *po* [QP [Q' *dva* [<sub>NP</sub> *rublja*<sub>PAUC</sub>]]]]

The paucals in Russian are instead themselves heads of QP, hence are incompatible with  $\emptyset$ .<sup>26</sup> This is why they can behave differently from  $\emptyset$ , assigning a special form which is generally syncretic with the GEN singular but occa-

<sup>24</sup> Bailyn (2012: section 5.2.3) also places Russian numerals in SpecQP, with a null Q head assigning GEN(Q), but only in the heterogeneous paradigm (i.e., when they do not decline); cased numerals are heads in Q. He thus states (p. 212) that "the Russian hybrid behavior reduces to the possibility that numerals like '5' can behave as heads, absorbing the Genitive case and leaving itself and its complement in the (lexical) case domain of a higher head." For an interesting recent treatment of quantifiers which also argues that, in the history of Germanic, heads became reanalyzed as specifiers, see Roehrs and Sapp 2016.

<sup>25</sup> When numerals agree they enter into the same structure as adjectives: [<sub>NP</sub> *pjati* / *dvum*<sub>DAT</sub> [<sub>NP</sub> *rubljam*<sub>DAT</sub>]].

<sup>26</sup> Bg in this regard differs from Russian, since all numerals behave similarly.



sionally distinct.<sup>27</sup> This was once dual number, but different languages dealt with the loss of the dual differently. Russian recast the form as a kind of case, extending it to ‘three’ and ‘four’ (which used to occur with plural nouns), and reanalyzed the GEN (plural) governed by the higher numerals (erstwhile nouns) as assigned by a new Q category, hence GEN(Q). Against this background, it is interesting to observe that the Bg *brojna forma* reflects a different solution: with the loss of case there could be no GEN(Q), but instead the *-a* dual was generalized and extended to be used after all numerals, as in (18b).

### 3.3. Definiteness Inflection

Definiteness in Bg, a DP language, is expressed inflectionally on the highest head in the extended nominal projection. Here are some examples:

- (27) a. *knigata*    b. *deteto*    c. *mâžât*  
           book<sub>DEF</sub>    child<sub>DEF</sub>    man<sub>DEF</sub>

- (28) a. *ženite*    b. *gradovete*    c. *mâžete*  
           women<sub>DEF</sub>    cities<sub>DEF</sub>    men<sub>DEF</sub>

At issue here are not the rules of placement—for more complex data and various possible analyses I refer the interested reader to Franks (2001)—but rather the form of the article. In general, it depends on the gender-number of the host word: *-ta* in (27a) being feminine, *-to* in (27b) being neuter, *-ât* in (27c) being masculine, and *-te* in (28) being plural. One cannot just look at the final vowel of the stem because *e* can be neuter, as in (27b), or plural, as in (28b, c).<sup>28</sup> However, when the stem ends in *a* the inflection must also be *-ta*, regardless of phi-features:

- (29) a. *selata*    b. *baštata*  
           villages<sub>DEF</sub>    father<sub>DEF</sub>

- (30) a. *xubavite sela*    b. *dobrijat bašta*  
           pretty<sub>DEF</sub> villages    good<sub>DEF</sub> father

<sup>27</sup> These have been much discussed in the literature, especially with respect to the problem of how adjectives within their scope behave. See Stepanov and Stateva (2018) for a comprehensive and insightful discussion.

<sup>28</sup> Also, consonantal stems, although typically masculine as in (27c), can also be feminine, in which case the definite inflection is stressed *-tá*: *noštá* ‘the night’.

- (31) a. **dvata** stola                      b. **dvamata** mâže<sup>29</sup>  
           two<sub>DEF</sub> chairs<sub>BF</sub>                      two<sub>DEF</sub> men

We see in (29) that because the stems end in *a* the suffix must be *-ta* but as (30) shows as soon as a modifier is introduced the suffix reflects its phi-features, i.e., plural for (30a) and masculine for (30b). On the other hand as (31) shows if that modifier happens to end in *a* then, once again, the definite suffix is *-ta*.

### 3.4. A Hundred is not a Lot

In light of this, consider how quantificational modifiers behave. We saw in (31) that if a numeral happens to end in *a*, then its articulated form is in *-ta*. Otherwise, it bears the *-te* suffix, reflecting the fact that it modifies a plural noun:

- (32) **dvete/trite**                      knigi/sela  
       two<sub>DEF</sub>/three<sub>DEF</sub> books/villages

Of course, it may not be so strange to see *dve* ‘two’ or *tri* ‘three’ with the *-te* inflection, since plurals in Bg often end in *e* or *i*. More surprising is the fact that *sto* ‘hundred’ also takes *-te*:

- (33) **stote/\*stoto** knigi/sela/mâže  
       hundred<sub>DEF</sub> books/villages/men

This is particularly striking since nouns in *o* take *-to*, not only when neuter as in (34a) but even when masculine, as in (34b):

- (34) a. **seloto**                                      b. **čičoto**  
           village<sub>DEF</sub>                                      uncle<sub>DEF</sub>

But compare (33) with what happens when the quantifier is *mnogo* ‘many’:

- (35) **mnogoto/\*mnogote** knigi/sela/mâže  
       many<sub>DEF</sub> books/villages/men

What is the difference here?

It is tempting to try to connect the distinction between *sto* and *mnogo* to another contrast these lexical items display: recall that although *sto* ‘hundred’

<sup>29</sup> *Dvama* is a special numeral used for modifying masculine humans (or groups containing at least one male), hence it can only take plural complements. Compare *dvata oficera* ‘two<sub>DEF</sub> (chess) bishops<sub>BF</sub> (lit. officers)’ with *dvamata oficeri* ‘two<sub>DEF</sub> officers<sub>PL</sub>’, cited by Xristozova (2012: 305). (In Bg, these forms exist only for ‘two’ through ‘six’.)

in (13c), like all other numerals, takes the *brojna forma*, *mnogo* ‘many’ and *malko* ‘few’ in (15) do not. Perhaps, then, there is a categorial difference between the two, and their articulated forms reflect this.<sup>30</sup> Unfortunately, the correlation is not sufficiently robust to allow us to reduce the contrast to a straightforward categorial contrast. The problem is that Bg has other quantificational elements ending in *o* that occur with the *brojna forma*, like *sto*, but which take the definite ending *-to*, like *mnogo*:

- (36) *njakolkoto*/\**njakolkote*    *vestnika*/\**vestnici*  
       several<sub>DEF</sub>                      newspapers<sub>BF</sub>/newspapers<sub>PL</sub>

It is nonetheless interesting that *njakolko* ‘several’ takes the *brojna forma* even though it is not a numeral per se. Bernard (1954: 36) states that this is because *njakolko* suggests a precise (or in any case, approximate) number.<sup>31</sup> That is, “several” means something like “two or three,” whereas “many” has no numeric correlate; *malko* ‘few’ and *poveče* ‘more’ are similar to *mnogo* in taking *-to* and eschewing the *brojna forma* in favor of the regular plural. It thus seems that we need both quantifiers and numerals, with *mnogo* ‘many’ belonging to the former group and *sto* ‘hundred’ to the latter, despite the phonological similarity.

If so, perhaps the generalization can be salvaged by exploiting a more fleshed out extended nominal projection system. Thus far, I have used NumP for the numeral in specifier position and referred to its containing phrase as a QP. A reasonable alternative, however, would be to reserve QP for quantifiers such as *mnogo* ‘many’ or *vsicki* ‘all’ and to use #P instead for numerals, as Ouwayda (2014) or Stepanov and Stateva (2018) do.<sup>32</sup> Under this scenario, an actual number still heads NumP but that NumP is the specifier of a #P rather than a QP. Elements like *vsicki* ‘all’ are canonical Qs, and elements like *pet* ‘five’ are canonical Nums, with NumP in Spec#P. The former inflect as heads, according to their final vowels (*mnogoto* and *vsickite*), while the latter inflect as agreeing plurals regardless of their own form (*stote* and *pette*). What this means for the analysis laid out in section 3.2 is that in Bg it must actually be a null # rather than a null Q that assigns the *brojna forma* and that the eventual locus of sortal classifiers is #. Further corroboration, as noted by Iliyana Krapova (p.c.), is that elements such as *mnogo* or *malko* cannot combine with

<sup>30</sup> In Franks (2001: 56) I suggested that “Q(uantifiers) taking *-to* are actually neuter nouns.”

<sup>31</sup> This sort of explanation probably won’t carry over to *kolkoto* ‘how many’, which seems to behave similarly. However, *kolkoto* is actually not an articulated interrogative and cannot, so far as I can tell, occur with the *brojna forma*. Rather, it only exists as a universal concessive, (co)relative, equative, or comparative form (cf. Franks and Rudin 2015).

<sup>32</sup> On QP in Bg cf. also Giusti and Dimitrova-Vulchanova 1996.

a sortal classifier in # (*\*mnogo/malko dúši studenti*), hence it makes sense to introduce them under QP rather than #P.<sup>33</sup> And, as expected, they can combine with mensural classifiers, since these are lower (*mnogo/malko butilki vino*). With this in mind, let us return to the problematic status of *njakolko*. It seems to me one might interpret its mixed behavior as calling for treatment as a kind of hybrid, perhaps analyzing it as starting in #, as a numeral and selecting for the *brojna forma*, but then moving up to Q, where it takes the definiteness ending *-to*, roughly as follows:

- (37) [DP D<sub>[+DEF]</sub>] [QP **njakolkoto** [Q' Q [#P [NumP *njakolko*]  $\emptyset$  [NP<sub>BF</sub> ]]]]

That being said, there is another peculiarity associated with *sto*, drawn to my attention by Iliyana Krapova and not to my knowledge previously observed. This concerns how speakers deal with the augmented forms *dvesta* 'two hundred' and *trista* 'three hundred' when they need to bear an article. Here, a curious thing happens, in that for many speakers the most natural definite version ends in *-te* rather than *-ta*.

- (38) a. *dvestate*  
two-hundred<sub>DEF</sub>                      b. *tristate*  
three-hundred<sub>DEF</sub>

This is the only place in Bg I know of where a word ending in *a* can fail to take *-ta* as its article suffix.<sup>34</sup> *Dvesta* and *trista* thus seem to behave like *sto*, with the definite form *stote*, and not like *dva*, with the definite form masculine *dvata* 'the two' (alongside *dvete* for feminine and neuter).<sup>35</sup> Even a pluralia tantum word like *xora* 'people', which is plural by every other diagnostic, takes *-ta*. The otherwise inviolate rule that *a* always wins thus finds a striking exception with numerals. In sum, whatever the analysis, this fact alone seems to warrant our assigning them to a special category.

<sup>33</sup> She also points out that *mnogo* can be modified, as in *kolko mnogo* 'how many', *naj-mnogo* 'the most', *tvârde/strašno mnogo* 'very many'. This implies that Q has a projection of its own, while the classifier is lower; they can even co-occur, as in *Kolko mnogo duši studenti dojdoxa?* 'How many (persons) students came?'

<sup>34</sup> Web searches also reveal instances of *dvestata* and *tristata*, although far fewer. However, this too is being regularized. Iliyana Krapova (p.c.) informs me that the newest dictionary produced by the Institute for Bulgarian Language (*Institut za bǎlgarski ezik*) now insists on these as the norm.

<sup>35</sup> Higher numbers based on "100" use *stotin* and predictably take *-te*: *četiristotinte* '400', *petstotinte* '500', etc.

### 3.5. Nominal Domain Phases

Another place to look for evidence of QP in Bg might be whether numerals count as new heads for the purposes of binding and c-command, as discussed in section 2 above for BCMS and Russian. There we saw that a numeral, unlike an adjective, served to establish a new category above NP, thereby rendering NP an opaque domain.

The problem here was that these are NP-languages rather than DP-languages. Despić's point about QP is that even in an NP-language addition of a new category, albeit not DP, could have the same kind of phase-creating effect. But if Bg, as is usually claimed, is a DP-language like English,<sup>36</sup> it is unclear that adding QP (or #P) above DP should affect the binding possibilities, in that we expect this language to pattern like English independently of the interpolation of a numeral. If so, then this test would not reveal any information relevant to our question of the status of Q in Bg. Judgments are admittedly quite vexed, perhaps too much to allow for any definitive conclusions. It turns out, however, that when more carefully examined the data do seem to support the contention that quantifiers can head their own projections in Bg.

Since this does not accord with the conclusions reached by LaTerza (2016), let us consider her paper as a point of departure. In attempting to extend Despić's account of BCMS to Bg (and Mac), LaTerza (section 2.2) makes the claim that these languages—contrary to expectations since they are DP- rather than NP-languages—show the same binding behavior as BCMS. If this were correct, then one might expect to find that numerals have the same kind of effect as they do in BCMS (and Russian). Here are some of LaTerza's Bg examples, modeled on those of Despić:

- (39) a. \***Ivanovijat**<sub>i</sub> papagal **nego**<sub>i</sub> uxapa včera. [LaTerza's (13b)]  
 Ivan's<sub>DEF</sub> parrot him bit yesterday  
 [Intended] 'Ivan<sub>i</sub>'s parrot bit him<sub>i</sub> yesterday.'
- b. \***Negovijat**<sub>i</sub> papagal uxapa **Ivan**<sub>i</sub> včera. [LaTerza's (13a)]  
 his<sub>DEF</sub> parrot bit Ivan yesterday  
 [Intended] 'His<sub>i</sub> parrot bit Ivan<sub>i</sub> yesterday.'
- c. \***Tozi negov**<sub>i</sub> papagal uxapa **Ivan**<sub>i</sub> včera. [LaTerza's (15a)]  
 this his parrot bit Ivan yesterday  
 [Intended] 'This parrot of his<sub>i</sub> bit Ivan<sub>i</sub> yesterday.'

<sup>36</sup> See Bošković (this volume) and references therein. Tasseva-Kurktchieva and Dubinsky (this volume), however, take issue with Bošković's characterization, arguing that Bg is a "weak" DP-language in that the nominal domain may project up to a DP or not, with distinguishing diagnostics.

The judgments she reports are identical to those for BCMS (8) and (9b), but there disjoint reference it will be recalled resulted from the possessors *Ku-sturicin* and *njegov* c-commanding out of the subject NP. If these are in DP, as in English, then what is the source of the disjoint reference effect in (39)? LaTerza considers two possibilities: one, which she rejects for good reason, is that the possessives in Bg are adjoined to DP (hence c-command out of it), and the other is that they raise covertly (i.e., in LF) to some high position from which they c-command the entire clause. Although this latter solution strikes me as promising, as it has the potential of unifying the binding possibilities in BCMS, Bg, and Mac, it really just recasts Despić's structure for BCMS by saying that in all these languages possessives must adjoin to the maximal projection in the nominal domain.<sup>37</sup> They would then be at the phase edge and c-command out of the subject NP, giving rise to Condition B and C effects uniformly. The question still remains of why this does not happen in other DP-languages, such as English; perhaps it has to do with the morphology of the possessives, which is adjectival in those languages, whereas English possessives are not even constituents, consisting as they do of a phrase in SpecDP and a genitive 's piece in D; on overt movement differences see also Bošković (2005).

Be that as it may, my informal inquiries suggest a number of problems with LaTerza's data. A major concern has to do with her use of the full pronoun rather than the clitic. For one thing, the string *papagalât nego uxapa* 'parrot<sub>DEF</sub> him bit', on which (39a) is based, is impossible to begin with (without heavily contrastively focusing *nego*). When confronted with this issue and asked to use a clitic instead, one of LaTerza's original informants provided (40) as acceptable with coreference (also commenting that if a clitic pronoun is used, then coreference is acceptable in the Bg translation of BCMS (10)):

<sup>37</sup> An argument she makes for this is that even embedded possessives show the disjoint reference effect in BCMS:

- (i) \*[NP [N Prijatelj] [NP **Markove**<sub>i</sub> majke]] je zagrlio **njega**<sub>i</sub>.  
 friend Marko's mother AUX<sub>3SG</sub> hugged him

[Intended] 'A friend of Marko<sub>i</sub>'s mother hugged him<sub>i</sub>.'

While indeed true, the lack of felicity has a different source than intended by LaTerza. As before, the question arises of what happens if a clitic is used instead of the tonic pronoun. It turns out that here coreference is good (I have slightly modified the example to make it less confusing):

- (ii) [NP [N Prijateljica] [NP **Markove**<sub>i</sub> majke]] **ga**<sub>i</sub> je zagrlila.  
 female-friend Marko's mother HIM AUX<sub>3SG</sub> hugged

'A (female) friend of Marko<sub>i</sub>'s mother hugged him<sub>i</sub>.'

This indicates that the problem in (i) has to do with using the tonic pronoun, and that in fact the embedded possessive can never c-command out of the higher NP. This is of course as expected.

- (40) **Ivanovijat<sub>i</sub>** papagal **go<sub>i</sub>** uxapa včera.  
 Ivan's<sub>DEF</sub> parrot HIM bit yesterday  
 'Ivan<sub>i</sub>'s parrot bit him<sub>i</sub> yesterday.'

This sort of judgment is corroborated by Iliyana Krapova (p.c.), who points out that the tonic form induces a disjoint reference reading even when LaTerza's sentences are corrected for word order. This is not true, however, if a clitic is used instead of the tonic pronoun. Compare Krapova's examples in (41):

- (41) a. \*Edin **nein<sub>i</sub>** papagal uxapa **neja<sub>i</sub>** včera.  
 one her parrot bit HER yesterday  
 [Intended] 'A parrot of her<sub>i</sub>'s bit her<sub>i</sub> yesterday.'
- b. Edin **nein<sub>i</sub>** papagal **ja<sub>i</sub>** uxapa včera.  
 one her parrot HER bit yesterday  
 'A parrot of her<sub>i</sub>'s bit her<sub>i</sub> yesterday.'

Along these lines, she also deems coreference possible in the following:

- (42) a. **Negovijat<sub>i</sub>** bašta **go<sub>i</sub>** smjata za mnogo inteligenten.  
 his<sub>DEF</sub> father HIM considers for very intelligent  
 'His<sub>i</sub> father considers him<sub>i</sub> very intelligent.'
- b. Baštata na **Marko<sub>i</sub>** **go<sub>i</sub>** smjata za mnogo inteligenten.  
 father<sub>DEF</sub> of Marko HIM considers for very intelligent  
 'Marko<sub>i</sub>'s father considers him<sub>i</sub> very intelligent.'

Crucially—and as predicted if the subject nominals are DPs in Bg but NPs in BCMS—there is a clear contrast here between the two languages. The following near minimal pair with Bg (42) shows that coreference is impossible in BCMS:

- (43) \***Markov<sub>i</sub>/Njegov<sub>i</sub>** otac **ga<sub>i</sub>** smatra veoma pametnim.  
 Marko's/his father HIM considers very intelligent<sub>INST</sub>  
 [Intended] 'Marko<sub>i</sub>'s/His<sub>i</sub> father considers him<sub>i</sub> very intelligent.'

In light of these corrected judgments, the effect of introducing a quantifier or numeral above the possessive in such Bg sentences would be immaterial, since there is nothing to be ameliorated. That is, so far as Condition B applied to clitic pronouns is concerned, Bg behaves like a DP-language. Matters, however, become more complicated when one considers R-expressions, and it is here that we find possible evidence for a QP.

Relying again on Krapova's judgments, let us probe Condition C more carefully. As a point of departure, she indeed rejects coreference in examples such as (39b), from LaTerza (2016) and repeated below:

- (44) \***Negovijat**<sub>i</sub> papagal uxapa **Ivan**<sub>i</sub> včera.  
 his<sub>DEF</sub> parrot bit Ivan yesterday  
 [Intended] 'His<sub>i</sub> parrot bit Ivan<sub>i</sub> yesterday.'

Interestingly, when a demonstrative is introduced, as in (45), coreference greatly improves; similarly, Krapova does not agree with the infelicitous judgment reported for (39c).

- (45) ?**Tezi negovi**<sub>i</sub> papagali uxapaxa **Ivan**<sub>i</sub> včera.  
 these his parrots bit Ivan yesterday  
 'These parrots of his<sub>i</sub> bit Ivan<sub>i</sub> yesterday.'

While, unsurprisingly, adding a numeral to the mix, as in (46a), does not diminish (45), it is striking that the numeral on its own has the same ameliorating effect as the demonstrative. This is shown in (46b).

- (46) a. ?**Tezi pet negovi**<sub>i</sub> papagala uxapaxa **Ivan**<sub>i</sub> včera.  
 these five his parrots<sub>BF</sub> bit Ivan yesterday  
 'These five parrots of his<sub>i</sub> bit Ivan<sub>i</sub> yesterday.'
- b. ?**Pette negovi**<sub>i</sub> papagala uxapaxa **Ivan**<sub>i</sub> včera.  
 five<sub>DEF</sub> his parrots<sub>BF</sub> bit Ivan yesterday  
 'The five parrots of his<sub>i</sub> bit Ivan<sub>i</sub> yesterday.'

Presumably, *tezi* 'these' indicates a DP and the quantifier indicates a QP (or #P) above the phrase containing the possessive pronoun, which is adjoined to NP, as in BCMS (8)–(12). It thus appears that embedding the subject inside a QP facilitates coreference, just like embedding it inside a DP does. Here is another, more complete paradigm provided by Krapova, which shows that not just numerals but any quantifier above the possessive makes coreference acceptable:

- (47) a. \***Nejnite**<sub>i</sub> problemi pritesnjavaxa **Marija**<sub>i</sub> mnogo.  
 her<sub>DEF</sub> problems troubled Maria much  
 [Intended] 'Her problems made Maria very uneasy.'



- (47) b. Tezi **nejni**<sub>i</sub> problemi pritesnjavaxa **Marija**<sub>i</sub> mnogo.  
 these her problems troubled Maria much  
 ‘These problems of hers made Maria very uneasy.’
- (48) a. Mnogoto **nejni**<sub>i</sub> problemi pritesnjavaxa **Marija**<sub>i</sub>.  
 many<sub>DEF</sub> her problems troubled Maria  
 ‘Her many problems made Maria uneasy.’
- b. Vsičkite **nejni**<sub>i</sub> problemi pritesnjavaxa **Marija**<sub>i</sub>.  
 all<sub>DEF</sub> her problems troubled Maria  
 ‘All her problems made Maria uneasy.’
- c. Tezi vsički **nejni**<sub>i</sub> problemi pritesnjavaxa **Marija**<sub>i</sub>.  
 these all her problems troubled Maria  
 ‘All these problems of hers made Maria uneasy.’

While (47) replicates the contrast exhibited in (44) versus (45), (48) demonstrates that introducing a quantifier above *nejni* ‘her’ similarly prevents it from c-commanding *Marija*, thereby avoiding the potential Condition C violation. Iliyana Krapova (p.c.) thus concludes that “the generalization seems to be that *nejni* cannot c-command *Marija*, and only when it reaches a higher Spec—SpecDP—it appears to be able to. It is protected in this sense by Dem, Q, etc.”

While this strikes me as the right conclusion in that we want to argue that the demonstrative or quantifier serves to establish a phase above *nejni* in (47b) and (48), we are nonetheless left with a serious problem: Why does the DP in (47a) not do the same? As suggested to me by Miloje Despić (p.c.), one can imagine accounts, indeed such as considered by LaTerza 2016, which involve covert movement of the possessive to a position from which it c-commands the entire clause—hence also the R-expression—and that this movement is blocked by the presence of an overt demonstrative or quantifier. This could be instantiated in a number of ways, which I do not explore here, because, however this is ultimately accomplished, puzzles remain. For the sake of discussion, let us suppose, as LaTerza (2016: 252) suggests, that in these languages “prenominal possessors uniformly raise at LF to the edge of their largest containing nominal.” When that nominal is the subject, this causes the possessor to c-command everything in the clause, just as quantified expressions in SpecDP must in English sentences such as (49) in order to give the bound-variable reading:<sup>38</sup>

<sup>38</sup> This requires LF movement, since the felicity of coreference in *John’s mother loves him* and *His mother loves John* (as well as ungrammatical *\*John’s mother lives himself*) shows that SpecDP does not c-command out of DP.

(49) Every boy's mother loves him.

Even so, since LF movement is notoriously unconstrained, what would cause the blocking effect of demonstratives and quantifiers? One possibility is that these somehow count as intervening operators. More credible, it seems to me, is that overt movement of the possessive phrase is involved, and that the intervening demonstrative or quantifier disrupts that movement. This could make sense if these expressions are specifiers (of DP for the demonstrative, of NumP or QP for the quantifier) and the possessive phrase moves to SpecDP from below (wherever it originates, but for the sake of explicitness let it merge in SpecPossP). So, movement to SpecDP roughly as in (50a) is felicitous, but similar movement is not possible in (50b):

- (50) a. [DP **possessive** D<sub>[+DEF]</sub> [POSSP [~~possessive~~ [NP ... ]]]]  
 b. \*[DP **possessive** D<sub>[+DEF]</sub> [QP **numeral/quantifier** [Q' Q [POSSP [~~possessive~~ [NP ... ]]]]]]

An overt demonstrative, on the other hand, would be incompatible with movement because SpecDP is already occupied:

- (51) [DP **demonstrative** D<sub>[+DEF]</sub> [POSSP [**possessive** [NP ... ]]]]

We then want possessives to adjoin to their containing phrase at LF, so that they scope out of that phrase, thereby giving rise to the observed binding theoretic effects. In this way, it is only possessives that are already high in the nominal domain that can c-command out of that domain. This of course requires *nejniti* in (47a) to be in SpecDP, whereas all the other (unarticulated) instances of *nejni* in (47) and (48) remain lower.<sup>39</sup>

Whatever the account of disjoint reference in (47a), as opposed to possible coreference when the possessive is lower, the mystery remains of why (40) and its ilk are acceptable.<sup>40</sup> That is, however we ultimately explain why having

<sup>39</sup> In Franks 2001 I compare several accounts of articulation in Bg, eventually arguing that the article appears on the head of the phrase that is the complement to D. This is not compatible with the movement approach described above. Instead, the specifier of the complement to D must move to SpecDP and undergo Spec-head agreement. It seems to me that this would similarly render SpecDP unavailable as an escape-hatch, giving rise to the kinds of Complex NP Constraint effects discussed for Bg by Bošković (this volume). (That is, as Bošković readily admits, the CNPC facts merely argue against an N-to-D movement analysis, but the truth is that everything we know about article placement in Bg, however accomplished, shows that the article is not in D, but rather an inflection on the head of some phrase as close as possible to D.)

<sup>40</sup> Two thoughts about how to approach this mystery are: either (i) the lack of c-command in (40) has to do with the position of the clitic (*vis-à-vis* a comparable R-expression), which in turn may be a matter of whether the goal is a head or a phrase; or (ii) it

just a DP, as in (47a), is not sufficient to avoid c-command for the purposes of Condition C, we are faced with the problem that replacing the R-expression with a clitic pronoun makes these good for Condition B.<sup>41</sup> Be that as it may, the data lead to two important conclusions. First, when one abstracts away from potential confounds and concentrates on unequivocal Condition B effects, the correct conclusion is that there is a real contrast between Bg and BCMS and that this contrast has to do with the depth of nominal structure (although just how to calculate that depth remains a puzzle). Second—and more relevant to the question posed by the title of this paper—when one explores putative Condition C effects in Bg (whatever their cause), one observes that quantifiers behave as if they project their own phrases above NP. In sum, although Bg behaves like a DP-language when one uses clitic pronouns to test for Condition B effects, so that introducing a numeral is irrelevant, more careful probing of the data reveals that numerals may matter for nominal structure in Bg. We conclude that the ability to suppress traditional Condition C effects demonstrates that numerals can indeed count as adding a distinct category above NP. And, more generally, we conclude that, just as in other Slavic languages, the evidence supports the postulation of a distinct QP projection in Bg as well.

Indiana University  
franks@indiana.edu

## References

- Andrejčin, Ljubomir, Konstantin Popov, and Stojan Stojanov. (1977) *Gramatika na bâlgarskija ezik*. Sofia: Nauka i izkustvo.
- Babby, Leonard. (1987) "Case, prequantifiers, and discontinuous agreement in Russian". *Natural language and linguistic theory* 5(1): 91–138.
- Bailyn, John. (2012) *The syntax of Russian*. New York: Cambridge University Press.
- Bernard, Roger. (1954) "Alcune osservazioni sul plurale secondo in bulgaro moderno." *Richerche slavistiche* 3: 30–42.
- Bošković, Željko. (2005) "On the locality of left branch extraction and the structure of NP." *Studia linguistica* 59: 1–45.
- . (2006) "Case checking vs. case assignment and the case of adverbial NPs". *Linguistic inquiry* 37: 522–533.

---

depends on timing, with c-command calculated (for some reason) for the clitic before LF movement of the possessive.

<sup>41</sup> It is also not obvious why tonic pronouns should pattern with R-expressions in requiring a further degree of embedding for coreference to become felicitous.

- Bošković, Željko. (2008) "What will you have, DP or NP?" Emily Elfner and Martin Walkow, eds. *NELS 37*. Amherst, MA: University of Massachusetts, Graduate Linguistic Student Association, 101–114.
- . (2014) "Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis". *Linguistic inquiry* 45: 27–89.
- . (2016) "Getting really edgy: On the edge of the edge". *Linguistic inquiry* 47: 1–33.
- . (this volume) "On affixal articles: An argument from Bantu for Bulgarian, Romanian, and Icelandic", 37–52.
- Bresnan, Joan. (2001) *Lexical-functional syntax*. Oxford: Blackwell.
- Chomsky, Noam. (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Cinque, Guglielmo and Iliyana Krapova. (2007) "A note on Bulgarian numeral classifiers". Gabriela Alboiu, Andrei A. Avram, Larisa Avram, and Daniela Isac, eds. *Pitar Moș: A building with a view. Papers in honour of Alexandra Cornilescu*. Bucharest: Editura Universității din București, 45–51.
- Corbett, Greville. (1978) "Numerous squishes and squishy numerals in Slavonic". *International review of Slavic linguistics* 3: 43–73.
- Despić, Miloje. (2011) *Syntax in the absence of Determiner Phrase*. Doctoral dissertation, University of Connecticut.
- . (2013a) "Binding and the structure of NP in Serbo-Croatian". *Linguistic inquiry* 44(2): 239–270.
- . (2013b) "A note on paucal, agreement and case". Alexander Podobryaev, ed. *Formal Approaches to Slavic Linguistics: The Second MIT Meeting*. Ann Arbor, MI: Michigan Slavic Publications, 57–71.
- Franks, Steven L. (1994) "Parametric properties of numeral phrases in Slavic". *Natural language and linguistic theory* 12: 570–649.
- . (1995) *Parameters of Slavic morphosyntax*. New York: Oxford University Press.
- . (1998) "Parameters of Slavic morphosyntax revisited: A minimalist retrospective". Željko Bošković, Steven L. Franks, and William Snyder, eds. *Formal Approaches to Slavic Linguistics: The Connecticut meeting*. Ann Arbor, MI: Michigan Slavic Materials, 134–165.
- . (2000) "Zametki o čislitel'nyx v slavjanskix jazykax." *Sâpostavitelno ezikoznanie* 25, 5–32.
- . (2001) "The internal structure of Slavic NPs, with special reference to Bulgarian". Adam Przepiórkowski and Piotr Bański, eds. *Generative linguistics in Poland: Syntax and morphosyntax*. Warsaw: Instytut Podstaw Informatyki PAN, 53–69.
- . (2002) "A Jakobsonian feature based analysis of the Slavic numeric quantifier genitive". *Journal of Slavic Linguistics* 10: 141–81.
- . (2003) "Case features, markedness, and quantification". Peter Kosta, Joanna Blaszczak, Jens Frasek, Ljudmila Geist, and Marzena Zygis, eds. *Investigations into formal Slavic linguistics. Contributions of the Fourth*

- European Conference on Formal Description of Slavic Languages (FDSL IV)*. Frankfurt am Mein: Peter Lang, 579–600.
- Franks, Steven L. (2009) “Case assignment in quantified phrases”. Tilman Berger, Karl Gutschmidt, Sebastian Kempgen, and Peter Kosta, eds. *HSK Slavische Sprachen: An International Handbook of their Structure, their History, and their Investigation*, Berlin: de Gruyter Mouton, 355–69.
- . (2017) *Syntax and Spell-Out in Slavic*. Bloomington, IN: Slavica Publishers.
- Franks, Steven L. and Catherine Rudin. (2005) “Bulgarian clitics as  $K^0$  heads”. Steven L. Franks, Frank Y. Gladney, and Mila Tasseva-Kurktchieva, eds. *Formal approaches to Slavic Linguistics: The Columbia meeting*. Ann Arbor, MI: Michigan Slavic Publications, 104–16.
- . (2015) “Invariant *-to* in Bulgarian”. Miriam Shrager, Edna Andrews, George Fowler, and Steven L. Franks, eds. *Studies in accentology and Slavic linguistics in honor of Ronald F. Feldstein*. Bloomington, IN: Slavica Publishers, 99–136.
- Giuliana Giusti and Mila Dimitrova-Vulchanova. (1996) “Quantified noun phrase structure in Bulgarian”. Jindřich Toman, ed. *Formal approaches to Slavic linguistics: The College Park meeting, 1994*. Ann Arbor, MI: Michigan Slavic Publications, 123–44.
- Hauge, Kjetil Rå. (1999) *A short grammar of contemporary Bulgarian*. Bloomington, IN: Slavica Publishers.
- LaTerza, Ivana. (2016) “Binding in English and South Slavic and the parameterized DP hypothesis”. *Linguistic inquiry* 47(4): 741–753.
- Löbel, Elisabeth. (1990) “D und Q als funktionale Kategorien in der Nominalphrase”. *Linguistische Berichte* 127: 232–64.
- Neidle, Carol. (1982/1988) *The role of case in Russian syntax*. PhD dissertation, Massachusetts Institute of Technology. [Revised version published by Kluwer: Dordrecht.]
- Nicolova, Ruselina. (2017) *Bulgarian grammar*. Berlin: Frank & Timme.
- Ouwayda, Sarah. (2014) *Where number lies: Plural marking, numerals, and the collective-distributive distinction*. Ph.D. dissertation, University of Southern California.
- Pancheva, Roumyana. (this volume) “How many flowers! So many colors! Number marking in cardinality exclamatives in Bulgarian”, 197–233.
- Pesetsky, David. (1982) *Paths and categories*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Roehrs, Dorian and Christopher Sapp. (2016) *Quantifying expressions in the history of German: Syntactic reanalysis and morphological change*. Amsterdam: John Benjamins
- Ross, John Robert. (1972) “The category squish: Endstation Hauptwort”. *Papers from the Chicago linguistic society* 8: 316–28.

- Simpson, Andrew. (2005) "Classifiers and DP structure in Southeast Asia". Guglielmo Cinque and Richard Kayne, eds. *The Oxford handbook of comparative syntax*. Oxford: Oxford University Press, 806–38.
- Stateva, Penka and Arthur Stepanov. (2016) "Agreement errors and structural distance: A corpus study of Bulgarian". *Zeitschrift für Slawistik* 61(3): 448–62.
- Stepanov, Arthur and Penka Stateva. (2018) "Countability, agreement and the loss of the dual in Russian". *Journal of linguistics*. [online at <https://doi.org/10.1017/S0022226718000130>]
- Suprun, A. E. (1959) *O russkix čislitel'nyx*. Frunze: Kirgizskij gosudarstvennyj universitet
- Tasseva-Kurktchieva, Mila and Stanley Dubinsky. (this volume). "On the NP/DP frontier: Bulgarian as a transitional case", 287–311.
- Xristozova, Galja. (2012) "Brojnata forma na sâštestvitelnite imena ot mâžki rod: Normi, pravila, narušenija". *Bulgarian language and literature* 54(3): 300–10.
- Zanon, Ksenia. (2015) *On hybrid coordination and Quantifier Raising in Russian*. PhD dissertation, Indiana University.

# Bulgarian Multiple *Wh* Relatives Revisited

Elena Dimova and Christine Tellier

На раменете на гиганти\*

*Abstract:* Bulgarian displays constructions that, at first glance, appear to be multiple free relatives (MFRs) in argument positions, as argued for the first time in Rudin's pioneering 1986 study. Taking her findings as a point of departure, this paper focuses more closely on the properties of fronted *wh*-phrases. It is proposed that the first *wh*-constituent functions as a Topic, while the second *wh* is the one that is selected by both the main and the embedded verb. This analysis supports the view that MFRs are indeed free relatives, but not MFRs in the strict sense, so they might be called Pseudo MFRs. Based on the morphological properties of *wh*-phrases (specifically, the obligatory affixation of the particle *-to*), we sketch an analysis of free relative formation and labeling which also accounts for the fact that an intervening Topic *wh*-constituent is transparent for the purposes of c-selection.

## 1. Introduction

Within the generative tradition, Catherine Rudin was, as far as we know, the very first to draw attention to the existence of multiple *wh* relatives, bringing forth examples such as (1) in Bulgarian (Rudin 1986: 163):

- (1) *Ženite grabnaxa koj kakvoto vidi.*  
women<sub>DEF</sub> grabbed<sub>3PL</sub> who what saw<sub>3SG</sub>  
'The women each grabbed whatever they saw.'

In her pioneering 1986 study, as well as in more recent papers (Rudin 2006; 2007a, b; 2008a, b; 2009, 2012), she analyzed these constructions in great detail, first in Bulgarian, but also in various Slavic languages and in Romanian. Rudin showed, on both semantic and syntactic grounds, that these structures are quite different from multiple *wh* correlatives of the type previously described

---

\* "On the shoulders of giants." While Catherine Rudin's work has been a great inspiration to us both, the first author wishes to especially thank her for her unfailing help, advice, encouragement, and kindness. Thanks are also due to an anonymous reviewer for very helpful comments. All errors and omissions are our own.

by Andrews (1975) and Wali (1982) in Indic languages and of the type also instantiated in Bulgarian and Slavic more generally. Indeed, correlatives are left-peripheral to the clause, and must contain an anaphoric demonstrative (the correlate, bolded)—see Izvorski 1996—as shown in (2), from Rudin (2008a):

- (2) Koj kolkoto može, **tolkova** praštajte. (multiple correlative)  
 who how-much can<sub>3SG</sub> that-much send<sub>IMP,2PL</sub>  
 ‘Whoever can manage however much, send that much.’

However, multiple *wh* relatives like those in (1)—and in (3a) below—do not display these properties: they occur in argument (and sometimes adjunct) positions, and they are incompatible with an anaphoric pronoun or demonstrative. This is illustrated in (3b), from Rudin (2007a: 291):

- (3) a. Praštajte koj kolkoto može. (multiple *wh* relative)  
 send<sub>IMP,2PL</sub> who how-much can<sub>3SG</sub>  
 ‘Everybody send as much as you can.’  
 b. Praštajte \*go/\***tolkova** koj kolkoto može.  
 send<sub>IMP,2PL</sub> it<sub>CL</sub>/that-much who how-much can<sub>3SG</sub>

Rudin’s conclusion with respect to constructions such as (1) and (3a) is that they are multiple free relatives (MFRs); as she pointed out, like free relatives (FRs), they display a matching effect: one of the *wh*-phrases must satisfy the selectional requirements of both the main and the embedded verb. In light of competing analyses of FRs (the so-called Head vs. Comp hypotheses, cf. Bresnan and Grimshaw 1978 and Groos and van Riemsdijk 1981, respectively), this raised the important question of what the structure of MFRs might be. In her early work (Rudin 1986: 177 ff), Rudin examined and rejected an analysis whereby the two *wh* constituents in (1) would occupy different positions, the second one in Comp and the first one outside the FR, functioning as an independent indefinite pronoun. She argued instead that the two *wh*-phrases are part of the same constituent, that is, both are in Comp. In more recent studies (Rudin 2007b, 2008a, b; 2012), Rudin adopts the view that the *wh*-phrases are best analyzed as multiple CP specifiers, much as she had proposed for multiple *wh* interrogatives in Rudin (1988).

Building on Rudin’s research, we take a somewhat different look at multiple *wh* relatives in Bulgarian. We adopt her conclusion that these constructions are different from correlatives and that both *wh*-phrases are part of the FR itself. However, we depart from her analysis in that, in our view, the first *wh*-phrase is not a CP specifier but rather a Topic. Indeed, as we will show,



it displays all the properties characteristic of topicalized constituents.<sup>1</sup> This is consistent with Rudin's conclusions to the extent that, in a cartographic or extended projection approach of the CP-clause, both Topics and so-called CP specifiers are part of the same constituent selected by the main verb. However, this leads us to view multiple *wh* relatives not as MFRs but as simple FRs preceded by a *wh*-topic (interpreted as a universally quantified expression, see section 4). We call such constructions Pseudo Multiple Free Relatives (PMFRs).<sup>2</sup>

We start by adducing arguments to show that the first (leftmost) *wh*-phrase in these constructions is indeed topicalized. In the last two sections of the paper, we will sketch an analysis of PMFRs cast within a minimalist approach to category labeling.

## 2. What is Selected?

Let us take as a point of departure one essential characteristic shared by FRs in most languages: the fronted *wh*-phrase must satisfy the selectional, categorial, and case requirements of both the main verb and the embedded verb, a property known as matching (Grimshaw 1977). If we take this as a definitional criterion, then strictly speaking a true MFR would be a relative containing more than one *wh*-phrase, where each *wh*-phrase satisfies the matching condition. However, as already observed by Rudin (1986: 169), only the second *wh*-phrase fulfills this requirement in (1), since *koj* 'who' is not selected by the main verb.

In fact, as it turns out this represents a typical property of multiple *wh* relatives in Bulgarian. As Dimova (2014) has shown, in many such constructions the main verb selects only one of the *wh*-phrases; further, the selected *wh*-phrase obligatorily occupies the second position. This is illustrated by the examples in (4) and by the contrasts in (5):

- (4) a. Vojnicite otivat na kogo(to) kâdeto e zapovjadal  
 soldiers<sub>DEF</sub> go<sub>3PL</sub> to who where AUX<sub>3SG</sub> commanded  
 ministârât.  
 minister<sub>DEF</sub>

'The soldiers go where the minister ordered each of them to go.'

<sup>1</sup> Rudin (1986: 127) showed that *wh* constituents in simple *wh* relatives are not Topics. Indeed, Rudin observes that simple *wh* relatives can contain Topics, which must precede the *wh* constituent. Since relatives with two topics are rare or nonexistent, this suggests that the *wh*-phrase itself is not a Topic. We agree with this conclusion. What our analysis suggests is that the Topic in relatives can also be a *wh*-phrase.

<sup>2</sup> See also Citko and Gračanin-Yuksek (2016), who conclude on independent grounds that true multiple *wh*-relatives do not exist.

- (4) b. Sâbraxme na kogo(to) kakvoto bjaxa izpratili.  
gathered<sub>1PL</sub> to who what AUX<sub>3PL</sub> sent<sub>3PL</sub>  
'We gathered whatever they had sent to each of us.'
- (5) a. Jam koj(to) kakvoto gotvi.  
eat<sub>1SG</sub> who what cook<sub>3SG</sub>  
'I eat whatever whoever cooks.'
- b. \*Jam koj(to) s kakvoto gotvi.  
eat<sub>1SG</sub> who with what cook<sub>3SG</sub>  
[Intended] 'I eat whatever whoever cooks with.'

The main verb selects an AdvP in (4a) and a DP in (4b). In (5), the main verb selects an inanimate DP; this requirement is fulfilled by the second *wh*-phrase in (5a), but is unfulfilled in (5b).<sup>3</sup>

The question immediately arises as to why the selected *wh*-phrase must systematically occupy the second position in these structures. As we will show, whenever two fronted *wh* constituents occur in these constructions, the first *wh*-phrase functions as a Topic located within the FR; as such, it is transparent for the purposes of selection by the main verb.<sup>4</sup> In the next sections we summarize the main arguments brought forth by Dimova and Tellier (2015) in support of this view.

### 3. The First *Wh*-Phrase is a Topic

How can we distinguish between Topics and non-Topics? One argument can be drawn from resumption by a clitic. It has been shown in the literature that in Bulgarian clitics can (although need not) resume dislocated Topic constituents (Arnaudova 2002, Krapova 2002, Franks and Rudin 2005). Following the

<sup>3</sup> Note that only the second *wh*-phrase obligatorily bears the particle *-to*. We assume that obligatory *-to* and optional *-to* differ in that only the former bears a [REL] feature. We return to this issue in section 5.

<sup>4</sup> As is well known, Topics do not interfere in the c-selectional process. Thus in a sentence like *It's obvious that Mary, he can't stand*, the topicalized DP does not hinder selection of a declarative TP by the C-head. Under a traditional IP-adjunction analysis of topicalization (Baltin 1982, Lasnik and Saito 1992), the transparency of Topics is easily accounted for. However, it is not so readily explained either under a cartographic approach, where several projections among which TopP separate a head from its selected complement (see Shlonsky 2006 for a discussion of this problem), or in a minimalist framework, where adjunction must boil down to a question of labeling or lack thereof (see Hornstein and Nunes 2008, Hornstein 2009). In section 6 we will provide an account of the transparency of Topics in the context of PMFRs.

terminology used in Krapova (2002) and Krapova and Cinque (2008a), we reserve the term Clitic Left Dislocated Topics (CLLD) for constructions as in (6a) where a left-dislocated constituent is linked to a resumptive pronoun. An example of clitic-less topicalization is given in (6b).

- (6) a. [TOP Na Ivan] mu dadox knigata. CLLD-Topic  
           to Ivan him<sub>CL</sub> gave<sub>1SG</sub> book<sub>DEF</sub>  
           ‘As for (to) Ivan, I gave the book to him.’
- b. [TOP Na Ivan] dadox knigata. Clitic-less Topic  
           to Ivan gave<sub>1SG</sub> book<sub>DEF</sub>  
           ‘As for (to) Ivan I gave the book.’

Though Topics do not require resumption, what is crucial for our purposes is that non-Topic constituents *cannot* be resumed by a clitic, as (7) illustrates:

- (7) [FOC NA IVAN] (\*mu) dadox knigata.  
           to Ivan him<sub>CL</sub> gave<sub>1SG</sub> book<sub>DEF</sub>  
           ‘TO IVAN, I gave the book.’ (not to Maria)

Turning now to multiple *wh* relatives, we note that a resumptive pronoun may occur within the relative clause. But while it may correspond to the first *wh* constituent (henceforth, Wh1), as in (8a), it cannot resume the second one (henceforth, Wh2); cf. (8b):

- (8) a. Vojnicite otivat kogo(to) kâdeto go izpratjat.  
           soldiers<sub>DEF</sub> go<sub>3PL</sub> whom where him<sub>CL</sub> send<sub>3PL</sub>  
           ‘The soldiers go where they send each of them.’
- b. Pitaše kâde(to) kogoto (\*go) vidi.  
           asked<sub>3SG</sub> where whom him<sub>CL</sub> see  
           ‘He asked whoever in every place he saw (him).’

We are thus led to posit that in (8a), Wh1 targets a Topic position. The fact that the periphery of FRs (and indeed that of headed relatives—cf. (9b), adapted from Rudin 1986: 127)—can host a Topic is independently supported by the example in (9a):

- (9) a. Jam Marija kakvoto gotvi.  
           eat<sub>1SG</sub> Maria what cook<sub>3SG</sub>  
           ‘I eat what Maria cooks.’

- (9) b. Tova e ženata naj-složnite pesni kojato peeše.  
 this is woman<sub>DEF</sub> most complex<sub>DEF</sub> songs which sang<sub>3SG</sub>  
 ‘This is the woman who sang the most complex songs.’

A second argument in favor of our proposal comes from the examination of non-topicalizable constituents. As we will discuss in the next section, a general restriction on Topics (and on the elements occupying the designated Topic position in Bulgarian) is that they must be construed with respect to a salient entity in the discourse. In this, true adjuncts (manner, reason) such as *100 pounds* or *how many* contrast with arguments and quasi-arguments (such as locatives and temporal adjuncts), since they do not presuppose a set of entities but quantify over amounts, a distinction that has been expressed in terms of referentiality, or “referential index” (see among others Aoun et al. 1987, Comorovski 1989, Cinque 1990, Rizzi 1990).

In Bulgarian, the Topic position cannot be occupied by a nonreferential phrase such as *sto kilograma* in (10):

- (10) \*Mislja sto kilograma če Ivan teži.  
 think<sub>1SG</sub> one hundred kilograms that Ivan weigh<sub>3SG</sub>  
 [Intended] ‘I think Ivan weighs one hundred kilograms.’

Similarly, a non-referential (manner) *wh*-phrase cannot occur in the first position within a PMFR, as the contrast in (11) shows:

- (11) a. Jam sushi kâde(to) kakto go napravjat.  
 eat<sub>1SG</sub> sushi where how it<sub>CL</sub> make<sub>3PL</sub>  
 ‘I eat sushi at each place however they make it.’  
 b. \*Jam sushi kakto kâdeto go napravjat.  
 eat<sub>1SG</sub> sushi how where it<sub>CL</sub> make<sub>3PL</sub>

This supports the view that the first *wh*-phrase in PMFRs is indeed a Topicalized constituent. Before turning to the structure of FRs, we must make clear in just what sense the first *wh* constituent in PMFRs is interpreted not only as a Topic but also as a universally quantified expression (as the English translations indicate).

#### 4. What’s in a (Wh-)Topic?

The proposal according to which a *wh*-phrase can be a Topic is not new. It has been put forth in the context of multiple *wh* interrogatives in various lan-

guages, including Bulgarian (see, e.g., Jaeger 2004, Krapova and Cinque 2008b, Dimova 2011).

The idea of a topicalized *wh*-phrase appears at first sight to run afoul of proposals which view *wh*-phrases as Foci, i.e., representing new information (or requests for new information in the case of questions; see, e.g., Rizzi 1997, Bošković 1998). This apparent contradiction is addressed by Jaeger (2004), who concludes that the topic of a question is what the sentence primarily requests information about, where “aboutness” refers to discourse saliency. As Jaeger discusses, the sentence in (12) is felicitous only in a context where the set of persons having been painted is salient in the discourse:

- (12) Kogo koj go e narisuval?  
 whom who him<sub>CL</sub> AUX<sub>3SG</sub> painted  
 ‘Who has been painted by whom?’

More generally, we understand “topicality” as referring to a relation (attribute-of, member-of, etc., see Ward and Prince 1991) between a constituent and an entity which is salient (although not necessarily previously mentioned) in the discourse, a notion that appears equivalent to Pesetsky’s (1987) D-linking. The topicality of a constituent is independent of its morphological makeup. In the case of a *wh*-phrase, it is independent of whether a restriction is expressed, as in intrinsically D-linked *which book*, or not, as in *what*.

Kiss (1993: 99) has pointed out that the wider scope *Wh*-operator in multiple *wh*-questions is interpreted as a “distributive universal quantifier, a synonym of *each*.” As Krifka (2001) discusses, quantifiers based on *each* presuppose a set of discourse-salient entities, hence are naturally topical. This lends credence to the idea that the first *wh*-phrase in PMFRs is indeed in a topic position. Quantified DPs with ‘each’ can also occur in Topic positions in Bulgarian as shown in (13):

- (13) Vsjaka ot statiite sâm ja čela dva pàti.  
 each of articles<sub>DEF</sub> AUX<sub>1SG</sub> it<sub>CL</sub> read two times  
 ‘As for each of the articles, I have read it twice.’

This is exactly the reading that the first *wh*-constituent also has in Bulgarian PMFRs. This is true in Rudin’s example in (1), and in all our examples.

Before sketching an analysis of Bulgarian PMFRs, we will briefly discuss the role of the particle *-to* in Bulgarian.

## 5. What is *-to*?

One striking fact about Bulgarian concerns the occurrence of the particle *-to*, which can be affixed to *wh*-words in certain contexts. This particle is precluded in *wh*-questions, but it is obligatory in all relatives (headed, free, correlatives), as well as equatives and comparatives.<sup>5</sup> Historically, invariant *-to* originates from a demonstrative, just as the definite article does. It has been analyzed as a definiteness marker/article, a modal element, a relative complementizer, or a nominalizer.<sup>6</sup> Here we take the view that *-to* is affixed to a *wh*-word in the lexicon and that it is not definite; rather, its function is to turn a *wh*-word with interrogative content into a non-interrogative operator. Our contention is that Bulgarian overtly marks a property shared by relatives in other languages, namely that the *wh*-phrase is non-interrogative.

We now briefly relate obligatory *-to* and the categorial specification or labeling of the relative itself. Consider first Chomsky's (2013: 45) labeling algorithm, stated in (14):

- (14) A syntactic object (SO) formed by merger of XP and YP can be labeled in one of two ways:
- a. SO is modified so that there is only one visible head (one of the constituents must move); or
  - b. XP and YP are identical in a relevant respect, providing the same label, which can be taken as the label of SO.

Case (a) is illustrated by successive cyclic movement: a *wh*-phrase merged to a CP with a non-*wh* C head yields a SO that cannot be labeled. Hence, the *wh*-phrase has to move (a discontinuous constituent—or the copy of the moved *wh*-phrase—is invisible for labeling). Hence the SO is labeled CP.

Case (b) is instantiated by indirect questions. In this case, the *wh*-phrase and the C head share a prominent feature, [Q], which can be taken as the label of the SO.

Returning to *-to*, let us suppose that it endows the *wh*-word with a [REL] feature. As mentioned above, this signals that the *wh*-word is non-interrogative. Suppose further that the [REL] feature must be checked with a C head bearing the same feature. What could the [REL] feature on a C head be? It could represent one of the possible values for C: [Q] for interrogatives, [DECL] for declaratives, and [REL] for clausal constituents that must obligatorily enter

<sup>5</sup> The particle *-to* is optional on *wh*-topics, where it appears to convey a distributional meaning. We leave this usage aside, concentrating on obligatory *-to*, which we take to be a distinct particle.

<sup>6</sup> For an extensive discussion on *-to*, see Franks and Rudin (2015).

into a predicative relation with either a nominal head (for headed relatives) or a correlate (for correlatives). In other words, a  $C_{REL}$  heads a CP that must be a predicate, never an argument. When a *wh*-phrase with affixed *-to* merges with CP, the feature [REL] (prominent and shared) will be selected as the label of the syntactic object, in accordance with (14b). At the point of *wh*-merge, then, free and headed relatives are indistinguishable and are labeled [REL], as (15) illustrates:

(15) Merge Wh: [REL Wh-*to* [CP  $C_{REL}$ ]]

If, however, the constituent labeled [REL] is merged with V, a predication relation cannot obtain. Since [REL] has two values depending on the constituent which bears it (non-interrogative on *wh*-phrases, PRED on C), [REL] on *wh* is selected. This in turn will designate the *wh*-phrase as the head of the free relative and the complement of the verb. In other words, the label of the constituent is determined by the *wh*-phrase bearing both the prominent [REL] feature compatible with complement status and the categorial label consistent with the c-selectional requirements of the verb.

## 6. How are PMFRs Labeled?

Consider now the derivation of a PMFR such as (4b), repeated below as (16):

(16) Sâbraxme na kogo(to) kakvoto bjaxa izpratili.  
 gathered<sub>1PL</sub> to who what AUX<sub>3PL</sub> sent<sub>3PL</sub>  
 'We gathered whatever they had sent to each of us.'

First, the *wh* constituent *kakvoto* merges with CP. As this constituent bears the feature [REL], which is prominent and shared with  $C_{REL}$ , it is the label of the syntactic object (17a,b). Next, another *wh* constituent (with topic-like properties), here *na kogo(to)*, merges to the REL syntactic object, yielding the configuration in (17c):

- (17) a. kakvoto<sub>[REL]</sub> [ $C_{REL}$  bjaxa izpratili kakvoto na kogo(to)]  
 b. [<sub>REL</sub> kakvoto<sub>[REL]</sub>] [ $C_{REL}$  bjaxa izpratili kakvoto na kogo(to)]  
 c. na kogo(to) [<sub>REL</sub> kakvoto bjaxa izpratili kakvoto na kogo(to)]

Since there is no prominent feature shared by the *wh*-phrase *na kogo(to)* and the REL constituent, the resulting syntactic object cannot be labeled by (14b). The other labeling possibility afforded by the algorithm is that stated in (14a), i.e., movement of *na kogo(to)* to a different position. Such movement is possible (and necessary) because the constituent is a Topic. Recall that, as is cus-

tomarily assumed in the minimalist literature, “C is shorthand for the region that Rizzi (1997) calls the left periphery” (Chomsky 2008: 143). Presumably, *na kogo(to)* bears a [TOP] feature: because of this, it must move further up the left periphery, yielding the configuration in (18):

(18) *na kogo(to) na kogo(to)* [REL *kakvoto* *bjaxa izpratili kakvoto na kogo (to)*]

As per (14b), there is now only one visible constituent, REL, which determines the label of the syntactic object in (18). This explains why Topics are transparent for the purposes of c-selection, in PMFRs and generally: since they can (and must) move upwards in the periphery, the constituent they merge with always retains its own categorial label.

## 7. Conclusion

Taking as our point of departure Rudin’s findings on multiple *wh* relatives in Bulgarian, we have focused more closely on the properties of both *wh*-phrases in these constructions. We have proposed that the first *wh*-constituent functions as a Topic, while the second *wh*-constituent is the one that is selected by both the main verb and the embedded verb. This has led us to the view that multiple *wh*-relatives are indeed as Rudin (1986) had first concluded free relatives, however not multiple free relatives in the strict sense. Based on the morphological properties of *wh*-constituents in Bulgarian relative clauses (specifically, the obligatory affixation of the particle *-to*), we have sketched an analysis of free relative formation and labeling which also accounts for the fact that an intervening Topic *wh*-constituent is transparent to selection by the main verb.

Université de Montréal  
 elena.dimova@umontreal.ca  
 christine.tellier@umontreal.ca

## References

- Andrews, Avery. (1975) *Studies in the syntax of relative and comparative clauses*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Aoun, Joseph, Norbert Hornstein, David Lightfoot, and Amy Weinberg. (1987) “Two types of locality”. *Linguistic inquiry* 18(4): 537–77.
- Arnaudova, Olga. (2002) “Clitic left dislocation and argument structure in Bulgarian”. Jindřich Toman, ed. *Formal approaches to Slavic linguistics: The second Ann Arbor meeting, 2001*. Ann Arbor, MI: Michigan Slavic Publications, 23–46.



