



**University of Belgrade
Faculty of Philology
Department of General Linguistics**

James Branch Cabell • BREWER

BeLiDa 1

Thematic Collection of Papers



Reviewers

- ALEKSANDRA BLATEŠIĆ, Faculty of Philosophy, University of Novi Sad
BORKO KOVAČEVIĆ, Faculty of Philology, University of Belgrade
BRANKA BARČOT, Faculty of Humanities and Social Sciences, University of Zagreb
DORĐE BOŽOVIĆ, Faculty of Philology, University of Belgrade
GORANKA ANTUNOVIĆ, Faculty of Humanities and Social Sciences,
University of Zagreb
GORDANA RISTIĆ, Faculty of Philosophy, University of Novi Sad
IGOR LAKIĆ, Faculty of Philology, University of Montenegro
JELENA BRAJOVIĆ, Faculty of Philology, University of Belgrade
JELENA TUŠEK, Faculty of Humanities and Social Sciences, University of Zagreb
JOACHIM MATZINGER, Institute of Habedfsburg and Balkan Studies,
University of Vienna
MAJA MILIČEVIĆ PETROVIĆ, Department of Interpreting and Translation,
University of Bologna
MARKO SIMONOVIĆ, Institute of Slavonic Studies, University of Graz
NATALIJA PANIĆ CEROVSKI, Faculty of Philology, University of Belgrade
NEVENA CEKOVIĆ, Faculty of Philology, University of Belgrade
RADOSAV PUŠIĆ, Faculty of Philology, University of Belgrade
RAJNA DRAGIĆEVIĆ, Faculty of Philology, University of Belgrade
SMILJA SRDIĆ, Faculty of Philology, University of Belgrade
SOFIJA BILANDŽIJA, Faculty of Philology, University of Belgrade
VESNA POLOVINA, Faculty of Philology, University of Belgrade
VIRNA KARLIĆ, Faculty of Humanities and Social Sciences, University of Zagreb

DEPARTMENT OF GENERAL LINGUISTICS
FACULTY OF PHILOLOGY
UNIVERSITY OF BELGRADE, SERBIA

BeLiDa 1

Thematic Collection of Papers

Editors

NATALIJA PANIĆ CEROVSKI

BORKO KOVAČEVIĆ

MILICA DINIĆ MARINKOVIĆ



Belgrade, 2022

Organizing Committee

Natalija Panić Cerovski, Associate Professor, Head of the Department of General Linguistics, Faculty of Philology, University of Belgrade

Jasmina Moskovljević Popović, Full Professor
Department of General Linguistics, Faculty of Philology, University of Belgrade

Maja Đukanović, Full Professor,
Department of General Linguistics, Faculty of Philology, University of Belgrade

Borko Kovačević, Associate Professor,
Department of General Linguistics, Faculty of Philology, University of Belgrade

Milica Dinić Marinković, Assistant Professor,
Department of General Linguistics, Faculty of Philology, University of Belgrade

Ivan Knjižar, Assistant Professor, Department of General Linguistics,
Faculty of Philology, University of Belgrade

Đorđe Božović, Assistant Professor, Department of General Linguistics,
Faculty of Philology, University of Belgrade

Milena Oparnica, Teaching Assistant, Department of General Linguistics,
Faculty of Philology, University of Belgrade

Emilija Milojević, Teaching Assistant, Department of General Linguistics,
Faculty of Philology, University of Belgrade

Jelena Budimirović, Teaching Assistant, Department of General Linguistics,
Faculty of Philology, University of Belgrade

Scientific Committee

Željko Bošković, Full Professor of Linguistics and Faculty in the Cognitive Science Program, University of Connecticut, USA

Igor Lakić, Full Professor of Linguistics, Faculty of Philology,
University of Crna Gora, Montenegro

Junichi Toyota, Professor, Department of Language and Culture,
Osaka City University, Japan

Maja Miličević Petrović, Professor, Department of Interpreting and Translation,
University of Bologna, Italia

Daniel Śtapek, Professor, Institute of Romance Studies, Department of Italian Philology,
Faculty of Philology, Jagiellonian University Krakow, Poland

Jagoda Granić, Professor, Faculty of Humanities and Social Sciences,
University of Split, Croatia

Vesna Polovina, Full Professor, Department of General Linguistics, Faculty of Philology,
University of Belgrade, Serbia

Snežana Gudurić, Full Professor, Department of Romance Studies,
Faculty of Philosophy, University of Novi Sad, Serbia

Nenad Tomović, Professor, Department of English Language, Literature and Culture,
Faculty of Philology, University of Belgrade, Serbia

EDITORS' PREFACE

For the Department of General Linguistics at the Faculty of Philology, University of Belgrade, Serbia, the day of October 1st 2021 was the special anniversary. On that day, thirty years ago, the first generation of students was enrolled at a newly established studying programme of General Linguistics at the Faculty of Philology. To mark this significant anniversary, members of the Department organized an international linguistics conference – *Belgrade Linguistics Days – BeLiDa 2021*, held on the 3rd and 4th of December 2021 at the Faculty of Philology, gathering more than 70 scholars from 28 universities and institutes from all around the world.

We were honoured and pleased to have as keynote speakers at *BeLiDa 2021* renowned linguists: **Željko Bošković**, Full Professor of Linguistics and Faculty in the Cognitive Science Program, University of Connecticut; **Brian D. Joseph**, Distinguished University Professor of Linguistics and The Kenneth E. Naylor Professor of South Slavic Languages and Linguistics at the Ohio State University; **Jagoda Granić**, Professor of Linguistics at the Faculty of Humanities and the Social Sciences, University of Split; **Boban Arsenijević**, Professor of Linguistics at the Karl-Franzens-University of Graz; **Vesna Polovina**, Professor of General Linguistics at the Faculty of Philology, University of Belgrade, one of the founders of the Department of General Linguistics and long-term Head of the Department; **Jasmina**

Moskovljević Popović, Professor of General Linguistics, who has been teaching at the Department from its founding.

We would also like to express our deepest gratitude to all *BeLiDa 2021* contributors, for their inspiring work in investigating various linguistic phenomena from diverse perspectives. The result of their rich and exciting insights is this book.

The book is a thematic collection of selected research papers presented at the conference, organised into three sections: *Current Trends in Grammar Studies*, *Discourse – From Conversation to Construction*, and *Language and Culture*.

The *Current Trends in Grammar Studies* section consists of the following eight chapters: Željko Bošković (University of Connecticut, USA) explores the (dis)agreement between the generative and the functional approaches as well as generative typology and traditional typology, James Joshua Pennington (Concordia College, USA) examines what priming effects reveal about velar alternations in South Slavic languages, Branislav Ivanović & Sofija Bilandžija (University of Belgrade, Serbia) from a contrastive and typological perspective analyse the system of strong conjugation in contemporary German and Norwegian, Tatjana Marvin (University of Ljubljana, Slovenia) presents an analysis of adnominal arguments in phrases headed by deverbal nouns in Slovenian, Zorica Kovačević (University of Belgrade, Serbia) examines the most frequent examples of syntactical stylistic analysis in Swedish in relation to advanced language learning and translating, Vesna Plesničar (University of Nova Gorica, Slovenia) presents the results of the corpus study of the ordering of selected attributive adjectives in Slovenian, Tijana Vesić Pavlović (University of Belgrade, Serbia) & Milena Jakić Šimšić (Institute for the Serbian Language of SASA, Belgrade, Serbia) investigate the effect of context on the (a)symmetry of Serbian adjective antonyms, Emilija Milojević (University of Belgrade, Serbia) examines the syntactic complexity of the written production of third-year students of French at the Faculty of Philology in Belgrade.

The *Discourse – From Conversation to Construction* section of this publication consists of the following seven chapters: Jana Jurčević

(University of Zagreb, Croatia) investigates socio-pragmatic and semiotic features of public toilet graffiti, in a broader domain of critical discourse analysis, Igor Ilić (University of Strasbourg, France) explores argument narrativization through discourse representation, Maja Brkljač (University of Belgrade, Serbia) examines the linguistic means used in headlines and leads and their role in the interpretation of news articles, Ana Lalić (University of Sarajevo, Bosnia and Herzegovina) focuses on overcoming the “bad data” problem, analysing the examples of spoken language in Italian epistolary discourse, Milena Oparnica & Natalija Panić Cerovski (University of Belgrade, Serbia) analyse diminutives within spoken corpora with respect to morphology and pragmatics, Milica Dinić Marinković (University of Belgrade, Serbia) investigates orthographically unmarked fictional dialogues and presents some means for their extraction based on the identification of formal linguistic features of strings, Vesna Polovina (University of Belgrade, Serbia) explores non-omission of deictic personal pronouns on a corpus of spoken conversational language, and the factors that might be influencing the choice between overt and omitted deictic personal pronouns.

The *Language and Culture* section brings three chapters: Junichi Toyota (Osaka City University, Japan) tackles the issue of Y-N words in Indo-European from various perspectives, Ana M. Jovanović (University of Belgrade, Serbia) investigates the Chinese concept /xiào/ (filial piety) from the cultural linguistics perspective, Danilo Savić (Institute for Balkan Studies SASA, Serbia) explores Illyrian ethnonyms and their supposed Albanian cognates.

On this occasion, we would like to express our warmest thanks to our colleagues at the Faculty of Philology, for their help and support throughout the process of organizing the conference, and throughout the process of publishing the book you are holding in your hands (or reading its e-edition from the screen) right now. We are immensely thankful to: Professor Iva Draškić Vićanović, Dean; Professor Nenad Tomović, Vice-Dean for Science; Ljiljana Petković, College of Deans Secretary; Gordana Đoković, Professor of Library Science; Jadranka Krsmanović, Head of the Financial and Accounting

Affairs Department, and Tanja Kafedžić; Nikola Jelić, from Legal Affairs Department; Branko Petrić, Aleksandar Blagojević, Dragan Vranešević, Nikola Tarasjev from IT Service, and Dragica Jović, for being a great organizer of the conference's buffet. We are also very thankful to Bogdan Milojević, for creating the conference website, and for designing the conference poster used for the cover page of this book.

We are sincerely grateful to all the members of the *BeLiDa's 2021* Scientific and the Organizing Committee, as well as to each and every one of the reviewers, for their meticulous work and insightful suggestions, which made *BeLiDa 1 – a Thematic Collection of Papers* the best that it could be.

At last but not least, we express our profound appreciation and heartfelt gratitude to all our teachers and students of the Department of General Linguistics at the Faculty of Philology – University of Belgrade, Serbia, to whom we dedicate the first book of *BeLiDa* proceedings.

Belgrade, June 2022

The Editors

*To our students and teachers
who have been learning, working, and growing together
with and within the Department of General Linguistics*

CONTENTS

| | |
|-----------------------|---|
| EDITORS' PREFACE..... | 5 |
|-----------------------|---|

CURRENT TRENDS IN GRAMMAR STUDIES

Željko Bošković

| | |
|---|----|
| FORMALISM <u>AND</u> , NOT <u>VS</u> , FUNCTIONALISM..... | 17 |
|---|----|

Joshua Pennignton

| | |
|--|----|
| CYCLICAL SUPPLETION IN SLAVIC: WHAT PRIMING EFFECTS REVEAL ABOUT VELAR ALTERNATIONS IN BOSNIAN, CROATIAN, MONTENEGRIN, SERBIAN | 93 |
|--|----|

Branislav Ivanović & Safija Bilandžija

| | |
|--|-----|
| SYSTEM OF STRONG CONJUGATION IN CONTEMPORARY GERMAN AND NORWEGIAN: A CONTRASTIVE AND TYPOLOGICAL PERSPECTIVE | 113 |
|--|-----|

Tatjana Marvin

| | |
|---|-----|
| ADNOMINAL ARGUMENTS IN PHRASES HEADED BY DEVERBAL NOUNS IN SLOVENIAN | 135 |
|---|-----|

Zorica Kovačević

| | |
|--|-----|
| ANALYSIS OF SYNTACTICAL STYLISTIC ELEMENTS IN SWEDISH | 159 |
|--|-----|

Vesna Plesničar

A CORPUS INVESTIGATION OF THE ORDERING OF SELECTED
ATTRIBUTIVE ADJECTIVES IN SLOVENIAN 173

Tijana Vesić Pavlović, Milena Jakić Šimšić

THE EFFECT OF CONTEXT ON THE (A)SYMMETRY OF SERBIAN
ADJECTIVE ANTONYMS 197

Emilija Milojević

COMPLEXITÉ SYNTAXIQUE À L'ÉCRIT EN L2: LE CAS
DU FRANÇAIS 219

DISCOURSE – FROM CONVERSATION TO CONSTRUCTION

Jana Jurčević

A GLIMPSE AT SEMIOTIC LANDSCAPES AND STUDENT
POPULATION OF ZAGREB'S FACULTY OF HUMANITIES
AND SOCIAL SCIENCES: SOCIO-PRAGMATICS OF PUBLIC
TOILET GRAFFITI 243

Igor Ilić

ARGUMENT NARRATIVIZATION THROUGH DISCOURSE
REPRESENTATION: EXAMPLE OF ONE PUBLIC APPEARANCE 271

Maja Brkljač

THE IMPORTANCE OF THE HEADLINE AND LEAD
FOR THE INTERPRETATION OF MEDIA ARTICLES 289

Ana Lalić

OVERCOMING THE "BAD DATA" PROBLEM: EXAMPLES OF SPOKEN
LANGUAGE IN ITALIAN EPISTOLARY DISCOURSE 307

Milena Oparnica & Natalija Panić Cerovski

THE USE OF DIMINUTIVES IN EVERYDAY COMMUNICATION 327

Milica Dinić Marinković

(SOME SIMPLE) MEANS FOR EXTRACTION OF
ORTHOGRAPHICALLY UNMARKED FICTIONAL DIALOGUE 347

Vesna Polovina
NON-OMISSION OF DEICTIC PERSONAL PRONOUNS 367

LANGUAGE AND CULTURE

Junichi Toyota
UNIFORMITARIANISM AND 'YES' AND 'NO' IN INDO-EUROPEAN
LANGUAGE 407

Ana Jovanović
孝 /XIÀO/ — THE CULTURAL LINGUISTICS PERSPECTIVE..... 421

Danilo Savić
SOME ILLYRIAN ETHNONYMS AND THEIR SUPPOSED ALBANIAN
COGNATES: TAULANTII, DELMATAE, DARDANI..... 449

*Current Trends in
Grammar Studies*

Željko Bošković*
University of Connecticut

FORMALISM AND, NOT VS, FUNCTIONALISM Plenary talk

Abstract

Baker & McCloskey (2007) examine the relationship between traditional typology and generative theoretical syntax. Since then typology has started to play an increasingly more important role within the latter camp to the point that we can actually talk about generative typology. Given that traditional typology is generally associated with functional approaches (see e.g. Nichols 2007), this then gives us two approaches to typology, which I will refer to as traditional (TT) and generative typology (GT).¹ In a way, then, typology is setting grounds for a potential rapprochement of the functional and the formal approach to language more generally. This paper will provide a number of remarks to this effect, within a unificational view where both of these approaches have a place, i.e. where they are in principle not in competition with each other.²

* zeljko.boskovic@uconn.edu

¹ Below, I will use TT and GT to refer either to traditional typology and generative typology or the practitioners of these approaches, a distinction which should be clear from the context. Additionally, because of the TT/GT distinction, the term typology will often be used neutrally below.

² The paper will thus also contain a more general discussion of these approaches (see Thomas 2020 for a recent overview). The two approaches have been in opposition for so long (and rarely communicating with each other, the only communication often involving one-sided attacks and dismissals), that it is really impossible to find anyone who is completely neutral between these two approaches. This paper is written from the perspective of a formalist who is looking for a rapprochement between the two approaches. The perspective will inevitably in some places give

To this end, the paper will show that many of what are assumed to be clear demarcation lines between the traditional generativist and the traditional functionalist/typological camp (and the two approaches to typology) are not clear, in fact may not be there at all—the two approaches are not in opposition as much as they used to be, and as the practitioners of the two camps still seem to think they are. In this respect, it will be shown that many of the perceived irreconcilable differences and antagonism between the two fields are there because, to put it a bit more abstractly, there are differences between the actual state of affairs in field X and the way field X is perceived by field Y, where the negative reaction of Y to X is based on Y's perception of X.

The paper will also discuss points of convergence between the two traditions. One recent point of convergence in fact concerns the emergence of generative typology. Several other points of convergence will be discussed, including the minimalist assumption that language is characterized by efficient design, which opens the door for bringing in functional considerations into formalist approaches like minimalism. More generally, the paper argues for an overall view of the field where the functional and formalist approaches are seen not as being in competition but as complementary to each other (much of which will be based on a re-evaluation of some of the fundamental issues regarding the field where the formalist and functionalist approaches have been assumed to be in opposition (more precisely, where the practitioners of the two camps have been actively antagonistic to one another). Concrete examples of complementarity will also be provided.

Keywords: traditional typology, generative typology, functional approach, formal approach, complementarity.

1. FORMALISM VS FUNCTIONALISM: NOT THAT MUCH OF A DIVIDE

This section is intended to show that the divide between the generativist and the functionalist/typological camp is bigger in the slogans that are used as characterizations of the respective camps than in actual

the discussion a somewhat subjective (but also activist) flavor (for another work from a formalist perspective which is still quite different from the current one, see Newmeyer 1998). An important note: there are various formal approaches which considerably differ from each other (functional approaches are also far from being monolithic); when the differences are important enough to affect the discussion, what is assumed by the formal approach will be the Chomskian tradition, broadly characterized by what is referred to as the Principles and Parameters or the Minimalist approach.

research practice. Much of it is due to misunderstandings of the slogans in question, as well as taking them too literally, at their face value, without actually trying to see what is behind them.

1.1. EVERY LANGUAGE IN ITS OWN TERMS

A position that is often associated with the traditional functionalist/typological camp by the generativists, in fact often looked at as an insurmountable and fundamental difference, is that every language should be described in its own terms. This is often perceived by generativists as a there-is-no-universal-grammar attitude.³ The position in question is actually not universally adopted in the TT camp (more on that below). But there is a bigger issue here. What modern typologists mean by this is not what the generativists think they do. There are actually two misconceptions at work here. The generativists assume that what is meant by this stance is what American structuralists, who were the originators of the position in question

³ The perception among generativists that the stand is a reflection of a no-universal grammar (UG) attitude is somewhat misguided due to a difference in the phenomena that are investigated. As noted in Baker (2015), the kind of phenomena typological works typically explore are not considered by generativists to be the prime source of UG universals; those concern more abstract properties involving phrase structure, structural dependencies, locality relations involved in movement, coreference (im)possibilities..., which traditional typologists in turn generally do not deal with. To illustrate, here is one such generalization from Bošković (2012) (based on earlier work regarding only Slavic by Uriagereka 1988, Corver 1992; (i) is restated in the standard implicational universal way in fn17, which also discusses other conditions on the possibility of (ii). Note that when checking (i), it is necessary to ensure that (ii) in the language considered does not involve a base-generated topic (something like “as for expensive (things), John likes expensive cars”) and/or NP ellipsis in the sentence-initial constituent (something like “as for expensive cars, John likes expensive cars”, where ‘likes’ or ‘John’ would likely be focalized). The most straightforward test to control for this would involve introducing an island between *expensive* and *cars*.)

- (i) Left-branch extraction of adjectives (and adjectival-like elements), as in (ii), may be allowed only in languages without definite articles.
- (ii) *Expensive_i, Mary sells [_{t_i} houses]

(see Boas 1911), meant by it, which is not true. One of the reasons why it is not true is rather simple: for TTs, the stance in question arose at least in part as a reaction to some of the generativists' views discussed below, which American structuralists obviously could not have reacted to. Furthermore, there are misconceptions in the TT camp regarding the generativists' views in question. In other words, the negative reaction of generativists regarding the stance in question is (at least in part) based on misconceptions regarding what TTs mean by this view, and the TT view in question is in turn (at least in part) based on misconceptions regarding certain generativists' views.

To start untangling the cobweb of misconceptions concerning the generativist's reaction to the view in question, it does not seem that the practitioners of the TT camp truly believe it. The typologists from that camp have made incredibly important contributions to the field at large in terms of Greenberg-style generalizations, which the practitioners of the generative camp are increasingly relying on. Reaching such generalizations would not have been possible if they truly believed the slogan in question. American structuralists did believe it, but as a result, they also did not engage in typological work (see Greenberg 1974). A number of typologists have actually attempted to demonstrate that typological work is still possible while adhering to the slogan in question. For brevity, I will focus on one such work, Haspelmath (2010), one of the reasons being that the generativists who do look into the issue seem to take it to be a TT cannon, i.e. to reflect the general state of affairs in the TT camp (which actually is not true, as we will see). Haspelmath (2010) attempts to demonstrate that typological work is possible while adhering to the slogan in question regarding grammatical categories.⁴ However, he also provides

⁴ The following quote from Haspelmath (2020) indicates, however, that he does not take the slogan in question as literally as generativists assume that traditional functionalists/typologists in general take it (in fact, generativists would likely label the underlined part below UG – as discussed in section 4, there is actually much less disagreement between the two camps regarding the notion of “UG” than what is widely assumed): “First, language description is true to the categories of each language, but is inspired by the accumulated knowledge of comparative linguistics” (Haspelmath 2020:14).

universal definitions of the relevant categories, which are applicable to all languages and which make typological work possible.

He treats them as artificial linguistic constructs and not real (in fact not part of the grammar of individual languages), but there is really no deep reason (and, more importantly, nothing in the general TT worldview) why they should not be considered real. Haspelmath makes a distinction between a language particular descriptive category, call it X, and its crosslinguistically applicable comparative concept, call it Y (which is used in typological generalizations). But there does not seem to be a real issue here – it is possible that in some language, X is exactly like Y, while in another language, where this is not the case, we have a more complex situation where $X=Y+Z$; so there is still Y in that language as well. As an illustration, consider the typological generalization in (1a) and the definition of the relevant element in (1b), an example of Y.

- (1) a. GENERALIZATION: In all languages, markers of future tense are less bound than markers of present tense or past tense, or equally bound, but never more so.
- b. DEFINITION: A future tense is a grammatical marker associated with the verb that has future time reference as one prominent meaning. (Haspelmath 2010: 671)

To make his point regarding X and Y, Haspelmath observes crosslinguistic differences regarding future tense, e.g. in Spanish it is also used to express probability (but not habituality), while in Lezgian it is also used to express habituality (but not probability). This shows future tenses are not synonymous crosslinguistically, which then necessitates making a distinction between a language particular descriptive category (X) and its crosslinguistically applicable comparative concept (Y). But what we really have here is the more complex $X=Y+Z$ situation, where Y is still always present. Haspelmath gives similar definitions of other concepts (e.g. question words and ergative case), with similar crosslinguistic differences, all of which instantiate the $X=Y+Z$ situation. Haspelmath considers Ys to be concepts created by linguists for the purpose of formulating

typological generalizations. While they are applicable to all languages, they are supposed to be artificial, i.e. not psychologically real and not part of particular language systems. But there is no real reason why those Ys could not be real (and in fact part of UG from the perspective of a formalist; note I am putting aside here the question of what the real primitives of UG in the relevant domain are, which is irrelevant to the general point made here). In fact, a number of TT works have expressed this view, see e.g. Gill (2016), Lander and Arkadiev (2016), Round and Corbett (2020) (for a criticism of Haspelmath's position in question, see especially Spike 2020). In some languages those abstract categories would happen to map straightforwardly to surface categories, and in others that would not be the case: in such a case we could have the $X=Y+Z$ situation (with Y applicable to all languages). This kind situation would become more obvious if it is accepted that the grammar of each language that is studied in its own terms is, as Baker (2015: 936) puts it, "abstract to some non-trivial degree" (which is what generativists generally accept) – this would result in more $Y=X$ situations and more generally make the Y-X relationship more transparent.

At any rate, the relevant concepts can be defined differently for each language, or in a way that would at the same time make them universally applicable (which would be more abstract; abstract does not need to mean not real and artificial⁵). The latter is anyway needed for typological work, which makes Occam's razor ('use what must be there as much as possible so as to avoid positing additional things') pertinent here. As Haspelmath (2010) observes, a number of non-generative typologists refute the view that every language must be described in its own terms (e.g. Dahl 1985, Bybee and Dahl 1989, Lehmann 1989). In this respect, Lehmann 1989: 142 says: "Describe your language in such a way that the maxim of your description

⁵ There is nothing strange in what is more abstract being psychologically real. Consider e.g. the concepts of allophones and phonemes. In a typical case of allophonic variation, what is psychologically real is the abstract phonemic level, which actually does not correspond to anything that is physically real, since the phoneme will always be physically realized as one of its allophones.

could serve, at the same time, as the principle of general comparative grammar – and thus, the maxim of description of any other language.” This will give us fewer mechanisms, which reflects Occam’s razor as a general scientific principle (in addition to being a prerequisite for doing crosslinguistic typological work). Haspelmath’s position seems to be a result of accepting a certain level of abstractness in doing typological work but not in doing analyses of individual languages, which essentially leads to separating the two into different fields (as Haspelmath 2010: 682 puts it, “the analysis of particular languages and the comparison of languages are thus independent of each other as theoretical enterprises”). Allowing the same level of abstractness for both, which would also be in the spirit of Occam’s razor, would, however, dissolve this distinction (see also Round and Corbett 2020).

There are other typologists with positions similar to Haspelmath’s (see especially Dyer 1997, Croft 2001, who antecede Haspelmath’s work). The above discussion would extend to them. In fact, generative typology does not really differ from Haspelmath’s position in that works in this tradition also essentially assume what I have referred to as X and Y above (so there is really no disagreement here), the only difference being that Y, which Haspelmath considers a linguist’s construct, is treated as real and in fact part of UG. Thus, the same point that was illustrated with (1) can be illustrated with any of Bošković’s (2008, 2012) generalizations regarding definite articles (see e.g. (i) in fn3, which is the counterpart of (1a)), and Bošković’s definition of *definite article* (which superficially shows similar variation across languages as future tense) in Bošković (2016b), which is stated in semantic terms (as is the case with many of Haspelmath’s Ys).

Haspelmath’s explicit distinction between what was referred to as X and Y above is nevertheless a welcome and useful warning that should be heeded; those Ys that both traditional typologists and generative typologists are using are very often not quite the same as Xs used in individual languages, a difference which does get overlooked, especially by the latter.

At any rate, this is a case where there is less disagreement between traditional and formal typologists in practice than what is generally

assumed; note that the main point is actually methodological – assuming the X/Y distinction is necessary methodologically to be able to do typological work, whether those Xs are real or not is a separate issue (where in fact there is no full field vs field disagreement; as Round and Corbett (2020) (see also Spike 2020) put it, this is an issue of a more general philosophical understanding of science, which is independent of the two approaches to language discussed here.

There is another aspect of the describe-every-language-in-its-own-right view which should be taken by the generativists as a methodological warning, be careful before jumping to conclusions that something is in UG – it's a warning not to follow without further checking a detailed investigation of a single language with a proclamation that it is all UG. Unfortunately, this tendency is still there among the generativists to some extent – I am not talking here about very abstract properties like investigations of e.g. c-command and domination where a single language can be used as an illustration basically for ease of exposition (see Epstein 1999) – I am not aware of any language where the notions of c-command/domination (which essentially means structure) do not hold, but detailed investigations of the structure of a single language which is immediately followed by a proclamation that all languages are like that without even checking the language next door (or simply by forcing other languages into the mold set up by that detailed investigation of one language without seriously looking at what does not fit).⁶ From this perspective, as a generativist, I understand and am sympathetic to the mantra look-at-languages-in-their-own-right.

At any rate, the view endorsed here is that there are a lot of similarities across languages, but also a lot of differences – the quest

⁶ There has been strong emphasis on investigations of understudied languages in the generativist camp in recent years. However, Germanic and Romance to some extent still hold a privileged place when it comes to UG proclamations of the kind noted above, with understudied languages often being used to confirm those proclamations (as a result of which they are often used as mold fitters) – there are likely political and sociological reasons behind this but discussing them is beyond the scope of this paper (see Bošković 2021b).

for the former should not ignore the latter (and the other way round). While there are extremes in both camps (those who in an American-structuralist style overemphasize and truly believe the mantra look at languages in their own right, which leads to missing crosslinguistic similarities, and those who overdo it in the opposite direction by ignoring crosslinguistic differences, forcing all languages into one of those 'privileged' languages from fn6), the majority of both TTs and GTs seem to hold the view expressed above, regarding TTs, see e.g. Gill (2016: 458–459), which underscores the lack of a fundamental difference between them in this respect.

The look-at-languages-in-their-own-right mantra is used by TTs as a reaction to what they seem to take to be an assumption held by everyone in the generative camp, in fact one of the defining beliefs of that camp, which is that there is a universal sentence structure holding for all languages, the underlying assumption being that if we were to look at languages in their own right it would become clear that there is no such thing.⁷ But there is a misconception here regarding the actual state of affairs within the GT camp (in fact, the situation here is similar to the misconception that the generativists have regarding the look-at-languages-in-their-own-right position). While it is true that the universal structure claim is often made in isolation by the generativists, in actual research practice it is often given up. In fact, the most radical departures come from generative typology works, where the claim in question is argued against on typological grounds (which can actually be interpreted as a point of convergence between the two

⁷ See in this respect Gill (2016: 458): “Many of us have developed our views of language at least in part in reaction to a dominant and sometimes domineering universalist approach that attempts to impose an aprioristic set of universal categories on languages with respect to which the categories in question are completely irrelevant. Our response was to reject such universal categories, while returning to the old American structuralists’ ideal of describing each language on its own terms (Boas 1911).” Much of what is referred to as (universal) grammatical categories in discussions of this kind in TT is framed in terms of (universal) clausal structure in generative literature. The discussion below will be stated in these terms.

approaches to typology). E.g., Todorović (2016), a typological study within the formalist tradition, argues against the universal presence of TensePhrase – in particular, she argues for a broad typological distinction between languages with and without Tense (following a suggestion from Bošković 2012), correlating this distinction with a number of properties).⁸ Similarly, Bošković (2008, 2012) gives a number of crosslinguistic generalizations where languages with and without definite articles (henceforth with/without articles) are shown to differ regarding numerous syntactic and semantic properties, which cannot be accounted for if the distinction between languages with and without articles is simply a matter of phonology, namely whether articles are overt or null.⁹ Based on this, Bošković argues that languages without definite articles do not simply have a null definite article, they lack it altogether, they in fact lack the DP projection (which is the only possibility for the structural placement of the definite article; other D/DP-related elements from a language like English can be located in other projections, in which case they show different behavior from the corresponding elements in English, see

⁸ Todorović also shows that the labels that traditional grammars use, and which TTs often rely on, can be very misleading; thus, she shows that what is traditionally called Aorist and Imperfectum tenses in Bulgarian and Serbo-Croatian (SC) are actually very different things, in fact not even the same categories – in Bulgarian these are indeed tenses, while in SC they actually represent aspect.

⁹ See fn3 for one such generalization; three more are given in (i).

- (i) a. Second-position clitic systems are found only in languages without articles.
- b. Only languages without articles may allow scrambling (section 1.3).
- c. Only languages with articles may allow clitic doubling.

The works in question give a number of other generalizations, concerning phenomena as varied as interpretation of superlatives, negative raising (see section 3), sequence-of-tense, pro-drop in the absence of agreement, head-internal relatives, polysynthesis, multiple wh-fronting, possessives, numeral classifiers, subject reflexives, number morphology, scope, negative constituents, adjunct extraction, and focalization (see also fn11 for additional semantic arguments).

Bošković 2012¹⁰). This position argues against the universal structure hypothesis, where we would expect all languages without articles to have phonologically null articles (it would simply be a phonological accident that articles are unpronounced in some languages). Now, variation regarding the presence of DP is not a universally adopted assumption within the generativist camp. This is actually tied to the issue noted in fn6: fundamental proposals of this sort are generally made on the basis of Romance and Germanic. Consequently, most of the time those who adopt the universal DP assumption (which comes with a “phonological accident”) adopt it with no further discussion (after all, Romance and Germanic have DP, so all languages must have it), or attempt to fit other languages into the Romance/Germanic DP mold while ignoring relevant differences, i.e. ignoring what does not fit.¹¹ Even worse, they do it at a rather significant cost. As discussed in Bošković (2012), extraction patterns out of the nominal domain are completely different in languages with and without articles (for a partial illustration, see (i) in fn3 and section 3.2). Locality restrictions on movement are currently stated within the phase theory. There are two mechanisms that can be used to capture crosslinguistic variation regarding extraction of the kind that is found in the nominal domain: assuming structural differences (as in the NP/DP proposal) or variation in the locality system, i.e. phases. The latter concerns variation within the computational system itself; the former, on the other hand, can be easily stated in lexical terms (in terms of a particular feature,

¹⁰ See also Fukui (1988), Corver (1992), Zlatic (1997), Chierchia (1998), Cheng & Sybesma (1999), Layons (1999), Willim (2000), Baker (1996), Trenkić (2004), Despić (2013b), Marelj (2011), Takahashi (2011), Jiang (2012), Talić (2013), Cheng (2013), Runić (2014), Kang (2014), Bošković & Şener (2014), Zanon (2015), among others, for no-DP analyses of individual languages without articles.

¹¹ The universal DP literature often cites Progovac (1995), who argued for a DP in SC based on certain alleged parallelisms in word order between SC and Italian, completely ignoring the fact that Despić (2011, 2013a) has subsequently quite conclusively shown that these parallelisms do not hold at all (for similar situations, see Franks 2019 regarding DP claims in La Terza 2016, and Bošković 2009 regarding Pereltsvaig 2007).

+definiteness). Most generativists assume that there is no variation in the computational system itself, which means that there should be no variation regarding phases – all variation should be tied to lexical properties (e.g. Borer 1984, Boeckx 2008). The issue here is that those who assume universal nominal structure would generally also assume that there should be no variation regarding phases (i.e. the computational system). But it is simply not possible to assume both (unless we ignore differences in extraction patterns). Either we have variation in the structure (structural variation) or in the locality system (which means the computational system) – something has to give. (The universal DP analysis is often seen as having an appealing universalist character, but that is actually not true: it leads to a non-universal locality, i.e. phasal, system.) Adopting universal structure thus has a rather significant consequence, which those who adopt it don't seem to be aware of (since they generally do not discuss the issue).

Another point is worth noting. It's often assumed that there is a universal structural hierarchy (referred to as functional sequence). Take the abstract structure in (2) to represent it.

$$(2) \quad [_{XP} [_{YP} [_{ZP} [_{KP}$$

The standard universalist approach to the structural sequence is not that the whole sequence is always projected (CP e.g. is not projected in raising infinitives like *she seems to like me*; even a more drastic case of non-projection of full clausal structure concerns restructuring infinitives, which are even smaller, see e.g. Wurmbrand 2001, Cinque 2004). Rather, the standard universalist approach (even this is actually not really widely accepted) is that the structural hierarchy in (2) is respected: (a) there cannot be a language where YP is higher than XP and (b) it is not possible to project KP, ZP, and XP without projecting YP. There is nothing in the NP/DP approach that goes against either of these. In fact, there is nothing in the NP/DP variation approach that goes against anything in the standard assumptions regarding sentence structure: requiring that every nominal domain must project to DP would not be any different from requiring that every clause be

a CP, including, e.g., restructuring infinitives, which are standardly assumed not to be CPs. The works which argue against the NP/DP approach on universalist grounds thus seem to be based on very non-standard assumptions regarding these universalist grounds.¹²

The main point of the above discussion is that in this particular case (universal structure) the difference between the two camps is smaller than it is assumed to be. TTs are concerned with the structural issue in question (although they don't state it in these terms); they are generally not concerned with theoretical issues pertaining to the

¹² Syntacticians occasionally assume DP is necessary for semantic reasons, to be able to interpret a nominal as an argument. Curiously, this is not what a semanticist would assume, see e.g. Chierchia (1998) and Dayal (2004), who crucially adopt the NP/DP distinction. In fact, there are rather strong semantic arguments against the universal-DP Hypothesis. The hypothesis assumes that the only difference between a language like English and a language like SC regarding articles is phonological: SC simply has a null *the*. A number of Bošković's generalizations actually concern semantic phenomena, which shows that this simply cannot be right. Consider also Jenks's (2018) and Despić's (2019) arguments regarding the anaphoric use of nouns. Jenks observes that a bare noun cannot be used in Mandarin in donkey anaphora contexts like (i), which is surprising if Mandarin has a definite article just like English, which just happens to be phonologically null (Mandarin requires a demonstrative on the anaphoric/bound reading of "donkey").

(i) Every farmer that has a donkey_i beats the donkey_i.

Consider also Despić (2019) on the anaphoric use of mass nouns, illustrated by (ii). SC (iii) cannot have the meaning English (ii) has, with *fruit* anteceded by *grapes* (that reading requires a demonstrative). Despić shows this is a more general difference between languages with and without articles, also noting that if the latter had a definite article, which would just happen to be phonologically null (so the only difference would be in phonology), this would be totally unexpected. The conclusion is that the difference between the two language types is deeper – it's not a matter of phonology but syntax and semantics – there is no null *the*/DP in languages without articles in the counterparts of constructions where languages like English use it.

(ii) We have been growing grapes for generations – and you know, we have made millions on the fruit.

(iii) Naše mesto već generacijama proizvodi belo grožđe. Sve dugujemo voću.
our town already generations produces white grape everything owe fruit-dat.
'Our town has been producing white grape for generations. We owe everything to that fruit.'

locality of movement. They generally assume that there is no universal structure (the look-at-languages-in-their-own-right mantra being a reflection of that stand) and that they differ in this respect from GTs, but they are actually not aware of the full range of views among the generativists. Many generativists who espouse the universal structure view are in turn unaware of the full range of consequences of that view (most of them would not want the computational/locality system parameterized, but they seem to be unaware that this is a consequence of the universal structure view). In this respect, it's worth noting that Cinque (1999) provides evidence that different adverbials are located in different projections, with more than 30 such projections in a fixed hierarchy partially given in (3).

- (3) [*frankly* Mood_{speech act} [*fortunately* Mood_{evaluative} [*allegedly* Mood_{evidential} [*probably* Mod_{epistemic} [*once* T(Past) [*then* T(Future) [*perhaps* Mood_{irrealis} [*necessarily* Mod_{necessity} [*possibly* Mod_{possibility} [*usually* Asp_{habitual} [*again* Asp_{repetitive(I)} [*often* Asp_{frequentative(I)} [*intentionally* Mod_{volitional} [*quickly* Asp_{accelerative(1)} [*already* T(Anterior) [*no longer* Asp_{terminative}...

The full universal structure view, which is what Cinque (1999) espouses (and which the functional sequence hypothesis discussed above would require) is that all these projections are present even in both clauses of *I said that he left*, a sentence with no adverbials, in any language. While Cinque does assume they are all present, this is not a widely held assumption (see also (4)). It is thus fair to say that the assumption that all projections from (3) are present in every clause, which is what the all-structure-is-universal hypothesis would entail, is not uniformly accepted in the GT camp.

The universal structure proclamation is still often made by generativists, but not really adhered to (which means not truly believed), which is especially clear regarding works on the left periphery of the sentence.¹³ To give some illustrations, most accounts

¹³ There is a fallacy of universality syndrome among generativists, where the word is sometimes jumped at and adopted without really thinking about it (or with

of the voiding of the *that*-trace effect in (4b) (e.g. Bošković 2016a, Erlewine 2020) would actually also void it in (4a) if the projection where the adverb is located in (4b) is also present in (4a).

- (4) a. *Who do you think that would leave Mary?
b. Who do you think that under no circumstances would leave Mary?

Erlewine (2016) shows that in Kaqchikel, where wh-phrases and indefinites have the same form (a common crosslinguistic pattern, see Haspelmath 1997), the relevant elements are fronted to the specific projections in the left periphery on both functions, with the first element interpreted as a wh-phrase and the second as an indefinite pronoun when two of them are present. It is pretty clear that indefinites cannot be undergoing this kind of fronting universally (either overtly or covertly), assuming this (even covertly) would create havoc regarding e.g. scopal interpretation of indefinites.

A similar non-universality conclusion follows from the works on languages that front all wh-phrases. Thus, Rudin (1988) and Bošković (2002) show that such multiple wh-fronting languages (MWF) differ regarding the landing site of MWF; the highest clausal projection (CP) or lower than that. Thus, despite superficial similarity, there are numerous differences between (5) and (6), e.g. regarding the penetrability of fronted wh-phrases, their ordering, the availability of single-pair answers, inversion, the possibility of fronted wh-phrases following subjects, which can all be accounted if SC MWF lands lower than Bulgarian MWF (see Bošković 2002).

- (5) Koj kakvo kupuva?
who what buys (Bulgarian)
(6) Ko šta kupuje?
who what buys (SC)

clear arguments to the contrary ignored, see e.g. fn11). Obviously, not everything is universal. But the fallacy of universality syndrome sometimes leads to pulling out the “universal” card too quickly, without thinking about its consequences in the particular context in which it is used.

If the lower and the higher wh-fronting projections were always universally present, given the standardly assumed shortest move requirement, wh-fronting would always have to go to the lower wh-fronting projection (furthermore, due to the well-known and standardly assumed freezing effect (see e.g. Rizzi 2006), further movement from this projection would not be possible). It follows then that Bulgarian, in fact any language that fronts wh-phrases to the same projection as Bulgarian (this is also what English does, the only difference being that English fronts only one wh-phrase) cannot have the lower wh-fronting projection that SC has.

Or consider Russian (7a). Russian has been argued to have a high NegPhrase, where negation in this high NegP has semantic effects but does not have the true meaning of negation, as the translation shows. This negation also does not license negative concord (Brown and Franks 1995). It is clear that we don't want this NegP to be present in English (where negation always means negation) or other negative-concord languages, e.g. Serbo-Croatian, where negation always licenses negative concord. (Zanon 2020 notes that (7a) would be used in a context where John promised to stop by at some point this week but did not specify the day; *Didn't John stop by today* is not possible in this context in English; instead, *Did John stop by today* would be used.)

- (7) a. Ne zaxodil li Sergey segodnja? (Zanon 2020)
 neg stop.by Q Sergey today
 'Did Sergey stop by today by any chance?'
 b. Zaxodil li Sergey segodnja?
 'Did Sergey stop by today?'

Rizzi (1997) argues that the traditional CP should be split into a number of projections, shown in (8), primarily based on Italian.

- (8) [_{ForceP} [_{TopP*} [_{IntP} [_{FocP} [_{TopP*} [_{FinP} [_{IP}]]]]]]]]]

While (8) is often cited as a universal structure holding for all languages, there is evidence even internal to Italian that all the structure from (8) cannot always be present, see in this respect the anaphor binding data from Petrossino (2018). Furthermore, Abels

(2003) provides an account of the general immobility of IPs dominated by CP (which holds crosslinguistically), illustrated by (9), where it is crucial that the CP here is not split at all.

(9) *_{[IP}His mother left]_i everyone believes [_{CP}that t_i]

It thus seems clear that CP cannot be always or uniformly split in either Kaqchikel, Russian, English, or MWF languages – there is no uniform split CP field that is present either crosslinguistically or in all constructions of a single language (Rizzi 1997: 314–315 actually acknowledges this possibility; in fact, the facts discussed above indicate that even the version of the universal structural hierarchy where it's not possible to project KP, ZP, and XP in (2) without projecting YP cannot be right). The above illustrations are really just the tip of the iceberg. There is a great deal of crosslinguistic variation regarding left periphery which pretty strongly argues against structural universality of the left periphery.

At any rate, the point here is that there is a plurality of views within the generative camp regarding the notion of universal structure, a proclamation that is often made but not really adhered to, the look-at-languages-in-their-own-right TT stand being (at least in part) a reaction to that notion, as a result of which there is actually less disagreement here than what is believed within the TT camp. It should also be noted that many grammatical categories whose universality is questioned in TT works in what is taken to be disagreement with generativism are not taken to be universal, or even real at all, in generative works (this e.g. includes the notion of *subject*).¹⁴

¹⁴ The issue here is that what TTs have been reacting to with the stance in question may have been true (to some extent) of the past research in the generative paradigm, but this is no longer the case (the development of GT did contribute to this). In fact, the current state of the field within generativism encourages investigation of crosslinguistic differences (contrary to the perception among TTs that it “actively discourages the investigation of such differences” (Gill 2016: 459)) as well as investigation of understudied languages to the point that it is almost a must on the job market (field methods classes are also becoming a must in the curricula of generative departments; in my department they are regularly offered, the last one was on Mandinka). So the reaction is still there, but what is

In fact, just like there is a plurality of views regarding universality in the case discussed above within the generative camp, there actually is also a plurality of views regarding non-universality in the TT camp. The quote from Gill (2016) from fn7 continues as follows: “However, some of us have gone beyond rejecting specific proposals for universal categories, such as subject, adjective, or whatever, and allowed our prejudices against such categories to lead us to deny the very possibility of universal categories. It is this latter move that seems to me to be an unwarranted overreaction... I have been outspoken against the Eurocentrically-motivated imposition of universal categories such as noun, verb, and their various phrasal projections on languages that offer no evidence for their presence. However, it does not follow from this that universal syntactic categories do not exist; it’s just a question of choosing the right ones... while pursuing linguistic diversity, it is important not to lose sight also of the ways in which languages may resemble each other, and of the possibility that all languages may embody a fundamental unity.” Gill (2016) is certainly not unique in this view within the TT camp, see e.g. Round and Corbett (2020) and Lander and Akadiev (2020) (the two issues of *Linguistic Typology*, 20.2 and 24.3, are actually very useful starting points for generativists who would like to familiarize themselves with the broader range of views regarding the issues under consideration within the TT camp).

There is a perceived all-or-nothing/either-or difference between the two camps – it’s all universal or nothing is universal. Note we are talking here about perception of X by Y, not the actual state of affairs in X (where X/Y stand for the two camps). In reality, what we are dealing with here is a matter of degree, i.e. how much is universal. This is very different from what’s perceived. Different TTs and different GTs differ regarding the exact degree, but this is a very different situation from an all-or-nothing difference that would hold across the two fields, with everyone in completely opposite corners, which would not leave any room for common ground or an opening for a dialogue, since one side would have to be 100% wrong. The degree difference in fact

being reacted to is actually no longer there (except in the slogans, as discussed above).

opens the door for what should be a productive dialogue regarding how much, and what exactly, is universal.

At any rate, the upshot of the above discussion is that the two camps are really unaware of the full range of views within the respective camps regarding the issues/slogans discussed in this section, and what is really behind them, which leads to the impression that there is more disagreement between the two camps than there really is.

There is a difference here between what is proclaimed and what is truly believed (as shown by the actual research practice): the generativists react negatively to the every-language-in-its-own-right mantra since they take it at face value (taken as such, it does make comparative work impossible and reflects a no-UG attitude (an issue I return to below)); while it was intended to be taken at face value by American structuralists, this is not the case with the practitioners of TT who adopt it nowadays; they in turn use this mantra partly in reaction to a particular bad practice of the generativists (pulling the UG card too easily when examining details of the structure of a particular language) and in reaction to a universal structure proclamation that the generativists make (they in fact also always make it in reaction to the every-language-in-its-own-right stand), though they do not really believe it, as revealed by the actual research practice.

In other words, much of perceived disagreement comes from misinterpretations of pragmatically motivated slogans, where the two sides react negatively to what the other side is saying because they don't realize that what is said does not straightforwardly reflect what is really believed. Just like the TTs don't really believe in the mantra under discussion in this section (otherwise they would not be engaging in typological work), the generativists don't really believe in the universal structure slogan (which is easy to show by looking at the actual research practice).

1.2. SYNTAX AS A TOOL

Another widely assumed difference between functionalists and formalists concerns their stands regarding the role of syntax. The

perception of the difference is so significant here that there are functionalist works where simply showing that something is a semantic or pragmatic (rather than syntactic) phenomenon is taken to argue against the generative approach in general. Functionalists generally rely on much more impoverished syntax than generativists. The reason for this is mostly methodological, which in turn concerns their primary goal: for them pragmatics (and semantics) is more basic than syntax; they often look at syntax simply as a tool for expressing pragmatic functions and semantic roles – as a result, they generally do not consider syntactic relations that go beyond the tool role of syntax (they also generally do not consider what is not possible, since their goal is to determine how to capture what is possible, i.e. how to express the needed pragmatic and semantic notions; this is in fact something they have in common with the semantic approaches that generative syntacticians rely on, a point I return to). A generativist pursues a different methodology here, which is again connected to their primary goal that in a way gives primacy to syntax: A generativist is interested in examining the full complexity of syntactic relations, unbounded by the tool role of syntax; they are interested in uncovering syntactic principles that determine well-formed and ill-formed sentences – pragmatics and semantics take the former and assign them interpretation/pragmatic use (for some relevant discussion of formalism vs functionalism in this context, see also Baker 2015: 21). Importantly, there is really no deep worldview difference here. The approaches to semantics and pragmatics generative syntacticians rely on also assume much poorer syntax than generative syntacticians do and are also not concerned with what is not possible syntactically. But the reason for this difference is simple: they investigate semantics and pragmatics, not syntax – they go into syntax only to the extent it is relevant to their concerns. Due to the nature of their inquiry, syntax is just a tool for them (and the same holds for syntacticians when it comes to semantics and pragmatics). There is no fundamental difference in worldview here, they just do different things. The same in fact holds for functionalists and generative syntacticians in this respect. What is taken to be a deep-seated difference in the worldview is actually just

a byproduct of them doing different things – the difference here is very similar to the difference between generative syntacticians and the approaches to semantics/pragmatics that generative syntacticians rely on. As a result, there is really no deep reason why many of the results reached in functionalist works could not be incorporated into generative works. This is not happening in practice due to the pervasive perception that the two approaches are so fundamentally incompatible that the practitioners of the two approaches generally do not read each other’s works, even when examining the same topic. They are in fact not incompatible, to a large extent they are complementary (just like generative syntax and the approaches to semantics/pragmatics that generative syntacticians rely on).¹⁵

Such complementarity can be easily illustrated. Consider the phenomenon of ergativity. There are numerous syntactic differences between a verb like *work* and a verb like *arrive* crosslinguistically, which in the generative approach (in the Principles and Parameters tradition) are accounted for by having *Mary* start the derivation in different positions in (10) and (11).

(10) $\text{Mary}_i \text{ [}_{\text{VP}} \text{ t}_i \text{ works]}$

(11) $\text{Mary}_i \text{ [}_{\text{VP}} \text{ arrived t}_i \text{]}$

A functionalist (e.g. DeLancey 2001) would complain that these structures do not explain why *work* and *arrive* differ in the relevant respect. This is certainly a valid complaint, and the ultimate explanation will likely not be syntactic – it may very well turn out to involve cognitive or communicative factors (for some discussion relevant to these issues, see Levin & Rappaport Hovav 1995, Kuno & Takami 2004). But that would not invalidate all the structural/syntactic reflexes of the *work/arrive* distinction that hold across a variety of

¹⁵ Of course, sometimes it is not clear whether a particular phenomenon should receive a functional or a formal explanation, just like sometimes it is not clear whether a particular phenomenon should receive a syntactic or a semantic explanation – there is no deep incompatibility here.

different phenomena crosslinguistically,¹⁶ and which the derivations in (10) and (11) unify. Providing a non-syntactic explanation for why *work* and *arrive* differ in the relevant respect can be complementary to the syntactic differences that generative syntax has uncovered in this respect.

We are dealing with a broader issue here: functionalists often raise very valid “why” questions which, when taken seriously, indicate we need more than just syntax (even in the broad sense the generativists understand it) for particular phenomena, but the non-syntactic answers to those why questions very often can be added to the syntactic accounts, which would then give us better, more comprehensive accounts of the relevant phenomena. (Instead, the functionalists often interpret we-need-more-than-syntax as we don’t need syntax at all, and then ignore the syntactic part; the generativists, on the other hand, should be faulted for not raising, and missing, those why questions (which includes ignoring possible functional answers to those questions)).

In the next section I will discuss the status of Greenberg-style typological generalizations regarding the concept of Universal Grammar. From a formalist perspective, at the right level of abstractness that also dissolves exceptions to them (see below) Greenberg-style generalizations reflect UG at work – this has in fact prompted the development of generative typology; from this perspective the practitioners of the traditional functionalist/typological camp have contributed a great deal to the formalist’s understanding of UG.

¹⁶ To mention a few, *ne*-cliticization in Italian and genitive of negation and *po*-phrases in Russian (for an overview, see Alexiadou, Anagnostopoulou, and Everaert 2004).

2. GREENBERG-STYLE GENERALIZATIONS AND UNIVERSAL GRAMMAR

2.1. ON THE STATUS OF TYPOLOGICAL GENERALIZATIONS

Above I have discussed Haspelmath's distinction between language particular descriptive category and its crosslinguistically applicable comparative concept, observing that the distinction is also adopted in generative typology, though with a difference regarding the status of the latter, which concerns the notion of UG. The notion is supposed to represent a significant difference between traditional and generative typology. However, we will see in this section (and section 4) that the difference regarding the notion of UG may also be smaller than what is generally assumed (i.e. there may not be real fundamental disagreement even here).

In fact, in many respects, again in actual practice, not the slogans associated with the respective approaches, the practitioners of the functionalist/typological camp seem to be bigger believers in universal grammar (see also sec. 4), and have contributed more to the notion (although they may deny it for reasons discussed below) than many generativists (I will refer to the two camps below as α for the former and β for the latter, strictly for expository reasons). The goal of many practitioners of the former is to use detailed investigations of individual languages to reach broad Greenberg-style crosslinguistic generalizations, while many practitioners of the latter use them (generally an investigation of an understudied language in this case)¹⁷ to argue against proposed crosslinguistic generalizations. In doing so, the former, who are generally anti-Chomskian, do what they often

¹⁷ While there has been a surge in the work on understudied languages within the generative approach, when theoretical issues are discussed, such work is mostly done either to confirm broader theoretical proposals made with respect to more widely studied languages like Germanic and Romance (see fn6) or to argue against proposed crosslinguistic generalizations and/or theoretical proposals (so they are often used either to confirm or disconfirm proposals made based on Germanic, Romance...), they are seldom used to make such proposals (there are of course exceptions, like Baker's work).

accuse Chomsky of doing, and the latter, who are broadly classified as Chomskians, do what Chomsky himself would never do: in order to make sense out of what seem to be chaotic data, to be able to see patterns, parts of the chaotic data, sometimes even good chunks of it, have to be put aside. Chomsky's work is full of such examples, but this is simply the way science works, this is what is done in any mature scientific discipline. To reach those Greenberg-style generalizations, the α practitioner does exactly that, those generativists who attempt to knock off proposed crosslinguistic generalizations based on a single counterexample from an understudied language, which increases the likelihood that something has been misanalysed, do exactly the opposite.

Now, there is a reason why α practitioners would deny the label I have given them, 'believers in universal grammar' (see also sec. 4). α and β practitioners read Greenberg's generalizations very differently: here are some examples of Greenberg's generalizations, with the relevant parts bolded:

- (12) a. When the descriptive adjective precedes the noun, the demonstrative and the numeral, **with overwhelmingly more than chance frequency**, do likewise. (#18)
- b. Where morphemes of both number and case are present and both follow or both precede the noun base, the expression of number **almost always** comes between the noun base and the expression of case. (#39)

To α , the bolded part means "exceptions". Universal Grammar is supposed to mean no exceptions, so this is not UG, in fact it argues against UG – even when we come close to it, it is not that.¹⁸ β , a

¹⁸ While many of Greenberg's original generalizations are stated as if they have no exceptions, the current stand on language universals within TT seems to be that none are exceptionless – they all have something like the bolded part from (12); see e.g. Bickel 2007, Nichols 2007; this, however, has not been actually shown for all of them. The reason for at least some of the exceptions may be the implicational "If-X-then-Y" form in which they are stated – exceptions may disappear if additional conditions in the *if*-clause are added (as noted by Baker

Chomskian UG practitioner, ignores the bolded part; we then have universal generalizations here – examples of UG. (It's an interesting switch regarding the normal scientific methodology of putting exceptions aside until they can be better understood, α -s who pursue it to reach generalizations like (12) now drop it, those β -s who knock off proposed generalizations by ignoring it (i.e. by pointing out exceptions to them) now endorse it).

But they are really both right and wrong. (12) is and is not UG.¹⁹ To β , what is supposed to be in UG is not generalizations like (12) but mechanisms that deduce them. In other words, generalizations like (12) would be theorems, not principles of UG. Very often, deductions of principles leave room for exceptions. The right deduction of (12), based on the mechanisms present in UG, should then leave room for the exceptions. Moreover, it should explain why they are rare.²⁰ This

& McCloskey 2007: 288), as in if X and Z then Y, or if they are stated as clear one-way correlations of the form if X then no Y. Consider in this respect the LBE generalization from fn3. If stated as in (i), it has exceptions, e.g. Japanese and Chinese, which lack definite articles but disallow LBE. However, it turns out agreement is also necessary for LBE (see Bošković 2012; thus, Serbo-Croatian has both agreeing and non-agreeing adjectives, only the former allow LBE). The statement in (ii), of the form if X and Z then Y, then takes care of Japanese and Chinese. They can also be taken care of with the weaker statement (which emphasizes what is disallowed, not what is allowed) in (iii), which is of the form if X then no Y (I am actually not aware of any exceptions to (iii)).

(i) If a language lacks definite articles it allows adjectival left-branch extraction (LBE).

(ii) If a language lacks definite articles and has agreeing adjectives, it allows LBE of such adjectives.

(iii) If a language has definite articles then it does not allow adjectival LBE.

¹⁹ Baker (2011) raises the question of the relationship between the β notion of UG and Greenberg-style language universals. The following discussion provides a partial answer to this question. Note also that I will not be concerned here with actual deductions of (12) (though see Cinque 2005 for relevant discussion of (12a) and Harley & Ritter 2002 for (12b)).

²⁰ Formal explanations for why something is rare but possible are in principle possible – see e.g. Baker & McCloskey (2007) on the rarity of VSO languages (the reason being that a constellation of factors is needed for that type; for a different

is what I mean by (12) is and is not UG. (12) is in UG, including the bolded part, but as a theorem. **But:** this is the case if (12) is deducible from the formal mechanisms of UG. A priori, we don't know: there could be formal explanations for (12a-b), or functional, or formal-as-a-reflection-of-functional-considerations explanations of the kind discussed below. The way to tease them apart is to try them all and see which one deduces (12), including the exceptions behind the bolded parts. It may in fact turn out that a formal/functional explanation combination is needed.

As an example of such combination, consider the generalization in (3) (see e.g. Sapir 1921, Alexander 1990, Bošković 2005).

(13) If a language has scrambling (informally, free word order), it has overt Case-marking.

The explanation for (13) quite clearly should be functional. To put it in informal terms, we need to know who does what to whom. If word order won't help in this respect, then we need Case.

But there is more to scrambling/freedom of word order than Case. One of Bošković's article generalizations in fact concerns scrambling, where only languages without (definite) articles may have scrambling (see (ib) in fn9). Informally, this means that languages with free word order tend to lack articles (there is more to scrambling than this; at any rate, to mention some typical scrambling languages: Russian, Serbo-Croatian, Latin, Japanese, Turkish, Hindi, Chukchi, Warlpiri – they all lack articles). It looks like, then, that both the presence of Case and the lack of articles matter for the freedom of word order. While the former has a straightforward functional explanation, it is difficult

way of capturing rare patterns based on exceptional mechanisms from the point of view of UG, which concerns person restrictions, see Stegovec 2019).

One should be careful with exceptions though; there is an inherent noise when working with descriptive works that can arise due to the errors/misreading of the original sources (see Baker & McCloskey 2007: 290); what appear to be exceptions may then turn out not to be exceptions, which in turn means that what appear to be only strong tendencies due to such exceptions may actually be exceptionless generalizations.

to see how the latter could be explained in functional terms, it seems to call for a formal account (cf. the account in Bošković 2008). What this means is that both functional and formal factors may be at work here, i.e. a combination of the two may be in order.

At any rate, exceptions to a descriptive typological generalization do not mean that we do not have UG at work. There clearly are universal properties of language, e.g. c-command, domination, the notion of structure – they hold for any language. However, even research within the generative paradigm has reached the point where broad typological generalizations of the kind the traditional typologists have been concerned with have become crucial for understanding the nature of language and what the generativists refer to as the faculty of language. There is, however, a significant difference in focus between TT and GT here. When the latter are looking for typological generalizations they are generally looking for typological gaps (see below for illustrations²¹). If typological gaps are mentioned at all in traditional typological works they are most of the time brushed off as accidental gaps, the focus in the works in this tradition generally being on explaining what is possible, not what is not possible. There are certainly many aspects of language that are best explained by looking at the function of language. Functional explanations are certainly appropriate, but it is harder for them to rule out cases. They most of the time provide a rationale for what is found, it is harder for them to completely rule out cases, which is what accounting for typological gaps, which concern what is not possible, would take. (The rationale they provide is generally based on functional pressures, but pressures are most of the time a matter of degree, they are not either-or characterizations.)

Traditional typologists actually often assume that nothing is truly impossible, but they don't really look for what is impossible – the

²¹ The article/adjectival LBE correlation from fn3 actually reveals a typological gap. There are four possible language types here, (d) represents a typological gap: (a) languages with articles and no LBE; (b) languages without articles and LBE; (c) languages without articles and no LBE; *(d) languages with articles and LBE.

position is there simply by assumption; as Nichols (2007) notes, they really have no interest in what is not possible²² – the generativists do, this is in fact one of their central interests (for reasons discussed below). The difference is even reflected in the way implicational universals are stated. While functionalists generally state them in the if-X-then-Y way, formalists often state them in the if-X-then-no-Y way (see fn18). The former emphasizes what is found, and the latter what is not found – it directly reveals typological gaps, while the former implies them – this difference reflects the primary interests of the two approaches: for TTs typological gaps are implied, one has to look for them since the statements don't make them obvious. This is not the case with GTs, since they are looking for them.

One revolutionary aspect of the early steps in the generative tradition that is often overlooked was caring about ungrammatical sentences. What the difference between traditional and generative typology regarding typological gaps, more precisely, the obsession of the latter with typological gaps, boils down to is in fact the Chomsky revolution in caring about ungrammatical sentences. The non-existing language type issue is just a higher-level instantiation of the same concern: caring about ungrammatical sentences (i.e. what is not possible), explaining why they are ungrammatical. The driving force of research in the generative tradition has in fact been to rule out sentences, not rule them in. After all, we don't have to do anything to rule in a sentence. Suppose there is no such thing as grammar, every principle, every mechanism, every condition, none of them exist. Every good sentence of every language is still “accounted for”, they

²² There are exceptions though: explorations of the range of possible variation within particular domains in *Canonical Typology* (see e.g. Brown, Chumakina, and Corbett 2013, Bond 2019, Round and Corbett 2020) come close to that given that determining what is possible implies knowing what is not possible (in fact, with its stance toward universality (see section 1.1) and determining what is possible, Round and Corbett (2020) come close to the larger picture view of generativists, which underscores that the one-camp-vs-another distinction is fluid, and becoming more and more a matter of a degree.

are all ruled in. You don't have to do anything to 'rule in' a sentence. But all bad sentences then become a problem.

Chomsky's *Syntactic Structures* (SS) in fact set the research agenda in generative tradition in this respect. The SS system was extremely powerful; I'm not aware of any sentence in any language that could not be captured within that system (the system even allowed movement of non-constituents) – that was in fact the problem with the system. It could do anything. The system naturally allowed for too many things, constraining it has all been about blocking things that are not found in natural languages, in other words, ruling out ungrammatical sentences.

The same reasoning, only at a level higher than sentence, extends to generative typology. Traditional typology is all about finding broad scale generalizations regarding what is found in languages, and then providing a rationale for them. Generative typology also looks for broad scale generalizations regarding what is found in human languages. But the goal is different, it is looking for what is found in order to discover what is not found. To illustrate, say property X has 5 logical possibilities to be realized in human languages, but only four are found. A traditional typology paper will typically give the generalizations regarding what is found in this domain (and if an account is provided, a traditional typologist would typically focus on explaining why we have those existing four types – if the non-existing type is mentioned, it is typically put aside as an accidental gap). A generative typology paper will, on the other hand, give that, but it will then zero down on the gap in the paradigm, on the type that is logically possible but not actually found, and ask why that is the case.

To be more concrete, consider Greenberg's generalizations in (14) (note that they are stated in the if-X-then-Y way).

- (14) a. If the relative expression precedes the noun either as the only construction or as an alternate construction, either the language is postpositional, or the adjective precedes the noun or both. (#24)

- b. If either the subject or object noun agrees with the verb in gender, then the adjective always agrees with the noun in gender. (#31)

(14a) tells us which adposition/N and adjective/N order is found when a relative precedes N in a language. What a generativist is now thinking is that what (14a) really says is that we don't find a prepositional language where a relative precedes the noun and the adjective follows the noun. The next step is to block the possibility of such a language. Similarly, upon seeing (14b), a generativist is thinking how do we force A-N agreement in gender in the presence of N-V agreement in gender?

To give a concrete example of such generative typology works, consider Messick (2016, 2017) regarding attitude reports like *Bill_i said that he_i is smart*, which reports on *Bill said: "I'm smart"*.²³ Some languages (Amharic, Zazaki) use 1st person pronoun to refer to the attitude holder – so we get “Bill_i said that I_i am smart”. Messick observes a new type, represented with Telugu and Nuer, which uses a 3rd person pronoun controlling 1st person agreement – so we get “Bill_i said that he_i am smart”. Some languages (Ewe) use a special logophoric pronoun here (logophors are found in embedded attitude reports and cannot be the matrix subject of out-of-the-blue sentences). Donno Soand Tamil represent an additional twist in using a logophor with 1st person agreement.

By examining additional languages, Messick establishes a typology of attitude reports, and crucially observes a typological gap: there is no language that uses 1st person pronoun and 3rd person agreement, as in *Bill_i said that I_i is smart*. Messick then provides a comprehensive syntax/semantics/morphology account of attitude reports that accounts not only for the existing patterns but also for

²³ I am putting aside *de-re* readings, where the attitude holder is unaware that the attitude is about them. On that reading, Bill could be drunk, watching himself on TV saying something smart, but being too drunk to recognize the person is himself Bill says: “he is smart” (instead of “I am smart”; some languages, e.g. Golan, mark such readings, which I ignore below, differently).

this typological gap. The details are not important, though it is worth noting that the crucial ingredient concerning a distinction between morphological and semantic features of agreement-controlling nominals comes from a traditional typology work, Corbett (1983). But what is particularly important is the illustration of the empirical scope of generative typology. It is about discovering what is not possible, it's about looking for typological gaps. Typological gaps are like ungrammatical sentences. This is where the soul of a generativists most happily resides.

The best way to understand why this is the case is to go back to *SS*. As noted above, there was nothing that the *SS* system couldn't do, which was taken to be a problem. Ironically, the generative revolution that started with *SS* set the broader agenda for the field (within this particular perspective), really defined its soul, in a way that appears to be quite contrary to *SS* though the appearances are somewhat deceiving. Already at the *SS* stage the field was defined as going after "all and only" the sentences of the grammar, it's just that the practical implementation of this motto with respect to the corner of the grammar of English discussed in *SS* was such that it emphasized "all", not "only".

Returning to typology, to illustrate very abstractly the difference between a traditional and a generative typologist in terms of the history of generative grammar, traditional typology would be something like the *SS* system, and generative typology something like what has happened after that; the former emphasizes what is found in languages, and the latter emphasizes, in fact is really obsessed with, what is not found. This is just a higher-level instantiation of the same obsession that is found on the level of a sentence, where the emphasis in the generative approach is on ruling out ungrammatical sentences. But generative typology is obviously enormously indebted to traditional typology.

The above is not intended to imply that functional approaches cannot capture gaps in typological paradigms. The difference here is more about the nature of inquiry, it's more methodological: formalists always try to capture such gaps (for the reasons noted above),

functionalists generally don't try to capture them because they are not concerned with them – the focus being on explaining what is possible.

There is a parallel to be made here. Interface considerations constantly pop up while doing syntactic investigations in the generative framework; as a result, no generative syntactician can afford to ignore semantics; they would be doing it to the detriment of their own research. While there has been a great deal of productive interaction between syntax and semantics at the syntax-semantics border (much of that research in fact involves collaborations between generative syntacticians and formal semanticists), the difference noted above regarding generative syntacticians and functionalists also arises with generative syntacticians and formal semanticists. As noted above, generative syntax is obsessed with ungrammatical sentences. The field is almost all about blocking what is not possible, ruling out ungrammatical sentences. Principles, constraints... almost everything that has been proposed is there to rule out sentences. Formal semantics is very different in this respect, and similar to functionalist approaches. The principal goal of a formal semanticist is to write a semantics for an acceptable sentence; in fact, one seldom finds cases where a semanticist would write a semantics for an unacceptable sentence with a goal to rule it out. Most proclamations of the sort *this-sentence-is-ruled-out-for-semantic-reasons*, because it cannot be interpreted, are actually made by syntacticians. All of this also has a counterpart on a higher, more abstract level of the theory itself – namely, with constraining the theory, which has in fact been the driving force of generative research since *SS*. Not surprisingly, semanticists are also much less obsessed with constraining the theory (in fact, syntacticians are quite uncomfortable with one of the most standard formal semantics tools, *type shifting*, the reason being that it is a very powerful mechanism that can do almost anything one wants it to do); in other words, the obsession with ungrammatical sentences and with constraining the theory seem to again go hand-in-hand here. Generative syntax differs from both formal semantics and functionalist approaches in these respects.

At any rate, typological generalizations should be treated the same way as all descriptive generalizations, which means they raise an immediate question why they hold. This in turn means that they need to be deduced, and the deduction should also explain why some patterns are rare (this also concerns any potential exceptions to the generalizations) and provide explanations for actual typological gaps (where they exist). A proper deduction of a typological generalization (whether formal or functional, or a combination of the two) should address all of these – this is how the success of a deduction of a typological generalization should be judged: the closer it gets to this goal, the better.

At any rate, typological generalizations need to be paid serious attention to – generativists sometimes ignore them to their own detriment. Take e.g. agreement. There are properties of agreement systems crosslinguistically which show that certain widely held theoretical assumptions regarding agreement are seriously on the wrong track. In the minimalist system of Chomsky (2000), functional heads *v* and *T*(ense) are the locus of object and subject agreement respectively (*vP* is right above *VP* and *TP* is higher than *vP*), with *v* and *T* undergoing agreement with the relevant nominals. There is, however, a serious problem with these assumptions revealed by Julien's (2002) typological study of verb morphology, based on over 500 languages. Agreement is commonly marked at the edge of the verb: while there are exceptions, it generally follows everything else on the verb.²⁴ Crucially, this includes object agreement, as in Itelmen (15).

- (15) n-ǎlčqu-z-um
 3.pl.sub-see-pres-1.sg.obj
 'They see me' (Bobaljik and Wurmbrand 2001)

²⁴ Bantu languages may provide an exception, with tense occurring between subject and object agreement markers; however, these markers have been argued to have a different status regarding the clitic/affix distinction (Bresnan & Mchombo 1987 argue object markers are actually clitics), which may be an interfering factor here.

This is quite surprising from the current minimalist theoretical perspective, where *v*, the head of a phrase right above VP, is the locus of object agreement: since *v* is lower in the structure than T, object agreement should occur inside of tense (i.e. be closer to the verb than tense). As Julien (2002) shows, (15) is in fact not a rare pattern, hence not something that should be easily dismissed.

The above can be taken as an illustration of the importance of typological work for generativists (given that paying attention to it is fairly recent; many generativists in fact still ignore it). Typological work needs to be paid serious attention to, sometimes it may help prevent the generativists from going down a wrong theoretical path, as the above discussion indicates regarding the current minimalist assumptions concerning agreement.

I'll now briefly discuss another example of generative typology, whose driving force is again a quest for typological gaps. The account will be compared to a functional approach. However, the two will not be opposed – while superficially they look very different, more abstractly, they share interesting points of convergence, which will be taken to indicate that convergence that comes from such radically different perspectives may provide real and significant insights. Going beyond convergence (and in the general spirit of this paper), we will in fact see that there is room for both approaches in the comprehensive account of the relevant phenomenon.

2.2. ON PERSON RESTRICTIONS

In many languages, person specification regulates co-occurrence of weak pronominal objects, a restriction referred to as the Person-Case Constraint (PCC) and assumed to come in two versions: STRONG PCC, where in the presence of indirect object (IO) direct object (DO) cannot be 1P or 2P (see Greek (16)), and WEAK PCC (found e.g. in Arabic), where if IO is 3P DO must be 3P.

- (16) Tha **me/se/tuton**/***me**/***se** stilune. *1P/2P/3P.IO>>1P/2P.DO
 fut 1_{GEN}/2_{GEN}/3_{M.GEN} 3_{M.ACC}/1_{ACC}/2_{ACC} send_{3PL}
 'They will send **him**/***me**/***you** to **me/you/him**.'

While less-known, PCC-like restrictions also exist for external/internal argument combinations. Stegovec (2019) argues for unifying the two restrictions, observing that external (EA)-internal (IA) argument PCC also comes in strong and weak version. STRONG EA-IA PCC is found in e.g. Christian Barwar and WEAK EA-IA PCC in e.g. Southern Tiwa. Stegovec observes that both strong and weak restriction can be unified for EA-IA and IA-IA pairs, e.g. for the weak one: when pronominal markers co-occur, if the S/IO is 3P the O/DO must be 3P.

Almost all formalist approaches to the PCC focus on IO/DO pairs, analyzing it in such a way that it is crucial that the first argument, IO, bears an inherent Case (the typical situation is that IO is dative and DO accusative). This obviously cannot be extended to EA-IA pairs. Furthermore, Stegovec (2019) shows that even IO-DO PCC is insensitive to the Case type of the first argument based on languages that allow both orders of IO/DO clitics. Thus, Slovenian shows PCC effects with DAT-ACC clitic order (17). If the clitic order is reversed, a reverse PCC arises: the IO's, not DO's, person is restricted (18).

- (17) Mama {**ti ga**}/*{**mu te**} bo predstavila. ✓2P_{IO}>>3P_{DO};*3P_{IO}>>2P_{DO}
 mom 2P_{DAT} 3P_{M.ACC} 3P_{M.DAT} 2P_{ACC} will introduce
 'Mom will introduce **him** to **you**/* **you** to **him**.'
- (18) Mama {**te mu**}/*{**ga ti**} bo predstavila. ✓2P_{DO}>>3P_{IO};*3P_{IO}>>2P_{DO}
 mom 2P_{ACC} 3P_{M.DAT} 3P_{M.ACC} 2P_{DAT} will introduce
 'Mom will introduce **you** to **him**/***him** to **you**.'

This shows that person restrictions (PR) are not limited to inherent-Case-on-the-first/higher-argument contexts and that the position in the syntax, not grammatical function, is what matters.

Stegovec also conducts a survey, which builds on Haspelmath (2004) and Albizu (1997) but significantly broadens their scope by including more than 100 languages from more than 20 families. While the survey reveals more patterns than have been previously reported, it's not the case that anything goes. PRs always have these properties:

- (19) a. The restriction always applies to the structurally lower marker.

- b. The restriction either forces the lower marker to be 3P or bans it from being either 1P or 2P.
 - c. A language can only have a reverse PR if it also has the standard one ((18) is found only in the presence of(17)).
 - d. If a language has both EA-IA and IA-IA PR, the latter cannot be weaker than the former; cf. (20), where the first pattern refers to EA-IA PR and the second to IA-IA PR—(20d) is a typological gap
- (20)a. WEAK+STRONG (Southern Tiwa, *Kiowa-Tanoan*)
- b. WEAK+WEAK (Alutor, *Chukotkan*)
 - c. STRONG+STRONG (Telkepe, *Semitic*)
 - d. STRONG+WEAK unattested

Stegovec provides a formal account that deduces the generalizations in (19) – it allows attested but not unattested patterns and also explains why some patterns are rare due to them using exceptional theoretical mechanisms (e.g. the pattern where only 3P>>2P is banned is found only in Salish). The gist of the account is that there is no crosslinguistic variation in argument structure, locality domains and the way agreement works. Relevant crosslinguistic variation comes from independently motivated variation in the internal structure of pronouns, certain movement possibilities (DO-IO order arises through movement) and the presence/properties of certain functional projections (so there are no PCC-specific mechanisms). What's crucial in the account is the structural placement of a particular head, *v* (which introduces external arguments); in particular, EA is higher and IAs are lower than *v* (so this is not simply an issue of argument hierarchy). Regarding the issue of whether particular pronominal elements are involved in a strong or a weak pattern, what matters is whether they are weak pronouns or clitics.

It is instructive to compare Stegovec's approach with a predecessor he relied on, Haspelmath (2004), which provides a functionalist perspective on the phenomenon (for ditransitives). For Haspelmath, PR restrictions are about frequency and grammaticalization. The basic idea is that there are prominence scales for person, where 1/2P is

ranked higher than 3P, and semantic roles, where recipient is ranked higher than theme. Person-role association is more natural when high (1/2) persons are associated with high (Recipient) role and low (3rd) person with low role (Theme). The gist is that the requirement for alignment of semantic-role prominence with person prominence disfavors ranking DOs higher than IOs in person. PCC then reflects alignment of grammatical function prominence with person prominence (the approach is most straightforward if *Weak PCC* is the core of PCC). At any rate, the gist of the approach is that the higher a pronominal element is on the thematic scale, the more likely it is to be 1/2P. Based on corpus studies, Haspelmath claims this is the case even in non-PCC languages as revealed by the frequency of usage (German is not a PCC language, allowing all pronoun combinations, but the most frequent one is 3P DO and 1/2P IO) – what is different with PCC languages is that this preference is grammaticalized (though it’s not really clear what is meant by ‘grammaticalized’, which is an important issue here; the following discussion will in fact address it). It seems difficult to capture all relevant variation (see Stegovec 2019) under this approach. Also, it would need to be shown that all types of PCC effects in PCC languages have their type counterparts in non-PCC languages, which doesn’t seem likely. However, syntactic accounts of PCC effects like Stegovec’s do not have anything to say about non-PCC languages. There may then be an opening for both approaches, where something along the lines of Haspelmath (2004) would be correct for non-PCC languages, where the effect is not syntactically encoded. Syntactic encoding of the effect (cf. the above remark about grammaticalization) would then lead to a Stegovec-style approach. This can be looked at more broadly as formalization of functional factors.²⁵ In fact, it would not be surprising to see such a situation arising quite often, where a formal issue arises as a reflex of a functional (or a broader cognitive) concern, getting a life of its own upon this kind of syntactic encoding,

²⁵ When functional considerations give rise to a formal mechanism, they may end up being incorporated in a rather abstract and formal way once the formal properties of the computational system of language are brought into the picture.

to the point that it can go quite beyond the original functional factor.²⁶ The reason why this kind of situations are not seen more often may be the formalist/functionalist division in the field, which minimizes the interaction that would be necessary to locate such situations.²⁷ (Another case where a combination of functional and formal factors may be involved, concerning free word order, was noted above; see also the discussion of ergativity in section 1.2.; For different suggestions regarding potential ways of combining the two, see Baker & McCloskey 2007 and Polinsky & Kluender 2007).

From this perspective, while there are differences between Haspelmath's and Stegovec's account there are also important similarities. Haspelmath crucially relies on semantic roles and their prominence, which corresponds to argument structure in Stegovec's account, where the relative prominence is reflected in the structural height of arguments (in other words, generativists also have a counterpart of Haspelmath's semantic-role prominency scale; in fact, they also have something similar for the person scale, which has been used in accounts of the PCC, e.g. Bejar & Rezac 2009, Anagnostopoulou 2003). The reverse PCC pattern, however, raises a problem for Haspelmath's account. Semantic-role prominence doesn't matter here, it in fact cannot be appealed to: what matters is the syntactic height (the first object can be shown to be structurally higher than the second). However, putting this issue aside, both the semantic-role hierarchy and the person hierarchy have their counterparts in generative mechanisms, and both are in fact used in Stegovec's account. Furthermore, in both Haspelmath's and Stegovec's account Case simply doesn't matter. Previous accounts of the PCC crucially relied on

²⁶ The lack of a complete correspondence between PCC and non-PCC languages hinted at above would then not be a problem.

²⁷ Another relevant case may concern the requirement that subjects in Chinese be interpreted as specific/definite. While English doesn't have this requirement, specific/definite subjects are much more frequent than indefinite ones in English, which could also be looked at as a syntactization/formalization (in Chinese) of a broader cognitive tendency that English reflects (for a more general discussion along these lines, see Hawkins 2010).

Case (but see Baker 2008). Neither Stegovec's nor Haspelmath's does. So, what we have here is a convergence on a particular issue from two very different perspectives. And the facts regarding reverse PCC quite conclusively show Case indeed doesn't matter. So, yes, we could dwell on the differences and emphasize them, but abstractly there are clear similarities between these accounts: when convergence comes from two rather different perspectives, that can be interpreted as a rather strong sign that those points of convergence are indeed on the right track. And in fact, as noted above, there may actually be room for both perspectives in a comprehensive account of the phenomenon, which could be an instantiation of a more general situation where functional factors get encoded syntactically; the relevant syntactic mechanisms then arise as a reflex of a functional factor, but their "syntactization" can take them quite beyond the original functional factor (see also section 5 for a more general discussion).

In this respect, functionalists often question whether a typological generalization requires a formal analysis of the data, but they also often use the term 'grammaticalization' in a way that in fact indicates such a need. More generally, what the functionalists often mean by grammaticalization (and what Haspelmath also means by it in his approach to person restrictions), is that X started out as a tendency due to functional/pragmatic reasons, but has "rigidified" into a hard rule, without really explaining the rigidification. This is a place where a formalist can step in to provide an actual account of the rigidification, including the details of exactly what X has hardened into, and why it has hardened into a particular format rather than a different logically possible format. This would lead to a productive and complementary investigation of a phenomenon from both perspectives that would in turn provide a more comprehensive understanding of the phenomenon. Paying closer attention to what is meant by something being "grammaticalized" can then not only be progress toward rapprochement between the two perspectives, but also make the complementarity of the two perspectives more obvious.

3. EFFICIENT DESIGN AND THE NATURE OF LANGUAGE

Above, I have discussed what the goals of a deduction of a typological generalization should be. There is a related point to be made. One of the tenets of the Minimalist Program is that language faculty is characterized by efficient design, which appears to be a contentious issue in functionalist circles. But there should be nothing controversial about this. Occam's razor as a research methodology in fact leads to the efficient design hypothesis – it is a dictum that your subject of inquiry should be only as complex as it needs to be, hence the efficient design hypothesis. But what do we know about efficient design? Binary computer language systems and Markovian finite state devices are pretty efficient and simple, why not them? The reason why not them is due to an obvious question that arises here: efficient design for what? The *what* question has to be answered, and this is where the two candidates in question fail. So the crucial question is efficient design for what? Is it the nature of interaction with non-linguistic interfaces (namely, conceptual-intentional and articulatory-perceptual), as is often assumed by generativists? Or the function of language? The two actually may not be very different (though generativists generally talk only about the former). The point made here is that to look for efficient design one actually needs to pay attention to functional considerations, efficient design should include what language is used for. Efficient design concerns the nature of language – interaction with the interfaces, what language is used for – they are fundamental to the nature of language.²⁸ The efficient design hypothesis, which

²⁸ It should be noted in this respect that Chomsky (2020) suggests that I-language (which is what the generativists mean by *language faculty/FL*) doesn't care about use (i.e. parsing) or communication, but it cares about expression of thought; the design of language should then capture the fundamental aspects of thinking. It is not, however, easy to make a clear distinction between communication and expression of thought (in fact, I am not doing it here); in a sense thinking involves communication with oneself; further, communication with oneself or with others, it involves expression of thought. On the other hand, use, i.e. parsing, is a different issue, I-language may indeed not care much about that – I refer the reader to

is often seen as a point of divergence between the generativist and the functionalist camp (though see Golumbia 2010), is then actually another potential point of convergence between the two.

Now, one of the tenets of minimalism is that as much as possible, and as directly as possible, should follow from bare output conditions (see Chomsky 1995 on this notion), i.e. the nature of language, which means that formal reasons that have nothing to do with the nature of language should be minimized. The latter often goes unacknowledged by generativists (though see Chomsky 2004). There is a related aspect of all this. As noted above, the nature of language inevitably includes functional considerations. While functional explanations for typological gaps might be more difficult to achieve than strictly formal explanations for reasons discussed above, a priori they should actually be preferred, on the grounds that they would be more likely to tie the explanations to something else, i.e. to explain the relevant properties of language by appealing to the nature of language. The above statement obviously overgeneralizes – each specific case of a typological gap and its explanation need to be looked at separately and evaluated with respect to the overall goal regarding what an explanation should achieve. But the point is that everything else being equal, functional explanations should be preferred on conceptual grounds, in fact based on one of the core minimalist tenets. Note also that what I have referred to above as the minimalist tenet, tie as much as possible, and as directly as possible, to the very nature of language, is essentially the Occam-razor strategy – simplify to what is unavoidable. So, Occam's razor, which is to say the nature of language, favors functional explanations. More generally, Occam's

Chomsky (2020) for relevant discussion, adding only one point: everything we know about syntax indicates structure building proceeds bottom-up (there have been occasional attempts at top-down syntax (Phillips 1996), but doing it that way raises numerous non-trivial issues). Parsing of course proceeds left-to-right. There is an obvious conflict here between the fundamental specification of FL and parsing, which indicates that FL indeed does not particularly care about the latter, as Chomsky suggests (Chomsky notes in this context issues that filler-gap dependencies raise for parsing).

razor then disfavors, and requires re-evaluation of, strictly formal mechanisms, like e.g. the EPP, which in its generalized form from Chomsky (2000) is not simply a property of a single head (Infl) but drives all movement (see Bošković's 2021a discussion of the EPP from this perspective). There is actually an implicit acknowledgment of the point that strictly formal mechanisms should be disfavored in Chomsky's (1995) position on Agr, where he argues for elimination of Agr (i.e. agreement phrases) on the grounds that this is a strictly formal element which does not contribute to interpretation (Chomsky 2000 also attempts to minimize the role of Case, which for him is also a strictly formal property). Chomsky's (1995) stand on Agr can be interpreted as a minimize-purely-formal-considerations strategy. The other side of that coin, generally ignored by the generativists, should lead to adopting a maximize-functional-considerations-strategy (the strategy in this context really follows from Occam's razor).

Consider now a more concrete example of what it means to minimize purely formal considerations. Consider the operation of movement, which immediately raises a question: why do we have movement in the first place? Chomsky's (2000: 120–121) position is that this has to do with “externally imposed legibility conditions”, i.e. it is due to “conditions imposed by the external systems”. What this means is that the reason for it is essentially functional, or more broadly non-syntactic: to be able to express notions that go beyond the basic argument structure (for which Merge alone suffices), i.e. more complex semantic notions involving issues like scope/scopal ambiguities, pragmatic notions like topic/focus interpretation, specificity... In other words, movement is there to express various semantic and pragmatic relations (in essence, everything that goes beyond the basic argument structure). A question that arises here is to which extent these notions have led to the development of formal requirements which then drive movement (see fn25)? In other words, do the functional reasons in question directly motivate movement, where movement would be directly interpretation driven, or are there formal requirements that essentially serve as intermediaries, leaving syntax free of semantic considerations? To make the question a little more concrete: when a

moves to SpecTopicP and receives topic interpretation, does α move there in order to receive such interpretation or is there a formal reason behind the movement, with α moving to satisfy this formal reason, as a result of which it is interpreted as a topic?²⁹

While the question may seem innocent in this particular case (in that there doesn't seem to be much of a difference between the two positions), it isn't in other cases. Consider e.g. the different behavior of Bulgarian and Japanese in multiple questions, where in Bulgarian all wh-phrases move to the interrogative SpecCP while in Japanese they all remain in situ. There are various ways of implementing this formally. However, without appealing to formal properties that would cause this difference, i.e. on the direct syntax-interpretation mapping approach, we are led to the conclusion that Bulgarian and Japanese wh-phrases are interpreted differently, i.e. they differ semantically (the gist would be that Bulgarian wh-phrases must function as operators binding a variable, which is not the case in Japanese, where wh-phrases are interpretable in-situ). This means that on the direct syntax-interpretation mapping approach, Bulgarian and Japanese questions cannot have the same LF (interestingly from this perspective, in French, which has optional wh-movement, wh-movement and wh-in-situ constructions have different semantics, see Boeckx 1999, Zubizarreta 2003). Note that positing LF/covert wh-movement in wh-in-situ languages like Japanese (e.g. Huang 1982), as a result of which Japanese and Bulgarian would have the same LF, would not work – we would still need a formal difference between Japanese and Bulgarian that would be responsible for the overt/covert movement difference, which would go against the spirit of the direct syntax-semantics mapping approach.

²⁹ There is a cartology/mapping debate regarding discourse notions like topic and focus within formalist approaches (see e.g. Rizzi 1997 and Neeleman & van de Koot 2008). I will not go into it here, apart from noting that the former involves more formal factors than the latter, which means that the minimize-formal-considerations strategy (on its own) would favor the latter, see Lacerda (2020).

The two cases noted above can thus be handled through either directly interpretation-driven movement³⁰ or formal requirements as intermediaries (which would essentially be an indirect syntax-semantics mapping approach), but with non-trivial implications under the latter regarding the semantics of questions/wh-phrases. Those non-trivial consequences, however, set the scene for a typological investigation; in particular, what under a purely formal approach of the kind argued for in e.g. Huang (1982) was considered to be a crosslinguistic distinction regarding whether wh-movement takes place overtly (Bulgarian) or covertly (Japanese) should now become a matter of typology of wh-phrases: there are languages where wh-phrases are only interpretable as operators/in moved position (Bulgarian), languages where they are interpretable as variables/in-situ (Japanese), and languages where they are interpretable as variables/in-situ only if another wh-phrase is interpreted in the moved position/as an operator (English). This essentially semantic typology in turn raises a question: to what extent is all this reflected in the morphological make up of wh-phrases and/or other functions that they may perform. Regarding the latter, it is well-known that in many languages wh-phrases can have a range of non-question functions (see especially Haspelmath 1997). There are several language types in this respect. One is represented by Serbo-Croatian, a multiple wh-fronting (MWF) language, which has a fully productive system where addition of an inseparable affix to a wh-phrase results in a series of meanings shown below. I will refer to this pattern as the *sub-wh system* (intended to indicate the morphological subset-superset relationship between the wh/question usage and other usages).

(21) *ko* ‘who’; *iko* ‘anyone’; *niko* ‘no one’; *neko* ‘someone’; *svako* ‘everyone’; *bilo ko* ‘whoever’

Note that the sub-wh system is different from the situation found in Chinese, where the exact same form can have a series of usages

³⁰ Quantifier raising, which is assumed to take place because quantifiers cannot be interpreted in situ, might be the best candidate for such movement.

(including *wh*), depending on the context it finds itself in, as well as the situation found in Japanese, where (except in the case of the counterpart of 'someone'), the element added to the stem can be detached from it (there is an special element present even on the *wh*-usage; in standard Japanese it is obligatorily detached, but in Okinawan Japanese it can be attached to the stem itself, which indicates we are not dealing here with a sub-*wh* system). Finally, English has a somewhat similar situation to SC, cf. *where, somewhere, everywhere, nowhere*. However, English is still quite different from SC; SC has a fully productive sub-*wh* system where the process in question is fully productive for all *wh*-phrases. This is not the case in English, as indicated by e.g. **somewho/*everywho/*nowho* or *somewhat/*everywhat/*nowhat*.

Consider now the list of MWF languages given in Bošković (2012): Polish, Czech, Russian, Slovenian, Ukrainian, Bulgarian, Macedonian, Romanian, Hungarian, Basque, Yiddish, and Mohawk. It turns out all these languages have a fully productive sub-*wh* system. Particularly interesting is Romance. Latin was a MWF language (Ledgeway 2012, Dadan 2019). Importantly, it had a fully productive sub-*wh* system, which got lost in all modern Romance languages except one: Romanian, the only modern Romance language that still has MWF.

(22) **Latin interrogative existential, etc...**

person quisali-quis
thing quid ali-quid
place ubi ali-cubi
time quandoali-quando

(23) **Italian**

person chi qualcuno
thing chequalchecosa/qualcosa
place dove in qualche luogo
time quandoqualque volta

(24) **Romanian**

person cine cine-va
thing cece-va
place undeunde-va
time cîndcînd-va

We then have a correlation between MWF and the sub-wh system, the former requiring the latter.

(25) If a language has multiple wh-fronting, it has a sub-wh-system.

Recall that on the direct syntax-semantics mapping, MWF indicates that wh-phrases in the language in question must function as operators binding a variable. What this then indicates is that this semantic property has a reflex in morphology: it requires a sub-wh system.

In summary, the minimalist tenet that as much as possible, and as directly as possible, should follow from bare output conditions, i.e. the nature of language, leads to preferring, in principle, functional over strictly formal explanations and more generally to a minimize-strictly-formal mechanisms strategy (where formal reasons that have nothing to do with the nature of language should be minimized). This in turn prefers (again in principle) direct syntax-semantics mapping, where movement is directly interpretation-driven, over indirect syntax-semantics mapping, where formal requirements would serve as intermediaries (and would drive movement). We have seen one area where the strategy has non-trivial consequences: it sets the scene for a semantic typology of wh-phrases (where the formal distinction overt vs covert wh-movement is recast as a difference in the semantics of wh-phrases), which in turn raises the question to which extent the relevant semantic differences are correlated with the morphological make up of wh-phrases and/or other (non-question) functions that the relevant elements may perform. We have seen that there may in fact be a correlation in the case of one particular type.

Before closing this section, I will briefly note a new argument for the direct syntax-semantics mapping approach (given its importance in the context of the current discussion). Consider the so-called Superiority effect, illustrated by (26a-b) ((27) gives their structure prior to wh-movement).

- (26) a. Who did you tell that she should buy what?
b. ?*What did you tell who that she should buy?
(27) You tell who that she should buy what.

In Minimalism, the effect from (26) follows from economy of derivation, which requires every condition to be satisfied through the shortest movement possible. Wh-movement has to take place in English. In (27), the requirement can be satisfied by moving either *who* or *what*. Since the former results in shorter movement (cf. the pre-movement structure in (27)), (26a) is preferred to (26b).

There is, however, evidence that Superiority is also semantically conditioned. Consider (28).

- (28) a. I wonder who bought what?
- b. *I wonder what who bought?
- c. Who wonders who bought what?
- d. Who wonders what who bought?

What cannot undergo wh-movement in (28a-b), a familiar superiority effect. However, (28d) is grammatical in spite of *what* undergoing wh-movement. Importantly, (28d) is not ambiguous: it is acceptable only if the second *who* takes matrix scope (i.e. as a matrix multiple question on the two *who*-s); it is unacceptable on the multiple indirect question reading, where *what* and second *who* are interpreted in the embedded SpecCP. Apparently, superiority effects are conditioned on the relevant wh-phrases being interpreted in the same interrogative SpecCP (*what* and second *who* are not interpreted in the same interrogative SpecCP on the allowed reading of (28d)). Now, superiority effects are a result of a formal derivational requirement of the computational system. A question then arises how to bring in semantic considerations into a requirement of the computational system itself. At the point of movement of *what* to the embedded SpecCP, which occurs before the matrix clause is built in (28d) (given that structure building proceeds bottom-up, the embedded clause is built before the matrix clause), (28b) and (28d) are in fact identical syntactically. Movement to the embedded SpecCP, then, needs to be made sensitive to the final semantics of these examples. In the indirect syntax-semantics mapping approach, where syntax is free of semantic considerations, it is very difficult to achieve this. On the basis of such examples, Golan (1993) suggests only constructions

with the same interpretation can be compared regarding economy of derivation. On the direct syntax-semantics mapping approach, where movement is directly driven by the relevant semantic considerations, those semantic considerations are directly involved in the relevant movements, hence the movements can be made sensitive to them. The effect in question may then provide an argument for the direct syntax-semantics mapping approach.

3.1. LANGUAGE ACQUISITION

As discussed above, Occam's razor leads to the efficient-design hypothesis. There is an important point in this respect that is often overlooked nowadays in generative circles: efficient design should make language easily learnable. Generative syntax in fact essentially started with language acquisition, i.e. what is often referred to as the logical problem of language acquisition: how come children can learn something as complex as language so easily. In principle, then, explicit concern for language acquisition should be a significant difference between GT and TT. Recent research in generative syntax has, however, simplified the syntactic design of language a great deal (at the risk of some exaggeration, a simple merge operation and a locality domain). From the perspective of a generativist, if we simplify the design too much, and say that only that is innate, a question arises how can the children learn the rest. The Occam-razor driven research in syntax has essentially raised an Occam-razor problem in another area, namely regarding language acquisition (which may have become a bit of a back-burner for a generative syntactician. In this respect, GT should lead to re-raising the issue of language acquisition – after all, the child needs to learn all the differences typologists are concerned with.) From this perspective, the above discussion regarding differences between languages with and without articles gains wider significance. There have been proposals regarding so-called octopus parameters, where one parameter has consequences for a number of phenomena. The DP/NP parameter (the word parameter is used for ease of exposition) is by far the biggest octopus ever, it is a veritable

monster (more than 20 morphological, syntactic, and semantic properties crucially depend on it, see fn9). What is important in the context of the current discussion is that it greatly eases the burden of language acquisition. Given all the generalizations where the NP/DP difference is involved, all the phenomena where the NP/DP difference reaches into, the difference gives the child so much, more than anything else. But how can the child learn the NP/DP difference? The NP/DP generalizations all involve potential triggers but most of them (maybe even all of them) are not plausible candidates. They either involve rather sophisticated phenomena, like the interpretation of superlatives, or phenomena that are simply not widely available crosslinguistically, like multiple *wh*-fronting. How about the definite article itself? Bošković (2010, 2016c) suggests the definite article is indeed the trigger for learning the whole monster.³¹ In this respect it should be noted that *the* is in fact the most frequent word in English. *The* then seems to be a perfect trigger for learning the monster. Importantly, Koulidobrova (in press) shows that definite article is actually the trigger for proper emergence of D-items in language acquisition. Early on, children make a lot of mistakes in the usage of D-related items in English, which also includes omitting *the*. Once *the* is properly learned, mistakes regarding other D-related items stop, which indicates that *the* is indeed the trigger for proper acquisition of D-related items, and more generally DP. Koulidobrova in fact interprets her data as supporting the DP/NP analysis.³² Everything

³¹ This implies that there cannot be a null definite article language, a language with null *the* would essentially be unlearnable. (In fact, Bošković's NP/DP generalizations presuppose that there is no null *the*.) Strong evidence that there is indeed no null *the* in languages without articles is provided by Despić (2019) and Jenks (2018), see fn11.

³² There have in fact been proposals that children go through the NP stage even in English, which would then be a default (Guasti et al 2008; Mathewson et al 2001). See also Petroj (2020) for an interesting parallelism between language acquisition and code switching (a process which involves alternation between different languages within a single utterance): she provides evidence for the NP/DP hypothesis based on code switching between a DP language (Romanian) and an NP language (SC).

thus starts falling into place once language acquisition is taken into consideration. Most importantly, we are addressing here the issue of efficient design when it comes language acquisition: learning *the* helps the child with so many things.

