

## Birds & Beasts: Shaping Events in Old Russian

### 1. Relevance relative to the call for proposals: Four levels

- (1) Družinu tvoju, knjaže, pticъ krily priodě, a zvěri krovъ polizaša.  
‘O prince, birds have set wings on your guardsmen, and beasts licked the blood.’

In this gruesome quote from *The Tale of Igor’s Campaign*, one of the masterpieces of Old Russian literature, Izjaslav reports that birds and beasts are feasting upon the bodies of the fallen after a massive defeat.<sup>1</sup> The birds and the beasts are engaged in two different kinds of events. Whereas the birds have carried out a goal-oriented action, setting their wings on the dead soldiers, the beasts have lapped at the blood, an event that is not necessarily goal-oriented. The Birds & Beasts Project explores the historical development of two crucially different ways of conceptualizing the shape of events in Russian, addressing key questions of linguistic typology by developing empirical databases and applying statistical methods to their analysis.

The Birds & Beasts Project addresses the shaping of events on four levels. On the level of **linguistic theory**, we seek to expand existing knowledge about the typology of verb classifier systems. On the **empirical level**, we propose creating a corpus of Old Church Slavic and Old/Middle Russian that will parallel the architecture of other corpora, facilitating linguistic analysis and comparison. On the **level of methodology**, we apply linguistic profiling to corpus data in order to explore form-meaning relationships and evaluate these relationships by means of statistical methods. On the **applied level**, this project has impact on the teaching and learning of Russian, a key component in the University of Tromsø’s strategy for building competence relevant to the Barents Region (“Nordområdene”).

### 2. Aspects relating to the research project: Shaping events

Human languages describe a variety of types of events, and these can be grouped as telic (goal-oriented) vs. atelic (not goal-oriented). The telic vs. atelic (cf. Greek *τέλος* ‘goal’) distinction, first formulated by Aristotle and ubiquitous in linguistics since the work of philosopher Zeno Vendler (1957), has had a major influence on linguistic theories of verbal aspect. Some verbs, like ‘give’, describe events that are inherently telic; setting wings on the guardsmen in (1) above is also a telic event, since it results in the arrangement of wings upon bodies. Other verbs, like ‘moan’, are inherently not goal-oriented (atelic), since they lack an inherent endpoint. There are ambiguous verbs that can yield either telic or atelic construals, such as ‘write’, which is telic in the case of ‘to write a book’, but atelic in descriptions of general activities as in ‘learning how to write’.<sup>2</sup>

Approximately half of the world’s languages (Dahl 1985; for a map see <http://wals.info/> and request Feature 65A: Perfective/Imperfective Aspect) distinguish perfective vs. imperfective aspect in verbs. Perfective forms describe events that are construed as discrete and complete, and imperfective forms describe unbounded states and activities. Slavic languages such as Russian mark the perfective vs. imperfective distinction with a complex system of about two dozen affixes.

One would expect perfective forms to express telic events and imperfective forms to express atelic events, and both of these expectations are confirmed. In Russian the phrase ‘she wants to write a book’ is *ona xočet napisat’ knigu*, using the prefix *na-* to mark as perfective the verb *napisat’* ‘write’. Contrast this with ‘she is learning how to write’ *ona učitsja pisat’*, where we have the unprefixated imperfective verb *pisat’* ‘write’.

<sup>1</sup> For over two centuries since *The Tale of Igor’s Campaign* was first published in 1800, there have been doubts as to its authenticity. However, Zaliznjak’s (2008) masterful analysis of a variety of factors including clitic placement, dialectal features and scribal errors has laid these doubts to rest.

<sup>2</sup> Notice that our use of “telic” implies that the situation in question has an inherent goal, but not necessarily that this goal is attained. We acknowledge that other scholars (e.g. Krifka 1989, 1998) use “telic” more restrictively so as to presuppose the actual attainment of the goal inherent in the situation.