- Baltin, Mark. (1982) "A landing site theory of movement rules". *Linguistic inquiry* 13(1): 1–38.
- Bošković, Željko. (1998) "Wh-phrases and wh-movement in Slavic". *Workshop on comparative Slavic morphosyntax*, Indiana University, Bloomington.
- Bresnan, Joan, and Jane Grimshaw. (1978) "The syntax of free relatives in English". *Linguistic inquiry* 9(3): 331–91.
- Chomsky, Noam. (2008) "On phases". Robert Freidin, Carlos Otero, and Maria-Luisa Zubizarreta, eds. *Foundational issues in linguistic theory. Essays in honor of Jean-Roger Vergnaud*. Cambridge, MA: MIT Press, 133–66.
- . (2013) "Problems of projection". *Lingua* 130: 33–49.
- Cinque, Guglielmo. (1990) *Types of A'-dependencies*. Cambridge, MA: MIT Press.
- Citko, Barbara and Martina Gračanin-Yuksek. (2016) "Multiple (coordinated) (free) relatives". *Natural language and linguistic theory* 34(2): 393–427.
- Comorovski, Ileana. (1989) "Discourse-linking and the Wh-island constraint". *NELS 19*, Amherst, MA: University of Massachusetts, Graduate Linguistic Student Association, 78–96.
- Dimova, Elena. (2011) "Spécificité et ordre des syntagmes Qu-multiples en Bulgare." *Actes du congrès annuel de l'Association canadienne de linguistique 2011*. University of New Brunswick, Fredericton.
- . (2014) "A new look at multiple free relatives: Evidence from Bulgarian". Paper presented at the 9th meeting of the Slavic Linguistics Society (SLS 9), University of Washington, Seattle, 19–21 September.
- Dimova, Elena and Christine Tellier. (2015) "On multiple free relatives that are not". Paper presented at FDSL 11, University of Potsdam, Germany, 2–4 December.
- Franks, Steven L. and Catherine Rudin. (2005) "Bulgarian clitics as K<sup>0</sup> heads". Steven L. Franks, Frank Y. Gladney, and Mila Tasseva-Kurkchieva, eds. *Formal approaches to Slavic linguistics: The South Carolina meeting, 2004*. Ann Arbor, MI: Michigan Slavic Publishers, 104–16.
- Franks, Steven L. and Catherine Rudin. (2015) "Invariant *-to* in Bulgarian". Miriam Shrager, Edna Andrews, George Fowler, and Steven L. Franks, eds. *Studies in accentology and Slavic linguistics in honor of Ronald F. Feldstein*. Bloomington, IN: Slavica Publishers, 99–136.
- Grimshaw, Jane. (1977) *English Wh-constructions and the theory of grammar*. Ph.D. dissertation, University of Massachusetts.
- Groos, Anneke and Henk van Riemsdijk. (1981) "Matching effects in free relatives: A parameter of core grammar". Adriana Belletti, Luciana Brandi, and Luigi Rizzi, eds. *Theory of markedness in generative grammar*. Pisa: Scuola Normale Superiore, 171–216.
- Hornstein, Norbert. (2009) *A theory of syntax*. Cambridge: Cambridge University Press.
- Hornstein, Norbert and Jairo Nunes. (2008) "Adjunction, labeling, and bare phrase structure". *Biolinguistics* 2(1): 57–86.

- Izvorski, Roumyana. (1996) "The syntax and semantics of correlative pro-forms". Kiyomi Kusumo, ed. *NELS* 26. Amherst, MA: University of Massachusetts, Graduate Linguistic Student Association, 133–47.
- Jaeger, Florian T. (2004) "Topicality and superiority in Bulgarian Wh-questions". Olga Arnaudova, Wayles Browne, Maria-Luisa Rivero, and Danijela Stojanović, eds. *Formal approaches to Slavic linguistics: The Ottawa meeting, 2003*. Ann Arbor, MI: Michigan Slavic Publications, 207–28.
- Kiss, Katalin É. (1993) "Wh-movement and specificity". *Natural language and linguistic theory* 11: 85–120.
- Krapova, Iliyana. (2002) "On the left periphery of the Bulgarian sentence". *University of Venice Working Papers in Linguistics* 12: 107–28.
- Krapova, Iliyana and Guglielmo Cinque. (2008a) "Clitic reduplication constructions in Bulgarian". Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins Publishing Company, 257–87.
- . (2008b) "On the order of Wh-phrases in Bulgarian multiple Wh-fronting". Gerhild Zybatow, Luka Szucsich, Uwe Junghanns, and Roland Meyer, eds. *Formal description of Slavic languages: The fifth conference, Leipzig 2003*. Frankfurt am Mein: Peter Lang, 318–36.
- Krifka, Manfred. (2001) "Quantifying into question acts". *Natural language semantics* 9: 1–40.
- Lasnik, Howard and Mamoru Saito. (1992) *Move  $\alpha$ : Conditions on its applications and output*. Cambridge, MA: MIT Press.
- Pesetsky, David. (1987) "Wh-in-situ: Movement and unselective binding". Eric Reuland and Alice ter Meulen, eds. *The representation of (in)definiteness*. Cambridge, MA: MIT Press, 98–129.
- Rizzi, Luigi. (1990) *Relativized minimality*. Cambridge, MA: MIT Press.
- . (1997) "The fine structure of the left periphery". Liliane Haegeman, ed. *Elements of grammar: Handbook of generative syntax*. Dordrecht: Kluwer Academic Publishers, 281–337.
- Rudin, Catherine. (1986) *Aspects of Bulgarian syntax: Complementizers and Wh constructions*. Columbus, OH: Slavica Publishers.
- . (1988) "On multiple questions and multiple Wh fronting." *Natural language and linguistic theory* 6: 445–501.
- . (2006) "Multiple Wh fronting, free relatives, correlatives, and the [ $\pm$ MFS] split". IULC lecture, Indiana University, Bloomington, Indiana.
- . (2007a) "Multiple Wh relatives in Slavic". Richard Compton, Magdalena Goledzinowska, and Ulyana Savchenko, eds. *Formal approaches to Slavic linguistics: The Toronto meeting, 2006*. Ann Arbor, MI: Michigan Slavic Publications, 282–307.
- . (2007b) "Multiple Wh fronting in correlatives and free relatives." Paper presented at Linguistic Society of America annual meeting, Anaheim, CA, 4–7 January.

- Rudin, Catherine. (2008a) "Pair-list vs. single pair readings in multiple Wh free relatives and correlatives". *Kansas Working Papers in Linguistics* 30: 257–67.
- . (2008b) "Multiple wh non-interrogatives." Paper presented at University of Wisconsin and University of Minnesota linguistics colloquia.
- . (2009) "The Bulgarian relative marker *-to*". Steven L. Franks, Vrinda Chidambaram, and Brian D. Joseph, eds. *A linguist's linguist: Studies in South Slavic linguistics in honor of E. Wayles Browne*. Bloomington, IN: Slavica Publishers, 403–22.
- . (2012) "Multiple Wh free relatives and correlatives". Talk presented at Slavic Linguistic Society 2012 summer workshop, Indiana University, Bloomington, IN, 14–17 May.
- Shlonsky, Ur. (2006) "Projection étendue et cartographie de SC". *Nouveaux cahiers de linguistique française* 27: 83–93.
- Wali, Kashi. (1982) "Marathi correlatives: A conspectus". *South Asian review* 6(3): 78–88.
- Ward, Gregory L. and Ellen F. Prince. (1991) "On the topicalization of indefinite NPs". *Journal of pragmatics* 16: 167–77.

# The Pluperfect in Bulgarian and Macedonian: From *Bai Ganyo* to the *Bombi*

Victor A. Friedman

*Abstract:* Macedonian differs from Bulgarian in the use of the pluperfect. This is in part due to the rise in Macedonian of a new pluperfect using 'have', which even sixty years ago had already supplanted the inherited pluperfect in 'be' in resultative contexts. In the course of the past half-century or so, however, the inherited 'be' pluperfect in Macedonian has retreated to the point that it is virtually obsolete in spoken Macedonian, as illustrated by the 2015 *Bombi*. As the translation of the Bulgarian novel *Bai Ganyo* into Macedonian shows, also relevant is the Macedonian elimination of the imperfective aorist and its replacement with either a perfective aorist or an imperfective imperfect. This alignment has contributed to taxis being inferred from context rather than marked by an explicit form. Another change has been the Macedonian use of the perfective imperfect as the irrealis mood of choice, replacing modal uses of the pluperfect. Overall, the development of the 'have' series has complicated the Macedonian verbal system vis-à-vis Bulgarian, but at the same time other parts of the Macedonian system have become simplified. In the matter of pluperfects, Macedonian has diverged from Bulgarian in ways that bring Macedonian closer to its neighboring Balkan languages in certain respects.

## 1. Introduction

In recent work Rudin (2015) has discussed important differences between Macedonian and Bulgarian morphosyntax.<sup>1</sup> In earlier work (e.g., Rudin et al. 1999, Rudin 2001), she likewise compared Bulgarian with Macedonian and other Balkan languages. In this article honoring her many years of fruitful, insightful, and accessible scholarship, I discuss another heretofore overlooked difference between Macedonian and Bulgarian: the functioning of the pluperfect. The basic corpus will be translations of Konstantinov (1895), the Bulgarian novel *Bai Ganyo*. In addition to providing numerous diagnostic examples, the text is especially appropriate for this Festschrift, since Catherine Rudin was part of the team that made an award-winning translation into English

---

<sup>1</sup> See also Friedman (2015), which would not have been written were it not for Rudin's stimulating work.

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin*. Bloomington, IN: Slavica Publishers, 2018, 121–30.

(Konstantinov 2010). At the end, I briefly discuss the situation in modern colloquial Macedonian as illustrated by the *Bombi* (2015), a corpus of wire-tapped conversations, the details of which are discussed below. What emerges from this is that the inherited pluperfect in 'be' is becoming increasingly obsolete in Macedonian, not only owing to competition with the new pluperfect in 'have', but also because taxis and mood are being expressed by other means. This can be connected to the Macedonian elimination of the imperfective aorist (Friedman 1993), as discussed in the conclusion. In terms of the pluperfect, therefore, Macedonian and Bulgarian, which began diverging in this respect no later than the fifteenth century, are continuing to diverge.

## 2. Pluperfects in Bulgarian and Macedonian

The starting point for the comparison is Old Church Slavonic (OCS), which in this respect can be taken to represent the original situation that diverged in the two modern languages. Basically, the OCS pluperfect was formed with the imperfective preterite (usually imperfect but also aorist—3sg *běše* vs. *bě*) of 'be' and the resultative participle in *-l*, which was formed only from the aorist (or infinitive) stem.<sup>2</sup> The perfective aorist of 'be' was already in competition with the old optative of 'be' (3sg *by* vs. *bi*) for forming the conditionals.<sup>3</sup>

Jumping ahead a thousand years or so, the inventories of Macedonian and Bulgarian have developed differently, and, moreover, the deployment of shared forms is different. It is this latter difference that is particularly interesting, since it points to underlying systemic differences between the two languages that have developed over time, and which, in fact, are continuing to develop. In terms of inventories, both languages developed an imperfect resultative participle (which is no longer a participle in Macedonian; for convenience the term *l-form* is used for both languages henceforth).<sup>4</sup> In Macedonian, both the aorist and the imperfect *l*-forms can be used with the imperfect of

<sup>2</sup> One of the many differences between Bulgarian and Macedonian in their pluperfects is the fact that Bulgarian still has *be* (OCS *bě*) as an optional choice for the 3sg auxiliary, but Macedonian does not.

<sup>3</sup> This competition was eliminated by the merger of /i/ and /y/ to /i/ in South Slavic in general. A further difference between Macedonian and Bulgarian is that the auxiliary continues to be conjugated in Bulgarian whereas in Macedonian the old 3sg *bi* is now an invariant participle.

<sup>4</sup> Torlak, some west Bulgarian, and some southeast Macedonian dialects never developed the imperfect *l*-form (see Friedman 2014: 112–13 for details). For purposes here, the standard languages are the main focus, and the dialects on which they are based represent significantly divergent systems within Balkan Slavic. The old resultative participle still has some participial uses in modern Bulgarian, but they are not of concern here.

'be' to form pluperfects, whereas in standard Bulgarian, only the aorist participle is used.<sup>5</sup> Unlike Bulgarian, Macedonian has developed a set of analytic perfect and pluperfect paradigms using the auxiliary *ima* 'have' in the present, imperfect, and unmarked past (*l*-past) plus the neuter verbal adjective (see Friedman 2014 for discussion). On the other hand, around the fifteenth century, Bulgarian began to use the perfect of 'be' (3sgM *bil e*) as an auxiliary for forming a pluperfect that competed with the old pluperfect (Dejanova 1970: 28–29). It is worth noting here that this use of the perfect of 'be' as a pluperfect auxiliary is attested in medieval Serbian/Croatian a century earlier than in Bulgarian (Dejanova 1970: 64–65), and it is also found elsewhere in Slavic (Dejanova 1970: 100). For the purposes of this article, it is the inherited pluperfect and, for Macedonian, the corresponding pluperfects with an imperfect *l*-form and the imperfect of *ima* 'have' (3sg *imaše*)+neuter verbal adjective that constitute the focus of comparison.<sup>6</sup>

It is fair to say that at this point in time (2018), Macedonian is well on the way to the complete obsolescence of the *beše* pluperfect. Speakers of the youngest generation in Skopje do not use it, and some do not even recognize it. Moreover, it is not necessarily being replaced by the *imaše* pluperfect, although this is sometimes the case. Frequently, however, taxic relationships are carried by context and other preterite forms. Meanwhile in Bulgarian inherited pluperfects are very much in use both in speech and in writing.

### 3. Pluperfects in *Baj Ganyo*

A comparison of a Macedonian translation of *Baj Ganjo*, made 50 years ago (Konstantinov 1967), is instructive as an illustration. Even though the corpus is relatively small, the differences are significant to the point of being diagnostic. The translator, Gjorgji Caca (1920–2006), came from a distinguished family in Bitola with ties to Kruševo, and thus the *ima* series was native to his Macedonian. Nonetheless, there is not a single example of an *ima* perfect or pluperfect in his translation of *Baj Ganjo*. Caca's translation shows a clear grasp of stylistic nuances, e.g., he uses the Macedonian southwesternism *zapuli* for Konstantinov's *zazjapa* 'stare at for no particular reason', and Caca, although by profession a constitutional judge and professor of law (e.g., Caca 1974), was also the founder of the Bitola National Theater.<sup>7</sup> Thus, there

<sup>5</sup> Normative Bulgarian grammar concerning the imperfect *l*-form does not correspond to actual usage, but for this article the normative formulation suffices (see Friedman 1986 and 2002 for discussion).

<sup>6</sup> The vexed question of evidential usages is beyond the focus of this article. See Friedman (2002) and references therein.

<sup>7</sup> <<http://www.pravdiko.mk/janko-tsatsa-1881-1968/#more-7659>> accessed 27 December 2017.

is no question regarding his language sensibilities, rather, the translation represents what could be taken for a conservative norm of his time, which could also have been influenced by the original Bulgarian. It is for these reasons that, although Caca's use of the Macedonian *beše* pluperfect agrees with Konstantinov's Bulgarian slightly more than half the time, the fact that almost half the time he makes other choices, without ever having recourse to the *ima* series, is indicative of the changes already underway in Macedonian at that time, and since then even more evident.

The bare statistics are the following: Konstantinov's Bulgarian original had a total of 35 pluperfects, and Caca's translation used the corresponding Macedonian pluperfect (i.e., with aorist stem *l*-form) in 19 instances and an imperfect *l*-form once.<sup>8</sup> Of the remainder, the Macedonian uses a simple preterite (either synthetic or unmarked [cf. Friedman 2014]) nine times and has an imperfect of 'be' plus verbal adjective in six cases. There is also one example where Bulgarian uses 'be' plus past passive participle where Caca used a *beše* pluperfect.<sup>9</sup> The instances where the two languages agree need not be illustrated, as it is the departures that are indicative of what have become the trends of the current Macedonian norm. The following examples, therefore, which illustrate Caca's departures from Konstantinov's Bulgarian, also illustrate the tendencies that have become ever more pronounced in the 50 years since his translation was published. Examples (1–7) illustrate some of the basic differences between Bulgarian and Macedonian for pluperfect usage. The English translations are from Konstantinov (2010). The Bulgarian original is given first, and the Macedonian translation second.<sup>10</sup>

Example (1) illustrates an imperfect *l*-form in Macedonian, which would be impossible in Bulgarian. The choice of an imperfect *l*-form in this context also indicates a difference between Bulgarian and Macedonian in that the Macedonian chooses to focus on a durative or iterative aspect of the original packing of the provisions.

<sup>8</sup> I should note here that Caca, like most translators of *Bai Ganyo*, translated the 1895 edition published in Konstantinov's lifetime, which consisted of 12 chapters. Six more chapters that Konstantinov, who was murdered in 1897, did not live to add to a second edition were added in Čipev's 1929 edition (Konstantinov 1929), so the figures here are only for the original 1895 work.

<sup>9</sup> There is also one instance where both Bulgarian and Macedonian use imperfect 'be' plus past passive participle/verbal adjective.

<sup>10</sup> Owing to the length of the examples, I have used boldface for the relevant forms rather than interlinear glossing. Since page numbers vary from edition to edition, I have supplied the references in the form of chapter number and sentence number, e.g., 5.14 = the 14th sentence in Chapter 5.

- (1) [B] Entuziazmât ni rasteše proporcionalno s izprazvaneto na šišetata i košnicite s provizija, s koito v izobilie **bjaxa se snabdili** praktičnite turisti. (5.14)

[M] Našeto vooduševuvanje rasteše proporcionalno so praznenjeto na šišinjata i košnicite so hrana, so koe izobilno **se bea snabdele** praktičnite turisti. (5.14)

‘Our enthusiasm grew in direct proportion to the emptying of the abundant wine bottles and food baskets that our practical travelers **had packed.**’

Example (2) shows a Macedonian unmarked past for the Bulgarian pluperfect. In this context, the narrator is feigning inference, but it is nonetheless striking that Bulgarian uses the inherited pluperfect rather than the newer form with *bil* as the auxiliary. As Friedman 1986, 2014 argues, the auxiliary *bil* signals the marked nonconfirmative, while the inherited pluperfect is neutral in this respect. In the Macedonian, the unmarked aorist carries the nonconfirmative nuance contextually, but also the taxis is contextual.

- (2) [B] Baj Ganjo ot bârzina, vidi se, **beše zabravil** da turi v disagite zakuska. (5.85)

[M] Ganjo, se gleda, od brzina **zaboravil** da stavi vo disagite jadenje. (5.85)

‘Ganyo **had**, in his haste, evidently **forgotten** to pack any food for the trip in his *disagi*.<sup>11</sup>

Example (3) is similar to example (2), except that here the narrator has chosen to present himself as omniscient, and the Macedonian uses a confirmative aorist. The Macedonian also does not contain the adverb ‘already’, but in both versions the imperfect *prevâznasjaše / vzdigaše* ‘was singing the praises of’ is taking place while Baj Ganjo takes off his *anterija* as well as during the subsequent action of seeking (lice).

- (3) [B] Dodeto v ednata staja studentât prevâznasjaše bogatstvata na baj Ganja, v sâsednata staja baj Ganjo **beše veče sâbljakâl** anterijata si i târseše v neja nešto s goljamo vnimanie, kato mâmoreše pod nosa si: (7.98)

<sup>11</sup> *Disagi*: a kind of saddlebag made from a rectangular piece of colorful, heavy, woven cloth. The ends are folded and sewn at each side to create two sacks that can be slung over the back of a pack animal or the shoulder of a person. It is clear from the context that the narrator is being facetious about Bai Ganyo’s having forgotten.



[M] Duri vo ednata soba studentot gi vozdigashe bogatstvata na baj Ganjo, vo drugata soba baj Ganjo ja **sleče** anterijata i baraše po nea nešto so golemo vnanie, mrmorejki pod nos: (798)

‘While in one room the student was singing the praises of Bai Ganyo’s riches, in the neighboring room Bai Ganyo **had** already **taken off** his padded jacket and was rooting around very carefully in it for something, muttering under his breath.’

Example (4) resembles example (2) in the use of an unmarked aorist to refer to an anterior action, but in this case it is also interesting to note the differences in aspect choice. The Bulgarian is an imperfective in *-a*, which could in principle, be either aorist or imperfect. An imperfective aorist in this context would present the action as repeated completed (i.e., the subject successfully avoided paying his rent on multiple occasions) or as an action for which the end point was not the focus. In Macedonian, however, the imperfective aorist was already more or less obsolete (cf. Friedman 1993), and thus the choice here would be imperfective imperfect or perfective aorist. Thus, in the Bulgarian the not paying is presented as an extended action or a repeated completed action, followed by the subject’s (Bodkov’s) success in borrowing money from his landlady despite not having paid rent. In the Macedonian, however, the nonpayment is in sequence with and is followed by the subject’s success in getting the money. In the context of the narration, the source of the information for both the (non)payment and the success is the omniscient narrator. Thus, reportedness is not the issue; rather, this is a remnant of the original function of the resultative participle, which is still part of its meaning in Macedonian (see Friedman 2014).

(4) [B] Toj uspjia da zaemne ot xazjajkata—na kojato ne **beše plaštal** dosta otdavna naema za stajata—deset guldena. (797)

[M] Toj uspea da zeme na zaem od stopankata—na koja ne **platil** dosta odamna kirija za sobata—deset guldeni. (797)

‘He managed to borrow from the landlady—to whom he **had** not **paid** his rent for some time now—10 gold pieces.’

Example (5), on the opposing page, illustrates differences in the modal use of the pluperfect of the type described as expectative unfulfillable by Kramer (1985). In the Bulgarian, the protasis is an ‘if’ clause with a perfective pluperfect and the apodosis uses the inherited Common Slavic conditional in *bi* rather than the Balkan conditional of future+past marking. In the Macedonian translation, the protasis and apodosis both use perfective imperfects in a classic Balkan conditional construction. While the use of a pluperfect in the protasis is not ungrammatical in Macedonian, it simply no longer occurs. The

choice of old versus new conditional in the apodosis is another area where Macedonian and Bulgarian tend to differ.

- (5) [B] Ako v tozi moment **beše izbuxnal** vulkan na Vitoša, edva li **bi slisal** našite pâtnici poveče, otkolkoto тази сцена... (10.132)  
 [M] Vo toj moment vulkan da **buvneše** na Vitoša, ne **kje gi zbrložeše** našite patnici povekje od taa сцена... (10.132)  
 ‘If a volcano **had erupted** on Vitosha at that moment, it **would** hardly **have astounded** our travelers more than that scene..’

Example (6) illustrates another quasimodal difference between Bulgarian and Macedonian. In the construction meaning ‘as if’ (Bulgarian *kato če*, Macedonian *kako da*), Macedonian uses a *da*-clause (analytic subjunctive) and unlike all other uses of *da* this one can occur with an aorist, as seen in the Macedonian translation, whereas Bulgarian uses an indicative construction, and *da*+aorist is impossible.

- (6) [B] Toj kato če **beše poveril** celija svoj intelekt na okrâžaxštija go štab prijateli! (11.355)  
 [M] Toj kako da go **doveri** siot svoj intelekt na štabot prijateli što go opkružuvaa! (11.355)  
 ‘It was as if he **had entrusted** his entire intellect to the staff of friends surrounding him!’

Example (7) is a typical instance where an intransitive Bulgarian pluperfect is translated by a Macedonian imperfect of ‘be’ plus verbal adjective, a relatively frequent device in the Macedonian.

- (7) [B] klozetite **bjaxa se prevârnali** i v peračnici (5.208)  
 [M] nužnicite **bea pretvoreni** i vo peralni (5.208)  
 ‘the lavatories **had been turned** into laundries’

#### 4. Pluperfects in the *Bombi*

Before turning to the conclusions, it is instructive to adduce data from a modern colloquial Macedonian text, the *Bombi* (2015). This material contains almost 26 hours of illegally wire-tapped natural conversation totaling close to 130,000 words or almost 300 pages of text from over 100 speakers.<sup>12</sup> The

<sup>12</sup> The wire-taps were ordered by the ex-Prime Minister Nikola Gruevski and his first cousin, the ex-chief of the secret police Sašo Mijalkov, and affected 20,000 people

speakers were all connected in some way with the former ruling party (VMRO-DPMNE), and they reveal breathtaking levels of corruption. In fact, chapter 11 of *Bai Ganyo*, “Bai Ganyo Holds Elections,” a fictionalized account of real corruption in nineteenth-century Bulgaria, pales in comparison to the excesses documented in the *Bombi*. Most of the speakers in the *Bombi* were educated in Skopje. With regard to pluperfects, it is striking that they are extremely rare: 13 are *imaše* pluperfects and only one is a *beše* pluperfect, and that one is uttered by Silvana Boneva (born 1959), a speaker from Strumica in eastern Macedonia with a strong eastern accent. The *Bombi* thus bear out the observation that the inherited pluperfect is becoming obsolete in modern Macedonian, and, moreover, to the extent that a pluperfect is used, it is the ‘have’ type. The ‘have’ series originated in southwestern Macedonia and was already moving north and east before codification in 1944. However, the use of the imperfect and *l*-form auxiliaries progressed much more slowly than the use of the present auxiliary (Friedman 1988). Thus, the fact that the one example of a *beše* pluperfect is from a Strumica speaker who was already over 50 at the time the recording was made would be consistent with the current situation in Macedonia.

## 5. Conclusion

Among the many ways in which Macedonian differs from Bulgarian is the use of the pluperfect. This is in part due to the rise in Macedonian of a new pluperfect using ‘have’, which even sixty years ago had already supplanted the inherited pluperfect in ‘be’ in resultative contexts (Friedman 2014). In the course of the past half-century or so, however, the inherited ‘be’ pluperfect in Macedonian has retreated to the point that, while it can still be used in literature, it is virtually obsolete in spoken Macedonian, as seen in the statistics from the *Bombi* (2015). However, as the translation of *Bai Ganyo* into Macedonian (Konstantinov 1967) shows, at issue is not simply the new pluperfect in ‘have’ replacing the old pluperfect in ‘be’, although this is part of the story. Another factor is the fact that the Macedonian aspectual system has eliminated the imperfective aorist and replaced it with either a perfective aorist or an imperfective imperfect, resulting in uses of one or the other where Bulgarian would require an imperfective aorist. This alignment, in turn, has contributed to taxis being inferred from context rather than marked by an explicit form. Another change has been the use of the perfective imperfect as the irrealis mode of choice, replacing modal uses of the pluperfect. Overall,

---

in Macedonia, including the entire elite ruling circle of the ex-Prime minister himself. Only those wire-taps of Gruevski’s own ministers and their associates (including Gruevski and Mijalkov themselves) were released as the *Bombi*. See Friedman Forthcoming for additional details.

the development of the 'have' series has complicated the Macedonian verbal system vis-à-vis Bulgarian, but at the same time other parts of the Macedonian system have become simplified. Moreover, taking into account the fact that Bulgarian conjugates old imperfective aorist/optative (conditional) of 'be' as well as the auxiliary 'want' in past/modal forms, where Macedonian has invariant particles, and that in Macedonian the new 'have' series is beginning to enter into modal oppositions, it is fair to say that overall the Bulgarian and Macedonian systems have diverged in ways that bring Macedonian closer to its neighboring Balkan languages in certain respects.<sup>13</sup>

University of Chicago  
vfriedm@uchicago.edu

## References

- Bombi. (2015) <http://prizma.mk/kompleten-materijal-od-site-bombi-na-opozitsijata> (accessed 27 December 2017).
- Caca, Gjorgji. (1974) "The constitutional development of the S(ocialist) R(epublic of) Macedonia". Translated by Victor A. Friedman. *Macedonian review* 5(1): 76–91.
- Dejanova, Marija. (1970) *Istorija na složnite vremena v bâlgarski, sârbo-hârvoatski i slovenski ezik*. Sofia: Bâlgarska akademija na naukite.
- Friedman, Victor A. (1986) "Evidentiality in the Balkans: Macedonian, Bulgarian, and Albanian". Johanna Nichols and Wallace Chafe, eds. *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex, 168–87. [*Advances in Discourse Processes, Vol. 20.*]
- . (1988) "Za složenite minati vremenja vo dihovskiot govor vo sporedba so makedonskiot literaturnen jazik". Božidar Vidoeski, ed. *Jazičnite pojavi vo Bitola i bitolsko deneska i vo minatoto*. Skopje: Makedonska Akademija na Naukite i Umetnostite, 193–200.
- . (1993) "The loss of the imperfective aorist in Macedonian: Structural significance and Balkan context". Robert A. Maguire and Alan Timberlake, eds. *American contributions to the Eleventh International Congress of Slavists*. Columbus, OH: Slavica Publishers, 285–302.
- . (2002) "Hunting the elusive evidential: The third-person auxiliary as a boojum in Bulgarian". Victor A. Friedman and Donald L. Dyer, eds. *Of all the Slavs my favorites: Studies in honor of Howard I. Aronson on the occasion of his 66th Birthday*. Bloomington, IN: Slavica Publishers, 203–30. [*Indiana Slavic Studies, 12.*]

<sup>13</sup> The Bulgarian conjugation of the auxiliary 'be' in the *l*-form has an analogue in the Macedonian conjugation of the *l*-form of 'have' as an auxiliary.

- Friedman, Victor A. (2014) *The grammatical categories of the Macedonian indicative*, 2nd, revised edition. Bloomington, IN: Slavica Publishers.
- . (2015) "Sometimes a cigar is just a cigar: Bulgarian and Macedonian QUIPs and their relatives". Ljudmila Popović, Dojčil Vojvodić, and Motoki Nomachi, eds. *U prostoru lingvističke slavistike/ Inside the space of Slavic linguistics*. Belgrade: University of Belgrade, 768–76.
- . Forthcoming. "Turkisms in the Macedonian "Bombs" of 2015: On language as flag in the Balkans in the 21st century". Donald L. Dyer, Mary Allen Johnson, and Brian D. Joseph, eds. *Naylor Lecture 20th Anniversary Volume*, University of Mississippi: Balkanistica.
- Konstantinov, Aleko. (1895) *Baj Ganjo*. Sofia: Daskalov.
- . (1929) *Baj Ganjo*. Sofia: T. F. Čipev.
- . (1967) *Baj Ganjo*. Translated into Macedonian by Gjorgji Caca. Skopje: Kultura.
- . (2010) *Baj Ganyo*. Translated into English by Victor A. Friedman, Christina E. Kramer, Grace E. Fielder, and Catherine Rudin. Edited by Victor A. Friedman. Madison, WI: University of Wisconsin.
- Kramer, Christina E. (1985) *Analytic modality in Macedonian*. Munich: Otto Sagner. [*Slavistische Beiträge* 198.]
- Rudin, Catherine. (2001) "Clitic pronoun ordering in the Balkan languages". Victor A. Friedman and Donald L. Dyer, eds. *Of all the Slavs my favorites: Studies in honor of Howard I. Aronson on the occasion of his 66th birthday*. Bloomington, IN: Slavica Publishers, 1–19. [*Indiana Slavic Studies*, 12.]
- . (2015) "Whatever: Wh-Universal constructions in Macedonian and Bulgarian". Donald L. Dyer, Brian D. Joseph, and Christina E. Kramer, eds. *Od Čikago i nazad: Papers to Honor Victor A. Friedman on the occasion of his retirement*. University of Mississippi: Balkanistica, 409–33. [*Balkanistica* 28.]
- Rudin, Catherine, Christina Kramer, Loren Billings, and Matthew Baerman. (1999) "Macedonian and Bulgarian *li* questions: Beyond syntax". *Natural language and linguistic theory* 17(3): 541–86.

# Genitive/Dative Case Puzzles in the Bulgarian DP\*

Iliyana Krapova and Guglielmo Cinque

*Abstract:* The paper discusses issues in the grammar of Case marking in the DP by focusing on two interrelated puzzles in the syntax of Bulgarian nominalizations. The first puzzle concerns the ban on strong pronouns to act as DP-internal subjects. We argue that this is due to a morphological Case conflict, and we also discuss some historical considerations bearing on the loss of the genitive in Bulgarian. The second puzzle we discuss concerns DPs headed by an object nominal which show no traces of a Case conflict. The paper offers some considerations bearing on the genitive-dative Case syncretism in the history of Bulgarian.

## 0. Introduction

In this paper we discuss two interrelated puzzles in the syntax of Bulgarian Case, puzzles that find a straightforward solution once we consider the grammar of Case marking of DP-internal arguments from a comparative perspective.

The first puzzle concerns a curious difference between lexical and pronominal arguments in event/process nominal DPs with regard to their availability to act as DP-internal subjects. To approach this puzzle, in section 1 we discuss some preliminary facts about DP-internal Case marking of event/process nominals in Bulgarian. As in many other languages, these have the same structural make-up as their respective verbal correspondents. In section 2 we propose an account in terms of Case conflict, and in section 3 elaborate parallels with comparative phenomena from several other languages. Finally, in section 4, we introduce the second puzzle, which concerns clitic doubling in ordinary object nominals and the type of possessive Case these DPs realize. In

---

\* We dedicate this paper to Catherine Rudin, whom we admire very much as a linguist and as a person. We would like to thank Steven Franks for discussing various relevant issues with us and an anonymous reviewer for highly valuable comments, which have helped us improve the paper. For the requirements of the Italian academic system Iliyana Krapova takes responsibility for sections 1, 2, and 3 and Guglielmo Cinque for sections 4 and 5.

particular, we address the question why the subject of an object nominal, unlike the subject of an event/process nominal, should not cause a Case conflict.

## 1. Event/Process Nominals and the First Case Puzzle

In this section, we review the distribution of lexical and pronominal phrases as arguments within the DP. We show that while lexical phrases can perform the entire range of DP-internal syntactic functions, possessive clitics and possessive adjectives are limited to functioning as DP-subjects. Strong pronouns on the other hand can never be DP-subjects. This complementary distribution constitutes the first puzzle of the Bulgarian DP that we are going to address in section 2.

### 1.1. Full DP-Arguments: Distribution and Interpretation

DP-internal arguments in Bulgarian corresponding to a subject, a direct object, or an indirect object of an event/process nominal<sup>1</sup> (in the sense of Grimshaw 1990) are expressed by one and the same preposition/case marker, *na* 'of'. The examples in (1) and (2) below illustrate their distribution and interpretation. Thus, subject *na*-phrases may realize the Agent argument of nominals derived from transitive and unergative verbs, as in (1a, b), or the Theme argument of nominals derived from unaccusative and passive verbs, as in (1c) and (1d). As (1a) also shows, *na*-phrases may also realize the Theme of a transitive nominal. Since DP-internal subjects and objects do not appear to have a fixed order,<sup>2</sup> (1a) is ambiguous between an Agent > Theme and a Theme > Agent reading:

- (1) a. opisanieto      na Ivan na Petko      [transitive N]  
 description<sub>DEF</sub> of Ivan of Petko  
 'Ivan's description of Petko/Petko's description of Ivan'
- b. laeneto      **na sâsedskoto**      **kuče**      [unergative N]  
 barking<sub>DEF</sub> of neighbor-Adj<sub>DEF</sub> dog  
 'the barking of the neighbor's dog'

<sup>1</sup> We do not discuss the internal composition of deverbal Ns or their distribution into structural types according to nominalizing suffix (e.g., *-ne*, *-nie*, *-cija*). For more details, the reader is referred to Markova 2007, Krapova and Cinque 2013, and references cited therein.

<sup>2</sup> Although we do not discuss here the issue of word order within the DP, we note that while both orders S>O and O>S are possible, the latter seems preferred.

- (1) c. pristiganeto **na srednoštija gost** [unaccusative N]  
 arrival<sub>DEF</sub> of midnight-Adj<sub>DEF</sub> visitor  
 ‘the arrival of the midnight visitor’
- d. arestuvaneto **na mafjotskija bos** [passive N]  
 arrest<sub>DEF</sub> of mafia-Adj<sub>DEF</sub> boss  
 ‘the arresting of the mafia boss’

Passive nominals are a subset of nominals ending in *-ne* (such as *arestuvane* ‘arresting’ in (1d), *predstavjane* ‘presentation’, *obsâždane* ‘discussion’, etc.). To introduce the Agent, they take a different preposition, namely *ot* ‘by’, the same as the one used in passive clauses. Compare (2a) and (2b) and see Cinque (1980), Krapova and Cinque (2013) for details and discussion:

- (2) a. arestuvaneto **na Toto Rina** (ot/\*na policijata)  
 arrest<sub>DEF</sub> of Toto Rina (by/\*of police<sub>DEF</sub>)  
 ‘the arresting of Toto Rina (by the police)’
- b. Toto Rina e arestuvan ot/\*na policijata.  
 Toto Rina is arrested by/\*of police<sub>DEF</sub>  
 ‘Toto Rina has been arrested by the police.’

The fact that passives (optionally) take an *ot* ‘by’-Agent is a useful diagnostic for distinguishing passive nominals from transitive ones also in cases involving multiple argument realization. See, e.g., (3) where the Theme is promoted to a DP-subject while the second *na*-phrase realizes the Goal (the DP-indirect object).<sup>3</sup>

- (3) *Theme > Goal*  
 predstavjaneto na Petko **na gostite** (ot domakina)  
 presentation<sub>DEF</sub> of Petko to guests<sub>DEF</sub> (by host<sub>DEF</sub>)  
 ‘Petko’s presentation to the guests’/  
 \*‘the presentation of the guests to Petko (by the host)’

Thus, a sequence of two *na*-phrases does not necessarily imply that we are dealing with a transitive configuration.

<sup>3</sup> An anonymous reviewer asks whether transitive nominals can also appear with a second *na*-phrase corresponding to a Goal argument. The answer is negative, as only passive nominals in Bulgarian can combine with dative arguments. Apart from Goals, Locative arguments are also possible with some derived nouns, e.g., *stoeneto na plaža* ‘staying on the beach’.



## 1.2. Possessive Clitics and Possessive Adjectives

DP-internal arguments can also be realized by possessive clitics or pronominal adjectives.<sup>4</sup> They show parallel syntactic behavior, as illustrated by the examples in (4–8). Thus, both *mu* ‘him<sub>DAT</sub>’ and *negov* ‘his’ may refer to Agents of transitive and unergative nominals (5–6), as well as to Themes of unaccusatives and passive ones (Tasseva-Kurktchieva 2004).<sup>5</sup> Importantly, pronominal possessives necessarily realize the DP-subject, as shown by (4), which unlike (1a) above is no longer ambiguous, as well as by (8), where neither possessive can refer to the Goal (indirect object) argument.<sup>6</sup> The parallel behavior of these possessive elements raises the question of Case: what type of abstract Case does the possessive clitic realize given that it bears dative morphology? We return to this issue in section 2.3.

- (4) a. opisanieto      **mu**      na Petko      [transitive Ns]  
 description<sub>DEF</sub> him<sub>CL</sub> of Petko  
 ‘his description of Petko’/ \*‘Petko’s description of him’

<sup>4</sup> A well-known property of possessive clitics in Bulgarian is that they appear strictly right-adjacent to the noun or the adjective bearing the definite article (Franks 2001: 59ff, a.o.), while pronominal possessive adjectives bear the definite article themselves and agree with the head noun in number and gender.

<sup>5</sup> Possessive clitics cannot substitute for a directional *na*-phrase PP either. This is seen in the contrast in (i):

- (i) a. kačvaneto na masite      e zabraneno  
 getting<sub>DEF</sub> on tables<sub>DEF</sub> is forbidden  
 b. \*kačvaneto im      e zabraneno  
 getting<sub>DEF</sub> to-them<sub>CL</sub> is forbidden

<sup>6</sup> As discussed in Franks and King (2000: 276f), possessive clitics “can never correspond to true datives.” That is, in (i), although the base verbs (*vlijaja* ‘influence’, *objasnavam* ‘explain’) from which these deverbal nouns are derived take dative complements, expressible as dative clitics or full *na*-phrases, the clitic *mu* cannot be interpreted in this function:

- (i) a. vlijanieto      mu  
 influence<sub>DEF</sub> him<sub>CL.DAT</sub>  
 ‘his influence’/ \*‘the influence on him’  
 b. objasnenieto      mu  
 explanation<sub>DEF</sub> him<sub>CL.DAT</sub>  
 ‘his explanation’/ ‘its explanation’/ \*‘the explanation to him’

This, it seems to us, would be hard to understand if the possessive clitic were a dative clitic. It can only be made sense of if it is a genitive clitic, standing in every case for the subject of the DP. For additional discussion see Krapova and Cinque (2013).

- (4) b. **negovoto** opisanie na Petko  
 his<sub>DEF</sub> description of Petko  
 ‘his description of Petko/ \*Petko’s description of him’
- (5) a. laeneto **mu** [unergative Ns]  
 barking<sub>DEF</sub> him<sub>CL</sub>  
 ‘his barking’  
 b. **negovoto** laene  
 his<sub>DEF</sub> barking  
 ‘his barking’
- (6) a. pristiganeto **mu** [unaccusative Ns]  
 arrival<sub>DEF</sub> him<sub>CL</sub>  
 b. **negovoto** pristigane  
 his<sub>DEF</sub> arrival
- (7) a. arestuvaneto **mu** [passive Ns]  
 arrest<sub>DEF</sub> him<sub>CL</sub>  
 b. **negovoto** arestuvane  
 his<sub>DEF</sub> arrest  
 ‘his arrest’
- (8) a. predstavjaneto **mu** na gostite  
 presentation<sub>DEF</sub> him<sub>CL</sub> to guests<sub>DEF</sub>  
 ‘his presentation to the guests’/  
 ‘\*the presentation of the guests to him’  
 b. **negovoto** predstavjane na gostite  
 his<sub>DEF</sub> presentation to guests<sub>DEF</sub>  
 ‘his presentation to the guests’/  
 ‘\*the presentation of the guests to him’

### 1.3. Strong Pronouns

DP-internal strong pronouns are introduced by the preposition *na*, though these oblique forms (e.g., *na nego* ‘of/to him’) differ sharply in distribution and interpretation with respect to both full *na*-phrases and pronominal possessives. This is shown in (9). Strong pronouns cannot refer to the Agent of transitive Ns (9a), the Agent of unergative Ns (9b), the Theme of unaccusative Ns

(9c), or of a passive N (9d). They are thus barred from realizing the subject of the DP:<sup>7</sup>

- (9) a. opisanieto      na prirodata na Ivan Vazov/ \***na nego** [Agent]  
 description<sub>DEF</sub> of nature<sub>DEF</sub> of Ivan Vazov/ of him  
 ‘Ivan Vazov’s/\*his description of nature’
- b. \*laeneto      **na nego**  
 barking<sub>DEF</sub> of him  
 ‘his barking’
- c. \*pristiganeto **na nego**  
 arrival<sub>DEF</sub> of him  
 ‘his arrival’
- d. \*arestuvaneto **na nego**  
 arrest<sub>DEF</sub> of him  
 ‘his arrest’

Strong pronouns can only function as DP-internal direct and indirect objects, as we show in (10): in (10a, a’) the *na* + strong pronoun realizes the DP-direct object corresponding to the Theme, while in (10b, b’) it realizes the indirect object corresponding to the Goal argument.

- (10) a. opisanieto      ti      **na Ivan/ na nego**  
 description<sub>DEF</sub> you<sub>CL</sub> of Ivan/ of him  
 ‘your description of Ivan/him’
- a.’ tvoeto opisanie      **na Ivan/ na nego**  
 your<sub>DEF</sub> description of Ivan/ of him  
 ‘your description of Ivan/him’
- b. predstavjaneto ti      **na Ivan/ na nego**  
 presentation<sub>DEF</sub> you<sub>CL</sub> to Ivan/ to him  
 ‘your presentation to Ivan/him’

<sup>7</sup> Note that this is a restriction that pertains to the DP only. In clauses, *na* + strong pronouns can function as indirect objects, preferably doubled by a dative clitic:

- (i) Ivan (mu) dade parite na nego.  
 Ivan him<sub>CL</sub> gave<sub>3SG</sub> money<sub>DEF</sub> to him  
 ‘Ivan gave the money to him.’

- (10) b. tvoeto predstavjane **na Ivan/ na nego**  
 your<sub>DEF</sub> presentation to Ivan/ to him  
 ‘your presentation to Ivan/him’

The distributional facts discussed so far are summarized in Table 1.

Table 1

Full <i>na</i> -DP-arguments	Possessive adjectives	Possessive clitics	<i>na</i> +strong pronouns	
OK	OK	OK	*	DP-subject
OK	*	*	OK	DP-object
OK	*	*	OK	DP-indirect object

The special behavior of *na* + strong pronouns, which constitutes the puzzle to be addressed in the next section, is the first of three converging properties that can be said to characterize the DP-internal subject of Bulgarian. These are given in (11):

- (11) a. Only DP-subjects fail to be expressed by the preposition *na* ‘of’ + a tonic pronoun (of the type *na men(e)* ‘of me’, *na teb(e)* ‘of you’, *na nego/ neja/tjax* ‘of him/her/them’).  
 b. Only DP-subjects can be rendered by a possessive adjective or by a (DP-internal) possessive clitic (of the type *na men(e)* ‘of me’, *na teb(e)* ‘of you’, *na nego/ neja/tjax* ‘of him/her/them’).  
 c. Only DP-subjects can be extracted. (A detailed discussion of this can be found in Krapova and Cinque 2013.)

## 2. Towards an Account

We now turn to an account of property (11a), i.e., why oblique forms of strong pronouns are barred from DP-subject position. The account will be couched in terms of a morphological Case conflict, a phenomenon well known from various languages, as we will illustrate in section 4. The formal execution of the account will follow the spirit of Kayne (2005) but will incorporate the basic tenets of Caha’s (2009) Case theory, especially his treatment of functional

prepositions (such as *na* in Bulgarian) as syntactically analogous to Case suffixes within the Case hierarchy.

### 2.1. DP-Internal Lexical Subjects and Caha's Case Theory

As a preliminary step of the analysis consider the following examples with deverbal nouns from Latin taken from Benveniste (1966), where both the object and subject argument are realized as genitives:

- (12) a. *genitivus obiectivus*<sup>8</sup>  
 neglegentia religionis (cf. *neglegere religionem*)  
 (the) neglect of religion (cf. 'to neglect religion')
- b. *genitivus subiectivus*<sup>9</sup>  
 adventus consulis (cf. *consul advenit*)  
 (the) arrival of (the) consul (cf. 'the consul arrives')

On the basis of examples like (12), we propose, in the spirit of Benveniste (1966), the following informal generalization of Case correspondences between the deverbal DP and the clause:

- (13) Nominative and Accusative of the clause are rendered in the corresponding deverbal nouns with genitive (*genitivus subiectivus* and *genitivus obiectivus*), arguably a structural Case in that it is independent of the particular theta-role assigned to the DP bearing it.

With this generalization in mind, we would like to argue that:

- *Na*-phrases (lexical or pronominal) realizing the DP-internal subject and object are underlyingly Genitive—*genitivus subiectivus* and *genitivus obiectivus*, respectively.
- *Na*-phrases (lexical or pronominal) realizing the DP-internal indirect object are inherent *datives*.

<sup>8</sup> "la fonction du génitif est de transposer en dépendance nominale la relation d'un accusatif régime d'un verbe transitif. C'est donc un *génitif de transposition...*" [the function of the genitive is that of transposing inside a nominal the accusative relation of a transitive verb...] (Benveniste 1966: 146).

<sup>9</sup> "cette fois la forme casuelle transposée en génitif n'est plus un accusatif, mais un nominatif." [this time the Case form transposed in the Genitive is no longer an Accusative but a Nominative] (Benveniste 1966: 147).

The syncretic nature of the corresponding clitic comes out very clearly if one compares its uses within the DP, where, as mentioned, it can only express the syntactic subject (whether an Agent, a Theme, or a Possessor), never a Goal (as opposed to clausal syntax, where this is the only available interpretation, cf. *Az mu dadox knjigata* 'I gave him the book') and never to what is realized as a genitive in other languages such as English *of* + DP, e.g., *He convinced us of his innocence, He thinks of you, He is ashamed/proud of me* (these *of* + DP genitive cases are rendered by prepositional phrases in Bulgarian other than *na*).

We thus conclude that clitics are morphologically syncretic between dative (in the clause) and genitive (within DPs). *Na*-phrases are instead exclusively dative in the clause (abstracting away from the directional *na*-phrases mentioned in fn. 5), but can express both dative and genitive in the DP. As a reviewer pointed out, the archaic dative non-clitic pronominal *nemu/nej* 'him/her<sub>DAT</sub>' cannot express the subject of the DP (*\*pomnja reakcijata nemu* 'Lit.: remember<sub>1SG</sub> reaction<sub>DEF</sub> him<sub>DAT</sub>'), thus reinforcing the idea that DP subjects are exclusively in the genitive Case.<sup>10</sup>

## 2.2. A Historical Detour

The proposal outlined briefly in section 2.1, according to which genitives and datives are structurally distinguished in the Bulgarian DP, differs from both traditional and more recent analyses (e.g., Pancheva 2004, Harizanov 2011, 2014), which argue that Bulgarian possessive clitics as well as full or pronominal *na*-phrases (at least those doubled by a clitic) value Dative Case.

As is well known, a rather conspicuous contact-induced change that occurred in Bulgarian is Case loss. One of its hallmarks was the merger (syncretism) of the genitive and the dative, which began already in Old Bulgarian (OB)/Old Church Slavonic (OCS) (starting from the 10th century onwards) and gradually led to the replacement of the genitive by the dative, first in the clitic paradigm and later on in the pronominal and nominal systems.

Relying on evidence from diachrony, possessor raising, clitic doubling, and the behavior of non-clitic possessors, Pancheva (2004) argues that already in the older stages of the language dative clitics, and by extension dative marked nonclitic pronouns as well as full DPs, valued abstract Dative Case,

<sup>10</sup> The same situation is found with the archaic relative Dative *wh*-form *komuto*:

- (i) a. *čovekât, komuto podarix knjigata*  
 man<sub>DEF</sub> whom<sub>DAT</sub> gave<sub>1SG</sub> book  
 'the man to whom I gave the book'
- b. *\*čovekât, komuto pomnja reakcijata*  
 man<sub>DEF</sub> who<sub>DAT</sub> remember<sub>1SG</sub> reaction<sub>DEF</sub>  
 'the man whose reaction I remember'

both in their possessive (DP-internal) use and in their use as indirect objects (clause-internally). Obviously, homophony with the genitive (morphological merger) cannot be invoked as the trigger for the change because the older stages of the language kept the two paradigms distinct at least until the 17th century. Krapova and Dimitrova (2016) have shown, based on an analysis of OB/OCS corpora, that the genitive and the dative were not used indistinguishably in the earliest written texts; cf. (14), where the dative and the genitive are clearly differentiated in their clausal (dative of interest) vs. DP-internal (possessive) usage. This is not expected under an analysis which postulates, given their synonymous usages inside the DP, that one and the same head noun could have both abstract Dative and abstract Genitive features in its lexical frame before the merger took place.

- (14) da pokryjotъ se **emou** děla **ego**  
 let COVER<sub>3PL</sub> REFL him<sub>DAT</sub> deeds<sub>NOM</sub> his<sub>GEN</sub>

‘Let his deeds be covered for him’

(Euch.Sin., 194, 68b, Minčeva 1964: 25)

Arguably, the OB dative was an inherent case, while the genitive was a DP-internal structural case, and the two also differed in other respects, including clitic vs. nonclitic paradigms. We believe that it was precisely the dative clitic (for 1st and 2nd person) and the 3rd person anaphoric/weak pronoun of the applicative construction of the type illustrated in (15) that gave rise to the process of genitive-dative case merger:

- (15) a. (Cod. Supr., 23, 127v, 2)

jako mečemŭ otŭsěčeši **mi** glavŭ  
 as sword<sub>INST</sub> cut off me<sub>CL.DAT</sub> head<sub>ACC</sub>

‘as you cut off my head with a sword’

- b. (Mt. 9:30, Cod. Mar.)<sup>11</sup>

i otvrěste se **ima** oči  
 and opened REFL them<sub>DU.DAT</sub> eyes<sub>NOM</sub>

καὶ ἠνεώχθησαν **αὐτῶν** οἱ ὀφθαλμοί  
 and were opened **of-them**<sub>GEN</sub> the<sub>NOM.PL</sub> eyes<sub>NOM</sub>

‘and their sight was restored’

lit. ‘and were opened to-them the eyes’

(Krapova and Dimitrova 2016: ex. (6a))

<sup>11</sup> In all of the examples to follow, the OB/OCS text will be accompanied by the Greek original, following Nestle (1904).

As is also the case with Indo-European in general, one conspicuous property of the OB clausal dative was that, provided necessary syntactic conditions (on which see Minčeva 1964 or Krapova and Dimitrova 2016), this Case could also signal a possessive relation (*dativus possessivus*) with certain inalienable nouns (kinship, body parts, and other relational nouns). Thus, applicative constructions, especially the ones in (15) involving a dative clitic or a weak pronoun (referring to an extra argument of the verb) and an inalienable object, provided the first context that triggered the overlap with the possessive genitive. In the specific case of Bulgarian, this was partly due to influence from New Testament Greek, where a similar process had already occurred in the opposite direction, namely genitive clitics which properly belonged to the possessed DP (see (15b)) could cliticize to the verb and end up in positions typically reserved for indirect-object and applicative datives. For details and more examples see Krapova and Dimitrova (2016).

On the view that syntactic change proceeds in small but discrete steps, it could be suggested that the OB DP was reanalyzed in such a way as to incorporate the linearly preceding dative clitic as an argument of the noun. This change must have occurred at the time when both definiteness marking and DP-internal second position effects were on the rise (e.g., the Non Initiality constraint discussed in Harizanov 2011).<sup>12</sup> The possessivization strategy that first affected inalienable nouns quickly led to the assumption of all genitive functions by the dative clitic (or weak pronoun) and opened the possibility for other pronouns as well as lexical nouns to combine with datives instead of genitives,<sup>13</sup> including nominalized DPs in which the dative was used as a structural case in place of the older *genitivus subiectivus* and *genitivus obiectivus* comparable to (13) from Latin.

To summarize, we believe that the advent of the dative should receive a syntactic explanation. The partial overlap between the syntactic functions of the dative and the genitive can be expressed formally in terms of morphological case underspecification. If anything, the evidence presented in Cinque

<sup>12</sup> This is visible from *varia lectiones* of one and the same document in which the clitic or the weak pronoun alternate between the prenominal and the postnominal position, the latter seemingly second within the DP. For examples see Dimitrova-Vulchanova and Vulchanov (2010).

<sup>13</sup> This conclusion then follows in the spirit of Meillet (1897: 151), who viewed the Bulgarian possessive dative as resulting from the postposition of the first/second person dative enclitics (*otrokŭ mi* 'my child', *bratŭ ti* 'your brother'), followed by third person anaphoric pronouns (*imę emu* 'name him<sub>DAT</sub> = his name', *tělo jemu* 'body him<sub>DAT</sub> = his body', *zaštiteľi imŭ* 'defender of them<sub>DAT</sub> = their defender') and finally by full dative NPs (e.g., *propovėdnikŭ živyimŭ i mrŭtvymŭ* 'preacher for the alive<sub>DAT</sub> and the dead<sub>DAT</sub>', Supr. 461: 10, ex. from Minčeva 1964: 52).



and Krapova (2009)<sup>14</sup> and in this paper leads us to adopt the second possible scenario that Pancheva (2004: 183) herself envisages theoretically, namely, that clausal indirect object clitics and possessive clitics have “distinct case features, [but are] realized by a single form because of homophony of the two exponents or because of complete underspecification for case of the single exponent.”

### 2.3. Addressing the First Puzzle: Case Realization in the Inflectional Domain of DP in Modern Bulgarian

In this section, we discuss the derivation of *na*-phrases as DP-internal arguments such as those in (1) above. In particular, we are interested in the type of Case *na*-phrases realize given that they are also underspecified for case and thus ambiguous between a Dative (on the clausal level) and a Genitive (on the DP level). Suppose first that the inflectional domain of DPs in Bulgarian (possibly in all languages) contains subject and object agreement Phrases ( $\text{Agr}_s$  and  $\text{Agr}_o$ , or comparable Case licensing positions) which assign/check Nominative and Accusative Case as they do in finite clauses. A classical case in point is that of Hungarian DPs described by Szabolcsi (1994) where, as illustrated in (16), subjects (e.g., *az elnök* ‘the president’ in (16b)) bear Nominative case under agreement with the head noun (*megfoszt-ás-a* ‘deprivation’). Moreover, DP-internal agreement is identical in morphemic realization (*-a/-ja*) to the verbal agreement morphology licensing Nominative Case in the corresponding clause (16a). In (17) we give, following Szabolcsi (1994), the (simplified) structural representation of (16b) with the  $\text{Agr}_s$  projection, which in Szabolcsi’s original rendering is represented as N(ominative) + I(nflection) and is deemed responsible for Nominative Case assignment/checking:

- (16) a. Edith megfoszt-**ja** az elnök-öt a jogai-tól.  
 Edith<sub>NOM</sub> deprive<sub>3SG</sub> the president<sub>ACC</sub> the privileges-from  
 ‘Edith deprives the president of his privileges.’

<sup>14</sup> In Cinque and Krapova (2009) and Krapova and Cinque (2013) we show that two “possessor raising” constructions must be distinguished in Modern Bulgarian: one which involves genuine movement (corresponding to the Romance genitive extraction), and another one which does not involve movement and thus corresponds to the externally merged dative clitic of Romance applicative constructions (like e.g., *Gli ho rotto la macchina* ‘I broke the car on him/I broke his car’). For details we refer to those papers. In languages that make a morphological distinction between genitive and dative (Italian, French, etc.) it is only genitive phrases that extract (cf. Cinque 1980):

- (i) Az *mu* gledax *t* arestuvaneto *t* po televizijata.  
 I him<sub>DAT</sub> saw<sub>1SG</sub> arrest<sub>DEF</sub> on TV<sub>DEF</sub>  
 ‘I saw his arrest on TV.’

- (16) b. az elnök megfoszt-ás-a a jogai-tól  
 the president<sub>NOM</sub> deprivation<sub>POSS.3SG</sub> the privileges-from  
 ‘the president’s deprivation of his privileges’  
 (Szabolcsi 1994: 233)

- (17) [DP az [Agr<sub>s</sub>(=N+I)P elnök ... [Agr<sub>s</sub>(=N+I) megfoszt-ás-a]]

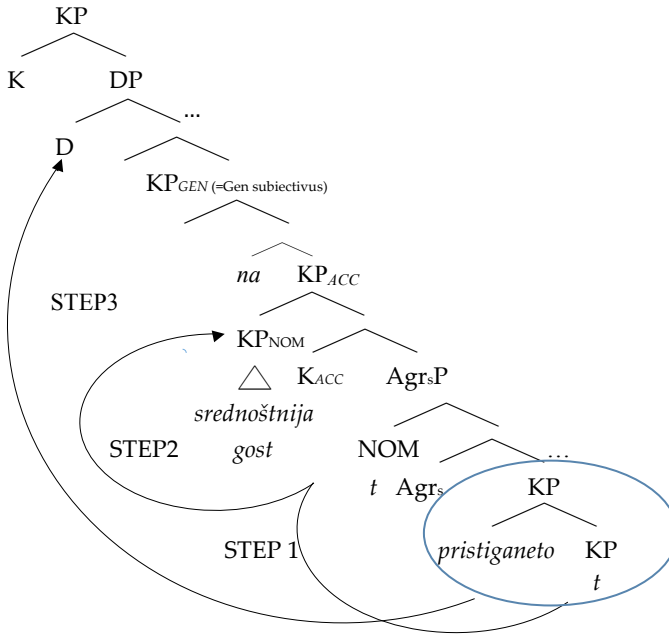
Although DP’s inflectional domain universally contains subject and object agreement Phrase (Agr<sub>s</sub>P, Agr<sub>o</sub>P), as we propose, languages differ in whether or not they express person agreement features morphologically. We can assume that languages with no person morphology on the head noun are still able to assign/check Nominative Case but this Case cannot be licensed in the absence of morphological distinctions. Caha’s (2009) nanosyntactic theory offers us an insight into the solution of this problem. According to Caha, Case is represented by a Case suffix or by a functional preposition, so if a language has no Case morphology for a certain abstract Case of the Universal Case Contiguity hierarchy given in (18) below, this language must resort to the insertion of a functional preposition in the same position where the respective Case suffix would be merged. We thus propose that the functional preposition *na* merges in the position that would correspond to the Genitive as the next higher Case above the Accusative. Note that although lexical nouns do not show any morphological Case distinctions, thus appearing to be syncretic between Nominative and Accusative, strong pronouns clearly show Accusative morphology when selected by prepositions, cf. *na/s/ot nego/neja* ‘of/with/from him<sub>ACC</sub>/her<sub>ACC</sub>’. Genitive is thus a composite Case made up of the Genitive preposition *na* + Accusative Case.