In light of this, consider the semantic role of the definite article. In formal semantics terms, iota-operator, which is the primary semantic job of the definite article, turns NP, which is of type $\langle e,t \rangle$, into type e ; what this essentially means is that it turns a predicate into an argument. This way, it makes it possible to integrate the NP into the clause/VP. On a more informal level, taking verbs and nouns to be the primary categories, we want to be able to integrate the two, which will give us a clause. This is in fact the job of an iota-operator, definite article being an instantiation of an iota-operator.³³ This certainly looks like a pretty important and extremely basic job; see in fact Progovac (2010) for an idea that this was all there was in one stage of evolution of syntax; according to Progovac, at one point there was only a single argument-predicate combination. Definiteness is then the easiest way of making a distinction between predicates and arguments, and of integrating the NP into the VP/clause. From this perspective (which is essentially functional), it does not seem surprising that definiteness is so important, that its tentacles reach into so many areas – it can in fact be considered part of efficient design.

3.2. ON THE METHODOLOGY OF INVESTIGATING CROSSLINGUISTIC VARIATION

Returning to efficient design, when studying it one should use whatever one can, syntactic phenomena, semantic phenomena, PF aspects, language acquisition, language change, different languages... the use-whatever-you-can methodology is in fact the standard stand in the generative approach, often reiterated by Chomsky, and a normal approach in natural sciences. There is a related point: it is often assumed in functionalist circles that the

³³ In a language without definite articles this would then be a purely semantic operation.

generative approach is so syntax-centric that just showing that some phenomenon is not syntactic, or has non-syntactic aspects, argues against the generative approach in general (see sec. 1.2). This relies on a view of the generative approach which may have been true long time ago, but certainly not any more.³⁴ The generative approach has reached the point where linguistic phenomena are looked at in their totality, not compartmentalized by subfields. A generative syntactician must constantly pay attention to semantics, morphology, phonology, language change, language acquisition, and look seriously into understudied languages and typology. Typology is particularly important here. We have reached the point where typology should be at the center of investigating what generativists refer to as UG. At this point of our understanding, broad crosslinguistic comparisons, Greenberg-style typological generalizations, are more enlightening regarding the nature of language and UG than detailed investigations of particular languages (the latter are a prerequisite for the former though the practitioners of the latter are often a reviewing stop sign for generative typology works for reasons discussed below).

Now, what has dominated the generative camp regarding crosslinguistic studies is the so-called microparametric approach, which compares languages that are otherwise very similar. (Kayne's work is one of its cornerstones; the typological approach (the term is used neutrally here, not to be equated with TT) is sometimes referred to as macro(parametric) approach, and contrasted with the micro(parametric) approach. For consistency with the literature I will use that term in discussion of the opposition.) The microapproach seems to be considered 'easier', but it really isn't. Consider languages A and B within the micro/macro approach distinction. In the microapproach, A and B, say Florentino and Trentino, varieties of Italian, share a lot of things; as a result, it may be easier to spot and focus on the differences. Within the macroapproach, languages A and

³⁴ Independently of this, even if focused on generative syntax, the attitude in question would be rather strange: generative syntax cannot account for palatalization in Slavic or bring peace to the Middle East, but that does not show that it is wrong.

B, say Russian and Chinese, differ in many respects; it may then be easier to spot and focus on the similarities. There is really no sense in which one approach is 'easier' than the other; but they do differ methodologically, with respect to what we focus on: differences or similarities, something that has often escaped attention. In many respects, the macroapproach can be more fruitful, and can help us locate the points of crosslinguistic variation that would otherwise be difficult to notice. The reason for this is that those similarities from the microapproach, which are often taken to represent invariant FL properties and considered principles of UG, can turn out to be points of variation from a broader point of view. There is, however, one situation where the macroparametric approach, taken as a search for similarities among very different/unrelated languages, is particularly useful which is generally neglected: when very different languages are compared it is difficult to pinpoint the reason for the differences unless we also find some similarities. Consider, e.g. the comparison of Mohawk and English in Baker (1996).³⁵ English allows extraction of nominal complements:

(29) Who did John see friends of?

Mohawk disallows such extraction. Baker points out that Mohawk is a polysynthetic language, while English is not, and provides an account of this contrast that ties it to this difference. The issue here is that there are a lot of differences between Mohawk and English, it is difficult to put the finger on the ultimate factor behind the contrast in question. A macroparametric/typological approach (taken as a search for similarities among very different/unrelated languages) can help us here. It turns out that Serbo-Croatian behaves just like Mohawk regarding (29). In fact, SC, which is in the respect Baker discusses different from Mohawk but like English (it's not a polysynthetic language, like English and unlike Mohawk), behaves remarkably like Mohawk and unlike English regarding extraction out of nominals. While SC disallows extraction of nominal complements, like Mohawk

³⁵ The comparison scale of Baker (1996) is much broader than what is illustrated by the discussion below, whose point is strictly methodological.

and unlike English, it allows left-branch extraction of possessors (i.e. extraction of the possessor in cases like *Mary's house*), which is actually also allowed in Mohawk but not English. If we are going to have a uniform account of these similarities/differences, what's responsible cannot be polysynthesis. There is, however, a similarity between SC and Mohawk which, as noted above, Bošković (2008, 2012) argues is a point of a broad typological difference that affects numerous syntactic, semantic, and morphological phenomena: SC and Mohawk lack definite articles, while English has them. In fact, Bošković (2012) provides an account of the relevant extraction differences between English and SC which would extend to Mohawk, since what crucially matters there is the presence vs lack of articles. In fact, Mohawk turns out to fit a number of Bošković's article generalizations: one of them, regarding left-branch extraction, was noted above; it also fits Bošković's superiority generalization (multiple wh-fronting languages without definite articles do not have fixed order of fronted wh-phrases), head-internal relatives island generalization (head-internal relatives display island-sensitivity in languages without definite articles, but not in languages with articles), and the number morphology generalization (number morphology may not be obligatory only in languages without articles).³⁶ The point here is that it is difficult to pinpoint the responsible factor when comparing languages which are very different, with no point of similarity that can be used to illuminate the issue, as was the case here with English and Mohawk. But once SC entered the picture, we got that point of similarity (between SC and Mohawk), which was in fact provided by a typological comparison. If the above discussion is on the right track, the reason for the English-Mohawk differences in question is not polysynthesis, but the presence vs lack of definite articles. In this respect it should be noted that Baker (1996) actually observes that polysynthetic languages in general

³⁶ See Bošković (2012) for what is meant by the phenomena referred to in these generalizations. Most of them are one-way correlations; SC is not relevant to the last two but e.g. Japanese, another language without articles that is not a polysynthetic language, patterns with Mohawk in these respects.

lack articles (this is again a one-way correlation, it does not mean languages without articles will all be polysynthetic). The polysynthesis parameter is then also tied to articles. (Baker in fact also suggests that the lack of articles implies the lack of DP.) But it was difficult to see what is responsible for the English-Mohawk differences in question, polysynthesis or definite articles, until languages without articles that are not polysynthetic were brought into the picture. I emphasize that the relevant crosslinguistic differences are not micro but macroparametric (i.e. typological) in nature: they illustrate the benefit of looking for similarities between (unrelated) languages that are otherwise very different. (In fact, it's worth noting that Bošković's NP/DP project started with a combination of the micro and macro (i.e. typological) approach: the first point of comparison were differences among Slavic languages (only Bulgarian and Macedonian have definite articles within Slavic) and in the history of Romance (Latin, which didn't have articles, vs modern Romance), and similarities between Slavic languages without articles and East Asian languages without articles).

At any rate, the gist of the difference between the microparametric approach, which dominates the generative program, and the macroparametric/typological approach (which has only recently started to gain traction within the generative paradigm hence has often been taken to characterize the functional approach in broad comparisons of generative and functional approaches), can be summarized as follows: (a) the former: X and Y are very similar, let's see where they differ (this is what is informative); (b) the latter: X and Y are very different, let's see where they are similar (this is what is informative).

There is obviously a place for both strategies. While this used to be a point of divergence in crosslinguistic studies between generativists and functionalists (with the former not taken to be particularly interested in employing both strategies), it is increasingly (with the emergence of generative typology) becoming another point of convergence.

There may be another reason for the (initial) research strategy difference (where the difference concerns what is emphasized) regarding the kind of languages that are looked at (similar or unrelat-

ed) between the two camps which has to do with the questions the two camps are primarily asking. The generativists generally focus on similarities across languages, asking the question why is it that there are so many similarities (because their goal is to investigate UG), while the functionalists often focus on differences, asking why there is so much crosslinguistic variation. This could actually be the reason for the difference in the methodology of crosslinguistic research, where the generativists emphasize looking at similar, related languages (which have a lot in common) while TTs emphasize comparisons of unrelated languages. But this is only an issue of focus, these two lines of inquiry are not incompatible. Both of these are right questions to ask (why there are so many similarities across languages and why there are so many differences across languages) – there are both a lot of similarities and a lot of differences across languages. Both need to be addressed.

In the Principles & Parameters approach, the differences fall within the domain of parameters (see Roberts 2019 for a recent discussion of parameters). The issue of focus is, however, clear here: there is much more research on the principles of UG (i.e. invariant properties) than on the parameters (i.e. crosslinguistic variation); the distinction may not have been that sharp initially, but has been increasing over time, as with the Minimalist Program. What may have contributed to this was the hypothesis that crosslinguistic variation should be confined to lexical properties, not the computational system (see sec. 1.1). The computational system itself has always been more of a focus of research than lexical properties, being regarded as an inherently more interesting subject of inquiry. Once the principles vs parameters distinction got translated into a computational system vs lexical properties distinction, the decline of research interest into the latter was inevitable. The emergence of generative typology, which by its nature looks at crosslinguistic variation, should, however, revive research interest into the latter (all this also ties with language acquisition, where the researchers have always been concerned with acquisition of parametric variation).

4. PICKY MIDDLE WAY

The most convenient way of doing typological work (TT or GT) is to base it on descriptive works dealing with particular languages. An issue that generative typology faces is that generative works typically examine issues that are not addressed in such descriptive grammars (like islands, see e.g. (i) in fn3), which furthermore also focus on what is possible in languages. To deal with this practical problem, Baker & McCloskey (2007: 294) suggest what they call “Middle Way”, “a style of research that would look at fewer languages than a typical typological study does, but at more languages than a typical generative study does”. What makes this particularly feasible is that generative research is generally theoretically guided (which in fact differentiates it from traditional typology). To illustrate, one of Bošković’s article generalizations concerns negative raising, where negative raising (in particular, licensing of strict clause-mate negative polarity items (SCNPIs) under negative raising) is possible only in languages with articles.³⁷ This is a rather surprising interaction, the kind of interaction that cannot be stumbled upon by accident. Indeed, there is a theoretical proposal that ties definiteness and negative raising. Thus, Gajewski (2005) treats negative raising predicates as plural definite descriptions – they essentially combine a modal base (set of accessible worlds) with a definite article (see Gajewski 2005, Bošković & Gajewski 2011 for details, which need not concern us here). Without a theoretical proposal that ties definiteness and negative raising, the potential correlation between articles and negative raising (which

³⁷ SCNPIs require clause-mate negation (cf. *he didn’t leave until yesterday* and **he left until yesterday*, the underlined element being a SCNPI), except when embedded under a negative raising predicate (compare *he didn’t believe that Ann would leave until tomorrow* and **he didn’t claim that Ann would leave until tomorrow*, *believe*, but not *claim*, being a negative raising predicate). SCNPI licensing under negative raising is disallowed in e.g. SC, Czech, Slovenian, Polish, Russian, Ukrainian, Turkish, Korean, Japanese, and Chinese, which lack definite articles, but allowed in e.g. English, German, Spanish, French, Portuguese, Romanian, and Bulgarian, which have articles.

from the point of view of descriptive grammars (which is rather surprising) would not have been checked.³⁸

In light of GT being theory-driven, using theory as its guide (which differentiates it from TT), I would like to slightly amend the research strategy suggested by B&M, renaming it *Picky Middle Way*. Picky Middle Way is guided by investigation of a particular theoretical mechanism, and picks examples from different, unrelated languages to illuminate that theoretical mechanism, without need to offer a detailed investigation of each language considered that would go beyond the mechanism in question; even when particular constructions are used, they are discussed only to the extent that they bear on the theoretical mechanism in question. Investigation of extraction out of moved elements in Bošković (2018) and out of conjuncts in Bošković (2020), whose goal was to show that, contrary to what is standardly assumed, such extractions are possible as well as to determine the exact contexts under which they are possible, can be taken as examples of Picky Middle Way since these works pick particular (crucially different) constructions from a variety of languages (Serbo-Croatian, Dutch, Janitzio P'urhepecha...) to examine particular topics which in turn illuminate the relevant theoretical issues, islandhood in this particular case.

A practical issue arises with the reviewing of Picky Middle Way works though. Many generativists spend their lifetime research doing detailed investigations of constructions of a single language. When faced with a “Picky Middle Way” work, where certain aspects of construction X from language Y are used to illuminate theoretical mechanism Z, they often require detailed discussion of all aspects of construction X (since this is how they write their papers), although those aspects have nothing to do with Z. This is a serious practical problem that the development of generative typology is facing.

³⁸ Needless to say, such theory-driven investigation of potential typological generalizations does not always work out, but when it does it also provides evidence for the relevant theoretical proposals (in this case it provides evidence for a semantic approach to negative raising (a contentious issue, see e.g. Collins and Postal 2014 for a syntactic account), in fact a particular semantic account).

Investigations of UG in the generative framework used to rely on detailed investigations of particular languages. In the old-habits-are-hard-to-change spirit, those who are used to such investigations are often the reviewing stop sign for typological works since they often require the same kind of detailed investigation for every individual language considered in a typological work that would be found in a work devoted to just one language.³⁹

5. FINAL DIFFERENCES AND REMARKS

In this section I will discuss three remaining issues, which concern what many consider to be the biggest differences between the two camps, the issues in question being left for the end of the paper in the hope that the preceding discussion has taken the edge of some of the perceived irreconcilability and antagonism between the two camps. Another reason why these issues (some of which were touched upon above) are discussed together here is that they are often lumped together under “faculty of language” (with the attitude you believe in all of it, i.e. the whole package, or not) – it’s, however, important to separate them to see where the real disagreement lies. In that vein, I will “decompose” the generativist’s faculty of language into three distinct notions: universal grammar, innatism, and domain specificity.

³⁹ As an example, I used the Japanese numeral floating construction in a typological investigation of extraction from conjuncts. A reviewer, clearly a Japanese syntax specialist used to writing papers involving detailed studies of particular constructions of Japanese, kept on insisting the paper should have a detailed discussion of the construction in question that would discuss in detail the issue of whether floating numerals should be analyzed the way Sportiche (1988) treats similar elements in English or as adjuncts although this was totally irrelevant to the issue under discussion – nothing would have changed regarding that issue regardless of which approach to the Japanese construction in question was taken. But that didn’t matter to the reviewer. The attitude was you cannot use construction X unless you discuss all aspects of construction X, period.

I will first discuss the notion of universal grammar, which should be the least controversial (the issue was also discussed in section 1.3).

In spite of the pronouncements that are associated with the respective camps, it is not clear that there is a real fundamental disagreement when it comes to “believing in universal grammar”. UG is a rather loaded term, let us put it aside for a moment because of that, and use a less loaded term “general properties of language”. Traditional functional typologists are concerned with uncovering general properties of language – this is in fact what Greenberg’s generalizations are all about. The same holds for generativists – they just call those general properties of language UG. But what is important is that both camps seem to believe that there are general properties of language. There are differences regarding how those general properties of language look like, but that is not the point here – the point is that both camps do hold that there are such things (otherwise they would not be trying to uncover them). We can call them general properties of language or UG, or α for that matter, the terminology itself does not affect the point.

Innatism is another widely held difference between the two camps, but it is not clear that that difference is real either. To see this, let us move to another difference between the two camps, which does seem more significant, namely what is behind those general properties of language: is it (whatever it is) domain specific (the generativists often use the term faculty of language – FL) or are we dealing with broader cognitive mechanisms, which are not domain specific. In principle, innatism can be attached to either position (many of those broader cognitive mechanisms that the functionalists appeal to are obviously innate), so at least in principle, innatism is not necessarily a fundamental difference between the two camps (again, the point is not how those things that would be held to be innate would look like, but that in principle at least there can be such things).

So it seems that the major, fundamental difference boils down to domain specificity vs broader cognitive mechanisms. But even there the difference is not as black-and-white as it is generally held to be. As discussed above, the minimalist tenet that as much as possible

should follow from the nature of language, which includes functional considerations and paying serious attention to the systems language is interfacing with (the articulatory-perceptual and conceptual-intentional, which themselves are not part of FL), leads to a maximize functional/bare-output conditions considerations and minimize formal considerations strategy (formal reasons that have nothing to do with the nature of language should be minimized), which in the bigger picture means that broader cognitive mechanisms (henceforth BCM) need to be paid attention to – generativists should not wave an FL flag as an excuse to ignore them. But the interaction between the relevant mechanisms is such that it actually blurs the FL/BCM line, to the extent that even in this case we cannot be talking about a clear line-in-the-sand demarcation between the two camps.⁴⁰

To illustrate, consider Kayne's (1994) Linear Correspondence Axiom (LCA), a proposal where a structural relation, asymmetric *c*-command (where *X* *c*-commands *Y* but *Y* does not *c*-command *X*) determines word order (i.e. if *X* asymmetrically *c*-commands *Y*, *X* precedes *Y* and everything *X* dominates precedes everything *Y* dominates). Follow up works like Kayne (2010), Di Sciullo (2015), Bošković (2021a) have considerably expanded the domain of the LCA. In fact, Bošković argues FL quite generally favors asymmetric relations, positing a general Maximize Asymmetric Relations (MAR) preference, which he shows unifies a number of phenomena,⁴¹ namely the diachronic loss of specifiers,⁴² their avoidance in language

⁴⁰ More generally, what we may be dealing with here, and which would blur the line in question, is a situation involving evolutionary developments in the organization of the human brain which would represent adaptations specifically for language, where pre-existing more general cognitive structures/mechanisms got adapted (or exapted) to be used specifically for language (for a recent general discussion, see Hauser, Chomsky, and Fitch 2002).

⁴¹ For reasons of space, the following discussion is rather packed; at any rate its goal is merely to give an impression how wide the scope of the broadening of Kayne's original LCA is (see the cited work for details).

⁴² To illustrate with questions/wh-movement, Dadan (2019) shows that the general direction of diachronic change is from wh-movement to wh-in-situ (we don't

acquisition, the LCA, the Phase Impenetrability Condition (which, among other things, is crucially involved in determining locality of movement), the no-Spec-without-complement aspect of bare phrase structure (where a head cannot take a specifier unless it also takes a complement), the crosslinguistic rarity of multiple wh-fronting (which may be expected to be quite common since it has been argued to have the most transparent and simplest syntax-semantics mapping), and the mysterious *who left* effect, where subject wh-movement cannot proceed through SpecIP, i.e. the subject position, although this position otherwise must be filled in English (see Bošković 2016a, Messick 2020 for relevant arguments). Bošković (2021a) also shows that MAR has broad theoretical consequences, e.g. regarding structure building, Case-licensing and the EPP. While Kayne's original LCA looked FL-specific, the broadening raised the question of whether MAR is an irreducible formal property of FL or it can be traced back to FL-external factors, even related to broader cognitive mechanisms.⁴³ The preference nature of MAR may in fact suggest the latter, if, as often assumed (e.g. Baker & McCloskey 2007), tendencies and preferences, as opposed to absolute properties, are more likely to have functional/broader cognitive than formal/FL-specific origins. Furthermore, Kayne (2010) observes that both parsing and language production are asymmetric, in that they show a beginning vs end asymmetry. If there is indeed a connection here, it would suggest that we may be dealing with a broader property of language, which goes beyond linguistic competence, i.e. knowledge of language, that

find changes from wh-in-situ to wh-movement); e.g. Old Japanese, Archaic Chinese, and Sanskrit were wh-movement languages, while Japanese, Chinese, and modern Indic languages are wh-in-situ languages. Bošković (2021a) shows that the MAR-based account of this uni-directionality of diachronic change also explains why certain cases of wh-fronting are more resistant to loss than others (this has to do with wh-fronting not targeting the same position in all languages).

⁴³ If we are indeed dealing with a broader property here, Bošković (2021a) can be taken as an illustration of how looking seriously into formal properties of language can help us elucidate those broader mechanisms.

the generativists are generally concerned with – the property would extend to performance, i.e. usage of that knowledge.

Now, Minimalist Program explores the possibility that UG, taken as a property of FL, is an optimal way of satisfying requirements that are imposed on FL by the external systems FL interfaces with. From this perspective, asymmetry can be encoded in UG even if its ultimate source is those external systems. On this view, asymmetry would essentially be imposed on UG as the optimal way of satisfying external system requirements. What is important for the general point made here is that this kind of situation blurs the line between FL (i.e. language-specific) and BCM (i.e. broader cognitive mechanism), which is often taken to be a clear line in the sand (in fact a differentiating line between the generative (more precisely, Chomskian) and the traditional functional camp): something can in a way be both (if it is language-specific as a reflection of broader cognitive mechanisms). In fact, this is a pretty common situation.⁴⁴ Consider the very basic question, why do we have movement?⁴⁵ It is there essentially for functional reasons – due to broader cognitive needs to express things (as noted above, movement enables us to express more complex semantic and pragmatic relations that go beyond basic argument structure; as often pointed out, movement is language-specific – nothing similar to movement is found in other symbolic systems, where the relevant cognitive needs are not present). They got encoded in UG due to the mechanism of movement, with a number of constraints on movement that do appear to be language specific. But the existence of movement is a reflection of broader cognitive mechanisms/needs (see also the discussion in the end of sec. 2.2).

As another illustration, suppose there is something like (30), which is a simplified version of Rizzi's (1997) split CP, where ForceP

⁴⁴ For a smaller-scale situation along these lines, see the discussion of person restrictions in section 2.2.

⁴⁵ Or whatever is used to encode the fact that elements often don't occur in positions where they are interpreted, as in *Him, Sue likes*, where *him* is interpreted as the object of *like* but does not occur in that position.

indicates the force of a sentence, TopP houses topicalized and FocP focalized elements. Bianchi & Frascarelli (2010) further argue for splitting TopP into projections for three different types of topics, aboutness, contrastive, and given topics, with the hierarchy in (31) ((32) shows aboutness topics precede contrastive topics)).

(30) [_{ForceP} [_{TopP} [_{FocP} [_{IP}]]]]

(31) [_{TopATP} [_{TopCTP} [_{TopGTP} ...]]]

(32) a. (As for) Rosa_{AT}, my next book_{CT} I will dedicate to her.
 b. *My next book, Rosa, I will dedicate to her.

A question that arises here is what is responsible for these hierarchies? If the answer is FL, a further question arises: why did FL pick these particular hierarchies? A functionalist would complain that saying that the fixed order of topics is due to a structural hierarchy does not explain anything, why do we have those structural hierarchies? That is a very legitimate complaint. We should try to understand this better, not simply stipulate a structural hierarchy. There are in fact proposals even within generative approaches that the ultimate reasons are semantic/pragmatic, see e.g. Neeleman & de Koot (2008) and Lacerda (2020).⁴⁶ If those structural hierarchies are there for reasons that are ultimately semantic/pragmatic, what do we mean by semantic/pragmatic? The natural answer is that what is meant by that is what Chomsky (1995) refers to as the conceptual-intentional system. But the conceptual-intentional system is FL-external, which means that we would then be dealing here with what I have referred to above as broader cognitive mechanisms. This is, then, another illustration of the give-and-take between FL and BCM which blurs the line between the two.

More generally, while it is not hard to draw a line between what the generativists refer to as FL and language-external mechanisms on the sound side (i.e. form side), it is much harder to do that on the

⁴⁶ These works adopt a different, mapping approach to discourse relations, but the details are not important to the general point made here, which is why e.g. topics precede focus, or why different types of topics are subject to a hierarchy (Neeleman & de Koot discuss the former, while Lacerda also discusses the latter).

meaning side. While not often explicitly articulated, it is generally implied that most of what formal semantics does is outside of overt and covert syntax, which means that it would belong to the conceptual-intentional system, i.e. it is outside of FL. Chomsky (2020), however, suggests that it is actually part of FL (for relevant discussion, see also Hauser, Chomsky, and Fitch 2002). The issue is in fact far from trivial,⁴⁷ but almost impossible to address in our current state of understanding – it has to do with what should be the demarcation line between language and our thoughts. This underscores how difficult it is to draw a clear line between the boundary of FL and the external systems it interfaces with (i.e. broader cognitive mechanisms) on the meaning side.

At any rate, the more general point made here is that generative and traditional functional approaches may not be as fundamentally opposed to each other as is generally widely believed or as the slogans standardly associated with these approaches suggest. What are standardly assumed to be clear lines in the sand demarcating these two approaches in many cases may not be there, or they are rather blurred. A wider realization that this is indeed the case should provide an opening for a rapprochement between the two camps. The first step in that direction should be talking to each other, something that is now sorely lacking. The goal of this paper is to make a small contribution to this effect.

6. CONCLUSION

The overarching point of this paper is that there should be less disagreement between the generative and the functional camp, as well as generative typology and traditional typology (which is often functionally based), than there actually is. Much of the disagreement

⁴⁷ To put it more bluntly, assuming there is FL, and that this is what linguists investigate, the issue is whether a formal semanticist like Partee is a linguist or a psychologist (it would be the former if formal semantics is part of FL, and the latter if it investigates the conceptual-intentional system).

comes from the two camps not interacting with each other, and misunderstandings of the slogans which are used to characterize the two camps, where the practitioners of the two camps are not really trying to see what is behind those slogans (the former is partly responsible for the latter); what also does not help in this respect is that with both camps, there are differences between what is said and what is really believed (as reflected in the actual research practice). There is a warzone out there, which is really unwarranted – there is actually little true opposition, what the two camps do is largely complementary, which is not easy to see because of the slogans that are constantly repeated but which should not be taken at face value. It's almost like a situation where a war lasts for so long that people forget what it is all about or don't see that the reasons for the war no longer hold. In this particular case, there is no real scientific reason for the war (that of course does not include human nature and sociological factors⁴⁸). Additionally, with the minimalist tenet that as much as possible should follow from the nature of language, which by necessity includes functional considerations (and in fact naturally leads to the maximize-functional-explanations strategy), and especially the emergence of generative typology, which also attempts to find Greenberg-style typological generalizations (and which is increasingly seen as essential in generative circles), we are actually reaching an exciting point of potential convergence, where the two camps should be talking more to each other, and interacting productively with each other. There will still be competition between the two approaches regarding analyses of particular phenomena: linguistic phenomena do not come with labels regarding whether a formal or a functional explanation is more appropriate for them. This is healthy competition of the kind we are all used to.⁴⁹ Importantly, there should still be

⁴⁸ Longing for good old times with fixed demarcation lines between good and bad guys (with clear good and bad guys) may also take time to overcome.

⁴⁹ E.g., it is often not clear whether a particular phenomenon should receive a syntactic or a semantic treatment within the same overall approach, a particularly common situation nowadays in the generative circles, where semantics has been increasingly gaining in importance (with a training in semantics essentially

competition but no incompatibility, i.e. antagonism that comes from opposing worldviews regarding the subject matter that would deny to the opposing camp the scientific right to exist.⁵⁰ In other words, the two camps should be increasingly seen as complementary, much like syntax and semantics or syntax and morpho-phonology are seen as complementary. Typology is really crucial here; it is in fact what is getting the formalists to engage with the functionalists – if you are a generative typologist you have to read functionalist works, you cannot ignore them. The problem before, which is to a great extent responsible for antagonism between the two camps, was precisely that they were ignoring each other. The emergence of generative typology is getting formalists not to ignore functionalists, to read the relevant functional literature; hopefully, the more traction and more importance typology gets in formal circles, which will inevitably lead to more typological work being done within the formalist camp, traditional functional-

becoming a must for a generative syntactician), which has turned out to be very productive. Typical theoretical syntax oriented department will house a semanticist, due to the correct belief that theoretical syntax and semantics are not fundamentally incompatible and that a syntactician should be familiar with semantics (at least to be able to recognize a semantic issue when it arises). Such a department does not house a functionalist (the same holds in the other direction), but there is no real (non-sociological) reason why it shouldn't – the situation here is, or should be, the same as with theoretical syntax and semantics. We should be moving from the current, "our" vs "their" department situation to a predominantly "our" vs predominantly "their" department situation. The field (and the training the students get) will be better for it. The students need to be exposed to both views, at least to be familiar with them (so that they can recognize when a formal/functional issue arises, i.e. when a particular investigation starts crossing into the other domain), just like a syntactician needs to be exposed to semantics, and a semanticist needs to be exposed to syntax. In all these cases we are dealing with complementarity, not mutually incompatible worldviews. The latter would rightly require complete department separations (essentially to avoid a warzone). We are still in the separation mode, but we shouldn't be.

⁵⁰ Upon some reflection, no sane generativist would claim that there is no aspect of language that can be explained by looking at the function of language. The antagonism between the two camps, however, can be so strong to override the "upon-some-reflection" requirement.

based typologists will find themselves in the same position. The surest way to antagonism is ignorance, i.e. ignoring, not interacting, not reading each other's works (or reading the works from the other camp only to talk to your own camp about them in order to make fun of the other "side" based on a cartoonish version of the other side). With the emergence of generative typology, which is rapidly gaining traction within the generative camp, typology in general will increasingly make this kind of ignoring more difficult, leading to hopefully more productive interaction between the two camps.

REFERENCES

- Abels, K. (2003). *Successive cyclicity, anti-locality, and adposition stranding*. (Publication No. AAI3104085) [Doctoral dissertation, University of Connecticut]. UCONN Library. <https://opencommons.uconn.edu/dissertations/AAI3104085>
- Albizu, P. (1997). *The syntax of person agreement*. [Doctoral dissertation, University of Southern California].
- Alexander, M. (1990). *Agreement configurations: Grammatical relations in modular grammar*. [Doctoral dissertation, MIT].
- Anagnostopoulou, E., Alexiadou, A., & Everaert, M. (2004). Introduction. In E. Anagnostopoulou, A. Alexiadou, & M. Everaert (Eds.), *The unaccusativity puzzle: Explorations of the syntax-lexicon interface* (pp. 1–21). Oxford: Oxford University Press.
- Anagnostopoulou, E. (2003). *The syntax of ditransitives: Evidence from clitics*. Berlin: Gruyter.
- Baker, M., & McCloskey, J. (2007). On the relationship of typology to theoretical syntax. *Linguistic Typology* 11, 285–296.
- Baker, M. (1996). *The polysynthesis parameter*. New York: Oxford University Press.
- Baker, M. (2008). *The syntax of agreement and concord*. Cambridge University Press.
- Baker, M. (2011). The interplay between Universal Grammar, universals, and lineage specificity. *Linguistic Typology* 15, 473–482.
- Baker, M. (2015). Formal generative typology. In B. Heine, & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (2nd ed., pp. 925–951). Oxford: Oxford University Press.

- Béjar, S., & Rezáč, M. (2009). Cyclic Agree. *Linguistic Inquiry* 40, 35–73.
- Bianchi, V., & Frascarelli, M. (2010). Is Topic a root phenomenon? *An International Journal of Theoretical Linguistics* 2, 43–88.
- Bickel, B. (2007). Typology in the 21st century: Major current developments. *Linguistic Typology* 11, 239–251.
- Boas, F. (1911). Introduction. In F. Boas (Ed.), *Handbook of American Indian languages* (pp. 1–83). Washington DC: Bureau of American Ethnology.
- Bobaljik, J., & Wurmbrand, S. (2001). Seven prefix-suffix asymmetries in Itelmen. *Proceedings of CLS* 37, 205–219.
- Boeckx, C. (1999). *Properties of interrogatives in French*. [Master's thesis, University of Connecticut].
- Boeckx, C. (2008). *Bare Syntax*. Oxford: Oxford University Press.
- Bond, O. (2019). Canonical Typology. In J. Audring, & F. Masini (Eds.), *The Oxford Handbook of Morphological Theory* (pp. 409–431). Oxford: Oxford University Press.
- Borer, H. (1984). *Parametric syntax*. Dordrecht: Foris.
- Bošković, Ž. (2002). On multiple wh-fronting. *Linguistic Inquiry* 33, 351–383.
- Bošković, Ž. (2005). Left branch extraction, structure of NP, and scrambling. In J. Sabel, & M. Saito (Eds.), *The free word order phenomenon: Its syntactic sources and diversity* (pp. 13–73). Berlin: de Gruyter.
- Bošković, Ž. (2008). What will you have, DP or NP? *Proceedings of NELS* 37, 101–114.
- Bošković, Ž. (2009). More on the no-DP analysis of article-less languages. *Studia Linguistica* 63, 187–203.
- Bošković, Ž. (2010). Another monster. Presented at *MayFest 2010: Bridging typology and acquisition*, University of Maryland.
- Bošković, Ž. (2012). On NPs and clauses. In G. Grewendorf, & T. E. Zimmermann (Eds.), *Discourse and grammar: From sentence types to lexical categories* (pp. 179–242). Berlin: de Gruyter.
- Bošković, Ž. (2016a). On the timing of labeling: Deducing Comp-trace effects, the Subject Condition, the Adjunct Condition, and tucking in from labeling. *Linguistic Review* 33, 17–66.
- Bošković, Ž. (2016b). On second position clitics crosslinguistically. In F. L. Marušič, & R. Žaucer (Eds.), *Formal Studies in Slovenian Syntax. In honor of Janez Orešnik* (pp. 23–44). Amsterdam: John Benjamins.
- Bošković, Ž. (2016c). On the typology of the nominal domain: Its syntax, semantics, and learnability. Presented at *Workshop on the Syntax and Semantics of the Nominal Domain*, Goethe-Universität Frankfurt.

- Bošković, Ž. (2018). On movement out of moved elements, labels, and phases. *Linguistic Inquiry* 49, 247–282.
- Bošković, Ž. (2020). On the Coordinate Structure Constraint, across-the-board movement, phases, and Labeling. In J. van Craenenbroeck, C. Pots, & T. Temmerman (Eds.), *Recent developments in Phase Theory* (pp. 133–182). Berlin: de Gruyter.
- Bošković, Ž. (2021a). Generalized Asymmetry. In P. Kosta, & K. Schlund (Eds.), *Keynotes from the International Conference on Explanation and Prediction in Linguistics: Formalist and Functionalist Approaches* (pp. 15–77). Berlin: Peter Lang.
- Bošković, Ž. (2021b). On the (im)practicality of the state: Why do I have to have a country? In G. Grewendorf (Ed.), *Chomsky on State and Democracy* (pp. 77–94). Nomos Verlag.
- Bošković, Ž., & Gajewski, J. (2011). Semantic correlates of the NP/DP parameter. *Proceedings of NELS 39*, 121–134.
- Bošković, Ž., & Şener, S. (2014). The Turkish NP. In P. C. Hofherr, & A. Zribi-Hertz (Eds.), *Crosslinguistic studies on nominal reference: With and without articles* (pp. 102–140). Leiden: Brill.
- Bresnan, J., & Mchombo, S. (1987). Topic, pronoun, and agreement in Chichewa. *Language* 63, 741–782.
- Brown, D., Chumakina, M., & Corbett, G. G. (2013). *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Brown, S., & Franks, S. (1995). Asymmetries in the scope of Russian negation. *Journal of Slavic Linguistics* 3, 239–287.
- Bybee, J. L., & Dahl, Ö. (1989). The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13, 51–103.
- Cheng, H. (2013). *Argument ellipsis, classifier phrases, and the DP parameter*. [Doctoral dissertation, University of Connecticut].
- Cheng, L., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30, 509–542.
- Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics* 6, 339–405.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MIT Press.
- Chomsky, N. (2000). Minimalist inquiries. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–155). Cambridge: MIT Press.

- Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *The cartography of syntactic structures Vol. 3* (pp. 104–131). Oxford: Oxford University Press.
- Chomsky, N. (2020). The UCLA lectures. lingbuzz/005485
- Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Cinque, G. (2004). “Restructuring” and functional structure. In A. Belletti (Ed.), *Structures and Beyond: The Cartography of Syntactic Structures* (pp. 132–191). Oxford: Oxford University Press.
- Cinque, G. (2005). Deriving Greenberg’s universal 20 and its exceptions. *Linguistic Inquiry* 36, 315–332.
- Collins, C., & Postal, P. (2014). *Classical NEG raising: An essay on the syntax of negation*. Cambridge: MIT Press.
- Corbett, G. G. (1983). *Hierarchies, targets and controllers: Agreement patterns in Slavic*. University Park: Pennsylvania University Press.
- Corbett, G. G. (2007). Canonical typology, suppletion, and possible words. *Language* 83, 8–42.
- Corver, N. (1992). On deriving left branch extraction asymmetries: a case study in parametric syntax. *Proceedings of NELS 22*, 67–84.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Dadan, M. (2019). *Head labeling preference and language change*. [Doctoral dissertation, University of Connecticut].
- Dayal, V. (2004). Number marking and (In)definiteness in kind terms. *Linguistics and Philosophy* 27, 393–450.
- DeLancey, S. (2001). *Lectures on functional syntax*. LSA Summer Institute, UC Santa Barbara.
- Despić, M. (2011). Syntax in the absence of determiner phrase. [Doctoral dissertation, University of Connecticut].
- Despić, M. (2013a). Intensifiers, focus, and clitics: is pronoun position truly an argument for D in Serbo-Croatian? In L. Schürcks, A. Giannakidou, U. Etxeberria, & P. Kosta (Eds.), *Nominal Constructions in Slavic and Beyond* (pp. 39–74). Berlin: Mouton de Gruyter.
- Despić, M. (2013b). Binding and the structure of NP in Serbo-Croatian. *Linguistic Inquiry* 44, 239–270.
- Despić, M. (2019). On kinds and anaphoricity in languages without definite articles. In A. Aguilar-Guevara, J. Pozas Loyo, & V. Vázquez-Rojas Maldonado (Eds.), *Definiteness across languages* (pp. 259–291). Berlin: Language Science Press.

- Di Sciullo, A. M. (2015). On the domain specificity of the human language faculty and the effects of principles of computational efficiency: contrasting language and mathematics. *Revista Lingüística* 11, 57–70.
- Dryer, M. S. (1997). Are grammatical relations universal? In J. L. Bybee, J. Haiman, & S. A. Thompson (Eds.), *Essays on language function and language type: Dedicated to T. Givón* (pp. 115–143). Amsterdam: John Benjamins.
- Epstein, S. (1999). Un-principled syntax: The derivation of syntactic relations. In S. Epstein, & N. Hornstein (Eds.), *Working Minimalism* (pp. 317–345). Cambridge: MIT Press.
- Erlewine, M. Y. (2016). Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language and Linguistic Theory* 34, 429–479.
- Erlewine, M. Y. (2020). Anti-locality and subject extraction. *Glossa* 5, 1–38.
- Franks, S. (2019). Binding and phasehood in south Slavic revisited. *Studies in Polish Linguistics* 14, 61–80.
- Fukui, N. (1988). Deriving the differences between English and Japanese. *English Linguistics* 5, 249–270.
- Gajewski, J. (2005). Neg-raising: Polarity and presupposition. [Doctoral dissertation, MIT].
- Gil, D. (2020). Describing languoids: When incommensurability meets the language-dialect continuum. *Linguistic Typology* 20, 439–462.
- Golan, Y. (1993). Node Crossing Economy, Superiority and D-linking. [Master's thesis, Tel Aviv University].
- Golumbia, D. (2010). Minimalism is functionalism. *Language Sciences* 32, 28–42.
- Greenberg, J. (1974). *Language typology: A historical and analytic overview*. Berlin: Gruyter.
- Guasti, T., Gavarro, A., de Lange, J., & Caprin, C. (2008). Article Omission across Child Languages. *Language Acquisition* 15, 89–119.
- Harley, H., & Ritter, E. (2002). Person and number in pronouns. *Language* 78, 482–526.
- Haspelmath, M. (1997). *Indefinite pronouns*. Oxford: Oxford University Press.
- Haspelmath, M. (2004). Explaining the ditransitive person-role constraint: A usage-based approach. *Constructions* 2, 1–49.
- Haspelmath, M. (2010). Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86, 663–687.
- Haspelmath, M. (2020). General linguistics must be based on universals (or nonconventional aspects of language). [lingbuzz/005158](https://lingbuzz.org/005158).
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

- Hawkins, J. (2010). Processing efficiency and complexity in typological patterns. In J. Jung Song (Ed.), *Oxford handbook of linguistic typology* (pp. 1–16). Oxford: Oxford University Press.
- Huang, C-T. J. (1982). Logical relations in Chinese and the theory of grammar. [Doctoral dissertation, MIT].
- Jenks, P. (2018). Articulated definiteness without articles. *Linguistic Inquiry* 49, 501–536.
- Jiang, L. (2012). Nominal arguments and language variation. [Doctoral dissertation, Harvard].
- Julien, M. (2002). *Syntactic Heads and Word Formation*. Oxford: Oxford University Press.
- Kang, J. (2014). On the Absence of TP and its Consequences: Evidence from Korean. [Doctoral dissertation, University of Connecticut].
- Kayne, R. (1994). *The antisymmetry of syntax*. Cambridge: MIT Press.
- Kayne, R. (2010). Why are there no directionality parameters? *Proceedings of WCCFL* 28, 1–23.
- Koulidobrova, E. (In press). DP or not DP: Acquisitional evidence from English. *Language Acquisition*.
- Kuno, S., & Takami, K. (2004). *Functional constraints in grammar: On the unergative-unaccusative distinction*. Amsterdam: John Benjamins.
- Lacerda, R. (2020). Middle-field syntax and information structure in Brazilian Portuguese. [Doctoral dissertation, University of Connecticut].
- Lander, Y., & Arkadiev, P. (2020). On the right of being a comparative concept. *Linguistic Typology* 20, 403–416.
- LaTerza, I. (2016). Binding in English and South Slavic and the parameterized DP hypothesis. *Linguistic Inquiry* 47, 741–753.
- Ledgeway, A. (2012). *From Latin to Romance: Morphosyntactic typology and change*. Oxford: Oxford University Press.
- Lehmann, C. (1989). Language description and general comparative grammar. In G. Graustein, & G. Leitner (Eds.), *Reference grammars and modern linguistic theory* (pp. 133–62). Tübingen: Niemeyer.
- Levin, B., & Hovav, M. R. (1995). *Unaccusativity at the syntax-lexical semantics interface*. Cambridge: MIT Press.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Marelj, M. (2011). Bound-Variable Anaphora and Left Branch Condition. *Syntax* 14, 205–229.
- Matthewson, L., Bryant, T., & Roeper, T. (2001). A Salish stage in the acquisition of English determiners: Unfamiliar “definites”. *Proceedings of SULA* 25, 63–71.

- McCloskey, J. (2000). Quantifier float and wh-movement in an Irish English. *Linguistic Inquiry* 31, 57–84.
- Messick, T. (2016). Pronouns and agreement in Telugu embedded contexts. *Proceedings of WCCFL 33*, 309–319.
- Messick, T. (2017). The morphosyntax of self-ascription: A Cross-linguistic Study. [Doctoral dissertation, University of Connecticut].
- Messick, T. (2020). The derivation of highest subject questions and the nature of the EPP. *Glossa* 5, 1–12.
- Neeleman, A., & van de Koot, H. (2008). Dutch scrambling and the nature of discourse templates. *Journal of Comparative Germanic Linguistics* 11, 137–189.
- Newmeyer, F. J. (1998). *Language form and language function*. Cambridge: MIT Press.
- Nichols, J. (2007). What, if anything, is typology? *Linguistic Typology* 11, 231–238.
- Pereltsvaig, A. (2007). On the universality of DP. *Studia Linguistica* 61, 59–94.
- Petroj, V. (2020). Ignoring Language Barriers: Romanian-Serbian Code-Switching. [Doctoral dissertation, University of Connecticut].
- Petrosino, R. (2018). The left periphery fragmented: Evidence from Italian. *Proceedings of NELS 48*, 221–230.
- Phillips, C. (1996). Order and structure. [Doctoral dissertation, MIT].
- Polinsky, M., & Kluender, R. (2007). Linguistic typology and theory construction: Common challenges ahead. *Linguistic Typology* 11, 273–283.
- Progovac, Lj. (1998). Determiner Phrase in a language without determiners. *Journal of Linguistics* 34, 165–179.
- Progovac, Lj. (2010). Syntax: Its evolution and its representation in the brain. *Biolinguistics* 4, 234–254.
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Dordrecht: Kluwer.
- Rizzi, L. (2006). On the form of chains: Criterial positions and ECP effects. In L. Cheng, & N. Corver (Eds.), *Wh-movement: Moving on* (pp. 97–133). Cambridge: MIT Press.
- Roberts, I. (2019). *Parameter hierarchies and Universal Grammar*. Oxford University Press.
- Round, E., & Corbett, G. (2020). Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology* 24, 489–525.

- Rudin, C. (1988). On multiple questions and multiple wh-fronting. *Natural Language and Linguistic Theory* 6, 445–501.
- Runić, J. (2014). A new look at clitics, clitic doubling, and argument ellipsis. [Doctoral dissertation, University of Connecticut].
- Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt.
- Spike, M. (2020). Fifty shades of grue: Indeterminate categories and induction in and out of the language sciences. *Linguistic Typology* 24, 465–488.
- Sportiche, D. (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19, 425–449.
- Stegovec, A. (2019). Person on the Edge: Lessons from Crosslinguistic Variation in Syntactic Person Restrictions. [Doctoral dissertation, University of Connecticut].
- Takahashi, M. (2011). Some consequences of Case-marking in Japanese. [Doctoral dissertation, University of Connecticut].
- Talić, A. (2013). Adverb extraction, specificity, and structural parallelism. *Canadian Journal of Linguistics* 60, 417–445.
- Thomas, M. (2020). *Formalism and functionalism in linguistics: The engineer and the collector*. New York: Routledge.
- Todorović, N. (2016). On the Presence/Absence of TP: Syntactic Properties and Temporal Interpretation. [Doctoral dissertation, University of Connecticut].
- Trenkić, D. (2004). Definiteness in Serbian/Croatian/Bosnian and some implications for the general structure of the nominal phrase. *Lingua* 114, 1401–1427.
- Uriagereka, J. (1988). On government. [Doctoral dissertation, University of Connecticut].
- Willim, E. (2000). On the grammar of Polish nominals. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step* (pp. 319–346). Cambridge: MIT Press.
- Wurmbrand, S. (2001). *Infinitives: Restructuring and Clause Structure*. Berlin: de Gruyter.
- Zanon, K. (2015). On hybrid coordination and quantifier raising in Russian. [Doctoral dissertation, University of Indiana, Bloomington].
- Zanon, K. (2020). Negated polar interrogatives in Russian. Presented at *Slavic Linguistics Society* 15. Indiana University.

- Zlatić, L. (1997). The structure of the Serbian Noun Phrase. [Doctoral dissertation, University of Texas at Austin].
- Zubizarreta, M. L. (2003). Intervention effects in the French wh-in-situ construction: Syntax or interpretation? In R. Nuñez Cedeño, L. Lopez, & R. Camero (Eds.), *A Romance perspective in language knowledge and use – Selected papers from the 31st Linguistic Symposium on Romance Language* (pp. 359–379). Amsterdam/Philadelphia: John Benjamins.

Željko Bošković

FORMALISME ET, PAS VS, FONCTIONNALISME

Résumé

Baker & McCloskey (2007) examinent la relation entre la typologie traditionnelle et la syntaxe théorique générative. Depuis lors, la typologie a commencé à jouer un rôle de plus en plus important au sein de ce dernier domaine au point que l'on peut en fait parler de typologie générative. Étant donné que la typologie traditionnelle est généralement associée à des approches fonctionnelles (voir par exemple Nichols 2007), cela nous donne alors deux approches de la typologie, que j'appellerai typologie traditionnelle (TT) et typologie générative (TG). D'une certaine manière, la typologie prépare donc le terrain pour un rapprochement potentiel de l'approche fonctionnelle et formelle du langage. Cet article fournira plusieurs remarques à cet effet, dans une vision unificatrice où ces deux approches ont chacune leur place, c'est-à-dire où elles ne sont pas en concurrence l'une avec l'autre.

À cette fin, l'article montrera que beaucoup de ce que l'on suppose être des lignes de démarcation claires entre le camp générativiste traditionnel et le camp fonctionnaliste/typologique traditionnel (et les deux approches de la typologie) ne sont pas claires, en fait peuvent ne pas être là du tout — les deux approches ne s'opposent plus autant qu'autrefois et s'opposent moins que les praticiens des deux camps semblent encore le penser. À cet égard, il sera démontré que beaucoup de différences irréconciliables perçues et de l'antagonisme entre les deux domaines sont là parce que, pour le dire d'une manière plus abstraite, il existe des différences entre la situation réelle dans le domaine X et la manière dont le domaine X est perçu par le domaine Y, où la réaction négative de Y à X est basée sur la perception de X par Y.

L'article discutera également des points de convergence entre les deux traditions. Un point de convergence récent concerne en effet l'émergence de la typologie générative. Plusieurs autres points de convergence seront discutés,

y compris l'hypothèse minimaliste selon laquelle le langage est caractérisé par une conception efficace, ce qui ouvre la porte à l'introduction de considérations fonctionnelles dans des approches formalistes comme le minimalisme. Plus généralement, l'article plaide pour une vision globale du domaine où les approches fonctionnelles et formalistes ne sont pas considérées comme concurrentes mais comme complémentaires (une grande partie de ce point de vue reposera sur une réévaluation de certaines questions fondamentales concernant le domaine où les approches formaliste et fonctionnaliste ont été supposés s'opposer (plus précisément, où les praticiens des deux camps se sont activement opposés). Des exemples concrets de complémentarité seront également fournis.

Mots-clés : typologie traditionnelle, typologie générative, approche fonctionnelle, approche formelle, complémentarité

James Joshua Pennington*, PhD
Concordia College
CNN, Semiring Inc.

CYCLICAL SUPPLETION IN SLAVIC: WHAT PRIMING EFFECTS REVEAL ABOUT VELAR ALTERNATIONS IN BOSNIAN, CROATIAN, MONTENEGRIN, SERBIAN

Abstract

This paper presents a wug experiment on BCMS, aimed at ascertaining whether priming effects can trigger the mutation of *k, g, h > c, z, s* upon the addition of the plural morpheme *-i* (cf. feminine dative *-i* and imperatival *-i*) or the retention of the final velar with the addition of the plural allomorph *-ovi*. The definition of suppletion taken here matches that of Corbett (2007), which defines stem suppletion as suppletion par excellence, in opposition to Mel'chuk (1994), who considers lexical replacement to be suppletion proper. Preliminary results suggest priming can affect choice of allomorph, while also indicating a hierarchy of metalinguistic factors at play in speaker choice, including syllable count, phonological neighborhood effects, among others.

Keywords: suppletion, morphonology, allomorphy, BCMS, wug experiment, priming, Slavic Second Palatalization.

* joshua@semiring.com

1. INTRODUCTION

Suppletion has generally occupied a marginalized position in morphological theory: some have considered it “scandalous” due to its defiance of morphological naturalness (Dressler 1985: 97); others wave a flag of caution, calling it “hostile terrain” (Corbett 2007: 8), due to its inability to be neatly captured through rule-based derivation typical of the predominant, 20th-century generative tradition of linguistics (often due to complex diachronic developments, as in the classic case from English *go*.PRS ~ *went*.PST [Janda & Joseph 2003: 109]). Nonetheless, despite the significant challenges suppletion poses to linguistic theory, both historical and synchronic linguists continue to be drawn to the enigmatic topic.

Recent advances have been made both in terms of the diachronic (Juge 2013) and synchronic approaches to suppletion (see Mel’čuk 1994, which views suppletion to be ideally full lexical replacement; Veselinova 2006; Corbett 2007, which, opposite to Mel’čuk 1994 views stem suppletion as the ideal form of suppletion; and Embick 2010). Still, a unified theory eludes us, as common exhortations for fresh approaches to solving this issue in these recent works include [my emphasis]:

“future work should concentrate on [**psycholinguistic**] issues as well” (Veselinova 2006: 108);

or:

“[...] the real issue is whether suppletion is a matter of degree, being the end-point of a scale of irregularity [...] One must hope that eventually there will be clear **psycholinguistic** evidence on this issue” (Corbett 2007: 16);

and:

“the crucial evidence adjudicating between these theories [of stem alternation] might have to come from **psycho- and neurolinguistic** studies of linguistic representation” (Embick 2010: 34).

In the last decade, linguists have heeded these calls, producing two important “wug” (Berko 1958) or nonce word experiments conducted on Russian vowel ~ zero stem alternations of the type *den’*

‘day’ ~ *dnja*.gen.sg (Gouskova and Becker 2013) and Russian verb stem palatalization (Kulinich et al. 2016). The results of these experiments resoundingly demonstrate that the respondents performed poorly in the application of stem alternations to nonce words, indicating that the alternation was not a viable grammatical transformation for these speakers. Similar results were found in a wug experiment on Czech plural allomorphy conducted by Bermel et al. (2017), who state: “If we thus ask the question, ‘which form should native speakers default to?’ there is no clear-cut answer” (ibid.: 4). Although, the results of their experiment more or less confirm this unpredictability, they do note that “stem patterns – representing, possibly, analogical similarity – are critical” (ibid. 16). One specific stem type, velar stems, is at the forefront of stem allomorphy in Slavic due to a phonological development in Late Common Slavic, known as the “Second (Regressive) Palatalization of Velars” (P2) (see Wandl 2020 for a comprehensive review of the topic), formalized phonetically as: $C[+_{\text{back}}] / _V[-_{\text{back}}] \rightarrow C[-_{\text{back}}]$. The variable application of this change is one of the primary features of Slavic dialectal differentiation. According to Kapović (2008), P2 spread through Common Slavic between 6th-7th cc. resulting in nominal stem allomorphy in South and West Slavic, e.g.: BCMS *momak*.nom.sg. ‘boy’ ~ *momci*.nom.PL ‘boys’ or Czech¹ *kluk* ~ *kluci* (the same change occurs in Polish, Sorbian, and Slovak in West, and in Bulgarian and Macedonian in South, excluding Slovenian). However, the change applied only variably throughout the East Slavic dialects (see Wandl 2020).

¹ At the time of the drafting of this article, I recruited a mere 37 respondents for my Czech experiments, and therefore this study will concentrate on my BCMS experiments, which had significantly higher respondent totals (n=179). Czech is occasionally used in a comparative context.

| Sg. | Pl. |
|---|-------------------------------|
| <i>o</i> -stems Proto-Slavic * <i>bogŭ</i> ‘God’ Skt. <i>bhágas</i> | Common Slavic * <i>bogi</i> |
| <i>u</i> -stems * <i>synŭ</i> ‘son’ Lith. <i>sūnùs</i> | Common Slavic * <i>synove</i> |

Figure 1. Before P2

| Sg. | Pl. |
|---------------------|-----------------|
| * <i>bogŭ</i> ‘God’ | * <i>bozi</i> |
| * <i>synŭ</i> | * <i>synove</i> |

Figure 2. After P2

| Sg. | Pl. |
|---------------------|---|
| * <i>bogŭ</i> ‘God’ | * <i>bodzi</i> ~ <i>bogove(i)</i> in West(South) * <i>bogi</i> in East, OR Loc. sg. <i>rukě</i> (cf. Rus dialectal <i>kel-</i> ‘whole’); but Ukr. <i>ruká</i> ~ Loc. sg. <i>rucí</i> |
| * <i>synŭ</i> | * <i>synove</i> |

Figure 3. Late Common Slavic

After the completion of P2, the initial consonantal allophony was phonologized, and then later morphonologized at morpheme boundaries, i.e. the phonological rule $C_{[+back]} / -V_{[-back]} \rightarrow C_{[-back]}$ stem internally was reinterpreted as only operable at morpheme boundaries as a morphonological rule $C_{[+back]} / -+V_{[-back]} \rightarrow C_{[-back]}$ given that the surrounding phonetic environment was not static. This is what many morphologists (namely Greville Corbett [2007]) call “stem suppletion”, which may have triggered the spread of the *u*-stem plurals in *-ove* (West

Slavic) and *-ovi* (South Slavic), presumably to eliminate stem variation. Further, evidence from the treatment of borrowings in BCMS, such as Western Rumelian Turkish *böbrek* > *bubreg*.NSG – *bubrezi*.NPL “kidney” demonstrate that the morphophonemic alternation was still operable at the time of the Ottoman invasion of the Balkans (14th century) on borrowed stems.² Compare this with the recent English borrowing *hotdog* into BCMS: *hotdog* / *hotdogovi*, *blog* / *blogovi* but never **hotdozi* or **blozi* for the plural, suggesting that the alternation is no longer operable on borrowed stems. However, stem variation upon addition of the plural morpheme [-i] still occurs in Slavic (e.g. variant plural forms *vukovi* ‘wolves’ ~ *vuci*, *duhovi* ‘spirits’ ~ *dusi*) and a good account of the situation in BCMS, namely, is sorely lacking. Browne & Alt (2005: 29) make descriptive remarks: “The basic masculine endings are those of *prozor* [nom. pl *prozori*] like most monosyllables and some disyllables, have the “long plural”, adding *-ov-* before plural endings (*-ev-* after palatals)”. Alexander (2006: 38) expands on this:

According to the general rule, monosyllabic masculine nouns add *-ov-* / *-ev-* in the plural, and polysyllabic ones do not. There are two sets of exceptions to this rule – **monosyllabic nouns which do not add it, and disyllabic ones which do. Some members of the first group, such as monosyllabic nouns denoting nationality names, are predictable; others must simply be learned.** The second group contains several disyllabic masculine nouns, usually with fleeting *-a* in their second syllable. **These plurals are not predictable, but must be learned.**

Given all this unpredictability encountered in the literature, a psycholinguistic approach, as provided in the present article, is useful in that it concentrates on identifying the strategies native speakers use in deciding whether or not to mutate a velar-final stem for new words (real or invented).

² This word was likely imported during the Ottoman conquest of the Balkans, 4-5 centuries after the phonological basis (P2) for the alternation was already lost (in Common Slavic). To offer a trivial example to show what happens with root-internal velars, the Turkish drink *rakı* (imported from Arabic ‘*araq* ‘sweat, wine’, entered South Slavic as *rakija*, long after the operation of P2, demonstrating that the alternation was only morphonological in nature.

2. METHODOLOGY

The study follows closely prior work on stem suppletion in Slavic, namely (as mentioned above): Gouskova and Becker (2013), Kulinich, Royle, and Valois (2016), and Bermel, Knittl, and Russell (2017). However, this is the first work to address the situation in BCMS, and especially using Berko’s (1956) “wug” test. In constructing my own wug³ experiment, my goal was to test whether priming affected speaker response, mirroring the presumption that speaker choice of stem suppletion is dictated by metalinguistic pressures (namely that our speech resembles what we hear and read most frequently). I constructed the questionnaires using the open source online linguistic survey tool *Ibexfarm* (developed by Alex Drummond and unfortunately shut down on Sep 30, 2021), with support for PC, MAC, and their respective tablet/smartphone OS.⁴ After constructing my experiment, I distributed an invitation link to a number of my native-speaking BCMS academic colleagues: I posted to numerous BCMS language pages on Facebook; and I sent out messages on several Slavic email listserves (including SEELANGS). The survey’s initial page (Fig. 4) collected demographic data from BCMS speakers (entered manually by keyboard), providing a breakdown according to hometown, sex, age, and education level. My respondents (n=179; see Fig. 4 for breakdown according to hometown) were then instructed to press the spacebar to fill in the blanks on the next page (Fig. 4).

³ When citing my wug forms here, in order to distinguish them from native BCMS words, I will denote them with a ♣, e.g. ♣ dork (an actual wug used in my experiments).

⁴ Fortunately, it is possible to migrate old Ibexfarm experiments to PCibex, which is hosted by the University of Penn and Jeremy Zehr at: <https://farm.pcibex.net/>

Dobar dan! Sprovodim istraživanje na jeziku koji je ranije bio poznat kao srpskohrvatski.

Potrebni su mi izvorni govornici da ispune kratki upitnik.

Upitnik je anoniman, a dobijeni podaci biće korišćeni isključivo u naučne svrhe.

Da biste učestvovali u upitniku potrebno je da imate najmanje 18 godina.

Za ispunjavanje upitnika je potrebno između 10 i 15 minuta.

Upitnik funkcioniše i na mobilnima, samo morate da okrenete ekran horizontalno.

Izuzetno bih vam bio zahvalan na pomoći!

Molimo navedite sljedeće informacije:

Koliko imate godina?

U kojem ste gradu odrastali? (npr. Beograd, Sarajevo, Zagreb)

Spol

M Ž

Da li imate više obrazovanje? (Izaberite „da“ čak i ako niste završili studije)

Da Ne

Na sledećoj stranici pritisnite razmaknicu da biste popunili praznine.

Hvala vam na učešću!

[→ Click here to continue](#)

Figure 4. Welcome page/demographic data collection

| | | | | | | | | | | | |
|------------------|----|-------------|---|---------------|---|-------------------|---|-----------------|---|----------------|----|
| Aleksinac | 1 | Arandelovac | 1 | Backa Palanka | 1 | Banja Luka | 2 | Bela Palanka | 1 | Belgrade | 58 |
| Bijelo Polje | 1 | Cacak | 2 | Čačak | 1 | Daruvar | 1 | Donja Ljubovia | 1 | Dubrovnik | 2 |
| Gornji Milanovac | 1 | Kovacica | 1 | Kragujevac | 2 | Kruševac | 1 | Leskovac | 1 | Loznica | 1 |
| Metković | 1 | Mostar | 1 | Nikšić | 1 | Niš | 8 | NONE | 3 | Nova Gradiška | 1 |
| Novi Sad | 5 | osijek | 2 | Pancevo | 1 | Pančevo | 1 | Paraćin | 2 | Ploče | 1 |
| Podgorica | 3 | Priboj | 1 | Pula | 1 | Rieška | 1 | Rijeka | 1 | Ruma | 1 |
| Sabac | 1 | Sarajevo | 9 | Sibenik | 1 | Šibenik | 2 | Skoplje | 1 | slavonski Brod | 2 |
| Smederevo | 2 | Sokobanja | 1 | Sombor | 1 | Sremska Mitrovića | 1 | Travnik | 1 | Trstenik | 1 |
| užice | 1 | valjevo | 2 | varaždin | 2 | varaždin | 1 | veliko Gradište | 1 | Veljevo | 1 |
| visoko | 1 | vlasenica | 1 | vranje | 2 | vrgorac | 1 | vrnjacka Banja | 1 | Zadar | 2 |
| Zagreb | 18 | Zemun | 1 | Zenica | 1 | Zlatibor | 1 | Zrenjanin | 4 | Zvornik | 1 |

Figure 5. BCMS respondents broken down according to hometown

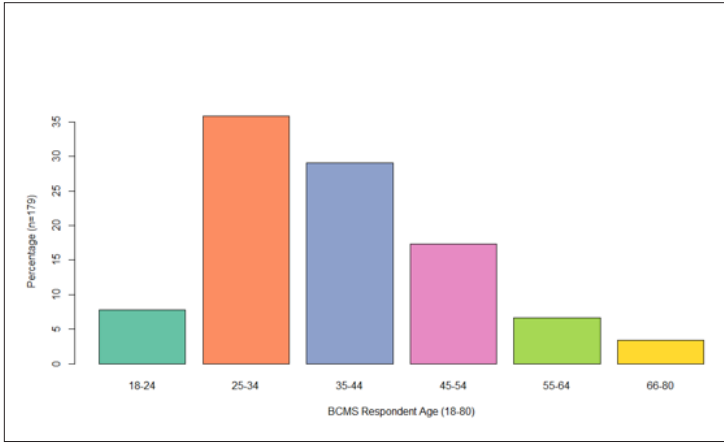


Figure 6. Respondent break-down according to age.

The second page then, only for EG1 and EG2 opens up to a series of blanks and the respondents are required to press spacebar to fill in the blanks (Fig. 7 and Fig. 8)

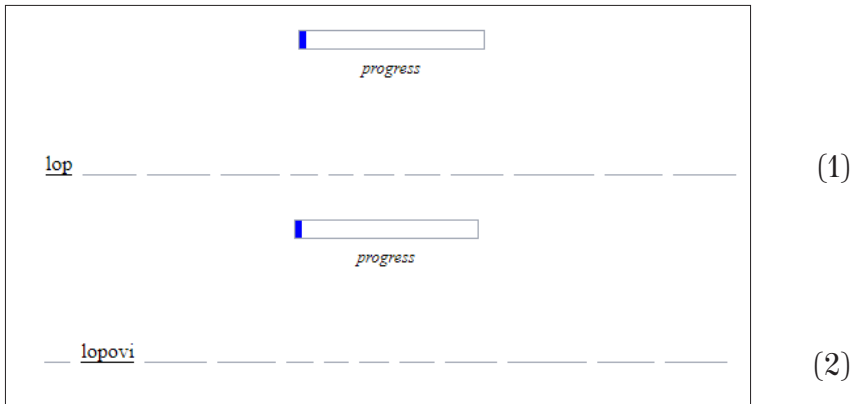


Figure 7. For EG1, two successive screens filling in the blanks with *-ovi* plurals

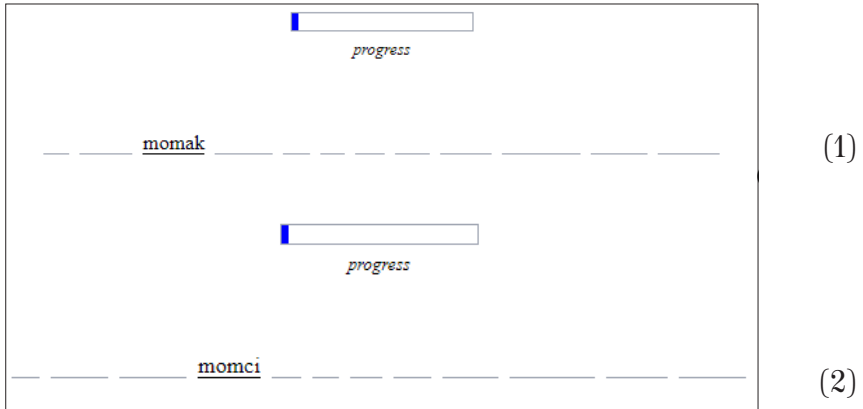


Figure 8. For EG2, two successive screens filling in the blanks with stem-mutated plurals

The idea behind this strategy was to have a Control Group (CG, $n=50$) simply choose from a list of plural options their preferred variant (with the aim of obviating the Observer’s Paradox [Labov 1972: 209]). Unlike Dąbrowska (2008) and Bermel et al. (2017), I do not place my wugs in any sentential context (Fig. 9), therefore removing the respondent’s ability to generalize plural endings based on semantic class similarity (see Fig. 9). All respondents have to work with is their innate L1 phonetics/phonotactics/morphonology. Once the respondents were “inside the test”, as it were, they were then asked to choose their preferred plural form for 15 wugs (Tab. 1), each created with phonological neighborhood context⁵ in mind.⁶ This experiment was repeated with two experimental groups primed (as in Fig. 7 and Fig. 8) only for plurals in *-ovi* (EG1 = ‘Experimental Group

⁵ See Marslen-Wilson’s (1987) seminal work on phonological neighborhoods.

⁶ We created wugs that obeyed BCMS phonotactics, introducing our task with native words, e.g. *momak* ‘boy’ ~ *momci*, recently borrowed yet still fairly unfamiliar *milkšejk* ‘milkshake’, before presenting our wugs. Such a scheme represents a gentle transition from familiar [native words] > unfamiliar [new borrowing] > totally unfamiliar [wugs], with the hope of reducing “wug shock” and keeping the alternation or lack of alternation, depending on the control group priming, active in the mind of the respondent.

1', n=45) and for *-ci*, *-zi*, and *-si* (E2 = 'Experimental Group 2', n=56). The following hypotheses were formulated:

| Wugs (total syllables) |
|----------------------------|
| 1. milkšejk (2) |
| 2. dork (1) |
| 3. storg (1) |
| 4. blah (1) |
| 5. krag (1) |
| 6. plumak (2) |
| 7. bovog (2) |
| 8. klesah ⁷ (2) |
| 9. spirobrag (3) |
| 10. liropih (3) |
| 11. brepazok (3) |
| 12. parahog (3) |
| 13. filosodak (4) |
| 14. flogisterobag (5) |
| 15. parahidrotorbak (6) |

Table 1. Invented wugs with syllable count in parentheses

⁷ I thank an anonymous reviewer who pointed out that *klesah* is also the aorist form of *klesati* 'to carve, cut stone'. However, it is not clear whether this affected speaker choice, as the majority were happy to mutate the final *-h* regardless of priming group. Interestingly, there was a significant increase ($\chi^2(1, N = 45) = 7.49, p < .05$) in the production of *klesahi*, without mutation in EG1 primed for *-ovi*.

(Screen presented after introductory task revealing both plural variants)

Sada vam je predstavljen niz hipotetičkih reči. Prva iz njih: *milkšejk*. Odaberite oblik množine:

1. *milkšejci*
2. *milkšejkovi*
3. *milkšejki*

(Screen presented after clicking on the preferred plural variant of *milkšejk*)

dork

(Next screen after pressing spacebar)

Odaberite oblik množine:

1. *dorki*
2. *dorci*
3. *dorkovi*

Figure 9. Presentation of wugs without any sentential context (instructions only).

The same process was repeated for EG1 and EG2, except that the priming section, where the respondents press the space bar to reveal singular/plural word form pairs was entirely replaced by real words with either *-ovi* plurals (for EG1) or palatalized velars +[i] (EG2).

Based on the prior research on variant palatalization in Slavic (mentioned above) and what the BCMS handbooks say, the following hypotheses have been formulated:

H1: Respondents in the CG who are not primed for a specific plural allomorph will exhibit higher levels of variation with wug words.

H2: Respondents in the CG (without priming for a variable) will take phonological proximity to existing words into account when forming plurals.

H3: Respondents in EG1 primed for *-ovi* will exhibit a higher preference for pluralizing wug words with *-ovi*.

H4: Respondents in EG2 primed for stem alternation will exhibit a higher preference for pluralizing wug words with *-ci*, *-zi*, and *-si* (< [k]+[i], [g]+[i], and [h]+[i], respectively).

H5: Longer syllable count will cause a dispreference for *-ovi*, which as leading Western BCMS handbooks (Browne and Alt [2005: 29] and Alexander [2006: 38] explain, tends to be preferred for monosyllabic stems, e.g. sg. *vuk*/ pl. *vukovi* (cf. however, dialectal/"bookish" pl. *vuci*; sg. *dan* 'day' / pl. *dani*⁸).

H6: Speakers will sparingly accept *-ki*, *-gi*, *-hi*, based on dative forms like *u Beogradanki* 'inside the Beogradjanka (the tallest building in Belgrade)', as opposed to the dispreferred but grammatically correct *u Beogradanci* (with the expected velar mutation).

3. RESULTS AND DISCUSSION

| WUG (! = unexpected preference) | CG (n=50), not primed | EG1 (n=45), primed for <i>-ovi</i> | EG2 (n=56), primed for velar mutation |
|------------------------------------|---|--|--|
| milkšejk | milkšejkovi = 46 milkšejci = 4 milkšeki = 0 | milkšejkovi = 39 milkšejci = 4 milkšejki = 2 | milkšejkovi = 52 milkšejci = 4 milkšejki = 0 |
| dork | dorkovi = 34 dorci = 14 dorki = 2 | dorkovi = 34 dorci = 7 dorki = 4 | dorkovi = 33 dorci = 23 dorki = 0 |

⁸ The second example *dan*, can be explained through the Late Common Slavic loss of the jers and a subsequent outcome in /a/ for all jers in BCMS (cf. Russian sg. *den'* / pl. *dni*). Of course, learners do not typically have access to this diachronic information, so, in the synchronic sense, *dani* counts as an exception.

| | | | |
|-----------|---|---|---|
| storg | storgovi = 35 storzi = 11 storgi = 4 | storgovi = 37 storzi = 4 storgi = 4 | storgovi = 43 storzi = 11 storgi = 2 |
| blah | blahovi = 28 blasi = 20 (!) blahi = 2 | blahovi = 26 blasi = 15 (!) blahi = 4 | blahovi = 35 blasi = 20 blahi = 1 |
| krag | kragovi = 35 krazi = 11 kragi = 4 | kragovi = 37 krazi = 6 kragi = 2 | kragovi = 47 krazi = 8 kragi = 1 |
| plumak | plumakovi = 2 plumkovi = 3 plumaci = 24 (!) plumci = 21 (!) plumaki = 0 plumki = 0 | plumakovi=0 plumkovi = 3 plumaci = 27 (!) plumci = 13 (!) plumaki = 2 plumki = 0 | plumakovi = 3 plumkovi = 5 plumaci = 16 (!) plumci = 31 (!) plumaki = 1 plumki = 0 |
| bovog | bovogovi = 4 bovozi = 42 bovogi = 4 | bovogovi = 0 bovozi = 42 bovogi = 3 | bovogovi = 2 bovozi = 53 bovogi = 1 |
| klesah | klesahovi = 6 kleshovi = 6 klesasi = 18 klesi = 14 klesahi = 6 | klesahovi = 3 kleshovi = 4 klesasi = 16 klesi = 10 klesahi = 12 (!) | klesahovi = 3 kleshovi = 8 klesasi = 28 klesi = 16 klesahi = 1 (!) |
| spirobrag | spirobragovi = 6 spirobrazi = 43 spirobragi = 1 | spirobragovi = 5 spirobrazi = 37 spirobragi = 3 | spirobragovi = 7 spirobrazi = 48 spirobragi = 1 |

| | | | |
|-----------------|---|---|---|
| liropih | liropihovi = 9 (!) liropisi = 40 liropihi = 1 | liropihovi = 5 liropisi = 37 liropihi = 3 | liropihovi = 10 (!) liropisi = 46 liropihi = 0 |
| | | | |
| brepazok | brepazokovi = 6 brepazoci = 29 brepasci = 13 brepaski = 0 brepazoki = 2 | brepazokovi = 6 brepazoci = 27 brepasci = 6 brepaski = 0 brepazoki = 6 | brepazokovi = 9 brepazoci = 37 brepasci = 8 brepaski = 0 brepazoki = 2 |
| | | | |
| parahog | parahogovi = 6 parahozzi = 44 parahogi = 0 | parahogovi = 4 parahozzi = 40 parahogi = 1 | parahogovi = 9 parahozzi = 47 parahogi = 0 |
| | | | |
| filosodak | filosodakovi = 2 filosodaci = 39 filosoci = 8 filosodaki = 1 filosotki = 0 | filosodakovi = 0 filosodaci = 33 filosoci = 8 filosodaki = 4 filosotki = 0 | filosodakovi = 3 filosodaci = 39 filosoci = 12 filosodaki = 2 filosotki = 0 |
| | | | |
| flogisterobag | flogisterobagovi = 7 flogisterobazi = 38 flogisterobzi = 2 flogisterobagi = 3 | flogisterobagovi = 5 flogisterobazi = 36 flogisterobzi = flogisterobagi = 4 | flogisterobagovi = 13 (!) flogisterobazi = 36 flogisterobzi = 4 flogisterobagi = 3 |
| | | | |
| parahidrotorbak | parahidrotorbakovi = 6 parahidrotorpkovi = 0 parahidrotorbaci = 41 parahidrotorpci = 1 parahidrotorpmki = 0 parahidrotorbaki = 2 | parahidrotorbakovi = 2 parahidrotorpkovi = 0 parahidrotorbaci = 35 parahidrotorpci = 3 parahidrotorpmki = 0 parahidrotorbaki = 5 | parahidrotorbakovi = 8 parahidrotorpkovi = parahidrotorbaci = 39 parahidrotorpci = 7 (!) parahidrotorpmki = 0 parahidrotorbaki = 2 |

Table 2. Results of CG, EG1, and EG2 wug experiments for BCMS masculine plural stem allomorphy

For H1, respondents in the CG performed as expected, with respondent preferences not contradicting the BCMS handbooks.

For H2, data was inconclusive across all groups.

For H3: Respondents in EG1 chose *-ovi* at a higher rate when primed with such existing plural forms, only for ♣*dork* (Fig. 10).

For H4: Respondents in EG2 chose to pluralize words with *-ci*, *-zi*, and *-si* (< [k]+[i], [g]+[i], and [h]+[i], respectively) when primed with only such existing plural forms, only for ♣*dork* (Figure 10).⁹

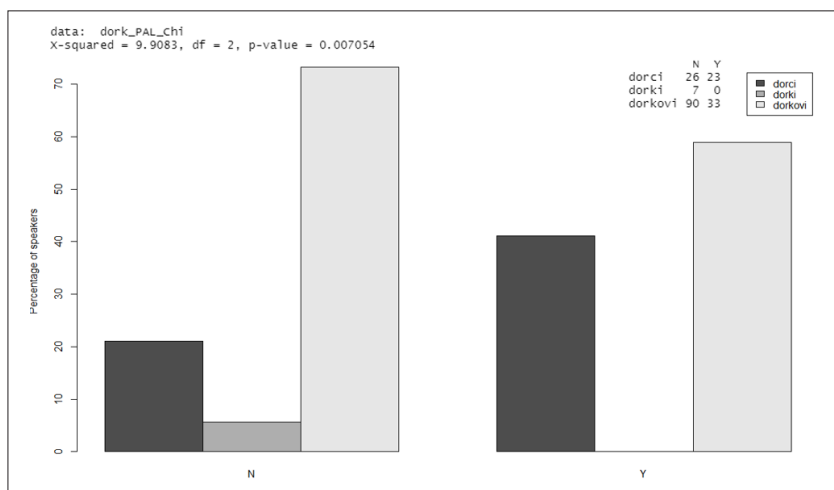


Figure 10. Does priming for velar mutation in the plural increase the selection of ♣*dorci*?

Although the respondent numbers are not as robust as desired, Bermel et al. (2017) make generalizations on Czech patterns with far fewer respondents (n=32). My numbers are nearly double that for the wug ♣*dork*. When primed with velar-mutated plural stems, my EG2 respondents double the acceptance rate of the velar-mutated plural ♣*dorci*.

H5: For monosyllabic and disyllabic wugs, priming had a dramatic effect on choice of plural variant, only for ♣*dork*. Data is inconclusive on speaker preference for other monosyllabic stems, which is ostensibly due

⁹ Chi-square and p-value both calculated using the R Project for Statistical Computing built-in function.

to constraint ranking of phonological factors (if one takes the Optimality Theoretic approach) or more probable, a hitherto undefined metalinguistic hierarchy involving analogical levelling(s) based on syllable count, phonological neighborhood effects, and fuzzy idiolectal semantics. However, the data clearly demonstrate that velar-mutated stems are preferred for tri-, quadri-, pentasyllabic, and larger stem-roots when forming the plural.

H6: As predicted, speakers do sparingly accept plurals in *-ki*, *-gi*, *-hi*, perhaps in analogy to forms like *u Beogradanki* ‘inside the Beogradjanka (the tallest building in Belgrade)’.

4. CONCLUSION

The Second Palatalization of velars in Slavic ceased to operate on the phonological level more than a millennium ago, and that while it was still being applied morphologically during Ottoman times in BCMS, modern BCMS shows variation to how mutation is applied to velar stems for new borrowings (e.g. *blogovi* and not **blozi*; *hotdogovi* and not **hotdozi*; but *piknik*, which has both *piknikovi* and *piknici*). What then, causes speakers to continue applying velar mutation to these polysyllabic wugs? The most obvious factor seems to be the stored plural forms for existing BCMS words that are in the phonological neighborhood of these borrowings and my wugs. Interestingly though, depending on how a respondent is primed, their choice can swing to either the polysyllabic *-ovi* plural or the mutated velar stem. Moreover, it is quite clear that polysyllabic stems choose mutation by default, e.g. ♣*flogisterobazi*, presumably to keep words shorter. But why do very few respondents accept ♣*flogisterobagi*? What compels speakers to insist on this velar-stem mutation? And who decided that *hotdog* would take the *-ovi* plural? My own wug, ♣*bovog* shows that speakers are still very happy to mutate this stem (if even to avoid the awkward sounding ♣*bovogovi*). There are many unanswered questions from this preliminary investigation, and subsequent experiments should target the structure of ♣*dork*, ♣*bovog*, or even ♣*klesah*, which shows a preference for unmutated ♣*klesahi* in CG1. In offering a preliminary account, I would suggest that perceived disyllabic forms like *hotdog* and *piknik* may be interpreted as

bi-elemental compounds¹⁰ and thus only the final element is considered in the pluralization strategy by speakers, with subsequent analogy to (rarely occurring) *dog* ‘dog’ (or more likely *bog* ‘god’) and *šejk* ‘sheikh’. However, this remains speculative and inconclusive, given the results (see Table 2) for ♣flogisterobazi = 38 vs. ♣flogisterobagovi = 7 (with presumed analogy to *bag* ‘(software) bug’) in the unprimed Control Group; one would expect the *-ovi* plural in the compound analysis. Moreover, ♣*parahoz*i is the unanimous choice (87%) by all speakers regardless of priming. And the rare acceptance of forms with *-ki*, *-gi*, *-hi* in possible analogy to *u Beogradanki*, suggest that the non-mutated short plural in *-i* is at least a potential outcome. Thus, following from the Neogrammarian tenet that “sound change, in so far [sic!] as it takes place mechanically, takes place according to laws that admit no exceptions” (Osthoff and Brugmann, 1878), such variation cannot be due to synchronically active phonological factors. There appears to be a hierarchy of metalinguistic factors at play here, involving syllable count, phonological neighborhood effects, and fuzzy idiolectal semantics, which should be the target of further investigation.

Finally, in contributing to the theoretical discussion of suppletion, I offer the following. How does a dead diachronic rule (like this velar mutation in Slavic, for instance) remain synchronically viable? For example, why haven’t speakers eradicated English *go* ~ *went* type suppletion, especially given the clues for this diachronic development are also long buried? The answer appears to lie in cross-generational language acquisition. Specifically, evidence from child language acquisition studies show that children mimic the speech of their parents by first memorizing irregular forms. But once they infer grammatical patterns, e.g. *-ed* = PAST, they then generalize the pattern to all applicable contexts. This overgeneralization (hypercorrection) then has to be unlearned as the child ages in order to meet the expectations of their speech community (Steinberg and Sciarini 2006: 31). This is what I mean by “cyclical” (as in my title): irregular patterns are learnt,

¹⁰ I thank an anonymous reviewer, who shared similar ideas about compounds, for asking for clarification here.

unlearned, and relearned within a single generation (directly feeding the cross-generational extension of this cycle). One might ask, why doesn't the child rebel against this irregularity in the third and final stage, in favor of a seemingly more reliable, maximally productive system? One strong factor appears to be statistical preemption (Boyd & Goldberg 2011), which states that "speakers learn not to use a formulation if an alternative formulation with the same function is consistently witnessed" (ibid.: 55). Language is after all, a social phenomenon, and part of being a member of a social group means playing by the same rules (which are often incorporated through these learning cycles, à la the emergent grammar of Hopper 1987).

REFERENCES

- Berko, J. (1958). The child's learning of English morphology. *Word* 14, 150–177.
- Bermel, N., Knittl, L., & Russell, J. (2017). Still on the brink: The productivity of Czech nominative plural desinences. In A. Makarova, S. M. Dickey, & D. Divjak (Eds.), *Each venture a new beginning: Studies in honor of Laura A. Janda* (pp. 1–17). Slavica Publishers
- Boyd, J. K., & Goldberg, A. E. (2011). Learning what NOT to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87, 55–83.
- Corbett, G. G. (2007). Canonical Typology, Suppletion, and Possible Words. *Language* 83(1), 8–42.
- Dąbrowska, E. (2008). The effects of frequency and neighborhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of memory and language* 58, 931–51.
- Dressler, W. U. (1985). Suppletion in word formation. In J. Fisiak (Ed.), *Historical semantics – Historical Word-Formation* (pp. 97–112).
- Gouskova, M., & Becker, M. (2013). Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory* 31, 735–765.

- Hopper, P. (1987). Emergent grammar. *Proceedings of the thirteenth annual meeting of the Berkeley Linguistics Society 1987*, 139–157.
- Janda, R. D., & Joseph, B. D. (2003). Introduction. In B. D. Joseph, & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 3–180). Blackwell Handbooks in Linguistics. Oxford: Blackwell.
- Juge, M. L. (2013). Analogy as a source of suppletion. In R. Kikusawa, & L. A. Reid (Eds.), *Historical linguistics 2011: Selected papers from the 20th International Conference on Historical Linguistics, Osaka, 25-30*, [*Current issues in linguistic theory* 326], 175–197.
- Kulinich E., Royle, P., & Valois, D. (2016). Palatalization in the Russian verb system: A psycholinguistic study. *Journal of Slavic linguistics* 24(2), 337–357.
- Mel'čuk, I. A. (1994). Suppletion: Toward a logical analysis of the concept. *Studies in language* 18, 339–410.
- Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson, & C. Jones (Eds.), *Historical linguistics II. Theory and description in phonology* (pp. 353–389). Amsterdam: North Holland.
- Osthoff, H., & Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen I*. Leipzig: S. Hirzel, iii–xx.
- Steinberg, D. D., & Sciarini, N. V. (2006). *An introduction to psycholinguistics* (2nd edition). Harlow: Pearson Longman.
- Steinberg, D. D. (1974). Untitled presentation on the suffix '-ity'. Linguistic Society of America Summer Meeting.
- Veselinova, Lj. (2006). *Suppletion in verb paradigms: Bits and pieces of a puzzle*. John Benjamins Publishing.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102

Джеймс Джошуа Пеннингтон

ЦИКЛИЧНЫЙ СУППЛЕТИВИЗМ В СЛАВЯНСКИХ ЯЗЫКАХ: ЧТО
ЭФФЕКТЫ ПРАЙМИНГА ГОВОРЯТ О ЧЕРЕДОВАНИИ ВЕЛЯРОВ
В БОСНИЙСКОМ, ХОРВАТСКОМ, ЧЕРНОГОРСКОМ,
СЕРБСКОМ ЯЗЫКАХ

Резюме

В данной статье представлен эксперимент с выдуманными словами (*wug experiment*) по боснийско-хорватско-черногорско-сербскому (БХЧС) языку, целью которого было выяснить, могут ли эффекты прайминга индуцировать у носителей БХЧС мутацию основного веляра -к, -г, -х > -ц, -з, -с при добавлении морфемы множественного числа -и (ср. женский датив -и и императив -и) или сохранение основного веляра с добавлением алломорфа множественного числа -ови. Принятое здесь определение концепции “супплетивизм” (*suppletion*) совпадает с определением Корбетта (2007), который определяет супплетивизм основы (*stem suppletion*) как супплетивизм главным образом, в отличие от Мельчука (1994), который считает лексическую замену собственно супплетивизмом. Предварительные результаты показывают, что прайминг может воздействовать на выбор алломорфа, а также указывают на иерархию металингвистических факторов в выборе носителя, в том числе количество слогов, эффекты фонологического соседства и другие.

Ключевые слова: супплетивизм, морфонология, алломорфия, БХЧС, *wug* эксперимент, прайминг, вторая славянская палатализация

Branislav Ivanović*

University of Belgrade – Faculty of Philology

Sofija Bilandžija**

University of Belgrade – Faculty of Philology

SYSTEM OF STRONG CONJUGATION IN CONTEMPORARY GERMAN AND NORWEGIAN: A CONTRASTIVE AND TYPOLOGICAL PERSPECTIVE

Abstract

Contrastive research on topics in Germanic morphology have in recent years been overshadowed by contrastive-typological research of other linguistic phenomena. The researchers have not always focused on a contrastive-typological-diachronic research on strong verbs or systems of strong conjugation, although the phenomenon is old and extremely complex. This paper takes a diachronic perspective in defining common features of strong verbs and strong conjugations in German and Norwegian, establishing thus a common denominator for present-day converging features. It uses a synchronic contrastive-typological perspective to establish and motivate diverging processes which result in a pronounced morphological asymmetry within the system. Although the diverging processes are numerous and heterogeneous, and the morphological asymmetry more than visible, the authors conclude, based on the criterial common features, that the systems of strong conjugation are synchronically part of the same typological paradigm.

Key words: strong conjugation, verbal categories, apophony, metaphony, introflexion, extroflexion, divergency, convergency

* branislav.ivanovic@fil.bg.ac.rs

* sofija.bilandzija@fil.bg.ac.rs

1. INTRODUCTION

So far, the researchers haven't been concerned with more comprehensive contrastive-typological research on the conjugation systems of German and Norwegian, two genetically, typologically and geographically close languages, although the data could prove to have significant theoretical and practical consequences. From the theoretical standpoint, one could establish the causes of gradual divergencies concerning the conjugation systems of German and Norwegian which show pronounced quantitative and qualitative morphological asymmetry from a synchronic perspective. Furthermore, once the similarities and differences have been established, one could clearly see a general morphological tendency within the conjugation systems of continental Germanic languages from a prospective angle. In addition to this, the practical implications of a contrastive-typological approach could have an explanatory value for an improved approach to diverse language teaching challenges concerning German and Norwegian as foreign languages.

There as on the contrastive-typological approach to the systems of strong conjugation in these two languages has been some what neglected stems from less attention that morphology, as a traditional linguistic discipline, has received in the last couple of decades in these languages, and the fact that contrastive studies have been focusing predominantly on other linguistic systems. The contrastive research on verbs has so far been more concerned with the defectiveness of Norwegian subjunctive mode, and with establishing equivalent structures in the instances of the absolute modal asymmetry concerning German subjunctives and corresponding structures in different types of Norwegian texts. It's important to note that this type of research is to a much lesser extent concerned with a morphological perspective, but with pragmatic, translation, and textual perspectives (Fabricius Hansen 2004: 119–155; Solfjeld 2009: 219–250). The matters of contrastive inflectional morphology are part of a comprehensive research program *ProGr@mm* (started in 2007), with Norwegian being only one of the contrasted languages, with

the primary goal to analyze German inflection from the perspective of other European languages and to establish a morphological propedeutics for native experts on German language, but also for experts from other countries. Taking the number of contrasted languages into consideration, as well as the project goals, it's clear that the project could not include a systemic diachronic-typological or prospective approach (Augustin 2009: 9, 11, 13–14).

2. COMMON GROUNDS OF STRONG CONJUGATION IN GERMAN AND NORWEGIAN

The Germanic language has two systems of conjugation: the strong and the dental system. The strong system is historically older and is a partial inheritance of Indo-European conjugation system, while the dental (or weak) is a Germanic innovation. Common features for both are synthetic paradigms and the reduction of categorial opposites within the Proto Indo-European, apart from the tripartite category of Person. The reduction of primary i.e. categorial oppositions resulted in a “morphological simplification” of the complete Germanic conjugation. The loss of dual within the category of Number has established the opposition between Singular and Plural, while the formal conflation of the subjunctive (i.e. conjunctive) with optative, and the loss of injunctive, have caused the tripartite modal opposition (indicative, subjunctive, imperative). Meanwhile, the IE temporal paradigm which comprised multiple members, and which according to several authors was more akin to aspect (Streitberg 1974: 276–281; Stanišić 2006: 252), got reduced to present and preterite tense in Germanic. The zero opposition within diathesis (Voice) concerns solely West Germanic and is brought about by complete disintegration of medial and passive voice, while Gotho-Nordic still shows some features of the mediopassive (Krahe 1969: 50–51; Ranke/Hofmann 1979: 68–69; Pudić 1972: 151–152).

The strong Germanic conjugation is predominantly thematic with a smaller number of athematic *mi*-verbs. This conjugation type is primarily characterized by four verbal stems, progressive loss of

reduplication, a special system of extroflexive markers, systemically preserved introflexion, West and North-Germanic metaphony, the possibility of simultaneous marking of grammatical categories by using both introflexive and extroflexive elements, as well as significantly lesser degree of syncretism in comparison with dental paradigms (Krahe 1967: 95–115).

2.1. INTROFLECTIVE ELEMENTS OF THE STRONG GERMANIC CONJUGATION

2.1.1. *The functionalization and persistence of apophony*

The four canonic verbal forms of the strong conjugation are based on *apophony*, the oldest known qualitative-quantitative vowel alternation present in inflection forms of the same word or etymologically related words. Its causes have to this day remained unclear but are often linked to accent features in ProtoIndo-European (Meier-Brügger 2002: 151–152; Beroić 1980: 78). When compared with IE, Germanic languages have developed, functionalized and systemized the seven IE rows of apophony (i.e., ablaut) as markers of Tense opposition with present tense vs. preterite (OHG. *bëran* 'carry' – *bar*, ON. *bera* – *bar*) as well as the opposition in Number (OHG. *bar* (Singular) – *bârum* (Plural), ON. *bar* (Singular) – *bárum* (Plural)), and in West Germanic also the category of Person (Krahe 1967: 100; Pudić 1972: 144). Due to the fact that it expresses clear grammatical meaning (marking of Tense), the apophony within the strong Germanic conjugation has the status of a positionally specific, but at the same time variable *implicit* or *replacive morph* and represents obligatory expression of the strong verbal *introflexion* within the language group as well as in contemporary language (Bilandžija 2017: 20; Schippan 1992: 83).

2.1.2. *The grammatical alternation of consonants*

With certain Germanic strong verbs, the opposition in Number can alternatively be marked by the *grammatical consonant alternation* (Ger. *grammatischer Wechsel*). This phenomenon comprises of a regular

change of Germanic voiced and devoiced fricatives which developed from IE tenues (IE. *p, t, k* > germ. *f/b, Þ/d, χ/g*), and it includes *rhotacism* (IE. *s* > Germ. *r*). These alternations are caused by intervocalic position of the IE tenues/fricative *s* and the stressed vowel following the fricative and different branches of the Germanic language have preserved and developed it in varying degrees (Krahe 1966: 87–88). In addition to being the secondary inflective marker of plural form, the grammatical consonant alternation can also be qualified as the secondary marker of the opposition between present and preterite, considering that it is a subject of obligatory *temporal determination*, appearing exclusively in the preterite and preterite participle forms¹:

OHG. *kiosan – kōs – kurun – gi-koran* 'to choose'
friosan – frōs – frurun – gi-froran 'to freeze'
 ON. *kiōsa – kaus – kōrom – kōrenn*
frīōsa – fraus – frōrom – frōrenn

2.1.3. West- and North-Germanic metaphony as marker of inflection

In contrast with apophony, the *metaphony* (i.e. Umlaut) appears extremely seldom as the singular marker of a specific categorial feature of a strong verb, e.g. 2. person singular of the Imperative in West Germanic (OHG. *helfan – hilf* 'to help' – 'help!'). In this case one is dealing with the *inorganic metaphony*, presumably appearing due to an analogy with the preliterate *i*-umlaut within the present tense (Streitberg 1974: 347). The inorganic *iR*-metaphony (Nedoma 2006: 35) extremely sporadically marks the categories of Tense, Person and Number within the North-Germanic (OI. *koma – kōm* 'to come – (I) come'). In this case also, one can see the analogy with the 2. and 3. person singular, where the causer of metaphony was present in Old Norse, in contrast to the 1. person (Ranke/Hofmann 1979: 67).

Since the metaphony represents a partial regressive vowel assimilation, it appears predominantly together with an extroflexive

¹ The comparative-historic approach uses the term *preterite participle*. The equivalent terms in research on contemporary linguistic processes, both within German and Nordic linguistic tradition, are *perfect participle* or *participle II*.

marker which provokes it, and which is the primary bearer of the categorial features. Except for the aforementioned cases, the West and North Germanic metaphony in the oldest periods of Germanic should be considered the allomorphy of the verbal stem. In contrast to apophony, the preliterate West and North Germanic metaphony in the strong verb system has never reached the status of a systemic and primary (thus also singular) inflective marker of a grammatical category.

2.2. THE PROGRESSIVE LOSS OF REDUPLICATION

The simplification of the Germanic strong conjugation also implies the progressive loss of the IE reduplication as the preterite marker. The reduplication syllable appears before the verbal stem and comprises of the initial consonant within the IE verb root and the vowel *e*, followed usually by further simplification of complex consonant combinations (Stanišić 2006: 263). It occurred within the paradigms of the verbs in the seventh row of apophony. It is best attested in Gothic taking into consideration the time when the first written documents appeared (mid-4th century) and appears significantly less in Old Norse (ON. *róa* 'row' – *rera* – *rerum* – *róinn*). The evidence of its existence in West Germanic are the forms with resistant reduplication in preterite singular of the athematic verb *tuon* 'do' until the Middle High German period (de Boor/Wisniewski 1973: 134–135):

OHG. *teta*, *tâti*, *teta*

MHG. *tete*, *tæ te*, *tete*

It is probable that the loss of reduplication in the Germanic preterite can be related to it being redundant as the Tense marker, considering that the apophony was persistent and the general tendency towards it being a systemic and functionalized feature as the unequivocal temporal marker.

2.3. CO-OCCURENCE OF THE INTROFLECTIVE AND EXTROFLECTIVE MARKERS IN THE STRONG CONJUGATION

One of the characteristics of the strong Germanic conjugation is also the general *possibility* of simultaneous realization of both the introflective and extroflective markers of verbal categories. This co-occurrence is *morphologically determined* and this systemic possibility shows *clear topology* within the conjugation system. The simultaneous occurrence of introflective and extroflective markers is in all Germanic languages related to *secondary markers of Tense and Mode*, and its nature is always complex considering the fact that it can be *double* (more common), encompassing preterite singular/plural and preterite participle, but also *triple* (less common), encompassing preterite plural and preterite participle, but it can also occur in preterite singular by means of analogy (Schmidt 1996: 196). The double one shows frequent combination of apophony and extroflective markers:

OHG. *gēban-gâb-i - gâb-un-gi-gëb-an* 'to give'
ON. *gefa - gaf-t- gaf-um - gef-inn*

The triple marking of grammatical categories shows without exception apophony and grammatical consonant alternation as the introflective elements in combination with secondary extroflective markers:

OHG. *fahan - fiang-i-fiang-um - gi-fang-an* 'to get, take, receive'
ON. *fá-fekk-t - feng-um - feng-inn*

3. CAUSES OF MORPHOLOGICAL DIVERGENCIES BETWEEN GERMAN AND NORWEGIAN STRONG CONJUGATIONS

The linguistic causes² that lead to divergencies between the German and Norwegian strong conjugations are not morphologically

² In addition to purely linguistic factors, there have been additional non-linguistic factors which had undoubtedly influenced divergencies between the western and northern Germanic languages. The most prominent one is the early migration of

motivated in its nature but should be considered within the framework of historical accentology and the historical phonetics of Germanic languages.

3.1. GERMANIC ACCENT AND SYNCOPE / APOCOPE PERIODS

One of the crucial innovations occurring while the Germanic sub-family was differentiating itself from the Indo-European is the loss of melodic and non-fixed accent. Germanic carries on with the dynamic (expiratory) stress and gradually makes it fixed to the root vowel (the acrostatic type of fixed accent). Definitive finalization of the Germanic accent as initially intensive was a lengthy process, and the proof for this statement is the correlative relationship between voiced and devoiced fricatives as in Werner's Law. The acrostatic character of Germanic accent, its expiratory nature, and the dissolution of its amphikinetic character in Germanic paradigms can be characterized as the oldest, as well as primary cause that had further, far-reaching consequences for the profiling of Germanic inflectional morphology.

