

Izvorni znanstveni rad

Rukopis primljen 30. 6. 2020.

Prihvaćen za tisak 12. 3. 2021.

<https://doi.org/10.22210/govor.2020.37.08>

**Philipp Wasserscheidt**<sup>1</sup>, **Marija Mandić**<sup>5</sup>

*philipp.wasserscheidt@hu-berlin.de, marija.mandic@bi.sanu.ac.rs*

**Nadine Vollstädt**<sup>1</sup>, **Ana Jovanović**<sup>2</sup>, **Ivana Tanasijević**<sup>3</sup>

*nadine.vollstaedt@googlemail.com, anajovano@gmail.com, ivana@matf.bg.ac.rs*

**Teodora Vuković**<sup>4</sup>, **Ivana Vučina Simović**<sup>2</sup>, **Uliana Yazhinova**<sup>1</sup>

*bravethea@gmail.com, ivanavusim@gmail.com, yazhinovaul@gmail.com*

**Andelka Zečević**<sup>3</sup>

*andjelkaz@matf.bg.ac.rs*

<sup>1</sup> Department of Slavic and Hungarian Studies, Humboldt-University of Berlin  
Germany

<sup>2</sup> Faculty of Philology, University of Belgrade, Serbia

<sup>3</sup> Faculty of Mathematics, University of Belgrade, Serbia

<sup>4</sup> Department for Slavic Studies, University of Zurich, Switzerland

<sup>5</sup> Institute for Balkan Studies, Serbian Academy of Sciences and Arts, Serbia

## Corpus-based analysis of spoken narratives. Introducing a corpus and a search tool

### Summary

This paper is concerned with the development of a synchronic corpus containing Serbian spoken narratives and its use for narrative analysis. The corpus (CRONUS – Corpus for the Research On Narratives and their Use in Speech) is optimised to study the structure and use of this discourse genre. First, data sources are presented, followed by corpus creation and access. The semi-spontaneous spoken narratives were orthographically transcribed, and the corpus deeply annotated, with special emphasis on the annotation of narrative sections following Labov's approach and the annotation of argument structure constructions in the sense of Construction Grammar. Three case studies demonstrate how morphological and constructional annotation can be effective for the exploration of narratives.

**Keywords:** narrative, corpus, spoken language, Serbian, deep annotation

## 1. INTRODUCTION

Narratives are one of the most important and one of the most often studied discourse genres (Bamberg & Andrews, 2004; Bruner, 1991; de Fina, 2003; Georgakopoulou, 2007; Labov & Waletzky, 2003[1967]; Polanyi, 1979, to name only a few). The foundation for the study of narratives was laid by William Labov (Labov & Waletzky, 2003[1967]). According to Labov, narratives are not random stories about the past, but are characterized by a specific temporal structuring and social function. Within the framework of his structuralist analytical tradition, narratives have been interpreted as complex linguistic constructions. However, conversational narratives are also important places of social action, providing material for the study of individual and collective negotiations of identity and social orientation (Bamberg, 2004; de Fina, 2003). Narratives, thus, hold a key position between the two important levels of linguistic organization and are mutually integrated into the semiotic continuum in two directions downwards and upwards. The first one defines the semantic, syntactic and morphological structure and the semiotic continuum is inconceivable without it. The second one constitutes the identity-forming themes which often only have their justification within these families of stories. Thus, narratives are also at the interface of various cognitive domains: they are part of language, culture, social organization and serve as storages and filters for experiences and perceptions.

A deeper understanding of narratives in this dual function, therefore, has a high interdisciplinary significance for linguistics, cultural studies, anthropology, but also for digital humanities and modern, corpus-linguistic methods and for their part, promising to advance the complex analysis of narratives.

Narratives are also of interest in Natural Language Processing (NLP) because they contain information about speakers, events, and the stance of the speaker towards these events. Applications in the field of Information Retrieval or Text Summarization thus must recognize and decode narrative structures. For some special text types, there is a number of tools that can handle this task, for example for the processing of job applications or the summarization of medical reports (Savova, Chapman, Zheng, & Crowley, 2011). However, most of the NLP approaches and tools use the term narrative in a very broad sense, as anything that combines two events in time or does not provide information in form of a chart. In most cases, they deal with written language and not with spoken discourse. For example, abstracts of scientific papers have been regarded as narratives (Prabhakaran, Hamilton, McFarland, & Jurafsky,

---

2016). Still other approaches interpret the term narrative similarly to what has been described as semantic frame. Chambers and Jurafsky (2008, 2010) analyse typical chains of actions connected to crime and court as narrative chains (the ‘prosecution frame’).

Spoken narratives, which this paper is concerned with, are complex communicative units, they can only be studied using deeply and consistently annotated spoken corpora. The construction and annotation of a spoken corpus is challenging in many ways. This is especially the case for Serbian, since, practically, no spoken corpus is available, and many tools have not yet been developed. The construction of a corpus of narratives has some additional challenges to it: as the theory building in narrative theory is still developing dynamically, annotation must be as flexible and theory independent as possible, without using coarse categories.

The original idea of the project that led to the corpus described in this paper was to create a database that enables the investigation of narratives as constructions in the understanding of Construction Grammar (Goldberg, 1995). There already exist several small corpora that focus on either written or spoken narratives. One of the oldest is the PetersonMcCabe Corpus (Peterson & McCabe, 1983) which is sampled from elicited child narratives. A comparable resource is the (written) N2 narrative corpus which has been developed for the purpose of automated language processing ([narrative.csail.mit.edu/n2/](http://narrative.csail.mit.edu/n2/)). Corpora that are dedicated to the investigation of spoken narratives are the French Oral Narrative Corpus ([frenchoralnarrative.qub.ac.uk](http://frenchoralnarrative.qub.ac.uk)) and the English Narrative Corpus (Rühlemann & O’Donnell, 2012).

Based on the available information, none of the mentioned corpora includes the elements that in the tradition of Labov have been called narrative components. Instead, mostly machine-readable levels have been interpreted and some levels of annotation that can be added semi-automatically. Since the purpose of the corpus described in this paper is the investigation of the constructional structure of narratives, we found it vital to include other levels of annotation, even if their status may be more spurious. Rühlemann and O’Donnell (2012) explicitly refute a componential analysis referring to the unclear status of the components itself. In our view, this uncertainty can only be overcome with the help of corpus analyses of those components. Swanson, Rahimtoroghi, Corcoran, and Walker (2014) show that an automated annotation of narrative components in the sense of Labov and Waletzky is, in principle, possible. So, one of the aims of the present corpus is to offer a resource that enables research into the characteristics of narratives and their components. Also, in contrast to the

---

existing corpora, only spontaneous and semi-spontaneous spoken data are included in this corpus.

This paper focuses primarily on the challenges and solutions that we have found in the representation of spoken data in Serbian. Both the aspects of the technical infrastructure and the conceptual levels of analysis, that together ensure an extensible, reusable and comparable corpus for the study of the structure of narratives, will be included. Several case studies will illustrate how this corpus can be used to study the different levels that together constitute a narrative.

In section (2) the data sources (2.1), the processing steps (2.2), the annotation layers available (2.3) and the access to CRONUS (2.4) will be introduced. Building on that, the different layers of annotation in section (3), starting with transcription (3.1), the token-level annotation (3.2) and continuing with narrative (3.3) and constructional (3.4) annotation will be addressed. Section (4) will exemplify the need for an open, multi-layer architecture by several case studies.