- (18) The Case sequence:  
 nominative—accusative—genitive—dative—instrumental—comitative  
 (Caha 2009: 10)

- (19) The functional sequence:  
 [Comitative [Instrumental [Dative [Genitive [Accusative [Nominative]]]]]]]

With these tenets in mind, we propose the step-by-step derivation in (20) on the next page for (1c) *pristiganeto na srednoštnija gost* ‘the arrival of the midnight visitor’ containing an unaccusative N. In the first step, the maximal projection of the subject *srednoštnija gost* (a KP<sub>NOM</sub>) raises to Spec,Agr<sub>s</sub>P, where it attempts to check the structural Case assigned there, Nominative. This Case however fails to be licensed because of the absence of person agreement on the N *pristiganeto* differently from Hungarian (16b); consequently the subject KP is forced to receive and be licensed by another Case. Within the extended

(20) *pristiganeto na srednoštnija gost* ‘the arrival of the midnight visitor’



nominal projection the only structural Case available is the Genitive Case, which in Bulgarian is represented by the preposition *na* assigning Accusative (like every other preposition of the language). This means that the subject has to raise to the Spec of  $KP_{ACC}$ . The preposition *na* is then merged in the next projection up, namely, in the head of  $KP_{GEN}$ , which licenses the Accusative. After that, following Kayne’s (2002, 2004) analysis of PPs (in head-initial languages), we take the remnant *pristiganeto* to move to a projection immediately above (we do not take a stand on the identification of this projection), which derives the correct order of head and arguments.

To summarize, we propose that Genitive KPs (bearing *genitivus subjectivus* in Benveniste’s terms) are part of the extended projection of the NP in Bulgarian, and are realized as *na*-phrases corresponding to the DP-internal subject argument via two Case licensing DP-internal positions. In section 5, we return to the reasons behind this conclusion, essentially having to do with the fact that they occupy A-positions as well as with their capacity to A-bind anaphors.

## 2.4. Explaining the Case Conflict with DP-Subject Tonic Pronouns

Recall that although the oblique form of the strong pronoun in Bulgarian cannot realize a DP-subject, it can realize a DP direct or indirect object. We now turn to an explanation of this restriction. Tonic pronouns in Bulgarian have two morphologically distinct forms: one for Nominative and one for Accusative. Given what was said above, namely, that the functional preposition *na* assigns the next morphological case down the hierarchy, i.e., Accusative, the approach adopted above predicts that morphological Accusative on the pronoun will be compatible with the Accusative assigned by the preposition, though not with the Nominative Case feature licensed by  $\text{Agr}_s$ . Thus, the reason why *na* + Accusative pronoun cannot act as a DP subject follows from a Case conflict.

Case conflict also rules out a strong pronoun appearing in its Nominative form, as in (21), since in this case Nominative morphology is compatible with the Nominative checked in  $\text{Agr}_s\text{P}$  but not with the Accusative assigned by the preposition. Hence, there is no morphological form which can be simultaneously compatible with both Case features. This rules out the possibility of a *na*+strong pronoun acting as a DP subject in cases such as (21):

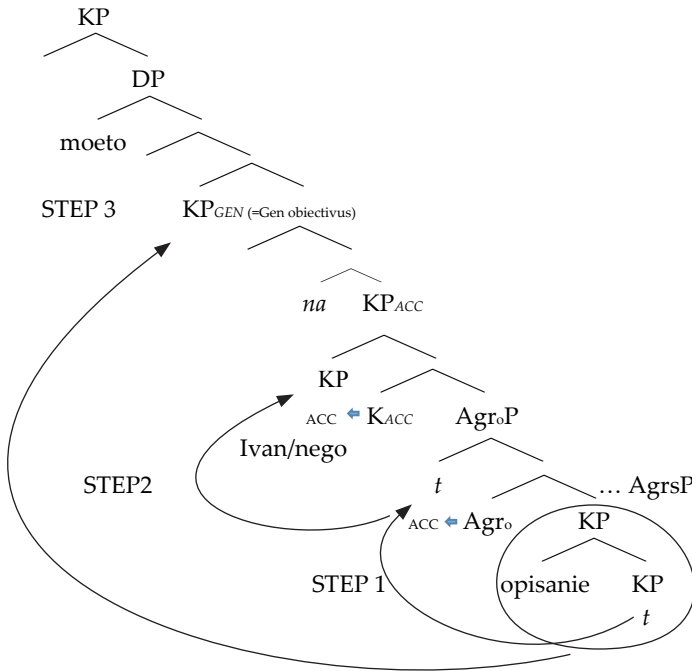
- (21) \**pristiganeto na nego/ na toj*  
 arrival<sub>DEF</sub> of him<sub>ACC</sub>/ of he<sub>NOM</sub>

No Case conflict arises with lexical nouns as in, e.g., *pristiganeto na Ivan* ‘Ivan’s arrival’ since, as mentioned above, in Bulgarian lexical nouns show no morphological distinctions, i.e., they are underspecified for morphological Case and are thus compatible with whatever Case gets assigned to them.<sup>15</sup>

The derivation of (direct) objects follows a similar pattern, assuming  $\text{Agr}_o\text{P}$  and a higher  $\text{KP}_{\text{GEN}}$  corresponding to *genitivus obiectivus* in Benveniste’s terms. As can be verified from the derivation of (22), no issue of Case conflict arises here. The reason is that the direct object checks Accusative in  $\text{Agr}_o\text{P}$  but—since (structural) Accusative is not licensed in this position in the Bulgarian nominal projection—the DP is forced to receive another Case, structural Genitive, which is composed of the functional preposition *na* assigning Accusative. This means that the direct object has to raise to  $\text{KP}_{\text{ACC}}$ , the Case assigned by the preposition *na*, merged above  $\text{KP}_{\text{ACC}}$ . Lexical DPs are not marked morphologically for any Case, so they are compatible with whatever Case is assigned to them when they move to  $\text{Spec},\text{KP}_{\text{ACC}}$  (in other words, they are protected from a potential Case conflict). Strong pronouns do not create a

<sup>15</sup> A reviewer raises the question of why clitics (e.g., *mu*) do not give rise to a Case conflict. We take them to also be underspecified. As noted, in the DP they express Genitive and in the clause they express Dative.

(22) *moeto opisanie na Ivan/hu nego* ‘my description of Ivan/him’



Case conflict either, since their Accusative case morphology will be compatible both with the Accusative Case checked in Agr<sub>o</sub>P and with the Accusative Case assigned by *na*. This predicts the availability of both full DPs and the *na*+pronoun sequence as direct objects. We return to indirect object (dative) arguments in section 4.

To summarize this section, Nominative and Accusative morphology is compatible with structural Genitive, the next Case up the hierarchy in (19), on the condition that the pronoun bears compatible morphology in the Agr position through which it passes on its way to the KP<sub>ACC</sub> required by *na*.

### 3. Case Conflicts Cross-Linguistically

The above account of the lack of subject properties of *na* + strong pronoun in Bulgarian is comparable to the (morphological) Case conflicts found in other contexts in other languages, such as the morphological Case conflict created by Topicalization in Norwegian (Taraldsen 1981), where subjects of clauses embedded under a bridge verb can be topicalized when they are not morphologically marked for Nominative or Accusative. We see this with non-pronominal DPs, e.g., proper names like *Per* ‘Peter’ in (23a) or 3rd person pronouns like *han*

'he/him', *dere* 'they/them' in (23b), which are not morphologically marked for Case hence can be topicalized. However, as (23c) shows, 1st and 2nd person pronouns, which have both a Nominative and an Accusative form, cannot be topicalized (in either form).

- (23) a. Per hadde de trodd [\_ville komme forsent]  
Per had they thought would arrive too late
- (23) b. han/dere hadde de trodd [\_ville komme  
'he/him'/'they/them' had they thought would arrive  
forsent]  
too late
- c. \*jeg/du/vi hadde de trodd [\_ville komme  
\*I<sub>NOM</sub>/you<sub>NOM</sub>/we<sub>NOM</sub> had they thought would arrive  
forsent]  
too late
- d. \*meg/deg/oss hadde de trodd [\_ville komme  
me<sub>ACC</sub>/you<sub>ACC</sub>/us<sub>ACC</sub> had they thought would arrive  
forsent]  
too late

This follows, as Taraldsen argues, if extraction is successive cyclic (passing through Spec,CP) and, as indicated in (24), the bridge verb assigns Accusative to the DP in Spec,CP on top of the Nominative assigned to it in Spec,IP:

- (24) Per<sub>i</sub> hadde de trod [CP t<sub>i</sub> [IP t<sub>i</sub> [\_ville komme forsent]]  
Nom/ACC →ACC Nom←  
'Per had they thought would arrive too late'

Only DPs which are not morphologically marked for either Nominative or Accusative Case can avoid the Case conflict. Hence those in (23a, b) succeed, but not those in (23c, d),<sup>16</sup> because neither the Nominative nor the Accusative

<sup>16</sup> The idea that bridge verbs (may) assign Accusative Case to a DP passing through the Spec of CP has also been proposed by Kayne (1980: 79f); also see the discussion in Bošković (1997: 50) on *wager*-class verbs, and Franks 2017: 116ff). Case assignment from matrix verb to an NP in COMP seems also appropriate for the following contrasts:

- (i) a. John who(m) I assure you [\_[\_to be the best]]  
b. \*I assure you [[ John to be the best]]
- (ii) a. Jean, que Marie croit [\_[\_être intelligent]], ...  
Jean who Mary believes to be intelligent, ...  
b. \*Marie croit [[Jean être intelligent]]



Only syncretic nominative and accusative forms, e.g., neuter forms, which happen to be compatible with both required Cases, result in absence of Case conflict.

- (27) a. Wezmę      którekolwiek przyślesz      [ $t_{ACC}$ ]. (Guz 2017: ex. (41/42))  
 take<sub>PERF.1SG</sub> whichever send<sub>PERF.2SG</sub>  
 ‘I’ll take whichever you send.’
- b. Wezmę      którekolwiek [ $t_{NOM}$ ] przyjdzie      pierwsze.  
 take<sub>PERF.1SG</sub> whichever comes<sub>PERF.3SG</sub> first  
 ‘I’ll take whichever comes first.’

The restriction on Accusative-marked subject pronouns has a wide cross-linguistic distribution. For Italian, for example, Cinque (1980; see also Cinque 2014) has shown that 1st and 2nd person pronouns too are incompatible with a subject interpretation (28a, b), although they are fine as DP-internal objects (28c); 3rd person pronouns on the other hand are fine under contrast, cf. (29). Note however that invoking contrast as an explanation would not suffice to account for the restriction on 1st and 2nd person pronouns, as in (28a, b), which remain equally ungrammatical even when the pronouns are contrastively stressed. Thus, it seems that the different behavior of *lui* ‘he/him’ in (29) is due to the fact that the Nominative and the Accusative forms of this pronoun are syncretic.

- (28) a. la nascita/ partenza di Gianni/ \*di me (non quella di te)  
 the birth/ departure of Gianni/ of me (not that of you)  
 ‘Gianni’s birth/departure / my birth/departure (not yours)
- b. la reazione di Gianni/ \*di me (non quella di te)  
 the reaction of Gianni/ of me (not that of you)
- c. la sua descrizione di me  
 the his description of me  
 ‘his description of me’ sua<sub>SUBJ</sub> > me<sub>OBJ</sub>  
 ‘\*my description of him’ \*him<sub>OBJ</sub> > me<sub>SUBJ</sub>
- (29) la nascita di LUI (non quella di lei)  
 the birth of him (not that of her)  
 ‘his birth’ (not hers)

English pronouns, on the other hand, except for *you* ‘2p.SG/PL’, have distinct forms for Nominative and non-Nominative case, so this language behaves as expected: pronouns are blocked from appearing as DP-internal subjects,



although they are available as DP-internal objects. Consider (30), judgments due to Steven Franks (p.c.):<sup>18</sup>

- (30) a. \*the arrival/ \*the reaction of him/me    [\*1/\*2/\*3 person pronoun]<sup>19</sup>  
       BUT  
       b. the/her description of him/me.

Although the data above from Bulgarian, as well as Norwegian, Slavic, German, Romance, and English, require a more in-depth treatment, all of these languages are apparently multiple Case checking languages, i.e., type B languages according to Bejar and Massam's (1999) typology. As Bejar and Massam note, Multiple Case checking is not only language-specific but construction specific, in that it is allowed only if there is no morphological conflict between the Cases assigned to the links of a chain. According to Bejar and Massam, this could be related to a PF requirement: if the chain created by movement contains conflicting case values, vocabulary insertion is blocked since there will be no item simultaneously consistent with every position in the chain. As mentioned in fn. 17, we do not take a stand here on the precise theoretical reason or which module of grammar is responsible for Case conflicts, leaving this issue for further research.

To summarize this section, we proposed that Bulgarian pronominal DP-internal subjects and objects spell out Genitive case, *genituous subiectious* and *genituous obiectious*, respectively. Adopting a Kaynean framework inspired by Caha's functional hierarchy, we argued that lexical Ns do not present a Case conflict since their morphology is compatible with the Case assigned by the preposition *na*. Subject tonic pronouns, on the other hand, do present a Case conflict since whichever morphological form is selected will conflict with the higher Accusative licensed by *na*.

<sup>18</sup> Differently from Italian, the syncretic form *you* does not yield a more acceptable result. It thus appears possible that a different factor may come to override the distinction in morphological case, namely the competition between the oblique form of *you* and the possessive adjective which is the more natural alternative strategy (cf. (i)):

- (i) your arrival/reaction

We thank Wayles Browne, Steven Franks, Thomas Grano, and Peter Cole for their judgements.

<sup>19</sup> Note that under coordination the violation disappears: *the arrival of me and my family* (Steven Franks, p.c.). Coordination can also rescue other would-be Case violations; cf. Grano (2006):

- (i) a. \*Me will leave tomorrow  
       b. You and me will leave tomorrow

#### 4. The Second Bulgarian DP Case Puzzle: Clitic Doubling

The above noted Case conflicts with strong pronouns disappear under clitic doubling. This issue will be discussed in the current section, the goals of which are twofold. We will briefly discuss the conditions on DP-internal clitic doubling in Bulgarian in order to pose the problem of what type of Case such structures realize. We then establish a correlation between true Datives, which we argue to be PPs, and doubled *na*-pronominals, which do not behave as such. In particular, as opposed to true Datives, they can front to the DP edge position. We also give two additional pieces of evidence that can help in distinguishing Genitive from Dative *na*-phrases within the Bulgarian DP. We then turn to a proposal about how to solve the second puzzle in relation to Agr-less DPs where clitic doubling of pronominals occurs.

##### 4.1. Event and Object Nominal and Clitic Doubling

Event/process nominals in Bulgarian, especially unaccusative and unergative ones, allow for clitic doubling of their DP-internal nominal arguments—see (31a), where *na Ivan* is doubled by the clitic *mu*.<sup>20</sup> But clitic doubled tonic pronouns are still ungrammatical, as (31b) shows, implying that the clitic cannot save a pronominal subject from a Case conflict.

- (31) a. pristiganeto **mu** **na Ivan**  
 arrival<sub>DEF</sub> him<sub>CL</sub> of Ivan  
 ‘Ivan’s arrival’
- b. \*pristiganeto **mu** **na nego**  
 arrival<sub>DEF</sub> him<sub>CL</sub> of him  
 [Intended] ‘his arrival’

<sup>20</sup> DP-internal direct or indirect objects cannot be clitic-doubled. Thus, *im* of (i) is not a coreferent clitic doubling *gostite* ‘the guests’ so (i) cannot get the interpretation ‘the presentation of the guests’:

- (i) predstavjaneto im na gostite  
 presentation<sub>DEF</sub> them<sub>CL</sub> to guests<sub>DEF</sub>  
 ‘their presentation to the guests’/ \*‘the presentation of the guests’

Note that this should in principle be possible given that the *na*-phrase can also realize the subject of a passive nominal like the one in (i). Obviously then, (i) shows that whenever a clitic is present within the DP, it must be interpreted as the DP-subject (including unaccusative and passive subjects, as in (6) and (7) above). This follows from the properties of the possessive clitic as formulated in (11b) and discussed in section 1.2.

However, object nominals (in the sense of Grimshaw 1990), whether inalienable (32a, a'), or alienable (32b, b'), present us with a puzzle: if the DP-subject is spelled out by a strong pronoun then clitic doubling becomes obligatory in order for the structure to be grammatical. Note that the clitic doubled pronoun can appear postnominally or prenominally:

- (32) a. *bratjata* \*(mu) **na nego** a'. **na nego** *bratjata* \*(mu)  
 brothers<sub>DEF</sub> him<sub>CL</sub> of him<sub>ACC</sub> of him<sub>ACC</sub> brothers<sub>DEF</sub> him<sub>CL</sub>  
 'his brothers'
- b. *knigite* \*(mu) **na nego** b'. **na nego** *knigite* \*(mu)  
 books<sub>DEF</sub> him<sub>CL</sub> of him<sub>ACC</sub> of him<sub>ACC</sub> books<sub>DEF</sub> him  
 'his books' 'his books'

The prenominal position of the clitic doubled strong pronoun can reasonably be identified with the DP edge, i.e., the absolute initial position of the entire DP, since as shown in (33), the doubled pronoun precedes all of the higher elements in the functional structure of the DP: the quantifier *vsički* 'all', (33a), the demonstrative, (33b), or the numeral, (33c):<sup>21</sup>

- (33) a. **Na tebe** *vsičkite ti* *knigi sa v spalnjata.*  
 of you<sub>ACC</sub> all<sub>DEF</sub> you<sub>CL</sub> books are in bedroom<sub>DEF</sub>  
 'All your books are in the bedroom.'
- b. **?Na nego** *tezi mu novi prijateli simpatični li sa ti?*  
 of him<sub>ACC</sub> these him<sub>CL</sub> new friends nice Q are you<sub>DAT</sub>  
 'Do you like these new friends of his?'
- c. **Na mene** *dvamata mi po-malki bratja sa*  
 of me<sub>ACC</sub> two<sub>DEF</sub> me<sub>CL</sub> more-young brothers are  
*arxitekti.*  
 architects  
 'My two younger brothers are architects.'

For the purposes of illustration, (34) gives the structural representation of the DP of (33c) after movement of the pronoun to the edge position indicated here as XP. We make no special assumptions about the projection hosting the clitic, so we label it simply as CLP (see Franks 2001 for a detailed discussion and

<sup>21</sup> Dimitrova-Vulchanova and Giusti (1999), Dimitrova-Vulchanova (2000), and Giusti and Stavrou (2008) argue that the preposed position of the doubled (noun or) pronoun is an A-bar position in the left periphery. Reasons not to agree with this conclusion are given below.

Tasseva-Kurktschieva 2004 for a proposal that the clitic heads PossP). In section 4.3. we will further suggest that the clitic and the strong pronoun start out together:<sup>22</sup>

- (34) [XP na mene [QP [DemP [DP dvama<sub>i</sub>-ta [ Num  $t_i$  [CLP mi [AP po-malki  
[bratja<sub>j</sub> [KP<sub>GEN</sub>  $t_{na}$  KP<sub>ACC</sub>  $t_{mene}$ ]  $t_j$ ]]]]]]

There are two reasons why obligatory clitic doubling with pronominal arguments presents a puzzle. First, this is not expected under the analysis in 2.3 since the accusative-marked pronoun would somehow be able to escape a Case conflict, at least as far as object nominals are concerned. Second, obligatory clitic doubling is a property of (strong) pronominal arguments only. Lexical nouns on the other hand can but need not double in either post- or prenominal position, as shown in (35):

- (35) a. bratjata (mu) na Ivan a'. na Ivan bratjata (mu)<sup>23</sup>  
brothers<sub>DEF</sub> him<sub>CL</sub> of Ivan of Ivan brothers<sub>DEF</sub> him<sub>CL</sub>  
'Ivan's brothers' 'Ivan's brothers'
- b. knigite (mu) na Ivan b'. na Ivan knigite (mu)  
books<sub>DEF</sub> him<sub>CL</sub> of Ivan of Ivan books<sub>DEF</sub> him<sub>CL</sub>  
'Ivan's books' 'Ivan's books'

A question arises at this point as to which abstract Case the oblique form of the strong pronoun realizes in view of the above clitic-doubling facts. A possible solution would be that when doubling becomes obligatory the pronoun does not bear Genitive any longer but bears instead a possessive Dative case

<sup>22</sup> In (34), QP and DemP are ordered in such a way as to reflect the fact that the universal quantifier precedes the demonstrative, as also seen in (i). The opposite order in (ii) is also possible but with a slightly different interpretation; cf. the English translations. Given this, it might be the case that there is an additional QP position below DemP.

- (i) Vsički tezi tvoi bratovčedi kâde da gi nastanja?  
all these your cousins where should them<sub>CL</sub> accommodate<sub>ISG</sub>  
'Where shall I accommodate all these cousins of yours?'
- (ii) Tezi vsički tvoi bratovčedi kâde da gi nastanja?  
these all your cousins where should them<sub>CL</sub> accommodate<sub>ISG</sub>  
'Where shall I accommodate all of these cousins of yours?'

<sup>23</sup> A reviewer states that in the absence of Clitic Doubling (35a') becomes ungrammatical. We do not agree with this judgment, although doubling is indeed sometimes preferred with kinship terms and part-whole nouns. The exact conditions remain to be established. Note that adding a doubling clitic in (38b) below does not save the sentence.

akin to the possessive dative of spoken French (Zribi-Hertz 2002), exemplified in (36a). This is only possible with pronouns also under doubling by a possessive adjective, as in (36b).

- (36) a. un ami à moi (Kayne 1975; Zribi-Hertz 2002: 156)  
 one friend to me  
 ‘a friend of mine’
- b. **mon** ami/livre à moi  
 my friend/book to me  
 ‘my friend/book’

A parallel situation is presented also by German, where a dative 3rd person pronoun or a dative-marked full DP doubles *sein* ‘his’, as in (37):

- (37) **ihm/dem Karl sein** Buch<sup>24</sup>  
 him<sub>DAT</sub>/the<sub>DAT</sub> Karl his book  
 ‘his book’

Bulgarian oblique pronouns could thus be argued to spell out Dative case under doubling only. This suggestion seems plausible at first because, although possessive adjectives in Bulgarian cannot be doubled (cf. *\*moeto dete na men* ‘\*my child of me’), the parallelism described in section 1.2. shows that the possessive clitic has identical distribution and interpretation to the possessive adjective. Thus, the difference between French/German and Bulgarian would be a matter of formal rather than functional language-specific choice. One might say that the doubling environments of Bulgarian are precisely the contexts that “keep memory” of the genitive-dative syncretism with vestiges of the Dative as inherent case with object nominals. It is, however, dubious that *na mene/na tebe/na nego* can ever be Datives when appearing DP-subjects and objects. In the next subsection, we present some relevant arguments against this at first tempting conclusion.

<sup>24</sup> In (colloquial) German this construction is quite restricted. Roland Hinterhölzl and Peter Paschke (p.c.) tell us that for them the initial dative can be a full DP and (more marginally for Peter Paschke) a 3rd person singular masculine pronoun but no other pronoun. Roehrs (2013: 59) appears to accept 3rd person pl. masculine pronouns as well.

## 4.2. Strong Pronouns Introduced by *na* do not Behave on a Par with *True Datives*

What we here label “true datives” refers to syntactic properties of DP-internal Goals (indirect objects). So let us now briefly investigate their properties to see whether the above proposal is on the right track.

### 4.2.1. True Datives are PPs

Quite generally, Goal arguments are not available with object nominals in Bulgarian. For example, *pismoto na decata* can never be interpreted as ‘the letter to the children’,<sup>25</sup> only as ‘the children’s letter’, so the Case marker *na* of Bulgarian is not parallel to the French preposition *à*, cf. *la lettre à Jean* ‘the letter to John’. This subsection will review the behavior of Goal arguments in passive nominalizations (see (3) in 1.1. above) as this is the only context where a “true dative” is available.

Goal arguments in Bulgarian have been analyzed as PPs (Slavkov 2008), and even though we cannot enter into this issue, we assume that they are PPs also DP-internally. One general property of PPs in Bulgarian is that they cannot be fronted within the DP (see the ungrammatical (38a) featuring the fronted argument PP *za тази книга* ‘about this book’). This behavior is obviously shared by the PP Goal argument *na decata* ‘to the children’ in (38b), which too cannot be fronted:

- (38) a. \*Mladijat žurnalist razkritikuva [DP [PP **za тази книга**]<sub>i</sub>  
 young<sub>DEF</sub> journalist criticized<sub>3PL</sub> [[for this book]  
 obštoprietoto mnenie *t<sub>i</sub>*].<sup>26</sup>  
 standard<sub>DEF</sub> opinion]  
 ‘The young journalist criticized the standard opinion about this book.’

<sup>25</sup> To express the dative relation other prepositions are used: *do* ‘to’, e.g., *pismoto do decata* ‘the letter to the children’, *za* ‘for’, e.g., *pomoštta za Ivan* ‘the help to Ivan’, or *kâm* ‘to, towards’, e.g., *priziv kâm prezidenta* ‘appeal to the president’. We are aware of only two object nominals that can combine with dative *na*: *podarâk* ‘present’ and *pametnik* ‘monument’. These may be only apparent exceptions if dative *na* is here selected by a silent participle, e.g., *present meant for N* or *monument dedicated to N*. Thanks to Steven Franks for raising this issue.

<sup>26</sup> Although we take KP, not DP, to be the highest projection of the extended nominal projection (with possible other projections above it, to host topic or focus KPs; cf. (34)), here we use the label DP rather than KP for ease of reference.

- (38) b. \*Učitelite razkritikuvaxa [DP [PP **na decata**]<sub>i</sub>  
 teachers<sub>DEF</sub> criticized<sub>3PL</sub> [[to children<sub>DEF</sub>]  
 bezrazbornoto razdavane na učebni materiali *t<sub>i</sub>* (ot  
 indiscriminate<sub>DEF</sub> distribution of study materials (by  
 sponsorite)].  
 sponsors<sub>DEF</sub>]  
 ‘The teachers criticized the random distribution of study  
 materials to children (by the sponsors).’

Compare now (38) with the grammatical (39), featuring a possessor *na*-phrase in (39a) and a DP subject of an event nominal in (39b), both of which can front:

- (39) a. Učitelite šte pokanjat [DP [PP **na decata**]<sub>i</sub> [Possessive]  
 teachers<sub>DEF</sub> will invite<sub>3PL</sub> of children<sub>DEF</sub>  
 roditelite (im) *t<sub>i</sub>*.  
 parents<sub>DEF</sub> them<sub>CL</sub>  
 ‘The teachers will invite the children’s parents.’
- b. Komissijata ne odobri [DP [PP **na Ivan**]<sub>i</sub> [DP-subject]  
 committee<sub>DEF</sub> not approved<sub>3SG</sub> of Ivan  
 slaboto (mu) predstavjane *t<sub>i</sub>*.  
 weak<sub>DEF</sub> him<sub>CL</sub> presentation<sub>DEF</sub>  
 ‘The committee did not approve the Ivan’s weak performance.’

If the proposal developed in section 2.3 is extended to lexical possessors with object nominals, then we may generalize that all KP Genitives can front while PP Goals cannot. This difference is unexpected if the position attracting the fronted phrase were a Topic or a Focus *A'* position, since fronting should be possible with all sorts of phrases, contrary to fact. We take it as confirmed then that the DP edge position is accessible to Genitive arguments only and that it qualifies as an *A*-position (*pace* Dimitrova-Vulchanova and Giusti 1999).

#### 4.2.2. True Datives Cannot be Rendered as Clitics

Another difference between genitives and true datives regards clitic behavior. As already mentioned in section 1.2, Goal (Dative) *na*-phrases, in contrast to (subject) Genitive ones, cannot be rendered by a clitic (nor by a possessive adjective, which we do not illustrate here): in (40b), *im* ‘them<sub>CL</sub>’ must refer to *knigi* ‘books’ of (40a). Note moreover that *im* and the *na*-phrase *na decata* ‘to children<sub>DEF</sub>’ cannot swap interpretation without inducing ungrammaticality. This gives us another argument that the type of Case the possessive clitic (and

adjective) realize cannot correspond to a true Dative—see also Franks and King 2000: 276f—but must correspond instead to a structural Genitive.

- (40) a. *razdavaneto na knigi na decata (ot učitelite)*  
 distributing<sub>DEF</sub> of books to children<sub>DEF</sub> (by teachers<sub>DEF</sub>)  
 ‘the distribution of books to the children (by the teachers)’
- b. *razdavaneto im na decata (ot učitelite)*  
 distributing<sub>DEF</sub> them<sub>CL</sub> to children<sub>DEF</sub> (by teachers<sub>DEF</sub>)  
 ‘the distribution of them (=books) to the children (by the teachers)’
- c. \**razdavaneto im na podarâci (ot učitelite)*  
 distributing<sub>DEF</sub> them<sub>CL</sub> of presents (by teachers<sub>DEF</sub>)  
 [Intended] ‘the distribution of presents to them (the children) (by the teachers)’

#### 4.2.3 Different Behavior with Respect to *na*-Drop

Another phenomenon which distinguishes between Dative and Genitive in Bulgarian is what has come to be called *na*-drop (Vakareliyska 1994). As noted by Vakareliyska and discussed by Slavkov (2008), unquestionable Goal (and affected) datives appear to allow omission of the preposition *na* under clitic left dislocation. See (41–42) from colloquial Bulgarian:

- (41) (Na) *nego ne sa mu kazali za tova.*  
 to him<sub>ACC</sub> not have<sub>3PL</sub> him<sub>CL</sub> told<sub>PRT.PL</sub> about that  
 ‘They haven’t told him about that.’
- (42) (Na) *nego včera mu sčupixa očilata.*  
 to him yesterday him<sub>CL</sub> broke<sub>3PL</sub> glasses<sub>DEF</sub>  
 ‘They broke his glasses on him.’

*Na*-drop is however impossible in non-affected contexts, which Cinque and Krapova (2009) have identified as Genitive, namely, those in which a clitic possessor is moved from within the DP alongside a full oblique pronominal, as follows:

- (43) \*(Na) *nego ti poluči li mu<sub>i</sub> [pimoto t<sub>CL</sub> t<sub>pronoun</sub>]?*  
 of him<sub>ACC</sub> you<sub>NOM</sub> received<sub>2SG</sub> Q him<sub>CL</sub> letter<sub>DEF</sub>  
 ‘Did you receive his letter?’



Whatever the explanation for the phenomenon of *na*-drop, we take the contrast between (41–42) and (43) to show that *na*, as a functional Case assigning preposition, cannot be deleted under movement of the oblique pronominal from within the DP headed by *pismoto* (see the next subsection), although this appears to be possible under clitic left dislocation involving a true PP headed by a silent preposition *na*.

There are many languages in which only Genitives can be extracted. Such is the case in Romance, where only Genitives which correspond to the external argument realized by Italian *di* ‘of’ + DP/*ne* can extract (cf. Cinque 1980, 2014). In Cinque and Krapova (2009) we discussed evidence showing the same to hold in Bulgarian in spite of the morphological syncretism between Genitive and Dative.

The evidence provided in sections 4.2.1–4.2.3 leads us to discard the hypothesis that *na*-phrases can ever be PPs within the DP. Although more research is needed to confirm this conclusion, so far it emerges that arguments of object nominals have the syntax of Genitive arguments of nominalizations rather than that of their Goal argument.

### 4.3. Back to the Puzzle

The puzzling possibility of spelling out the DP subject/possessor of an object nominal with a clitic doubled strong pronoun is compounded by the fact that clitic doubling is not obligatory with lexical DPs. Recall (32) repeated here as (44a, b); see also Dimitrova-Vulchanova and Giusti (1999), Giusti and Stavrou (2008):

- (44) a. *bratjata/ knigite \*(mu) na nego*  
*brothers<sub>DEF</sub>/ books<sub>DEF</sub> him<sub>CL</sub> of him<sub>ACC</sub>*  
 ‘his brothers/books’
- b. *bratjata/ knigite (mu) na Ivan*  
*brothers<sub>DEF</sub>/ books<sub>DEF</sub> him<sub>CL</sub> of Ivan*  
 ‘Ivan’s brothers/books’

At this point we do not have a definitive explanation for the contrast between (44a) and (44b), although obligatory doubling with pronominals is reminiscent of Kayne’s (2001: 192) observation that “pronominal arguments that are structurally Case-marked in French must be doubled by a clitic”.<sup>27</sup>

<sup>27</sup> Also see the obligatory clitic doubling with pronouns in Romanian (ia) and Spanish (ib) clauses:

- (45) Jean \*(me) connaît moi.  
 Jean me<sub>CL</sub> knows me  
 ‘Jean knows me.’

We will tentatively assume, following Kayne (2001: 192), that the clitic and its pronominal double are merged together in argument position as a “big KP” (cf. Uriagereka’s 1995: 81 or Franks and Rudin’s 2005 “big DP”),<sup>28</sup> as shown in (46). Translated into our terms, this means that the clitic is merged in the Spec of our KP<sub>GEN</sub> of (22) and subsequently raises to its post-determiner position stranding [*na nego*], which may also front (possibly, as a remnant) to the edge of the entire KP; cf. (47):

- (46) [KP<sub>GEN</sub> mu [na nego]]  
 him<sub>CL</sub> of him
- (47) a. knjigite mu<sub>k</sub> [t<sub>k</sub> [na nego]]  
 books<sub>DEF</sub> him<sub>CL</sub> of him  
 b. [t<sub>k</sub> [na nego]]<sub>i</sub> knjigite mu<sub>k</sub> t<sub>i</sub>

The Bulgarian pronominal clitic doubling in (46–47) recalls the obligatory doubling of Genitive possessive pronouns in Venetian (48),<sup>29</sup> in the Molise dialect of Capracotta (49), and in Greek (50):<sup>30</sup>

- (48) a. \*(so) mama **de eo**  
 his mother of him  
 ‘his mother’

- 
- (i) a. Am văzut-\*(o) pe ea. (Steriade 1980: 283)  
 have seen-her P her  
 ‘I saw her.’  
 b. \*(La) vió a ella. (Mayer 2003: 15)  
 (her) saw<sub>3SG</sub> P her  
 ‘He/she saw her.’

<sup>28</sup> The idea that they belong to a big DP receives some support from the fact that *mu* and *nego* (or *Ivan*) necessarily share the same referential index, being thus exempt from a Condition B (or C) violation (cf. Roehrs 2013: section 3.4).

<sup>29</sup> In Venetian (and other Veneto dialects) this doubling is obligatory with kinship terms.

<sup>30</sup> The structure in (46) will plausibly carry over to Venetian (i) and Greek (ii):

- (i) [DP \*(to) [de ti]]  
 (ii) [DP \*(mu) [emena]]



the same abstract Cases but differ in terms of which morphological cases they spell out and in which types of DPs abstract cases emerge.

University of Venice  
 krapova@unive.it  
 cinque@unive.it

## References

- Béjar, Susana and Diane Massam. (1999) "Multiple case checking". *Syntax* 2: 65–79.
- Benveniste, Emile. (1966) "Pour l'analyse des fonctions casuelles: le génitif latin". Emile Benveniste, ed. *Problèmes de linguistique générale*, 1. Paris: Gallimard, 140–48.
- Bošković, Željko. (1997) *The syntax of non-finite complementation: An economy approach*. Cambridge, MA: MIT Press.
- Caha, Pavel. (2009) *The nanosyntax of case*. Ph.D. dissertation, University of Tromsø: CASTL.
- Cinque, Guglielmo. (1980) "On extraction from NP in Italian". *Journal of Italian linguistics* 5: 47–99. [(1995) Guglielmo Cinque, ed. *Italian syntax and universal grammar*. Cambridge: Cambridge University Press, 7–53.]
- . (2014) "Extraction from DP in Italian revisited". Enoch O. Aboh, Maria Teresa Guasti and Ian Roberts, eds. *Locality*. New York: Oxford University Press, 86–103.
- Cinque, Guglielmo and Iliyana Krapova. (2009) "The two "possessor raising" constructions of Bulgarian". Steven L. Franks, Vrinda Chidambaram, and Brian D. Joseph, eds. *A linguist's linguist: Studies in South Slavic linguistics in honor of E. Wayles Browne*. Bloomington, IN: Slavica Publishers, 123–48.
- Citko, Barbara. (2004) "On headed, headless, and light-headed relatives". *Natural language and linguistic theory* 22 (1): 95–126.
- Davies, William and Stanley William Dubinsky. (2003) "On extraction from NPs". *Natural language and linguistic theory* 21(1): 1–37.
- Dimitrova-Vulchanova, Mila. (2000) "Possessive constructions and possessive clitics in the Bulgarian DP". Frits Beukema and Marcel den Dikken, eds. *Clitic phenomena in European languages*. Amsterdam: John Benjamins, 121–46.
- Dimitrova-Vulchanova, Mila and Giuliana Giusti. (1999) "Possessors in the Bulgarian DP". Mila Dimitrova-Vulchanova and Lars Hellan, eds. *Topics in South Slavic syntax and semantics*. Amsterdam: Benjamins, 163–92.

- Dimitrova-Vulchanova, Mila and Valentin Vulchanov. (2010) "An article on the rise. Contact-induced change and the rise and fall of N-to-D movement". Anne Bretbarth, Christopher Lucas, Sheila Watts, and David Willis, eds. *Continuity and change in grammar*. Amsterdam: Benjamins, 335–54.
- Franks, Steven L. (1995) *Parameters of Slavic morphosyntax*. New York: Oxford University Press.
- . (2001) "The internal structure of Slavic NPs, with special reference to Bulgarian". Adam Przepiórkowski and Piotr Bański, eds. *Generative linguistics in Poland: Syntax and morphosyntax*. Warsaw: Instytut Podstaw Informatyki PAN, 53–69.
- . (2017) *Syntax and spell-out in Slavic*. Bloomington, IN: Slavica Publishers.
- Franks, Steven L. and Tracy Holloway King. (2000) *Clitics in Slavic*. New York: Oxford University Press.
- Franks, Steven L. and Catherine Rudin. (2005) "Bulgarian clitics as K<sup>0</sup> heads". Steven L. Franks, Frank Y. Gladney, and Mila Tasseva-Kurkchieva, eds. *Formal approaches to Slavic linguistics: The South Carolina meeting, 2004*. Ann Arbor, MI: Michigan Slavic Publications, 104–16.
- Giusti, Giuliana and Melita Stavrou. (2008) "Possessive clitics in the DP. Doubling or dislocation?" Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: Benjamins, 389–433.
- Grano, Thomas. (2006) "Me and her" meets "he and I": Case, person, and linear ordering in English coordinated pronouns. Honors Thesis, Stanford University. (<http://pages.iu.edu/~tgrano/uht.pdf>)
- Grimshaw, Jane. (1990) *Argument structure*. Cambridge, MA: MIT Press.
- Guz, Wojciech. (2017) "Resumptive pronouns in Polish *co* relative clauses". *Journal of Slavic linguistics* 25(1): 95–130.
- Harizanov, Boris. (2011) "NonInitiality within spell-out domains: Unifying the post-syntactic behavior of Bulgarian dative clitics". Nicholas LaCara, Anie Thompson, and Matthew A. Tucker, eds. *Morphology at Santa Cruz: Papers in Honor of Jorge Hankamer*. Santa Cruz, CA: Linguistics Research Center, 1–30.
- . (2014) "Clitic doubling at the syntax-morphophonology interface: A-movement and morphological merger in Bulgarian". *Natural language and linguistic theory* 32(4): 1033–88.
- Higginbotham, James. (1983) "Logical form, binding, and nominals". *Linguistic inquiry* 14: 395–420.
- Kayne, Richard S. (1975) *French syntax*. Cambridge, MA: MIT Press.
- . (1980) "Extensions of binding and case-marking". *Linguistic inquiry* 11: 75–96

- Kayne, Richard S. (2001) "A note on clitic doubling in French". Guglielmo Cinque and Giampaolo Salvi, eds. *Current studies in Italian syntax. Essays offered to Lorenzo Renzi*. Amsterdam: Elsevier-North Holland, 189–212.
- . (2002) "On some prepositions that look DP-internal: English *of* and French *de*". *Catalan journal of linguistics* 1: 71–115.
- . (2004) "Prepositions as probes". Adriana Belletti, ed. *Structures and beyond*. Oxford: Oxford University Press.
- . (2005) *Movement and silence*. Oxford: Oxford University Press.
- Krapova, Iliyana and Guglielmo Cinque. (2013) "The case for genitive case in Bulgarian". Lilia Schürcks, Anastasia Giannakidou, Urtzi Etxeberria, eds. *The nominal structure of Slavic and beyond*. Berlin: De Gruyter Mouton, 237–74.
- Krapova, Iliyana and Tsvetana Dimitrova. (2016) "The genitive-dative syncretism in the history of Bulgarian: Towards an analysis". *Studi slavistici* XII: 181–208.
- Manzini, Maria Rita and Leonardo Savoia. (2005) *I dialetti italiani e romanci. Morfosintassi generativa*. 3 vols. Alessandria: Edizioni dell'Orso.
- Markova, Angelina. (2007) *Deverbal nominals in Bulgarian: A syntactic analysis*. M.A. thesis, Universitat Autònoma de Barcelona.
- Mayer, Elisabeth. (2003) "Optional direct object clitic doubling in Limeño Spanish". Miriam Butt and Tracy Holloway King, eds. *Proceedings of the LFG06 Conference*. Stanford: CSLI Publications.
- Meillet, Antoine. (1897) *Recherches sur l'emploi du genitif-accusatif en vieux slave*. Paris: É. Bouillon.
- Minčeva, Angelina. (1964) *Razvoj na datelnija pritežatelen padež v bǎlgarskija ezik*. Sofija: Izdatelstvo na BAN.
- Nestle, Eberhard. (1904) *Hē kainē diathēkē*. London: British and Foreign Bible Society.
- Pancheva, Roumyana. (2004) "Balkan possessive clitics: The problem of case and category". Olga Mišeska Tomić, ed. *Balkan syntax and semantics*. Amsterdam: Benjamins, 175–219.
- Roehrs, Dorian. (2013) "Possessives as extended projections". *Working papers in Scandinavian syntax* 91: 37–112
- Slavkov, Nikolay. (2008) "Formal consequences of dative clitic doubling in Bulgarian ditransitives: An applicative analysis". *Journal of Slavic linguistics* 16(1): 139–66.
- Steriade, Donka. (1980) "Clitic doubling in the Romanian WH constructions and the analysis of topicalization". *Papers from the Sixteenth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 282–97.
- Szabolcsi, Anna. (1994) "The noun phrase". Ferenc Kiefer and Katalin E. Kiss, eds. *The syntactic structure of Hungarian*. San Diego, CA: Academic Press, 179–274. [*Syntax and Semantics* 27.]

- Taraldsen, Knut Tarald. (1981) "Case-conflict in Norwegian topicalization". *NELS 11*. Amherst: University of Massachusetts, Graduate Linguistic Student Association, 377–98.
- Tasseva-Kurktchieva, Mila. (2004) "Possessives, Theta Roles, and the Internal Structure of Bulgarian DPs". Olga Arnaudova, E. Wayles Browne, María Luisa Rivero, and Danijela Stojanović, eds. *Formal approaches to Slavic linguistics: The Ottawa meeting, 2003*. Ann Arbor, MI: Michigan Slavic Publications, 251–69.
- Uriagereka, Juan. (1995) "Aspects of the syntax of clitic placement in Western Romance". *Linguistic inquiry* 26: 79–123.
- Vakareliyska, Cynthia. (1994) "Na-drop in Bulgarian". *Journal of Slavic linguistics* 2 (1): 121–50
- Vogel, Ralf. (2001) "Case conflict in German free-relative constructions: An optimality-theoretic treatment". Gereon Müller and Wolfgang Sternefeld, eds. *Competition in syntax*. Berlin: Mouton de Gruyter, 341–75.
- Zribi-Hertz, Anne. (2002) "On the asymmetrical but regular properties of French possessive DPs". Martine Coene and Yves D'Hulst, eds. *From NP to DP: the expression of possession in noun phrases*, vol. 2. Amsterdam: Benjamins, 141–63.

# The Use of Long and Short Definite Articles in Bulgarian

John Leafgren

*Abstract:* The Bulgarian definite article is a morpheme attached to the end of the first word of a definite noun phrase. When the noun in this noun phrase is masculine and singular, two forms of the article, one long and one short, are possible. According to prescriptive grammars, the long form is to be used for definite noun phrases in the nominative case (subjects and predicate NPs linked to subjects by a verb of “being”). The short form, on the other hand, is to be used in other positions, i.e., for definite noun phrases in the objective case (objects of verbs and prepositions). It has often been observed, however, that these rules of distribution are not generally followed in the spoken language, at least colloquially. The current study uses corpus analysis to examine the actual use of long versus short masculine singular definite articles in Bulgarian. The data range from nearly completely “correct” usage in formal written texts to almost complete avoidance of the long form definite article in colloquial oral conversations. In these extreme sources it is the exceptions that are most significant, and the study addresses the existence of shared features among the occurrences of these exceptions. At least equally interesting, however, are the data sources that fall between these two extremes. It was in these intermediate sources that the most intensive search was conducted for patterns, based on phonological, syntactic, and discourse-pragmatic features, in the distribution of the long and short forms.

The Bulgarian definite article is postpositive, attaching to the end of the first word of a definite noun phrase. Compare indefinite *kniga* ‘(a) book’ and *nova kniga* ‘(a) new book’ in (1a) to definite *knigata* ‘the book’ and *novata kniga* ‘the new book’ in (1b).

- (1) Definite versus Indefinite Noun Phrases
  - a. Indefinite: *kniga* ‘(a) book’    *nova kniga* ‘(a) new book’
  - b. Definite:    *knigata* ‘the book’    *novata kniga* ‘the new book’

Interesting discussions about the status of the Bulgarian definite article in terms of whether it is a suffix, a particle, or an enclitic can be found in Elson 1976, Stojanov 1965, Gyllin 1982, and Mayer 1987.<sup>1</sup>

<sup>1</sup> Elson concludes that the Bulgarian definite article has both suffixal and non suffixal characteristics, and that it is definitely not an enclitic. Stojanov believes it is clearly



The definite article agrees with the head of the noun phrase in number and, in the singular, in gender. In (1) we see the feminine singular form of the article *-ta*. The neuter singular form is *-to*. The plural forms for all genders are *-te* and *-ta*, with a phonological distribution: *-te* occurs after front vowels, *-ta* elsewhere. In the masculine singular there are also two forms of the definite article (each with a second orthographic variant used after soft consonants). These two forms are distinguished by the presence or absence of a final consonant *-t*. Compare the “long” forms with the *-t* in (2a) to the “short” forms without the *-t* in (2b).

(2) Long versus Short Masculine Singular Definite Articles

- a. Long definite: sinât ‘the son’ mladijat sin ‘the young son’
- b. Short definite: sina ‘the son’ mladija sin ‘the young son’
- c. Indefinite: sin ‘(a) son’ mlad sin ‘(a) young son’

The prescribed norm in the literary language is to use the long form of the article when a definite masculine singular noun phrase is in a nominative case position (syntactic subjects, predicate noun phrases linked to a subject by a verb of “being”, and appositions to these subjects and predicates). The short form, on the other hand, is to be used in other syntactic positions, i.e., for definite noun phrases in the objective case (objects of verbs and prepositions, and appositions to these objects). These prescriptions themselves are found in volume 2 of the Academy Grammar (1983: 120–21, 169), and are described and/or acknowledged in works by many scholars, e.g., Stojanov (1965: 15, 35–36), Aronson (1968: 50), Andrejčin, Popov, and Stojanov (1977: 133, 157), Andrejčin (1978: 185), Maslov (1981: 144), Chvany (1983: 81), Mayer (1984: 34–35; 1988: 57; 1992: 29), Scatton (1984: 41, 165, 316), Hauge (1999: 31), and Alexander (2000: 65).

Most of these same sources as well as some others, however, point out the limitations of the prescriptive rules, noting that they are applied only in writing or only in very careful or formal language use, that the choice between the long and short forms of the article is in fact a facultative one in the colloquial language (some further report that it is the short form that predominates in the spoken language). Statements like these can be found in Aronson (1968: 51), Andrejčin (1978: 176–77), Scatton (1980: 210), Maslov (1981: 144, 168), Scatton (1984: 41, 165, 316), Hauge (1999: 31), and Alexander (2000: 71). The question then arises concerning the extent to which speakers and writers of Bulgarian really follow the prescribed usage in formal speech and in various types of written communication, and also concerning the possible existence of factors

---

a morpheme, but a type of morpheme distinct from prefix, root, suffix, and ending. Gyllin, on the other hand, argues that the article is, in fact, an inflectional ending. And Mayer rejects all the traditional categories—he sees the Bulgarian article as a “noun phrase marker” rather than as an enclitic, a particle, or any sort of morpheme.

which might influence the distribution of the seemingly facultative use of the long and short forms in the colloquial language.

In his 1992 article “Long and short forms of the masculine definite article in spoken standard Bulgarian,” Gerald Mayer presents the findings of a study, the goal of which was to address these questions. In this study Mayer first had 40 educated Bulgarians who knew English translate orally 17 sentences designed to elicit masculine definite articles in the relevant syntactic positions from English into Bulgarian. He then had each translate 17 similar sentences in writing. Overall, Mayer reports improper usage in 30.9% of the orally produced forms and 16.8% of the written forms. (I use the terms “proper” and “improper” as a convenience and purely in relation to the officially prescribed rules. I in no way mean to imply that everyday colloquial speech is inherently inferior to an official standard.) In both the oral and written translations Mayer found a relatively higher proportion of improper usage in appositive noun phrases and in predicative constructions. Of great significance, although perhaps not surprising, is the fact that while most of the improper forms in the oral translations involved substituting short forms for the prescribed long forms on subjects and nominative predicates, in the written translations a type of hypercorrection predominated—most of the improper forms were long articles used for objects. The lowest rate of improper usage was found in the objects of prepositions, and the data suggest, Mayer writes in his conclusion, that there might be a greater tendency toward improper usage in adjectives than in nouns.

Mayer’s study is interesting and highly suggestive. One might question, however, whether translation responses will necessarily reflect the state of affairs in more natural language use. That is, since translation is a conscious linguistic exercise, it might be the case that such test subjects consciously or subconsciously have their guard up, so to speak, and are more inclined toward prescribed standard usage or even toward hypercorrection than they are in the natural use of language in its purely communicative function. By “natural” use of the language I do not mean only everyday oral colloquial use. Rather, I regard as “natural” any use of language in its primary communicative function, whether informal or formal in register, and whether oral or written in mode of delivery. What I wish to do is go beyond only data consisting of language produced for the sake of producing language.

What is needed is an analysis of the distribution of long and short form definite articles using a corpus-based approach in which the corpus or corpora are made up of naturally occurring language use. Mayer himself in an earlier study in 1984 took a step in this direction by collecting 1,234 examples of masculine singular definite articles from newspaper and magazine sources. Here he found only two improper forms, and one of these he says might be “either a legitimate error or a misprint” (1984: 38).

The goal of the current study is to expand on this corpus-based examination of the Bulgarian definite article, increasing the number of examples analyzed and using and comparing a variety of types of databases—both oral and written, both formal and informal in register. Initial frequency rates of long versus short articles in the varying syntactic positions provide a quick picture of the degree to which the prescribed distribution is being followed in various types of communicative contexts. An analysis of the “mistakes” in sources that almost perfectly follow proper usage rules, an analysis of the relatively few long forms (used properly and improperly) found in data in which the long form is almost never found, and most importantly, a close examination of the distribution in data where language users fall between these two extremes will create a more comprehensive picture of the actual use of these forms.

Table 1 identifies the seven types of data sources I examined and gives the total number of examples from each. Here and elsewhere I have combined figures for nouns and adjectives. I performed all analyses separately for these two parts of speech, but I found no significant differences. For example, there was an overall rate of proper use for nouns of 92.8%, as opposed to 91.8% for adjectives (with this difference somewhat more pronounced in the less formal data than in the more formal material).

The two oral databases are transcripts, which have been generously shared electronically at the sites indicated in the source list. The less formal of these is Krasimira Aleksova’s conversational material, and the more formal comes from Parliamentary debates. From the first of these I excluded from consideration the speech of the person making the recordings, since this speech was normalized in the transcripts. The informal written database consists of transcripts from an electronic chatroom, also accessible electronically. Finally, there are four separate, more formal written data sets. In order of increasing formality, these are the directly reported dialogue of the characters in two novels and a number of short stories, the non-dialogic portions of these

**Table 1.** Data Bases

Type	Data Base	Number of Masc. Sg. Def. Articles
a. Informal Oral	Colloquial Conversation	946
b. Formal Oral	Parliamentary Debates	380
c. Informal Written	Chatroom Transcripts	467
d. More Formal Written	Literary Dialogue	1039
	Literary Prose	1508
	Journalism	2000
	Scholarly Prose	1000
All Data		7340

same works, journalistic data accessed via the Internet, and, most formal, a data set consisting of scholarly prose. The precise identities of these sources can be found in the source list.

Table 2 gives the rate of occurrence of properly used masculine singular definite articles. Here and elsewhere in the study unless otherwise indicated I use the term “subject” in a general sense to refer to true subjects, predicate nominatives, and appositions to these and the term “object” to refer to all other syntactic positions. Thus, according to prescribed rules, definite masculine singular “subjects” should have the long form of the article, and those that are “objects” should have the short form. The percentage indicates the proportion of proper usage. After each percentage I give the actual number of occurrences as a fraction. For example, in the colloquial conversation data 1.9% of masculine singular definite noun phrases in syntactic subject position (or 6 out of 324) have the proper long form of the definite article (i.e., 318 have the short form).

Both the informal oral and the formal written sources display highly consistent definite-article use. In the former, the colloquial conversations, we find almost exclusively short forms, regardless of syntactic status. And in the four more formal written databases we find almost exclusively proper usage, i.e., long subject forms and short object forms. Of most interest in these sources are the exceptional forms—the few proper long subjects in the colloquial conversations, improper short subjects in the formal written data, and improper long objects in all of the sources. Attention will be turned to these exceptional forms later. First, however, I examine usage in the chatroom and Parliamentary data. It is in these sources, formal oral and informal written, that we find the most variation in selection of short- and long-form articles, and it is here ,

**Table 2.** Proper use of the Masculine Singular Definite Article

	Long Subject	Short Object
<b>Informal Oral</b>		
Colloquial Conversation	1.9% (6/324)	98.9% (615/622)
<b>Formal Oral</b>		
Parliamentary Debates	28.6% (34/119)	97.3% (254/261)
<b>Informal Written</b>		
Chatroom Transcripts	8.3% (9/108)	100.0% (359/359)
<b>More Formal Written</b>		
Literary Dialogue	96.8% (328/339)	99.3% (695/700)
Literary Prose	98.2% (439/447)	99.7% (1058/1061)
Journalism	99.4% (694/698)	99.8% (1299/1302)
Scholarly Prose	100.0% (261/261)	99.7% (737/739)

therefore, that one might find some pattern in the seemingly facultative variation.

I analyzed the masculine definite articles in the Parliamentary and chatroom transcripts, checking eleven features for a possible connection to the tendency toward the selection of long versus short forms. These features ranged from phonological to grammatical to discourse-level, and each was one I thought logically might have a connection to the article form, or at least was something I felt should be verified and ruled out as a factor. For each of these eleven features I did, in fact, find a statistical correlation to article selection. Some of these correlations are strong and almost certainly point to a relevant factor, while in others the correlation is much weaker and likely is coincidental within the data examined.

Because the difference between the long and short form article is the mere presence versus absence of a single dental stop, one might wonder whether a phonological feature in the environment of the article might be playing a role in the distribution of the two forms. More specifically, one might hypothesize that a similarity or difference in the point or manner of articulation in the preceding or following environment might increase or decrease the likelihood that a writer, or especially a speaker, would include the *-t*. This hypothesis led to four of the features I examined in the data.

It seemed particularly important to check the manner of articulation of the following sound for a possible influence, since, although he himself reports not finding evidence of such influence in his translation data, Mayer (1992: 35) notes that the Fourth Orthographic Commission in 1921 put forth the rule that the long article was to be used before vowels and the short form before consonants, a rule which was officially in force for two years. An analysis of the chatroom data supports Mayer's suggestion that this phonological rule never took hold. But if there really is a phonological effect, it makes more sense to look for it in the oral data. And in the Parliamentary debate transcripts we do in fact find a correlation. There is indeed a certain difference in occurrence of the article forms dependent on the degree of obstruction of the airstream in the production of the following sound. The highly obstructed dental stop *-t* of the article form is more likely to occur when the following sound is not highly obstructed. But the crucial distinction is not precisely between consonants and vowels but rather between the highly obstructed obstruents on the one hand and the less obstructed resonants and vowels on the other. The details are given in Table 3 on the opposing page.

In the Parliamentary transcripts 41.2% of masculine singular definite subjects in position before a word beginning with a vowel have the long form. This compares to only 23.6% of those in position before obstruents. 41.2 to 23.6 converts to a ratio of 1.75:1. I use this type of ratio to compare the possible impact of the various features examined. Note that for now I am only looking at subjects; I return to objects later. We might also note, in connection with the

**Table 3.** Manner of articulation of the following sound and the frequency of occurrence of long (vs. short) articles on subjects in Parliamentary data

	<b>Long Form</b>
Before Vowel	41.2%
Before Resonant	39.1%
Before Obstruent	23.6%
	<b>41.2:23.6 = 1.75:1</b>

figures in table (3), that only one out of seven, or 14.3%, of masculine singular definite subjects in sentence-final position (i.e., before complete/terminal obstruction, had the long form. If this percentage turns out to be accurate with a larger database, such position would exhibit even less use of long forms than before obstruents.

In looking at the point of articulation of the following sound, it was again in the oral Parliamentary data, rather than in the written chatroom data, that a striking pattern appeared. First, there appeared to be no correlation when it was a vowel that followed—41.7% of masculine singular definite subjects before a front vowel were in the long form, compared to 40.0% before non-front vowels. But when the following word began with a consonant, a consistently increasing likelihood emerged as the point of articulation shifted to the back of the mouth. See the figures in Table 4.

