

Bulgarian Dialectology as Living Tradition: A Digital Resource of Dialect Speech

Ronelle Alexander

University of California, Berkeley

Linguists need data in order to perform analysis. Most data that are the base of linguistic analyses come from standard languages. This is partly because of the often unconscious thinking that a “language” is equivalent, both ideally and in actual practice, to the standard codified form of that language that is taught in schools, used by the media, and presented through belletristic literature and other means as the best and most expressive way to speak and write the language in question. More prosaically, it is also because the data of standard languages are the most accessible. Detailed reference works (grammars and dictionaries) are (generally) readily available, as are (at least if the linguist is working on his or her own language) native speakers.

Dialectal data, by contrast, are often – consciously or not – discounted as archaic, inferior or backward. Even those who are not hampered by this prejudice and fully recognize the immense riches provided by dialectal data, for both diachronic and typological studies, are faced with a block, in that dialectal data are simply much harder to come by, at least in the fully systematic and comparable way needed by the linguist who will perform a particular analysis. Only printed materials are available to the general public, and even the dialectologist who has made any one actual field recording has only that recording: that is, s/he lacks the ability to return to the informant to set up listening tests and elicit speaker judgments the way one can with an informant of the standard language.

Nevertheless, at least with respect to Bulgarian, there is a plethora of print data available to the scholar interested in dialectal classification and data. These resources include Stojko Stojkov’s overview of Bulgarian dialects,¹ many excellent monograph descriptions of individual dialects and dialect dictionaries,² including the new “ideographic” dictionary,³ and dialect atlases. There are several contributions in the latter category, the most well known of which is the four-volume *Bălgarski dialekten atlas*, with each of the four volumes devoted to a quadrant of today’s Republic of Bulgaria. These four volumes are the result of hundreds of hours of fieldwork by a team of dedicated scholars and students.

Between 1956 and 1961 they canvassed 395 villages throughout Bulgaria, noting down words that responded to questions on a lengthy list prepared by Stojkov.

This material was then collated and presented on maps: Volume 1 (southeastern Bulgaria, published in 1964) contains 277 maps; Volume 2 (northeastern Bulgaria, published in 1966) contains 290 maps; Volume 3 (southwestern Bulgaria, published in 1975) contains 314 maps; and Volume 4 (northwestern Bulgaria, published in 1981) contains 393 maps. Each map is devoted to a particular question within the realm of phonology, morphology or the lexicon, with the vast majority of the answers focused on individual words.⁴ Information is given by means of symbols on maps, one for each investigated village. Sometimes the information is binary, with a symbol of one color denoting the presence of an element and a symbol of another color denoting its absence; other times a number of variables are depicted, with symbols of different colors or shapes representing the different variables. In addition, the “commentaries” to these maps give large amounts of supplemental information by listing some of the actual data (responses to questionnaire items) which form the database for the map in question.

Four additional atlases have appeared, treating speech varieties located beyond the current borders of Bulgaria. Two are constructed according to the format of the central four-volume atlas: one treats dialects spoken in the Bosilegrad region of southeastern Serbia,⁵ and the other treats dialects spoken in the Serres-Drama region of northern Greece.⁶ The other two are constructed in a very different manner. First, each takes as its base the map in Stefan Mladenov’s 1929 *Geschichte der bulgarsichen Sprache*, thus including Macedonian and southeast Serbian dialects under the rubric “Bulgarian.” Second, each depicts information not by giving data for individual investigation points but rather as broad isoglosses. Both bear the title *Bălgarski dialekten atlas, obobštavašt tom*, and both state as their goal the desire to prove the “unity of the Bulgarian language.” The first, published in 1988, contains only 25 maps with little supplemental information. The second, published in 2001, bears the subtitle *Fonetika, akcentologija, leksika*, and is much more detailed. It contains 172 maps devoted to phonology, 88 maps devoted to accentual matters, and 108 maps devoted to lexical issues; all maps are handsomely produced with bright, clear colors. The choice to present material as isoglosses rather than as individual points (for which the data can then be verified), however, makes it more difficult to recover (or to research further) the actual data on which these isoglosses are based.