## 2. THE CORPUS

Version 1.0 of the corpus contains 57 transcripts of narratives with 8 embedded narratives in 50 documents. The corpus contains 25,320 words (lexical words, without punctuation marks or similar) and 3,867 types produced by 18 speakers. The narratives range from 24 to 1,000 words of length and are parts of ethnographic interviews conducted by the Balkan Institute of the Serbian Academy of Sciences and Arts between 2004 and 2008. The documents are stored in separate XML files according to the TEI standard. Each document is stored with comprehensive metadata about speakers, interviewers, place and date of recording as well as other metadata concerning the preparation of the text like transcribers and correctors. Speaker metadata are stored in separate files. The texts are transcribed diplomatically (section Transcription ) and deeply annotated (sections 3.2–3.4).

### 2.1. Data source

The source for the narrative analysis was the Digital Archive of the Institute for Balkan Studies of the Serbian Academy of Sciences and Arts (DABI), which stores several thousands of hours of interviews, photos and video material collected among multi-ethnic and multilingual communities in South-East Europe over the course of two decades (Ilić [Mandić] & Đurić Milovanović, 2012; Sikimić, 2012). It was decided

---

to focus on the South Banat, particularly on the narratives collected within the Serbian community in the village of Omoljica. Since 2004, the team of the Institute for Balkan Studies and its associates have collected hundreds of audio recordings among different speech and ethnic communities in the South Banat and produced a body of scholarly work on the South Banat communities. The scope of this paper does not allow us to cite all relevant papers but just a tight selection, cf. on Bulgarians Vučković (2007) and Sikimić and Nomaći (2016), on Hungarians Wasserscheidt (2010) and Bala (2010, 2015), on Roma Sikimić, Hristov, and Golubović (2012) and on Romanians Sorescu Marinković (2016). The choice of Omoljica was motivated by its tradition of multiculturalism and interethnic encounters, as well as by its tumultuous history of colonisation and urbanisation, which influenced the village's oral history and made it a fruitful field for oral narrative research.

Omoljica is a village located within the municipality of Pančevo in South Banat on the bank of the Danube, in the Serbian Province Vojvodina. Serbian settlers have been registered since the 15<sup>th</sup> century. At the beginning of the 18<sup>th</sup> century, when the Hapsburgs seized all territories north of the rivers Danube and Sava, Omoljica was colonised by German settlers, with the German colonisation reaching its peak between 1765 and 1770. Soon after, Serbian and smaller numbers of Croatian families from different regions (1789–1791) and Romanians (1781–1805), mainly from today's Romanian Banat, were settled (on the colonisation, see Haag, 1938 and Pecinjački, 1985).

Nowadays, Omoljica is located within the multi-ethnic municipality of Pančevo. According to the last Serbian Population census in 2011, Omoljica had 6,309 inhabitants (Statistical Office of the Republic of Serbia, 2013). The majority are Serbs (90.66%; 5,720), with the minorities including Roma (1.77%; 112), Macedonians (1.2%; 81), Hungarians (1.2%; 76), Romanians (1.14%; 72), Bulgarians (0.33%; 21) and Germans (0.28%; 18). The village is surrounded by settlements with varying ethnic composition, e.g., Ivanovo (Hungarian and Bulgarian), Vojlovica (Hungarian, Serbian and Slovak), Skorenovac (Hungarian, Bulgarian and Serbian) and Starčevo (Serbian and Croatian).

The ethnological fieldwork on oral histories and local traditions in Omoljica has been conducted by Marija Mandić (Ilić) and colleagues between 2004 and 2008, mainly with the Serbs and Romanians. More than 50 interlocutors were interviewed in that period. For this project, 7 main narrators were selected (4 men and 3 women), all speakers of Serbian whose age ranges from 45 to 85. The collocutors were mainly

older people who keep the oral tradition alive and have all different sorts of narrative themes and genres in their narrative repertoires. Additionally, 11 narrators also took part in the conversations or even in the narration.

## 2.2. Processing steps

The sample for the corpus was compiled through the hand selection of personal narratives from DABI. A researcher with long-standing experience in narrative analysis selected parts of existing and not yet transcribed recordings containing sequences with at least the Labovian minimal narrative structure "A then B". In the next step, all selected recordings have been transcribed by eight trained linguists with EXMARaLDA (Schmidt & Wörner, 2014) using the HIAT transcription system (Ehlich & Rehbein, 1976). Each transcription has been checked by two correctors (inter-annotator agreement, however, has not been consistently measured) and then exported to a TEI XML file. GATE (Cunningham, Tablan, Roberts, & Bontcheva, 2013) was chosen as the annotation tool. Annotation schemas have been written for all annotation levels which will be discussed below. Annotation was executed automatically (POS), semi-automatically (pauses) and manually (all other). The corpus files have then been exported to stand-off GATE XML, which in turn has been converted to TEI XML file using eXist ([exist-db.org](http://exist-db.org)).

## 2.3. Annotation layers

The main aim of corpus design was to enable research into the selected narratives from the perspective of Construction Grammar (e.g., Fillmore, Kay, & O'Connor, 1988; Goldberg, 1995). As theoretical premise, the traditional approach of Labov and Waletzky (1967) was used. Their analysis of narratives was based on the identification of (narrative) clauses and led to the introduction of narrative components such as orientation, evaluation or complicating action (see section 3.3). The narrative components were annotated by hand. As it were, many narratives include other narratives that serve the function of one of the components. For example, a speaker can provide a short narrative instead of an orientation, indicating that the resulting situation that occurred as outcome of the (embedded) narrated events will serve as starting point for a new narrative. Therefore, the distinction between main narratives and embedded narratives with their respective components was made.

The construction grammar perspective on the other hand urged us to include information about the linguistic surface. Research on narratives revealed that, among others, tense, aspect, person, number, clause type (state vs. event) and discourse

---

elements are important information at the surface level (Polanyi, 1982). These have been coded on the level of morphosyntactic description (see section 3.2). Clause types were not annotated so far, although they are the main analytical vehicle for Labov and Waletzky. Instead, argument structure constructions (Goldberg, 1995) were annotated. The two notions are to some extent similar, although the notion of construction is wider than the definition of a clause (see section 3.4).

If narratives are regarded as cognitive routines, it is very likely that they also display some sort of frequency effects (Arnon & Snider, 2010). Since reaction time or similar measures cannot be produced for this kind of data, we focused on the fluency of the narrative production. If a narrative has been produced repeatedly or the event described in the narrative is recurring frequently, we would expect a higher degree of entrenchment (Schmid, 2017) indicated by the number of pauses and lapses during production. Therefore, a special layer, where unfilled pauses, lengthened sounds, repairs and repetitions are annotated is also included in the corpus.

Narratives are often not recognizable without a proper understanding of the context in which they have been produced (Polanyi, 1982). In order to cope with this, we annotated a wider context, which gives information as to the motivation for the narration, the chosen type of narrative and the like.