An attempt to find a correlation between the stem-final consonant of the noun or adjective and the use of the long versus short form of the article attached to this stem was not productive. The greatest discrepancies found among the data are given in Tables 5 and 6 on the following page. These correlations between point and manner of articulation on the one hand and long versus short form use on the other are quite weak and do not appear to exhibit a relevant pattern. The Parliamentary and chatroom figures are comparable for manner of articulation, and I have combined them in Table 5. For point of articulation in Table 6 I used only Parliamentary data in an effort to find the

**Table 4.** Point of articulation of the following sound and the frequency of occurrence of long (vs. short) articles on subjects in Parliamentary data

	<b>Long Form</b>
Before Labial Consonant	9.5%
Before Dental Consonant	31.7%
Before Palatal Consonant	33.3%
Before Velar Consonant	40.0%
	<b>40.0:9.5 = 4.21:1</b>

**Table 5.** Manner of articulation of the stem-final consonant and the frequency of occurrence of long (vs. short) articles on subjects in Parliamentary and chatroom data

	<b>Long Form</b>
After Stop	15.7%
After Non-stop	20.3%
	<b>20.3:15.7 = 1.29:1</b>

**Table 6.** Point of articulation of the stem-final consonant and the frequency of occurrence of long (vs. short) articles on subjects in Parliamentary data

	<b>Long Form</b>
After Dental Consonant	30.0%
After Non-dental Consonant	23.7%
	<b>30.0:23.7 = 1.27:1</b>

strongest possible candidate for significance—the chatroom data produced no recognizable pattern.

We see in Table 5 that the final stop in articles occurs less frequently when the preceding consonant is a stop, seemingly parallel to the figures for pre-obstruent position in the table in (3). And one can reason that the fact shown in Table 6 that the article-final dental occurs more frequently after another dental consonant has a sort of logic to it based on ease of pronunciation (not having to move the tongue to a different point of articulation). But in both cases the correlations are weaker than any of the remaining features examined.

Another phonological feature investigated is that of stem length. As I began this examination I had two conflicting hypotheses: Perhaps long articles would be more frequent on longer stems, since longer stems occur with relatively greater frequency in more formal registers, where proper usage is more likely to be found (assuming the speakers in Parliament were not speaking in a single, unchanging register, and likewise for the chatroom writers). Or perhaps fewer long forms would actually be found on longer stems, assuming apocope is more likely to occur in the more relaxed pronunciation when an article is far from word-level stress (which would happen more often in longer words). The results in Table 7, on the opposing page, suggest that the second hypothesis is more likely to be accurate than the first.

Several of the other features investigated as potential factors influencing the distribution of long and short articles are syntactic in nature, having to do with some characteristic of the relationship between the verb and the definite noun phrase. Perhaps most obviously suspect among these features is linear

**Table 7.** Stem length and the frequency of occurrence of long (vs. short) articles on subjects in Parliamentary and chatroom data

	<b>Long Form</b>
Stems with 1 or 2 syllables	20.2%
Stems with 3 or more syllables	11.8%
	$20.2:11.8 = \mathbf{1.71:1^2}$

order: Are preverbal subjects more likely to have long-form articles than postverbal subjects? The logic behind this question rests on the fact that SVO is the predominant, neutral word order in Bulgarian. Because of this, postverbal subjects are atypical and therefore may subconsciously feel less “subject-like”, leading to an increased tendency to use a short-form article. The figures in Table 8 support this hypothesis and produce a fairly strong ratio of almost two to one. Note that for these purposes I looked at only “true subjects”, excluding predicate nominatives (which so far I have been including in the generic term “subjects” on the basis of their being included with “true subjects” in the prescribed rule for long form usage). The predicate constructions are excluded because they do not fit into the logic of the hypothesis concerning typical preverbal position.

A fairly strong correlation can also be found when one examines voice and occurrence of long versus short articles on true subjects. The subjects of active-voice verbs are more than twice as likely to use the long form than are the subjects of passive-voice verbs. Again, this is logical, since in the active-voice subjects are more subject-like. They are both grammatical and logical subjects, as opposed to the subjects of passive voice verbs, which are grammatical subjects but logical objects. See the figures in Table 9 on the following page. I have excluded predicative constructions from consideration here as irrelevant to the basis of the correspondence. The chatroom material is also

**Table 8.** Linear order and the frequency of occurrence of long (vs. short) articles on true subjects in Parliamentary and chatroom data

	<b>Long Form</b>
Preverbal Subject	23.5%
Postverbal Subject	12.3%
	$23.5:12.3 = \mathbf{1.91:1}$

<sup>2</sup> The percentages in the chatroom data are even more unbalanced than these combined rates. Here zero of thirteen (0.0%) masculine singular definite subjects with three or more syllables in the stem use the long form of the article, as opposed to nine of 97 (9.3%) for shorter stems.



excluded on the basis of insufficient data—only two passive constructions with masculine singular definite subjects were found in this corpus (both with short form articles).

**Table 9.** Voice and the frequency of occurrence of long (vs. short) articles on true subjects in Parliamentary data

	<b>Long Form</b>
Active Voice	30.8%
Passive Voice	14.8%
<b>30.8:14.8 = 2.08:1</b>	

Another verbal category examined, transitivity, produced weaker results. One might reason that true subjects of transitive verbs would feel more “subject-like”, and thus be more likely to be used with long-form articles, since they tend to be, overall, more active as Agents than are the subjects of intransitive verbs, and since the very presence of a direct object in the case of transitive verbs might, through contrast, strengthen the perception of the subject’s status as such. But that reasoning appears to be faulty. In actual fact there was, in both the Parliamentary and especially the chatroom data, the opposite correlation, although not a very strong one. The figures are given in Table 10.

**Table 10.** Transitivity and the frequency of occurrence of long (vs. short) articles on true subjects in Parliamentary and chatroom data

	<b>Long Form</b>
Transitive Verb	12.5%
Intransitive Verb	18.6%
<b>18.6:12.5 = 1.49:1</b>	

The final grammatical feature explored is simply that of true subject versus nominative predicate. As in the case of transitivity, my initial predictions proved incorrect. I had expected true subjects, which are often active Agents, and which are more likely to occur in preverbal position than are nominative predicates, to more frequently have the long form of the article. As we see, however, in Table 11 on the opposing page there is actually a relatively weak correlation between predicate nominatives and an increased likelihood of using the long-form of the article.

One discourse-level feature turned out to have a rather strong correlation to the distribution of long- and short-form articles. While looking at the transcripts of the Parliamentary debates, a couple very clear instances stood

**Table 11.** True subject versus nominative predicate and the frequency of occurrence of long (vs. short) articles on subjects in Parliamentary and chatroom data

	<b>Long Form</b>
True Subjects	17.9%
Nominative Predicates	26.9%
<b>26.9:17.9 = 1.50:1</b>	

out, where a speaker would begin his or her turn at the podium using the prescribed rules for the selection of the forms and then part way through the oration would cease doing so completely or partially. It was as if some speakers began either consciously or perhaps from nervousness to speak very formally and properly but then would slide into a more colloquial, natural style of speaking as they got into the flow of their speech. A formal count produced the figures in Table 12. I counted as “early” in the discourse turn anything found in the first six sentences. Analyzing this feature in the chatroom data would be pointless, since speaker turns here were almost exclusively one, two, or three sentences in length.

**Table 12.** Position in the discourse turn and the frequency of occurrence of long (vs. short) articles on subjects in Parliamentary data

	<b>Long Form</b>
Early in Discourse Turn	44.4%
Not Early in Discourse Turn	22.1%
<b>44.4:22.1 = 2.01:1</b>	

Finally, we come to the factor that seems to be the single most influential determinant of the likelihood of using long- or short-masculine singular definite objects—idiolectal variation among individual language users. Some speakers and writers are clearly more likely to use long forms of the article than are others. In some cases this may be due to education (not necessarily in a quantitative sense), dialect, familial environment or perhaps even certain personality traits, e.g., perfectionism or personal inclination toward perceived propriety. The strength of this factor does not negate the value of some of those already discussed. The other features define patterns in the distribution of the long and short forms among those who use both. The idiolectal factor simply reflects the fact that the inclination toward or away from one of the forms is stronger in some language users than in others.

I say that the idiolectal factor seems to be the single most influential determinant. The data make it quite clear that this is the case, but I have not

come up with a method of calculating this statistically so as to produce ratios that are comparable to those given above. A somewhat less systematic set of observations will have to suffice. In the chatroom data, for example, 11 of the 15 writers provide examples of masculine definite articles in the first place. Among these 11, an average of 8.3% of definite masculine singular subjects have the long form of the article, but this average is largely due to the fact that three of the writers use the long form at least half of the time (five use only short forms, and more than half of the long-form subjects occur in the data from just two writers). In the Parliamentary material there are 34 instances of long-form subjects. These are spread out over 18 of the 23 speakers whose recorded speech includes masculine singular definite subjects. Four of the speakers consistently follow the prescribed rules for their subjects, using only long forms. These, however, are speakers whose contributions to the debates include three or fewer such instances. If we confine ourselves to the 10 speakers whose orations include five or more masculine singular subjects, we see great differences in likelihood of using a long form—one at nearly 60%, four at approximately 40%, two at or near 20%, one at 14%, and two at less than 10%.

Table 13 summarizes and ranks the features that have been discussed as potential influences on selection of definite article form. The ranking and figures in this table are oriented toward likelihood of selection of the prescribed long form. The 11 potential factors are listed in order of decreasing likelihood of influence. Chi-square probability figures are given parenthetically in italics.<sup>3</sup>

Attention so far has focused on the forms in the data where the most variation was found—subjects in the Parliamentary and chatroom material. I conclude with a brief look at whether the factors listed as possible influences in the hierarchy in Table 13 appear to be playing a role in the remaining relevant forms—the few properly long subjects in the colloquial conversation data, the improperly short subjects in more formal written data, and the improperly long objects in all the sources. Because we are now looking at relatively few examples in these categories, firm conclusions cannot be drawn. I mention only the correlations or lack of correlations that are most conspicuous in terms of supporting or calling into question the rankings in Table 13.

First, recall that there are only six instances of long form subjects in Aleksova's colloquial oral corpus. Clearly significant is the fact that all six long subjects occur in the speech of a single speaker (among the approximately eighty represented). This certainly supports idiolect as an influential factor. Other support is given by the fact that long forms followed by velars outnumber those followed by labials three to one, all six examples are in active voice

<sup>3</sup> Smaller numbers here are better. For example, 0.041 means there is a 4.1% chance that there is no relationship between the feature and article form, 0.475 means there is a 47.5% chance. Conventional acceptance of statistical significance is 0.05 or less.

**Table 13.** Possible factors working in favor of selection of long form article for masculine singular definite subjects (based on Parliamentary and chatroom data)

More likely influences		( $p=0.041$ – based on the nine speakers whose speech included the most examples)
1.	Idiolectal variation	[Parliament only] ( $p=0.045$ )
2.	Point of articulation of following sound	e.g., before a velar the long form is 4.21 times more likely than before a labial
3.	Voice	with active voice the long form is 2.08 times more likely than with passive voice
4.	Position in discourse turn	within the first six sentences the long form is 2.01 times more likely than later
5.	Linear order	in preverbal position the long form is 1.91 times more likely than in postverbal position
Less likely influences		( $p=0.058$ )
6.	Manner of articulation of following sound	e.g., before a vowel the long form is 1.75 times more likely than before an obstruent
7.	Stem length	with stems of one or two syllables the long form is 1.71 times more likely than with stems of three or more syllables
8.	“True” subjects versus predicates	nominative predicates are 1.50 times more likely to have the long form than are true subjects
9.	Transitivity	with intransitive verbs the long form is 1.49 times more likely than with transitive verbs
10.	Stem-final manner of articulation	after a stop the long form is 1.29 times more likely than after a non-stop
11.	Stem-final point of articulation	after a dental consonant the long form is 1.27 times more likely than after a non-dental
		[Parliament only] ( $p=0.475$ )

clauses, five of the six occur early in the discourse turn (and this speaker does have some long speaker turns) and five of the six are preverbal. The observations that five of the six occur in positions before an obstruent and that four of the six occur on stems that do not end in stops argue against the associated features from Table 13 as relevant factors. Analysis of the remaining features produced nothing noteworthy.

There are 23 improperly short subjects in the more formal written data. I analyzed these forms looking at the features listed in Table 13 in a sort of reverse way. That is, I attempted to discover whether these more formal writers were failing to use the long form of the article in situations where there was an absence of those features found to be conducive to proper long use in the informal writing and the formal speech. Most notable here is the feature associated with linear ordering. Only three of the 23 short-form subjects are in preverbal position. And in two of these three, the short-form in question is attached to the subject of a subordinate clause, which itself is the direct object of a higher clause verb, which may make it feel like an object to the writer. The only other factor listed in Table 13 to receive fairly clear positive support from the improperly short subjects in the more formal written data is that concerning stem length. In Table 13 we see that subjects with stems of one or two syllables are 1.71 times more likely than longer stems to receive the long form of the article in the chatroom and Parliamentary transcripts. Of the 23 improperly short subjects in the formal written data, accordingly, nine have stems of three or more syllables. This is disproportionately high, when one takes into account an overall ratio of approximately 6:1 short to long stems for masculine singular subjects.<sup>4</sup>

The formal written improper short subjects call into question transitivity (feature no. 9 in Table 13) as a factor relevant to the distribution of long and short articles. The chatroom and Parliamentary data suggest that subjects of intransitive verbs are one and a half times more likely to receive the long form of the article than are subjects of transitive verbs. One might then expect that it would be, conversely, the subjects of transitive verbs that would be more likely to be used with the improper short form in the generally more proper formal writing. What we find, however, is that not a single one of the improperly short forms here occurs with a transitive verb.

Relatively weak support for the manner of articulation of the stem final consonant (no. 10 in Table 13) and minor negative evidence for stem-final point of articulation (no. 11 in Table 13) are also suggested by the improperly short

<sup>4</sup> This ratio is based on the count of masculine singular definite subjects in the chatroom and Parliamentary databases, where 193 such subjects have stems of one or two syllables, while only 34 have three or more syllables.

form written subjects.<sup>5</sup> These short written subjects provide no insight into the remaining features summarized in Table 13.

Lastly, I call attention to the possibility that the factors suggested as being conducive to the selection of the long form of the definite article in the Parliamentary and chatroom data in Table 13 might also be playing a role in the use of the improperly long (hypercorrect) objects, which can be found in all of the data types except the chatroom transcripts. To explore this possibility, I looked at each of the 27 long-form objects in the data in light of each of the features listed in Table 13, with the exceptions of no. 8 (True subjects versus predicates), as obviously irrelevant to objects, no. 3 (Voice), as hardly likely to be relevant to objects, and no. 9 (Transitivity), as too heavily skewed in favor of positive transitivity when looking only at clauses with objects. Of the remaining features proposed as possible factors affecting the distribution of long and short form articles only one, idiolectal variation, finds further affirmation in the long object examples. Of the eight long objects found in the artistic literature, six are in the work of a single author. In the Parliamentary debates three of the seven long objects are used by a single speaker (who also uses exclusively long subjects). And in the colloquial conversational data, one of the approximately 80 speakers is responsible for five of the seven long-objects (the same speaker who used the six long-subjects). The long-object examples do not provide noteworthy support for or evidence against the status of any of the remaining features as influences on the selection of article forms—all show somewhat weaker correlations than those seen in the subject data from the Parliamentary and chatroom material.

We can thus conclude that although the choice between long and short forms of the definite article is in a sense a facultative one, there are some discernible patterns. The prescribed rules tend to be followed in formal written language and ignored in colloquial oral communication. And both in the exceptions to these generalizations and in the seemingly less predictable usage found in formal speech and informal writing, several factors (phonological, syntactic, idiolectal, and discourse-structural) can be identified as correlating to patterns in the selection of these facultative forms.

University of Arizona  
leafgren@email.arizona.edu

<sup>5</sup> 19 of the 23 short form written subjects, a 4.75:1 ratio, have stem-final non-stops, as opposed to a 2.81:1 ratio among masculine singular definite subjects in the chatroom and Parliamentary data overall. And 18 of the 23, i.e., 3.60:1, have stem-final dentals, as opposed to a 1.67:1 overall ratio.

## Data Sources

- Aleksova, Krasimira. (1994) "Korpus ot razgovoren bâlgarski ezik". Available at: <http://folk.uio.no/kjetilrh/bulg/Aleksova/>.
- Cačev, Kliment. (1990) "Znoj i xladen vjatâr" (5–14), "Dobâr den, život" (27–60), "Razkaz za geroja" (61–69), and "Po zalez slânce" (70–78). *Znoj i xladen vjatâr*. Sofia: Voенно izdatelstvo.
- Duridanov, Ivan, ed. (1991) *Gramatika na starobâlgarskija ezik*. Sofia: Bâlgarskata Akademija na naukite, 19–31.
- Džonova, Marina. "Bulgarian chatroom talk". Available at: <http://folk.uio.no/kjetilrh/bulg/ICQ/>.
- Georgieva, Elena. (1987) *Slovoed na usložnenoto prosto izrečenie*. Sofia: Bâlgarskata Akademija na naukite, 5–39.
- Gerova, Darina. (1990) "Obvinenie v ubijstvo" (5–30), "Bludnijat sin" (79–86), and "Grobove v krepostta" (87–91). *Ikoni za neovjarvašti*. Sofia: Izdatelstvo "Xristo G. Danov".
- Gugulanova, Ivanka. (2005) *Bâlgarskite pričastija i deepričastija v slavjanski kontekst*. Sofia: Universitetsko izdatelstvo "Sv. Kliment Oxridski", 9–28.
- Igov, Svetlozar. (1991) *Istorija na bâlgarskata literatura 1878–1944*. Sofia: Bâlgarskata Akademija na naukite, 7–21.
- Markova, Ilinda. (1988) *Strax ot kotki*. Sofia: Bâlgarski pisatel.
- Mavrodieva, Ivanka. "Debati v Narodnoto sâbranie". Available at: <http://folk.uio.no/kjetilrh/bulg/Parliament/>.
- News.bg*. <http://news.bg/>, 119 articles from October and November 2003 and from May 2004.
- Sevov, Koljo. (1990) *Ostrovât*. Sofia: Narodna mladež.
- Standart News*, Internet Edition. <http://www.standartnews.com>, 120 articles from August and September 2002.

## References

- Academy grammar*, vol 2. (1983) (See Stojanov, Stojan, Kalina Ivanova, Petâr Pašov, and Valentin Stankov, eds.)
- Alexander, Ronelle. (2000) *Intensive Bulgarian: A textbook and reference grammar*, Vol. 1. Madison, WI: University of Wisconsin Press.
- Andrejčin, Ljubomir. (1978) "Iz vâprosite i praktikata na členuvaneto v bâlgarskija knižoven ezik". Petâr Pašov, comp. *Pomagalo po bâlgarska morfologija: Imena*. Sofia: Nauka i izkustvo, 176–85.
- Andrejčin, Ljubomir, Konstantin Popov, and Stojan Stojanov. (1977) *Gramatika na bâlgarskija ezik*. Sofia: Nauka i izkustvo.
- Aronson, Howard I. (1968) *Bulgarian inflectional morphophonology*. The Hague: Mouton.

- Chvany, Catherine V. (1983) "On 'definiteness' in Bulgarian, English, and Russian". Michael S. Flier, ed. *American contributions to the Ninth International Congress of Slavists, Kiev, September 1983, volume 1: Linguistics*. Columbus, OH: Slavica Publishers, 71–92.
- Elson, Mark J. (1976) "The definite article in Bulgarian and Macedonian". *Slavic and East European journal* 20(3): 273–79.
- Gyllin, Roger. (1982) "Morfoložična klasifikacija na opredelitelniža člen v bälgarskija ezik". *Sâpostavitelno ezikoznanie* 7(3): 16–22.
- Hauge, Kjetil Râ. (1999) *A short grammar of contemporary Bulgarian*. Bloomington, IN: Slavica Publishers.
- Maslov, Ju. S. (1981) *Grammatika bolgarskogo jazyka*. Moscow: Vysšaja škola.
- Mayer, Gerald L. (1984) "The use of long and short forms of the masculine definite article in Bulgarian". *International journal of Slavic linguistics and poetics* 29: 33–41.
- . (1987) "The morphological categorization of the Bulgarian definite article". *International journal of Slavic linguistics and poetics* 35–36: 143–50.
- . (1988) *The definite article in Contemporary Standard Bulgarian*. Wiesbaden: Otto Harrassowitz.
- . (1992) "Long and short forms of the masculine definite article in spoken standard Bulgarian". *Sâpostavitelno ezikoznanie* 17(2): 29–38.
- Scatton, Ernest A. (1980) "On the shape of the Bulgarian definite article". Catherine V. Chvany and Richard D. Brecht, eds. *Morphosyntax in Slavic*. Columbus, OH: Slavica Publishers, 204–11.
- . (1984) *A reference grammar of modern Bulgarian*. Columbus, OH: Slavica Publishers.
- Stojanov, Stojan. (1965) *Členuvane na imenata v bälgarskija ezik*. Sofia: Nauka i izkustvo.
- Stojanov, Stojan, Kalina Ivanova, Petâr Pašov, and Valentin Stankov, eds. (1983) *Gramatika na sâvremennija bälgarski knižoven ezik: Tom II, morfoložija*. Sofia: Bälgarskata Akademija na naukite. (=Academy grammar, volume 2).



# Heads and Dependents in Bulgarian

Petya Osenova

*Abstract:* This paper discusses the notion of head from different perspectives. More specifically, it outlines the various descriptions of the properties of heads as well as the factors that play an important role in determining heads in phrases. Such factors are: the specific theoretical framework (in both constituent and dependency paradigms), the involvement of the various linguistic levels (morphosyntax and semantics), the part-of-speech (lexical or functional), and the domain of operation (word-level or phrasal). These factors, as well as the diagnostics formulated by Zwicky (1985), are considered in detail with a focus on Bulgarian data. I conclude that Bulgarian favors compromising head approaches like mutual selection between the head and its dependent.

## 1. Introduction

The notion of head, although controversial, has proven to be very important for modeling linguistic data. Discussions of its nature and diagnostic criteria have been always lively (especially in the 1980s). A common-sense (and thus somewhat simplistic) definition of head might be as follows: the main word in a phrase that selects other words/expressions or is modified by other words/expressions; the head is also the element that determines the relation of the phrase to the other phrases in the sentence.

However, there are obvious problems that arise with this formulation. It happens very often that we find different elements as heads within the same constituent construction (such as in NN phrases, appositions, etc.). There are also other debatable issues with respect to this definition, e.g., whether it is expressed only by lexical elements or by a phrase, etc. The concept of head varies across theories. Despite its not always quite clear nature, the notion of head has proven to be a crucial component within both constituency-based and dependency-based models. In both, heads are important for investigating word order, subcategorization, modification, etc. They are suitable instruments for

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin*. Bloomington, IN: Slavica Publishers, 2018, 183–95.

modeling local (e.g., agreement, government) and nonlocal (e.g., topicalization) language phenomena.

This paper does not aim to provide a complete description and discussion on the interpretation of the notion of head in the different linguistic theories. Rather, it focuses on presenting the known problems in a conceptual way and interpreting them in light of the Bulgarian language.

The paper is structured as follows: in section 2, I list some perspectives on heads; section 3 presents specific diagnostics for determining what a head is; section 4 summarizes the various types of head realizations with a focus on Bulgarian. Section 5 concludes the paper.

## 2. Perspectives on Heads

Theories based on the classical Chomskyan generative framework use *X*-bar phrases, in which the projections of the lexical head to a saturated phrase are derived through intermediary *X*-bar levels, i.e., they expect some further element(s) in order to become a complete phrase. This marking indicates that the element is still not a saturated phrase. A simple example is the phrase *the man*, where the noun *man* is projected to an *N*-bar (*N'*), thus indicating that the phrase is not complete. When the definite article is added, the phrase is finished as an NP or a DP (depending on the theoretical model adopted): [NP/DP the [N' [N man]]].

Hoeksema (1992) mentions that, similarly to *X*-bar theory, Categorical Grammar prefers the technical (operational) understanding of a head in contrast to the semantic, distributional, or morphosyntactic one. According to this view, the head is the element that determines the category of the phrase. For example, in a VP the head is a verb, and in a PP the head is a preposition, etc. Needless to say, this is the common view in most linguistic theories. Note that this perspective favors functional heads, i.e., auxiliary words as heads (especially prepositions in PPs and auxiliaries and modals in VPs). The alternative view that will be outlined below favors lexical categories as heads, not auxiliaries. Thus it takes a more semantics-oriented perspective. Such a view is often supported by dependency theories.

From a typological point of view, the position of the head within a phrase is considered a parameter across different languages. It can be phrase-initial, phrase-final, or varying. I support the view that this parameter is applicable to particular linguistic constructions within languages, but not to the languages themselves. For example, in Bulgarian the position of the head noun is not entirely fixed. It is initial in a phrase with a prepositional modifier:

- (1) **sok** ot malini  
juice from raspberries  
'raspberry juice'

But it is final in a phrase with an adjectival modifier:

- (2) udoben **stol**  
comfortable chair  
'a comfortable chair'

In PPs, the head is always initial:

- (3) **v** gradinata  
in garden<sub>DEF</sub>  
'in the garden'

Like NPs, the position of the head within VPs varies. In spite of the fact that the canonical word order in Bulgarian is SVO, the verbal head can precede or succeed its arguments and adjuncts. For example, see the realizations in a head-complement phrase:

- (4) a. **Četa** kniga.                      b. Kniga **četa**.  
read<sub>1SG.PRS</sub> book                      book read<sub>1SG.PRS</sub>  
'I am reading a book.'

In a head-subject phrase similar word order patterns can be observed:

- (5) a. Ivan **tiča**.                      b. **Tiča** Ivan.  
Ivan runs                      runs Ivan  
'Ivan runs.'

In a head-adjunct phrase the position of the adjunct can vary as well:

- (6) a. **Tâguvam** dâlgo.                      b. Dâlgo **tâguvam**.  
grieve<sub>1SG.PRS</sub> long                      long grieve<sub>1SG.PRS</sub>  
'I have been sad for a long time.'

In all dependency-based theories the notion of head is crucial, although its interpretations are diverse. For example, Mel'čuk (2003) uses two terms, *head* and *governor*. The first notion refers to a head that is internal to the phrase, while the second one refers to a head that is external to the phrase and

governs it. Thus, in the phrase *udoben stol* ‘comfortable chair’, the head is *stol*, while in the phrase *viždam udoben stol* ‘see<sub>1SG.PRS</sub> comfortable chair’, the verb *viždam* is the governor of the NP. Mel’čuk also claims that the notion head itself is unsuitable for dependency theories since it has been used and formulated mainly for constituent-based theories.

According to Hudson (1990), who developed Word Grammar, X-bar theory somewhat reflects dependency grammar ideas, but there are also some significant differences. For example, in dependency grammars there are only lexical heads, since the syntactic relations are made among words, while in constituency-based grammars there are also phrasal heads, which occur as a function of the argument saturation mechanism. Another difference we can identify is the relationship between the head and its dependent: while in dependency-based theories this relation is direct, in constituent ones it is mediated due to the binary nature of trees.

There are linguistic theories, such as Head-driven Phrase-structure Grammar (HPSG) (Pollard and Sag 1994), where the head is central. Here the feature **HEAD** has many implications: it is a part-of-speech indicator and an internal head of the phrase, as well as an external head that interacts with the heads of other phrases. In HPSG, heads are realized through propagation from the lexical sign up to a saturated phrase or sentence. **HEAD** is part of the syntactic feature **CAT** (=category). **HEAD** has features such as gender, number, case, tense, and aspect. Some of the features can be shared between syntax and semantics, because they are also part of **CONT**(=content). Such features are gender, number, aspect, tense, and mood. In this theory, **HEAD** is defined for lexical as well as for phrasal signs.

Let us take an example with the word *dârvoto* (tree<sub>DEF</sub> ‘the tree’), presented in the feature structure on Fig. 1:

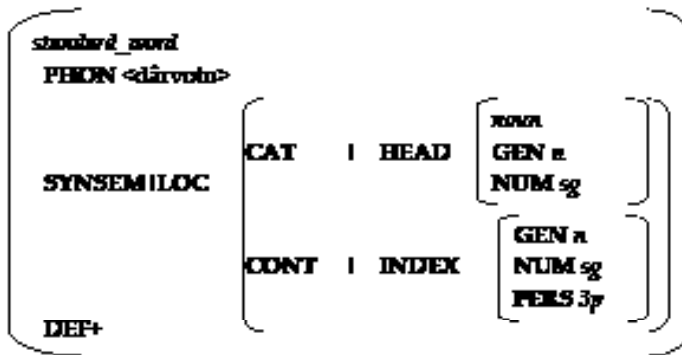


Figure 1

This is a description of a lexical sign which has a **PHON** (=phonetic) form and **SYNSEM** (=syntactic-semantic) characteristics. These characteristics are considered **LOC** (=local). Within the nested syntactic characteristics **CAT** (=category) there is the feature **HEAD**. Its value in this case is *noun*. The *noun* has two syntactic characteristics: **GEN** (=gender) with a value *n(euter)* and **NUM** (=number) with a value *sg* (=singular). Note that within the semantic characteristics **CONT** lies the semantic feature **INDEX**. It can be referential (as in this case) or non-referential (as the expletive *it*). This information, however, is omitted here. The feature **INDEX** also has characteristics **GEN** and **NUM** whose values are the same as the values of the syntactic ones. This means that some features can be part of the syntactic as well as the semantic domain. Additionally, we find the feature **PERS** (=person), which is considered to be semantic. The boolean **DEF** (=definiteness) feature with a positive value is a characteristic of the sign and put at the same level as the features **PHON** and **SYNSEM**. It is determined by a language-specific principle which is not discussed here.

The HPSG approach is typical of constituent-based theories, where the lexical heads coincide with parts-of-speech. Some of these lexical heads can project phrasal heads. Phrasal heads, for their part, can also project other phrases recursively. Of course, such projections are governed by specific principles: for example, the verbal lexical head first takes its complements projecting a VP. This VP becomes a phrasal head that can take its subject (if any) and then adjuncts. Thus, the verb is the lexical head for its arguments. Let us consider the verb [V *vižda*] ‘see<sub>3SG.PRS</sub>’ in [VPC *vižda dârvo*] ‘see<sub>3SG.PRS</sub> a tree’.<sup>1</sup> The verb is a lexical head for its complement *dârvo* ‘tree’. Subsequently, [VPC *vižda dârvo*] becomes a phrasal head for the subject *momče* ‘boy’. Then the subject-head phrase becomes a phrasal head for the adjunct *v parka*, ‘in park<sub>DEF</sub>’: [VPA [VPS *momčeto*] [VPC *vižda dârvo*]] *v parka*, ‘boy<sub>DEF</sub> see<sub>3SG.PRS</sub> tree in park<sub>DEF</sub>’.

The same holds for the nominal domain where modifiers are viewed as adjuncts. Initially, the bare noun in Bulgarian is the lexical head (for example, the noun *slânca* ‘sun’ in the phrase *jarko slânca*, ‘bright sun’). However, when another adjective is added on the left recursively (e.g., *toplo* ‘warm’), it is added already to the phrasal head *jarko slânca*, ‘bright sun’. The result is: [*toplo* [*jarko slânca*]] ‘warm, bright sun’. In all these phrases, heads are purely syntactic features. The main principle in HPSG (which is somewhat similar to X-bar theory) is the Head-feature principle, which translates information from the head to the mother phrase.

At the feature inheritance level, however, heads are divided into syntactic and semantic types. Thus, in phrases of type “head-complement” and

<sup>1</sup> Please note that in this version of HPSG the lexical head first takes its complements, then the subject, and finally the adjuncts. The phrasal notation is as follows: VPC = verb-complement phrase, VPS = verb-subject phrase, and VPA = verb-adjunct phrase.

“head-subject,” where the head is typically a verb or a verbal phrase, these two types of heads coincide. This means that the verb or the verbal phrase promotes both the syntactic and the semantic properties to the higher phrase projection. In phrases of the type “head-adjunct,” however, the responsibilities of the head are divided into syntactic (the verb or the noun) and semantic (the adjunct). Thus, in this latter case the inheritance to the phrase takes into account syntactic properties that come from the syntactic head as well as semantic constraints that come from the adjunct phrase .

### 3. Popular Linguistic Tests for Determining Heads

Linguists have attempted to define diagnostics for identifying heads. One of the most popular views belongs to Zwicky (1985), later debated by Hudson (1987). It should be noted that the former was working in a constituent theory (HPSG), while the latter in a dependency theory (Word Grammar). Zwicky suggested the following diagnostic criteria:<sup>2</sup> (1) Which element is the semantic argument, i.e., the constituent whose meaning has the status of argument in relation to some functor? (2) Which element is the subcategorizand, i.e., the constituent that is subcategorized with respect to its sisters? (3) Which element is the morphosyntactic locus, i.e., the constituent on which any inflections that are relevant to the mother phrase are located? (4) Which element is the governor, i.e., the constituent that determines the morphosyntactic form of some dependent sister? (5) Which element is the determinant of concord, i.e., the constituent with which some other constituent must agree? (6) Which element is the distributionally equivalent constituent, i.e., the constituent whose distribution is similar to that of the mother? (7) Which is the obligatory constituent, i.e., the one which has to be present if the mother is to be categorized as it is? and (8) Which element is the ruler (=head) of dependency theory, i.e., the word on which other words depend?

All the above diagnostics have problems. They refer to properties of both a morphosyntactic and a semantic nature. Also, very often the phrasal projections require the inheritance of linguistic properties from various elements within the phrase, not just the head. For example, diagnostics 1, 6, and 7 are semantic, while 2, 3, 4, and 5 are morphosyntactic. Another problem is the typological difference among languages and among phrases. For example, point 3 would not apply to non-inflectional languages, such as Chinese, and point 5 is not relevant to phrases with a verbal head and an uninflected adverb as an adjunct.

Later, Zwicky (1993: 293) realized that the notion of head is not homogeneous and that it denotes at least three phenomena: a semantic functor, a base,

<sup>2</sup> Note that this list of diagnostic criteria is based on Hudson’s (1987: 110–11) response article, since he enumerates the criteria in a comfortably concise manner.

and a head. The functor is associated with the semantic head in the sense that it is usually the modifier (“adjunct” in HPSG) that propagates the semantic properties further up to the mother phrase. The base is similar to Mel’čuk’s external head, or “governor,” and the head reflects agreement and the governing morphological characteristics with respect to elements which are not part of the phrase. Hudson (1987) omits one of Zwicky’s original diagnostics. This is number 5, agreement (morphosyntactic agreement is referred to as “concord”). For Hudson this criterion has nothing to do with heads. However, he adds two more diagnostics, adjacency and grammatical relations. Hudson views the head as a kind of grammatical relation, similar to subjects and objects, but at a higher level of generalization (i.e., abstraction). Although he applies Zwicky’s diagnostics, Hudson does not agree with him on what a head should be.

Let us apply Zwicky’s diagnostics to two Bulgarian phrases: the head-adjunct type NP (*udoben stol* ‘comfortable chair’) and the head-complement type VP (*gledam televizija* ‘watch<sub>1SG.PRS</sub> TV’). For the NP, the following can be observed with respect to his diagnostics: (1) the semantic argument is the noun, while the functor is the adjective that narrows the possible denotation of the phrase; (2) this diagnostic is not applicable to this phrase; (3) the morphosyntactic locus is the noun although inflection occurs on the adjective as well; (4) this characteristic is not applicable to this phrase; (5) the agreement controller is the noun; (6) the distributive equivalent is the noun; (7) the obligatory element is the noun; and (8) the head with respect to dependency grammar is the noun. Thus, the indisputable head is the noun. However, if instead of an adjective some determiner or quantifier is used, the situation becomes more complex, because, first of all there are approaches that treat determiners and quantifiers as subcategorizers and governors of the noun. Concerning numerals, Bulgarian differs from the other Slavic synthetic languages, since the predicate agrees with the noun, not the numeral.<sup>3</sup> Compare the following:

- (7) a. Mnogo/ Pet            [NP srpskih pisaca] je otišlo            [Serbian]  
       many / five<sub>NEUT</sub>        [Serbian writers] is went<sub>NEUT.SG</sub>  
       u inostranstvo.  
       in abroad  
       ‘A lot of/ Five Serbian writers went abroad.’
- b. Mnogo/ Pet            [NP srâbski pisateli] otidoxa            [Bulgarian]  
       many / five            Serbian writers<sub>PL</sub> went<sub>PL</sub>  
       v čužbina.  
       in abroad

<sup>3</sup> We leave aside cases with numerals from 1 to 4 in the synthetic Slavic languages. This example is taken from Zlatić (1997: 70).

Apart from agreement of the predicate with the noun, other factors also suggest that the noun is the agreement controller. For example, anaphoric pronouns also agree with the noun. If we continue the story from (7b), we could say:

- (8) Sreštnaxme gi tam.  
 met<sub>IPL.PST</sub> them<sub>PL.CL</sub> there  
 ‘We met them there.’

On the other hand, there are counterarguments to the noun being the head. In cases like *mnogo dojdoxa*, ‘many (of them) came’, the ellipsis reflects either a quantified noun (*mnogo xora*, ‘many people’) or a quantified PP (*mnogo ot xorata*, ‘many of the people’). Furthermore, count forms of masculine non-person nouns are not invariant but depend on the numerals that quantify them (\**stola*, ‘chairs<sub>COUNT</sub>’ vs. *pet stola*, ‘five chairs<sub>COUNT</sub>’). Last but not least, and also in connection with count forms, in phrases like *pet litra mljako*, ‘five liters<sub>COUNT</sub> milk’, it seems that the constituency is [[*pet litra*][*mljako*]] rather than [[*pet*][*litra mljako*]].<sup>4</sup> To sum up, agreement is not a valid test, given that government and subcategorization favor the numeral instead of the noun. Franks and Pereltsvaig (2004) propose at least two possible projections of the NP—as a QP or DP—and subsequent to their proposal Bulgarian was standardly classified as a predominantly DP language, in which agreement of the subject and predicate is derived from the noun, not the numeral.

In Bulgarian, demonstratives and other D<sup>0</sup>s behave like adjectives with respect to inflection, but they occur in a special position in the NP (usually the leftmost place), and unlike adjectives, they lack the possibility of recursion. However, the diagnostics do not clearly identify D as a head. Nevertheless, a DP analysis remains an alternative, if no unified analysis of phrases with articles and demonstratives or indefinites is sought.

Thus, when applied, Zwicky’s diagnostics can explicitly identify more than one head, because they employ various ways of selecting an element as a head. The same problem of having two possibilities occurs in appositional phrases of the type *djado Ivan* ‘grandfather Ivan’. These problems seem to be more universal than language-specific for Bulgarian.

Within the VP, the verb generally satisfies all the head diagnostics and thus is undoubtedly the head of the phrase. However, when the predicate includes auxiliaries or semantically bleached verbs (e.g., *šte četa*, ‘will read<sub>1SG.PRS</sub>’, *štjax da četa*, ‘would (to) read<sub>1SG.PRS</sub>’, *vzimam vrâx*, ‘take<sub>1SG.PRS</sub> peak’,

<sup>4</sup> I do not dwell here on the behavior in Bulgarian of nouns with quantity semantics, such as ‘liter’, ‘bottle’, ‘cup’, ‘glass’, or ‘group’. Discussion of this issue can be found in Osenova (2014), where these nouns are viewed as syntactic heads that govern the following NP.



meaning ‘I prevail’), identifying the head becomes less simple. One possibility is that the auxiliary is the head (a purely syntactic approach), while the other is that the verb is the head (a semantics-oriented approach). Modal verbs can vary between these two alternatives. Both approaches have their pros and cons. The first one is structurally and syntactically more natural, while the second one relies on semantics and seeks to minimize structural differences among languages.

#### 4. Head Realization Types

These considerations suggest that heads are a multidimensional phenomenon. For that reason, their realization is related to many characteristics, including at least the following: (1) the number of heads in the phrase (one, many, or none); (2) the level of linguistic representation (morphosyntactic or semantic); (3) the domain of operation (lexical, phrasal element, or both); and (4) part-of-speech (functional or lexical). Let us consider each of these in the next subsections.

##### 4.1. Number of Heads in the Phrase

Some dependency theories (like constituent-based ones) stipulate that each phrase has a head. However, such a view creates problems for coordinated constructions, in which there seem to be multiple heads, and for headless phrases, where formally the leftmost element is the head. In HPSG, the concept of a headless phrase is very popular, but at the same time the head is divided into a syntactic and a semantic one at a more abstract level of grammar modeling. In Hudson’s Word Grammar, the coexistence of multiple heads is validated. In dependency theories the leftmost conjunct is often viewed as the head, but this decision is technical. Going back to HPSG, the idea of shared headedness is also promoted; it licenses the mutual selection between a head and a dependent. In Chomskyan transformational grammars, the head of an NP can be either the noun or the determiner, the latter building a determiner phrase (DP). DP is licensed syntactically by the fact that articles are separate words in some Germanic languages (like English, German) as well as in some Romance ones (like Spanish, Italian). However, in Bulgarian the definite article, despite its phrasal affix nature, is part of the word (head or dependent). Thus, the mechanism of mutual selection can serve as a compromise to handle the discrepancy between the semantic status of the mother NP (definite) and the non-fixed contribution of definiteness coming either from the head in unary phrases (e.g., *ženata*, ‘woman<sub>DEF</sub>’) or from the dependent in binary phrases (e.g., *xubavata žena*, ‘pretty<sub>DEF</sub> woman’). Such a compromise is possible due to the ambiguous outcome of diagnostics 2, 3, and 6.

When we consider indefinites, a further problem arises. If following Nicolova (2008) the existence of a zero indefinite article is accepted (as in  $[\text{NP } \emptyset \text{ žena}]$ ), then the problem of the determiner being a head is even bigger in constraint-based theories like HPSG, since such theories work with declarative and thus non-procedural mechanisms, where the zero operator is avoided. For that reason I accept such phrases as NPs.

In the process of identifying a head, we may also observe a suppressed hierarchy (underspecification of the relations among the elements). This is the case with so-called flat structures, where all the arguments (subject and complements) as well as the adjuncts are realized simultaneously. This strategy can be applied given the following conditions: when there are too many possible heads; when there are language phenomena that are difficult for modeling in binary terms (coordination, apposition, etc.), and when analyzing non-configurational and usually VP-less languages. In HPSG, a flat structure is temporarily used within the head-complement phrase where all the complements are realized at once and in coordinated phrases that are headless. Another issue is substantivization (e.g., *bolnijat beše pregledan ot doktora*, ‘patient<sub>DEF</sub> was examined by doctor<sub>DEF</sub>’). This phenomenon, which is productive in Bulgarian, collapses diagnostics 3, 5, and 6, showing that the dependent can become a head under some circumstances.

## 4.2. Levels of Linguistic Representation

Linguistic theories are often divided into monostratal and multistratal ones. The former attribute properties of the head to one item only, while the latter divide the contribution of headedness among the elements within the phrase (head and dependents). For example, Mel’čuk’s Meaning-Text theory works with additional head-oriented notions than just head. HPSG works with syntactic and semantic heads: in head-subject and head-complement phrases both coincide while in head-adjunct phrases the dependent is the semantic head.

## 4.3. Domain of Operation

Dividing the heads into lexical and phrasal ones (as in X-bar theory or HPSG) is suitable for Bulgarian due to the zero indefinite article, which makes a noun into a lexical head (*žena* ‘woman’, *stol* ‘chair’). Such a view also supports a lexicalist approach to grammar, in which the verb in a valency lexicon syntactically generates its own participants. Additionally, it is consistent with pro-drop languages. Thus, the lexical head in Bulgarian can be any of the following: an intransitive verb with a pro-drop or non-realized subject (*tičam* ‘run<sub>1SG.PRS</sub>’), a transitive verb with a pro-drop or non-realized subject, and also with non-realized complements (*četa* ‘read<sub>1SG.PRS</sub>’).

The accusative clitic is sensitive to the way in which its contribution is interpreted. If analyzed as a lexical realization (V or N) in a sentence with a doubling direct object *četa ja [knigata]* ‘read<sub>1SG.PRS</sub> her<sub>ACC.CL</sub> book<sub>DEF</sub>’, the sequence *četa ja* is defined as a lexical projection, while *knigata* projects an NP denoting the complement. If analyzed as a syntactic realization, the sequence *četa ja* is considered a verbal head with a complement, but then the trouble remains of what to do with the full-fledged complement *knigata* (‘book<sub>DEF</sub>’). Franks and Rudin (2005), among others, propose that such case clitics in Bulgarian are heads of a KP (case phrase) forming a unit with the doubled DP. The authors also take the clitic-as-agreement approach instead of the clitic-as-argument one. Of course, this is just one of the possible solutions and it shows the complexity of the problem. This solution is very close to the lexical approach in the sense that the clitic does not occupy a real argument position; it just reveals some agreement features before the realization of the real complement.

The analysis of phrasal heads depends on the process of argument realization. For example, in HPSG, where the order of saturation is “complements > subject > adjuncts,” phrasal heads are phrases of the type head-subject (for a future head-adjunct projection), head-complement (for a future head-subject or head-adjunct projection), or head-adjunct (for a future recursive head-adjunct projection).

#### 4.4. Part-of-Speech

Here we can imagine three possible situations in which determining the part-of-speech (or category) of a head may be difficult. The first is when there is ambiguity in the category of the head. The second is when the head represents a functional category. The third is when a phrase contains multiple constituents with the same part-of-speech, thus relying on strongly semantic factors to determine which one is the head.

The first situation is best exemplified by an ambiguity in the nominal domain, namely, the rivalry between the noun and the determiner. In the VP domain the ambiguity revolves around auxiliaries and main verbs (modals can go either way), e.g., *bjax kazal* ‘had<sub>1SG.PST</sub> said<sub>M.SG</sub>’, *dobâr sâm* ‘good am<sub>1SG.PRS</sub>’, *moga da dojda*, ‘can<sub>1SG.PRS</sub> (to) come<sub>1SG.PRS</sub>’. The second case refers to complementizers and subordinators as heads; these can subcategorize their clauses in a semantically vacuous way. In HPSG, there is also an alternative where these parts-of-speech just mark the clause they introduce, but do not serve as heads. The third situation is best exemplified by cases of NN. Even with clearly divided quantity (e.g., *čša voda* ‘glass water’) or appositive structures (e.g., *profesor Petrova* ‘professor Petrova’), the head might not be stable and fixed on the same structural element; for more on this topic see Osenova (2014).

## 5. Conclusion

The notion of head is very important to linguistic theory, but at the same time it depends on many and diverse factors, such as one's specific theoretical model, the types of elements within a phrase, the hierarchy of phrasal projections, and so forth.

On the one hand, the concept of head can be used alongside other closely related notions (governor, functor, specifier, base, etc.), but alternatively it can incorporate more than one meaning. The notion of head is a gradable and changeable one within constructions. The most important thing is that, across syntactic domains and among the linguistic levels of modeling, its interpretation should remain consistent.

The Bulgarian data seem to support the idea of mutual selection between the head and the dependent due to the various linguistic contributions of both types of elements to the mother node. Bulgarian also shows diversity in head-initial and head-final structures. This diversity is due to the canonical word orders within various phrases but also to phenomena like ellipsis, substantivization, and so forth.

Sofia University  
petyaosanova@hotmail.com

## References

- Franks, Steven L. and Asya Pereltsvaig. (2004) "Functional categories in the nominal domain". Olga Arnaudova, E. Wayles Browne, María Luisa Rivero, and Danijela Stojanović, eds. *Formal approaches to Slavic linguistics: The Ottawa meeting, 2003*. Ann Arbor, MI: Michigan Slavic Publications, 109–28.
- Franks, Steven L. and Catherine Rudin. (2005) "Bulgarian clitics as  $K^0$  heads". Steven L. Franks, Frank Y. Gladney and Mila Tasseva-Kurktchieva, eds. *Formal approaches to Slavic linguistics: The South Carolina meeting, 2004*. Ann Arbor, MI: Michigan Slavic Materials, 106–18.
- Hoeksema, Jack. (1992) "The head parameter in morphology and syntax". Dickey Gilbers and Sietze Looyenga, eds. *Language and cognition 2*. Groningen: De Passage, 119–32.
- Hudson, Richard. (1987) "Zwicky on heads". *Journal of linguistics* 23: 109–32.
- . (1990) *English word grammar*. Cambridge, MA: Blackwell.
- Mel'čuk, Igor. (2003) "Levels of dependency in linguistic description: Concepts and problems". Vilmos Agel, Ludwig M. Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer, Henning Lobin, eds. *Dependency and valency. An international handbook of contemporary research*. vol. 1. Berlin: Mouton de Gruyter: 188–229.

- Nicolova, Ruselina. (2008) *Bâlgarska gramatika. Morfologija*. Universitetsko izdatelstvo "Sv. Kl. Ohridski".
- Osenova, Petya. (2014) "Classifier noun phrases of the type N1N2 in Bulgarian". Olga Spevak, ed. *Noun valency*. Amsterdam: John Benjamins Publishing Company, 141–60.
- Pollard, Carl and Ivan Sag. (1994) *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press,
- Zlatić, Larisa. (1997) *The Structure of the Serbian noun phrase*. PhD dissertation, University of Texas at Austin.
- Zwicky, Arnold. (1985) "Heads". *Journal of linguistics*: 21–129.
- . (1993) "Heads, bases, and functors". Greville Corbett, Norman M. Fraser, and Scott McGlashan, eds. *Heads in grammatical theory*. Cambridge: Cambridge University Press, 292–315.

## ***How Many Flowers! So Many Colors!* Number Marking in Cardinality Exclamatives in Bulgarian\***

Roumyana Pancheva

*Abstract:* Masculine nouns in Bulgarian inflect for three numbers: singular, plural, and “count.” The count form appears in structures with numerals (e.g., *three colors*), in cardinality questions (e.g., *How many colors?*), and declaratives (e.g., *that many colors*), but is prohibited in combination with *many*, and in cardinality exclamatives (e.g., *How/ So many colors!*), where only the plural form is acceptable. That exclamatives pattern differently from their interrogative and declarative counterparts is particularly surprising because they are formed with the same *wh-/th-* pronouns seemingly directly combining with the noun phrase. This paper offers an analysis of the distinction in number marking in cardinality expressions in Bulgarian. It argues that the composition of *wh-/th-*pronouns, numerals, and *many* with noun phrases is mediated by one of two nonovert degree expressions, MEAS and MEAS<sub>SG</sub>. The former imposes a semantic plurality on its nominal complement, the latter a semantic singularity, encoded by the count form. Underlying this distinction are two modes of cardinality measurement: estimation and counting. Exclamatives concern cardinality measures based on estimation, not counting, and thus require the noun phrase to be plural.

---

\* Catherine Rudin’s work has greatly influenced my thinking about Bulgarian syntax. One of my earliest published papers, on the links between *wh*-movement and focus movement, was directly inspired by Rudin (1986), and my long-standing interest in the syntax of ‘than’-clauses and of free relatives also has its origins in Catherine’s foundational book. I have also been fortunate to collaborate with Catherine on the syntax of focus in *li*-questions, so for me that corner of Bulgarian grammar is shaped by her perspective as well. In this paper I address another topic in A’-dependencies in Bulgarian—the structure of exclamatives—although I consider the internal structure of exclamative phrases rather than their clausal syntax. Thanks to Iliyana Krapova for the initial discussion that convinced me to make my contribution to this volume be on this topic and to Steven Franks and two anonymous reviewers for their comments. Thanks also to Vesela Simeonova and Steven Franks for help with locating some of the references on the Bulgarian count form.

## 1. Introduction

English places restrictions on the type of *wh*-phrases that can appear in exclamatives, in contrast to questions. Only degree-denoting *wh*-expressions, in the form of *how* and *what* (e.g., *how tall*, *how many colors*, *what deserts*), can form exclamatives. The restricted distribution allows for the identification of syntactic and semantic properties that distinguish between the two types of *wh*-clauses. Bulgarian, on the other hand, allows a wider range of *wh*-expressions in exclamatives. In fact, the syntax of exclamative A'-dependencies in the language appears to be identical to that of *wh*-questions. This makes it difficult to identify properties that are specific to exclamatives and to establish cross-linguistic generalizations concerning this clause type.

There is one syntactic environment, however, where exclamatives in Bulgarian stand apart from their corresponding *wh*-questions, namely, number marking on the noun accompanying quantity-denoting *kolko* 'how many/much'. In combination with numerals some Bulgarian nouns appear in a special count form (*brojna forma*), distinct from the singular and plural. The same count form of the noun is used in 'how many' questions. Yet in 'how many' exclamatives the nouns can only be plural. Given that the form of the *wh*-expression *kolko* 'how many/much' is the same in exclamatives and questions, the differential number marking, count vs. plural, is puzzling.

The same number marking facts are found in another type of exclamative clause, based on demonstrative *tolkova* 'that many/much'. The English counterpart of this expression, as it is used in cardinality exclamatives, is *so many* rather than *that many*, which is used in declaratives. Arguably this is so because the *so*-proforms are degree denoting (e.g., *so tall*, *so many colors*, *such deserts*), the counterpart of the degree-denoting *wh*-expressions admissible in exclamatives. In Bulgarian, where the cardinality expression is *tolkova* 'that many/much' in both exclamatives and declaratives, there is nevertheless a distinction in the number marking on the noun. Whereas in declarative sentences, the relevant nouns take the count form in combination with *tolkova*, in exclamatives with *tolkova* the same nouns appear in the plural form.

This paper aims to provide an analysis of the distinction in number marking between cardinality exclamatives and their question and declarative counterparts in Bulgarian. I propose that cardinality exclamatives are formed with a null measure expression, which also links *mnogo* 'many/much' to NPs and which, just like *mnogo*, is only acceptable in cardinality structures with plural nouns. Cardinality questions and declaratives may include this measure expression, with the same result on nominal number marking, or they may be formed with a different measure expression which requires the count form. While analyzing cardinality expressions in exclamatives, questions, and declaratives, the paper also provides a description of the distribution of

the count form more generally and proposes that it is a semantically singular number marker.

### 1.1. *Wh*-Exclamatives and *Wh*-Questions

Word order in *wh*-exclamatives in Bulgarian mirrors that of *wh*-questions. The *wh*-phrase needs to be fronted to the left periphery of the clause, as seen in (1–4), and if there is more than one *wh*-phrase, all have to undergo movement, with superiority respected as in (2). Subjects cannot intervene between the *wh*-phrase and the verb; they can either precede the *wh*-phrase or appear post-verbally as in (3). What appears to be left-branch extraction is allowed in case the *wh*-pronoun originates in a predicative adjective, as in (4).

- (1) *Kakva kniga e napisala* {!/?}  
 what-kind book be<sub>3SG</sub> written<sub>FEM.SG</sub>  
 ‘What kind of book she wrote!’ / ‘What kind of book did she write?’
- (2) a. *Kolko studenti kolko statii publikovaxa* {!/?}  
 wh-quantity students wh-quantity articles published  
 ‘So many students published so many articles!’ /  
 ‘How many students published how many articles?’
- b. \**Kolko statii kolko studenti publikovaxa* {!/?}  
 wh-quantity articles wh-quantity students published
- (3) (Vie) *kolko knigi (\*vie) imate (vie)* {!/?}  
 you<sub>PL</sub> wh-quantity books have<sub>3PL</sub> you<sub>PL</sub>  
 ‘How many books you have!’ / ‘How many books do you have?’
- (4) *Kolko e visoka Marija* {!/?}  
 wh-quantity be<sub>3SG.PRES</sub> tall<sub>FEM.SG</sub> Maria  
 ‘How tall Maria is!’ / ‘How tall is Maria?’

While word-order is, of course, only the surface manifestation of the underlying syntactic structure, the facts in (1–4) are nevertheless suggestive of a close link, if not full identity, between the structure of the *A'*-dependency in *wh*-questions and *wh*-exclamatives (with differences in form limited to the type of complementizer and prosody). For extensive discussion of the syntax of *wh*-questions in Bulgarian see Rudin (1988, 1986/2013), among many others. Rudin (1986/2013) also includes some remarks on the syntax of *wh*-exclamatives in Bulgarian.



The facts of Bulgarian are not surprising. Many analyses posit structural commonalities between *wh*-questions and *wh*-exclamatives (Michaelis and Lambrecht 1996, Michaelis 2001, Zanuttini and Portner 2003, a.o.). Nevertheless, there are differences between the two types of *wh*-clauses, most notably having to do with the specific *wh*-expressions allowed: *what* and *how* form matrix exclamatives in English but *who*, *when*, *where*, *why* do not. Other languages, including Bulgarian, allow a wider class of *wh*-expressions in exclamatives. For a detailed discussion of cross-linguistic differences in *wh*-exclamatives see Villalba (2008) and Nouwen and Chernilovskaya (2015).

## 1.2. So-Exclamatives and *That*-Declaratives

Similarities also exist between *so*-exclamatives and declarative clauses containing demonstrative *that* in place of degree *so*. In Bulgarian both roles are played by demonstrative pronouns (sometimes called *th*-pronouns), resulting in an ambiguity. Just as the *wh*-clauses in (1–4) give rise to a *wh*-exclamative and a *wh*-question, the sentences in (5–8) are ambiguous between a *so*-exclamative and a *that*-declarative (putting aside intonation). In the latter case, the *th*-pronouns are referential, possibly accompanied by a pointing gesture in a demonstrative use, or interpreted anaphorically. Both exclamatives and declaratives allow the *th*-expression to be fronted to the left periphery of the clause but do not require such movement.

- (5) *Takava kniga e napisala* {!/.}  
that-kind book be<sub>3SG</sub> written<sub>FEM.SG</sub>

‘Such a book she wrote!’ / ‘She wrote that kind of book.’

- (6) *Tolkova studenti publikovaxa tolkova statii* {!/.}  
th-quantity students published th-quantity articles

‘So many students published so many articles!’ / ‘That many students published that many articles.’

- (7) *Vie imate tolkova knigi* {!/.}  
you<sub>PL</sub> have<sub>3PL</sub> th-quantity books

‘You have so many books!’ / ‘You have that many books.’

- (8) *Tolkova e visoka Maria* {!/.}  
th-quantity be<sub>3SG.PRES</sub> tall<sub>FEM.SG</sub> Maria

‘Maria is so tall!’ / ‘Maria is that tall.’

*So*-exclamatives are less commonly discussed in the literature, but they appear to be closely related to their *wh*-counterparts. For instance, in English

*so*-exclamatives conform to the degree restriction found in *wh*-exclamatives. In Bulgarian too, *so*-exclamatives and *wh*-exclamatives behave the same with respect to number marking in cardinality nominals, as discussed in the next section.

