OLD CHURCH SLAVONIC *BYTI* PART ONE: GRAMMATICAL PROFILING ANALYSIS

Hanne M. Eckhoff, Laura A. Janda, and Tore Nesset, UiT The Arctic University of Norway

1. Introduction

In Old Church Slavonic (OCS), as in many languages, the 'be' verb, *byti*, is special in several ways, with a particularly wide range of forms and uses. Scholarly studies of OCS verbs tend to deflect focus from *byti*, setting it apart from all other verbs as "exceptional." Van Schooneveld (1951) argued that *byti* was in fact not one verb, but two, namely an aspectual pair, an opinion that is often repeated in literature on the aspectual behavior of the verb in both OCS and the modern Slavic languages.[1] We examine this controversy in the light of a new empirical method—grammatical profiling.

Our argument is organized as follows. In section 2 we present the paradigm of *byti* and the scholarly debate to date on its status as a single verb vs. an aspectual pair. Next we introduce the PROIEL corpus that our study is based on in section 3, followed by a presentation of the grammatical profiling method in section 4. Section 5 describes our grammatical profiling analysis of *byti* with respect to other OCS verbs. We close with conclusions in section 6. All the data and statistical code used in this analysis are available at: hdl:10037.1/10074.

2. Why is *byti* special?

Indeed, *byti* is special in most respects. It is by far the most frequent verb in OCS, comprising 13.8% (2,428 examples) of all verb attestations in our dataset (see section 3). *Byti* is also special from a formal point of view: it is historically cobbled together from two verbs (*es-*, *bū-*; see Vasmer 1976: 159, 405; Shevelov 1965: 96, 238), and the alternation between the two stems is

1. Some scholars maintain that verbal aspect in Russian, for example, is actually an inflectional category, in which case there are not pairs of verbs, but single lexemes inflected for aspect (see Percov 1998 and Janda 2007 for references and further discussion). However, for the purposes of this article, we will consider an aspectual pair to consist of two verbs.

«романистом» — рассказчиком у «блатных». В обоих рассказах акт повествования становится буквальной и образной стратегией выживания, однако именно в «Заклинателе змей» Шаламов ставит под вопрос нравственную сторону этой стратегии.

Данная статья утверждает, что повествование и выживание приобретают характер сложной творческой деятельности, с помощью которой Шаламов пытается переосмыслить меняющуюся роль «лагерной» прозы.

still clearly visible in the OCS forms.[2] *Byti* corresponds to two verbs in Greek: the stative *eimi* 'be' and *gignomai* 'become' which expresses a change of state. This relationship to two distinct Greek verbs has led several scholars (see section 2.1) to question the unity of OCS *byti* as a verb, since a stative 'be' meaning can be associated with the *es*-group, whereas a change-of-state 'become' meaning can be associated with the *bū*-group. Due to its historical origins, the verb has a more complicated inventory of subparadigms than other OCS verbs, as documented in Table 1.

Table 1: The *byti* paradigm organized according to stem and morphology

	Morphological subparadigm	*es*-group	*bū*-group
"duplicate" forms for *es*-group and *bū*-group	present	*jestъ*	*bǫdetъ*
	imperfect	*běaše*	**bǫdeaše*[3]
	aorist	*bě*	*by(stъ)*
	present participle	*sǫšt-*	*bǫdǫšt-*
unique forms	subjunctive	*bi*	
"shared" forms	past participle		*byvъ*
	imperative		*bǫdi*
	infinitive		*byti*
	l-form		*bylъ*

Each cell in Table 1 represents an entire subparadigm, with the third person singular form shown for subparadigms that inflect for number and person (present, imperfect, aorist, subjunctive); the stem for the present participle; the masculine singular (nominative) form for the past participle and *l*-form; the second person singular for the imperative; and the uninflected infinitive form. The columns of the table separate the forms traditionally associated with the *es*-group from those associated with the *bū*-group. Each row presents a morphological subparadigm distinguished by its formal characteristics. The etymological distinction between the *es*-group and the *bū*-group is somewhat blurred for the subparadigms of the *es*-imperfect, *es*-aorist, and subjunctive, all of which begin with a *b*- which has probably been incorporated from the *bū*-group (Shevelov 1965: 96, Rix 2001: 242).

From Table 1 we see that there are three kinds of subparadigms for *byti*: subparadigms with "duplicate" forms for both the *es*-group and the *bū*-group (present, imperfect, aorist, present participle); a subparadigm with only *es*-group forms (the subjunctive/conditional, which is unique to *byti*); and several subparadigms with only *bū*-group forms (the past participle, imperative,

2. Here we follow the tradition of positing two origins for *byti*, but recognize that the grouping of forms according to an *es*-group vs. a *bū*-group is a simplification.

3. Not attested in canonical Old Church Slavonic.

infinitive, and *l*-form). This last set of subparadigms can also be thought of as the "shared" forms in the sense that they are common to both the *es*-group and the *bū*-group, since these are all forms that a typical verb has (as opposed to the subjunctive, which is atypical); without these forms, the *es*-group would have a defective paradigm.

We use the combination of group and subparadigm labels to distinguish among the "duplicate" forms thus: *es*-present vs. *bū*-present, *es*-imperfect vs. *bū*-imperfect, etc.

Table 1 makes it look as if there is quite some potential for aspectual contrast between the *es*-group and the *bū*-group, but it also conceals a number of important issues concerning the "duplicate" forms. In the imperfect, the *es*-group *běaše* has a defective subparadigm consisting only of third person forms, and the elusive *bū*-group **bǫdeaše* is not attested in our corpus or in canonical OCS at all. The relationship between the *es*-aorist and the *bū*-aorist is contested, as documented in section 2.1. As concerns the present participle, the *bū*-group *bǫdǫšt-* is relatively rare, and thus does not provide much contrast with the *es*-group *sǫšt-*; note also that Dostál (1954: 148–149) considers *bǫdǫšt-* to be an adjective, not a participle, probably on the grounds that it is normally attested in attributive or nominalized position, not as a conjunct participle, as we see in our dataset.