However, that is only part of the story. Slavic languages have developed secondary imperfective forms that serve as specifically telic imperfectives, usually expressing either processes or repeated events, as in *perepisyvat* ‘rewrite’ (the secondary imperfective of the perfective *perepisat* with the same meaning), used to describe a rewriting process or a repeated action of rewriting. A special characteristic of Russian is that one can additionally use perfective forms to express atelic events. For example it is possible to take the imperfective verb *stonat* ‘moan’ and add the prefix *po-* to create the perfective verb *postonat*, which means ‘moan for a while, do some moaning and then stop’. Russian has a very productive capacity for creating what we call “procedural” verbs (also known as Aktionsarten) that place boundaries on activities and states, making it possible to present them as perfectives. The Modern Russian translation of the second clause in (1) contains the parallel procedural verb *polizat* with the prefix *po-* added to the imperfective *lizat* ‘lick’ to create the atelic perfective ‘lick for a while’. However, procedurals of this type were rare in Old Church Slavic (a collection of 9th-11th century texts that represent a close equivalent to a protolanguage of the Slavs), and it is uncertain exactly how and when procedurals emerged in Old Russian (cf. Dickey 2007). Thus, the Old Russian perfective verb in (1), *polizaša* might have two interpretations: a telic one meaning ‘licked up’, and an atelic one meaning ‘licked for a while’. Indeed Old Russian dictionaries disagree: Sreznevskij (1893) offers only the telic interpretation, while the Academy Dictionary (Avanesov et al. 1975-) equivocates, offering both options.

The evolution of a system including both telic and atelic perfectives and imperfectives in Russian has important implications for the typology of verb classifier systems. We seek to broaden the understanding of such systems both by exploring the origins of the system in Old Church Slavic and Old/Middle Russian, and by contrasting the Modern Russian system to non-related systems that differ from it in crucial ways.

## 2.1. Background and status of knowledge: What needs to be done?

The Birds & Beasts project presents four areas in which we have cutting-edge competence and are thus poised to contribute new knowledge and promote scientific renewal. The areas are: the **typology of verb classifier systems**, the building of **linguistic corpora**, the implementation of **statistical models to support linguistic theory**, and **language pedagogy**.

Whereas noun classifier systems are fairly well understood and have been identified in many languages, particularly in South America and Asia (Aikhenvald 2000: 98-124; for a map of world-wide distribution, see <http://wals.info/> and request Feature 55A: Numeral Classifiers), linguists have only recently become aware of **verb classifier systems**. McGregor (2002: 404) states that verb classification “has not yet been incorporated into mainstream linguistic knowledge as a category that might be expected in a language”, but that it is “a far from exotic phenomenon” and “[d]oubtless it is not confined to the relatively few languages in which it has been hitherto described, though the extent of its distribution across the world’s languages remains to be charted.” Our research has launched the new proposal that Russian be considered a verb classifier language (Janda et al. to appear), marking a radical departure from traditional scholarship on Russian aspect. The Birds & Beasts project will build upon and deepen this line of research by adding two key dimensions: a **diachronic perspective** to achieve a better understanding of how such a system could arise, and a **contrastive perspective** to highlight key features of the Russian system by comparing it to languages (notably Sámi and Norwegian) which differ from it in interesting ways. We will pay special attention to the **atelic perfectives** and **telic imperfectives**, both diachronically and contrastively, since our previous studies focused on the role of prefixes in forming telic perfectives.

On the empirical level, Old Church Slavic and Old/Middle Russian have traditionally been studied on the basis of small sets of examples, gathered by individual scholars working with different and often implicit criteria. Although this approach has provided a solid understanding of Old/Middle Russian phonology and morphology, the development of the Slavic aspectual system is still mired in controversy. While some researchers claim that the Modern Russian system was

inherited from Common Slavic, others claim that it took on its current shape as late as in the 16<sup>th</sup>-17<sup>th</sup> centuries (Bermel 1997, Nørgård-Sørensen 1997). In order to shed new light on the issue, quantitative analysis of more extensive data is required. In other words, what is needed is a large **electronic corpus of Old and Middle Russian**.