This type of accent has provoked the lenition of full vowels in medial and final syllables in West and North Germanic and has resulted in *syncope* – the loss of vowels in medial positions, but also complete medial syllables. In addition to this, the nature and position of the new Germanic accent caused the processes of *apocope* – the loss of vowels, but also whole syllables in final position. Syncope and apocope are common conjugation features of both Germanic language groups but the *significantly different timing* of these processes in West and North Germanic can undoubtedly be considered. the cause of the divergencies.

Germanic tribes from their primary territories (southern Scandinavia, present day Denmark and northern Germany) in completely different directions, such as north-eastern Europe, islands in the Atlantic Ocean, as well as Middle and Southeastern Europe. It was relatively early in their development that different directions of migration, insular dislocation and geographical distance brought about weaker and weaker direct linguistic contact (Haarmann 2004: 37–38, 150; Schmidt 1996: 43, 55–58).

In contrast to German, which had preserved full unstressed vowels in atonic syllables until 1050, more precisely until the beginning of Middle High German, the north Germanic branch had been influenced by the syncope much earlier, during the later Proto Nordic period (V–VIII centuries). The syncope started with short Proto Nordic vowels, while the long vowels in medial syllables got secondarily shortened in order to go through syncope later. The same is true for the north Germanic apocope. It begins much earlier than the Middle High German period, also during the later Proto Nordic period, and is co-occurrent with syncope. The mechanism of Proto Nordic apocope is identical to that of the syncope: it is the short vowels that disappear first in final position, and the others then go through the shortening and secondary loss (Ranke/Hofmann 1979: 37).

The said phonetic laws which were caused by the initially intensive Germanic accent had brought about not only the significantly earlier *quantitative syllabic deprivation* in North Germanic, but had direct consequences for the earlier, but more pronounced simplification of the synthetic verbal inflection due to the weakening and loss of extroflexive markers. These phonetic processes have resulted in a relatively early *loss of certain grammatical categories*, and *increased degree of syncretism* in North Germanic in comparison with High German. This morphological asymmetry has been preserved to this day since newer periods within the history of Norwegian have not showed a secondary restitution of extroflexive markers.

3.2. EARLY NIVELLATION OF THE GRAMMATICAL CONSONANT ALTERNATION

It's a well-known fact that different Germanic languages show a varying degree of grammatical consonant alternation, also within the system of strong conjugation. Based on written documents we know that voiced and devoiced fricatives were nivellated early in Gotho-Nordic (towards the devoiced), while the alternation was best preserved on Old Saxon and High German (Pudić 1974: 70).

It is still unclear why the grammatical consonant alternation was nivellated so early in North Germanic, but the process can with

certainly correlate with extremely early and intensive syncope and apocope. They have often resulted in *monosyllabic infinitives* (as in contemporary NO. *dra* 'to drag, travel') that have lost the consonant ('g') which could have gone through the alternation, with preterite and preterite participle that were in later stages formed in analogy with the monosyllabic infinitive, and no longer had the possibility to retain the consonant alternation. As the monosyllabic infinitive developed so early, several verbs in contemporary Norwegian can synchronically show only *latent grammatical consonant alternation* which can also be characterized as *pseudoepithetic*.

In contrast to Norwegian, West Germanic has nivellated the grammatical consonant alternation at a much slower rate, the infinitives in High German have retained their full form in all phases, and that is the reason this introflective phenomenon shows a much higher degree of *diachronic resistance* with a continuous tendency towards *quantitative simplification* in later stages of language history. After the Early New High German period (1350–1650) had ended, the grammatical alternation persisted with few strong verbs in contemporary language (GER. *erkiesen* – *erkor* – *erkoren* 'to choose', *schneiden* – *schnitt* – *geschnitten* 'to cut', *ziehen* – *zog* – *gezogen* 'to drag').

3.3. ABSENCE OF PRELITERARY METAPHONY IN NORTH GERMANIC PRESENT TENSE

Although the preliterate metaphony was provoked by the same factors in West and North Germanic, in the latter it didn't occur in the present tense singular forms (Ranke/Hofmann 1979: 67). Its absence is potentially caused by early processes of extroflective apocope, i.e., the fact that the factor provoking the combinatorial change of the root vowel had been lost much earlier than in West Germanic. Although the metaphony is not a qualitative vowel alternation which functions as a grammatical morph, but rather an allomorph, the North Germanic present tense paradigm shows much more reduction in morphology in comparison with Old High German:

Table 1. Differences in Old Icelandic and Old High German paradigms (indicative mood)

| OIsl. <i>gefa</i> 'to give' | OHG. <i>gëban</i> |
|--|---|
| 1. <i>gef</i> 1. <i>gefum</i> 2. <i>gefr</i> 2. <i>gefið</i> 3. <i>gefr</i> 3. <i>gefa</i> | 1. <i>gibu</i> 1. <i>gebemes</i> 2. <i>gibis</i> 2. <i>gebet</i> 3. <i>gibit</i> 3. <i>gebant</i> |

3.4. FUNCTIONALIZATION OF AN ORIGINALLY WEAK MARKER WITH STRONG PRETERITE PARTICIPLE

Germanic languages could form preterite participle forms in two ways: with the help of IE. **-ne/no-* with strong verbs showing apophony and possible reduplication, and with the help of IE. **-te/to-* with verbs with dental preterite. Since the dental participle marker is older than the weak verb group, one can assume that it could also occur with strong preterite participles, especially in those cases where the participles had converted to nouns or adjectives (Streitberg 1974: 195–196, 287).

The German strong preterite participles have been consequent in keeping the IE. **-ne/no-* marker, contemporary *-en*, while the nowadays systemic occurrence of the weak participle marker *-(e)t* within strong verb paradigms in verbal complexes is not explained as an analogy with weak participles, but as the fact that it is historically the inflected neuter participle form (Næs 1965: 173, Hanssen et al. 1975: 82). Synchronically speaking, the difference is striking: GER. *gebunde-en*: NOR. *bund-et* 'bound'³. Without the diachronic approach, the Norwegian strong participles show signs of *morphological descentence* towards the weak conjugation⁴.

³ In contemporary Norwegian this is the non-congruent participle form, used in complex tenses and passives. Reflexes of the old participle ending *-inn* (ON. *kominn, fundinn, spunnninn*), today: *-en*, can still be seen in when them as c/fem. forms of participles are used as attributes, as well as in definite and plural forms, e.g. *skreven, skrevne*.

⁴ Participle sending in *-et* are considered to be strong participles, while many strong verb participles today are essentially weak (Næs 1965: 173; Faarlund et al 1997: 486).

4. RECENT PROCESSES OF STRUCTURAL CONVERGENCIES BETWEEN THE GERMAN AND NORWEGIAN CONJUGATIONS

The high degree of syncretism, provoked by the syncope and apocope, had resulted in a general tendency towards analytic verbal forms in both languages. The primary tense opposition between the present and preterite tense becomes more complex by the appearance of the perfect, past perfect and future tenses, and additionally first and second conditional in Norwegian as integral components of the system of tenses. The analytic forms appear additionally within the infinitive and become more and more related to the voice and tense categories (present/perfect infinitive, active/passive). Both languages increase the number of auxiliary verbs (Ger. *haben*, *sein*, *werden*, *bekommen* and *gehören*, Nor. *ha*, *væ re*, *bli*, with the addition of the originally modal *skulle* and *ville*). Both languages show the tendency towards expressing the passive voice analytically, with co-occurrent semantic specification (Ger. *werden-*, *sein-*, *bekommen-* and *gehören-* passive, Nor. *bli-*, *væ re-* and *få-* passive, with the certain preservation of the synthetic modal *s-passive*).

5. MORPHOLOGICAL ASYMMETRY FROM A SYNCHRONIC PERSPECTIVE

Present-day morphological asymmetry between German and Norwegian strong conjugations can be considered both from a quantitative and a qualitative perspective, but one has to remark that they overlap. The morphological simplification in Norwegian results in a qualitative asymmetry within the strong verb conjugation in the examined languages. Quantitatively the asymmetry is present in the number of verbal categorial features, the oppositions that are present, the number of possible extroflexives, but also the inventory of strong verbs. It's important to note that the system of strong conjugation in German, formally speaking, preserves the reflexes of a more conservative language stage to a higher degree.

5.1. ABSOLUTE VS. PARTIAL SYNCRETISM OF PERSON. ASSYMETRY OF NUMBER. THE INVENTORY OF EXTROFLECTIVES.

The quantitative asymmetry is present in the categories of Person and Number, but also considering the number of potential extroflexive elements. In contrast to Norwegian, where the categories of Person and Number are *de facto* lost, they have been preserved in German. The non-existence of Person in Norwegian is a consequence of the nivellation of extroflexive markers and the *absolute recurrence* of just two markers: *-(e)r* and *{-ø}* in all the “finite” paradigms. Compared with Norwegian, German shows a *partial recurrence* of Person and Number markers: *-e*, *-(e)st*, *-(e)t*, *-(e)n* and *-ø*. The result of this quantitative asymmetry in the number of extroflexive provokes an *absolute qualitative syncretism* of Person and Number in Norwegian, and a *partial qualitative syncretism* of Person and Number in German, while the degree of syncretism in German can be additionally reduced by allomorphy of the verbal stem (transition of *e* to *i*, umlaut):

NO. *jeg/du/han/hun/den/det/vi/dere/de gi-r, gav-ø*
 GER. *ichgeb-e, du gib-st, er gib-t, wir/sie/Siegeb-en, ihrgeb-t*
ich/er/sie/esgab-ø, dugab-st, wir/sie/Siegab-en, ihrgab-t

5.2. REDUCTION OF SUPPLETIVISM WITHIN NORWEGIAN STRONG CONJUGATION

Suppletivism is a peripheral phenomenon in Germanic strong conjugation. It is originally connected to the athematic *mi*-verb: OHG. *bim/bin*, ON. *em* '(I) am'. The canonical forms of infinitive and present tense are formed from IE. roots **(e)s-*, **bhu-/bheu-*, while the preterite is formed from the IE. root **ues-* (GOT. *wisan*, OHG. *wesan*) and show signs of rhotacism:

ON. *em – var – várum – verit*
 OHG. *sin – bim – was – warun – giwesan*

In contrast to contemporary German, which has preserved the suppletive forms of all verbal roots from the protolanguage, followed by the grammatical alternation of consonants (*sein – bin – war –*

gewesen), the present tense stem **bhu-/bheu-* from the protolanguage was lost exceptionally early. Neither the infinitive is formed from the *(*e*)*s*-stem from the protolanguage, since the process of intensive nivellation of the infinitive towards the plural preterite forms had already started around year 1200, something that had brought about the loss of rhotacism (Ranke/Hofmann 1979: 65). In contrast to German, the suppletivism in contemporary Norwegian is restricted exclusively to the present tense paradigm without a synchronically explicit rhotacism: *jeg/du/... er*.

5.3. MANIFEST VS. LATENT GRAMMATICAL CONSONANT ALTERNATION

As a secondary introflective marker of Tense and Number, the grammatical consonant alternation has in both languages quantitatively been reduced to a smaller number of strong verbs. The differences can synchronically be said to follow this set of rules: in contemporary German it presents as *predominantly manifest* (exception: *hauen – hieb* 'to hit, beat') because it occurs in all basic forms. In Norwegian it presents as *a latent alternation with the restitution of the primary consonant*. From a strict synchronic perspective one can interpret it as *consonant pseudoepithesis*. The latent character of the grammatical alternation arrived when the full forms of infinitive got reduced to monosyllabic forms without the primary consonant that could have alternated with the consonant in the preterite form:

GER. *schneiden – schnitt–geschnitten* 'to cut', *ziehen – zog–gezogen* 'to drag'
NOR. *gå – gikk–gått* 'to go'

5.4. FLUCTUATION IN PRETERITE FORMS

The preterite fluctuation is an extremely complex phenomenon, especially in contemporary Norwegian where it's still much more pronounced than in German. There are several causes responsible

for this phenomenon, which leads to different types of preterite fluctuation in contemporary language.

5.4.1. Homotype fluctuation as a result of vowel nivellation

The main cause of this type of preterite fluctuation is the fact that strong verbs in both languages have in later stages of language development gone through spontaneous processes of *vowel nivellation*. This term denotes that root vowels became the same in the singular and plural forms, which in turn leads to the apophony no longer functioning as the marker of Number (in both languages), leading to a more explicit profiling of Tense. Vowel nivellation in both languages could be twofold: 1. present-day preterite forms have preserved the vowel from former singular forms, and 2. present-day preterite forms have preserved the vowel from former plural forms. This has resulted in fewer canonical forms both in German and Norwegian from four to three (infinitive – preterite – perfect participle). The vowel nivellations are crucial in the system of strong conjugation since they are historically the last major change that has led to the profiling of present-day strong conjugation in both languages.

In contemporary German one can consider the vowel nivellation had ended by around 1650, marking thus the end of Early New High German, with sporadic stabilization until the beginning of 18. century (Srđić 2008: 69; Hartweg/Wegera 2005: 164). In contrast to this, the vowel nivellation in the Norwegian strong verb system shows a more pronounced *synchronic dynamism*, which also today provokes a frequent *homotypical fluctuation*, since its results in two forms of preterite. In contemporary German, the homotypical fluctuation is no longer systemic – it is a rare phenomenon occurring with the verbs *dreschen*, *heben*/*schwören*:

NOR. *bite* – *bet/beit* 'to bite', *fyke*–*føk/fauk* 'to blow', *klyve*–*kløv/klauv* 'to climb'

GER. *dreschen*–*drosch/drasch* 'to thresh', *heben* – *hob/hub* 'to lift', *schwören* – *schwor/schwur* 'to swear'

5.4.2. Homotypical fluctuation with consonant restitution

A special subtype of homotypical fluctuation appears on contemporary Norwegian with those verbs who have previously undergone intensive diachronic tendency of infinitive shortening towards monosyllabic structure. Present-day preterite forms fluctuate between a synchronically motivated form and a form with a consonant restitution (from previous infinitive forms). This phenomenon is system wise not specific for contemporary German:

NOR. *be – ba/bad* 'to prey, to ask', *dra – dro/drog* 'to drag', *lyge–løy/laug* 'to lie'

5.4.3. Heterotypical fluctuation as a result of descendance towards weak conjugation

This subtype of preterite fluctuation is inherent to both languages and is caused by a general tendency of strong verbs adapting or transitioning of to weak conjugation, which is nowadays considered prototypical in all Germanic languages. Considering the fact that this type of fluctuation is present in both languages, one does not consider it to be a pure morphological asymmetry, but one can determine that two opposing tendencies can be observed in the two languages: it is *more frequent* in Norwegian, but at the same time, in contrast to German, it is *less likely* to reflect a semantic difference. The result of the heterotypical preterite fluctuation is a parallelism between an older, strong form and a newer, weak inflection form and has no correlation with lexical semantic contiguity between verbs in German and Norwegian:

NOR. *Jeggravde/grov et hull*. 'I dug a hole'
GER. *Siewebte/wobaneinemTeppich*. 'shewove a carpet'

The preterite fluctuation shows that the strong conjugation in Norwegian is less stable, that it shows pronounced synchronic dynamism as well as a more pronounced transitioning towards the weak conjugation.

5.5. STRONG PARTICIPLE CIRCUMFIX

There is a complete morphological asymmetry between the two languages considering the strong participle circumfix *ge-...-(e)n* as it is a German characteristic. This circumfix is different from the weak participle circumfix *ge-...-(e)t*, which means that this infinite form shows a much more obvious markedness in relations to the infinitive and preterite forms in case of identical vocalism in basic verb forms. The Norwegian perfect participles have not developed a systemic circumfix considering the fact that the former Germanic derivational element *ga-*, *gi-*, *ge-* (with perfective meaning) was specific for Gothic, West Germanic and Anglo-Saxon, according to the earliest written documents (Krahe/Meid 1969: 37–38). Based on this, the morphological opposition between the preterite and perfect participle is less defined, and can in some cases even show syncretism:

NOR. *falle – falt – falt* 'to fall', *holde – holdt – holdt* 'to hold'
 GER. *fallen – fiel – gefallen* 'to fall', *heben – hob – gehoben* 'to lift'

In addition to this, the present-day strong participle marker *-et*, which is a historical rest of inflected neuter form, makes all strong participles in Norwegian formally close to weak participle endings (*-t*, *-d*, *-dd*, *-dt*). Based on this, one can conclude that Norwegian strong conjugation is less morphologically marked in contrast to the weak conjugation.

5.6. THE PHENOMENON OF *T*-EPITHESES IN NORWEGIAN STRONG PRETERITE

A number of strong Norwegian verbs add a final unmotivated dental ending *-t*. In these cases, the dental ending doesn't have the status of a grammatical marker, but is a case of *epithesis*, probably emerging by means of analogy with a prototypical dental preterite. Synchronically speaking, epithesis can be qualified as a type of „weakening“ of otherwise strong preterital forms:

NOR. *falle – falt* 'to fall', *vinne – vant* 'to win', *spinne – spant* 'to spin'

From a synchronic perspective the *t*-epithesis can easily be wrongly identified either with latent grammatical alternation, but also with consonant restitution. It differs from consonant restitution by the fact that *t*-epithesis appears with verbs that have preserved full infinitive forms. It differs from latent grammatical alternation by means of etymology: if former full infinitive forms used to have the dental consonant *d*, and present-day preterite has a *t*, one has to have latent grammatical alternation in mind.

5.7. MORPHOLOGICAL (UN)MARKEDNESS OF MODAL OPPOSITION

Although modally syncretic forms do exist, the trifold Mode opposition between indicative – subjunctive – imperative still exists in contemporary German and is much more consistently marked than in Norwegian. The causes for pronounced asymmetry within the category of Mode should certainly be identified with the maximal reduction of the Norwegian extroflexive system, at the same time as the apophony is specialized for marking the Tense, which have brought about a pronounced defectiveness of Norwegian verbal Mode.

The divergencies concerning the subjunctive arise from the 13. century onwards. German shows a tendency towards multiplying an originally binary subjunctive paradigm and begins to develop periphrastic forms also within the subjunctive paradigm. The Norwegian system is at the same time experiencing a pronounced reduction in the number of extroflexive markers, pronounced simplification of the rows of apophony and their nivellation, but is also going through a gradual loss of preterite subjunctive (Seip 1955: 321). The intensive and early processes where extroflexive markers went through apocope have resulted in that present-day present subjunctive in Norwegian is morphologically unmarked and shows syncretism with present infinitive. This has diachronically led to the loss of free syntactic use, and the subjunctive has survived in sentence structures with a predominantly optative meaning, where one also has to consider the process of phraseologisation:

NOR. *Leve kongen!* 'long live the king' *Gudvelsignedeg!* 'god bless you'

In contrast to Norwegian, German subjunctive paradigms are morphologically marked by primary (*-e*, *-en*), but also secondary extroflexive markers (*-est*, *-et*) with obligatory metaphony of the vowels *a*, *o* and *u* in preterite subjunctive, showing thus allomorphy (*käme*, *löge*, *trüge*). In addition to this, a few strong verbs in German have preserved *recessive preterite subjunctive forms* (*hülfe*, *stünde*, *würfe*).

Also imperative in Norwegian is less marked than imperative in German. The verbs with monosyllabic infinitive forms show syncretism, while the verbs with full infinitive form imperative by losing the infinitive marker *-e*. In contrast to German, the imperative forms in Norwegian never experience vowel metaphony, which in German marks more clearly not only the category of Mode, but also of Person. In addition to this, there is an optional and archaic extroflexive marker *-ein* German for 2. person singular, while the 1. person plural and distant forms are expressed by present subjunctive (adhortative subjunctive):

NOR. *Gimeg boka!*

Hjelpmeg!

GER. *Gib mir das Buch! Gebt mir das Buch! Geben Sie mir das Buch!*

Hilf mir! Helft mir! Helfen Sie mir!

Seien wir jetzt ruhig!

Based on this, one can conclude that the Norwegian strong conjugation is morphologically speaking “spared for” allomorphy, but also for secondary modal extroflexive markers.

6. CONCLUDING REMARKS

Contemporary German and Norwegian are two genetically, geographically and typologically related languages, but nevertheless show a high degree of morphological asymmetry within strong verb conjugation. This asymmetry is a result of historically divergent

development of the two languages. The markers of strong Germanic conjugation were originally uniform and can positionally be classified as internal and external. Internal or introflective markers of strong Germanic conjugation are ablaut (apophony), umlaut (metaphony or regressive vowel assimilation) and regular grammatical consonant alternation. Ablaut is to be considered as specific replacive morph which indicates the category of tense. These features have invariant topology, i.e., they appear in precisely allocated positions within strong verbal paradigms.

External markers comprise of a specific system of inflection markers without dental elements, and they mark the verbal categories of person, number, tense, and mode. The inflectional system of the strong Germanic conjugation shows pronounced allomorphy. One marker can appear in multiple inflection forms, and that is why the strong conjugation also comprises of syncretic forms. Also suppletive forms can appear within the conjugation, but they are rare.

The authors have contrasted the aforementioned features and principles in German and Norwegian and have established the tendency towards weak conjugation in Norwegian. This is all due to the loss of subjunctive, the disappearance of old inflectional endings and the destruction of the paradigm but is also due to the fact that old sound changes have a less prominent role within verbal paradigms. On the other hand, German shows a tendency towards stability within the strong conjugation, preserving several sound changes, as well as strong inflectional markers. Due to that, on the synchronic level, the morphology of strong conjugation clearly differs from both weak and mixed conjugations.

REFERENCES

- Augustin, H. (2009). Die Konzeption der Flexionsmorphologie des Deutschen aus kontrastiver Perspektive in der Lernplattform ProGr@mm. *Studia Linguistica* 27(1), 9–16.
- Berić, V. (1980). *Osnovi istorijske gramatike nemačkog jezika sa osvrtom na uporedne germanske jezike*. Novi Sad: Institut za strane jezike.

- Bilandžija, S. (2017). *Slaganje u skandinavskim jezicima – prototip i periferija*. Beograd: FOKUS – Forum za interkulturnu komunikaciju.
- de Boor, H., & Wisniewski, R. (1973). *Mittelhochdeutsche Grammatik*. Berlin-New York: Walter de Gruyter.
- Fabricius Hansen, C. (2004). Wessen Redehintergrund? Indirektheitskontexte aus kontrastiver Sicht (Deutsch-Norwegisch-Englisch). In: *Tempus/Temporalität und Modus/Modalität im Sprachvergleich* (O. Leirbukt et al., eds.), Tübingen: Stauffenberg, 119–155.
- Faarlund, J.T. et al. (1997). *Norskreferansegrammatikk*. Oslo: Universitetsforlaget.
- Haarmann, H. (2004). *Lexikon der untergegangenen Sprachen*. München: Verlag C. H. Beck.
- Hanssen, E. et al. (1975). *Norrønn grammatikk*. Oslo: Universitetsforlaget.
- Hartweg, F., & Wegera, K. P. (2005). *Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit*. Tübingen: Max Niemeyer Verlag.
- Krahe, H. (1966). *Germanische Sprachwissenschaft. Einleitung und Lautlehre*. Berlin: Walter de Gruyter & Co.
- Krahe, H. (1967). *Germanische Sprachwissenschaft. Formenlehre*. Berlin: Walter de Gruyter & Co.
- Krahe, H., & Meid, W. (1969). *Germanische Sprachwissenschaft. Wortbildungslehre*. Berlin-New York: Walter de Gruyter.
- Meier-Brügger, M. (2002). *Indogermanische Sprachwissenschaft*. Berlin-New York: Walter de Gruyter.
- Nedoma, R. (2006). *Kleine Grammatik des Altisländischen*. Heidelberg: Universitätsverlag Winter GmbH.
- Næs, O. (1965). *Norsk grammatikk. Elementære strukturer og syntaks*. Oslo: Fabritius.
- Pudić, I. (1974). *Gotski jezik I. Istorijaska gramatika*. Beograd: Zavod za udžbenike i nastavna sredstva Srbije.
- Ranke, F., & Hofmann, d. (1979). *Altnordisches Elementarbuch*. Berlin-New York: Walter de Gruyter.
- Schippan, Th. (1992). *Lexikologie der deutschen Gegenwartssprache*. Tübingen: Max Niemeyer Verlag.
- Schmidt, W. (1996). *Geschichte der deutschen Sprache*. Stuttgart: S. Hirzel – Wissenschaftliche Verlagsgesellschaft.
- Seip, Didrik Arup (1955). *Norsk språkhistorie til omkring 1370*. Oslo: H. Aschehoug&Co.

- Solfjeld, K. (2009). Redewiedergabe in verschiedener Form – ein Vergleich Deutsch-Norwegisch. *Hermes. Journal of Language and Communication Studies* 43, 219–250.
- Srdić, S. (2008). *Uvod u rano novovisoko nemački jezik*. Beograd: Filološki fakultet.
- Stanišić, V. (2006). *Uvod u indoevropsku filologiju*. Beograd: Čigoja štampa.
- Streitberg, W. (1974). *Urgermanische Grammatik. Einführung in das vergleichende Studium der altgermanischen Dialekte*. Heidelberg: Carl Winter – Universitätsverlag.

Branislav Ivanović
Sofija Bilandžija

DAS SYSTEM DER STARKEN KONJUGATION IN DER
DEUTSCHEN UND NORWEGISCHEN GEGENWARTSSPRACHE
AUS KONTRASTIV-TYPOLOGISCHER SICHT

Zusammenfassung

Obwohl starke Verben und starke Konjugation als sehr alte und typisch germanische morphologische Phänomene zu gelten haben, sind sie bisher in der einschlägigen Literatur aus der kontrastiv-typologisch-diachronen Sicht nur unzureichend beschrieben worden. Die vorliegende Arbeit stellt eine Vereinigung von zwei nur oberflächlich gegensätzlichen Betrachtungsweisen dar, dank denen äußerst komplexe und verschiedenartige Prozesse der sprachlichen Konvergenzen und Divergenzen zwischen den deutschen und norwegischen starken Verbalsystemen erhellt werden konnten. Mithilfe der diachronen Betrachtungsweise wurde eine gemeinsame deutsch-norwegische Grundlage der starken Konjugation festgestellt, auf der die heutigen konvergenten Prozesse beruhen, während mithilfe der synchronen Betrachtungsweise Divergenzen im heutigen System der starken deutschen und norwegischen Verben beschrieben wurden. Obwohl Divergenzen, die auf verschiedene Ursachen zurückzuführen sind, zwischen der gegenwärtigen deutschen und norwegischen starken Konjugation sehr zahlreich und heterogen sind, was in der äußerst ausgeprägten morphologischen Asymmetrie zum Vorschein kommt, kann aufgrund der vorliegenden Untersuchung festgestellt werden, dass die beiden Systeme dem gleichen typologischen Paradigma angehören.

Schlüsselwörter: starke Konjugation, verbale Kategorien, Ablaut, Umlaut, innere Flexion, äußere Flexion, Divergenzen, Konvergenzen

Tatjana Marvin*
University of Ljubljana

ADNOMINAL ARGUMENTS IN PHRASES HEADED BY DEVERBAL NOUNS IN SLOVENIAN

Abstract

The paper deals with the issue of word order in Slovenian nominalizations with three adnominal arguments (agent, theme, and recipient) appearing simultaneously as postmodifiers in a nominal phrase whose head is a deverbal process/event noun. The main goal is to derive the fixed order of arguments: theme genitive >> agent genitive >> recipient dative. The article presents an analysis in which the order is achieved by introducing the VoiceP into the nominalization structure with transitive and unergative verbs and by movement of the theme argument to a position above the agent and recipient arguments.

Keywords: deverbal nominalization, adnominal argument, word order, Slovenian

1. INTRODUCTION

In Slovenian, all three arguments, agent, theme, and recipient can appear simultaneously as postmodifiers in a nominal phrase whose head is a deverbal process/event noun derived from a ditransitive verb. If these arguments are all expressed by noun phrases, their word

* tatjana.marvin@guest.arnes.si

order (WO) is fixed as in (1): *head >> theme genitive >> agent genitive >> recipient dative*.^{1,2}

- (1) podeljevanje [nagrada] [mednarodne žirije] [najboljšim filmom]
awarding awards.GEN international jury.GEN best movies.
DAT
“presenting awards to the best movies by the international jury”

The primary goal in this article is to show that the WO in (1) cannot be obtained if we adopt other proposals for Slavic languages dealing with adnominal arguments in process nominalizations (Zlatić 1997, Dvořák 2011, Šarić 2018). The argument is laid out in two steps. First, previous proposals for structures with two adnominal arguments are presented (section 2), followed by the analysis of the problem that arises when the third argument is added (section 3). In the last part, a possible solution is presented (sections 4 ad 5).

2. TWO ADNOMINAL DP ARGUMENTS IN OTHER SLAVIC LANGUAGES

We first consider cases with two adnominal arguments, i.e. various combinations of theme, agent and recipient noun phrases, and show how these are treated in the literature. We focus primarily on the analysis of the Serbian and Czech data.³

¹ We did not find any examples of this kind in the literature on Serbian that we examined. Šarić (2018: 79) notes that Serbian nominalizations can probably express three arguments as well, but native speakers find them difficult to process.

² The examples with three adnominal arguments are rarely used by Slovenian speakers, probably because of the difficulty of processing and for stylistic reasons. However, they are possible and have been mentioned in Slovenian literature, e.g. in Žele (2001: 31). An empirical study with data obtained from native speakers of Slovenian would certainly be desirable as a next step in this research. Unfortunately, such a study is beyond the scope of this paper and is left for future research.

³ We do not deal with Russian data here because Russian process nominals do not allow two genitive arguments (only result nominals do in Russian). See Rappaport (1998) for a detailed analysis.

2.1. THEME AND RECIPIENT

Zlatić (1997: 212) observes that the genitive theme precedes the dative recipient in deverbal nominalizations, as in (2).

- (2) donacija novca bolnicama *donacija bolnicama novca
 donation money.GEN hospitals.DAT
 Zlatić (1997)

In her analysis, the order is a consequence of case assignment: the genitive assigned by the noun is an instance of structural case, while dative case is an instance of inherent case. She proposes a rule according to which the NP bearing structural case precedes all other phrasal elements. She further argues that two genitives are not allowed in deverbal nominals in Serbian, as only one element can fill the position for structural case assignment. Thus, one can have a genitive agent as postmodifier if this is the only genitive argument, (3), but not if the theme is present – in such cases the agent is expressed as a possessive, (4). We return to Zlatić’s analysis in Section 2.2, where we present a critical assessment in Šarić (2018).

- (3) protest ovog studenta
 protest this.GEN student.GEN
- (4) a. *opis Jovana Amerike
 description John.GEN America.GEN
- b. *opis Amerike Jovana
 description America.GEN John.GEN
- c. Jovanov opis Amerike
 John’s decription America.GEN
 Zlatić (1997)

Dvořak (2011) also discusses the order of adnominal arguments and draws conclusions regarding case assignment based on Czech data. Czech is particularly interesting when comparing the order of arguments in sentences with their order as adnominal arguments. Ditransitive verbs can be divided into two groups with respect to the order of two objects in the sentence: 1) DAT-ACC verbs, where the unmarked WO is dative>>accusative, e.g. *dát*, *darovat* ‘give as a gift’,

věnovat 'inscribe/dedicate', *poslat* 'send' and 2) ACC-DAT verbs, where the unmarked order is accusative >> dative, e.g. *podřídít* 'subordinate, accommodate', *vystavit* 'expose', *zasvětit* 'devote', *svěřit* 'entrust', *zanechat* 'leave', Dvořák (2011). The two groups are exemplified in (5) and (6):

- (5) Karel poslal Marii dopis. // #...poslal dopis Marii.
Charles.NOM sent Mary.DAT a letter.ACC
- (6) Karel podřídil svoje plány Marii. // #... podřídil Marii svoje plány.
Charles.NOM adjusted his plans:ACC Mary.DAT
Dvořák (2011)

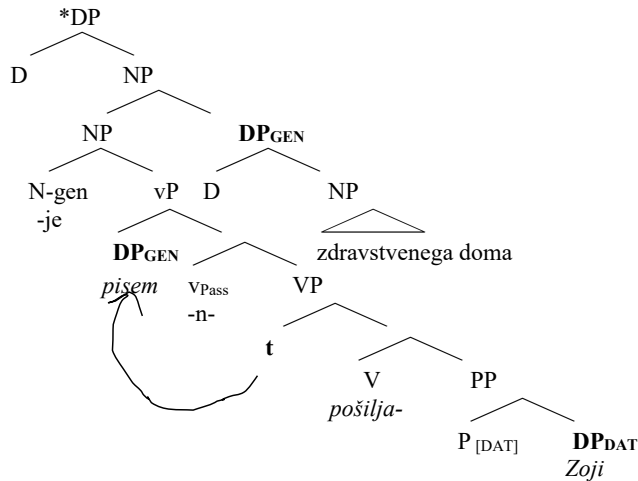
In deverbal nominalizations, however, the WO difference is neutralized. As can be seen from (7) and (8), the theme argument bearing genitive case (underlying accusative direct object) always precedes the recipient argument bearing dative (underlying dative indirect object) regardless of their positions in the sentence (cf. 5, 6).

- (7) Poslání dopisu Marii (se Karlovi nevyplatilo).
sending.NOM.SG letter.GEN Mary.DAT (refl Charles.DAT not-paid-off)
"Sending Mary a letter (didn't pay off to Charles)."
- (8) Podřízení plánů Marii (se Karlovi nevyplatilo).
adjusting.NOM.SG plans.GEN Mary.DAT (refl Charles.DAT not-paid-off)
"Adjusting plans to Mary (didn't pay off to Charles)."
Dvořák (2011)

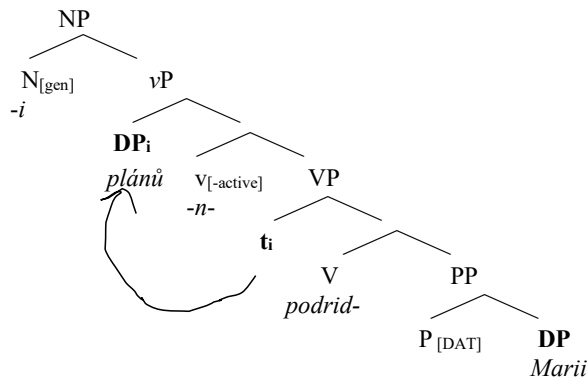
Dvořák assumes that nominalized verbs have a nominalizing head attached to the extended verbal projection with a defective little *v*, which has no case-marking capacity (following Abney 1987, Borer 1999, Alexiadou 2001 among others). The DP merged in the internal argument position (complement of V) receives only the object theta role, but not case. The movement of the DP to a postnominal position where genitive is assigned is therefore necessary (to spec *v*P in her

analysis).⁴ The dative DP, on the other hand, remains in situ, since its theta marking and case-assignment occur simultaneously. The structures for the two nominalizations (together with the relevant movements) are given in the trees below.

(9) Structure for (7), Dvořak (2011)



(10) Structure for (8), Dvořak (2011)



⁴ Dvořak argues that the Czech postnominal genitive is a structural case assigned by a noun to the first DP that it c-commands.

2.2. AGENT AND RECIPIENT, AGENT AND THEME

We now turn to adnominal combinations involving the agent noun phrase. Šarić (2018) notes that it is not always the case that the genitive argument precedes the dative one in adnominal environment and provides cases like (11).

- (11) pretnja lopovu šefa policije
threat thief.DAT chief.GEN police.GEN
“a threat to the thief by the chief of police”
Šarić (2018)

This observation cannot be directly compared with Zlatić's in (2). In both cases we are dealing with a dative recipient; however, the genitive noun phrase is associated with theme in (2) but with agent in (11).⁵

An interesting observation regarding two postnominal phrases in genitive case is found in Kovačević (2013) and is further explored in Šarić (2018). In contrast to Zlatić, these two works provide data showing that Serbian process nominals do allow two genitive arguments, as exemplified in (12).⁶

- (12) osvajanje Rima velikog generala
conquest.NOM Rome.GEN great.GEN general.GEN
“the conquest of Rome by the great general”
Šarić (2018)

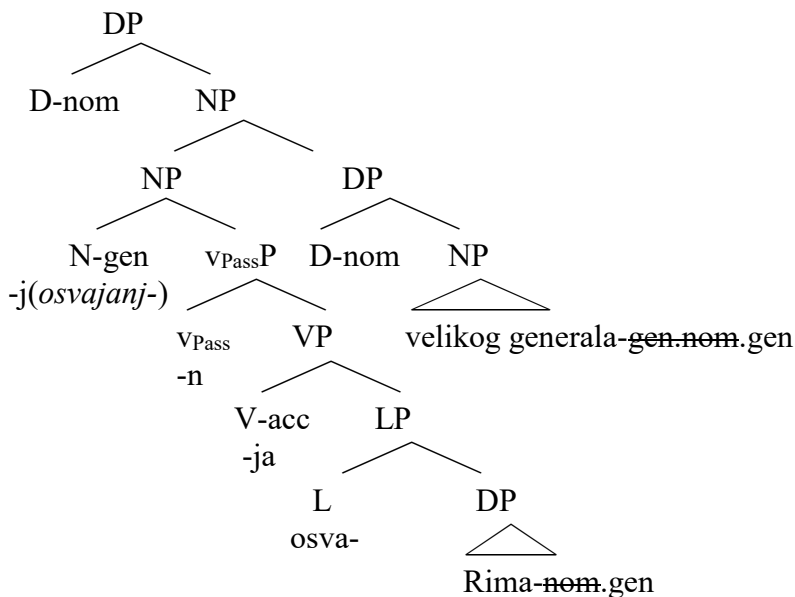
The structure Šarić proposes for process nominals with two genitive arguments is as in (13). The theme argument is introduced as a complement to the root, while the agent is introduced as an adjunct to NP. In Šarić's view (building on Alexiadou 2001), the little *v* head is passive, so it does not project the agent argument, but leaves open the

⁵ Šarić seems to agree with Zlatić on the relative order of theme-genitive >> recipient-dative as in (2), at least she does not indicate otherwise.

⁶ Kovačević (2013) notes that the possibility of expressing the agent argument in genitive case depends upon the possibility of expressing the agent by a possessive adjective – only when the latter is not an option, genitive can be used (e.g. for agents consisting of at least two words or for plural nouns).

possibility of expressing it as an adjunct (either a genitive DP or an adjunct PP “sa strane”). In Šarić’s approach, the order theme-genitive >> agent-genitive falls out of the structure.⁷

(13) Structure for (12), Šarić (2018)



3. THREE DP ADNOMINAL ARGUMENTS

We now turn to Slovenian examples with three adnominal arguments, repeated in (14), and show that an upgrade of the two proposals is needed to derive the correct WO.⁸

(14) podeljevanje nagrad mednarodne žirije najboljšim
filmom

⁷ Šarić (2018) employs the framework of case assignment as proposed in Pesetsky (2013).

⁸ Here, we are only concerned with the reading in which the second genitive phrase (*mednarodne žirije* »international jury«) is understood as the agent of the awarding event and not perhaps the postmodifier to the genitive phrase *nagrada* »awards«, which would lead to the reading »the awards that belong/are associated with the international jury«. Both interpretations are possible and this ambiguity

awarding awards.GEN international jury.GEN
 best movies.DAT
 “presenting awards to the best movies by the international jury”

We first give a brief overview of the structure of Slovenian ditransitive verbs, from which the process nominals in question are derived (section 3.1). The structure of ditransitives and their arguments is significant because in this analysis we adopt an approach to nominalizations that preserves the verbal structure found in the sentence, Distributed Morphology (Halle and Marantz 1993, Marantz 1997, and subsequent work).⁹ We then turn to the nominalizations derived from ditransitive verbs in which three arguments are expressed adnominally (sections 3.2 and 3.3.).

3.1. SLOVENIAN DITRANSITIVE VERBS AND THEIR NOMINALIZATIONS

Marvin and Stegovec (2012) argue that the structure of Slovenian ditransitive sentences depends on the order of the Direct Object (DO) and Indirect Object (IO) arguments and also on the type of verb. The proposal is summarized in the following table:

Table 1: The possible structures for different word orders, Marvin and Stegovec (2012)

| WO: Dat>>Acc (Applicative) | | WO: Acc>>Dat (PDC) | |
|----------------------------|--|---|--|
| <i>give</i> -type verbs | <i>send</i> - and <i>throw</i> -type verbs | <i>give</i> -type verbs | <i>send</i> - and <i>throw</i> -type verbs |
| low applicative | low or high applicative | Prepositional Dative Construction (PDC) | Prepositional Dative Construction (PDC) |

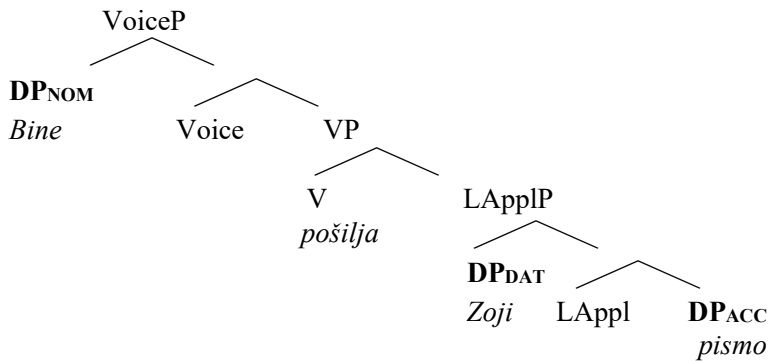
is certainly one of the reasons why native speakers rarely use three adnominal arguments with deverbal nouns.

⁹ Both Šarić (2018) and Dvořak (2011) adhere to the the framework of Distributed Morphology.

In the word order DAT>>ACC, Slovenian ditransitive sentences yield two possible meanings with the corresponding structures, the low and the high applicative one, spelled out in trees like those in (15a and 15b). In (15a), where the dative object is the (intended) recipient of the direct object, the structure is the so-called low applicative as in Pytkkänen (2002, 2008). This interpretation is equivalent to the English Double Object Construction (DOC). In (15b) (nonexistent in English), where the dative object is the beneficiary of the event of sending the letter (but not a recipient of the letter), the structure proposed is the so-called high applicative.^{10,11}

- (15) Bine pošilja Zoji pismo.
 Bine.NOM send.PRES Zoja.DAT letter.ACC

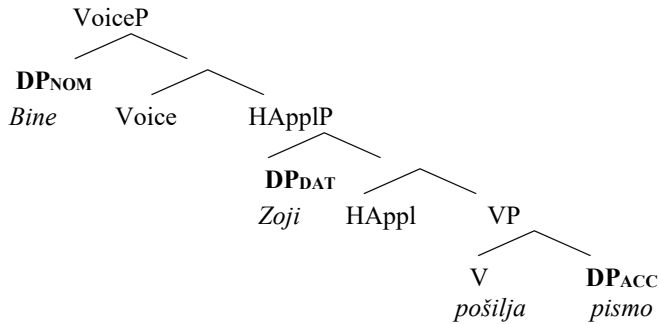
- a. “Bine is sending Zoja a letter.”
 (low applicative; dative object=recipient)



¹⁰ In both structures the dative DP asymmetrically c-commands the Theme DP; given Bruening’s (2001) analysis, we therefore expect a frozen scope in both the low and high readings, which is exactly the case in Slovenian, as Marvin and Stegovec (2012) show.

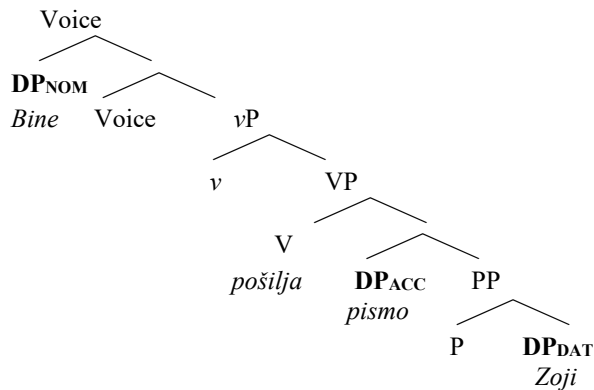
¹¹ The type of verb is also relevant, as *give*-type verbs never yield a high applicative structure in the DAT>>ACC order. This issue is not relevant to the analysis in this paper, so we leave it aside.

- b. “Bine is sending a letter for Zoja (so Zoja wouldn’t have to do it).”
 (high applicative; dative object=beneficiary)



When the word order is ACC>>DAT, the structure is not applicative, but the so-called Prepositional Dative Construction (PDC), in which the dative DP is a complement of a (phonologically null) prepositional head P, as in (16).^{12, 13}

- (16) Bine pošilja pismo Zoja.
 Bine.NOM send.PRES letter.ACC Zoja.DAT
 “Bine is sending a letter to Zoja.”



¹² See Gračanin-Yuksek (2006) for a similar proposal for Croatian.

¹³ The PDC (ACC>>DAT) and the low applicative/DOC (DAT>>ACC) are very similar in meaning, but their structural properties differ crucially in scope and binding of possessives. See Marvin and Stegovec (2012) for details.

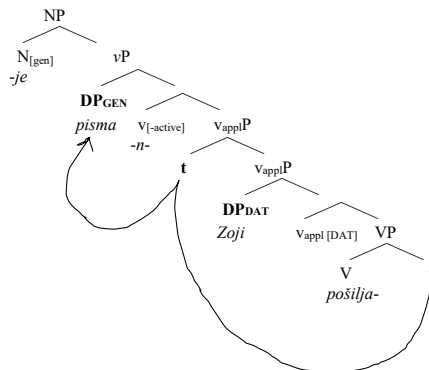
3.2. PROBLEM 1: ORDER OF THEME AND RECIPIENT/BENEFICIARY¹⁴

Leaving the agent genitive aside for the moment, we find that the same situation occurs in Slovenian as in Czech: the WO of arguments in Slovenian sentences with ditransitives is relatively free (ACC >> DAT or DAT >> ACC), while in adnominal environments it is fixed (GEN >> DAT). In other words, the theme in low and high applicative sentence structure is found in a lower position than the recipient/beneficiary, but in a higher position in process nominalizations derived from ditransitive verbs (17).

- (17) a. pošiljati Zoji pismo (low or high applicative)
to send Zoja.DAT letter.ACC
b. pošiljanje pisma Zoji
sending letter.GEN Zoja.DAT

Here, we follow Dvořák and propose movement of the theme argument over the dative argument, shown in the trees below. We provide only one tree for low and high applicative structure, since in both cases the same movement of the theme over the recipient/beneficiary is necessary.

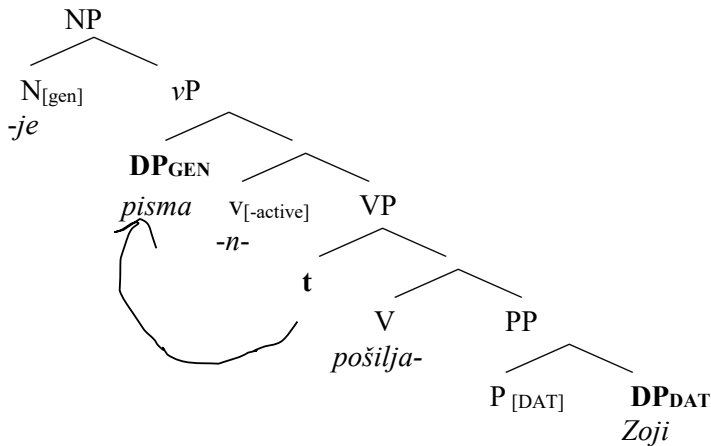
(18) Unified structure for low and high applicative



¹⁴ We add the meaning of beneficiary (to that of recipient) because in Slovenian sentences with DAT>>ACC order, the dative argument can be ambiguous between recipient and beneficiary. However, this does not affect the movements that occur when deriving the order within the process nominals from ditransitive verbs.

For PDC, where the theme argument is higher than the recipient/beneficiary to begin with, we follow Dvořák (2011) and introduce movement. The latter does not affect the relative order of the two arguments, but is still necessary for case reasons.

(19) Structure for PDC



3.3. PROBLEM 2: THE POSITION OF THE AGENT ARGUMENT

We now turn to the more problematic cases, i.e. the ones with double genitives in combination with dative, (20).

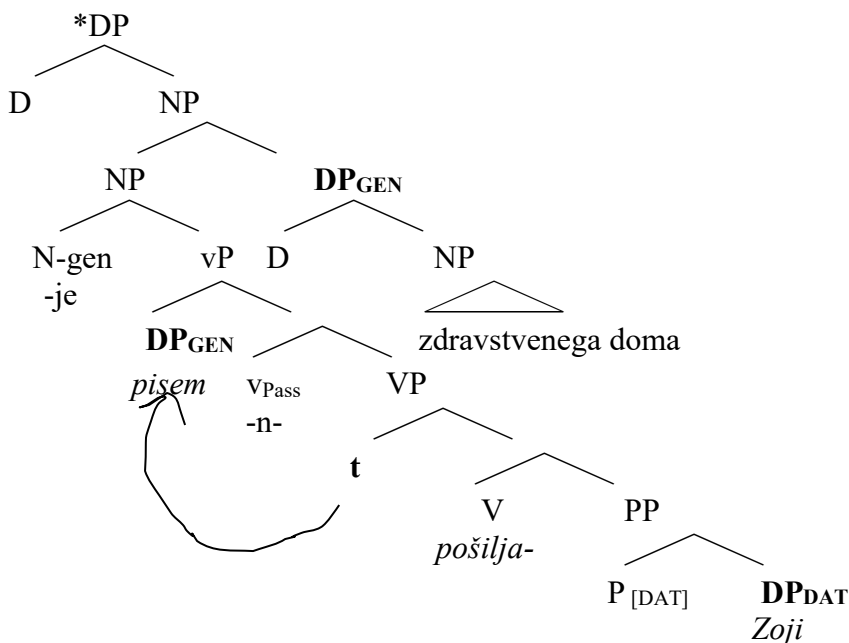
- (20)a. pošiljanje pisem zdravstvenega doma Zoji
 sending letters.GEN_{TH} medical center.GEN_{AG} Zoja.DAT
- b. *pošiljanje pisem Zoji zdravstvenega doma
 sending letters.GEN_{TH} Zoja.DAT medical center.GEN_{AG}

If only two genitives are involved (theme and agent), then Šarić’s proposal in (13) seems a reasonable option for Slovenian nominalizations.¹⁵ However, her analysis does not lead to the correct

¹⁵ Slovenian deverbal process nominalizations pattern with Serbian ones in many ways. The most important is the fact that they allow two genitives (unlike many other languages that do not).

WO if we add the dative argument (recipient/beneficiary). Suppose we adopt her structure, place the dative argument in the positions proposed by Marvin and Stegovec (2012), and then follow Dvořák's procedure of moving the internal argument to a position higher than the dative argument when forming the process nominal. In this way we get the correct order Head >> Genitive-theme >> Dative-recipient, but still do not get the correct position of the agent. The latter should be between the theme and the recipient/beneficiary (20a), but in this structure, it is the last element in row, being merged as an adjunct in the rightmost position. Here, we give an example of the PDC ditransitive construction integrated into Šarić's tree for the double genitive construction; exactly the same situation occurs with the high and the low applicative (the agent appears last in the linear sequence).

(21) Incorporating a ditransitive verb into the structure with the genitive agent as adjunct



4. PROPOSAL

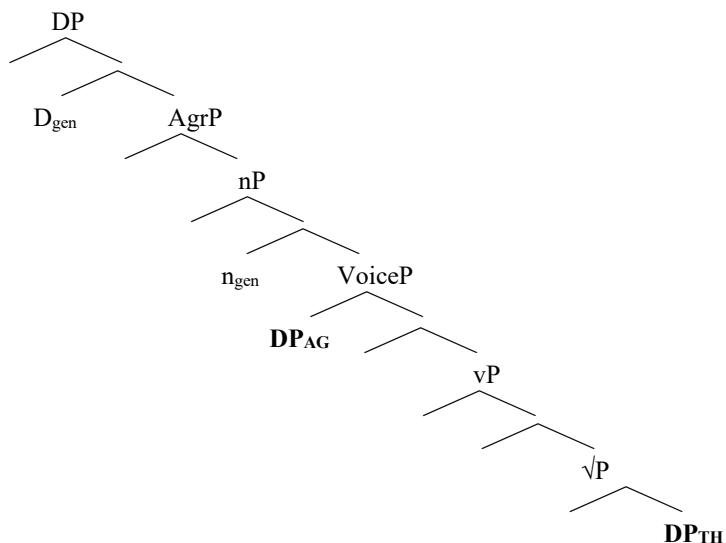
In this part we present a proposal that solves the WO problem of the three adnominal arguments.¹⁶ The approaches that we present in the paper (Šarić 2018, Dvořak 2011) are all based on Alexiadou's (2001) proposal, which eliminates the agent introducing little *v* from process/event nominals. Alexiadou claims that the little *v* in such nominals is intransitive, while Šarić and Dvořak go one step further, positing a little *v* that is referred to as »passive« (Šarić) or »[-active]« (Dvořak). Thus, the agent can only be expressed as an adjunct (in Alexiadou's approach as a PP in Šarić's analysis of Serbian as a PP or as an adjunct genitive NP). In contrast to these authors, we claim that process/event nominals contain VoiceP, which introduces the agent in its specifier position with transitive and unergative verbs and that the correct WO can be derived by way of movement. Here, we follow the proposal in Cuervo (2003), in which the little *v* is an event introducer (of the types *v*DO, *v*BE or *v*GO), while Voice is responsible for the syntactic and semantic licensing of the external argument. Specifically, Voice relates the external argument (projected as its specifier) to the event described by the verbal phrase that Voice takes as its complement.¹⁷

In transitive and unergative verbs VoiceP appears on top of *v*P, which in these verbs is of type *v*DO or *v*BE (in the sense of Cuervo 2003). VoiceP is not present with unaccusatives, because the unaccusative little *v* (*v*GO) and Voice are incompatible. The basic structure for nominalizations from transitive verbs is as follows:

¹⁶ In this paper, we leave aside the issue of case assignment, since the focus here is on WO. Moreover, the approaches compared here use very different theories of case assignment, although they all treat the derivation of process nominals in a similar way (Distributed Morphology approach).

¹⁷ Voice combines with its complement *v*P via a semantic rule called Event Identification and adds the external argument as a participant of the event (Kratzer 1996).

(22) DP



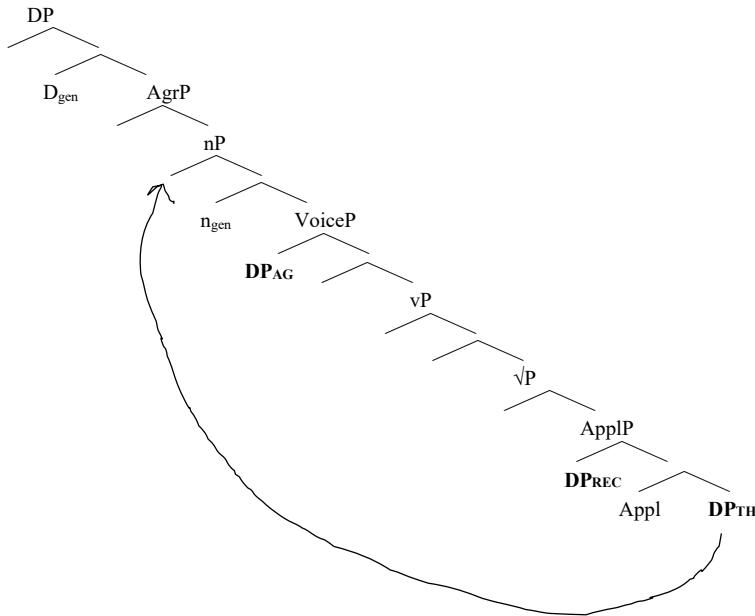
The object DP gets its theta role from the little *v*, but since this *v* is defective within the nominalizing environment (see e.g. Dvořák 2011), it cannot provide case for the object. Thus, the object DP must be moved to a position where it can be in a local relation to a genitive case-valuing nominal head, which is *n* in Dvořák's proposal. Here, we propose that the object DP be moved to spec *nP*, where it receives genitive case.¹⁸ The subject DP remains in situ, receiving the theta role from Voice and case from D (genitive in this case is checked at a distance). The root picks up the morphology and ends up in Agr, so the final linear order is the nominal head followed by the object DP and the subject DP.

A similar structure is proposed for ditransitive verbs with the following difference: the recipient/beneficiary DP gets its dative case inherently, from its theta role assigner, which can be a low applicative

¹⁸ In Dvořák's proposal, where there is no Voice, it moves to spec *vP*, where it is assigned the genitive by N (equivalent to *n* in this tree). For Dvořák, the Czech postnominal genitive is a structural case assigned by a noun to the first DP that it c-commands and it need not be accompanied by theta assignment.

head (in DOC), a high applicative head (in high applicatives) or a silent preposition (in PDC). In all three cases, the agent DP and the recipient/beneficiary DP remain in situ, while the theme DP moves to spec *nP*, which results in the correct WO of adnominal arguments $DP_{TH} \gg DP_{AG} \gg DP_{REC}$. Here, we provide an example with the low applicative (DOC); the other two structures yield the same WO after the movement of the theme DP (underlying direct object).

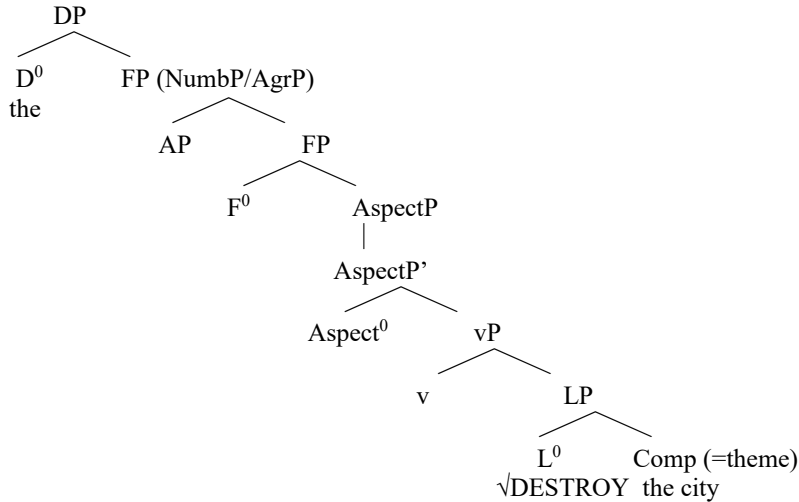
(23) Deriving the WO via movement



5. REVISITING ALEXIADOU (2001)

We now revisit Alexiadou’s original proposal and show that the arguments for eliminating the agent introducing little *v* from the structure of process nominals are not justified for Slovenian. Alexiadou (2001) presents a proposal in which the head *v* is present in the structure of nominals with the process/event reading, (24).

(24) Alexiadou's (2001) structure for process nominals



In her proposal, the little *v* crucially does not project an agent; it is an intransitive *v* [-external argument]. For this reason, it is also unable to assign accusative case. By focusing on the position of the agent noun phrase, we now review the arguments for Alexiadou's proposal and show that they do not hold for Slovenian. This is independent support for keeping the agent in the VoiceP projection rather than placing it into an adjunct position when deriving Slovenian process nominals from transitives and unergatives.

5.1. IMPOSSIBILITY OF TWO GENITIVES IN PROCESS/EVENT NOMINALS

Alexiadou (2001) shows that in Greek (and in several other languages), the co-occurrence of two genitive phrases with process nouns is impossible and that the agent can only be expressed by a PP¹⁹:

¹⁹ Two genitives are possible in result nominals, but in such cases the second genitive denotes a possessor and not an agent.

(i) metafrasi tis tragodias tu Kakridi
 the translation the tragedy.GEN the Kakridis.GEN

(25) i katastrofi tis polis apo tus Italus
the destruction the city.GEN by the Italians
“the destruction of the city by the Italians”

(26) *i katastrofi tis polis ton Italon
the destruction the city.GEN the Italians.GEN
Alexiadou (2001)

This is not true for Slovenian and Serbian, as can be seen from the examples in this paper, in Kovačević (2013), and in Šarić's work (examples (1), (12)).

5.2. PROCESS/EVENT NOMINALS FROM UNERGATIVES

Alexiadou claims that process nominals can be derived from unaccusatives, but not from unergative predicates.²⁰ When an unergative nominalization appears with the genitive argument, (27), the latter is not interpreted as an agent, but as a possessor and the nominal as a result nominal and not a process nominal. This is supported by the impossibility of a PP by-phrase expressing the agent in Greek examples, (27).

(27) to kolimpi tu Jani /*apo to Jani
the swimming the John.GEN/*by the John

Again, Alexiadou's claim cannot be generalized to Slovenian, since process nominals can be freely derived from unergative verbs, (28).²¹

(28) tekanje otrok po igrišču cel dan
running kids.GEN on playground whole day

„Kakridis's translation of the tragedy“

²⁰ It is interesting to note that in the analysis of Grimshaw (1990) and Borer (1993) it is impossible to derive nominalizations from unaccusative verbs.

²¹ The adverbial modification is added to show that we are indeed dealing with the process reading of the nominal.

Alexiadou mentions the possibility of unergative verbs in Russian process nominals, citing examples from Schorlemmer (1995). Nevertheless, she concludes that this is not a problem for her analysis, noting only that one could argue that Russian does not distinguish between unergatives and unaccusatives. There is, however, ample evidence in research on Russian for the existence of a distinction between the two classes of verbs, e.g. Schorlemmer (1995), Harves (2009). For a discussion of Slovenian unaccusative verbs existing as a group with special properties, see Marvin and Ilc (2016) and Simonović and Mišmaš (2022). We believe that the data from Slavic languages (at least Slovenian and Russian) argue against generalizing Alexiadou's claim that process nominals cannot be derived from unergative verbs.²²

5.3. A SINGLE GENITIVE ARGUMENT MUST BE A THEME IN PROCESS NOMINALS

Finally, Alexiadou's proposal is based on the observation that the single genitive argument within process nominals is necessarily interpreted as a theme. Agents, when syntactically realized, must be realized as PPs (and are treated as adjuncts). In (29), an example with a process nominal, the genitive is necessarily understood as the object of the observation.²³

- (29) i paratirisi tis Marias
 the observation the Mary.GEN

Again, this observation does not hold for Slovenian, as we find many process nominals where the single genitive argument is ambiguous between a theme and an agent. In (30), for example, the

²² The claim here is not that Alexiadou's account should be rejected outright, but that her analysis, which applies to the languages treated in her paper, does not apply to Slovenian.

²³ The genitive can appear with the nominalization in its plural form, but is then not understood as an agent, but as a possessor.

(i) i paratirisi tis Marias ine panda akrivis.
 the observations the Mary are always exact.PL

genitive *lovcev* “hunters” can be understood as agent or as theme, in both cases with the process reading of the nominal.

- (30) *streljanje lovcev cel dan*
shooting hunters.GEN whole day

In some cases, a single genitive noun phrase cannot be interpreted as a theme, but is interpreted as an agent. An example is given in (31).

- (31) *upiranje otrok brezčutnim učiteljem*
resisting children.GEN insensitive.DAT teachers.DAT
“children’s resistance to insensitive teachers”

6. CONCLUSION

In this paper we have tried to solve the problem of deriving the correct word order in Slovenian nominalizations with three adnominal arguments (agent, theme, and recipient) appearing simultaneously as postmodifiers in a nominal phrase whose head is a process deverbal noun. When these arguments are all expressed by noun phrases, their order is fixed: theme genitive >> agent genitive >> recipient dative.

First, we have shown that the order of the three arguments cannot be derived simply by adopting other proposals for Slavic languages dealing with adnominal arguments in process nominalizations (Zlatić 1997, Dvořák 2011, Šarić 2018). Second, a proposal was presented in which the correct order is achieved by introducing the VoiceP in the nominalization structure with transitive and unergative verbs and by movement of the theme argument above the agent and recipient arguments. To provide independent support for this view, Alexiadou’s (2001) original proposal was reconsidered and it was shown that in Slovenian, the arguments for Alexiadou’s elimination of the agent introducing little *v* from the structure of process nominals cannot be sustained.

REFERENCES

- Abney, S. P. (1987). *The English Noun Phrase in Its Sentential Aspect*. PhD thesis, MIT.
- Alexiadou, A. (2001). *Functional structure in nominals: Nominalization and ergativity*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Borer H. (1993). *Parallel Morphology*. Ms. University of Massachusetts at Amherst.
- Borer H. (1999). *The form, the forming and the formation of nominals*. Ms. USC.
- Bruening, B. (2001). QR obeys superiority, frozen scope and ACD. *Linguistic Inquiry* 32: 233–273.
- Cuervo, M. C. (2003). *Datives at large*. PhD thesis, MIT.
- Dvořák, V. (2011). Inherent case and locality requirement: Evidence from ditransitives and their nominalizations. In: *Proceedings of the 34th Annual Penn Linguistics Colloquium*, Philadelphia: University of Pennsylvania Working Papers in Linguistics, 95–104.
- Gračanin-Yuksek, M. (2006). Double object construction in Croatian, Arguments against Appl0, In: *Formal approaches to Slavic linguistics* 13, Michigan Slavic Publications: Ann Arbor.
- Grimshaw, J. (1990). *Argument Structure*. Cambridge, Mass.: MIT Press.
- Halle, M. and A. Marantz (1993). Distributed Morphology and the Pieces of Inflection. In: *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (K. Hale and S. J. Keyser, eds.), Cambridge, MA: MIT Press, 111–76.
- Harves, S. (2009). Unaccusativity. In *Die slavischen Sprachen/The Slavic Languages* [Handbuchen zur Sprach- und Kommunikationswissenschaft Vol. 1], (T. Berger, S. Kempgen, P. Kosta and K. Gutschmidt, eds.), Berlin: De Gruyter Mouton, 415–430.
- Ilc, G. and T. Marvin (2016). Unaccusatives in Slovenian from a cross-linguistic perspective. In: *Formal studies in Slovenian syntax: In honor of Janez Orešnik* (F. Marušič, and R. Žaucer, eds.), Amsterdam: John Benjamins, 145–166.
- Kovačević, B. (2013). „Именице са темом и агенсом исказаним генитивом у српском језику”. *Српски језик XVIII*, 663–668.
- Kratzer, A. (1996). Severing the External Argument from Its Verb. In: *Phrase Structure and the lexicon* (J. Rooryck and L. Zaring, eds.), Dordrecht: Kluwer, 109–138.

- Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In: *University of Pennsylvania working papers in linguistics* 4 (A. Dimitriadis et al., eds.), Philadelphia: University of Pennsylvania. 201–225.
- Marvin, T. and A. Stegovec (2012). On the syntax of ditransitive sentences in Slovenian. *Acta Linguistica Hungarica* 59 (1-2), 177-203.
- Pesetsky, D. (2013). *Russian case morphology and the syntactic categories*. Cambridge, MA: MIT Press.
- Pylkkänen, L. (2002). *Introducing arguments*. PhD thesis, MIT.
- Pylkkänen, L. (2008). *Introducing arguments*. Cambridge, Mass: MIT Press.
- Rappaport, G. (1998). *The Slavic Noun Phrase*. Ms. available at: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.31.2503&rep=rep1&type=pdf>
- Schoorlemmer, M., (1995). Participial passive and aspect in Russian. PhD thesis, Utrecht University.
- Simonović, M. and P. Mišmaš (2022). Lowest theme vowels or highest roots? An 'unaccusative' theme-vowel class in Slovenian. *Glossa: a journal of general linguistics* 7(1).
- Šarić, A. (2018). *Nominalizations, double genitives and negation*. Berlin: Mouton de Gruyter.
- Zlatić, L. (1997). *The structure of the Serbian noun phrase*. PhD thesis, University of Texas at Austin.
- Žele, A. (2001). *Vežljivost v slovenskem jeziku: s poudarkom na glagolu*. Ljubljana: ZRC SAZU.

Татјана Марвин

АДНОМИНАЛЬНЫЕ АРГУМЕНТЫ ВО ФРАЗАХ, ВОЗГЛАВЛЯЕМЫХ
ОТГЛАГОЛЬНЫМИ СУЩЕСТВИТЕЛЬНЫМИ В СЛОВЕНСКОМ ЯЗЫКЕ

Резюме

В статье рассматривается вопрос об образовании порядка слов в словенских номинализациях с тремя адноминальными аргументами (действующее лицо, тема и реципиент), выступающими одновременно в качестве постмодификаторов в именной группе, головной частью которой является процессное отглагольное существительное. Если все эти аргументы выражены именными группами,

их порядок фиксирован: головное существительное >> тема в родительном падеже >> агент в родительном падеже >> реципиент в дательном падеже. Основная цель статьи - показать, что невозможно вывести такой порядок, непосредственно применяя решения относительно адноминальных аргументов в процессных номинализациях, предложенные для других славянских языков (Златич 1997, Дворжак 2011, Шарич 2018). В статье предлагается анализ, при котором правильный порядок трех адноминальных аргументов в словенском языке достигается путем введения VoiceP в структуру номинализации с переходными и неэргативными глаголами и перемещением аргумента темы над аргументами агента и реципиента.

Ключевые слова: отглагольная номинализация, адноминальное дополнение, порядок слов, словенский язык

Zorica Kovačević*

University of Belgrade – Faculty of Philology

ANALYSIS OF SYNTACTICAL STYLISTIC ELEMENTS IN SWEDISH

Abstract

Sentence structure is one of the most important components of stylistic analysis since it primarily represents textual readability and comprehensibility. Languages with predominantly strict word order, such as Swedish, show various stylistic characteristics. Stylistic analysis is affected by numerous aspects e.g. parataxis, hypotaxis, topicalization, duplication, elliptical sentences, sentence structure in formal and informal texts combined with additional syntactic features. Stylistic variations of a given text can also improve language learning. The aim of this study is to examine the most frequent examples of syntactical stylistic analysis in Swedish and to emphasize their importance for advanced language learning and translating.

Key words: Swedish syntax, standard language, translation, advanced language learning

1. INTRODUCTION

Previous contrastive studies between the Scandinavian language communities and Serbian community do not contain any remarks on stylistic studies and the main tendencies in the stylistic research in Scandinavia. This is also the reason why it often is a great challenge in

* zorica.kovacevic@fil.bg.ac.rs

teaching and contrasting Swedish and Serbian to clarify the differences in stylistic traditions in these two linguistic and cultural communities. The tradition of stylistic studies in our language community has so far dealt mainly with special cases in literary texts, the Swedish and the Scandinavian stylistic traditions in general have been dealing also with non-literary functional styles in stylistic analysis for the past half century.

In relation to the type of stylistic studies, normative stylistics is the oldest stylistic discipline in Scandinavia. Thus, for example, in Swedish history of linguistic and stylistic studies during the 17th century, names and studies such as Bergbo's book on the importance of the Swedish language and challenges in improving its status in Europe at that time (Bergbo, 1658: *Thet Swenska Språketz Klagomål*) stand out. Further on, in studies on speeches in central Sweden (Stiernhjelm, 1643: *Gambla Swea- och Götha-Måles Fatebur*) or Columbus's handbook for the preservation of the Swedish language (Columbus, 1674–1678: *En Svensk Orde-Skötsel*) – all authors speak, in fact, about the excellence of the Swedish language (*svenskans förträfflighet*), but also about the methods that would help Swedish gain a well-deserved place among other languages in the world. Columbus, for example, points out that the power of the (Swedish) language is evident both in its originality and in its use in the modern age (Kovačević 2021: 22).

Somewhat later in the 17th century, the classicist approach became increasingly popular, relying not only on the thesis of a language originality factor that influences its strong position in the language community, but also on factors such as clearer communication of learned people of the time. Descriptive and functionalist approaches in stylistic studies have become very common in Scandinavia, especially in Sweden. They have, nonetheless, never excluded normative-rhetorical stylistics, but focused more on language use, situational factors, efficiency and less on systematized rhetoric (Enkvist 1973: 77–78). The approach to stylistic studies in the Scandinavian languages for the past fifty years has over time become predominantly functionalist, as best seen in Lagerholm's observation (2008: 23) that the answer to the question of what stylistic studies include is that they investigate

all types of texts, bearing in mind that each text is determined by a certain style (Kovačević 2021: 23).¹

In his comment about the natural connection between languages and linguistic approaches to stylistics Cassirer (1999: 21) uses examples from the history of Swedish and its stylistic studies specifically pointing out that *language* and *style* are two words that are often used together. The predecessor to an annual publication that dealt with stylistics issues, *Nysvenska studier* (New Swedish Studies), was called *Språk och stil* (Language and Style) until 1920 and again in 1990s. In stylistic surveys, Cassirer remarks that the authors often switch between the terms *language* and *style* without clearly stating whether they intend to mark the difference. It is also obvious that the terms are close to each other. Attempts have been made to distinguish between them by pointing out the difference between language errors and style aberrations (Askeberg in *Stilbegreppet* (1954) and his work on the definition of the concept *style*) or by naming linguistic phenomena that extend stylistically beyond units such as clause and sentence or by defining style as statistically significant (linguistic) deviation from a norm. Definitions like the latter one, however, in Cassirer's opinion make the term *style* unnecessary where one could settle for the term *language*. If the term *style* is to be retained at all, it should be used in situations where the term *language* is not sufficient. The description of the term style is often based on a linguistic analysis, but that is usually not enough. Stylistics also wants to investigate all possible correlations between a type of sentence and the content expressed in the sentence and, last but not least, the effect this relationship

¹ Cassirer (1999: 60) also agrees that stylistic analysis should be applied throughout language and content and not only in connection with literary genres. It has undeniably been to the detriment of the stylistics that experts have devoted themselves so much to the unique (literary) texts instead to the ordinary and trivial. Therefore, almost all stylistic analysis has been about unique cases of combinations where an ingenious author has achieved certain effects in an unprecedented way. Stylistics has therefore faced great difficulties in formulating its general rules.

produces. Style thus becomes a term that denotes one's experience of the language and stylistics a method for a thorough text description.

2. WORD ORDER AND THE FORMAL FUNDAMENT IN SWEDISH

In a historical overview, the Old Norse language was mainly stable.² It consisted of somewhat free word order like other older Germanic languages, but certain basic principles were already present, such as the first or the second position of a finite verb, and the object following the verb. Strict word order was apparent already during the second half of 8th and the first half of 11th centuries and also present at the beginning of the administrative style of the Scandinavian provincial laws. Strictly speaking, the word order was very early close to the one in the Scandinavian languages today, i.e. the F/v1/n/a/V2/N/A structure.³

The formal fundament field in Swedish (Swe. *fundamentfältet*) is crucial for information structuring. This is the part of the independent clause that precedes the main finite verb (v1) becoming its presentational focus:

- (1) *Sven* (F) läser (v1) en bok. (*Sven* is reading a book.)
- (2) *Sven, klassens bästa elev, som tänker ägna sig åt att studera språkvetenskap* (F) läser (v1) en bok. (*Sven, the best pupil in the class, who intends to study linguistics* is reading a book.)

This field usually contains all types of constructions, from nominal phrases to complete clauses. This in return affects the length of the Swedish fundament or the quantity of the whole sentence, in Swedish often described as its “weight” (Swe. *tyngd*), respectively “left

² Old Norse lasted from about 8th to 14th centuries in Scandinavia and coincided with the Viking Age.

³ The F/v1/n/a/V2/N/A sentence structure consists of the following features: F – the formal fundament field (Swe. *fundament* and/or *fundamentfältet*); v1 – finite verb; n – pronominal phrase/subject; a – adverbial phrase/sentence adverb; V2 – infinite verb; N – pronominal phrase/object; A – adjective.

sided weight” (vänstertyngd) for sentences that consist of a massive fundament field and “right sided weight” (högertyngd) for the ones that contain a relatively light fundament field compared to the rest of the sentence. Therefore, there is more information in the beginning of the sentence if the fundament contains more elements. Lagerholm comments that much attention should be given to the length and the number of words the fundament contains and gives two specific examples (constructions that belong to the fundament fields are here purposely written in italic):

- (3) *Även om Pär Lagerkvist (1891–1974) i sin utveckling visar stor konsekvens och enhetlighet – till exempel i sin medvetna stilvilja och sitt ständiga återvändande till livsåskådningsmässiga, religiösa grundproblem – kan det för översiktens skull vara lämpligt att urskilja fyra perioder i hans diktning [...]. (Lagerholm 2008: 131)⁴*
- (4) *Varför firar vi våren med eld på Valborgsmässoafton? Någon entydig förklaring finns inte. Men det är troligt att elden tillsammans med oväsen från grytlock och annat, använts för att skrämma rovdjur på flykten i samband med att kreaturen släppts på vårbeta. Seden är vanlig på många håll i nordliga länder. I Sverige känner vi den sedan 1600-talet. (Lagerholm 2008: 132)⁵*

Example (3) is excerpted from a book about literary history at university level and shows the length of the clauses used in the formal fundament field. The purpose here is to explain and investigate the

⁴ “Although Pär Lagerkvist (1891–1974) in his development shows great consistency and uniformity – for example in his conscious will to style and his constant return to basic philosophical, religious views – it may be appropriate for the sake of overview to distinguish four periods in his poetry [...].”

⁵ “Why do we celebrate spring with fire on Walpurgis Night? There is no unambiguous explanation. But it is probable that the fire, together with noise from pot lids and other things, was used to scare predators on the run, in connection with the cattle being released out for spring. The custom is common in many places in northern countries. In Sweden we have known it since the 17th century.”

significant features of the writer's poetics before naming the main periods in his work. Example (4) is cited from a tourist brochure, its style is less complex and the fundament field in each sentence consists of relatively small constructions (a question word, a nominal phrase, a prepositional phrase etc.).

In general, depending on whether the fundament is extensive or not, the left/right weight of the sentences does affect textual complexity and readability. Informal texts tend to contain fundament fields that consist of fewer words. Such texts are easier to read and understand and they often correspond to spoken language. On the other hand, formal texts consist mostly of sentences with predominantly left side weight. Such texts are very difficult to read and understand. Lagerholm comments that, according to Margareta Westman (1974), less formal Swedish texts have an average of 3.1 words per fundament field. The material in Westman's research consisted of newspaper articles, brochures, textbooks and debate texts. The field of fundament is more extensive in very formal writing which is confirmed by Landqvist's (2000) study of legal Swedish. According to this, the average fundament length is 6.0 words in the Swedish state regulations (2008: 132).⁶

3. STYLISTIC FEATURES AND TEXTUAL CHARACTERISTICS

In text linguistics, the first place in the sentence is usually called a *theme* and the field closer to the end of a sentence is called a *rheme*. In

⁶ Lagerholm refers to two studies, Margareta Westman's 1974 doctoral thesis *Brukprosa* (Utility prose) and Hans Landqvist's book from 2000 *Författningssvenska. Strukturer i nutida svensk lagtext i Sverige och Finland* (Legal Swedish. Structures in modern Swedish legal texts in Sweden and Finland.). Original in Swedish: "Enligt Westman (1974) har bruksprosatexter i genomsnitt 3,1 ord per fundament. Brukprosa i hennes undersökning bestod av tidningstext, broschyrer, läroböcker och debattexter. Att fundamenten är längre i mycket formell skrift bekräftas av Landqvists (2000) undersökning av författningssvenska. Enligt denna är den genomsnittliga fundamentlängden 6,0 ord i svenska författningar" (Lagerholm 2008: 132).

most cases, already known information is on the theme side and new information on the rheme side (the topic of the sentence). If a new information cannot be provided, different linguistic means are used in Swedish to avoid starting the sentence with a new information. In other words, various constructions can be used to avoid putting the new information in the first place in a sentence:

- (5) “En springande pojke kommer. (A running boy is coming.)
- (5.a) *Det kommer* en springande pojke. (*There comes* a running boy.)
- (5.b) *Det är* en springande pojke som kommer. (*It is* a running boy coming.)
- (5.c) *Nu är det så att det kommer* en springande pojke. (*Now it's so that there comes* a running boy.)

In addition, in Swedish formal texts long fundament field carries the theme of the sentence, e.g.:

- (6) *Från utrikes ort infört gods, som är avsett till proviantiering och utrustning av fartyg eller luftfartyg eller, enligt vad där- om särskilt stadgas, till försäljning till flygpassagerare på tullflygplats och som tillika är belagd med tull eller annan in- förselavgift, må efter verkställd tulltaxering, utan att dylika avgifter erläggas men, såvitt avser andra varor än rusdrycker och tobaksvaror, mot ställande av säkerhet för belopp som mot- svarar avgifterna, uppläggas i särskilt magasin under vård av godsets ägare. (Lagerholm 2008: 208)⁷*

Contrary to the cases where new pieces of information are being presented closer to the end of a sentence, the cases of *topicalization* (Swe. *topikalisering*; *fundamentering*) contain the main semantic

⁷ “Goods imported from abroad intended for the provision and equipment of ships or aircrafts or, as specifically provided, for sale to passengers at airport’s customs and which are also subject to customs duties or other import duties, without additional fees, in the case of goods other than alcoholic beverages and tobacco products, in accordance with provision of security for amounts corresponding to the fees, are placed in a special storage under the supervision of the owner of the goods.”

theme already at the beginning of a sentence. Topicalization in Swedish is implemented directly by placing the needed segment in the fundamental field with the finite verb (v1) following.

- (7) Jag gick på bio för två dagar sedan. (I went to the cinema two days ago.)
- (7.a) På bio gick jag för två dagar sedan. (It was to the cinema I went two days ago.)
- (7.b) För två dagar sedan gick jag på bio. (Two days ago, I went to the cinema.)

The cases of topicalization in Swedish reflect differently on written and spoken Serbian and in specific situations it is very challenging to convey the fact that certain segments of a sentence are of greater importance than the others, e.g.:

- (8) Jag tycker om honom. (I like him.)⁸
- (9) *Honom* tycker jag om. (It is him I like.)⁹
- (10) *Honom* vet jag att du inte kan stå ut med.¹⁰ (Him I know you can't stand.)
- (11) *Skicka bort pojken till internatskola* ville de inte.¹¹ (To send the boy away to boarding school is something they didn't want.)

In all of the abovementioned examples it is possible to produce an adequate translation by using additional words or different word order so that the same effect can be provided in the target language. Nevertheless, oral translation of the examples could also rely on verbal features (stress, tone etc.) in order to convey the exact meaning from the source language often without changing the word order

⁸ In Serbian, in a rather neutral meaning: On mi se dopada.

⁹ In Serbian, in a specific meaning: *On* mi se dopada. / *Upravo on* mi se dopada.

¹⁰ *Swedish. A Comprehensive Grammar* (2013: 547); in Serbian: (Upravo) *njega* ne podnosiš, to znam. / Znam da (upravo) *njega* ne podnosiš.

¹¹ *Swedish. A Comprehensive Grammar* (2013: 546); in Serbian: Nisu hteli da pošalju dečaka u internat. / (?) *Poslati dečaka u internat* je nešto što nisu hteli. The latter example seems to be slightly less common.

(see example (10)). In many cases there are even two or several other solutions in translation, yet not all of them can be equally appropriate (see example (12)).

Lagerholm emphasises that the notion *parataxis* denotes sentences dominated by coordination (Swe. *samordning*) and that the style is then paratactic. When a language is instead dominated by subordination (Swe. *underordning*), this is called *hypotaxis*, and the style is then hypotactic. Parataxis and hypotaxis are a purely quantitative phenomena determined by the frequency of subordinate clauses. This can be specified, for example, by the number of clauses per 1,000 words, clauses per graphic sentence or clauses per syntactic sentence. Lagerholm also comments that it has been argued that parataxis is typical of spoken Swedish language and more comprehensible texts, while hypotaxis is typical of formal writing in Swedish (2008: 126).¹² Translation exercises often include the cases of parataxis with identical expressions, such as:

- (12) Han blev *tröttare och tröttare*.¹³ (He was getting *more and more* tired.)
- (13) Det finns *folk och folk*. (There are people *of all kinds*.)¹⁴
- (14) Frisk och sund. (Safe and sound.)¹⁵

Examples (12) and (13) show that in some languages paratactic segments can include different parts of speech (e.g. adverbs instead of verbs) or that they even can be avoided. Example (14) consists of

¹² In our comment this is the citation we have used in its Swedish original: "När meningar domineras av samordning kallas den tendensen för paratax, och stilen är då parataktisk. När språket i stället domineras av underordning kallas detta för hypotax, och stilen är då hypotaktisk. Paratax och hypotax är ett rent kvantitativt fenomen och avgörs av bisatsfrekvensen. Denna kan till exempel anges i antal bisatser per 1 000 ord, bisatser per grafisk mening eller bisatser per syntaktisk mening. [...] Traditionellt har det hävdats att paratax är typiskt för talspråk och ledig, enkel prosa medan hypotax är typiskt för formell skrift."

¹³ *Svenska Akademiens grammatik*, Vol.4 (1999: 901); in Serbian: Umarao se *sve više*.

¹⁴ In Serbian: Ljudi ima *i ovakvih i onakvih*.

¹⁵ In Serbian: Živ i zdrav.

rather fixed paratactic expressions that in different languages have different form, but correspond semantically.

The *duplication* or the *doubling* (Swe. dubblering) of a clause element by means of pronominal word is common in spoken Swedish. Most frequently duplicated in this way are indefinite or article-less forms. The effect of duplication is to emphasise the free element placed in the extra position outside the clause. Duplication in Swedish sentences can occur in different parts of a sentence, but most often in its final part. Final duplication is one of the most prominent features in Swedish used mainly in dialects and colloquial style which is something that should be pointed out in translation exercises. Here are some examples of duplication in Swedish:

- (15) I Uppsala, *där* trivdes jag under min studietid. (In Uppsala, *there* I enjoyed my life as a student.)¹⁶
- (16) Att tala ryska, *det* är ju inte lätt, *det inte!* (Speaking Russian, *that* isn't easy, *that!*)¹⁷
- (17) Han är rolig, *Olof*. (He is funny, *Olof*)¹⁸
- (18) Jag kan inte göra det, *jag* heller. (Me? *I* can't do that either.)¹⁹

In the first two examples (15) and (16) one or more cases of duplication can be observed both in the middle and in the end of the sentence. Examples (17) and (18) represent the cases of final duplication. An important point in teaching translation of the abovementioned and other similar examples of duplication in Swedish is the fact that most examples are impossible to translate literally, because of the contextual variants that need to be transferred to the target language in a different way. Repetition of constructions in translation in places

¹⁶ *Swedish. A Comprehensive Grammar* (2013: 549); in Serbian: (*Upravo*) u Upsali sam uživala u studentskim danima.

¹⁷ *Swedish. A Comprehensive Grammar* (2013: 549); in Serbian: (*Pa i*) nije tako lako govoriti ruski!

¹⁸ *Swedish. A Comprehensive Grammar* (2013: 548); in Serbian: (*Onaj*) Ulof, (*taj vam*) je baš zabavan.

¹⁹ *Swedish. A Comprehensive Grammar* (2013: 548); in Serbian: To čak ni ja ne mogu da uradim.

where they are found in Swedish examples often appears as unnatural, sometimes even incomprehensible, which is an important information in language and translation teaching and learning.

Exceptionally small part of teaching and learning material contains information that the final duplication in Swedish is a phenomenon related to the spoken language and most often a feature that is present in dialectal speech and rarely in written Swedish. Lindström (2011: 211) points out that the pronominal duplication is more established in the Finnish-Swedish dialects around Helsinki, but the final duplication in different forms is characterized by, if not all Swedish dialects, then a majority of them. This means that in teaching translation from Swedish, in addition to introducing future translators to the concept of final duplication and its forms, it is necessary to make a note regarding the domains of how these constructions are used. Consequently, translation of these constructions includes various non-linguistic factors, specifically cultural, historical, social and other factors in the translation process. Having in mind that other languages may also contain constructions with final repetition, but almost never in the same sense and in the same speech situations as in Swedish, a new question may occur: what kind of translation is acceptable in order to achieve intelligibility, grammatical accuracy of translation and communicativeness in relation to the audience of the target language (Kovačević 2020: 246).

Certain types of constructions can etymologically or logically be seen as ellipses of more extensive expressions, but should simultaneously still represent complete constructions.

(19) Att han inte ger sig! ((It is so wonderful that) he is not giving up!)²⁰

(20) Prat! (Tittle-tattle!)²¹

(21) Får ej övertäckas. (Do not cover.)²²

²⁰ *Svenska Akademiens grammatik*, Vol. 4 (2010: 975); in Serbian: (*Izvršno je to što*) *on ne odustaje!*

²¹ *Svenska Akademiens grammatik*, Vol. 4 (2010: 977); in Serbian: *Prazne priče!*

²² *Svenska Akademiens grammatik*, Vol. 4 (2010: 966); in Serbian: *Ne prekrivati.*

Here too it is important to understand that expressions of this kind are not translated word for word, because they often contain subtle semantic differences or represent fixed phrases in a language. In one of the referential Swedish grammars, *Svenska Akademiens grammatik, Vol. 4*, some of the examples are presented as particular expressions which merely resemble elliptical constructions but are particularly interesting and challenging for conveying into Serbian for example, as the target language as seen in the examples (19), (20) and (21).

4. CONCLUDING REMARKS

According to Cassirer, stylistic analysis is certainly not something that we learn merely because of itself; it is a tool to better understand life and a way to master it. “As soon as we get to know the technology of how we are affected by manipulation, we are able to expose it and rule with the help of languages (1999: 14)”.²³ Communicative translation – in the sense of the exact contextual meaning of the original which is readily accepted and comprehended by the target language readership (Newmark 1988: 41) – requires not only operational knowledge of the translator or interpreter, but also the ability to convey the most appropriate stylistic features in the process of interlinguistic mediation between two languages.

Analysis of syntactical stylistic elements in Swedish proves the initial assumption that they should also, at a communicative level, be interpreted as cultural linguistic structures: the Swedish formal fundament field indicates not only the level of comprehensibility and readability of a text, but also its functional style (administrative, colloquial, literary etc.). In Swedish, the position of a new information in a sentence can affect its capacity to be transferred into another language. Concurrently, in sentences dominated by coordination some parts of the information can even be omitted in translation. The cases

²³ “Endast om vi känner tekniken för hur vi blir påverkade med språkens hjälp kan vi avslöja manipulationen och göra oss själva dess herrar.”

of duplication, even if they are established more dialectally, represent an important feature because of their unique nature that in many aspects does not correspond with the nature of similar constructions in other languages. Elliptical constructions also often require cultural analysis and adaptation in a target language.

In conclusion, a course in stylistics at advanced levels of language proficiency is of great importance for future translators and interpreters. The course in the Scandinavian (Danish, Norwegian and Swedish) stylistics at University of Belgrade, Faculty of Philology starts as a review of the basics of grammatical and lexical structures and writing rules in general, as well as the review of dictionaries and other reference books. This is followed by an in-depth overview of the Danish/ Norwegian/ Swedish written language and syntactic registers in a translation perspective. Within the framework of descriptive stylistics, the relationships between speech and writing and between free and formal written language and stylistic means are then examined. The course also investigates the development of the standard Scandinavian languages during the 20th century while the discussion also includes the topics of language standardization and democratization in the modern Scandinavian societies.

REFERENCES

- Cassirer, P. (1999). *Stilistik & stilanalys*. Stockholm: Natur och kultur.
- Enqvist, N. E. (1973). *Stilforskning och stilteori*. Lund: CWK Gleerups Bokförlag.
- Holmes, P. and I. Hinchliffe. (2013). *Swedish. A Comprehensive Grammar*. New York: Routledge.
- Kovačević, Z. (2020). Kulturno posredovanje u nastavi prevodenja na primeru švedskih konstrukcija s finalnim udvajanjem. In: *Prevodenje kao veština i kao umetnost: zbornik u čast Branimiru Živojinoviću povodom devedesetogodišnjice rođenja* (J. Kostić-Tomović et al., eds.), Beograd: FOKUS, 243–249.

- Kovačević, Z. (2021). *Skandinavija i mi. Prevodenje kao jezičko, književno i kulturno posredovanje*. Beograd: Udruženje naučnih i stručnih prevodilaca Srbije.
- Lindström, J. (2011). Samtalsgrammatik på dialekt? In: *Interaktionell dialektologi* (G. Bockgård, Nilsson J., eds), Uppsala: Författarna och Institutet för språk och folkminnen, 210–211. https://www.researchgate.net/publication/281291014_Samtalsgrammatik_pa_dialekt [21-11-2021].
- Lagerholm, P. (2008). *Stilistik*. Lund: Studentlitteratur.
- Newmark, P. (1988). *A Textbook of Translation*. New York, London: Prentice Hall.
- Teleman, U., S. Hellberg and E. Andersson. (2010). *Svenska Akademiens grammatik*, Vol. 4. Stockholm: Norstedts.

Зорица Ковачевић

АНАЛИЗА СИНТАКТИЧКОСТИЛСКИХ ЕЛЕМЕНАТА У ШВЕДСКОМ ЈЕЗИКУ

Резиме

Структура реченице је једна од најважнијих компоненти стилске анализе, због тога што се односи пре свега на читљивост и разумљивост текста. Језици са строго устаљеним редом речи, попут шведског и осталих скандинавских језика, испољавају и специфичне стилске особености. На стилску анализу могу утицати бројни аспекти у говорном и писаном шведском језику, нпр. паратакса и хипотакса, топикализација, удвајање, елиптичне реченичне конструкције, као и структура реченице у формалним и неформалним текстовима у комбинацији са додатним синтаксичким карактеристикама. Стилске варијације датог текста такође могу унапредити учење језика и стицање знања о различитим језичким функцијама у специфичним језичким ситуацијама. Циљ овог рада је да се испитају најчешћи примери који се јављају у вежбама синтаксичке стилске анализе у шведском језику на напредном нивоу учења и у превођењу, као и то да се додатно образложи и истакне њихов значај.

Кључне речи: синтакса шведског језика, стандардни језик, превођење, учење језика на напредном нивоу

Vesna Plesničar¹
University of Nova Gorica

A CORPUS INVESTIGATION OF THE ORDERING OF SELECTED ATTRIBUTIVE ADJECTIVES IN SLOVENIAN

Abstract

The paper investigates the question of attributive adjective ordering in Slovenian against the background of the cartographic model of natural language syntax (e.g., Cinque 1994, 2010, Scott 2002, Shlonsky 2004, Laenzlinger 2005, Ramaglia 2014). Using the nearly 1.2-billion-word Gigafida 2.0 corpus, we conducted a study in which we check the frequency of attested orders of selected attributive adjectives belonging to thirteen semantic categories and determine whether we can adequately predict language use if we adopt a cartographic model as a working theoretical framework, specifically the adjective hierarchy proposed in Scott (2002). The results show that the probability of encountering an order that violates the hierarchy is in general extremely small compared to the probability of encountering an order that respects the hierarchy, which indicates that the order of adjectives attested in the Slovenian corpus is by-and-large compatible with the proposal that the order is governed by a hierarchy of adjective projections.

Key words: generative grammar, cartography, attributive adjective string, language use, corpus

¹ The author acknowledges the financial support from the Slovenian Research Agency (research core funding No. J6-7282).

1. INTRODUCTION

It is very likely that in languages with prenominal adjectives the order of adjectives will be as in (1a) and not as in (1b).

- (1) a. small green Chinese vase
b. Chinese green small vase
Sproat and Shih (1991: 565)

What is probably even more interesting or important is that preferences for one order over another apply regardless of whether these are languages with pre- or post-nominal adjectives (e.g., Martin 1969, Hetzron 1978, Dixon 1982, Sproat and Shih 1991, Shlonsky 2004, Scontras and Kachakeche 2020).² Preferences for one order over another are present in Slovenian too. Below, we present two proposals for adjective ordering in Slovenian, made within the traditional grammar framework. One is by Vidovič-Muha (1978) and one by Toporišič (2000). The models differ to some extent in the subclassification of adjectives, i.e., their clustering into the semantic classes. Since these two models make different predictions as to how adjectives are supposed to be ordered in language use, and they also lack a comparative, crosslinguistic aspect, we merely present them here and use them as a basis for the prediction that order restrictions exist in Slovenian too.

According to Vidovič-Muha's (1978) proposal, shown in (2), SUBJECTIVE COMMENT adjectives are placed furthest from the noun, followed by adjectives expressing MEASURE, ATMOSPHERE, COLOR, STATE, and MATERIAL. The adjectives closest to the noun are classifyingly used adjectives, i.e., adjectives denoting KINDS.

- (2) **SUBJECTIVE COMMENT** (*lep* 'beautiful', *grd* 'ugly', *dober* 'good', *slab* 'bad') > **MEASURE(MENT)** (*velik* 'big', *majhen* 'small', *globok* 'deep', *plitev* 'shallow', *dolg* 'long', *kratek* 'short', *težak* 'heavy', *lahek* 'light', *mlad* 'young', *star* 'old')

² These are all orderings with neutral stress and intonation, i.e., the basic orderings, without comma intonation and without special interpretations.

>**ATMOSPHERE** (*jasen* ‘clear’, *meglen* ‘foggy’, *deževen* ‘rainy’) >**COLOR** (*rdeč* ‘red’, *zelen* ‘green’, *bel* ‘white’) >**STATE** (*bolan* ‘ill’, *jezen* ‘angry’, *pečen* ‘baked’, *rezljan* ‘carved’, *zašit* ‘sewed’) >**MATERIAL** (*lesen* ‘wooden’, *železen* ‘iron’, *kovinski* ‘metallic’) >**KIND** (*ladijski* ‘ship’, *otroški* ‘child’) >**N**

According to Toporišič (2000), the hierarchy is as in (3), with the highest position occupied by adjectives for PLEASING, followed by OBJECTIVELY ATTRIBUTED PROPERTY, AGE/PERIOD, COLOR, PARTICIPLE, ORIGIN/NATIONALITY, and with adjectives denoting KINDS as the category closest to the modifying noun.

(3) **PLEASING** (*ljubek* ‘cute’) >**GENERAL ADJECTIVE; OBJECTIVELY ATTRIBUTED PROPERTY** (*divji* ‘wild’, *monden* ‘cosmopolitan’, *pameten* ‘smart’, *lep* ‘beautiful’, *grd* ‘ugly’, *majhen* ‘small’, *težek* ‘heavy’) >**AGE/PERIOD** (*mlad* ‘young’, *star* ‘old’, *nov* ‘new’, *moderen* ‘modern’) >**COLOR** (*zelen* ‘green’, *siv* ‘grey’, *rdeč* ‘red’, *lila* ‘violet’) >**PARTICIPLE** (*razbit* ‘broken’, *boleč* ‘painful’, *zarjavel* ‘rusted’, *zaželen* ‘desired’) >**ORIGIN/NATIONALITY** (*pariški* ‘Parisian’, *kitajski* ‘Chinese’, *gozden* ‘forest’, *cerkven* ‘church’, *državen* ‘state’) >**KIND**(*mali* ‘little’) >**N**

The model that has taken shape within the formal linguistic theory and has been very prominent in the field of adjective order in multiple adjective string research is the cartographic model of natural language syntax. According to this model, crosslinguistic tendencies for relatively rigid adjective ordering can be accounted for if we assume that adjectives are merged in specifiers of various dedicated adjectival functional projections merged above the noun (e.g., Cinque 1994, 2010, Scott 2002, Shlonsky 2004, Laenzlinger 2005, Ramaglia 2014). As a result, adjectives in multiple adjective strings cannot appear in just any of all possible orders but follow each other in a specific sequence dictated by the functional hierarchy. All deviations from the basic and arguably universal pre-nominal order of adjectives are considered a consequence of either adjective phrase

movement, where an adjective is moved for reasons of topic or focal marking, or different structural positions of the merge.