Balkanistica 28 (2015)

Unfortunately, much of this rich set of data is inaccessible to scholars outside Bulgaria. This is not just because all the material is monolingual or because the transcription system represents phonological values with symbols different from those used in the West. More significantly, it is because data found within these different sources are not easily comparable. If the question one seeks to answer is one for which the dialect atlas has included a map, one is fortunate; even then, however, one is frustrated by the fact that not all four volumes of the atlas include maps devoted to the same question or with material laid out in the same manner. Thus, for instance, an innovative project whose aim was to create XML markups of the atlas data, and to use this to track regional variation and lexical diffusion at quite precise levels, had to be abandoned because of “inconsistencies in the data set.”⁷ Furthermore, if one wishes to study a question not addressed by the atlas, one must comb through many different sources, each constructed in a different manner. The fact that atlas coverage is limited to the word level means that anyone wishing to study questions of syntax, phraseology, or discourse – all of which require longer stretches of data – cannot do so at any systematic level with the current resources available.

It is for this reason that I have constructed, together with colleagues both in Bulgaria and in the U.S., the interactive database entitled *Bulgarian Dialectology as Living Tradition* (BDLT), whose goals are to make such dialectal material available, to make it available not only in transcription but also as audio recordings, and to make all the material accessible to a broad range of users. At the core of the database are audio recordings of natural conversation with dialect speakers, in 177 separate segments (“texts”) representing 68 villages from across the entire landscape of Bulgaria. Each text has been transcribed twice — once using the Cyrillic set of symbols accepted by Bulgarian dialectologists, and once using a modified international transcription — and then translated into English; finally, each text is annotated for relevant linguistic features in such a way as to allow data searches at many different levels. The website will also include more traditional scholarly material, in the form of a prose overview of Bulgarian dialects, and prose summaries of the distinguishing features of each of the separate local dialects exemplified within the database

A more general but equally important goal is to return the focus of Bulgarian dialectology to that of living speech as it was recorded in a natural conversational context. At the current level of digital technology, it is now relatively easy to make actual field recordings readily accessible to the public. The availability of such

recordings not only allows those who were not present at the moment of recording to experience that moment as fully as possible, but also helps to demystify dialectology, both for the general public and for linguists who work exclusively with standard languages. Furthermore, it gives an additional level of access for those dialectologists who, though accustomed to working with a broad span of variation, may still have had relatively little contact with dialectal speech in its natural context.

The audio segments which form the basis of this collection have been selected from a large archive of field recordings with two basic criteria in mind: each text should illustrate the most salient features of the dialect in question, and each should constitute a well-formed instance of discourse and give some insight into traditional life or worldview. Other than that, the texts are unedited and provide natural conversational data. This allows researchers access to linguistic phenomena above the level of the word (including phrase-level phenomena such as clitic strings and compound verb forms, as well as phenomena of syntax and sentence intonation), and to discourse-level phenomena which are valuable for conversational analysis. Although the method of elicitation was not questionnaire-based, and therefore cannot attain the systematicity of a dialect atlas, the richness of the data at all other levels makes the material a highly valuable supplement to the *Bulgarian Dialect Atlas*.

The value of such material is obvious. Nearly all scholarly works which utilize dialect data quote that data from print sources; in many instances the quoted examples have themselves been quoted from an earlier work by another scholar. Our material, however, is available in the actual context of real speech. Furthermore, the material is being transcribed and analyzed collaboratively by the very researchers who recorded it in the field; this intensely interactive process, full of cross-checking, assures a high degree of accuracy. Over and above this, the availability of the actual audio recordings on the site will allow users to listen directly to the recordings and to verify our transcriptions and analyses for themselves.

The idea of presenting the material in its present format is my own, but the enterprise is a collective effort at all levels. For the construction of the database and further development of site design, I am indebted to Quinn Dombrowski and Cammeron Girvin in Berkeley, and for the data themselves I am indebted to my Bulgarian colleagues, above all Vladimir Zhobov and Georgi Kolev, who organized the field trips that took place between 1990 and 2013,⁸ who allowed *Balkanistica* 28 (2015)

inclusion within the database of material they had recorded on their own in the late 1980s (before I had become involved in joint fieldwork), who did the in-depth research necessary to choose the correct sites, and who, subsequent to the field trips, listened to the entire corpus of field recordings and made both the selections of texts to be included in the database and the initial transcriptions of these texts. The overall supervision of the project, and the preparation of the data after the initial transcription, is largely my responsibility, but for data entry itself I have been aided by a talented group of students through UC Berkeley's Undergraduate Research Apprenticeship Program.