In sum, it appears that *byti* has "duplicate" forms that can potentially contrast only in the present and aorist. This verb has a unique subjunctive subparadigm, largely lacks a proper imperfect, but otherwise has a normal collection of forms. As we see in section 2.1, scholars are divided over how to make sense of the *byti* paradigm, particularly with respect to the imperfect and aorist forms.

2.1 Previous scholarship on the *byti* paradigm

We can view previous scholarship in terms of two hypotheses:

(1) One-verb hypothesis: *byti* is a single verb that has a more complicated paradigm than other verbs.

(2) Two-verb hypothesis: *byti* is an aspectual pair, with perfective and imperfective paradigms.

The clearest statement of a position on this controversy comes from van Schooneveld (1951), who argues that *byti* is in fact an aspectual pair of verbs (hypothesis 2), with opposing imperfective and perfective paradigms as outlined in Table 2. Key to the controversy concerning the status of *byti* is the subparadigm of the *es*-aorist.

Van Schooneveld's view of *byti* as an aspectual pair has been highly influential and is reflected in several later works on Late Common Slavic (Schenker 1995: 144), OCS (Remneva et al. 1999: 77, Krivčik and Možejko

Table 2: Van Schooneveld's distribution of forms of *byti* as aspectually opposed verbs

	imperfective *es-*	perfective *bū-*
present	*jestъ*	*bǫdetъ*
imperfect	*běaše*	**bǫdeaše*
aorist	*bě*	*by(stъ)*

1974: 141), and even modern Slavic languages (Junghanns 1997, Błaszczak 2009). Overall, the majority of works on OCS classify the *es*-aorist *bě* as an aorist, though typically with some reservations. Van Wijk (1931: 226) notes that it is used as an imperfect ("als ein Imperfekt verwendet wird"). This view is echoed by Selіščev (1951: 169): "These forms were used in the meaning of the imperfect" ("Эти формы применялись в значении имперфекта"), Ivanova (1977: 145) "these aorist forms based on the imperfect are used in the meaning of the imperfect" ("эти формы аориста от основы имперфекта употребляются в значении имперфекта"), and Večerka (1984: 141) "they have the form of the aorist, but the meaning of the imperfect" ("jsou co do formy aorist, co do významu imperfektum"). Others, while still classifying *bě* as an aorist, comment on the form as problematic: Lunt (2001: 138) notes "some degree of confusion" between the *es*-aorist *bě* and the imperfect *běaše*; Xaburgaev (1974: 282–283) calls the *es*-aorist an "aorist with an "imperfective" base" ("аорист с 'имперфективной' основой"). Gasparov (2001: 139) refuses to classify the *es*-aorist at all, instead listing it as a separate subparadigm: "Alongside regular paradigms for Aor. and Imp., *byti* had a third, hybrid aorist-imperfect form. It was built on the stem of the imperfect *bě-*, but used the formative suffix and the endings of the sigmatic aorist."

Alternatively, and despite the formal identity of the *bě* subparadigm as an aorist, some scholars simply classify the *es*-aorist as an imperfect (Diels 1932: 276; Vaillant 1948: 228, 298; Dostál 1954: 150; Meillet 1965: 274–275; Trubetzkoy 1968: 179; Schmalstieg 1976: 135; Elkina 1960: 168; Gardiner 1984: 80). None of these dissenters from the aorist interpretation offers any comparable counterargument to van Schooneveld, but together they can be considered proponents of the one-verb hypothesis (hypothesis 1), because they posit one set of forms for the imperfect and one for the aorist rather than "duplicate" forms.

Van Schooneveld's argument that OCS *byti* was in fact an aspectual pair of verbs rests on several problematic assumptions and does not take into account the full range of facts and data available. Contrary to the view of most standard grammars, which state that the *es*-imperfect and the *es*-aorist are semantically indistinguishable (cf. Lunt 2001: 156), van Schooneveld argues that the two subparadigms do differ semantically, in that the *es*-imperfect expresses simultaneity, whereas the *es*- aorist does not.[4] He supports his argu-

ment with a few examples, but there is no indication that he has tested his hypothesis against a large dataset.

Van Schooneveld's approach entails that the *bū*-present is in fact a perfective present. He makes this claim (van Schooneveld 1951: 103), but supports it with no evidence.

Furthermore, van Schooneveld simply disregards the "shared" forms, and does not present any views about the properties of the infinitive (*byti*), the imperative (*bǫdi*) and the *l*-form (*bylъ*). These forms are all from the *bū*-group, but it is implausible that they should be restricted only to perfective meanings, and such a hypothesis is not supported by Greek correspondences (see Part Two). Van Schooneveld leaves us with two options, both of which are unattractive: either we can assign the "shared" forms to the *bū*-group only, which would leave the *es*-group with no infinitive, imperative, or *l*-form; or we can claim that *byti* was biaspectual in the "shared" forms, but aspectually distinct in the others.

Finally, we argue that van Schooneveld's strictly structuralist approach gives an impoverished description of the verb. The reality of *byti*'s behavior is much more complex than what can be captured by assigning absolute contrastive values such as "perfective" and "imperfective" to various subparadigms.

Our approach is to test the two hypotheses against a large set of corpus data on *byti*. Rather than making a priori assumptions, we will use statistical methods to discover the patterns in the use of *byti* and thus allow a richer description to emerge from the data.

3. The PROIEL corpus and the *byti* data

In this section we present the PROIEL corpus and our dataset of *byti* examples extracted from that corpus. Note that our entire dataset, plus the statistical code used to analyze this data, is available at hdl:10037.1/10074, and the reader is encouraged to inspect the data and validate the results.