Following Cinque (2010) there are two separate sources for attributive adjectives: direct modification and indirect modification or (reduced) relative clause (RC). These two sources have different interpretative properties and different merge positions; merge positions for indirect (reduced RC) modification, associated with, i.a., intersective, literal interpretation, are above the positions hosting direct modification adjectives, associated with, i.a., non-intersective, possibly idiomatic interpretation.

In the present paper, the focus is on indirect modification adjectives, i.e., adjectives interpreted intersectively. In intersective adjective-noun combinations, the adjectives are interpreted literally and maintain truth-conditional independence, unlike in the case of direct modification, where an adjective-noun combination gives rise to a non-intersective reading; compare the reading in (4a) with the one in (4b).

- (4) Peter je velik jedec.
Peter is big eater.
a. Intersective: Peter is tall and an eater.
b. Non-intersective: Peter eats a lot.

That these are in fact two different structural positions is illustrated in (5), where both indirect and direct modification are instantiated simultaneously by the same adjective *velik* 'big'. The sentence is grammatical only if the adjective that is further away from the noun is interpreted intersectively ('tall and an eater'), and the adjective that is closer to the modifying noun is interpreted subsectively ('someone who eats a lot'). An attempt to interpret the adjective that is further away from the noun subsectively and the one that is closer to the noun intersectively leads to ungrammaticality.

- (5) Peter je velik velik jedec.
Peter is big big eater
'Peter is a tall big eater.'

Since direct modification adjectives are merged in a separate field of the NP complex, in the part of the hierarchy that is closer to the noun than adjectives deriving from a relative clause, they are not part of the sequence that is the subject of our research. We can infer from the example in (6) that both parts of the hierarchy are arranged in a specific order, moreover, we can assume that the order is the same in both parts.

- (6) a. Peter je velik star prijatelj.
Intersective: 'Peter is a friend who is tall and old.'
Non-intersective: 'Peter is a great long-time friend.'
- b. Peter je star velik prijatelj.
Intersective and non-intersective: 'Peter is an old great friend.'

While (6a) is ambiguous between two interpretations, namely an intersective one 'a friend who is tall and old' and a non-intersective one 'a great long-time friend', (6b) can only have one, the interpretation in which the first adjective is interpreted intersectively and the second non-intersectively, namely 'a great friend who is old'.

The most elaborated version of adjective hierarchy proposed within the cartographic model of natural language syntax, which can explain the order of both direct and indirect modifiers, is the Universal hierarchy of adjectival functional projections proposed in Scott (2002).³ According to Scott (2002: 114), the hierarchy of thirteen

³ Although Scott (2002) does not at any point explicitly define that the hierarchy he proposes is about the ordering of indirect modifiers, this can be inferred, first, from the examples he excludes from the discussion, see (i) and (ii), and second, from all the data he provides in support of the proposed Universal hierarchy of adjective phrase-related functional projections.

(i) senatorial industrial investigation

(ii) industrial senatorial investigation

As far as (i) and (ii) are concerned, it is claimed that these are cases with no basic ordering, since ordering is chosen upon the intended interpretation; (i) refers to the senate's investigation of the industry, (ii) to the industry's investigation of the senate (Scott 2002: 92). As for Scott's material in support of the universal

semantic categories of adjectives relevant for our corpus research is as illustrated in (7).

- (7) SIZE>LENGTH>HEIGHT>SPEED>DEPTH>WIDTH>WEIGHT>
TEMPERATURE>WETNESS>AGE>SHAPE> COLOR>MATERIAL> NP

As a currently highly prominent model in the domain of complex NP research, the cartographic model has attracted a lot of criticism, ranging from the problem of innateness, origin, and functional projection proliferation to the problem of rigidity (e.g., Svenonius 2008, Truswell 2009, Scontras et al. 2017, Leivada and Westergaard 2019, Larson 2021). Given the conclusions based on large databases (cases of actual use of multiple adjective strings), the concerns of corpus studies focusing on the rigidity problem, or the so-called empirical undergeneration problem, seem particularly relevant (Wulff 2003, Truswell 2009, Kotowski and Hartl 2019, Trotzke and Wittenberg 2019). Since a rigid order is assumed as a direct output of adjectival functional hierarchy within the cartographic model, the data that do not attest to restrictions in ordering among the proposed semantic categories could be problematic for this strictly syntactic approach.

2. MOTIVATION AND RESEARCH QUESTION

There are several different motives for testing the predictions of Scott's (2002) cartographic model of adjective phrase-related functional projections. First, while it has been claimed that order restrictions apply to Slovenian complex NPs (Vidovič-Muha 1978, Toporišič 2000), there is a lack of experimental studies in the domain of adjective ordering constraints, especially studies that would verify the predictions of the cartographic model (attempts – pilot studies Plesničar 2017; forced-choice experiment, Plesničar 2018; self-paced reading experiment). Second, there is the advantage of corpus studies (compared to experiments of type e.g., forced-choice), namely, the

hierarchy, all adjectives in his examples have the properties of indirect modifiers, as defined in Cinque (2010).

feasibility of comparing a large number of adjectives within a wide range of semantic categories. And finally, although much attention has recently been paid to testing adjective order constraints with the corpus approach, no corpus study has really directly examined the predictions of the cartographic model, even though the cartographic model is currently dominant in the domain of adjective order research, and Scott's (2002) proposal, moreover, is one of the most fine-grained of all offered so far. Generally, corpus studies test the predictions of Scott's cartographic model either indirectly, against other semantic classifications, or within a limited set of semantic categories and nouns (Wulff 2003, Truswell 2009, Scontras et al. 2017, Kotowski and Hartl 2019, Trotzke and Wittenberg 2019).

In addition to the fact that no corpus study goes into a detailed review of the cartographic model offered by Scott (2002), there is another significant problem with corpus studies on adjective order, namely the problem of conflicting outcomes.

On the one hand, it is argued that a bipartite division between subsective and intersective adjectives is sufficient to explain the data since there is a clear tendency for subsective adjectives to dominate intersective adjectives and no ordering preferences were found among subsective and intersective multiple adjective strings (Truswell 2009).⁴ On the other hand, interestingly, a very influential recent study, Scontras et al. (2017), which examines the ordering of semantic categories that are a subset of the categories proposed by Scott (2002), using corpus research merely to validate a behavioral

⁴ In Truswell (2009) the subsective interpretation is claimed to be equivalent „to the conjunction of the property denoted by *N* with the property denoted by *Adj* related to a comparison class largely determined by *N*” (Truswell 2009: 526). According to the above criteria considered by Truswell (2009), adjectives for AGE or even adjectives for COLOR can be, just like adjectives for SIZE, misconceived as subsective. Consider, for example, the adjective *turquoise*, which is very likely to be defined as something between blue and green when compared to prototypical blue and as blue when compared to red. We will take the position that the adjectives discussed in the present study are intersective, and their interpretation depends on their inherent vagueness and context-sensitivity (see e.g., Klein 1982, Partee 1995, Morzycki 2015).

measure of ordering preferences attained through a direct measure of subjectivity, reaches the same conclusion as Scott (2002). According to Scontras et al. (2017) adjectives for SIZE and LENGTH, grouped into the class of DIMENSION, are located furthest from the noun, followed by adjectives for AGE, SHAPE, COLOR, and MATERIAL, as shown in (8) below.

- (8) DIMENSION (SIZE, LENGTH) >AGE>SHAPE>COLOR>
MATERIAL adapted by Scontras et al. (2017)

To contribute to this ongoing debate, we conducted a corpus study in which we sought to check the frequency of attested orders of selected attributive adjectives of thirteen semantic categories and determine whether we can adequately predict language use if we adopt a cartographic model as a working theoretical framework, specifically the adjective hierarchy proposed in Scott (2002). We assume that if default adjective ordering in an NP is a result of universal hierarchically ordered NP structure, then the principles that govern the process of building complex NPs should be observed as differences in the number of occurrences between what would then boil down to syntactically adequate and syntactically anomalous complex NP structures.

3. MATERIALS AND METHODS

To be able to assess as accurately as possible whether the prediction of the model proposed by Scott (2002) holds for Slovenian, we decided to examine a wide range of semantic categories from Scott's hierarchy and investigate how they are ordered with respect to each other in the largest corpus of written Slovenian, i.e. the nearly 1.2-billion-word reference corpus *Gigafida 2.0*⁵ (deduplicated). In (9) we list all thirteen semantic categories used in our corpus study, their corresponding

⁵ Available at: <http://www.gigafida.net>.

adjectives, and their relative frequencies⁶; although irrelevant here, the semantic categories are listed in the order from Scott (2002).

- (9) **SIZE:** *velik* 'big' (2,667,041), *majhen* 'small' (586,904), *ogromen* 'huge' (54,879), *droben* 'tiny' (53,775), **LENGTH:** *dolg* 'long' (502,666), *kratek* 'short' (384,723), **HEIGHT:** *visok* 'high' (801,128), *nizek* 'low' (275,075), **SPEED:** *hiter* 'fast' (251,343), *počasen* 'slow' (36,521), **DEPTH:** *globok* 'deep' (67,679), *plitek* 'shallow' (6,254), *plitek* 'shallow' (2,248), **WIDTH:** *širok* 'wide' (191,419), *ozek* 'narrow' (68,106), **WEIGHT:** *težek* 'heavy' (264,669), *lahek* 'light' (127,823), **TEMPERATURE:** *vroč* 'hot' (99,915), *mrzel* 'cold' (28,777), *hladen* 'cold' (75,857), **WETNESS:** *suh* 'dry' (74,922), *moker* 'wet' (21,017), **AGE:** *star* 'old' (80,209), *mlad* 'young' (711,377), **SHAPE:** *kvadraten* 'square' (67,222), *okrogel* 'round' (61,360), **COLOR:** *rumen* 'yellow' (73,551), *moder* 'blue' (100,430), **MATERIAL:** *lesen* 'wooden' (70,476), *plastičen* 'plastic' (36,920), *kovinski* 'metallic' (31,859)

The corpus was searched using the No Sketch Engine browsing tool. Not to miss any of the potentially relevant cases, we opted for the lemma search, as Slovenian lemmas can have many different word forms. Although we anticipated that this would produce some extra noise as the lemma search includes all forms of the word in the result, including comparative and superlative adjective forms, we deemed this to be outweighed by the greater number of hits and thus by the greater number of potentially relevant cases. (10) provides a template of the lemma query used.

- (10) [tag="P.*" & lemma="..."] [tag="P.*" & lemma="..."]

⁶ Frequency and length have been shown in previous studies to be factors that can influence the adjective order in multiple adjective strings (e.g., Wulff 2003, Scontras et al. 2017, Kotowski and Hartl 2019, Trotzke and Wittenberg 2019). However, it is not entirely clear why frequency and length affect adjective order in some cases more than in others. For the present study, we tried to control for frequency and length, although this proved to be feasible only to a certain extent, both within and between semantic categories.

We tested all possible combinations of two adjectives from the list in a total of 882 queries. The total number of concordances obtained was 14.916. The obtained data were then exported to excel files, where we first edited and then cleaned them, i.e., marked the excluded data with the appropriate exclusion criterion tag. We used pre-prepared criteria for cleaning the data, all of them are presented in paragraph 3.1. After cleaning, we had 9.156 concordances left for analysis. We should note at this point that a decision to exclude a particular example from the analysis was inevitably subject to a certain degree of subjectiveness, more prominent when deciding, for example, whether an adjective is used as a classifying adjective or in a collocation, and less, or not at all in the case of incorrect tagging or when an adjective was used in a conjunction with a further adjective (e.g. *red and long*). Since some combinations yielded a very large number of hits, such as the combination *velikokrogel* 'large round' (617 hits) or *dolglesen* 'long wooden' (300 hits), and since reviewing so many hits would be too time-consuming, we set an upper limit on the number of cases examined for each combination: when the number of concordances for the combination of two adjectives exceeds 100, the number of total relevant examples was calculated from the percentage of 100 examples examined.

3.1. CRITERIA FOR DATA CLEANING

The exclusion criteria range from cases of classifying use, collocations, corpus tag errors, adjective modifications, comparative and superlative forms, to conjunction and proper name use. An occurrence was excluded from further analysis as soon as it met one exclusion criterion; but the vast majority of exclusions actually met two or more exclusion criteria. One such example is shown in section 3.1.5.

3.1.1. *Classifying use and collocation*

Our data sample contained many adjective-noun combinations with classifyingly used adjectives (e.g., *kratke hlače* 'shorts', *mlad krompir* 'new potatoes') as well as collocations (e.g., *suharoba* 'woodenware',

okrogla obletnica ‘round number anniversary’). As a classifying use, we label only clear-cut examples of the type *mlad krompir* ‘new potatoes’; everything else was labeled as collocation, although it may be, by definition (e.g., Gantar 2007), closer to a set phrase (e.g., *mrzli stric* ‘first uncle’, *okrogla miza* ‘panel discussion’) and could as such be considered as an indirect modifier, in the sense of Cinque (2010). Examples of collocation and classifying use from the corpus are given in (11) and (12).

(11) Collocation:

| | | |
|---------------------------|----------|---------------|
| visokih | okroglih | obletnicah |
| high | round | anniversaries |
| ‘big round anniversaries’ | | |

(12) Classifying use:

| | | |
|----------------------|-------|-----------|
| majhne | mlade | krompirje |
| small | young | potatoes |
| ‘small new potatoes’ | | |

3.1.2. Tag error

We encountered two types of tagging errors. The first group includes incorrect tags, which could be due to an error in the original text, as in example (13), or an error that could be the result of relatively uncommon part-of-speech sequences, such as the adjective-noun-noun sequence in (14a) or the adjective-modal-noun in (14b). In (13) the word *stari* is tagged as an adjective, but it is clear from the context and from the grammatical agreement that *stari* is actually a misspelling of the noun *stvari* ‘things’.

(13) Incorrect tag, due to an error (e.g., typo) in the original text:

| | |
|---------------|-------|
| drobne | stari |
| tiny | old |
| ‘tiny things’ | |

In (14a) the word *dolg* preceded by the adjective *velik* ‘big’ and followed by the noun *razlog* ‘reason’ is incorrectly tagged as an adjective (*dolg* ‘long’), when it should be tagged as a noun (*dolg* ‘debt’).

The situation is comparable in (14b), where the context reveals that *lahko* is not used in its function of an adjective (*lahek* 'light') but in its function of a modal expression (*lahko* 'can').

(14) Wrong tag, due to uncommon sequence:

- a. velik dolg razlog
big debt reason
'big debt is the reason'
- b. veliko lahko veselje
great can joy
'the joy can be great'

The second group consists of tags that are not specific enough. In (15) the adjectives *velikimi* 'large' in *visoka* 'high' are not part of one and the same NP complex, but they modify two different noun phrases. *Velikimi* 'large' modifies an unpronounced but understood noun *državami* 'countries', *visoka* 'high' modifies noun *cena* 'price'.

(15) Tagging not specific enough:

- cena druženja z velikimi visoka
price socializing with big high
'the price of aligning oneself with big (i.e. powerful) countries
is high'

3.1.3. Adjective modification and adjectives in comparative and superlative form

We excluded from further consideration cases in which the adjective was modified, e.g., by an adverbial or quantificational phrase, or when one of the adjectives appears in the comparative or superlative form. It has been noted for Romance that in the superlative form, adjectives can appear higher than normal (e.g., Cinque 2010, Loccioni 2020)⁷. Extrapolating from these findings, we assume that Slovenian

⁷ Although Cinque (2010) and Loccioni (2020) highlight only the difference between basic adjective forms and superlatives, we assume that this also applies to comparatives, as comparatives have also been argued to differ semantically from the positive/base adjective forms. According to Klein (1982), comparatives

adjectives in superlative and comparative forms are also in different structural positions than Slovenian adjectives in their positive forms. We did not, however, exclude cases where both adjectives appear in the superlative or both appear in the comparative form, assuming that in these cases, the relative ordering should be the same as between positive-form adjectives, just with different points of origin.

Examples of adjectival modification are given in (16a) and (16b). In (16a) the adjective *velika* 'big' is modified by the adverbial phrase *zelo* 'very', and in (16b) the adjective *široke* 'wide' is modified by the quantificational phrase *8 cm* '8 centimeters'. Examples of adjective-adjective-noun sequences that were excluded due to one of the adjectives being in the comparative or superlative form are in (17). In (17a) the first adjective in the sequence is in the comparative and the second in the base form, and in (17b) the first adjective is in the comparative and the second in the superlative.

(16) Modification:

- | | | | | | |
|----|-------------------------------------|---------------|---------------|------------------|-----------------|
| a. | <i>zelo</i> | <i>velika</i> | <i>visoka</i> | <i>pozitivna</i> | <i>številka</i> |
| | very | big | high | positive | number |
| | 'very large high positive number' | | | | |
| b. | <i>8 cm</i> | | <i>široke</i> | <i>rumene</i> | <i>cvetove</i> |
| | 8 centimeters | | wide | yellow | flowers |
| | '8 centimeters wide yellow flowers' | | | | |

(17) Adjective in comparative or in superlative form:

- | | | | |
|----|----------------------------|-----------------|----------------|
| a. | <i>večja</i> | <i>lahka</i> | <i>kolesa</i> |
| | bigger | light | wheels |
| | 'larger light wheels' | | |
| b. | <i>višjo</i> | <i>največjo</i> | <i>hitrost</i> |
| | higher | biggest | speed |
| | 'higher top/maximum speed' | | |

are complex predicates consisting of an adjective and a degree variable, and as such they are not vague, unlike their base adjective counterparts.

3.1.4. Conjunction and proper name

The use of adjectives in conjunction results in special comma intonation, which could affect anotherwise fixed order of adjectives. We therefore exclude cases of enumeration, and cases where one of two successively used adjectives is part of conjunction. (18a) presents an example of enumeration, and in (18b) the adjective *modrimi* ‘blue’, i.e. the second adjective in the adjective string *širokimi modrimi* ‘wide blue’, is a part of the conjunction phrase *modrimi in belimi* ‘blue and white’.

- (18) Enumeration and adjective as part of a conjunction phrase:
- a. star mlad visok nizek hiter počasen drag
old young tall low fast slow expensive
poceni težek lahek
cheap heavy light
 - b. širokimi modrimi in belimi črtami
wide blue and white stripes
‘wide blue and white stripes’

We also excluded cases in which one of the adjectives is a proper name or a constituent part of a proper name. Two such examples are presented in (19a) and (19b), respectively. In (19a) the whole adjective string is a proper name, and in (19b) only the second adjective in the adjective-adjective string is a part of the proper name.

- (19) Proper name or a constituent part of a proper name:
- a. Velikega Globokega
big deep
‘Veliko Globoko’ (the name of a settlement in south-eastern Slovenia)
 - b. široke Velike Pode
broad large plateau
‘broad Large Plateau (Veliki Podi is the name of a plateau cunder Mount Grintovec)’

3.1.5. Multiple exclusion criteria

The string *višji Velika* ‘higher big’ in (20) met four exclusion criteria, so it had to be excluded on multiple counts.

- (20) Modification, conjunction, an adjective in comparative or in superlative form, a constituent part of a proper name:
nekoliko višji Velika in Zadnja Mojstrovka
somewhat higher big and back Mojstrovka
‘slightly higher Velika (Mojstrovka) and Zadnja Mojstrovka’

The identified exclusion criteria are as follows, first, the adjective *višji* ‘higher’ is in comparative adjective form, second, it is modified by the adverb *nekoliko* ‘somewhat’, third, the adjective *Velika* ‘big’ is a constituent part of a proper name (*Velika Mojstrovka* peak), and fourth, it is also part of a conjunction phrase, consisting of *Velika* ‘big’ and *Zadnja* ‘back’.

4. RESULTS AND DISCUSSION

We analyzed the cleaned data set, 9.156 examples in total. The first step was to compare the number of occurrences with expected and unexpected adjective orders in general. In Figure 1 visual presentation of the compared samples is given.

The Sapiro-Wilk test of normality shows that neither the sample of occurrences with the expected adjective-adjective order ($W = 0.30185$, $p < 0.001$) nor the sample of occurrences with the unexpected adjective-adjective order ($W = 0.44809$, $p < 0.001$) were normally distributed. Based on this outcome a non-parametric test for independent samples, the Mann-Whitney U test, was used to test the null hypothesis that, for randomly selected frequency values from the expected and unexpected order samples, the probability of the frequency of the expected order being greater than that of the unexpected one is equal to the probability of the frequency of the unexpected being greater than that of the expected one. The number of occurrences with the expected order (Min = 0.00, Max = 670.14,

Median = 2.00, Mean = 19.28) is higher than that with the unexpected order (Min = 0.00, Max = 31.00, Median = 0.00, Mean = 1.484). The test indicates that this difference is statistically significant, $U(N_{\text{expected}} = 411, N_{\text{unexpected}} = 441) = 138255, p < 0.001$.

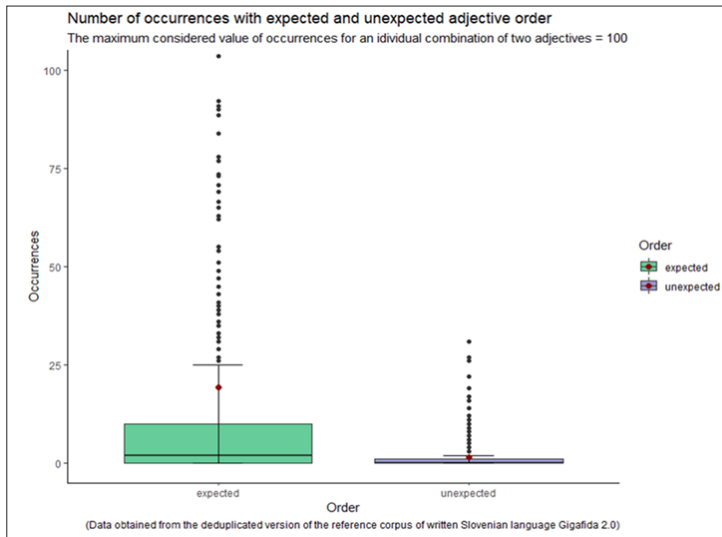


Figure 1: Expected vs. unexpected adjective order

Low medians and means indicate that in the corpus there is a large proportion of adjective-adjective combinations for which there is no data. This result is not surprising, as combinations of two adjectives are not particularly common constructions in general, and among them, there are some that are less common or less meaningful than others, compare for example the HEIGHT-DEPTH combination, for which it is not entirely clear what real-world context it could be used to describe, with the combination SIZE-COLOR, for which applicable real-world contexts are very easy to imagine. As shown in Figure 2, the proportion of unattested adjective order is higher in the case of the unexpected order, specifically, it is twice as high when compared to the proportion in the case of the expected order. This outcome is expected assuming the adequacy of Scott's (2002) model, according to which, ideally (if speakers don't form ungrammatical constructions

and if our data cleaning successfully excluded all independently explainable reversals), we should get no concordances in the case of the unexpected adjective order.

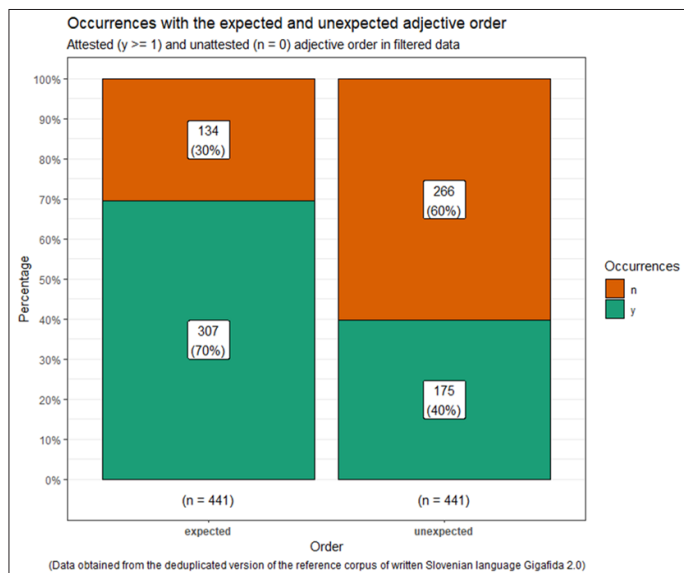


Figure 2: Attested vs. unattested adjective order with the expected and unexpected adjective order

In Figure 3 an overview of the data by individual category is given. In addition to a very clear distinction between expected and unexpected orders across all thirteen semantic categories, it also shows great variability between semantic categories according to the number of examples in the case of expected order. While the number of relevant examples is greater in the case of adjectives from the top and the bottom of the hierarchy (i.e. left and right edge in Figure 3) and there are also fewer deviations from the expected ordering in these cases, the picture is not as clear for the middle part of the hierarchy, where the number of relevant examples is considerably smaller.

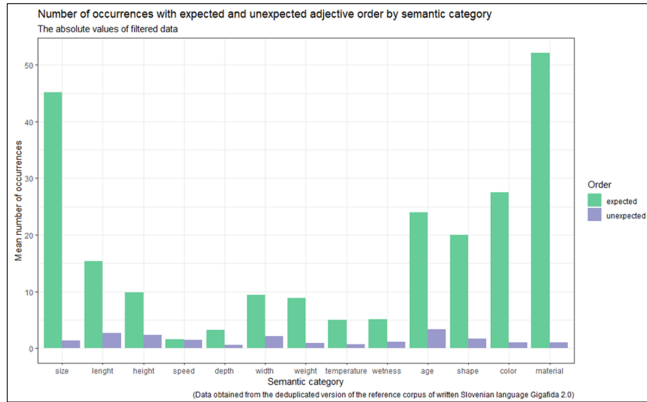


Figure 3: Expected vs. unexpected adjective order by semantic categories

A detailed review of the data reveals that some categories are more „obedient“ than others. As can be seen from the percentage values of the data presented in Table 1, in which the rows refer to the first position and the columns to the second position in the adjective-adjective-noun complex. Semantic categories AGE, WEIGHT, WIDTH, and SPEED depart the most from the expected behavior. (The greener the shade, the higher the percentage; the redder the shade, the lower the percentage.)

Table 1: The percentage values of occurrences per combination of two semantic categories

| | size | length | height | speed | depth | width | weight | temperature | wetness | age | shape | color | material |
|-------------|------|--------|--------|-------|-------|-------|--------|-------------|---------|-----|-------|-------|----------|
| size | N/A | 83% | 93% | 46% | 96% | 96% | 91% | 88% | 96% | 87% | 98% | 100% | 99% |
| length | 17% | N/A | 74% | 53% | 100% | 79% | 58% | 92% | 91% | 54% | 89% | 91% | 99% |
| height | 7% | 26% | N/A | N/A | 67% | 59% | 36% | 75% | 86% | 54% | 79% | 85% | 99% |
| speed | 54% | 48% | 100% | N/A | 60% | N/A | 86% | 100% | 100% | 53% | 100% | 100% | 100% |
| depth | 4% | N/A | 33% | 40% | N/A | 32% | 50% | 85% | 67% | N/A | 75% | 95% | 95% |
| width | 4% | 21% | 41% | 100% | 68% | N/A | N/A | N/A | 83% | 41% | 100% | 94% | 99% |
| weight | 9% | 42% | 64% | 14% | 50% | 100% | N/A | 60% | 63% | 55% | 100% | 100% | 100% |
| temperature | 13% | 8% | 25% | N/A | 15% | 100% | 40% | N/A | 74% | 83% | N/A | 98% | 100% |
| wetness | 4% | 9% | 14% | N/A | 33% | 17% | 38% | 26% | N/A | 33% | 80% | 100% | 100% |
| age | 13% | 46% | 46% | 47% | 100% | 59% | 45% | 17% | 67% | N/A | 86% | 95% | 100% |
| shape | 2% | 11% | 21% | N/A | 25% | N/A | N/A | 100% | 20% | 14% | N/A | 68% | 78% |
| color | 0% | 9% | 15% | N/A | 5% | 6% | N/A | 2% | N/A | 5% | 33% | N/A | 87% |
| material | 1% | 1% | 1% | N/A | 5% | 1% | N/A | N/A | N/A | 0% | 22% | 13% | N/A |

Except for the semantic category AGE, these are at the same time the categories for which the corpus exhibits only a few relevant cases (for example, for the category WETNESS there are only 5 occurrences in combination with the category WIDTH, 2 in combination with the category DEPTH and only 1 in combination with the category SPEED). If we look at the absolute values of the data presented in Table 2 (in which colorshading reflects numbers of occurrences, not numbers of occurrences correctly predicted), this is a broader problem present throughout the middle part of the hierarchy, spanning roughly the categories from SPEED to WETNESS.

Table 2: The number of occurrences per combination of two semantic categories

| | size | length | height | speed | depth | width | weight | temperature | wetness | age | shape | color | material |
|-------------|------|--------|--------|-------|-------|-------|--------|-------------|---------|-----|-------|-------|----------|
| size | N/A | 35 | 42 | 13 | 106 | 73 | 52 | 42 | 68 | 456 | 832 | 1071 | 2087 |
| length | 7 | N/A | 26 | 42 | 8 | 155 | 21 | 122 | 42 | 13 | 8 | 105 | 315 |
| height | 3 | 9 | N/A | N/A | 2 | 41 | 4 | 3 | 32 | 49 | 44 | 28 | 302 |
| speed | 15 | 38 | 17 | N/A | 3 | N/A | 6 | 4 | 1 | 9 | 1 | 5 | 5 |
| depth | 4 | N/A | 1 | 2 | N/A | 9 | 2 | 11 | 2 | N/A | 33 | 60 | 38 |
| width | 3 | 42 | 29 | 3 | 19 | N/A | N/A | N/A | 5 | 14 | 39 | 51 | 156 |
| weight | 5 | 15 | 7 | 1 | 2 | 3 | N/A | 3 | 10 | 16 | 6 | 13 | 379 |
| temperature | 6 | 10 | 1 | N/A | 2 | 1 | 2 | N/A | 86 | 19 | N/A | 59 | 73 |
| wetness | 3 | 4 | 5 | N/A | 1 | 1 | 6 | 31 | N/A | 6 | 4 | 5 | 33 |
| age | 70,5 | 11 | 42 | 8 | 3 | 20 | 13 | 4 | 12 | N/A | 18 | 60 | 733 |
| shape | 13 | 1 | 12 | N/A | 11 | N/A | N/A | 1 | 1 | 3 | N/A | 27 | 146 |
| color | 4 | 11 | 5 | N/A | 3 | 3 | N/A | 1 | N/A | 3 | 13 | N/A | 113 |
| material | 15 | 4 | 4 | N/A | 2 | 1 | N/A | N/A | N/A | 3 | 42 | 17 | N/A |

By and large, the results of our corpus investigation thus align with the prediction that the order of adjectives is governed by the hierarchy of adjective phrase-related functional projections, since the overall probability that we will encounter an order that violates the hierarchy, compared to the probability that we will encounter an order that respects the hierarchy, is extremely small. As expected, the proportion of combinations with no occurrences or a lower number of occurrences is higher in the case of counter-hierarchy adjective orders. If we consider the percentage values from Table 1 above, the categories that are almost invariably ordered according to expectations are SIZE, LENGTH, HEIGHT, DEPTH, TEMPERATURE,

WETNESS, SHAPE, COLOR, and MATERIAL. Due to the low number of occurrences for combinations of adjectives from the middle part of the hierarchy – SPEED, DEPTH, WIDTH, WEIGHT, TEMPERATURE, WETNESS – firm conclusions about the order of these semantic categories are harder to draw.

CONCLUSION

The results of this study provide partial evidence in support of the claim that the order of adjectives is governed by the hierarchy of adjective phrase-related functional projections (Cinque 1994, 2010, Scott 2002, Shlonsky 2004, Laenzlinger 2005, Ramaglia 2014). Analyzing adjective-adjective combinations of thirteen semantic categories in the Gigafida 2.0 corpus of written Slovenian we can predict with great certainty the relative orders in the case of six semantic categories, namely SIZE, LENGTH, HEIGHT, SHAPE, COLOR, MATERIAL. The semantic categories DEPTH, TEMPERATURE, and WETNESS could potentially also be included in this list; however, the scarcity of relevant examples proves to be a problem in this case. The orderings are inconsistent with the predictions of the adjectival hierarchy in the case of four semantic categories, namely SPEED, WEIGHT, AGE, and WIDTH. Finally, if we look critically at the database used, we see that for certain aspects of the adjective-adjective-noun complex, the Gigafida 2.0 corpus proves insufficient despite its relatively large size of nearly 1.2 billion words, as the lack of data for certain domains of the proposed adjectival hierarchy complex in the corpus simply makes it impossible to evaluate Scott's (2002) hierarchy as a whole. To fill the gap, future research will need to find alternative ways to zoom in on the categories of the central part of the hierarchy.

REFERENCES

- Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In: *Paths towards Universal Grammar: Studies in Honor of R. S. Kayne* (G. Cinque et al., eds.), Washington, DC: Georgetown UP, 85–110.
- Cinque, G. (2010). *The syntax of adjectives. A comparative study*. MIT Press.
- Dixon, R. M. W. (1982). *Where have all the adjectives gone? and other essays in semantics and syntax*. Berlin, Germany: Mouton.
- Gantar, P. (2007). *Stalnebesednezveze v slovenščini: korpusnoprístup*. Ljubljana: Založba ZRC, ZRC SAZU.
- Hetzron, R. (1978). On the relative order of adjectives. In: *Language Universals* (H. Seiler ed.), Tübingen: Narr, 165–184.
- Klein, E. (1982). The interpretation of adjectival comparatives. *Journal of Linguistics* 18: 1, 113–136.
- Kotowski, S. and H. Hartl. (2019). How real are adjective order constraints? Multiple prenominal adjectives at grammatical interfaces. *Linguistics* 57. 395–427.
- Laenzlinger, C. (2005). French adjective ordering: perspective on DP-internal movement types. *Lingua* 115, 645–689.
- Larson, R. K. (2021). Rethinking cartography. *Language* 97:2, 245–268.
- Leivada, E. and M. Westergaard. (2019). Universal linguistic hierarchies are not innately wired. Evidence from multiple adjectives. *PeerJ* 7: e7438.
- Loccioni, N. (2020). The syntax of superlative phrases in Romance. In: *Romance Languages and Linguistic Theory 16: Selected papers from the 47th Linguistic Symposium on Romance languages (LSRL)* (I. Vogel, ed.), Newark, Delaware, 172–186.
- Martin, J. E. (1969). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior* 8, 697–704.
- Morzycki, M. (2015). *Modification. Key Topics in Semantics and Pragmatics*. Cambridge University Press.
- Partee, B. H. (1995). Lexical semantics and compositionality. In *Language: An invitation to cognitive science*, (Gleitman, L. R. and M. Liberman eds.), The MIT Press, 311–360.
- Plesničar, V. (2017). Adjective order restrictions of attributive adjectives in Slovenian. In *12th Slavic Linguistic Society Annual Meeting: A book of abstracts* (Repaňsek, L. and M. Šekli eds), Ljubljana: Založba ZRC, ZRC SAZU, 159–160.

- Plesničar, V. (2018). Order Restrictions of Attributive Adjectives in Slovenian: A Self-Paced Reading Study. In *Psycholinguistics in Flanders: A book of Abstracts*, Ghent: Ghent University, 58.
- Ramaglia, F. (2014). The syntax and semantics of attributive adjectives in Romance. *Lingue e Linguaggio* 2, 131–158.
- Scontras et al. (2017). Subjectivity predicts adjective ordering preferences. *Open Mind: Discoveries in Cognitive Science* 1: 1, 53–65.
- Scontras, G. and Z. Kachakeche. (2020). Adjective ordering in Arabic: Post-nominal structure and subjectivity-based preferences. *Proceedings of the Linguistic Society of America* 5: 1, 419–430.
- Scott, G.-J. (2002). Stacked adjectival modification and the structure of nominal phrases. In: *The cartography of syntactic structures* 1, (G. Cinque ed.), 91–120.
- Shlonsky, U. (2004). The forms of Semitic noun phrases. *Lingua* 144: 12, 1465–1526.
- Sproat, R. and C. Shih. (1991). The cross-linguistic distribution of adjective ordering restrictions. In: *Interdisciplinary approaches to language. Essays in honor of S. Y. Kuroda* (C. Georgopolous and R. Ishihara eds.), Dordrecht: Kluwer Academic Publishers, 565–593.
- Svenonius, P. (2008). The position of adjectives and other phrasal modifiers in the decomposition of DP. In: *Adjectives and adverbs: Syntax, Semantics and Discourse* (L. McNally and C. Kennedy eds.), New York: Oxford University Press, 16–42.
- Toporišič, J. (2000). *Slovenskaslovnica* [Slovenian Grammar]. 4th edition. Maribor: Obzorja.
- Trotzke, A. and E. Wittenberg. (2019). Long-standing issues in adjective order and corpus evidence for a multifactorial approach. *Linguistics* 57. 273–282.
- Truswell, R. (2009). Attributive adjectives and nominal templates. *Linguistic Inquiry* 40: 3, 525–533.
- Vidovič-Muha, A. (1978). Merila pomenske delitvene za imenske pridevniške besede [...]. *Slavistična revija* 26: 3, 253–276.
- Wulff, S. (2003). A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8. 245–282.

Vesna Plesničar

A CORPUS INVESTIGATION OF THE ORDERING OF SELECTED
ATTRIBUTIVE ADJECTIVES IN SLOVENIAN

Summary

The paper presents the results of corpus study conducted to test the predictions of the most fine-grained version of the, arguably, universal hierarchy of adjective phrase-related functional projections (Scott 2002), which arose within the cartographic framework, currently dominant in the field of adjective order restriction research (e.g., Cinque 1994, 2010, Scott 2002, Shlonsky 2004, Laenzlinger 2005, Ramaglia 2014). Although often criticized for being too rigid and detailed, Scott's model has never been subjected to a detailed examination, i.e., the validity of the predictions this model makes about language use has never been tested in any detail (Wulff 2003, Truswell 2009, Scontras et al. 2017, Kotowski and Hartl 2019, Trotzke and Wittenberg 2019). The present study thus makes an important contribution to the ongoing debate, particularly in this direction. Using the nearly 1.2-billion-word Gigafida corpus we check the frequency of attested orders of selected attributive adjectives of thirteen semantic categories and the results indicate that the order of adjectives attested in our Slovenian corpus is overall compatible with the proposal that the order is governed by a hierarchy of adjective projections. There are some obscurities in the case of individual categories, which we believe are associated with the low number of examples obtained, a problem that should be addressed in future research.

Key words: generative grammar, cartography, attributive adjective string, language use, corpus

Tijana Vesić Pavlović*

University of Belgrade, Faculty of Mechanical Engineering

Milena Jakić Šimšić

Institute for the Serbian Language of SASA, Belgrade

THE EFFECT OF CONTEXT ON THE (A)SYMMETRY OF SERBIAN ADJECTIVE ANTONYMS

Abstract

The paper deals with the symmetry of antonymic relationships of 22 polysemous Serbian adjectives, i.e. 11 posited antonym pairs, when their specific senses (primary, secondary concrete, secondary abstract) are activated within the context of a sentence or a phrase. We use data from two empirical studies conducted with participants who had the task to replace the adjective in a sentence or a phrase with its antonym. It is shown that, in both tasks, the existence and level of symmetry in terms of mutual elicitation and strength of associative relationship depends on the context in which the adjective is activated.

Key words: antonymy, adjectives, Serbian language, symmetry, sentence context, phrase context

* tvesic@mas.bg.ac.rs
milena.jakic@isj.sanu.ac.rs

1. INTRODUCTION

Although antonymy is a well-investigated linguistic phenomenon, various debates are still raised as to the nature of the antonymic relationship and the features which a lexeme should possess to be subsumed under the term antonym. Two major approaches to defining antonymy have been outlined in the pertinent literature. One states that antonymy should be treated as a lexical-categorical relation, while the other, the cognitive prototype approach, treats it as a conceptual relation (Storjohann 2016). The latter approach emphasises the role of context, arguing that antonymy is a category which exhibits prototypicality effects, involving a continuum with good and less good representatives (e.g. Kostić 2016, Paradis et al. 2009).

The behaviour of antonymous pairs across contexts has been discussed in previous studies. Namely, some antonym pairs tend to be more strongly connected and hence are regarded as better opposites than others (Storjohann 2016). An important issue relates to the existence of polysemy and the possibility for an antonym relation to hold for all senses of antonyms. Rasulić (2016: 178) underlines that “antonymy relates particular, not necessarily all senses of one lexeme to another”, further drawing attention to the fact that, frequently, there are “asymmetries in the extended senses of the two members of an antonym pair”, since one member of the pair commonly has “richer semantic extensions than the other” (Rasulić 2016: 180).

In this paper, we deal with the symmetry of antonymic relationships of polysemous Serbian adjectives when they are used in different contexts, that is, when their specific senses are activated in an empirical task with sentences and phrases as stimuli. We focus on 11 posited antonym pairs from the perspective of their lexicographically defined primary meaning, with the aim of exploring whether the posited opposition is stable across different contexts, that is, the three selected adjective senses and in two different types of tasks.

The paper is organised as follows. In the second section, we dwell on the findings of previous studies related to antonym symmetry and dependence on context. Then we proceed with specifying the aim

of the paper in the third section and describing the materials and methods used in the study in the fourth section. The fifth presents the findings of the research, while the final section elaborates on the results and offers concluding remarks.

2. THEORETICAL CONSIDERATIONS

J. Deese (1964) was the first to argue and empirically demonstrate that mutual elicitation of words in a free association task may serve as a strong indicator of antonymy. Evidence shows that some antonym pairs tend to be stable across word senses, while other pairs do not (Murphy 2003: 33); hence, according to Murphy (2003: 34), stability of some antonym pairs across senses and contexts serves as good evidence that those are canonical antonyms. The issue of antonym symmetry has been raised often due to its relevance in lexicography, and pertinent studies have shown that the lexicographic treatment of antonyms differs, implying that different dictionaries employ different approaches to this issue (Jakić 2015; Paradis & Willners 2007; Šarić 1994).

The strength of the relationship of antonyms in a pair has been attested through another phenomenon. If one member in an antonym pair acquires a new sense, the other member may also develop a similar sense, i.e. their original opposition may remain in the new domain (Paradis et al. 2009: 415). Rasulić (2020) analysed corpus instances of occurrence of 10 pairs of English canonical antonyms (*high/low, long/short, broad/narrow, deep/shallow, thick/thin, heavy/light, hard/soft, large/small, fast/slow, hot/cold*) in their semantic extension to investigate the potential of antonymy for dynamic meaning construal. She found that, in semantically creative instances of antonym use, antonym relations can be projected so that in the other adjective in the pair, an extended sense is activated, referred to as 'dormant sense' (Rasulić 2020: 147) (e.g. *high priest – low priest* (dormant sense); *cold statistics – hot statistics*); this projection is induced by context (Rasulić 2020: 149).

Various authors have pointed to the importance of context for the phenomenon of antonymy. M. L. Murphy argues that antonym variability may lead us to conclude that “antonymy relates senses or word uses rather than relating the words themselves” (Murphy 2003: 34). G. Murphy and Andrew (1993) empirically explored whether different contexts, e.g. different nouns used with adjectives, elicited different antonyms. Fourteen English adjectives were shown to respondents in isolation and in the context of a noun (*dry* vs. *dry wine*, *dry cake*). Respondents’ task was to first provide an antonym when the adjective was given in isolation and then when it was paired with nouns. It was shown that the elicited antonyms differed depending on the presence or absence of context. Deignan (1999) used corpus data to investigate whether the antonymy relation between adjective pairs such as *hot/cold* or *warm/cool* in literal senses holds in metaphorical senses. She found that these adjectives were relatively rarely used as antonyms in metaphorical senses. In a corpus-based study, Rasulić (2004) also established that the antonymous relationship holding between the literal senses of adjective pairs *high/low* and *visok/nizak* may not be replicated in their metaphorical senses. In the empirical research by Jakić Šimšić and Vesić Pavlović (2020) on 22 Serbian adjectives shown to respondents in their three senses within a sentence, it was found that different senses (primary, secondary concrete, secondary abstract) affected the number of antonyms given by respondents in an elicitation task.

In empirical tasks, context plays a significant role in mutual elicitation of antonyms; namely, “the less contextually constrained the pairings are, the more strongly they will elicit one another in context-free elicitation experiments” (Paradis et al. 2009: 415). The results of a previous empirical study on 394 adjectives of Serbian language (Jakić Šimšić 2021), shown to respondents without any context, indicate that certain adjective pairs exhibit a strong degree of symmetry while in others the level of symmetry is quite low, possibly due to the existence of alternative lexemes with similar meaning in the linguistic system which may serve as more appropriate antonyms. It is argued that different factors influence antonym pair asymmetry

in isolation, such as the number of adjective meanings (polysemy), meaning probability, the existence of alternative lexemes with the same or similar meaning (synonyms, near synonyms) etc. Some authors propose that markedness may also play an important role in the strength of associative relationship between antonyms so that the marked member of an antonymous pair may elicit the other member much more frequently than the reverse (e.g. *crn* elicits *beo* much more frequently than *beo* elicits *crn*; Todić 2016: 99).

3. RESEARCH AIM

Bearing in mind the results of previous studies on the importance of context for antonymy, we wish to empirically investigate the effect of context on adjective antonym symmetry in Serbian. Under antonym symmetry, in this paper we imply the strength of the associative relationship between the members of the antonym pair in both directions (Jakić Šimšić 2021). As for the notion of context, it should be borne in mind that, in this study, it refers to different senses in which the adjective was activated, but also to different contexts in which the adjective was shown to respondents, i.e. the sentence and phrase context.

The aim of the research is: (1) to determine whether the given adjectives reciprocally elicit each other in each of the investigated contexts (primary, secondary concrete, secondary abstract sense); (2) to establish whether the strength of symmetry varies depending on the sense in which the adjective is activated and (3) to compare the results obtained in two separate tasks – one, where adjectives were activated within a sentence and the other, where adjectives were activated within a phrase.

4. MATERIALS AND METHODS

We analyse 22 polysemous Serbian adjectives, i.e. 11 pairs from the perspective of their primary meaning: *pun* – *prazan* [full – empty],

širok – *uzak* [wide – narrow], *dubok* – *plitak* [deep – shallow], *čist* – *prljav* [clean – dirty], *lak* – *težak* [light – heavy], *veliki* – *mali* [big – small], *gust* – *redak* [thick – thin], *topao* – *hladan* [warm – cold], *visok* – *nizak* [tall – short], *dug* – *kratak* [long – short] and *brz* – *spor* [fast – slow]. Three senses were selected for each adjective relying on the referent dictionary of Serbian (RSJ 2007): primary, secondary concrete, and secondary abstract sense. Since some of the selected adjectives did not have the listed secondary concrete or abstract sense in the used referent dictionary, this yielded 62 senses in total¹.

Based on the descriptions of the senses extracted from the dictionary, we constructed sentences and phrases which served as stimuli in two empirical studies with native speakers of Serbian². They were instructed to replace the underlined adjective in a sentence or a phrase with its antonym.

For Task 1, we constructed stimuli in which the given adjectives were used in a sentence which activated one of the senses (e.g. *Kamen je upao u dubok bunar.* [A stone fell into a deep well]; *Trčali su po dubokom snegu.* [They were running through deep snow]; *Znali su malo o dubokoj prošlosti.* [They knew little about the deep past])³. Participants in Task 1 were native speakers of Serbian (N=81), first-year psychology students of the University of Belgrade. There were 82.7% female and 17.3% male respondents in the sample. Mean age of respondents in the first sample was 20 (M=20.43, SD=4.94).

For Task 2, based on the same adjective senses extracted from the dictionary, we constructed phrases in which the adjective was used in a particular context – primary sense: *dubok bunar* [a deep well], secondary concrete: *dubok sneg* [deep snow] and secondary abstract:

¹ This pertains to the adjectives *uzak*, *prljav* and *hladan*, for which we had no secondary concrete sense, and the adjective *slow*, with no secondary abstract sense.

² A detailed account of the process of selection of adjective senses from the RSJ dictionary can be found in Jakić Šimšić and Vesić Pavlović 2020 (59–61).

³ For the full list of sentence stimuli, see Appendix 1.

duboka prošlost [the deep past]⁴. Phrases served as stimuli in the second empirical study conducted on the second sample. Participants in this task were native speakers of Serbian (N=59) belonging to general population. The gender structure of this sample included 64.4% female and 35.6% male respondents, with the mean age of 34.6 (M=34.59, SD=17.88).

In data analysis, we performed the following steps. First, based on the collected data, we established a list of antonyms stated as dominant by the respondents in both tasks and calculated the percentage in which the dominant antonym was stated by the respondents in both tasks. Second, we singled out the cases in which adjectives from the posited pairs mutually elicited each other, as well as those in which elicitation was not reciprocal. In the next step, we calculated the difference between the stated percentages in cases where symmetry was recorded to establish the strength of symmetry (relying on methodology implemented in the previous study by Jakić Šimšić 2021). Finally, we compared the results obtained in two tasks with the aim of investigating the relevance of context in which the adjective was used (primary, secondary concrete, secondary abstract sense) and the effect of the type of task (sentence-stimuli vs. phrase-stimuli) on the occurrence of antonym symmetry or asymmetry in the collected data.

5. RESULTS

The obtained findings are presented in tables which contain the percentage in which the dominant antonym was stated by the participants and an additional column in which we calculated the difference between the percentages in which dominant antonyms were stated for both adjectives in cases where symmetry was recorded.

⁴ For the full list of phrase stimuli, see Appendix 2.

5.1. ANTONYM SYMMETRY WHEN THE ADJECTIVE IS USED IN ITS PRIMARY SENSE

When the adjective is activated in its primary sense in a sentence, symmetry is recorded in all posited pairs but one (*dug – kratak*). The adjective *dug* elicited the antonym *kratak* in 100% of cases, but *kratak* dominantly elicited the antonym *dugačak* (71.6%). Still, this may be viewed as seeming asymmetry since *dug* and *dugačak* may be regarded as forms of the same adjective. The percentage in which the dominant antonym was stated by respondents is very high and it ranges from 100% to 83.9% (Table 1).

Table 1: Antonym symmetry when the adjective is activated in primary sense in a sentence

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (X) | (%) | Difference |
|---------------|----------------------|------|---------------|----------------------|------|------------|
| <i>dubok</i> | plitak | 100 | <i>plitak</i> | dubok | 100 | 0 |
| <i>pun</i> | prazan | 100 | <i>prazan</i> | pun | 100 | 0 |
| <i>čist</i> | prljav | 100 | <i>prljav</i> | čist | 98.8 | 1.2 |
| <i>brz</i> | spor | 96.3 | <i>spor</i> | brz | 93.8 | 2.5 |
| <i>uzak</i> | širok | 96.3 | <i>širok</i> | uzak | 93.8 | 2.5 |
| <i>nizak</i> | visok | 100 | <i>visok</i> | nizak | 96.3 | 3.7 |
| <i>veliki</i> | mali | 98.8 | <i>mali</i> | veliki | 93.8 | 5 |
| <i>lak</i> | težak | 100 | <i>težak</i> | lak | 91.3 | 8.7 |
| <i>topao</i> | hladan | 98.8 | <i>hladan</i> | topao | 88.9 | 9.9 |
| <i>redak</i> | gust | 98.8 | <i>gust</i> | redak | 83.9 | 14.9 |

In the case of activating the adjective in its primary sense in a phrase, we obtain similar results as in the above-case (Table 2). There is symmetry in all pairs but one (*dug – kratak*; the dominant antonym given for the adjective *kratak* is again *dugačak*). The percentage in which the dominant antonym was stated is also high, ranging from 100% to 74.9%.

Table 2: Antonym symmetry when the adjective is activated in primary sense in a phrase

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (X) | (%) | Difference |
|---------------|----------------------|------|---------------|----------------------|------|------------|
| <i>veliki</i> | mali | 93.2 | <i>mali</i> | veliki | 94.8 | 1.6 |
| <i>pun</i> | prazan | 98.3 | <i>prazan</i> | pun | 100 | 1.7 |
| <i>uzak</i> | širok | 96.6 | <i>širok</i> | uzak | 94.9 | 1.7 |
| <i>nizak</i> | visok | 98.3 | <i>visok</i> | nizak | 96.6 | 1.7 |
| <i>čist</i> | prljav | 96.6 | <i>prljav</i> | čist | 100 | 3.4 |
| <i>dubok</i> | plitak | 94.9 | <i>plitak</i> | dubok | 100 | 5.1 |
| <i>topao</i> | hladan | 96.6 | <i>hladan</i> | topao | 86.4 | 10.2 |
| <i>lak</i> | težak | 94.9 | <i>težak</i> | lak | 83 | 11.9 |
| <i>brz</i> | spor | 98.3 | <i>spor</i> | brz | 84.7 | 13.6 |
| <i>redak</i> | gust | 96.6 | <i>gust</i> | redak | 74.9 | 21.7 |

5.2. ANTONYM SYMMETRY WHEN THE ADJECTIVE IS USED IN ITS SECONDARY CONCRETE SENSE

When the adjective is used in the secondary concrete sense in a sentence, there is symmetry in all of the pairs (Table 3)⁵. The percentage in which adjectives were mutually elicited ranges from 100% to 55.5%.

⁵ Still, bearing in mind that there were three adjectives without the secondary concrete sense, as mentioned earlier.

Table 3: Antonym symmetry when the adjective is activated in secondary concrete sense in a sentence

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (X) | (%) | Difference |
|---------------|----------------------|------|---------------|----------------------|------|------------|
| <i>spor</i> | brz | 100 | <i>brz</i> | spor | 98.8 | 1.2 |
| <i>pun</i> | prazan | 95 | <i>prazan</i> | pun | 97.5 | 2.5 |
| <i>težak</i> | lak | 79 | <i>lak</i> | težak | 81.5 | 2.5 |
| <i>mali</i> | veliki | 100 | <i>veliki</i> | mali | 96.3 | 3.7 |
| <i>nizak</i> | visok | 98.8 | <i>visok</i> | nizak | 91.3 | 7.5 |
| <i>redak</i> | gust | 100 | <i>gust</i> | redak | 83.9 | 16.1 |
| <i>plitak</i> | dubok | 98.8 | <i>dubok</i> | plitak | 82.7 | 16.1 |
| <i>dug</i> | kratak | 100 | <i>kratak</i> | dug | 55.5 | 44.5 |

Similar to the above case, when the secondary concrete sense is activated in a phrase, there is symmetry in all of the pairs (Table 4). The percentage in which adjectives were mutually elicited ranges from 100% to 71.2%.

Table 4: Antonym symmetry when the adjective is activated in secondary concrete sense in a phrase

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (X) | (%) | Difference |
|---------------|----------------------|------|---------------|----------------------|------|------------|
| <i>spor</i> | brz | 91.5 | <i>brz</i> | spor | 89.8 | 1.6 |
| <i>težak</i> | lak | 74.6 | <i>lak</i> | težak | 71.2 | 3.4 |
| <i>mali</i> | veliki | 100 | <i>veliki</i> | mali | 96.6 | 3.4 |
| <i>nizak</i> | visok | 100 | <i>visok</i> | nizak | 94.9 | 5.1 |
| <i>pun</i> | prazan | 98.3 | <i>prazan</i> | pun | 91.5 | 6.8 |
| <i>redak</i> | gust | 94.9 | <i>gust</i> | redak | 83 | 11.9 |
| <i>dug</i> | kratak | 98.3 | <i>kratak</i> | dug | 78 | 20.3 |
| <i>plitak</i> | dubok | 100 | <i>dubok</i> | plitak | 78 | 22 |

5.3. ANTONYM SYMMETRY WHEN THE ADJECTIVE IS USED IN ITS SECONDARY ABSTRACT SENSE

When the adjective is activated in its secondary abstract sense in a sentence, symmetry is recorded in one half of pairs, with the percentage in which the dominant antonym was stated ranging from 91.3% to 46.9% (Table 5).

Table 5: Antonym symmetry when the adjective is activated in secondary abstract sense in a sentence

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (X) | (%) | Difference |
|---------------|----------------------|------|---------------|----------------------|------|------------|
| <i>mali</i> | veliki | 90.1 | <i>veliki</i> | mali | 82.7 | 7.4 |
| <i>prljav</i> | čist | 53 | <i>čist</i> | prljav | 64.2 | 11.2 |
| <i>težak</i> | lak | 88.9 | <i>lak</i> | težak | 76.5 | 12.4 |
| <i>dug</i> | kratak | 91.3 | <i>kratak</i> | dug | 77.8 | 13.5 |
| <i>topao</i> | hladan | 88.9 | <i>hladan</i> | topao | 46.9 | 42 |

There are five pairs in which there is no symmetry, with two subcases:

- (1) one adjective elicits the other from the posited lexical antonym pair in varying percent, but the second does not (Table 6).

Table 6: Antonym asymmetry when the adjective is activated in secondary abstract sense in a sentence (subcase 1)

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (Z) | (%) |
|---------------|----------------------|------|--------------|----------------------|------|
| <i>nizak</i> | visok | 87.6 | <i>visok</i> | dubok | 59.2 |
| <i>plitak</i> | dubok | 49.4 | <i>dubok</i> | blizak | 38.3 |
| <i>uzak</i> | širok | 95 | <i>širok</i> | kratak | 43.2 |

- (2) both adjectives from the pair elicit other adjectives as antonyms, i.e. not the ones posited as members of the lexical antonym pairs (Table 7).

Table 7: Antonym asymmetry when the adjective is activated in secondary abstract sense in a sentence (subcase 2)

| Stimulus (X) | Dominant antonym (Z) | (%) | Stimulus (Y) | Dominant antonym (W) | (%) |
|--------------|----------------------|------|---------------|----------------------|------|
| <i>pun</i> | tih | 18.5 | <i>prazan</i> | značajan | 33.3 |
| <i>redak</i> | čest | 70.4 | <i>gust</i> | opušten | 21 |

When the adjective is activated in its secondary abstract sense within a phrase, there are 8 pairs in which symmetry is recorded, which is a difference compared to the case in which the adjective is activated in the same sense within a sentence (Table 8). The percentage in which the dominant antonym is stated is fairly similar to that of the first task and ranges from 95% to 30.5%.

Table 8: Antonym symmetry when the adjective is activated in secondary abstract sense in a phrase

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (X) | (%) | Difference |
|---------------|----------------------|------|---------------|----------------------|------|------------|
| <i>prljav</i> | čist | 64.4 | <i>čist</i> | prljav | 64.4 | 0 |
| <i>pun</i> | prazan | 30.5 | <i>prazan</i> | pun | 33.9 | 3.4 |
| <i>težak</i> | lak | 91.5 | <i>lak</i> | težak | 79.7 | 11.8 |
| <i>dug</i> | kratak | 88.1 | <i>kratak</i> | dug | 72.9 | 15.2 |
| <i>mali</i> | veliki | 88.1 | <i>veliki</i> | mali | 72.8 | 15.3 |
| <i>topao</i> | hladan | 93.2 | <i>hladan</i> | topao | 57.6 | 35.6 |
| <i>nizak</i> | visok | 84.7 | <i>visok</i> | nizak | 49.1 | 35.6 |
| <i>uzak</i> | širok | 95 | <i>širok</i> | uzak | 52.5 | 42.5 |

Consequently, there are fewer posited adjective pairs that do not exhibit symmetry compared to the case in which the adjective is activated in this sense in a sentence; there are two pairs in which there is no symmetry (Table 9). If we compare these findings with those in the task with sentences as stimuli, it can be observed that all five pairs in which symmetry was present in the sentence-stimuli task also prove to be symmetrical in the task with phrases as stimuli. Two pairs in which there is asymmetry in the phrase-stimuli task also showed asymmetry in the sentence-stimuli task.

Table 9: Antonym asymmetry when the adjective is activated in secondary abstract sense in a phrase

| Stimulus (X) | Dominant antonym (Y) | (%) | Stimulus (Y) | Dominant antonym (Z) | (%) |
|---------------|----------------------|------|--------------|----------------------|------|
| <i>plitak</i> | dubok | 47.5 | <i>dubok</i> | blizak | 22 |
| <i>gust</i> | redak | 32.2 | <i>redak</i> | čest | 64.4 |

To summarise the above-presented findings: in the case of activating their primary sense, a vast majority of adjectives within the posited pairs reciprocally elicit each other in both the sentence context and phrase context, i.e. the results show existence of symmetry in 10 adjective pairs, while the case of the antonym pair *dug – kratak* may be deemed as only seeming asymmetry. When the adjective is used in the secondary concrete sense, whether in a sentence or a phrase, there is symmetry in all 8 posited pairs. Finally, when the adjective is used in the secondary abstract sense, there are some differences, depending on whether it is activated in a sentence or a phrase. Symmetry was found in 5 out of 10 pairs in the case when sentences were used as stimuli, and in 8 pairs out of 10 when phrases were used as stimuli.

Now we proceed with a discussion on the differences in the percentage in which the dominant antonym was stated by the respondents in two members of the pair, which may serve as an indicator of the strength of symmetry in the posited antonymous pairs. If the difference is low, this implies that the adjectives were

mutually elicited in a similar percent, which points to a relatively stable symmetry between them. We will draw on the data on the calculated difference, presented in the final column of Tables 1, 2, 3, 4, 5 and 8, i.e. all the tables which contain adjective pairs in which symmetry was found. We interpret the strength of symmetry through three levels of symmetry: high (up to 5% difference), medium (between 5.1 and 16.1) and low symmetry (16.2 onwards).

In the case of activating the adjective in primary sense in the sentence context (Table 1), we can see that the level of symmetry is dominantly high (7 pairs), followed by medium (3 pairs). Results are similar in the phrase context (Table 2): high level of symmetry in 6 pairs, followed by medium (3 pairs) and low (1). When the adjective is used in secondary concrete sense in the sentence context (Table 3), the level of symmetry is again dominantly high (5 pairs), followed by medium (2 pairs) and low (1 pair); the results are almost the same when the adjective is activated in the phrase context (Table 4): high (5 pairs), medium (1 pair) and low (2 pairs). Finally, when we activate the adjective in the secondary abstract sense within a sentence (Table 5), symmetry ranges from medium (4 pairs) to low (1 pair). The results are slightly different when this adjective sense is activated within a phrase (Table 8): all three levels of symmetry are present, high symmetry in 2 pairs, medium in 3 pairs and low in 3 pairs. Although the number of pairs in which symmetry is present is bigger in this task compared to the task with sentence-stimuli, we can see that the level of symmetry is mostly medium and low.

6. CONCLUDING REMARKS

The focus of our paper was to explore the effect of context on the preservation of symmetry in antonymous adjective pairs in Serbian, relying on the data from two empirical studies. The obtained findings point towards a significant and diverse role of context with this respect.

First of all, the number of pairs that exhibit symmetry is overall the highest when the adjectives are activated in primary and secondary

concrete senses, while a lower number of pairs exhibit symmetry in the secondary abstract senses. This applies to both specific contexts in which antonyms were shown, sentence-stimuli and phrase-stimuli.

In the cases where asymmetry was recorded, there are some differences between two tasks. In the task using sentence-stimuli, there are cases of asymmetry where mutual elicitation occurs in one direction and cases where it does not occur at all. It may be argued that the reasons behind this kind of asymmetry predominantly lie in the absence of the corresponding sense in one of the adjectives posited as pair members or in both adjectives (*Otpevala je to punim glasom.* [She sang this in full voice.] vs. **Otpevala je to praznim glasom.* [She sang this in empty voice.] or *To su prazne reči.* [These are empty words.] vs. **To su pune reči.* [These are full words.]). That is why, in these cases, respondents in Task 1 dominantly stated antonyms *tih* (*Otpevala je to tihim glasom.* [She sang this in silent voice.]) and *značajan* (*To su značajne reči.* [These are significant words.]). Still, it is possible that, if the context in which the adjective is shown is more narrow, the preservation of the original opposition seems like a viable option, which is shown by the answers of the respondents from Task 2 related to this same adjective pair. Namely, the dominant antonym in the respondents' answers for this pair was *pun glas* – **prazan glas*, *prazne reči* – **pune reči*, although there is no mention of these senses in the referent dictionary. However, for some other adjective pairs, respondents in both tasks, i.e. regardless of whether the adjective was shown in the context of a sentence or a phrase, chose the same antonyms which did not preserve the original symmetry (e.g. sentence-stimulus: *Znali su malo o dubokoj prošlosti*, phrase-stimulus: *duboka prošlost* – in both cases, the dominantly stated antonym was *blizak*[close]).

Another issue that was discussed is whether the strength of symmetry, expressed through the difference in which both antonyms were elicited, varies depending on the activated adjective sense. When using the adjective in primary sense, the level of symmetry is dominantly high in both tasks; the same holds in the case of secondary abstract sense, but in both tasks there appear cases of medium and

low symmetry as well. Finally, when it comes to secondary abstract sense, the level of symmetry is dominantly medium and low, with a slight difference in the task with phrases as stimuli, where two cases of high symmetry also occur. Hence, it is inferred that the associative relationship in these pairs is the strongest when the adjective is activated in the primary sense, slightly drops when the secondary concrete sense is activated, and is the lowest in the secondary abstract sense. This holds for both tasks.

Based on the results, it may be argued that context, in terms of different senses of an adjective, has an effect on adjective antonym symmetry. The effect of the context in which the adjective is shown to respondents, whether it is a sentence (a wider context) or a phrase (a narrower context), remains an open issue, which deserves further research attention in future. Another open issue is a possibility that, in some cases, a strong antonymous relationship between the pair members established in their primary meanings may lead the respondents to assume that the symmetrical relationship holds in the cases where the referent dictionary does not mention a corresponding sense of the other pair member at all. The possibility of the aforementioned dynamic meaning construction in this type of empirical tasks also deserves to be explored in more detail.

The findings of the study may bear practical relevance for lexicographers, for instance, to include specific qualifiers when stating an antonym for a given word which would explain whether it refers to all senses of the word or a specific sense only. Still, the presented research results are significantly limited by a small number of adjective pairs used and a restricted number of their senses. Hence, future studies should focus on the effect of context on antonym symmetry using larger sets of adjectives and various classes of polysemous adjectives, which would be activated in a variety of contexts. Additionally, a more precise methodology for investigating antonym symmetry in different contexts should be developed, combining different measures used to assess symmetry, so as to yield more reliable findings.

REFERENCES

- Deese, J. (1964). The Associative Structure of Some Common English Adjectives. *Journal of Verbal Learning and Verbal Behaviour*, 3(5): 347–357.
- Deignan, A (1999). Metaphorical polysemy and paradigmatic relations: A corpus study. *Word* 50(3): 319–338.
- Jakić, M. (2015). Formalno označavanje antonimije u Rečniku SANU. *Zbornik Matice srpske za filologiju i lingvistiku* LVIII(1): 155–178.
- Jakić Šimšić, M. and T. Vesić Pavlović (2020). Značenjska suprotnost prideva srpskog jezika u rečeničnom kontekstu. *Zbornik Matice srpske za filologiju i lingvistiku* LXIII(2): 55–74.
- Jakić Šimšić, M. (2021). Simetrija antonimske veze prideva srpskog jezika u testu kontrolisanih asocijacija. *Naš jezik* LII(1): 45–61.
- Kostić, N. (2016). On conceptual nature of antonymy: Evidence from corpus-based investigations. *Journal of Foreign Language Teaching and Applied Linguistics*, Spring, 175–188.
- Murphy, G. and J. Andrew (1993). The conceptual basis of antonymy and synonymy in adjectives. *Journal of Memory and Language* 32(3): 301–319.
- Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*. Cambridge: Cambridge University Press.
- Paradis, C. and C. Willners (2007). Antonyms in dictionary entries: methodological aspects. *Studia Linguistica* 61(3): 261–277.
- Paradis, C., C. Willners and S. Jones (2009). Good and bad opposites. Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon* 4(3): 380–429.
- Rasulić, K. (2004). *Jezik i prostorno iskustvo: Konceptualizacija vertikalne dimenzije u engleskom i srpskohrvatskom jeziku*. Beograd: Filološki fakultet.
- Rasulić, K. (2016). *Exploring Semantic Relations*. Beograd: Filološki fakultet.
- Rasulić, K. (2020). High Expectations and Low Blows: Antonymy and Dynamic Meaning Construction. *Belgrade English Language and Literature Studies*, 12 (1): 137–161.
- RSJ: Rečnik srpskoga jezika. Novi Sad: Matica srpska, 2007.
- Šarić, Lj. (1994). Antonimija u strukturi jednojezičnih rječnika. *Filologija* 22–23: 269–274.
- Storjohann, P. (2016). Sense relations. In: *The Routledge Handbook of Semantics* (N. Riemer, ed.), New York: Routledge, 248–265.
- Todić, B. (2016). Priroda asocijativne veze antonima, *Naš jezik* XLVII/3–4: 89–103.

Тяiana Весич Павлович и Милена Якич Шимшич

ЭФФЕКТ КОНТЕКСТА НА (А)СИММЕТРИЮ ПРИЛАГАТЕЛЬНЫХ-АНТОНИМОВ В СЕРБСКОМ ЯЗЫКЕ

Резюме

Данная работа исследует антонимическое отношение 22 полисемичных сербских прилагательных, т. е. 11 пар прилагательных-антонимов из перспективы первичного значения, когда три их значения (первичное, вторичное и вторичное отвлеченное) активируются в задании в рамках предложения или словосочетания. Мы опираемся на данные двух эмпирических исследований, в которых респонденты, носители языка, имели задание привести антоним к подчеркнутому прилагательному, активированному в соответствующем значении в предложении или словосочетании. Работа преследует следующие цели: (1) установить, возникает ли взаимная элицитация прилагательных в рамках данных пар в любом из активированных значений, (2) установить, варьирует ли степень антонимичности в зависимости от значения, в котором прилагательное активировано и (3) сравнить полученные результаты в двух заданиях – первом, в котором прилагательные активированы в предложении, и другом, в котором прилагательные активированы в словосочетании. Когда прилагательные активируются в первичном и вторичном конкретных значениях, несмотря на тип задания, оказывается, что симметрия имеется во всех анализируемых парах. Однако, когда прилагательное активируется во вторичном отвлеченном значении, количество пар, у которых симметрия наблюдается, уменьшается в обоих заданиях. В задании, в котором стимулами были предложения, симметрия возникает в половине пар; в случаях асимметрии различаем две ситуации: одна, в которой одно прилагательное вызывает в разном проценте другой элемент пары, но другое не вызывает (напр. *nizak* – visok, но *visok* – dubok), и другая ситуация, когда оба прилагательных, составляющих пару, вызывают другие прилагательные в качестве антонимов (напр. *pun* – tih, *prazan* – značajan). В задании со стимулами-словосочетаниями, количество пар, в которых наблюдается симметрия, – меньше, чем в случае задания со стимулами-предложениями. Когда прилагательные активируются в первичном и вторичном конкретных значениях, степень антонимичности – высокая в обоих заданиях, в то время как во вторичном отвлеченном значении уровень симметрии несколько различается в зависимости от типа задания. В заключении мы коснемся роли контекста в поддержании симметрии антонимической пары и продискутируем о возможных причинах наблюдаемой симметрии.

Ключевые слова: антонимия, прилагательные, сербский язык, контекст предложения, контекст словосочетания

APPENDIX 1
THE LIST OF SENTENCES USED AS STIMULI IN TASK 1.

- | | |
|---|---|
| 1. Stela je brz konj. | 32. Ovde je reka prilično plitka. |
| 2. Na poslu su bili baš brzi. | 33. To je zbog njegove plitke pameti. |
| 3. Pomoć davljenicima bila je veoma brza. | 34. Na stolu je bila prazna kesa. |
| 4. Došao je u veliku šumu. | 35. Poslali su prazan kamion. |
| 5. Ove cipele su mi velike. | 36. To su prazne reči. |
| 6. Njagovo znanje je veliko. | 37. U kesi su prljave košulje. |
| 7. Eno ga onaj visoki momak. | 38. Bavio se samo prljavim poslovima.. |
| 8. Kroz oblake se videlo visoko sunce. | 39. Uzeo je punu čašu. |
| 9. Čuo je njen visoki glas. | 40. Sve sobe su pune. |
| 10. Pili su gusto južno vino. | 41. Otpevala je to punim glasom. |
| 11. Na horizontu se pojavio gust dim. | 42. Prošao je prstima kroz svoju retku bradu. |
| 12. Situacija je bila prilično gusta. | 43. Kroz retku maglu nazirale su se kuće. |
| 13. Kamen je upao u dubok bunar. | 44. Njegov deda bio je redak junak u ratu. |
| 14. Trčali su po dubokom snegu. | 45. Tako je spor u pokretima. |
| 15. Znali su malo o dubokoj prošlosti. | 46. U gradu je često spora vožnja. |
| 16. Više volim dugu kosu. | 47. Ugledao je teško kamenje. |
| 17. Haljina mi je duga. | 48. Bila je u haljini od teške svile. |
| 18. Izmorila ga je duga bolest. | 49. Igrao je uvek teške uloge. |
| 19. Imao je kratak rep. | 50. Popila je toplo mleko. |
| 20. Ovi rukavi su mi kratki. | 51. Uzmi ovu toplu haljinu. |
| 21. Nastade kratka tišina. | 52. Posmatrao ju je toplim pogledom. |
| 22. Podigla je svoj laki kofer. | 53. Provela nas je kroz uzak hodnik. |

- | | |
|---------------------------------------|--|
| 23. Hodao je u kaputu od lakog štofa. | 54. Velika književnost nije za uzak krug čitalaca. |
| 24. Išao je lakim korakom. | 55. Pila je hladnu limunadu. |
| 25. Sedeo je u svojoj maloj sobi. | 56. Tako je hladan prema meni. |
| 26. Sve reči pisao je malim slovima. | 57. Pored stola su čiste čarape. |
| 27. To je čovek male pameti. | 58. Pogledala je u čisto nebo. |
| 28. Ne voli niske devojke. | 59. Imali su čist obraz. |
| 29. U sobi je bio veoma nizak plafon. | 60. Imao je široka ramena. |
| 30. Doveo mi je snaju niskog roda. | 61. Iz hodnika se ulazi u široku prostoriju. |
| 31. Obuo je plitke cipele. | 62. Usledila je široka debata. |

APPENDIX 2

THE LIST OF PHRASES USED AS STIMULI IN TASK 2.

- | | |
|---------------------|----------------------|
| 1. brz konj | 32. plitka reka |
| 2. brz na poslu | 33. plitka pamet |
| 3. brza pomoć | 34. prazna kesa |
| 4. velika šuma | 35. prazan kamion |
| 5. velike cipele | 36. prazne reči |
| 6. veliko znanje | 37. prljava košulja |
| 7. visoki momak | 38. prljavi poslovi |
| 8. visoko sunce | 39. puna čaša |
| 9. visoki glas | 40. puna soba |
| 10. gusto vino | 41. pun glas |
| 11. gust dim | 42. retka brada |
| 12. gusta situacija | 43. retka magla |
| 13. dubok bunar | 44. redak junak |
| 14. dubok sneg | 45. spor u pokretima |
| 15. duboka prošlost | 46. spora vožnjavv |
| 16. duga kosa | 47. težak kamen |

- | | |
|--------------------------|---------------------------------|
| 17. duga haljina | 48. teška svila |
| 18. duga bolest | 49. teška uloga |
| 19. kratak rep | 50. toplo mleko |
| 20. kratki rukavi | 51. topla haljina |
| 21. kratka tišina | 52. topao pogled |
| 22. lak kofer | 53. uzak hodnik |
| 23. kaput od lakog štofa | 54. uzak krug čitalaca |
| 24. lak korak | 55. hladna limunada |
| 25. mala soba | 56. hladan čovek (prema nekome) |
| 26. malo slovo | 57. čiste čarape |
| 27. mala pamet | 58. čisto nebo |
| 28. niska devojka | 59. čist obraz |
| 29. nizak plafon | 60. široka ramena |
| 30. snaja niskog roda | 61. široka prostorija |
| 31. plitke cipele | 62. široka debata |

Emilija Milojević*

University of Belgrade – Faculty of Philology

COMPLEXITÉ SYNTAXIQUE À L'ÉCRIT EN L2: LE CAS DU FRANÇAIS¹

Abstrait

Cet article examine la complexité syntaxique de la production écrite des étudiants de troisième année de français langue étrangère. L'objectif de la recherche est de déterminer quelle est la complexité de la langue écrite des répondants en termes de différentes variables quantitatives proposées pour mesurer objectivement la complexité syntaxique, ainsi que de voir quel est le lien entre la qualité générale du texte produit d'une part, et la complexité syntaxique d'autre part. Après l'analyse statistique faite à l'aide du logiciel SPSS, les résultats montrent que la langue écrite des étudiants de troisième année est modérément complexe. Des essais jugés de meilleure qualité se sont avérés syntaxiquement plus complexes que les essais de qualité inférieure. Les mesures telles que la longueur moyenne des unités T correctes et le nombre moyen des unités T correctes se sont avérées les plus utiles pour capter les différences entre les deux groupes d'étudiants.

Mots-clés : complexité syntaxique, unité T, production écrite en L2, français langue étrangère, acquisition de la L2, qualité du texte, indice de développement

* emilija.milojevic@fil.bg.ac.rs

¹ Ce papier est le résultat du travail sur le mémoire du master intitulé "Analyse de la complexité syntaxique dans la production écrite des étudiants de troisième année de français". Cet article n'en présente qu'une partie, légèrement modifiée.

1. INTRODUCTION

Pendant longtemps les chercheurs avaient besoin d'un indice de développement dans le domaine de l'acquisition de la L2 pour mesurer objectivement la compétence d'un élève (Larsen-Freeman & Strom, 1977). Dans le domaine de l'acquisition de la langue maternelle, il existe déjà des indicateurs de développement syntaxique proposés, comme l'indice de syntaxe productive et la longueur moyenne des énoncés. Le besoin d'introduire un tel indice de développement dans le domaine de la L2 est justifié par l'opinion qu'il existe des tendances générales de développement communes à tous les élèves de la L2, comme c'est le cas avec la L1 (Larsen-Freeman & Strom, 1977). Les chercheurs se sont concentrés sur le développement syntaxique en L2 suivant l'idée que la capacité de combiner des propositions plus courtes et plus simples en phrases plus longues et plus complexes est étroitement liée au niveau de développement du langage (Monroe, 1975) et représente un fort indicateur de la compétence en L2.

Il existe de nombreuses définitions de la complexité syntaxique dans la littérature. Nous allons nous limiter ici à la définition de Wolfe-Quintero et al. (1998) qui définissent la complexité syntaxique en termes de l'accès rapide et l'utilisation d'un large éventail de structures linguistiques de base et de structures sophistiquées, tandis que l'absence de complexité signifie que l'apprenant a l'accès seulement à un nombre limité de structures de base. De la notion de complexité syntaxique ainsi définie, il ressort clairement qu'un texte complexe se caractérise par une plus grande variation de structures syntaxiques utilisées. Toutefois, comme le remarque De Clercq (2016), les recherches sur la complexité syntaxique prennent souvent une vue réductrice du sujet, vu que les chercheurs prêtent attention seulement à la complexité structurelle en mesurant la longueur de certaines unités linguistiques et incluent rarement la diversité des structures dans leurs recherches.

Plusieurs études ont essayé de déterminer la relation entre la complexité syntaxique et la qualité générale du texte produit (e.g., Beers & Nagy 2007; Yang et al. 2015; Kyle & Crossley 2018), entre la

complexité syntaxique et la compétence en L2 (voir Ortega 2003 pour la synthèse de ces études), ainsi qu'entre la complexité syntaxique et les jugements humains de la qualité du texte (Crossley & McNamara 2014).

2. COMMENT MESURER LA COMPLEXITÉ SYNTAXIQUE?

Des différentes unités de mesure ont été proposées dans la littérature pour mesurer objectivement la complexité syntaxique, et leur utilisation varie selon les chercheurs. Les mesures de complexité syntaxique visent généralement à quantifier une ou plusieurs dimensions de la complexité : variété de structures syntaxiques, longueur de l'unité, degré de complexité structurelle de certaines structures syntaxiques et quantité et type de coordination, subordination et intégration (Bulté & Housen, 2012: 35).

Lorsque Hunt (1965) a proposé l'unité T (*T-unit*)² comme indice de développement syntaxique, cette mesure a été l'une des mesures les plus "populaires" et les plus utilisées, même si certains chercheurs mettent en question la pertinence de cette unité (voir par exemple Biber et al. 2011 pour une évaluation critique de l'unité T). Dans la langue maternelle, le développement syntaxique se reflète dans la capacité de compresser autant d'informations que possible en aussi peu de mots que possible – ce qui conduit à une augmentation progressive de la longueur de l'unité T (Gaies, 1980). La logique est la même pour la L2 – les apprenants ajoutent de nouveaux éléments syntaxiques et prolongent l'unité T, ce qui signifie le progrès dans l'acquisition et l'usage plus avancé de la langue. Kyle & Crossley (2018) remarquent pourtant que toutes sortes de structures peuvent augmenter la longueur de l'unité T, mais que les mesures proposées jusqu'à présent ne captent pas le type de ces structures.³ Or, la longueur de l'unité

² Hunt (1965) a défini l'unité T de la manière suivante: «la phrase principale avec toutes les phrases subordonnées et les structures non phrastiques qui s'y trouvent ou qui y sont attachées».

³ Les mêmes auteurs (2018) font une distinction entre les indices fins de complexité syntaxique qui devraient être utilisés pour capter la structure interne de l'unité

linguistique nous indique seulement qu'il y a de l'élaboration, mais ne nous informe pas sur le type de l'élaboration en question (*Ibid.*).

Deux autres mesures souvent utilisées par les chercheurs sont la longueur moyenne des phrases et la longueur moyenne des propositions. Cependant, la pertinence de ces deux mesures a été mise en question vu qu'il a été montré que "les élèves au niveau initial de compétence coordonnent excessivement les propositions, de sorte que l'unité T est un meilleur indicateur de maturité linguistique que la longueur moyenne des phrases" (Monroe, 1975: 1026).

Si nous calculons le nombre moyen de propositions par l'unité T, nous obtenons une autre unité de mesure, qui représente l'indice de subordination. Plus le nombre de propositions par l'unité T est élevé, plus la subordination est présente dans la production des apprenants (Monroe, 1975). Pourtant, cette mesure n'est sans fautes non plus, vu que les indices de subordination ciblent seulement l'un des trois principales dimensions d'organisation syntaxique – le niveau de phrase, tandis qu'ils ne ciblent ni le niveau de proposition ni le niveau syntagmatique⁴ (Bulté & Housen, 2012: 37).

L'utilisation de mesures proposées est également problématique pour d'autres raisons. Beers & Nagy (2009: 196) trouvent que le nombre de propositions par unité T et la longueur moyenne de proposition mesurent deux différents aspects de l'usage de la syntaxe. Bulté & Housen (2012) dans leur synthèse de quarante études sur la L2 trouvent que les chercheurs utilisent un nombre très limité de mesures et souvent mesurent une même dimension de la complexité syntaxique plusieurs fois (voir Norris & Ortega 2009 pour l'explication de redondance des mesures). Un nombre limité de mesures et l'analyse

T et les indices traditionnels qui ne sont pas assez sensibles pour capter ces différences (*fine-grained indices vs. large-grained indices*).

⁴ Certains auteurs (Norris & Ortega 2009 ; Lu & Ai 2015 ; Kuiken & Vedder 2019) voient la complexité syntaxique comme une notion multidimensionnelle et distinguent plusieurs dimensions de complexité, parmi lesquelles se trouvent la complexité par subordination, la complexité générale, la complexité par élaboration de syntagme.

de petites quantités de données sont la conséquence du manque d'outils d'analyse automatique (Lu, 2011).

Cependant, quelques-unes parmi les mesures mentionnées se sont avérées plus utiles que les autres. Certains auteurs (Ortega 2003 ; Lu 2011) ont identifié les unités de mesure qui augmentaient de façon linéaire et qui pourraient être de bons indices de développement, parmi lesquelles se trouvent les mesures que nous allons utiliser dans notre recherche : la longueur moyenne d'unité T, de proposition et de phrase (voir Ortega 2003 ; Lu 2011 pour la liste complète de mesures). Les recherches antérieures sur la complexité syntaxique étaient concentrées notamment sur l'acquisition de l'anglais comme L2 et "ont révélé quelques tendances développementales générales, qui sont, dans le cas du français comme L2, le mieux documentés pour des apprenants suédophones" (De Clercq, 2016:4). Le but de ce papier est de documenter et d'analyser la production écrite des apprenants du français serbophones.

3. LES QUESTIONS DE RECHERCHE

Dans ce papier, nous allons essayer de répondre à trois questions suivantes :

1. Dans quelle mesure la langue écrite des étudiants de troisième année de français est-elle complexe au niveau de la syntaxe?

Nous allons répondre à cette question en utilisant les mesures largement adoptées par les chercheurs dans le domaine de l'acquisition de la L2 : longueur moyenne de proposition (w/c), d'unité T (w/T), de phrase (w/S), indice de subordination (c/T), indice de coordination (T/S). Dans la deuxième partie de l'analyse, aux mesures mentionnées, nous ajouterons aussi : longueur moyenne des unités T correctes et nombre moyen des unités T correctes.

2. Y a-t-il une relation directe entre la qualité du texte rédigé évaluée par le professeur du FLE et la complexité syntaxique?

Les chercheurs sont d'accord qu'un texte plus complexe ne signifie toujours un texte de meilleure qualité, mais qu'il devrait y avoir une relation entre la qualité et la complexité. Faute d'une définition précise de la qualité du texte, les chercheurs ont leurs propres critères selon lesquels ils jugent de la qualité du texte. Parmi ces critères se trouvent: réalisation de la tâche, organisation, développement des idées, cohérence, choix des mots, syntaxe (Kyle & Crossley 2018), accent mis sur le sujet, détails et élaborations, utilisation de la langue et des mots, ton du texte (Beers & Nagy 2007), contenu, organisation, vocabulaire, utilisation de la langue et orthographe (Crossley & McNamara 2014). La nature exacte de la relation entre la qualité générale du texte produit et la complexité syntaxique n'a pas encore été établie avec certitude. Beers & Nagy (2007: 187) voient cette relation de la manière suivante: "la capacité à produire des phrases complexes est probablement mieux comprise comme une condition nécessaire mais pas une condition suffisante pour écrire des textes de haute qualité". Les mêmes auteurs (2007: 196) concluent que cette relation dépend du genre de texte et des mesures spécifiques employées. Malgré le fait que les résultats d'études visant à déterminer le lien entre ces deux dimensions de langue soient inconsistants, nous allons vérifier dans notre corpus si les essais classés dans le groupe supérieur selon la qualité générale du texte et d'après l'évaluation faite par le professeur de FLE expérimenté sont également plus complexes que les essais jugés de qualité moindre.

3. Quels sont les meilleures mesures pour évaluer la complexité syntaxique?

Malgré les critiques diverses de mesures proposées, nous allons essayer de déterminer quelles sont les meilleures mesures pour évaluer la complexité syntaxique de la production écrite de nos répondants et pour capter les différences entre les deux groupes qui produisent des textes de différente qualité.

4. MÉTHODOLOGIE

Pour répondre à ces questions, nous avons collecté 123 essais écrits par les étudiants de français de troisième année à la Faculté de Philologie à Belgrade. La langue maternelle de ces étudiants est le serbe; ils se sont inscrits à la Faculté de Philologie avec un certain niveau de connaissance de français et le niveau de compétence attendu en L2 en troisième année est B2. Les essais ont été réécrits sur ordinateur pour faciliter l'analyse et ont ensuite été divisés en deux groupes en fonction de la qualité générale du texte rédigé. Les essais ont été classés dans le groupe supérieur et inférieur selon l'évaluation faite par le professeur de FLE qui a utilisé les critères suivants: vocabulaire, morphosyntaxe, cohérence. En fonction de ces critères, l'évaluateur a attribué de 5 à 10 points à chaque essai. Les essais ayant obtenu 5, 6 et 7 points ont été classés dans le groupe inférieur de qualité, et les essais ayant obtenu 8, 9 et 10 points ont été classés dans le groupe supérieur de qualité.

Les travaux écrits ont été choisis parce qu'ils étaient facilement accessibles et il a semblé à l'auteur que la production écrite serait plus complexe que la production orale⁵, compte tenu de la possibilité de faire des pauses à l'écrit, de réfléchir, de planifier et de revenir sur ce qui a été écrit. Ces essais ont été rédigés lors de la partie écrite de l'examen de français contemporain. Il s'agit des essais argumentatifs que les étudiants écrivent sur un des sujets proposés en L2 et disposent généralement de 90 minutes pour rédiger le texte. Comme il n'y a souvent aucune limite du nombre minimum et maximum de mots, les essais collectés varient considérablement en longueur. L'analyse statistique a été faite à l'aide du logiciel SPSS.

⁵ Biber et al. (2011) critiquent la manière dont les chercheurs ont traditionnellement mesuré la complexité syntaxique en L2. Selon eux, pour mesurer la complexité syntaxique de la production écrite en L2, les chercheurs ont utilisé les unités de mesure qui correspondent plus à la langue orale qu'à la langue écrite. Ils trouvent que la langue orale et la langue écrite sont complexes de manières différentes, et par conséquent, il est nécessaire de trouver des mesures adéquates et différentes pour évaluer la complexité syntaxique des deux.

5. RÉSULTATS

Nous allons d'abord présenter les résultats d'analyse du corpus entier des textes d'étudiants de troisième année de français et ensuite les résultats d'analyse du corpus divisé en deux groupes selon la qualité du texte rédigé.

5.1. RÉSULTATS D'ANALYSE DU CORPUS ENTIER

Dans le tableau dessous sont présentés les résultats d'analyse du corpus entier des textes. Pour interpréter les résultats obtenus, il a été nécessaire de les comparer aux résultats d'études similaires à la nôtre, vu qu'il n'existe pas encore une échelle des valeurs proposées pour chacune de ces mesures à chaque stade acquisitionnel. Pour savoir si l'unité T de 14,15 mots est suffisamment longue pour que ce langage soit compris comme complexe, nous nous retournons aux travaux de chercheurs qui ont mené des expériences similaires, avec des mesures similaires.

| | Minimum | Maximum | Moyenne | Écart type |
|-----|---------|---------|--------------|------------|
| w/C | 6.32 | 12.70 | 8.55 | 1.24 |
| w/S | 10.72 | 37.20 | 18.49 | 4.30 |
| w/T | 9.17 | 25.40 | 14.15 | 2.99 |
| C/T | 1.11 | 3.10 | 1.67 | 0.33 |
| T/S | 1.00 | 2.00 | 1.31 | 0.19 |

Tableau 1. Les résultats d'analyse du corpus entier pour les cinq mesures utilisées⁶

Lintunen et Mäkilä (2014) ont mesuré les différences entre la complexité de la langue orale et la langue écrite chez les élèves du lycée dont la L1 est le finnois et dont le niveau de connaissance

⁶ L'abréviation w/c (words per clause) représente la longueur moyenne de proposition; w/S (words per sentence) représente la longueur moyenne de phrase; w/T (words per T-unit) représente la longueur moyenne d'unité T; c/T (clauses per T-unit) représente le nombre de propositions par unité T; T/S (T-units per sentence) représente le nombre d'unités T par phrase.

attendu de l'anglais L2 est intermédiaire (niveau B1 supérieur et niveau B2 inférieur). Pour mesurer la complexité du langage écrit, les répondants ont dû rédiger un essai informel de 150 à 250 mots sur l'un des trois sujets proposés. Les chercheurs concluent que la production écrite des répondants est complexe sur la base des données suivantes: la longueur moyenne de phrase (w/S) est 16,53; la longueur moyenne d'unité T (w/T) est 14,57 et la longueur moyenne de proposition (w/C) est 6,25. Si nous comparons ces données avec les données du tableau ci-dessus, nous voyons clairement la similitude de la longueur moyenne d'unité T, mais aussi la plus grande longueur de proposition et de phrase chez les apprenants de français comme L2 que chez les apprenants d'anglais comme L2.

Wolfe-Quintero et al. (cité dans Lintunen et Mäkilä, 2014) ont comparé les résultats de différentes études sur la complexité et ont obtenu les données suivantes pour différents niveaux de connaissance des élèves : la longueur moyenne des phrases variait de 8,5 mots aux niveaux inférieurs à 23,59 mots aux niveaux supérieurs de connaissance; la longueur moyenne des unités T de 6,0 à 23,0 et la longueur moyenne des propositions de 5,20 à 10,83. Si nous comparons ces données avec les données du tableau ci-dessus, où la longueur moyenne de proposition est 8,55, la longueur moyenne de phrase 18,49 et la longueur moyenne d'unité T est 14,15, il serait clair que d'après les variables la longueur moyenne de phrase et de proposition, la langue des étudiants de français reflète la complexité syntaxique enregistrée à des niveaux de connaissance plus élevés, tandis que selon la variable la longueur moyenne d'unité T leur langue correspondra plus à la complexité enregistrée au niveau de connaissance intermédiaire (la longueur moyenne de 14,15 mots est environ au milieu de l'intervalle 6,0-23,0).

Si nous comparons les résultats qu'a obtenus Monroe (1975)⁷ avec nos résultats, nous pourrions conclure que la langue de nos

⁷ En utilisant la méthode de faire réécrire des textes aux répondants de différents niveaux de compétence en français, il a obtenu les résultats suivants: les étudiants de troisième et quatrième années (w/c 6,43, w/T 9,64, w/S 12,27, c/T 1,50, T/S

étudiants est plus proche de la langue des diplômés de Monroe en termes de variables suivantes : la longueur moyenne d'unité T et de phrase et l'indice de subordination. En ce qui concerne l'indice de coordination de nos répondants, il ressemble davantage à celui du groupe d'étudiants de troisième et quatrième année de Monroe (1,31 contre 1,27). Il est surprenant que la longueur moyenne de proposition de nos étudiants soit supérieure à celle des locuteurs natifs (8,55 contre 7,44). Pourtant, la langue des locuteurs natifs se distingue de celle de tout autre groupe en termes de l'indice de subordination et de coordination. Les locuteurs natifs ont beaucoup plus de propositions subordonnées et moins de propositions coordonnées dans leur production. Nous pourrions présumer que c'est dans cette direction que l'interlangue des apprenants va se développer.

Après avoir comparé les résultats de notre analyse avec les résultats d'études similaires, nous pourrions conclure que la langue écrite de nos répondants est d'une complexité intermédiaire, vu qu'il y a de l'espace pour qu'elle devienne encore plus complexe et proche de la langue des locuteurs natifs.

5.2 RÉSULTATS D'ANALYSE DU CORPUS DIVISÉ EN DEUX GROUPES

Afin de répondre à la deuxième question de recherche, nous allons présenter les résultats d'analyse comparative de production écrite de deux groupes d'étudiants, répartis en groupes inférieur (66 essais – marqué N dans les tableaux) et supérieur (57 essais – marqué V dans les tableaux) d'après l'évaluation de la qualité générale du texte rédigé.

Comme indiqué dans le tableau 2, les étudiants du groupe inférieur ont écrit des propositions d'une longueur moyenne de 8,39 mots, tandis que les étudiants de niveau supérieur ont écrit des propositions légèrement plus longues, avec la longueur moyenne de 8,73 mots. Après le test U de Mann Whitney, la valeur $p=0,163$ ($p>0,05$)

1,27), les diplômés (w/c 7,11, w/T 12,27, w/S 19,29, c/T 1,73, T/S 1,57), les locuteurs natifs du français (w/c 7,44, w/T 17, w/S 17, c/T 2,29, T/S 1).

a été obtenue, ce qui signifie que les différences obtenues entre les deux groupes en termes de longueur moyenne des propositions ne sont pas statistiquement significatives.

| | | Descriptive Statistics | | | | | | | | |
|------|-----------------------|------------------------|-----------|-----------|-----------|----------------|-----------|------------|-----------|------------|
| nivo | | N | Minimum | Maximum | Mean | Std. Deviation | Skewness | | Kurtosis | |
| | | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| N | mean words per clause | 66 | 6.53 | 11.17 | 8.3892 | 1.10955 | .587 | .295 | -.329 | .582 |
| | Valid N (listwise) | 66 | | | | | | | | |
| V | mean words per clause | 57 | 6.32 | 12.70 | 8.7348 | 1.36321 | .867 | .316 | 1.190 | .623 |
| | Valid N (listwise) | 57 | | | | | | | | |

Tableau 2. Différences entre les deux groupes en termes de longueur moyenne de proposition

La longueur moyenne de phrase pour les étudiants du groupe inférieur est 17,84 mots, alors qu'elle est plus grande pour les étudiants du groupe supérieur et s'élève à 19,24 mots (tableau 3). La différence entre le groupe supérieur et inférieur s'est avérée statistiquement significative après le test U de Mann Whitney, avec la valeur $p = 0,035$.

| | | Descriptive Statistics | | | | | | | | |
|------|-------------------------|------------------------|-----------|-----------|-----------|----------------|-----------|------------|-----------|------------|
| nivo | | N | Minimum | Maximum | Mean | Std. Deviation | Skewness | | Kurtosis | |
| | | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| N | mean words per sentence | 66 | 10.72 | 37.20 | 17.8365 | 4.60742 | 1.268 | .295 | 3.593 | .582 |
| | Valid N (listwise) | 66 | | | | | | | | |
| V | mean words per sentence | 57 | 11.91 | 31.12 | 19.2367 | 3.80517 | .642 | .316 | .715 | .623 |
| | Valid N (listwise) | 57 | | | | | | | | |

Tableau 3. Différences entre les deux groupes en termes de longueur moyenne de phrase

La longueur moyenne d'unité T des étudiants du groupe inférieur est 13,64 mots, tandis que l'unité T du groupe supérieur contient en moyenne 14,79 mots (tableau 4). Après le test U de Mann Whitney, aucune différence statistiquement significative n'a été obtenue entre le groupe inférieur et supérieur en termes de longueur moyenne d'unité T ($p = 0,073$).

| | | Descriptive Statistics | | | | | | | | |
|------|-----------------------|------------------------|-----------|-----------|-----------|----------------|-----------|------------|-----------|------------|
| nivo | | N | Minimum | Maximum | Mean | Std. Deviation | Skewness | | Kurtosis | |
| | | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| N | mean words per T-unit | 66 | 9.17 | 23.11 | 13.6439 | 2.61204 | .807 | .295 | 1.304 | .582 |
| | Valid N (listwise) | 66 | | | | | | | | |
| V | mean words per T-unit | 57 | 9.45 | 25.40 | 14.7937 | 3.29824 | 1.273 | .316 | 2.261 | .623 |
| | Valid N (listwise) | 57 | | | | | | | | |

Tableau 4. Différences entre les deux groupes en termes de longueur moyenne d'unité T

L'indice de subordination dans le groupe inférieur est 1,64 et dans le groupe supérieur 1,71. Dans ce cas aussi, il y a une légère différence entre les groupes – le groupe supérieur a plus de phrases dépendantes dans ses essais (tableau 5). Le test U de Man Whitney n'a pas trouvé de différences statistiquement significatives entre les deux groupes en termes de nombre moyen de propositions par unité T ($p = 0,393$).

| | | Descriptive Statistics | | | | | | | | |
|------|---|------------------------|-----------|-----------|-----------|----------------|-----------|------------|-----------|------------|
| nivo | | N | Minimum | Maximum | Mean | Std. Deviation | Skewness | | Kurtosis | |
| | | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| N | mean clauses per T-unit (subordination ratio) | 66 | 1.11 | 2.67 | 1.6369 | .29910 | 1.020 | .295 | 1.684 | .582 |
| | Valid N (listwise) | 66 | | | | | | | | |
| V | mean clauses per T-unit (subordination ratio) | 57 | 1.21 | 3.10 | 1.7080 | .35954 | 1.563 | .316 | 3.534 | .623 |
| | Valid N (listwise) | 57 | | | | | | | | |

Tableau 5. Différences entre les deux groupes en termes de nombre moyen de propositions par unité T

Il est intéressant de noter que les deux groupes ont un indice de coordination presque identique – d'environ 1,31 (tableau 6). Aucune différence statistiquement significative n'a été trouvée entre les deux groupes en termes de nombre moyen d'unités T par phrase. La valeur $p = 0,596$ a été obtenue après le test U de Mann Whitney.

| | | Descriptive Statistics | | | | | | | | |
|------|--|------------------------|----------------------|----------------------|-------------------|-----------------------------|-----------|------------|-----------|------------|
| nivo | | N Statistic | Minimum Statistic | Maximum Statistic | Mean Statistic | Std. Deviation Statistic | Skewness | | Kurtosis | |
| | | | | | | | Statistic | Std. Error | Statistic | Std. Error |
| N | mean T-units per sentence (coordination ratio) | 66 | 1.00 | 2.00 | 1.3053 | .20421 | 1.003 | .295 | 1.297 | .582 |
| | Valid N (listwise) | 66 | | | | | | | | |
| V | mean T-units per sentence (coordination ratio) | 57 | 1.00 | 1.71 | 1.3129 | .16556 | .651 | .316 | -.010 | .623 |
| | Valid N (listwise) | 57 | | | | | | | | |

Tableau 6. Différences entre les deux groupes en termes de nombre moyen d'unités T par phrase

La détermination d'une différence statistiquement significative uniquement pour la longueur moyenne de phrase peut suggérer que les deux groupes ne diffèrent pas significativement en termes de syntaxe, mais que les principales différences entre les groupes se situent à des niveaux langagiers inférieurs à la syntaxe, mais aussi que les mesures utilisées ne sont pas adéquates pour capter ces différences.

