When we went digital and seven other stories about Slavic historical linguistics in the twenty-first century

Tore Nesset

Abstract. In this overview article, I seek to identify and discuss some tendencies in Slavic historical linguistics in recent years. Rather than presenting an extensive catalogue of studies on miscellaneous topics, I focus on three general issues, viz. how Slavic historical linguistics is developing in response to new theoretical ideas, methodological innovation and “new” data. The article explores case studies from the syntax, morphology and phonology of a number of Slavic languages and tells eight stories about Slavic historical linguistics in the twenty-first century.

1. Introduction – the big picture

It has been a long time since Slavic linguistics was Slavic historical linguistics, but historical work in Slavic is still alive and kicking in the twenty-first century – as I hope this overview will show. My account, by necessity incomplete and to some extent subjective, will be based on an inclusive view of what historical linguistics is. However, I will not discuss ongoing and recent changes in the modern Slavic languages, and thus not explore phenomena such as the emergence of the so-called new vocative in Russian (Danièl' 2009 and Andersen 2012) or recent changes in particles in conversational Czech (Fried 2009). Instead I will focus on diachronic and synchronic analyses of older stages of the Slavic languages. My overview will not cover problems concerning standard languages such as the relationship between East Slavic and Church Slavic in Kievan Rus’ (Uspenskij 2002 and Živov 1996). I will furthermore leave out studies devoted to particular texts, thus for instance ignoring Zaliznjak’s (2008a) masterful study of the authenticity of Slovo o polku Igoreve. New textbooks (e.g. Galinskaja 2014 and Nesset 2015), handbooks (Kempgen et al. 2014) and grammars (e.g. Polivanova 2013 and Krys’ko 2000-2006) will not be reviewed, and I will limit myself to discussing works published in the year 2000 or later. Even with these limitations the field is large and heterogeneous. Instead of trying to cover everything, I will identify some trends and discuss some representative works in more detail.

Before starting, let us take one step back and ask what is going on in linguistics in general – and in the world outside. The first thing that comes to mind is that we live in a digital age. How has the digital revolution influenced linguisticis? To what extent have these influences impacted historical linguistics? Are we experiencing a “quantitative turn” (Janda 2013)? In section 2, addressing these questions I discuss new digital resources and give some examples of how they have been used in Slavic historical linguistics. A further example is given in section 6 on aspect.

Another major trend in modern linguistics can be labeled the “social turn”, beginning with the birth of modern socio-linguistics in the 1960s and 1970s. A possible echo of this social turn in Slavic historical linguistics is the revived interest in contact phenomena in recent years. As Timberlake (2014: 165-4) puts it, Slavic “provides an ideal context for studying how language contact has worked over an extended period.” Section 3 of the present article gives some examples of recent studies in Slavic historical linguistics from the perspective of language contact.
What happens when science gets access to new data? This question is discussed in section 4, where I explore the importance of the birch bark letters for Slavic historical linguistics.

New theoretical ideas also change what questions are asked and how linguistic analysis is carried out. Section 5 on grammaticalization theory, section 7 on Construction Grammar, section 8 on language typology and section 9 on phonology address a number of examples of this type.

This article tells eight stories about Slavic historical linguistics in the twenty-first century. As summarized in section 10, they offer a multifaceted picture of a fertile subfield of Slavic linguistics.

2. Slavic historical linguistics in the digital age – a quantitative turn?

The curse of the historical linguist is scarcity of data. The few examples one has managed to find may be suggestive, but nevertheless all too often too few to demonstrate robust tendencies. Does the digital age give reasons for hope? I argue that digital resources under development call for cautious optimism, and I will show how such resources can shed new light on contested issues in Slavic historical linguistics.

Recent years have witnessed a number of projects developing historical corpora of Slavic languages. Here are some examples (but not an exhaustive list):

(1) a. The Russian National Corpus (historical subcorpus)

b. The Czech National Corpus (diachronic section)

c. The PROIEL corpus of Old Indo-European languages including OCS
d. The TOROT (Tromsø Old Russian and OCS Treebank)
e. The Manuskript corpus of Slavic and Russian texts

f. The RRuDi corpus (Regensburg Russian Diachronic Corpus)

In addition, there are numerous digital editions of various important texts that are available and searchable. Examples include the Kiev, Suzdal and Galician-Volhynian chronicles made available by the Institute of the Russian language at the Russian Academy of Sciences and a number of texts available on David J. Birnbaum’s platform obdurodon.org, e.g. the Old Russian Primary Chronicle (Ostrowski ed. 2003) and the Codex Suprasliensis (prepared in collaboration with the Bulgarian Academy of Sciences).2