PROIEL (http://foni.uio.no:3000/; created at the University of Oslo) is a parallel corpus of Ancient Greek, Old Church Slavonic, Classical Armenian, Gothic and Latin. The Old Church Slavonic portion of PROIEL consists primarily of *Codex Marianus*, supplemented by portions of *Codex Zographensis* (both tetraevangelia) and *Codex Suprasliensis* (saints' lives and homilies); all three date from approximately the tenth-eleventh centuries, and all belong to the canon of texts that defines Old Church Slavonic. At the time when we

4. This is also van Schooneveld's general definition of the difference between the imperfect and the aorist in OCS, the imperfect is the marked partner and is [+past, +simultaneous], whereas the aorist is the unmarked partner and is just [+past]. Detailed studies of the semantics of the OCS imperfect and aorist make this analysis seem unlikely because the imperfect is not restricted to denoting eventualities that are simultaneous with another eventuality, but has a generalized imperfective semantics on a par with that found in, e.g., Ancient Greek (Amse-De Jong 1974, Eckhoff and Haug under submission).

extracted our data (September 2011), the total size of the Old Church Slavonic portion of PROIEL was approximately 62,000 words.[5]

Our dataset consists of 2,428 attestations of *byti* from *Codex Marianus*, *Codex Suprasliensis*, and *Codex Zographensis*. Each attestation is represented as a row in the database listing the context (sentence), along with various identifiers for the source, location in the text, morphosyntactic features and argument structure, and the Greek parallel verb (where relevant). In other words, every sentence has a unique code that identifies its source (e.g. *Marianus*) and location (e.g. Mark 2.18), plus numerous tags reflecting the morphological form of the verb and its syntactic context (arguments and their case marking, adverbs, etc.). The columns in the database present the various codes and tags.

4. Grammatical profiling and correspondence analysis

We use a linguistic profiling technique and the statistical method of correspondence analysis to probe the *byti* data for patterns. These methods are described here along with a discussion of how we handle the challenges posed by *byti*'s unusual collection of forms and the aim of objectively evaluating the one-verb vs. two-verb hypotheses.

Linguistic profiling is a means for operationalizing theoretical questions concerning the form-meaning relationship in language and the structure of linguistic categories. Profiling methods belong to a usage-based approach to linguistics (Langacker 2013), according to which generalizations emerge from language data. This approach assumes that meaning is the central motivation for all language phenomena, yielding the expectation that differences in behavior (distribution of forms) are motivated (though not specifically predicted) by differences in meaning. As a result, it is possible to use the distribution of forms to probe their meanings.

Linguistic profiling draws upon a long history of work on the behavioral profiles of words (Firth 1957, Harris 1970, Hanks 1996, Geeraerts et al. 1999, Speelman et al. 2003, Divjak and Gries 2006, Gries and Divjak 2009), employing comprehensive sets of tags for a wide range of linguistic factors, among them morphological, syntactic, semantic and lexical factors. By contrast, targeted profiling methods such as grammatical and constructional profiling focus on morphological and syntactic features respectively, thus providing a tighter focus and data that are more tractable for interpretation.

We apply the profiling methods that are most relevant to our investigation of aspectual and semantic behavior of *byti*, namely grammatical profiling and constructional profiling. This article focuses on grammatical profiling (see section 4.1), while Part Two (this volume) additionally uses constructional

5. This is admittedly a small corpus and thus vulnerable to the idiosyncracies of the scribes involved. Strictly speaking the results are thus valid only for this corpus.

profiling, described therein. The data that the grammatical profiles yield must be represented in a fashion that facilitates their interpretation, and for this purpose we use correspondence analysis, as described in section 4.2.

4.1 Grammatical profiling

Grammatical profiling examines the relationship between the frequency distribution of inflectional forms and linguistic categories. Grammatical profiling is inspired by the observation that some words are used more frequently in some forms of their paradigm than others. Perhaps the earliest relevant study of grammatical profiling was done by Karlsson (1985, 1986), who showed that the corpus attestations of most Finnish words involve a restricted number of paradigm forms, a finding confirmed in a more recent statistical analysis of Finnish verbs (Arppe 2001, 2005). Similar findings of skewed distributions of paradigm forms have been found for English verbs by Newman and Rice (Newman 2008; Newman and Rice 2006; Rice and Newman 2005). An important generalization emerges: each individual word has a unique grammatical profile, a unique set of preferences for distribution across its paradigm forms.

Janda and Lyashevskaya (2011) used grammatical profiling to reveal a strong relationship between the inflected forms of Russian verbs in the Russian National Corpus (ruscorpora.ru) and the expression of perfective vs. imperfective aspect. They found that the grammatical profiles of perfective verbs were significantly different from the grammatical profiles of imperfective verbs.[6] In other words, the frequency distribution (relative distribution across present, past, imperative, infinitive) of perfective verbs was clearly different from that of imperfective verbs. Grammatical profiling has further been used to test Dostál's classification of OCS verbs as perfective vs. imperfective (Eckhoff and Janda 2013, see also section 5) and to probe gender stereotypes in Russian as reflected by use of masculine vs. feminine past tense forms (Kuznetsova 2013: Chapter 3).

An essential issue for the present study is the aspectual status of *byti*: is it a single verb (hypothesis 1), or is it an aspectual pair of verbs (hypothesis 2)? Given that it is already known that there is a relationship between the grammatical profiles of verbs and aspect in both Russian (Janda and Lyashevskaya 2011) and OCS (Eckhoff and Janda 2014), it is reasonable to use grammati-

6. The result was statistically significant with a robust effect size (N = 5951250, chi-squared = 947756, df = 3, p-value < 2.2e-16, Cramer's V = 0.399). The aggregate grammatical profile of imperfective verbs has its peak in non-past forms with 47%, followed by the past (33%), infinitive (17%) and imperative (3%). The peak in the grammatical profile of perfective verbs is the past, with 63%, followed by the infinitive (22%), the non-past (12%) and the imperative (3%). The Janda and Lyashevskaya study specifically excluded grammatical forms that are restricted by aspect in modern Russian, namely gerunds and participles, and also excluded the verb *byt'* 'be'.

cal profiles to probe the aspectual behavior of *byti* as well. In this study we look at the frequency of *byti* forms in its subparadigms (see Table 1) and see how these compare with other OCS verbs.