### 1.3. Number Marking in Cardinality Exclamatives, Declaratives, and Questions

Cardinality *wh*-exclamatives and *wh*-questions in Bulgarian differ, despite the identical form of their *wh*-pronouns. *Kolko doklada* ‘how many papers’ in (9a), where the noun has count morphology, yields a question only; *kolko dokladi* ‘how many papers’ in (9b), with a plural noun, forms an exclamative only, in the normative language (e.g., Stoyanov 1993: 108, Pašov 2011: 69).<sup>1</sup>

- (9) a. *Kolko doklada predstavixa studentite vi*  
 wh-quantity paper<sub>COUNT</sub> present<sub>3PL.PAST</sub> students<sub>DEF</sub> your  
 na konferencijata?  
 at conference<sub>DEF</sub>  
 ‘How many papers did your students present at the conference?’
- b. *Kolko dokladi predstavixa studentite vi*  
 wh-quantity paper<sub>PL</sub> present<sub>3PL.PAST</sub> students<sub>DEF</sub> your  
 na konferencijata!  
 at conference<sub>DEF</sub>  
 ‘How many papers your students presented at the conference!’

In cardinality declaratives and exclamatives, number morphology on the nominal plays the same disambiguating role. *Tolkova doklada* ‘that many papers’ in (10a), with a noun in the count form, results in a declarative only; *tolkova dokladi* ‘so many papers’ in (10b), with a plural noun, is interpreted as an exclamative in the normative language.<sup>2</sup>

- (10) a. *Studentite vi predstavixa tolkova doklada*  
 students<sub>DEF</sub> your present<sub>3PL.PAST</sub> th-quantity paper<sub>COUNT</sub>  
 na konferencijata.

<sup>1</sup> In the colloquial language (9b) can also be interpreted as a question (with suitable intonation). Still an asymmetry exists, because even colloquially (9a) cannot be interpreted as an exclamative.

<sup>2</sup> There is variation in the colloquial language, as in the case of *wh*-questions vs. *wh*-exclamatives, with the same asymmetry (see footnote 1). (10b) can be both a declarative and an exclamative, but (10a) can only be a declarative.

at conference<sub>DEF</sub>

‘Your students presented that many papers at the conference.’

- b. Studentite vi predstavixa *tolkova dokladi*  
 students<sub>DEF</sub> your present<sub>3PL.PAST</sub> th-quantity paper<sub>PL</sub>

na konferencijata!

at conference<sub>DEF</sub>

‘Your students presented so many papers at the conference!’

The differential number marking in exclamatives vs. questions and declaratives has been noted before (Stoyanov 1993, Cinque and Krapova 2007, Pašov 2011, Hristozova 2012, Franks this volume) but as far as I know there has been no formal analysis. This paper sets out to provide one.

## 2. Background on Number Marking in Nominals

Bulgarian masculine nouns make a three-way distinction in number: they have a so-called count form, in addition to a singular and a plural form. Feminine and neuter nouns do not have a count form. The plural inflection is varied: apart from the general and gender-neutral *-i* suffix, it involves sub-regularities and irregular suffixes, some of them specific to masculine nouns, as well as occasional changes in the stress pattern, e.g., *sin-sinové* ‘son(s)’ and vowel-zero alternations, e.g., *orél-orlí* ‘eagle(s)’. The count inflection is regular: it involves the *-a* suffix, which can predictably surface as *-ja*, and which doesn’t change the stress pattern or involve stem changes. The morphological distinction is productive, whether the nouns take the most general *-i* plural, the regular *-ove* plural for mono-syllabic masculine nouns, or one of the irregular plural suffixes, as seen in Table 1. There are only a few nouns that do not have a count form.<sup>3</sup>

<sup>3</sup> E.g., the personal noun *bašta* ‘father’ (regular plural *bašti*) and *djado* ‘grandfather’ (irregular plural *djadovci*) have no count forms. They are atypical masculine nouns because they end in a vowel. The nonpersonal *pât* ‘time, occasion’ also doesn’t have a count form (its regular plural is *pâti*).

**Table 1.** Number distinctions in masculine nonpersonal and personal nouns

	singular	plural	count	
nonpersonal	kon	koné	kónja	'horse'
	pât	pâtišta	pátja	'road'
	krak	kraká	kráka	'leg'
	cvjat	cvetové	cvjata	'color'
	slon	slónove	slóna	'elephant'
	orél	orlí	oréla	'eagle'
personal	mâž	mâžé	máža	'man'
	sin	sinové	sína	'son'
	kmet	kmétove	kméta	'mayor'
	pevéc	pevcí	pevéca	'singer'

Although the three-way distinction—singular, plural, and count—is morphologically productive with both types of masculine nouns, normative grammar prohibits the use of the count form with personal nouns. However, there is variation when it comes to the colloquial language, whether spoken or written. The next sections discuss the norms and the observed colloquial variation.

### 2.1. Norms: Masculine Nonpersonal Nouns

The count form is only used for masculine nonpersonal nouns in combination with exact and approximate numerals (except for *edin* 'one') and quantity *wh-/th*-expressions like indefinite *njakolko* 'several', negative indefinite *nikolko* 'no, not any', interrogative *kolko* 'how many', free relative *kolkoto* 'how(ever) many', and demonstrative/anaphoric *tolkova* 'that many', (11–12). The quantity expressions in (12) are all morphologically related.<sup>4</sup>

<sup>4</sup> Etymologically, *kolko* and *tolkova* (and their free relative counterpart *kolkoto*) are inherited from Proto-Indo-European interrogative/relative *\*k<sup>w</sup>o-* and the demonstrative *\*to-*, respectively, in combination with a morpheme that was likely derived from *\*h<sub>2</sub>el-* 'to grow' and which in Old Slavic was *-li-* (Georgiev 1979: 556–57). Latin cognates of the *wh-/th*-quantity pronouns are *quālis* and *talis*. The modern stem *-lko* is derived from *-li-* and the adjectival suffix *-k<sup>v</sup>* (Georgiev 1979: 556–57); cf. *velik<sup>v</sup>* 'big, great'. The prefixes *nja-* and *ni-*, added to the *wh*-quantity stem *ko-lko*, contribute existential and negative existential meaning, respectively (cf. *nja-kâde* 'some-where', *ni-kâde* 'no-where').

- (11) {deset / desetina / pet-šest} cvjata  
 ten ten-or-so five-six color<sub>COUNT</sub>
- (12) {njakolko / nikolko / kolko / kolkoto / tolkova}  
 several not-any wh-quantity<sub>Q</sub> wh-quantity<sub>FR</sub> th-quantity  
 cvjata  
 color<sub>COUNT</sub>

Quantifiers, and interrogative and demonstrative pronouns that do not encode cardinality require the plural or singular form, (13).

- (13) a. {edni / vsički / njakoi / koi / koito / tezi} cvetove  
 some<sub>PL</sub> all some<sub>PL</sub> which<sub>Q,PL</sub> which<sub>FR,PL</sub> these color<sub>PL</sub>
- b. {edin / vseki / njakoj / koj / kojto / tozi} cvjat  
 one each some<sub>SG</sub> which<sub>Q,SG</sub> which<sub>FR,SG</sub> this color<sub>SG</sub>

Perhaps surprisingly, given that they too are concerned with quantity, *mnogo* ‘many/much’, *malko* ‘few/little’, and their comparative and superlative forms combine with the plural rather than the count form, (14) (Stoyanov 1993: 108, Tasseva-Kurktchieva 2006, Cinque and Krapova 2007, Pašov 2011: 69–71, Hristozova 2012, Stateva and Stepanov 2016, Mikova 2017, Franks 2018).<sup>5</sup>

<sup>5</sup> Tasseva-Kurktchieva (2006) marks *nikolko* ‘no, not any’ in combination with a count noun as ungrammatical, and an anonymous reviewer agrees. My own judgments are different. Below are some relevant examples, from an internet search. No examples of the corresponding *nikolko levove* ‘no lev<sub>PL</sub>’ or *nikolko kone* ‘no horse<sub>PL</sub>’ were found. *Nikolko* in nominals is rare, and possibly this low frequency is behind the difference in judgments.

- (i) S *nikolko* leva njama da se uveličat zaplatite na  
 with not-any lev<sub>COUNT</sub> not-will SUBJ REFL increase<sub>3PL</sub> the-salaries of  
 lekarite.  
 the-doctors  
 ‘Doctors’ salaries will not increase by even a single lev.’ (*lev* is the Bulgarian currency)
- (ii) Za sto leva točno *nikolko* konja njama da kači,  
 for hundred lev<sub>COUNT</sub> exactly not-any horse<sub>COUNT</sub> not-will SUBJ add  
 garantiram ti.  
 guarantee<sub>1SG</sub> you  
 ‘For hundred leva, it will add no horsepower, I guarantee you.’ (on making improvements to a car)

- (14) a. {mnogo / poveče / povečeto / naj-mnogo} cvetove  
 many/much more most the-most color<sub>PL</sub>  
 b. {malko / po-malko / naj-malko} cvetove  
 few/little fewer/less the-fewest/the-least color<sub>PL</sub>

## 2.2. Norms: Masculine Personal Nouns

The norm for masculine personal nouns combining with numerals and *wh-/th-*quantity expressions is the plural form. The numerals themselves take a special suffix (e.g., unmarked *tri* vs. masculine personal *trima* ‘three’). The suffix *-(i)ma* is widely used for the lower numerals, ‘two’ to ‘six’, but less so for higher numerals.<sup>6</sup> Suffixed numerals are only available in the context of masculine personal nouns (see Hurford 2003, Cinque and Krapova 2007 for the suggestion that the suffix is a bound numeral classifier). The patterns are illustrated in (15–16).

- (15) trima mâže / dvama kmetove / petima lekari  
 three<sub>MASC.PERS</sub> man<sub>PL</sub> two<sub>MASC.PERS</sub> mayor<sub>PL</sub> five<sub>MASC.PERS</sub> doctor<sub>PL</sub>
- (16) {njakolko / ... / kolko(to) / tolkova} {mâže / kmetove /  
 several wh-quantity<sub>Q(FR)</sub> th-quantity man<sub>PL</sub> mayor<sub>PL</sub>  
 lekari}  
 doctor<sub>PL</sub>

When combining with non-numeral quantifiers and *many* and *few* the personal and nonpersonal nouns do not differ, and appear in the plural form. Compare (17) with (14).

- (17) {mnogo / ... / malko / ...} {mâže / kmetove / lekari }  
 many/much few/little man<sub>PL</sub> mayor<sub>PL</sub> doctor<sub>PL</sub>

## 2.3. Variation: Masculine Non-Personal Nouns

Non-personal nouns show considerable variation in departing from the norms in colloquial registers. The use of the plural instead of the count morphology for nonpersonal nouns is noted even in grammars which otherwise tend to be prescriptive in favor of more formal styles. Stoyanov (1993: 108) lists the doublet forms in (18), and in fact notes that the plural form is preferred to the

<sup>6</sup> In the case of *sedem* ‘seven’ and *osem* ‘eight’ where, for phonological reasons, the form is rarely, if ever, used. Sometimes the approximative form ending in *-ina* is used instead of the masculine personal form for these numerals.

count form. All such nouns use the general *-i* suffix to form the plural; indeed, it seems to me that the irregular plural suffixes are more easily blocked by the count suffix. But at best this is a tendency, and “incorrect” plural forms can be found with all types of masculine nonpersonal nouns; see (19) and (20a–b), from the Bulgarian National Corpus (BNC)<sup>7</sup> and (20c), from an internet search, where the nouns form irregular plurals. Importantly, these are not isolated examples, and they can appear with numerals (apart from *dva* ‘two’<sup>8</sup>) and with quantity *wh-/th*-expressions.<sup>9</sup>

- (18) tri {{prozoreca / prozorci} / {orela / orli} / {ovena /  
three window<sub>COUNT</sub> window<sub>PL</sub> eagle<sub>COUNT</sub> eagle<sub>PL</sub> ram<sub>COUNT</sub>  
ovni}}  
ram<sub>PL</sub>

- (19) a. Ne moga dori da kaža kolko kone smenix  
not can<sub>1SG</sub> even subj say<sub>1SG</sub> wh-quantity horse<sub>PL</sub> change<sub>1SG</sub>  
po pătja.  
on road<sub>DEF</sub>

‘I can’t even say how many horses I changed while I was on the road.’

- b. dostavjal xrana, paša i pari za 8000 duši  
bring<sub>3SG.PAST</sub> food fodder and money for 8000 people  
i tolkova kone  
and th-quantity horse<sub>PL</sub>

‘he used to bring food, fodder, and money for 8000 people and that many horses.’

- c. Polovinata ot konvoja kapna ot umora, njakolko  
half of konvoj<sub>DEF</sub> drop<sub>3SG.PAST</sub> from fatigue several  
kone padnaxa.  
horse<sub>PL</sub> fall<sub>3PL.PAST</sub>

‘Half of the convoy suffered extreme fatigue; several horses fell down.’

<sup>7</sup> <http://dcl.bas.bg/bulnc/en/>

<sup>8</sup> The count form is derived historically from the dual form. Likely this is the reason that the count form is obligatory with *dva*. In addition to number, *dva* also agrees with the NP in gender. It is unambiguously marked masculine (the feminine and neuter form is *dve*), and is the only numeral apart from *edin* ‘one’ to be marked for gender.

<sup>9</sup> I find all examples, from (18) to (22), acceptable, provided the discourse is colloquial.

- (19) d. Namerixa se samo deset kone za [v]sički ni, a  
 found<sub>3PL.PAST</sub> REFL only ten horse<sub>PL</sub> for all us and  
 poveče njamaše.  
 more there-wasn't  
 'Only ten horses were found for all of us, and there were no more.'
- (20) a. *Kolko pâtišta vodjat do stenite na krepostta?*  
 wh-quantity road<sub>PL</sub> lead<sub>3PL.PRES</sub> to walls<sub>DEF</sub> of fortress<sub>DEF</sub>  
 'How many roads lead to the walls of the fortress?'
- b. *Sâštestvuvat tolkova pâtišta kolkoto odelni duši*  
 exist<sub>3PL.PRES</sub> th-quantity road<sub>PL</sub> wh-quantity separate souls  
 'However many different souls there are, there are that many roads.'
- c. *Sofia e razpoložena čudesno ... na krâstopât pone na*  
 Sofia is situated wonderfully on crossroad at-least of  
*sedem, osem pâtišta.*  
 seven eight road<sub>PL</sub>  
 'Sofia is wonderfully situated [...] on the crossroad of at least seven, eight roads.'

Such variation between count and plural forms can be found even within the same sentence; see (21) from Hristozova (2012: 307) where the 'incorrect' plural form is used in one case (plural *vârrove* instead of count *vârxa* 'summits') but not in another (count *kontinenta* 'continents').

- (21) *Alpinistât Džordan e pokoril sedemte vârrove na*  
 the-alpinist Jordan is conquered the-seven summit<sub>PL</sub> of  
*sedemte kontinenta.*  
 the-seven continent<sub>COUNT</sub>  
 'The mountain climber Jordan conquered the seven summits of the seven continents.'

Another factor for the acceptability of the plural form is the presence of intervening adjectives (Pašov 2011: 70, Hristozova 2012, Stateva and Stepanov 2016). (22) is a "violation" (Hristozova 2012: 308) where the plural *centrove* 'centers' is used instead of the normative count *centâra*. The attributive modifiers are plural. Stateva and Stepanov (2016) treat such cases as agreement attraction errors and show that the use of the plural form increases with the increase in the number of intervening adjectives (with distance measured in terms of structural nodes).



- (22) *njakolko golemi obštinski centrove*  
 several large<sub>PL</sub> county<sub>PL</sub> center<sub>PL</sub>  
 ‘several large county centers’

To conclude, contexts with numerals and *wh-/th*-quantity expressions readily allow variation between the count and the plural form in the colloquial language, although the count form is prescribed for nonpersonal nouns. Nevertheless, the interchangeability is unidirectional: the contexts in (13a) and (14) require the plural and do not permit the count form (occasional examples with the count form are very rare and possibly reflect idiolectal variation).<sup>10, 11</sup>

#### 2.4. Variation: Masculine Personal Nouns

There is also variation in the number marking of masculine personal nouns with numerals and *wh*- and *th*-quantity expressions in the colloquial language. Often one finds the count form instead of the plural (e.g., in Pašov 2011: 69).

<sup>10</sup> In partitives a count form may appear in place of the expected plural. Consider (i) (Pašov 2011: 70), where the NP is not directly selected by the numeral. According to the norms, the NP needs to be plural (*prepisi* ‘copies’). On the other hand, Hristozova (2012: 38) gives the partitive in (ii), with a plural NP (*uroci* ‘lessons’), as an example of an incorrect use, suggesting that the count form *uroka* ‘lessons’ should be used instead. Clearly there is variation in this area as well. (I find both forms acceptable.)

- (i) *Originalât e zaguben, no se pazjat pet ot negovite prepisa.*  
 the-original is lost but REFL keep<sub>3PL.PRES</sub> five of its copy<sub>COUNT</sub>  
 ‘The original is lost but five of its copies are being preserved.’
- (ii) *pet ot naj-trudnite uroci*  
 five of most-difficult<sub>DEF</sub> lesson<sub>PL</sub>  
 ‘five of the most difficult lessons’

<sup>11</sup> One can find examples of the use of the count form with *mnogo* ‘many/much’ as well as with *njakoi* ‘some’ and *tezi* ‘these’, though they are very few. The examples in (i) are from an internet search (they do not sound acceptable to me). A search of BNC yielded no such forms, but had many examples of *many* with plural masculine nonpersonal (and personal) nouns, as the norm dictates.

- (i) a. ... *polučete kato podarâk zabavna ximikalka s mnogo*  
 receive<sub>2PL.IMP</sub> as present fun pen with many/much  
*cvjata v neja.*  
 color<sub>COUNT</sub> in it  
 ‘... receive as a free gift a fun pen that has many colors.’
- b. *Bojan Kostov pâk e s naj-mnogo glasa— 331...*  
 Bojan Kostov INTERJ is with est-many/much vote<sub>COUNT</sub> 331  
 ‘Bojan Kostov has the most votes: 331.’

The count forms in (23–24), from BNC, are acceptable for me, as would be their normative plural forms (with the general plural *-i*). Examples where the count form substitutes an irregular plural form can be found, but they are rarer, and to me at least they do not sound that great, e.g., the count form in (25), from BNC, which is used instead of the irregular plural *mâže* ‘men’. Monosyllabic nouns that take the regular *-ove* plural (*popove* ‘priests’) sound more acceptable to me when used in their count form, as in (26), from BNC.

- (23) I kakvo šte praviš ti s tvoite sto vojnika  
and what will do<sub>2SG</sub> you with your hundred soldier<sub>COUNT</sub>  
sreštu sto xiljadi?  
against hundred thousand  
‘And what will you do with your hundred soldiers against a hundred thousand soldiers?’
- (24) Tja pokani *njakolko studenta* da posetjat klinikata ì...  
she invited several student<sub>COUNT</sub> SUBJ visit clinic<sub>DEF</sub> her  
‘She invited several students to visit her clinic.’
- (25) Kojto piše za istorijata na Oxrid, trjabva edro da  
who<sub>FR</sub> write about history<sub>DEF</sub> of Ohrid must notably SUBJ  
otbeleži *njakolko mâža*, koito opropastixa vsičkite si  
recognize several man<sub>COUNT</sub> who ruin<sub>3PL.PAST</sub> all<sub>DEF</sub> REFL  
kapitali po narodnoto delo.  
capital on national<sub>DEF</sub> cause  
‘Whoever writes about the history of Ohrid should strongly recognize several men who spent all their possessions for the national cause.’
- (26) *Njakolko popa* otslužili molitvi ...  
several priest<sub>COUNT</sub> serve<sub>3PL.PAST</sub> prayers  
‘Several priests said prayers...’

Hristozova (2012: 307) gives examples where in the same sentence one numerically quantified masculine personal NP has the “wrong” count form (count *sina* instead of plural *sinove* ‘sons’) while others have the normative plural form.

- (27) Djado            Teodosij ima *dvama*            *sina*,            četirima  
 Grandfather Teodosij has two<sub>MASC.PERS</sub> son<sub>COUNT</sub> four<sub>MASC.PERS</sub>  
*vnuci*,            četirima            *pravnuci*            i            *dvama*  
 grandson<sub>PL</sub> four<sub>MASC.PERS</sub> great-grandson<sub>PL</sub> and two<sub>MASC.PERS</sub>  
*prapavnuci*.  
 great-great-grandson<sub>PL</sub>  
 ‘Old Teodosij has two sons, four grandsons, four great grandsons, and  
 two great-great-grandsons.’

I have not found attested examples of the count form of personal nouns after *mnogo* ‘many’. There doesn’t seem to be variation of the type that nonpersonal nouns allow marginally (see fn. 11).

## 2.5. Summary

Masculine nouns make a three-way morphological distinction between singular, plural, and count forms. The count form is only available in combination with numerals and with *wh-/th-*quantity expressions. In such contexts count inflection is the norm for nonpersonal nouns, while plural inflection is the norm for personal nouns, but often the two are used interchangeably with both types of nouns in the colloquial language. *Mnogo* ‘many/much’ in all its degree forms combines with plural nouns, personal or nonpersonal. Table 2, on the opposing page, gives a summary (\*/? marks the restricted, possibly idiolectal, variation; √norm marks the prescribed norm, and the rest reflects common use).

Mikova (2017) notes a change in normative grammars with respect to the prescribed number marking on personal nouns. In 1945 both count and plural forms were listed as acceptable, in 1983 a preference was expressed for the plural form, particularly after numerals with the suffix *-(i)ma*, and in 2012 only the plural form was considered acceptable. On the other hand, the normative grammars consistently recognize only the count form of nonpersonal nouns as “correct,” though in 1983 the forms in (18) were listed as doublets. We can conclude that a change in progress has been underway, whereby an older grammar of count marking on personal nouns is replaced by a grammar where such nouns are marked plural. A similar change must be underway in the case of nonpersonal nouns, though the progress of the newer plural-marking grammar has been slower or more recent and is not yet acknowledged in normative grammars. The present-day variation in the colloquial language reflects the effects of the older and newer grammars in competition.

**Table 2.** Number distinctions for masculine nouns including normative and colloquial varieties

	nonpersonal <i>cvjat</i> 'color'		personal <i>kmet</i> 'mayor'	
	<i>cvjata</i> <sub>COUNT</sub>	<i>cvetove</i> <sub>PL</sub>	<i>kmeta</i> <sub>COUNT</sub>	<i>kmetove</i> <sub>PL</sub>
numerals	√norm	√	√	√norm
{ <i>nja</i> -/ <i>ni</i> -} <i>kolko</i> 'several' / 'not any'	√norm	√	√	√norm
<i>kolko</i> ( <i>to</i> ) 'how many' (free rel.)	√norm	√	√	√norm
<i>tolkova</i> 'that many', 'so many'	√norm	√	√	√norm
( <i>naj</i> -) <i>mnogo</i> 'many' ('the most')	*/?	√norm	*	√norm
<i>poveče</i> ( <i>to</i> ) 'more' ('most')	*/?	√norm	*	√norm

### 3. Exclamatives

Against the background in section 2, consider again number marking in exclamatives vs. questions and declaratives. Because neuter *cvete* 'flower' does not have a count form, the fragments in (28a) and (28b) are ambiguous. But for masculine nonpersonal *cvjat* 'color' there is no ambiguity in the normative language: the plural brings about the exclamative reading, (29), while the count form yields the question or declarative reading, (30).

- (28) a. *Kolko cvetja* {!/?}  
wh-quantity flower<sub>PL</sub>  
'How many flowers {!/?}'
- b. *Tolkova cvetja* {!/?}  
th-quantity flower<sub>PL</sub>  
'So many flowers!' / 'That many flowers.'
- (29) a. *Kolko cvetove!*  
wh-quantity color<sub>PL</sub>  
'How many colors!'
- b. *Tolkova cvetove!*  
th-quantity color<sub>PL</sub>  
'So many colors!'
- (30) a. *Kolko cvjata?*  
wh-quantity color<sub>COUNT</sub>  
'How many colors?'
- b. *Tolkova cvjata.*  
th-quantity color<sub>COUNT</sub>  
'That many colors.'

In colloquial usage, there is variation between the plural and count form in questions and declaratives, as seen in (19) and (20). Two other examples are (31), from an internet search, and (32), from BNC, with a plural (*cvetove*) instead of a count noun (*cvjata*) of *cvjat* ‘color’.

- (31) Razgledajte kartinkata i vnimatelno prebrojte *kolko*  
 look-at picture<sub>DEF</sub> and carefully count wh-quantity  
*cvetove* različavate.  
 color<sub>PL</sub> differentiate  
 ‘Look at the picture and carefully count how many different colors  
 you see.’
- (32) Roklite na ženite bjaxa našareni s *tolkova*  
 dresses<sub>DEF</sub> of women<sub>DEF</sub> were colored with th-quantity  
*cvetove*, kolkoto izobšto imaše.  
 color<sub>PL</sub> wh-quantity<sub>FR</sub> even there-were  
 ‘The women’s dresses had that many colors on them, however many  
 colors even existed.’

In light of (19), (20), (31), and (32), it is clear that in the colloquial language the plural form of nouns allows both an exclamative and a question or a declarative reading of cardinal *wh-/th*-expressions. The variation is one-sided though. While questions and declaratives allow plural nouns in lieu of count nouns, exclamatives do not allow count nouns. Table 3, on the opposing page, gives a summary.

Since the exclamative and interrogative *kolko* ‘how many’ are morphologically the same, it is surprising that they differ with respect to the number marking on the noun they combine with. The same holds for the pair of *th*-expressions: they are identical in form, yet declarative *tolkova* ‘that many’ combines with the count form, or optionally with the plural, while the *so*-exclamative only allows the plural. What sets exclamatives apart from their interrogative/declarative counterparts? Exclamative *wh-/th*-expressions also stand out among the wider class of nominal cardinality expressions such as those with numerals, numeral quantifiers *njakolko* ‘several’ and *nikolko* ‘no, not any’, and free relative *kolkoto* ‘how many’, which too allow variation in number marking.

**Table 3.** Number distinctions for masculine nouns in questions and declaratives vs. exclamatives including normative and colloquial varieties

	nonpersonal <i>cvjat</i> 'color'		personal <i>kmet</i> 'mayor'	
	<i>cvjata</i> <sub>COUNT</sub>	<i>cvetove</i> <sub>PL</sub>	<i>kmeta</i> <sub>COUNT</sub>	<i>kmetove</i> <sub>PL</sub>
<i>kolko</i> 'how many' (question)	√norm	√	√	√norm
<i>tolkova</i> 'that many' (declarative)	√norm	√	√	√norm
<i>kolko</i> 'how many' (exclamative)	*	√norm	*	√norm
<i>tolkova</i> 'that many' (exclamative)	*	√norm	*	√norm

### 3.1. A Null *Mnogo* 'Many'?

A possible line of analysis is that exclamatives contain a nonovert *MNOGO* 'many/much'. On this view, in (29) it is not the *wh-/th-*pronoun that determines the number marking on the noun, but a nonovert *MNOGO*, as in (33b). The count form is unacceptable, because nonovert *MNOGO*, like its overt counterpart, does not combine with count-marked nouns. The covert element is in small caps.

- (33) a. *question/declarative: count or plural*  
 {*kolko* / *tolkova*} {*cvjata* / *cvetove*}  
 wh-quantity th-quantity color<sub>COUNT</sub> color<sub>PL</sub>
- b. *exclamative: plural*  
 {*kolko* / *tolkova*} *MNOGO* *cvetove*  
 wh-quantity th-quantity many/much color<sub>PL</sub>

A nonovert *MNOGO* may also seem attractive because of a semantic property that is obvious enough to be noted in traditional grammars (Pašov 2011: 69). The exclamatives in (29) convey that the number of colors is large (for the context at hand); the question and the declarative in (30) do not carry such an implication. If a child has a set of 120 colored pencils but draws a picture that only has the colors red, blue, and yellow, the exclamatives in (29) would not be felicitous. However, if told that the child drew a picture with a very small number of colors, it would be felicitous to ask the question in (30a) or to answer it with the declarative in (30b). This meaning component—exceeding the standard degree on the relevant scale—is known as evaluativity (Rett 2015).

Cardinality exclamatives are always evaluative while cardinality questions and declaratives need not be. A covert *MNOGO* can be the reason exclamatives are evaluative. The positive form of ‘many’ (‘many colors’) is evaluative, conveying that the number of colors is large.

An overt intensifier *mnogo* may be added to cardinality exclamatives without a noticeable difference in meaning, since cardinality exclamative phrases already express a high number. The same is true about the addition of the degree intensifier *very* in English.<sup>12</sup>

- (34) a. Kolko *mnogo* cvetove!  
 wh-quantity many/much color<sub>PL</sub>  
 ‘How (very) many colors!’
- b. Tolkova *mnogo* cvetove!  
 th-quantity many/much color<sub>PL</sub>  
 ‘So (very) many colors!’

The intensifier *mnogo* can also be added to questions and declaratives but with a noticeable effect on meaning. Its presence contributes evaluativity.<sup>13</sup> And, in apparent support of the idea that the obligatory plural marking in cardinality exclamatives is due to a null *mnogo*, and that the overt intensifier *mnogo* requires the presence of a null *mnogo*, questions and declaratives with an overt *mnogo* only accept the plural form of nouns.<sup>14</sup>

- (35) a. Kolko *mnogo* cvetove?  
 wh-quantity many/much color<sub>PL</sub>  
 ‘How very many colors?’

<sup>12</sup> The degree intensifier ‘very’ is also *mnogo* in Bulgarian.

(i) Marija e *mnogo* visoka i osven tova tiča *mnogo* bârzo.  
 Maria is very tall and besides this runs very fast  
 ‘Maria is very tall and she also runs very fast.’

<sup>13</sup> This effect of *mnogo* ‘many’ in questions was noted by Rett (2008: 111, 2015: 150). The proposal in Rett (2006) is not meant to account for this effect.

<sup>14</sup> As noted in fn. 11, *mnogo* may appear with the count form in rare cases, and the same holds for *kolko/ tolkova mnogo* expressions; see (i), from an internet search yielding four distinct results for the string *kolko mnogo cvjata*.

(i) Vižite ošte kolko *mnogo cvjata* očkavam.  
 see<sub>2PL.IMP</sub> more wh-quantity many color<sub>COUNT</sub> expect<sub>1SG</sub>  
 ‘See how many more colors I am expecting.’

- (35) b. Tolkova *mnogo* cvetove.  
 th-quantity many/much color<sub>PL</sub>  
 ‘That very many colors.’

Positing a null *MNOGO* may resolve the issue of number marking, but it does not immediately account for evaluativity. Positive degree adjectives in general (e.g., *tall*) are evaluative, and the role of contributing this aspect of meaning is often attributed to the presence of a null degree *POS-*, in complementary distribution to comparative *-er*, superlative *-est*, and degree *how* and *that* (e.g., Cresswell 1976). *Pos-* encodes the meaning of exceeding a contextual standard on the scale associated with the adjective. Thus, *Mary is pos-tall* means that Mary’s height is above the standard of height in the given context; *pos-many colors* expresses that the number of colors exceeds the contextual standard for a large number. But in *wh-/th*-expressions, *kolko* and *tolkova* should preclude the presence of *pos-* as they fill the same position. More needs to be said about the internal structure of such expressions before we could adopt the idea of a null *MNOGO*.

Moreover, evaluativity is a general feature of all *wh*-exclamatives that contain degree expressions and not just of cardinality exclamatives. The examples in (36) are evaluative, with or without *very*. The same would hold for their counterparts in Bulgarian (not illustrated here).

- (36) a. How (very) much wine we drank!  
 b. How (very) tall you are!

Clearly, a closer look is needed at the issue of evaluativity in exclamatives and its possible source. But first I will consider another precedent for positing null structure in exclamatives.

### 3.2. The Degree Restriction in Exclamatives

A notable feature of English *wh*-exclamatives is that they accept only a subset of *wh*-pronouns: *what*, manner *how*, and degree *how* (e.g., *how tall*, *how many*). As Rett (2011) notes, potential exclamatives with perfectly natural interpretations are ungrammatical (see (37), her ex. (14d–f)).

- (37) a. \*Who that lovely woman married! (...He’s so acerbic!)  
 b. \*Where she goes out partying! (...It’s so seedy!)  
 c. \*When she gets out of bed in the morning! (...I eat lunch at that hour!)



If one wanted to express surprise at the identity of the individual who the lovely woman married, (37a) should be the way to do it, but the exclamative is not well-formed. The corresponding *wh*-question would be grammatical.<sup>15</sup> Similar considerations apply to the other examples in (37).

Michaelis and Lambrecht (1996) and Rett (2011) suggest that English exclamatives are restricted to degree interpretations, which precludes *who*, *where*, *when*, and *why*: these *wh*-pronouns do not range over degrees. In contrast, *what* ranges over both individuals and degrees, and while only the individual reading is available in questions (Zanuttini and Portner 2003, Rett 2011, 2015), only the degree reading is found in *wh*-exclamatives. Rett (2011) gives the example in (38) (*her* (17)). Even in the absence of a degree predicate, its content is about a gradable quality of the desserts, e.g., being tasty or exotic. It cannot express surprise that John baked a particular set of deserts (a baklava and a tikvenik) instead of an expected other set (an apple pie and a blueberry pie).

(38) (My,) What desserts John baked!

The degree restriction extends to all (matrix) exclamatives, even those that do not involve a *wh*-dependency, yet it is not present in declarative exclamations, suggesting that it cannot be attributed to the speech act of exclamation.<sup>16</sup> The structure of exclamatives must be responsible. Rett (2011) proposes that English exclamatives contain a null measure function (M-OP) mapping individuals to degrees in the absence of an overt degree predicate (e.g., *what M-OP desserts*). M-OP is valued contextually. In (38) M-OP can be associated with a scale of deliciousness, richness, etc.

<sup>15</sup> However, these *wh*-expressions are acceptable in embedded contexts under what are sometimes called “exclamative predicates”, as noted in Michaelis (2001), Rett (2011), Nouwen and Chernilovskaya (2015). Analyzing these as embedded questions, and restricting exclamatives to only matrix contexts, is appealing, but is not without difficulties.

- (i) a. You wouldn’t believe who that lovely woman married!
- b. I am amazed where she goes out partying!
- c. You wouldn’t believe when she gets out of bed in the morning!

We can identify a scalar meaning, yet it does not concern gradable properties of individuals, but a scale of likelihood. Thus (ia) expresses surprise that the woman married the specific person (an individual, not a degree reading) and additionally conveys that the person is the least likely for her to have married among the relevant alternatives.

<sup>16</sup> Rett (2011: ex. (24b) and (25b)) shows that only the declarative exclamation in (ii) can express surprise at the fact that Sue likes banana bread; in (i), in the form of a question, the surprise is about the degree of her love of banana bread.

- (i) (Boy,) Does Sue like banana bread!
- (ii) (My,) Sue likes banana bread!

There is a suggestive link between the putative null *mnogo* in cardinality exclamatives and the null measure function M-OP. Yet there are also difficulties with equating the two. First, a quantity measure function must already be present in all cardinality *wh-/th*-expressions as well as in nominal phrases with numerals. Second, the conditions under which a null M-OP is obligatorily triggered in exclamatives but not in questions or declaratives remain unclear, and Rett (2011), while noting the problem, does not offer a solution. And finally, the degree requirement does not hold for exclamatives cross-linguistically, with some languages allowing *wh*-pronouns that do not range over degrees to form exclamatives (Nouwen and Chernilovskaya 2015). The last issue is particularly relevant, because Bulgarian allows a wider range of *wh*-exclamatives. Below are some examples of exclamatives (from an internet search) that would not be well-formed as English matrix exclamatives, and that do not express a surprise at the degree to which an individual has a gradable property but rather at the identity of the referent of the *wh*-expression.

- (39) Bože gospodi, *kakvo* namerix v edin arxiviran doklad na  
 God Lord what find<sub>1SG.PAST</sub> in one archived report of  
 Johanes Han ...  
 Johannes Hahn

Lit. 'Oh my God, what I found in an archived report of Johannes Hahn!'

- (40) Lele *koj* ni bie stanaxme za smex!  
 INTERJ who us beat<sub>3SG.PRES</sub> become<sub>1PL.PRES</sub> for laughter

Lit. 'Wow, who beat us! We've become a laughing stock!'

(on the occasion of the loss of the Bulgarian national volleyball team to Germany)

Given these facts, it is difficult to maintain that a null measure function like M-OP plays a central role in Bulgarian *wh*-exclamatives. If it is available in the first place, it does not have to be present. Plus, our specific concerns are with cardinality expressions, and these must independently involve a measure function to turn the predicate of individuals (the denotation of the NP) into a predicate of degrees whose degree argument is then saturated by numerals or bound by degree quantifiers.

### 3.3. Back to Evaluativity

Rett (2015: 163,167) notes that in addition to positive adjectives like *tall* and positive quantity expressions like *many*, indefinite quantity nominals like those

in (41) (her ex. (33a)) are also evaluative. This underscores the point that evaluativity should not be built into the semantics of the nonovert *POS-* morpheme, commonly posited in the representation of positive degree expressions, as here the degree argument is bound by an overt *some* so *POS-* cannot be present.

(41) Doug owns some number of shoes.

In earlier work, Rett (2008) characterized evaluativity as the contribution of a null degree modifier *eval*, limiting the role of the degree quantifier in positive *tall* and *many* to existential quantification (the counterpart of *some*). Rett (2015) argues instead that evaluativity arises as an implicature in the case of expressions that would otherwise be trivial: if *tall* simply means to have a degree of height, it, like *some number*, would be uninformative. The strengthening of meaning results in the interpretation that the degree predicate holds to a high degree. Evaluativity is pragmatically accomplished rather than lexically encoded in the degree quantifier *POS-* or modifier *eval*.

With respect to exclamatives, Rett (2015) proposes that evaluativity also arises as the result of an implicature. Exclamatives contribute the meaning of speaker's surprise and so their content needs to be noteworthy. In that context, the literal semantic content of the exclamative is strengthened to a meaning concerning an unusually high degree. Consider the illustration in (42) (Rett 2011, 2015). In combination with an illocutionary exclamative operator, existential quantification obtains over the degree variable contributed by the measure expression *many*. The weak meaning is overcome through an implicature and is strengthened to degree intensification.

- (42) a. How many shoes you have!  
 b. EXCL-FORCE ( $\exists d$  [you have *d*-many shoes])

The upshot of this discussion is that no extra structure needs to be posited in degree exclamatives to account for their evaluativity, according to Rett (2011, 2015). However, this does not help us resolve our original question as to why exclamatives differ from other *wh-/th-*quantity expressions in requiring the plural form of NPs and not accepting the count form. Pragmatic strengthening cannot directly be responsible for the selection of one type of number inflection over another.

The next section aims to examine closely the semantic composition of cardinality expressions, to see whether a null *MNOGO* may be posited for exclamatives, and if so, to elucidate the details of its relation to the measure functions independently found in expressions with overt *mnogo* as well as with numerals and with question/declarative *wh-/th-*quantity pronouns.

## 4. The Morpho-Semantics and Syntax of Cardinality Expressions

### 4.1. Q-Adjectives and *Wh-/Th*-Quantity Expressions

*Many* and *much*—called ‘Q(uality)-adjectives’ in Bresnan (1973)—play a measurement role. A common approach to their semantics (e.g., Hackl 2009) posits that they incorporate a measure function: they combine with a predicate of individuals (the denotation of the NP *color(s)*), and they map an individual of which the predicate is true (a portion of color or a plurality of colors) to a degree, i.e., to a unit of measurement on a cardinality (*many*) or noncardinality (*much*) quantity scale. On this view, the lexical semantics of Q-adjectives is very similar to that of adjectives like *tall*. However, the distribution of Q-adjectives is broader: for instance, they also appear as differentials in comparatives (*much taller, many more colors*) and in other environments where adjectives cannot (Schwarzschild 2006, Rett 2014, 2018, Solt 2015, among others). For this reason, the role of introducing the measure function in quantity nominals is sometimes attributed to a null element rather than to the Q-adjectives. I represent the null element in the extended nominal functional sequence as Meas(ure) with the lexical semantics in (43) (essentially as in Rett 2018: ex. (29) and similar to Solt 2015: ex. (35); cf.  $\text{MON}^0$  in Schwarzschild 2006). Meas includes an underspecified measure function  $m$ , which yields cardinality or noncardinality measures depending on other properties of the nominal structure, e.g., number marking on NP, the type of binder of the degree argument. A semantically plural NP would typically determine that the measurement involves the dimension of number rather than any other quantity dimension.

- (43)  $[[\text{Meas}]] = \lambda P_{\langle e,t \rangle} \lambda d \lambda x [P(x) \ \& \ \mu(x) \geq d]$
- a.  $[[\text{Meas colors}]] = \lambda d \lambda x [\text{colors}(x) \ \& \ \mu(x) \geq d]$     where  $\mu = \text{number}$
- b.  $[[\text{Meas color}]] = \lambda d \lambda x [\text{color}(x) \ \& \ \mu(x) \geq d]$     where  $\mu = \text{volume / surface size}$

The expressions in (43a, b) have the type of gradable adjectives,  $\langle d, et \rangle$ . Therefore, in principle, degree quantifiers like *-er, -est*, and *wh-/th*-pronouns that can range over degrees, like *how* and *that*, could combine directly with [-MEAS wine(s)] and bind its degree variable. This, I suggest, is the case for Bulgarian *wh-/th*-quantity pronouns *kolko* and *tolkova*; see (44). The interpretation of this structure is straightforward: *kolko* is a *wh*-degree indefinite (of the type of individuals, predicates, or quantifiers—all approaches to *wh*-words have been pursued in the literature and we do not need to make a choice here), and *tolkova* denotes a definite degree. The *wh-/th*-expressions themselves are not specified for cardinality or noncardinality dimensions; they are compatible with both interpretations in (43a, b). The individual argument of MEAS is

existentially bound by a nonovert determiner or a mechanism of existential closure, as commonly assumed.

(44) [*kolko / tolkova* [Meas NP]]

English *how* and *that* may not saturate the degree argument of [Meas NP], and neither may English degree quantifiers *-er* and *-est*, nor their Bulgarian counterparts *po-* and *naj-*. I will assume here that the reason is morpho-syntactic, concerning the category distinction between NPs and the expressions that can appear in their extended projections, such as lexical adjectives and Q-adjectives. So even though [Meas NP] has the same  $\langle d, et \rangle$  type as *tall*, its nominal category precludes the merge of *how*, *that*, and degree quantifiers, which otherwise combine with *tall*. To appear in cardinality nominal structures, these expressions need to merge with a Q-adjective first.

Q-adjectives have the semantics in (45) (cf. Schwarzschild 2006: ex. (124), Solt 2015: ex. (32), Rett 2018: ex. (25), which differ in the details but share key aspects of this meaning)—they are gradable predicates of degree intervals, i.e., predicates of intervals with an extra degree argument.

(45) [[*many / much*]] =  $\lambda d \lambda D_{\langle d, t \rangle}$  [the size of  $D \geq d$ ]

The Q-adjective phrase merges with [MEAS NP], as in (46). Before the Q-adjective phrase and [MEAS NP] compose semantically, the individual variable of the latter needs to be existentially bound. The interpretation of the structure in (46) is as in (47); compare with (43a, b).

(46) [[Q-adjP *many/much*] [Meas NP]]

(47) a. [[[Q-adjP *many*] [Meas colors]]]  
=  $\lambda d'$  [the size of  $\{d: \exists x [\text{colors}(x) \ \& \ |x| \geq d]\} \geq d'$ ]

b. [[[Q-adjP *much*] [Meas color]]]  
=  $\lambda d'$  [the size of  $\{d: \exists x [\text{color}(x) \ \& \ \mu(x) \geq d]\} \geq d'$ ]

*How*, *that*, and the degree quantifiers saturate the degree argument of Q-adjectives. The structure behind English cardinality *wh*-questions and declaratives is as in (48a), in contrast to their Bulgarian counterparts in (44). In positive forms of Q-adjectives such as *many colors* and *much color* a POS- degree quantifier merges as the degree argument of the Q-adjective, as in (48b), and in comparative and superlative forms, *-er* and *-est* do so.

- (48) a. [[Q-adjP *how/that many/much*] [Meas NP]]  
 b. [[Q-adjP POS- *many/much*] [Meas NP]]

The Bulgarian Q-adjective *mnogo* ‘many/much’ may appear in the structure in (48b) and be interpreted in the same way as its English counterpart, and the same is true for its comparative and superlative forms. However, the *wh-/th-* quantity expressions *kolko* and *tolkova* do not appear in the structure in (48a) with an overt *mnogo*, only in the structure in (44). Recall that when *mnogo* surfaces overtly with *kolko* and *tolkova*, the result is an evaluative question or declarative, as in (35), with an additional inference that the quantity meets or exceeds a contextual standard for a large quantity. In English, the structure in (48a) does not result in an evaluative interpretation. Therefore, the overt *mnogo* accompanying *kolko* and *tolkova* in Bulgarian is not the Q-adjective *mnogo* ‘many/much’ but the intensifier *mnogo* ‘very’. The meaning of this degree intensifier is as in (49). It introduces a *pos-* quantifier binding the degree argument of its sister adjective, contributing evaluativity and measures the degree interval in excess of the standard. The latter aspect of meaning underlies the syncretism between the intensifier and the Q-adjective *mnogo*.

- (49) [[*mnogo* intensifier]] = [[*very*]] =  
 $\lambda A_{\langle d, et \rangle} \lambda d \lambda x \exists d' [A(d')(x) \ \& \ d' > d_s \ \& \ \text{the size of } \{d'' : d' \geq d'' > d_s\} \geq d]$
- a. [[*mnogo* intensifier *visok*]] = [[*very tall*]] =  
 $\lambda d \lambda x \exists d' [x\text{'s height} \geq d' \ \& \ d' > d_s \ \& \ \text{the size of } \{d'' : d' \geq d'' > d_s\} \geq d]$
- b. [[*mnogo* intensifier [Q-adjP *mnogo* Q-adjective [Meas colors]]]] = [[*very many colors*]] =  
 $\lambda d \lambda x \exists d' [\text{the size of } \{d_1 : \exists x [\text{colors}(x) \ \& \ |x| \geq d_1]\} \geq d' \ \& \ d' > d_s \ \& \ \text{the size of } \{d'' : d' \geq d'' > d_s\} \geq d]$

(49a, b) illustrate the composition of the intensifier with a lexical adjective and with a pre-nominal Q-adjective; the latter is the structure of questions, declaratives, and exclamatives with an overt *mnogo*, as in (34) and (35), with the *wh-/th-* pronouns *kolko* and *tolkova* binding the degree variable that is the measure of the size of the interval in excess of the standard. If instead a *pos-*quantifier binds that variable, an evaluative interpretation arises that the size of that interval is large.

Only one *mnogo* is pronounced in (49b), likely because of the identical form of the intensifier and Q-adjective. This also happens in positive structures: the counterpart of *very many colors* is not \**mnogo mnogo cvetove* but *osobeno mnogo cvetove* ‘particularly many colors’. Yet there is also an alternative that cannot be ruled out: the intensifier *mnogo* could be combining not with the structure in (46) but with the one in (44), the structure without a Q-adjective. The se-

mantic types of the two are identical and both are semantically suitable for combination with the intensifier, with degree quantifiers, and with *wh-/th-* degree pronouns. What ruled out (44) in the case of English *how*, *that*, and degree quantifiers and in the case of Bulgarian degree quantifiers was the mismatch of category. But while there are good empirical reasons to claim that *very* only combines with adjectives, whether lexical or Q-adjectives, it is not clear that the same is true for the Bulgarian intensifier *mnogo*. Thus, it is possible that another structure is behind evaluative questions, declaratives, and exclamatives with an overt *mnogo*, in addition to (49b), namely (50).

- (50)  $[[mnogo \text{ intensifier [Meas colors]]] =$   
 $\lambda d \lambda x \exists d' [colors(x) \ \& \ |x| \geq d' \ \& \ d' > d_s \ \& \ \text{the size of } \{d'': d'' \geq d' > d_s\} \geq d]$

To summarize, English cardinality question, declarative, and exclamative nominals have the structure in (51). Their Bulgarian counterparts without overt *mnogo* have the structure in (52a), and those with overt *mnogo* have either the structure in (52b) or the one in (52c).

- (51)  $[[Q\text{-adjP } \textit{how/that (very) many} \text{ [Meas NP]]]$

- (52) a.  $[kolko / \textit{tolkova} \text{ [Meas NP]]]$   
 b.  $[kolko / \textit{tolkova} \text{ [mnogo intensifier [Q-adjP mnogo Q-adjective [Meas NP]]]]]$   
 c.  $[kolko / \textit{tolkova} \text{ [mnogo intensifier [Meas NP]]]$

Now that we have an explicit syntax and semantics for *wh-/th-* cardinality nominals, we can see that there is no reason to attribute evaluativity to a null *MNOGO*. The same structure in (51) yields nonevaluative questions and declaratives and evaluative exclamatives in English. The facts should be the same for Bulgarian (52a). Adding an overt *mnogo* results in evaluativity in both questions and declaratives and in exclamatives in either of the structures in (52b, c). One could of course posit that evaluativity in exclamatives comes from the obligatory merge of a null intensifier, but this will apply to English as much as it will to Bulgarian.

We can now turn to the question of how these structures relate to number marking on the NP. For measurement along a cardinality dimension the complement of *MEAS* needs to be semantically plural, i.e., denote a predicate of singular and plural individuals. That cardinality measurement depends on semantic plurality was suggested in Hackl (2001) and linked there to a generalization that measure functions are order preserving: if two individuals are ordered with respect to a dimension, their respective degrees on the relevant scale are similarly ordered. Concerning the dimension of number, a plural NP but not a singular count NP would allow for a nontrivial, order-preserv-

ing mapping of individual sums to degrees on the scale of natural numbers. On the assumption that morphologically singular NPs denote predicates of singular individuals, at least in English and Bulgarian where NPs are not number neutral, singular-marked NPs will be prohibited as complements of MEAS. Empirically this is correct. In combination with *how/that many* and *kolko/tolkova*, singular count NPs are prohibited. (51) and (52) require plural NPs.

## 4.2. Numerals

One approach to the semantics of numerals is to say that they denote numbers: the numeral *five* denotes the number 5, etc. How then is *five colors* formed and interpreted? The answer is to posit a nonovert expression which incorporates a cardinality measure function and then have the numeral saturate its degree argument. A null counterpart of the Q-adjective *many* is sometimes suggested for that role. But as we discussed above, Q-adjectives should not be treated as relations between degrees and individuals. Such a function is better served by MEAS. Accordingly, MEAS can also be implemented as the link between numerals and NPs, as in (53). A determiner, possibly null, further binds the individual variable (*the five colors, five colors*); as was assumed to be the case for *the many colors* and *many colors*.

- (53) a.  $[[\text{five}]] = 5$   
 b.  $[[\text{five colors}]] = [[\text{five Meas colors}]] = \lambda x [\text{colors}(x) \ \& \ \mu(x) \geq 5]$  where  $\mu = \text{number}$

This approach to numerals crucially relies on combination with plural NPs. Given that it is the same Meas that supplies the measure function, as in the cases discussed in the previous section, the expectation is that singular-marked count NPs will be precluded from structures with numerals. This is empirically so in both English and Bulgarian.

## 4.3. The Count Form of NPs

The count form in cardinality expressions with *wh-/th-*pronouns and numerals may not appear in (52a) or (53b). These structures require a plural NP. A different nonovert measure expression must be responsible for the count inflection on NPs.

There is cross-linguistic variation in whether numerals combine with plural or singular NPs. In Finnish, Turkish, and other languages numerals combine with singular-marked (bare) NPs and not with plural-marked NPs. Partly because of this, Ionin and Matushansky (2006) propose that numerals need to compose with NPs that denote predicates of atomic individuals, i.e.,



singular count NPs. Bale et al. (2011) have questioned this account, arguing that in some languages bare NPs are not semantically singular but number neutral, i.e., they denote predicates of singular and plural individuals, much like plural NPs in English. Yet, I suggest that the Bulgarian count form is precisely a case of a semantically singular NP combining with numerals (and *wh-/th-* quantity pronouns). In this, I depart from tradition, as grammars treat the count form as a special plural marker (Pašov 2011: 69, Hristozova 2012: 301), and this is also the proposal in Ouwayda (2014).

Specifically, I propose that the count form spells out singular number and objective case. It is the direct counterpart of the genitive singular form of Russian nouns in combination with paucal numerals ‘two’, ‘three’, and ‘four’, as in (54) (Pesetsky 2013: 1, adjectives omitted).<sup>17</sup> The Russian and the Bulgarian form are historically related (Duridanov 1993: 141; Stepanov and Stateva 2018).

- (54) èti                    dva                    stol-a  
 these<sub>NOM.PL</sub> two<sub>MASC.NOM</sub> table<sub>MASC.GEN.SG</sub>  
 ‘these two tables’

In support of the proposal that the count form has singular number and objective case, consider the following fact. Bulgarian nouns do not mark case overtly, except for singular masculine nouns. These are the only nouns which overtly distinguish between a nominative and a non-nominative, or objective, form. The distinction emerges in definite nominals, as seen in (55). And the singular objective form for masculine nouns is the same as the count form.

- (55) a. {Konjat                    / mâžât                    / prizrakât}                    padna  
 horse<sub>DEF.SG.NOM</sub> man<sub>DEF.SG.NOM</sub> ghost<sub>DEF.SG.NOM</sub> fall<sub>3SG.PAST</sub>  
 ‘{The horse / the man / the ghost} fell.’
- b. Vidjax                    {konja                    / mâža                    / prizraka}  
 see<sub>1SG.PAST</sub> horse<sub>DEF.SG.OBJ</sub> man<sub>DEF.SG.OBJ</sub> ghost<sub>DEF.SG.OBJ</sub>  
 ‘I saw the horse / the man / the ghost.’

Although semantically singular, the meaning of the count inflection differs from that of the singular number marker, which, I suggest, encodes a

<sup>17</sup> Alternative accounts suggest that the Russian NP marks paucal number (Bailyn and Nevins 2008, Pereltsvaig 2010, and others); or that it marks neither number nor case but is an expression of a functional category related to atomization and countability (Stepanov and Stateva 2018). I do not have the space to defend the singular-marking account here, though note its simplicity, as it posits that morphology transparently reflects interpretation. Pesetsky’s (2013) proposal that the Russian form is not marked for number is compatible with my suggestion that the form is semantically singular.

presupposition that the DP denotes a single entity (Sauerland et al. 2005). This presupposition precludes singular-marked NPs from combination with numerals in general, since unless the numeral is *one* the presupposition of the singular morpheme will not be satisfied. The count form, lacking such a presupposition, can combine with numerals, following the semantics of Ionin and Matushansky (2006).

I illustrate Ionin and Matushansky's (2006) semantics below, but modify it to accommodate the idea that numerals denote numbers (for them numerals are predicate modifiers). A null degree modifier needs to be posited, as in (56) (cf. Ionin and Matushansky 2006: 318–19). I will call it  $\text{Meas}_{SG}$  to reflect the fact that it combines with semantically singular NPs and makes available atomic units for counting.  $\text{Meas}_{SG}$  combines with a degree  $d$ , whose value is provided by the numeral and a predicate of atomic individuals  $P$  and returns a predicate of plural individuals that has  $d$ -many atoms. It is important that  $P$  be a predicate of singularities, otherwise a partition into plural individuals would satisfy (56) and *five colors* could mean a plurality of colors with five parts, each part being of unspecified cardinality.<sup>18</sup>

- (56)  $[[\text{Meas}_{SG}]] = \lambda P \lambda d \lambda x \exists S [\Pi(S)(x) \ \& \ |S| \geq d \ \& \ \forall s [s \in S \rightarrow P(s)]]$

A set of individuals  $S$  is a partition  $\Pi$  of a plural individual  $x$  iff the members of  $S$  exhaust all nonoverlapping parts of  $x$

Numerals universally denote numbers, but in English they combine with *Meas* and thus with a plural NP, whereas in Bulgarian they may combine both with *Meas* and thus with plural NPs, and with  $\text{Meas}_{SG}$  and thus with the semantically singular count NPs. Variation that exists across languages here exists in a single language. In normative registers, NPs combine with  $\text{Meas}_{SG}$  when their head noun is masculine nonpersonal and with *Meas* when the head noun is feminine or neuter. The situation with masculine personal nouns is more complex, because of variation with the *-ma* suffix, which is optional, in addition to the variability between plural and count form.

In the idealized grammar of the current normative language, the *-ma* suffix is obligatory for masculine personal nouns, and the nouns themselves are plural. The co-occurrence of the *-ma* suffix with the count form, encountered in the colloquial language, is a remnant of an older grammar, normatively acceptable in 1945 and 1983 according to Mikova (2017). Presumably in that older grammar *-ma* is a masculine personal agreement marker. In the newer normative grammar, *-ma* signals a classifier structure. A classifier is necessary to turn a semantically plural NP into a predicate of atomic individuals, which

<sup>18</sup> The cardinality-plurality link discussed earlier is still satisfied here.  $\text{MEAS}_{SG}$  in effect makes predicates of singularities plural before measuring their cardinality. It thus has a more complex semantics than *Meas*.

is then an appropriate argument to  $MEAS_{SG}$ . In this grammar all masculine nouns, personal and nonpersonal, need to combine with  $MEAS_{SG}$ , but a classifier corresponding to *-ma* is available only for masculine personal nouns. This is in agreement with Cinque and Krapova (2007), who propose that *-ma* is a suffixal classifier doubling the features of an overt or covert classifier, as can be seen in (57) (their (4b)) (see also Hurford 2003).<sup>19</sup>

- (57) trima                    (dúši)    aktjori  
 three<sub>MASC.PERS</sub>    person    actor<sub>MASC.PL</sub>  
 ‘three actors’

The expression *dúši* behaves like a numeral classifier: it appears only after numerals and *wh-/th-*quantity expressions.<sup>20</sup> Another example of a numeral classifier structure is in (58). *Broj* ‘count’ is a classifier for both nonpersonal and personal nouns, and like *dúši*, it can only appear with numerals and *wh-/*

<sup>19</sup> Ouwayda (2014) proposes that the *-ma* suffix on the numeral and the count *-a* suffix on the NP are merged in the same functional projection, #, and so are in complementary distribution.

<sup>20</sup> *Dúši* is formally plural but this probably reflects its origin in the feminine noun *dušá* ‘soul’. It also likely has formal masculine features, so it can spell out the masculine personal classifier which triggers the suffix *-ma* on the numeral. *Dúši* can be used in reference to female individuals and to groups including them (see (i) and (iii) from an internet search), but it does not combine directly with female-denoting NPs, whether they are formally feminine or neuter; see (ii). Neuter nouns denoting male individuals are reasonably acceptable, see (iii), though such uses are very rare in the contemporary language; most results were from the 19th century.