I had originally conceived the project as an electronic publication, comprising streaming audio with accompanying text files. The choice to move to a digital format has made the preparation of the material more complex and time-consuming; however, the eventual advantage to researchers of the digital format more than justifies the time and effort involved (on the part of project workers) and the wait (on the part of those who will be consulting the material). A further advantage of this project is that the digital interface was developed using the open-source content management system Drupal, which means that the format is easily reproducible. We plan to produce documentation of the entire process involved in the construction of this project. Thus, although the current project includes only material recorded within the Republic of Bulgaria, we hope that other Balkan linguists may one day emulate the model and produce similar collections of dialectal material from closely (or not so closely) related Balkan languages.

The homepage of the site is available at [bulgariandialectology.org]. Informational material about the site and its organization is open to the public now, and the actual textual material will be made available gradually, as annotations are completed. The first section of the material is slated to be opened to the public in late 2015. Data preparation and entry will continue as long as needed, and additional sections will be released in stages, as soon as the material in each section is fully entered, annotated, and verified.

The database is organized as follows. The basic unit is the Text. Each text is an excerpt carefully selected from the entirety of our field recordings. As noted earlier, two basic criteria guided the selection of texts. The first and most obvious was that the text should exemplify as many of the salient features of the local dialect as possible. The second was that it should be a well-formed piece of discourse. In a few instances these are narratives, either a personal experience or a folktale. Most conversations, however, concern the way things were done in “the

old days.” Guiding the conversation in this direction is an effective way not only to garner valuable ethnographic information (about holiday customs, agricultural practices, food preparation, and the like), but also functions to direct the speaker’s attention away from present-day life (and thus to minimize the influence of the standard language).

Most texts concentrate on the speech of a single informant. Because elderly women tend to be the “keepers of tradition,” we spoke most often with them, but we also spoke with elderly men as well. In some instances the voices of other speakers are heard on the tape. Usually they are in the background, although in some cases there are two (and rarely three) speakers in dialogue with one another. In certain instances these segments were chosen deliberately because the topic and the liveliness of the conversation provided especially vivid examples of natural dialectal speech.

The audio segments have been digitized in MP3 format. In keeping with our goal of providing natural speech, there has been minimal editing of the tapes. This means that natural environmental noises such as crowing roosters or bleating goats are sometimes heard, as are (unfortunately) traffic noises. Although we have attempted to keep such interference to a minimum, the overriding criterion has been to choose a text that is “good” in terms of linguistic features and traditional content, and not to worry about the occasional truck, goat or rooster.

Once each text was selected and the corresponding audio file created, the text was transcribed using a modified version of the International Phonetic Alphabet; once the transcription was finalized, a script was run on each of these to produce a parallel Cyrillic version, using the symbols accepted within Bulgarian dialectology. Our principle has been to transcribe everything on the tape, thus including non-verbal cues such as laughter, coughing and onomatopoeic sounds: this has been done to enable conversational analysis of the texts as well as the more traditional linguistic analysis. For tapes made prior to 1990, in which only the informant’s voice was recorded, it has been necessary to reconstruct the investigator’s questions in order to give the semblance of a natural conversation

Each text was then divided into lines, numbered for the purposes of data retrieval. Each new turn in the conversation occupies a separate line. Longer segments by the same speaker are broken up into lines of between nine and twelve words; line breaks correspond, wherever possible, to natural rhythmic or syntactic breaks. The speaker is identified at the beginning of each line. Informants are not named, but are rather identified simply by lower-case letters. The first (or only)

informant in any one village is (a), the second is (b), and so on. Investigators are identified by their first and last initials, with full names given at the beginning of the text.

Each line of the text is then translated into English, in a manner which attempts to render not only the content but also the style of the original, so as to convey something of the flavor of the situation. Literal translations of individual words are given as part of the interlinear glosses; and the prose commentaries to each text will comment, when relevant, on the ethnographic content, the situational context of recording, or both.

Each text bears the name of the village where it was recorded; if more than one excerpt was chosen from the material of a village, the texts are numbered. Thus, text “Bansko” is the single example in the corpus of material from the town of Bansko, whereas text “Eremija 5” is the fifth of six texts in the corpus from the village of Eremija. The site has been designed so that texts can be displayed in several different views; once the site is active, buttons at the top of the text will allow the user to toggle between these views. The most primary of these is Token display. This display contains the spoken line transcribed using internationally recognizable characters, the English translation of each line, interlinear grammatical and lexical information for each token in each line, and a time code to aid in locating the audio of any one line in the accompanying audio file. Next is Line display, the goal of which is to create an evenly spaced line (this is done by omitting the interlinear identifications). Because this display contains only the translation and the original text, it provides a much easier and more natural way to read the texts for content. The third display, Cyrillic Line display, is intended for the Bulgarian academic audience. Since this audience is familiar with dialectal speech and accustomed to reading transcripts in the notation used only by Bulgarianists, this display includes only the spoken line transcribed in this notation, and the time codes to aid in locating any one line within the accompanying audio file.