In order for such resources to remedy the problem of scarcity of data, the resources must (a) be large, and (b) have good annotation. With regard to size, the PROIEL and TOROT corpora are representative examples. Taken together they comprise about 160,000 words of OCS and approximately 180,000 words of Old Russian/Middle Russian.3 To the uninitiated reader this may seem substantial, but compared to existing corpora of modern languages (e.g. the Russian National Corpus with more than 265 million words) the historical corpora are quite small. Therefore, the problem of scarcity of data is hardly

---

1 I will not discuss the distinction between “corpus” and “electronic edition”, which is of no consequence for my line of argumentation. The Kiev, Suzdal and Galician-Volhynian chronicles can be accessed from http://www.ruslanas.ru/agens.php?id=res. The Primary Chronicle, the Codex Suprasliensis and a number of other texts are available at http://obdurodon.org/. Notice that throughout the article I use the term “Old Russian” (with one r) instead of the more traditional “Old Russian”, since we are dealing with the ancestor of all the modern East Slavic languages, not just Russian. Arguably, the epithet “old” in “Old Russian” is redundant, but I use it to avoid confusion with (modern) Russian.

2 Words with syntactic annotation as of July, 2016 (Hanne Eckhoff, personal communication).
eliminated. However, it is possible to detect robust tendencies for frequent linguistic patterns, while at present the existing historical corpora have less to offer for the study of infrequent phenomena.

When it comes to annotation, the corpora mentioned in (1) are lemmatized and have solid part of speech and morphological annotation. This makes it possible to search for all forms of a particular lexeme and specify searches for inflectional categories (cases, tenses, etc.). In addition, the PROIEL and TOROT resources include syntactic annotation, which facilitates searches for particular syntactic constructions.

How can corpus data shed new light on controversial issues in Slavic historical linguistics? By way of example, consider a recent study of the OCS verb byti ‘be’ (Eckhoff, Janda and Nesset 2014a-b). In a seminal article, van Schooneveld (1951) proposed that byti is best analyzed as two verbs, which form an aspectual pair. Eckhoff et al. (2014a-b) use corpus data and statistical analysis to test this “Two-verb hypothesis” and compare it to the more traditional “One-verb hypothesis”, whereby byti represents one single verb with a more complicated paradigm than other verbs. Byti is a highly frequent verb, so it was possible to extract a dataset of 2,428 examples from the PROIEL corpus, which were compared to other verbs in OCS extracted from the same corpus. The complete dataset comprised more than 17,500 examples. In Eckhoff et al. 2014a, the frequency distribution of the various inflected forms was used to test the two hypotheses. Unfortunately, this test turned out to be inconclusive, although it weakly favored the One-verb hypothesis (Eckhoff et al. 2014a: 494–495 and 2014b: 498). Eckhoff et al. (2014b) then went on to analyze the use of byti across grammatical constructions, and this test yielded strong evidence in favor of the One-verb hypothesis, insofar as “the grammatical behavior of byti is neatly integrated into the overall picture of the OCS verb inventory, and in this context byti is best interpreted as a single verb rather than a pair of verbs because the aspectual contrast we would expect to find with an aspectual pair is lacking” (Eckhoff et al. 2014b: 522). While in-depth discussion of the details of the statistical analysis is beyond the scope of the present article, it seems clear that studies where corpus data are subjected to statistical analysis have the potential of shedding new light on controversial questions in Slavic historical linguistics. However, given the limited size of existing historical corpora, statistical analysis is only feasible for relatively frequent phenomena.

3. Contact phenomena – a social turn?

The study of language contacts has always played an important role in Slavic historical linguistics. Contacts with Finno-Ugric and other languages in the east, Germanic in the west as well as the contact in and around the Balkan Sprachbund in the south have shaped the Slavic languages as we know them today. Early borrowing has been important in the argumentation concerning contested issues in Slavic historical linguistics such as the primordial home of the Slavs (see e.g. Holzer 2014). However, it seems that contact phenomena have received renewed interest in recent years (see e.g. Wiemer and Wählhi 2012). It is symptomatic that Wiemer and Hansen (2012) relate a number of grammaticalization phenomena in Slavic to language contact and that Dickey (2011) relates the development of aspect to language contact. We will return to
the relationship between language contact and grammaticalization in section 5 and to aspect in section 6. In the present section, I will limit myself to discussion of three case studies of language contact – one from East Slavic, one from South Slavic, and one from West Slavic.

The substrate influence of Finno-Ugric languages has been debated for a long time. Among the most frequently mentioned cases of Finno-Ugric influence on Contemporary Standard Russian are the following (Grenoble 2012: 584):

(2) a. Vowel reduction ("akan’e"), which many researchers relate to influence from Moksha Mordvinian (see Veenker 1967: 29–35 and Haarmann 2014:1195–1196)

b. Sentences without copula in the present tense, e.g. On student ‘He is a student’.

c. The predicative possessive construction, e.g. U menja (est’) kniga ‘I have a/the book’.