4.2 Correspondence analysis

A database representing the distribution of grammatical profiles is hard to interpret in its raw form. Correspondence analysis is an effective means of representing this type of data in a meaningful way (Baayen 2008: 128–136). Correspondence analysis is a statistical technique that takes a matrix of data, computes the distances between the rows and the columns in the matrix, and then plots the distances in two dimensions that maximize the amount of variance that can be accounted for in the data. In other words, the mass of detail in a database can be compressed into a simple map that shows the differences between items much in the way that a road map shows the distances between cities. The grammatical and constructional profiles of *byti* are recorded in detail in the 2,428 lines of the database, corresponding to the number of examples extracted from the PROIEL corpus. We use correspondence analysis of grammatical profiles to show what the distribution of OCS verbs looks like overall, and then examine the position of *byti* in that distribution.

4.3 An objective approach to *byti*

All statistical models by necessity impose assumptions on the structure of data, and these limitations often force us to make uncomfortable choices from the start. In the input data, *byti* can be represented either as a single verb with a very rich set of subparadigms, or as two separate verbs with different sets of subparadigms. Clearly, this initial choice will influence the results of the analysis. In order to make our study more objective, we have decided not to settle for a single approach, but rather to use both starting assumptions. In this way we can compare the results and the answer to our question can emerge naturally from the data.

The problem is at its most acute when we want to compare the grammatical profile of *byti* with those of other verbs. Regardless of the starting assumption, we lose some detail, since *byti* has forms that other verbs do not share. We are therefore limited to comparing a subset of *byti*'s subparadigms with those of other verbs.

5. Grammatical profiling—how does *byti* relate to other verbs?

The study of the grammatical profiles of OCS verbs by Eckhoff and Janda (2014) provides a context for evaluating the behavior of *byti*. That study used as input 9,694 verb forms representing 129 verbs in the PROIEL corpus, but excluded *byti* as well as all verbs with less than twenty attestations. The input to the correspondence analysis was the set of grammatical profiles for the 129 verbs. The grammatical profile of each verb is comprised of the number of at-

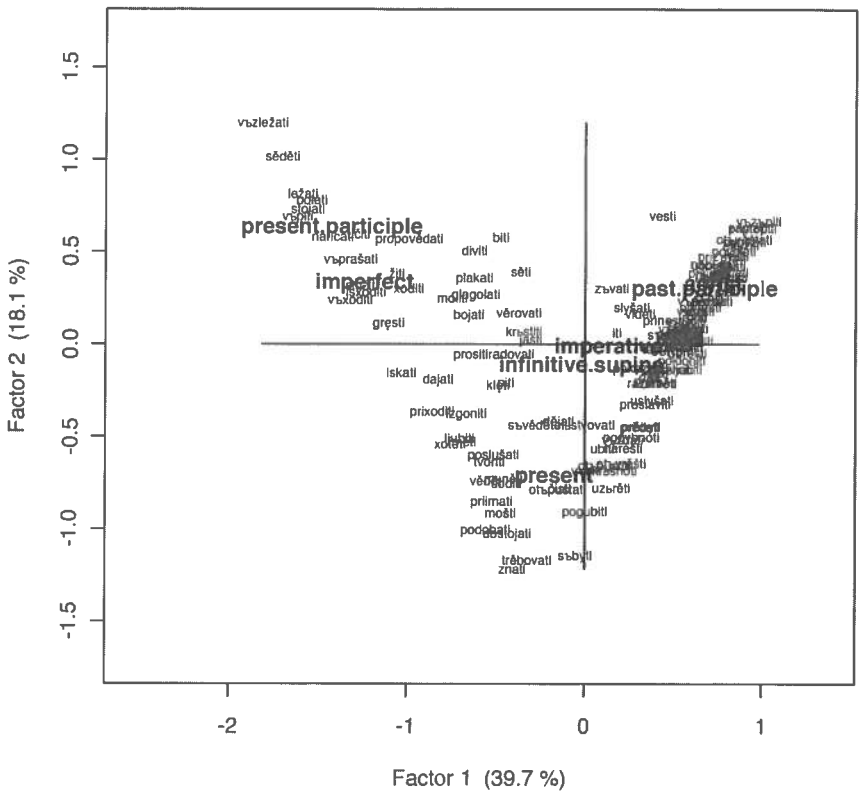

Figure 1: Correspondence analysis of OCS verbs in Eckhoff and Janda 2014

tested forms for that verb for each of its subparadigms. For example, the verb *tvoriti* 'make' has the following grammatical profile: aorist 0, imperative 14, imperfect 12, infinitive/supine 23, past participle 0, present 99, present participle 26.[7] The output of the correspondence analysis is the plot in Figure 1. The 129 verbs (which are the rows in the dataset) are printed in small type, while the seven subparadigms (which are the columns in the database) are printed in larger type.[8]

Figure 1 shows the distribution of the verbs (and subparadigms) according to the two most important mathematically deduced factors that are found in the data. Figure 1 can be thought of as a map of the distances between the

7. The *l*-form was excluded from the Eckhoff and Janda (2014) study on the grounds that it was restricted to a range of fairly different periphrastic constructions; the *l*-form is thus excluded also from the grammatical profiles in Part One, but not from the study of constructions in Part Two.

8. Given the large number of verbs, many are illegible on this diagram. A full list of verbs with their Factor 1 values is available at: hdl:10037.1/10074.

verbs, where the most important dimension is the horizontal one (Factor 1). The lines on the map show where the calculations divide the verbs according to the values of the factors as less than zero or greater than zero.