The annotation layers we included are thus as can be seen in Table 1:

**Table 1.** Levels of annotation

**Tablica 1.** Razine anotacije

No. / Br.	Name / Naziv	Description / Opis	Annotation / Anotacija
1	Narrative type	(Types of) narratives and their context(s) [Values: Personal, Collective, Hypothetical, Habitual, Communal, Implied, Anecdote, Independent context]	Manual
2	Narrative element	Components of the narratives according to Labov [Values: Abstract, Coda, Communicative Context, Complicating Action, Evaluation, Orientation, Resolution]	Manual
3	Embedded narrative element	Components of embedded narratives [Values: Abstract, Coda, Communicative Context, Complicating Action, Evaluation, Orientation, Resolution]	Manual

No. / Br.	Name / Naziv	Description / Opis	Annotation / Anotacija
4	Discourse constructions	Constructions that organise discourse [Value: List Construction]	Manual
5	Clause-level constructions	Argument structure constructions [Values: Transitive, Intransitive, Identification, Characterization, Existential, Localization, Intransitive Motion, Ditransitive, Caused Motion, Other]	Manual
6	Morpho-syntactic Description (MSD)	Syntactically relevant morphological categories [Values: MULTEXT-East-Tagset for Serbian]	Automatic
7	Pauses	Deferred speech [Values: Unfilled, Lengthened, Turn Taking, Repairs, Planning]	Automatic
8	Sentence	Sentence boundaries according to the transcription	Automatic
9	Token	Tokenization according to ANNIE tokenizer in GATE	Automatic

## 2.4. Corpus access

The complete data is available as XML files on the website [www.spokencorpus.eu/cms/cronus](http://www.spokencorpus.eu/cms/cronus). However, the investigation of narratives still requires a qualitative analysis, the main work of which is the comparison of different types of narratives or different components and their properties. In order to make this possible, CRONUS can be also queried online via a specially programmed search mask, available at <http://poincare.matf.bg.ac.rs/~andjelkaz/diwna/>. The DIWNA tool allows users to perform searches on the following levels:

- Narrative types
- Narrative components (embedded and unembedded)

Within narrative types and narrative components, users can search for:

- Morpho-syntactic Description (MSD)
- Clause-level constructions

The search mask consists of selection boxes or selection lists for all levels except the morphosyntactic one, so that a simple search is possible:



**Figure 1.** Home screen of the online tool DIWNA  
**Slika 1.** Početni ekran internetskoga alata DIWNA

The output can be provided as full text i.e., the entire sequence corresponding to a specific search query is output. This makes it possible to read complete narratives. A comparison can be made on all levels in a parallel window, compare Figure 2.

**Figure 2.** Example analysis: Comparison of transitive constructions in complication vs. orientation in personal narratives  
**Slika 2.** Analiza primjera: usporedba tranzitivnih konstrukcija u komplikaciji i orijentaciji osobnoga pripovijedanja

### 3. TRANSCRIPTION AND ANNOTATION

Transcription and annotation will be explained in brief. The annotation included morpho-syntactic features, narrative components and argument structure constructions.

#### 3.1. Transcription

The selected recordings were "diplomatically" transcribed by a trained linguists and checked by two correctors each. A diplomatic transcription uses, as far as possible, the writing system available in the respective language (regardless of whether it is a phonological, morphological or etymological orthography), but adapts the forms to the actual pronunciation. Typical forms for Serbian colloquial language are, for example, the shortening of the vowel combination *ao* in participles (*došo* instead of *došao* 'came') or other words (*ko* instead of *kao* 'like'). We used the system HIAT (Half-interpretative working transcription; Ehlich & Rehbein, 1976; Rehbein, Schmidt, Meyer, Watzke, & Herkenrath, 2004). It is only minimally intrusive and allows easy reading of the transcripts. HIAT provides several potential encodings. However, we have decided to proceed diplomatically here as well: we have included all the information that is potentially relevant for analysis but limited the information to the necessary. Specifically, the highlighted features are listed in Table 2:

**Table 2.** Elements of HIAT used in the transcription

**Tablica 2.** Elementi HIAT-a primijenjeni u transkripciji

Element / Element	Transcription / Transkripcija	Tag / Oznaka	Example / Primjer
Pauses	• or (Xs)	PAUSE	O:vaj • • uglavnom • • gde ((1.5s)) je bilo tako
Abortions	/	-	A imate i izbe/ i recentn/• najnovije izbeglice?
Interruptions	...	-	zavisi koja doba godišnje i ...
Elongations	:	-	Izmotavali se šalili se na svoj račun i na račun Nemaca.
Non-comprehensible parts	(inc)	-	Ali to su (inc) se desi
Guesses	()	-	nije (Bosa/) ((laughs))
Non-linguistic actions	((laughs)) ((coughs))	META	ɔ ((coughs)) Kad su digli sa tim šleperima

Punctuation marks are used to represent an illocutionary mode. To increase readability, we have decided to capitalize the beginning of the sentence according to Serbian orthography, as well as proper names. Abortions were also marked, given that they can sometimes provide insight into the thinking process, signalling confusion, change of mind or sometimes even mark a complete change in the course of the narrative. On the level of prosody, filled and unfilled pauses are noted. Apart from the verbal production and prosodic elements, the transcripts contain information about other vocal cues audible in the recordings, such as cough or laughter, as some of them can be very important in the narrative analysis (Goodwin, 2015: 199), often representing the attitude of the speaker. Information about overlaps is accessible using the time-points in EXMARaLDA.

We preferred diplomatic transcription over standard-oriented transcription. Since transcription is generally the bottleneck in the processing of spoken data, the work invested should be used as effectively as possible. Above all, most information from the audio file should be preserved. Since the characteristics of the spoken language, which can be reproduced from transcripts using diplomatic transcription, are recognised by the transcriber during the process and represent valuable information for the analysis, it would be a waste of resources not to record it if a standard-oriented transcription is used. Furthermore, the used HIAT system is relatively intuitive and easy to use even by less trained transcribers. There are further arguments for a diplomatic transcription: on one hand, there may be dialectal or colloquial morphological forms which have no trivial equivalent in the standard language. On the other hand, it is relatively easy to automatically annotate the standard forms for trivial correspondences by means of a lexicon and a tagger.

### 3.2. Token-level annotation

MULTEXT-East Tagset (Version 5.0, Erjavec, 2012) with its specifications for Serbian and Croatian was used for the annotation of parts of speech and morphological categories. MULTEXT-East is a tagset that focuses on syntactically relevant morphological categories and is well equipped to represent the morphological wealth of the Slavic languages. The Serbian/Croatian specifications have potentially more than a thousand combinations: the Serbian translation of George Orwell's '1984', which has been used as the basis for the development of MULTEXT-East specifications and lexicons (Erjavec et al., 2003), already contains 906 different morpho-syntactic descriptions (MSDs; Krstev, Vitas, & Erjavec, 2004). The resources provided by the ReLDI project, especially the Serbian lexicon srLex (version 1.0,



---

We have added some dialectal and colloquial forms to the dictionary srLex to increase the accuracy of the morphological annotation. This addendum can be downloaded on [www.spokencorpus.eu/cms/cronus](http://www.spokencorpus.eu/cms/cronus). Furthermore, the codifications specific to the HIAT transcription standard were added to the lexicon, as these elements are naturally not available within the MULTEXT-East specifications. In the example, this concerns the markings of the pauses (marked in HIAT with a dot •) and the non-linguistic or meta-linguistic events (such as *cough*).

### 3.3. Narratives

Starting point for the annotation of narratives and narrative components in particular was Labov's concept of the personal experience narrative (PEN) and the structural framework of PEN which consists of six components (Labov, 1976: 363–374; Labov & Waletzky, 2003[1967]: 93–102):

1. *Abstract*

Answers the question 'What is the story about?'; placed at the beginning of the narrative.

2. *Orientation*

Answers the question 'Who, where, when, what?'; provides background information relevant to the narrative.