In a series of articles summarized in Weiss 2012, Weiss has proposed that another important example of the Finno-Ugric substrate influence on Russian is the so-called Serial Verb Construction, i.e. examples such as sjadem-podumaem ‘we will sit down and think’. Against Aikhenvald (2006), Weiss (2012: 613) argues that the Russian construction meets the typological criteria for serial verb constructions. In particular, in Russian (a) the morphological outfit of both verbs is identical in terms of tense, mood and other grammatical categories, (b) both verbs share the same subject, (c) auxiliary morphemes (e.g. the subjunctive clitic by) are shared by both verbs, and (d) the order of the two verbs may be reversed. In view of the fact that the Serial Verb Construction most likely existed already in Common Finno-Ugric, while the construction is not characteristic of Slavic languages other than Russian, Weiss argues that we are dealing with the result of a Finno-Ugric substrate in Russian. Since a syntactic borrowing of this kind would require extensive language contacts and bilingualism over time, Weiss (2012: 637–638) speculates that the source could be the Finno-Ugric tribes of the Merja and Murom, which inhabited the area around Moscow.

A long-standing issue in Slavic Historical linguistics is the development of articles. A recent contribution to this field is Breu 2012, who addresses the emergence of articles in Molise Slavic. This is the South Slavic variety spoken in the Region of Molise in Southern Italy by descendants of immigrants from Dalmatia who came to Italy some 500 years ago (Breu 2012: 275). What makes Molise Slavic particularly interesting is the fact that it has developed a full-fledged indefinite article, but no definite article – a fact that places Molise Slavic in an unusual and perhaps unique position in Europe (Wiemer and Wälchli 2012: 31). The indefinite article has developed from the numeral ‘one’, which in Molise Slavic is jena (masculine singular nominative long form, Breu 2012: 279). Breu shows that jena has developed the referential, non-referential and generic functions one would expect from indefinite articles in “article languages”. Since Molise Slavic has been in contact with Italian for centuries, it seems likely that the emergence of an indefinite article in Molise Slavic is a contact phenomenon. Breu (2012: 310) notes the strong tendency for Slavic languages not to develop

5 The construction is also known as “double verbs”. The origin of the most well-known example, the formula žil-był from fairytales, is contested, but will not be discussed here (see Petrusin 2007 for critical discussion).

6 Breu (2012) also discusses Upper Sorbian, but I will not review this part of his article, since the data from Molise Slavic are sufficient to illustrate the relevance of language contact for Slavic historical linguistics.
articles, and hypothesizes that this “diachronic constant” can only be overridden in situations of extensive language contacts.

In a slim volume containing five articles, Berger (2008) revisits a number of issues in the history of Czech – both phonological and grammatical – from a contact perspective. Among the phonological phenomena considered is the development of fixed stress in West Slavic, for which Berger (2008: 18) postulates a Proto West Slavic system with immobile stress on any syllable except the last. According to Berger, under influence from German and Hungarian this system developed into initial stress in Czech and Slovak, whereas Polish, which he argues was less strongly influenced by German, developed penultimate stress. With regard to grammatical issues, Berger (2008: 68–69) clarifies the criteria for how impact through language contact can be established and classifies the relevant phenomena according to how likely they can be related to language contact. Throughout the book, Berger stresses that internal factors and external factors (e.g. language contact) are not mutually exclusive, but rather go hand in hand as explanations of language change.

4. Birch bark letters – how do new data change the field?

Although it has been more than fifty years since the first birch bark letter was excavated in Novgorod, it took time before they became acknowledged as a large and important source of data in Slavic historical linguistics. However, today the birch bark letters have an obvious and central place in university courses of East Slavic historical linguistics, and they are discussed extensively in contemporary textbooks (e.g. Galinskaja 2014: 67–70, Nesset 2015: 291–301). It is time to ask how the birch bark letters have changed Slavic historical linguistics.

The birch bark letters are meticulously documented and analyzed in Zaliznjak’s (2004) massive monograph of more than 850 pages. The first part of the book, comprising approximately 200 pages, is a grammatical sketch of the Old Novgorod dialect, which includes chapters on phonology, inflectional morphology, and syntax, as well as word-formation and lexicon. In the second part Zaliznjak offers 400 pages of detailed analyses of individual birch bark letters, organized chronologically in five periods. The last 200 pages or so contain a number of indexes that make the book extremely useful as a reference work.