- (i) V salona e imalo 5 dúši –vsički ženi.  
 in salon<sub>DEF</sub> be<sub>3SG.PRES</sub> have<sub>SG.PART</sub> 5 person all women  
 ‘There were five people in the salon—all of them women.’
- (ii) \*tri(-ma) dúši {aktrisi / momičeta}  
 three<sub>(MASC.PERS)</sub> person actor<sub>FEM.PL</sub> girl<sub>NEUT.PL</sub>
- (iii) Obadete se do 23.01.2011 g. [...]s informacijata— [...]s  
 call REFL until 23-Jan-2011 year with information<sub>DEF</sub>  
 kolko dúši momčeta i momičeta šte učastvat.  
 wh-quantity person boy<sub>NEUT.PL</sub> and girl<sub>NEUT.PL</sub> will participate  
 ‘Call until 23 January 2011 with the following information: how many boys and girls will participate.’

*th-* quantity pronouns.<sup>21, 22</sup> Its complement NP must be plural. In the absence of *broj* the normative language dictates that masculine nonpersonal NPs appear in the count form.

- (58) tri broja {bileti / \*bileta } {zajci /  
 three count<sub>COUNT</sub> ticket<sub>MASC.PL</sub> ticket<sub>MASC.COUNT</sub> rabbit<sub>MASC.PL</sub>  
 \*zaeka }  
 rabbit<sub>MASC.COUNT</sub>  
 ‘three tickets’ / ‘three rabbits’

The attested and impossible structures for numerals other than *edin* ‘one’ and masculine nouns in the current normative language is given in (59). All masculine NPs combine with  $MEAS_{SG}$ .

- (59) Masculine NPs, normative language
- |                               |   |                                      |
|-------------------------------|---|--------------------------------------|
| a. *Numeral                   | $MEAS_{SG}$                               | $NP_{PL}$                            |
| b. *Numeral                   | $MEAS_{SG}$                               | $NP_{SG}$                            |
| c. $\sqrt{\text{Numeral-}ma}$ | $MEAS_{SG} CL_{MASC.PERS}$                | $NP_{PL}$ (personal nouns)           |
| d. $\sqrt{\text{Numeral}}$    | $MEAS_{SG} CL_{MASC.NON-PERS} -a_{COUNT}$ | $NP_{PL}$ (non-personal nouns)       |
| e. $\sqrt{\text{Numeral}}$    | $MEAS_{SG}$                               | $NP -a_{COUNT}$ (non-personal nouns) |

$MEAS_{SG}$  precludes direct combination with plural-marked NPs (59a) since it needs a predicate of atomic individuals. Singular-marked NPs are prohibited because the presupposition of the singular number marker only allows

<sup>21</sup> The classifier *broj* is possible with personal nouns, see (i), but likely because of competition with *dúši* it tends to primarily appear with nonpersonal nouns.

- (i) kandidatât e bil naučen râkovoditel na 3 broj diplomanti, ...  
 candidate<sub>DEF</sub> is been scientific advisor to 3 count<sub>COUNT</sub> graduate<sub>MASC.PL</sub>  
 ‘the candidate has been the advisor to three graduates ...’

<sup>22</sup> *Broj* has a use as a regular noun with the meaning of ‘number’, in which case it can appear without numeral expressions.

- (i) dâržavnata im izdržžka zavisi ot broj studenti, koito  
 state<sub>DEF</sub> their benefit depends on number<sub>DEF</sub> student<sub>PL</sub> who<sub>PL</sub>  
 obuĉavat  
 educate<sub>3PL</sub>  
 ‘their state benefit depends on the number of students who they educate’

combination with numeral *one*, (59b). A classifier turns plural-marked NPs into predicates of atomic individuals, and so (59c) is acceptable. The personal noun is plural-marked, the numeral is suffixed with *-ma*, and the classifier may or may not be pronounced, as in (57), though it is always present. A classifier structure is available for nonpersonal nouns as well, as in (59d), but the classifier *broj* (as in (58)) is always pronounced. The classifier turns the plural-marked NP into a predicate of atomic individuals, suitable for combination with  $MEAS_{SG}$ . Finally, the count form is semantically singular and so it meets the requirement of  $MEAS_{SG}$ , (59e).

In sum, masculine NPs in the normative language do not obtain cardinality measures through  $MEAS$ .  $MEAS$  measures plural individuals without direct access to the atoms of the plurality. A (precise) cardinality measure is assigned to the plurality without counting through a process we can call estimation (following O'Connor and Biswas 2017). A useful comparison is with container and measure pseudo-partitives (*a basket of cherries, three pounds of cherries*). Container/measure nouns partition the plurality into nonoverlapping parts (concretely or abstractly), map the parts to units (conventional units like *pound* or contextual units like *basket*), and then count or measure the units (*three pounds, a large basket*). In the case of *mnogo cvetove* 'many colors' or *five colors* in English, cases that involve  $MEAS$ , the pseudo-partitive unit is particularly abstract: a quantity. No reference is made to atoms (individual colors), what is measured is the size of the collection of colors.

On the other hand,  $MEAS_{SG}$  yields cardinality measures through counting, i.e., through reference to the atoms of the pluralities. The restriction on Bulgarian masculine NPs, in the normative language, is that cardinality measures be obtained by counting, and so with  $MEAS_{SG}$ . This must have been the case too in the older grammar (described in 1945 and 1983; see Mikova 2017), remnants of which are still observed today.  $MEAS_{SG}$  combined with personal nouns in the absence of a classifier, but the noun was in the count form, and *-ma* was just an agreement marker.

For current colloquial Bulgarian I propose that the system is as in (60).

(60) Masculine NPs, colloquial language

- |   |                                  |                               |
|---|----------------------------------|-------------------------------|
| a. *Numeral                               | $MEAS_{SG}$                      | $NP_{PL}$                     |
| b. *Numeral                               | { $MEAS_{SG}$ / $MEAS$ }         | $NP_{SG}$                     |
| c. $\sqrt{\text{Numeral}}$ ( <i>-ma</i> ) | $MEAS_{SG}$ $CL_{MASC.PERS}$     | $NP_{PL}$ (personal nouns)    |
| d. $\sqrt{\text{Numeral}}$                | $MEAS_{SG}$ $CL_{MASC.NON-PERS}$ | $NP_{PL}$ (nonpersonal nouns) |
| e. $\sqrt{\text{Numeral}}$                | $MEAS_{SG}$                      | $NP_{-aCOUNT}$                |
| f. $\sqrt{\text{Numeral}}$                | $MEAS$                           | $NP_{PL}$                     |

It retains (59a, b, c, d) but the classifier agreement marker *-ma* becomes optional. (59e) is generalized to all masculine nouns, not just the nonpersonal ones, as in (60e). Both changes involve weakening of the personal/nonpersonal distinction, which remains encoded only on the classifiers. Finally, combination with MEAS is permitted to masculine nouns, as in (60f). With *-ma* optional and the masculine personal classifier not necessarily overt, structures (60c, f) are pronounced the same, so evidence against positing MEAS with masculine personal nouns is no longer available. But once MEAS can be used with personal nouns, its use is extended to nonpersonal nouns as well, given that the personal/nonpersonal distinction is now obligatorily expressed only on the classifiers and the personal classifier is null. Once the importance of the personal vs. nonpersonal marking is undermined, the special status of masculine nouns is also undermined: they no longer have to obtain cardinality measures through counting and can combine with MEAS just like feminine and neuter nouns.

## 5. Back to Exclamatives

The extensive discussion of the grammar of cardinality expressions in the previous section was necessary because as far as I know there is no existing analysis of the pattern of distribution of the plural and count forms. Now that we have a theory of the structure and meaning of these forms, we can turn to the question that we started with: what explains the fact that the count form is possible in interrogative and declarative *wh-/th-* expressions (and other numeral contexts) but not in exclamative *wh-/th-* expressions.

I posit that cardinality exclamatives must be formed with MEAS rather than with MEAS<sub>SG</sub>. This amounts to saying that exclamatives obtain cardinality measures through estimation, not counting. While I cannot offer a definitive proof here, intuitively this seems right to me. Cardinality exclamatives express surprise that the referenced quantity is large and not surprise at the number value of the cardinality measure. Recall that Bulgarian is more flexible than English in allowing individual readings for exclamatives, and accordingly a wider range of *wh-*expressions, as in (39–40). One could imagine that individual-like readings (i.e., readings about number values) will be available for cardinality exclamatives in Bulgarian as well, i.e., that *How many colors!* in (29) could be used to express surprise at the particular number value of the cardinality of colors. If we expect an artist to draw a sketch using only one color and then we see that she used three colors instead, it would be felicitous to express surprise at the number three in this context, but the Bulgarian exclamative in (29) cannot be used to convey that meaning. Cardinality estimation with MEAS<sub>SG</sub> focuses on the number of individual atoms in the plurality, but what is needed in an exclamative is focus on the size of the plurality irrespective of the atoms, which is what MEAS delivers.

Given that exclamatives are formed with  $MEAS$ , exclamative *kolko* and *tolkova* combine only with plural NPs, just like overt *mnogo* does. Interrogative and declarative *kolko* and *tolkova* can combine with  $MEAS_{SG}$  and thus with the count form or with  $MEAS$  and thus with the plural form. This is where the difference in number marking between the two types of numeral *wh-/th*-expressions comes from. See (61–62).

- (61) *wh*-interrogative and *th*-declarative cardinal nominals (without overt *mnogo*)
- a. [ *kolko* / *tolkova* [ $MEAS$  NP]]
  - b. [ *kolko* / *tolkova* [ $MEAS_{SG}$  NP]]
- (62) *wh*- and *th*- exclamative cardinal nominals (without overt *mnogo*)
- [ *kolko* / *tolkova* [ $MEAS$  NP]]

There is no null *MNOGO* in exclamatives (a possibility discussed in section 3.1). What is responsible for the plural marking on NPs in exclamatives is the same measure-function containing nonovert  $MEAS$  found with questions and declaratives in the case of feminine and neuter nouns, as well as with masculine nouns in the colloquial language (as in (60f)). Masculine nouns also have available another null measure-function containing expression,  $MEAS_{SG}$ . Exclamatives are notable in relying on the estimation structure for cardinality measurement and resisting the counting structure.

## 6. Conclusions

Bulgarian exclamatives formed with the numeral *wh-/th*-pronouns *kolko* and *tolkova* differ from their interrogative and declarative counterparts in only accepting the plural form of masculine nonpersonal NPs. According to the normative language, interrogative *kolko* has to combine with the count form of masculine nonpersonal NPs and so does declarative *tolkova*. Given that the pronouns are identical in form, the differential number marking calls for an explanation. And the contrast is particularly surprising, given that there is significant variation in the colloquial language when it comes to number marking after interrogative *kolko* and declarative *tolkova*: both personal and nonpersonal nouns can appear in either the count or the plural form, yet exclamative *kolko* and *tolkova* resist such variation.

Given that there seems to exist no comprehensive analysis of number marking in Bulgarian, the paper set out to provide one. It was argued that *wh-/th*-pronouns, like numerals, connect with NPs through the help of two non-overt degree expressions,  $MEAS$  and  $MEAS_{SG}$ . The former imposes a semantic plurality on its NP complement (in cardinality contexts), the latter a semantic

singularity. Underlying this distinction are two modes of cardinality measurement: estimation and counting. Both singular-marked and count-marked NPs in Bulgarian were argued to denote predicates of atomic individuals, i.e., to be semantically singular, but the former was also said to be associated with a further presupposition that the DP denotes a single entity. The proposal that count NPs are semantically singular departs from usual assumptions about this form in grammars and linguistic analyses. With these parameters in place, the pattern of number marking in cardinality expressions follows, both in the normative language and its colloquial variety. The change between the two registers involves a weakening of the personal/nonpersonal distinction in masculine nouns and a corresponding erosion of the special status of masculine nouns in the language.

The resistance of exclamatives to count-marked NPs follows by the proposal that exclamatives are formed on the basis of MEAS, i.e., they concern a cardinality measure based on estimation, not counting. This is related to the evaluativity aspect of the meaning of exclamatives, which persists even in a language like Bulgarian that does not restrict its exclamatives to degree readings only but also allows individual readings.

University of Southern California  
pancheva@usc.edu

## References

- Bailyn, John and Andrew Nevins. (2008) "Russian genitive plurals are impostors". Andrew Nevins and Asaf Bachrach, eds. *Inflectional identity*. Oxford: Oxford University Press, 237–70.
- Bale, Alan, Michaël Gagnon, and Hrayr Khanjian. (2011) "Cross-linguistic representations of numerals and number marking". *Proceedings of SALT 20*, 582–98.
- Bresnan, Joan. (1973) "Syntax of comparative clause construction in English". *Linguistic inquiry*, 4: 275–344.
- Cinque, Guglielmo and Iliyana Krapova. (2007) "A note on Bulgarian numeral classifiers". Gabriela Alboiu, Andrei A. Avram, Larisa Avram, and Daniela Isac, eds. *Pitar Moş: A building with a view: Papers in honour of Alexandra Cornilescu*. Bucharest: Editura Universităţii din Bucureşti, 45–51.
- Cresswell, Max. (1976) "The semantics of degree". Barbara Partee, ed. *Montague grammar*. New York: Academic Press, 261–92.
- Duridanov, Ivan. (1993) *Grammar of the Old Bulgarian language*. Sofia: Bulgarian Academy of Sciences Press. [Original: Дуриданов, Иван. Граматика на старобългарския език. Sofia: Издателство на българската академия на науките.]



- Franks, Steven. (this volume) "A Bulgarian solution to the Slavic Q question?", 93–120.
- Georgiev, Vladimir, ed. (1979) *Bulgarian etymological dictionary*. Vol. II. Sofia: Bulgarian Academy of Sciences Press. [Original: Георгиев, Владимир И. 1979. Български етимологически речник. Том II. Издателство на Българската академия на науките.]
- Hackl, Martin. (2001) "Comparative quantifiers and plural predication". Karine Megerdooimian and Leora Anne Bar-el, eds. *Proceedings of WCCFL XX*. Somerville, MA: Cascadilla Press, 234–47.
- . (2009) "On the grammar and processing of proportional quantifiers: *most* versus *more than half*". *Natural language semantics* 17: 63–98.
- Hristozova, Galja. (2012) "The count form of masculine nouns—norms, rules, violations". *Bulgarian language and literature*, 54(4): 300–09. [Original: Христозова, Галя. 2012. Бройната форма на съществителните имена от мъжки род—норми, правила, нарушения. Български език и литература, година LIV, книжка 4, 300–09.]
- Hurford, James. (2003) "The interaction between numerals and nouns". Frans Plank, ed. *Noun phrase structure in the languages of Europe*. Berlin: Mouton de Gruyter, 561–620.
- Ionin, Tania and Ora Matushansky. (2006) "The composition of complex cardinals". *Journal of semantics* 23(4): 315–60.
- Michaelis, Laura. (2001) "Exclamative constructions". *Language typology and language universals* 2: 1038–50.
- Michaelis, Laura and Knud Lambrecht. (1996) "The exclamative sentence type in English". Adele Goldberg, ed. *Conceptual structure, discourse and language*. Stanford, CA: CSLI, 375–89.
- Mikova, Lora. (2017) "History of codification of the count form". *Bulgarian language* 64: 114–24. [Original: Микова, Лора. 2017. История на кодификацията на бройната форма. Български език 64: 114–24.]
- Nouwen, Rick and Anna Chernilovskaya (2015) "Two types of exclamatives". *Linguistic variation* 15(2): 201–24.
- O'Connor, Ellen and Priyanka Biswas. (2017) "Dual modes of measurement in language". Unpublished ms. University of Southern California.
- Ouwayda, Sarah. (2014) *Where number lies: Plural marking, numerals, and the collective-distributive distinction*. PhD dissertation, University of Southern California.
- Pašov, Petâr. (2011) *Bulgarian grammar*. Plovdiv: Hermes Press. [Original: Пашов, Петър. 2011. Пловдив: Българска Граматика. Издателска къща „Хермес“.]
- Pereltsvaig, Asya. (2010) "As easy as two, three, four?" E. Wayles Browne, Adam Cooper, Alison Fisher, Esra Kesici, Nikola Predolac, and Draga Zec, eds. *Formal approaches to Slavic linguistics: The Cornell meeting, 2009*. Ann Arbor, MI: Michigan Slavic Publications, 417–34.

- Pesetsky, David. (2013) *Russian case morphology and the syntactic categories*. Cambridge, MA: MIT Press.
- Rett, Jessica. (2006) "How many maximizes in the Balkan sprachbund". Masayuki Gibson and Jonathan Howell, eds. *Proceedings of SALT 16*. Ithaca, NY: CLC Publications, 190–207.
- . (2008) *Degree modification in natural language*. PhD dissertation, Rutgers University.
- . (2011) "Exclamatives, degrees, and speech acts". *Linguistics and philosophy* 34: 411–42.
- . (2014) "The polysemy of measurement". *Lingua* 143: 242–66.
- . (2015) *The semantics of evaluativity*. Oxford: Oxford University Press.
- . (2018) "The semantics of many, much, few, and little". *Language and linguistics compass* 12: 1–18.
- Rudin, Catherine. (1986/2013) *Aspects of Bulgarian syntax: Complementizers and wh-constructions*. Columbus, OH: Slavica Publishers.
- . (1988) "On multiple questions and multiple wh-fronting". *Natural language and linguistic theory* 6: 445–501.
- Sauerland, Uli, Jan Anderssen, and Kazuko Yatsushiro. (2005) "The plural is semantically unmarked". Stephan Kesper and Marga Reis, eds. *Linguistic evidence: Empirical, theoretical, and computational perspectives*. Mouton de Gruyter, 413–34.
- Schwarzschild, Roger. (2006) "The role of dimensions in the syntax of noun phrases". *Syntax* 9: 67–110.
- Solt, Stephanie. (2015) "Q-adjectives and the semantics of quantity". *Journal of semantics*, 32: 221–73.
- Stateva, Penka and Arthur Stepanov. (2016) "Agreement errors and structural distance: A corpus study of Bulgarian". *Zeitschrift für Slawistik* 61(3): 448–62.
- Stepanov, Arthur and Penka Stateva. (2018) "Countability, agreement, and the loss of the dual in Russian". *Journal of linguistics* 54(4): 779–821.
- Stoyanov, Stoyan, ed. (1993) *Grammar of the Contemporary Bulgarian language*. Vol. 2: Morphology. Sofia: Bulgarian Academy of Sciences Press. [Original: Стоянов, Стоян (гл. ред.) (1993) Граматика на съвременния български книжовен език. Том 2: Морфология. София: Издателство на българската академия на науките.]
- Tasseva-Kurktchieva, Mila. (2006) "The categorial status of quantifiers in Bulgarian: Evidence for DP over QP". James E. Lavine, Steven L. Franks, Mila Tasseva-Kurktchieva, and Hana Filip, eds. *Formal approaches to Slavic linguistics: The Princeton meeting, 2005*. Ann Arbor, MI: Michigan Slavic Publications, 378–93.
- Villalba, Xavier. (2008) "Exclamatives: A thematic guide with many questions and few answers". *Catalan journal of linguistics* 7: 9–40.

Zanuttini, Raffaella and Paul Portner. (2003) "Exclamative clauses: At the syntax-semantics interface". *Language* 79(1): 39–81.

# A Glimpse into the Acquisition of Bulgarian Morphosyntax: Pronominal Clitics

Teodora Radeva-Bork

*Abstract:* This paper discusses the main properties of accusative clitics and clitic doubling of accusative objects in Bulgarian from the perspective of first-language acquisition. The results from the experimental studies with 79 monolingual Bulgarian children aged 2;2–4;3 are illuminating with respect to the early emergence of clitic grammar, the placement rules of clitics, their agreement properties, and the established asymmetry between single and double cliticization.

## 1. Introduction

The aim of this paper is to give another perspective on Bulgarian morphosyntax by way of analysing the very early stages of its development in young monolingual children. What principles govern Bulgarian morphosyntax at the beginning of language acquisition? Are there any fundamental differences between child and adult grammar? What insights can the early stages of acquisition give us about the nature and driving mechanisms of the final stage of morphosyntax? These are some of the general questions the paper addresses by looking at a specific target, namely, direct-object clitics and clitic doubling in the acquisition of Bulgarian morphosyntax.

Although clitics have been an attractive area of research for several decades, little is known about their properties and syntactic principles in the early stages of grammar. Research on the acquisition of Slavic clitics especially is still scarce, with only studies on Croatian and Polish (cf. Stiasny 2006 and Tryzna 2010). The research presented in this paper aims to fill the existing gap by presenting original L1 data from Bulgarian, and subsequently derive some theoretically relevant generalizations based on data analysis. The empirical study of Bulgarian pronominal clitics is illuminating in relation to (i) the debatable nature of clitics as arguments vs. agreement, (ii) the major principles of clitic placement, and (iii) the relation between the phenomena of single and double cliticization.

I begin with a general discussion of the properties of single and double clitics in Bulgarian, followed by a presentation of the construct and results of

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin*. Bloomington, IN: Slavica Publishers, 2018, 235–50.

the acquisition studies. Finally, I outline the most significant theoretical implications made on the basis of the empirical data.

## 2. Properties of Single Clitics and Clitic Doubling

For reasons of space and since the main focus of the paper is the acquisition of clitics, I cannot present a full theoretical review of previous research on clitics and clitic doubling that justifies the influential, seminal work done in the field. For further information see, among many others, Tomić 1996, Avgustinova 1997, Dimitrova-Vulchanova 1999, Bošković 2002, Krapova and Cinque 2008, and Guentchéva 2008. For present purposes, I only briefly sketch the main properties of single clitics and clitic doubling (CD) in Bulgarian, focusing exclusively on direct-object clitics and direct-object clitic doubling, since these are the environments tested in the study.

Bulgarian direct-object clitics appear in an eight-member paradigm. They are marked for case, number, and person as well as gender for third person singular. Since the target grammar disallows empty objects, clitics in object positions must be obligatorily pronounced. The clitics are verb-adjacent, which distinguishes them from other Slavic object clitics (e.g., Serbo-Croatian), which are second-position (or Wackernagel) clitics. The default position of Bulgarian clitics is proclitic, i.e., to the left of their phonological host, which is invariably the verb, as shown in (1a, b). The relevant clitics are italicized in the examples.

- (1) a. *Majkata*            *gi*            *celuva* često.  
           mother<sub>DEF.NOM</sub>    them<sub>CL.ACC</sub>    kisses    often  
           ‘The mother kisses them often.’
- b. *Majkata* često *gi* *celuva*.
- c. \**Celuva* često *gi* *majkata*.
- d. \**Majkata* *gi* često *celuva*.

If no element precedes the clitic (or the clitic cluster) in a sentence initial position, the clitic must follow the verb. This is imposed by a phonological constraint known as the Tobler-Mussafia effect, which requires that the clitic occur post-verbally as an enclitic (2a–b), or lean on a host in a sentence-initial position (2c).

- (2) a. \**Gi*            *celuva* često.  
           them<sub>CL.ACC</sub>    kisses    often  
           ‘She/he kisses them often.’

- (2) b. *Celuva gi često.*  
 c. *Često gi celuva.*

In line with Franks and Rudin (2005) and Franks (2009), I consider Bulgarian clitics to be  $K(\text{ase})^0$  heads that move to Agr. This is grounded on the idea that Bulgarian has retained the structure of nominals found in older Slavic, where nominal expressions were maximally Kase Phrase (KP) with clitics instantiating  $K^0$ . So while clitics in Romance are  $D^0$ , in Slavic they are  $K^0$ .

Based on Chomsky's (1995) model of bare phrase structure, in which an ambiguous element can be merged as a maximal projection but subsequently moves as head, Franks (2009) suggests that object clitics are in AgrO but that they have moved there from argument positions. This is in line with Rudin's (1997) view of clitics as agreement markers, with the difference that they merge as theta-marked arguments in theta-positions and subsequently move as  $K^0$  heads to Agr. This is reflected in the syntactic structure for Bulgarian clitics given in (3):

- (3)  $[_{\text{AgrOP}} [_{\text{AgrO}} [_{\text{K}} \textit{clitic} + \textit{AgrO}]] [_{\text{VP}} \textit{verb } t_{\text{KP/K}}]]$  (Franks 2009: 197)

Another significant property of Bulgarian pronominal clitics is that they can participate in clitic doubling, i.e., clitics can overtly double a verbal argument, which acts as their associate, inside the same clausal domain. The clitic bears the same *phi*-features and case as the associate. The associate can be a full pronoun, a DP, a CP, or a *wh*-word. Both direct and indirect objects can be doubled. Since the only relevant distribution for the study is the one where direct object DPs are clitic-doubled, only this type is illustrated below. The clitic and the associate are italicized.

- (4) a. *Mača* go spečeli po-dobrijat otbor.  
 match<sub>DEF</sub> it<sub>CL.ACC</sub> won better<sub>DEF</sub> team  
 b. Po-dobrijat otbor go spečeli *mača*.  
 better<sub>DEF</sub> team it<sub>CL.ACC</sub> won match<sub>DEF</sub>  
 c. Spečeli go po-dobrijat otbor *mača*.  
 won it<sub>CL.ACC</sub> better<sub>DEF</sub> team match<sub>DEF</sub>  
 d. Spečeli go *mača* po-dobrijat otbor.  
 won it<sub>CL.ACC</sub> match<sub>DEF</sub> better<sub>DEF</sub> team  
 'The better team won the match.'

The doubling requirement is in no way contingent on the position of the object. All possible constellations are attested with doubling in Bulgarian. Constituent order in the language is flexible, theoretically allowing for all

possible word order combinations, i.e., SVO, OVS, VSO, etc., and this is in turn reflected in the positioning of doubled objects in CD constructions, cf. (4).

CD in Bulgarian is not obligatory across the board (in difference to Macedonian), and it is triggered by a complex set of conditions mainly associated with discourse (regarding the specificity and definiteness of doubled objects), syntactic structure, predicate types, and information structure.<sup>1</sup> This leads to the suggestion that different types of CD in Bulgarian can be distinguished (cf. Radeva-Bork 2012). One of the most prominent types, and the one that was the focus of the present experiment, is object identification. Object identification is typically found in transitive constructions, in which both nominal constituents are animate and of the same number, cf. (5) and (6).<sup>2</sup> The unmarked word order and the interpreted order in the absence of a doubling clitic is SVO. If the initial constituent is to be interpreted as a direct object, CD is obligatory.

- (5) *Učitelkata*            \*(*ja*)            narisuva    deteto.  
 teacher<sub>DEF.SG.FEM</sub>   her<sub>CL.ACC</sub>    drew        child<sub>DEF.SG</sub>  
 'The child drew the teacher.'

- (6) *Marija*    nikoj        ne \*(*ja*)        celuna.  
 Maria<sub>SG</sub>   nobody<sub>SG</sub>   neg   her<sub>CL.ACC</sub>    kissed  
 'Nobody kissed Maria.'

Here it is the particular word order, i.e., subject not first, that triggers the obligatory use of CD as a means of disambiguation between *učitelkata* in (5) and *Marija* in (6) being subjects as non-clitic-doubled arguments or objects as clitic-doubled arguments. Additionally, it is important to note that a strict discourse licensing is at play here, since only specific and definite objects can be doubled in such environments.

<sup>1</sup> For a detailed account of the interplay of the conditions determining the occurrence of CD in Bulgarian see Krapova and Cinque (2008). The authors discuss cases of obligatory CD with specific predicate types such as psych and physical perception predicates with dative or accusative experiencers, feel-like constructions, and other contexts.

<sup>2</sup> Similar sentences are sometimes analysed as clitic left dislocation structures (see, for example, Krapova and Cinque 2008). I consider these to be instances of clitic doubling proper, for argumentation see Radeva-Bork (2012).

### 3. The Child Data

This section presents first-language acquisition data from three experimental studies with 79 monolingual Bulgarian children aged 2;2–4;3,<sup>3</sup> tested in two kindergardens in Varna, Bulgaria. In addition, 38 adult controls were tested (cf. Radeva-Bork 2012). The experiments examined the production of direct object clitics (3rd person singular and plural forms) in both preverbal, i.e., proclitic, and post-verbal, i.e., enclitic, positions, as well as the comprehension of clitic doubling environments.

#### 3.1. Experiments on Single Clitics

The single-clitic data come from two elicited production experiments: Experiment 1 examined the production of direct object clitics in post-verbal positions (enclisis), and Experiment 2 investigated the acquisition of direct object clitics in pre-verbal position (proclisis). Experiment 1 included an elicited production task with 46 children (19 girls and 27 boys), aged 2;2–4;3. Each child was tested on five transitive verbs, *gušvam* ‘hug’, *ritam* ‘kick’, *butam* ‘push’, *celuвам* ‘kiss’, *sresвам* ‘comb’, in four conditions, depending on the elicited accusative clitic–masculine (*go*), feminine (*ja*), neuter (*go*), and plural (*gi*). All answers obligatorily included an enclitic, as shown in (7).

- (7) Experimenter: *What did Borko do to the ball?*  
 Child 2;2: *Ritna ja.*  
*kicked<sub>3SG</sub> it<sub>CL.ACC.3SG.FEM</sub>*  
 ‘He kicked it.’

In Experiment 2, 17 monolingual children (10 girls and 7 boys), aged 2;5–4;3, were tested on 16 items using the same direct object clitic forms as in Experiment 1 with four transitive verbs, *ritam* ‘kick’, *celuвам* ‘kiss’, *jam* ‘eat’, and *običam* ‘love’. Here it was necessary to use different contexts from the ones eliciting enclisis in Experiment 1, since the syntactic environments for proclitics in Bulgarian are different (cf. section 2), and it was important to make sure that only clitics in preverbal positions would be used by the children when answering the control questions. The eliciting contexts, which satisfied this condition and allowed the use of proclitics only, were subordinate *because*-clauses, coordination with *and*, *iskam da* ‘want to’ plus accusative object clitic, *šte* ‘will’ plus accusative object clitic, and *samo* ‘only’ plus accusative object clitic. Some of these contexts are shown in (8a, b):

<sup>3</sup> 2 years and 2 months; 4 years and 3 months.



- (8) a. Experimenter: *What do you think Rabbit will do with the carrot?*  
 Child 2;9: Šte go izjade.  
 will<sub>3SG</sub> it<sub>CL.ACC.3SG.MASC</sub> eat  
 'He will eat it.'
- (8) b. Experimenter: *What does Little Froggy want to do to Mummy Frog?*  
 Child 3;3: (Iska) da ja cunka.  
 (wants) to her<sub>CL.ACC</sub> kiss  
 '(She wants) to kiss her.'

### 3.2. Experiment on Clitic Doubling

This experiment investigated the acquisition of direct object CD in 16 monolingual children (5 boys and 11 girls), aged 2;5–4;2. The study utilized a comprehension experiment based on a picture-matching task. One specific type of obligatory CD in Bulgarian was tested: transitive constructions in which both nominal constituents are animate and of the same number, with the non-canonical OVS word order with a fronted object that has to be invariably doubled by an agreeing clitic in order to be identified as the internal argument of the sentence. In such cases the direct object clitic is the only grammatical means to identify the internal argument. In order to ensure that the discourse conditions in the experiment are satisfied, we made sure that all doubled objects are specific and definite by presenting them in a specific context beforehand.

Four transitive verbs were used to describe the actions in the pictures: *xvaštam* 'catch', *celuвам* 'kiss', *nastâpвам* 'step on someone's foot', and *sresвам* 'comb'. Each of the verbs was used with three accusative clitic forms: masculine (*go*), feminine (*ja*), and neuter (*go*), giving a total of 12 test items. Using a bigger number of items with young children in such an experimental setting is often problematic, as children easily get distracted after the first 10 items and do not focus on the task any more. Similar numbers of items have been used in previous experiments on CD, such as in Varela (1988) for Spanish or in Kapia (2010) for Albanian. Each item in the present experiment consisted of three pictures. An example of the elicitation procedure is given in (9).

- (9) Test sentence (spoken, presented after the pictures):

*Tatkoto* \*(*go*) celuna Marija.  
 father<sub>DEF</sub> him<sub>CL.ACC</sub> kissed Maria  
 'Maria kissed the father.'

Picture 1. Maria kissed the father. (*correct answer*)

Picture 2. Borko kissed the father. (*distracter*)

Picture 3. The father kissed Maria. (*reverse action*)

Although the word order of the test sentence is OVS, it can mistakenly be interpreted as SVO were it not for the clitic, which has the same *phi*-features as the doubled object and clearly fulfils the role of an object agreement marker (cf. Radeva-Bork, 2012). Thus the expectation is that if children go for the correct interpretation, in which “Maria” is the subject of the sentence, there is evidence that they comprehend CD in an adult-like manner. If, however, children interpret “the father” as the sentence subject, this would indicate that CD is not yet fully acquired.

### 3.3. Results

#### 3.3.1. Single Clitics

The results from the elicitation studies indicate that Bulgarian children show an adult-like mastery of clitic syntax in the initial stages: they produce clitics at the age of 2;2 and obey the clitic requirements of verb-adjacency by producing both proclitic and enclitic constructions from the clitic onset.

Children from 2;2 to 3;0 produced direct object clitics 58% of the time, increasing to 80% for age group 3;1–3;7, and reaching almost adult-like performance with 99% at 3;8–4;3. The adult control group produced clitics invariably at 100%. The high clitic production rates already at the onset of clitic acquisition give evidence for the lack of a clitic omission stage in Bulgarian, especially when compared to omission rates in the so-called clitic omission languages (cf. Radeva-Bork 2012, Babyonyshev and Marin 2006) such as French and Italian (cf. Jakubowicz and Rigaut 2000 and Schaeffer 2000). As they age, there is an increase in children’s clitic production in Bulgarian. Bulgarian patterns with Spanish, for which similar low rates of clitic omission have been established (cf. Castilla, Pérez-Leroux, and Eriks-Brophy 2008).

The results also show that children have an adult-like mastery of clitic syntax and produce both proclisis and enclisis from the clitic onset, i.e., from 2;2 onwards. Here absolute findings are derived—all of the 17 participants in the test showed an adult-like mastery of the syntax of verbal clitics and placed the elicited clitics correctly, i.e., in preverbal position, 100% of the time. There were literally no placement errors.

At this point, five principal observations can be made on the basis of the production experiments on single direct object clitics:

- (i) Direct object clitics in Bulgarian emerge as early as 2;3. Clitic production is robust and increases steadily across age.

- (ii) In a cross-linguistic perspective Bulgarian patterns with Spanish, Greek, Romanian, and Croatian,<sup>4</sup> where clitic production is unproblematic and early. This is expected under the Unique Checking Constraint (UCC) (cf. Wexler 1998 and his subsequent works) which predicts the lack of high clitic omission in the early stages of non-participle-clitic agreement languages, since no double checking of uninterpretable features is required.
- (iii) Children seldom choose the grammatically correct but pragmatically odd option of using full DPs instead of clitics. The substitution rates are negligible.
- (iv) Concerning the morphosyntactic shape of clitic forms, only 11 out of 46 children made some kind of clitic form errors with an average of 8%. No consistent preference for a specific clitic form was established. This confirms Babyonyshev and Marin's (2006) observations for Romanian. Furthermore, the findings are in line with the absence of ill-formed clitic production in, for example, Spanish and Catalan.
- (v) Regarding the placement properties of clitics in child grammar, these are target-like from early on since there were literally no errors and children showed an adult-like mastery of clitic syntax obeying the requirements of verb-adjacency and producing both proclitic and enclitic constructions from the clitic onset, i.e., from 2;2 onwards.

### 3.3.2. Clitic Doubling

Turning to the results of the CD experiment, CD seems to be a late phenomenon—the children's overall success rate is 54%, with 47% at 2;8 and 63% at 4;1. It seems that Bulgarian children do not reach adult-like performance of CD even by the age of 4 years and that CD acquisition at this stage still relates to optionality of use. From 3;0 onwards we find 79% clitics vs. 51% CD, and from 4;0 onwards we observe 99% clitics vs. 63% CD.<sup>5</sup>

<sup>4</sup> Cf. the studies in Wexler et al. (2004), Marinis (2000), Babyonyshev and Marin (2006), and Stiasny (2006).

<sup>5</sup> As an reviewer points out, some potential confounding variables in the CD experiment may be: The construction involves multiple clauses, which could be a source of overall processing difficulty; the word order used is non-canonical, which could also lead to processing difficulty that ends up in a wrong interpretation; the overall comparison of the results from the two experiments has to be taken with caution due to well-known asymmetries found in production vs. comprehension tasks. Nevertheless, since this is the first study of the comprehension of CD with Bulgarian children, and CD is a notoriously difficult phenomenon to test in children (and adults), the results of this study are relevant for outlining some initial observations and illuminative for further investigations of the CD phenomenon.

An examination of the overall results of the CD study allows for the following observations:

- (i) Bulgarian children show interpretive knowledge of clitic doubling constructions. They are sensitive to object marking by means of clitic doubling since in many cases they manage to identify the fronted clitic-doubled arguments as the syntactic objects of the sentences.
- (ii) Yet, it may be that CD constructions are variably ambiguous for children in the early stages since their performance on the task is successful slightly more than half of the time (at an average of 54%). Even at the age of 4 years, children's grasp of CD differs from that of adults (cf. 63% for the 4-year-olds and 100% for the control adults). However, since children do show knowledge of CD in the test, it may be the case that the variation in their performance is due to processing difficulties, i.e., children do not have access to that knowledge at all times.
- (iii) The finding that children have knowledge of doubling contexts in Bulgarian in the early stages but retrieve this knowledge optionally is in line with Schaeffer 1995 with regard to Dutch. She finds that Dutch children show optional scrambling of direct and indirect objects. The construction exists in the early stages, but children seem to have no knowledge of its obligatoriness.
- (iv) Age seems to be a factor, as there is a certain degree of improvement in the performance of the task (a rise from 47% at 2;8 to 63% at 4;1).<sup>6</sup>

## 4. Theoretical Implications

### 4.1. Emergence of Clitic Grammar

Generally, we find proof that the same threefold distinction that is found in the adult grammar (cf. Rudin 1997) exists already in the early stages: in the child data we found null objects with overt clitic agreement, overt object DPs with null clitic, and overt-object DPs with overt-object agreement clitic resulting in CD.

### 4.2. Clitic Placement

Bulgarian children show an adult-like mastery of clitic syntax, obeying the clitic requirements of verb-adjacency and producing both proclitic and enclitic constructions from the clitic onset. This finding is in line with research

<sup>6</sup> Mean ages.

on other languages, e.g., French (Grüter 2006), Greek (Marinis 2000), Romanian (Babyonyshev and Marin 2006), Spanish and Croatian (Stiasny 2006), where position errors occur only at very low rates, if at all. If acquisition proceeds in the same manner with both second position and verbal clitics, i.e., object clitics are never misplaced, we could conclude that the kind of placement requirement (verbal or second position) to be acquired plays no role in the acquisition of clitic placement across languages. In line with Stiasny (2006), it is then possible to assume that the same syntactic processes determine main aspects of clitic placement and that surface differences can be attributed to independent, parameterized requirements. In other words, clitic placement across languages may be governed by similar kinds of rules, irrespective of the clitic placement requirements of their grammar for second-position or verb-adjacent clitics.

### 4.3. Clitics as Agreement

The findings of the CD experiment may be taken as indication that pronominal clitics in child grammar are analysed as object agreement markers, since the doubling clitic was the only means of differentiating between internal and external arguments in the test situations. This finding has direct relevance to the debate of the nature of clitics as arguments or agreement and provides support for the analysis of clitics as object agreement as in Rudin (1997) and Franks (2009), building on Franks and Rudin (2005) (for object clitics being K heads that move to Agr).

In terms of a possible diachronic development and in view of the syntactic resemblances between Bulgarian and Macedonian, it seems that if Bulgarian object clitics exhibit at least *some* object agreement properties, they may have a transitional character and may undergo the same development as Macedonian clitics by strengthening their agreement properties (for further details see Radeva-Bork 2012). This also gives an interesting generalization about the difference between pronominal clitics in Macedonian and Bulgarian. In Franks's view Bulgarian clitics are  $K^0$  heads that move to Agr, while Macedonian clitics are essentially Agr<sup>0</sup> heads.

This is not to say that pronominal clitics in Bulgarian are object-agreement markers across the board (although I see this as a very possible diachronic development). Rather, I suggest that Bulgarian clitics in CD constructions can fulfill different functions depending on word order facts and information structure (see again the detailed discussion of various CD environments in Krapova and Cinque 2008). An analysis in which clitics are treated differently is justified because as Franks notes "it is not true that every instance of a clitic,

even within a single linguistic system, should be treated the same" (Franks 2009: 199).<sup>7</sup>

Franks (2009: 189) also argues that diachronically, Macedonian pronominal clitics are more advanced than their Bulgarian counterparts. Macedonian pronominal clitics are closer to being complete agreement markers than the Bulgarian pronominal clitics. Since both languages are syntactically very similar, it is possible to assume that Bulgarian clitics will develop diachronically in line with their Macedonian counterparts. That is, in a situation of transition Bulgarian clitics are still at the beginning and their Macedonian counterparts have progressed further. Under this view, it is indeed possible to posit that Bulgarian clitics show at least some properties of object-agreement markers and over time may develop following the same path as Macedonian clitics in becoming stronger object-agreement markers.<sup>8</sup>

It is appropriate here to remark on the success scores in the elicitation study, ranging between 47% and 63% across the age groups. Whereas there is an indication of children's interpretative knowledge of object clitics as object-agreement markers, children seem to be unable to retrieve their knowledge of CD in an adult-like manner. In the context of the few previous studies of the acquisition of CD, the overall results of an average 54% for Bulgarian are not surprising. For Greek it has been shown that CD appears after single clitics and that there is individual variation in the data (Marinis 2000). Albanian children do not show adult-like mastery of CD until the age of 4;0 (Kapia 2010).

Of course, the question is what exactly is it that prevents children from implementing their knowledge of CD contexts (since the study results show that children indeed have such knowledge) in an adult-like manner. In my previous work (cf. Radeva-Bork 2012, 2014), I suggest that the children's inability to retrieve this kind of knowledge is not due to a grammatical deficit but rather to certain linguistic externally relevant properties related to clitic doubling as an interface phenomenon between (at least) syntax and discourse. Section 4.4 below elaborates further on the possible source of the acquisitional asymmetry between single clitics and clitic doubling.<sup>9</sup>

<sup>7</sup> In a somewhat similar spirit, Franks (2006 and other works) suggests that the various Bg clitics occupy different functional heads. He assumes a traditional structure in which AgrSP dominates TP and puts forward the following: auxiliaries with the exception of 3sg *e* are positioned highest in AgrS, *e* is in T<sup>0</sup>, the pronominal clitics are adjoined to T<sup>0</sup>, and the verbal participle is placed immediately to the right of T.

<sup>8</sup> For the diachrony of clitics see Pancheva (2005) and Dimitrova-Vulchanova and Vulchanov (2008). For comparisons between Bulgarian and Macedonian see Dimitrova-Vulchanova and Tomić (2009).

<sup>9</sup> Of course, I am also aware of the general production-comprehension asymmetry typical for child language acquisition.

#### 4.4. An Early Asymmetry Between Single and Double Clitics

The timing of the phenomena of single and double cliticization in child grammar provides valuable insights into the internal syntactic structure of clitic constructions. In view of the Uniformity Hypothesis (Sportiche 1996), which postulates that the constructions of single clitics, CD, and CLLD involve the same underlying structure, we expect the phenomena of single clitics and CD to be simultaneous processes in child language. Concerning the observations that, on the one hand, single clitics in Bulgarian emerge at around 2;2 with productive use at 2;6 and, on the other hand, CD is adult-like at 63% even by age 4;2, it is evident that single and double cliticization are not simultaneous processes in child grammar (similar findings for Greek in Marinis 2000). This indicates that single clitics and CD do not share the same underlying structures, which is in contradiction to the Uniformity Hypothesis but in conformity with analyses treating single clitics, CD, and dislocation constructions as having different properties (e.g., Anagnostopoulou 1994). Furthermore, since CD appears after the acquisition of single clitics, we can take that as evidence that CD has a more complex structure than single clitics. The same has been shown by Marinis (2000) based on Modern Greek acquisition data.

A logical question concerns the nature of the established asymmetry in the acquisition of the two related phenomena. Why is it that pronominal clitics in doubling environments appear later than single clitics in an error-free way?

I suggest that the asymmetry in the acquisition of single clitics and CD is not grounded in a grammatical deficit, but may be attributed to differences in the processing of information at and from the interfaces of syntax and discourse, which are related to the processing of CD constructions. In other words, the instable retrieval of knowledge may be due to interface-coordination difficulties in the case of CD. Children do have knowledge of CD but apparently they cannot retrieve it all of the time, which results in optionality in the output.<sup>10</sup> It has been shown cross-linguistically that grammatical phenomena requiring access to knowledge from different interfaces such as clitic doubling in Albanian (Kapia 2010), scope assignment in Mandarin Chinese (Zhou 2011), and scrambling in Dutch (Schaeffer 2000) and Ukrainian (Mykhaylyk and Ko 2010) are optional (or delayed) in child grammar.

Children's non-adult linguistic performance is traceable to the load of computations at multiple interfaces and specifically at the interfaces of syntax and discourse/pragmatics,<sup>11</sup> which have been shown to be problematic in ac-

<sup>10</sup> An additional factor, as pointed out by a reviewer, is the general optionality of CD in many Bulgarian constructions in the input.

<sup>11</sup> In Radeva-Bork 2014, I analyse multiple interface phenomena such as double cliticization as more costly, since they are associated with a higher computational load at

quisition. Montrul (2009) suggests that linguistic properties at interfaces are inherently more complex than linguistic properties internal to a specific domain (syntax, phonology etc.). The present findings show that syntactic properties are fully acquired whereas interface properties trigger residual optionality effects in the early stages. Properties of the syntax-discourse interface are more costly to compute than narrow syntactic properties. This may explain the difference in the acquisition of single clitics and CD.

In a broader perspective, the study results are indicative of a modular architecture of the faculty of language with syntax present from the age of 2;2, and the syntax-discourse interface not present in a target-like manner even by the age of 4;2. The results of the study give a strong hint that interfaces are not monolithic, i.e., equally represented, in child grammar.

## 5. Conclusion

This paper provides another glimpse into Bulgarian morphosyntax by analysing structures with single and double clitics in the very early stages of grammar development. The experimental results are mostly illuminating with respect to the early emergence of clitic grammar, the placement rules of clitics, their agreement properties, and the established asymmetry between single and double cliticization.

On the basis of the data it can be argued that: (i) child L1 learners fundamentally do not differ from adult native speakers in how they represent and process clitic grammar, as we find the same threefold distinction in child grammar as in the adult language (cf. Rudin 1997); (ii) the phrase marker in child grammar is developed, and it seems that pronominal clitics exhibit at least some object-agreement properties which provides support for their analysis as object agreement as in Rudin (1997) and Franks (2009), building on Franks and Rudin (2005) for object clitics being  $K^0$  heads that move to Agr; (iii) the kind of clitic placement requirement (verbal or second position) to be acquired cross-linguistically plays no role, and the same syntactic processes determine major aspects of clitic placement; (iv) single clitics and double cliticization present two fundamentally different phenomena associated with distinct underlying syntactic structures (contrary to Sportiche's 1996 Uniformity Hypothesis); and (v) double cliticization, being an instance of a multiple inter-

---

the interfaces, i.e., computation of a higher number of interpretations than for single interface phenomena. The parser's constant error-free accessibility to the target interpretation choosing among competing interpretations requires additional procedural resources. This may be the source of the optionality in the child output. A structure like CD induces a number of (at least two) competing interpretations and the access to the target one is not invariably guaranteed. This results in optional knowledge retrieval in children due to their limited processing resources in the early ages.



face phenomenon, requires higher computation costs and is associated with optionality in the initial stages.

University of Potsdam

teodora.radeva-bork@uni-potsdam.de

## References

- Anagnostopoulou, Elena. (1994) *Clitic dependences in Modern Greek*. Ph.D. dissertation, University of Salzburg.
- Avgustinova, Tania. (1997) *Word order and clitics in Bulgarian*. Ph.D. dissertation, Saarland University. [Dissertation in Computational Linguistics and Language Technology.]
- Babyonyshev, Maria and Stefania Marin. (2006) "Acquisition of pronominal clitics in Romanian". *Catalan journal of linguistics* 5: 17–44.
- Bošković, Željko. (2002) "Clitics as nonbranching elements and the linear correspondence axiom". *Linguistic inquiry* 33(2): 329–40.
- Castilla, Anny P., Ana Teresa Pérez-Leroux, and Alice Eriks-Brophy. (2008) "Syntax and the lexicon in early omission of Spanish clitics". Harvey Chan, Heather Jacob, and Enkeleida Kapia, eds. *Proceedings of the 32nd annual Boston University conference on language development*. Somerville, MA: Cascadilla Press, 72–83.
- Chomsky, Noam. (1995) *The minimalist program*. Cambridge, MA: MIT Press.
- Dimitrova-Vulchanova, Mila. (1999) "Clitics in the Slavic languages". Henk van Riemsdijk, ed. *Clitics in the languages of Europe*. Berlin: Mouton de Gruyter, 83–122.
- Dimitrova-Vulchanova, Mila and Olga Tomić. (2009) "The structure of the Bulgarian and Macedonian nominal expression: Introduction". Mila Dimitrova-Vulchanova and Olga Tomić, eds. *Investigations in the Bulgarian and Macedonian nominal expression*. Trondheim: Tapir Academic Press, 1–23.
- Dimitrova-Vulchanova, Mila and Valentin Vulchanov. (2008) "Clitic doubling and Old Bulgarian". Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 105–32.
- Franks, Steven L. (2006) "Another look at *li* placement in Bulgarian". *The linguistic review* 23(2): 161–211.
- . (2009) "Macedonian pronominal clitics as object agreement markers". Steven L. Franks, Vrinda Chidambaram, and Brian D. Joseph, eds. *A linguist's linguist: Studies in South Slavic linguistics in honor of E. Wayles Browne*. Bloomington, IN: Slavica Publishers, 189–221.
- Franks, Steven L. and Catherine Rudin. (2005) "Bulgarian clitics as  $K^0$  heads". Steven L. Franks, Frank Y. Gladney, and Mila Tasseva-Kurkchieva, eds.

- Formal approaches to Slavic linguistics: The South Carolina meeting, 2004.* Ann Arbor, MI: Michigan Slavic Publications, 106–18.
- Grüter, Therese. (2006) *Object clitics and null objects in the acquisition of French*. Ph.D. dissertation, McGill University.
- Guentchéva, Zlatka. (2008) “Object clitic doubling constructions and topicality in Bulgarian”. Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 203–23.
- Jakubowicz, Celia and Catherine Rigaut. (2000) “L’acquisition des clitiques nominatifs et des clitiques objets en français”. *Canadian journal of linguistics/Revue canadienne de linguistique* 45: 119–57.
- Kapia, Enkeleida. (2010) *The role of syntax and pragmatics in the structure and acquisition of clitic doubling in Albanian*. Ph.D. dissertation, Boston University.
- Krapova, Iliyana and Guglielmo Cinque. (2008) “Clitic reduplication constructions in Bulgarian”. Dalina Kallulli and Liliane Tasmowski, eds. *Clitic doubling in the Balkan languages*. Amsterdam: John Benjamins, 257–87.
- Marinis, Theodora. (2000) “The acquisition of clitic objects in modern Greek: Single clitics, clitic doubling, clitic left dislocation”. Nanna Fuhrhop, ed. *Papers on language change and language acquisition*. Berlin: Zentrum Allgemeine Sprachwissenschaft, 260–83. [*Papers in Linguistics* 15(2).]
- Montrul, Silvina. (2009) “On linguistic complexity, interfaces, and incomplete acquisition”. Paper presented at *Mind-Context Divide*, University of Iowa, 30 April–2 May, 2009.
- Mykhaylyk, Roksolana and Heejeong Ko. (2010) “Optional scrambling is not random: Evidence from English-Ukrainian acquisition”. Merete Anderssen, Kristine Bentzen, and Marit R. Westergaard, eds. *Variation in the input: Studies in the acquisition of word order*. Dordrecht: Springer, 207–40. [*Studies in theoretical psycholinguistics*.]
- Pancheva, Roumyana. (2005) “The rise and fall of second-position clitics”. *Natural language and linguistic theory*. 23(1): 103–67.
- Radeva-Bork, Teodora. (2012) *Single and double clitics in adult and child grammar*. Frankfurt: Peter Lang.
- . (2014) “Computations at the interfaces in child grammar”. Peter Kosta, Steven L. Franks, Teodora Radeva-Bork, and Lilia Schürcks. *Minimalism and beyond: Radicalizing the interfaces*. Amsterdam: John Benjamins, 307–13. [*Language Faculty and Beyond*.]
- Rudin, Catherine. (1997) “Agr-O and Bulgarian pronominal clitics”. Martina Lindseth and Steven L. Franks, eds. *Formal approaches to Slavic linguistics: The Indiana meeting, 1996*. Ann Arbor, MI: Michigan Slavic Publications, 224–52.
- Schaeffer, Jeannette. (1995) “On the acquisition of scrambling in Dutch”. Dawn MacLaughlin and Susan McEwen, eds. *Proceedings of the 19th Annual*

- Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 521–32.
- . (2000) *The acquisition of direct object scrambling and clitic placement*. Amsterdam: John Benjamins.
- Sportiche, Dominique. (1996) “Clitic constructions”. Johan Rooryck, and Laurie Zaring, eds. *Phrase Structure and the Lexicon*. Dordrecht: Kluwer, 213–76.
- Stiasny, Andrea. (2006) *The Acquisition of Clitics in Croatian and Spanish and its Implications for Syntactic Theory*. Ph.D. dissertation, University of Michigan.
- Tomić, Olga M. (1996) “The Balkan Slavic nominal clitics”. Aaron Halpern, and Arnold Zwicky, eds. *Approaching second: Second position clitics and related phenomena*. Stanford, CA: CSLI Publications, 511–36.
- Tryzna, Marta (2010) “Acquisition of object clitics in child Polish: Evidence for D-Linking”. *Proceedings of WIGL 2010*. Madison, WI: Linguistics Students Organization, 71–84. [LSO Working Papers in Linguistics 8.]
- Varela, Ana. (1988) *Binding in Spanish: A theoretical and experimental study*. Ph.D. dissertation, University of Connecticut.
- Wexler, Ken. (1998) “Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage”. *Lingua* 106: 23–79.
- Wexler, Ken, Anna Gavarró, Vincenç Torrens. (2004) “Feature checking and object clitic omission in child Catalan and Spanish”. Reineke Bok-Bennema, Bart Hollebrandse, Brigitte Kampers-Manhe, and Petra Sleeman, eds. *Romance languages and linguistic theory 2002: Selected papers from ‘Going Romance’, Groningen 2002*. Amsterdam: John Benjamins, 253–69.
- Zhou, Peng. (2011) *Interface conditions in child language: A view from Mandarin Chinese*. Ph.D. dissertation, Macquarie University.

# Binding and Labeling in Bulgarian\*

Lilia Schürcks

*Abstract:* This article looks at the colloquial forms of the personal reflexive pronoun in Bulgarian and draws upon their problematic accommodation into Chomsky's classical Binding Theory (1980, 1981, 1982, 1986a, 1986b). The recent labeling account of Chomsky (2013, 2015) is explored with the purpose of finding an explanation for the issue at hand. The notion of markedness is revised, and its major role in the spellout of Bulgarian reflexives is highlighted.

## 1. Introduction

In this paper, I analyze Bulgarian reflexives in terms of Chomsky's labeling algorithm (2013, 2015). I first present the data of interest in Bulgarian. I then show that Bulgarian reflexives are problematic for Chomsky's Binding Theory (1980, 1981, 1982, 1986a, 1986b) because they do not obey c-command relations in the way it predicts. I then develop a labeling account based on the features [ $\pm$ refl] and [ $\pm$ phi], building on my previous work. I further propose that the notion of markedness plays a crucial role in the spellout of Bulgarian reflexives: the more marked a reflexive is in terms of the above-mentioned features, the earlier it is spelled out. I conclude with a summary and an agenda for future research.

## 2. Bulgarian Reflexives

Bulgarian has two sets of reflexive pronouns. There are colloquial personal reflexive pronouns that distinguish person, number, and gender, as well as a set of prescriptive forms that, by contrast, lack these distinctions. The prescriptive forms are shown in (1):

\* I dedicate this work to Catherine Rudin, who has always been a cheerful advocate and loyal friend. I would like to thank Steven Franks, Brian Joseph, and a reviewer for their valuable suggestions and comments. Also, there is some speaker disagreement about the judgments reported here, and particularly for the colloquial long forms *nego si* and *negov si*, Iliyana Krapova only finds local c-commanding antecedents acceptable. It should thus be borne in mind that all data and their analysis reflect the intuitions of the author.

## (1) Personal Reflexive Pronouns

Long Forms	Clitics	
	Acc/Dat	Dat/Gen
sebe si 'self'	se	si
'myself, yourself, etc.'	'myself, yourself, etc.'	
for all persons and cases in sg and pl	for all persons in sg and pl; also 'my, your, his, her, etc.'	

The following examples illustrate the use of the forms above:

- (2) a. Tja se reši.  
'She is combing herself.'
- b. Ivan si dade počivka.  
'Ivan gave himself a break.' (used non-idiomatically)

The accusative use is shown in (2a) and the dative one in (2b). In nonfocused sentences both the long form and the clitic can be used, as in (3a). The long form *sebe si* is accusative and has a contrastive reading in (3b):

- (3) a. Vidjax se/sebe si v ogledaloto.  
'I saw myself in the mirror.'
- b. Vidjax SEBE SI/\*SE v ogledalo, a ne Ivan.  
'It was myself who I saw in the mirror, not Ivan.'

The colloquial personal reflexive pronouns are presented in (4).

With regard to the notions of the classical Binding Theory, the prescriptive forms in (1) have a typical anaphoric character and the personal pronouns show a pronominal character. The colloquial personal reflexive pronouns in (4), however, have questionable status—they are neither anaphors nor pronouns.<sup>1</sup>

<sup>1</sup> For a more detailed study of *nego si* see Franks (2013). In this research the colloquial form is correctly considered to be a Balkan anaphor.

## (4) Long Forms

Sg	Pl
<i>mene si</i> 'myself'	<i>nas si</i> 'ourselves'
<i>tebe si</i> 'yourself'	<i>vas si</i> 'yourselves'
<i>nego si</i> 'himself'	<i>tjax si</i> 'themselves'
<i>neja si</i> 'herself'	
<i>nego si / neja si</i> 'himself / herself / itself'	

We observe exactly the same patterns with the possessive reflexive forms, both the prescriptive possessive reflexive pronouns in (5a) and the colloquial possessive reflexive pronouns in (5b). The former are devoid of person, number, and gender features, while the latter make differentiations in form based on these features:

## (5) a. Possessive Reflexive Pronouns

Long Forms	Clitic
<i>svoj</i> (sg masc) 'my, your, etc.'	<i>si</i> 'my, your, etc.'
<i>svoja</i> (sg fem) 'my, your, etc.'	
<i>svoe</i> (sg neut) 'my, your, etc.'	
<i>svoi</i> (pl) 'my, your, etc.'	