3. *Complicating action*

Answers the question 'Then what happened?'; provides the plotline which refers to events following a chronological order.

4. *Evaluation*

Answers the question 'Why is this story worth telling?'; conveys the narrator's point of view.

5. *Resolution*

Answers the question 'How did it all end?'; usually follows or coincides with the evaluation.

6. *Coda*

Answers the question 'That's it?'; signals that the narrative is finished and switches from narrated time to time of narration.

According to this framework, the complicating action is technically the only obligatory nucleus of a personal narrative. It consists of so-called narrative clauses, which represent the chronological order of the narrated event and thus cannot change

---

order (Labov & Waletzky, 2003[1967]: 84–88). Free clauses, on the other hand, can occur more or less freely throughout the narrative, and can exchange places with other clauses without changing the narrative's original semantic interpretation (Labov & Waletzky, 2003[1967]: 89).

Although Labov pointed out that his oral narratives were obtained as a response to questions during an interview, he did not include interaction in the description of oral narration. Post-Labovian research later moved from the study of narrative as a closed text to the study of narrative-in-context (Georgakopoulou, 2006: 123). It has become clear that many narratives do not fit completely into Labov's model, e.g., autobiographies, reports, chronicles, habitual narratives, small stories or hypothetical narratives (Bamberg & Andrews, 2004; Bamberg & Georgakopoulou, 2008; Georgakopoulou, 2006, 2007; Johnstone, 1993). Nevertheless, Labov's model has been applied widely in the analysis of different cultural communities and narrative genres (Bell, 1991; de Fina, 2003; Holmes, 1998; Linde, 1993; Polanyi, 1979, 1985).

In the case of interview discourse, which is also the main type of data in CRONUS so far, we come across dialogical narratives elicited for the purpose of research. Most often, there is a main narrator, while others are contributing to the storytelling: researchers ask questions or encourage the narrator, other participants supplement the story, emphasise parts of it, approve, object or repeat the words of the main narrator.

The narrative components were annotated in teams of two so that possible uncertainties and discrepancies could be discussed immediately. The annotation of narrative components is not an easy matter. The Inter-Annotator Agreement is reported to be relatively low and there are many cases where a decision is difficult to make (Swanson et al., 2014). The annotation employed for narratives and narrative structures, thus, does not aim at producing an essentialism based on a theory that has been criticised for its limitations and has already been further developed. Rather, it is believed that the theory of Labov and Waletzky is a good heuristic starting point for a more comprehensive description of the Serbian narratives. The annotation is not fixed but will change in accordance with the feedback from the analysis itself. The current annotation thus uses more or less intuitively accessible prototypes for both narrative types and narrative components.

From a structural point of view, principal narratives (frame stories) and embedded narratives, which are part of a bigger frame narrative and serve as one of its structural components, were differentiated. By using two criteria – the structure of

---

the narrative and its social role – different narrative genres were distinguished, namely personal experience narratives, anecdotes, collective narratives, communal narratives, habitual narratives, and hypothetical narratives (see Table 4 for the distribution of narratives in the corpus according to genres).

**Table 4.** Types of narratives in the corpus and their numbers

**Tablica 4.** Vrste pripovijedanja i njihova zastupljenost u korpusu

Narrative genre / Vrsta pripovijedanja	Number / Broj	Narrative genre / Vrsta pripovijedanja	Number / Broj
Personal experience narratives	29	Collective narratives	11
Anecdotes	6	Communal narratives	8
Hypothetical narratives	2	Habitual narratives	23

*Personal experience* narratives (PEN) were defined by Labov and Waletzky (1967: 81) as a verbal technique for recapitulating an individual experience of the narrator or other people.

*Anecdotes* represent a short humorous narrative form which refers to an individual person and/or an incident. This genre is usually described as transitory between a biographical account and a fable (Blache, 1999; Hranjec, 1990; Ilić [Mandić], 2007). As such, it is usually first told as PEN, then it takes over some established folklore motifs and merges them with PEN, and eventually becomes part of the community narrative repertoire. Although anecdotes differ from PEN in their social role, they use the same narrative structure.

In this spoken corpus, several narrative genres refer to collective experience. Collective, communal and habitual narratives are distinguished. *Collective* narratives are based upon collective cultural memory and belong to the community's narrative repertoire (Ilić [Mandić], 2014a). When referring to collective agency, they play an important social role in the negotiation of the group's identity. *Communal* narratives also employ collective agency and refer to events which are believed to have some importance for the given community. However, their social role is limited: it is not sure whether they are widespread, if they are going to survive the generational memory, and become part of cultural memory. Canefe (2004) argues that 'communal past occupies an otherwise overlooked grey area between individual recollection and

communal or national reconstruction of history'. *Habitual* narratives refer to iterative cultural and social practices like customs (Ilić [Mandić], 2014b: 257–259). They often employ shifting agency in one narrative, depending on the distance which the narrator takes to recalled events, i.e., whether he/she participated in, observed or just heard about a cultural script.

*Hypothetical* narratives were also identified in the material. They are narratives of projected events, placed in the realm of possibility, and are structurally close to some types of small stories as described by Georgakopoulou (2006, 2007).

### 3.4. Constructions

The integration of an annotation level for constructions originates in the purpose of the project in which the corpus has been developed. The project focuses on the possibility to describe narratives as complex constructions. Although from the perspective of Construction Grammar, constructions exist on various levels of linguistic description (and narrative components and narratives itself are also constructions), this level of annotation focuses on clause-level constructions. The rationale behind this is to test whether a construction-based analysis can replace the traditional Labovian approach which focuses on narrative clauses. The advantage of constructions lies in the fact that they are more flexible than the conception of clauses. In contrast to a clause-based analysis, they allow for basic narrative elements that do not contain verbs and hence would not bear the status of a clause. Whether this approach is useful still has to be tested.

On the clause level, argument structure constructions and constructions with nominal predicates were annotated. On the level of the discourse, the list construction [X *i* Y; 'X and (then) Y'] has been tagged, which is important for narrative analysis since it often frames the complicating action. During the ongoing research, there may be other discourse-level constructions that will be added. Constructions were annotated by hand by one researcher and checked by a second. However, we annotated only the most frequent constructions as described in (Wasserscheidt, 2016: 191). These are the following, here shortly explicated with the semantic roles involved:

Transitive [Agent Action Patient], Intransitive [Agent Action], Identification [Theme COPULA Identity], Characterization [Theme COPULA Property], Existential [BE Theme], Localization [Place COPULA Theme], Intransitive

---



Motion [Agent Action Path], Ditransitive [Agent Action Benefactive Patient], and Caused Motion [Agent Action Patient Path].

Problematic issues for the annotation of sentential constructions are well known and involve mainly the independence of constructions in relation to similar ones. The question is thus whether a shortened or stretched construct still counts as an exemplar of the construction or whether it should already be regarded as an independent construction. For the annotation, we resorted to a verb centred solution: if the missing argument is a frequent and thus expectable argument of the verb, the full construction was chosen (as in (2) below). If not, another construction was selected.

- (2) Ovaj prodaje ovo • Ovaj kupuje • ne zna-m.  
 This sells this This buys NEG know-1SG.  
 [<sub>TRANSITIVE</sub> This one sells this.] [<sub>TRANSITIVE</sub> That one buys ... [I don't know].]

While the first utterance in (2) is clearly a transitive construction, the second lacks a patient. However, since the speaker obviously planned to add an example of goods that have been bought, but cannot come up with one, and since the verb *kupovati* 'buy' is predominantly transitive, we categorized the construct as transitive.