The most important outcome of this analysis is the distribution of verbs according to Factor 1, which concurs with Dostál's (1954) aspectual classification of OCS verbs at the rate of 97%. In other words, given only the distribution of grammatical forms of verbs and asked to plot those verbs, the correspondence analysis yields Factor 1 which accounts for 39.7% of the variance in the data and divides them into two groups (those with negative Factor 1 values vs. those with positive Factor 1 values) that strongly suggest an aspectual distinction. Verbs that Dostál classes as imperfectives such as *vъzležati* 'lie (at table)', *sěděti* 'sit', *vъpiti* 'cry' have negative Factor 1 values, whereas verbs that Dostál classes as perfectives such as *vъzъpiti* 'cry out', *pristǫpiti* 'step up to', *zaprětiti* 'threaten, rebuke' have positive Factor 1 values. This finding gives compelling support for Dostál's claim that there was an aspectual pair system in OCS, in contrast to scholars who suggest a more recent provenance for aspect in Slavic (Borodič 1953; Růžička 1957; Budich 1969; Bermel 1997; Nørgård-Sørensen 1997).

As mentioned above, *byti* was excluded from the Eckhoff and Janda (2014) study due to its unusual contested paradigm and the sheer numbers of its attestations, which presented the danger of skewing the data since nearly one of seven of all verb attestations in OCS are of *byti*. Let us now add *byti* to the correspondence analysis and see what happens. Because we wish to be as objective as possible, we do this in two alternative rounds, representing the two alternative hypotheses: round 1 represents the one-verb hypothesis and round 2 represents the two-verb hypothesis.

5.1 Round 1: *byti* according to the one-verb hypothesis

As we see from section 2.1, scholars who adhere to the one-verb hypothesis identify the *es*-aorist as an imperfect. If we use that designation for the *byti* data and add it to the correspondence analysis, we get the output presented in Figure 2, which is very similar to that in Figure 1, with only *byti* added (see circle in Figure 2).

In order to make the data conform to the standards used in Eckhoff and Janda 2014, Figure 2 excludes the attestations of subjunctive forms and *l*-forms of *byti*, as well as all *bū*-present and *bū*-present participle forms, and merges the *es*-aorist with the *es*-imperfect. Thus 416 *byti* attestations are excluded and 2,307 rows of *byti* data are added to the data on the remaining verbs. Under this analysis, *byti* emerges as a rather garden-variety imperfective verb with a Factor 1 value of -0.5. Its nearest neighbors include the states *ljubiti* 'love', *xotěti* 'want', *iměti* 'have', as well as activities such as *poslušati* 'obey', *tvoriti* 'make', and *svědětelьstvovati* 'witness'. In other words, when we look at the meanings of the neighboring verbs, this plot makes a lot of

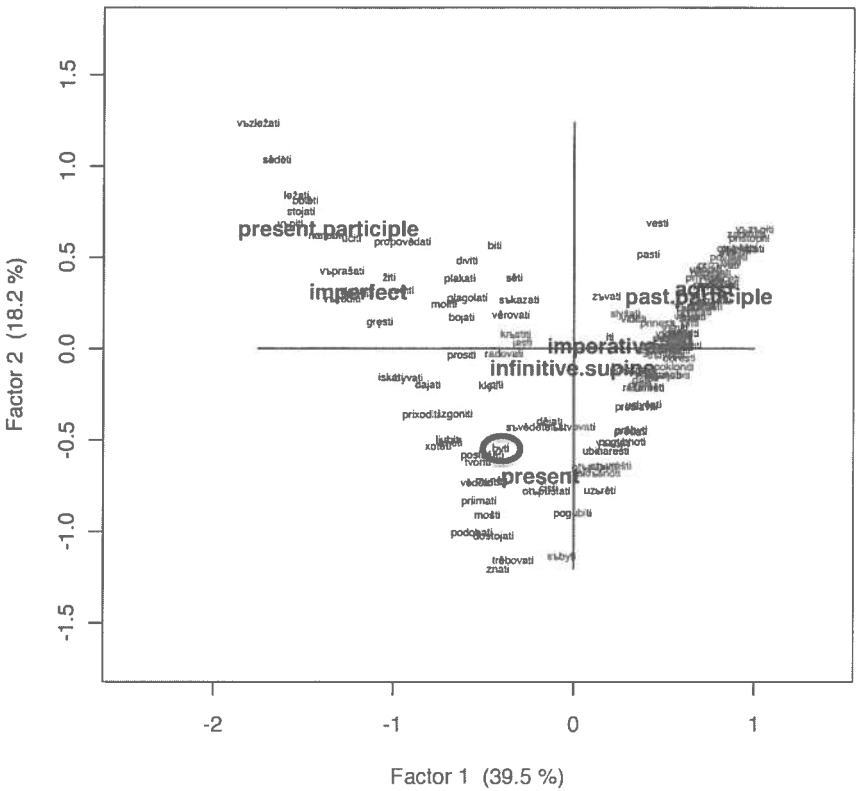

Figure 2: Correspondence analysis of OCS with *byti* included according to the one-verb hypothesis

sense, suggesting that *byti* can be felicitously interpreted as a single imperfective verb.[9]

5.2 Round 2: *byti* according to the two-verb hypothesis

If we instead split *byti* into two verbs, we get a different result. Under this assumption, we have: an *es*-group verb with present, present participle, imperfect and aorist subparadigms; and a *bū*-group verb with present, present participle, aorist, past participle, infinitive, and imperative subparadigms. The subjunctive must still be excluded, since no other verb has a subjunctive form.

9. An alternative, also compatible with the one-verb hypothesis, is to argue that *byti* belongs to a small residue of aspectually neutral verbs, as argued by Eckhoff and Haug (under submission). Most verbs exclude either the aorist + past participle or the imperfect + present participle, or are very strongly skewed in the distribution of these forms. Eckhoff and Haug find that, along with *byti*, there are only forty-two OCS verbs for which both types of forms are well attested, among them verbs like *biti* 'beat', *věrovati* 'believe', *dělati* 'do'.