Corpus linguistics holds a strong position in Russian and Slavic linguistics, and the resources for the modern Slavic languages are excellent, in particular, the Russian National Corpus represents a powerful tool for the study of Modern Russian. However, much work remains before we have satisfactory resources for earlier language stages. The Old Church Slavic corpus resources are limited, and the Old/Middle Russian resources are scattered and have limited markup. For Old Church Slavic, the only corpus with lemmatization and full morphosyntactic annotation is the PROIEL corpus, which consists of about 70 000 words. Several Old/Middle Russian corpus resources exist or are under development, but none of them have linguistic markup beyond lemmatization and morphological annotation.<sup>3</sup> In cooperation with the Russian National Corpus and the Bulgarian Academy of Sciences, we propose to expand the PROIEL corpus to a unified and freely accessible corpus of Old Church Slavic, Old Russian and Middle Russian. The corpus will have not only lemmatization and morphology, but also syntactic annotation in the same dependency grammar format, and with specialized annotation for the verbal derivational morphology that is the main objective of the Birds & Beasts project. Syntactic annotation is crucial for us to be able to use the corpus in constructional profiling, a methodology that is pivotal in the Birds & Beasts project (cf. section 2.3 below). The broad, generalized annotation will make the corpus a unique resource of international interest, both for linguistic studies and for the further development of automatic and semi-automatic tools for the analysis of Old Church Slavic and Old and Middle Russian.

Previous work on the PROIEL corpus has made available a large database of Old Church Slavic forms paired with grammatical information, and this architecture can easily be adapted to the morphological peculiarities of Old Russian texts. We also have accumulated considerable experience with and guidelines for annotating early Slavic syntax, which makes a quick corpus expansion very feasible.

With the advent of large electronic corpora, a need for methodological renewal has arisen in linguistics: new quantitative methods are required that facilitate analysis of large datasets from corpora. In response to this challenge, in the CLEAR (Cognitive Linguistics: Empirical Approaches to Russian) research group at the University of Tromsø we have developed a suite of strategies, called **“linguistic profiling”**, for using corpus data to explore form-meaning relationships in language. These strategies are inspired by behavioral profiling (Divjak and Gries 2006; Gries and Divjak 2009), and have been tailored to focus on specific theoretical issues. Linguistic profiling includes a variety of techniques that make it possible to operationalize hypotheses about the meanings of linguistic forms by quantifying relevant patterns in corpus data. Our data is analyzed by means of **statistical models** available through the “R” open source software project (<http://cran.r-project.org/>). Despite some interest in recent years, the use of statistical models is still in its infancy in Russian and general linguistics and practically absent in Slavic historical linguistics, and the linguistic profiling methodology needs to be further refined and tested against new data sets.

In language pedagogy, university courses in Modern Russian have changed radically in recent years, but nothing much has happened with Old Church Slavic and Old/Middle Russian. Students of Modern Russian have a number of multimedia resources at their disposal in addition to traditional grammars and textbooks. Old/Middle Russian and Old Church Slavic, on the other hand, are taught

---

<sup>3</sup> The most important resources are the following: The Russian National Corpus’ upcoming Old Russian subcorpus, which is lemmatized and has morphological annotation; the Manuskript corpus at the State University of Udmurtia (<http://www.manuscripts.ru/>), which has a number of texts with lemmatization and partial morphological information; the digital library of Old Russian texts (particularly chronicles) of The Institute of Russian Language of the Russian Academy of Sciences, with lemmatization and morphological annotation. There is also the Regensburg diachronic corpus of Old Russian, which has a set of partially lemmatized texts from different stages of Russian.

the same way they have been taught since the Neogrammarian August Leskien published his *Handbuch der Altbulgarischen Sprache* in 1871. In other words, students read excerpts of texts and are presented with textual and grammatical commentary, backed up with traditional grammars and dictionaries. We wish to exploit our corpus to bring Old Church Slavic and Old/Middle Russian pedagogy up to date, because it makes it possible to develop online teaching materials where students can create their own datasets and analyze them. A corpus with syntactic analysis will also make it possible to move away from the traditional focus on phonology and morphology alone. We also wish to present the corpus online in such a way that it will be a reading aid for students and scholars who are not specialists in Early Slavic, but want a diachronic or typological angle on some linguistic phenomenon. These practical applications will make the Birds & Beasts corpus a resource of international interest.