Passive constructions that might be seen as transformed transitive or ditransitive constructions have been annotated as *Other*. There are also a lot of 'spoken' constructions and uses of single NPs in order to answer question (3) or re-emphasize elements (4). All of them have been categorized as *Other*, as well.

- (3) Tri lanca zemlje.  
 [<sub>OTHER</sub> Three hides land.]
- (4) Pet godina • • • dečaćić u kapi • a pored lika petokraka ovolika.  
 Five years boychild in cap and beneath picture fivestar such  
 [<sub>OTHER</sub>(He lived) Five years.] [<sub>OTHER</sub>A boy with a cap.] [<sub>OTHER</sub>And next to the picture a five-pointed star, such a big one.]

In particular, the constructions typical of spoken language deserve closer examination in the future.

#### 4. CASE STUDIES

Three short case studies will show how the corpus can be used, how a corpus-based analysis can generate new discoveries about narratives and, finally, how the use of the Labovian categorization can provide interesting insights.

#### 4.1. The use of person over narrative types and components

The different components of narratives in general are expected to have different communicative functions. For example, the orientation is expected to set up a common ground between the narrator and the listener, while the purpose of the evaluation is to underline the significance of the narrative and to justify the discourse time used by the speaker to produce the narrative (Fludernik, 2014: 97). These functions likely correlate with component-specific morphological features. These features are partly necessary in order to code the function of the component and to allow the listener a proper interpretation. For example, both orientation and evaluation build a communicative relation between speaker and listener in that the speaker anticipates the listener's background knowledge and evaluates whether the establishment of the intended common ground was successful. This is evident from the use of the second person singular or plural (e.g., *Znate?* 'Do you know?') which for the listener serves as cue to interpret the information provided next to this element as background information and to add it to the mental space of the narrated event. On the other hand, the components' features also depend on their reference to the situations and events narrated. In this vein, it is expected that, for example, personal experience narratives contain more singular (first or third person) verb forms than collective narratives. Person, therefore, is a point of interest for research on the narrative corpus. Person is encoded within the MULTEXT-East tags for verbs. Serbian has three persons (first, second, third) in two numbers (singular, plural).

One way to look at the different narrative components as well as the different types of narratives from the perspective of the category person, is to consider the role of the narrator. Their role can be both the agent of the narration, as well as the narrator or commentator. If the narrator is also the protagonist, the narratives are expected to occur in the first person. If the narrator is only the storyteller, as in collective narratives, reference to their own attitude towards the story will most likely happen in the evaluation, if anything. In order to test this, the three most frequent narrative genres were selected, all verbs were counted – or, in the case of analytic forms, auxiliaries – occurring in first person singular and compared this figure to the overall number of verbs.

---

**Table 5.** Use of 1. person in narrative elements in personal, collective and habitual narratives (ratio between 1SG and all verbs, figures above 0.200 marked bold)

**Tablica 5.** Uporaba prvoga lica u pripovjednim elementima u osobnim, kolektivnim i habitualnim pripovijedanjima (omjer broja prvoga lica i ukupnoga broja glagola, vrijednosti iznad 0,2 su podebljane)

Narrative type / Vrsta pripovijedanja	Orientation / Orijentacija	Complication / Komplikacija	Evaluation / Evaluacija
Personal	0.164 (27/165)	<b>0.251</b> (67/167)	<b>0.261</b> (43/165)
Collective	0.096 (12/125)	0.030 (2/67)	<b>0.224</b> (11/49)
Habitual	0.100 (9/90)	<b>0.212</b> (32/151)	0.132 (9/68)

The result indicates that there is a significant difference between the overall use of the first person singular in the three narrative types – personal, collective and habitual ( $\chi^2 = 18.9956$ ,  $p = 0.000075$ ). There is also a significant difference between the overall use of the first person singular in the three components – orientation, complicating action and evaluation ( $\chi^2 = 13.1857$ ,  $p = 0.00137$ ). First person singular is thus unusual in the complicating action of collective narratives and in the orientation of all types of narratives. As expected, first person is not used in the complication of collective narratives, but figures in the evaluation of collective stories. An interesting difference between collective and habitual narratives is noted; while the former presents events that usually not happened to the narrator, first person is used in evaluation exactly as frequent as expected in general (the overall expected ratio of 1Sg vs. all main verbs in evaluation is 0.223). On the other hand, habitual narratives use first person singular in their complicating action, but not in the evaluation.

#### 4.2. Constructions in different narrative components

The different functions of the components are also likely to influence the kind of argument structure used in them. Naturally, the complicating action is expected to show mainly argument structures that express actions or motion. The orientation, on the other hand, has the function to give background information, which often

provides some kind of categorization. It was, therefore, expected that the orientation has more nominal predicates. In order to see what types of constructions (see section 3.4) were used in the narrative elements, their distribution within each individual narrative element in relation to the other constructions was analysed and then compared the amount of usage of the individual constructions across the narrative elements.

**Table 6.** Distribution of constructions within the individual narrative elements (in percent, figures above 14.00 marked bold)

**Tablica 6.** Raspodjela konstrukcija u pojedinačnim pripovjednim elementima (u postotcima, vrijednosti iznad 14 % su podebljane)

Construction / Konstrukcija	Abstract / Abstrakt	Orient / Orijentacija	Complic / Komplikacija	Eval / Evaluacija	Resol / Rasplet	Coda / Koda
Transitive	<b>27.09</b>	<b>21.97</b>	<b>30.56</b>	<b>27.35</b>	<b>26.59</b>	<b>23.48</b>
Ditransitive	2.96	1.32	2.73	2.60	2.31	2.61
Intransitive	<b>18.72</b>	11.28	<b>16.05</b>	12.59	<b>20.81</b>	<b>17.39</b>
Intransitive motion	3.45	3.48	7.61	2.17	6.94	1.74
Caused motion	0.49	1.08	2.14	0.87	0.58	1.74
Localization	3.45	4.32	1.55	1.59	2.31	1.74
Identification	3.94	9.12	3.09	6.22	7.51	7.83
Characterization	4.43	8.88	2.97	9.84	3.47	4.35
List	4.93	11.16	<b>14.03</b>	11.29	5.78	5.22
Existential	<b>15.27</b>	6.84	2.02	5.21	6.94	<b>15.65</b>
Other	<b>15.27</b>	<b>20.53</b>	<b>17.24</b>	<b>20.26</b>	<b>16.76</b>	<b>18.26</b>
<i>Sum:</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>100</i>

Table 6 shows the percentage distribution of the specific construction types in the individual narrative elements within this corpus. For example, the abstract contains 27.09% transitive, 2.96% ditransitive, 18.72% intransitive constructions, etc. The highest percentage of annotated constructions for all narrative elements are transitive constructions, followed by either intransitive constructions, or those that

have been marked as *Other*. However, there is some variation in the distribution of constructions across narrative components. As expected, there is a larger percentage of "stative" characterization, identification and localization constructions in the narrative element orientation (in sum 22.32%) as compared to the other narrative components, which points to the description of the contextual setting in terms of person, location etc. The difference is even more notable, if we look at the distribution of the individual constructions within the narrative elements in Table 7. The table indicates that, for example, 7.27% of all transitive constructions occur in the abstract, 24.17 in the orientation, 33.95 in the complication and so on.