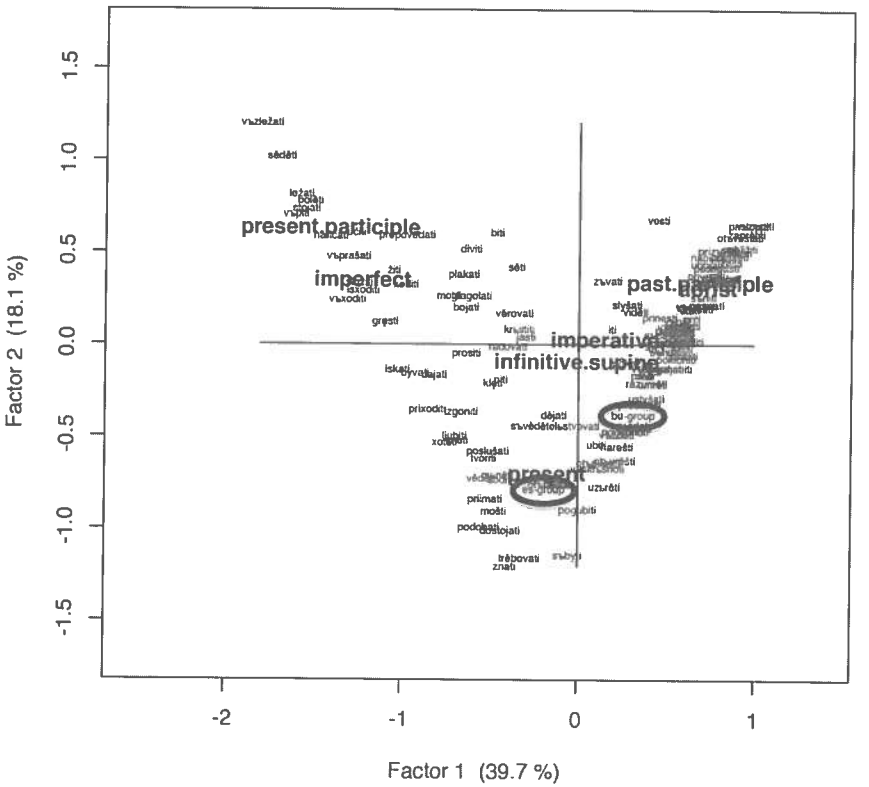

Figure 3: Correspondence analysis of OCS with *byti* included according to the two-verb hypothesis

The plot resulting from the correspondence analysis appears in Figure 3.

In this analysis, *byti* does indeed appear to split apart into two verbs, with the *es*-group landing with other imperfective verbs in the negative zone for Factor 1, and the *bū*-group landing among perfective verbs in the positive zone. This would seem to support van Schooneveld's argument and thus also hypothesis 2. Alternatively one could argue that this split is an artifact of the way in which the data were manipulated, segregating the subparadigms as described above. Furthermore, the split brings with it some inconvenient problems. If we have two verbs, one of them is rather strange: the *es*-group verb lacks an infinitive, imperative, and past participle, and has a defective imperfect subparadigm. These deficits are not due to lack of attestation, but are caused by our very definitions of these verbs. The split also assumes that the morphological identity of forms gives them a correct classification, which is clearly controversial for at least two of the subparadigms. Does it really make sense to class the *bū*-group present as a (perfective) present, or should it be

accorded special status as a future tense? Is the *es*-aorist really an aorist and is it really semantically distinct from the *es*-imperfect?

6. Conclusion

Our goal was to determine whether OCS *byti* 'be' is best characterized as a single verb or as an aspectual pair of verbs. We have approached the competing hypotheses about the status of *byti* in as objective a fashion as possible, starting from tangible data on the forms of the verb and contexts of its attestations.

We used grammatical profiling and correspondence analysis to test the hypotheses that OCS *byti* was a single verb vs. an aspectual pair of verbs. Based on previous studies of OCS and Russian, we know that verbs that differ in aspect also differ in their grammatical profiles. We therefore ran correspondence analyses comparing the grammatical profile of *byti* to those of other OCS verbs. This made it possible to contextualize the behavior of *byti* in comparison with other OCS verbs. In order to maximize objectivity, the comparison was conducted in two rounds, representing the two alternative hypotheses.

When we start by assuming that *byti* is a single verb, *byti* lands in the neighborhood of clearly imperfective verbs, some of which also share stative semantics. However, it was necessary to exclude several subparadigms in order to compare *byti* to other verbs. When the subparadigms are split according to their origin from the *es*-group or the *bū*-group, on the other hand, the forms appear to behave as an aspectual pair. However, this second model is flawed in that it presumes an *es*-group verb that has a very defective paradigm (with no imperative, infinitive, or past participle), fails to address any possible contrasts that could be expressed by the "duplicate" forms, and the division across the negative and positive Factor 1 values is arguably an artifact of the segregation of the subparadigms in the input. Thus whereas round one supports the one-verb hypothesis and round two supports the two-verb hypothesis, these results are not conclusive. Furthermore, this part of the analysis rests only on the morphological forms without taking into account their meanings, which is possible only in the context of their constructions.

An interesting fact about both the one-verb analysis in round 1 and the two-verb analysis in round 2 is that even though we have added a large portion of data to the calculations (*byti* adds nearly 24% more data), the outcome for the other verbs is barely perturbed, which suggests that either hypothesis is compatible with other verb types in the system. It is theoretically possible to model *byti* as either a single verb or as a verb pair, but correspondence analysis alone is not sufficient to resolve the controversy between our two hypotheses. The input data in both analyses were impoverished. In the single-verb analysis all subjunctive, *l*-forms, *bū*-present, and *bū*-present participle forms were excluded. In the two-verb analysis, the two latter subparadigms were included, but instead the analysis suffered from the proposal of a verb lacking forms that are otherwise not absent from verbal paradigms. Both analyses are

unable to directly address the status of the *es*-aorist and the *bū*-present and contrast their behavior with their purported counterparts (the *bū*-aorist and the *es*-present), or to assess the behavior of the shared forms (the infinitive, imperative, and *l*-form). In order to examine *byti* in more detail, we turn to a constructional profiling analysis in Part Two.

Acknowledgement: This research was supported in part by grant number 222506 from the Norwegian Research Council.