The birch bark letters have brought to the attention of Slavic historical linguists a number of interesting phenomena, including:

(3) a. The absence of traces of the second palatalization
   b. The enigmatic nominative singular ending ‘e
   c. The distribution of clitics
   d. The tense system

Examples such as this from birch bark letter no. 247 testify to the absence of the second palatalization (Zaliznjak 2004: 239–40):

(4) А замъке къле а дьри къл […]
   ‘But the lock is whole and the doors are whole […]

While other varieties of Slavic have the affricate /c/ before /ě/ due to the second palatalization in the word for ‘whole’ (cf. modern Russian целый), the Old

7 Notice that Zaliznjak 2004 is available electronically at http://gramota.ru/, which also includes a searchable archive of birch bark letters from Novgorod and other cities. Birch bark letters excavated between 2001 and 2014 are described and analyzed in Janin et al. (2015).
Novgorod dialect preserved /k/ as in кълне and кълън in the example above. What is the explanation for this? While the most likely interpretation is perhaps that the second palatalization simply never reached the northeastern periphery of the Slavic language area, an alternative interpretation cannot be excluded. Vermeer (2000) proposes instead that it was the monophthongization of diphthongs that reached Novgorod too late to trigger the second palatalization.

According to this scenario, the word for 'whole' still contained a diphthong in Novgorod when the second palatalization was active. According to Vermeer, monophthongization took place in Novgorod when the second palatalization was no longer active. While Vermeer’s analysis is interesting and elegant, there does not, however, appear to be much independent evidence to support it.

Another interesting characteristic of the example замъкe кълнe ‘the lock is whole’ in (4) is the nominative singular ending –е, as opposed to the expected –ъ. The –е ending not only lacks parallels in other varieties of Slavic, but also seemingly did not trigger the first palatalization, as shown in замъкe where the ending is preceded by the velar stop /k/ instead of the affricate /č/. While this phenomenon raises many questions and can be interpreted in a variety of ways, no generally accepted solution seems available at present. However, in a recent study Olander (2012), who gives a thorough review of the literature on the topic, argues that Pre-Slavic *ə as gave Proto-Slavic *а, which turned into –е in the Old Novgorod dialect, while producing –ъ in all other varieties of Slavic.7

Although the study of clitics has been a blooming field in Slavic linguistics (see e.g. Franks and King 2000), the East Slavic languages have not received as much attention as their sister languages in the west and the south. This is partly because the modern East Slavic languages do not display rich inventories of clitics, but also because the available data from Old Russian has been limited. The birch bark letters have to some extent changed this situation, as pointed out by Zaliznjak (2008b: 3). Zaliznjak, who draws on data from birch bark letters and other sources, proposes the following hierarchy of Old Russian enclitics:

7 For a general overview of Proto-Slavic inflectional morphology, see also Olander 2015.

(5)

a. же
b. ли
c. бо
d. ти
e. бы
f. dative pronouns: мн, ти, си, ны, вы, на, ва

g. accusative pronouns: мя, тя, ся, ны, вы, на, ва, и, ю, е
h. auxiliary verbs: емь, еси, есть, etc.

All these clitics are integrated in the first phonetic word of the clause according to Wackernagel’s law. In examples with multiple clitics, the clitics occur in the order predicted by the hierarchy in (5); items higher up in the hierarchy precede items further down. Thus, in the following example же precedes си:

(6) Крести же си въ церкви святое Софьи. (Primary Chronicle 988 AD)

‘He was baptized in the Church of St. Sophia.’

However, two factors complicate the picture. First, Zaliznjak’s generalizations work best for texts that are close to the spoken language, which once again testifies to the significance of the birch bark letters. Second, Zaliznjak’s analysis is complicated by what he calls ритмико-синтаксические барьеры ‘rhythmic-
syntactic barriers’ (Zaliznjak 2008b: 47). Such barriers do not tolerate clitics to their left, and therefore force clitics to be placed closer to the end of the clause than Wackernagel’s law would lead us to expect. Here, Zaliznjak’s analysis may provide valuable material for more theoretically inclined Slavic historical linguists. However, Zaliznjak (2008b) himself does not refer to any recent theoretical works on clitics, neither in Slavic, nor in general linguistics.

According to Zaliznjak (2004), after the loss of the aorist and imperfect there were three tenses in the Old Novgorod dialect, viz. the present, past and the pluperfect. Andersen (2006c) takes issue with this analysis, and argues that the birch bark letters provide evidence for two more tenses – the perifrastic future and the future perfect. First, Andersen considers examples with the phase verb počnu + infinitive, and argues that they “call for straight future readings” (Andersen 2006c: 74). Second, Andersen analyzes the budu + I-form construction, which Zaliznjak (2004: 134 and 177) has described as a presuppositional mood. Andersen instead suggests that budu + I-form represents a future perfect, describing something that happened before a point of time in the future. If Andersen’s analysis is correct, the upshot is that “the Old Novgorod tense system, as it is reflected in the birch bark letter corpus, was not different from that of other Old Russian dialects” (Andersen 2006c: 86).