**Table 7.** Distribution of individual constructions across narrative elements (in percent, figures above 30.00 marked bold)

**Tablica 7.** Raspodjela konstrukcija prema pripovjednim elementima (u postotcima, vrijednosti iznad 30 % su podebljane)

Construction / Konstrukcija	Abstract / Abstrakt	Orient / Orijentacija	Complic / Komplikacija	Eval / Evaluacija	Resol / Rasplet	Coda / Koda	Sum / Zbroj
Transitive	7.27	24.17	<b>33.95</b>	24.97	6.08	3.57	100
Ditransitive	9.23	16.92	<b>35.38</b>	27.69	6.15	4.62	100
Intransitive	9.27	22.93	<b>32.93</b>	21.22	8.78	4.88	100
Intransitive motion	5.43	22.48	<b>49.61</b>	11.63	9.30	1.55	100
Caused motion	2.70	24.32	<b>48.65</b>	16.22	2.70	5.41	100
Localization	9.59	<b>49.32</b>	17.81	15.07	5.48	2.74	100
Identification	4.57	<b>43.43</b>	14.86	24.57	7.43	5.14	100
Characterization	4.81	<b>39.57</b>	13.37	<b>36.36</b>	3.21	2.67	100
List	3.17	29.52	<b>37.46</b>	24.76	3.17	1.90	100
Existential	18.13	<b>33.33</b>	9.94	21.05	7.02	10.53	100
Other	5.77	<b>31.84</b>	27.00	26.07	5.40	3.91	100

It is shown that no less than 43.43% of all identification and 49.32% of all localization constructions used in the narratives can be found in the orientation. Constructions predominantly denoting some kind of action, on the other hand, can

be found mostly in the complication. This component consists by 59.09% of the first five "non-static" constructions in Table 6. These numbers are largely due to transitive and intransitive constructions. However, for the intransitive motion and the caused motion construction, this means that almost half of all occurrences are used in the complication. The orientation, on the other hand, consists by only 39.13% of these construction types.

Another example of meaningful variation between the narrative components are list constructions. Here, the highest percentage was also found in the complication. List constructions function on the discourse level, where the listing of successive events is connected by the conjunction *i* 'and' and serves to mark the temporal sequencing of the events. Interesting is furthermore the comparatively high number of existential constructions in both the abstract and the coda (see Table 6) – which, in a sense, indicate the situation in the initial and final situations and frame the narrative action.

### 4.3. Tense in different narrative genres

Since the identified types or genres of narratives are used to communicate events that have different relations to the time of narration, it is expected that they also differ in the tenses they use. Collective and communal, but also personal narratives refer exclusively to past events and are likely to display a higher use of past tense. It is, however, a common feature to all European languages that past events can also be told in present tense – which is called narrative or historic present. This is most characteristic for anecdotes but may also be used in other narrative genres.

Serbian has a rich system of verb tenses: one tense for present, two for future (I and II), and four for past – perfect, pluperfect, aorist and imperfect. The last three are not used often in daily speech, especially the imperfect (there are, however, other Serbian dialects that use aorist regularly, see Veljović, 2015). The Serbian perfect tense covers the functions of all the English past tenses and is the most commonly employed past tense in general. The aorist refers to terminated and/or completed actions which usually happened immediately prior to the moment in which they are described.

We analysed the use of the tenses in the components complication and resolution in different non-embedded narrative genres. The most commonly used tenses over all genres happen to be present (50.73%) and perfect (47.61%).

---

**Table 8.** Use of verb tenses in different narrative genres (in percent, figures above 50.00 are marked bold)

**Tablica 8.** Uporaba glagolskih vremena u različitim pripovjednim vrstama (u postotcima, vrijednosti iznad 50 % su podebljane)

Tense / Vrijeme	Collective / Kolektivni	Communal / Komunalni	Habitual / Habitualni	Anecdote / Anegdota	Personal / Osobni
n	54	37	148	36	206
Perfect	<b>77.78</b>	<b>83.78</b>	7.43	33.33	<b>64.56</b>
Present	22.22	16.22	<b>91.89</b>	<b>66.67</b>	32.04
Aorist	0	0	0	0	3.40
Future	0	0	0.68	0	0

The past tense dominates in collective, communal and personal experience narratives. The relatively high number of present and aorist verbal forms in the personal experience narratives compared to collective and communal narratives can be explained by the use of the narrative (historical) present and by the use of direct or indirect speech, where verbs denoting the act of speaking (*I say, he says* etc.) are either in the narrative present or aorist. In contrast to these genres, habitual narratives almost exclusively employ present tense forms. Anecdotes are an interesting case: although based on stories which happened in the past, they preferably use narrative present.

## 5. CONCLUSION

This paper describes the first attempt to develop a corpus of spoken narratives that uses the Labovian structural model and enables analysis with a deep annotation. The work is not yet finished, and improvements have to be made at all levels (transcription, annotation and representation). Nevertheless, we were able to show that this resource can be used to achieve valuable results and generate new insights.

Specifically, the combined analysis of morpho-syntax, syntactic constructions and narrative components can be used to reveal various characteristics of narratives that a) show the function of the individual narrative components and b) enable us to describe individual narrative genres in more detail. The corpus could thus potentially help to close the gap between the structure of narratives that Labov postulated and that has been applied in many analyses and the empirical difficulty of identifying them.

It is planned to enlarge the database from the hitherto modest size. Above all, more narrators are to be involved in order to be able to carry out meaningful, supra-individual analyses. The corpus will also contain additional annotation levels, such as syntactic functions and information structure. One of these additional layers might include prosodic features such as intonation, stress, tempo etc. Between narrative clauses, we often find embedded orientation and other comments to the story that are clearly set off because they belong to separate intonational levels (Fludernik, 2002: 45). Intonation is also used to differentiate between the story telling and its enclosing activities (Selting, 1992). A frame-semantic annotation also renders useful, since narratives essentially build on (individual, collective or general) knowledge.

## ACKNOWLEDGEMENTS

The corpus has been developed as part of the project funded by the German Academic Exchange Service (DAAD) with funding from the German Ministry of Foreign Affairs and the Serbian Ministry of Education, Science and Technological Development, within the programme for bilateral cooperation PPP. The paper is also result of the project (No. 178010) Language, Folklore, Migrations in the Balkans, funded by the Serbian Ministry of Education, Science and Technological Development.

We are indebted to three anonymous reviewers for their valuable comments. We remain responsible, of course, for any shortcomings of the present paper.

## REFERENCES

- Arnon, I., & Snider, N.** (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67–82.
- Bala, M.** (2010). Tamo gde se kulture susreću: naracije o sopstvu kroz uticaj različitosti. Životne priče mađarske dijaspore u Vojlovici [Where cultures meet: Narratives of the self through the influence of diversity. Life stories of the Hungarian diaspora in Vojlovica]. In L. Subotić, & I. Živančević-Sekeruš (Eds.), *Susret kultura: knjiga I [Encounter of cultures: Book I]* (pp. 395–402). Novi Sad: Filozofski fakultet.
- Bala, M.** (2015). Diskurzivna konstrukcija etničkog identiteta u usmenim autobiografijama bukovinskih Mađara [Discursive construction of ethnic identity in oral autobiographies of Bukovinian Hungarians]. In V. Lopičić, & B. Mišić Ilić (Eds.), *Jezik, književnost, diskurs: književna istraživanja [Language,*
-