## (5) b. Colloquial Possessive Reflexive Pronouns

<b>Long Forms</b>			
<b>Masc</b>	<b>Fem</b>	<b>Neut</b>	<b>Pl</b>
moj si	moja si 'my' (for all genders and pl)	moe si	moi si
tvoj si	tvoja si 'your' (for all genders and pl)	tvoe si	tvoi si
negov si	negova si 'his' (for all genders and pl)	negovo si	negovi si
nein si	nejna si 'her' (for all genders and pl)	nejno si	nejni si
negov si	negova si his/her/its (for all genders and pl)	negovo si	negov si
naš si	naša si 'our' (for all genders and pl)	naše si	naši si
vaš si	vaša si 'your' (for all genders and pl)	vaše si	vaši si
texen si	tjaxna si 'their' (for all genders and pl)	tjaxno si	texni si

The featural composition of the forms participating in binding structures is as follows:

- (6) sebe si (REFL REFL-CL)  
 svoj (REFL<sub>POSS.MASC</sub>)  
 si (REFL-CL<sub>DAT</sub>)  
 se (REFL-CL<sub>ACC</sub>)  
 si (REFL-CL<sub>POSS</sub>)  
 nego si (PRON<sub>3SG.MASC</sub> REFL-CL)  
 negov si (PRON<sub>POSS.3SG.MASC</sub> REFL-CL<sub>POSS</sub>)  
 nego (PRON<sub>3SG.MASC.ACC</sub>)  
 na nego (PRON<sub>3SG.MASC.DAT</sub>)  
 negov (PRON<sub>POSS.3SG.MASC</sub>)  
 go (CL<sub>3SG.MASC.ACC</sub>)  
 mu (CL<sub>3SG.MASC.DAT</sub>)

In Bulgarian the dividing line between anaphors and pronouns is not clear; (6) represents the lexical items involved in binding. The form *nego si* patterns with reflexives and pronouns, whereas in some contexts, the form *nego* appears where reflexives are expected. This results in a striking overlap of forms. The forms *svojata* and *si* are not interchangeable in all contexts:

- (7) a. Vseki vze knigata *si*.  
 'Everybody took their own book.'  
 b. Vseki vze *svojata* kniga, samo Marija vze тази na Ivan.  
 'Everybody took their own book, only Marija took Ivan's.'

In other words, the clitic forms usually occur in neutral contexts while *svoj* is always stressed in some way.

Other examples where the prescriptive clitic forms (*si*, *se*) are often replaced by clitics containing phi-features—*mi* (1sg), *ti* (2sg), *ni* (1pl), etc.—are shown in (8):

- (8) a. Kazax na sestra *mi* da dojde. (colloquial)  
 'I told my sister to come.'  
 b. Kazax na sestra *si* da dojde. (prescriptive)  
 'I told my sister to come.'

If we take the English anaphor *himself* and the pronoun *him* as a point of departure, the Bulgarian anaphors and pronouns look like this:

- (9) ↗ short-distance function *sebe si*, *nego si*, *se*, *si*  
 himself  
 ↘ long-distance function *nego si*
- (10) ↗ short-distance function *svoj*, *svoj si*, *negov si*, *si*  
 his → long-distance function *negov si*  
 ↘ the possessive pronouns *negov*, *mu*

Let us consider (9) first. The English *himself* fulfills both the short-distance function and the long-distance function, functions carried by different forms in Bulgarian. The short-distance function is fulfilled by the long forms *sebe si* and *nego si* and the short forms *se* and *si*. These patterns are illustrated below in (11) and (12). The English counterparts below each Bulgarian example, here



and elsewhere, serve as translations and make for useful points of comparison:<sup>2</sup>

- (11) a. Ivan mrazi sebe si/ nego si.  
 'Ivan hates himself.'  
 (i) sebe si = Ivan  
 (ii) nego si = Ivan
- b. Ivan se mrazi.  
 'Ivan hates himself.'  
 (i) se = Ivan
- (12) a. Ivan govori na sebe si/na nego si.  
 'Ivan talks to himself.'  
 (i) sebe si = Ivan  
 (ii) nego si = Ivan or a higher subject
- b. Ivan si govori.  
 'Ivan talks to himself.'  
 (i) si = Ivan

In (11a) and (12a), the form *sebe si* is obligatorily bound by *Ivan*. In the same examples, *nego si* can also be bound by *Ivan*. In (11b) and (12b), the subject *Ivan* binds the short forms *se* and *si*. The long-distance function of *himself* is fulfilled by *nego si*. This is illustrated in (13):

- (13) a. Javor<sub>1</sub> kaza, če Ivan<sub>2</sub> zadade na Petâr<sub>3</sub> vâpros za sebe si<sub>\*1,2,\*3</sub>.  
 'Javor said that Ivan asked Petâr a question about himself.'  
 (i) \*sebe si = Javor  
 (ii) sebe si = Ivan  
 (iii) \*sebe si = Petâr
- b. Javor<sub>1</sub> kaza, če [Ivan<sub>2</sub> zadade na Petâr<sub>3</sub> vâpros za nego si<sub>1,2,3</sub>].  
 'Javor said that Ivan asked Petâr a question about himself.'  
 (i) nego si = Javor  
 (ii) nego si = Ivan  
 (iii) nego si = Petâr

<sup>2</sup> It must be noted that not all speakers of Bulgarian accept the distribution shown in (11) through (16) of these reflexive forms. My conclusions, therefore, hold only for speakers who accept the data as presented here. I leave for another study the very interesting question of how to analyze the distributional patterns of speakers with different judgments and what to make of the differences in judgments.

(13) shows again that *sebe si* is always bound by the subject of the minimal clause, *Ivan*. As expected, *nego si* can be bound by any agreeing phrase in the sentence.

The English pronoun *his* corresponds to both the reflexive possessives, i.e., *svoj, svoj si, negov si, si* and pronouns, i.e., *negov* and *mu*. This is illustrated by the examples in (14) and (15):

- (14) a. Ivan vidja svojata/negovata si kniga.  
'Ivan saw his book.'
- b. Ivan vidja knigata si.  
'Ivan saw his book.'
- c. Ivan si vidja knigata.  
'Ivan saw his book.'
- (15) a. Ivan vidja negovata kniga.  
'Ivan saw his book.' (Ivan's book, or someone else's book)
- b. Ivan vidja knigata mu.  
'Ivan saw his book.' (Ivan's book, or someone else's book)
- c. Ivan mu vidja knigata.  
'Ivan saw his book.' (only someone else's book)

In all cases of (14), the subject *Ivan* obligatorily binds the respective possessive forms, i.e., *svojata, negovata si, si*. In (14c), the possessor *si* has moved out of the DP [<sub>DP</sub> *knigata si*] and has become part of the verbal complex. In (15), the long form *negovata* and the short form *mu* can be bound either by *Ivan*, or by another person in the discourse. The long-distance function of *negov si* is demonstrated in (16):

- (16) a. Ivan<sub>1</sub> kazva, če Petâr<sub>2</sub> xaresva svojata<sub>\*1,2</sub> kniga.  
'Ivan says that Petâr likes his book.'
- (i) \*svojata = Ivan  
(ii) svojata = Petâr
- b. Ivan<sub>1</sub> kazva, če Petâr<sub>2</sub> xaresva negovata si<sub>1,2</sub> kniga.  
'Ivan says that Petâr likes his book.'
- (i) negovata si = Ivan  
(ii) negovata si = Petâr

In (16), the form *svojata* is bound only by the subject of the minimal clause, *Petâr*, while *negov si* can be bound either by the long-distance agreeing phrase, *Ivan* or the local agreeing phrase *Petâr*.

### 3. Why Is There a Problem for the Theory?

Classical Binding Theory involves a classification of NPs based on the features [ $\pm$ anaphoric] and [ $\pm$ pronominal]. If we concentrate on the overt NPs, the cross-classification in terms of these features together with sample items from English is presented in (17):

(17)	[+anaphoric, -pronominal]	himself, each other
	[+anaphoric, +pronominal]	—
	[-anaphoric, +pronominal]	he, him
	[-anaphoric, -pronominal]	Peter, the girl

It is clear that the combination [+anaphoric, +pronominal] does not have any overt manifestation. Chomsky (1982: 78–89) claims that this holds true for most languages, and that only the empty element PRO has this type of feature specification (see also Chomsky 1986b). The explanation is that an overt category with this lexical content would be ungoverned in terms of Principles A and B of the classical Binding Theory and consequently would violate the Case Filter. The feature specifications of the lexical elements presented above relate to classical Binding Theory in the following way:

(18) Principle A: A [+anaphoric] NP must be bound in its governing category (or in a local domain D).

Principle B: A [+pronominal] NP must be free in its governing category (or in a local domain D).

Principle C: A [-anaphoric, -pronominal] NP must be free.

It is not obvious that (17) is valid for all languages. Thráinsson (1991) presents arguments that the reflexives in Scandinavian languages do not fit into the classification in (17). Everaert (1986, 1991) argues that, apart from the regular binding conditions, “contextual binding principles” are needed. Let us recall some of the empirical evidence speaking against classifying the Bulgarian overt NPs in terms of the oppositions [ $\pm$ anaphoric] and [ $\pm$ pronominal]. Consider the examples in (19) and (20):

- (19) Ivan<sub>1</sub> kazva, že Petâr<sub>2</sub> mrazi sebe si<sub>\*1,2</sub> / nego si<sub>1,2</sub>  
 Ivan says that Petâr hates REFL REFL-CL / pron<sub>3SG.MASC</sub> REFL-CL  
 / nego<sub>1,\*2</sub>.  
 / pron<sub>3SG.MASC</sub>  
 ‘Ivan says that Petâr hates him(self).’  
 (i) sebe si = Petâr (not Ivan)  
 (ii) nego si = Ivan or Petâr  
 (iii) nego = Ivan or someone else

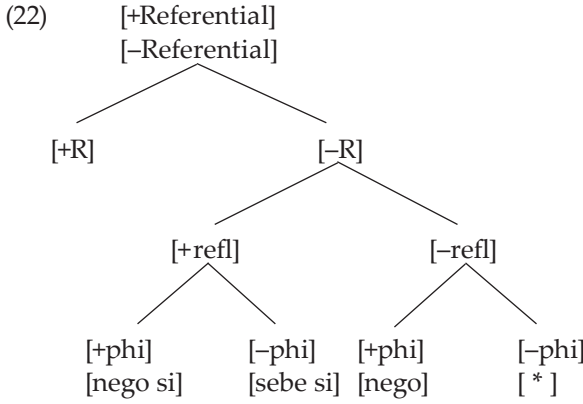
- (20) Ivanovijat<sub>1</sub> bašta<sub>2</sub> kritikuva sebe si<sub>\*1,2</sub> / nego si<sub>1,2</sub>/  
 Ivan’s father criticizes REFL REFL-CL / pron<sub>3SG.MASC</sub> REFL-CL/  
 nego<sub>1,\*2</sub>.  
 pron<sub>3SG.MASC</sub>  
 ‘Ivan’s father criticizes him(self).’  
 (i) sebe si = bašta  
 (ii) nego si = Ivan or bašta  
 (iii) nego = Ivan or someone else

In (19), we see that *nego si* can pattern with the reflexive *sebe si* (it is bound by *Petâr*), thus behaving like a local anaphor, and patterns equally with the pronoun *nego*, which can refer to *Ivan*. In the same way in (20), *nego si* can be bound by the c-commanding phrase *bašta*, as well as by the non-c-commanding phrase *Ivanovijat*. Hence we face the problem of how to classify *nego si*. In terms of the oppositions in (17) and Binding Theory in (18), this element is [+anaphoric, +pronominal]. But as noted above this type of combination, i.e., [+anaphoric, +pronominal] is ruled out for overt elements.

I thus abandon the NP classification [ $\pm$ anaphoric] and [ $\pm$ pronominal] presented in (17), since it is not effective for the appropriate analysis of the Bulgarian forms. In more general terms, binding relations reflect the relationship between an antecedent (contained in the c-commanding phrase) and a dependent element. In other words, we talk about a relationship established between two NPs. My claim is that the Bulgarian forms start out marked with the further specification of the categorial feature N, i.e., [ $\pm$ Referential], as in (21):

- (21) [ $\pm$ referential] or [ $\pm$ R]

The further evolution of the feature content and their PF realization into the forms *sebe si*, *nego si*, and *nego* are shown in (22):



I take reflexivity,  $[\pm\text{refl}]$ , also to be a further specification of the categorial feature N. Then N can be specified as either  $[-\text{refl}]$  or  $[\text{+refl}]$ . Along with the two types of categorial features, the phi-features  $[\pm\varphi]$  are also active for analyzing binding in Bulgarian. As already noted above, the derivation for all forms comes from the numeration with the feature content  $[\pm R]$ , whereas the features  $[\pm\text{refl}]$  and  $[\pm\varphi]$  are added in the process of derivation. If we include the feature oppositions  $[\pm\text{refl}]$  and  $[\pm\varphi]$ , the feature specifications of the binding forms in Bulgarian are as in (23):

- (23) a.  $[\text{+refl}, \text{+}\varphi]$   
 b.  $[\text{+refl}, \text{-}\varphi]$   
 c.  $[\text{-refl}, \text{+}\varphi]$

(23a) shows the rule representation of reflexives like *nego si*, (23b) demonstrates that for reflexives like *sebe si*, and (23c) is the rule representation for pronouns like *nego*. (23c) illustrates the feature content of the pronouns which are the elsewhere forms. The reflexives in (23a) and (23b) are the marked case, with (23b) more marked than (23a). Here we observe typological markedness, namely, asymmetrical properties of otherwise equal linguistic elements. Typological markedness presents causal relationships among cross-linguistic asymmetries reflecting the encoding of function into grammatical form.

My next theoretical step is that each NP (as we see later on, syntactic objects) involved in binding consists of the two types of categorial distinctions,  $[\pm\text{Referential}]$ . These claims are summarized in (24):

- (24) •  $[\pm\text{Referential}]$  and  $[\pm\text{refl}]$  are further specifications of the categorial feature N.  
 • In binding configurations, the forms start as  $[\pm\text{Referential}]$  as part of a c-selection process



## (27) Binding Theory

## a. [+R]:

(i) If  $\alpha$  has the feature characteristics [+R], interpret it as disjoint from every c-commanding syntactic object.

## b. [-R]:

(ii) If  $\alpha$  has the feature characteristics [+refl,  $-\varphi$ ], interpret it as coreferential with some subjects in  $D^*$ .

(iii) If  $\alpha$  has the feature characteristics [+refl,  $+\varphi$ ], interpret it as coreferential with some syntactic objects with  $\varphi$  features matching those of a.

(iv) If  $\alpha$  has all the feature characteristics [-refl,  $+\varphi$ ], interpret it as disjoint from all c-commanding SO in  $D^*$ .

\*  $D$  is construed as the domain of application, that is, the TP or DP phase (the minimum clause or the DP).

The local domain  $D$  (the respective phase) is identified only for the [+refl,  $-\varphi$ ] and [-refl,  $+\varphi$ ] forms.

The examples in (28–31) below illustrate how the reformulated Binding Principles (27) work for Bulgarian. The sentences in (28) reflect the phenomena generalized in Principle (27b.ii); the sentences in (29), (30a) and (30c) cover Principle (27b.iii); the sentences in (30b) and (30d) illustrate Principle (27b.iv), and finally (31) shows the effect of Principle (27a.i):

- (28) a. Ivan mrazi sebe si.  
'Ivan hates himself.'
- b. Ivan se mrazi.  
'Ivan hates himself.'
- c. Ivan xaresva svojata kniga.  
'Ivan likes his book.' (Ivan's book)
- (29) a. Ivan mrazi nego si.  
'Ivan hates himself.'
- b. Ivan xaresva negovata si kniga.  
'Ivan likes his book.' (Ivan's book)
- (30) a. Ivanovijat brat kritikuva nego si.  
'Ivan's brother criticizes himself/him.'
- (i) nego si = himself = brat (brother)  
(ii) nego si = him = Ivan

- (30) b. Ivanovijat brat kritikuva nego.  
 ‘Ivan’s brother criticizes him.’ (Ivan or someone else)
- c. Ivanovijat brat kritikuva negovata si kniga.  
 ‘Ivan’s brother criticizes his book.’
- (i) negovata si = brat (the brother’s)  
 (ii) negovata si = Ivan (Ivan’s)
- d. Ivanovijat brat kritikuva negovata kniga.  
 ‘Ivan’s brother criticizes his book.’
- (i) negovata = brat (the brother’s)  
 (ii) negovata = Ivan (Ivan’s) or someone else’s
- (31) a. Ivan mrazi Petâr.  
 ‘Ivan hates Petâr.’
- b. Ivan kritikuva prijatelja si.  
 ‘Ivan criticizes his friend.’

In (28), the forms *sebe si*, *se*, and *svojata* are the marked case, i.e., they have the feature characteristics [+refl,  $-\varphi$ ] and they are coreferential with the subject *Ivan* in the local domain D, the TP. In (29), the forms *nego si* and *negovata si* possess the feature characteristics [+refl,  $+\varphi$ ] ( $\varphi = 3\text{sg}$ ) and they corefer with the subject *Ivan* ( $\varphi=3\text{sg}$ ). Thus the  $\varphi$  features of the correspondent form must be identical with the  $\varphi$  features of the syntactic objects. In (30a) and (30c), the forms *nego si* and *negovata si* have the feature characteristics [+refl,  $+\varphi$ ] ( $\varphi=3\text{sg}$ ) and they corefer with the non-c-commanding phrase *Ivanovijat*, with the same range of  $\varphi$  features ( $\varphi=3\text{sg}$ ). In (30b) and (30d), the forms *nego* and *negovata* have the feature characteristics [ $-\text{refl}$ ,  $+\varphi$ ] ( $\varphi=3\text{sg}$ ) and they must not have an antecedent (contained in the c-commanding syntactic objects) in D, i.e., the TP phase. Finally, the forms *Petâr* and *prijatelja* in (31) have the feature characteristics [+R] and are disjoint from every c-commanding syntactic object (*Ivan*) in D, the TP phase.

#### 4. The Labeling Algorithm and Binding

Let us consider the way a certain form  $\alpha$  enters in a binding configuration in Bulgarian. For more clarity, consider the examples in (32), where the feature characteristics of the syntactic objects in question are also shown:



- (32) Ivan mrazi sebe si / nego si / nego.  
 Ivan hates REFL REFL.CL / PRON<sub>3SG.MASC</sub> REFL.CL / PRON<sub>3SG.MASC</sub>  
 ‘Ivan hates him(self).’  
 sebe si [+refl, -φ]  
 nego si [+refl, +φ]  
 nego [-refl, +φ]

My hypothesis for binding in Bulgarian is that the forms involved in binding  $\alpha$  (see (32) *sebe si*, *nego si*, *nego*) start out as a syntactic object where initially only the [ $\pm$ Referential] (or more specifically [-Referential]) feature oppositions are involved. It is only later, in the process of multiple Merge that the other feature oppositions are added, that is, [ $\pm$ refl] and [ $\pm\phi$ ]. This particular set of features is added in the process of syntactic computation, since the binding configuration involves looking for an antecedent in a certain domain (a TP or DP). First, the [ $\pm\phi$ ] features are added in the process of multiple Merge operations. It is only later that the [ $\pm$ refl] features are acquired.

In a modern reformulation of syntactic theory, Chomsky (2013, 2015) abandons familiar notions of X-bar theory and specifically its theoretical stipulation that all phrases are inherently endocentric. Thus, Chomsky argues that the notion of “projection,” as it has been traditionally understood, is a theory-internal concept and as such should be treated with a proper amount of scientific skepticism. He replaces X-bar theory in favor of Merge, a syntactic operation that combines objects which are themselves either heads (H) or phrases (XP). Therefore both bar-levels (X') and Specifiers are no longer relevant.

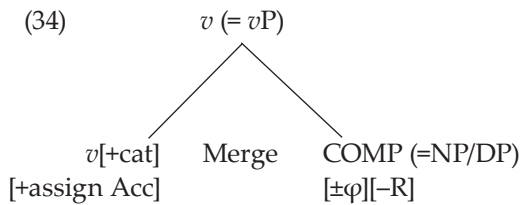
In combining syntactic objects, Merge creates unordered (set theoretic) sets that are potentially infinitely recursive. Lastly, Merge is of two types: External and Internal. External Merge combines a syntactic object with a new syntactic object, whereas Internal Merge combines a syntactic object with an already-Merged syntactic object:

- (33) Merge (X, Y)  $\rightarrow$  {X, Y} External Merge  
 Merge (Z, Y) where  $Y \in Z \rightarrow$  {Y, {<sub>Z</sub> X, Y}} Internal Merge

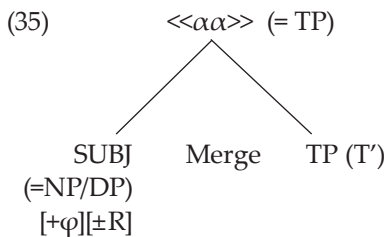
The standard Chomskyan framework of generation theorizes that strings created by Merge must be transferred to interfaces which interpret the strings as instructions. The two interfaces are the Conceptual-Intentional interface and the Sensory-Motor interface. Generation in this manner is strictly cyclic, being limited to what are called phases, namely *vP*, CP, and DP. During the transfer operation, syntactic structures are labeled by an algorithmic process that allows them to be recognized and interpreted by the interfaces. Merge, which operates indiscriminately, does not itself label but merely combines ob-

jects (or sets of objects). Therefore, the interfaces are responsible for deciding what strings are grammatical and what strings crash.

Chomsky describes this labeling algorithm, which happens during the transfer operation, in these terms: for structures of the type  $\{H, XP\}$ , labeling proceeds naturally—the algorithm determines that  $H$  is the label of the entire structure. For example, given the Merging of two items, a verb and its complement, the set formed will be  $\{v, \text{COMP}\}$ . The labeling algorithm will automatically choose the head  $v$  as the label. Going back to applying that to binding relations in Bulgarian,  $v$  will search for a complement with the features  $[-\text{Referential}]$  and, as shown in (34), the set will be  $\{v, \{\text{COMP}[\pm\varphi] [-\text{Referential}]\}\}$ :



According to Chomsky, the interesting types are structures such as  $\{H, H\}$  (two heads, which I do not treat here since they are not relevant for binding) and  $\{XP, YP\}$  (two full phrases). The items must find shared features that can be used as a label. The labeling algorithm, then, must consist of a minimal search operation to target and make salient these features. In the case of, for example, a subject (DP) and its predicate (TP), the salient features are  $\varphi$ -features (interpretable on DP and uninterpretable on TP, traditionally), and hence subjects agree in terms of  $\varphi$ -features with their predicates. The resulting label is  $\langle\alpha\alpha\rangle$  in (35):



In (34) and (35), we witness multiple MERGE operations:

- (36) The binding relations are established at the point at which Transfer (CI or conceptual-intentional) and Transfer (SM or sensory motor) operations apply

I adopt the view of previous minimalist frameworks (Chomsky 1993, 1995) that a given form *a* (*sebe si*, *nego si*, *nego*) in (32) merges/moves for Case checking. The Case feature being [-interpretable], is checked and erased. At first, the only formal features that are available, that is, the categorial features [ $\pm$ Referential], are checked but remain accessible for future operations. Assuming that there is a multiple checking operation in the process of syntactic computation, the other feature oppositions, [ $\pm$ refl] and [ $\pm\varphi$ ], are added and also checked. The [ $\pm\varphi$ ] features are acquired in the process of merging the form [-Referential] and the verb. Hence in the syntax, syntactic objects like those in (34) and (35) are labeled. Since only the oppositions [ $\pm$ refl] and [ $\pm\varphi$ ] turn out to be active for considering the binding phenomena of Bulgarian (because they are added later in the process of labeling), we continue by considering only these two pairs.

Consequently, we obtain theoretically four different syntactic objects defined by different combinations of the feature characteristics, namely:

- |                |                |
|----------------|----------------|
| (37) a. SO1    | b. SO2         |
| [+refl]        | [+refl]        |
| [- $\varphi$ ] | [+ $\varphi$ ] |
| c. SO3         | d. SO4         |
| [-refl]        | [-refl]        |
| [+ $\varphi$ ] | [- $\varphi$ ] |

Universal Grammar provides the four possible lists of features in (37). Out of these four possibilities, only the first three, SO1, SO2, and SO3, are sets which are processed by Transfer operations as overt forms. The fourth option SO4 is recognized as either *pro* or *A* trace.

$\varphi$ -features, when present, are checked and deleted but not erased in the course of the syntactic computation. Both  $\varphi$ -features and categorial features are accessible for the Transfer operations. Consequently, the syntactic objects will be labeled as the lexical items *sebe si* in (38), as *nego si* in (39), and as *nego* in (40):

- (38) *sebe si* ← [ $\pm$ refl]  
                  [- $\varphi$ ]
- (39) *nego si* ← [ $\pm$ refl]  
                  [+ $\varphi$ ]
- (40) *nego* ← [-refl]  
                  [+ $\varphi$ ]

It is the three feature sets in (38–40) which will be processed by T (CI = conceptual-intentional system) and T (SM = sensory motor system). In other words, the binding interpretive procedure will recognize only the following sentences (see (32)):

- (41) a. Ivan mrazi sebe si.  
'Ivan/He hates himself.'
- b. Ivan mrazi nego si.  
'Ivan/He hates himself.'
- c. Ivan mrazi nego.  
'Ivan/He hates him.'

Consequently, only legitimate label sets are allowed to be interpreted at the interfaces.

### 5. Degree of Markedness Spell-Out Principle

An important issue to be addressed here is the order in which the different forms in Bulgarian, presented with the combination of feature characteristics [ $\pm$ refl], and [ $\pm\varphi$ ], are labeled. We have seen that the variety of forms (lexical items) relevant for binding in Bulgarian are the following:

- (42) a. [ $+\text{refl}, -\varphi$ ]  $\rightarrow$  sebe si (REFL REFL-CL)  
svoj (REFL<sub>POSS</sub>.MASC)  
si (REFL-CL<sub>DAT</sub>)  
se (REFL-CL<sub>ACC</sub>)  
si (REFL-CL<sub>POSS</sub>)
- b. [ $+\text{refl}, +\varphi$ ]  $\rightarrow$  nego si (PRON<sub>3SG.MASC</sub> REFL-CL)  
negov si (PRON<sub>POSS.3SG.MASC</sub> REFL-CL<sub>POSS</sub>)
- c. [ $-\text{refl}, +\varphi$ ]  $\rightarrow$  nego (PRON<sub>3SG.MASC.ACC</sub>)  
na nego (PRON<sub>3SG.MASC.DAT</sub>)  
go (CL<sub>3SG.MASC.ACC</sub>)  
mu (CL<sub>3SG.MASC.DAT</sub>)

The question arises now about the order in which these objects are spelled out. In earlier work (Schürcks 2003, 2008, 2014), I proposed that this order bears directly on the degree of markedness and specificity. The following Degree of Markedness Spell-Out Principle, (43), emerges expressing the relationship between the degree of markedness and Spell-Out:

- (43) The more marked a structural representation is, the earlier it is spelled out. The unmarked structural representation is the last to be spelled out.

In other words, the more unmarked a representation is, the more delayed its process of Spell-Out is. In terms of the labeling approach, I reformulate the Degree of Markedness Spell-Out Principle as in (44):

- (44) The Transfer CI (= conceptual-intentional system) and Transfer SM (= sensory motor system) operations for [+refl,- $\varphi$ ] forms (*sebe si, svoj, si, se*) precede those for [+refl,+ $\varphi$ ] (*nego si, negov si*). The subsequent Transfer operations will produce the SO *nego* [-refl,+ $\varphi$ ].

Taking into account the new principle (44), the forms (38–40) are three syntactic objects consisting of the same set of features, and the structural description of any of the three rules is contained in that of the others. Out of these syntactic objects, the Transfer (CI) and Transfer (SM) operations for [+refl,- $\varphi$ ] forms—*sebe si, svoj, si, se*, that is, (38)—apply first. Consequently, this syntactic object is also the first to be processed. The next one in the row will be (39), followed by (40). The SOs in (40) will be the last to be processed.

The sensory motor system ordering in terms of precedence is illustrated below in (45) (the sign '>' means processed earlier than or precedes the processing of):

- (45) [+refl,- $\varphi$ ] > [+refl,+ $\varphi$ ] > [-refl,+ $\varphi$ ]

In (45), Transfer (CI) and Transfer (SM) operations apply first for the syntactic object [+refl,- $\varphi$ ], while operations are delayed for the syntactic object [-refl,+ $\varphi$ ]. If we translate the representations into the respective range of the lexical items in (46), then the scale of Spell-Out ordering is the following:

- (46) *sebe si* > *nego si* > *nego*

Thus the Spell-Out (or precedence) of the different forms relevant for binding in Bulgarian is ranked according to the respective Transfer (CI) and Transfer (SM) operations.

## 6. Conclusion

In this paper, I have revised the Degree of Markedness Principle, formulated in my previous work (Schürcks 2003, 2008, 2014). The concept of typological markedness has led to many controversial views and is widely considered to be a vague term for tackling syntactic phenomena (Haspelmath 2006).

However, I have applied the newest development of the Minimalist Program (Chomsky 2013, 2015) in order to explain the precedence of the Spell-Out forms and relate it to markedness. Some research questions certainly remain:<sup>3</sup>

1. Can we analyze binding phenomena completely within a framework clearly breaking from the phrase structure tradition?
2. Are both operations Merge and Transfer part of UG? Can we account for binding considering only Merge?
3. How do the notions of hierarchical structure and labels co-occur in binding? What exactly is the relevance of c-command (in terms of Chomsky 2013, 2015)?

I leave these for future work.

Freie Universität Berlin, Germany  
schuercks@gmail.com

## References

- Bošković, Željko. (2016) "On the timing of labeling: Deducing comp-trace effects, the subject condition, the adjunct condition, and tucking in from labeling". *The linguistic review* 33: 17–66.
- . (2017) "On movement out of moved elements, labels, and phases". *Linguistic inquiry* 49: 247–82.
- Chomsky, Noam. (1980) "On binding". *Linguistic inquiry* 11: 1–46.
- . (1981) *Lectures on government and binding*. Dordrecht: Foris.
- . (1982) *Some aspects and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.
- . (1986a) *Barriers*. Cambridge, MA: MIT Press.
- . (1986b) *Knowledge of language*. New York: Praeger.
- . (1993) "A minimalist program for linguistic theory". Ken Hale and Samuel J. Keyser, eds. *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press, 1–52. [Also published in N. Chomsky, *The Minimalist Program*, 167–217. Cambridge, MA: MIT Press (1995)].
- . (1995) *The minimalist program*. Cambridge, MA: MIT Press.
- . (2013) "Problems on projection". *Lingua* 130: 33–49.
- . (2015) "Problems on projection: Extensions". Elisa di Domenico, Cornelia Hamann, and Simona Matteini, eds. *Structures, strategies and beyond: Studies in honour of Adriana Belletti*. Amsterdam: John Benjamins.

<sup>3</sup> Bošković 2016 and Bošković 2017 certainly contribute to their solution.

- Everaert, Martin. (1986) *The syntax of reflexivization*. Dordrecht: Foris.
- . (1991) “Contextual determination of the anaphor/pronominal distinction”. Jan Koster and Eric Reuland, eds. *Long-distance anaphora*. Cambridge: Cambridge University Press, 77–118.
- Franks, Steven L. (2013) “Bulgarian *nego si* is a Balkan anaphor”. *Linguistique balkanique* 52: 3–28.
- Haspelmath, Martin. (2006) “Against markedness (and what to replace it with)”. *Journal of linguistics* 42(1): 25–70.
- Schürcks, Lilia. (2003) *Binding and Bulgarian*. PhD dissertation, University of Groningen. [Groningen Dissertations in Linguistics GRODIL 44.]
- . (2008) “Binding and discourse: Where syntax and pragmatics meet”. *Potsdam Linguistic Investigations* 23. Frankfurt am Main: Peter Lang.
- . (2014) “Syntax presupposes, morphology disposes: Markedness and NP typology”. Lilia Schürcks, Anastasia Giannakidou, and Urtzi Etxeberria, eds. *The nominal structure in Slavic and beyond*. Berlin-London: De Gruyter Mouton, 391–412. [Studies in Generative Grammar 116.]
- Thráinsson, Höskuldur. (1991) “Long-distance reflexives and the typology of NPs”. Jan Koster and Eric Reuland, eds. *Long-distance anaphora*. Cambridge: Cambridge University Press, 49–75.

# Two Declarative Complementizers in Bulgarian\*

Vesela Simeonova

*Abstract:* This paper investigates the syntactic and semantic properties of two declarative complementizers in Bulgarian. The analysis supports and extends recent developments in the theory of complementation such as Moulton 2009, which regard noun clausal complements as a property and not a proposition. Evidence from Bulgarian allows for a unified analysis of both noun and propositional attitude complements. This paper also sheds light on the long-debated relation between factivity and relativization, arguing against proposals that regard factives as covert relatives (Kayne, 2008, Krapova, 2010). I propose instead that the underlying property that unifies them is the notion of exemplification.

## 1. Introduction

The rich properties of complementizers and their role for the syntactic and semantic composition of the embedded clauses in the languages of the Balkan area have been a topic of research for decades, for example Rudin (1982/2013), Rudin (1988), Joseph (1985, 2016), Dobrovie-Sorin (1989), Farkas (1989), Rivero (1994), Roussou (1994, 2010), Varlokosta (1994), Tomić (1996), Krapova and Karastaneva (2000), Ammann and Van Der Auwera (2004), Isac and Jakab (2004), Arsenijević (2009), Becker (2010), Krapova (2010), Baunaz (2014, 2016), Boye and Kehayov (2016).

---

\* I thank Catherine, who gave me feedback on an ancient ancestor of this paper, my 2013 undergraduate honours thesis, and has helped me, supported me, and inspired me in multiple ways on many other linguistic and non-linguistic occasions. I am grateful also to Carlos de Cuba, thanks to whom I was able to appreciate this puzzle (and who supervised that ancient project), and to Iliyana Krapova for laying it out for Bulgarian in Krapova 2010. I thank Ana Arregui and Marisa Rivero for many crucial discussions on the ongoing project in the fall of 2017 while it shapeshifted (from a syntax paper) into this. For other comments that might have affected the course and shape of this project at various stages, I am grateful to Karsten Koch, Andres Salanova, Keir Moulton, Dennis Ott, the Carleton Reading group, the Slavic Linguistic Society 2016 audience, and two anonymous reviewers of this volume. Finally, I thank Steven Franks for giving me the opportunity to write this in honor of Catherine.

Steven L. Franks, Vrinda Chidambaram, Brian D. Joseph, and Iliyana Krapova, eds. *Katerino Mome: Studies in Bulgarian Morphosyntax in Honor of Catherine Rudin*. Bloomington, IN: Slavica Publishers, 2018, 271–85.



My paper contributes to this pool of research an investigation of two complementizers that head embedded declarative finite clauses in Bulgarian. The peculiar distribution of these complementizers can inform theories of content, propositional attitude, relativization, noun modifying clauses, and shed light on the quirky nature of emotive factive predicates. Here, I merely scratch the surface in the hope of raising more awareness of the impact the Bulgarian puzzle can have. I do so from a different point of view from the one in the first extensive discussion of the core data by Krapova (2010), which was focused on relatives and the nature of emotives.

The core data are presented in section 2. In section 3 I argue that complementizers are not semantically transparent but serve an important role in semantic composition. I discuss some theoretical implications and crosslinguistic variation in section 4. Section 5 concludes.

## 2. Core Data: Two Declarative Complementizers

Bulgarian has two morphologically distinct declarative complementizers, *deto* and *če*.<sup>1</sup> Only *če* introduces embedded clauses under propositional attitude predicates, as in (1), as well as noun modifying clauses, (2). *Deto* is not allowed in these environments.

- (1) Ivan misli / kaza [če/\***deto** Marija e tuk].  
 Ivan thinks / said če/deto Maria be<sub>PRES,3SG</sub> here  
 ‘Ivan thinks/said that Maria is here.’

- (2) idejata, **če/\*deto** Ivan e tuk.  
 idea<sub>DEF</sub> če/deto Ivan is here  
 ‘the idea that Ivan is here.’

Only *deto* (in addition to relative *Wh*-pronouns) is used in relative clauses, as in (3).

- (3) idejata, **kojato/deto/\*če** Ivan zapisa v beležnika  
 idea<sub>DEF</sub> which/deto/če Ivan wrote in notebook<sub>DEF</sub>  
 ‘the idea that Ivan wrote down in the notebook’

<sup>1</sup> Bulgarian has also a number of non-declarative complementizers or particles, such as *da* (subjunctive, polarity), *li* (interrogative), *dali* (interrogative), *ako* (conditional). This paper only discusses the declarative ones. A reviewer asks what prevents the use of *če* in complements of verbs like *iskam* ‘want’. I follow Roussou (2010) in assuming that *da* is a polarity particle selected by volitional verbs, while *če* does not have polarity properties.

There is only one environment in which both *deto* and *če* are allowed: a subset of matrix verbs that select a finite declarative clause, restricted particularly to emotive factive predicates, (4). This peculiar fact has been noted in the English linguistic literature as early as Rudin (1982), and is recently discussed in detail in Krapova (2010) and de Cuba (2017).

- (4) Ivan sâžaljava [če/deto Maria e tuk].  
 Ivan regrets če/deto Maria be<sub>PRES.3SG</sub> here  
 ‘Ivan regrets that Maria is here.’

This paints a curious picture of the distribution of the two complementizers. There exist proposals, e.g., Roussou 2010, for a similar situation in Modern Greek, that certain complementizers directly encode factivity (see also Baunaz 2016 for a similar claim for Bulgarian). But in Bulgarian, factivity alone does not regulate the choice of *deto*, because factive nouns do not take *deto* (2), and relative clauses, as in (4), are not necessarily factive. Another approach to the phenomenon, Krapova 2010, posits that emotive factive clauses are in fact relative clauses and that *deto* is simply a relativizer. However, that approach would predict that *če* would not be allowed in emotives, as it is not allowed in relatives, (3), but (4) shows that emotives do not pattern as relatives in terms of the complementizer distribution. The goal of the paper is to offer a new account of the phenomenon, which does not directly rely on factivity or relativization, but rests instead on the notions of content and exemplification.

### 3. Analysis

The gist of the analysis is that *če* is a content complementizer and *deto* is an exemplifying complementizer in the sense of Kratzer 2006, notions developed below. Developing ideas from Moulton 2009, this captures the distribution of *če* under nouns and verbs particularly encoding content. The proposal also explains the connection between factivity and relativization, noticed at least since Kiparsky and Kiparsky 1970, and more recently notably in Aboh 2005—in terms not of reducing either one to the other, but rather to abstracting both to the more general notion of exemplification.

#### 3.1. The Notion of Content

The analysis of complementizers developed here is based on the concept of content, as in PROPOSITIONAL CONTENT (though see Rawlins, 2013 for a more general version), a descendant of SUBJECT MATTER in philosophy (Lewis, 1988a, 1988b, 2003). The concept of content is used to relate propositional attitude

predicates, such as *think*, *believe*, *know*, to what is thought, believed, known, e.g., the proposition *John is a spy*:

- (5) Bill thinks [<sub>CP</sub> that John is a spy].

A seminal semantic theory on how this relation is realized is proposed in Hintikka 1969 (following Frege in spirit) and schematized in (6): the matrix verb is a quantifier over possible worlds. It takes the embedded clause as its propositional argument  $p$  and returns all the belief worlds (marked by *Dox* for doxastic) of the matrix subject,  $x$ , in which the proposition is true.

- (6)  $\llbracket \text{believe} \rrbracket = \lambda p. \lambda x. \lambda w. \forall w' (w' \in \text{Dox}(x)(w) \rightarrow p(w') = 1)$

However, this view has recently been challenged by both philosophers and linguists. As Moltmann (2013b) puts it, “[p]ropositions as mind-independent abstract objects raise serious problems such as their cognitive accessibility and their ability to carry essential truth conditions” (p. 679). Or, as Moulton (2009) puts it, rumors can be mean, but propositions cannot be. This casts doubt on the idea that the embedded clause denotes a proposition.

One of the alternative solutions begins with Kratzer 2006 and is being developed in a growing body of literature, e.g., Moulton 2009, 2015; Kratzer 2013, 2016; Uegaki 2015, 2016; Elliott 2017 (for another approach see Moltmann, 2013a, 2013b). The core concept of this approach is that expressions of propositional attitude do not describe propositions directly but instead linguistic elements with propositional content. This means that the embedded clause is not a (semantic) argument of the matrix verb. This is reflected in a minimal semantics of attitude verbs whereby *believe* denotes a property of events:

- (7)  $\llbracket \text{believe} \rrbracket = \lambda s. \text{believe}(s)$

The content is recovered via a function that takes a proposition and identifies it as the content of a contentful individual, presented in (8).<sup>2</sup>

- (8)  $f_{\text{CONT}}(x) = \{w : w \text{ is compatible with } x\}$  Moulton (2009): ex. 17

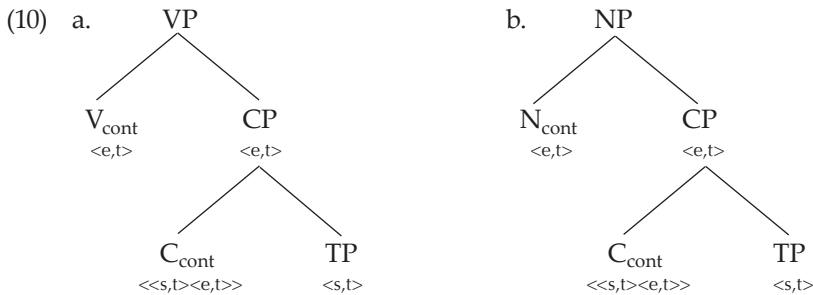
<sup>2</sup> There is no change in the ontology. Things with content, e.g., *idea*, though sometimes subscripted with a  $c$ , are regarded as having the same semantic type as things without content, e.g., *apple*. Only content nouns can be modified by a clause, (i), which identifies their content. Nouns like *apple* do not denote contentful individuals, and content cannot be predicated of them, (ii).

- (i) I like the idea that John is a spy.  
(ii) \*I like the apple that John is a spy.

The function is hosted by the complementizer *that*, (9) (Kratzer, 2006; Moulton, 2009, 2015).<sup>3</sup>

$$(9) \llbracket C_{\text{cont}} \rrbracket = \lambda p. \lambda x_c. f_{\text{cont}}(x_c) = p \quad \text{Moulton (2009): ex. 18}$$

Thus the embedded *that*-clause is a predicate of contents, not a proposition. This allows for analogous and straightforward composition of content clauses both in the nominal domain (content nouns) and in the verbal domain (attitude predicates) via predicate modification. This is shown by Elliott (2017) for attitude predicates, (10a) and (11) (see Kratzer, 2006, 2016, Hacquard, 2006 for different approaches), and by Moulton (2009) for content nouns, (10b) and (12). In (11), under the assumption that eventualities are of the same semantic type as individuals (see Elliott (2016) for detailed reasoning), we get a saying event whose content is “Shirley is upset” (and whose agent is later merged as Abed). In (12), the content noun *story* combines with the CP, again by predicate modification, to give an entity such that it is a rumor and its content is that “Edna is a thief.”



(11) Abed says [CP [CP that Shirley is upset]. Elliott (2017): ex. 83

Derivation:

$$\llbracket \text{say} \rrbracket = \lambda e. \text{saying}(e)$$

$$\llbracket \text{CP} \rrbracket = \lambda x. \text{cont}_{w_0}(x) = \lambda w'. \text{Shirley is upset in } w'$$

$$(\llbracket \text{say} \rrbracket)(\llbracket \text{CP} \rrbracket) = \text{byPM} \lambda e. \text{saying}_{w_0}(e) \& \text{cont}_{w_0}(x) = \lambda w'. \text{upset}(S)(w')$$

Elliott (2017): ex. 85

<sup>3</sup> Approaches differ in what exactly the projection hosting the content function is. For Kratzer (2006) and Moulton (2009, 2015) it is the complementizer *that*, as shown in (9). For others, it is Mood (Kratzer, 2016; Bochnak and Hanink, 2017) (a high Mood above C in Kratzer, 2016 and a low Mood under TP in Bochnak and Hanink, 2017), or a more abstract second CP layer scoping over CP (Elliott, 2017) (similar to cP in de Cuba, 2007, 2017). I believe the data and arguments presented here support the single C layer. As for Mood, see remarks in section 4.

- (12) ‘the story that Edna was a thief’

Derivation:

 $[[\text{story}]] = \lambda x_c.\text{story}(x)$  $[[\text{CP}]] = \text{K } \lambda x.[f_{\text{cont}}(x) = \lambda w^0.\text{thief}(E.)(w^0)]$  $[[\text{story}]]([[\text{CP}]]) = {}_{\text{byPM}}\lambda x.\text{story}(x) \& f_{\text{CONT}}(x) = \lambda w'.\text{thief}(E)(w')$ 

After Moulton (2009): ex. 20

This approach has an advantage over alternatives in that it allows a unified account of attitude-encoding linguistic objects of different syntactic categories—verbs and nouns. However, it raises some questions about the syntactic derivation of *that* clauses. Moulton (2009) provides a number of arguments for regarding the *that* clause modifying nouns as an adjunct and not a complement, which justifies the predicate modification based approach.

### 3.2. The Bulgarian Complementizers

I propose that the Bulgarian complementizers are syntactically and semantically rich and play an active role in semantic composition by encoding intensional content (*če*) and exemplification (*deto*).

#### 3.2.1. *Če* is the Content Complementizer

I analyze Bulgarian *če* as the complementizer in (9), the overt bearer of the content function in (8). This explains the data from both the verbal and nominal domain. It appears under attitude predicates (including emotives) because they relate to content, cf. (13)=(1):

- (13) Ivan misli / kaza [če/\*deto Marija e tuk].  
 Ivan thinks / said če/deto Maria be<sub>PRES.3SG</sub> here  
 ‘Ivan thinks/said that Maria is here.’

*Če* can appear in clauses modifying content nouns, as in (14)=(2), and not other nouns, cf. (15), because these nouns can relate to content.

- (14) idejata, če/\*deto Ivan e tuk  
 idea<sub>DEF</sub> če/deto Ivan is here  
 ‘the idea that Ivan is here’
- (15) \*picata, če Ivan e tuk.  
 pizza<sub>DEF</sub> če Ivan is here  
 [Intended] ‘the pizza that Ivan is here’

At the same time, this view correctly predicts that *če* cannot introduce relative clauses, regardless of whether they are modifying content nouns, (16)=(3), or non-content nouns, (17).

(16) idejata, **kojato/deto/\*če** Ivan zapisa v beležnika  
 idea<sub>DEF</sub> which/deto/če Ivan wrote in notebook<sub>DEF</sub>  
 ‘the idea that Ivan wrote down in the notebook’

(17) picata, **kojato/deto/\*če** Ivan izjade.  
 pizza<sub>DEF</sub> which/deto/če Ivan ate  
 ‘the pizza that Ivan ate’

These facts all support the idea that *če* is the overt realization of the content complementizer in Bulgarian.

### 3.2.2. *Deto* as an Exemplifying Complementizer

Though used in relative clauses, *deto* cannot be empty, as the standard assumption about relativizers goes, because it appears also with emotives, as noted by Krapova (2010):

(18) Ivan sâžaljava [če/**deto** Marija e tuk].  
 Ivan regrets če/deto Maria be<sub>PRES.3SG</sub> here  
 ‘Ivan regrets that Maria is here.’

If *deto* did not have any semantic contribution, the embedded clause would remain a proposition and would not have a way of composing with the verb—or at least not without positing extra, silent material.

Second, *deto* cannot refer to content. We have seen that *deto* under content nouns cannot tap into their content, (19), but can relativize them, (20) (as well as any other noun, (17)).

(19) idejata, **če/\*deto** Ivan e tuk  
 idea<sub>DEF</sub> če/deto Ivan is here  
 ‘the idea that Ivan is here’

(20) idejata, **kojato/deto/\*če** Ivan zapisa v beležnika  
 idea<sub>DEF</sub> which/deto/če Ivan wrote in notebook<sub>DEF</sub>  
 ‘the idea that Ivan wrote down in the notebook’

In addition, as shown, *deto* cannot appear under verbs that clearly relate to content, (13).

- (21) Ivan misli / kaza [če/\***deto** Marija e tuk].  
 Ivan thinks / said če/deto Maria be<sub>PRES.3SG</sub> here  
 ‘Ivan thinks/said that Maria is here.’

So, *deto* cannot be empty, and cannot host the content function.

I propose that *deto* is the overt realization of Kratzer’s (2006) “factive” or exemplifying complementizer (as a reference to Baylis (1948), who coined the term). Exemplification is another important notion in support of dispensing with propositions as arguments, particularly with respect to truth. It rests on the Wittgensteinian idea that facts are not merely true propositions but instead particulars, parts of the world, which exemplify propositions (Kratzer, 2002). As Kratzer puts it, “a situation that (is a fact that) exemplifies a proposition *p* is a situation that does not contain anything that does not contribute to the truth of *p*” (Kratzer, 2012: 166). I propose that *deto* carries the exemplifying function:

- (22)  $\llbracket C_{ex} \rrbracket = \lambda p.\lambda e. \text{exemplifies}(p)(e)$  Kratzer (2006): (14)

The function takes a proposition and returns an eventuality that exemplifies this proposition. To compare the relation of exemplification with that of content, note that the propositional content of a rumor or claim does not contribute to its truth. Facts, on the other hand, are slices of the world that immediately inform truth. This is how factivity is represented by exemplification, doing away with the idea of propositions as directly relating to truth. This explains the distribution of *deto* in emotive factives, (18)=(4).<sup>4</sup> Finally, composing relative clauses by exemplification:

<sup>4</sup> Two notes are due. One, cognitive factive predicates do not take *deto*, or at least not as easily, and are subject to more speaker variation than emotive factives. I leave this matter open, suggesting that it might have to do with the fact that only emotives have been called the class of “true factive” verbs. Two, the interpretation of clauses headed by *če* under factive verbs: in line with observations on other languages, e.g., Roussou 2010, I suggest that the factivity interpretation of the embedded clause is different with *če* than with *deto*. For example, in contexts where there is a discrepancy about the subject’s view of reality and the speaker’s, such as (iii), *deto* is not felicitous, while *če* is still allowed:

- (iii) Ivan has Alzheimer’s. For some time he has been fantasizing that he has a car. Yesterday I visited him and...  
 beše mnogo razstroen, **če/#deto** kolata mu e otkradnata.  
 be<sub>PST.3SG</sub> very upset<sub>SG.MASC</sub> če/deto car<sub>DEF</sub> him be<sub>PRES.3SG</sub> stolen  
 ‘...he was very upset that his car was stolen.’

(23)  $\llbracket \text{the } [_{\text{CP}, \langle e, t \rangle} \text{ pizza } [_{\text{CP}, \langle e, t \rangle} \text{ deto I ate } ] ] \rrbracket = \lambda x. \lambda x: \text{pizza}(x) \& \text{EXEMPLIFY}(\text{eat})(x)$

I assume the copy theory of movement and that the lowest copy is interpreted inside the relative clause (Chomsky, 1993; Fox, 2002). This means that at the TP node there is no trace but rather a saturated proposition (*I ate pizza*) (and also, no need for  $\lambda$ -abstraction). The complementizer takes the proposition and returns a fact that exemplifies it. It combines (again) via predicate modification with the noun *pizza* (under the assumption of no type difference between individuals and eventualities).

At the DP level, we get the unique individual such that it is a pizza and it exemplifies my pizza-eating event. This seems to get the relative right—even if I had eaten other things in addition to the pizza, they do not exemplify my pizza eating. On the other hand, had I eaten more than one pizza (uniqueness condition), the matrix *pizza* would not exemplify my pizza-eating event, as it does not suffice to make it true.

### 3.2.3. Summary

In the four scenarios analyzed here, noun modifying clauses, relative clauses, and embedded clauses to attitude verbs, and the special case of emotives, the embedded CP has retained its semantic type as a predicate and combines via predicate modification. The differences in interpretation between CPs relating to content—be they modifying nouns or verbs—and CPs in relatives and true factives come down to the different complementizers, as schematized in the following trees:

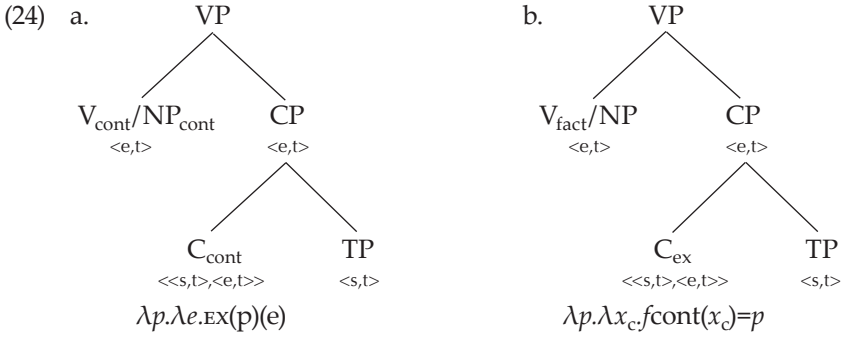
---

In fact, the factivity can sometimes be canceled in English as well, for example under negation:

- (iv) I am not happy that John arrived—in fact, he did not manage to arrive because he missed his plane.

In the light of these examples, I suggest that the factivity interpretation with *če* comes about via an implicature, only arising from the embedding verb, and *če* does not itself contribute anything to the factive interpretation (a desirable approach, since *če* is used also under verbs like ‘believe’, ‘think’, etc.). This retains the single denotation of *če* as a content complementizer while capturing the intuition that the factivity with *če* is somewhat weaker or cancelable, and makes the analysis potentially extendable to similar data and facts in Greek (Roussou 2010).





This tidies up the initially complex landscape of the distribution of Bulgarian complementizers, while in line with and lending further support to current views on propositional attitudes, propositions, and copies.

#### 4. Implications

In this paper, I have made the case for two types of *that* clauses: content CPs and exemplifying CPs. Neither of them is a propositional argument—a theoretically desirable approach—and neither is a (semantic) argument of the matrix verb, again in line with current developments in the literature.

These proposals relate to a number of existing views on the nature of *that* clauses. For example, according to recent developments on CPs as content clauses, e.g., Elliott 2017, all *that* clauses relate to content. Such a view would have to seek an alternative derivation of relative clauses which clearly do not relate to content, and the double nature of emotives.

According to another line of work, a single type of *that* clauses exists, but it is the relative clause. It was first proposed by Kayne (2008) and has much support (with varying technical implementations), e.g., Aboh (2005); Arsenijević (2009); Krapova (2010); Caponigro and Polinsky (2011); Haegeman (2012).

My proposal preserves and captures the differences between the two kinds of noun-modifying clauses—content, *the idea that he is here*, and relative, *the idea that he wrote down*. At the same time, it salvages the valuable insights of the Kaynean hypothesis, such as the idea that all CPs are adjuncts. I have merely implemented this idea from the syntax to the semantics by composing all CPs via predicate modification.

The analysis presented here captures the distribution of *deto* while ruling out the accidental homophony hypothesis whereby *deto* in relative clauses and under emotives would be two separate lexical entries. Cross-linguistic evidence for patterns very similar to those observed in Bulgarian provides reasons to avoid the potential *deto* homophony view. Other languages that exhibit similar behavior to Bulgarian with respect to relative and content clauses in-

clude Modern Greek (Varlokosta 1994; Roussou 1994, 2010; Baunaz 2014, 2016), Swedish and Finnish (de Cuba 2017), Icelandic (Thráinsson 2007), Basque (De Rijk 2008), Durban Zulu (Halpert 2012), potentially Serbo-Croatian (Baunaz 2016 but see Arsenijević 2009 for a different view), Marathi, Asante Jwi, and Meiteilon (Baunaz 2016), Khmer (Comrie and Horie 1995), and Balkan Romani (Matras and Tenser 2016). While the general pattern in these languages seems to be strikingly similar, studying the fine-grained variation among them can inform and advance the theory of content advocated here.

In this paper, I have given complementizers a lot of work to do. They are solely responsible for the different interpretations of the four cases presented here. In view of this, the question arises of how to approach languages that do not show morphological distinctions in complementizers. Such languages are the Romance languages (Kayne 2008, 2014; Baunaz 2014), Thai (Siriwittayakorn 2018), Russian (Baunaz 2016), Gungbe and Saramaccan (Aboh 2006), Persian (Farudi 2007), Adyghe (Caponigro and Polinsky 2011), and Brabant Dutch (Broekhuis and Dekkers 2000, cited in Arsenijević 2009). Some of them are known to encode differences similar to those discussed here for Bulgarian, elsewhere, for example, in Mood for Romance. For this reason I suggest that the distinctions outlined here are represented not via homophonous complementizers but in a different morphosyntactic element in the extended C-spine.

Finally, this paper is a peek at the nature of relatives. It lends support to the copy theory of movement and views that dispense with predicate abstraction in the composition of relatives, e.g., Fox (2002).

## 5. Conclusion

This paper presents a simple, yet powerful composition of clausal embedding. It is simple because it assumes just one semantic type for clauses (while at the same time in accord with the view that there are no propositional arguments and with the type assimilation of clauses and individuals) and just one rule for composing all *that* clauses in natural language: Predicate Modification. It is powerful because it preserves the crucial differences between relatives and content clauses that alternative approaches struggle with.