The phenomena listed in (3) above and many other linguistic patterns found in the birch bark letters indicate how this new body of data has changed the field. The new data has sparked vivid discussions of competing hypotheses and raised questions, some of which are yet to receive generally accepted solutions. However, the birch bark letters are also significant on a more general level, since they have implications for the shaping of the modern Russian language as a whole. Although Novgorod lost the political struggle with Moscow in the middle ages and ended up as a small provincial town, it does not follow from this that the impact of the Old Novgorod dialect was negligible in the shaping of the modern Russian language. A case in point is the second palatalization. As is well known, Contemporary Standard Russian does not show any traces of the second palatalization in the nominal paradigm (cf. e.g. dative/locative forms like руке of рука ‘hand’), whereas Belarusian and Ukrainian have kept the palatalization in the corresponding forms. The question is why the analogical leveling took place in Russian, but not in Belarusian and Ukrainian. A possible answer is that Russian was more susceptible to influence from the Old Novgorod dialect documented in the birch bark letters.10

5. Grammaticalization theory – new light on old data?

Grammaticalization is a field with long traditions; the term goes back to Meillet (1912), and the classical definition in (7) was proposed by Kuryłowicz (1965 [1975], see also Lehmann 1995):

(7) Grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status, e.g. from a derivative formant to an inflectional one. (Kuryłowicz 1965 [1975]: 52)

For a recent interview with Andrej Zaliznjak where he discusses the role of the Old Novgorod dialect in the shaping of the Russian language, the reader is referred to http://www.onlinetv.ru/video/1607/?playFrom=248.
Grammaticalization received renewed interest toward the end of the twentieth century (see e.g. Heine et al. 1991 and Hopper and Traugott 1993), and has established itself as an important, but at the same time controversial subfield of linguistics in the twenty-first century (Joseph 2004, Dickey 2012). Hence, the present article would be incomplete without a brief discussion of grammaticalization. Two questions are important: What has grammaticalization theory to offer Slavic historical linguistics? Conversely, what can Slavic historical linguistics offer grammaticalization theory?

As for the first question, grammaticalization theory provides a unified framework for describing and explaining a wide variety of phenomena that would otherwise not have been put on a common denominator. Relevant phenomena include:

(8)  
   a. "Long forms" of adjectives
   b. Articles
   c. Future tense
   d. Resultatives
   e. Modal auxiliaries
   f. Passive

If we take grammaticalization theory seriously, we must look for general principles that all these developments adhere to. Furthermore, we must ask whether there is one shared driving force underlying all these developments. Clearly, such considerations may lead to rethinking the traditional analyses of many phenomena in Slavic historical linguistics.

As an example of the usefulness of grammaticalization theory, let us consider the development of the future tense across Slavic. While Common Slavic did not have an inflected or periphrastic future tense, a number of different systems have developed in the modern Slavic languages due to grammaticalization. With regard to the source of the future tense, Andersen (2007) distinguishes between four types:11

(9)  
   a. De-inceptive auxiliaries: East Slavic, e.g. Old Russian počinu 'I begin' + infinitive
   b. De-modal auxiliaries: South Slavic, e.g. OCS xotēti 'want' + infinitive
   c. Copular auxiliary + l-participle: Slovenia/Croatian dialects and parts of the Polish language area, e.g. Slovene bom stavil 'I shall place'
   d. De-existential auxiliary + infinitive: West Slavic, e.g. Polish będę pracować 'I will work'

Later developments have obfuscated the neat east – west – south distribution of types (9a, b, d). For instance, the de-existential type of (9d) has replaced the de-inceptive type of (9a) in most of the East Slavic language area (Andersen 2007: 132).

All developments in (9) are examples of grammaticalization, since full-fledged lexical verbs turn into auxiliary verbs with specific grammatical functions. In some cases, the grammaticalization has gone further. Notably, the de-modal auxiliary verb has developed into an enclitic in Croatian and Serbian and a proclitic in Macedonian and Bulgarian (Wiemer 2012: 744–745). Ukrainian

---

11 This exposition is somewhat simplified. Examples are from Andersen 2007: 132–135. In Andersen 2006b: 12–13 a slightly different classification is offered, but the substance of the analysis is the same.
has developed a synthetic future based on Common Slavic jeti ‘take’ (Wiemer 2012: 745), which Andersen (2007: 132) subsumes under the de-inceptive type.

Unifying all the developments in (9) as examples of one phenomenon – grammaticalization – helps us see the spatial and temporal relationships between them more clearly. Importantly, Andersen’s analysis has implications for the study of aspect. He notes that de-inceptive auxiliaries combine with imperfective verbs only, which implies that the perfective-imperfective distinction must have been in place before the emergence of the periphrastic future (Andersen 2006a: 238 and 2007: 132). We will return to this point in section 6.

A debated issue in grammaticalization theory in recent years is the relationship between grammaticalization and language contact (Wiemer and Wälchli 2012: 25–27). What is the impact of contact on grammaticalization? The grammaticalization of the so-called recipient passive in Czech and Upper Sorbian sheds light on this question. As illustrations of recipient passives Giger (2012: 560) offers examples with Czech dostat ‘get’ (10) and Upper Sorbian dostać ‘get’ + a participle (11):

(10) Dostal jsem doporučen pobyt na venkově.