Within texts, the most basic unit of analysis is the Token. A token is an individual word or unit of speech (including exclamations, hesitations sounds, false starts and the like). Each individual token is tagged in three different ways. The first and most basic set of tags annotates “Grammatical Categories,” and gives the basic lexical gloss for most words. These tags, which are displayed on the Text page itself (in Token display), appear directly under the token in question. The following chart gives the different tags which may be assigned; see below for

explications. Tokens with primarily grammatical meaning do not receive a lexical gloss; all others do. For instance, the token *jà* (the first person singular nominative pronoun) is given only the grammatical tags [*nom 1sg*] whereas the token *volòve* (the indefinite plural form of the masculine noun meaning “ox”) is given both the lexical gloss [*ox*] and the grammatical tags [*pl m*]. False starts (e.g., *nə* when the speaker began to say a verb with this prefix and then went on to say something else), as well as non-verbal cues like laughter or coughing, are given the tag [...] .

Case	Number	Gender	Def	Person	Verb Form	Function	Clitic	Deictic	Other	Pragmatic
nom	sg	m	def	1sg	pres	aux	clt	prox	adj	adrs
acc	pl	f		2sg	aor	comp		med	an.num	bkch
dat	pl.t	n		3sg	impf	cop		dist	adv	excl
voc	ct	indcl		1pl	fut	interr			coll	disc
				2pl	imv	neg			conj	hes
				3pl	inf	refl			exist	hort
					cond	rel			fut.pst	ost
					L.part				imprs	
					L.part.impf				name	
					P.part				place	
					ger					
					vbl.n					

The system of tags has been derived on the basis of the dialectal material to be tagged; conventions for assigning these tags have been written up in full detail for the project team and may be made publicly available at some point. The notes below simply spell out the several tag names and add extremely brief commentaries on the most salient facts about their usage.

1. Case. [*nom*]: nominative, [*acc*]: accusative, [*dat*]: dative, [*voc*]: vocative.

Except for the tag [*voc*], which is assigned only to nominal forms with a vocative desinence, these tags are used primarily for personal pronouns. Noun forms which exhibit oblique case marking are given the tag [*acc*], except for the very rare dialectal instances of dative plural, which are given the tag [*dat*].

2. Number. [*sg*]: singular, [*pl*]: plural, [*pl.t*]: pluralia tantum, [*ct*]: count form

3. Gender. [*m*]: masculine, [*f*]: feminine, [*n*]: neuter, [*indcl*]: indeclinable (no gender or number marking)

Gender tags are assigned to all nouns, to all third-person singular personal pronouns (subject and object) and to all singular adjective forms. No gender is assigned to plural personal pronouns or plural adjectives.

4. Definite. [*def*]: definite

This tag marks the definite article. Further information, if relevant for the dialect in question, is coded under “Deictic.”

5. Person. [*1sg*]: first singular, [*2sg*]: second singular, [*3sg*]: third singular, [*1pl*]: first plural, [*2pl*]: second plural, [*3pl*]: third plural

6. Verb form. [*pres*]: present, [*aor*]: aorist, [*impf*]: imperfect, [*fut*]: future, [*imv*]: imperative, [*inf*]: [truncated] infinitive, [*cond*]: conditional, [*fut.pst*]: “future in the past,” [*L.part*]: active participle (“L-participle”), [*L.part impf*]: the imperfect active participle, [*P.part*]: passive participle, [*ger*]: gerund, [*vbl.n*]: verbal noun

7. Function. [*aux*]: auxiliary, [*comp*]: complementizer, [*cop*]: copula, [*int*]: interrogative, [*neg*]: negation, [*refl*]: reflexive, [*rel*]: relative

The tag [*comp*] identifies the clause connector *da*. Although admittedly technically imprecise, it is adopted here as a cover term.

8. Clitic. [*clt*]: clitic

This tag is added to all clitic pronoun objects, the interrogative form *li* and all present-tense forms of the verb *săm*.

9. Deictic. [*prox*]: proximal, [*med*]: medial, [*dist*]: distal

These tags are used in dialects which distinguish three degrees of distance either in the definite article, in demonstrative pronouns and certain adverbs, or both. Forms in relevant dialects with the marker *-s* (as in *seloso* ‘the [this here] village,’ *soə* ‘this’) are labeled as “proximal,” and forms with the marker *-n* (as in *selono* ‘the [that there] village,’ *naə* ‘that’) are labeled as “distal.” In such dialects the marker *-t* (as in *seloto* ‘the village,’ *toz* ‘this’), elsewhere the only marker for

definiteness, receives the tag *[med]* to indicate that it participates in the three-way opposition.