Le nombre moyen d'unités T grammaticalement correctes et la longueur moyenne d'unités T correctes sont des unités de mesure qui peuvent indiquer des différences entre le groupe inférieur et supérieur qui ne sont pas forcément liées à la complexité syntaxique, mais aussi à l'exactitude. Ce sont les mesures qui ont été proposées pour remplacer la mesure la plus utilisée – la longueur moyenne d'unité T. Puisque les erreurs sont fréquentes dans la langue des apprenants, l'indice de développement devrait refléter cet aspect de l'interlangue (Gaies, 1980). Les étudiants d'un niveau de connaissance plus élevé devraient produire des unités T correctes plus longues et avoir un nombre élevé des unités T correctes dans leur travail par rapport aux étudiants d'un niveau de connaissance inférieur (Gaies, 1980). Larsen-Freeman & Strom (1977) rapportent aussi une tendance d'augmentation du nombre d'unités T correctes avec l'augmentation du niveau de compétence.

Ici, nous avons décidé que l'unité T correcte serait l'unité T qui contient des erreurs mineures qui sont supposées être le résultat de la performance et non de la compétence. Ce sont des fautes mineures, comme changer la place d'une lettre ou utiliser la mauvaise

lettre, mais seulement lorsque ce remplacement ne donne pas un mot complètement différent. L'absence ou le remplacement de lettres contenant des informations grammaticales importantes était considéré comme une erreur : -s pour le pluriel, -e, -t, -ent, etc. pour la personne, nombre et temps des verbes, puis diverses erreurs de nature morphologique, morphosyntaxique et syntaxique (erreurs de conjugaison, de concordance des temps, d'accord, de choix du temps, d'usage des conjonctions et des prépositions etc.), ainsi que de nature sémantique (choix du mot, registre, néologismes, etc.).

En ce qui concerne le nombre d'unités T correctes, les essais du groupe inférieur contiennent 872 unités T correctes, sur un total de 1532 unités T. Le pourcentage d'unités T correctes dans le groupe inférieur est donc 57 %, tandis que les 43 % restants sont des unités T avec au moins une erreur. Les essais du groupe supérieur contiennent 1124 unités T correctes, sur un total de 1471 unités T. Le pourcentage des unités T correctes dans le groupe supérieur est 76 % et les unités T avec au moins une erreur représentent 24 % du nombre total d'unités T dans les essais du groupe supérieur. Un essai du groupe inférieur contient en moyenne 23,21 unités T, dont 13,21 sont complètement correctes, tandis qu'un essai du groupe supérieur contient en moyenne 25,81 unités T, dont 19,72 sont complètement correctes. Ces données sont résumées dans le tableau 7.

| | Nombre d'unités T correctes | Nombre de toutes les unités T | Pourcentage d'unités T correctes | Nombre moyen d'unités T correctes par essai | Nombre moyen de toutes les unités T par essai |
|---------------------------|-----------------------------|-------------------------------|----------------------------------|---|---|
| Textes de moindre qualité | 872 | 1532 | 57% | 13.21 | 23.21 |
| Textes de haute qualité | 1124 | 1471 | 76% | 19.72 | 25.81 |

Tableau 7. Différences entre les deux groupes en termes de nombre moyen d'unités T correctes

Puisqu'il a été déterminé que la distribution diffère de la distribution normale ici également, le même test non-paramétrique (Mann Whitney U test) a été utilisé pour la vérification si les différences obtenues sont statistiquement significatives. La valeur $p < 0,001$ a été obtenue, ce qui

suggère que les différences entre les deux groupes en termes de nombre d'unités T correctes sont statistiquement significatives.

En ce qui concerne la longueur moyenne d'unités T correctes (tableau 8), ces valeurs ont été obtenues en divisant le nombre total de mots d'unités T correctes par le nombre d'unités T correctes. Dans les essais du groupe inférieur, la longueur moyenne d'unités T correctes est 11,62 et dans les essais du groupe supérieur elle est 13,52. Le tableau ci-dessous montre les longueurs moyennes de toutes les unités T et les longueurs moyennes des unités T correctes pour les deux groupes. Puisqu'il a été déterminé que cette distribution est également différente de la normale, le test U de Mann Whitney a été effectué et une valeur $p < 0,001$ a été obtenue, ce qui dans ce cas aussi suggère que les différences entre le groupe inférieur et supérieur pour la variable longueur moyenne d'unités T correctes sont statistiquement significatives.

| | Longueur moyenne d'unités T correctes | Longueur moyenne de toutes les unités T |
|---------------------------|---------------------------------------|---|
| Textes de moindre qualité | 11.62 | 13.64 |
| Textes de haute qualité | 13.52 | 14.79 |

Tableau 8. Différences entre les deux groupes en termes de longueur moyenne d'unités T correctes

6. DISCUSSION

En ce qui concerne l'ensemble du corpus des textes, l'analyse de la complexité syntaxique à travers les différentes variables quantitatives et la comparaison des résultats avec d'autres études similaires (Lintunen & Mäkilä 2014 ; Monroe 1975) ont révélé que la langue écrite des étudiants de troisième année est modérément complexe et que le développement syntaxique ultérieur va probablement inclure l'allongement de l'unité T, l'augmentation de l'indice de subordination et la diminution de l'indice de coordination, ce qui a été noté dans la langue des locuteurs natifs dans la recherche de Monroe (1975) et dans les recherches sur les itinéraires acquisitionnels en français L2

par Bartning & Schlyter (2004) qui remarquent un passage d'énoncés simples, de juxtaposition et de coordination au niveau débutant à l'apparition de subordination et de structures intégrées et elliptiques au niveau avancé.

En ce qui concerne la répartition des essais dans les deux groupes, les résultats montrent que les différences de complexité existent, même si seulement certaines d'entre elles sont statistiquement significatives. Le groupe supérieur était mieux classé que le groupe inférieur pour toutes les variables quantitatives, bien que des différences dans la complexité syntaxique statistiquement significatives n'aient été trouvées que pour trois des sept variables examinées (longueur moyenne de phrase, longueur moyenne d'unités T correctes et nombre moyen d'unités T correctes). Cela peut signifier que les mesures utilisées ne sont pas suffisamment sensibles pour capter les différences subtiles entre les deux groupes au même stade de développement en L2.

Nous pourrions supposer que les différences dans la production écrite entre les deux groupes n'aient rien à voir exclusivement avec la syntaxe, mais aussi avec des niveaux de langue inférieurs, comme le suggèrent des différences statistiquement significatives entre les deux groupes pour les variables longueur moyenne d'unités T correctes et nombre moyen d'unités T correctes. Ces deux mesures, qui combinent la complexité syntaxique et l'exactitude, suggèrent que les deux groupes diffèrent notamment au niveau morphologique et morphosyntaxique, au niveau du vocabulaire et de l'orthographe, vu qu'il y avait un grand nombre d'erreurs dans les essais concernant la forme des verbes, l'emploi des articles, l'usage des temps, la concordance des temps, l'accord, le choix des subordonnants et des pronoms etc. et que ces erreurs sont beaucoup plus présentes dans les essais classés dans le groupe inférieur. Nous pourrions supposer que les différences morphosyntaxiques et sémantiques soient plus perceptibles au sein des groupes au même stade d'acquisition de L2, tandis que la complexité syntaxique est principalement une caractéristique développementale de l'interlangue, de sorte que des différences significatives dans la complexité syntaxique des textes

peuvent être mieux perçues entre des stades de développement et des niveaux de connaissance clairement distincts, par ex. en comparant les travaux d'étudiants de première, deuxième, troisième et quatrième année.

Lorsqu'il s'agit de répondre à la question de l'interdépendance de la qualité du texte et de la complexité syntaxique, les résultats suggèrent qu'au niveau général, cette relation existe, vu que des essais classés dans le groupe supérieur selon la qualité générale du texte se sont avérés syntaxiquement plus complexes que les essais de qualité inférieure. Pourtant, notre recherche, qui a seulement établi une relation générale entre les deux dimensions, ne peut pas être comparée aux recherches mentionnées à l'introduction (Beers & Nagy 2007; Yang et al. 2015; Kyle & Crossley 2018) parce que ces chercheurs ont analysé la relation entre les mesures spécifiques et individuelles et la qualité du texte pour voir quelle mesure est le meilleur prédicteur de la production écrite jugée de haute qualité.

Une question importante est si les mesures utilisées sont adéquates pour déterminer la complexité syntaxique d'un texte et pour capter les différences entre les deux groupes d'étudiants. En ce qui concerne les mesures strictement liées à la complexité syntaxique, la différence statistiquement significative entre les groupes a été obtenue uniquement pour la variable longueur moyenne de phrase. Bien que des unités grammaticales plus longues signifient souvent une utilisation plus avancée de la langue, parce que plus d'informations sont stockées dans une unité grammaticale, la longueur de phrase ne signifie pas toujours une langue plus complexe ou une langue de meilleure qualité. D'une part, des propositions peuvent être juxtaposées ou coordonnées au sein d'une longue phrase, ce qui ne peut être considéré comme un langage complexe, et d'autre part, des phrases extrêmement longues rendent souvent le texte incompréhensible. La proposition peut aussi être allongé de différentes manières, en élaborant le syntagme nominal ou en ajoutant d'éléments verbaux. Dans les futures recherches, il sera nécessaire de faire une distinction entre les différentes manières d'élaboration d'une unité linguistique et d'utiliser des mesures spécialisées pour chacune de dimensions de la

complexité syntaxique. Pour capter des différences subtiles entre les deux groupes au même stade de développement en L2, des mesures plus fines seront nécessaires. Dans notre recherche, les mesures telles que la longueur moyenne des unités T correctes et le nombre moyen des unités T correctes se sont avérées les plus utiles pour capter les différences entre les deux groupes d'étudiants, ce qui est conforme aux hypothèses de Gaies (1980), et aux études de Larsen-Freeman & Strom (1977) et de Larsen-Freeman (1978).

7. CONCLUSION

L'article analyse d'abord l'ensemble de la production écrite des étudiants de troisième année de français selon des variables quantitatives qui mesurent objectivement la complexité syntaxique. Il a été déterminé que la langue des apprenants est modérément complexe et que son développement ultérieur est susceptible d'aller dans le sens d'un allongement de l'unité T, d'une augmentation de l'indice de subordination et d'une diminution de l'indice de coordination. Ensuite, le lien entre la complexité syntaxique et la qualité générale du texte a été analysé, et ce lien a été confirmé au niveau général, parce que les essais jugés de meilleure qualité étaient aussi plus complexes au niveau de la syntaxe.

Les enseignants de FLE pourraient bénéficier des résultats de cette étude et d'autres recherches sur le développement de la complexité syntaxique à l'écrit et son lien avec la qualité du texte produit parce que ces résultats leur permettront de suivre de près le processus d'acquisition chez leurs étudiants et de savoir s'ils s'approchent des niveaux de complexité et des caractéristiques syntaxiques remarqués aux niveaux avancés de la compétence en L2. Nous avons vu qu'il y a une relation importante entre la complexité syntaxique et la qualité, même si la complexité au niveau de la syntaxe reste seulement un parmi plusieurs facteurs qui contribuent à la qualité du texte. Néanmoins, les enseignants de FLE pourraient insister davantage sur l'aspect syntaxique de la production écrite des étudiants dans le

but d'améliorer la qualité de leurs écrits. Comme le note Crowhurst (1983:10), "un enseignement tel que la combinaison de phrases qui vise – et réussit à – accroître la fluidité syntaxique peut également conduire à une amélioration de qualité d'écriture générale".

RÉFÉRENCES

- Bartning, I. & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2 [Acquisitional routes and stages of development in L2 French]. *Journal of French Language Studies*, 14, 281 - 299. [10.1017/S0959269504001802](https://doi.org/10.1017/S0959269504001802).
- Beers, S. F., Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?. *Read Writ* 22, 185–200. <https://doi.org/10.1007/s11145-007-9107-5>
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45, 5–35.
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & F. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Philadelphia/Amsterdam: John Benjamins.
- Crossley, S. A., & McNamara, D. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66-79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Crowhurst, M. (1983). Syntactic Complexity and Writing Quality: A Review. *Canadian Journal of Education / Revue Canadienne de l'éducation*, 8(1), 1–16. <https://doi.org/10.2307/1494403>
- De Clercq, B. (2016). Le développement de la complexité syntaxique en français langue seconde : complexité structurelle et diversité. *SHS Web of Conferences*, 27, 07006. [10.1051/shsconf/20162707006](https://doi.org/10.1051/shsconf/20162707006).
- Gaies, S. (1980). T-Unit Analysis in Second Language Research: Applications, Problems and Limitations. *TESOL Quarterly*, 14(1), 53-60. [doi:10.2307/3586808](https://doi.org/10.2307/3586808)

- Hunt, K.W. (1965). Grammatical Structures Written at Three Grade Levels. National Council of Teachers of English Research Report No. 3, National Council of Teachers of English, Urbana.
- Kuiken, F. & Vedder, I. (2019). Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *Int J Appl Linguist.* 29: 192– 210. <https://doi.org/10.1111/ijal.12256>
- Kyle, K., & Crossley, S. A. (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. *Modern Language Journal*, 102(2), 333-349. <https://doi.org/10.1111/modl.12468>
- Larsen-Freeman, D. & Strom, V. (1977). The construction of a second language acquisition index of development. *Language Learning*, 27: 123-134. <https://doi.org/10.1111/j.1467-1770.1977.tb00296.xli>
- Larsen-Freeman, D. (1978). An ESL Index of Development. *TESOL Quarterly*, 12(4), 439–448. <https://doi.org/10.2307/3586142>
- Lintunen, P. & Mäkilä, M. (2014). Measuring Syntactic Complexity in Spoken and Written Learner Language: Comparing the Incomparable?. *Research in Language*. <https://doi.org/10.1515/rela-2015-0005>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36-62. <https://doi.org/10.5054/tq.2011.240859>
- Lu, X. & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*. 29. [10.1016/j.jslw.2015.06.003](https://doi.org/10.1016/j.jslw.2015.06.003).
- Monroe, J. (1975). Measuring and Enhancing Syntactic Fluency in French. *The French Review*, 48(6), 1023-1031. Retrieved September 1, 2021, from <http://www.jstor.org/stable/389177>
- Norris, J. & Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics - APPL LINGUIST*. 30. 555-578. [10.1093/applin/amp044](https://doi.org/10.1093/applin/amp044).
- Ortega, L. (2003). Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing. *Applied Linguistics - APPL LINGUIST*. 24. 492-518. [10.1093/applin/24.4.492](https://doi.org/10.1093/applin/24.4.492).
- Yang, W., Lu, X. & Weigle, S. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*. 28. [10.1016/j.jslw.2015.02.002](https://doi.org/10.1016/j.jslw.2015.02.002).
- Wolfe-Quintero K., S. Inagaki & H.-Y. Kim. (1998). The grammatical complexity measures. In Wolfe-Quintero, Inagaki, Kim. *Second Language*

Development in Writing: Measures of Fluency, Accuracy, and Complexity.
(69–99) Honolulu: University of Hawaii Press.

Emilija Milojević

SYNTACTIC COMPLEXITY IN L2 WRITING: THE CASE OF FRENCH

Summary

In this paper we examine the syntactic complexity of the written production of third-year students of French at the Faculty of Philology in Belgrade. Students' essays were collected and analyzed using SPSS. Firstly, the entire sample was analyzed according to the most commonly used variables for objective measurement of syntactic complexity: mean length of clause, mean length of sentence and mean length of T-unit, mean number of clauses per T-unit and T-units per sentence. Then the same variables, with additional ones – mean number of correct T-units and mean length of correct T-units – were used to measure the syntactic complexity of lower and higher-quality papers evaluated by the experienced teacher of French as a foreign language. It was determined that the interlanguage of third-year students is moderately complex. It turned out that the language of students who wrote higher-quality essays is more complex than the language of students who wrote lower-quality essays, but statistically significant differences were found only for some variables: mean length of sentence, mean number of correct T-units and mean length of correct T-units. The two measures that proved to be the most useful in pointing out differences between texts of different overall quality are mean number of grammatically correct T-units and mean length of grammatically correct T-units.

Keywords: syntactic complexity, T-unit, L2 writing, French as a foreign language, second language acquisition, text quality, index of development

*Discourse – From
Conversation to Construction*

Jana Jurčević*
University of Zagreb

A GLIMPSE AT SEMIOTIC LANDSCAPES
AND STUDENT POPULATION OF
ZAGREB'S FACULTY OF HUMANITIES
AND SOCIAL SCIENCES
Socio-pragmatics of Public Toilet Graffiti

Abstract

The main focus of this study are semiotic landscapes of Faculty of Humanities and Social Sciences, at the University of Zagreb, particularly the graffiti found in its public toilets. These graffiti bear a lot of socio-cultural meaning and significance, which is contained in their discursive structures and their social functions. In order to observe the (re)production of ideologies and identity performance in the student population of this specific part of Zagreb's university, the analytical emphasis was put on socio-pragmatic and semiotic features of graffiti in a broader domain of critical discourse analysis. The following four hypotheses emerged from preliminary research: i. toilet graffiti under scrutiny represent group identity of the student population of this faculty, ii. this student population forms a community of practice that mediates between an individual and the society, iii. these graffiti also reveal social and cultural values, problems of politics, and life in general, iv. discourse structure of latrinalia (i.e., toilet graffiti) is mostly dialogical, action oriented, and multimodal. Finally, there were two main goals of this study: primarily to show the interconnectedness of all three levels of social context, and secondly to

* jjurcevic@ffzg.hr

bring to the awareness that we are not bare consumers of space, place, signs and language, but also their creators, interpreters and critics.

Key words: Critical discourse analysis, socio-pragmatics, linguistic ethnography, graffiti, latrinalia, identity, ideology, student population

1. INTRODUCTION

This study focuses on semiotic landscapes (Jaworski and Thurlow 2010), particularly the graffiti found in the public toilets of the Faculty of Humanities and Social Sciences, at the University of Zagreb. These graffiti, also known as latrinalia (Dundes 1966), bear a lot of socio-cultural meaning and significance which is contained in their discursive structures and their social functions (Wales and Brewer 1976, Nwoye 1993, Whiting and Koller 2007). In order to observe the (re)production of ideologies and identity performance in the student population of this specific part of Zagreb's University, the analytical emphasis was placed on socio-pragmatic and semiotic features of graffiti, in a broader domain of critical discourse analysis.

The graffiti in the analysed corpus were categorised on the basis of their content, form, genre, and specific location. The methodology used in the analysis consisted from Grice's cooperative principle (1980), tools from conversation analysis (Sacks, Schegloff and Jefferson 1974), speech act theory (Austin 1976, Searle 1980), and van Leeuwen's social actor theory (1996).

Finally, the interpretation of data has been done in respect to three levels of social context: micro context (the very toilet), meso context (the faculty), and macro context (Croatian and global society).

The main objectives of this study are –

1. Laying down the methodological foundations and theoretical background for investigating discursive and social potential in the agentive power of graffiti.
2. Demonstrating different layers of social reality which are tightly intertwined, and which ultimately affect each other.

3. Highlighting the relevance of communities of practice, and specific speech communities in society.
4. Utilising linguistic ethnography for the purposes of critical discourse analysis and adequate contextualization.

There are four main hypotheses that emerged during the preliminary research, which we aim to demonstrate to be valid on the basis of example analysis. They are the following –

1. Analysed *latrinalia* represents a group identity of the student population at the Faculty of Humanities and Social Sciences in Zagreb.
2. This community of practice (COP) (Wenger, 1998) mediates between an individual and the broader society.
3. It also reveals social and cultural values, problems in politics, and life in general.
4. Its discourse is structured mostly dialogically, action-oriented (filled with speech acts), and is multimodal.

2. THE FIELD OF LINGUISTIC AND SEMIOTIC LANDSCAPES

Investigations of linguistic and semiotic landscapes (LLs & SLs) do not go very far back in the overall history of linguistics, at least not under that name. We could approximately place the emergence and flourishing of scientific interest in LLs and SLs in the late '90s and early '00s, but there are a few caveats to bear in mind. Firstly, the interest in linguistic and semiotic landscapes started with the rise of urbanisation, mass media, and saturation of public places with various signs. Secondly, ethnographic approaches to communication (such as these) have already existed but usually as a part of other traditions (e.g., sociology) or they were ahead of their time and never went beyond theoretical modelling. The best example of such a seminal model is Hymes's ethnography of speaking.

In order to define both linguistic and semiotic landscapes, it is important to see how landscape alone is conceptualised. As seen by Jaworski and Thurlow (2010: 3), landscape would:

“...not to be confined to the mediated representations of space in art and literature. It is a broader concept pertaining how we view and interpret space in ways that are contingent on geographical, social, economic, legal, cultural and emotional circumstances...” (Jaworski & Thurlow, 2010: 3)

Following that, we can also say that the concept of linguistic landscapes, just as well as the concept of landscape alone, has been broadened. The traditional view of the field and the first usage of the term linguistic landscapes (LLs) was given by Landry and Bourhis in 1997.

“The language of public road signs, advertising billboards, street names, place names, commercial shop signs, and public signs on government buildings combines to form the linguistic landscape of a given territory, region, or urban agglomeration.” (Landry and Bourhis, 1997: 25)

Later on, with the rise of interest in the field, and with the divergence of topics, the view and the very definition of the field of LLs started to broaden and change. Apart from focusing mostly on multilingual signs and language policies of different societies reflected in LLs, studies began investigating the links between landscape and identity, social order and power, not only at the institutional level but also from the individual perspective.

The evolution of the field didn't stop here. The final step in forming LLs and SLs can be connected to the emergence of multimodal approaches to communication and discourse analysis. That is of course the domain of semiotics which had to be included in order for the analysis of any (contemporary) landscape to be complete. As Kress and van Leeuwen say:

“In the era of multimodality, semiotic modes other than language are treated as fully capable of serving for representation and communication. Indeed, language, whether as speech or as writing, may now often be seen as ancillary to other semiotic modes: to the visual for instance. Language may now be 'extravisual'. The very facts of the new communicational landscape have made that inescapably the issue.” (Kress & van Leeuwen, 2001: 46)

Following that, we can agree with Jaworski and Thurlow that semiotic landscapes would be “...in the most general sense, any

(public) space with visible inscription made through deliberate human intervention and meaning making.” (Jaworski & Thurlow, 2010: 2)

Finally, it is worth mentioning the ground breaking approach of Ben-Rafael (2009) and his co-authors (2004, 2006) in studying linguistic landscapes of Israel. They brought about a turn in this field by adopting a bottom-up approach. Instead of starting from the institutions (the top), they started their analysis from the micro level, from an individual. They have also taken into account multicultural aspects of globalisation typical for our day and age, and its influence on linguistic and semiotic landscapes. Finally, the authors referred to Erving Goffman’s presentation of self and Pierre Bourdieu’s capital flow in society, which proved to be very beneficial for adequate interpretation of data, and for connecting various levels of analysis.

2.1. CONTEXT OF THE ANALYSED SL – IDENTITY OF ZAGREB’S FACULTY OF HUMANITIES AND SOCIAL SCIENCES AND ITS PUBLIC PERCEPTION

The official website of the faculty¹ provides a formal description of this institution. In its presentation to the public, it states that its mission is to become a national leader and a global partner in producing and passing on the knowledge which shapes a more humane society. To accomplish that, it offers a diversity of study programs in various disciplines, a long tradition in student teaching, and a favourable environment for building an academic career.

The history of Zagreb’s Faculty of Humanities and Social Sciences dates to 1669, making this institution one of the oldest faculties in Europe. Since then, it has been developing in many ways, but in 1948 it went through a big reform and structural reorganisation. Today, the faculty offers 42 undergraduate and 33 graduate programs, and each academic year it enrolls around 1500 new students. Finally, it has more than 700 employees with more than 500 people as academic

¹ Available at URL: <https://web2020.ffzg.unizg.hr/o-nama/misija-i-vizija/> (27th January 2022).

staff, and 600 part-time lecturers. This alone makes it the biggest faculty of the University of Zagreb, and in the whole country.

Since 1961 the faculty has been located in a building of 15 616 m², placed in the broader centre of Zagreb. Offices of professors and departmental administration are spread throughout the corridors in the basement and on three upper floors of the building. At the same time the biggest lecture rooms (D1-D6) are placed on the ground floor, in two different wings of the building. Smaller lecture rooms on the first and the second floor can be found mostly in tract A, while the third floor has the additional E and F tract.

Because the focus has been placed on the student population and public toilets, the most important locations will prove to be the ground floor toilets, where an abundance of graffiti can be found.² We presume that is the case because of the biggest circulation of people on the ground floor of the main building, where, apart from the biggest lecture rooms, there is the students' club³, the cafeteria⁴, the new library⁵ entrance, the faculty administration (dean's office, human resources etc.)⁶, and a bookstore.

It was those reasons that led us to focus exclusively on the ground floor of the main building, and its public toilets. We haven't analysed

² Red oval shapes show the locations of public toilets at the faculty.

³ On a layout shown in Figure 1, the students' club is marked with a green ellipsis. It is located on the ground floor, between the entrance and lecture room D4, on the left side of the building. Its position and activities held there are quite significant for graffiti production in this area. The club serves different beverages at the bar, has an area for reading, playing chess and cards. Club also organises different types of social events such as student parties, poetry reading, small concerts etc. Their main principle is volunteering and their motto is "I read. I think. I turn the world upside down." It is important to note here that the unisex toilet nearest to the club has the most elaborate graffiti and (probably) the most graffiti in general, in comparison to all the other toilets in the building.

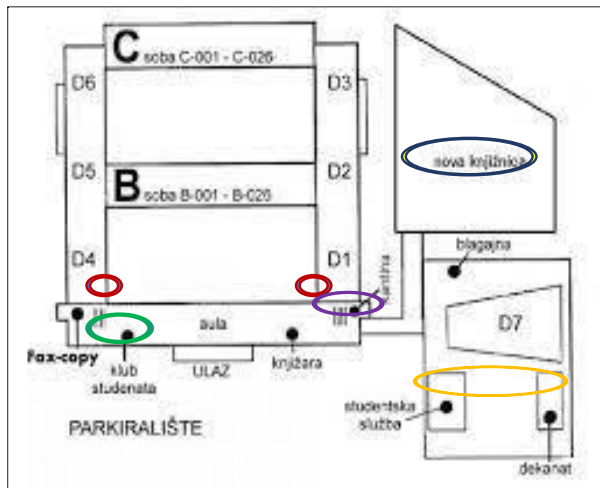
⁴ Cafeteria is marked by a purple ellipsis and is located in the basement of the building.

⁵ The new library is an adjacent building marked with a yellow ellipsis.

⁶ The faculty administration is marked with an orange ellipsis.

toilets in the library because they are (almost) devoid of latrinalia, as well as the toilets for the staff.

Figure 1. Ground floor of the faculty



Apart from the identity of the faculty provided by the institution itself, it is also significant to see the students' perspective and their role in its making.

Everywhere in the world, the student population has always been a motivating force in society; the avant-garde in fighting for freedom and equality. A rebellious force criticising society and trying to solve political problems. This is especially the case for the Faculty of Humanities and Social Sciences, which played a large part in Croatian politics. For instance, the Croatian spring movement in 1971. However, the crucial year in the recent history of the faculty was the year 2009. It was the year of students' protest and their occupation of the faculty in their fight for the right of free education. Some of the graffiti made in that period still stand in the toilets of the faculty, so the following summary of the events will also provide us with a context for the SL analysis.

It all started in the spring of 2009, with the students' occupation of the faculty building as a protest against the Croatian government's

plans to reduce public funding for higher education. This lasted for 35 days at Zagreb's Faculty of Humanities and Social Sciences, but it also spread to the University of Zadar, as well as other bigger cities, and their faculties or universities. During the occupation, everyone was free to enter and leave the building but regular classes were not held. Instead, the students organised an alternative program consisting of lectures, public discussions, workshops, film screenings, etc. At the time, the central body for official decision-making was set up to run the faculty and negotiate with the government. It was an assembly of all interested students and citizens who had equal rights to speak and vote on discussed matters. All decisions were made in a direct democratic manner, by public voting at a daily meeting called Plenum. Hence the graffiti "Go to Plenum!". The second occupation of the faculty was somewhat shorter than the first one, as it lasted for two weeks in the fall semester of 2009, and was terminated by the Plenum votes.

After the protests, a new model of tuition payment in higher education was adopted. It has been decided that the enrolment in the first year of studies would stay free, but the percentage of participation in paying tuition in the following years was to be evaluated according to the student's credits (ECTS).

These were the most famous, most effective, and the longest protests of the student population in the recent history of the faculty. These events have left a mark on the group identity of this population, as well as a mark on the public perception of the institution as a whole. Because of that, we believe that this short overview gives a good insight into a group of people collected around the same goals, life views, and values. This provides us not only with a context for graffiti interpretation but also with the fact that here we are indeed dealing with a specific community of practice (COP), as well as a specific group identity and its discourse. In short, we can say that this group consists of young people, with mostly liberal, sometimes libertarian, and leftist political inclinations. They usually come from middle-class families, and many of them are connected to different youth subcultures.

Finally, the public view of this institution, the perspective of a broader society, of the “outsiders” also adds to the context. This perspective can most easily be spotted in media discourse, but also in some of the graffiti (as we shall see). When discussed publicly and informally, it is not uncommon to hear a stance that people from the Faculty of Humanities and Social Sciences in Zagreb are commies, lazy people, party animals, and hippies. But also, that they are the avant-garde and freedom fighters who will stand aside the working class and the oppressed.

2.2. COMMUNITIES OF PRACTICE AND SPEECH COMMUNITIES

In the previous section, we have noted that the student population of the Faculty of Humanities and Social Sciences in Zagreb can be considered a specific community of practice. It has also been shown how this community revolves around a specific group identity. But can we consider this group also as a specific speech community? We strongly believe that we can, because this very speech has its own distinctiveness and functions combining youth culture jargon with academic discourse, with common ground and common goals in the background of communication processes.

Community in general can be identified in terms of space, place, affiliation, practices, and any combination of these. The term itself is usually used to designate a social unit larger than a household. It can also refer to national, international, and/or virtual (online) groups.

Community of practice (COP) was elegantly defined by Etienne Wenger (1998) who thought of them as aggregates of people that come together around mutual engagement in some common endeavour. Practices of COPs members (ways of talking, beliefs, values, power relations) emerge from their joined activities. Some examples of COPs would be co-workers, regular guests at a pub, a nuclear family, and it seems also – the student population. Furthermore, COPs can be large or small, long-lasting, or short-lived, and can be deeply embedded in other communities. We should also bear in mind that an individual

usually participates in multiple communities of practice, and that their individual identity is in turn shaped by all of them. Finally, we can agree with the authors Eckert and McConnell-Ginet (1992: 9) who claim that:

“The community of practice is where the rubber meets the road – it is where observable action and interaction do the work of producing, reproducing and resisting the organisation of power in society...”

A speech community is distinguished by a shared system of interaction and symbols, and its users are aware that they share it amongst themselves. These communities usually become more self-aware when there is a crisis, but in such circumstances, they also come into collective awareness. It happens when they themselves are under a threat or when they pose a threat to dominant power structures. The greatest names associated with the concept of a speech community were Dell Hymes and John Gumperz. Hymes (1962) defined a speech community as a community sharing knowledge of rules of conduct and interpretation of speech. He also noted that a speech community should be considered a fundamental concept in sociolinguistics because it serves as a relation between language, speech, and social structure. Gumperz (1968: 219) thought of speech communities as social constructs and defined them as:

“...any human aggregate characterised by regular and frequent interaction by means of shared body of verbal signs and set off from similar aggregates by significant differences in language usage.”

Today we can distinguish between two main approaches to the speech community analysis. The first one is sociolinguistic, investigating semantic, pragmatic, and conversational features of language. The second one follows the line of critical discourse studies and sociology of language focusing on ideology (re)production, identity, and power relations. In this study, we have opted for both, because they complement each other well and together give a better insight into the problematics. This combination of critical linguistic and ethnographic approaches resulted in methodological bricolage

which enveloped the aforementioned tools and concepts adopted from various authors.

2.3. GRAFFITI AND LATRINALIA

The word graffiti comes from the Italian term *graffito* meaning a small engraving, a scratch (Al-Khawaldeh, et al., 2017). It is a very broad term that labels various types of non-institutional but intentional inscriptions or images on public surfaces, which (re)produce meaning. Furthermore, Abel and Buckley (1977: 3) suggest that graffiti are a form of communication that is both personal and free of the everyday social strains that normally prevent people from giving uninhibited reign to their thoughts. Finally, Chilwa (2008: 274) describes graffiti as:

“...any form of writing or images on the walls or surfaces of public buildings, parks, toilets, buses or trains, usually bearing some political or sexual contents ...”

Observing graffiti as a particular genre, it is inevitable not to notice their hybrid nature. They are composed of written language with characteristics of speech, and this is sometimes accompanied by drawings, stickers, and the usage of different colours. Finally, the fact that this type of communication is mediated (through public surfaces) makes it anonymous, quite democratic, and dialogic.

There are several types of graffiti, but the most known are tags and toilet graffiti. Tags are usually short graffiti signatures that are used as urban territorial markers. They are typical for the hip-hop subculture, and they provide spatial orientation and build group identity.

On the other hand, toilet graffiti is mostly used for communicating thoughts, beliefs, feelings, social status, and social roles. Alan Dundes, an anthropologist at the University of California, Berkeley, coined a specific term for this type of graffiti: *latrinalia*, expanding on another English word for a toilet – a *latrine*. It first appeared in his study, published in 1966 called “Here I sit – A study of American *latrinalia*”, in which he analysed public toilet graffiti and gave valuable insight into functions and the significance of this type of graffiti. Dundes also gave

a useful typology of different forms of latrinalia which usually appear as advertisements, commands or demands, instructions, comments or personal introspective musings.

Since Dundes, many other scholars have studied this topic from different perspectives and came to various findings. For instance, Gadsby (1995) studied humour in graffiti, while Sheivandi et al. (2015) observed the usage of alliteration, rhyme and puns as attention gaining mechanisms. Stocker et al. (1972) hypothesised that graffiti are accurate representations of consensual values of a community in which they are made. Opposing that theory, Gonos et al. (1976) suggested that the content and the frequency of graffiti can vary depending on the current relevant and dominant values of society. It is also fair to mention some of the studies considering gender roles. Kinsey and associates (1953) were one of the first who analysed specifically latrinalia of sexual content and found that men are their main producers. In 1975, Far and Gordon repeated this study and had very interesting finds. They have noticed an increase in graffiti in women's toilets from 25 to 44%. Still, male latrinalia with sexual content stays prevalent.

Modern-day investigations of latrinalia and graffiti in general, tend to combine content analysis and critical discourse analysis. We will follow this trend and try to build on it with several useful tools from socio-pragmatics, linguistic ethnography, and semiotics.

3. METHODOLOGY

Our methodology represents a bricolage, a patchwork of analytical tools of previously mentioned fields:

- Critical discourse analysis (CDA) provided us with an insight into background ideologies and power relations.
- Socio-pragmatics contributed with information on identity performance (group/individual), institutional influences, meaning interpretation and production, and contextualization.

- Linguistic ethnography proved valuable in data collection and systematisation.
- Semiotics was required for the analysis and interpretation of the multimodal aspects of latrinalia.

These analytical tools in turn revealed the aforementioned discursive structures. Those will be shown in the examples.

3.1. DATA COLLECTION AND CATEGORIZATION

Data was collected during the fall semester of the academic year 2019/2020. Graffiti was documented by photography. Toilets intended for working staff were not investigated, as well as the ones in the library (which is a new adjacent building).

There were no problems with data collection concerning institutional interventions such as cleaning/removing the graffiti. Some may see the institutional stance towards latrinalia as positive and enticing, while others might think of it as neglect. We'd like to believe it is the former, not the latter, or that it has at least evolved from being seen as vandalism to a valuable form of expression and communication.

Data was categorised on the bases of four criteria because of the need for a deeper qualitative analysis, and more accurate interpretation of findings:

1. **Content** revolved around the topics of sex, politics, health, culture, identity, gender, etc. As categorization criteria, content, as well as meaning, told us a lot about the culture and the society in which it has been produced.
2. **Location** category was concerned with the position of the toilet – which floor is it on, how far is it from the entrance, the students' club, or the library, and which gender it is intended for. This was relevant because it provided the information on the frequency of toilet usage as well as its users/population. Here, the position of the very graffiti in the toilet was also taken into account, which pointed us to the (approximate)

time of graffiti production (centrally positioned graffiti are usually older than the ones on the periphery of a wall or a door). Their placement can also help with the interpretation of implicatures and indirect speech acts.

3. **Modality** category referred to the type of script and language, colours, drawings, or stickers used in graffiti formation. Implementation of different modalities can suggest the writer's intentions, and insinuate something about their identity and values.
4. **Genre** (e.g., commentary, dialogue, advertisement) informed us about the microphysics of power fluctuating through discourse, and the positioning of the social actors in the overall social network. It also informed us of the author's motivation for a specific graffiti production.

Apart from data categorization, different types of **discursive structures** (of latrinalia) revealed their dialogic nature, the relevance of illocutionary force and type of a speech act (Austin 1976, Searle 1980), and the importance of Grice's cooperative principle (1980). The following structures will also be shown through the previously selected examples from a larger corpus, where we will see:

- a) latrinalia with no answer,
- b) question/statement latrinalia with an answer(s),
- c) modification of the existing graffiti (addition, replacement, deletion).

These structures roughly correspond to the findings of conversation analysis (Sacks, Schegloff and Jefferson 1974).

Finally, considering data it is inevitable to mention their creators. The types of **actors** that we see here have two main qualities. They are usually anonymous (unnamed, rarely named (van Leeuwen, 1996) and equal in their right and possibility to participate in the communication. Those attributes will structure a network, not a hierarchy but a very specific community of practice.

This study hasn't dealt with quantitative data analysis because our research goals as well as the hypotheses benefited more from

qualitative analysis and critical data assessment. It is quite important to note here that counting graffiti can be very complex and potentially misleading. Why is that? Primarily, graffiti are constantly going through the process of change – they are being erased, rewritten and modified as we speak, so it is difficult to tell how many of them there really are. Secondly, if we do not know their authors, or at least the population associated with their production the very numbers won't mean much. Thirdly, we believe that quantitative research on graffiti production might be more useful if done as a longitudinal study, which would show the fluctuations in graffiti usage through time. But that will be left for some future endeavours.

4. ANALYSIS AND EXAMPLES

Because of the complexities of discursive structures of latrinalia, the analysis was conducted on three different levels, the level of -

Micro context – the very toilet (and the stall) at the faculty, with the focus on the individual, personal identity performance, communication practices from an anonymous position.

Meso context – the context of the Faculty of Humanities and Social Sciences as a community of practice (COP), with its ideologies, beliefs and values.

Macro context – Croatian and global society, exemplified in latrinalia in topics of religion, politics and culture (usually quite intertextual).

The following examples and their shortened analysis have been extrapolated from the overall corpus analysis as the most representative ones, in order to illustrate the discursive processes played out in the background. Due to the limited space provided by the genre of scientific article, it was not possible to present the entire data analysis of the collected corpus. Instead we have opted for the selection of the examples, nevertheless it is possible (for anyone interested) to access the original data. For all the readers who are interested, there is a link for Instagram profile which features toilet graffiti from Faculty of

Humanities and Social Sciences in Zagreb – <https://www.instagram.com/ffzgw/?hl=hr>.

The profile is a public profile, so anyone registered on Instagram can view documented material. Its maker and moderator is anonymous. Finally, everybody is invited to contribute with their own photos of graffiti from that place. For now, the profile contains 939 posts/photographs of latrinalia and has 2307 followers.

4.1. MICRO LEVEL ANALYSIS AND EXAMPLES

Example No #1 –

Statement: *“I have a girlfriend for already a year and all my friends think I’m straight.”*

Answer: *“You obviously have no real friends.”*

Considering the content of this graffiti, it is evident that it is a personal confession concerning sexual orientation. Discursive structure of this latrinalia came in the form of a statement (which is a representative speech act) with an answer (which is a verdictive speech act). Actors were anonymous which opened the space for less tact and politeness in communication. Although there is no emotional support demonstrated in this exchange, Grice’s cooperative principle with its maxims has been adhered to in terms of relevance, modality, quantity and perhaps even quality. This exchange was monomodal because it relied only on writing, while the genre of the graffiti – a dialogue, as well as its location – the unisex toilet on the first floor, revealed the networked social structure of this community and equality of its members.

Example No #2 –

Statement: *“I’m drunk and I have to go to class.”*

No answer(s).

Content of this graffiti would roughly be a representation of an inner state and outer state of affairs. Both clauses here come in a form of speech acts – representatives. The topic in question can be

characterised as substances and student life. The graffiti was located in the unisex toilet on the first floor, but from the Croatian original we can infer that the author is female. The structure of the graffiti is a statement with no answer, in the terms of a genre it could be categorised as a micro narrative, which is monomodal – relying only on written language.

Example No #3 –

Statement: *“I’m vegan.”*

Multiple answers: *“Me to! – So am I. – I’m a waiter. – I’m Italian. – I’m an atheist. – And I’m writing on the doors. – And I eat little animals. – I don’t give a damn.”*

This would be a typical example of presentation of self, in the form of a statement with multiple answers. The very statement can in turn be characterised as a representative speech act which provides us with the information on someone’s life philosophy⁷. The structure of this latrinalia is very rich in terms of conversation analysis and cooperation principle. It is obvious we have multiple anonymous actors participating in this interaction, exchanging turns, building on other people’s statements and staying relevant or irrelevant in regards to the topic. Genre we are dealing with here would be a dialogue and communication monomodal, relying only on script.

4.2. TRANSITION FROM MICRO TO MESO LEVEL – EXAMPLES

Example No #4 –

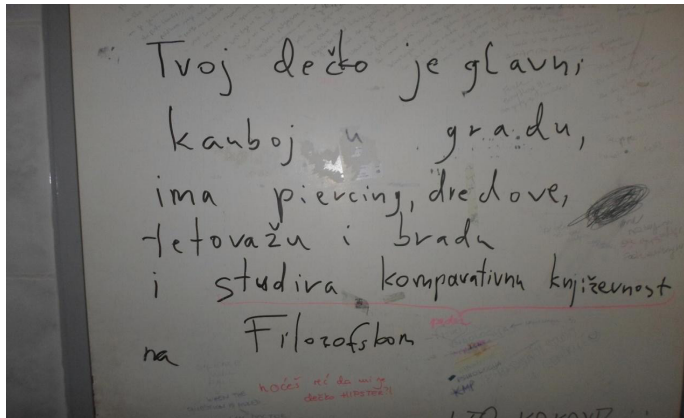
Short narrative: *“Your boyfriend is the main sheriff in town, he has a piercing, dreads, a tattoo and a beard, and he studies Comparative literature at the Faculty of Humanities and Social Sciences.”*

Answers/Comments: *“Are you telling me my boyfriend is a hipster?”*
“Gay” – written underneath a curly bracket marking *“...studies comparative literature”*.

⁷ Here you can also notice the mixing in of the meso level, e.g., veganism & LGBTIQ problematics as ideologies typically supported by this Faculty’s student population.

The example describes and ironizes a stereotype of a student from this faculty. It also demonstrates the awareness of a specific community and its perception from the broader public. In terms of speech acts, these would all be representatives. From the following picture of this graffiti in figure 4, it is obvious that this instance of communication was monomodal (i.e., relying only on script). It was also dialogic because of the interventions made around that graffiti, most prominent ones written in red and pink, and listed in the example as answers/comments.

Figure 4. Example 4



Example No #5 –

Statement: “*You doodled all over the place, you sons of b***es!*”⁸

No answer.

The example comes in the form of a verdictive speech act, and is a great example of self-irony. It (again) demonstrates an awareness of a group identity. It is monologic but intended for this specific audience to read, it is anonymous, agentive and more similar to speech rather than writing.

⁸ This has been censored by the authors of this study. In the original graffiti there are no asterisks.

Example No #6 –

Statement: “I survived my first presentation <3”

Answer: “Congratulations”

The example shows a representative (the statement) and an expressive speech act (the answer). Goal accomplished here would be group cohesion via support, and the topic would (again) be the one of a student life. Considering the modality of this latrinalia, we could say it shows some signs of multimodal correspondence because of the usage of heart emoticon which conveys the mood of the speaker in mediated communication. This is also a good example of transferal of writing conventions from cyberspace into the physical realm of correspondence.

4.3. MESO LEVEL ANALYSIS AND EXAMPLES

Example No #7 –

Statement: “I love the secular Faculty of Philosophy”

No answer.

This example shows recurring graffiti throughout the faculty's toilets. It comes in the form of an assertive speech act, and it points to a specific ideology, the one of secularism which is typical for this COP.

Figure 5. Example 7



Example No #8 –

Order/instruction: “*Go to plenum*”

Statement: “*One world, one fight. Knowledge isn’t a commodity!*”

Date and signature: *15th May ‘09 Luci*”

This graffiti shows a directive speech act (the order to go to plenum) and a slogan (which follows), appealing to group identity values and it can be considered a call for action. The novelty of this latrinalia lies in the fact that it is not anonymous, and on top of that it has been dated, so we know the exact day of its making – it was during the time of students’ protest for the right to free education, in the spring of 2009.

Example No #9 –

Statement: “*Commies!*” No answer.

Statement: “*ANTI-FA*” No answer.

Statement: No answer.

Statement: “*Anarchy*” No answer.

These 4 different yet similar graffiti demonstrate ideologies and values of this population. There were no answers or modifications written beside or over the graffiti, all of them are anonymous and all of them can be seen as representative speech acts. They mostly rely exclusively on script, but there is also one symbol implemented here – the piece symbol, which leads us to the domain of semiotics.

Example No #10 –

Statement (on a red sticker(s) at the toilet paper dispenser and a flusher saying):

“*Now you know what to do with your diploma.*”

No answer.

In this example we can see a successful utilisation of multimodality and the usage of implicatures by breaking some of the maxims of Grice’s cooperative principle. Here, the very location or placement of these stickers is of great importance, because it contextualises this indirect speech act and it enables its intended interpretation. The topic is again the one of a student life.

4.4. MACRO LEVEL ANALYSIS AND EXAMPLES

Example No #11 –

Statement: “*Jesus is coming, pretend you’re working.*”

No answer.

The graffiti in question has a strong intertextual element. It alludes to Croatian, predominantly Christian society and its hypocrisy. We can also see the usage of a representative and a directive speech act. Main topics of this latrinalia would be ideology and religion, but in rhetorical terms there is irony as well as pronounced criticism.

Example No #12 –

Statement: “*I’m just watching you*” + a big drawing of an eye + a tag.

No answer.

This example comes in the form of a representative speech act. Besides from that we have utilisation of intertextuality (an eye alluding to the Panopticon, or Orwell’s big brother) and there is also a multimodal aspect to this because of the act of combining text with a drawing and with a tag. The example is shown in figure 6. Finally, although it doesn’t have answers or comments around it, this latrinalia is still addressing a specific audience and interacting with it.

Figure 6. Example 12



Example No #13 –

Statement/declaration: *“In the name of Croatian Democratic Union and Social Democratic Party of Croatia, I christen this toilet bowl and the Constitutional court of the Republic of Croatia.”*

No answer.

Here we can see the usage of a declarative speech act, but not with the intention of changing reality but with the intention of ridiculing it (satire). There was no actual christening act that happened there, because the person doing the act didn't have the authority for that (i.e., he wasn't a priest), following from that there was no illocutionary force in the act and no perlocution. Instead, the satirical element has been realised and it brought with it a criticism dressed in humour. The entire statement is also quite intertextual (a reader needs to be familiar with Croatian politics and culture to interpret it correctly).

No #14 –

Statement: *“The only good system is a sound system.”*

Answer: *“What about the solar system U Hippy c**t?”*

The final example comes in the form of a statement/answer. In terms of speech acts, the first statement is an assertive while the answer is a verdictive speech act. By its content we can presume it is connected to youth subcultures (i.e., the sound system problematics and hippies), but it surpasses them by a witty answer which comes in the form of a rhetorical question and invokes the topic of environmentalism. Finally, we have the usage of a vulgarism which gives this discourse an informal overlay and insight to micro dynamics of power, which obviously still exists in this non-hierarchical, networked social structure.

5. INTERPRETATION AND DISCUSSION – POWER STRUCTURES

Structures of (social) power are reflected in the structures of discourse. This is also applicable to the case of latrinalia.

As we have seen –

1. Speech act type and its illocutionary force can point us to an authority which legitimises social action, moreover it shows us the agentive power of language by means of its perlocution (i.e., its consequences upon reality).
2. Responses and discussions between actors show us dominant ideologies in this community as well as its power structure, which is primarily networked, not hierarchical.
3. Deletion and addition of existing latrinalia can also be interpreted as a (speech) act of silencing which conveys refusal, resistance and disagreement. It also reveals a ludic and dialogic nature of this type of graffiti.

Anonymity has also proven to be extremely important. Mediated nature of this type of communication has made it unburdened by social and auto-censorship and free from social sanctions when breaking the norms. It has made communication via latrinalia democratic: whoever wants can become its active participant. Finally, because of its public and democratic character, latrinalia is also quite dialogic (not monologic as mass communication).

From all of that it is easy to see the importance of context in meaning formation, communication and interpretation. But to be more precise, these are the advantages of the analysis conducted at three different levels of social context:

1. Micro level analysis was necessary because of the anonymity and freedom of expression given by a toilet stall.
2. Meso level analysis provided us with the information on the ideological values of the institution (where toilets are located) and its presentation in the broader society.
3. Macro level examples gave us insight to the common knowledge, values and problems of a specific society, as well as its current trends (intertextuality of latrinalia).

Finally, we have noticed several gender differences which are notable to mention, but also require further investigation. It has been witnessed that unisex and women's toilets have the most latrinalia in

general. We presume that is the case because they are frequented more than the other toilets because of their placement in the faculty building and the mixing of genders in the unisex toilet. Graffiti in these toilets also present multimodality, a broader range of topics and a more liberal discourse.

On the other hand, latrinalia is rare in men's toilets in comparison to unisex and women's toilets. Discourse is much more aggressive which presents in insults, directive speech acts, curses, vulgarisms, homophobia and nationalism. The topic of sports is also quite common, namely soccer.⁹

6. CONCLUSION(S)

There are several conclusions that revealed themselves upon this study.

Firstly, we have witnessed that schools, as well as universities form *speech communities* (Hymes 1962, Gumpertz 1968) and/or communities of practice (COP) (Wenger 1998). Secondly, ideologies, values and group identities are communicated through toilet graffiti (apart from the more official channels such as newspapers, journals, pamphlets and flyers). Furthermore, those speech communities mediate between an individual and a broader society. They also reveal broader social and cultural values of a society, as well as its problems and power struggles. On top of it all they present a niche for marginalised groups to have their own say (e.g., LGBTQIA community). Finally, as a discourse type, latrinalia can be described as democratic, dialogic, multimodal, public yet anonymous form of mediated communication.

⁹ Observations concerning gender differences were made on the basis of a larger corpus of data collected, not just on the basis of the examples presented in this paper.

6.1. FUTURE PROSPECTS

Linguistic ethnography in combination with critical discourse studies, semiotics, sociolinguistics and pragmatics has demonstrated to be a very beneficial concoction for getting insights into the interface of communication and society. From that interface, many contemporary problems emerged, as well as their solutions.

Today, it almost seems as a given that our next endeavour should be the investigation of cyber landscapes and their linguistic, semiotic and social aspects. That comes hand in hand with dealing with globalisation trends in LLs and SLs (e.g., the domination of English language and language hybrids such as *Crenglish* – Croatian + English).

Finally, all future studies can be complemented with quantitative methods (e.g., counting graffiti of specific content), but also with polls/questionnaires and interviews, depending on the research goals of the author.

7. REFERENCES

- Abel, E. and B. Buckley (1977). *The Handwriting on the Wall: Toward the Sociology and Psychology of Graffiti*. Westport, CT: Greenwood.
- Al-Khawaldeh, N. N., I. Khawaldeh, B. Bani-Khair and A. Khawaldeh (2017). "An Exploration of Graffiti on University's Walls: A Corpus-based Analysis study", *Indonesian Journal of Applied Linguistics*, 7(1): 29–42.
- Austin, J. L. (1976). *How to do Things with Words: The William James Lectures delivered at Harvard University in 1955*. Oxford: OUP.
- Ben-Rafael, E., E. Shohamy, M. H. Amara and N. Trumper-Hecht (2004). *Linguistic Landscape and Multiculturalism: A Jewish–Arab Comparative Study*. Tel Aviv: Tami Steinmetz Centre for Peace Research.
- Ben-Rafael, E., E. Shohamy, M.H. Amara, N. Trumper-Hecht (2006). 'Linguistic landscape as symbolic construction of the public space: The Case of Israel', *International Journal of Multilingualism*, 3(1): 7–30.
- Chiluwa, I. (2008). "Religious Vehicle Stickers in Nigeria: A Discourse of Identity, Faith and Social Vision". *Discourse and Communication*, 2(4): 371–387.

- Daniels, S. and D. Cosgrove (1988). "Introduction: Iconography and Landscape". In Cosgrove and Daniels (eds), *The Iconography of Landscape*. Cambridge: CUP, pp. 1–10.
- Dundes, A. (1966). "Here I sit: A Study of American Latrinalia", *Kroeber Anthropological Society Papers*, 31: 91–105.
- Gadsby, J. M. (1995). *Looking at the Writing on the Wall: A Critical Review and Taxonomy of Graffiti Texts*. Unpublished master's thesis at the University of Boston.
- Gonos, G., V. Mulkern and N. Poushinskiy (1976). "Anonymous Expression: A Structural View of Graffiti". *Journal of American Folklore*, 89(351): 40–48.
- Grice, H. P. (1980). „Logika i razgovor." U: Aleksandar Pavković (ur.), *Svest i saznanje: Ogleđi iz savremene analitičke filozofije*. Beograd: Nolit.
- Gumperz, J. (1968). "The Speech Community". In A. Duranti (ed.): *Linguistic Anthropology: A reader*, 1: 66–73.
- Hymes, D. (1962). "The Ethnography of Speaking." In: Gladwin & Sturtevant (eds), *Anthropology and Human Behavior*. Washington DC: Anthropological Society of Washington, pp. 13–53.
- Jaworski, A. and C. Thurlow (2010). "Introducing Semiotic Landscapes". In Jaworski & Thurlow (eds), *Semiotic Landscapes: Language, Image, Space*. London; New York: Continuum, pp. 1–40.
- Kinsey, A. C., W. B. Pomeroy, C. E. Martin and P. H. Gebhard (1953). *Sexual Behaviour in the Human Female*. Philadelphia: W.B. Saunders.
- Kress, G. and T. van Leeuwen (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- Landry, R. and R. Y. Bourhis (1997). "Linguistic Landscape and Ethnolinguistic vitality: An Empirical Study". *Journal of Language and Social Psychology*, 16: 23–49.
- Leeuwen, T. van (1996). "The Representation of Social Actors". In Carmen Rosa Caldas-Coulthard & Malcolm Coulthard (eds.): *Texts and Practices: Readings in Critical Discourse Analysis*. London: Routledge.
- Nwoye, O. G. (1993). "Social Issues on Walls: Graffiti in University Lavatories", *Discourse and Society*, 4 (4): 419–442.
- Sacks, H., E. A. Schegloff and G. Jefferson (1974). "A simplest systematics for the organization of turn-taking for conversation", *Language*, 50: 696–735.
- Searle, J. (1980) *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: CUP.

- Sheivandi, L., A. Taghinezhad, A. Alishavandi and S. Ranjbar (2015). "Exploring the Linguistic Aspects of Iraninans' Graffiti". *Journal of Applied Linguistics and Language Research*, 2(5): 62–73.
- Stocker, T. L., L. W. Dutcher, S. M. Hargrove and E. A. Cook (1972). "Social Analysis of Graffiti", *Journal of American Folklore*, 85(338): 356–366.
- Wales, E. and B. Brewer (1976). "Graffiti in the 1970's", *Journal of Social Psychology*, 99: 115–123.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: CUP.
- Whiting, S. and V. Koller (2007). „Dialogues in solitude: the discursive structures and social functions of male toilet graffiti". Working paper No. 126, Centre for the Study of Language in Social Life, Lancaster University.

Jana Jurčević

EIN BLICK AUF SEMIOTISCHE LANDSCHAFTEN UND
STUDENTEN POPULATION DER FAKULTÄT FÜR GEISTES- UND
SOZIALWISSENSCHAFTEN AN DER UNIVERSITÄT ZAGREB –
Soziopragmatik von Graffiti auf öffentlichen Toiletten

Zusammenfassung

Die Studie konzentriert sich auf die semiotischen Landschaften der Universität Zagreb, insbesondere auf die Graffiti, die in öffentlichen Toiletten der Fakultät für Geistes- und Sozialwissenschaften gefunden wurden.

Diese Graffiti, auch bekannt als Latrinalia (Dundes 1966), tragen eine große soziokulturelle Bedeutung, die in ihren diskursiven Strukturen und sozialen Funktionen enthalten ist. Um die (Re-)Produktion von Ideologien und Identitäten Leistungen in dieser Studenten Population zu identifizieren, wurde der analytische Schwerpunkt auf soziopragmatische Merkmale von Graffiti in einen breiteren Bereich der kritischen Diskursforschung und der sprachlichen Ethnographie gelegt (Hymes 1968).

Graffiti wurden nach Inhalt, Form, Genre und spezifischen Ort kategorisiert. Die in der Analyse verwendete Methodik bestand aus Genossenschaftsprinzipien (Grice 1980), Werkzeugen aus der Gesprächsanalyse (Sacks, Schegloff und Jefferson 1974) und der Sprechakttheorie (Austin 1976, Searle 1980).

Die Interpretation der Daten erfolgte in Bezug auf drei verschiedene Ebenen des sozialen Kontextes: Mikroebene (die Toilette), Mesoebene (die Fakultät) und Makroebene (Gesellschaft).

Die vorläufige Nachforschung gab folgenden Hypothesen: i. die analysierte Latrinalia darstellt die Gruppenidentität von Studenten Population der Fakultät für Geistes- und Sozialwissenschaften in Zagreb, ii. diese Studenten Population repräsentiert die Gemeinschaft der Praxis dass zwischen einer Individuum und die Gesellschaft vermittelt, iii. zudem decken sie (die Graffiti) die soziale und kulturelle Werte, die politische Probleme und das Leben insgesamt auf, iv. Diskursstruktur der Latrinalia ist dialogisch, handlungsorientiert und multimodal.

Schließlich gibt es zwei Hauptziele in dieser Studie – i. die Vernetzung unterschiedlicher sozialer Kontext Ebenen aufzuzeigen, und (ii.) bewusst zu machen, dass wir nicht bloße Konsumenten von Raum, Ort, Zeichen und Sprache sind, sondern auch deren Schöpfer, Interpreten und Kritiker.

Schlüsselwörter: Kritische Diskursanalyse, Soziopragmatik, Sprach Ethnographie, Graffiti, Latrinalia, Identität, Ideologie, Studentenschaft

Igor Ilić*
UR 1339 LiLPa
University of Strasbourg

ARGUMENT NARRATIVIZATION THROUGH DISCOURSE REPRESENTATION: EXAMPLE OF ONE PUBLIC APPEARANCE

Abstract

It seems that direct and indirect speech, terms used in the schooling process, are not a topic of interest even in grammar textbooks¹ (e.g. Klikovac 2010, Stanojčić & Popović 2012). With this fact in mind, one could think that this is not a topic worthy of interest in the research community. However, the aim of the present paper is to refute such a belief. One acquires very early that both of these speech types serve as a transmitter of other people's words, the latter being transmitted in their original form using three models, whereas within the reported speech these words are being transformed and integrated in the enunciator's statement. The term discourse representation will be used in the present paper for both types of speech, and their simplified typology based on three main features – direct, indirect and free, will be presented as well. A corpus which is composed of a transcript of a 80-minute political speech is explored and its analysis has shown numerous very descriptive, sometimes plastic, examples of represented discourse that serve not as an indicator of what another speaker had stated but as a picture of the situation in which the main speaker is on one hand and the presence or the absence of the other speaker's credibility, or even the main speaker's credibility on the other. The argument narrativization rests in fact on representation, a sort

* ilicigor@hotmail.rs

¹ In Serbia.

of staging of other people's words – which is described within the concept of polyphony – and this representation is a cohesive factor since it either introduces a part of argumentation or connects one part with another related part of argumentation, i.e. the represented discourses speak on behalf of the speaker.

Key words: discourse representation, narrativization, polyphony

1. UNRAVELING

How does one language phenomenon keep being extremely familiar to the scientific as well as to the non-scientific community and yet still dubious in regard to its function(s) within the language? One could naturally think of that well-known phrase *Secrets are best hidden in plain sight*, and not without a reason. On the other hand, it is needless to point out another aspect of this phenomenon, and that is its frequency, given the fact that in the second of the two preceding sentences the phenomenon is already present. To say discourse representation (in further text DR), a term used in line with N. Fairclough, is less a way of 'starting with a clean slate' and more a way of defining a broadly studied field in order to include and underline some pragmatological aspects that weren't under the linguist's microscope – "[...] there is always a decision to interpret and represent it [what was said or written] in one way rather than another" (Fairclough 1995: 54). Other familiar terms – 'reported speech' in English, 'discours rapporté' in French, '(ne)upravni govor' in Serbian, tend naturally to associate this type of discourse with a statement of what a person said or wrote thus somewhat transforming the efforts to describe it into a matter of stylistics. With this in mind, another doubt one could have concerns the delimitation of the phenomenon: how can 'words of others' be noticed, and eventually analysed, especially if the bakhtinian maxime of multilayered enunciation is applied?

It may be easier to answer this question in the case of written language, since language is stabilised there. However, within the framework of this dualism between spoken and written language, we

should not, as Koch & Oesterreicher (2001) point out, limit ourselves to distinguishing two completely distinct poles. Rather, it is a continuum in which the point of discrimination is no longer the code (phonic or graphic), but another aspect that could be described as the conception or the spatio-temporal realisation of a verbal exchange, which the authors call immediate or remote communication. In other words: (a) a written text would belong to the graphic code and to distant communication, (b) an oral utterance to the phonic code and to immediate communication, while (c) a written text spoken orally would belong to the phonic code, but to distant communication and (d) an oral utterance retranscribed to writing would belong to the graphic code, but to immediate communication. We would like to point out, along with Koch & Oesterreicher, that this type of encompassing dichotomy allows us to take into account the hybrid cases mentioned in (c) and (d), which, in terms of DR, result in situations in which α tells *viva voce* what β has written to them, or the latter writes what the former has told them orally.

From the conceptional point of view, the ‘text’ we have based ourselves on – a presidential address – is in the communicative distance, since the speaker relies on a scriptural support prepared in advance. We are therefore not in a context of spontaneous speaking, however there is a certain amount of improvisations, which we will see in some examples. We are not in an explicit dialogue either, since the speaker is addressing an audience that is there to listen given the very status of the empirical subject. Since our aim is to understand more about the use, and even the utility, of DR from a pragmatic perspective, we focus on this aspect and leave aside not only the social, but also the rhetorical baggage of the speaker, of which we are fully aware, but which, for the moment, does not seem essential to our analysis. When we say ‘speaker’ (*locuteur*), we mean ‘abstract entity responsible for a given utterance’, in particular in the tradition of O. Ducrot (1984).

Given the fact that our corpus is in Serbian, it seems appropriate to review a general definition of our subject taken from Peco & Stanojčić (1972):

(1) **DIREKTNI** (ili **UPRAVNI** ili **NEPOSREDNI**) **GOVOR**,

1. prenošenje poruke, saopštenja trećeg lica u obliku u kojem je to rečeno. Kako se njime prenosi neizmenjen govor, onaj koji to čini uvodi slušaoca u saopštenje upotrebljavajući prvo nezavisnu rečenicu koja sadrži neki glagol govorenja, mišljenja, iskazivanja rečima ili sl. i posle pauze, koja odvaja govor trećeg lica od govora govornog lica, navodi reči trećeg lica. [...]² (id: 72–73)

(2) **NEUPRAVNI GOVOR**,

način govora u kojem se nečije reči ne daju onako kako su izgovorene, nego samo iznosi njihov osnovni smisao. Tada pretvaramo upravni govor u nepravni, a rečenična interpunkcija je kao kod izričnih rečenica. [...]³ (id: 242)

The French-speaking reader will find an almost identical equivalent definition for French in Riegel et al. (2014: 1009–1016) or in Le Goffic (1993: 268–270), which underlines a certain tendency, i.e. a certain vision of the DR phenomenon. Is there another possible way?

2. POSSIBLE FRAMEWORK

First, let us look more closely at the implications of the definition (1). *Conveyance of a message*: these words essentially point out the usefulness, even the finality, of the process, but is the purpose of a speaker representing a discourse a mere transmission, a mere report

² **DIRECT** (or **GUIDED** or **PROXIMATE**) **DISCOURSE**,

¹ conveyance of a message, a statement of a third person following the form in which it was uttered. Given that this form is used to convey unchanged discourse, the person doing it introduces the listener to the statement by using firstly an independent clause with *verba dicendi*, *verba putandi*, some kind of expression with words or something similar, and after the pause that separates the speech of the third person from the speech of the enunciator, cites the words of the third person [...] [our translation]

³ **UNGUIDED DISCOURSE**,

speech type in which words of another speaker are not given in the way they were pronounced, only their basic sense is presented. In this case we transform guided into unguided speech, and the punctuation is the same as in declarative clauses. [...]

disinterested of any contextualisation? *The message of a third person*: isn't the third person an ambiguous term, given the cases cited by the author of the article where the speaker seems to be quoting himself, but also, even if it is somewhat trivial to recall, the cases where the speaker seems to be quoting his interlocutor – the 'second' person or you? *The message follows the form in which it was uttered*: this implies a more or less complete respect for the original utterance of the speaker who is quoted. Although we have not yet managed to find similar results for Serbian, it is worth mentioning that analyses of DR in spoken French have shown that a significant number of DRs produced are not *retransmitted*, but probably *staged* (Rosier 2008: 19). As for the form more precisely, we ourselves noticed during the analysis carried out for the purposes of our Master's thesis (Ilić 2021) that speakers sometimes completely modify the speech of others, thus making it incompatible with the original statement. It is not clear why within the same definition (Peco & Stanojčić 1972: 72–73) we find under 2. the *free unguided speech*, which is there related to artistic language, but the author uses an interesting criterion in defining it. Indeed, we read that it is “[...] direktni govor kojim se u umetničkom jeziku ističe pripovedač koji – navodeći tuđe reči kao deo svoga govora, – ističe i svoj stav prema tome što te reči kazuju” (our modified emphasis). We are approaching a fairly conclusive description here since we are given an interpretive characterization, but let us observe before proceeding the definition (2).

To give not the form of the words, but their basic sense, or in other terms their ideational content: the discrimination that this definition makes legible between direct and indirect speech on one hand, and which is based precisely on having either a complete authentically quoted utterance or a kind of paraphrasing reformulation of the same utterance also joins its syntactic translation – the author explains further in definition (2) that the quoted utterance is a declarative subordinate proposition, its punctuation therefore being identical to that of declarative sentences, whereas for direct speech it is explained that there would be two independent propositions. If we focus on this aspect in the Normative Grammar of the Serbian Language (Piper

2014: 492–521), we come across a mention of DR in the paragraph devoted to declarative sentences. The authors point out that these are used in indirect discourse, whilst asyndetic sentences are employed for direct discourse (*id.*: 497).

Let us now take an example from our corpus and test it against the definitions cited:

(3) [...] samo želim da vam kažem da su stvari mnogo komplikovanije od onih kakvim ih predstavljamo jer nisu stvari hoćeš da priznaš nećeš da priznaš nećeš i neće ali imamo milion stvari kompromisnih koje moramo da rešimo [...](00:42:32)

[...] I just want to say that things are a lot more complicated than what we present them to be because things are not like you want to recognize it or you don't want to recognize it I won't and we won't but we have million compromise matters that we have to solve [...]

It is immediately noticeable that the citing cotext – the one in which the quoted speech is inserted – prepares the ground: the speaker makes a statement that is supposed to summarize an image they have just created and represents two DRs to support their thesis. Given that there is no “independent clause with *verba dicendi*, *verba putandi*, some kind of expression with words or something similar” (*cf.* (1) above), these two DRs could be qualified as *free direct discourses*, as it seems to us, for the first of the two, that the group *jer nisu stvari/because things are not* is far from being a usual introducer of the DR. This is the first problematic point with regard to the definitions because there is no mention of free direct speech, which is not surprising since descriptions of this phenomenon are late in French as well (Rosier 2008). It can also be stated without hesitation that it is not a question of “conveyance of a message, a statement of a third person following the form in which it was uttered”, because on one hand the presupposed genuine utterance did not take place (as it is) and on the other hand the third person is consequently non-existent. In fact, this is what Roulet defines as *diaphonic* (*diaphonique*) discourse (1985: 73), which occurs when the speaker represents an apparent exchange between themselves and an interlocutor. In this case, it is a *potential* diaphony because the exchange did not really

take place and an *implicit* one because the exchange is represented without any introducer.

Another element of definition (1) above seems useful, even if it is impossible to reconstruct the author's thought in this passage – *and after* the pause that separates the speech of the third person from the speech of the enunciator, *cites the words of the third person*. The pause would be perceptible only in spoken language, its written equivalent being a simple space possibly with a colon. And indeed, the speaker makes a pause (audible in the audiovisual recording of the document) between the two DRs and changes the tone of the second DR somewhat, thus marking this diaphonic setting.

This apparently effective existence in (3) of another *voice* is part of what Ducrot or E. Roulet (1985) generally characterise as *polyphony*, namely that in enunciation seen as the process of creating an utterance there is a superposition of several voices (Ducrot 1984: 171–233). It should be noted that in Ducrot's descriptions, the DR is seen as the trace of a double enunciation, i.e. the utterance bears witness to the presence of two speakers, but it is nevertheless one of the two forms of polyphony that Ducrot distinguishes. The other case involves the presence of another discursive being to whom the speaker does not explicitly attribute words – the *enunciator* (*énonciateur*). These two entities are also sufficient for our own analysis because they are naturally compatible with the DR: on one hand a *speaker* who is presented as responsible for the utterance and on the other hand an *enunciator* who shows through the utterance without actually being marked in the utterance (*ibid.*). The reader will observe this phenomenon in (4) below, where the underlined sequence among others is entirely attributable to an enunciator whose position would be equivalent to the content of the sequence. The speaker does not take this attitude on their own, but only quotes it in order to make the previous sequence explicit, hence a *zato što/because*.

(4) [...] ne smete da govorite zamislite o svojim ekonomskim uspesima koje potvrđuje Eurostat zato što Srbija kao parija mora da bude naviknuta na to

mesto jer Srbija sme da pobeđuje u košarci ponekad u odbojci i vaterpolu a ovo drugo Srbija ne sme da pobeđuje [...] (00:40:26)

[...] imagine you can not talk about your successes in economy confirmed by Eurostat because Serbia like a pariah must be accustomed to that place because Serbia may win in basketball sometimes in volleyball and water polo but in this other aspect Serbia mustn't win [...]

In his modular approach to discourse, Roulet distinguishes several dimensions, one of which we are particularly interested in – the enunciative dimension (*cf.* Roulet 1997), and we have also retained the way in which this author sequences the different discourses in order to use it, in our case, to mark the DRs. When annotating the corpus, we placed them in square brackets [] and preceded them with the source of the voice, for example L for speaker (*locuteur*). The criterion Roulet uses to make a first differentiation is that of the formulation of the speeches of others (*cf.* Roulet 1999) – unlike Fairclough (1995: 55) who distinguishes five parameters including the mode, corresponding roughly to the type of DR in the canonical terms, and including more or less the less usual examples in relation precisely to the canon. It is only when the DR is explicitly *formulated* that the traditional direct/indirect dualism can be introduced. If the DR is *not formulated*, it remains to be determined whether it is designated (*désigné*), i.e. the statement indicates that a previous speech act has taken place (*cf.* (5) below), or whether it is implied (*implicité*), i.e. the statement implies and relies on previous speech without comprising a citation (*cf.* (6) below).

(5) [...] a onda kada sam primereno odgovorio L [] onda je rekao [...]
(00:50:07)

[...] and then when I responded appropriately L [] he said [...]

(6) [...] C [pa ne možemo znate mnogi od njih imaju dosijea neki su činili pre neki posle toga različita krivična dela i prekršaje pa sad nije u redu da oni budu u blizini predsednika] C [] je li i onda sam napravio [...] (00:54:12)

[...] C [oh you know we just can't we have files on lots of them some of them committed different crimes and delicts before that some of them after and so it's not appropriate to have them near the president] C [] is it so and then I made [...]

From the examples in our corpus quoted so far ((3), (4), (5) and (6)), it is clear, to say the least, that these utterances are likely to stretch any attempt at theoretical framing to the extreme. Nevertheless, they can be described using the standards mentioned above, because in this work it is more important for us to understand why speakers invest themselves so much in discourse representation – wouldn't it be easier, more direct and more economical to make their positions explicit?

3. EXPLORATION

All of the questions arose on the base of a corpus made out of a speech that Serbian president A. Vučić gave in Serbian in front of the Serbian Parliament in June 2021. It is a political speech consisting of closely 50 minutes of material that has been completely transcribed for the purpose of the analysis. The time indicated in brackets behind each example corresponds to the beginning of the quoted sequence in the recording of the speech available online. The spelling is respected as it facilitates the reading of the transcript, but there is no punctuation given that it is spoken language. We could be criticised for intentionally citing more complex examples in the introduction, so let's start with a 'more classic' case.

(7) [...] u tom trenutku vi shvatite da više ne razgovarate ni o čemu nekako su nas vratili uspeli da nas vrata Borelji i ostali do nestalih osoba da pre toga smo vodili dijalog L+K [] oko zajednice srpskih opština šta god da ste rekli od svih pravila L [pacta sunt servanda] do toga da L [sve što je odlučeno ne može da se menja potpisano od tri strane potpisnice i da mora da bude ispunjeno da bi postojalo poverenje u nastavak procesa] na to je njegov odgovor bio da K [ustavni sud Kosova je to odbio] i da K [njega to ne zanima] a onda je [...] (00:35:05)

[...] in that moment you realize that you're not discussing anything anymore they somehow managed to restore the focus Borrell and others to the missing persons oh yes before that we had a dialogue L+K [] about the community of Serb municipalities no matter what you said of all the rules L [pacta sunt servanda] to the fact that L [everything that was decided can not be changed signed by three signatory parties and that it has to be accomplished so that we believe in the continuation of the process] his

answer was that **K** [the constitutional court of Kosovo has disputed it] and that **K** [he's not interested in it] and then he [...]

It is necessary, in order to allow a good contextualisation, to quote slightly longer passages, which will be the case with the following two examples, but there is no more economical way to show the effects we intend to describe. In (7) the speaker tries to show that the dialogue between the two parties is not working, and that it is not the fault of the speaker's side. The latter narrates an episode by first designating a speech (*vodili dijalog/had a dialogue*) and then quoting examples of what they said in two direct speeches, thus underlining their efforts already mentioned by *šta god da ste rekli/no matter what you said*. The L2 refuses to cooperate, to dialogue, which is supported by the two indirect discourses at the end of the example. The L puts the L2 on stage in order to better illustrate what they have personally experienced in their own view of things.

(8) [...] morate da razumete da sve što rade su trikovi samo da bismo priznali nezavisnost Kosova a deklaraciju o miru možete da potpišete samo sa suverenom i nezavisnom državom sad moje pitanje šta da mu kažete ako kažete ne onda ste vi protiv mira a oni su za mir i onda morate da uđete u objašnjenja L [] koja traju po 20 minuta pola sata zašto to tako ne može ali vi vidite iza svake rečenice koju izgovaraju L2 [] to je samo L2 [nateraće Amerika i EU Srbiju da prizna nezavisnost Kosova a mi im nećemo čak ni ponuditi ništa] vidite ne postoje ni želje za razgovor a to je ono što me brine mi smo spremni da štitimo tu najvišu vrednost uz slobodu a to je mir ali nismo [...] (00:57:55)

[...]you have to understand that everything they do is tricks only to get us recognise the independence of Kosovo so you can sign the declaration of peace only with a sovereign and independent country now my question is what should you respond to him if you say no then you are against peace they are for peace and then you have to make 20 30 minute long explanations about why you can't do such a thing but you see behind every sentence they pronounce it's just L2 [USA and EU will force Serbia recognise the independence of Kosovo and we won't even offer them anything] you see there is not even willingness to talk and that's what concerns me we are ready to protect that most important value beside freedom and that's peace but we're not [...]

Example (8) is similar to the previous sequence in that we have the same L-L2 pairing, but unlike (7) it can hardly be said that the direct speech imputed to L2 in the second part of the example is based on an explicitly presented statement. Moreover, the L indicates this by more or less specifying that it is his own interpretation 'between the lines' in *ali vi vidite iza svake rečenice koju izgovaraju/but you see behind every sentence they pronounce*. The L must therefore have understood the L2's intentions through their utterance during the conversation in order to summarise and represent them in their own speech in this DD, which could be described as free since its only lexical link with the previous part is *to je samo/it's just*. This representation further degrades the image of the L2, while the context weighing on the L's perspective tells us about the latter's apparent tactics. It is worth mentioning here the distinction made by Ducrot (1984: 199-203) between the speaker as such (*locuteur en tant que tel*), who is thus presented as responsible for the utterance, and the speaker as a being of the world (*locuteur en tant qu'être du monde*), who is the complete person behind the discourse. This second differentiation, after the first one between the empirical author and the speaker as there is no default sign of equivalence between the two, helps to better account for these effects on the speaker's image described above. The speaker who describes a scene like the one in (8) and corroborates it with a DR does not give us information about themselves as a physical person, but about themselves as a simple speaker who is open to discussion and tries to discuss with L2, but they do not receive any answer and their efforts are doomed to fail because of the L2.

(9) [...] dakle bezbroj rizika ima bezbroj problema naše ponašanje mora da bude odgovornije ozbiljnije mi moramo narodu da govorimo istinu da ga ne obmanjujemo da ne pričamo bajke zato što nam predstoje izbori neka pričaju bajke oni koji ga lažu sve vreme znate kako izgledaju ti njihovi razgovori O+F [] sa strancima O [mi moramo da smenimo ovog diktatora Vučića vidite on ne želi da prizna nezavisno Kosovo] onda ih oni pitaju F [a je l' vi želite] O [ne ne pa mi mislimo znate dragi drugi sekretaru i treća sekretarice i prvi zameniče druge sekretarice mi želimo da vas obavestimo da je najveći problem u razgovorima sa Prištinom nedostatak transparentnosti znate i demokratskog kapaciteta] ovi ljudi gledaju šta ove budale pričaju pa

kažu F [pa dobro sve je u redu nego nam recite kakvo vi rešenje vidite konačno] O [pa znate ta transparentnost i taj demokratski kapacitet je strašno važan vi nam pomozite da pobedimo Vučića a mi ćemo da se dogovorimo sa Albancima brzo za 15 dana mesec dana] F [u redu je to ljudi da ćete vi da se dogovorite sa Albancima nego šta će da bude to rešenje rezultat vašeg dogovora] O [pa nemojte sad mnogo da nas pitate nismo još spremni do ovog nivoa nismo došli da razgovaramo sa vama ali je najvažnije da nam pomognete da se ratosiljamo diktatora a onda će sve da bude drugačije i lakše] i zamislite kada imate takve neodgovorne tipove sa jedne strane i ovako neodgovorne ljude u Prištini sa ovakvim pristupom o čemu sam već pričao u šta to sve može da se pretvori u šta to sve može da se pretvori samo u jednom jedinom danu [...] (01:10:49)

[...] so there are countless riskscountless problems our approach must be more responsible more serious we have to tell the truth to the people and not deceive it not tell fairy tales because the elections are approaching leave fairy tales to those who lie to people all the time do you know whattheir conversations with foreign diplomates look like O [we must replace this dictator Vučić you see he doesn't want to recognise the independent Kosovo] then they ask them F [and do you want it] O [well no no we think you know dear Mr second secretary and Mrs third secretary and Mr first deputy of the Mrs second secretary we want to inform you that the biggest problem in the conversations with Priština is the lack of transparency you know and of the democratic capacity] those diplomates look at what these idiots are saying and respond F [well good everything is alright but can you present us your final solution] O [well you know that transparency and that democratic capacity is really crucial help us win against Vučić and we'll make a deal with Albanians quickly in 15 days or in a month] F [that's okay you'll make a deal with Albanians but what will be the result of your deal] O [well don't ask us now we're not yet ready at this point we did not come to talk to you but the most important thing is that you help us get rid of the dictator and then everything will be different and easier] and imagine when you have this kind of irresponsible individuals on one side and those irresponsible individuals in Priština with this kind of approach that I have already explained what can become of that what can all of that become in just one day [...]

Indeed, example (9) is one of the most complex, as the L enters into a digression depicting a whole conversation almost worthy of a theatre scene. After stressing the importance of a certain behaviour *naše ponašanje mora da bude odgovornije/our approach must be more responsible*, the L has to validate this instruction, contrast it in a

dialectical confrontation with another behaviour. This is why the L launches into an episode where, through the DRs, they make see and understand this behaviour. This episode is already in a 'difficult' context posed by the L with other DRs which it would be too cumbersome to reproduce here, but which would further facilitate the understanding of the L's behaviour. Once the introduction is made *ti njihovi razgovori sa strancima/their conversations with foreign diplomats*, and this discourse is designated, the second DR, which is a free direct discourse, makes things explicit and gives us a voice representing the opposition, marked O. This free DD characterizes the L as a being of the world and problematizes its position by containing, among other things, a polemical negation (Ducrot 1984: 217) *on ne želi da prizna/he doesn't want to recognize* supposedly countering the opinion of a supposed content *mi želimo/we do want*. The following direct speech attributed to F is based exactly on this point, and the free direct speech attributed to O immediately afterwards responds to it. This response is all the more interesting because we find a characterization of F as a being of the world also in *drugi drugi sekretaru.../dear Mr second secretary...*, these characterizations by a progressive reversal ending up characterizing the main speaker as such in a good way. By discrediting these potential speakers in the continuation of this diaphonic representation, the L regains and emphasizes their credibility, also by summarizing these DRs in the final part of sequence (9) by *i zamislite kada.../and imagine when*.

(10) [...] zadovoljan sam što smo juče sa delom ljudi koji su van parlamenta dakle drugačije misle imali dobar ozbiljan sadržajan razgovor i mnogi od njih su uprkos protivljenjima našoj politici rekli da P [su spremni da pomognu svojoj državi ne nama kao političkim subjektima ne personalno ali svakako da pomognu svojoj državi] ja im izražavam zahvalnost na odgovornom pristupu [...] (01:13:55)

[...] I'm pleased that we had a good serious meaningful conversation yesterday with a part of the group that isn't in the parliament they think differently so to say and lots of them said in spite of their objections towards our policy that P [they are ready to help their country not us as political subjects not personally but non the less help their own country] and I express my gratitude to them for their responsible approach [...]

The last example seems, at first sight, simpler than the previous ones, but here too there is some confrontation. The speakers represented in the indirect discourse are opposed to an enunciator whose attitude could be expressed by the content *da pomognu političarima na vlasti/to help the politicians in charge*, but it is difficult to say whether the negation *ne nama.../not us...* is a quotation or the attitude underlined by the L. We are inclined to say that it is also a content represented by the words *uprkos protivljenjima.../in spite of their objections...* In the end, this representation by a somewhat paradoxical return (*zahvalnost na odgovornom pristupu/gratitude for their responsible approach*) confirms again the validity of the argument and the good intentions of the L.

4. CONCLUDING REMARKS

After examining all the examples in the corpus, it can be affirmed on one hand that what they show can be fitted into the framework posited by Vincent & Dubois (1997: 131), namely that the direct/indirect bicephalous structure persists and that direct DRs are prevalent – 85% of the occurrences in the French corpus of the cited authors, 44% in our Serbian corpus. However, this shift needs to be studied further on a multitude of examples in order to draw definite conclusions. On the other hand, we can also agree with these authors that ‘conveying a message, simply reporting words’ is not, in the majority of cases in our corpus, the function of the DR. This process often seems to be, as Vincent (2006: 128) points out, a tool of persuasion, because the examples we have presented here show to what extent speakers can, against a certain linguistic economy, especially in (9), use different tools in order to make the listener understand the elements they want to demonstrate.

If we deconstruct the definitions analysed in part 1, we can assume to some extent that they are based on written language, except possibly ‘the listener’ and ‘the pause’, which are elements of spoken language. It should be pointed out that the studied examples show that this type

of traditional definitions of discourse representation, whether it is the case of Serbian or French, fail in the pragmatical frame, which implies the need for redefining the phenomenon in a more comprehensive way thus creating more stable theoretical tools for further analysis. Besides the longly debated question of the unicity of the uttering subject (*unicité du sujet parlant*) – establishing that the one who speaks is not necessarily and automatically the one who is spoken (about) (*c.f. e.g.* Ducrot 1984: 171) –, the examples emphasize yet again the question we tried to call attention to in this work, and that is the need of discourse representation strategies. The speaker’s tendency to represent in various contexts is marked in our corpus to the point that we asked ourself who is really speaking. It seems to be evident at this stage that nearly every time the speaker in the corpus represents discourses, there is a point to be made argumentatively speaking, and we describe this conditionally as argument narrativization given that the represented discourses in context are ‘on their own’, i.e. the speaker generally does not provide any elaborate interpretation of the RDs, the RDs *speak on behalf of the speaker*.

In a complementary perspective, we would like to draw the reader’s attention to an aspect that is important for representation in spoken language in our opinion, and that is the prosodic component, i.e. the actual material voice of the speaker. There are examples in our corpus of what L. Perrin observes when he writes that “[...] le locuteur s’emploie à montrer, reproduire mimétiquement, pasticher un discours objet ou un point de vue, à en produire une sorte de réplique qu’il prétend plus ou moins fidèle ou approximative [...]” (2005: 183). This kind of reproduction can be based on the lexical component (see (11) below), but also solely on intonation.

(11) [...] i pet puta ponovio **K** [kada kada kada] i to jedino pitanje **K** [kada] [...] (00:33:26)

[...] and repeated five times **K** [when when when] and that single question **K** [when] [...]

In the ongoing research for our doctoral thesis, we hypothesized that the speaker can use the elements of intonation to express his

or her attitude towards a represented content. S. Günthner (1999) provides an initial exploration of this subject. Indeed, this author investigates the way in which speakers communicate their points of view on the dialogues they report in order to understand which prosodic means are used by the speaker to express their agreement or disagreement with a represented statement. Although her findings on the use of prosodic features and voice quality to produce plurivocal discourses apply to everyday conversations in German, we believe that this phenomenon needs to be studied in Serbian as well as in French or English, but this would require a separate article just to begin with.

REFERENCES

- Bahtin, M. (1980). Marksizam i filozofija jezika. Nolit.
- Carel, M., Ducrot, O. (2009). Mise au point sur la polyphonie. *Langue française*, n° 164 (4), 33–43.
- Ducrot, O. (1984). Le dire et le dit. Minuit.
- Faireclough, N. (1995). Critical discourse analysis: The critical study of language. Longman.
- Günthner, S. (1999). Polyphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics*, 31 (5), 685–708.
- Ilić, I. (2021). Le discours représenté à l'oral (dir. A. Kuyumcuyan). Mémoire de master préparé et soutenu à l'Université de Strasbourg.
- Klikovac 2012: Кликовац, Д. (2010). Граматика српског језика за основну школу. Креативни центар.
- Koch, P., Oesterreicher, W. (2001). Langage parlé et langage écrit. *Lexikon der Romanistischen Linguistik*, I, 2, 584–627.
- Le Goffic, P. (1993). Grammaire de la phrase française. Hachette supérieur.
- Peco, A., Stanojčić, Ž. (dir.) (1972). Srpsko-hrvatski jezik (t. 1). *Enciklopedijski leksikon Mozaik znanja*. Interpres Beograd.
- Perrin, L. (2005). Polyphonie et séquence écho. *Dialogisme et polyphonie : Approches linguistiques*, 173–185.
- Piper 2014: Пипер, П. (2014). Нормативна граматика српског језика. Матица српска.

- Riegel, M., Pellat, J.-C., Rioul, R. (2014). *Grammaire méthodique du français* (5. éd.). Presses Universitaires de France.
- Rosier, L. (2008). Le discours rapporté en français. *Ophrys*.
- Roulet, E. (1985). L'articulation du discours en français contemporain. P. Lang.
- Roulet, E. (1997). L'organisation polyphonique et l'organisation inférentielle d'un dialogue romanesque. *Cahiers de linguistique française*, 19, 149–179.
- Roulet, E. (1999). La description de l'organisation du discours: Du dialogue au texte. Didier.
- Stanojčić & Popović (2012). Станојчић, Ж., Поповић, Љ. Граматика српског језика за гимназије и средње школе. Завод за уџбенике.
- Vincent, D., Dubois, S. (1997). Le discours rapporté au quotidien. Nuit Blanche éditeur.
- Vincent, D. (2006). Polyphonie et interaction. *Le sens et ses voix, dialogisme et polyphonie en langue et en discours*, 28, 127–142.

Igor Ilić

LA NARRATIVISATION D'ARGUMENTS VIA LA REPRÉSENTATION DE DISCOURS : EXEMPLE D'UNE INTERVENTION PUBLIQUE

Résumé

Le présent article tend à montrer en quoi consiste l'objectif d'un locuteur par rapport à la représentation de discours dans le cadre d'une argumentation. À partir d'un cadre théorique reposant sur les travaux de Ducrot (*cf. e.g.* Ducrot 1984) et ceux du cercle genevois de Roulet (*cf. e.g.* Roulet 1985), un corpus constitué d'une allocution d'un président serbe a été retranscrit et annoté en DR et ce corpus a permis de constater un phénomène qui peut être décrit sous réserve comme une narrativisation, à savoir que le locuteur essayant de prouver une certaine position représente des discours soit pour introduire cette position, soit durant l'argumentation, soit pour corroborer, en tout cas pour faire comprendre une position opposée à la sienne à l'aide du DR. Plusieurs exemples ont été tirés du corpus et analysés en détails afin de décrire ce procédé qui est loin bien sûr d'être généralisable, mais qu'il faudrait davantage étudier dans de différentes situations de communication.

Mots clés : représentation de discours, narrativisation, polyphonie

Maja Brkljač*

University of Belgrade – Faculty of Philology

THE IMPORTANCE OF THE HEADLINE AND LEAD FOR THE INTERPRETATION OF MEDIA ARTICLES

Abstract

The research examines the role of headlines and leads in the interpretation of articles, along with the linguistic means they use and their purpose. It aims to draw conclusions on whether these headlines and leads present the news accurately and what effect they have on the reader. The paper analyzes headlines about three events – the start of the Covid19 epidemic in Serbia, the anti-Serb graffiti in Zagreb, and the protests against the new Montenegrin Law on Religious Freedom – considering how online media have changed the way of reporting as opposed to print media.

Key words: the media, lead, headline, online portal, media objectivity, information, clickbait

INTRODUCTION

Mass media have a major influence on the general public and a significant impact on the general public's opinion, due to their public nature and availability to large numbers of people all around the globe. As McQuail claims "mass media are channels that carry mass

* majabrkljac95@gmail.com

communication and almost all research into the latter is based on the assumption that the media have significant effects on the affairs of people” (1994: 327). Media language, or more precisely, newspaper language, may be said to be a discourse on its own, as it has features and characteristics that separate it from other types of discourse. According to traditional beliefs, media reporting should always be objective, with no evaluations or insinuations made by reporters that could influence the reader’s attitudes. However, the existence of media objectivity has been called into question.

Sociologist Stuart Hall argues that the cultural leadership and hegemony are exercised mainly through media (1982). He claims that the media are involved in the politics of signification, in which they produce images that give particular events particular meanings. He adds that the media do not simply mirror reality, but rather represent it, or re-create it. Hall also suggests that the media have the power to signify events in a particular way, which points to the possible multiple versions of reality, created by the choices that article authors make. Social, cultural, economic and political factors determine the way in which the world will be presented in the media. Fairclough claims that ideological representations are generally implicit rather than explicit, since ideology is most effective when its workings are least visible (1989: 85).

Van Dijk (1993) lists some of the discursive structures that may contribute to the manipulation of beliefs:

- headlines and leads of news reports express semantic macrostructures defined by journalists and thus may underline the preferred macrostructures of mental models (e.g. a demonstration may be presented as an act of violence or the democratic right of the demonstrators);
- implications and presuppositions serve to assert information that may or may not be true in a subtle way;
- passive structures and nominalizations may be used to hide or downplay certain events;
- the lexical expression may influence opinions and knowledge in the mental models of the recipients;

- metaphors are a powerful tool to make abstract mental modes more concrete (e.g. the abstract notion of *immigration* becomes more real in the metaphor *waves of immigrants*, which creates the image of drowning in immigrants in the recipients' minds, thus creating a fear of them).

This research focuses on the analysis of articles published on various online portals in Serbia and it aims to gain perspective into how adequately the headlines and leads actually present the news piece, to see if they promote certain ideologies or contain emotionally charged language (and if yes, why), as well as to determine the means used to attract the audience and their differences depending on the character of the media. The headlines and leads refer to three different events that happened in Serbia over the course of three weeks, in the spring of 2020.

It is important to note that the goal of this research is to focus on the analysis of the headlines and leads and their possible effects on the readers' interpretations of the texts, and not to provide any type of judgement about the sources and/or events mentioned.

NEWSPAPER ARTICLE CHARACTERISTICS

According to James G. Stovall (2004), all journalistic writing should share four characteristics:

1. accuracy – all writers, not just journalists, are expected to present their information accurately;
2. completeness – refers to presenting the information in a context, in a clear and coherent way, while providing all the necessary answers that the audience needs for processing and understanding the information;
3. precision – refers to using accurate grammar structures and appropriate words;
4. efficiency – all journalists should use the fewest words to present the information clearly and truthfully.

Each newspaper article consists of various parts, but the aforementioned characteristics should be present in all of them in order to achieve objectivity. Van Dijk (1988b) provides a model of a *news schema*, according to which each news story has several constituent parts that can be categorized into different schematic categories.

According to this model, each news report consists of a *summary* and a *news story*. The headline and the lead make up the summary of the news story, i.e., they present the main events. The lead refers to the introductory sentences or introductory paragraph of an article, which serve to introduce its topic or purpose and attract the reader's attention. The news story encompasses the details of the situation (main events, its consequences, but also the background of the story – previous events, circumstances, history) and the comments (verbal reactions of other people or comments of the authors themselves). Van Dijk states that some of these schematic categories are mandatory, some are optional, and some of the categories are recursive, e.g., when an article is about several events, each with their own background and consequences (1986). The headline and the lead together give an overview of the text and express its semantic macrostructure (Van Dijk, 1988b: 53). The headline is obligatory and always comes before the lead, while the lead is common, but not always present. These categories appear in the text according to the value of the information they are providing, meaning that every newspaper article should progress from the most relevant pieces of information to the least important ones. For example, the headline always comes first and it is the first thing the readers notice. Only if the headline catches their attention, will they proceed with reading the rest of the article.

G. J. Digirolamo and D. L. Hintzman conducted an experiment during which they recorded people's reaction to an object stimulus that was presented to them five times, but either the first or the last picture was a mirror reverse. Upon seeing both photos side by side, most of the subjects claimed they only saw one of them – the one they had seen first, regardless of the remaining four, proving that the first impression can have a lasting effect (Digirolamo & Hintzman, 1997).

This suggests that headlines are important in that sense as well, since they are usually the first piece of information that people receive, which can a priori influence their perception of what can be expected in the article itself. What ensues from this is that the headline and the lead play a major role in news article interpretation.

Conventionally, the headline and the lead serve to summarize the essence of the text. As Bell puts it “the headline is an abstract of the abstract. The lead pares the story back to its essential point, and the headline abstracts the lead itself” (1991: 150). Being shorter of the two, the headline only contains the core of the story, while the lead conveys more details about the event. Bell states that “the lead focuses the story in a particular direction. It forms the lens through which the remainder of the story is viewed. This function is even more obvious for the headline” (1991: 152). This means that just by reading the headline and the lead the reader gets an idea on the article content and possibly forms a wrong opinion if the headline and the lead are biased (e.g., when they express a secondary topic of the article), especially since not all the readers will continue reading the text after having read the headline and the lead. Also, since the headline is the summary of a summary of a text, we must conclude that it is partly a reflection of the author’s perspective, since there can be many different headlines for every article, so the headline should not be equated with the main topic of the text (Brown & Yule, 1983: 73).