‘I was recommended a stay in the countryside.’

(11) Wón dostanje wot wšětkich pomhane.1

‘He is helped by everyone.’

Giger (2012: 560) analyzes the recipient passives in West Slavic as the result of influence from German, which has constructions with bekommen ‘get’, kriegen ‘get’ and erhalten ‘get’ + a participle: Sie bekommt/kriegt/erhält den Katalog zugeschickt ‘She gets the catalogue sent’. Although on the face of it the situations in Sorbian and Czech appear parallel, there are important differences. Giger (2012: 579) concludes: “While in Sorbian the construction perfectly resembles its German counterpart (including stylistic differences between dialectal use and use in the standard language), in Czech there are some peculiarities in the choice of the full verbs used in the recipient passive that do not agree with either Standard German or East Middle German dialects (as far as the situation in the latter is known).” The reason behind the strong parallelism between Upper Sorbian and German may be the extensive language contact with a high degree of bilingualism, which may have facilitated parallel grammaticalization in the two languages (Giger 2012: 576, Wiemer and Wälchli 2012: 26). The contact between Czech and German is arguably less extensive, and this may have given rise to a more independent development of the recipient passive in Czech, once the construction had been transferred from German. Regardless of whether these hypotheses hold up to further scrutiny or not, the West Slavic recipient passives show that Slavic historical linguistics has the potential to inform grammaticalization theory.

6. Aspect

An important characteristic of a Slavist with a modicum of self-respect is his/her obsession with the grammatical category of aspect. While the study of aspect in the modern Slavic languages has yielded a number of important publications in recent years (e.g. Dickey 2000, Zaliznjak and Smelj 2000, Janda et al. 2013), significant advances have also been made in Slavic historical aspectology.
A long-standing issue is the question as to when and how the aspectual distinction between the perfective and the imperfective arose and developed. Was this distinction in place already in Common Slavic (Dostál 1954), or is it of much more recent origin (see e.g. Bermel 1997 and Nørgård-Sørensen 1997)? In the previous section, we saw that Andersen’s (2006a-b and 2007) analysis of the periphrastic future lends support to an early birth of the perfective-imperfective distinction. In what follows, we will consider another argument from a recent study by Eckhoff and Janda (2014), which also illustrates the potential of corpus data and statistical analysis for Slavic historical linguistics (section 2 above).

Based on a dataset of more than 15,000 OCS verb attestations extracted from the PROIEL corpus (described in section 2 above), Eckhoff and Janda carried out a detailed analysis of the frequency distributions of various grammatical forms. Correspondence analysis, a statistical model that groups the verbs using as few dimensions as possible, shows that the OCS verbs fall into two neat groups that by and large conform to the classification into perfective and imperfective verbs proposed by Dostál (1954). This result corroborates the hypothesis of an early provenance of the perfective-imperfective distinction in Slavic.

While the studies by Andersen and Eckhoff and Janda discussed above suggest an early birth of the perfective and imperfective aspects, it does, of course, not follow that the difference between the two aspects was the same in Late Common Slavic as it is today. Dickey (2000) has shown that there are important differences among the Slavic aspectual systems, so it seems clear that the development of the perfective-imperfective distinction must have followed different paths in different parts of the Slavic language area (Dickey 2015). An example is atelic perfectives, i.e. verbs such as modern Russian читать 'read for a while', which are widespread in East Slavic, but less so in West Slavic. Dickey (2008) argues that atelic perfectives represent a relatively late development in Russian, which suggests that the perfective-imperfective distinction has undergone expansion over time so as to cover more and more types of verbs. It is worth mentioning that one of the factors that according to Dickey (2011 and 2015) has contributed to the east-west division in Slavic aspect is language contact between Germanic and Slavic in the west (see my discussion of contact phenomena in section 3 above).

7. Constructions and Construction Grammar

What is the basic unit of language? In the twentieth century, most linguists probably would have voted for the morpheme or the word, or maybe the sentence. However, towards the end of the twentieth century the notion of “construction” received more and more attention, inter alia through the work of Fillmore (e.g. 1988) and Goldberg (1995 and 2006). To what extent has the emergence of Construction Grammar had an impact on Slavic historical linguistics?12

Construction Grammar regards as constructions all form-meaning pairs that are unpredictable and/or sufficiently frequent to be stored as separate units in the speakers’ mental grammar (Goldberg 2006: 5). In view of this very inclusive definition, it comes as no surprise that Slavists have always been working on constructions, and we have seen examples of that earlier in this overview, e.g. Weiss’ (2012) analysis of the Serial Verb Construction in Russian (see section 3 above). In the following, I will explore a recent contribution by Mirjam Fried,

12 For an up-to-date overview of Construction Grammar and historical linguistics in general, see Hilpert 2015: 359–361.
which illustrates the potential of Construction Grammar as a framework for the study of the history of the Slavic languages.