- literature, discourse: Literary research*] (pp. 153–162). Niš: Filozofski fakultet Univerziteta u Nišu.
- Bamberg, M.** (2004). Narrative discourse and identities. In J. C. Meister, T. Kindt, W. Schernus, & M. Stein (Eds.), *Narratology beyond literary criticism* (pp. 213–237). Berlin, New York: Walter de Gruyter. Retrieved from <http://www.clarku.edu>
- Bamberg, M., & Andrews, M.** (2004). *Considering counter-narratives: Narrating, resisting, making sense. Studies in narrative* (Vol. 4). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bamberg, M., & Georgakopoulou, A.** (2008). Small stories as a new perspective in narrative and identity analysis. *Text & Talk*, 28(3), 377–396.
- Bell, A.** (1991). *The language of news media. Language in society* (Vol. 17). Oxford: Blackwell.
- Blache, M.** (1999). The anecdote as a symbolic expression of the social and cultural milieu of journalists. *Folklore*, 110, 49–55.
- Bruner, J.** (1991). The narrative construction of reality. *Critical Inquiry*, 18, 1–21.
- Canefe, N.** (2004). Communal memory and Turkish Cypriot national history: Missing links. In M. N. Todorova (Ed.), *Balkan identities: Nation and memory* (pp. 77–102). New York, NY: New York Univ. Press.
- Chambers, N., & Jurafsky, D.** (2008). Unsupervised learning of narrative event chains. *Proceedings of ACL/HLT 2008*.
- Chambers, N., & Jurafsky, D.** (2010). A database of narrative schemas. *Proceedings of LREC-2010*, Malta.
- Cunningham, H., Tablan, V., Roberts, A., & Bontcheva, K.** (2013). Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS Computational Biology*, 9(2). <https://doi.org/10.1371/journal.pcbi.1002854>
- de Fina, A.** (2003). *Identity in narrative: A study of immigrant discourse. Studies in narrative* (Vol. 3). Amsterdam: Benjamins.
- Ehlich, K., & Rehbein, J.** (1976). Halbinterpretative Arbeitstranskriptionen (HIAT). *Linguistische Berichte*, 45, 21–41.
- Erjavec, T.** (2012). MULTEXT-East: Morphosyntactic resources for Central and Eastern European languages. *Language Resources and Evaluation*, 46(1), 131–142.
-

- Erjavec, T., Krstev, C., Petkevič, V., Simov, K., Tadić, M., & Vitas, D.** (2003). The MULTEXT-East: Morphosyntactic specifications for Slavic languages. *Proceedings of the 2003 EACL Workshop on Morphological Processing of Slavic Languages*.
- Fillmore, C. J., Kay, P., & O'Connor, C.** (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64, 501–538.
- Fludernik, M.** (2002). *Towards a 'natural' narratology* (1<sup>st</sup> ed.). Abingdon, Oxon: Taylor & Francis.
- Fludernik, M.** (2014). Conversational narration – oral narration. In P. Hühn, J. C. Meister, J. Pier, & W. Schmid (Eds.), *Handbook of narratology* (pp. 93–104). Berlin, Boston: De Gruyter.
- Georgakopoulou, A.** (2006). Thinking big with small stories in narrative and identity analysis. *Narrative Inquiry*, 16, 129–137.
- Georgakopoulou, A.** (2007). *Small stories, interaction and identities. Studies in narrative 8*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Goldberg, A. E.** (1995). *Constructions: A construction grammar approach to argument structure. Cognitive theory of language and culture*. Chicago etc.: The University of Chicago Press.
- Goodwin, C.** (2015). Narrative as talk-in-interaction. In A. de Fina, & A. Georgakopoulou (Eds.), *The handbook of narrative analysis* (pp. 197–218). West Sussex: Wiley-Blackwell.
- Haag, R.** (1938). *Ortsgeschichte von Omoljica (Homolitz). Grossgemeinde in der Woiwodina (Südbanat) des Königreiches Jugoslawien 1766-1938*. Omoljica.
- Holmes, J.** (1998). Narrative structure: Some contrasts between Maori and Pakeha story-telling. *Multilingua. Journal of Cross-Cultural and Interlanguage Communication*, 17(1), 25–58. <https://doi.org/10.1515/mult>
- Hranjec, S.** (1990). Semantičko polje usmene anegdote [The semantic field of the oral anecdote]. In *Rad XXXVII. kongresa Saveza udruženja folklorista Jugoslavije [Work of the XXXVII. Congress of the Association of Folklorists of Yugoslavia]*. Plitvička jezera.
- Ilić [Mandić], M.** (2007). Folklorna anegdota: između individualnog i kolektivnog – terenska istraživanja Srba u Medini (Mađarska) [Folklore anecdote: Between individual and collective – Field research of Serbs in Medina (Hungary)]. *Glasnik Etnografskog instituta SANU*, 55(1), 217–231.
-

- 
- Ilić [Mandić], M. (2014a). Collective narrative: The narrative on Croatian language from academic to far-right discourses in Serbia. *Journal of Comparative Research in Anthropology and Sociology*, 5(1), 49–73.
- Ilić [Mandić], M. (2014b). *Discourse and ethnic identity. The case of the Serbs from Hungary. Studies on Language and Culture in Central and Eastern Europe (Slccee)* (Vol. 23). München – Berlin: Verlag Otto Sagner.
- Ilić [Mandić], M., & Đurić Milovanović, A. (2012). Digital archive of the Banat vernaculars and culture: Fieldwork and perspectives. In *Proceedings of the Regional Conference "Research, Preservation and Presentation of Banat Heritage: Current State and Long Term Strategy* (pp. 151–155). Vršac: Gradski muzej Vršac.
- Johnstone, B. (1993). Community and contest. Midwestern men and women creating their worlds in conversational storytelling. In D. Tannen (Ed.), *Oxford studies in sociolinguistics. Gender and conversational interaction* (pp. 62–80). Oxford University Press.
- Krstev, C., Vitas, D., & Erjavec, T. (2004). MULTEXT-East resources for Serbian. In *Proceedings of the 7<sup>th</sup> International Multi-Conference INFORMATION SOCIETY IS 2004: Volume B: Language Technologies*. Jožef Stefan Institute, Ljubljana, Slovenia.
- Labov, W. (1976). *Language in the inner city: Studies in the Black English vernacular* (2<sup>nd</sup> ed.). Oxford: Basil Blackwell (Original work published 1972).
- Labov, W., & Waletzky, J. (2003[1967]). Narrative analysis: Oral versions of personal experience. In C. B. Paulston, & G. R. Tucker (Eds.), *Sociolinguistics. The essential readings* (pp. 74–104). Blackwell Publishing. (Source: Narrative Analysis: Oral Versions of Personal Experience. In J. Helm (Ed.), *Essays on the verbal and visual arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. Seattle: University of Washington Press. 1967, 12–44.).
- Linde, C. (1993). *Life stories: The creation of coherence. Oxford studies in sociolinguistics*. New York: Oxford University Press.
- Ljubešić, N., Klubička, F., Agić, Ž., & Jazbec, I.-P. (2016). New inflectional lexicons and training corpora for improved morphosyntactic annotation of Croatian and Serbian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia.
- Pecinjački, S. (1985). *Pančevački district 1717–1773 [Pančevo district 1717–1773]*. Novi Sad: Filozofski fakultet, Institut za istoriju.
-