Furthermore, traditionally, every piece of news should provide answers to the following questions: Who? What? When? Where? Why? How? Van Dijk refers to the works of Garst and Bernstein, who claim that according to the normative rules of news writing, the lead must contain answers to all these questions. (Garst & Bernstein, 1982 in Van Dijk, 1986). However, this is not an explicit rule, as leads do not express all the information about participants, actions or events, locations, or other properties of news events, but rather the information about main actors, main event, main location, i.e., the macropropositions of the text (Van Dijk, 1986).

As all the articles analyzed in this research were taken from online portals due to practical reasons, we must not forget to mention *clickbait* headlines. Clickbait headlines are links on the Internet which aim to get as many people as possible to click on them and visit the website. Various means are used in clickbait links, such as exaggeration, leaving out the vital part of information in the title, using more sensationalist language in order to grab someone's attention, etc. It can be argued that *clickbaiting* has gained momentum in journalistic reporting for different reasons, including different business models that depend on the number of page visitors for revenue. To put it simply, clickbait headlines are created with the sole purpose of earning page views on a website and that goal is often achieved through the use of language that is more emotional than informational, often misleading, and meant to spark curiosity, while non-clickbait headlines present the main idea of the article.

In an interview for BBC, Peter Preston, the former editor of *The Guardian*, claimed that the journalists should find a way to get the maximum level of interest, as they are the ones who should be serving their readers. He also added that there is a danger of *dumbing down* of content on news websites, as "the criterion for including the story on the website is determined by the number of clicks", which poses a danger and, according to Preston, leads to focusing on less important news (2015). In the same interview, Damian Radcliffe, honorary research fellow at Cardiff University's School of Journalism, claimed that "headline writing is an art" because it is something that "draws people in" (2015).

ANALYSIS

The articles were taken from the most popular web portals in Serbia – *N1*, *Nova*, *Danas*, *RTS*, *Prva*, *B92*, *Blic*, *Alo*, *Kurir*, *Telegraf*. Apart from their popularity, one of the criteria for choosing the following sources is their difference in character, i.e., their informativeness. For instance, *RTS* is the portal of the national television of the Republic of Serbia, while *Alo*, *Kurir*, and *Telegraf* are tabloids, or the so-called

yellow press. The articles were accessed via search engine *naslovi.net*, as this way it was possible to search for headlines by day, making it easier to compare how different media reported on the same event. The events analyzed happened between June 14 and June 28, 2020. The three events this paper will focus on are the appearance of Darija Kisić Tepavčević, the Deputy Director of the *Batut* Institute¹, on the RTS Morning Show in the midst of Covid-19 epidemic in Serbia, the anti-Serbian graffiti in Zagreb, and the new Montenegrin Law on Religious Freedom, which resulted in mass protests all around the country. For the sake of easier comparison, these three topics will be separated during the analysis, but it should be noted that they are not always mutually exclusive.

The analysis that follows will try to determine how different sources reported on the mentioned topics, what means (linguistic and others) they used to attract readers, as well as whether the title summarized the article objectively and accurately. The context of each of the events will be explained for easier understanding. The goal of this analysis is not to judge someone's reporting, but only to answer the aforementioned questions.

COVID-19 EPIDEMIC IN SERBIA

On March 15, 2020 Serbia declared a state of emergency due to the spread of the new virus Covid-19 and a huge increase in the number of cases on a daily basis. Various strict measures were introduced in order to stop the spread of the virus, many of which made it impossible for people to work or move freely. However, only 11 days prior, parliamentary elections were called and they were supposed to take place on June 21. There were doubts concerning whether the elections would take place, as it was impossible to predict how long the epidemic was going to last and whether the elections would pose a health threat for the citizens of Serbia. The state of emergency was lifted on May 6,

¹ Institute of Public Health of Serbia *Dr Milan Jovanović Batut* is an expert institution which provides advice, support and guidance for the Serbian government and all public health departments in the country.

after it was declared that the situation was under control and that life could slowly return to normal. Most of the measures have been lifted, and it was decided that the parliamentary elections would be held when they were originally scheduled.

The Deputy Director of the *Batut* Institute, Dr. Darija Kisić Tepavčević, appeared on the RTS Morning Show on June 14, stating that the number of patients in Belgrade had increased in the past few days, due to people gathering in larger numbers. She adds that the situation is under control in the rest of the country, that the clinical picture of the newly infected patients is mild, and asks the citizens to act responsibly, since the holiday season has begun.

Web portals that reported on her appearance are *RTS*, *N1*, *Danas*, *Nova*, *Kurir*, and *Telegraf*. *RTS* had the following headline and lead:

H: There are no newly infected people in most parts of the country – why most new cases are from the area of Belgrade.

L: The coronavirus is not giving up, most of the newly infected are in Belgrade. Protection measures and responsible behavior are key, says Dr. Darija Kisić Tepavčević for RTS.

When comparing the article to the headline and lead, we can say that both the headline and the lead summarize the article adequately, providing key points and allowing the reader to find out more by reading the whole article. Still, the headline is a sort of clickbait, as there is a question *why most new cases are from the area of Belgrade*. We can assume that readers will get curious to find out and will therefore click on the link and open it to learn more.

N1 and *Danas* both copy the article, citing *RTS* as the source, but they change the headline:

H: Kisić Tepavčević: There are no newly infected people in most parts of the country

These two online portals publish the same article, with the same headline, quoting Dr. Kisić Tepavčević. Despite the headline being a quote, if the readers only read the headline, they could be misled to think the situation has improved overall and that the risk is almost

gone, which indicates that the headline without the lead would not adequately convey the point of the article. It is worth mentioning that both *RTS* and *N1* posted the video of the guest appearance, which allowed the readers to check the accuracy of the articles.

Portal *Nova* published the same article, but a shorter version, with different headline and lead:

H: Kisić: The majority of new cases in Belgrade due to celebrations

L: The vast majority of newly infected with the coronavirus are from the area of Belgrade, and epidemiological research has shown that the infection was transmitted indoors, at celebrations and at larger gatherings, said the deputy director of the Batut Institute, Darija Kisić Tepavčević.

Portal *Nova* cited the second part of Dr. Kisić's statement, which reoriented the direction of seeing the news – while *N1* and *Danas* focused on a small number of the newly infected in Serbia, portal *Nova* highlighted the fact that a large number of people in Belgrade caught the virus at celebrations and large gatherings, and the extensive lead that follows the headline confirms this view.

Telegraf reported on this only briefly, focusing on the increase in the number of new cases in Belgrade as opposed to the rest of the country.

H: Belgrade is the new hotbed of coronavirus: Everyone caught the virus in the same way, in almost the same place.

L: Kisić Tepavčević pointed out that more than 75 percent of all newly confirmed cases do not have any symptoms of infection.

Telegraf also focuses on the growing numbers, calling Belgrade *the hotbed* of the virus. This is another clickbait, as we can assume the readers would want to find out what place it is where everyone in Belgrade got infected. This rather alarming title is accompanied by a lead with the completely opposite tone, which informs the readers that the clinical picture of 75% newly infected is mild or asymptomatic. However, even this title and lead adequately display the text, given the fact that it is quite abbreviated compared to those on other portals.

Kurir states the following:

H: NO COVID CASES IN MOST OF SERBIA, BUT IT'S A WHOLE OTHER STORY IN BELGRADE! Dr. Daria found out why! HERE IS WHERE THE TRANSMISSION HAPPENS!

Kurir gives us the most dramatic perspective of all. The majority of the headline is in capital letters in order to catch the readers' attention and create the feeling that the text is about something shocking. The author contrasts the parts of Serbia that are virus-free and Belgrade which is *a whole other story*, insinuating that the situation in Belgrade is quite alarming. The readers are invited to read the text and find out why, as well as where the virus is most transmitted, which makes this title the most obvious clickbait of all, especially considering that the title itself does not provide us with any specific information.

All in all, none of the headlines deceive the reader in the true sense of the word. However, by seeing how the same event (which was even recorded on camera and available to everyone who wanted to watch it) was reported differently, and even more, how sometimes the headline changed while the text remained the same, we can conclude that there are factors outside the news that affect in what way it will be shared with an audience (journalist, portal policy, external factors, etc.). In this case, the focus of the headline shifts to different aspects of the news: 1) there are no new cases in most parts of Serbia; 2) the number of new cases is growing in Belgrade; 3) the largest number of new cases got infected at celebrations. Most of the headlines are accompanied by leads that somewhat balance them out and mention the key details missing from the headline, but we must bear in mind that we don't always click on the news and get the chance to read the lead, so we simply acquire the information from the headline. Certain expressions and capital letters used in these headlines and leads underscore the gravity of the epidemic, and clickbait appears even on the portal of the Radio Television of Serbia.

ANTI-SERBIAN GRAFFITI IN ZAGREB

A graffiti of a hanged Serbian family appeared at the periphery of Zagreb on June 14, which was widely condemned by the public, both in Serbia and in Croatia. The graffiti appeared on a wall close to a kindergarten and it shows people hanging from the tree, along with the text *Serbian family tree*. This was the third provocation of this type in just a few days.

N1 reports about the new graffiti in a short article, including the photo of the graffiti, and states that the Prime Minister of the Republic of Croatia Andrej Plenković and the Minister of the Interior Davor Božinović condemned such provocations. The headline and the lead are the following:

H: Another message of hatred, graffiti with hanged Serbs appeared in Zagreb.

L: A new graffiti insulting Serbs has appeared in Zagreb. The graffiti is located in Sveta Klara in Zagreb, right next to the children's playground.

RTS reports on this event too, stating in the news piece what were the previous two provocations:

H: Zagreb, the third threatening message sent to Serbs

L: Another threatening message addressed to Serbs appeared in Zagreb. This is the third message of its kind in the last few days.

Blic publishes the same article, with a different headline and no lead:

H: The third scandalous graffiti on the streets of Zagreb: Graffiti HANGED SERBIAN FAMILIES on the wall near the kindergarten (PHOTO)

Alo also changes the headline:

H: NEW, DISGUSTING MESSAGE IN ZAGREB Ustashas have been taught since they were little, a sick burden of hatred towards Serbs (PHOTO)

L: There is a tree from which people are hung, and next to which is written "Serbian family tree".

As the article itself is rather short on every portal, all headlines assert the main topic of the text. The clearest headlines are those found

in *N1* and *Blic*. However, the headline in *Blic* contains sensationalist elements – the word *PHOTO* in capital letters, which invites the readers to click on the link and see the graffiti. Furthermore, the adjective *scandalous* is used to describe the graffiti and the phrase *HANGED SERBIAN FAMILIES* is in capital letters too, which strengthens its tone and emphasizes the mentioned scandal.

The *RTS* headline does not inform the readers what happened. *The third threatening message addressed to Serbs* may relate to many things, not necessarily to the graffiti written. Even the lead does not provide this piece of information, so the readers are forced to read further if they want to know what kind of message it is. This is an obvious example of omitting key information from the headline, or an example of a clickbait.

The same can be said for the title of the *Alo* portal, with a few additions. First, in this case, the lead provides additional information about the event. Second, in the title we clearly see the author's subjectivity in the choice of the adjective used (*disgusting*) and the conclusion the author drew, which is that children in Croatia are taught to develop negative feelings and attitudes towards Serbs. The author bases this attitude on the fact that the graffiti appeared next to the children's playground.

According to the *Serbian Journalists' Code of Ethics*, there must be no spreading of any kind of stereotype in journalism, and personal prejudices of journalists must not be present in their work, and neither should any colloquial, derogatory and imprecise naming of a certain group. By reporting this way, the journalist encourages the readers to accept these views, thus promoting discrimination with their free interpretation and violating the *Code*. Every journalist must be aware of the danger of discrimination that can be spread by the media and should do everything in their power to avoid it. Discrimination of this type is mainly based on race, gender, age, sexual orientation, language, religion, political and other opinion, national or social origin².

² Accessed at <http://www.savetzastampu.rs/latinica/uploaded/Prirucnik%20U%20JAVNOM%20INTERESU%20final.pdf>

PROTESTS IN MONTENEGRO

The new Montenegrin Law on Religious Freedom, which stipulates that the state becomes the owner of all religious buildings built before December 1918 (when Montenegro became part of the Kingdom of Serbs, Croats and Slovenes) – if religious communities do not prove ownership, caused numerous protests around Montenegro. This topic was widely covered by the media, so in this case we will not look at one event, but at those headlines and leads that are openly biased, as they may help us reach certain conclusions relevant for this research. The difference in the media quality is most obvious regarding this event. *Danas*, *N1*, *RTS* and *Nova* published texts with neutral headlines and leads. Every headline is informative and every lead serves its purpose to provide more details. Also, they summarize the articles adequately and the reader would not get the wrong idea by reading only the headline and the lead, instead of the whole text. These four portals published a significantly lower number of articles that deal with this topic than tabloids, where we can find the following headlines and leads:

H: THE PEOPLE ARE RISING, LITANIES STARTED IN MONTENEGRO: Bijelo Polje, Berane, Pljevlja, Plužine stood up in defense of the sanctuary ...

L: After three months and the fight against the coronavirus pandemic, liturgies for the defense of the sanctuary of the Serbian Orthodox Church continue in Montenegro today (*Alo*, June 6, 2020)

H: CHAOS IN PODGORICA Milo's police forces broke into the Cathedral, the people are guarding Archpriest Mirčeta

L: Drama in the capital of Montenegro (*Alo*, June 14, 2020)

H: ALL AS ONE FOR THE DEFENSE OF THE SANCTUARY A video from Podgorica showing the UNBREAKABLE STRENGTH OF THE PEOPLE AT LITANIES (*Blic*, June 15, 2020)

H: THERE IS NO END TO REPRESSION! Father Željko stays behind bars, the reason is amazing! The people are chanting THROW AWAY THE BADGE AND SHAME ON YOU! VIDEO

L: Danilovgrad parish priest Željko Čalić was detained for 72 hours, and the people gathered in front of the Prosecutor's Office are indignant because of this repression. (*Kurir*, June 15, 2020)

H: MONTENEGRO RISES Litanies in Podgorica and Budva LED BY BISHOPS, Archbishop of Bijelo Polje summoned to the Security Center for questioning (*Blic*, June 21, 2020)

H: BRUTALITY IN MONTENEGRO Police shoot a man during the arrest in Budva, TEAR GAS ALL OVER PODGORICA (VIDEO) (*Blic*, June 24, 2020)

H: MILO'S SUBJECTS, HOW MUCH MORE? HARASSMENT OF A PRIEST AND HIS SON ON THE MONTENEGRIN BORDER!

L: KONSTANTIN DOJIĆ AND HIS UNDERAGE SON STAYED AT THE BORDER WAITING FOR SOMEONE TO RESOLVE HIS COMPLAINT. (*Alo*, June 26, 2020)

Each of the headlines is obviously a clickbait – we see capital letters, words *photo/video*, as well as emotionally charged vocabulary (*harassment, brutality, repression, chaos, unbreakable strength, all as one*, etc.). Also, the headlines create a negative attitude towards the president of Montenegro, Milo Đukanović, as he is implied to be responsible for the harassment of the Serbian people, while the Montenegrin police are called *Milo's subjects*. A number of the headlines somewhat supports the protests, presenting them as a sort of revolution (*Montenegro rises, all as one, defending the sanctuary*, etc.), highlighting patriotic feelings by implying the difference between *us*, the Serbian people, who defend what is sacred, and *them* who harass and repress. This subjectivity directly violates the *Serbian Journalists' Code of Ethic*.

What is interesting is that, although these headlines are explicitly biased, the leads that follow and the actual articles are more neutral. A possible reason for this is that the headline is simply exaggerated so that readers, shocked, could go to the portal and read the whole article, which, as we have already mentioned, is the goal of clickbait.

CONCLUSION

Almost all the means of manipulation mentioned by Van Dijk (1993) can be seen in the analyzed headlines – implications, presuppositions, the use of passive, nominalization, open attitude of the author, use of

emotionally-charged words, etc. There is an obvious difference in the reporting of tabloids and other media, especially when it comes to the second and third topics, which contain aspects of nationalism.

As far as coronavirus reporting is concerned, we have noticed that portals report differently on the same event, as well as that the author of the text determines what will be presented as the primary news topic. This is further evidenced by the fact that various portals published identical articles with different headlines. The appearance of clickbait headlines was noticed, as well as taking quotes for headlines. In general, none of these titles are highly manipulative or subjective.

The second group of articles resulted in the use of emotionally-charged vocabulary and more obvious subjectivity, especially in the tabloid media. The authors present their own biased conclusions and opinions without any relevant support. Such reporting can be extremely dangerous, as it may result in creating tensions and negative attitudes towards the citizens of the Republic of Croatia, because some media not only report on the event, but also express a very negative attitude towards a nation that readers can adopt.

When reporting on the protests in Montenegro, the authors use sensationalism. We can argue that these titles try to awaken the national consciousness of the Serbian people and to provoke certain emotions towards those who threaten their national identity. As in the previous case, the goal of these titles is not only to present the event, but to cause a feeling of shock and create attitudes in the readers. The subjectivity in these titles is distinctly explicit.

In general, non-tabloid media still follow the traditional way of reporting and creating headlines to some degree, although there is no doubt that each author has their own view, according to which they write the headline. Without exception, tabloids use means to turn headlines into clickbait headlines. Subjectivity is more explicit in the tabloids, even though the topic itself is an important factor in determining its presence.

Going back to Van Dijk's and Bell's definitions of headlines and leads, we see that there is a difference between theory and practice. While the headlines vary significantly, the leads mostly provide us

with all the necessary information about the given news. Even in tabloids, where headlines are sensationalist and subjective, the lead often has a completely different tone. In some cases, the title and the lead are not related or represent two (sometimes opposite) points from the text, so the headline does not actually always prove to be the summary of the lead. Nevertheless, even if the lead is more objective than the headline, we must bear in mind that we cannot see the lead unless we click on the headline, so the headline remains the primary source of information, which can influence our views.

Still, before we confirm the author's intention to manipulate the readers, we must keep in mind that online media have significantly changed the way headlines and articles are created and that research on traditional print media cannot be fully applied to them. Portals report differently than the print media. The main difference is that the portal can publish more articles on a topic per day, that is, each new piece of information about the event can be (and mostly is) separate news, because the portal profits from the number of visitors. Also, portals have the opportunity to post a relevant video along with the article (e.g., the guest appearance of Dr. Kisić Tepavčević), which enables readers to verify the facts they are served. We can see this in the example of *N1* and *RTS* – even if their headlines were completely subjective, we could assume that their main intention was not to deceive their readers.

To sum up, after seeing the headlines, leads and all the articles, it is most probable that the biggest difference can be seen in the headlines, but the reason for that is most likely the wish to attract as many readers as possible, rather than to manipulate and deceive. If it were not the case, there would be significant differences in every element of the articles. This does not mean that there is no subjectivity, so we should still be careful, think critically and be well informed about a topic before we make our judgment and form an attitude. Studying the influence of the media is certainly a broad topic and it is necessary to constantly re-examine our conclusions and look at information from different perspectives, so that we have a chance

to achieve relative objectivity and not become passive recipients of the media content.

REFERENCES

- Bell, A. (1991). *The language of news media*. Oxford: Blackwell.
- Brown, G. & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press
- Digirolamo, G. J. & Hintzman, D.L. *Psychonomic Bulletin & Review* (1997) 4: 121. <https://doi.org/10.3758/BF03210784>
- Fairclough, N. (1989). *Language and power*. London: Longman.
- Frampton, B. (2015, September 14). *Clickbait: The changing face of online journalism*. Accessed at <http://www.bbc.com/news/uk-wales-34213693>.
- Hall, S. (1982). The rediscovery of "ideology": Return of the repressed in media studies. In: M. Gurevitch, T. Bennett, J. Curran and J. Woollacott (Eds.). *Culture, society and the media*. London: Methuen.
- McQuail, D. (1994). *Mass communication theory: an introduction*. London: Sage Publ.
- Spasić, J., & Pešić, J. (2019). *U JAVNOM INTERESU; Priručnik za novinare, urednike i student novinarstva o najčešće kršenim odredbama Kodeksa novinara Srbije*. Beograd, Srbija: Savet za štampu. Accessed at <http://www.savetzastampu.rs/latinica/uploaded/Prirucnik%20U%20JAVNOM%20INTERESU%20final.pdf>
- Stovall, J. G. (2004). *Journalism: who, what, when, where, why, and how*. London: Pearson.
- Van Dijk, T. A. (1986) News Schemata. In: S. Greenbaum & Cooper, (Eds.) *Studying Writing. Linguistic approaches* (pp. 155–186) Beverly Hills, CA: Sage.
- Van Dijk, T. A. (1988b). *News as discourse*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Van Dijk, T. A. (1993). *Principles of critical discourse analysis*. *Discourse & Society*, 4(2). 249–283. doi: <https://doi.org/10.1177/0957926593004002006>

Майя Бркляч

ВАЖНОСТЬ ЗАГОЛОВКА И ЛИДА ДЛЯ ИНТЕРПРЕТАЦИИ СТАТЕЙ СМИ

Резюме

Средства массовой информации играют важную роль в формировании общественного мнения из-за их доступности и присутствия во всех аспектах жизни, а их объективность уже давно подвергается сомнению. Это исследование направлено на то, чтобы понять роль заголовков в интерпретации новостей, языковых средств, представленных в заголовках, и их цели. Цель исследования — сделать вывод, адекватно ли и точно ли представляют заголовки и лиды тексты, и какое влияние они оказывают на читателей. Тексты, являющиеся предметом анализа, поступают из разных СМИ и относятся к трём событиям первой половины 2020-ого года — началу Эпидемии коронавируса в Сербии, появлению антисербских граффити в Загребе и протестам против принятия нового Закона о свободе вероисповедания в Черногории. В исследовании также рассматривается появление кликбейтных заголовков, т. е. насколько онлайн-СМИ отличаются от печатных СМИ и насколько изменилась роль самых заголовков, от информирования читателя до привлечения как можно большего числа читателей. Анализ статей позволяет сделать вывод о том, что кликбейтные заголовки и субъективизм почти всегда присутствуют в таблоидах, реже в других СМИ, что говорит о важности развития медиаграмотности читателей.

Ключевые слова: СМИ, лид, заголовок, онлайн-портал, объективность СМИ, информация, кликбейт

Ana Lalić*

University of Sarajevo, Faculty of Philosophy

OVERCOMING THE “BAD DATA” PROBLEM: EXAMPLES OF SPOKEN LANGUAGE IN ITALIAN EPISTOLARY DISCOURSE

Abstract

This research paper focuses on the “bad data” problem that exists in historical pragmatics and discusses whether epistolary discourse might help to overcome it. The “bad data” problem is defined as a problem encountered by researchers in historical pragmatics that have to rely on written data in order to research spoken language in past eras. The goal of this research is to examine if epistolary discourse is adequate in overcoming this problem. The corpus is based on the *Lettere di Levante* (Letters from the East), a collection of 88 letters written in the first half of the fifteenth century. In this research, we focus on one letter dated July 14, 1428. In order to determine the aspects of spoken language found in the letter, the frameworks devised by Jucker (2008), Koch and Oesterreicher (1985), and Culpeper and Kytö (2010) were applied. The results seem to suggest that this particular corpus contains examples of spoken language, albeit a particular subtype – diplomatic discourse.

Key words: historical pragmatics, corpora, Italian language, koiné, spoken language source, diplomatic discourse

* ana.lalic@ff.unsa.ba

1. INTRODUCTION

This paper focuses on a specific problem in the field of historical pragmatics, the “bad data” problem, which deals with selecting an adequate corpus when conducting research on spoken language used in the past. It is suggested that an Italian epistolary corpus from the 15th century could serve as an adequate surrogate, given the lack of archival records. In order to prove this hypothesis, I provide a general outline of the “bad data” problem, and then share more about the analysis of one letter from the *Lettere di Levante* corpus in order to show the way in which such letters can reveal examples of spoken language from that historical period. The limitations of this corpus are also addressed.

2. DEFINING THE “BAD DATA” PROBLEM

Historical pragmatics is defined as a science that “focuses on language use in past contexts and examines how meaning is made. It is an empirical branch of linguistic study, with focus on authentic language use in the past” (Taavitsainen & Fitzmaurice, 2007: 13). In fact, it is a science that applies principles of both pragmatics and historical linguistics. The main scope of this science is concerned with linguistic usage from the past which offers examples of different language elements, such as implications, presuppositions, conversational maxims, politeness, or speech acts (Paternoster, 2015: 19). Historical pragmatics was born within the ambit of pragmatics as proposed by Andreas Jacobs and Andreas H. Jucker (1995), and over the past three decades it has managed to move from a marginal position to a more central one (Taavitsainen & Jucker, 2010: 4). Up until the 1990s, the reconciliation of pragmatics with historical linguistics seemed almost impossible because the latter was based on the philological analysis of written texts, predominantly literary texts. On the other hand, pragmatics was concerned with spontaneous, face-to-face interaction (Jucker, 2008: 895). Since pragmatics looks at both “live” and spoken language, it is clear that historical pragmatics should conduct research

on language that was spoken in the past. However, due to the paucity of records of spontaneous language from the past, historical pragmatics must rely on written corpora. This is what Labov (1994: 11) has termed the “bad data” problem.

Labov (1994: 11) identifies the problem of using written texts as corpora. He begins from the fact that written texts survive by happenstance, thanks to a mix of historical events and accidents. Secondly, the language used in these documents probably differs from the vernacular language used by the writers, who would presumably have expressed themselves in a manner that corresponded to the linguistic standards of their time. Many of the documents are filled with hypercorrect forms, a mix of dialects, and orthographic errors; and another problem is that we rarely know anything about the writer’s social status or the social structure of their community. Finally, even though one can read what was written, we can very rarely deduce what might have been understood by the reader. Due to all of these reasons, Labov states that historical pragmatics represents “the art of making the best use of bad data” (1994: 11).

Research conducted since Labov (1994) follows the same line of thinking. Recent research in the field of historical pragmatics, as shown in *The Historical Perspective in Pragmatics* by Jacobs and Jucker (1995), addresses the problem of corpus selection. Other research on the topic has been done by Culpeper and Kytö (2000), Jucker (1994, 2000, 2008), and Taavitsainen and Fitzmaurice (2007), all of whom echo the same sentiment: corpora in historical pragmatics are flawed, but there are ways in which they may be used when doing research. The following chapter outlines some of the ways in which the “bad data” problem might be overcome.

3. METHODOLOGY IN HISTORICAL LINGUISTICS – OVERCOMING THE “BAD DATA” PROBLEM

In this chapter, I focus on various strategies that might allow us to overcome the “bad data” problem, or to at least “make good use of” the material we have access to.

Thanks to the work of Koch and Oesterreicher (1985), which predates Labov's contributions by almost a decade, written language is no longer considered the polar opposite of spoken language. In fact, Koch and Oesterreicher (1985)¹ propose an alternative to the spoken language-written language dichotomy. They recommend the use of a graphic/phonic code dichotomy, which implies a range of possible communicative acts that might be identified on the continuum between immediate language use and distant language use. In short, this dichotomy would imply that, for example, funeral services (which are spoken) are distant, but correspondence (which is written) is immediate. Essentially, all communication can be placed on a continuum ranging from immediate to distant (Koch & Oesterreicher, 1985: 23). Thanks to this dichotomy, written texts may be used when researching spoken language as they are not texts of communicative distance. According to an interpretation by Taavitsainen and Jucker (2010:10), this model allows an analyst to define precise characteristics of communicative immediacy and distance, and to modify the crude dichotomy of simply distinguishing between written and spoken language.

Furthermore, Taavitsainen and Jucker (2010: 7) suggest two solutions to the problem:

1. The selection of a corpus should adequately represent the spoken language.
2. Even written language has communicative aspects and deserves to be the subject of research on historical pragmatics.

Essentially, the first solution implies that there are written records that approach spoken language, in that they mimic it and take on the same characteristics as spoken interaction. The second solution admits that there are limitations to using written corpora as sources when considering spoken language, but that, even though written language is also used for communication, its communicative

¹ See also Koch (1999).

mechanisms should be researched in their own right, even if they do not completely mirror spoken language.

When it comes to discussing the selection of corpora, much progress has been made by Culpeper and Kytö (1999, 2000, 2006, 2010), who propose a useful terminology that aims to distinguish between the various orality levels of written texts that approach spoken language (Culpeper & Kytö, 2010: 17). They call such texts “*speech-related*” and they identify three categories:

1. Speech-like – corpora that mimic the spoken language (for example, private correspondence) when the interlocutors are not in close physical proximity to one another;
2. Speech-based – corpora that are based on spoken language (for example, transcripts of judicial proceedings);
3. Speech-purposed – literary texts that mimic spoken language (for example, dialogues in literary or prose works, or theatrical pieces).

However, historical pragmatics must still rely on texts that do not completely capture the complexities and characteristics of spoken communication. Even though historical pragmatics faces substantial challenges, it is a flourishing field, and the existence of the *Journal of Historical Pragmatics* published by John Benjamins since the year 2000 testifies to this fact. Nevertheless, relatively little research has been done on the Italian language spoken in past eras, given that most of the research conducted so far has concerned itself with English or German.

Nonetheless, it is worth making special mention of some advances made in recent years. The importance of the topic has even been recognized by the *Associazione per la Storia della Lingua Italiana* (ASLI), which organized a conference on this very topic in 2018. The book of proceedings, *Pragmatica storica dell’italiano. Modelli e usi comunicativi del passato (Historical Pragmatics of the Italian Language. Models and communicative usage in the past²)*, was published

² All translations in the article are the author’s.

in 2020 and edited by Gabriela Alfieri, Giovanna Alfonzetti and Daria Motta. Prior to that, Annick Paternoster (2015) researched examples of historical politeness in a corpus of classical Italian literary works in *Cortesi e scortesi. Percorsi di pragmatica storica da Castiglione a Collodi (Polite and Impolite. The Courses of Historical Pragmatics from Castiglione to Collodi)*. Furthermore, Margherita di Salvo and Cristina Muru (2014) edited the volume titled *Dragomanni, sovrani e mercanti. Pratiche linguistiche nelle relazioni politiche e commerciali del Mediterraneo moderno (Dragomans, Sovrans, and Merchants. Linguistic Practices in Political and Commercial Relations in the Modern Mediterranean)*, which deals entirely with the pragmatic problems encountered when researching Mediterranean languages. Even though research done by Enrico Testa (2014) does not deal with historical pragmatics in the true meaning of the concept, he does present findings on various spoken language varieties throughout the history of the Italian language in his work, *L'italiano nascosto (Hidden Italian)*. It is also worth noting that Held (2005, 2010), Colella (2012), Alfonzetti and Spampinato Beretta (2012), Ghezzi and Molinelli (2016), and Culpeper (2017) all research different phenomena that may be observed in the Italian language through a lens of historical pragmatics.

3.1. LETTERS AS SOURCES OF SPOKEN LANGUAGE

In this section, I outline some of the characteristics of epistolary correspondence that make it a suitable surrogate for spoken language. Taavitsainen and Jucker (2007) claim that letters represent a particular type of corpus because they function on two levels – the primary level and the incorporated level – which means that letters represent a communicative act in itself, but they may also contain records of past communication. Even though letters do not always accurately represent spoken language, they show characteristics of face-to-face interaction such as salutations, and politeness formulae; and they contain clearly recognizable speech acts such as questions, apologies, offers of thanks, requests, or promises (Jucker, 1994: 535).

The amount of research on letter writing as a means of communication is substantial, but only a few are cited here at this time (Fitzmaurice, 2002; Nevala, 2004; Hakanen & Koskinen, 2009; King, 2011). While most of the research focuses on personal letters, the focus of my research is on diplomatic discourse, and it is expected that such discourse will differ from personal letter writing in terms of the characteristics displayed and formats followed.

3.1.1. The structure of a letter

According to the *Encyclopedia of Rhetoric*, letter writing belongs to the *ars dictaminis*, which is “the variety of medieval rhetoric that provided instruction in the composition of letters and other epistolary documents” (Camargo, 2006: 16). According to the same author (Camargo, 2006: 16), by the mid-twelfth century, most letters comprise the following parts:

1. *salutatio* – a greeting
2. *captatio benevolentiae* – securing “goodwill” (also called *exordium*, *arenga* or *proverbium*)
3. *narratio* – a statement of facts
4. *petitio* – a request
5. *conclusio* – a summary or complimentary close

I would also add a sixth part, the dating formula found at the end of a letter (Sarri, 2018: 121–122), sometimes called *datatio*.

As for the general structure of Ragusan diplomatic letters, Zovko (2014: 27–28) writes that letters contain conventional and recognizable parts, without specifying what they are. She also writes that the central part contains detailed instructions for the ambassadors on the attitudes they should adopt when dealing with a certain problem or question (Zovko, 2014: 27).

4. THE CORPUS – THE *LETTERE DI LEVANTE* COLLECTION

The corpus I have chosen for my study consists of 88 letters from the period 1402 to 1463. They were all addressed to ambassadors of the Republic of Ragusa who held posts in the Bosnian Kingdom. These letters are part of the *Lettere di Levante* (*Letters of the East*) collection, currently held in the Dubrovnik State Archives in Croatia. The letters were written in an Italian koiné³ – the official language used by members of the Ragusan chancellery. For the purpose of this paper, I have chosen the following letter: 14.7.1428, Lett. di Lev., X, 95v. (July 14, 1428, *Lettere di Levante*, X, 95v.). This letter is suitable for the analysis because it presents the typical structure without any deviations, because of historical significance, and presence of speech acts. Furthermore, due to its brevity, I can analyze it in its entirety. This 29-line letter congratulates King Tvrtko II Kotromanić and his future wife, Doroteja Gorjanska, on their wedding, and it will be demonstrated that even a short letter of this kind can contain examples of spoken language.

5. ANALYSIS OF THE LETTER AND ITS STRUCTURE

| Italian original | English translation |
|--|--|
| Rector di Ragusa col suo consiglio alli nobili et saui ser Nicola Mat. de Georgi (et) ser Marin Ja. de Gondola ambascadori al re di Bosna dilecti cittadini nostri salute. | The Ragusan Rector – along with his council of noble and sage sirs – sends his greetings to Nicola Mat. de Georgi and Marin Ja. de Gondola, Ambassadors to the Bosnian king and our esteemed citizens. |

³ *Koiné* is a term used to define a supraregional language in Italian chancelleries in use from the 13th century (Lubello, 2014: 229). This language variety abandons local language characteristics and gradually stabilizes a uniform model, comprehensible in Italian-language speaking areas (Tavoni, 1992: 47–48).

This excerpt is the opening of the letter. It states that the Republic of Ragusa sends greetings to ambassadors Nicola de Georgi and Marin de Gondola, and makes use of the following formulae: the *salutatio* and the *captatio benevolentiae*. It is evident from the use of the *syntagme* “*alli /.../ salute*” that it is a salutation to these gentlemen. I would also submit that the adjectives *nobili* (noble) and *savi* (sage) represent the act of securing the goodwill of the ambassadors, to whom the letter is addressed, by complimenting them and praising their abilities.

| Italian original | English translation |
|--|---|
| Per le letere vostre fatte di la adi VI luglio presente le qual adi XIII ^o del detto reciuessimo, ne scriuete che adi sei de questo per uno corier che mandassimo, scriuessimo alla signoria vostra le qual vostre lettere non abiamo ricevute et molto di nostro piacer seria stado auanti queste auer le ricevute per esser informati meglio. Et de piu ne auisate, come aueti cominzado a cortizare lo re per lo partir vostro di la secondo vi e scritto. Et anche come a mandato suo zentilomeni in contra la nouiza et se aparechia de far lo noze. | You wrote to us from there in your letters dated July 6, which we received on the fourteenth of the same month that you wrote to us, through a courier that we sent, but we never received your letters and it would have been our great pleasure to have received these letters earlier so that we could have been better informed. And, furthermore, you let us know that you began to court the king when you left, as you had been instructed. And also how the king sent his noblemen to meet the bride and how he is preparing for his wedding. |

This excerpt is the *narratio*, where by the Republic of Ragusa refers to events which happened prior to the time that the letter was written. Namely, that the ambassadors had sent letters which were not received in time and that they had heard that the king was preparing for his wedding.

| Italian original | English translation |
|---|---|
| <p>/.../ vogliamo che dobiate ritornar in drie ala maiesta del re. Et se ancora riceuando questa fosseno apreso de luy vogliamo pur che com esso re dobiate remanir e caualcar et andare allegrarui de la festa sua fazando scusa et digando che la signoria nostra di Ragusa non sapiando ne abiando di certo in che ancora volesse far la festa de la nouiza et noçe.</p> | <p>We would like for you to return to His Majesty the King. And if you are still with him when you receive this letter, we would like you to remain with this king, to ride with him and rejoice because of his celebration, and to apologize and say that our <i>signoria</i> of Ragusa did not know for certain that he was going to celebrate his wedding.</p> |

This excerpt contains the *petitio*, or the expression of a request. In fact, in this section, the ambassadors are being commissioned to perform certain duties and behave in certain ways. Even though the entire central part describes the attitude the ambassadors are to adopt, I chose this paragraph in order to illustrate the official attitude of the Republic toward the Bosnian king – that his wedding celebrations bring joy to its neighboring country and that the ambassadors are to express this pleasure.

| Italian original | English translation |
|--|--|
| <p>Et di la aspetate nostra risposta. Abiamo accettato de pagar yperperos CC che vi a datto Vochaç Mora. Daremo modo che auereti altri denari di la.</p> | <p>And there you shall await our reply. We have agreed to pay the 100 perpers that were given to you by Vukac Mura. We shall find a way for you to receive more money.</p> |

This is the *conclusio*, the final remarks which close the message and give an order to the ambassadors for them to perform the final task.

| Italian original | English translation |
|--|--|
| <p>Datum Ragusii, die XIII jullii 1428</p> | <p>Dated July 14, 1428, in Ragusa.</p> |

The final formula included in this letter is the *datatio*, which closes the document and states when and where the letter was written.

5.1. ANALYSIS OF SPOKEN LANGUAGE CHARACTERISTICS

In this section, I analyze the characteristics of spoken language, as contained within the letter. I looked for examples of the aforementioned speech acts⁴ that can be found in letters and serve as examples of spoken language, and I have identified the following:

– Greeting:

According to Searle (1969: 67) salutations mark the beginning and end of conversations, they are often ritualized, and are void of all propositional content. Such speech acts often involve the use of a limited number of crystallized formulae (Jucker, 2017: 40). This speech act can be identified in the following example:

- (1) Rector di Ragusa col suo consiglio alli nobili et saui ser Nicola Mat. de Georgi (et) ser Marin Ja. de Gondola ambascadori al re di Bosna dilecti cittadini nostri salute.

(The Ragusan Rector – along with his council of noble and sage sirs – sends his greetings to Nicola Mat. de Georgi and Marin Ja. de Gondola, Ambassadors to the Bosnian king and our esteemed citizens.)

In this example, the speech act of greeting is recognizable because of the use of the formula *salute a* (greetings to), as underlined in the text. It is also evident that the speech act is accompanied by appellatives, a common feature among greetings, since the ambassadors' names are used (Jucker, 2017).

– Reprimanding:

Reprimanding can be defined as an illocutionary speech act, whereby the speaker indicates disapproval with or a negative feeling

⁴ Since speech act theory and its application in diachrony is not the main focus of this paper, it is not discussed as a theoretical approach. However, the speech acts identified in the corpus will be considered briefly. For more on the topic of speech acts, see Austin (1962), and Searle (1969, 1976, 1979). For more on the application of speech act theory to historical pragmatics, see Jacobs and Jucker (1995), Arnovick (1999), Bertucelli Papi (2000), Kohnen (2008), and Lalić (2020).

toward the state of affairs described in the proposition, and for which they hold, either directly or indirectly, the hearer accountable (Trosborg, 1995: 311). According to House and Kasper (1981: 159–167), four criteria may be used to identify complaints: an action is mentioned explicitly or implicitly, the negative valuation of the action is explicitly stated, the hearer’s active involvement is stated explicitly or implicitly, and the negative valuation of the hearer is also expressed explicitly or implicitly. I identified one complaint in the letter:

(2) /.../ le qual vostre lettere non abiamo ricevute et molto di nostro piacer seria stado auanti queste auer le ricevute per esser informati meglio.

(/.../ but we never received your letters and it would have been our great pleasure to have received these letters earlier so that we could have been better informed.)

I would consider this excerpt a reprimand because it states the Republic of Ragusa’s displeasure with the behavior of its ambassadors. The action – the fact that the letters were not received – is explicitly stated; the negative valuation of the action is also explicitly stated (the consequences were negative because the Rector and Councils were not informed in time). The hearer is explicitly involved because the reprimand is directed at the ambassadors. I would also submit that the ambassadors are being held accountable because the complaint is being directed at them (explicitly) and the missing letters are referred to as “your letters” (*vostre lettere*).

– Apologizing:

According to Austin (1962), the speech act of apologizing consists of: 1) the speaker showing remorse for a past action, and 2) a promise to behave better in the future. It is easily recognizable because of the use of typical formulae. In order for a speech act to be classified as such, it must meet the following criteria: an act was committed by the speaker, the hearer suffers or the speaker believes that they are suffering because of the act, and the speaker feels remorse. An apology can be found in the following example:

- (3) /.../ fazando scusa et digando che la signoria nostra di Ragusa non sapiando ne abiando di certo in che ancora volesse far la festa de la nouiza et noçe.
(/.../and to apologize and say that our *signoria* of Ragusa did not know for certain that he was going to celebrate his wedding.)

The act in question is that the Republic of Ragusa did not know in time that the king was getting married and they fear that he may be offended. Such an offense, and the cause of the hearer's suffering, may be what the Republic is apologizing for. The apology issued by the Republic of Ragusa through the ambassadors is an expression of remorse for the offense. The speech act is recognizable because the typical formula for apologizing is used, in this case: *fare scusa* (*fazando scusa* in the text, *facendo scusa* in modern Italian). What can be noted is that, besides the apology to the Bosnian king, the Republic of Ragusa also offers some explanations for its past behavior.

– Making a request

Searle (1979: 13) defines requests as speech acts that get the hearer to do something. Flöck and Geluykens (2015: 16) describe three levels of requests: direct, conventionally indirect, unconventionally indirect (or only indirect). In this letter, I extrapolated requests belonging to the first level and they are recognizable because of typical forms, such as the imperative, which are often associated with requests (Flöck and Geluykens, 2015: 16). Here I draw attention to the following examples:

- (4) /.../ vogliamo che dobiate ritornar /.../
(/.../ we want you to return /.../)
- (5) Et ariuando la nouiza nello terreno di Bosna subito con presteza per vostre lettere ne auisate del suo ariuar.
(And when the bride arrives in Bosnia, we want you to let us know right away about her arrival in your letter.)
- (6) Et di la aspetate nostra risposta.
(And there you shall await our reply.)

In example number four (4), a request is recognizable because of the use of the verb *volere* (to want) in the first person plural, which indicates that the following action is an order given by the state. In the next two examples, (5) and (6), requests are recognizable in the use of the imperative mood, *avisate* (let someone know) and *aspetate* (wait), which indicate that an order is being given.

As was expected, requests are the most numerous type of speech acts found in the letter because they are, by definition, instructions to the ambassadors. Other types of requests may exist in this corpus, but they are not present in this particular letter and their presence and modalities will be the subject of further research.

– Congratulating

According to Searle (1969: 67), congratulations are expressed when there is a happy event that relates to the hearer, when the speaker believes that the event is happy for the hearer, and when the hearer is content with the event. Here, the offer of congratulations counts as an expression of the speaker's pleasure with a past event. The following may be identified as a congratulatory message:

(7) /.../ siando avisata de le noze et festa de la nouiza vostra /.../
dobbiamo esser con la maiesta vostra ad allegrarssi et far festa
con essa de la sua festa et noze si come cordiali zelatori et
amici de la vostra serenita.

(/.../ seeing as we were informed about your wedding and your
nuptial celebrations /.../ we must be with Your Majesty and
rejoice and celebrate with you your feast and your wedding
as cordial supporters and friends of your serenity.)

The event in question is the wedding between King Tvrtko II Kotromanić and his bride, Doroteja Gorjanska. Weddings and wedding celebrations are usually considered happy events for the people getting married, and it is only natural that congratulations are in order on such occasions. This excerpt demonstrates the pleasure of the speaker through the use of the verb *allegrarssi* (to rejoice), which is *allegrarsi* in modern Italian, and the expression *far festa* (to

celebrate), both of which demonstrate that the speaker is pleased with the event and wishes to share in the hearer's happiness.

6. CONCLUSION

It is clear that this letter follows the structure of written correspondence in the Middle Ages, as I demonstrated the use of all six of the essential elements of a typical medieval letter. Furthermore, even on the basis of one short letter from the sample, a number of speech acts could be identified: greeting, reprimanding, directing someone to do something, congratulating and apologizing. I also analyzed these speech acts using frameworks devised specifically for spoken language. This letter shows that directives are the predominant speech acts, which was also to be expected, since the letters contain directives issued to the ambassadors. The goal of future research will be to examine a larger corpus and determine the characteristics of the speech acts, their structures, and frequencies of use. It is expected that the use of speech acts will be in accordance with the structure and conventions of diplomatic correspondence.

Upon closer analysis, the structure of the letter clearly shows the limitations of the corpus. Diplomatic discourse follows specific rules and has certain structures, which implies that these letters do not necessarily, or at all, reflect the spoken vernacular, but rather a specific diplomatic register which is characterized by a certain set of particular rules that differ from the vernacular. Also, the rigid structure implies that the spontaneity of the writers' expression cannot be detected as readily as in personal letters. Given the fixed structure of a diplomatic letter, it is not possible to state that this particular corpus overcomes the problem fully. However, I would submit that the correspondence approaches the spoken language in such a way that it allows us to research the characteristics of diplomatic discourse during this period. Despite the lack of more suitable corpora, Italian diplomatic epistolary discourse may be considered a suitable surrogate.

Source

Državni arhiv u Dubrovniku, 14.7. 1428, *Lettere di Levante*, X, 95v.

REFERENCES

- Alfieri, G. et al. (2020, eds.). *Pragmatica storica dell'italiano. Modelli e usi comunicativi del passato*. Firenze: Franco Cesati.
- Alfonzetti, G. & Spampinato Beretta, M. (2012). Gli insulti nella storia dell'italiano. Analisi di testi del tardo medioevo. In: *Pragmatique historique et syntaxe. Historische Pragmatik und Syntax*, (Wehr, B. and F. Nicolsi, eds.), Frankfurt: Peter Lang, 1–21.
- Arnovick, L. K. (1999). *Diachronic Pragmatics. Seven Case Studies in English Illocutionary Development*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Oxford University Press.
- Bertucelli Papi, M. (2000). Is a diachronic speech act theory possible?. *Journal of Historical Pragmatics*, 1, 57–66.
- Camargo, M. (2006). Ars dictaminis. In: *Encyclopedia of Rhetoric* (Sloane, T. O., eds.), Oxford: Oxford University Press, 60–61.
- Colella, G. (2012). Performativi espliciti in italiano antico. In: *Pragmatique historique et syntaxe. Historische Pragmatik und Syntax: Actes de la section du même nom du XXXIe Romanistentag allemand. Akten der gleichnamigen Sektion des XXXI. Deutschen Romanistentags* (Wehr, B. and F. Nicolsi, eds.), Frankfurt: Peter Lang, 23–46.
- Culpeper, J. (2017). The influence of Italian manners on politeness in England, 1550–1620. *Journal of Historical Pragmatics*, 18 (2), 195–213.
- Culpeper, J. & Kytö, M. (1999). Modifying pragmatic force: hedges in a corpus of early modern English dialogues. In: *Historical Dialogue Analysis* (Jucker, A. H. et al., eds.), Amsterdam: John Benjamins, 293–312.
- Culpeper, J. & Kytö, M. (2000). Data in historical pragmatics: Spoken interaction (re)cast as writing. *Journal of Historical Pragmatics*, 1:2, 175–199.
- Culpeper, J. & Kytö, M. (2006). “Good, good indeed, the best that ere I heard”: Exploring lexical repetitions in the Corpus of English Dialogues 1560–1760. In: *Dialogic Language Use/Dimensions du dialogisme/Dialogischer Sprachgebrauch, Mémoires de la Société Néophilologique de Helsinki, Vol. 66* (Taavitsainen, I. et al., eds.), Helsinki: Société Néophilologique, 69–85.

- Culpeper, J. & Kytö, M. (2010). *Early Modern English Dialogues: Spoken Interaction as Writing*. Cambridge: Cambridge University Press.
- Fitzmaurice, S. M. (2002). *The Familiar Letter in Early Modern English. A pragmatic approach*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Flöck, I. & Geluykens, R. (2015). Speech Acts in Corpus Pragmatics: A Quantitative Contrastive Study of Directives in Spontaneous and Elicited Discourse. In: *Yearbook of Corpus Linguistics and Pragmatics: Current Approaches to Discourse and Translation Studies* (Romero-Trillo, J., eds.), New York: Springer Publishing, 7–37.
- Ghezzi, C. & Molinelli, P. (2016). Politeness markers from Latin to Italian: Periphery, discourse structure and cyclicity. *Journal of Historical Pragmatics*,17(2), 307–337.
- Hakanen, M. & Koskinen, U. (2009). From “friends” to “patrons”. Transformations in the social power structure as reflected in the rhetoric of personal letters in sixteenth- and seventeenth-century Sweden. *Journal of Historical Pragmatics*,10:1, 1–22.
- Held, G. (2005). Politeness in Italy. In: *Politeness in Europe* (Hickey, L., eds.), Clevedon/Buffalo/Toronto: Multilingual Matters, 292–305.
- Held, G. (2010). Supplica la mia parvidade... Petitions in medieval society – a matter of ritualised or first reflexive politeness?. *Journal of Historical Pragmatics*,11:2, 194–218.
- House, J. & Kasper, G. (1981). Politeness markers in English and German. In: *Conversational routine: explorations in standardized communication situations and prepatterned speech* (Coulmas, F., eds.), Berlin/New York: Mouton de Gruyter, 157–185.
- Jacobs, A. & Jucker A. H. (1995). The Historical Perspective in Pragmatics. In: *Historical Pragmatics* (Jucker, A. H., eds.), Amsterdam: John Benjamins Publishing Company, 3–33.
- Jucker, A. H. (1994). The feasibility of historical pragmatics. *Journal of Pragmatics*, 22: 533–536.
- Jucker, A. H. (2000). English historical pragmatics: Problems of data and methodology. In: *English diachronic pragmatics* (Martino, di, G. and M. Lima, eds.), Napoli: CUEN, 17–55.
- Jucker, A. H. (2008). Historical pragmatics. *Language and Linguistics Compass* 2.5: 894–906.

- Jucker, A. H. (2017). Speech Acts and Speech Act Sequences: Greetings and Farewells in the History of American English. *Studia Neophilologica*, vol. 89: 39–58.
- King, J. (2011). Variation through time and text type: The nature of direct and indirect requests in Early Modern Spanish. *Spanish in Context*, no. 8: 272–294.
- Koch, P. (1999). Court Records and Cartoons Reflections of Spontaneous Dialogue in Early Romance Texts. In: *Historical Dialogue Analysis* (Jucker, A. H., eds.), Amsterdam/Philadelphia: John Benjamins, 399–430.
- Koch, P. & Oesterreicher, W. (1985). Sprache der Nähe — Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36: 15–43.
- Kohnen, T. (2008). Tracing directives through text and time: Towards a methodology of a corpus-based diachronic speech-act analysis. In: *Speech Acts in the History of English* (Jucker, A. H. and I. Taavitsainen, eds.), Amsterdam: John Benjamins Publishing, 295–310.
- Labov, W. (1994). *Principles of Language Change. Volume 1: Internal Factors*. Oxford: Blackwell.
- Lalić, A. (2020). La realizzazione degli atti linguistici nei testamenti medievali. *Italica Belgradensia 1*: 53–70.
- Lubello, S. (2014). Cancelleria e burocrazia. In: *Storia dell'italiano scritto. III. Italiano dell'uso* (Antonelli, G.etal., eds.), Roma: Carocci, 225–437.
- Nevala, M. (2004). Inside and out. Forms of address in seventeenth- and eighteenth-century letters. *Journal of Historical Pragmatics*, 5:2, 271–296.
- Patronoster, A. (2015). *Cortesi e scortesi. Percorsi di pragmatica storica da Castiglione a Collodi*. Roma: Carocci.
- Salvo, di, M. & Muru, C. (2014, eds.). *Dragomanni, sovrani e mercanti. Pratiche linguistiche nelle relazioni politiche e commerciali del Mediterraneo moderno*. Pisa: Edizioni ETS.
- Sarri, A. (2018). *Material Aspects of Letter Writing in the Graeco-Roman World: 500 BC – AD 300*. Berlin/New York: Mouton de Gruyter.
- Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Searle, J. R. (1976). A Classification of Illocutionary Acts. *Language in Society* 5: 1–23.
- Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

- Taavitsainen, I. & Fitzmaurice, S. (2007). Historical pragmatics: What it is and how to do it. In: *Methods in Historical Pragmatics* (Fitzmaurice, S. & Taavitsainen, I., eds.), Berlin/New York: Mouton de Gruyter, 11–36.
- Taavitsainen, I. & Jucker, A. H. (2007). Speech act verbs and speech acts in the history of English. *Methods in Historical Pragmatics* (Fitzmaurice, S. & Taavitsainen, I., eds.), Berlin/New York: Mouton de Gruyter, 107–138.
- Taavitsainen, I. & Jucker, A. H. (2010). Trends and developments in historical pragmatics. In: *Historical Pragmatics* (Jucker, A. H. & Taavitsainen, I., eds.), Berlin/New York: Mouton de Gruyter, 3–30.
- Tavoni, M. (1992). *Storia della lingua italiana. Il Quattrocento*. Bologna: il Mulino.
- Testa, E. (2014). *L'italiano nascosto. Una storia linguistica e culturale*. Torino: Piccola Biblioteca Einaudi.
- Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints, and apologies*. Mouton de Gruyter: Berlin.
- Zovko, V. (2014). Metode i tehnike komunikacije između vlasti i poslanika u pregovorima oko proširenja dubrovačkih granica. *Anali Dubrovnika*, 52(1): 21–49.

Ana Lalić

COME SORPASSARE IL PROBLEMA DEI “DATI ERRATI”. ESEMPI DELLA LINGUA PARLATA IN DISCORSO EPISTOLARE IN ITALIANO

Riassunto

In questa ricerca focalizziamo il problema dei “dati errati” nella pragmatica storica. Il problema dei “dati errati” è definito come il problema dei ricercatori nel seno della pragmatica storica che devono appoggiarsi sui dati scritti nelle ricerche della lingua parlata in prospettiva diacronica. Se prendiamo in considerazione che fonti registrate delle epoche passate non esistono, i ricercatori dipendono dalle fonti che si avvicinano alla lingua parlata. L'obiettivo della ricerca è esaminare se il discorso epistolare sia adeguato a sorpassare il problema dei dati errati. Il corpus consiste di 88 lettere dalla collezione *Lettere di Levante* della prima metà del Quattrocento. Tutte le lettere sono redatte dalla Repubblica di Ragusa e dirette agli ambasciatori nel Regno di Bosnia. Nel determinare le caratteristiche della lingua parlata, utilizziamo il quadro escogitato da Jucker (2008), Koch e Oesterreicher

(1985) e Culpeper e Kytö (2010). I risultati della ricerca implicano che questo particolare *corpus* si avvicina alla lingua parlata, ma a un sottotipo particolare – il discorso diplomatico.

Parole chiave: pragmatica storica, corpus, lingua italiana, koinè, fonte di lingua parlata, discorso diplomatico

Milena Oparnica*

Natalija Panić Cerovski

University of Belgrade – Faculty of Philology

THE USE OF DIMINUTIVES IN EVERYDAY COMMUNICATION

Abstract

The main goal of this paper is to analyse diminutives within spoken corpora with respect to morphology and pragmatics. We analyse the formation process of diminutives, as well as their pragmatic functions within the discourse.

The corpus being used for this research has been collected manually and contains transcripts of everyday communication among students (about 250,000 tokens), as well as transcripts of TV talk shows on politics (about 173,000 tokens). For the purpose of the extraction of diminutives we have created a simple script in Python.

Additionally, we are interested in exploring whether there is a difference in the use of diminutives within two different spheres of discourse – political discourse and everyday communication among students.

Keywords: diminutives, morphopragmatics, pragmatic functions, corpus analysis, spoken discourse

* milena.oparnica@fil.bg.ac.rs
natalija.panic@fil.bg.ac.rs

1. INTRODUCTION

The term *diminutive* is usually defined as a category which denotes smallness and affection. However, diminutives have many different pragmatic functions within the discourse (cf. e.g. Savickiene & Dressler, 2007). In addition to expressing affection and endearment, diminutives are used with pejorative and augmentative meanings. Pragmatic meanings are usually interpreted as an extension of the basic semantic meaning, i.e. *smallness*, along with its connotations related to emotions. On the contrary, others view pragmatic meanings as fundamental (cf. Dressler & Merlini Barbaresi, 1994; Jurafsky, 1996).

Since diminutives are widely used in Serbian, our goal was to analyse the spoken corpus in order to find the answers to the following questions:

- a) Are there differences in the use of noun diminutives in the different sub-corpora of spoken language?
- b) What are the pragmatic meanings of diminutives within the corpus?
- c) What are the most frequent suffixes?

For the purpose of this research, we have used manually collected transcripts of spoken discourse related to two different spheres of discourse: students' communication and political talk shows.

2. DIMINUTIVES IN SERBIAN

Serbian is an inflection-rich Indo-European language. Some authors (cf. Grickat, 1995) claim that different parts of speech in Serbian can be diminutized. Grickat (1995) claims that noun diminutization (e.g. „satić”, *small watch*) refers to space dimension, adjective diminutization refers to qualitative dimension (e.g. „slankast”, *a bit salty*), and verb diminutization (e.g. „leškariti”, *lying down a bit*) refers to time and qualitative dimension. However, our focus within this

paper will be noun diminutives. Therefore, we present here Serbian noun morphology related to the formation of diminutives.

In Serbian, nouns are traditionally classified with regard to masculine, feminine and neuter gender. Nominal diminutives are formed with suffixes, which can be divided to those forming masculine, feminine and neuter gender nouns.

| Masculine | Feminine | Neuter |
|---|--|---|
| – (ič/ar/er/elj/ulj/et)ak: “plam-ičak” <i>little flame</i> “šum-arak” <i>grove</i> “put-eljak” <i>path</i> “smot-uljak” <i>roll</i> “cvet-ak” <i>little flower</i> | -(č)ica: “trav-čica” <i>little blade of grass</i> “trav-ica” <i>little grass</i> | -(an/aš/en/eš)ce: “tel-ašce” <i>little body</i> “tel-ence” <i>little calf</i> “ogledal-ce” <i>small mirror</i> |
| -(č)ić: “kamen-čić” <i>pebble</i> | -ca: “reč-ca” <i>particle</i> | -če: “čaj-če” <i>tea-AFFECTION</i> |
| -ac: “brat-ac” <i>little brother</i> | -ka: “trav-ka” <i>blade of grass</i> | |

Table 1. Diminutive suffixes with examples

Usually, diminutive suffixes preserve grammatical features of nouns. However, feminine, masculine and neuter nouns sometimes take the same diminutive suffixes. For instance, “šuma” *forest* is feminine in Serbian. Common feminine diminutive suffix is *-ica*. However, in addition to f. “šum-ica” *little forest*, we can form a diminutive with the suffix *-arak*, and thus form a masculine noun. In the dictionary (Николић, 2011), the noun m. “šum-arak” *little forest* is defined as *little, usually lower, young and thick forest*. Within the same dictionary, the noun f. “šum-ica” *little forest* is defined as *little, young forest*. Additionally, the diminutive noun m. “brež-uljak” *little hill*, is derived from the neuter noun “breg” *hill*.

A great majority of diminutives ending in *-ak* and *-če* have their forms in *-ić*, *-čić* or *-ica* if feminine (Đurić, 2004, p. 151). Đurić (2004, p. 151) claims that those diminutives in *-ak* and *-če* have *hypocoristic, emotional or poetic meaning*, if not lexicalized, while those in *-ić*, *-čić* or

-*ica* primarily refer to the physical meaning of smallness, developing additional, secondary, pragmatic meanings. Within our paper, we will explore the meanings of diminutives found in our corpus.

Nominal diminutives can also be derived from verbal stems. For instance, the diminutive “smot-uljak” from the transitive perfective verb “smotati” *to fold*, means *something small and folded*. It should be noted that even though the verb “smotati” *to fold* does not contain diminutive meaning, the noun “smotuljak”, referring to the object of folding, always has a diminutive meaning. There are different levels of productivity for different diminutive suffixes. In this paper we do not investigate levels of productivity of diminutive suffixes in Serbian, since we are focused here primarily on the usage of diminutives and their frequency in conversations. The productivity of diminutive suffixes is a very important and interesting subject to investigate and it will be tackled in our future work.

3. CORPUS AND METHOD

The spoken corpus we used has been collected manually. It consists of about 423,000 tokens: the part of private everyday conversations (students and young people) transcripts contains about 25,000 tokens, and the part of public conversations (TV talk shows) transcripts contains about 173,000 tokens (Knjižar, 2019).

Given the fact that the corpus is unannotated, we devised a simple python script for extracting diminutives based on regular expressions. First, we defined regular expressions for all the suffixes that exist in grammar books (cf. e.g. Stanojčić & Popović, 2004). The expression in (A) matches neuter diminutives, (B) matches feminine and (C) matches masculine diminutives.

- A. `re.match('(.*(en|(a|e)š)ce(ta|tu)?$)|(.*če(ta|tu)?$)|(.*čad(i|ima)?$)')`
- B. `re.match('(.*ic(a|ma|e|i|o|u)$)|(.*^[aeiou]k(a|i|e|o|u|.ma))$')`
- C. `re.match('(.*ić(a|u|em|om|i|ima|e)?|.*^[aeiou]c(ima|i)|.*ak$|.*^[aeiou]k(a|u|kom)$|.*^[aeiou]c(ima|i)|.*ak$|.*^[aeiou]k(a|u|kom)$')`

Second, given that regular expressions find a great number of tokens which are not diminutives (e.g. “priče” *stories*, “juče” *yesterday*, “kolače” *cookies*), we used an electronic dictionary of about 2000 diminutives (Vitas & Krstev, 2012; Krstev, 2008), to find matches. We created a simple stemmer in order to normalize the results (see examples 4–7 for neuter diminutives), before the comparison to the dictionary is executed.

- D. `norm = series_diminutives.str.replace("ence(ta|tu)?$", "ence", regex=True)`
- E. `norm = series_diminutives.str.replace("ešce(ta|tu)?$", "ešce", regex=True)`
- F. `norm = series_diminutives.str.replace("ce(ta|tu)?$", "ce", regex=True)`
- G. `norm = series_diminutives.str.replace("če(ta|tu)?$", "če", regex=True)`

Finally, we manually checked all the results in order to find diminutives that exist in our corpus, but do not exist within the dictionary. Additionally, we manually checked whether diminutives were lexicalized or not, and removed the lexicalized from the list (see Table 2).

4. RESULTS

We have found about one hundred and fifteen diminutive tokens in our corpus (~ 0.0002% of all tokens). The following graph shows the frequency distribution of diminutive suffixes. First, we will analyse examples from the perspective of morphology, and second, from the perspective of pragmatics.

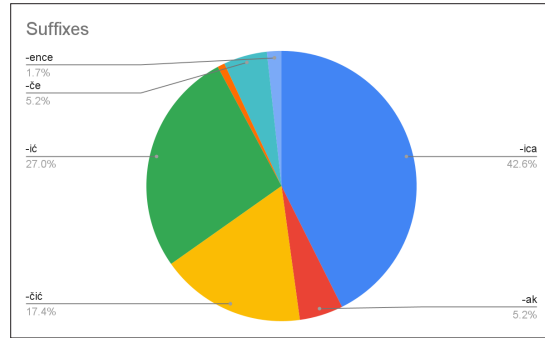


Figure 1. The distribution of diminutive suffixes within the corpus

The following is the list of diminutives found within the corpora, their absolute and relative frequencies, as well as the type of corpus they appeared.

| lemma | a_frequency | r_frequency | suffix | corpus |
|------------------|-------------|-------------|--------|------------|
| šamarčić | 1 | 0.0006% | -čić | political |
| staklič | 1 | 0.0006% | -ič | political |
| šefič | 1 | 0.0006% | -ič | political |
| nesposobnjaković | 1 | 0.0006% | -ič | political |
| detaljčić | 1 | 0.0006% | -čić | political |
| papirić | 2 | 0.0005% | -ič | colloquial |
| gradić | 1 | 0.0002% | -ič | colloquial |
| dečkić | 1 | 0.0002% | -ič | colloquial |
| mužić | 1 | 0.0002% | -ič | colloquial |
| vetrić | 1 | 0.0002% | -ič | colloquial |
| ćevapčić | 2 | 0.0005% | -čić | colloquial |
| vinjačić | 1 | 0.0002% | -ič | colloquial |
| konjačić | 1 | 0.0002% | -ič | colloquial |
| parkić | 1 | 0.0002% | -ič | colloquial |

| | | | | |
|------------|----|---------|-------|------------|
| stočić | 15 | 0.0035% | -čić | colloquial |
| draguljčić | 1 | 0.0002% | -čić | colloquial |
| štapić | 3 | 0.0007% | -ić | colloquial |
| komadić | 1 | 0.0002% | -ić | colloquial |
| stikerčić | 1 | 0.0002% | -ić | colloquial |
| evrić | 3 | 0.0007% | -ić | colloquial |
| mišić | 1 | 0.0002% | -ić | colloquial |
| đavolak | 2 | 0.0005% | -ak | colloquial |
| sokić | 2 | 0.0005% | -ić | colloquial |
| osmejak | 1 | 0.0002% | -ak | colloquial |
| mesečak | 1 | 0.0002% | -ak | colloquial |
| kamičak | 2 | 0.0005% | -ak | colloquial |
| ormarić | 1 | 0.0002% | -ić | colloquial |
| prstić | 1 | 0.0002% | -ić | colloquial |
| lančić | 2 | 0.0005% | -ić | colloquial |
| testić | 1 | 0.0002% | -ić | colloquial |
| parčić | 1 | 0.0002% | -ić | colloquial |
| mrvica | 2 | 0.0005% | -ica | colloquial |
| pčelica | 1 | 0.0002% | -ica | colloquial |
| tetkica | 1 | 0.0002% | -ica | colloquial |
| čašica | 3 | 0.0007% | -ica | colloquial |
| kućica | 2 | 0.0005% | -ica | colloquial |
| kesica | 1 | 0.0002% | -ica | colloquial |
| dušica | 1 | 0.0002% | -ica | colloquial |
| dečica | 2 | 0.0005% | -ica | colloquial |
| kiflica | 2 | 0.0005% | -ica | colloquial |
| detence | 2 | 0.0005% | -ence | colloquial |

| | | | | |
|-------------|---|---------|------|------------|
| letvica | 1 | 0.0002% | -ica | colloquial |
| gitarica | 2 | 0.0005% | -ica | colloquial |
| batica | 1 | 0.0002% | -ica | colloquial |
| pljačkica | 1 | 0.0002% | -ica | colloquial |
| tačkica | 4 | 0.0009% | -ica | colloquial |
| krmenadlica | 1 | 0.0002% | -ica | colloquial |
| pundica | 1 | 0.0002% | -ica | colloquial |
| bubica | 1 | 0.0002% | -ica | colloquial |
| voćkica | 1 | 0.0002% | -ica | colloquial |
| cipelica | 2 | 0.0005% | -ica | colloquial |
| štiklica | 1 | 0.0002% | -ica | colloquial |
| klupče | 1 | 0.0002% | -če | colloquial |
| torbica | 1 | 0.0002% | -ica | colloquial |
| resica | 2 | 0.0005% | -ica | colloquial |
| nogica | 2 | 0.0005% | -ica | colloquial |
| zvezdica | 2 | 0.0005% | -ica | colloquial |
| vežbica | 3 | 0.0007% | -ica | colloquial |
| sličica | 1 | 0.0002% | -ica | colloquial |
| knjižica | 1 | 0.0002% | -ica | colloquial |
| jarče | 2 | 0.0005% | -če | colloquial |
| klinkica | 1 | 0.0002% | -ica | colloquial |
| restoranče | 1 | 0.0002% | -če | colloquial |
| tenkić | 2 | 0.0005% | -ić | colloquial |
| spravica | 1 | 0.0002% | -ica | colloquial |
| prozorče | 1 | 0.0002% | -če | colloquial |
| mukica | 1 | 0.0002% | -ica | colloquial |
| prvenče | 1 | 0.0002% | -če | colloquial |

| | | | | |
|----------|---|---------|-------|------------|
| grudvica | 1 | 0.0002% | -ica | colloquial |
| čikica | 2 | 0.0005% | -ica | colloquial |
| brdašce | 1 | 0.0002% | -ašce | colloquial |

Table 2. Diminutives, their absolute frequency and type of corpus they appeared in

4.1. MASCULINE DIMINUTIVES

Masculine noun diminutive forms we have found in our corpus ended in *-ak* and *-(č)ić*. 5.2% of all diminutive types ended in *-ak*, while 44.4% of all diminutive types ended in *-(č)ić*. Even though traditional grammar recognizes suffix *-ac* as a masculine diminutive suffix, we have not found diminutives formed with *-ac* within the corpus. As Jovanović (2004) notes, it seems that the suffix *-(a)c* ceased to be used for the formation of diminutives and its current use seems to be limited to literature and otherwise marked context.

Here, we provide examples found in the corpus.

- (1) Bio sam pravi mali **đavol-ak**.

I was a real little **devil-DIM**.

When the speaker says he was a little devil in Serbian, the hearer usually perceives him as being mischievous, naughty, but in a likeable, funny, playful way. The usual connotation for the use of this diminutive form is child-related behaviour. Hence, the negative connotation of being a “devil” is not present. Here, the diminutive form of the noun “đavo” intensifies the *cuteness* of being a little devil and at the same time diminishes the unpleasantness and intensity of the qualification of being a devil.

- (2) Pa može negde, neka kafana ili nek ili ne znam tako neki **restoran-čić**, videćemo,...

Well, yes, we can go somewhere, to a tavern or some or I don't know some **restaurant-DIM**, we'll see...

Here, the diminutive form of the noun “restaurant” is being used in order to denote that it can be *something similar to the restaurant*, but something which is less formal than a restaurant. The speaker refers, in other words, to some cosy place where food and drinks are served, and where guests can drop by without a previous plan and reservations, in an informal way, and thus they perceive such a place as more comfortable.

- (3) ... i uvek imate mali osmej-ak...
... and you always have that little **smile-DIM**...

A *little smile-DIM* denotes here a compliment for a hearer. The diminutive form of the noun “osmeh” *smile* added an additional meaning of agreeable quality, warmth, cuteness and affection. Moreover, a speaker in this example uses an adjective “little” to describe the “smile-DIM” which strongly suggests that “smile-DIM” is not felt as a smile that is not broad, but rather as a smile that is nice, cute, agreeable.

- (4) Brisač, **erkić**.
Wiper blade, **die-DIM**.

We have also found an interesting example where the noun diminutive suffix is added to the verb stem of the verb “crći” *to die* (4) to express the fact that something stopped working. Here, it seems to have the function of softening the negative statement and a bit of humorous effect, as well. It also seems to establish a friendly context for the interaction.

4.2. FEMININE DIMINUTIVES

42.6% of all diminutive types were those ending in *-ica*. Traditional grammar recognizes the diminutive suffix *-ka*, in addition to *-ica*. However, those examples seem to be mostly lexicalized and we have not found them within the corpus. Đurić (2004), for instance, mentions “snajka” *daughter-in-law* as a diminutive, even though it is a highly lexicalized diminutive and most speakers of the Serbian language probably would not perceive it as a diminutive.

- (5) Oćete vinjač-ić jedan, **rakij-icu**? vinjač-ić?
Would you like a shot of vinjak-DIM, of **rakija-DIM**? Of
vinjak-DIM?

Using diminutive forms in the context of offering drinks and beverages is very usual in Serbian. Here, the diminutive forms unveil levels of intimacy and informality between speakers, which is the most usual pragmatic effect of the use of diminutive forms of nouns denoting beverages or even food in this context.

- (6) Što me vi navučete svaki put... ali pola **čaš-ice**, pola **čaš-ice**
(smeh)
Oh, you lure me into this every time... but, please, half of a
glass-DIM, half of a **glass-DIM** (laughs)

Speaker answers to the offer to have a drink with a diminutive form of the noun “čaša” *glass*. However, this is a highly lexicalized diminutive form with the meaning of a small glass that is commonly used for spirits. Repetition is a means of intensification (cf. Martin & White, 2005, p. 144). Here, the speaker emphasises that he only wants the half of the glass repeating the noun phrase *pola čašice*.

4.3. NEUTER DIMINUTIVES

Within the corpus, we have found two diminutive suffixes: *-će* (5.2% of all diminutive types) and *-(en)ce* (1.7% of all diminutive types), *-(aš)ce* (0.9% of all diminutive types)

- (7) Kad sam raspoložena odem tamo... Na ono **brd-ašce**.
When I am in a good mood I go there... To this **hill-DIM**.

Here, the meaning seems to be related to affection, the positive feelings of the speaker related to this place, the little hill. We may also conclude that the speaker is a female person (*raspoložen-a* – adj, nominative case, feminine – “in a good mood”). It would be interesting to explore how the sex of the speaker, as well as the sex of the recipient affect the use of diminutives in the Serbian language. Makri-Tsilipakou

(2003) found that men use diminutives more extensively than women within the Greek language.

- (8) Imali smo organizaciju da idemo negde u neko **restoran-će** kao.

We've been thinking about going to **a restaurant-DIM** or so.

In (8), a diminutive “restoranće” *restaurant* has a modal function as well – it means that they thought about going to a restaurant *or so*. The use of “kao”- *or so* (cf. Половина и Панић Церовски, 2013; Panić, 2010) immediately after the noun “restoranće”, thus relating to it, intensifies the meaning of an unspecified type of the service where food and drinks are served. It does not necessarily mean that the restaurant is small, but that it can be anything similar to, but less formal than, a restaurant. The same situation is presented in the example (2), where the other diminutive suffix (-*čić*) is used, but with the same scope of additional meaning.

5. THE USAGE OF DIMINUTIVES

Within this section, we will present and analyse examples from the corpus. We are interested in pragmatic functions that diminutives have within different types of discourse in our corpus: informal, private conversations and public, and thus more formal conversations.

5.1. THE POLITICAL DISCOURSE

The total number of diminutive tokens we have found within political discourse is 5 (0,00003% of all tokens within the corpus of political discourse).

- (9) ... a recimo, jedan sitan, mali, **detaljčić**, ako... ovaj... se izveze jedna fabrika vrednosti od 30 miliona evra,...
... and for example, one tiny, little, **detail-DIM**, if.. am.. one factory, 30 million euros worth, is being exported...

In the example (9), the speaker doesn't mean that the detail is small. Quite the opposite, this detail seems to be very important, it is related to the exportation of a factory that is worth 30 million euros. It seems that the speaker uses the diminutive form here in order to obtain a humorous effect and to highlight the importance of the detail, i.e. the contrary meaning of the lexeme "detaljčić" *detail-DIM*. It is interesting to note that, in this example, the motive noun *detail* already brings the humorous effect, especially with adjectives "sitan" *tiny* and "mali" *little* before the noun. Using the diminutive here emphasises the humorous effect, with a bit of irony or cynicism.

(10) da budete kafe-kuvarica treba da poznajete nekog šefa tamo, šef-ića u nekoj stranci, da vas zaposli.

If you want to work in a catering service, you need to know some manager there, **manager-DIM** in a political party.

In (10) we observe the repetition of the noun "šef" *manager*, with the second one being a diminutive form of the motive noun. Using the diminutive seems to be related to emphasising the pejorative meaning and the disapproving tone towards the subject matter of the whole utterance (cf. Martin & White, 2005). In fact, the use of this diminutive form suggests small significance of that position viewed from the perspective of political importance, but at the same time, reflects that the importance of it is construed through the fact that being on this position, someone is given the power to decide on ordinary people's job position.

(11) A šta bi značio ovaj mali, kako bih to nazvala, šamar-čić kako vi to sad hoćete da podignete svoj rejting u stranci ili svoju poziciju u stranci?

And what would be the meaning of this little (slap in the face), what would I call it, a **tiny slap-DIM** in the face, how do you wish to increase your rating within the political party?

The example in (11) does not indicate a tiny slap-DIM in the face at all. On the contrary, using the diminutive form emphasises the negative attitude the speaker has towards the proposition. The

speaker is a TV anchor and her guest is a politician. Even though we do not observe the whole context in (11), and we do not know who slapped who, and what was the reason, we do observe this negative attitude.

The diminutives in (9) and (11) could be changed to augmentatives as well. Dressler and Merlini Barbaressi (2001) use similar examples in Italian as evidence that the meaning of diminutives cannot be constructed on the basis of semantic feature [small]. Rather, it is based on the pragmatic feature [fictive], and its more specific meaning [non-serious].

5.2. THE CONVERSATIONS AMONG STUDENTS

The total number of diminutive tokens we have found in this part of the corpus is 108.

- (12) Oćete **vinjać-ić** jedan, **rakij-icu?** **vinjać-ić?**
Would you like a **brandy-DIM**, **schnapps-DIM?** **Brandy-DIM?**

In (12) a speaker is asking his friends what they would like to drink. He is using diminutive forms of drink names to denote cordiality, affection and endearment.

- (13) Pa da, šta sada, oni imaju to od 4 do 6, medialunas i **čaj-ić**.
Yes, so what, they have it from 4 to 6, medialunas and **tea-DIM**.

Similarly, in the example (13), a speaker explains habitual actions in Argentina and uses a diminutive form “čajić” tea-DIM. Diminutive form here denotes affection.

- (14) Petsto pedeset? [...] ali... nekako [...] mislim da je to **pljačkica**.
Five hundred fifty? [...] but.. somehow [...] I think it is a **robbery-DIM**.

The speaker explains that the specified amount of money is a lot of money, and thus it is *a kind of robbery*, although not a real robbery. The example in (14) thus has a pragmatic function of a hedge.

- (15) Nekad imam neku malu izložbicu ja, moju autorsku...
Sometimes, I have a small exhibition-DIM me, my own...

A speaker in (15) mitigates the force of an utterance with a diminutive “izložbica” *small exhibition*. The use of diminutive form in this case seems to be a sign of unpretentiousness, humbleness, absence of arrogance.

As we have seen in examples (1), (3), (9) and (11), an adjective “mali” *little* is used along with a diminutive form. In conversation, it does not represent a pleonasm. It suggests that when speakers use diminutive forms, their primary meaning is not related to smallness, but to the hypocoristic, pejorative meaning, mitigating the force of an utterance, or other pragmatic meanings.

6. CONCLUDING REMARKS

Both within political discourse and within private, student conversations, diminutives seem to have a role in expressing the communicative effect of humour and they are often used to express irony as well.

By the means of irony and cynicism, diminutives in the political discourse section of our corpus expressed most frequently pejorative meaning, disapproval and negative attitude. However, within the conversations among students, we have found cases of diminutives denoting hospitality, affection and endearment (examples 12 and 13), hedges (14), as well as examples when diminutives were used to mitigate the force of an utterance (15), to ease the situation (4), conveying at the same time humorous effects. According to the fact that the private, student conversations in our corpus were friendly interactions without strong opposing or disputes between interactants, unlike political TV talk shows, the use of diminutive forms to convey

negative attitude, or open disapproval is less present in that part of the corpus. In other words, the effects and connotations of the use of diminutive forms have shown the tendency to be different depending on the type of the conversation, the relations between participants and their perceived role, as well as the topics that were discussed. With respect to morphology, concerning the derivation process of diminutives, our research has shown that the most frequent suffixes were *-ica* for feminine gender, *-ić* for masculine nouns, and *-će* for neuter diminutives, where 42.6% of all diminutive types were those ending in *-ica*.

Finally, it would be interesting to explore the productivity of those suffixes as well as to explore the semantic and pragmatic meanings of diminutives on other corpora. In addition, it would be important to examine in what way the sex of the speaker, as well as the sex of the recipient affect the use of diminutives in the Serbian language.

REFERENCES

- Barbaresi, L. M., & Dressler, W. U. (2020). Pragmatic explanations in morphology. *Word Knowledge and Word Usage*, 405.
- Dressler, W. U., & Merlini Barbaresi, L. (2001). Morphopragmatics of diminutives and augmentatives. *Perspectives on Semantics, Pragmatics, and Discourse: A Festschrift for Ferenc Keifer*. Amsterdam: John Benjamins Publishing Company, 43–58.
- Đurić, R. (2004). Deminutivni sufiksi u srpskom i njihovi prevodni ekvivalenti u engleskom. *Zbornik Matice srpske za filologiju i lingvistiku* 47(1), 147–164.
- Јовановић, В. (2010). *Деминутивне и аугментативне именице у српском језику*. Београд: Институт за српски језик САНУ.
- Krstev, C. (2008). *Processing of Serbian. Automata, texts and electronic dictionaries*. Faculty of Philology of the University of Belgrade.
- Makri-Tsilipakou, M. (2003). Greek diminutive use problematized: Gender, culture and common sense. *Discourse & Society*, 14(6), 699–726.
- Martin, J. R. & White, P. R. (2005). *The language of evaluation: Appraisal in English*. London: Palgrave Macmillan.

- Николић, М. (ур.) (2011). *Речник српскога језика*. Нови Сад: Матица српска.
- Panić, N. (2010). Тако, 'нако, nebitno. [Serbian approximators: *tako*, 'naко, *nebitno*]. *Језик и друштво* (ур. Јулијана Вучо, Јелена Филиповић), едиција *Филолошка истраживања данас*, Филолошки факултет, Београд, 221–235.
- Половина, В. & Панић Церовски Н. (2013). Функције апроксиматора у српском разговорном дискурсу. [Functions of approximators in Serbian conversational discourse]. *Српски језик, књижевност, уметност. Традиција и иновације у савременом српском језику*. (књ. 1). (ур. Милош Ковачевић). Крагујевац: ФИЛУМ, 157–168.
- Ponsonnet, M. (2018). A preliminary typology of emotional connotations in morphological diminutives and augmentatives. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 42(1), 17–50.
- Savickiene, I. & Dressler, W. U. (2007). Introduction. In I. Savickiene & W. U. Dressler (Eds.), *The Acquisition of Diminutives* (pp. 1–12). Amsterdam: Benjamins.
- Schweiger, S., Barbaresi, A., Korecky-Kröll, K., Ransmayr, J., & Dressler, W. U. (2019). Diminutivvariation in österreichischen elektronischen Korpora. In B. Lars, K. Herbert & A. K. Fischer (Eds.), *Dimensions of Linguistic Space: Variation – Multilingualism Conceptualisations Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung* (pp. 147–162). Wien: Peter Lang
- Spasovski, L. (2012). *Morphology and pragmatics of the diminutive: evidence from Macedonian* (Doktorska disertacija, Arizona State University). Preuzeto sa <https://core.ac.uk/download/pdf/79564048.pdf>
- Stanojčić, Ž. & Popović, Lj. (2004) Gramatika srpskog jezika: udžbenik za I, II, III i IV razred srednje škole. Beograd: Zavod za udžbenike i nastavna sredstva.
- Vitas, D., & Krstev, C. (2012). Processing of corpora of Serbian using electronic dictionaries. *Prace Filologiczne*, 63, 279–292.

Milena Oparnica
Natalija Panić Cerovski

L'USAGE DES DIMINUTIFS DANS LA COMMUNICATION QUOTIDIENNE

Résumé

Le présent article vise à analyser les diminutifs dans les corpus de la langue parlée, depuis les perspectives morphologique et pragmatique. Nous analysons le processus de la formation des diminutifs, ainsi que leur fonction pragmatique au sein du discours.

Le corpus utilisé pour la présente recherche a été collecté manuellement et contient les transcriptions de la communication quotidienne des étudiants (environ 250 000 tokens), ainsi que les transcriptions des émissions de débat télévisées sur la politique (environ 173 000 tokens). Afin d'en extraire des diminutifs, nous avons créé un script simple en Python.

Nous avons davantage exploré s'il existe une différence dans l'usage des diminutifs dans ces deux sphères distinctes, à savoir le discours politique et la communication quotidienne des étudiants.

Dans le discours politique ainsi que dans le discours privé – la conversation estudiantine –, les diminutifs semblent avoir le rôle d'exprimer l'effet communicatif de l'humour ; ils sont fréquemment utilisés pour exprimer aussi de l'ironie.

Par le biais de l'ironie et du cynisme, les diminutifs dans la section politique de notre corpus ont exprimé le plus fréquemment la signification péjorative, la désapprobation et l'attitude négative. En revanche, dans le cadre de la communication parmi les étudiants, nous avons trouvé les cas des diminutifs dénotant l'hospitalité, l'affection, la sympathie, l'enclosure (*hedges*), ainsi que les exemples dans lesquels les diminutifs étaient utilisés pour atténuer la force de l'élocution et pour détendre la situation, tout en transmettant l'effet humoristique. Vu que les conversations privées, estudiantines, de notre corpus étaient des interactions amicales sans forts désaccords ou disputes entre les participants – contrairement aux émissions de débat télévisées sur la politique –, l'usage des formes diminutives pour transmettre l'attitude négative ou la désapprobation ouverte est moins présent dans cette partie du corpus. Autrement dit, les effets et les connotations de l'usage des formes diminutives ont montré la tendance d'être différents, en fonction du type de la conversation, des relations entre les participants et de la perception de leurs rôles respectifs, ainsi que des sujets des discussion.

Dans le domaine de la morphologie, en ce qui concerne le processus de la dérivation des diminutifs, notre recherche a démontré que les suffixes les plus

fréquents étaient *-ica* pour le genre féminin, *-ić* pour les noms masculins, ainsi que *-če* pour les diminutifs neutres ; dans l'ensemble, 41,6 % de tous les diminutifs terminaient en *-ica*.

Enfin, ce serait intéressant d'explorer la productivité de ces suffixes, ainsi que les significations sémantique et pragmatique des diminutifs dans les autres corpus. En outre, ce serait important d'examiner de quelle manière le sexe du locuteur, ainsi que celui de l'interlocuteur affectent l'usage des diminutifs dans la langue serbe.

Mots-clés: diminutifs, morphopragmatique, fonctions pragmatique, analyse de corpus, langue parlée

Milica Dinić Marinković*

University of Belgrade – Faculty of Philology

(SOME SIMPLE) MEANS FOR EXTRACTION OF ORTHOGRAPHICALLY UNMARKED FICTIONAL DIALOGUE

Abstract

The main aim of this research paper is to investigate whether it is possible to extract orthographically and typographically unmarked fictional dialogue from unannotated corpora with a simple regex query, and to what extent. In order to give an answer to this question, the paper provides a brief review of recent approaches and proposed solutions to this problem, with special attention on work with unannotated corpora. The simple means for extraction proposed here are based on the identification of overt linguistic features of character strings in orthographically and typographically unmarked fictional dialogues, and different ways of their conversion into patterns for the extraction from unannotated corpora. As a research corpus we used Sally Rooney's novel "Normal people", both the original version (English) and Serbian translation.

Keywords: unmarked fictional dialogue, unmarked direct speech representation, overt linguistic features of a fictional dialogue, extraction patterns from unannotated corpora, Sally Rooney's fictional dialogue

* milica.dinic.marinkovic@fil.bg.ac.rs

1. INTRODUCTION

One of the simple exercises for students in any corpus linguistics course is to form a regular expression¹ that will match direct speech strings/fictional dialogue in a corpus of prose fiction. This is a pretty convenient task because the solution to this problem lies in the identification and defining of orthographical and typographical conventions for direct speech representation in a given corpus and does not require corpus annotation.

Although typography of fictional dialogue markers varies in regard to specific convention – publishing style, "fashion" of the epoch, as well as regional or national tradition, in these diachronic and/or diatopic variations we could easily find the most frequent typographical markers of direct speech. For SerBoCroatian publishing tradition these markers are quotation marks and long dashes, so, according to this, there are two simple types of regular expressions for direct speech strings matching. In the case of quotation marks, one could use the following simple regular expressions.

E.g. (1) „.+?”

E.g. (2) „[A-Za-z²]+ [0-9\.,\?!]*”

But the question is what happens when orthographical and typographical markers of a fictional dialogue in the text are missing?

¹ *Regular expression*, or *regex*, is a pattern that consists of a sequence of characters, and it is used for formal description of character strings, or sequences of character strings in a text. The main usage of the regex is a specification of search patterns in some corpus/text, and all the character strings that correspond to the query in form of a regex are called *matching strings*. (For more about regex see Jurafsky & Martin, 2021).

² This range of characters corresponds to ASCII character encoding. For the Serbian language there are two different ranges of characters, depending on the alphabet, i. e. character code which is being used in the text – Cyrillic or Latin.

1.1. BACKGROUND, RESEARCH QUESTIONS, AND GOALS

We are witnessing the trend of intentionally neglecting traditional punctuation conventions in contemporary literature (e. g. Sally Rooney, Bernardine Evaristo, Milena Marković), among which Sally Rooney's novel *Normal People* is undoubtedly significant as a globally popular phenomenon³. According to its status, this novel could easily be a corpus for students' exercises. Yet, since this novel lacks quotation marks in fictional dialogues (as do other novels by the same author), it is impossible for students to fulfill the seemingly simple task of matching direct speech strings in this text.

This particular challenge caught my attention during the course of reading the novel *Normal people* when I realized I had not been noticing the lack of quotation marks for about 20 pages, although my perception of dialogues had not been impaired by that. This experience confirms several things.

Firstly, this partly confirms the stance of Cormac McCarthy (the author who also doesn't use quotation marks in his writing) that:

"There's no reason to blot the page up with weird little marks. I mean, if you write properly you shouldn't have to punctuate."

Sally Rooney herself stated:

*"I can't remember ever really using quotation marks. I didn't see any need for them, and I don't understand the function they perform in a novel."*⁴

Besides the clarity of writing, viewing this matter from the psycholinguistic perspective implies that the linguistic characteristics of a fictional dialogue have a clear representation in our mind, a

³ Although various sorts of planned neglect of the different orthographical and typographical conventions have been presented in the history of literature, particularly in the epoch of Russian Futurism, it has not been the mainstream. There are only several, mostly Irish writers, who keep minimalist usage of punctuation in order to achieve maximal clarity in writing, e. g. James Joyce, Samuel Beckett, and Roddy Doyle.

⁴ <https://www.palatiniate.org.uk/the-end-of-conventional-punctuation/>

representation that does not depend on typographical markers of fictional dialogues but on features of a different kind.

However, these observations of novelists are significant for the investigation and findings on the nature and characteristics of written language. According to the literature on features of written language (Halliday, 1989; Halliday & Martin, 1993), it is doubtless that the differences between typical oral and typical written language are fundamental, not only regarding their form but in their function and content as well. Of course, like in every social phenomenon, between these prototypical instances of written and oral forms of language use, there is a whole spectrum of various sorts of mixed or hybrid forms. The literary form of novel surely belongs to a hybrid form of written language, because it contains instances of constructed oral communication represented in fictional dialogues.

As Abercrombie stated, “(...) *the aim of writing is not, usually, to represent actual spoken utterances which have occurred*”⁵. When Halliday examines which features of oral language are represented in written form, he concludes:

“So the omission of prosodic features from written language is, in some respects and under certain circumstances, a genuine deficiency. There is, on the other hand, a device that is used in order partially to overcome this deficiency; this is the device of punctuation.” (1989: 32)

Further, by analyzing the functions of punctuation, Halliday concluded that the main purpose of punctuation is not to represent some of the prosodic features of oral language, but that it evolved within the written language for the purpose of achieving greater clarity of the written text itself (*Ibid.*: 32–39). The history of emerging and conventionalization of quotation markers confirms this standpoint⁶.

Bearing in mind the writers’ statements, as well as the finding of research on written language, one could anticipate that in the future the number of writers who avoid conventional punctuation may rise, while corpus and computational methods for extraction are based

⁵ Abercrombie, 1965, as cited in Halliday, 1989: 31.

⁶ On the history of punctuation markers see Houston 2013.

either on punctuation characteristics of character strings or on very sophisticated tools and time-consuming construction of resources.

Therefore, the research questions of this paper are: (1) what would be the simple means for the extraction of an orthographically and typographically unmarked fictional dialogue from unannotated corpora, and (2) whether it is possible to extract unmarked fictional dialogues from unannotated corpora by simple regex (almost) matching, and to what extent?

In order to answer these questions, the main aims of this paper are (1) identification of linguistic features of strings in orthographically and typographically unmarked fictional dialogue, and (2) construction of patterns for their extraction from unannotated corpora.

As a research corpus we used Sally Rooney's novel "Normal people", both the original version (English) and Serbian translation⁷.

2. *FICTIONAL DIALOGUE* IN RECENT LINGUISTIC RESEARCH

Fictional dialogue is usually defined as "passages of character-character conversation within a literary text" (Nykänen & Koivisto, 2016: 1), or „passages of characters' direct speech in prose fiction text" (Kurfalı & Wirén, 2020: 105). Cited definitions represent a narrow understanding of this narration mode. Broader, and, in our opinion, more precise determination of fictional dialogue includes all the characters' interactions which stand opposite to the narrator's telling the story⁸. That includes, for example, an e-mail correspondence of characters', their SMS or chat conversation, and also internal dialogue.

Besides the term fictional dialogue, the following terms are being used (though less frequently): *imaginary dialogue* (Kinzel & Mildorf,

⁷ Only for the purpose of the research presented in this paper .txt form of published books have been made. In further processing of text, AntConc software was used. For details on the analyzed corpus and used tools, see the citations at the end of this paper.

⁸ As in any other phenomenon, there are all sorts of different manners of characters' interaction representation, e. g. „free indirect style, *style indirect*

2012), *constructed dialogue, hypothetical dialogue* (Semino et. al., 1999), *characters' discourse* (Ek & Wirren 2019).

The phenomenon of a fictional dialogue occupies the attention of researchers in various linguistic fields – Stylistics and Corpus Stylistics (Toolan, 1985; 1987; Leech & Short, [1981], 2007; Oostdijk, 1990; Semino & Short, 2004; Axelson, 2009; Ковачевић, 2013; Mahlberg et. al., 2019), Conversation Analysis (Bahtin, 1980; Burton, 1980; Thomas, 2012), Representology (Ковачевић, 2015), Literary Linguistics (Nykänen & Koivisto, 2016), Natural Language Processing (Byszuk et. al., 2020; Wirén, Ek & Kasaty, 2020; Kurfalı & Wirén, 2020; Ek & Wirén, 2019; Weiser & Watrin, 2012; Yeung & Lee, 2016 & 2017; Axelson, 2009; Oostdijk, 1990).

The growing interest in fictional dialogue within the field of natural language processing is driven by the need for automatic extraction of fictional dialogue for the purpose of providing a large amount of empirical data for systematic linguistic analyses. There are various research questions that depend on systematic collections of linguistic data of this kind, such as (1) questions of narrative structure, which include: distinguishing narration and speech, keeping track of addresses, identification, and modeling of fictional characters; (2) stylistic characteristics of fictional dialogue; (3) conversation analysis key questions concerned with the relation, similarities, and differences between real-life conversation and constructed conversation represented in fictional dialogue, (4) getting data on and analysis of (written) speech-like language from historical periods⁹, and (5) questions about the very nature of fictional dialogue, which include description, determination, and positioning in regards to both spoken conversation and other narrative means in fiction.

libre, represented speech and thought, quasi-direct discourse, and combined discourse" (Toolan, 2001: 119), „free indirect speech" (Oostdijk, 1990: 236), but these phenomena are out of reach of the research presented in this paper. For terminological apparatus for different representations of speech in Serbian see Ковачевић, 2013 and Ковачевић, 2020.

⁹ About challenges in historical pragmatics regarding the issue of collecting suitable data see Lalić, A. *Surpassing the "bad data" problem: Italian epistolary discourse as a source of spoken language* in this volume.

3. SOME RECENT APPROACHES TO FICTIONAL DIALOGUE AUTOMATIC EXTRACTION

As Axelsson (2009) accurately distinguished, challenges for automatic extraction of fictional dialogue lie upon numerous linguistic, both theoretically fundamental and practical, unsolved problems. The first problem that Axelsson detected is the very definition of fictional dialogue, as its scope may vary in regard to concrete research questions (2009: 191). The other two major issues that she addressed originate from the treatment of prose fiction in referent corpora (BNC 2001, 2007), both in terms of representativeness of the samples that were included, and in terms of lack of direct speech annotation/tagging.

Approaches to the problem of automatic extraction of orthographically and typographically unmarked fictional dialogue could be systematized in the following way.

In most cases, the starting point is the process of annotation, but the scope and levels of annotation vary greatly. Some researchers use lemmatized corpora with in-depth MSD¹⁰ annotation¹¹ (Ek & Wirén, 2019;), as well as different levels of semantic annotation (Wirén *et. al.*, 2020). This annotation process could be manual or semi-automatic. Others conduct manual annotation of targeted markers, e. g. speech framing verbs, sentence type, and/or structural elements (Oostdijk, 1990), or carry out manual annotation of the macrostructure of text (Ek & Wirén, 2019; Wirén *et. al.*, 2020).

A central step in recent approaches to fictional dialogue automatic extraction is a machine learning process, i.e. machine training. These processes are always, in the first instance, based on typographically unambiguously marked direct speech, which serves as material for machine training. In the next step, some researchers remove typographical markers (Kurfalı & Wirén, 2020), or use a multilingual approach and deep learning with *Bidirectional Encoder*

¹⁰ MSD stands for morpho-syntactic descriptions.

¹¹ Although extraction could be accomplished regardless of MSD tagging, MSD annotation is done due to further linguistic analysis.

Representations from Transformers (Devlin *et al.*, 2018) so that machine classifiers do not rely on typographic markers (Bysuzuk *et. al.*, 2020).

Contrary to dominantly machine learning approaches, there is a more linguistic approach for automatic extraction of direct speech. This method proposed the creation of a formal unlexicalised grammar of direct speech representations and the automatic construction of an e-dictionary of lexical elements that can introduce direct speech instances. In the study of Weiser & Watrin (2012) unmarked direct speech instances are described as syntactic structure, and finite state automata were built (in Unitex software) according to recognized patterns of structure. E-dictionary, which consists of speech framing verbs, was used as direct speech “hunter”.

4. ANALYSIS PART 1: OVERT LINGUISTIC FEATURES OF UNMARKED FICTIONAL DIALOGUE

In work with unannotated corpora it is impossible to conduct an analysis on a word level. Hence the investigation of overt linguistic features of unmarked fictional dialogue in our research has been done on the sentence level. During the analysis of overt linguistic features of strings in instances of fictional dialogue in the analyzed corpus, and according to solutions proposed by Weiser & Watrin (2012)¹², the following components emerged as significant: (1) structural elements of an unmarked fictional dialogue, (2) order of structural elements in instances of unmarked fictional dialogue, (3) lexico-grammatical features of extracted instances, and (4) their sentence punctuation.

4.1. Structural elements of an unmarked fictional dialogue that are singled out are:

- unmarked direct speech (UDS)
- comma (,)
- speech framing verb (SFV)
- reference to the character – personal name or pronoun (RTC)

¹² With appropriate adjustments.