University of Ottawa  
vesela.simeonova@uottawa.ca

## References

- Aboh, Enoch. (2005) "Deriving relative and factive clauses". *Contributions to the thirtieth incontro di grammatica generativa*. Venezia: Libreria Editrice Cafoscarina, 265–85.
- . (2006) "Complementation in Saramaccan and Gungbe: The case of C-type modal particles". *Natural language & linguistic theory*, 24(1): 1–55.
- Ammann, Andreas and Johan Van Der Auwera. (2004) "Complementizer-headed main clauses for volitional moods in the languages of South-Eastern Europe: A Balkanism?" *Balkan syntax and semantics*: 293–314.
- Arsenijević, Boban. (2009) "Clausal complementation as relativization". *Lingua*, 119(1): 39–50.
- Baunaz, Lena. (2014) "On the various sizes of complementizers". *Probus*, 27(2): 193–236.
- . (2016) "Deconstructing complementizers in Serbo-Croatian, Modern Greek, and Bulgarian". Christopher Hammerly and Brandon Prickett, eds. *NELS 46*. Amherst: University of Massachusetts, Graduate Linguistic Student Association, 69–77.
- Baylis, Charles. (1948) "Facts, propositions, exemplification, and truth". *Mind*, 57(228): 459–79.
- Becker, Martin. (2010) *Mood in Rumanian*. Amsterdam: John Benjamins Publishing Company.
- Bochnak, Ryan and Emily Hanink. (2017) "Factivity in embedded clauses in Washo". Paper presented at 91st annual meeting of the Linguistic Society of America (LSA 91), Austin, TX, 8 January 2017.
- Boye, Kasper and Petar Kehayov. (2016) *Complementizer Semantics in European Languages*. Berlin: de Gruyter Mouton. [*Empirical approaches to language typology*, 57.]
- Broekhuis, Hans and Joost Dekkers. (2000) "The minimalist program and optimality theory: Derivations and evaluations". Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, eds. *Optimality theory: Phonology, syntax, and acquisition*. Oxford: Oxford University Press, 386–422.
- Caponigro, Ivano and Maria Polinsky. (2011) "Relative embeddings: A Circassian puzzle for the syntax/semantics interface". *Natural language & linguistic theory*, 29(1): 71–122.
- Chomsky, Noam (1993) "A minimalist program for linguistic theory". Kenneth Hale and Samuel Jay Keyser, eds. *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press, 1–52.
- Comrie, Bernard and Kaoru Horie. (1995) "Complement clauses versus relative clauses: Some Khmer evidence". Abraham Werner, Talmy Givón, and Sandra A. Thompson, eds. *Discourse grammar and typology: Papers*

- in honor of John W. M. Verhaar*. Amsterdam: John Benjamins Publishing Company, 65–76.
- de Cuba, Carlos. (2007) *On (non) factivity, clausal complementation, and the CP-field*. PhD dissertation, Stony Brook University.
- . (2017) “Noun complement clauses as referential modifiers”. *Glossa*, 2(1).
- De Rijk, Rudolf. (2008) *Standard Basque: A progressive grammar*. Cambridge, MA: MIT Press.
- Dobrovie-Sorin, Carmen. (1989) “Emploi du subjonctif et structure de la phrase en roumain”. *Revue des langues romanes*, 93.
- Elliott, Patrick. (2016) “Severing everything: Substitution failures, and their consequences for the theory of argument structure”. Paper presented at the UCLA syntax round table, 25 April 2016.
- . (2017) *Elements of clausal embedding*. PhD dissertation, University College London.
- Farkas, Donka. (1989) “On the morpho-syntax of subjunctive clauses in Rumanian”. *International journal of Rumanian studies*, 6(2): 7–19.
- Farudi, Annahita. (2007) “An antisymmetric approach to persian clausal complements”. Unpublished ms., University of Massachusetts.
- Fox, Danny. (2002) “Antecedent-contained deletion and the copy theory of movement”. *Linguistic inquiry*, 33(1): 63–96.
- Hacquard, Valentine. (2006) *Aspects of modality*. PhD dissertation, Massachusetts Institute of Technology.
- Haegeman, Liliane. (2012) *Adverbial clauses, main clause phenomena, and the composition of the left periphery: The cartography of syntactic structures*. Oxford: Oxford University Press.
- Halpert, Claire. (2012) *Argument licensing and agreement in Zulu*. PhD dissertation, Massachusetts Institute of Technology.
- Hintikka, Jaakko. (1969) “Semantics for propositional attitudes”. *Models for modalities: Selected essays*. Dordrecht: D. Reidel, 87–111.
- Isac, Dana and Edith Jakab. (2004) “Mood and force features in the languages of the Balkans”. Olga Mišeska Tomić, ed. *Balkan syntax and semantics*. Amsterdam: John Benjamins, 315–38.
- Joseph, Brian D. (1985) “Complementizers, particles, and finiteness in Greek and the Balkans”. *Folia Slavica*, 7(3): 390–411.
- . (2016) “The semantics and syntax of complementation markers as an areal phenomenon in the Balkans, with special attention to Albanian”. Kasper Boye and Petar Kehayov, eds. *Complementizer Semantics in European Languages*. Berlin: de Gruyter Mouton, 265–92. [Empirical approaches to language typology, 57.]
- Kayne, Richard. (2008) “Antisymmetry and the lexicon”. *Linguistic variation yearbook*, 8: 1–31.

- . (2014) "Why isn't *this* a complementizer?" Peter Svenonius, ed. *Functional structure from top to toe: The cartography of syntactic structures*. Oxford: Oxford University Press, 188–231.
- Kiparsky, Paul and Carol Kiparsky. (1970) "Fact". Manfred Bierwisch and Karl Heidolph, eds. *Progress in linguistics: A collection of papers*. The Hague: Mouton, 143–73.
- Krapova, Iliyana. (2010) "Bulgarian relative and factive clauses with an invariant complementizer". *Lingua*, 120(5): 1240–72.
- Krapova, Iliyana and Tsena Karastaneva. (2000) "Complementizer positions in Bulgarian". *Bulgarian studies at the dawn of the 21st century: A Bulgarian-American perspective*. Sofia: Gutenberg, 93–104.
- Kratzer, Angelika. (2002) "Facts: Particulars or information units?" *Linguistics and philosophy*, 25(5): 655–70.
- . (2006) "Decomposing attitude verbs". *Handout on semantics archive*.
- . (2012) *Modals and conditionals: New and revised perspectives*. Oxford: Oxford University Press.
- . (2013) "Modality for the 21st century". Stephen R. Anderson, Jaques Moeschler, and Fabienne Reboul, eds. *The Language Cognition Interface / L'Interface Language Cognition*. Geneva-Paris: Librairie Droz, 179–99.
- . (2016) "Evidential moods in attitude and speech reports". *Slides presented at UConn Colloquium*, 9 September 2016.
- Lewis, David. (1988a). "Relevant implication". *Theoria*, 54(3): 161–74.
- . (1988b). "Statements partly about observation". *Philosophical papers*, 17(1): 1–31.
- . (2003) "Things qua truthmakers". Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra, eds. *Real metaphysics: Essays in honour of DH Mellor*. London–New York: Routledge, 25–38.
- Matras, Yaron and Aron Tenser. (2016) "Complementizers in Romani". Kasper Boye and Petar Kehayov, eds. *Complementizer Semantics in European Languages*, Berlin: de Gruyter Mouton, 341–75. [*Empirical approaches to language typology*, 57.]
- Moltmann, Friederike. (2013a). *Abstract objects and the semantics of natural language*. Oxford: Oxford University Press.
- . (2013b). "Propositions, attitudinal objects, and the distinction between actions and products". *Canadian journal of philosophy*, 43(5–6): 679–701.
- Moulton, Keir (2009). *Natural selection and the syntax of clausal complementation*. PhD dissertation, University of Massachusetts.
- . (2015) "CPs: Copies and compositionality". *Linguistic inquiry*, 46(2): 305–42.
- Rawlins, Kyle. (2013) "About 'about' ". *Semantics and linguistic theory*, 23: 336–57.
- Rivero, Maria Luisa. (1994) "Clause structure and V-movement in the languages of the Balkans". *Natural language & linguistic theory*, 12(1): 63–120.

- Roussou, Anna. (1994) *The syntax of complementisers*. PhD dissertation, University of London.
- . (2010) "Selecting complementizers". *Lingua*, 120(3): 582–603.
- Rudin, Catherine. (1982) *Aspects of Bulgarian syntax*. Phd thesis, Indiana University.
- . (1988) "Finiteness and opacity: Evidence from the Balkans". Michael Hammond, Edith A. Moravcsik, and Jessica Wirth, eds. *Studies in syntactic typology* 17. Amsterdam: John Benjamins Publishing Company, 37–52.
- . (2013) *Aspects of Bulgarian syntax: Complementizers and WH constructions*. Columbus, OH: Slavica Publishers.
- Siriwittayakorn, Teeranoot. (2018) "The processing of relative clause attachment as a tool for resolving a problem in typology of relative clauses: Preliminary evidence from Thai data". *Journal of psycholinguistic research* 47(1): 179–193.
- Thráinsson, Höskuldur. (2007) *The syntax of Icelandic*. Cambridge: Cambridge University Press. [Cambridge Syntax Guides.]
- Tomić, Olga Mišeska. (1996) "The Balkan Slavic clausal clitics". *Natural language & linguistic theory*, 14(4): 811–72.
- Uegaki, Wataru. (2015) *Interpreting questions under attitudes*. PhD dissertation, Massachusetts Institute of Technology.
- . (2016) "Content nouns and the semantics of question-embedding". *Journal of semantics*, 33(4): 623–60.
- Varlokosta, Spyridoula. (1994) "The properties of *pu*-complements in Modern Greek". Irene Philippaki-Warbuton, Katerina Nicolaidis, and Maria Sifianou, eds. *Themes in Greek linguistics: Papers from the First International Conference on Greek Linguistics*. Amsterdam: John Benjamins Publishing Company, 61–8.

# On the NP/DP Frontier: Bulgarian as a Transitional Case\*

Mila Tasseva-Kurktchieva and Stanley Dubinsky

*Abstract:* This paper will show that Bulgarian sits on the edge of Bošković's (2012) classification of languages into "DP languages" and "NP languages." Where, on Bošković's view, languages can be clearly categorized into DP or NP languages, Bulgarian appears to project a DP, but not always. When it does so it behaves (mostly) like a DP language. And, when it doesn't, it appears to be an NP language. Bulgarian thus falls categorically into neither of the two classes, and its properties indicate yet another important parameter along which languages may vary—there are strong DP languages (English), non-DP languages (Chinese), and weak DP languages (Bulgarian).<sup>1</sup>

## 1. Introduction

Bošković's (2012) classification of DP/NP languages relies on 18 diagnostics applied to a wide range of languages. Here, we limit our discussion to those diagnostics that are applicable to Bulgarian. In sections 2, 3, and 4, we show that some indicate a DP-status for the language, some an NP-status, and some depend on the presence of the definite determiner. In section 5, we review the structure of DP in Bulgarian, and show how the structure of Bulgarian DPs leads to the results described in sections 2–4.

Our goal in this paper is not to argue either for or against a concrete analysis of nominal structure in Bulgarian, but only to examine Bulgarian through the lens of Bošković's diagnostics to determine where it is situated in the NP/DP dichotomy. Bošković himself implicitly classifies Bulgarian as a DP language on the basis of its having a definite article. Others, including Dimitrova-Vulchanova (2000), Dimitrova-Vulchanova and Giusti (1998, 1999), Arnaudova (1996) have done so explicitly. We note, however, that Bulgarian does not have the equivalent of an indefinite determiner, the closest approximation

---

\* Catherine Rudin's seminal work on *wh*-movement in Bulgarian has influenced a great number of syntacticians working on the language, including ourselves. We are truly grateful to her for her contributions to the field.

<sup>1</sup> Bulgarian may not be alone in this regard. See Pereltsvaig 2013 and 2015 for arguments that Russian is a sometimes-NP sometimes-DP language, as well as Von Hofwegen for an analysis of Lithuanian along these lines.

being *edin*, the numeral ‘one’.<sup>2</sup> However, as Dimitrova-Vulchanova and Mišeska-Tomić (2009) observe, while *edin* can superficially function as an indefinite article, it is more accurately analyzed as a numeral adjectival modifier. Thus, the presence of a definite article in Bulgarian, coupled with the absence of anything analogous to the indefinite article leads to the plausible conjecture (which we will seek to show) that definite nominal phrases are DPs and that indefinite nominal phrases are not.<sup>3</sup>

We will see that Bulgarian looks to be a well-behaved DP language with respect to clitic doubling and obligatory number morphology (section 2). However, when one looks at Bulgarian in light of Neg raising and subject expletives, it behaves like an NP language (section 3). Of greater interest, in section 4, is the behavior of Bulgarian with respect to majority superlative readings, exhaustivity presupposition for possessives, and extraction out of NP subject islands. There, we find that the language appears to be a DP language when the nominal phrases in question have a definite determiner but not otherwise.

## 2. DP Language Properties in Bulgarian

This section examines the behavior of Bulgarian with respect to clitic doubling and obligatory number morphology. Here, the behavior of Bulgarian suggests that it might be a DP language.

Bošković (2012) shows clitic doubling to be a property of DP languages. In this regard, Bulgarian allows clitics to co-occur with (in)direct objects and possessors, and requires them with experiencer dative subjects. Thus, in (1a) Bošković 2012: 28, the accusative object clitic *go* optionally co-occurs with the object *pismo*, and in (1b) the dative clitic *mu* obligatorily co-occurs with the experiencer dative subject *na studenta*.<sup>4</sup>

Bošković (2012) claims that if a language allows (some) nouns to be interpreted as plural in the absence of plural morphology, then that language must

<sup>2</sup> For a discussion of similar phenomena regarding the numeral ‘one’ (i.e., *exad*) in Hebrew, see Borer 2004, chapter 7.

<sup>3</sup> For a different view, see Schick (2000) and Geist (2013), who argue that the numeral ‘one’ functions as either a specific or nonspecific indefinite determiner.

<sup>4</sup> The phenomenon of clitic doubling with indirect objects and possessives is discussed in Rudin 1997, Dimitrova-Vulchanova 1995, Dimitrova-Vulchanova and Hellan 1999, and Schick 2000. Furthermore, Schick (2000) shows that specificity is a necessary condition for clitic doubling, such that if the nominals in (1a) and (1c) are indefinite/non-specific, then clitic doubling is not allowed, as we see in (i) and (ii).

(i) Ivan (\*go) napisa pismo.  
Ivan it wrote letter

(ii) Na student (\*mu) e studeno.  
to student him is cold



be an NP language. For instance, in an NP language such as Japanese, *hon* 'book' may mean (depending on context) 'a book' or 'books'. Bulgarian does not appear to be an NP language by this criterion, in that it requires plural morphology on plural nouns, as in example (2).

- (1) a. Ivan (go) napisa pismo-to.  
Ivan it wrote letter<sub>DEF</sub>  
'Ivan wrote the letter.'
- b. Na student-a \*(mu) e studeno.  
to student<sub>DEF</sub> him is cold  
'The student is cold.'
- (2) a. Marija kupi kniga  
Maria bought book/\*books  
'Maria bought a book.'
- b. Marija kupi knigi  
Maria bought books/\*book  
'Maria bought books.'

### 3. NP Language Properties in Bulgarian

This section explores ways in which Bulgarian fails two DP-language tests. First, we show that Bulgarian prohibits negative raising. Next, we show that Bulgarian fails to have expletive subjects, which is predicted if NP languages do not project TP, as Bošković claims.

Bošković (2012: 4–5) suggests that "languages without articles disallow Negative Raising (NR), and languages with articles allow it." NR is best exemplified by the licensing of strict clause-mate NPIs. Thus, the English NPI *in a long time* must have a licenser in the same minimal clause, as in (3). In (3a), we see that it cannot occur without clausal negation (or some other clause-mate NPI licenser). Evidence that English allows NR is seen in (3b), where the NPI is licensed by negation in the higher clause. Given that *in a long time* requires strict clause-mate licensing, this is evidence that *not* has raised out of the complement clause.<sup>5</sup>

In Bulgarian, evidence for NR is somewhat harder to come by. This is, in part, because many NPI phrases are homophonous with their non-NPI coun-

<sup>5</sup> Of course, the grammaticality of sentences such as (3b) depends on the matrix verb's allowing NR. Some verbs, such as *claim*, do not allow it, and (i) is ungrammatical for this reason.

(i) \*John does not claim [that Mary has stayed/come here [NPI in a long time]].

terparts. For example, *dâlgo vreme* can mean ‘in a long time’ (NPI) or ‘for a long time’ (non-NPI), depending on context. One non-homophonous NPI/non-NPI pair is *nikakâv* ‘no kind’ and *njakakâv* ‘some kind’. Embedding the NPI in a complement clause provides the paradigm needed to test whether Bulgarian allows Neg Raising. The NPI, shown in (4a) in its plural form *nikakvi*, requires a Neg licensor within its clause, and absence of negation renders the sentence ill-formed. If Bulgarian were a Neg Raising language like English, then we should expect that negation will license the NPI from the main clause as it does in the English example (3b). The ungrammaticality of (4b), then, is evidence that Bulgarian does not allow Neg Raising, and in this regard behaves like an NP language.

- (3) a. Mary has \*(not) stayed/come here [<sub>NPI</sub> in a long time].  
 b. John does \*(not) believe [that Mary has stayed/come here [<sub>NPI</sub> in a long time]]
- (4) a. Ivan vjarva, če Marija \*(ne) e poxarčila nikakvi pari.  
 Ivan believes that Maria (not) is spent none money  
 ‘Ivan believes that Maria has not spent any money.’  
 b. \*Ivan ne vjarva, če Marija e poxarčila nikakvi pari.  
 Ivan not believes that Maria is spent none money  
 ‘Ivan does not believe that Maria has spent any money.’

Extending his NP-DP dichotomy to the sentential level, Bošković suggests that NP languages do not have a TP either. One of the consequences of a reduced functional layer is that those languages would fail to have expletive subjects.<sup>6</sup> Conversely, the TP projection in a DP language should require expletive subjects to be used when no other nominal satisfies the EPP. This is certainly the case for English, as in (5), where the sentences are ungrammatical without their expletive subjects.

Following this generalization, one would expect to find expletive subjects in Bulgarian if it were classified as a DP language. However, an examination of comparable data shows this not to be the case. As seen in (6), sentences lacking an overt lexical subject do not need an expletive to fill the subject position. In fact, the insertion of an expletive renders these sentences ungrammatical. From these data we can see that Bulgarian clearly does not behave as a DP language with respect to expletive subjects.

<sup>6</sup> It is noted by a reviewer of this chapter that Italian doesn’t have expletive subjects either, but is understood to be a DP-language.

- (5) a. \*(It) has been raining all day.  
 b. \*(There) is water in the carafe.
- (6) a. (\*To) vali cjal den.  
 it raining all day  
 'It has been raining all day.'  
 b. (\*Tam) ima voda v kanata  
 there is water in carafe  
 'There is water in the carafe.'

#### 4. DP-Dependent Properties in Bulgarian

This section shows that the DP-ness of Bulgarian nominals depends on the presence of the definite determiner. Bulgarian licenses Majority superlative readings (4.1) and the Exhaustivity presupposition for possessives (4.2) only when the NP has a definite determiner. Conversely, extraction out of NP subject islands (4.3) is only possible in the absence of the definite determiner.

##### 4.1. Majority Superlative Readings

Bošković (2012: 7) makes the claim that “only languages with articles allow the ... majority superlative reading” (MSR), as shown in Slovenian example (7). Here, *največ ljudi* ‘most people’ is not interpreted as ‘more than half the people’ but rather as ‘the greatest number of people’ irrespective of the predicate. Thus if ‘the greatest number of people’ drink beer, then it does not follow necessarily that most of them do.<sup>7</sup> Bošković contrasts Slovenian with German,

<sup>7</sup> Note that there is a possibility, based on comparisons with other Slavic languages, that *največ* is actually an adverb meaning ‘mostly’, yielding an interpretation for (7) of ‘Mostly, people drink beer’. In Bulgarian the cognate lexical item functions clearly as an adverbial and appears not to be relevant as a diagnostic for the phenomena in question here.

- (i) Xorata/toj pijat/pie bira naj-veče.  
 people/he drink/drinks beer mostly  
 ‘People/he drink/drinks beer, mostly.’
- (ii) Povečeto ot tjax pijat bira.  
 more<sub>DEF</sub> of them drink beer  
 ‘Most of them drink beer.’

As (i) shows, *naj-veče* does not collocate with the noun that one might presume it to quantify. In (ii), we see that in order to quantify the pronominal subject ‘they’, an entirely different lexical item and construction is needed. We would thus question

a purportedly DP-language with articles, in which the same sentence (8) allows for both interpretations.

- (7) Največ ljudi pije pivo. (Slovenian, NP language)  
 most people drink beer  
 \*'More than half the people drink beer.' (Majority reading, MSR)  
<sup>v</sup>'More people drink beer than drink any other beverage.'  
 (Plurality reading, PR)
- (8) Die meisten Leute trinken Bier. (German, DP language)  
 the most people drink beer.  
<sup>v</sup>'More than half the people drink beer.' (MSR)  
<sup>v</sup>'More people drink beer than drink any other beverage.' (PR)

Rather than finding the availability of the MSR to be language-dependent, we have found, at least where English and Bulgarian are concerned, that the MSR is restricted to contexts in which a definite determiner fails to have scope over the relevant quantifier. The discussion here will show that DP is always present in English, but that a quantifier can be either outside or inside DP. This leads to two possible outcomes (MSR or PR), depending on where DP is projected. We will show, further, that the morphosyntax of the Bulgarian definite determiner forces a MSR when present, and only allows a PR when DP is not projected at all.

In English, the presence of a determiner and its position relative to the quantifier gives rise to different (majority and plurality) readings. Thus, when *most* is not dominated by *the*, as in (9a), it yields a majority reading (MSR), and when *the* does dominate *most*, as in (9b), it has a plurality reading (PR).

- (9) a. Ivan ate **most of the candies**.  
 = Ivan ate > 50% of the candies. (MSR)
- b. Ivan ate **the most candies**.  
 = Ivan ate more candies than anyone else. (PR)
- c. Ivan ate **most candies**.  
 = Ivan ate > 50% of the types of candy. (MSR)

We might illustrate the contrast in the following way. Example (9a) is true if Ivan ate more than half the candies. Here then, he ate most of the candies. If Andrea ate three-fourths of the candies, then Andrea also ate most of the

---

whether the Slovenian example above provides evidence one way or the other as to the availability of the majority superlative reading with a subject NP.

candies. However, it is not the case that Ivan ate the most candies. Andrea did. Example (9b) is true if Ivan ate 40 out of 100 candies, Andrea ate 30 out of 100 candies, and Mark ate 20 out of a 100 candies. Here, Ivan ate the most candies but he didn't eat most of the candies.<sup>8</sup> In addition to the MSR/PR contrast, we find differences in regard to a type/token distinction. The type reading is available when *most* does not dominate *the*. In (9a), *most* has scope over *the*, and the sentence has a preferred token majority reading (e.g., Ivan could have eaten more than half of a bowl of Hershey's Kisses). If the verb were changed, e.g., from *ate* to *tried*, the preferred reading might be a type majority reading. In (9b), *the* has scope over *most* and has a plurality reading, either type or token (e.g., Ivan could have eaten the greatest number of individual candies or types of candy). Example (9c) has a preferred type majority reading (e.g., Ivan could have eaten more than half of the types of candy available). In our discussion of DP/NP properties, type/token distinctions do not play a critical role.

Bulgarian presents a different set of facts, with two quantifiers involved in majority and plurality readings. Because *naj-veče* functions as an adverb and not a true quantifier in Bulgarian (and possibly so in Serbo-Croatian as well), its distribution appears not to be relevant at all for a test of MSR (see note 3, above). Our analysis of MSR in Bulgarian is thus based upon the distribution of the comparative *poveče* 'more' and the superlative *naj-mnogo* 'most' (lit: 'est-many'). As (10a) and (10b) illustrate, *poveče* is compatible in a comparative context ('more candies than Maria') but not in a superlative context (#'more candies among the students'). In contrast, *naj-mnogo* is not felicitous in a comparative context (#'most candies than Maria), but is fine as a superlative ('most candies among the students'). Note that *naj-mnogo* in (10b) only carries a plurality reading. That is, it means 'He ate more candies than anyone else' [PR], and not 'He ate > 50% of the candies' [MSR].

Interesting things happen when these two quantifiers interact with the definite determiner *-to*, as in (11). Here, we find that *povečeto* is no longer a comparative and induces a Majority Superlative Reading. In this context, *naj-mnogoto* is marginal. The relevant contrast is thus between (10b), where

<sup>8</sup> Complicating the picture for English, and outside the scope of this discussion, is the fact that the collocation *the most N* is ungrammatical in subject position (unlike the case with German). Thus, while (i) is grammatical, with the expected majority reading, (ii) is not.

- (i) Most of the people know Ivan.
- (ii) \*The most people know Ivan.

However, we know that this to be a syntactic fact rather than a semantic one, since it is the case that if (i) and (ii) are passivized, as in (iii) and (iv), then the contrast observed in (9) obtains.

- (iii) Ivan is known by most of the people.
- (iv) Ivan is known by the most people.

*naj-mnogo* means ‘the most’ (PR), and (11a), where *povečeto* means ‘most of’ (MSR). Evidence that *povečeto* is not a comparative is seen when it is placed in a comparative context. Compare (11a) and (11c).

- (10) a. Izjade **poveče** / \***naj-mnogo** bonboni ot Marija.  
ate<sub>3SG</sub> more SUPERL-many candies than Maria  
‘He ate more/\*most candies than Maria.’
- b. Izjade \***poveče** / **naj-mnogo** bonboni izmeždu učenicite.  
ate<sub>3SG</sub> more SUPERL-many candies among students<sub>DEF</sub>  
‘He ate \*more/(the) most candies among the students.’ (PR)
- (11) a. Izjade **povečeto** bonboni.  
ate<sub>3SG</sub> most<sub>DEF</sub> candies  
‘He ate most of the candies.’ (MSR)
- b. ?Izjade **naj-mnogoto** bonboni.  
ate<sub>3SG</sub> SUPERL-many<sub>DEF</sub> candies
- c. \*Izjade **povečeto** bonboni **ot Marija**  
ate<sub>3SG</sub> most<sub>DEF</sub> candies than Maria

We can see how these two readings are stable and contrast in (12). In (12a), he not only ate a greater number of candies than anyone else (PR), but also ate more than 50% of them (MSR). In (12b), anomalously, he not only ate more than 50% of the candies (MSR), but also ate a greater number than anyone else (PR). The infelicity of (12b) arises from the fact that if he ate *povečeto* ‘more than half’, then he must certainly also have eaten *naj-mnogo* ‘a greater number than anyone else’. Here, the second clause is irrelevant and entailed by the first.

- (12) a. Izmeždu učenicite, ne samo će izjade **naj-mnogo**  
among students, not only that ate<sub>3SG</sub> SUPERL-many  
bonboni, no izjade **povečeto** bonboni.  
candies, but ate<sub>3SG</sub> most<sub>DEF</sub> candies  
‘Among the students, not only did he eat the most candies, but he ate most **of** the candies.’

- (12) b. <sup>#</sup>Izmeđuđu ućenicite, ne samo će izjate **povećeto**  
 among students, not only that ate<sub>3SG</sub> more<sub>DEF</sub>  
 bonboni, no izjate **naj-mnogo** bonboni.  
 candies, but ate<sub>3SG</sub> SUPERL-many candies
- <sup>#</sup>Among the students, not only did he eat most of the candies, but he ate the most candies.

To summarize the distribution of MSR and PR in English and Bulgarian, we can make the following observations. For English, MSR obtains when *most* is not dominated by *the*, as in (9a) above, and PR occurs when *the* does dominate *most*, as in (9b). For Bulgarian, MSR is available only in the presence of the determiner *-to*, and PR normally obtains when *-to* is not present. The observations above are formalized here with bracketed representations (13) and (14).

(13) **English**

- a. [DP **the** [QP most [NP candies]]] = PR  
 b. [QP most of [DP **the** [NP candies]]] = MSR

(14) **Bulgarian**

- a. [QP naj-mnogo [NP bonboni]] = PR  
 b. [DP poveće<sub>1</sub>-to [QP poveće<sub>1</sub> [NP bonboni]]] = MSR

We might generalize across these seemingly disparate patterns by observing that MSR obtains only in the presence of DP, but is blocked when D commands all positions of the chain projected by Q, whether spelled out or not. In English, DP is always projected, and MSR becomes available when QP is merged above DP, as we see in (13b). Here, the quantifier *most* might be generated above DP or moved there, but in either case its spelled-out position is above *the*. In Bulgarian, DP is projected optionally, and MSR is unavailable when it is not present, as in (14a). When DP is projected, *poveće* is generated within the scope of *-to*, but undergoes head movement, such that its spelled-out position is no longer commanded by *-to*.

#### 4.2. Exhaustivity Presupposition in Possessives

Partee (2006) observes that English possessor-numeral-noun expressions, as in (15), presuppose that the possessor owns exactly (exhaustively) the number of items referenced in the noun phrase. In NP languages, e.g., Chinese (16), possessives do not induce exhaustivity (i.e., Zhangsán can have more than three sweaters). Using this contrast as a basis, and comparing several

languages, Bošković (2012) concludes that “possessors may induce an exhaustivity presupposition only in DP languages.”

- (15) Zhangsan’s three sweaters. (English, DP language)  
[Bošković 2012: (52)]
- (16) Zhangsan de [san jian maoxianyi] (Chinese, NP language)  
Zhangsan <sub>DEPOSS</sub> three CL sweater [Bošković 2012: (53)]  
‘Zhangsan’s three sweaters’

This generalization, while somewhat insightful, does not adequately describe the conditions under which exhaustivity readings are available, at least not in English. Exhaustivity presuppositions are available in English, a DP language, only when a possessor/agent or a definite article has scope over the quantifier, as in (17a). These same conditions result in a PR (plurality reading) in section 4.2, above. If the noun phrase has no definite determiner or the possessive does not have scope over the quantifier, as in (17b), then the exhaustive reading is not available.

- (17) a. **Susan’s/the** three portraits are in the gallery.  
(Exhaustive = only three portraits)
- b. Three portraits (of **Susan’s/Susan**) are in the gallery.  
(Non-exhaustive = at least 3 portraits)

Turning to exhaustivity in Bulgarian, the appearance of the definite article (correlating with the optional projection of DP) licenses the exhaustive presupposition. In other words, the exhaustivity presupposition is available only in the presence of the definite article and irrespective of other indicators of definiteness/uniqueness. Consider the paradigm in (18–20). The first two examples have a pre-nominal possessive NP, before a quantifier in (18) and following the quantifier in (19).<sup>9,10</sup> Example (20) has a post-nominal possessive

<sup>9</sup> (18) is restricted in discourse to cases in which the possessive is contrastively focused. For some speakers, even a contrastively focused possessive cannot precede an indefinite quantifier, as seen in (18b).

<sup>10</sup> Sentences parallel to (19), with a possessive clitic, also trigger an exhaustive interpretation. However, this clitic can only co-occur with a definite article, as in (i); it does not play any role in the contrasts we are examining here.

- (i) Trite mu portreta sa v galerijata. (Exhaustivity available)  
the.three his portraits are in gallery<sub>DEF</sub>
- (ii) \*Tri mu portreta sa v galerijata.  
three his portraits are in gallery<sub>DEF</sub>



PP. In each of these examples, the possessive NP and the possessive PP may be interpreted as possessive, agent, or theme, but the exhaustive presupposition is only available in the presence of the determiner *-te*.

- (18) a. **Ivanovite** tri portreta sa v galerijata.  
 (Exhaustivity available)  
 Ivan<sub>POSS,DEF</sub> three portraits are in gallery<sub>DEF</sub>
- b. ?**Ivanovi** tri portreta sa v galerijata (Only non-exhaustive)  
 Ivan<sub>POSS</sub> three portraits are in gallery<sub>DEF</sub>
- (19) a. **Trite Ivanovi** portreta sa v galerijata.  
 (Exhaustivity available)  
 three<sub>DEF</sub> Ivan<sub>POSS</sub> portraits are in gallery<sub>DEF</sub>
- b. **Tri Ivanovi** portreta sa v galerijata. (Only non-exhaustive)  
 three Ivan<sub>POSS</sub> portraits are in gallery<sub>DEF</sub>
- (20) a. **Trite portreta na Ivan** sa v galerijata.  
 (Exhaustivity available)  
 three<sub>DEF</sub> portraits of Ivan are in gallery<sub>DEF</sub>.
- b. **Tri portreta na Ivan** sa v galerijata  
 (Only non-exhaustive)  
 three portraits of Ivan are in gallery<sub>DEF</sub>.

The differences, then, between Bulgarian and English are summarized in examples (21–25), below.

- (21) a. [DP **Susan's** [QP three [NP portraits]]]  
 (Exhaustive entailed)
- b. [DP [QP three [NP portraits [PP of Susan's]]]]
- (22) a. [DP **the** [QP three [NP portraits [PP of/by Susan]]]]  
 (Exhaustive entailed)
- b. [DP [QP three [NP portraits [PP of/by Susan]]]]
- (23) a. [DP tri<sub>1</sub>-**te** [QP *t*<sub>1</sub> [NP Ivanovi portreta]]]  
 (Exhaustive allowed)
- b. [QP tri [NP Ivanovi portreta]]

- (24) a. [<sub>FocP</sub> [[Ivanovi<sub>i</sub>] **te**<sub>2</sub>] [<sub>DP</sub> *t*<sub>2</sub>] [<sub>QP</sub> tri [<sub>NP</sub> *t*<sub>1</sub> portreta]]]]  
 (Exhaustive allowed)  
 b. [<sub>FocP</sub> Ivanovi] [<sub>QP</sub> tri [<sub>NP</sub> *t*<sub>1</sub> portreta]]]
- (25) a. [<sub>DP</sub> tri<sub>1</sub>-**te**] [<sub>QP</sub> *t*<sub>1</sub>] [<sub>NP</sub> portreta [<sub>PP</sub> na Ivan]]]  
 (Exhaustive allowed)  
 b. [<sub>QP</sub> tri] [<sub>NP</sub> portreta [<sub>PP</sub> na Ivan]]]

In English: iff *the* or a genitive argument has scope over QP, exhaustivity is triggered. In Bulgarian: iff DP is projected, exhaustivity is licensed. Thus, it is not exactly the case that possessors may induce an exhaustivity presupposition only in DP languages. Rather, the DP projection can license or trigger exhaustivity, with languages being divergent on the conditions under which this obtains.<sup>11</sup>

### 4.3. Extraction from NP Subjects

Another phenomenon that shows the role of DP projections in Bulgarian, independent of semantic factors is extraction out of subject position. In English there is a robust contrast in the acceptability of extraction from subject and object positions, with extraction from object being largely permitted and extraction from subject being uniformly prohibited. Example (26) illustrates this. The acceptability of extraction out of object NPs depends only on definiteness, with indefinite NPs allowing it and definite NPs not (26a). Extraction out of subject NPs, in contrast, is ungrammatical regardless of definiteness, as shown in (26b) [Davies and Dubinsky 2003: 24].<sup>12</sup>

- (26) a. Who<sub>1</sub> did Sharon read [a/every/\*that/\*my recent book about *t*<sub>1</sub>]?  
 b. \*Who<sub>1</sub> did [a/every/that/my recent book about *t*<sub>1</sub>] bore Herman to tears?

<sup>11</sup> Related, but not quite parallel, facts are reported for Russian in Pereltsvaig (2013: 207–08). There it is shown (in the presence of cardinal and ordinal number modifiers) that interpretations are dependent on the relative scope of the number elements. In Bulgarian, however, at least with respect to quantifiers and agent/possessors, the relative order of elements is not implicated in the exhaustive reading, since it is only available in the presence of the determiner, regardless of word order.

<sup>12</sup> It is observed that (i) is perfectly grammatical.

(i) Who did Sharon write her latest book about?

However, while this does appear to counterexemplify the claim made above, it is shown in Davies and Dubinsky 2003 that extraction in this sentence is licensed through abstract incorporation of the head of the object of *write* to the verb.



not block extraction. Thus example (32) shows acceptable extraction out of an infinitival subject, contrasting with example (30b).

- (31) a. [Za kogo] četoxa [njakolko knigi za kogo]?  
 about whom read<sub>3PL</sub> several books  
 ‘About whom did they read several books?’
- b. [za kogo] [njakolko knigi za kogo] ne sa v  
 about whom several books not are in  
 bibliotekata?<sup>13</sup>  
 library<sub>DEF</sub>  
 ‘(About whom were several books not in the library?)’
- (32) [na kakvo] misliš [če [da otide na kakvo] beše  
 to what think<sub>2SG</sub> that to go was  
 važno za nego].  
 important for him  
 ‘To what do you think that to go was important for him?’

Bulgarian does however restrict extraction out of nominals in the presence of an overt DP phrase (projected by *-te*). Note that this restriction is not necessarily connected with semantic definiteness, since extraction out of semantically definite (e.g., possessed) nominals is indeed allowed when DP is not projected and prohibited when it is. And this same effect is found both in object and subject positions. In (33a) we see that extraction out of objects is possible with an indefinite quantifier or a possessor. Similarly, in (34a), extraction out of an indefinitely quantified or possessed nominal is possible. In each case, though, extraction becomes unacceptable when the definite determiner is added, as in (33b) and (34b), showing that the DP projection, on its own, blocks extraction.<sup>14</sup>

<sup>13</sup> (31b) requires contrastive focus to be perfectly acceptable, but such focus does not rescue (26b).

<sup>14</sup> Acceptability judgments for these sentences were collected from two groups of Bulgarian native speakers, one group that resides in Bulgaria and one group that has lived outside Bulgaria for at least 10 years. It was found that the acceptability judgments of nonresident Bulgarian speakers were in general more inconsistent. Judgments collected from resident Bulgarian speakers in contrast were quite clear and strongly supported the claims made here. It is further the case that some of the examples, such as (33a) and (34a) with *negovi* are improved if the left-dislocated PP is contrastively focused. The ungrammatical examples, (33b) and (34b), are unaffected by changes in intonation.

- (33) a. Za koja vojna četoxa njakolko/negovi knigi?  
 about which war read<sub>3PL</sub> some/his books  
 ‘About which war did they read several/\*his books?’
- b. \*Za koja vojna četoxa njakolkoto/negovite knigi?  
 about which war read<sub>3PL</sub> some<sub>DEF</sub>/his<sub>DEF</sub> books  
 (‘About which war did they read the-some/the-his books?’)
- (34) a. Za Cezar li njakolko/negovi knigi pomognaxa na  
 about Caesar Q some/his books helped to  
 studentite da naučat istorija?  
 students<sub>DEF</sub> to learn history  
 (‘Is it about Caesar that some/his books helped the students learn history?’)
- b. \*Za Cezar li njakolkoto/negovite knigi pomognaxa na  
 about Caesar Q some<sub>DEF</sub>/his<sub>DEF</sub> books helped to  
 studentite da naučat istorija?  
 students<sub>DEF</sub> to learn history  
 (‘Is it about Caesar that some-the/his-the books helped the students learn history?’)

## 5. Summary and Analysis

While Bošković (2012) sought to categorize languages according to a projects DP/doesn't project DP parameter, it would appear that such a categorization oversimplifies things greatly. Bulgarian, for its part, belongs to the category neither, as we have shown in substantial detail above.<sup>15</sup> Taking relevant diagnostics from Bošković 2012, we showed in Sections 2 and 3 that Bulgarian exhibits both DP language properties (Clitic doubling and Obligatory number morphology) as well as NP language properties (Neg raising and Subject expletives). Section 4 showed three ways in which the behavior of Bulgarian nominals is determined by the presence or absence of the definite article.

<sup>15</sup> The claim that we've made here, namely, that Bulgarian is a sort of weak DP language, is in line with claims made in Despić 2015, in which it is shown that the cross-linguistic distribution of reflexive possessives correlates with the way definiteness is expressed in different languages. Languages which encode definiteness postnominally, such as Bulgarian, or not at all, such as Russian, allow them; while languages which have prenominal definiteness marking, such as English, do not. See also Reuland 2011.

**Majority superlative readings (4.1):** Bulgarian behaves as a DP language when the definite article is present and like an NP language when it is not, leading to the hypothesis that DP is projected in the presence of the definite article and that Bulgarian has bare NPs (PosPs, and QPs) otherwise.

**Exhaustivity presupposition for possessives (4.2):** Bulgarian possessives do not have a definiteness feature. So exhaustivity interpretations are limited to cases where DP is projected and [DEFINITE] is valued and positive.

**Extraction from subject position (4.3):** In Bulgarian, extractions from subject and object are both possible but only in the absence of the definite article, leading to the hypothesis that Bulgarian subjects are only sometimes DPs and only block extraction when they are.

Table 1 reviews these properties, contrasting the behavior of Bulgarian with English.

**Table 1.** Bulgarian vis-à-vis NP-DP properties

Phenomena	Ideal DP languages	Ideal NP languages	Bulgarian
2.1 Clitic doubling	√	×	√
2.2 Obligatory number morphology	√	×	√
3.1 Neg raising	√	×	×
3.2 Subject expletives	√	×	×
4.1 Majority superlative reading available	√	×	DP √ NP ×
4.2 Exhaustivity presupposition in possessives	√	×	DP √ NP ×
4.3 Extraction out of NP subject islands	×	√	DP × NP √

**5.1. The Structure of DP in Bulgarian**

Recall (section 1) that Bulgarian does not have the equivalent of an indefinite determiner. The only potential candidate, *edin* ‘one’, does not pattern like the English indefinite article *a(n)* as it only co-occurs with [+specific] nominal phrases. Additionally, *edin* can co-occur with the definite determiner, as in *ed-*

*nijat dobâr student* ‘the one good student’.<sup>16</sup> If *edin* headed DP, then it should not co-occur with ‘the’. Also, *edin* behaves more like a modifier than a determiner. As evidence for this, we see that a focused adjective can appear outside the scope (i.e., to the left) of *edin*, as in (35a). Conversely, the same adjective may not appear in this position if the definite determiner *-jat* is present, as in (35b). The contrast in (35) suggests that (i) adjectivals cannot move across DP, and (ii) *edin* does not head DP.

- (35) a. *dobâr*    *edin* student    [contrastive focus on *dobâr*]  
           good    one    student  
           ‘one **good** student’
- b. \**dobâr*    *edinijat* student [contrastive focus on *dobâr*]  
           good    one<sub>DEF</sub> student  
           (the one **good** student)

Additional evidence, from split NPs, supports the hypotheses that (i) DP is an optional projection in Bulgarian and (ii) indefinite NPs don’t project it. It is a well-documented fact that split NPs arise in a number of languages, including German, Bulgarian, Polish, and Serbo-Croatian (Fanselow and Ćavar 2002). An example of this phenomenon is given here in (36), where either the adjective *nova* ‘new’ or the noun *kola* ‘car’ may be moved out of the NP and into a focus position before the verb complex *e kupil* ‘he has bought’. Example (37) is parallel, except that (37b) is unacceptable with a definite article affixed to the adjective. When the object is definite (i.e., projects a DP, under our account), it is possible to extract the adjective in order to form a split NP with focus on the first element. Extraction of the noun, though, is disallowed. This contrast can be accounted for if DP is only projected when a nominal is +definite and if DP blocks extraction out of its complement (i.e., *kola* cannot move directly out of NP in (37b) without crossing DP).<sup>17</sup>

<sup>16</sup> Note that *edinijat dobâr student* does not mean ‘the one good student’ in the sense that there is only one good student. To say ‘The one good student from my class came to my office hour [there being only one good student]’, we would use the expression *edinstvenijat dobâr student* ‘the one and only good student’.

<sup>17</sup> There are two cases of apparent split NPs which seem to violate the account given here. They are shown here in examples (ia) and (ib). The central reason for not taking these examples as instances of the split NP phenomenon is that there is no well-formed nominal out of which the extractions could take place. That is, the phrase \**kolata nova* is itself ill-formed. Consider the derivation in (ii), which would be the source of the split NPs in (ia) and (ib). It is ill-formed in that only the highest element within a nominal (*nova* here) may move to D. Given the interpretations of (ia) and (ib), it is clear that the adjective *nova* is a depictive predicate modifying the object.

- (36) a. *nova* e kupil [NP **nova** kola] (ne stara).  
 new is bought new car not old  
 'He bought a *new* car, not an old one.'
- b. *kola* e kupil [NP nova **kola**] (ne lodka).  
 car is bought new car not boat  
 'He bought a new *car*, not a boat.'
- (37) a. *novata* e kupil [DP **novata** kola] (ne starata).  
 new<sub>DEF</sub> is bought new<sub>DEF</sub> car not old<sub>DEF</sub>  
 'He bought **the** *new* car, not the old one.'
- b. \**kola* e kupil [DP novata **kola**] (ne lodka).  
 car is bought new<sub>DEF</sub> car not boat  
 ('He bought **the** new *car*, not a boat.')

Having discussed the optionality of DP projections, we turn our attention now to the structure of DP itself. A possible structure of DP would involve *ta* as an affixal head of DP, and head movement of an element within its scope to provide a carrier for the affix. Some supporting evidence involves the agreement of coordinated modifiers with a following noun and their interaction with the determiner. In (38a), a pair of coordinated possessives modifies *kâšta* 'house'. Here, possession is shared. (38b) is equivalent to (38a), but definite. There, only the first of two coordinated possessives can host *-ta*. In (38c), each of the coordinated NPs is contained in its own DP projection and carries a definite determiner *-ta* on its possessive modifier. In (38d), the coordinated possessor phrase cannot have a definite determiner attached to it. The derivation in (38b) involves movement of the first conjunct into DP, where it is affixed to *-ta*. Example (38d), which would involve movement of the entire conjoined possessor phrase into DP is ungrammatical. This suggests that movement into DP involves head rather than phrasal movement.<sup>18</sup>

- (i) a. *nova* e kupil kolata.  
 new is bought car<sub>DEF</sub>  
 'He bought the car, *new*.'
- b. *kolata* e kupil nova.  
 car<sub>DEF</sub> is bought new  
 'He bought *the* car, new.'
- (ii) [DP [D ta ] [NP [AP nova ] [NP kola ]]] →  
 \*[DP [D [N kola ] ta ] [NP [AP nova ] [NP kola ]]]

<sup>18</sup> One additional possibility is the case in which each conjunct has its own definite determiner, (i). Given that the interpretation of (i) is identical to that of (39c), it clearly





- (40) b. David govori s *baštata* gord s tova dete.  
 David spoke with father<sub>DEF</sub> proud of this child  
 'David spoke with the *father* who is proud of this child.'
- c. \*David govori s [gord s tova deteto/ja] bašta.  
 David spoke with [proud of this child<sub>DEF</sub>] father  
 ('David spoke with the father who is proud of this child.')

Not only must the definite determiner attach to the head of the AP, but when the AP is the leftmost element in the nominal expression, the head of that phrase moves across any higher constituent (such as a possessive clitic) to merge with the definite determiner. Consider first the addition of a possessive, which can appear either before the AP (41a) or after it (41b). Here, *-ija* attaches to the leftmost element of the nominal, giving either *Aninija* (41a) or *gordija* (41b). Consider what happens if the possessive is a clitic occurring to the left and outside of the nominal expression, i.e., *ì* 'her' as in (42). Here, the head of the AP must move across the clitic to get to DP. Example (42) thus suggests movement of the head of AP out of the nominal expression, to the head of DP, as shown in (43).<sup>19</sup> That *gord* is displaced from the AP and moves across a possessive clitic to get into the higher DP is evidence that it is derived through head movement, and that the definite determiner occupies the head of DP.

- (41) a. David govori s **Aninija** gord sâs svoja sin bašta.  
 David spoke with Anna's<sub>DEF</sub> proud of self's son father
- b. David govori s gordija sâs svoja sin **Anin** bašta.  
 David spoke with proud<sub>DEF</sub> of self's son Anna's father  
 'David spoke with Anna's father, who is proud of his own son.'
- (42) David govori s gordija ì sâs svoja sin bašta.  
 David spoke with proud<sub>DEF</sub> her of self's son father  
 'David spoke with her father, who is proud of his own son.'
- (43) [DP [D [A gord ] -ija ] [PossP ì [NP [AP ~~gord~~ sâs svoja sin ] bašta]]]

<sup>19</sup> Note that the structure proposed in (43) has AP as an NP-adjunct, contra claims in Bošković 2005, Corver 1992, and Dimitrova-Vulchanova 2000 that the lexical structure of Bulgarian nominals parallels English, with AP dominating NP. Nothing in the analysis presented here rests on that distinction.

Based on the facts presented above, we might conclude that the definite determiner heads DP in Bulgarian, and that some element in its complement undergoes head movement to D.

There are other facts which suggest an alternative analysis, one in which the element bearing the definite affix (or agreeing with it) occupies Spec,DP. In a narrow range of cases, phrasal movement into DP appears to be possible. In (44), *svetlo* 'light' or *mnogo* 'very' modify *sinja* 'blue'. Here the definite determiner can attach to the entire expression *svetlo sinja* or *mnogo sinja* (44a), but not to the first element *svetlo* or *mnogo* (44b).

- (44) a. svetlo/mnogo sinjata kniga  
light/very blue<sub>DEF</sub> book  
'the light/very blue book'
- b. \*svetloto/mnogoto sinja kniga  
light<sub>DEF</sub>/very<sub>DEF</sub> blue book  
'(the light/very blue book)'

The fact that *svetlo* has a neuter ending (i.e., *-o*) rather than a feminine ending (i.e., *-a*, which would agree with *kniga*) shows it to behave like an adverb, modifying *sinja*. These data would support an analysis such as that in (45), where the *ta*-marked element is a phrase occupying Spec,DP. Example (45a) shows DP having a null head and *ta* as an agreement affix on the element in Spec,DP, and (45b) has *ta* as head of DP with an element moving to Spec,DP to host the affix at Spell-out. The analyses presented in Dimitrova-Vulchanova and Giusti 1999 and Arnaudova 1996 are closer to that shown in (45a).

- (45) a. [<sub>DP</sub> [<sub>AP</sub> nova]-ta [<sub>D'</sub> [<sub>D</sub> Ø ] [<sub>NP</sub> [<sub>AP</sub> nova] [<sub>NP</sub> kola]]]]  
b. [<sub>DP</sub> [<sub>AP</sub> nova] [<sub>D'</sub> [<sub>D</sub> ta ] [<sub>NP</sub> [<sub>AP</sub> nova] [<sub>NP</sub> kola]]]]

Our analysis of Bulgarian, respective of its being a DP language, an NP language, or something between does not rely on the correctness of any one of the three structures presented here (i.e., (43), (45a), or (45b)). Our main claims are that DP is optionally projected, that DP is always [+definite], and that the presence/absence of DP affects the behavior of nominal expressions with respect to the phenomena discussed in section 4. These properties are summarized in Table 2 on the following page.

**Table 2.** Bulgarian variable NP-DP properties

Phenomena	Bulgarian
4.1 Majority superlative reading available	DP ✓ NP ✗
4.2 Exhaustivity presupposition in possessives	DP ✓ NP ✗
4.3 Extraction out of NP subject islands	DP ✗ NP ✓

**5.2. DP as a Functional Projection in Bulgarian**

Pesetsky and Torrego (2007) (revising Chomsky 2000/2001), propose a feature-sharing version of Agree, as stated in (46). This version of Agree accounts for all four logical combinations of the featural properties of Valued and Interpretable, as in (47), revising the previously biconditional relation by which a feature is interpretable iff it is valued. This separation of valuation and interpretability opens the door to an analysis of the divergent behavior of English and Bulgarian determiners and to an explanation for why DP is only projected in Bulgarian when the definite determiner is selected.

(46) **Agree (feature-sharing version)** [Pesetsky and Torrego 2007: (5)]

- (i) An unvalued feature F (a probe) on a head H at syntactic location  $\alpha$  ( $F\alpha$ ) scans its c-command domain for another instance of F (a goal) at location  $\beta$  ( $F\beta$ ) with which to agree.
- (ii) Replace  $F\alpha$  with  $F\beta$ , so that the same feature is present in both locations.

(47) **Types of features** [Pesetsky and Torrego 2007: (9)]

- uF val uninterpretable, valued      iF val interpretable, valued
- uF [ ] uninterpretable, unvalued      iF [ ] interpretable, unvalued

In English, the head of DP can be valued or unvalued. For example, when DP is headed by a definite article, e.g. *the book*, it is inherently valued as [+definite]. Here feature sharing does not obtain. In other cases, D carries an interpretable, unvalued feature for definiteness,  $iF [ ]$ . This feature scans its domain, and takes its value (+ or -) from its goal, as in (48). Thus, the head of DP

takes the value of its [DEFINITE] feature from another element in the nominal expression, *two* in (48a) and *my* in (48b).

- (48) a. [DP D<sub>[DEF]</sub> [QP two<sub>[DEF-]</sub> [NP books]]] →  
           [DP D<sub>[DEF-]</sub> [QP two<sub>[DEF-]</sub> [NP books]]]  
 b. [DP D<sub>[DEF]</sub> [PossP my<sub>[DEF+]</sub> [NP books]]] →  
           [DP D<sub>[DEF+]</sub> [PossP my<sub>[DEF+]</sub> [NP books]]]

In Bulgarian, the head of DP is always valued for the feature [+definite] and does not participate in Agree. When there is no definite article, as in (49a), there is no DP. When the definite article *-te* appears, as in (49b–f), it correlates with definiteness, entailed by an interpretable and valued definite feature. This explains why Bulgarian is unrestricted with respect to what element moves to the *-te* affix. As (49b–f) shows, this movement may involve possessors, nouns, adjectives, quantifiers, or numerals, suggesting it is prosodically motivated so as not to strand the affix (something akin to Lasnik’s 1981 Stray Affix Filter). Whether it is to head or Spec of DP is inconsequential, since in neither case will feature-sharing obtain.

- (49) a. *moi kartini* [PossP *moi* [NP *kartini*]]  
       *my pictures*  
 b. *moite kartini* [DP *moi-te*<sub>[DEF+]</sub> [PossP *moi* [NP *kartini*]]]  
       *my<sub>DEF</sub> pictures*  
 c. *kartinite* [DP *kartini-te*<sub>[DEF+]</sub> [NP *kartini*]]  
       *pictures<sub>DEF</sub>*  
 d. *novite kartini* [DP *novi-te*<sub>[DEF+]</sub> [NP *novi* *kartini*]]  
       *new<sub>DEF</sub> pictures*  
 e. *vsičkite kartini* [DP *vsički-te*<sub>[DEF+]</sub> [QP *vsički* [NP *kartini*]]]  
       *all<sub>DEF</sub> pictures*  
 f. *dвете kartini* [DP *dve-te*<sub>[DEF+]</sub> [NumP *dve* [NP *kartini*]]]  
       *two<sub>DEF</sub> pictures*

## 6. Conclusion

This paper has shown that Bulgarian is neither a DP language nor an NP language. While it can project a DP, it doesn’t always. It thus behaves with respect to some phenomena as a DP language, with respect to others as an NP language, and with respect to yet others as neither. Bulgarian thus points to yet another important parameter along which languages may vary: strong DP languages (English), non-DP languages (Chinese), and weak DP languages

(Bulgarian). Specifically, we found that Bulgarian seems to be a DP language with respect to Clitic doubling and Obligatory number morphology, and an NP language when Neg raising and subject expletives are considered. For other phenomena, such as Majority superlative readings, Exhaustivity presupposition for possessives, and Extraction out of NP subject islands, Bulgarian shows DP-like behavior when nominals have a definite determiner, but not otherwise.

University of South Carolina  
TASSEVAK@mailbox.sc.edu  
dubinsky@sc.edu

## References

- Arnaudova, Olga. (1996) "The Bulgarian noun (adjective) movement to D". *University of Trondheim working papers in linguistics* 28: 1–28.
- Borer, Hagit. (2005) *Structuring sense*. Vol. I: *In name only*. Oxford: Oxford University Press.
- Bošković, Željko. (2005) "On the locality of left branch extraction and the structure of NP". *Studia linguistica* 59(1): 1–45.
- . (2012) "On NPs and clauses". Günther Grewendorf and Thomas Ede Zimmermann, eds. *Discourse and grammar: From sentence types to lexical categories*. Berlin: de Gruyter Mouton, 179–242.
- Chomsky, Noam. (2000) "Minimalist inquiries". Roger Martin, David Michaels, and Juan Uriagereka, eds. *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press, 89–156.
- . (2001) "Derivation by phase". Michael Kenstowicz, ed. *Ken Hale: A life in language*. Cambridge, MA: MIT Press, 1–52.
- Corver, Norbert. (1992) "Left branch extraction". *NELS* 22. Amherst: University of Massachusetts, Graduate Linguistic Student Association, 67–84.
- Davies, William and Stanley Dubinsky. (2003) "On extraction from NPs". *Natural language and linguistic theory* 21(1): 1–37.
- Despić, Miloje. (2015) "Phases, reflexives, and definiteness". *Syntax* 18(3): 201–34.
- Dimitrova-Vulchanova, Mila. (1995) "Clitics in Slavic". *Studia linguistica* 49(1): 54–92.
- . (2000) "Possessive constructions and possessive clitics in the English and Bulgarian DP". Frits Beukema and Marcel den Dikken, eds. *Clitic phenomena in European languages*. Amsterdam: John Benjamins, 121–46. [*Linguistik Aktuell/Linguistics today*]
- Dimitrova-Vulchanova, Mila and Giuliana Giusti. (1998) "Fragments of Balkan nominal structure". Artemis Alexiadou and Chris Wilder, eds. *Pos-*

- sessors, predicates and movement in the determiner phrase. Amsterdam: John Benjamins, 333–60.
- Dimitrova-Vulchanova, Mila and Giuliana Giusti. (1999) "Possessors in the Bulgarian DP". Mila Dimitrova-Vulchanova and Lars Hellan, eds. *Topics in South Slavic syntax and semantics*. Amsterdam: John Benjamins, 163–92.
- Dimitrova-Vulchanova, Mila and Olga Mišeska Tomić. (2009) "The structure of Bulgarian and Macedonian nominal expressions: Introduction". Mila Dimitrova-Vulchanova and Olga Mišeska Tomić, eds. *Investigations in the Bulgarian and Macedonian nominal expressions*. Trondheim: Tapir Academic Press, 1–24.
- Fanselow, Gisbert and Damir Ćavar. (2002) "Distributed deletion". Artemis Alexiadou, ed. *Theoretical approaches to universals*. Amsterdam: John Benjamins, 65–107.
- Geist, Ljudmila. (2013) "Bulgarian *edin*: The rise of an indefinite article". Uwe Junghanns, Dorothee Fehrmann, Denisa Lenertová, and Hagen Pitsch, eds. *Formal description of Slavic languages: The ninth conference*. Frankfurt am Main: Peter Lang, 125–48.
- Lasnik, Howard. (1981) "Restricting the theory of transformations". Norbert Hornstein and David Lightfoot, eds. *Explanation in linguistics*. London: Longmans, 152–73.
- Partee, Barbara. (2006) "A note on Mandarin possessives, demonstratives, and definiteness". Betty J. Birner and Gregory Ward, eds. *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*. Amsterdam: John Benjamins, 263–80.
- Pereltsvaig, Asya. (2013) "Noun phrase structure in article-less Slavic languages: DP or not DP?" *Language and linguistics compass* 7(3): 201–19.
- . (2015) "The functional structure of the nominal domain". Antonio Fábregas, Mike Putnam, and Jaume Mateu, eds. *Contemporary linguistic parameters*. London: Bloomsbury, 303–31.
- Pesetsky, David and Esther Torrego. (2007) "The syntax of valuation and the interpretability of features". Simin Karimi, Vida Samiian, and Wendy K. Wilkins, eds. *Phrasal and clausal architecture: Syntactic derivation and interpretation*. Amsterdam: John Benjamins, 262–94.
- Reuland, Eric. (2011) *Anaphora and language design*. Cambridge, MA: MIT Press.
- Rudin, Catherine. (1997) "AgrO and Bulgarian pronominal clitics". Martina Lindseth and Steven Franks, eds. *Formal approaches to Slavic linguistics: The Indiana meeting, 1996*. Ann Arbor, MI: Michigan Slavic Publications, 224–52.
- Schick, Ivanka. (2000) "Clitic doubling constructions in Balkan-Slavic languages". Frits Beukema and Marcel den Dikken, eds. *Clitic phenomena in European languages*. Philadelphia: John Benjamins, 259–92.

Van Hofwegen, Janneke. (2013) "Extraction or splitting? How A' movement phenomena in Lithuanian can shed light on the syntax-phonology interface and the NP-DP debate". Unpublished ms., Stanford University.