- Peterson, C., & McCabe, A.** (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum Press.
- Polanyi, L.** (1979). So what's the point? *Semiotica*, 25(3/4), 207–241.
- Polanyi, L.** (1982). Linguistic and social constraints on storytelling. *Journal of Pragmatics*, 6, 509–524.
- Polanyi, L.** (1985). *Telling the American story: A structural and cultural analysis of conversational storytelling*. *Language and being*. Norwood, NJ: Ablex Publishers Corp.
- Prabhakaran, V., Hamilton, W., McFarland, D., & Jurafsky, D.** (2016). Predicting the rise and fall of scientific topics from trends in their rhetorical framing. *Proceedings of the 54<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL)*. Berlin, Germany.
- Rehbein, J., Schmidt, T., Meyer, B., Watzke, F., & Herkenrath, A.** (2004). Handbuch für das computergestützte Transkribieren nach HIAT. *Arbeiten zur Mehrsprachigkeit*, Folge B (56).
- Rühlemann, C., & O'Donnell, M. B.** (2012). Introducing a corpus of conversational stories. Construction and annotation of the Narrative Corpus. *Corpus Linguistics and Linguistic Theory*, 8(2), 313–350. <https://doi.org/10.1515/cllt-2012-0015>
- Savova, G. K., Chapman, W. W., Zheng, J., & Crowley, R. S.** (2011). Anaphoric relations in the clinical narrative: Corpus creation. *Journal of the American Medical Informatics Association: JAMIA*, 18(4), 459–465. <https://doi.org/10.1136/amiajnl-2011-000108>
- Schmid, H.-J.** (2017). *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. *Language and the human lifespan series* (1<sup>st</sup> ed.). Berlin, Washington, DC: de Gruyter Mouton; American Psychological Association.
- Schmidt, T., & Wörner, K.** (2014). EXMARaLDA. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 402–419). Oxford University Press.
- Selting, M.** (1992). Intonation as a contextualization device. In P. Auer, & A. Di Luzio (Eds.), *Pragmatics & Beyond New Series. The Contextualization of Language* (Vol. 22, p. 233). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.22.16sel>
- Sikimić, B.** (2012). Timski terenski rad Balkanološkog instituta SANU. Razvoj istraživačkih ciljeva i metoda [Team field work of the Balkan Institute of SASA.
-

- Development of research goals and methods]. In M. Ivanović Barišić (Ed.), *Terenska istraživanja – poetika susreta* [*Field research – The poetics of encounters*] (pp. 167–198). Belgrade: Etnografski institut SANU.
- Sikimić, B., Hristov, P., & Golubović, B.** (Eds.) (2012). *Labour migrations in the Balkans. Studies on language and culture in Central and Eastern Europe* (Vol. 18). München: Sagner.
- Sikimić, B., & Nomaći, M.** (2016). Jezički pejzaž memorijalnog prostora višejezičnih zajednica: banatski Bugari Palčani u Srbiji [Linguistic landscape of the memorial space of multilingual communities: Banat Bulgarians Palčani in Serbia]. *Južnoslovenski filolog*, 72(1–2), 7–31.
- Sorescu Marinković, A.** (2016). Contemporary Romanian migrations to Serbia: Stages, actors, reasons. *Balkanica Posnaniensia*, 23, 151–165.
- Statistical Office of the Republic of Serbia. (2013). *Population census 2011. Census of population, households and dwellings in the Republic of Serbia. Religion, mother tongue and ethnicity. Data by municipalities and cities*. Belgrade: Statistical Office of the Republic of Serbia.
- Swanson, R., Rahimtoroghi, E., Corcoran, T., & Walker, M.** (2014). Identifying narrative clause types in personal stories. In K. Georgila, M. Stone, H. Hastie, & A. Nenkova (Eds.), *Proceedings of the 15<sup>th</sup> Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)* (pp. 171–180). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-4323>
- Veljović, B.** (2015). Oblici aorista u govoru Tutina i okoline [Forms of aorists in the speech of Tutin and its surroundings]. In *Srpski jezik, književnost, umetnost: zbornik radova sa IX. međunarodnog naučnog skupa, održanog 24. i 25. 10. 2014. na Filološkoumetničkom fakultetu Univerziteta u Kragujevcu* [*Serbian Language, Literature, Art: Proceedings of the IX International Scientific Conference, held on 24 and 25 October 2014 at the Faculty of Philology and Arts, University of Kragujevac*] (pp. 67–83). Kragujevac: Filološko-umetnički fakultet.
- Vučković, M.** (2007). Drugi u diskursu Marka Gurana [Others in the discourse of Mark Guran]. In P. Asenova et al. (Eds.), *Bălgarskite ostrovi na Balkanite* [*Bulgarian islands on the Balkans*] (pp. 205–218). Sofia: Figura.
- Wasserscheidt, P.** (2010). Mehrsprachigkeit und Sprachwandel bei Ungarischsprechern im serbischen Banat. In C. Voß (Ed.), *Ottoman and Habsburg legacies in the Balkans: Language and religion to the North and to the South of the Danube River* (pp. 197–246). München, Berlin: Sagner.
- Wasserscheidt, P.** (2016). *Bilinguales Sprechen: Ein konstruktionsgrammatischer Ansatz*. Berlin: Freie Universität Berlin.

**Philipp Wasserscheidt**<sup>1</sup>, **Marija Mandić**<sup>5</sup>

*philipp.wasserscheidt@hu-berlin.de, marija.mandic@bi.sanu.ac.rs*

**Nadine Vollstädt**<sup>1</sup>, **Ana Jovanović**<sup>2</sup>, **Ivana Tanasijević**<sup>3</sup>

*nadine.vollstaedt@googlemail.com, anajovano@gmail.com, ivana@matf.bg.ac.rs*

**Teodora Vuković**<sup>4</sup>, **Ivana Vučina Simović**<sup>2</sup>, **Uliana Yazhinova**<sup>1</sup>

*bravethea@gmail.com, ivanavusim@gmail.com, yazhinovaul@gmail.com*

**Anđelka Zečević**<sup>3</sup>

*andjelkaz@matf.bg.ac.rs*

<sup>1</sup> Odsjek za slavistiku i hungarologiju, Humboldtovo sveučilište u Berlinu  
Njemačka

<sup>2</sup> Filološki fakultet, Sveučilište u Beogradu, Srbija

<sup>3</sup> Matematički fakultet, Sveučilište u Beogradu, Srbija

<sup>4</sup> Odsjek za slavistiku, Sveučilište u Zürichu, Švicarska

<sup>5</sup> Balkanološki institut, Srpska akademija znanosti i umjetnosti, Srbija

## **Korpusna analiza govornoga pripovijedanja. Prikaz korpusa i alata za pretragu**

### **Sažetak**

Rad se bavi razvojem sinkronoga korpusa, koji sadrži govorno pripovijedanje na srpskome jeziku, te uporabom toga korpusa za pripovijednu analizu. Korpus (CRONUS – Korpus za istraživanje pripovijedanja i njegove uporabe u govoru) je optimiziran za proučavanje strukture i uporabe žanra usmenoga pripovijedanja. U radu su predstavljeni izvori podataka te stvaranje korpusa i pristup korpusu. Usmeno pripovijedanje iz istraživačkih intervjua je transkribirano, a korpus duboko anotiran, s osobitim naglaskom na anotaciju pripovjednih dijelova prema Labovu te prema konstrukcijama argumentnih struktura Konstrukcijske gramatike. Tri studije slučaja pokazuju kako se morfološke i konstrukcijske anotacije mogu učiniti plodnima za istraživanje govornoga pripovijedanja.

**Ključne riječi:** govorno pripovijedanje, korpus, govoreni jezik, srpski jezik, dubinska anotacija