10. Other. *[adj]*: adjective, *[an.num]*: animate number, *[adv]*: adverb, *[coll]*: collective, *[conj]*: conjunction, *[exist]*: existential, *[imprs]*: impersonal, *[name]*: personal name, *[place]*: toponym

The tag *[exist]* identifies the forms *ima* ‘there is’ and *n’ama* ‘there isn’t,’ in all tense forms; the tag *[imprs]* identifies forms which appear only in 3rd singular, such as *može* ‘it is possible’ or *trjabva* ‘it is necessary,’ again in all tense forms; and the tag *[an.num]* identifies numbers of the form *dvama*, marked as referring to humans only.

11. Pragmatic. *[adrs]*: address, *[bkch]*: backchanneling, *[excl]*: exclamation, *[disc]*: discourse, *[hes]*: hesitation, *[hort]*: hortative, *[ost]*: ostensive

All the above tags mark particles used to structure discourse in various ways.

The function of the second set of tags is to associate each token with a particular “Lexical Headword,” the purpose of which is to allow the grouping together of all phonetic realizations within the database of a single lemma. Wherever possible, this tag is a word found in standard Bulgarian dictionaries: the *Bălgarski tălkoven rečnik* (ed. by Andrejčĭn *et al.*) is used as the basic reference to determine the form of this tag. In the case of dialectal words which are not listed in standard dictionaries, a “dialectal lexeme” is derived according to normal phonological and morphological rules of Bulgarian. This dialectal lexeme is then cited in standard dictionary form (singular indefinite for nouns, masculine singular for adjectives, first-person singular present for verbs, etc.). It is marked with a following asterisk to indicate that it has been derived in this manner, and annotated with a reference to the existing dialect lexicon(s) in which it has been recorded. Instances in which dialectal speakers use a word which exists in the standard language, but with a significantly different meaning than in the standard definition, will also be marked and annotated.

At some later point, when the full set of dialectal lexemes has been entered, we plan to create a separate set of tags in order to categorize dialectal semantics. These tags will be organized on a separate field called “Semantic Relations,” and will consist in each instance of the standard Bulgarian word which best fits the meaning (or usage) of the dialectal word (or usage). As opposed to the Lexical

Balkanistica 28 (2015)

Headword tag, whose purpose is to group together all possible phonetic realizations of a single word, the Semantic Relations tags will allow the grouping together of all dialectal expressions of a single concept in instances where dialectal usage differs significantly from that of the standard language. Obviously, not all tokens will get both tags. Nearly every token will be given a Lexical Headword tag: only hesitation sounds, false starts, and the like are excluded. Clearly, however, only some tokens will be given Semantic Relations tags.

The third set of tags, called “Linguistic Traits,” covers the broadest range of information, and is intended to provide data about traits of interest at various levels of linguistic structure. Most of the tags, though they are phrased in synchronic terms, describe the results of diachronic processes, and index the traits normally catalogued in dialect descriptions. The list is organized in the manner of a full dialect description, with subsets corresponding to the different levels of linguistic structure. The largest set of tags concerns phonology: these tags mark different reflexes of the major Common Slavic vowels and consonants, as well as other vocalic phenomena such as elision and lengthening, and other consonantal phenomena such as palatalization, epenthesis, prothesis, and changes in various consonant clusters. Another set of tags concerns details of nominal and verbal stress patterns, and particularly interesting accentual phenomena such as the well-known dialectal “double accent.” The list of tags concerning morphology is also quite detailed. With respect to nouns and adjectives, these tags mark different dialectal reflexes of the definite article and of certain plural forms, or signal the presence of various derivational suffixes. With respect to verbs, they concern not only conjugational peculiarities, but also phenomena connected with individual verbal forms or aspectual usage; there are also tags signaling the presence of particular prefixes and suffixes.