It is important to emphasize that, since the research question of this paper is whether it is possible to extract unmarked fictional dialogue from unannotated corpora, we could only lean on instances of dialogue already marked by the writer/narrator. This means that analyzed instances of fictional dialogue do not fit the narrow concept of fictional dialogue, which excludes instances of the narrator's framing of characters' conversation from the fictional dialogue, together with references to the characters.

4.2. The (linear) order of structural elements in unmarked fictional dialogue

| | | | | |
|----------------------------------|-----------|----------------|-------------|---------------------|
| [And I hear you did very well] | [,] | [she] | [says]. | |
| 1. | 2. | 3. | 4. | (NP ¹³) |
| UDS | , | RTC | SFV | |
| | | | | |
| [A čujem i da si odlično uradio] | [,] | [nastavila je] | [Lorejn]. | |
| 1. | 2. | 3. | 4. | (NLj) |
| UDS | , | SFV | RTC | |
| | | | | |
| [He was top of the class] | [,] | [says] | [Marianne]. | |
| 1. | 2. | 3. | 4. | (NP) |
| UDS | , | SFV | RTC | |
| | | | | |
| [Najbolje u celom razredu] | [,] | [dodala je] | [Merijen]. | |
| 1. | 2. | 3. | 4. | (NLj) |
| UDS | , | SFV | RTC | |

In the Serbian part of the analyzed corpus the order of structural elements is fixed – [UDS] [,] [SFV] [RTC], while in the English part of the corpus we find an alternating order of the last two elements – RTC and SFV, which can occupy third or fourth position in this type of fictional dialogue. The reason for the inalterable order of structural elements of fictional dialogue in the Serbian part of the corpus is the specific canonical form of speech framing verbs in fictional dialogue (see section 4.3), which does not allow a change in word order, and therefore in order of structural elements.

¹³ Abbreviation for analyzed corpus: NP stands for the English version (*Normal People*), NLj stands for the Serbian version (*Normalni ljudi*).

4.3. LEXICO-GRAMMATICAL FEATURES OF SOME STRUCTURAL ELEMENTS

On the lexico-grammatical level, the least variable aspect of fictional dialogue instances represents the narrator's framing of the characters' conversation, which is achieved by using *speech framing verbs*, *lexical quotatives*, and *quotative markers* (Panić Cerovski & Ivanović, 2016: 143), or, traditionally *verba dicendi*.

These words do not just form a specific paradigmatic lexical set, but they also appear in a specific canonical form within the fictional dialogue (at least this is the case with the analyzed corpus).

In the English version of the novel, speech framing verbs appear in 3rd person Sg simple present tense more frequently than in the past tense (e. g. „to say” – says: freq. 751, said: freq. 480). In the Serbian version, speech framing verbs appear only in one form – 3rd person Sg past tense inverted form, e. g. *rekao je* (=he said), *upitao je* (=he asked), *iznenadila se* (=she got surprised), *nastavio je* (=he continued). This is a specific canonical form, determined by the initial position in the embedded clause, that governs the position of reference to the character.

4.4. SENTENCE PUNCTUATION

When we take into account the characters of strings, it is impossible to exclude punctuation from consideration. Since it is clear that the full stop could not be considered a marker of any kind in the phenomenon we investigate here, we have analyzed the usage of exclamation and question marks in our corpus.

Conducted analyses show that the exclamation marks exclusively appear within a fictional dialogue, or at the boundaries of direct speech instances, which was expected. Indeed, it would be very odd for the writer to yell at the readers. Humor aside, it is not surprising, since there is no place for imperative or exclamatory sentences in the narration – direct commands, requests, warnings, or expressing of strong emotions, or other speech acts that these sentences represent.

Apart from the exclamation marks, question marks also belong exclusively to the strings of fictional dialogue in the analyzed corpus. There are two types of (quasi) exceptions that have been recognized. These are:

- (1) occurrences of „?” in email texts that the female protagonist reads, and
- (2) occurrences of „?” in the protagonists’ internal dialogue.

According to the narrow concept of fictional dialogue – as the passages of characters’ conversation within the text of prose fiction, these instances should be excluded. But, as we stated in section 2, we consider that everything which is not the writer’s telling of the story belongs to a fictional dialogue. In addition, these two types of context – email text, and internal dialogue, belong to the mode of character interaction, not to the narration.

Although exclamation and question marks exclusively appear within a fictional dialogue, or at the boundaries of direct speech instances, most instances of fictional dialogue in the analyzed corpus do not contain these punctuation marks. Thus, even though the patterns for the extraction containing these marks will not match something other than fictional dialogue, the majority of the instances of fictional dialogue will stay unrecognized.

5. ANALYSIS PART 2: PATTERNS (FROM FEATURES) FOR EXTRACTION

Formal linguistic features that are presented and described in the previous section are converted into patterns, i. e. regular expressions for the matching and extraction of fictional dialogue in the analyzed corpus.

The first regex that was tested is based on structural elements and their order in instances of fictional dialogue (see 4.1 and 4.2).

Regex (1): $^{\wedge}([A-Z])\backslash w+ []^* [A-Za-z]^+, [A-Za-z]? (says|asks|...)$
 $[A-Za-z]?$

This regex corresponds to the following pattern.

[new line] [word start with capital]
[word] {0-as many as there are words in a sentence} [word] [,]
[reference to a character] {0,1}
[sequence of speech framing verbs in suitable form in
disjunctive relation]
[reference to a character] {0,1}

Thus, every instance of fictional dialogue (those that belong to the representation of the conversation between characters in a novel) always starts with a new line followed by a word starting with a capital. Between the initial word in a passage, and an obligatory element – a word followed by a comma, it could be null or many appearances of words (special character * stands 0 or more of them). After the word followed by a comma, there comes the element which is not obligatory in that precise position (in English), and that is why the special character "?" stands in Regex (1). This structural element (a reference to a character) could take the position (slot) before or after the speech framing verb, but the presence of both speech framing verbs and references to the characters are obligatory structural elements in this pattern of fictional dialogue instances.

Regex (1) matches, for example, the following instances of fictional dialogue:

*The lads are fairly late, says Lisa.
Momci bogami baš kasne, prokomentarisala je Liza.*

*If they don't show up I will actually murder Connell, says Rachel.
Ako se ne pojave, ima da ubijem Konela, rekla je Rejčel.*

*Da li ikada sretneš Pegi u Dublinu? upitala ga je.
Do you ever see Peggy in Dublin? she says.*

*Jesi dobro? upitala ga je Lorejn.
Are you alright? says Lorraine.*

Regex (1) matches only those passages of typographically unmarked fictional dialogue which contain the narrator's interference, i. e. framing. In the analyzed corpus instances that contain speech

framing verbs appear in about every 3rd – 5th paragraph of fictional dialogue, or maybe it is more precise to say – in about every 3rd – 5th turn within the characters' discourse, as Rooney uses paragraph breaks to indicate turns in conversation (see excerpts in text boxes below).

Text Box 1: Excerpt from the NP

- (1.p) She shut her eyes. I do like you, **she said**.
(2.p) Well, if you met someone else you liked more, I'd be pissed off, okay? Since you ask about it. I wouldn't be happy. Alright?
(3.p) Your friend Eric called me flat-chested today in front of everyone.
(4.p) Connell paused. She felt his breathing. I didn't hear that, **he said**.
(5.p) You were in the bathroom or somewhere. He said I looked like an ironing board.
(6.p) Fuck's sake, he's such a prick. Is that why you're in a bad mood?
(7.p) She shrugged. Connell put his arms around her belly.
(8.p) He's only trying to get on your nerves, **he said**. If he thought he had the slightest chance with you, he would be talking very differently. He just thinks you look down on him.

Text Box 2: Excerpt from the NLj

- (1.p) Zažmurila je. Pa sviđaš mi se, **rekla je**.
(2.p) A ja bih popizdeo kad bi upoznala nekoga ko bi ti se više dopao. Eto to bi bilo. Kad si me već pitala. Ne bi mi uopšte bilo svejedno. Da znaš. Okej?
(3.p) Tvoj drugar Erik mi je pred svima rekao da sam ravna ko daska.
(4.p) Nije joj odmah odgovorio. Osećala je njegov dah. Ja to nisam čuo, **rekao je**.
(5.p) Bio si u kupatilu ili šta znam gde. Rekao je da izgledam kao daska za peglanje.
(6.p) On je drkadžija, jebo te. Jesi zbog toga loše raspoložena?
(7.p) Slegnula je ramenima. Obgrlio ju je oko struka.
(8.p) On samo hoće da te iznervira, **rekao je**. Kad bi kapirao da ima i najmanje šanse kod tebe, pričao bi sasvim drugu priču. On samo misli da ga potcenjuješ.

Thus, the majority of the instances of fictional dialogue cannot be recognized. On the other hand, we think that tracking of writer's interference in the fictional dialogue, and catching patterns of it could give valuable results. It seems that the phenomenon in question, besides being driven by textual cohesion (see Polovina, 1999), is of

a discourse-pragmatic nature similar to one that takes place during the course of conversation (Polovina, 1988; Savić & Polovina, 1989), and especially as it is described in the chapter *Non-omission of deictic personal pronouns* in this volume (Polovina, 2022).

It is clear that the presented regex is too long and that it does not meet the condition of minimalism and elegance, nor is it possible to form a regex of adequate size according to all the conditions that should be defined by it. Besides that, this regex also matches the strings which are not instances of fictional dialogue, but they appear on the boundaries of it, such as:

He waits for the coughing to subside, and then says¹⁴: *What does he do to you?*
Sačekavši da mu prođe napad kašalj, rekao je: *A šta ti on to radi?*

The solution for these particular cases for the English version is:

Regex (2): "(says|asks...):", or the following pattern:

[sequence of speech framing verbs in suitable form in disjunctive relation] [:]

This implies, as do the other examples too, that the solution to the problem lies in the combining of structural elements into the smaller distinct (sub)sets, instead of constructing one (long) regex with the potential to cover as many cases as possible.

6. CONCLUSION: THE REACH/LIMITS OF SIMPLE MEANS

The main question of the research presented in this paper is whether it is possible to extract unmarked fictional dialogue from unannotated corpora by simple regex (almost) matching, and to which extent. In order to give the answer to this question identification of overt linguistic features of strings in orthographically and typographically

¹⁴ The underlined part of the sentence is the one that is matched by the Regex (1).

unmarked fictional dialogue, as well as the construction of patterns for their extraction from unannotated corpora has been done.

It emerged that the main problem of unmarked fictional dialogue matching is the inability to define the boundaries of a fictional dialogue. Because of this, we cannot use a regex for matching the exact and whole strings of fictional dialogue, nor is it possible to match (and extract) all the instances of fictional dialogue in the analyzed corpus. The reasons for this limitation lie in two areas – in the area of conceptual determination of a fictional dialogue, and in the area of linguistic description for the purpose of natural language processing. From the conceptual point of view, the boundaries of a fictional dialogue are not invariably defined. At one end of a continuum, there are researchers who take into account only instances of conversational dialogue, while on the other end of the continuum, there are researchers who consider that all the instances which are not direct writer's words, i. e. narration belong to a fictional dialogue. As far as the formal description of boundaries of a fictional dialogue is concerned, one could be able to define the end of certain types of fictional dialogue, but we are not able to define the beginning of an unmarked direct speech passage in an unannotated corpus/text.

In addition to the incapability of defining boundaries of an unmarked fictional dialogue, there is yet another limitation. The conducted analyses presented in this paper show that in a work with unannotated corpora it is possible to detect only those passages of a typographically unmarked fictional dialogue which contain narrator interference, i. e. framing. And one should bear in mind that every 3rd–5th paragraph of fictional dialogue contains speech framing verbs.

Also, there is a challenge with a very long and robust form of regular expressions that have the potential to match more instances of fictional dialogue. This particular challenge could be overcome by dividing and organizing patterns into smaller distinct subsets. This means that the procedure of matching and extracting fictional dialogue instances by using simple regex queries needs to be repeated several times with different smaller regexes. Of course, not all instances of direct speech would be recognized, but they are not recognized with

more sophisticated means either. However, this type of task is well suited for students' exercises in corpus linguistics, and there is a lot of material for the practicing of regex formation.

REFERENCES

- Axelsson, K. (2009). Research on fiction dialogue: Problems and possible solutions. In: Jucker, AH, Schreier, D, Hundt, M. (Eds.), *Corpora: Pragmatics and Discourse*. Amsterdam: Rodopi, (189–201).
- Bahtin, M. (1980). *Marksizam i filozofija jezika*. Beograd: Nolit.
- Burton, D. (1980). *Dialogue and Discourse. A Sociolinguistic Approach to Modern Dialogue and Naturally Occurring Conversation*. London: Routledge and Kegan Paul.
- Byzuk, J., Wozniak, M., Kestemont, M., Lesniak, A., Łukasik, W., Šela, A. & Eder M. (2020). Detecting direct speech in multilingual collection of 19th-century novels. In: *Proceedings of 1st Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)*, European Language Resources Association (ELRA), 100–104.
- Ek, A. & Wirén, M. (2019). Distinguishing narration and speech in prose fiction dialogues. In: *Digital Humanities in the Nordic Countries 4th Conference (DHN)*, Copenhagen, Denmark, 124–132.
- Halliday, M., A. K. (1989). *Spoken and Written language*. Oxford University Press.
- Halliday, M., A. K. & Martin, J. R. (ur.), (1996 [1993]). *Writing Science, Literary and Discourse Power*. London: The Falmer Press.
- Houston, K. (2013). *Shady characters: the secret life of punctuation, symbols, & other typographical marks*. First edition. New York – London: W.W. Norton & Company.
- Jurafsky, D. & Martin, J. H. (2021). *Speech and Language Processing* (3rd edition – draft). Available at <https://web.stanford.edu/~jurafsky/slp3/2.pdf>
- Kinzel, T., & Mildorf, J. (Eds.) (2012). *Imaginary Dialogues in English: Explorations in a Literary Form*. Heidelberg: Universitätsverlag Winter, 9–30.
- Ковачевић, М. (2013). О терминосистему туђег говора, In: *Лингвистика као србистика*, Vol. 1, (pp. 67–99). Универзитет у Источном Сарајеву, Филозофски факултет Пале.

- Ковачевић, М. (2015). *Стилска и грамашка стилских фигура*. Београд: Јасен.
- Ковачевић, М. (2020). Типови говора у роману Џејн Остин „Разум и осјећајност”. *Филолош-часопис за језик, књижевност и културу*, (22), 116–142.
- Kurfali, M. & Wirén, M. (2020). Zero-shot cross-lingual identification of direct speech using distant supervision. *Proceedings of LaTeCH-CLfL 2020*, 105–111.
- Leech, G., & Short, M. (2007). [1981]. *Style in fiction: A linguistic introduction to English Fictional prose*. 2nd ed. Harlow, Essex: Pearson Education.
- Mahlberg, M., Wiegand, V., Stockwell, P., & Hennessey, A. (2019). Speech-bundles in the 19th-century English novel. *Language and Literature*, 28(4), 326–353. <https://doi.org/10.1177/0963947019886754>
- Nykänen, E., & Koivisto, A. (2016). Introduction: Approaches to fictional dialogue. *International Journal of Literary Linguistics* 5(2), 1–14.
- Oostdijk, N. (1990). The language of dialogue in fiction. *Literary and Linguistic Computing* 5(3), 235–241.
- Panić Cerovski, N., & Ivanović, B. (2017). Glagoli govorenja kao markeri emfaze – slučaj pseudocitiranja. *Komunikacija I Kultura Online*, 7(7), 142-154. <https://doi.org/10.18485/kkonline.2016.7.7.10>
- Savić, S., & Polovina, V. (1989). *Razgovorni srpskohrvatski jezik*. Novi sad: Filozofski fakultet, Institut za južnoslovenske jezike.
- Semino, E., Short, M., & Wynne, M. (1999). Hypothetical Words and Thoughts in Contemporary British Narratives. *Narrative*, 7(3), 307–334.
- Semino, E., & Short, M. (2004). *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.
- Polovina, V. (1988). The Basic "Verba Dicendi" and their cohesive role in spoken conversation. *Acta Linguistica Hungarica*, 38(1/4), 193–200.
- Polovina, V. (1999). *Semantika i tekstlingvistika*. Београд: Ћигоја штампа.
- Polovina, V. (2021). Non-omission of Deictic Personal Pronoun, In: Polovina, V. & Panić Cerovski, N. (Eds.), *BeLiDa 2021* (pp. 19–20). Belgrade: Faculty of Philology.
- Thomas, B. (2012). *Fictional Dialogue: Speech and Conversation in the Modern and Postmodern Novel*. Lincoln, NE: University of Nebraska Press.
- Toolan, M. (1985). Analyzing Fictional Dialogue. *Language and Communication*, 5(3), 193–206.

- Toolan, M. (1987). Analysing Conversation in Fiction: The Christmas Dinner Scene in Joyce's: "Portrait of the Artist as a Young Man." *Poetics Today*, 8(2), 393–416.
- Toolan, M. (2001). *Narrative: a critical linguistic introduction*. (2nd ed.). London: Routledge.
- Weiser, S. & Watrin, P. (2012). Extraction of unmarked quotations in Newspapers; A Study Based on Direct Speech Extraction Systems. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC12)*, 559–562.
- Wirén, M., Ek, A., & Kasaty, A. (2020). Annotation Guideline No. 7: Guidelines for annotation of narrative structure. *Journal of Cultural Analytics*. doi: 10.22148/001c.11772
- Yeung, C. Y., & Lee, J. (2016). An annotated corpus of direct speech. In: Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA), 1059–1063.
- Yeung, C. Y., & Lee, J. (2017). Identifying speakers and listeners of quoted speech in literary works. In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, Vol 2: 325–329. Asian Federation of Natural Language Processing.

Analyzed Corpus¹⁵:

- NP – Rooney, S. (2018). *Normal People*. London: Faber & Faber Ltd.
- NLj – Runi, S. (2019). *Normalni ljudi*. Beograd: Geopoetika izdavaštvo.

Used Tools:

- Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

¹⁵ Corpus in .txt form was constructed only for the purpose of the research.

Милица Динич Маринкович

(НЕКОТОРЫЕ ПРОСТЫЕ) СРЕДСТВА ДЛЯ ИЗВЛЕЧЕНИЯ
ОРФОГРАФИЧЕСКИ НЕМАРКИРОВАННОГО ХУДОЖЕСТВЕННОГО
ДИАЛОГА

Резюме

Одно из простых упражнений для студентов любого курса корпусной лингвистики состоит в том, чтобы сформировать регулярное выражение, которое будет соответствовать вымышленным строкам диалога / прямой речи в некотором корпусе художественной прозы. Это довольно удобная задача, поскольку решение этой проблемы заключается в идентификации и определении орфографических и типографских соглашений для представления прямой речи в данном корпусе и не требует аннотации корпуса. Но вопрос в том, что происходит, когда эти орфографические и типографские маркеры в тексте отсутствуют?

Чтобы дать ответ на этот вопрос, в статье представлен краткий обзор последних подходов и предлагаемых решений этой проблемы, с особым вниманием к работе с неаннотированным корпусом. Выбранные подходы и предлагаемые решения основаны на процессах машинного обучения, различных наборах эвристик для автоматического извлечения, а также аннотации корпусов или создании электронных словарей. Тем не менее, основная цель этой исследовательской работы состоит в том, чтобы выяснить, возможно ли вернуть эту проблему в рамки простых студенческих упражнений и в какой степени.

Таким образом, простые средства извлечения, которые мы предлагаем, основаны на идентификации лингвистических особенностей строк в орфографически и типографически немаркированных художественных диалогах и различных способах их преобразования в шаблоны для извлечения из неаннотированных корпусов. В качестве исследовательского корпуса мы использовали роман Салли Руни «Нормальные люди», как оригинальную версию (на английском языке), так и сербский перевод.

Ключевые слова: немаркированный художественный диалог, немаркированная прямая речевая репрезентация, лингвистические особенности художественного диалога, образцы извлечения из неаннотированных корпусов, художественный диалог Салли Руни

Vesna Polovina*

University of Belgrade – Faculty of Philology
Plenary talk

NON-OMISSION OF DEICTIC PERSONAL PRONOUNS

Abstract

Whereas linguistic literature is abundant concerning the use or non use of personal pronouns as anaphoric units, especially from the syntactic perspective, the non-omission of deictic, 1st and 2nd pronouns, both singular and plural, has received less attention. This is not surprising, since 3rd person pronouns seem to be more frequent units in texts, and typically tie with other syntactic elements. But there seems to be no agreement as to what factors influence their omission (or non-omission), syntactic, phonological, or discourse-pragmatic. Also, the fact that rich agreement languages show tendency to allow deletion of pronouns is not an undisputable fact, since some such languages actually do not allow it, and some poor agreement languages actually do, in some contexts at least. The primarily deictic pronouns in Serbian, a rich agreement language, however, prove to be very frequent units, at least in spoken conversational language. Even in linguistic contexts when the agreement verbal form is overtly indicating person, i.e. when the pronouns could be easily omitted, they are being used. The possible explanations, then, naturally, fall within discourse, pragmatic functioning of these pronouns. Yet it is not easy to determine why non-omission of these pronouns occurs so often. Based on the examples from a corpus of spoken conversational language, we will try to determine their use, i.e. the factors that might be influencing the choice between overt and omitted deictic personal pronouns.

Keywords: non-omission, deictic personal pronouns, conversation, discourse.

* polovinav@fil.bg.ac.rs

1. INTRODUCTION

1.1. ON PERSONAL PRONOUNS – DEICTIC AND NON-DEICTIC PRONOUNS

The pronouns and deixis have been subjects of a great number of studies in not only linguistic, but psycholinguistic, sociolinguistic, computational linguistics, anthropological and philosophical literature. It seems that 3rd person pronoun has been more studied, in view of the fact that it poses the problem of anaphor resolution, i.e. the antecedent–anaphor relation, primarily from a syntactic point of view but also dependent on some functional and discourse factors (Barss 2003). While the third person pronoun is deemed often syntactically undetermined and non-specific, the first and second person pronouns are generally not considered problematic in the sense that they are semantically specific, i.e. referring to the speaker and listener, as Bath puts it, they denote “speech role distinctions rather than distinctions between referents” (Bath 2004: 10). A distinction is being made here between what pronouns ‘denote’ – “direct participants”, – and what other words ‘denote’ – “conceptual content” (Oliva 2018: 72). The sometimes contradictory statements in literature about the semantic content of the personal pronouns are probably due to terminology such as ‘denote’, ‘refer’, ‘signify’, etc. (Lyons 1977: 174–224).

This differentiation between the first and second person pronoun on the one hand and third person pronouns on the other, was already pointed out in Benveniste (1966), who also wrote that an utterance containing the pronoun *je* belongs to the level of pragmatics, and to different types of texts, for example hardly ever to appear in a long scientific text, but almost impossible to avoid in “un court texte parlé” (*ibid.*, 252). There is a “process of discourse appropriation” taking place by means of personal pronouns and all the elements that agree with them, including the verb forms. The language (*langage*) is an instrument used to confide, order, question, inform and provokes a certain response (“comportment”), and if we define discourse as the “*langage mis en action*”, “*nécessairement entre partenaires*”, there is, in

behaviorist's terms, an exchange of stimulus et response. This capacity allows for the fact that subjectivity is established by pronouns *ja* and *tu*, and other "void" forms. The conscience of self is possible only if there is a contrast with another person. (*ibid.*: 258–260). In other words Benveniste claims that "*Le langage propose en quelque sorte des formes "vides" que chaque locuteur en exercice de discours s'approprie et qu'il rapporte à sa "personne", définissant en même temps lui-même comme je et un partenaire comme tu*" (*ibid.* 263).

This interactional situation and its pragmatic impact on the use of pronouns and similar deictic forms is more or less confirmed and further developed in more modern approaches. Conversation is constantly being developed contextually and situationally, as underlined by interactionalist point of view (Hausendorf 2003). There are grammatical signals to signify whose role (speaker/listener/third person's) is active at the moment, plus "demonstration and perception *ad oculos et ad aures*", i.e. visual, acoustic and kinetic signs, separately or together combining to maintain the communication (Hausendorf 2003: 261).

These ideas concerning pragmatic and interactionalist side of the use of deictic signals, such as first and second person pronouns in spontaneous conversations, are highly relevant for our study. It is important to stress that spontaneous conversations evolve round many topics and subtopics that change quickly, that interlocutors' contributions are not planned in advance and in that sense, it is a type of discourse that is quite obviously being co-created verbally by interlocutors – "again and again, as economically as possible and highly inconspicuously" (Hausendorf 2003: 261).

1.2. THE OMISSION AND NON-OMISSION OF PERSONAL PRONOUNS

The interest in pronouns and pronominal systems is closely connected to various approaches to the omission and obligatoriness of pronouns. This is particularly important in typology of languages as pro-drop, null-subject, allowing the lack of overt pronoun, in

which the verb forms as morpho-syntactic categories mark the same 'person' information as the pronoun. In rich agreement languages, as is Serbian (Stevanović 1989, Piper et al. 2005), a bound pronominal affix on the verb can also indicate the referent, thus showing the same information twice. Due to the agreement between the pronoun and the verb form, in such languages one would expect, at least in majority cases, that the morphological form of the verb allows that the pronoun be omitted. But there is no clear-cut division into types of languages according to this criterion. Cole (2010) analyzing, on a sample of six languages, whether covert or overt subjects occur following different types of antecedents, also gives a variety of cases which show that the overt pronoun subjects are possible even "in consistent null subject languages where there is overt subject verb agreement, namely Greek, Serbian and Spanish", whereas for example, in languages with no agreement verb morphology null subjects also occur (Cole 2010: 284).

The agreement itself presents a variety of cases concerning grammatical marking of categories of person, number and gender. Even in South Slavic languages, for example, in which the agreement between person pronouns and verb forms is mainly similar, there are differences. For instance, in Slovenian, that has kept the dual as grammatical number, there are two forms for second person plural, masculine and feminine (Đukanović 2009). In Serbian, for example, the syntactic gender agreement for second person plural is shown only through the agreement in the nominative case of adjectives used in sentences with second person plural pronoun subjects, otherwise the agreement is purely semantically based, especially with first and second person singular (more on interesting examples for Serbian in Popović 1991, 2000, Moskovljević 1983).

Though the pronominalisation can refer back to non-subject grammatical roles, the dominant syntactic role to be referred with an anaphoric pronoun is the subject role. But, even with non-subject role there are many instances of anaphoric use of pronouns (and agreement verb forms). An illustrative example in English is *Few people amaze Brittany*, where the grammatical object is more likely to

be pronominalized further in a sentence/text, since the verb meaning is such that the topic tends to attach to the non-subject syntactic role (Rohde and Kehler 2014: 919). In majority of cases the subject coincides with the topic, a functional, information category, and the anaphoric chain with pronouns and/or agreement morphological markers is created so as to signal the continuation of the topic. Whether the subject position is the default place for the topic in all the languages of the world is not the matter that we can go into in our study, but it is worth mentioning not only because of the obvious reason that subject 1st and 2nd deictic pronouns agree most clearly with verb forms, but also because we have not included in this research any of the non-subject syntactic forms of 1st and 2nd personal pronouns in Serbian that occur in our corpus, such as “logical subjects” occurring with some verbs (for example: *sviđati se* – “to like someone” which has “semantic subject” in an oblique, non-nominative case: *Meni se sviđa Jovan* – “Me (oblique) like Jovan”, or as “qualifiers”: *Nas dvoje smo odlučili* – “Us (oblique) two decided”).

The linguistic literature on referential chain/pronominalization usually points out the chain generally starts with full NPs, most often in subject position, and continues with pronouns (up to a certain, psychologically based distance) or null subjects, or continues with person marked verb forms. Beside the above mentioned preference for antecedents in higher syntactic position (subject as opposed to object, etc.) in languages with rich agreement, “there have been also proposals that overt pronouns in such languages “prefer antecedents in a lower position” (cf. Carminati’s 2002 *Position of Antecedent Hypothesis*, cited in Herbeck 2018: 173). But that, and other similar matters of tackling the referential chain concerning people and things with third person pronouns and verb forms is not within the domain in this study (A comparative overview of Serbian vs English concerning referential continuum with the third person pronouns can be found in Šajinović 2020).

1.3. FIRST AND SECOND PERSON PRONOUNS IN STUDIES OF CONVERSATION

There is an ever-growing literature on the use of deictic personal pronouns in conversational language from pragmalinguistic point of view. Several facts have been documented by theoretical, psycholinguistic and corpus driven studies of deictic personal pronouns (Grenoble 1998, Lenz 2003, Oliva et al 2013, Hernandez et al. 2011, Kragh et al. 2013)

In conversation the frequency of first person pronoun seems to be one of those undisputable facts. In Spanish, the frequency of 1st person subject (*yo*) is followed by 2nd person (*tu*), than 1st person plural (*nosotros*), and 2nd person plural (*usted*) in conversational language (Oliva et al, 2013, 77–79), with different percentages but the same order in our corpus. In spite of the differences between types and size of corpora in this and other studies (Tamaredo 2018, Detges 2013, for example), the general tendencies follow the same pattern.

Secondly, the pragmatic functions of pronouns are generalized into two functions: “*topic-shift*, if it indicates a contrast with respect to the foregoing old discourse topic” and “stance formulae” (*I think, I believe, etc.*) (Detges 2013: 34–45, Benvenist 1966: 264, among others). Detges gives a detailed justification for these uses of the pronouns, and analyzes interesting examples, such as the following:

Moi mes parents mon pere etait sous-chef de gare.
'I my parents my father was second head of station'.

The author’s explanation is that *moi* ‘functions as a conceptual “anchor” – that the “speaker is the most accessible discourse referent in the situation and is therefore maximally suited as a starting point for the elaboration of the topic” (Detges 2013: 35). For other authors (for example, Oliva, 2013: 31–34), the notions of informativeness and saliency are key notions used to explain the discourse relevance and function of deictic personal pronouns.

The second function of the use of deictic personal pronouns is the “attitude”, “stance formula”. This is often linked to a couple of verbs that are most frequent in texts/discourses, especially in

conversational language: *think, believe, know* and *say, tell*, etc. The explanation goes along the following lines: “stance formulae (*I think, I believe*, etc.), express the “speaker’s viewpoint, thereby lending it a special pragmatic weight” (Detges’ example in French is: *Moi, je trouve que c’est pas normal*).

In connection with these frequent verbs, there are clear examples of pragmaticalization of their expression. According to Detges, from “a syntactic and a prosodic point of view, the so-called disjoined or tonic pronouns of Modern Spoken French are neither necessarily disjoined nor stressed elements any more”, from which he assumes that such pronouns, especially the first-person form *moi*, represent the process of “cliticization” (Detges 2013: 34). It is due to the rhetorical, systematic over-use of the constructions undergoing a process of rhetorical devaluation, thereby losing their contrastive potential.

However, in spite of the similarities between languages on a discourse and pragmatic level, it is still difficult to fully, or precisely, explain the functioning of deictic pronouns: What is the reason that the speakers use them so often, even in the presence of other person signals, such as person marked verb? How does a speaker make him/herself “more prominent” or give themselves “extra weight”, or turn themselves into “focus” within a discourse? What is behind the various examples that authors mention, or even get repeated in other paper and studies? For example, Klajn quotes the example given by Bar-Hillel (Bar-Hillel 1954: 367–368, cited in Klajn 1985: 24) of a man who, if he wanted to be “precise” would not be able to ask for breakfast from his wife without using some indexical such as “I”, “here”, “now”. What exactly the words such a man would use is beside the point, since we now quote this same example to illustrate how specific contexts, thought up or observed/experienced by an author, can be found in literature that deals with deictics.

In our analysis of the Serbian conversational language, our chief aim was to answer some questions on the use or non-use of deictic person pronouns with person marked verb forms. The corpus consisted of three types of communication: spontaneous conversations between friends and relatives, mostly students but also participants varying

in age recorded among family members and relatives, recordings of television interviews where most of the guests were politicians or public figures such as sports coaches, people from entertainment industry, and the recordings of parliament sessions. Though the largest corpus was that of spontaneous conversations, the range of tokens for the other two corpora was about 60% of the main corpus (altogether there were 188312 tokens in the whole corpus).

2. THE ANALYSIS OF THE CORPORA

2.1. QUANTITATIVE DESCRIPTION OF DEICTIC PRONOUNS

The striking feature in conversational language is the fact that 1st and 2nd pronouns occur among the most frequent words in corpora. Since interlocutors are present in speech situation, this frequency is at odds with the fact that interlocutors need to refer to themselves so often, when there are other cues, such as voice quality, prosody, kinemic and proxemic signals, and in languages such as Serbian, syntactic agreement markers on verbs in great majority of cases. In order to check whether similar picture presents itself in informal and more formal dialogic contexts, we counted the number of occurrences in Casual conversation, TV interviews and Parliamentary communication. In Table 1 the frequency of each pronoun is expressed as percentage of occurrences within the total number of words/tokens in the relevant corpus:

Table 1. The frequency of first and second person pronouns in three types of spoken discourse¹

| Casual conversation | TV interviews (debates) | Parliamentary discourse |
|---------------------|-------------------------|-------------------------|
| ja – 1,43% | ja – 0,71% | ja – 0,51% |
| ti – 0,35% | ti – 0,01% | ti – (0,00%) |
| (mi – 0,12%) | (mi – 0,43%) | (mi – 0,79%) |
| (vi – 0,01%) | (vi – 0,30%) | (vi – 0,25%) |

¹ These results have also been presented in Panić Cerovski & Polovina (in press).

As can be seen from Table 1., the first person pronoun is by far the most dominant deictic personal pronoun in conversation, twice as much as in TV interviews and almost three times more frequent than in parliamentary discourse. As expected, the second person pronoun *ti* is almost non-existent in more formal contexts, whereas the plural forms, especially the first person plural *mi* rise in frequency with more formal contexts. This picture of the use of personal deictic pronouns in the corpus needs further clarification, in terms of verbs they collocate with, some of them occurring also frequently, with or without pronouns, but marked for person by their endings.

2.1.1. Types of verbs with high frequency

In all three types of corpora there are some verbs in 1st and 2nd subject agreement forms that stand out as the most used, with or without the 1st and 2nd pronouns, and they can be grouped roughly into *verba dicendi* and *intelligendi* used as modal verbs and corpus specific verbs. The first place in the frequency list are always the verb forms: *sam, ste, smo* (“am, are”), used either as a copulative verb or as an auxiliary. Near the top of the list are generally two ‘full’ verbs: *znati* – “know” and *misliti* – “think”. Following them we find in the corpus of conversational language: *imati* – “have”, *moći* – “can/be able”, *kazati* – “say/tell”, *morati* – “must/have to”, *videti* – “see”, *ići* – “go”, *razumeti* – “understand/see”, *reći* – “say”; in the corpus of TV interviews: *kazati* – “say/tell”, *imati* – “have”, *moći* – “can/be able”, *morati* – “must/have to”, *pitati* – “ask”, *reći* – “say”, in parliamentary discourse: *imati* – “have”, *moći* – “can/be able”, *morati* – “must/have to”, *kazati* – “say/tell”, *reći* – “say”, *govoriti* – “speak”, *moliti* – “(beg) ask for permission/attention”, *verovati* – “believe”.

Verba intelligendi et dicendi. Benveniste (1966) wrote that there is only a semblance of the verb forms with three persons paradigm, because when attention is focused on some verbs conjugated in the present tense, for example *je mange, tu manges, il mange*, the sense of that verb, the action of *manger*, seems to remain the same in all three cases. But if one looks at the verbs such as *je crois (que...) je presume*,

je conclus, etc., no action is being “described” – it is an attitude that is expressed concerning the utterance that follows. Such subjectivity occurs only with first person (Benveniste 1966: 264). If some such verb is used with the second person pronoun: *tu supposes qu’il est parti*, again it is just a *verbatim* repetition of the what the previous speaker said in a sort of argumentative discourse, whereas in third person use *il suppose que ...* it is a statement (Benveniste 1966: 265). This is confirmed by the further analysis of these two verbs in our corpus.

First of all, some of the verbs, especially *znati* – “know” and *misliti* – “think”, have 1st and 2nd person forms used as discourse markers in their shortened forms: *'naš* and *mis'im*:

... pa tu prolaze automobili, znaš^{2nd sing} tamo u Leskovcu... (“and there pass the cars, [you] know there in Leskovac...”)

... a jednom je bilo nešto što je 'naš ono kad vidiš da je muva ... (“and once there was [y']know sort of when you see he’s hooking up her...”)

Both *znaš* and *'naš* usually precede a short assertion and invite the listener to accept a brief introduction into a state of affair or situation (Polovina 1994). This function comes from the general use of this verb as a sort of introduction even in its full form as the following example shows:

B: Znaš^{2nd sing} ti^{2nd sing} koju sam ja imala foru sa odg, odg... (“Know you what kind of cool thing I had with my ans, answer...”)

Even if the verb was used with question particle *li*, in its full interrogative form *Znaš li ti ...* it is not a genuine question, but rather a signal that there is an interesting story to follow. The interlocutor could even answer: “No, I don’t know”, which would probably sound a bit unkind, but the speaker would proceed with their story.

In our count of the verbs *znaš* represents 15% of the 2nd sing person marked verb forms but we excluded *'naš* from counting since in that phonetic form it cannot be considered a verb, but as a proper discourse marker.

A similar thing happens with the verb *misliti*, “think” since in one of its uses the 1st person singular form *mislim* has been pragmatized as a discourse marker, most often pronounced in its shortened version: *misim*, so much so that one of the not many abbreviations used by SMS texters in Serbian, especially the younger generation, often write it as *msm* (Polovina 2019: 113–121).

G: Poenta je što me ne boli uopšte, *misim*, nadam se da ne boli, *misim* šta da me boli.

(“G: The point is it doesn’t hurt at all, *I mean*, I hope it doesn’t hurt, *I mean*, why would it hurt.”)

T: Nije to jedina opasnost, *misim*, može da ti se stvori neko stanje.

(“T: It’s not the only risk, *I mean*, some situation can come about to you.”)

Since this form cannot be considered a verb of cognition, but a particle used mostly as a discourse marker to signal the ensuing reformulation, or self-correction, it has been excluded from the total of counted 1st person verb forms, as opposed to *misliti* used as a verb of cognition. As a verb of cognition it is used to express attitude or belief of the speaker towards the statement that follows in the complement clause:

S: Ili spavaćim kolima.

(“S: Or in the sleeping car [of the train]”)

K: Da/ treba videti i to, da li ima?

(“K: Yes/ one should see that too, is there one?”)

A: Ma ne spavaćim kolima! / prvo, idete znaš kad je / mi putujemo celog dana, od koliko // od devet sati ima voz.

(“A: Well not the sleeping car! / first y’know when it’s / we travel the whole day, from // there is a train at nine o’clock”)

M: **Ja mislim** da ima i ranije. **Mislim** da ima ranije.

(“M: I think there is one earlier. [I] think there is one earlier.”)

Another verb that was excluded from the counting was the verb *reći* in the 1st person singular form aorist: *rekoh* > *reko* which is shortened in the same way as the participle *rekao* when used for preterite: *je sam rekao* > *reko*, which again often appears in the conversational language

mostly as a discourse quotative marker: *on je reko oko pola sedam* (“He said about half past six”), *ja reko pa reko znam tata je to imao* (“I said well said I know daddy had it”), ... *a, reko ja čekam Ivanu tamo...* (“Oh, I said I am waiting for Ivana there”), and actually could often be translated into English with “and *I go* ‘I am waiting for Ivana there’”). (More on quotatives see Panić Cerovski & Đukanović & Kovačević 2012; Панић Церовски 2013; Panić Cerovski & Ivanović, 2016.)

There are slightly more than 2% of 1st person verb form of *kazati* (“say”, “tell”) in the total of person marked forms (Table 2.) which also shows some properties of discourse markers, as in:

A: U malim bio
 (“A: was in small [pots])

B: Al’ dobro *kažem*^{1st sing} bio je i u najvećoj zato što sam vadila zemlju.
 (“B: But ok [I] *say* was in the biggest one too because I took out some soil”)

The conversation topic is “a plant that did not grow properly”. The verb *kažem* is at the beginning of the last turn of that segment about the plant, and follows two other introductory words “but ok”, slightly adversative, and together they precede an opposite statement implying: “no, not only the small pots but also in big pots”.

Both 1st singular and 2nd person singular and plural of this verb have the function of signaling that the repetition of something previously said will follow, and together with the 3rd person verb form in present tense, this verb is a kind of quotative verb, similar to Russian *grit* or *gyt*, or *govorju* (1st sg) or *grju*, which is “pervasive in colloquial Russian”, and called a new “evidential particle” by Grenoble (1998: 142).

Verbs specific for the context. It is worth mentioning that in everyday casual communication between friends there are verbs that are frequent but do not appear in other corpora: *voleti* – “like”, *sećati se* – “remember”, for example. This is because there are some frequent themes in casual conversation: what each of the interlocutors likes or dislikes, they talk about themselves and other people, things, events: *ali volim kad on priča a ja slušam*. “I like when he talks and I listen”, *Ja i bananu volim / one, više one zelenije* – “I also like banana / those, more those greener”, *ona ja nju jako volim znaš a ali* – “she, I like her

very much b but”. Or as a way to remind themselves of a story, event, situation: *ja se sećam iz istorije još u osnovnoj školi* – “I remember from history even from the primary school”, *sećam se pričala si* – “I remember you talked”, *e se sećaš kad smo se ja i ti jednom* – “hey d’ye remember once when you and I”.

There are also verbs used with a specific meaning, relevant for the context, that characterize the corpus more than the first ten most frequent. For example, in TV interviews, the verb *varati*, the basic meaning of which is “to cheat”, is used by several speakers in a set form *ako se ne varam* – “if I am not wrong”. In Parliamentary discourse the use of *moliti* – “beg” occurs in context of asking for some extra time for speech: *Još samo jedna rečenica, molim Vas.* – “One more sentence, please”, a criticism of opponents speech: *I, znate, molim Vas, molim Vas, nemojte, nemojte na taj način govoriti o ministrima.* – “And, you now, I beg, I beg of you, do not, do not speak about ministers in that way”, general warning context: *Dakle, molim vas, razmislite* – “Therefore, I beg of you, think it over”. On the other hand, in conversational language, casual speech between friends, this verb has more versatile function: as a formulaic expression M: *Šta, molim!?* – “What, beg your pardon? (What did you say)?”, or just *Molim?* – “Pardon?”, a demand for correcting listener’s behaviour/speech: *E, nemoj da si moja mama, molim te..* – “Don’t be like my mom, please”, etc.

2.2. QUANTITATIVE DESCRIPTION OF THE NON-OMMISSION OF DEICTIC PERSONAL PRONOUNS

Since Serbian is a language with agreement markers on verbs we were interested in checking how often it is not omitted. We counted the number of deictically marked verb forms altogether and then pronoun plus marked form. The following table gives percentage of deictic pronoun plus verb marked forms within the total number of deictically marked 1st and 2nd verb forms.

Table 2. The percentage of deictic Pronoun + V^{agr}

| | Casual conversation | TV interviews | Parliamentary discourse |
|-----------------------------|---------------------|---------------|-------------------------|
| JA + V 1 st sing | 38,44% | 30,1% | 19,37% |
| TI + V 2 nd sing | 18,78% | / | / |
| MI + V 1 st pl | 15,16% | 29,4% | 18,39% |
| VI + V 2 nd pl | 22,93% | 8,54% | 11,25% |

Table 2. shows the percentage of use of the person pronouns that appeared with verbs clearly marked for person. There are variants of the certain verb forms, however, that could not be taken into account. One form, for example, the so-called *potencijal* (conditional) in Serbian grammars, is not always produced with the necessary person ending *bih* (“would”) but is pronounced as *bi* so that it becomes the same as the form for the 2nd and 3rd person singular and 3rd plural, and such cases could not be included in our counting. As a third person form (*bi mogao, bi mogla* “would be able”) it is quite frequent in the corpus, much more frequent than 1st and 2nd person form.

Slightly different is a nearly syncretic form of the modal *moći*, the 1st person singular in the present tense: *moгу*. Though it is the same as 3rd person plural in writing, it differs in spoken language due to different accents, and therefore has been included in the counting. It is more often than not used with another form, so-called *da + present*, which always has the agreement ending, thus showing twice the marking of the first person at a short distance: *Mogu^{1st} da objasnim^{1st}* ([I] can^{1st}, *da+explain^{1st}*, “I can explain it”), *Ne mogu^{1st} da stignem da jedem^{1st}* ([I] cannot^{1st} *da+manage^{1st} da+eat^{1st}*, “[I] cannot get to eat”), and in conversation sometimes even with the omitted *da*: *Mogu^{1st} ga vidim^{1st}*, ([I] can^{1st} him see^{1st}, “I can see him”).

We also ignored the very few examples of 2nd person sing marked verbs in TV interviews and parliamentary interviews from calculation of percentage, since the 2nd singular pronoun did not occur at all, whereas the verb forms appeared only a couple of times. For example, in the corpus of parliament communication the 2nd person verb form appeared only in a quoted proverb:

Ja znam da postoji posledica, poslovice [...] „Kad **si**^{2nd sing} zadužen **nisi**^{2nd sing} slobodan“

(“I know of a proverb ‘When [you] are in debt, [you] are not free’”).

And more interestingly in a paragraph of very critical lines by a representative of opposition directed towards the then president of the country:

... i, Borise Tadiću dok god **podržavaš**^{2nd sing} ovakve ljude, koji prave statute, zastave, himne, koji prave državu u državi mi gubimo silno vreme. nema od tebe^{2nd sing pron oblique} Borise ništa.

(“and, Boris Tadić, as long as [you] **support** such people, who are making states, flags, hymns, who are making a state within a state we are losing immense time, no use of you Boris” ...)

Though formal situation would demand the use of honorific 2nd person plural pronoun and 2nd person plural verb form, nevertheless, the transition to 2nd person singular pronoun is a well-known marked form when the speaker gets angry with someone, even a person of higher social status, meaning some kind of “loss of control” over proper behaviour (Polovina 1983). In this, parliamentary context it is highly marked and adds extra critical and pejorative emphasis.

2.3. THE DISTRIBUTION AND PRAGMATIC FUNCTION OF DEICTIC PRONOUNS PLUS DEICTIC VERB FORMS

In Serbian conversation the non-omission seems not to be the preferable choice for speakers, since the percentage is greater for verb forms used without pronouns. The speakers can easily talk even about one of them as a topic without using pronouns. So how is the topic of *I* or *you* introduced? Is it directly introduced with the pronoun used in a sentence at the beginning of the topic, or some other units? Or do speakers use deictic pronouns as a supportive device, as Bath puts it: “agreement markers are used obligatorily and consistently in these bound-pronoun languages, independent personal pronouns are used only optionally, and have only a supportive role” (Bath 2004: 24). Or they use ‘supportive devices’, i.e. such language units as appositive

noun phrases, vocatives, noun phrases in a matrix clause, etc. for “denoting some of the identifying characteristics of these performers (referents of personal pronouns) in certain non-prototypical contexts (like oath-taking or telephone conversation)” (Bath 2004: 124).

In the following passage we will illustrate two things: 1) that a vocative, a “supportive” device can be used to call out the next speaker, even though it is not necessary since the non-verbal situation can easily indicate who the next speaker is, and 2) that if that person is also the topic, well established, the “fit looks” of the called-out interlocutor in this example, there is no need for the use of any pronoun:

B: Može sa ledom... pa kako si^{2nd} Miloše? Dobro izgleda^{2nd}. Nešto si^{2nd} se promenio. Ne znam^{1st} šta...

(“B: With ice please ... so how are you Miloš (vocative)? You look good. You’ve changed somehow. I don’t know what)

M: Jesam^{1st} se prolepšao?

(“M: Have I become handsomer?)

B: Pa ne znam^{1st}, nešto si^{2nd} se promenio...

(B: Well I don’t know, you’ve changed somehow...)

M: Ne znam^{1st}, bildujem^{1st}. Sad sam^{1st} malo veći...

(“M: I don’t know, doing builds-up. I’m a little bigger...)

B: A ugojio si^{2nd} se sad, promenio frizuru..

(B: And you got some weight, changed your hair...)

S: A frizura! Možda, možda!

(“S: Oh the haircut! Maybe, maybe!)

M: Nije to, bildujem^{1st}.

(M: Not that, am doing build-ups)

In this tight-knit segment there is only a vocative: *Miloše* at the beginning, as a sort of calling out the addressee, and then the topic is introduced: his “good looks”. Since there are other lexical and pragmatic ties and the turns are short, there is no need for any pronouns, only the verbs marked for person are used.

But casual conversations between friends and relatives in our corpus are longer, and the non-omitted pronouns plus verb forms present a challenge to explain, especially if they are felt as syntactically not obligatory. What we found was that there are some typical functions of the use of deictic personal pronouns with person marked verbs. In the following we suggest what these functions and contexts of their use most commonly occur.

3. FIRST PERSON SINGULAR PRONOUN AND FIRST PERSON SINGULAR VERB FORMS

3.1. THE TAKE/UP (APPROPRIATION) OF A TOPIC – ANALOGY AND CONTRAST

Our corpus shows that there are two main positions and types of context in which both the 1st and 2nd person pronouns plus verb agreement forms occur. It is most clear in the case of 1st person. Since in conversations the turns are rarely of considerable length, and the interchange of speakers occurs often, the speaker must have a way of giving a signal to their contribution. It may be with a number of markers: discourse markers, exclamations, use of set phrases, etc, but also combined with using the pronoun *I*. The typical position of this pronoun is at the beginning of a turn. These turns are linked to the previous ones mostly by the relations such as 1) analogy or 2) contrast:

1) analogy

B: znam^{1st sing} da me je cimala sa nekog u fazonu daj kao daj taj broj, [...]
“[I] know that she rang me up from some she was like give like give that number”

C: da, znam^{1st} / i ja^{1st} kad sam^{1st} joj poslala poruku da nije... /
“C: yes, [I] know / also when I sent her a message that she didn’t.../”

In this segment about problems of communicating with a mutual friend over the phone, there is a clear case of analogy, the confirmation that the speaker C. knows of the problem because when she also sent a message there was no communication, the same as it happened to B.

2) contrast

The following examples shows a type of contrastive relation between the turns:

A: A još bolje/mislim/na nju mogu manje da računam nego na tebe zato što ona radi...

(“A: And even better/ I mean/ I can count less on her than on you because she works...”)

S: A **ja**^{1st} ništa ne **radim**^{1st}!

(“S: And I don’t do anything!”)

The contrastive response of speaker S. is not really an introduction of a topic, but the cancellation of the implicature coming from the previous turn in which the speaker implied that S. is not a person one can count on much, but even less is the person she “who works”, the implicature being that S. does not work, and therefore the contrastively used 1st person pronoun in S.’s turn could not be omitted in this context. Here the contrast is emphatic.

The following example shows a more typical context of contrast in conversations between friends:

C: ... tako da nekad ono isključim^{1st} [...] nekad ili okrenem^{1st} na drugu stranu znaš da... [...] a nekad ga jednostavno **ugasim**^{1st} pa ono

(“C: so that sometimes I turn it off [...] or I turn myself to the other side, you know [...] and sometimes I simply turn it off and so”).

D: **ja**^{1st} ga ne **gasim**^{1st}

(“D: I don’t turn it off”)

C: pa i **ja**^{1st} opšte dugo dugo **nisam**^{1st} gasila, sad sam^{1st} nešto ...

(“C: well neither did I turn it of for a along time, but now I have sort of ...”)

The topic is the habit of switching the light off when going to bed. Each of the interlocutors present his / her habit that are opposite. This continuation of a topic by personal contribution of the speakers in this manner can be characterized as elaboration of a topic (Detges 2013: 35), but we could add some more observations on topic change at this moment, since it might better explain the positioning of the first person singular pronoun plus marker agreement verb form.

The point is that in a longer conversation there are frequent topic changes, which often develop into subtopics, and then subsubtopics, so the most relevant distribution of certain discourse markers, including the position of deictic pronouns, can be better seen when we treat paragraph units (though most transcriptions of conversation are usually confined to turns only), i.e. units around a topic/subtopic. When such comparisons are looked at, we can see the clearer picture of the change of topics and use of deictic pronouns. Here is one example, in which we marked end of a paragraph with the symbol (¶), and skipped short turns that are not necessary for this illustration (marked as [...] *n* turns)

1. J: A šta si^{2nd} danas radila sa bakom?
 (“J: What did [you] do with your grandma?”)
2. O: Mesila sam^{1st} kolače. [...]

(“O: “[I] made cookies”)

 [...] 7 turns on making the cookies ¶-1
3. J: Pa dobro jesi^{2nd} ti^{2nd} mesila sa bakom ove s jabukama?
 (“J: Well ok did you make with grandma these with apples?”)
4. O: Jesam^{1st}, ¶-2 (*obraća se baki, S.*) Lepo sam ti rekla da staviš više brašna, a ti^{2nd} nisi^{2nd}... ¶-3
 (“O: Yes. (*Turning to S, grandma*) [I] did tell you to put more flour, and you didn’t..”)
5. S: [...] Stvarno, odnela mi Jeca vagicu, pa kad nemam vagice
 (“S: Really, Jeca took my kitchen scales, so when [I] don’t have the scales”)
 [...] 1turn ¶-4
6. O: Imali smo *mi* neku, pa smo *mi* / pa / kao nešto nije radila / pa smo je popravljali. *Ja* sam ono izvukla pa nije moglo da se uvuče.
 (“O: We had one, and we have / well / it somehow didn’t work / so we were repairing it. I took that thing out and couldn’t put it back”)
7. V: Pa to li si ti^{2nd} pokvarila? ¶-5a
 (“Oh it was you who broke it?”)

- [...] 4 turns continuing on the broken kitchen scales ¶-5
8. J: Pa ima sad vagica kuhinjska [...] ... **Ja** ne mogu da mesim bez vage.
9. (“J: Well there are now kitchen scales ... [...] I can’t make cakes without the scales”)
[...] 1 turn and a pause ¶-6
[...] 5 turns on the high quality of a cake. ¶
10. S: **Kako ti** Jovanka stigneš da sve tako dobro radiš i da odlično kuvaš i da dobijaš nagrade za nauku.
11. (S: “How do you, Jovanka, manage to do things so well and cook well and get awards for science?”)
12. J: **Ja** volim kuvati. To mi je kao neki rad u laboratoriji. **I** za to treba imati ljubavi i strpljenja.
13. (“J: I like cooking. It’s sort of like working in a laboratory. One should have love and patience for that too.”)
S: **Aja** nemam mnogo strpijenja. [...]
14. (“Well, I do not have much patience [...]”)
15. [...] 4 turns on how S dislikes cooking. ¶-7

The general topic of cooking was introduced through a question and answer (turns 1 and 2) without the use of any deictic pronoun. Turn 3 is “elaboration”, again question on a subtopic “cooking of the apple cookie”, and within it a critical remark of the O. to her grandmother S. for forgetting to put enough sugar in the apple cookies, which can be considered as a sub/subtopic, embedded within the two wider topics. The use of the pronoun *ti* + *marker verb* can be said to be a type of individuation, but also fit the further elaboration of the general topic. Turn 5 is cohesive with the previous paragraph, but also signals a new topic: “broken kitchen scales”. Turns 6 and 7 are again new subtopic – with *mi* (“we”), *ja* (“I”) and *ti* (“you”) + person marked verb forms again used to tie up with previous turns but giving new perspectives on “who broke the kitchen scales”, then in 8 a new subtopic with *ja* introducing the “importance of scales for cooking”, and finally the invitation with a vocative and emphatic *ti* within the utterance expressing

the admiration for J's excellent cooking, and two *ja* in a contrastive context: the speaker J. "likes cooking", the speaker "S. dislikes it".

In most of these utterances where deictic pronoun + deictic verb form were combined, their purpose was to take up or "appropriate" a subject and further elaborate from the interlocutors' standpoints, mostly as characters in the events they are talking about. This becomes even clearer in narrative contributions, usually starting with one type of situation and then each of the interlocutors present their "own story".

3.2. APPROPRIATION OF STORY THEMES

In spontaneous conversation the interlocutors very often give reports and stories on what happened to them in the past, share their experiences (Norrick 2000, Jefferson 1978, Mandelbaum 2013, Polovina 2019), some are about themselves and some about third persons. It is therefore to be expected that first person pronoun often occurs in them. Concerning deixis the often quoted Bühler's transfer of *origo* proper into the realm of imagination is illustrated by "mind's" eye or ear in everyday language. Fricke (2003) points out that the speaker, once s/he assumes the speaker role, is in a position to provide the local *origo* or "to intentionally allocate secondary *origos* to intrinsically arranged entities, be these perceptual or imaginary" (Fricke 2003: 88). Since much of the storytelling in our corpus deals with personal experiences of the speaker, it is not surprising that many of the combined uses of 1st person pronoun + 1st person verb form were found in such contexts.

The following excerpt from our corpus of spontaneous conversation will illustrate two typical uses of first person pronoun plus first person verb form, 1) to give an analogous story by the speaker, and 2) to posit himself as a character who participated in a story dialogue, and therefore must quote his own words. This example is taken from a conversation of four students in which the main topics are the "way they remember things", "the way they study

for the exams” and their “university and high school experiences”. Because of the length of this excerpt we skipped some short turns of other participants – comments, expressions of surprise, agreements, exclamations (by marking them as “[... short turn]”). The main topic of the D’s narrative, here segmented into “paragraph” units, marked with the symbol (¶) at the end of each paragraph, is expressed in the first lines (1, 2, 3): “once he had a crib sheet at a history test”, then proceeds with “why he made crib sheet for that test” (lines 4 – 9), “what happened during the test” (lines 9 – 13), and finally “the dialogue between him and the teacher who caught him out” (lines 14 – 18):

1. D: *ja^{1st} sam^{1st} jednom kod Ćuftice a to je bilo jedini put da sam^{1st} iz istorije imao pušku [...]*
 (“D: I once at Ćuftica’s and that was the only time in history class that [I] had crib notes [...]“)
 [... 3 short turns]
2. D: *i onda nam je on dao kontrolni [...]*
 (“D: And then he gave us a control test [...]“)
 [... 2 short turns]
3. D: *gluposti... i sećam^{1st} se bilo ono nešto rani srednji vek ono franačka država ... Hlodoveh neki...*
 (“D: Silly ... and [I] remember something like Middle Ages some Frankish state ... some Chloderic ...”)
 [... 1 short turn]
4. D: *e da, i ja^{1st} tu i sad imamo kontrolni / ¶-1 a ja^{1st} baš naučio dobro i reko^{1st} samo sam^{1st} uzo godine,*
 (“D: And so, and I here and now we have the test / and I really learnt good an’ [I] said [I] just took the years“)
5. *stavio na papir, samo da se podsetim^{1st} godina ja^{1st} otprilike znam, lupam, ajde da je neki*
 (“wrote them on paper, just to remind myself of the years / I sort of knew / for example / ok that)
6. *papa i/zdao neku bulu osamsto neke godine.*
 (“a pope issued a decree in eight hundred and some year“)
 [... 1 short turn]

7. *D: ... znači deveti vek ali ne znam^{1st} tačno godine / ja^{1st} znam_{1st} sve to lepo / stoji onako kako*
 (“D: ... so the ninth century but [I] don’t know the exact year / I know everything / there like how“)
8. *kada zašto sve ali ne znam^{1st} tačno dal je osamsto šezdes pete ili osamsto sedamdes neke /*
 (“when, why, everything, but [I] don’t know if it’s eight hunderd sixty five or eight hunderd seventy“)
9. *znam^{1st}, samo čisto da imam^{1st} brojke napisane. ¶-2 Ja^{1st} napišem^{1st} to i stavim^{1st} ispod i on ništa*
 (“[I] know things / just to have the numbers put down/ I write that down and put it beneath an’ he nothing“)
10. *ja^{1st} radim^{1st} kontrolni, završim^{1st} sve, nisam^{1st} ni koristio taj papir, što je najbolje ,*
 (“I do the test / finish it / everything/ didn’t even use that paper, that’s the best of all“)
11. *i samo sam^{1st} ovako izvadio čisto da proverim^{1st} jednu godinu i on me provali.*
 (“An [I] only took it out like this just to check a year and he found me out.“)
12. *i kao... a reko je ne znam^{1st} kao – ako vidim^{1st} nekoga dajem^{1st} mu keca,*
 (“and like ... an’ he said I don’t know like – if [I] see someone [I] give him a zero“)
13. *i ja^{1st} reko- jao bože šta mi ovo trebalo*
 (“an’ I go – oh God what did [I] need it for“)
 [... 3 short turns] ¶-3
14. *D: a poenta što sam se ja^{1st} toliko unervozio- e nećeš majke ti profesore to nije moje*
 (“and the point is that I got so nervous / hey you won’t damn it – teacher it’s not mine“)
15. *i ja^{1st} krenem^{1st} tako da lažem^{1st} providno, a on fazon zna čovek ono ...*
 (“and I start lying so obviously, and he like the man knew ...“)
 [... 1 short turn]

16. *D: ovaj, aaa, e, a to to mi bio jedini argument koji sam^{1st} mogo.../*
C: da/ ...ja^{1st} reko, a vi ste baš
 (“And well, and that was the only argument [I] could sort of ...
 I said, you teacher are really”)
17. *imali čas pre nas, ja^{1st} reko-pa profesore to sigurno ostalo od*
nekog iz, iz drugo jedan, on kao.
 (“had another class before us, I said, well that must have been
 left by someone from another class”)
18. *Pa tetkica čistila sad i nije, ja^{1st} – pa ne znam^{1st} jaaaa^{1st}.*
 (“Well, the cleaning lady cleaned now but she didn’t, I don’t
 know not meeee”) ¶-4
 (Smeh – “laughter”)

The speaker D. is continuing the series of reminiscences that the four interlocutors have been engaged into in the conversation. In ¶-1 he introduces “his story”, and by using 1st person pronoun + V^{1st} activates himself as a chief character, and then gives the general setting of the story. The next segment of the story ¶-2 starts again with the pronoun *I*, and there is an interesting combination of *origo* replacement in line 14: *i ja^{1st} tu i sad imamo^{1st} kontrolni* – literally “and I here and now we have the test”, which is a short reformulation of the setting and the main character, and a conclusive line for introduction. This is followed, in the same turn (line 9) by an utterance beginning with *ja^{1st} + V^{1st}* to start ¶-3, a kind of psychological justification of his action of taking crib notes to the test: *a ja^{1st} baš naučio dobro* (“and I learned quite well”), *ja^{1st} otprilike znam^{1st}* (“I know approximately ... the year”), *ja^{1st} znam^{1st} sve to lepo* (“I know all that nicely”), thus emphasizing that “he did study, and did know”, implying he did not really need the crib notes. The central part ¶-4 (lines 9-14) of what happened during the test begins with two uses of JA + V^{agree} for the main actions: *Ja^{1st} napišem^{1st} to i stavim^{1st} ispod* (“I wrote that down and put it down under”), and *ja^{1st} radim^{1st} kontrolni, završim^{1st} sve, nisam^{1st} ni koristio taj papir* (“and I do the test, finish it all, didn’t even use that paper”).

These first three segments of the story show that the use of personal pronoun can serve both as a means of segmenting a narrative

structure into its expected parts and simultaneously connecting them, since even though it would be possible to use only the verb agreement form, the story would not easily proceed from the role “I” in the setting, “I” in the psychological description of the main character’s motivation, and “I” as the performer of the main actions.

The last segment ¶-4, “the dialogue between him and the teacher who caught him out”, is the culmination of the narrative – the teacher caught D. having the crib sheet, and a stressful dialogue ensued, where the need to use *on* (“he”) and *ja* (“I”) is pragmatically necessary. In a reported dialogue quotative markers: *reko*, (colloquial form of aorist lacking the first person marker – “said, say”), *krenem* (“start, go”), *fazon, kao* (“sort of”, “like” in lines 15 and 17), are used in this part of the narrative in combination: 1st person pronoun + marked verb forms + quotative markers. A reported dialogue with longer introduction of direct speech would certainly go against the need to create a vivid culminating segment.

Thus, the first person pronoun plus first person marked verb are distributed at the beginning of the segments of narrative structure, helping to segment that structure into paragraph units, “actively placing the speaker” Fricke (2003: 70) as the main character of the story.

3.3. SECOND PERSON SINGULAR PRONOUN AND SECOND PERSON SINGULAR VERB FORMS

As noted in literature for other languages (Helbrecht 2003, Oliva 2013) the use of the second person pronoun singular is much less frequent in conversational language than the first person pronoun. But it is nevertheless used often enough to posit it among relatively frequent words, and therefore worth examining. Most often one thinks about the use of this pronoun in terms of theories of politeness. It is the fact that some languages have grammatical sensitivity to this phenomenon, and that in certain contexts the speakers of Japanese and languages similar in this respect avoid using personal pronouns

when addressing other persons and “rather use status and kinship terms, titles and other complex nominal expressions” (Helmbrecht, 197). However, for our study of the conversational language, we must point out that in a corpus of casual conversation between friends, relatives, young people such as students, politeness must be taken as a very general principle of organization of communication, since the cooperativeness and closeness of the relations between the analyzed speakers does not give much material for some explicit hierarchy of politeness.

It is quite difficult sometimes to determine the pragmatic function of use of 2nd person pronoun in singular. Even if it is about the generalized *ti* in an utterance like: *Ti*^{2nd} *imaš*^{2nd} *i druge programe*. (“You have also other programs”), said by one teenager to another who previously had some negative remarks about the program he was using, the question is – why did the speaker use both the pronoun and the person marked verb? He could have used the marked verb only: *Imaš*^{2nd} *i druge programe*. The generalized meaning would remain. Was it because he wanted to add an implication “so go and find those other programs, and then you will have solved the problem”? In order to try to answer that question we analyzed examples from our corpus, and we suggest some pragmatic functions of *ti* in conversational language.

Most of the cases show that even without special prosodic emphasis, the largest number of uses imply a contrast of *ti* and the speaker, *ti* and other people, *ti* and everyone else.

M: [...] možemo da idemo sutra u poštu
 (“M. [...] we can go to the post office tomorrow”)

N: Možemo
 (“N: We can”)

M: Kad ti^{2nd} *ideš*^{2nd} *na fakultet?*
 (“M: “When are you going to the faculty?””)

N: Pa moram da budem do tri tamo
 (“N: Well I should be there about three”)

Both M. and N. are planning a visit to the post office, but in order to arrange at what time to go, M., who knows her own time schedule, uses the pronoun *ti* here, implying: “I know when I can go, but not when you can go”. Without the pronoun, the question could sound a bit disconnected with the topic of planning, more casual, and could even be understood as a continuation of some previous topic.

One more example of the closeness of interlocutors and emphatic use of both *ti* + verb^{2nd} in a jokingly formulated question, with the contrast implied, and emphasis on the “reproach”:

M. ...pa dobro jesi^{2nd} ti^{2nd} normalna, pa kak... kako ja sad ovo da pijem?
 (“M. now really are you normal, well how...how am I to drink this?”)

In the following example, the contrast is between *ti* and other people:

B: [...] upisaću ja njoj četiri ali ću je pitati kad dode [...] daj brzo i privatne časove/ i inače su mi treba ..
 (“B: [...] I will write down mark four but I’ll ask her when she’s back [...] so get me private lessons quick [...] I needed those anyway”)

C: da; nevezano za...
 (“C: Yeah, not because of ...)

D: dobro ti^{2nd} si^{2nd} bar svesna toga al brate ono
 (“D: OK, you are at least aware of that but oh bro ...)

The conversation is between high school students, and B. is narrating an event when she got assessment by a teacher who said she would give B. mark four, and promised to examine her again, so B. took private classes to be sure not to fail when re-examined. Again, the use of *ti* by speaker D. could be explained as implication that other people in such situation would not be aware that they need extra classes, but that B. was, which also serves as a conclusive and supportive comment after the story.

As for the generalized *ti* it is possible that the use of both pronoun and marked verb contribute some meaning pragmatically, but it is not an easy task to determine for all the cases a common characterization. We mentioned the example of a teenager saying “you have also other

programs”, but apart from saying there is some sort of individuation within the generalized meaning of the sentence, not much else can be concluded. One more example from the corpus:

B: znaš drugo su one cipele dok je beba još mala / već drugo je kad imaš ti četiri godine naš ono ... drugo je sad kad imaš nešto što se šnira ...

(“B: you know one thing is the shoe wear while the baby is still small / but another thing is when you are four years you know ... it’s different when you have something you must tie the laces” ...)

Whereas in the “*you have other programs*” example we could suggest some nuance of hortative meaning, in the previous example about “shoe wear”, the utterance context does not allow such interpretation. There may be a simple case of added emphasis on the general condition “when you are four”, since similar argument has already been given in a previous turn.

3.4. FIRST AND SECOND PERSON PLURAL PRONOUNS AND MARKED AGREEMENT VERB FORMS

Most of the discussions of first person plural in literature bring in the question of whether it is inclusive or exclusive plural, i.e. whether the use of *we* denotes speaker and listener(s) or speaker/listener(s) and third person(s), or speaker and third persons. However in our corpus it seems that this issue is not bothering much the interlocutors, since we have not found any demand for clarification, although it is possible to imagine that in case of a confusion any of the listeners can ask for clarification “Who do you mean by ‘we’?”. This can be explained partly by the fact that interlocutors are friends and have common social background, or that the speakers, if they think it necessary, as some examples will show, do give additional information. As one author puts it “Covert inclusivity can only be inferred from the situational context: if a speaker A welcomes a hearer B with the words *I’m glad we two could meet*, it can be assumed that *we two* means ‘A and B’.or as in *We/ Us linguists are a crazy bunch* or *We/ Us three have to be leaving now*. (Hernandez 2011: 144). Similarly to the functioning of the first

person singular, approximately 50% of all the occurrences of *mi* is somewhere at the beginning of a turn, serving a rhetorical function of initiating analogues experience:

D: a mi^{1st pl} kod Emice kad smo^{1st pl} radili sastave / Vanja je uvek pre nego što, otprilike uvek znali smo na temu nešto uglavnom i svaki put pre pismenog [...] i Vanja kupi „Moju tajnu/ Moju sudbinu“ nešto pročitati i bukvalno uzme zaplet rečenice i to i napiše sastav i uvek je dobijala dobre ocene kad je to radila, uvek...
 (“D: and we at Emica’s when we did essays / Vanja would always before we, sort of always knew the topic something and every time before the essay writing Vanja buys “*My secret/ My destiny*” reads something and literally takes over the plot sentences and so on and writes the essay and she always got good marks when she did that, always”)

This is the third “experience” in the line of anecdotes, stories from school, and again we can see the pattern of the introduction with the connective *a* plus first person pronoun plus person marked verb to refer to the class of D. and then the story proceeds about a classmate (Vanja), a member of the “we”/group, who cleverly read a literary piece and then used its sentences/ideas in her essay writing.

The following example illustrates another typical occurrence of *we*:

D: a pazi, pazi – Milan, ja, ovde Nevena i ovde / Ema sela tu je bilo kao slobodno a iza nje Maja i Kostićka. Znači Milan, ja, Nevena i ovde sada Ema se okrenula njima kao / [...] znači okrenuta je njima [...] a mi bukvalno ja sam ja samo meni je Milan bukvalno dao Neveni moju vežbanku da proveriti i nas troje smo^{1st pl} svi dobili petice, [...]

(“D: but look, look – Milan, me, here Nevena and here / Ema sat there it was sort of empty and behind her Maja and Kostić. So, Milan, me, Nevena and now here Ema looking towards them / ... and **we** literally I only I only towards me and Milan literally gave Nevena my notebook to check and the three of us all got five”) [...]

There is obviously an effort on the part of the speaker D. to first describe the situation very clearly as to where the *we/three of us* were seated in contrast to Ema (the teacher), which is repeated twice and then proceeds to the main “action” during the class (“my notebook from Vlada to Tamara”), and finishes off with “the three of us all got five (the highest mark)”. Even though this is perhaps an example of the

speaker's extreme concern to define where the "group" she belongs to were seated, it is typical for talking about experiences of a group including speaker and third people in conversational contexts.

Another way of defining the group to which the speaker belongs, if necessary, is by use of *we* + defining NP/PP/S that defines the group: *Mi što smo*^{1st pl} *navikli da živimo*^{1st pl} ..., ("We who are used to living...") *Majke mi, mi kući jedemo*^{1st pl} *rukama*, ("Honestly, we at home use hands to eat"), *Pogotovo mi koji se bavimo*^{1st pl} *time*, ("Especially we who do that job"), *E tako da mi imamo*^{1st pl} *bosanskog porekla*, ("And so we have Bosnian origin"), *Mi ovako, mi iz Leskovac ovako zavrćemo*^{1st pl} ("We like this, we from Leskovac accentuate like this"). These are contexts in which *we* plus agreement verb forms is sometimes obligatory for syntactic reasons, but even without wider context it is obvious that most of these utterances are either emphatic conclusive remarks, or elaboration of a topic.

4. CONVERSATIONAL LANGUAGE VS TV INTERVIEWS AND PARLIAMENTARY DISCOURSE

Since our primary goal was to study the use of pronouns in conversational language, the comparison with spoken television interviews and parliamentary communication was undertaken to see what similarities and differences can be observed, beside the mere frequency of the use of the deictic pronouns and their non-omission in those contexts.

One general observation is as expected, the second person singular pronoun in more formal context is practically non-existent. This does not mean that in another corpus, for example, TV talk shows, the relationship between the host and the guests cannot be more friendly, and then the use of those forms could appear more natural. The relatively frequent use of the first person singular seems to be more connected to the anchor's turns in TV interviews, and the speech acts and modal utterances in parliamentary discourse. But this general statement could certainly be qualified further if these

two corpora were analyzed in detail. For example, we find that *ja* + *mislím* “I think” might be connected with more critical turns, as in the following example from parliamentary discourse:

... Ja () mislim da je, apsolutno, nekorektno na političkoj sceni Srbije, političke stranke kvalifikovati da su dobre ako žele da saraduju sa vama a ne, ukoliko tu saradnju ne žele.

(“I think that it is, absolutely, unfitting at the Serbian political scene, to qualify the political parties as good if they want to cooperate with you, and not good, if they don’t want to have that cooperation”)

Or emphasizing an individual for a critique:

Gospodine ministre, tužno je to je pričao, to nije smešno, Vi niste bili prisutni, tako da Vas molim da se ne smejete.

(“Mister Minister, it is sad what he told, it’s not funny, **you** were not present, so that I beg you not to laugh”).

Or used to initiate a conclusive remark:

Ja mislim da smo mi u ovom slučaju (), praveći jedan širi aranžman, koji se ne tiče samo NIS-a, nego je mnogo značajniji strateški, zapravo napravili dobar izbor za Srbiju. Hvala!

(“I think that **we** have, making a wider arrangement, not only for NIS, but much more strategically significant, made a good choice for Serbia. Thank you!”)

In parliamentary discourse with opposing political parties and opinions there are also emphatic uses of deictic pronouns plus marked verbs, serving exactly that need, to emphasize the opposing groups:

[...] Ovo je pisao neko iz DSS-a pošto vi nikada nećete imati predsednika Republike! Očigledno! A mi koji se spremamo za tu funkciju zaista bismo želeli da znamo šta to od nas zahtevate, da vam kažemo da li možemo da ispunimo vaše uslove ili ne! [...]

(“[...] This was written by someone from DSS, since **you** are never going to produce the president of the Republic! Obviously! And **we** who are preparing for that position we would really like to know what you are asking from us to do [...]”)

These cases not fundamentally different from conversation, but situations are different and impact the style and function of use of

deictic pronouns and verb forms. In TV interviews the relationship between the TV host and guests is respectful, the former being more dominant in asking questions and assigning the topics. In parliamentary discourse the communication is restricted by administrative rules and the 'turns' are of limited duration, but usually much longer than in conversation or TV interviews. There is no free exchange, since each member of the parliament must get permission to speak from the presiding person, and might be warned to get back to an in advance established topic, if necessary. Also, while in conversation and TV interviews there is a sense of solidarity and cooperation, the parliamentary discourse presupposes opposition between parties, whereas in conversation and interviews there is less motivation to oppose the opinion of the other.

Another difference between the three corpora, which should be more balanced in some further research, is the fact that the corpus of conversational language includes dozens of different speakers, the TV interviews are mostly those with one host plus one guest, and the host is in majority of those interviews the same individual. Some possible social differences between the speakers, such as age and gender could not be compared either, even less the individual differences, which might characterize certain speakers in different contexts.

5. CONCLUSIVE REMARKS AND FURTHER RESEARCH

Our analysis of the corpus of conversational language in informal and more formal communication situations has shown that there are strong similarities with other rich agreement languages such as Spanish, French and Russian in terms of the variations in expressing subjectivity, attitude towards the ensuing complement sentences after modal verbs or verbs of cognition and speech, for example. Some of the most frequent verbs have also undergone the process of pragmaticalization, and have become discourse/pragmatic markers in one of their uses. Another feature of the use of deictic person pronouns and agreement verb forms is their distribution – they occur

within the process of elaboration of a topic, or as “take-ups” of a topic, either by analogy or contrast, but not within the initiating turns (at least not without some supportive devices).

We insisted on the deictic pronoun not being omitted when the relevant verb form is used, but it would be interesting to study further the use of “logical subjects” in stressed versus non-stressed form (*Nama je velika čast* “To us it is a great honour” vs. *Velika nam je čast* – “It is our great honour”) as well as some variants that occur in the conversational language. We also noted that prosodically these deictic pronouns are not accented, but behave more as proclitics (as observed for French, Detges 2013). Only a few seem to be more accented (e.g. *Pa to li si ti^{2nd} pokvarila?* – “Oh it was you who broke it?”). It would be interesting to compare prosodic and possibly kinesics aspects of communication with these diverse uses – with overt, non-omitted pronoun and without it.

One more general topic from the point of view of discourse and pragmatics would be a thorough comparison between the use of all the person pronouns deictic and non-deictic third person, since even with the deictic forms plus agreement verb forms, they seem to follow the general rule, that in case they are used (for the take-up of the topic, as explicit introduction into a speech act, a conclusive remark, or other reasons), the following utterance would include a number of omitted deictic pronouns, unless there is a need for disambiguation if two or more characters are included in the story or conversational segment. Transitions to the next segments of conversation, new topics/subtopics, expressed in paragraphs, would quite often prompt a new use of deictic pronouns plus agreement verb form.

Generally speaking, the findings confirm that the primary function of non-omission of deictic pronouns in Serbian is pragmatic in nature and a discourse matter, linked with psychological need of the speaker to re-establish their role and other people’s roles in it, so as to clearly indicate their stance towards a statement, or towards the parts of the discourse itself.

REFERENCES

- Barss, A. (2003). *Anaphora. A Reference Guide*. Malden-Oxford-Melbourne: Blackwell Publishing Ltd.
- Benveniste, E. (1966). *Problèmes de linguistique générale*. Paris: Gallimard.
- Carminati, M. N. (2002). *The processing of Italian subject pronouns*. [Doctoral dissertation, University of Massachusetts at Amherst, Amherst, MA].
- Cole, M. (2010). Thematic null subjects and accessibility. *Studia Linguistica* 64(3), 271–320.
- Detges, U. (2013). First person strong pronouns in spoken French. A case study in cliticization. In J. K. Kragh, & J. Lindschouw (Eds.), *Deixis and Pronouns in Romance Languages* (pp. 33–48). Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Fricke, E. (2003). *Origo*, pointing, and conceptualisation – what gestures reveal about the nature of the *origo* in face-to-face interaction. In: F. Lenz (Ed.), *Deictic Conceptualisation of Space, Time and Person* (pp. 69–93). Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Fritz, T. A. (2003). “Look here, what I am saying!” Speaker deixis and implicature as the basis of modality and future tense. In F. Lenz (Ed.), *Deictic Conceptualisation of Space, Time and Person* (pp. 135–152). Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Grenoble, L. (1998). *Deixis and Information Packaging in Russian Discourse*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Hausendorf, H. (2003). Deixis and speech situation revisited: The mechanism of perceived perception. In F. Lenz (Ed.), *Deictic Conceptualisation of Space, Time and Person* (pp. 249–270). Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Helmbrecht, J. (2003). Politeness distinctions in second person pronouns. In F. Lenz (Ed.), *Deictic Conceptualisation of Space, Time and Person* (pp. 185–202). Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Herbeck, P. (2018). Deriving Null, Strong and Emphatic Pronouns in Romance *Pro*-Drop Languages, In P. Patel-Grosz, P. G. Grosz, & S. Zobel (Eds.), *Pronouns in Embedded Contexts at the Syntax-Semantics Interface* (pp. 171–214). Cham: Springer International Publishing.
- Hernández, N. (2011). Personal pronouns. In N. Hernández, D. Kolbe, & M. E. Shulz (Eds.), *A Comparative Grammar of British English Dialects*.

- Modals, Pronouns and Complement Clauses* (pp. 53–192). Berlin-Boston: Walter de Gruyter.
- Hernández, N., Kolbe, D., & Schulz, M. E. (2011). *A Comparative Grammar of British English Dialects. Modals, Pronouns and Complement Clauses*. Berlin-Boston: Walter de Gruyter
- Jefferson, G. (1978). Sequential aspects of storytelling in conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 219–248). New York: Academic Press.
- Jungbluth, K., & Da Milano, F. (2015). *Manual of Deixis in Romance Languages*. Berlin-Boston: Walter de Gruyter.
- Klajn, I. (1985). *Ofunkciji i prirodizamenica*. Beograd: Institut za srpskohrvatski jezik.
- Kragh, J. K., & Lindschouw, J. (Eds.). (2013). *Deixis and Pronouns in Romance Languages*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Lenz, F. (ed.). (2003). *Deictic Conceptualisation of Space, Time and Person*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Lyons, J. (1997). *Semantics I*. Cambridge: Cambridge University Press.
- Mandelbaum, J. (2013). Storytelling in conversation. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 492–507). Malden, MA: Wiley-Blackwell.
- Norrick, N. R. (2000). *Conversational narrative: storytelling in everyday talk*. Amsterdam: John Benjamins B.V.
- Oliva, M. A. I., & Serrano, M. J. (2013). *Investigating Variation in Spanish Pronoun Subjects*. Bern: Peter Lang AG.
- Panić Cerovski, N., Đukanović, M., & Kovačević, B. (2012). Novi markeri citiranja u srpskom i slovenačkom jeziku. In *Nauka i savremeni univerzitet – Filologija i univerzitet* (pp. 121–131). Niš: Filozofski fakultet Univerziteta u Nišu.
- Panić Cerovski, N., & Ivanović, B. (2017). Glagoli govorenja kao markeri emfaze – slučaj pseudocitiranja. *Komunikacija I Kultura Online*, 7(7), 142–154. <https://doi.org/10.18485/kkonline.2016.7.7.10>
- Panić Cerovski, N. i Polovina, V. (in press). Centralna deiksa u razgovornom diskursu. In S. Gudurić, M. Stefanović, & J. Dražić (Eds.), *Jezici i kulture u vremenu i prostoru 10*. Novi Sad: Filozofski fakultet.
- Patel-Grosz, P., Grosz, P. G. and Zobel, S. (Eds.). (2018). *Pronouns in Embedded Contexts at the Syntax-Semantics Interface*. Cham: Springer International Publishing AG.

- Pešikan, M. (1967). O sistemu zameničkih reči. *Naš jezik* XVI(4). Beograd, 245–267.
- Polovina, V. (1983). Upotreba jednine i množine ličnih zamenica u obraćanju sagovorniku u savremenom srpskohrvatskom jeziku. *Naučni sastanak slavista u Vukove dane XIII*, 185–195.
- Polovina, V. (1988). The Basic “Verba Dicendi” and their cohesive role in spoken conversation. *Acta Linguistica Hungarica*, 38(1/4), 193–200.
- Polovina, V. (1994). Glagol “znati” u razgovornom jeziku. *Zbornik Matice srpske za filologiju i lingvistiku* XXXVII, 511–523.
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience* 29(8), 912–927.
- Šajinović Novaković, J. (2020). *Jezik, izraz, upućivanje*. Doboj: Štampa.
- Tamaredo, I. (2020). *Complexity, Efficiency, and Language Contact, Pronoun Omission in World Englishes*. Bern: Peter Lang.
- Ђукановић, М. (2009). *Упоредна семантичка анализа заменичких речи у словеначком и српском језику*. Београд: Филолошки факултет.
- Московљевић, Ј. (1983). О конгруенцији глагола са напоредном именичком синтагмом. *Психологија: часопис Савеза друштва психолога Србије* 16(4), 72–78.
- Панић Церовски, Н. (2013). Цитирање у формалном и неформалном дискурсу. In А. Вранеш, & Љ. Марковић (Eds.), *Културе у дијалогу. Филологија и интеркултуралност* (Vol. I, pp. 177–196).
- Пипер, П., Антонић, И., Ружић, В., Танасић, С., Поповић, Љ., & Тошовић, Б. (2005). *Синтакса савременог српског језика: Проста реченица*. Београд-Нови Сад: Институт за српски језик САНУ, Београдска књига, Матица српска.
- Половина, В. (2019). *Сигнали дискурса*. Београд: Филолошки факултет.
- Поповић, Љ. (1991). Хонорифичка и семантичка конгруенција при учтеном обраћању. *Књижевност и језик* 38(1), 38–53.
- Поповић, Љ. (2000). Бивалентни контролори конгруенције, проблем лексикографског описа конгруенције граматичког и семантичког слагања. *Научни састајак слаvista у Вукове дане* 29(1), 65–80.
- Стевановић, М. (1989). *Савремени српскохрватски језик II (граматички системи и књижевнојезичка норма)*. Београд: Научна књига.

Vesna Polovina

NON-OMISSION DES PRONOMS PERSONNELS DÉICTIQUES

Résumé

Tandis que la littérature linguistique abonde en articles concernant l'usage ou le non-usage des pronoms personnels en tant qu'unités anaphoriques, surtout du point de vue de la syntaxe, la non-omission des pronoms déictiques – à savoir de la 1^{ère} et de la 2^{ème} personne du singulier et du pluriel – a reçu une attention moindre. Ces pronoms, principalement déictiques en serbe – langue riche en accords –, s'avèrent en effet être des unités très fréquentes, au moins dans la langue parlée dans le cadre d'une conversation. Ils sont utilisés même dans les contextes linguistiques où la forme verbale indique ouvertement la personne, c'est-à-dire où les pronoms pourraient être facilement omis. Les explications possibles résident alors naturellement dans le fonctionnement discursif, fonctionnel de ces pronoms (Benveniste, 1966). Récemment, on découvre une littérature croissante sur l'usage des pronoms personnels déictiques dans la conversation (espagnol, portugais, français etc.) du point de vue pragmatolinguistique.

Notre analyse du corpus du langage parlé dans les situations de communication non-formelles ou plutôt formelles a démontré qu'il existe des similarités considérables avec les autres langues riches en accords, telles que l'espagnol, le français et le russe, par exemple, en ce qui concerne les variations dans l'expressions de la subjectivité et dans l'attitude envers les complétives régies par les verbes modaux, les verbes épistémiques ou les verbes d'énonciation. Quelques-uns parmi les verbes les plus fréquents à la première ou à la deuxième personne (*mislim, kažem, znaš*) ont également subi le processus de pragmatolinguistique, de sorte qu'ils sont devenus marqueurs discursifs/pragmatiques. Une autre caractéristique de l'utilisation des pronoms personnels déictiques et des formes verbales comportant l'accord est leur distribution : ils apparaissent dans le cadre de l'élaboration d'un sujet, ou de la reprise d'un sujet – soit par l'analogie, soit par le contraste –, mais jamais dans les tournures d'ouverture d'un sujet (au moins sans dispositifs de support).

Mots-clés: non-omission, pronoms personnels déictiques, conversation, discours

Language and Culture

Junichi Toyota*
Osaka City University

UNIFORMITARIANISM AND ‘YES’ AND ‘NO’ IN INDO-EUROPEAN LANGUAGE¹

Abstract

The ‘yes’-‘no’ (Y-N) words do not always exist, and the Celtic languages, for instance, lack them. In other IE languages, the N word etymologically stems from a negative marker, but there are two main sources for the Y word, e.g. the copula and a complementiser or a subordinate clause. Geographically, the copula type is found in northern Europe, whereas the complementiser type, in southern and eastern Europe, except for Greek and dialects of Slovenian. The distributional pattern of the Y-N words is relatively easily formulated, but this paper examines it in terms of uniformitarianism. Due to the colloquial nature of the Y-N words, it is difficult to study them historically; however, examining what is happening now in the Celtic languages enables us to ascertain what must have happened in other branches of the IE languages where historical records are scarce. Celtic languages are developing their own Y-N words, along with a loan from English, representing both gradual changes found in uniformitarianism and abrupt innovations/changes often observable in catastrophism. In addition, some social factors such as shifts in religion, might have affected the development, thus suggesting another example of catastrophism. Therefore, by looking at the Y-N words, it is possible that a new perspective in historical change can be gained.

Key words: uniformitarianism, catastrophism, yes, no, language contact, copula, complementiser

* toyotaj@lit.osaka-cu.ac.jp

¹ Abbreviations used in this paper are as follows: COP = copula; NEG = negative marker; PRS = present; Q = question marker; SG = singular; VN = verbal noun; 1 = first person.

INTRODUCTION

‘Yes’ and ‘no’ are normally taken for granted in our daily use of language, and little attention has been given to them in linguistic research; in particular, historical studies. Their etymological sources have not been thoroughly studied, and they are merely mentioned in various dictionaries. The ‘yes’-‘no’ (Y-N) words are not mere replies, and they can provide us with a rich historical perspective into how people communicated in the past, even beyond the recorded history. Thus, research concerning them can give us insights into various areas of linguistic studies. This paper attempts to incorporate uniformitarianism to analyse the evolution of the Y-N words and shed light on various issues in historical changes more comprehensively among the Indo-European (IE) languages in Europe. Uniformitarianism involves both gradual shifts in form and function as well as abrupt changes, i.e. catastrophism. By incorporating both types, it is possible to predict what future changes can occur. Thus, comparative analysis based on uniformitarianism can cover a wide range of changes.

This paper is organised as follows: the nature of uniformitarianism is introduced, describing both gradual and abrupt changes. Followed by this, Y=N words in the Indo-European languages are reviewed, including their etymology. A distributional pattern of the Y-N words is also illustrated. Then the Celtic languages are analysed in detail, including their ongoing changes. Finally, uniformitarianism is reviewed in terms of the Y-N words in the Indo-European languages.

UNIFORMITARIANISM AND LANGUAGE CHANGE

Uniformitarianism was popularised by Lyell (1830–33) in the 19th century in the field of geology, but it was soon applied to other fields of study on historical principle. It refers to an interpretation of changes in the past by means of processes that are currently observable. Gordon (2013: 82), for instance, defines it as follows:

The uniformitarian principle assumes that the behavio[u]r of nature is regular and indicative of an objective causal structure in which presently

operative causes may be projected into the past to explain the historical development of the physical world and projected into the future for the purposes of prediction and control. In short, it involves the process of inferring past causes from presently observable effects under the assumption that the fundamental causal regularities of the world have not changed over time.

Lyell's claim has four main points, i.e. natural laws, process of changes, rate of processes and physical state, and it was claimed that they have been constant across time and space. Thus, changes are considered slow and steady, and the direction of change is not inexorable. However, some of them have been challenged, initially by Gould (1965). As Ager (1993: 81) states, the present may not be long enough to observe intricate mechanisms of changes in the past. Furthermore, the rate of processes is also questioned, claiming that it may not be uniformly gradual through time (cf. Smith & Pun 2006). Catastrophism may be also applicable, i.e. some changes in the past must have been sudden, short-lived and violent, which may not be easily observable now and thus, the process is not always uniformly gradual, but rather, a gradual process punctuated with some sudden changes.

Uniformitarianism has been applied to linguistics (Christy 1983; Walkden 2019), and it is somehow normally assumed in historical linguistics. For instance, Proto-Indo-European, a reconstructed original language of modern Indo-European languages, is known to have the active-stative alignment (cf. Gamkrelidze and Ivanov 1995), which is not found in its modern daughter languages. However, this alignment is commonly found among the Caucasian languages and some of the native languages in North America such as Siouan, Iroquoian and Na-Dene languages. Thus, the presence of this structure among the modern languages ensures applicability of active-stative alignment to the reconstructed Proto-Indo-European.

As for gradualness of changes, a pattern of historical change known as grammaticalisation is commonly assumed to complete its unidirectional cycle of gradual change in 2,000 to 3,000 years. However, some factors can affect the gradualness, or reverse the unidirectionality. For instance, language contacts are known to shorten

the process of grammaticalisation (Heine and Kuteva 2005), and its cycle is reduced to one tenth of its original length, i.e. 200 to 300 years. Thus, the reverse chain of changes known as degrammaticalisation also occurs in a relatively short period of time, i.e. 200-300 years, in comparison with normal grammaticalisation, and some social factors such as social identity or religious beliefs are a driving force behind degrammaticalisation. Burridge (2004), for instance, reports a case of degrammaticalisation in Pennsylvanian German, i.e. a change from a future tense auxiliary *welle* ‘will’ into a lexical verb *wotte* ‘want’, replacing the original lexical verb *winsche* ‘wish’. This change happened because speakers felt that it was blasphemy for common people to talk about the future, since it is believed that only god knows what future may hold for them. Thus, the future tense was purposely avoided, and the same expression was used for a modest wish for the future.

Another case of uneven change is found in the evolution of language. Toyota (2012) argues that there is a drastic change at an earlier stage in evolution of human language, claiming that the emergence of verbs from nouns took much longer, and once the binary pair of noun and verb was firmly established, the normal unidirectional grammaticalisation could take place, and as presented in Heine and Kuteva (2007: 111), the changes after the noun-verb binary pair is predictable. If the age of human language is considered around 100,000 to 150,000, as Toyota (2012: 111) argues, ‘about nine-tenths of the evolution (considering the age of language as 100,000 years) were spent on creating a binary feature [between noun and verb].’ Thus, the gradual change was observable only in the past 10,000 to 12,000 years, and there was a ‘violent’ change that shaped the basic outline of our modern languages.

In the rest of the current paper, this principle is used to analyse how the Y-N words were developed.

Y-N WORDS IN INDO-EUROPEAN LANGUAGES AND THEIR ETYMOLOGY

The Y-N words are normally believed to exist in every language, but this is not always the case. When it comes to the IE languages, there are different sources for the Y-N words, and some even lack these words. Table 1 illustrates a sample of the Y-N words in the IE languages. Celtic languages lack these words, but they are currently going through radical changes. These languages are analysed in details in a later section.

Table 1. Representatives of Y-N words

| | YES | NO |
|----------|---------------------|-----------------|
| Germanic | <i>yes, ja</i> | <i>no, nein</i> |
| Slavic | <i>da, tak, ano</i> | <i>ne, nie</i> |
| Baltic | <i>taip</i> | <i>ne</i> |
| Romance | <i>sì</i> | <i>non</i> |
| Greek | <i>ne</i> | <i>ochi</i> |
| Celtic | – | – |

Etymologically, N-words are derived from a negative marker in each language. Variations can be found concerning the etymon for Y-words. There are four main sources of the Y-words, which are copula, demonstrative, temporal/conditional phrase and conjunction/complementiser. However, even within a single branch of the IE languages, diversity can be found, as demonstrated in Table 2. For instance, Slavic languages normally take the conjunction/complementiser as a source for the Y words, but the source in Proto-Indo-European (PIE) differs, e.g. **doh* ‘thus, like so’ for Bulgarian *da*, but **tako* ‘thus, so’ for Polish. It is believed that the choice is related to the choice of religion, i.e. Slavic languages initially used PIE **tako* ‘thus, so’ for the Y-word, but with the spread of the Orthodoxy in the

region of East and South Slavic languages, a new form **doh* ‘thus, like so’ spread. Thus, the West Slavic languages have a variation in the Y-words. Apart from these, language contact played a role of creating Y-words. Romanian *da*, for instance is a loan from the South Slavic languages. Romanian is a part of the Balkan Sprachbund, sharing the South Slavic grammatical characteristics in general. Romanian *dais* argued as a Romance trait by some, e.g. Massey (2008), claiming that Latin *ita* ‘thus, so’ is the source. However, judging from other common grammatical features within the South Slavic languages, *da* is normally considered as a loan word. Czech and Slovak has their own *ano* ‘yes’, but also use *jo* ‘yes’ in colloquial speech. This is a loan from Hungarian. Likewise, Slovenian uses *ja*, a loan from German, along with *da*.

Table 2. Etymological sources for Y words

| | | |
|-------------------------------|------------------------|---|
| a. Temporal, conditional | | |
| | Albanian <i>po</i> | > PIE <i>*h₁é</i> ‘then, at that time’ |
| b. Conjunction/complementiser | | |
| | Bulgarian <i>da</i> | > PIE <i>*doh</i> ‘thus, like so’ |
| | Slovak <i>áno</i> | > <i>to je ono</i> ‘that is it’ |
| | Polish <i>tak</i> | > PIE <i>*tako</i> ‘thus, so’ |
| | German <i>ja</i> | > Proto-Germanic <i>*jǣ</i> ‘so’ |
| c. Demonstrative | | |
| | Lithuanian <i>taip</i> | > <i>taī</i> ‘that’ + adverbial suffix <i>-aīp</i> |
| | Italian <i>si</i> | > Latin <i>sice</i> > <i>sī</i> + <i>ce</i> , from PIE <i>*só</i> ‘this, that’ + PIE <i>*h₁e-</i> ‘demonstrative particle’ |
| | French <i>oui</i> | > Latin <i>hoc ille</i> ‘this he’ |
| d. Copula | | |
| | English <i>yes</i> | > Proto-Germanic <i>*jǣ</i> ‘so’ + <i>*sīe</i> ‘be (it)’ (‘so be it’) |
| | Greek <i>nei</i> | > <i>einei</i> ‘be’ |

| | | |
|---------------|-------------------------|-----------------------------------|
| e. Loan words | | |
| | Romanian <i>da</i> | > South Slavic language <i>da</i> |
| | Czech, Slovak <i>jo</i> | > Hungarian <i>jo</i> |

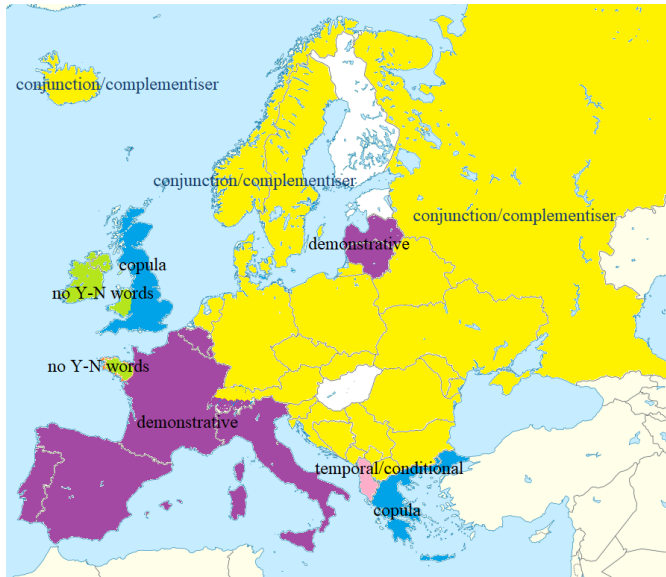


Figure 1. Distribution of Y-N words (keys: blue, copula; yellow, conjunction/ complementiser; purple, demonstrative; pink, temporal/conditional; green, no Y-N words)

As for the distribution of the etymological sources, the Romance languages except Romanian, form one area with the demonstrative source, along with the Baltic states. The rest of mainland Europe has the conjugation/complementiser source. The copula source is found peripherally in Europe. Also note that the Celtic languages without the Y-N words are also found in the peripheral. This language family is examined in detail in the next section.