Fried (2015) analyzes the historical development of the present active participle in Old Czech in the direction of an adjective. Three functions are relevant (Fried 2015: 11):

12) a. Predicative: přišel jsem "(I came) believing"
   b. Attributive: přišel-člověk "a believing person"
   c. Nominalized: přišel-č "a believer"

All three functions are well attested in Old Czech, but the predicative function in (12a) became marginalized, while the attributive and nominalized functions came to dominate Modern Czech (see Fried 2015: 11–12 and 151). Fried analyzes this gradual shift of syntactic category from verb (participle) to adjective in terms of a tension between word meaning and syntactic constructions, whereby the verbal meaning of the word gets suppressed by the nominal syntactic constructions the erstwhile participles occur in. Two theoretical points are important. First, Fried argues that the locus of the change is the construction, since the shift from verb to adjective cannot be understood properly without taking the relevant syntactic constructions into account. Second, it is argued that the rich representations of Construction Grammar facilitate a precise and explicit analysis of the interaction between lexical meaning and syntactic construction – and the gradual change of this interaction over time.

8. Language typology

A worst-case scenario for Slavic historical linguistics would be to analyze the Slavic languages as if they were the only languages in the world. Toward the end of the twentieth century, language typology became a major subfield of linguistics – also in the Slavic world. As an illustration of the cross-fertilization between typology and Slavic historical linguistics, I will discuss recent work on the Old Russian pluperfect. In Old Russian texts we find two kinds of pluperfects, which Ščinava (2013: 187 and 189) labels "regular" and "extra-complex" ("сверхсложный"):

13) a. "Regular" pluperfect: Imperfect of быти + l-participle (e.g. бяше приложилъ 'had been on friendly terms'
   b. "Extra-complex": Perfect of быти + l-participle (e.g. сказалъ есть былъ 'you had said')

The term "extra-complex" reflects the fact that this type consists of three verb forms, an auxiliary + l-participle of быти + l-participle of the main verb, although in many text examples the auxiliary is omitted (i.e. сказал есть был instead of сказал есть был).

The meaning and use of the extra-complex pluperfect has been the topic of a lively debate in recent years (Petruxin and Ščinava 2006 and 2008, Ščinava 2007 and 2013, Ševeleva 2007 and 2008), partly because the birch bark letters have increased the body of available examples (Ščinava 2013: 191). Ščinava and

13 Since many readers probably equate language typology with the typology of grammatical systems, it is worth pointing out that lexical typology, i.e. "the systematic study of cross-linguistic variation in words and vocabularies" (Koptjevskaja-Tamm 2012: 373), has received considerable interest in recent years – also in the Slavic world (e.g. Reznikova et al. 2012). For a diachronic analysis in this field, see Rakhilina (2007), who addresses the development of Russian verbs denoting motion in water.
Petruhin argue that the extra-complex pluperfect developed a number of meanings for which they use the Russian cover term "неактуальное прошедшее". In English, the relevant meanings are sometimes referred to as "discontinuous past" (see e.g. Plungian and van der Auwera 2006, Sičinava 2013: 40), and this usage will be adopted in the following. By way of example, consider the following sentence from Xaždenie igumena Danila (cited after Petruhin and Sičinava 2006: 203, Sičinava 2007: 114, and Sičinava 2013: 197):

(14) И в той келии Лазарь болъл, ту же умеръ быль.

'And in that cell Lazarus was ill, in there he had died.'

Although the result of dying normally cannot be cancelled afterwards, this is exactly what happened with Lazarus, who was brought back to life after he had died. Accordingly, Petruhin and Sičinava interpret (14) as an example of "cancelled result", the most frequent subtype of discontinuous past for the Old Russian extra-complex pluperfect (Sičinava 2013: 193–194).

The typological perspective is crucial for Petruhin and Sičinava's analysis, because the development of "cancelled result" and other discontinuous past meanings is widespread for (extra-complex) pluperfects cross-linguistically (Sičinava 2013: 28–42), in particular in verb systems such as the Old Russian, where the perfect tense is unstable or disappears (Sičinava 2013: 165). However, the typological perspective informs not only the analysis of the Old Russian pluperfects, but is also helpful for the analysis of analogous constructions in other Slavic languages. Sičinava (2013: 127 and 158) argues that there is a "European area" of "extra-complex" pluperfects comprising many varieties of Romance and Germanic languages, as well as Basque, Breton and Albanian, and an "Asian area" of similar constructions covering a number of Iranian and Turkic languages, as well as languages in the Volga area, the Caucasus and the Balkans. Sičinava (2013: 165–166) submits that the Slavic extra-complex pluperfects represent an intermediate case, which has some features typical of the European area, but share other features with constructions from the Asian area. In particular, the Slavic constructions share the tendency to lose the auxiliary with corresponding constructions from the Asian area, while semantically the Slavic constructions are closer to the European area, for which the development of discontinuous past meanings is characteristic.