A much smaller set of tags signals syntactic phenomena. These include tags concerning remnants of the case system (other than the normal *casus generalis* of far western dialects, which is tagged [*acc*] under Grammatical Categories), unusual shifts of gender and unexpected usage of aspectual forms. Finally, these tags are used to identify phrasal phenomena of interest. According to the current structure of the system, tagging must be done at the level of the token: technical limitations currently preclude the tagging of multiple-token phrases. Nearly everything of interest at the phrasal or syntactic level, however, can be tagged on the most relevant token. For instance, the presence of the renarrated mood can be signaled with the appropriate tag on the L-participle form that occurs without an

auxiliary; the presence of doubled pronoun objects can be signaled with the appropriate tag on the full object form; unusual clitic ordering can be signaled with the appropriate tag on one or more of the clitics in question; and phrasal accent groups that involve a shift of accent onto a clitic can be signaled with the appropriate tag on the clitic form which bears accent.

The manner in which the Linguistic Traits portion of the site is set up also allows for the creation of tags to mark unexpected instances of particularly interesting usage. These are being added as they are encountered in the process of data preparation.

Texts have been organized into five different groups, the first giving a representative sampling of dialects across the Bulgarian dialectal landscape, and each subsequent group filling out the picture more. Each group will be released to public view once all three sets of tags have been fully entered and verified. At a later point, we hope to tag the texts at a level higher than the individual token, and to categorize traits of discourse structure and narrative devices, and to produce an index of thematic content. In this way, the texts will be maximally useful not only to scholars of discourse, poetics and the like, but also to those interested in accessing the rich ethnographic content of the texts.

By providing linguists both with sizeable stretches of natural speech, and with the ability to listen repeatedly to any segments of these speech samples, the BDLT database will allow linguists to perform analyses of dialectal material that have heretofore been not possible. Two sample studies have been made of this material, the results of which are currently in press.⁹ One study is of Bulgarian dialectal accentuation, specifically speech segments which are characterized by the presence of more than one accent within a unit that normally bears only one accent. Analysis of the BDLT database yields evidence of not only the well-known “double accent,” but also of two other types. Common to them all, in addition to the presence of more than one accent, is the fact that each is associated with a particular well-defined sequence (the “conditioning frame”), and that such accentuation is variable. Isolating and defining the conditioning frame for each type, and then tabulating within the texts for each village not only all the instances where the specific accent type occurred, but also all the instances where it could have occurred but did not, allowed the question of variability to be addressed with considerably more precision. The second study is of vowel reduction. It is well known that this phenomenon characterizes Bulgarian dialects to varying degrees, yet it has not been possible to study it with any degree of precision. Now, using the *Balkanistica* 28 (2015)

audio files that are part of BDLT, we have been able to describe the different levels of quantitative vowel reduction in different dialects with considerable precision, and to examine the interconnections between quantitative degrees of reduction and qualitative features of the same vowels.

The coming of the digital age has changed many aspects of our lives. In this instance, it has allowed the creation of a research tool that will lead to a much better understanding of Bulgarian dialectology, and an even better appreciation of the rich living tradition embodied not only in Bulgarian dialects and the world outlook they represent, but also in the long and distinguished history of the study of Bulgarian dialects.

Notes

1. Stojkov, Stojko. *Bălgarska dialektologija*. The classic edition is the second, published in 1968. A third edition, with expanded bibliography and extensive notes prepared by Maksim Mladenov, was published in 1993, and a fourth edition appeared in 2002.
2. Most of these appeared in the series *Bălgarska dialektologija* (11 volumes in all, 1962 - 1983) and *Trudove po bălgarska dialektologija* (12 volumes in all, 1965-1984).
3. *Ideografski dialekten rečnik na bălgarskija ezik*, ed. by Todor Bojadžiev *et al.* Volume 1, covering the letters A-D, appeared in 2012; subsequent volumes are in preparation.
4. There are a few maps devoted to morphosyntax, such as “presence vs. absence of case forms in masculine animate nouns.”
5. Božkov, Rangel. *Bălgarski dialekten atlas: Severozapadni govori v caribrodsko i bosilegradsko*. 1986.
6. Ivanov, Jordan. *Bălgarski dialekten atlas: Bălgarski govori ot egejska Makedonija I, Dramsko, sjarsko, valoviško*. 1972.
7. See the archived comments at [<http://www.quinndombrowski.com/projects/bulgarian-dialect-atlas>]. Last accessed 17 December 2014.
8. The majority of the data were gathered on field trips sponsored by IREX grants, one in 1993 and one in 1996 ; I am grateful for this support. The volume reporting the results of the second of these trips is available electronically at [<http://escholarship.org/uc/item/9hc6x8hp>].
9. Vladimir Zhobov and Ronelle Alexander, *Bulgarian Dialects: Living Village Speech in the Digital Age*. Forthcoming from Sofia University Press.