REFERENCES

Amse-de Jong, Tine. *The Meaning of the Finite Verb Forms in the Old Church Slavonic Codex Suprasliensis. A Synchronic Study*. The Hague: Mouton, 1974.

Arppe, Antti. "Focal points in frequency profiles—how some word forms in a paradigm are more significant than others in Finnish." *Proceedings of the 6th Conference on Computational Lexicography and Corpus Research, June 28–30, 2001*. Birmingham, UK: University of Birmingham, 2001.

———. "Morphological features as 'context' in distinguishing semantically similar words." *Proceedings from the Corpus Linguistics Conference Series, vol. 1, Third Biennial Corpus Linguistics Conference, July 14–17, 2005*, Birmingham, UK: University of Birmingham, 2005.

Baayen, R. Harald. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press, 2008.

Bermel, Neil. *Context and the Lexicon in the Development of Russian Aspect*. Berkeley, CA: University of California Press, 1997.

Błaszczak, Joanna. "Differential Subject Marking in Polish: The Case of Genitive vs. Nominative Subjects in 'X was not at Y'-constructions." *Differential Subject Marking*. Ed. Helen de Hoop and Peter de Swart. Dordrecht: Springer, 2009. 113–149.

Borodič, V. V. "K voprosu o formirovanii soveršennogo i nesoveršennogo vida v slavjanskix jazykax." *Voprosy jazykoznanija* 2 (1953): 68–86.

Budich, Wulf. *Aspekt und Verbale Zeitlichkeit in der I. Novgoroder Chronik*. Graz: Akademische Druck- u. Verlagsanstalt, 1969.

Diels, Paul. *Altkirchenslavische Grammatik. I. Teil: Grammatik*. Heidelberg: Carl Winters, 1932.

Divjak, Dagmar and Stefan Th. Gries. "Ways of trying in Russian: Clustering behavioral profiles." *Corpus Linguistics and Linguistic Theory* 2 (2006): 23–60.

Dostál, Antonín. *Studie o vidovém systému v staroslověnštině*. Prague: Státní pedagogické nakladatelství, 1954.

Eckhoff, Hanne, Laura A. Janda and Tore Nesset. "Old Church Slavonic *byti* Part Two: Constructional Profiling Analysis." *Slavic and East European Journal* 58.3 (2014): 498–525.

Eckhoff, Hanne M. and Dag Haug. "Aspect and prefixation in Old Church Slavonic." Under submission.

Eckhoff, Hanne M. and Laura A. Janda. "Grammatical Profiles and Aspect in Old Church Slavonic." *Transactions of the Philological Society* 112.2 (2014): 231–258. DOI: 10.1111/1467-968X.12012.

Elkina, Nina Maksimovna. *Staroslavjanskij jazyk*. Moscow: Gosudarstvennoe učebno-pedagogičeskoe izdatel'stvo, 1960.

Firth, John R. "A synopsis of linguistic theory 1930–1955." *Studies in Linguistic Analysis*. Ed. John R. Firth. Oxford: Blackwell, 1957. 1–32.

Gardiner, S. C. *Old Church Slavonic*. Cambridge: Cambridge University Press, 1984.

Gasparov, Boris. *Old Church Slavonic*. (Languages of the World/Materials 338). Munich: LINCOM EUROPA, 2001.

Geeraerts, D., S. Grondelaers, and D. Speelman. *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: Meertens Instituut, 1999.

Gries, Stefan Th. and Dagmar Divjak. "Behavioral profiles: a corpus-based approach towards cognitive semantic analysis." *New Directions in Cognitive Linguistics*. Ed. Vyvyan Evans and Stephanie S. Pourcel, Amsterdam: John Benjamins, 2009. 57–75.

Hanks, Patrick. "Contextual dependency and lexical sets." *International Journal of Corpus Linguistics* 1 (1996): 75–98.

Harris, Zellig S. *Papers in structural and transformational linguistics*. Dordrecht: Reidel, 1970.

Ivanova, T. A. *Staroslavjanskij jazyk*. Moscow: Vysšaja škola, 1977.

Janda, Laura A. "Inflectional morphology." *Handbook of Cognitive Linguistics*. Ed. Dirk Geeraerts and Hubert Cuyckens. Oxford: Oxford U Press, 2007. 632–649.

Janda, Laura A. and Olga Lyashevskaya. "Grammatical profiles and the interaction of the lexicon with aspect, tense and mood in Russian." *Cognitive Linguistics* 22.4 (2011): 719–763.

Junghanns, Uwe. "On byt' (and byti)." *Formale Slavistik (Leipziger Schriften zur Kultur-, Literatur-, Sprach- und Übersetzungswissenschaft* 7). Ed. Uwe Junghanns and Gerhild Zybatow. Vervuert Verlag: Frankfurt am Main, 1997. 251–265.

Karlsson, Fred. "Paradigms and word forms." *Studia gramatyczne VII. Ossolineum* (1985): 135–154.

———. "Frequency considerations in morphology." *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39 (1986): 19–28.

Krivčik, V. F. and N. S. Možejko. *Staroslovjanskij jazyk*. Minsk: Vyšejšaja škola, 1974.

Kuznetsova, Julia. *Linguistic Profiles. Correlations between Form and Meaning*. PhD Dissertation, University of Tromsø, 2013.

Langacker, Ronald W. *Essentials of Cognitive Grammar*. Oxford: Oxford University Press, 2013.

Lunt, Horace G. *Old Church Slavonic Grammar*. Berlin: Mouton de Gruyter, 2001.

Meillet, Antoine. *Le slave commun*. Paris: Librairie Honoré Champion,1965.

Newman, John. "Aiming low in linguistics: Low-level generalizations in corpus-based research." *Proceedings of the 11th International Symposium on Chinese Languages and Linguistics*, National Chiao Tung University, Hsinchu, Taiwan. May 24 2008. http://www.ualberta.ca/~johnnewm/Aiming%20Low.pdf.

Newman, John and Sally Rice. "Transitivity schemas of English EAT and DRINK in the BNC," in S. Th. Gries and A. Stefanowitsch (ed.). *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin and New York: Mouton de Gruyter, 2006. 225–260.