While the extra-complex pluperfects illustrate the value of language typology for Slavic historical linguistics, they also testify to the relevance of Slavic historical linguistics for language typology. Without a proper understanding of the Slavic constructions, the connection between the European and Asian areas of extra-complex pluperfects would remain unclear.

9. Phonology

Phonology has always been a cornerstone in Slavic historical linguistics. Of particular importance, perhaps, is the strong tradition of prosodic studies from Stang (1957) via Ilč-Svityč (1963) and Dybo (1981) to Zaliznjak (1985, see also the more theoretically oriented monograph by Bethin 1998). A recent contribution to this field is Zaliznjak's (2014) monograph Древнерусское ударение. The first part of the book offers a brief but up to date overview of
Zaliznjak’s approach to stress in Old Russian with numerous tables and maps, as well as an instructive overview of relevant sources. The second part is an extensively annotated dictionary akin to Zaliznjak’s celebrated Grammatičeskij slovar’ russkogo jazyka (1977).

Recent years have also seen some renewed interest in Balto-Slavic prosody in an Indo-European context. A case in point is Olander (2009), who proposes a new analysis of accentual mobility in Balto-Slavic based on the phonological properties of the desinential syllables (Olander 2009: 2–3). Pivotal in Olander’s approach is what he refers to as the “Mobility Law”, which encompasses a change from high to low pitch in the final position of a phonological word (Olander 2009: 202). Although this approach has a number of advantages, it “requires a certain amount of analogical levelling in order to account for the facts”, as the author himself admits (Olander 2009: 202).

In the previous sections, we have seen several examples where new theoretical ideas give birth to innovative analyses when applied to well-known data. Parallel stories are found in the field of phonology as well. By way of example, consider Padgett’s (2003) analysis of the sound change from [ky, gy, xy] to [k’i, g’i, x’i] in Old Russian. One may ask if this change was triggered by the vowel or the consonant, but Padgett’s analysis shows that this is not the right question to ask. Instead, he opts for a systemic approach influenced by Fleming’s (1995) Dispersion Theory and couched in the broader framework of Optimality Theory (Prince and Smolensky 2004). For present purposes, a detailed account of the theoretical approach is not called for, because the gist of Padgett’s analysis can be explained without reference to theoretical technicalities. The basic idea is that languages strive to make contrasts as easily perceptible as possible – what Padgett (2003: 50) refers to as “goodness of contrast”. In simple terms, the difference between K’i and Ku is easier to hear than the difference between Ky and Ku. (K here stands for any velar consonant). The change from Ky to K’i thus maximized the goodness of contrast in the phonological system, a change that can be represented in terms of reranking of constraints in Optimality Theory.

Another classical problem of Slavic historical linguistics that has received some attention in recent years is the fall and vocalization of the jers (yers). Kavitskaya (2002, 2005), who follows Timberlake (1983a–b), relates the jers to the compensatory lengthening processes that took place in Late Common Slavic. Jers in final syllables were particularly prone to be deleted, and when the jers disappeared the preceding vowel was lengthened. Kavitskaya (2002, 2005) analyzes this as the phonologization of phonetic duration. Vowels in open syllables tend to have longer duration than vowels in closed syllables. When the word final jer in words like sъnъ ‘dream’ fell, the remaining jer ended up in a closed syllable, where it had an unusually long duration. Because of its unexpected long duration, according to Kavitskaya (2002, 2005) the remaining jer was reanalyzed as a “normal” non-jer vowel. The analysis of the jers in terms of compensatory lengthening shows that the jers formed disyllabic units where the length of the first jer depended on the fall of the word-final jer. Nesset (2016) argues that these disyllabic units may be analyzed as trochaic feet and discusses the implications of this idea for the analysis of the fall and vocalization of the jers.
10. Concluding remarks: Theory, method and data

In this overview article, I have told eight stories about Slavic historical linguistics in the twenty-first century. All storytelling is to some extent subjective and constrained by the storyteller’s personal interests and competence, so I have no illusions of having created an objective or complete picture of the situation. However, my account has brought out some salient tendencies – and has demonstrated that Slavic historical linguistics is alive and kicking.

In view of the fact that science in general and linguistics in particular is an interplay between theory, method and data, it is perhaps not surprising that most of my stories boil down to (combinations of) three archetypes: (a) new theoretical concepts spark new analyses, (b) new methods facilitate new analyses, and (c) new data form the basis for new analyses. While all these stories are important, I find the rapid methodological developments we are witnessing particularly significant. This is why this article is called “when we went digital”. If the twentieth century was the century of theoretical speculation in linguistics, maybe the twenty-first will be the century of methodologically innovative, empirically driven approaches?

References


UiT – The Arctic University of Norway
Department of Language and Culture
NO-9037 Tromsø
Norway
tore.nesset@uit.no