 CELTIC LANGUAGES AND ONGOING CHANGES

Celtic languages behave differently in respect to responses and rejoinders among the IE languages, since they do not possess the Y-N words. Without the Y-N words, speakers of the Celtic languages answer a question by repeating a main verb, as exemplified in (1). The main verb *téann* in (1a) is a dependent form of *téigh* ‘go’ in the present tense, and an affirmative reply is made simply by repeating a main verb, i.e. (1b). A negative reply is made with the addition of a negative marker *ní* to the affirmative answer, i.e. (1c). However, some forms of the Y-N words can be found in colloquial speech, and they are currently developing what is going to be ‘yes’ and ‘no’. The current state of the Y-N words is shown in Table 3. What is commonly found among them is the copula phrase ‘it is’, e.g. Irish *is sea* ‘it is’ or its shortened form *seo*, and a loan from English, e.g. *yeah* or *aye* in Scottish Gaelic. Note that even in Breton, a single language not spoken in the British Isles, an influence from English is visible in *ya*. Also, Irish uses a copula phrase even for the negative answer, i.e. *ní hea* ‘it is not’, while ‘no’ is commonly loaned from English *no*.

Irish

- (1) a. *An dtéanntú go dtí an scoil inniu?*
 Q go to to the school today
 ‘Do you go to the school today?’
- b. *Téim*
 go.PRS.1SG
 ‘Yes.’ (lit. ‘I go.’)
- c. *Ní théim*
 NEG go.PRS.1SG
 ‘No.’ (lit. ‘I don’t go.’)

Table 3. Y-N words under development in Celtic languages
(Toyota 2009: 489)

| | | YES | NO |
|-----------|-----------------|-----------------------------------|-----------------------------|
| Goidelic | Irish | <i>is ea; séo; yeah</i> | <i>ní hea</i> |
| | Manx | <i>abbyr eh; gyn ourys; seadh</i> | N/A |
| | Scottish Gaelic | <i>gu deabh; aigh</i> | N/A |
| Brittonic | Breton | <i>ya; eo</i> | <i>nann; ket</i> |
| | Cornish | <i>usi; eah; usy; ya</i> | <i>nag eus; na; nag usy</i> |
| | Welsh | <i>ie; oes</i> | <i>na</i> |

The choice of the copula phrase is perhaps due to its frequency. Among the IE languages, the copula as a source for the Y-word is attested, but rather rare. Due to the grammatically unique features, the Celtic languages are dependent on the copula phrases to refer to what is normally expressed by independent lexical verbs in other IE languages, e.g. (2). Thus, the lack of lexical verbs such as *like* (e.g. (2a)) or *have* (e.g. (2b)) is covered by the copula phrases. In addition, various tense and aspectual structures are covered by the verbal conjugation, but additional peripheral constructions are formed with the copula in Irish, e.g. (3). Thus, utility of the copula concerning the tense-aspect is found among the Celtic languages, as also observed in other IE languages, but it also covers the lack of basic lexicons.

Irish

- (2) a. *Tá a fhios agam a hainm*
 COP its knowledge at.me her name
 'I know her name.' (lit. 'I have its knowledge her name.')
- b. *Tá leabhar agam.*
 COP book at.me
 'I have a book.' (lit. 'a book is at me.')

Irish

- (3) a. *Tá mé ag staidéar Gaeilge*
 COP I at study.VN Irish
 'I am studying Irish.' (lit. 'I am at studying.')
- b. *Tá mé tar éis staidéar Gaeilge*
 COP I after study.VN Irish
 'I have studied Irish.' (lit. 'I am after studying.')

Due to the intense contacts with English, or French in the case of Breton, the loan of English *yes* and *no* is understandable. Field (2003: 38), for instance, proposes a borrowability hierarchy as shown in (4), stating that the content items are the most likely candidate to be borrowed, and the second likely candidate is function words. Although the Y-N words are not often included in the list of loan words, the Y-word is sometimes borrowed even outside of the Celtic languages, e.g. Czech, Slovak, Slovenian and Romanian as already discussed earlier.

- (4) content item > function word > agglutinating affix > fusional affix

Without the borrowing, the Celtic languages have started developing their own Y-words based on the copula. This is a recent development, and judging from the speed of its evolution, it could be a case of replication (cf. Heine and Kuteva 2005), i.e. using the loan word from English as a stimulus, it is possible to argue that the newly-developed phrases in Table 3 are replicated.

UNIFORMITARIANISM AND Y-N WORDS IN IE LANGUAGES

What is observable in the Celtic language now can shed light on the past development of the Y-N words in the IE languages. The choice of the copula in the Celtic languages is due to its frequency, and judging from this pattern, the Y-word is derived from a frequently used word or phrase. The use of the demonstrative must have been, thus, frequently used among the Romance languages. Likewise, judging

from the source, subordination and conjunctions were relatively frequent in earlier Slavic and Germanic languages. These changes occurred at their own pace, and it took a couple of millennia to see the change, i.e. the transition was gradual. Due to the lack of colloquial data, restriction to a certain register cannot be tested. However, according to the etymons, the subordination and conjugations must have been frequently used even in the colloquial register a thousand years ago or so.

Contrary to the gradual change, some modifications occurred according to catastrophism. Loan words are one such case, but a note has to be made on *da* in East and South Slavic languages. These languages used to use *tak*, an option still used in Polish and Belarussian, *da* was adopted along with the Cyrillic letters and Orthodox belief. Thus, this change was not expected according to a normal course of changes, but the religion-related factors forced abrupt innovations.

REFERENCES

- Ager, D. V. (1993). *The nature of the stratigraphical record* (3rd ed.). Chichester: John Wiley & Sons.
- Burridge, K. (2004). Changes within Pennsylvania German grammar as enactments of Anabaptist world view. In: *Ethnosyntax: Explorations in grammar and culture* (N. Enfield, ed.), Oxford: Oxford University Press, 207–230.
- Christy, T. C. (1983). *Uniformitarianism in linguistics*. Amsterdam: John Benjamins.
- Field, F. W. (2002). *Linguistic borrowing in bilingual contexts*. Amsterdam: Benjamins.
- Gamkrelidze, T. V. & V. V. Ivanov (1995). *Indo-European and Indo-Europeans: A reconstruction and historical analysis of a proto-language and proto-culture* (part 1): text. (English version by Johanna Nichols). Berlin: Mouton de Gruyter.
- Gordon, B. L. (2013). In defence of uniformitarianism. *Perspectives on Science and Christian Faith* 65: 79–86.

- Heine, B. & T. Kuteva (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Heine, B. & T. Kuteva (2007). *The genesis of grammar*. Oxford: Oxford University Press.
- Lyell, Charles 1830–33. *Principles of Geology*. London: John Murray (reprinted by Penguin, 1998).
- Massey, K. A. (2008). A Latin etymology for Romanian *da* = yes. *Ianua: Revista Philologica Romanica* 8: 93–100.
- Smith, G. A. & A. Pun (2006). *How Does Earth Work: Physical geology and the process of science*. New Jersey: Pearson.
- Toyota, J. (2009). On the evolutionary history of ‘yes’ and ‘no’. In: *Studies in language and cognition* (J. Zlatev et al., eds.), Newcastle upon Tyne: Cambridge Scholar Publishing, 485–498.
- Walkden, G. (2019). The many faces of uniformitarianism in linguistics. *Glossa* 4: 1–17.

Junichi Toyota

PRINCIPE UNIFORMITARISTE ET « OUI » ET « NON » DANS LES LANGUES INDO-EUROPÉENNES

Résumé

Les mots « oui »-« non » (O-N) n'existent pas toujours, et les langues celtiques, par exemple, en manquent. Dans d'autres langues indo-européennes, le mot N provient étymologiquement d'un marqueur négatif, mais il y a deux sources étymologiques principales pour le mot O, la copule et un complémentateur ou une clause subordonnée. Géographiquement, le type copule se trouve dans le nord de l'Europe, tandis que le type complémentateur, dans le sud et l'est de l'Europe à l'exception du grec et des dialectes du slovène. Le modèle de distribution des mots O-N est relativement facile à formuler, mais cet article l'examine en termes d'uniformitarisme. En raison de la nature colloquiale des mots O-N, il est difficile de les étudier historiquement; cependant, l'analyse de ce qui se passe maintenant dans les langues celtiques nous permet de déterminer ce qui a dû se passer dans d'autres branches des langues indo-européennes où les archives historiques. Les langues celtiques développent leurs propres mots O-N, ainsi qu'un mot emprunté à l'anglais, représentant à la fois des changements graduels trouvés dans l'uniformitarisme et

des changements brusques souvent observables dans le catastrophisme. De plus, certains facteurs sociaux tels que les changements de religion pourraient avoir affecté le développement, suggérant ainsi un autre exemple de catastrophisme. Par conséquent, en examinant les mots O-N, il est possible qu'une nouvelle perspective du changement historique puisse être acquise.

Mots clés: uniformitarisme, catastrophisme, oui, non, contact linguistique, copule, complémentateur

Ana M. Jovanović*

University of Belgrade – Faculty of Philology

孝 /XIÀO/ — THE CULTURAL LINGUISTICS PERSPECTIVE

The subject of this paper is the concept of 孝 /xiào/ (“filial piety” [CED, 1373]), which was, as many state (see for example Chan, 2004), one of the most important concepts that has governed interpersonal and family relations in China during its several millennia long history.

This paper will give a brief introduction of existing research on its origin, philosophical interpretation, and cultural significance, and try to answer the following questions: how productive is the word 孝 /xiào/? What culturally significant categories pertaining to concrete real-life behaviour can be extracted just considering the meaning and usage of these lexical items? What lexemes can serve as antonyms of 孝 /xiào/ in certain contexts and what are its closely related concepts?

Key words: 孝 /xiào/, filial piety, cultural concept, productiveness, categories, antonyms, similar concepts

O. INTRODUCTION

The concept 孝 /xiào/ has been widely researched and the literature examining it from different perspectives abounds both in China and abroad. However, as Xu and Fang (2021: 66) state: “There remains a gap in the research of this notion in linguistic and cultural studies”¹.

* jovanna762004@yahoo.com; ana.m.jovanovic@fil.bg.ac.rs

¹ Xu and Fang’s purpose is, as they say, to use “the analytical tools of ‘cultural conceptualisations’ (...) to deconstruct the notion of *xiao* in terms of its re-schematisation process across time and space from ancient China to contemporary

This paper is just another humble attempt to investigate it from the linguistic and Cultural Linguistics point of view and is primarily driven by my interest in this notion.

As already said, this paper will investigate culturally very strongly coloured concept 孝 /xiào/, that is by itself a so-called ‘cultural category’. Although I shall not give an overview of the Cultural Linguistics research on this topic², the definition of cultural category that I adopted is given by other authors and goes as follows: “Cultural categories are those culturally constructed conceptual categories that are primarily reflected in the lexicon of human languages.’ (Sharifian, 2017b, pp. 3–4)” (Xu, Fang 2021: 72). For this reason, I decided to use dictionaries in my research, as they represent the lexical depository of a language.

Lexical corpus used in the present research consists of two monolingual and two Chinese–English dictionaries. The reasons why I have chosen those dictionaries and the total amount of lexemes listed, will be given for each of them separately:

(1.) Zhongguo shehui kexueyuan yuyan yanjiusuo cidian bianjishi bian [中国社会科学院语言研究所词典编辑室编]. (2016/2018). *Xiandai Hanyu cidian* (di 7 ban) [现代汉语词典(第7版)/Dictionary of Modern Chinese (7th edition)]. Beijing: Shangwu yinshuguan [北京: 商务印书馆] (in further text referred to as DMC7). Starting from its earliest editions in 1960 and 1965, and its first formal edition in 1978 (DMC7 2016/2018: 3) all the way to the 7th edition which was used in this research, its compilation and all subsequent revisions have been guided by the team of the most notable Chinese linguists working under the Chinese Academy of Social Sciences (CASS), making it one of the most, if not the most important monolingual dictionary of Standard Chinese language. Although the “凡例/Guide to the dictionary usage” (DMC7 2016/2018: 2–5) do not mention the exact

Chinese diasporas” (Xu, Fang 2021: 66, 67) among others. Cultural and linguistic research on filial piety has also been conducted, for example, by Yuan (2014).

² For further reading on Cultural Linguistics see, for example, the works of Farzad Sharifian and the volumes he edited.

number of lexemes listed in it, given the explanation that only minor changes had been made to the previous 6th edition printed in 2012 (only 400 new words added, some old ones deleted and the like) (DMC7, 第7版说明/ Introduction to the 7th edition 2016/2018: 5), we can conclude that the number of listed lexemes is more or less similar to that of the 6th edition, and amounts to more than 69,000 (DMC7, 2012年第6版说明/ 2012 Introduction to the 6th edition 2016/2018: 12).

(2.) Wang, T. *et al.* [王涛等编著/ed.]. (1987/2006). *Zhongguo chengyu da cidian*. [中国成语大辞典/Dictionary of Chinese Idioms]. Shanghai: Shanghai cishu chubanshe [上海: 上海辞书出版社] (in further text referred to as DCI). The number of idioms listed in it is more than 18,000 (DCI, 序言/Preface; 凡例/ Guide to the dictionary usage 2006:1). This is just but one of the many specialized dictionaries of idioms that exist nowadays. However, I chose this one as part of my corpus because the famous Chinese linguist Wang Li, in its Preface, gave it a positive review.

(3.) Beijing waiguoyu daxue yingyu xi cidianzu bian [北京外国语学院英语系词典组编]. (1997/2006). *Han-ying cidian (xiuding ban suoyin ben)* [汉英词典(修订版缩印本)/A Chinese-English Dictionary (Revised edition)]. Beijing: Foreign Language Teaching and Research Press [北京: 外语教学与研究出版社] (in further text referred to as CED). The number of listed lexemes is approximately 80,000 (CED, 前言/Preface: no page). This dictionary has been reprinted seventeen times, which is, in my opinion, proof enough of its wide acceptance and use, and the reason why I chose to enlist it in my research corpus.

(4.) Chen, X. [陈欣望] (ed./编著). (1991). *Hanying yulin* [汉英语林: 成语、典故、谚语、褒贬语、隐语、委婉语、古语、俗语、警句、诗辞熟句/A Dictionary of Chinese idioms and phrases, proverbs and allusions, eulogistic and derogatory terms, enigmas and euphemisms, famous and popular sayings, sparkling sentences and well-known lines in ancient poems, lyrics and literary compositions with English translation]. Shanghai: Shanghai Jiao Tong University Press [上海: 上海交通大学出版社] (in further text referred to as CEYL). In my opinion, the number of 34,643 entries (CEYL 1991: 1) and a wide range of language material gathered from various sources,

spanning from the philosophical and historical texts, famous poems, traditional and modern novels and even contemporary newspapers and magazines (CEYL 1991: 1, 2) makes it a unique reference book that offers much broader view of the language used by the Chinese of different educational level and fields of interest. Therefore, I chose to incorporate it in the language corpus.

In order to find out how many lexemes in the chosen corpus are actually formed by the word 孝/xiào/, I went through the total amount of at least 201,640 dictionary entries³. The results of my study will be presented below.

Including the Introduction and the Conclusion, this paper consists of five parts: in the first part titled “The origin and philosophical interpretations of 孝 /xiào/”, I give a short overview of the meaning, cultural significance, and importance of this concept in China as it was explained and thoroughly studied by many scholars both in China and abroad. As I do not intend to go deeper into all the possible interpretations, its ethical, sociological, and historical meanings, and importance of 孝 /xiào/, the paper will present only a small portion of works considered relevant to the present study. The second part of the paper titled “Corpus analysis findings” presents the results of my research. The third part “Antonymic and the concepts closely related to 孝 /xiào/” brings forth the question of the antonyms of 孝 /xiào/ as well as the relationship between 孝 /xiào/ and other culturally significant concepts in China, such as 忠 /zhōng/, 悌 /tì/ and others. In that part I shall also present some preliminary ideas waiting to be thoroughly studied in the future.

³ Worth noting is that except for the DMC7, for which I was able to obtain a digital version, which significantly eased the search and increased its accuracy, all other dictionaries were checked manually, page by page, entry by entry, which was not only time consuming, but also prone to possibly higher error degree. Therefore, it must be noted that the real number of lexical items might be slightly different than that presented in this study, although not substantially so that can significantly change the results regarding the productivity of this word.

1. THE ORIGIN AND PHILOSOPHICAL INTERPRETATIONS OF 孝 /XIÀO/

The concept 孝 /xiào/ is culturally constructed concept (Xu, Fang 2021: 65–66) which appears in the classical text “*Xiao Jing*, translated as the *Classic of Filial Piety*” from “around 4th century BC”, and is composed of “18 short chapters documenting what Confucius said about the notion of *xiao*”, in the conversation with one of his disciples (Xu, Fang 2021: 66).

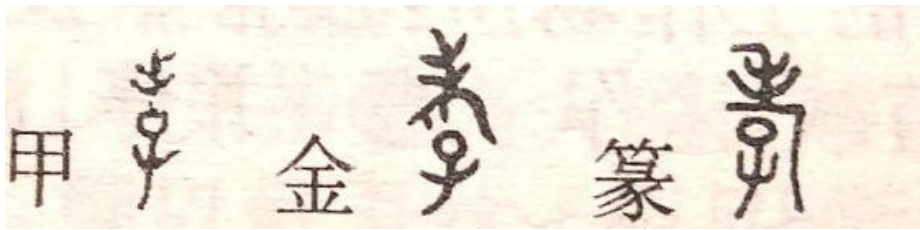
The character 孝 /xiào/, however, is much older, and is found even “in the oracle bones inscriptions (甲骨文) from the Shang dynasty” (Zhang 2020: 80) that ruled China from appr. 17th to 11th century BC (ACHC 1986: 6)⁴. Researchers generally agree about this (see for example Wang 2020). Detailed account on the socio-economic conditions that enabled the rise and development of this concept, and its numerous entailments is given by Tu (2010: 157) who states that its rise was closely related to the creation of the family, i.e., “a patriarchal clan society”⁵ in which everyone’s lineage was known (as opposed to the situation in the previous matriarchal society). Its ascension towards the sphere of moral virtue is linked to the establishment of the feudal society which happened during the Western Zhou dynasty (approx. 11th century – 771 B.C.) (ACHC 1986: 10), for which nowadays evidence in the form of inscriptions in bronze is abundant (Tu 2010: 157). Although already used as a tool for regulating the society in the period of Western Zhou, 孝 /xiào/ formally entered the sphere of politics during the Han dynasty (206 B.C – 220 A.D.) (ACHC 1986: 22, 28), due to the Neo-Confucianist Dong Zhongshu, and the fact that through the works of Confucius, Zengzi, Mencius and other Ruists, it was already so developed that could be used by the rulers of the Han

⁴ The time span for each dynasty or historical period mentioned in this paper is given according to Jian, B., X. Shao and H. Hu. (1986). *A Concise History of China*, Beijing: Foreign Languages Press, in further text referred to as ACHC. This source will be stated in the body of the text, immediately after the information that comes from it.

⁵ Throughout the text, translation of citations from the papers originally written in Chinese is given by the author of this paper.

“to pacify and rule over all under the Heavens” (Tu 2010: 158). In her research of the *Analects* and *Mencius*, as two of the most important Ruist classics, Chan (2004: 180) listed eleven specific “duties” that a filial son must perform including, but not limited to caring, esteem, obedience, proper treatment, performing sacrificial offerings upon parents’ death and having children so that they can continue the practice of ritual sacrifice. All of those must in their essence have the feeling of “affection for one’s parents” (Ibid).

As for the development of the character 孝 /xiào/, in his *Dictionary of the Origin and the Development of Chinese Characters*, Gu Yankui (2003: 246) gives the following pictures depicting the change of character’s form from the earliest oracle bones script, via inscriptions on bronze to the small seal script:



Oracle bones script Inscriptions on bronze Small seal script

According to Zhang (2020: 80), in its earliest form, it was composed of “the upper part – a long streak of hair representing an old man, and the lower part – his son who serves as a support, thus representing the respect and care” one shows towards his or her elders. Xu and Fang (2021: 66) state that looking from the morpho-semantic point of view “xiao (孝) is originally about sons supporting their aging parents”, but “it is also about humbleness, courtesy, righteousness, morality, law-abiding, loyalty, prosperity, nobleness, and the worship of the Lord or the Son of Heaven [... and] ultimately about achieving a harmonious relationship between the Heaven and the Earth” (ibid, 83).

So, what exactly is the social and ethical meaning of 孝 /xiào/? Zhang (2020: 80) states that “at the very beginning it was just a

consciousness one had about his/her need to respect the elders and the ancestors”, but that it “grew into a specific culture with the rise of Ruism”. According to Wang (2020: 28), as a concept that “embodies the Ruist culture”, it is not only related to the “respect and taking care for one’s living parents” and ancestors, which no doubt is its basic connotation reflected also in the fact that one must visit ancestors’ graves in order to pay them respect during “the four major traditional holidays”, i.e. “the Lunar New Year’s Eve, the Tomb Sweeping Day, the Double Ninth Festival and the Festival of the Dead Spits”, but is also extended towards all the elders in the society and the society as a whole, to which one must be loyal and devout. As Xu and Fang (2021: 66), among numerous other researchers throughout the history highlight, it is “one of the top virtues” in China, same as “仁 (*ren*, benevolence), 义 (*yi*, righteousness), 礼 (*li*, propriety), 智 (*zhi*, wisdom), 信 (*xin*, faith) and 忠 (*zhong*, loyalty)”, which is, as the 孝经 (*Xiaojing*, *The Classic of Filial Piety*) puts it “considered to be moral norms prescribed by the Heaven” (ibid, 67). It can also be understood (see Wang’s argument 2020: 28 about the form of character 孝 /xiào/ in the small seal script) as a “respect and gratefulness of offspring towards their parents for the loving care they were given” (Wang 2020: 28) and is often “re-schematised” in modern societies as “reciprocal support and mutual assistance financially and emotionally across generations” (Xu and Fang 2021: 83).

From the very beginning, the concept 孝 /xiào/ was more than just a virtue everyone should strive to achieve. Very early on, it has been properly “standardized” and even institutionalized firstly through the “norms of propriety 礼 [lǐ/]” and then, with the development of the legal society during the Qin dynasty, through the legal system (Zhang 2020: 80). Even though it has to be admitted that it was overwhelmingly used by the aristocracy and the emperor in order to “subdue their people”, the fact that “not being filial” was throughout the ages considered to be if not the biggest of all crimes, then one of the major ones, and that it even nowadays still is in some form extant in the current laws of the country (Zhang 2020: 81), is proof

enough of its significance in shaping the Chinese civilization and culture as a whole. As it is already mentioned above, Han dynasty was the first one to “use 孝 /xiào/ to pacify and rule over all under the Heavens”, and the importance of this concept in the institutional and political sense is clear enough from the fact that during that same Han dynasty it became the inseparable part of the exams for recruiting civil servants (Tu 2010: 158). All the later dynasties up to the last one, the Qing dynasty, “paid special attention to the transmission of *xiao*” (Zhang 2020: 80). After the fall of dynastic feudal system in 1911, according to Tu (2010: 159), famous reformers from the May 4th Movement, such as “Chen Duxiu, Hu Shi, Lu Xun and others, strongly criticized the feudal ethical code of 孝 /xiào/”, which was therefore forced to reform and accommodate to the newly established social circumstances.

As for the modern understanding of the notion 孝 /xiào/, Xu and Fang (2021: 82) state that some of its multiple meanings that developed during the course of time, such as, for example “worshiping the Lord or the Son of Heaven”, disappeared, primarily “due to the Cultural Revolution”, as those meanings were considered to be “acts or thoughts of feudal superstition”, and only the meaning related “with parent – child relationship within the family domain” has survived until today. As mentioned above, even in some state laws and regulations nowadays, there can be found articles regarding “the obligation of sons and daughters to take care of one’s parents” (Zhang 2020: 81). What is also very interesting is to see in what degree this culturally important concept survives when confronted with the Western cultural and behavioural matrices in second generation of Chinese diaspora in the USA, Australia, and other countries. If we are to judge according to some reports (see for example Yang 2011: no page), this concept has a tendency of disappearing from the minds of younger generation Chinese living in the societies predominantly governed by the Western culture and codes of behaviour, due to their exposure to the Western understanding that children must “respect” their parents, but not necessarily have “the moral obligations” towards

them (Yang 2011: n.p.).⁶ Even in China, as Poškaitė (2014: 111) states, there is “the confusion or disagreement over the understanding of the duty of filial piety between older and younger generations.” She says that, according to the sociological studies conducted by Wu Fei (2011), the young believe “material support” is expression enough of their “filiality”, while the parents are often not satisfied with just that (ibid. 112).

Given the plentiful articles on philosophical, ethical, behavioural, sociological, and other meanings of 孝 /xiào/, and due to the limitations of space, I will not go further on presenting and reviewing the existing literature, with hope that what has been said above will be enough for the reader to understand my motif to investigate this concept from the Cultural Linguistics’ point of view.

2. CORPUS ANALYSIS FINDINGS

As it has already been said in the Introduction, for the purpose of this research I checked more than 201,640 dictionary entries listed in the abovementioned four dictionaries and found out that the number of entries (compounds, idioms, proverbs etc.) formed from the word⁷ 孝 /xiào/ in them is only 64 (entry that appears in several dictionaries is counted only once). In other words, it makes up only 0.0317% of the entries listed in those dictionaries⁸. If we look at each of them

⁶ For the generational gap and differences in understanding this concept in Australia, see excellent study conducted by Xu and Fang (2021).

⁷ Many monosyllabic words in Modern Chinese language are at the same time root morphemes too. For the purpose of this paper, it is not relevant whether we consider 孝 /xiào/ to be the root morpheme, monosyllabic word, or we talk about it as the “character”, which is the reason why the explanation of the difference between them will not be given. In this paper the author will be mostly using the term “word” when referring to 孝 /xiào/.

⁸ It is interesting to note that Yuan (2014) did a similar research, limited though to one dictionary only, namely *The Big Dictionary of Chinese language* (汉语大词典). She found out that there are 150 items formed by the word 孝 /xiào/, which is

separately, the situation is as follows: DMC7 with more than 69,000 entries (DMC7, 2012年第6版说明/2012 Introduction to the 6th edition 2016/2018: 12) lists only 21 lexemes (approximately 0,03% of all) formed by the word 孝 /xiào/. DCI with more than 18,000 entries (DCI, 序言/Preface; 凡例 /Guide to the dictionary usage 2006: 1), lists only 12 lexemes that have the word 孝 /xiào/ as their component, which makes no more than 0,067% of the total number of entries. In the CED, with approximately 80,000 (CED, 前言 /Preface: no page) dictionary entries, only 27 (that is approximately 0,034%) contain the word 孝 /xiào/, while in the CEYL that number is 36 out of 34,643 (CEYL 1991: 1) entries it stores (which makes no more than 0,10 % of the collected lexical material). Although still a very small number, the last one of four gives proportionally the largest number of compounds, idioms, sayings etc., which is probably related to its wide range of language material and different types of collected linguistic data.

Contrary to what is expected from such a culturally important concept, 孝 /xiào/ has a very low productivity rate, which Liu (2009: 97) relates to it not being a basic–vocabulary word. Let us compare it to the body part word 心 /xīn/ (“1. the heart; 2. heart; mind; feeling; intention;” (...) [CED, 1381]) as an example of a very productive word that forms 399 lexemes in the 5th edition of DMC and the CED (Jovanović 2014: 4). Such a great number is no doubt a result of it belonging to the basic vocabulary and its usage in everyday life not only to denote abstract, but also concrete concepts. That is all quite different from the concept 孝 /xiào/, which is limited to the traditional, feudal ethics.

more than double the amount of those found in the four dictionaries I examined. In my opinion, the reason for such a big number is because, as she explained, she counted “every separate meaning of polysemic word as a separate lexeme” (Yuan 2014: 15). I do not agree with her counting methodology, since separate senses of polysemic lexemes cannot be considered separate words. Therefore, I think that the number she states in her paper must not be taken at face value.

What does 孝 /xiào/ mean? Let's look at the following explanation:

1. *filial*; 2. *filial piety*; *filial obedience*; 3. *the conventional mourning rites for a deceased elder member of one's family*; 4. *mourning apparel* [CED, 1373].⁹

In all of the 64 lexical items found in the corpus, character 孝 /xiào/ carries one (or more if the lexeme is polysemous) of the above cited meanings. In order to structure this not so big set of lexemes, I've tried to offer a possible classification¹⁰ of the set of words formed by 孝 /xiào/, with a purpose of highlighting the whole scope of its usage

⁹ Meanings of the lexical items are cited from the dictionaries that served as a corpus for my research. The source will be stated in the brackets, after the meaning, using the shortened name of the dictionary and the page on which it can be found (for example: [CED, 1373]). To make it clearer to the non-speaking Chinese reader, longer items, like idioms and proverbs, will firstly be given word by word literal translation of each component (CED will be used for that purpose), and after that the meaning given in the dictionary. If certain lexical item appears only in monolingual dictionaries, translation into English will be given by the author of this paper, according to the Chinese explanations and English translation of the words used in them.

¹⁰ Another possible classification of lexemes with 孝 /xiào/ is given by Yuan (2014: 15) and goes as follows: First three categories “are based on real-life practices, behaviors, understanding and feelings related to 孝 /xiào/”. The first category consists of nouns that serve to “denominate person who is filial”; second category consists of verbs that name activities and “practices related to filial behavior”; the third category is, again, comprised from nouns that relate to the “disposition, moral character and temperament”. Adjectives make the fourth category that is related to the “ways others evaluate someone's behavior” (ibid.: 15) and are mostly positive in their connotations (ibid.: 16); the fifth category is the most numerous one and consists of lexemes and expressions related to “mourning and funeral ceremonies” (ibid.: 15), in which the purpose of using 孝 /xiào/ is to “avoid the usage of words 死 (/sǐ/, to die, death), 歿 /mò/ (to die)” and the like, which are tabooed in Chinese culture (ibid.: 16). Yuan (ibid.: 15) divides this category into two subgroups, i.e., words expressing abstract, and words expressing more concrete meanings. The sixth category, which is related to “ideology and its restraining effects” (ibid.: 16), “can be further sub-categorized into the names or titles ruling class bestows upon those who are filial (...), abstract nouns pertaining to some things related to 孝 /xiào/ inside the [feudal] system [...], and words whose origin is related to some famous stories regarding filiality” (ibid.: 17).

in Chinese language and culture. Similarly to other researchers (see for example Yuan, 2014), who also tried to classify the set of lexemes formed by this culturally important concept, I hold that classifying lexemes into categories has twofold importance: firstly, it makes it easier for foreign students to form a holistic picture of the usage and importance of this concept, which is rarely achieved solely by learning individual items separately when we come across them; and secondly, as any other classification, it can help us better understand the areas in which they influence the thought and behaviour of the ordinary language users. For this paper, I extracted the following categories¹¹:

I. The status of 孝/xiào/ inside the system of moral values:

- 1) 百善孝为先 *bǎi shàn xiào wéi xiān* [hundred – virtuous – filial piety – be – before, first] “Filial piety is the root of all goodness” [CEYL, 523/524]
- 2) 万恶淫为首，百善孝为先 *wàn è yín wéi shǒu, bǎi shàn xiào wéi xiān* [ten thousand – evil – licentious – be – the head/first, hundred – virtuous – filial piety – be – first], “Of all crimes, lewdness is the worst; of all virtues filial piety is the best” [CEYL, 110]

In the analyzed corpus, only two proverbs are identified as belonging to this category, which place the concept of 孝 /xiào/ high above all other virtues, as we have seen was the case during the long history of the development of this concept that was introduced in the part 1. of this paper.

II. Person’s moral character:

- 3) 不孝 *bùxiào* “be an unfilial son or daughter; act contrary to filial piety” [CED, 101]
- 4) 不孝（肖）子孙 *bù xiào (xiào) zǐ sūn* [not – filial – children and grandchildren] i.e., “unworthy descendants” [CEYL, 236]

¹¹ As is the case with any other classification, the above is just my attempt to classify seemingly disparate words. When naming the categories, I had in mind the easiest and most understandable way of representing the meanings of the whole group of lexical items.

- 5) 不忠不孝 *bù zhōng bù xiào* [not – loyal – not – filial] i.e., “to be neither loyal to one’s country nor filial to one’s parents” [CEYL, 241]
- 6) 全忠全孝 *quán zhōng quán xiào* [completely – loyal – completely – filial] i.e., “to be complete both in fidelity and filial piety” [CEYL, 627]
- 7) 孝道 *xiàodao* “(inf.) be a good son or daughter” [CED, 1373]
- 8) 孝顺 *xiào shun* “show filial obedience” [CED, 1373]
- 9) 忠信孝悌 *zhōng xìn xiào tì* [loyal – confidence, trust, faith – filial piety – love and respect for one’s elder brother] i.e., “loyalty, sincerity, filial piety and fraternal love” [CEYL, 991]
- 10) 孝心 *xiàoxīn* “filial sentiments; filial devotion” [CED, 1373]
- 11) 忠臣孝子 *zhōng chén xiào zǐ* [loyal – minister – a filial son] i.e., “loyal statesmen and filial sons” [CEYL, 990]

This category has 22 items (variants of the same lexical entry excluded), of which only nine are listed as examples. All of the lexical items in this category either serve to characterize whether someone is having or lacking the moral character of 孝 /xiào/ (examples 3, 4, 5, 6, 11), describe someone’s behavior (example 7, 8) or just name the virtue of filiality or other closely related to it (9, 10).

III. The correlative nature of the relationship between the family members:

- 12) 母慈子孝 *mǔ cí zǐ xiào* [mother – kind – son – filial] i.e., “A kind mother brings up children dutiful to their family” [CEYL, 502]
- 13) 父慈子孝 *fù cí zǐ xiào* [father – kind – son – filial] i.e., “A kind father makes a filial son” [CEYL, 354]
- 14) 若要子女孝，自己孝父母 *ruò yào zǐ nǚ xiào, zì jǐ xiào fù mǔ* [if – want to – sons and daughters – filial, oneself – filial – parents] i.e., “The way in which you treat your parents has the greatest formative effect on your children’s behaviour” [CEYL, 972]

- 15) 贤父孝子 *xián fù xiào zǐ* [worthy – father – a filial son] i.e., “the affectionate father and filial sons” [CEYL, 1007]

These four idioms and proverbs state what seems to be the popular wisdom regarding achieving the virtue of 孝 /xiào/, drawn from the experience, and are clearly the echo of understanding that one should live by their own example and not expect their children to have virtues they themselves do not possess.

IV. Instructions derived from experience:

- 16) 棒头上出孝子 *bàng tóu shàng chū xiào zǐ* [stick – on – produce – a filial son] i.e., “Severe whipping was the most effective means of bringing up filial sons” [CEYL, 1583]
- 17) 父严子孝 *fù yán zǐ xiào* [father – strict – son – filial] i.e., “when father sternly disciplines his children, the children should be filial and respect him” [DCI, 391]
- 18) 孝心须费力，忠臣得赔命 *xiào xīn xū fèi lì, zhōng chén dēi péi mìng* [filial – heart – must – exert great effort, loyal – minister – must – stand a loss – life] i.e., „Filial service demands every effort; loyalty to the prince demands life itself” [CEYL, 785]
- 19) 移孝作忠 *yí xiào zuò zhōng* [change – filial piety – regard as – loyal] i.e., “to substitute filial piety with loyalty to the country” [CEYL, 1516]
- 20) 久病床前无孝子 *jiǔ bìng chuáng qián wú xiào zǐ* [long – sickness – bed – front – not have – a filial son] i.e., “in cases of chronic sickness, there are no dutiful children at the bedside” [CED, 642]

Example 16) is clearly based on the experience and has the purpose of giving parents an advice of how to bring up filial offspring. I will not comment the correctness of the method, but only state that what we deal with here is feudal ethics and the whole phenomena should be regarded in the light of that understanding. Similar to it is example 17). Examples 18) and 19) correlate the virtue of 孝 /xiào/, which is more basic, to the higher level, that is the loyalty to the sovereign, as

a natural extension of it. Example 20) brings forth the situation that could be true everywhere, regardless of the cultural matrix.

V. Obligations, customs, and various behavioral models

I divided this category into the following sub-categories:

V/1. Obligations towards ancestors:

- 21) 不孝有三，无后为大 *bù xiào yǒu sān, wú hòu wéi dà* [not – filial – have – three, not have – offspring – be – big] i.e., “There are three forms of unfilial conduct, of which the worst is to have no descendants” [CEYL, 236]

V/2. Mourning apparel and customs related to mourning:

- 22) 穿孝 *chuānxiào* “be in mourning; wear mourning” [CED, 188]
23) 带孝 *dàixiào* “wear mourning for a parent, relative, etc.; be in mourning” [CED, 234]
24) 吊孝 *diàoxiào* “(inf.) visit the bereaved to offer one’s condolences; pay a condolence call” [CED, 275]
25) 满孝 *mǎn//xiào* “to finish the period of mourning for the deceased elder member of the family” [DMC7, 875]
26) 披麻带孝 *pīmá – dàixiào* [wrap around – hemp – wear mourning for a parent, relative, etc.] i.e., “wear the hemp garments of mourning” [CED, 914]
27) 热孝 *rèxiào* “(usu. used in) ~在身 [zài shēn] wear mourning for one’s grandparents, parents, or husband” [CED, 1021]
28) 身穿重孝 *shēn chuān zhòng xiào* [body – wear – heavy – mourning apparel] i.e., “to be in sackcloth” [CEYL, 865]
29) 守孝 *shǒuxiào* “observe a period of mourning for one’s deceased parent” [CED, 1137]
30) 孝服 *xiàofú* “1. mourning apparel 2. a conventional period of mourning (for a deceased elder member of one’s family)” [CED, 1373]
31) 谢孝 *xièxiào* “after the period of mourning for a parent, visit and thank those friends and relatives who have offered condolences” [CED, 1380]

32) 重孝 *zhòngxiào* “in deep mourning (usu. after the death of one’s parent)” [CED, 1643]

V/3. Expected behaviour (in the family and beyond):

33) 尽孝 *jìnxiào* “fulfill one’s duty to one’s parents; display filial piety towards one’s parents” [CED, 624]

34) 居家孝友 *jūjiā xiào yǒu* [reside – home – filial – friendly] i.e., “to be filial and fraternal in one’s domestic life” [CEYL, 976]

35) 为臣死忠, 为子死孝 *wéi chén sǐ zhōng, wéi zǐ sǐ xiào* [be – minister – die – loyal, be – son – die – filial] i.e., “in feudal society, ministers should be loyal to their sovereign up to the point that they should not spare their life, while sons/children should be ready to give up their life in order to be filial to their parents” [CDI, 1323/1324]

36) 孝道 *xiàodào* “filial duty” [CED, 1373]

37) 孝敬 *xiàojìng* “1. show filial respect to (one’s elders) 2. give presents to (one’s elders or superiors) to show one’s respect; pay a tribute of respect to” [CED, 1373]

38) 孝幔 *xiàomàn* “the curtain or screen before a bier” [CED, 1373]

39) 孝行 *xiàoxíng* “filial behaviour” [CED, 1373]

V/4. The way the bereaved persons address themselves (in the old society):

40) 孝女 *xiàonǚ* “1. (old) bereaved daughter (a term used in an obituary or on a tombstone; (...))” [CED, 1373]

41) 孝男 *xiàonán* “(old) bereaved son (a term used in an obituary or on a tombstone)” [CED, 1373]

Example 21) from the sub-category V/1 is a well-known saying of Mencius (Ivanhoe 2004: 191) that has determined the most important task of every human being in the Ruist ethics, i.e., the continuation of one’s family name, and is typical, in that sense for the patriarchal society. All of the examples belonging to the sub-category V/2 are related to mourning rituals, either in the sense of just stating the fact that one is in mourning, describing what kind of clothes one is

wearing (which in itself is very interesting, as we can see that it is considered proper if one is wearing very plain clothing), name the requisites (such as curtain) used by the coffin or visiting those who came to offer their condolences during the mourning period. In that sense, all these lexical examples are very informative, especially to those not belonging to Chinese cultural matrix, as we can learn a lot about specific customs and behavioural codes related to this important life event. In these examples, as is already mentioned earlier in this paper, 孝 /xiào/ is used as a sort of “substitute” to culturally tabooed words (see Yuan 2014:16). Similarly, lexical examples belonging to the sub-category V/3 prescribe proper behaviour inside the family and beyond, while items in V/4 are old expressions used by grieving children in obituaries or tombstones, as it is already clear from the dictionary explanations.

3. ANTONYMIC AND THE CONCEPTS CLOSELY RELATED TO 孝 /XIÀO/

The obvious antonymic concept to 孝 /xiào/ is its direct negation, that is example no. 3) which will be, for the sake of convenience, repeated here:

- 3)¹² 不孝 *bùxiào* “be an unfilial son or daughter; act contrary to filial piety” [CED, 101]

However, corpus analysis shows that there are also two other words, namely 逆 /nì/ (“1. contrary; counter; inverse; converse; 2. go against; disobey; defy; 3. traitorous; rebellious; (...)” [CED, 876-877]) and 孽 /niè/ (“evil; sin” [CED, 883]; “1. evil, evil person; 2. sin; 3. (lit.) not loyal or not filial” [DMC7, 957]), which are used as its antonyms in certain fixed phrases and expressions. Although DMC7 as the 3rd meaning of 孽 /niè/ gives the meaning that can be considered directly antonymic to 孝 /xiào/, no such meaning is given for the word 逆 /nì/, and yet, we find both in the following lexical entries:

¹² All lexical items that appear earlier in the text are repeated by their original numbers.

- 42) 逆子 *nìzǐ* “unfilial son” [CED, 877]
43) 孽子 *nièzǐ* “1. (old.) son of a concubine; 2. an unfilial son; an unworthy descendant” [CED, 883]
44) 臣不反君，子不逆父 *chén bù fǎn jūn, zǐ bù nì fù* [minister – not – counter – monarch, son – not – counter – father], i.e. “A minister can’t go against his king, a son can’t go against his father” [CEYL, 535]
45) 奸臣逆子 *jiān chén nì zǐ* [treacherous – minister – traitorous – son] i.e., “disloyal ministers and unfilial sons” [CEYL, 694]
46) 逆子馋臣 *nì zǐ chán chén* [traitorous – son – greedy – minister] i.e., “unfilial sons and disloyal and traitorous statesmen” [CEYL, 1291/1292]
47) 叛臣逆子 *pàn chén nì zǐ* [rebel against – minister – traitorous – son] i.e., “traitorous officials and unfilial sons” [CEYL, 1289]
48) 忤逆不孝 *wǔ nì bù xiào* [disobedient – traitorous – not – filial] i.e., “obstinate and undutiful” [CEYL, 910]; “not obedient to parents and not filial” [DCI, 1376]

It seems that words 逆 /nì/ and 孽 /niè/ when paired with other morphemes are much stronger in meaning than plain 不孝 /bù xiào/ and are used to highlight how bad, or even evil it is if someone lacks the quality of being filial.

Another thing to be discussed in this part is that the findings of the corpus analysis I conducted prove once again what researchers (see for e.g., Ivanhoe 2004: 196, Zhang 2020: 80; Wang 2020: 29) have claimed about the connections between the concept of 孝 /xiào/ and other related and culturally important concepts, especially that of 忠 /zhōng/ (“loyal, devoted, honest” [CED, 1638]). Lexical items from our corpus in which they co-occur are the following:

- 49) 忠孝 *zhōngxiào* “loyalty and filial piety” [CED, 1638]
5) 不忠不孝 *bù zhōng bù xiào* [not – loyal – not – filial] i.e., “to be neither loyal to one’s country nor filial to one’s parents” [CEYL, 241]

- 6) 全忠全孝 *quán zhōng quán xiào* [completely – loyal – completely – filial] i.e., “to be complete both in fidelity and filial piety” [CEYL, 627]
- 35) 为臣死忠，为子死孝 *wéi chén sǐ zhōng, wéi zǐ sǐ xiào* [be – minister – die – loyal, be – son – die – filial] i.e., “in feudal society, ministers should be loyal to their sovereign up to the point that they should not spare their life, while sons/children should be ready to give up their life in order to be filial to their parents” [CDI, 1323/1324]
- 18) 孝心须费力，忠臣得赔命 *xiào xīn xū fèi lì, zhōng chén děi péi mìng* [filial – heart – must – exert great effort, loyal – minister – must – stand a loss – life] i.e., “Filial service demands every effort; loyalty to the prince demands life itself” [CEYL, 785]
- 19) 移孝作忠 *yí xiào zuò zhōng* [change – filial piety – regards as – loyal] i.e., “to substitute filial piety with loyalty to the country” [CEYL, 1516]
- 11) 忠臣孝子 *zhōng chén xiào zǐ* [loyal – minister – a filial son] i.e., “loyal statemen and filial sons” [CEYL, 990]
- 50) 忠君孝亲 *zhōng jūn xiào qīn* [loyal – monarch – filial – parent] i.e., “loyalty to one’s ruler and filiality to one’s parents – the feudal concept in old China” [CEYL, 990]
- 51) 忠孝两全 *zhōng xiào liǎng quán* [loyal – filial – both – complete] i.e., “have both, the loyalty to one’s country and the filiality to one’s parents” [CDI, 1758]
- 52) 忠孝难两全 *zhōng xiào nán liǎng quán* [loyal – filial – hard – both – complete] i.e., “One can hardly be complete both in fidelity and filial piety” [CEYL, 990]

In all the above examples, the concepts 忠 /zhōng/ and 孝 /xiào/ are closely related, with the former being conceived as the expected continuation or expansion of the latter beyond one’s family to the country and the sovereign as its representative. It is worth mentioning that for an individual, 忠 /zhōng/ meant that it was more important “to obey the feudal ruler and not the country and people” (Zhang 2020: 81). Ivanhoe (2004: 196) states that “Confucian thinkers regularly

assert that filial piety is the proper paradigm for the subject – ruler relationship as well as the child – parent relationship”. As for the establishment of the connection between the two concepts, according to Zhang (2020: 80) it “is even more [strongly] established by the Ruist classic *Xiaojing*”. Tu (2010: 158) explains that this connection was even more deepened after the Neo-Confucianist Zhang Cai wrote about them, i.e., “[on that basis] established the unity between them”. Indeed, over the long history of the Chinese feudal society 孝 /xiào/ became a powerful weapon for reigning over the people, and the loyalty towards the ruling class and the emperor was established through the connection with the concept of 忠 /zhōng/ (Zhang 2020: 84). Possible reason for the connection, it is believed, is the instalment of 孝 /xiào/ as one of the most important subjects in the exams for civil service (Tu 2010: 158). Although it has lost its original meaning and power, this idea of the close connection between the “filial piety” and the “loyalty” exists even nowadays in some fixed expressions of historical origin, as can be seen from the examples above.

Apart from 忠 /zhōng/, 孝 /xiào/ is combined, albeit in much lesser degree, with the concept of 悌 /tì/ (“love and respect one’s elder brother.” [CED, 1216]). The fact that 孝 /xiào/ is closely related to 悌 /tì/ is proven by the fact that “in the *Analects* [...]” both of them “have been described as the root of the virtue of *ren*” (Chan 2004: 176). And 仁 /rén/, (“1. benevolence; kindheartedness; humanity(...)” [CED, 1028]) is by far one of the most important concepts in Confucian (or Ruist) ethics. Love towards brothers and mutual respect is, as Tu (2010: 157, 158) states established during the Spring and Autumn and Warring States periods¹³ and was “the basic obligation between blood relations in the family”(ibid.: 158). Considering abovementioned Chan’s (2004) comment, it should not surprise us that no combination of 孝 /xiào/ and 仁 /rén/ is found in our corpus, although we might have expected it, and even though, as Tu (2010: 158) states, “the relationship between 仁 [/rén/] and 孝 [/xiào/] was thoroughly explained anew by Er Cheng”,

¹³ Spring and Autumn period: (772 BC – 481 BC); Warring States period: 403 BC –221 BC (ACHC 1986: 14).

who was the Neo-Confucian that lived in the Northern Song dynasty (960–1126) (ACHC 1986: 55). The lexical items with both 孝/xiào/ and 悌/tì/ are the following:

- 53) 孝悌 *xiàotì* “show filial piety to one’s parents and love and respect to one’s elder brothers” [CED, 1373]
- 54) 孝悌恭亲 *xiào tì gōng qīn* [filial – love and respect one’s elder brother – respectful – parent] i.e., “to be filial toward one’s parents and fraternal towards brother” [CEYL, 786]
- 9) 忠信孝悌 *zhōng xìn xiào tì* [loyal – confidence, trust, faith – filial piety – love and respect for one’s elder brother] i.e., “loyalty, sincerity, filial piety and fraternal love” [CEYL, 991]

In example 54) apart from “filiality” and “love and respect for one’s elder brother”, two other concepts appear, namely 恭/gōng/ (“respectful; reverent” [CED, 416]) and 亲/qīn/ (which, among others have the following meanings: “1. parent; 2. blood relation; next of kin; 3. relative; (...) 6. close; intimate; dear; (...)” [CED, 982]; while in Classical Chinese its meanings were, among others: “1. close, intimate; 2. relative, especially parent; (...)” [Wang 2000/2007:1248, 1249]). In example 9) apart from 孝/xiào/, 悌/tì/, and 忠/zhōng/, there is also a concept 信/xìn/ (with the meaning “confidence; trust; faith” [CED, 1389]). *Wang Li Guhanyu Zidian* [Wang Li’s *Dictionary of Classical Chinese*] (2000/2007: 28) gives it the following explanation: “1. to speak the truth, not lie (...) [which is] extended to mean ‘honest, one who does not deceive’ and ‘to believe, to have confidence in, to be trustworthy’, and as an adjective ‘real’. (...)”. Wang (2020: 29) states that the relationship between 孝/xiào/ and 信/xìn/ was explained by Confucius in his *Analects* where “信 [xìn/] appears 38 times”, and was one of the four major subjects that Confucius taught, specifically related to the reliability of spoken and written words (Wang 2020: 29).

Another concept that is usually closely linked to 孝/xiào/ is 慈/cí/ (“1. kind; loving; (...)” [CED, 200], or in Classical Chinese: “1. to love; 2. be filial to parents and wait upon them” [Wang 2000/2007: 324]). According to the analyzed corpus the following notions are also paired with 孝/xiào/: 友/yǒu/ (“1. friend; 2. friendly” [CED, 1522]) and

in Classical Chinese “friend”, which later “extended its meaning to verbal: ‘1. make friends with’; 2. love between brothers”; and “[again] extended its meaning to ‘dear, beloved” [Wang 2000/2007: 99].]; 义 /yì/ (“1. justice; righteousness; 2. righteous; equitable; just; 3. human ties; relationship (...)” [CED, 1485]), for which Wang (2000/2007: 962/963), among others gives the following meanings: “1. what the society considers as appropriate ways and conduct; 2. meaning, sense; 3. outward appearance, etiquette, moral standard; (...)”; 贤 /xián/ (“1. virtuous and able; worthy; 2. a worthy person; an able and virtuous person; (...)” [CED, 1343]; while its meaning in Classical Chinese was: “1. to show ability and morals, virtue; 2. one who shows ability and virtue; (...)” [Wang 2000/2007: 1332]); 节 /jié/ (“(...) 8. moral integrity, chastity (...)” [CED, 608]); 廉洁 /liánjié/ (“honest and clean; incorruptible” [CED, 748]), out of which 廉 lián in Classical Chinese had the meaning “proper, correct, careful, conscientious” [Wang 2000/2007: 278]. Lexical items in which there is a combination of these concepts with the concept of 孝 /xiào/ are the following:

- 12) 母慈子孝 *mǔ cí zǐ xiào* [mother – kind – son – filial] i.e., “A kind mother brings up children dutiful to their family” [CEYL, 502]
- 13) 父慈子孝 *fù cí zǐ xiào* [father – kind – son – filial] i.e., “A kind father makes a filial son” [CEYL, 354]
- 55) 孝子慈孙 *xiào zǐ cí sūn* [a filial son – kind – grandson] i.e., “refer to filial children and grandchildren” [CDI, 1421]
- 34) 居家孝友 *jū jiā xiào yǒu* [reside – home – filial – friendly] i.e., “to be filial and fraternal in one’s domestic life” [CEYL, 976]
- 56) 孝义廉洁 *xiào yì lián jié* [filial – righteous – honest] i.e., “filiality and righteousness, incorruptibility and integrity” [CEYL, 785]
- 57) 孝子贤孙 *xiào zǐ xián sūn* [filial – son – worthy – grandson] i.e., “(to be) sons reverent to parents and grandsons dutiful to ancestors” [CEYL, 785]
- 15) 贤父孝子 *xián fù xiào zǐ* [worthy – father – a filial son] i.e., “the affectionate father and filial sons” [CEYL, 1007]

- 58) 贤妻孝女 *xián qī xiào nǚ* [worthy – wife – filial – daughter] i.e., “faithful wife and filial daughters” [CEYL, 1007]
59) 孝子节妇 *xiào zǐ jié fù* [a filial son – chastity – wife] i.e., “filial sons and virtuous widows (or wives)” [CEYL, 785]
60) 孝义廉洁 *xiào yì lián jié* [filial – righteous – honest] i.e., “filiality and righteousness, incorruptibility and integrity” [CEYL, 785]
61) 忠孝廉节 *zhōng xiào lián jié* [loyal – filial – honest – chastity] i.e., “loyalty, filial piety, integrity and chastity” [CEYL, 990]

Examples 12) and 13) show the relation between the kindness of parents and filiality of children. According to Tu (2010: 157, 158) the obligation of “father being kind and son filial” was established as early as in the Spring and Autumn and the Warring States periods. As can be seen from the above listed examples dutifulness, righteousness, faithfulness, respect, chastity, honesty and the like, are highly regarded virtues expected from an individual, and as certain ethical and cultural norms entered the lexicon of Chinese language in the form of idioms.

In the text above some findings of my corpus analysis were presented. It must be said though that an interesting topic for further research would be the ways 孝 /xiào/ and the lexical items it forms are translated into Serbian language. I believe that it would also be informative to see whether such or similar concepts exist in Serbia and if they do what are the differences and similarities between the two cultures in their conceptualization.

4. CONCLUSION

The results of the study of more than 201,640 dictionary entries listed in four dictionaries of Chinese language (two monolingual and two bilingual) presented in this paper can be summed up as follows: 1. word 孝 /xiào/ as a culturally very important concept has extremely low productivity, as it appears in 64 items, occupying only

0.0317% of the examined lexicon. Normally, such a low productivity is not expected from culturally important concepts, but on the other hand, given that the concept 孝 /xiào/ belongs to the feudal ethics, it is neither surprising nor strange that it does not participate in the formation of new words in the modern Chinese language and society. The reason why just the small number of words containing 孝 /xiào/ is used in Modern Chinese is given by Yuan (2014:17) and is “the result of de-construction of traditional feudal ethics under the pressure of modernization”.

For the sake of convenience to all of those who would like to better understand or learn them, the paper classified the lexemes formed by the word 孝 /xiào/ into different categories that pertain to the position of the concept in the system of moral values, person’s moral character, correlative ties with the conduct of other family members, instructions for behaviour gained from the experience, customs, rituals, obligations, and behavioral models. I also investigated the words containing lexemes 逆 /nì/ and 孽 /niè/, that very often serve as antonyms to those containing 孝 /xiào/, with the purpose of highlighting how “evil” and “sinful” it is not to “be filial”, which the lexeme that contains only the negation of 孝 /xiào/, i.e., 不孝 /bùxiào/ does not strengthen enough. Apart from already known relations with other culturally important concepts, such as 忠 /zhōng/, 悌 /tì/, 信 /xìn/ and 慈 /cí/, this paper lists other concepts that seem important, given that they co-appear with 孝 /xiào/, such as: 友 /yǒu/, 义 /yì/, 贤 /xián/, 节 /jié/ and the bi-syllabic word 廉洁 /liánjié/. Finally, I should stress that the ways word 孝 /xiào/ itself and the lexical items it builds is translated into Serbian could not be, due to the limitation of space, the topic of this paper. That and the question of differences between the ways relationship toward parents (especially in their old age) is conceptualized and consequently lexicalized in Serbian and Chinese language and culture will be the topic of some other research in the future.

REFERENCES

- Chan, A.K.L. (2004). Does *xiao* come before *ren*? In: *Filial Piety in Chinese Thought and History* (A.K.L. Chan and S. Tan, eds.), London and New York: RoutledgeCurzon Taylor & Francis Group, 154–175.
- Chan, S.Y. (2004). Filial piety, commiseration and the virtue of *ren*. In: *Filial Piety in Chinese Thought and History* (A.K.L. Chan and S. Tan, eds.), London and New York: RoutledgeCurzon Taylor & Francis Group, 176–188.
- Gu, Y. [谷衍奎]. (2003). *Hanzi yuanliu zidian* [汉字源流字典 /Dictionary of the Origin and the Development of Chinese Characters]. Beijing [北京]: Huaxia chubanshe [华夏出版社].
- Ivanhoe, P. J. (2004). Filial piety as a virtue. In: *Filial Piety in Chinese Thought and History* (A.K.L. Chan and S. Tan, eds.), London and New York: RoutledgeCurzon Taylor & Francis Group, 189–202.
- Jian, B., X. Shao and H. Hu. (1986). *A Concise History of China*. Beijing: Foreign Languages Press. (in the text referred to as ACHC)
- Jovanović, A. (2014). Similarities and Differences in Conceptualizing “Heart” in Chinese and Serbian and their Practical Use in Language Teaching. In: *Studies on the Teaching of Asian Languages in the 21st Century* (A. Küçükler and H. İçen, eds.), Newcastle upon Tyne: Cambridge Scholar Publishing, 1–22.
- Liu, J. [刘洁]. (2009). Cong Hanyu zici kan xiao wenhua de neihan [从汉字字词看孝文化的内涵 /Connotation of Filial Piety Culture Seen from the Chinese characters]. *Chengdu daxue xuebao (shekeban)* [成都大学学报 (社科版)] 2: 97–99.
- Poškaitė, L. (2014). Filial Piety (*xiao* 孝) in the Contemporary and Global World: A View from the Western and Chinese Perspectives. *Asian Studies II* (XVIII), 1: 99–114.
- Tu, A. [涂爱荣] (2010). Zhongguo xiaodao wenhua de lishi zhuxun [中国孝道文化的历史追寻 /Historical overview of the development of the Chinese culture of filial duty]. *Xueshu luntan* [学术论坛] 9: 156–159. [CNKI, date of last visit: 24th September 2021]
- Wang, F. [王飞明] (2020). Ji yu hanzi gouzao wei zaiti de wenhua chuanbo yanjiu—yi xiao zi wei li [基于汉字构造为载体的文化传播研究——以孝字为例 /The Structure of Chinese Characters as a Medium of Cultural Dissemination—the case of *xiao*]. *Hanzi wenhua* [汉字文化]10 (256): 28–29. [CNKI, date of last visit: 20th September 2021]

- Wang, L. [王力著編/ed.]. (2000/2007). *Wang Li Guhanyu Zidian* [王力古漢語字典/*Wang Li's Dictionary of Classical Chinese*]. Beijing [北京]: Zhonghua shuju [中華書局].
- Xu, Zh. and N. Fang (2021). Re-schematisation of Chinese *Xiao* (Filial Piety) Across Cultures and Generations. In: *Cultural Linguistics and World Englishes*, (M. Sadeghpour and F. Sharifian, eds.), Singapore: Springer Nature Singapore Pte Ltd.: 65–85.
- Yang Z. [杨子岩]. (2011). “Xiao” zi de Zhong Xi liang yi [“孝”字的中西两义 /Chinese and Western Interpretation of the character “xiao”]. *Renmin ribao haiwai ban* [人民日报海外版]. August 10th, 006: no page. [CNKI, date of last visit: 24th September 2021]
- Yuan Y. [袁媛] (2014). “Xiao” cizu de wenhua toushi [“孝”词族的文化透视 /Cultural Perspectives of the Word Family of “Filial Piety”]. *Journal of Hubei Engineering University* [湖北工程学院学报] 34/2: 14-17. [CNKI, date of last visit: 26th March, 2022]
- Zhang, W. [张玮恒] (2020). “Xiao” zi gujin tan [“孝”字古今谈 /Ancient and Contemporary Account on the Character “Xiao”]. *Hanzi wenhua* [汉字文化] 19 (265): 80 – 81. [CNKI, date of last visit: 20th September 2021]

Corpus:

- Beijing waiguoyu daxue yingyu xi cidianzu bian [北京外国语大学英语系词典组编]. (1997/2006). *Han-Ying cidian (xiuding ban suoyin ben)* [汉英词典 (修订版缩印本)/*A Chinese-English Dictionary (Revised edition)*]. Beijing: Foreign Language Teaching and Research Press [北京: 外语教学与研究出版社] (in the text referred to as CED).
- Chen, X. [陈欣望] (编著/ ed.). (1991). *Hanying yulin* [汉英语林:成语、典故、谚语、褒贬语、隐语、委婉语、古语、俗语、警句、诗辞熟句/*A Dictionary of Chinese idioms and phrases, proverbs and allusions, eulogistic and derogatory terms, enigmas and euphemisms, famous and popular sayings, sparkling sentences and well-known lines in ancient poems, lyrics and literary compositions with English translation*]. Shanghai: Shanghai Jiao Tong University Press [上海: 上海交通大学出版社] (in the text referred to as CEYL).
- Wang, T. *et al.* [王涛等编著/eds.]. (1987/2006). *Zhongguo chengyu da cidian*. [中国成语大辞典/*Dictionary of Chinese Idioms*]. Shanghai: Shanghai cishu chubanshe [上海: 上海辞书出版社]. (in the text referred to as DCI).

Zhongguo shehui kexueyuan yuyan yanjiusuo cidian bianjishi bian [中国社会科学院语言研究所词典编辑室编]. (2016/2018). *Xiandai Hanyu cidian* (di 7 ban) [现代汉语词典(第7版)/*Dictionary of Modern Chinese* / (7th edition)]. Beijing: Shangwu yinshuguan [北京: 商务印书馆] (in the text referred to as DMC7).

Ana M. Jovanović

孝 /XIÀO/ — POSMATRAN IZ PERSPEKTIVE KULTUROLOŠKE
LINGVISTIKE

Sažetak

Koncept 孝 /xiào/, koji se na srpski prevodi prema njegovom engleskom parnjaku kao „sinovljeva poslušnost“, jedan je od najvažnijih pojmova tradicionalne kineske kulture. U ovom radu dat je najpre kratak pregled filozofskih i socioloških tumačenja ovog pojma, a potom su predstavljeni rezultati našeg istraživanja leksičkog korpusa, koji čine četiri rečnika kineskog jezika, od kojih su dva monolingvalna, a dva bi-lingvalna (kinesko–engleska rečnika). U radu smo pokušali da damo odgovore na sledeća pitanja: kakva je produktivnost reči 孝 /xiào/, tj. koliko je leksičkih jedinica (složenica, idioma, izreka itd.) u našem korpusu izgrađeno od nje? Koje lingvo-kulturološke kategorije možemo izdvojiti i prema njima razvrstati leksičke jedinice koje u sebi kao gradivnu morfemu sadrže 孝 /xiào/? Da li se, pored očiglednog antonima, tj. reči 不孝 /bùxiào/, dobijene negacijom ovog pojma i neke druge lekseme mogu smatrati njegovim antonimima? Poslednje pitanje kojim se bavimo jesu veze koncepta 孝 /xiào/ i drugih konceptata koji su smatrani vrlinama u okviru žuističke etike, kao što su 忠 /zhōng/, 悌 /tì/, 信 /xìn/, 慈 /cí/ i drugi.

Ključne reči: 孝 /xiào/, sinovljeva poslušnost, kulturološki koncept, produktivnost, kategorije, antonimi, srodni koncepti

Danilo Savić*

Institute for Balkan Studies SASA

SOME ILLYRIAN ETHNONYMS AND THEIR SUPPOSED ALBANIAN COGNATES: *TAULANTII, DELMATAE, DARDANI***

This paper explores the relation between the names of three Palaeo-Balkan tribes and their previously proposed cognates in Albanian: *Taulantii* to Alb. *dallëndyshe* ‘swallow’, *Delmatae* to Alb. *delme* ‘sheep’, and Dardani to Alb. *dardhë* ‘pear’. In various ancient sources, these tribes and their territories are labelled as “Illyrian”, but the linguistic and ethnographic scope of this term is obscure. Furthermore, only the first pair is unproblematic from the standpoint of regular Albanian sound change: Alb. *dallënd-* may continue an earlier *Taulant-*. On the other hand, the etymology of *delme* does not necessarily apply to *Delmate*. Finally, *dardhë* and *Dardani* are not related.

Keywords: Illyrian, Albanian, Thracian, etymology, onomastics

Before dealing with the problem at hand, a few remarks on the use of the term “Illyrian” are in order¹. I follow the view that Illyrian linguistic data should be sought primarily the south-east of Roman Dalmatia,

* danilo.savic@bi.sanu.ac.rs

** This paper results from my work on the Institute for Balkan Studies SASA financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Agreement 451-03-9/2021-14/200170, in vigour since 05.02.2021). Its content benefited from the suggestions of my colleagues Đorđe Božović and Orsat Ligorio of the University of Belgrade. Its shortcomings are, of course, my responsibility.

¹ For detailed discussions on the linguistic and ethnographic uses of the term, see Katičić (1976: 154–158), Eichner (2004: 96–103), Dzino (2014), Lippert & Matzinger (2021: 117–124), Ligorio & Savić (fthc.).

roughly from the Naro (Neretva) in the north to the Drilon (Drin) in south, and in parts of Roman Macedonia, roughly from the Drin in the north to the lands between the Adriatic coast and the lake Lychnidos (Ohrid) in south. This is based on the classification of the indigenous anthroponymy of Illyricum into regions with a characteristic distribution of certain groups of personal names as established by Radoslav Katičić (1976: 179–184). In this classification, the borders outlined above correspond to the so-called South-Eastern or Illyrian anthroponymic region². “Illyrian” is an appropriate label in this case for several reasons. The territories encompassed by the Illyrian region mostly belonged to the Illyrian kingdom, which emerged during the 3rd century BCE and was conquered by the Romans in 168 BCE. Names of several of its rulers, known from literary sources, are also epigraphically attested here, e.g. Ἄγρων, *Gentius*, Τεύτα. Lastly, Roman sources place what appears to be a peregrine community named *Illyrii proprie dicti* somewhere between Epidarus (Cavtat) and Lissus (Lezhë)³. In view of these circumstances and the general acceptance of Katičić’s classification, the territorial extent of the Illyrian language has generally been equated with that of the Illyrian anthroponymic region⁴. The obvious flaw of this framework is that it relies mainly on anthroponymic systems, whose borders may differ from the linguistic ones. In spite of this, it provides a more precise point of reference for the study of Illyrian than the previous approach, most notably found in the works of Hans Krahe (1925, 1929, 1955) and Anton Mayer (1957, 1959), whose notion of Illyrian encompassed all of Illyricum,

² Note that Katičić primarily deals with personal names stemming from Dalmatia, while those from the areas south of the Drin are collected and analysed by Carlo de Simone (1993).

³ Pliny the Elder (3.144) and Pomponius Mela (2.55). For the historical context and possible interpretations of the *Illyrii proprie dicti*, see Katičić (1964) and Papazoglu (1965: 177–179).

⁴ Cf. Polomé (1982: 867–869), Wilkes (1996: 70–87), Grbić (2016), de Simone (2018), Falileyev (2020: 900–914), Lippert & Matzinger (2021: 134–142), Ligorio & Savić (fthc.).

i.e., Roman Dalmatia and Pannonia – both of which are linguistically diverse territories.

This is not to exclude the possibility that Illyrian was spoken outside of Katičić's South-Eastern region. Another plausible source would be the so-called Delmato-Pannonian anthroponymic region – home to the Delmatae, among other tribes. This region covered the Dalmatian coast from Rider (near Šibenik) to Salona (Solin), and most of central Dalmatia up to the valley of the Sava. The language of this region might have been Illyrian, but the question ultimately remains open (Katičić 1976: 183, Falileyev 2020: 905–908)⁵. The same goes for Dardania, whose population was designated as Illyrian in a number of literary sources (Papazoglu 1978: 210–218), while its attested anthroponymic repertoire is of Delmato-Pannonian stock with a supposedly older, but thin, layer of Illyrian names in the southwest, and predominantly Thracian in the east (Katičić 1976: 181, Papazoglu 1978: 221–245, Јома 2010)⁶. Although these areas are on the periphery of the Illyrian kingdom and the Illyrian anthroponymic region, it remains possible that Illyrian was spoken there at some point in history.

For these reasons, it is not a priori unreasonable to suggest that the ethnonyms *Taulantii*, *Delmatae*, *Dardani* all stem from the Illyrian language, even within the system established by Katičić. The case for the latter two was briefly addressed above. Concerning the Taulantians, their territory was well within the South-Eastern or Illyrian region. They inhabited the hinterland between Epidamnus/Dyrrachium (Durrës) and Apollonia, and are one of the best known Illyrian tribes (Mayer 1957: 331). However, in contrast to the Delmatae and the Dardanians, the name of the Taulantians seems to have disappeared quickly after the Roman conquest of the Illyrian

⁵ It is in this context that one may speak of loanwords of Illyrian origin in Serbo-Croatian, cf. material in Matasović (1995) and Ligorio & Savić (fthc.).

⁶ On the borders of Dardania see Papazoglu (1978: 187–209, 224–225). Papazoglu (1978: 241) also believed that there existed a separate Dardanian group of personal names, but this is most certainly incorrect (Јома 2010).

kingdom⁷. The territories of the Illyrian anthroponymic region and of the Illyrian kingdom coincide with most of present-day northern and central Albania. This and other factors have led a number of scholars to propose that Albanian descends from Illyrian (Katičić 1976: 184–188), but the hypothesis is not without significant problems (Lippert & Matzinger 2021: 161–167).

In any case, the three presumably Illyrian ethnonyms seem to find quite striking parallels in the Albanian lexicon:

Taulantii – Alb. *dallëndyshe* ‘swallow’

Delmatae – Alb. *delme* ‘sheep’

Dardani – Alb. *dardhë* ‘pear’

These potential pairs of cognates have long been known in scholarly literature⁸. The comparison between *Taulantii* and Alb. *dallëndyshe* is perhaps supported by the following account by Stephanus of Byzantium: Ἄβροι, ἔθνος πρὸς τῷ Ἀδρίᾳ Ταλαντίνων (sic!) προσεχὲς τοῖς Χελιδονίοις, ὡς Ἐκαταῖος – “Abroi, a Taulantian people on the Adriatic bordering the Chelidonioi, according to Hecataeus.” The information from this passage relates to the late 6th or early 5th century, as evidenced by the mention of Hecataeus. The Χελιδόνιοι obviously have a Greek name derived from Gk. χελιδών ‘swallow’. While this is no definitive proof, it is difficult to ignore the fact that the Χελιδόνιοι are somehow connected to the Taulantians. It is also possible that Χελιδόνιοι is the Greek translation of the Illyrian name (Mayer 1959: 331). Furthermore, Alb. *dallënd-* may derive from *Taulantii* mostly in accordance with Albanian sound laws (Eichner 2004: 107–108); *-ysh(e)* is diminutive suffix, cf. for example *këlysh* ‘young animal, cub’. For Alb. *a* from **au* one may adduce *ag* ‘twilight’

⁷ Apart from their mention in the passages of Pliny and Mela (fn. 4), their name does not reappear under Roman administration.

⁸ See Mayer (1959: 36–37, 61, 112) and Çabej (1976: 105–107, 111) with earlier references. Mayer supports all of these comparisons, while Çabej does not give his opinion explicitly, except in the case of *Taulantii* and *dallëndyshe* where he agrees with Mayer.

< **h₂eug-*, cf. Gk. ἀγή ‘light’, or *ar* ‘gold’ < Lat. *aurum* ‘id.’⁹ For Alb. -*ll-* from intervocalic **l* one may adduce *hell* ‘spit, icicle’ < **skoł-o-*, cf. Gk. σκῶλος ‘pointed pole’, or *shpellë* ‘cave’ < Gk. σπήλαιον. An unstressed vowel can result in Alb. *ë* under a number of conditions (de Vaan 2018b: 1737), cf. *shëndet* ‘health’ < Lat. *sanitatem* ‘id.’ or *upeshkëp* ‘priest’ < Lat. *episcopus*. Albanian -*nd-* may result from an older *-*nt-*, as in *kuvend* ‘assembly’ < Lat. *conventum* ‘meeting’, although this is not always the case, cf. *ërgjent* ‘silver’ < Lat. *argentum* ‘id.’ (Landi 1988: 116–117). A more apparent problem arises in view of Illyr. *t-* for Alb. *d-*. One may propose an assimilation **taulant-* > **tauland-* > **dauland-* (Mayer 1939: 91–92), but there is no way to affirm this or other solutions. A hapax Δαυλάντιον ἔθνος (Nonnus, *Dionysiaca* 44.1) is of interest here, but its input remains unclear.

While the exact relation between Illyr. *t-* and Alb. *d-* would benefit from more precision, it appears quite certain that Alb. *dallënd-yshe* continues Illyr. *Taulantii*. Whether as a loanword or as an inherited form, it is difficult to say: the change **au* > Alb. *a* does not seem particularly old, since it affected Latin loanwords as well (cf. Alb. *ar* seen above, and Alb. *pak* ‘few’ < Lat. *paucus* ‘id.’). There is no plausible Indo-European etymology. Mayer’s connection with the root **teuh₂₋* ‘to swell’ (LIV²: 639) is semantically and formally untenable. Orel (1998: 55) considers *dallëndyshe* to be “a relatively recent compound” consisting of **dalluan* > Alb. *dalluar* (participle of *dalloj* ‘to split’) and *dysh* ‘split in two parts’. His solution is unsatisfactory, since it ignores the diminutive function of *-ysh(e)*, as well as the very plausible link with the Illyrian form. The comparison with the Germanic words for swallow (ON *svala*, OHG *swalawa*, etc.) hinges on the disputed development of **su-* into Alb. *d-* (favoured by Kortlandt 1998: 37, excluded by Demiraj 1997: 48, Orel 2000: 82). More importantly, the Germanic forms are without Indo-European cognates and have no clear etymology (Kroonen 2013: 495).

⁹ The form Ταλαντ- is not necessarily genuine, as it may be corrupted by Gk. τάλαντος, τάλαντεύω.

The case of *Delmatae* and Alb. *delme* ‘sheep’ is less straightforward, with some authors rejecting the connection between the two (Demiraj 1997: 127–128, Orel 1998: 58). The lack of compelling evidence in ancient literary sources that *Delmatae* is derived from a word meaning ‘sheep’ has been noted earlier (Katičić 1976: 173). The Greek geographer Strabo (7.5.5) supposedly glossed the name of their capital Δέλιμιον· πεδίων μηλόβοτον ‘pasturage for sheep’¹⁰. As noted by Katičić, the context clearly indicates that Δέλιμιον was razed by the Romans and turned into a πεδίων μηλόβοτον. It is also uncertain whether the Albanian etymology is applicable to *Delmatae* as well.

The form *delme* is attributed to the Geg dialect, with *dele* being the standard Tosk word. However, the presence of derivatives in *delm-* across different dialects suggests an originally wider distribution of *delme* (Çabej 1976: 111, Demiraj 1997: 127). Both forms are usually traced back to the root **d^heh₁-(i-)* ‘to suckle’ (LIV²: 138–139), but there is no agreement upon one particular proto-form. *Dele* would then be one of many *l*-derivatives of this root, such as Lat. *filius* ‘son’ < **d^hih₁-l-* < **d^hh₁-il-*, Gk. θῆλυς ‘female’, Skt. *dhārú-* ‘sucking’ < **d^heh₁-lu-*¹¹. The root vowel is most often derived from a PALb. diphthong **ai* with different possible PIE sources (Demiraj 1997: 127 and Orel 1998: 58 with earlier references).

Demiraj reconstructs **d^heh₁-i-lieh₂-*, whereby **eh₁i* presumably gives PALb. **ai* > Alb. *e*, merging with other PIE *i*-diphthongs (cf. Demiraj 1997: 45). The closest formal parallel (without the suffix *-i-*) is found in Lith. *dėlė* ‘leech’ and *pirm(a)-dėlė* ‘cow which bears a calf for the first time’ < **d^heh₁-lieh₂-*, cf. also *pirm(a)-dėlys* ‘first-born’ < **d^heh₁-lio-* (Fraenkel 1962: 87, Derksen 2015: 120–121). However, the development of **eh₁i* (and of **h₁ei*, **ei*) into Alb. *e* is not necessarily correct, as most examples are ambiguous (cf. Vermeer 2008: 595, de

¹⁰ Strabo’s form possibly emendable to the more common Δελμίνιον/Lat. *Delminium* (Mayer 1957: 118).

¹¹ On the morphology and the semantics of **d^heh₁-i-* ‘to suck’ and its relation with **d^heh₁-* ‘to put’ (LIV²: 136) see de Vaan (2018a). The root **d^heh₁-(i-)* has three basic meanings: ‘to suck milk from a teat’, ‘to nurse’, ‘to lactate’ (de Vaan 2018a: 177), all of which may surface in individual languages. Also, cf. the similar discussion in Martirosyan (2010: 230–231).

Vaan 2018b: 1738). Thus, one may in theory derive Alb. *be* ‘oath’ either from *b^heid^h*- like Gk. *πείθω* ‘persuade’ (Demiraj 1997: 45), or from *b^hoid^h*- like Lat. *foedus* ‘alliance’ (Orel 2000: 12). For Alb. *e* < PALb. **ai* < PIE **oi*, a standard example is *shteg* ‘path’ < **stoig^h*- (Gk. *στοίχος* ‘row’). On the other hand, Alb. *dimër* ‘winter’ may continue **g^heim*- like Gk. *χειμών*, Lith. *žiemà* ‘id.’ (Orel 2000: 13), or **g^him*- like Skt. *himá-* ‘cold, frost’; according to Demiraj (1997: 133), unstressed **ei* gave *i* in *dimër*.

A better alternative for Alb. *dele* may then be **d^hh₁-oi-lieh₂*- with an exact formal and, perhaps, semantic match in SCr. *dòjilja* ‘nurse’, cf. SCr. *dòjiti* ‘breast-feed’, OCS *doiti* ‘to nurse’ < **d^hh₁-oi-*.

Furthermore, a diphthong is not necessary if one assumes a zero-grade **d^hh₁-lieh₂*-, without the suffix **-i-*. This would give PALb. **daliā* > *dele*, where **a* > *e* is explained by *i*-mutation (on which see e.g. Demiraj 1997: 47, de Vaan 2018b passim). To my knowledge, the only formal parallel would be Arm. *dal* ‘colostrum, beestings’ that has been reconstructed as **d^hh₁-li(e)h₂*- (Martirosyan 2010: 231), but the semantics are perhaps less compelling than in the previous case.

Whatever the origin of its root vowel, *dele* is unproblematically understood as a *l*-derivative of the root **d^heh₁-(i-*), with the suffix **-io-*. This interpretation supports the view that *delme* was secondarily derived from *dele*, via the suffix **-mo-*, as it would otherwise present an isolated and an unusual formation in PIE. A comparable situation is apparently seen in Alb. *djalë* ‘boy’, pl. *djem* < *djelm*. Demiraj (1997: 134) explains this through the existence of two competing stems, one with sg. *djal-* and pl. *djel-*, other with sg. *djalm-* and pl. *djelm-*, which later merged into one. *Djalë* probably derives from the same root as *dele*, but a precise reconstruction is unknown (Demiraj 1997: 134–135, Orel 1998: 67). If this is correct, *delme* and *djelm* would be the only two forms of the root **d^heh₁-(i-*) in which **-mo-* is used. A similar morphological relation was suggested for Alb. *zjarrë* ‘fire’ and the pl. *zjerm*, whence perhaps an analogical sg. *zjarm* (de Vaan 2004: 80–82). In view of the plural stems *djelm-* and *zjerm-*, it is tempting to see *delm-* as a singularised plural¹². Taken together, these three

¹² Cf. a similar proposal by Çabej (1976: 111).

forms may suggest a common morphological pattern that developed in Proto-Albanian, where *-mo- is a formant of the plural.

Regarding the relation between *delme* and *Delmatae*, it is difficult to say whether the latter can continue a form with a diphthong. There is not much material to confirm or deny this. A diphthong may be conserved in the ethnonym *Daesitiates* (Mayer 1957: 104), neighbours of the *Delmatae*, or in the personal names in *Laid-/Laed-*, common in Illyrian and Delmato-Pannonian regions (Mayer 1957: 203, de Simone 1993: 60–61). On the other hand, **d^hh₁-l-m-* seems phonologically plausible for *Delmatae*. But, as in the case of *delme*, such a form is unexpected in PIE. If the two are indeed related, one could expect a secondary addition of the suffix *-mo- in *Delmatae* as well. It is worthwhile noting that *Delm-* appears in anthroponymy, cf. *Delmanna* (AE 1983 742, Dalmatia), *Dalmana* (IMS 4 89, Upper Moesia). The second element is either the frequent Illyrian and Delmato-Pannonian anthroponymic stem *Annā-* (Mayer 1957: 46–47, de Simone 1993: 52–53), or simply the suffix *-āno-*. Apart from the same stem, the two names show a variation between *a* and *e*, of unknown origin, also seen in various attested forms of *Delmatae* (Mayer 1957: 117). While the use of *-mo- on a *l*-derivative of the same root would be an interesting correspondence between Albanian and Illyrian, there is no evidence to support that this suffix was used for comparable purposes, nor that the meaning ‘sheep (vel.sim.)’ can be ascribed to the *Delmatae*. More importantly, given the state of the material, one may even derive *Delmatae* from the root **d^heh₁-* ‘to put’ (cf. OCS *dělo* ‘deed’, SCr. *dj lo* ‘id.’ < **d^heh₁-lo-*). Evidently, any comparison of *Delmatae* with *delme* requires an undesirable amount of speculation. On the other hand, an etymological link between the two forms is not impossible in theory. Although onomastic items derived from animal names are not uncommon, cf. *Taulantii* discussed above, *Ulcinium* < **ul^ku-* ‘wolf’ (SCr. *Ulcinj/Alb. Ulqin*), Thrac. *Βέβρυκες* < **b^heb^hr-u-* ‘beaver’ (tribe in Bithynia), one must also take into account the possibility that *Delmatae* and *delme* may come from the same root while having a different meaning. The problem is best left open for now.

The comparison between *Dardani* and Alb. *dardhë* ‘pear’ is not without problems as well. As mentioned before, the south-western parts of Dardania were seemingly considered Illyrian in early sources, while their anthroponymy is Delmato-Pannonian in Roman times. In the remaining parts of Dardania, Thracian anthroponymy is prevalent. It is then not a coincidence, perhaps, that ancient Balkan forms in *Dard-/Derd-* occur almost exclusively in a Thracian context¹³. Moreover, these are often compounds showing Thracian derivational patterns, cf. the personal names *Dardi-sanus*/Δαρδι-ζανις, Δερζι-ζενις, *Derzi-tralis*, and the place name Δαρδά-παρα. The second elements -ζανις/-ζενις ‘born of’ ‘born of’ < **gēnh₁*- (Gk. -γενής), -*tralis* ‘?’ (Dana 2014: 378–379), -παρα ‘river (?)’ (Georgiev 1977: 187–188) are all well attested in Thracian onomastics. Forms in *Derd-* must belong here in view of Thracian graphical fluctuations between *a* and *e* (Yanakieva 2018: 41), an example of which is -ζανις/-ζενις¹⁴. The -*z*- in *Derz-* is then best explained as a result of palatalisation. In this case, forms such as *Derzenus* conceivably derive from Δερζιζενις by way of haplology. The name *Derzō* is then probably a hypocoristic of *Derzenus*. This survey shows that *Dard-/Derd-* was a productive stem in Thracian onomastics, and that it is safe to assume its Thracian provenance. Here it should be noted that a similar name Δέρδας is particularly frequent in Epirus and Macedon¹⁵, while some examples are scattered

¹³ Cf. material in Detschew (1957: 117–119), and in Dana (2014: 112, 123–125); forms in *Derd-* are more frequent than the ones in *Dard-*. The Dardanians of Asia Minor are probably of Thracian descent, having migrated there from Thrace proper towards the end of the 2nd millennium BCE, along with other Thracian tribes, such as the Bithynians and the Mysians (cf. Brixhe & Panayotou 1994: 184). I have no opinion regarding the Apulian *Dardi* mentioned by Pliny the Elder (3.104).

¹⁴ Note that Georgiev (1977: 78) rather sees -ζανις as an *o*-grade variant comparable to Gk. -γοϋος. I do not know whether Thrac. *a/e* variants have anything to do with *Delm-/Dalm-*.

¹⁵ Six attestations in Epirus, mostly in Buthrotum (LGPN IIIa: 121), and ten in Macedon (LGPN IIIa: 89).

across Greece¹⁶, but none are found in Thracian territories. If Δέρδας is related to the Thracian names presented above, I would assume that it was imported from Thrace, since the stem *Derd-* is not productive in Epirus and Macedon. Furthermore, the highest concentration of examples comes from Macedon, which borders the western fringes of Thrace.

None of the names mentioned so far has been attributed to Illyrian, nor does such a proposal seem plausible. One Δάρδανος is found in Dyrrachium (LGNP IIIa: 117). Some instances of *Dardanus* are known in Dalmatia, but these come from Narona (ILJug 2 654) and Salona (ILJug 3 2041), both large settlements with diverse population. Furthermore, *Dardanus* may simply be derived from the ethnonym *Dardani*, such as Δαρδάνα τὸ γένος Δαρδάναν (LGNP IIIa: 116). Such attestations have no bearing on the problem.

I conclude that the names in *Dard-/Derd-* should be regarded as Thracian. This is relevant in view of different etymological proposals regarding Alb. *dardhë* ‘pear’. Çabej (1976: 107) derives *dardhë* from **d^horg^h-*, comparable to SCr. *drìjen* ‘cornel’, OHG *dirn-baum*, OIr. *draigen* ‘blackthorn, sloe’, maybe Gk. *τερχνός* ‘twig’, all from **d^h(e)rg^h-no-*. Note that in this case only the Albanian form does not take the suffix *-no-. Demiraj (1997: 121–122) proposes *dardhë* < **d^horHg^h-eh₂-* comparing it to *dredhë* ‘strawberry’, which he derives from **d^hrHg^h-eh₂-* (1997: 144), and further compares it with Lat. *frāgum* ‘id.’ only. However, the Latin form could be unrelated to *dredhë* and is possibly non Indo-European (de Vaan 2008: 339). Finally, Orel (1998: 56) considers that *dardhë* is derived from the verb *derdh* ‘to tip out, pour’ PIE *d^herg^h-* ‘to turn’ (LIV²: 146), but the semantics are not convincing. While Çabej’s etymology seems preferable, all of the mentioned proposals require **g^h* to explain Alb. *dh*. If this is the case, then the Thracian forms in *Dard-/Derd-* cannot be compared with

¹⁶ Δαρδα in Boeotia (LGNP IIIb: 106; female name?), Δέρδας twice in Thessaly (LGNP IIIb: 110; once as a patronym Δερδαία). Cf. Krahe (1929: 41–42), Mayer (1957: 108–109), de Simone (1993: 67–68). Krahe and Mayer present the relevant literary attestations as well. Given the geographical and historical context in which it occurs, Δέρδας sometimes seen as a Macedonian name.

Alb. *dardhë*, because **gʰ* and **gʰ* give Thrac. *z* or *s* depending on the orthography of the individual form, cf. *Dardisanus*/Δαρζιζενις seen above, or Thrac. -διζα ‘(walled?) settlement’ < **dʰ(e)igʰ*- related to Gk. τεῖχος ‘wall’, Av. *da za-* ‘id.’ Given the state of Illyrian linguistic remains, and the geographical distribution of the onomastic material, it is highly improbable that *Dard-/Derd-* can be taken as an Illyrian loan in Thracian. Additionally, the change **gʰ* > *d* cannot be accounted for in Illyrian. Therefore, Alb. *dardhë* is in all probability unrelated to *Dardani*¹⁷. The latter is perhaps better compared with Skt. *dardurá-* ‘frog, pipe’, Lith. *dardėti* ‘to rattle, chatter’, Gk. δάρδα· μέλισσα ‘bee’, sometimes read μόλυσμα ‘stain’, δαρδαίνει· μολύνει ‘to stain’ (both from Hesychius), but this merits a separate discussion, particularly in view of the aberrant semantics of the Greek glosses. If δάρδα is indeed understood as ‘bee’, the starting point for all these forms may be ‘noise, chatter (vel.sim.)’.

The following conclusions may be drawn from the previous discussion. The ethnonym *Taulantii* is indeed related to Alb. *dallëndyshe* ‘swallow’, but its origin in Albanian is currently unexplained. Maintaining the relationship between *Delmatae* and *delme* is quite difficult, although theoretically not impossible. More data from the Illyrian side would be helpful in resolving this question. On the other hand, *Dardani* is better understood as a Thracian form, comparable to Thracian names such as *Dardisanus*, *Derzitrallis*, etc. Furthermore, *Dardani* and *dardhë* are not related, since the proposed etymologies for the Albanian form cannot explain Thracian *Dard-/Derd-*.

Abbreviations and symbols

* — reconstructed form

> — yields

Lat. — Latin

Lith. — Lithuanian

¹⁷ I left out of consideration the comparison of Alb. *dardhë* with Gk. ἄχερδος ‘pear’, which is most probably of non Indo-European origin (Beekes 2010: 182). Any common form for these two words would still be incompatible with the Thracian material.

| | |
|---|---------------------------|
| < — comes from | OCS — Old Church Slavonic |
| AE — <i>Année épigraphique</i> | OHG — Old High German |
| Alb. — Albanian | PIE — ON — Old Norse |
| Arm. — Armenian | Proto-Indo-European |
| Av. — Avestan | pl. — plural |
| Gk. — Greek | Thrac. — Thracian |
| Illyr. — Illyrian | SCr. — Serbo-Croatian |
| IMS — <i>Inscriptions de la Mésie supérieure</i> | sg. — singular |
| ILJug. — <i>Inscriptiones Latinae in Jugoslavia</i> | Skt. — Sanskrit |

REFERENCES

- Beekes, R. S. P. (2010). *Etymological Dictionary of Greek*. Leiden.
- Brixhe, C. & Panayotou, A. (1994). Le thrace. In: *Langues indoeuropéennes* (Bader, F. ed.). Paris, 179–203.
- Çabej, E. (1976). *Studime etimologjike I*. Prishtina.
- Dana, D. (2014). *Onomasticon Thracicum. Répertoire des noms indigènes de Thrace, Macédoine Orientale, Mésies, Dacie et Bithynie*. Athens.
- Demiraj, B. (1997). *Albanische Etymologien*. Amsterdam.
- Derksen, R. (2015). *Etymological Dictionary of the Baltic Inherited Lexicon*. Leiden.
- Detschew, D. (1957). *Die thrakischen Sprachreste*. Vienna.
- Dzino, D. (2014). 'Illyrians' in ancient ethnographic discourse. *Dialogues d'histoire ancienne* 40/2, 45–65.
- Eienher, H. (2004). Illyrisch – die unbekannte Sprache. In: *Die Illyrer. Archäologische Funde des 1. Vorchristlichen Jahrtausends aus Albanien* (Bejko, L. ed.). Asparn-Zaya, 92–117.
- Falileyev, A. (2020). The silent Europe. *Palaeohispanica: Revista sobre lenguas y culturas de la Hispania antigua* 20/2: 887–919.
- Fraenkel, E. (1962). *Litauisches etymologisches Wörterbuch*. Heidelberg.
- Georgiev, V. (1977). *Trakite i tehniyat ezik*. Sofia.
- Grbić, D. (2016). Greek, Latin and Paleo-Balkan Languages in Contact. *Rhesis* 7/1: 32–41.
- Katičić, R. (1976). *Ancient Languages of the Balkans*. The Hague.
- Kortlandt, F. (1998). Reflexes of Indo-European Consonants in Albanian. *Orpheus* 8: 35–37.

- Krahe, H. (1925). *Die Alten balkanillyrischen geographischen Namen*. Heidelberg.
- Krahe, H. (1929). *Lexikon altillyrischer PPersonennamen*. Heidelberg.
- Krahe, H. (1955). *Die Sprache der Illyrier. Erster Teil: Die Quellen*. Wiesbaden.
- Kroonen, G. (2013). *Etymological Dictionary of Proto-Germanic*. Leiden.
- Landi, A. (1988). *Gli elementi latini nella lingua albanese*. Salerno.
- LGPN IIIa = Fraser, P. M. & Matthews, E. (1997). *A Lexicon of Greek Personal Names: Vol. IIIa. Peloponnese, Western Greece, Sicily, and Magna Graecia*. Oxford.
- LGPN IIIb = Fraser, P. M. et al. (2000). *A Lexicon of Greek Personal Names: Vol. IIIb. Central Greece: From the Megarid to Thessaly*. Oxford.
- Ligorio, O. & Savić D. fthc. Illyrian and Slavic. In: *Brill's Encyclopedia of Slavic Languages and Linguistics Online* (Greenberg, Marc L. & Grenoble Lenore A. eds.).
- Lippert, A. & Matzinger, Joachim. 2021. *Die Illyrer. Geschichte, Archäologie und Sprache*. Stuttgart.
- LIV² = Rix, H. et al. (ed.). (2001²). *Lexicon der indogermanischen Verben*. Wiesbaden.
- Matasović, R. (1995). Skokove 'ilirske' etimologije. *Folia onomastica Croatica* 4: 89–101.
- Mayer, A. (1939). *Illyrisches*. KZ, 66/1: 75–127.
- Mayer, A. (1957). *Die Sprache der alten Illyrier. Band I: Einleitung. Wörterbuch der illyrischen Sprachreste*. Vienna.
- Mayer, A. (1959). *Die Sprache der alten Illyrier. Band II: Etymologisches Wörterbuch des Illyrischen, Grammatik der illyrischen Sprache*. Vienna.
- Martirosyan, H. (2010). *Etymological Dictionary of the Armenian Inherited Lexicon*. Leiden.
- Orel, V. (1998). *Albanian Etymological Dictionary*. Leiden.
- Orel, V. (2000). *A Concise Historical Grammar of the Albanian Language*. Leiden.
- Papazoglu, F. (1965). Les origines et la destinée de l'Etat illyrien. *Historia* 14: 143–179.
- Papazoglu, F. (1978). *The Central Balkan Tribes in Pre-Roman Times*. Amsterdam.
- Polomé, E. (1982). Balkan languages. In: Boardman, J. et al. (ed.), *The Cambridge Ancient History*. London, 866–888.
- Simone, C. de. (1993). L'elemento non greco nelle iscrizioni di Durazzo ed Apollonia. In: *Grecs et Illyriens dans les inscriptions en langue grecque*

- d' Epidamnè-Dyrrhachion et d' Apollonia d' Illyrie*. (Cabanès, P. ed.). Paris, 35–75.
- Simone, C. de. (2018). Illyrian. In: *Handbook of Comparative and Historical Indo-European Linguistics 3* (Klein J. et al. eds.). Berlin, 1867–1872.
- Vaan, M. de. (2004). PIE *e in Albanian. *Die Sprache* 44/1: 70–85.
- Vaan, M. de. (2008). *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden.
- Vaan, M. de. (2018a). On the homonymy of 'put' and 'suck' in Proto-Indo-European. *Indo-European Linguistics 7*: 176–193.
- Vaan, M. de. (2018b). The phonology of Albanian. In: *Handbook of Comparative and Historical Indo-European Linguistics 3* (Klein J. et al. eds.). Berlin, 1732–1748.
- Vermeer, W. (2008). The prehistory of the Albanian vowel system: a preliminary exploration. *Studies in Slavic and General Linguistics 32*: 591–608.
- Wilkes, J. J. (1996). *The Illyrians*. Oxford.
- Yanakieva, S. (2018). The Thracian Language. *Orpheus* 25: 26–68.

*

Лома, С. (2010). Домородачко становништво Косова и Метохије у римско доба у светлу антропонимије. In: *Косово и Метохија у цивилизацијским шоковима: међународни тематски зборник 3* (Атлагић, Марко ed.). Kosovska Mitrovica, 19–40.

Danilo Savić

QUELQUES ETHNONYMES ILLYRIENS ET LEURS COGNATS SUPPOSÉS
EN ALBANAIS : *TAULANTII*, *DELMATAE*, *DARDANI*

Résumé

Les noms de trois tribus des Balkans anciens ont été comparés avec des formes possiblement apparentées en albanais : *Taulantii* avec alb. *dallëndyshe* 'hirondelle', *Delmatae* avec alb. *delme* 'brebis', *Dardani* avec alb. *dardhë* 'poire'. Plusieurs sources anciennes intègrent ces tribus et leurs territoires au complexe « illyrien », mais

l'apport linguistique de ce terme est obscur. Par ailleurs, seulement la première paire n'est pas problématique en ce qui concerne les développements phonologiques en albanais et en illyrien : on peut dériver alb. *dallënd-* de *Taulant-* sans grande réserve. La comparaison directe entre *Delmatae* et *delme* est indémontrable, car elle impose plusieurs spéculations. Pourtant, il reste possible que les deux formes remontent à la même racine. De l'autre côté, *dardhë* et *Dardani* ne sont pas apparentés. La distribution des formes en *Dard-/Derd-* permet de les attribuer à la langue thrace. Dans ce cas, *Dardani* n'est pas compatible avec les étymologies proposées pour *dardhë*.

Mots clés : illyrien, albanais, thrace, étymologie, onomastique

UNIVERSITY OF BELGRADE
FACULTY OF PHILOLOGY
DEPARTMENT OF GENERAL LINGUISTICS

BeLiDa 1

Publisher

Faculty of Philology, University of Belgrade

Proofreading

Milena Oparnica
Emilija Milojević
Ana Arsić

Graphic design

Biljana Živojinović

Printed by

Čigoja štampa, Beograd, Srbija, 2022.

ISBN 978-86-6153-695-3

<https://doi.org/10.18485/belida.2022.1>

Copyright © Faculty of Philology, University of Belgrade, 2022

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

81'362(082)

81-115(082)

BELIDA 1 : thematic collection of papers / editors Natalija Panić Cerovski, Borko Kovačević, Milica Dinić Marinković. – Belgrade : University, Faculty of Philology, 2022 ([Beograd] : Čigoja štampa). – 463 str. : ilustr. ; 24 cm

Na vrhu nasl str.: Department of General Linguistics, Faculty of Philology University of Belgrade, Serbia. – Tiraž 30. – Str. 5–8: Editors' preface / the editors. – Napomene i bibliografske reference uz tekst. – Bibliografija uz svaki rad. – Резюме; Riassunto; Résumé.

ISBN 978-86-6153-695-3

а) Компаративна граматика – Зборници б) Компаративна лингвистика – Зборници

COBISS.SR-ID 75863305



9 788661 536953