Nørgård-Sørensen, Jens. "Tense, aspect and verbal derivation in the language of the Novgorod birch bark letters." *Russian Linguistics* 21 (1997):1–21.

Percov, N. V. "Russkij vid: slovoizmenenie ili slovoobrazovanie?" Ed. M. Ju. Čertkova. *Tipologija vida. Problemy, poiski, rešenija*. Moscow, 1998.

Remneva, M. L., V. S. Savel'ev, I. I. Filičev. *Cerkovnoslavjanskij jazyk*. Moscow: Izdatel'stvo MGU, 1999.

Rice, Sally and John Newman. "Inflectional islands." Presentation at the 9th International Cognitive Linguistics Conference, Yonsei University, Seoul, Korea, 2005. http://www.ualbera.ca/~johnnewm.

Rix, Helmut. *Lexikon der indogermanischen Verben: Die Wurzeln und ihre Primärstammbildungen*. Dr.-Ludwig-Reichert-Verlag: Wiesbaden, 2001.

Růžička, Rudolf, 1957. *Der Verbalaspekt in der altrussischen Nestorchronik*. Berlin: Akademie-Verlag.

Schenker, Alexander M. *The Dawn of Slavic*. New Haven: Yale University Press, 1995.
Schmalstieg, William R. *An Introduction to Old Church Slavic*. Cambridge, MA: Slavica Publishers, 1976.
Seliščev, A. M. *Staroslavjanskij jazyk. Čast' vtoraja*. Moscow: Gosudarstvennoe učebno-pedagogičeskoe izdatel'stvo/URSS, 1951/2004.
Shevelov, George Y. *A Prehistory of Slavic*. New York: Columbia University Press, 1965.
Speelman, Dirk, Stefan Grondelaers, and Dirk Geeraerts. "Profile-Based Linguistic Uniformity as a Generic Model for Comparing Language Varieties." *Computers and the Humanities* 37.3 (2003): 317–337.
Trubetzkoy, Nikolaus S. *Altkirchenslavische Grammatik*. Graz: Hermann Böhlaus, 1968.
Vaillant, André. *Manuel du vieux slave. Tome I: Grammaire*. Paris: Institut d'études slaves, 1948.
van Schooneveld, Cornelis H. "The aspect system of the Old Church Slavonic and Old Russian verbum finitum *byti*." *Word* 7.2 (1951): 96–103.
van Wijk, Nicolaas. *Geschichte der altkirchenslavischen Sprache. Erster Band: Laut- und Formenlehre*. Berlin: Walter de Gruyter, 1931.
Vasmer, Max. *Russisches etymologisches Wörterbuch*. Erster Band: A–K. Heidelberg: Carl Winter, 1976.
Večerka, Radoslav. *Staroslověnština*. Prague: Státní pedagogické nakladatelství, 1984.
Xaburgaev, G. A. *Staroslavjanskij jazyk*. Moscow: Prosveščenie, 1974.

Аннотация

Ханне М. Экхофф, Лора А. Янда и Туре Нессет
Старославянский глагол *быти* (часть первая: анализ грамматических профилей)

В статье обсуждается спорный вопрос о том, был ли старославянский глагол *быти* одним глаголом несовершенного вида с необычно большим количеством словоформ, или же мы имеем дело с видовой парой, отражающей две этимологические основы (*es-* и *bū-*). Предлагается объективное эмпирическое исследование вопроса, основанное на статистическом анализе 2428 примеров употребления глагола *быти* в сопоставлении с 9694 примерами 129 других глаголов. Предлагаемый анализ способствует точной локализации *быти* в глагольной лексике старославянского языка. При сопоставлении используется метод «грамматического профилирования», описывающий частотное распределение словоформ каждого глагола. Анализ проведен в два этапа. На первом этапе мы исходили из того, что *быти* является единым глаголом, в то время как на втором этапе *быти* рассматривался как глагольная пара. Оба варианта анализа дают правдоподобные результаты, и несмотря на то, что одного грамматического профилирования недостаточно для того, чтобы разрешить вышеназванное противоречие, такое профилирование формирует базу для дальнейшего анализа, который будет представлен во второй части исследования, где *быти* представлен как один глагол.

OLD CHURCH SLAVONIC *BYTI* PART TWO: CONSTRUCTIONAL PROFILING ANALYSIS

Hanne Eckhoff, Laura A. Janda, and Tore Nesset, UiT The Arctic University of Norway

1. Introduction

In Part One (this volume) we reviewed the history of the controversy over whether Old Church Slavonic (OCS) *byti* 'be' is a single verb (the single-verb hypothesis) or an aspectual pair of verbs (the two-verb hypothesis). We presented a grammatical profiling analysis of *byti* in comparison with other OCS verbs. While there were fewer problems with the single-verb analysis (since it did not entail positing a verb with no imperative, infinitive, or past participle), both the single-verb hypothesis and the two-verb hypothesis found modest support in the grammatical profiling analysis. In this article we follow up with an analysis of the grammatical constructions of *byti*. On the basis of corpus data and statistical analysis we argue that *byti* is best represented as a single, though complex, verb.

The structure of our argument is as follows. In section 2 we present the forms and Greek correspondences for *byti* and discuss what these mean for the controversy concerning the status of this verb. Section 3 situates this study with respect to the use of radial categories in cognitive linguistics, as well as the use of cognitive linguistics in both Slavic and general linguistics. In that section we also introduce the method of constructional profiling. We detail our study of the constructional profile of *byti* in section 4 and close with conclusions in section 5. All the data and statistical code used in this analysis are available at: hdl:10037.1/10074.

2. *Byti*: forms and Greek correspondences

OCS *byti* has a particularly rich set of forms, largely due to its history (see details in Part One). Table 1, which summarizes these forms, is here reproduced from Part One for the reader's convenience. The terms presented in this table (*es*-group, *bi̯ū*-group, etc.) are used to identify the relevant forms in the remainder of this article.