## 2.2. Approaches, hypotheses and choice of method: The Verb Classifier Hypothesis and linguistic profiling

Russian has developed an aspectual system that has both telic and atelic perfectives and imperfectives, and we aim to probe this development by tracking its progress through Old/Middle Russian. We hypothesize that the telic and atelic perfectives parallel two main types of classifiers. In order to understand our Verb Classifier Hypothesis, it is best to start with nouns, and the most relevant parallel is with numeral classifier systems (which are a type of noun classifier system, the second most common after gender). A language like English (which does not have obligatory numeral classifiers) has count nouns that describe discrete objects like *candle*, *stick*, *leaf* as opposed to mass nouns that describe substances like *wax*, *wood*, *dirt*. The two types of nouns behave differently: count nouns have a singular vs. plural distinction that mass nouns lack (*\*dirts*) and mass nouns appear in quantifier constructions (*a pound of wax*). A language like Yucatec Maya with a numeral classifier system focuses on substances and in a sense treats most of its nouns like mass nouns. In this language the word ‘wax’ refers both to the substance and to candles. In the presence of numerals, a language of this type uses two kinds of classifiers: **sortal** classifiers that identify units, usually on the basis of a typical shape; and **mensural** classifiers that refer to quantities imposed on masses. Sortal quantifiers yield expressions like [one long-thin wax] = ‘one candle’, [two long-thin wood] = ‘two sticks’, [three long-thin banana] = ‘three bananas’, [one flat banana] = ‘one banana leaf’. Mensural classifiers produce expressions like [one chunk dirt] = ‘one chunk of dirt’, [one bit banana] = ‘one bit of banana fruit’.

Our Verb Classifier Hypothesis is that **Russian has a verb classifier system that is isomorphic to a numeral classifier system**. There are compelling parallels with Russian verbs: most base verbs are imperfective and thus refer to unbounded states and activities (the verbal parallel of masses, Janda 2004), and use of the perfective is a type of quantification (parallel to use of numerals (Jakobson 1957/1971: 136; cf. also Smith 1991, Dahl 1985)). Furthermore, Russian perfectivizing prefixes have different functions, a **sortal** one that identifies events according to the shape of the path they take (as in *napisat’* ‘write’ literally “onto-write”, *vpisat’* ‘insert in a text’ literally “into-write”, *ukrast’* ‘steal’ literally “away-steal”) and a **mensural** one that imposes boundaries (as in the procedurals *popisat’* ‘write for a while’, *polizat’* ‘lick for a while’). The sortal part of the hypothesis has been tested and confirmed in a series of large-scale statistical studies (Janda et al. forthcoming), whereas we owe the mensural part of the hypothesis to a recent insight by Stephen M. Dickey (personal communication). We propose that the Russian verb classifier system is an innovation: We know that the system was different in the earliest attested Slavic sources, and the received wisdom is also that the Old Russian system was different from that found in Old Church Slavic. We propose that procedurals expanded in about the 16th century in Russian (cf. Dickey 2007, Dmitrieva 1991) and this event consolidated the Russian aspectual system as we know it today. The Birds & Beasts project will connect the dots in this development by making it possible to trace its progress from Old Church Slavic through Old/Middle Russian to Modern Russian.

The Birds & Beasts project will apply and expand our repertoire of linguistic profiling techniques. Profiling techniques include grammatical profiling (examining the distribution of grammatical forms; Janda & Ljashevskaya 2011), constructional profiling (examining the distribution of grammatical constructions; Janda & Solovyev 2009; Sokolova et al. forthcoming), semantic profiling (examining the distribution of semantic properties; Janda & Ljashevskaya forthcoming), and radial category profiling (examining the distribution of types across the nodes of a network of meanings; Neset et al. 2011). The data in these studies have been analyzed using a variety of statistical models: hypothesis testing tools such as chi-square and effect size (Cramer's V), Fisher Test; clustering and classification models such as correspondence analysis and hierarchical clustering; and logistic regression. Samples of our data and "R" command code are publicly available at: [http://emptyprefixes.uit.no/methodology\\_eng.htm](http://emptyprefixes.uit.no/methodology_eng.htm). We have recently begun to experiment with some new statistical models, known as "c-trees", "random forests" (Strobl et al. 2009) and "naive discriminative learning" (Baayen et al. 2011), which we plan to implement in the Birds & Beasts project. We will continue to post our data and code to free-access websites, setting a standard for transparency and statistical analysis in the field of linguistics.

### **2.3. Project plan, project management, organization and cooperation: building corpus and organizing case studies**

The plan for the Birds & Beasts Project includes two main components: the **Old Church Slavic, Old Russian and Middle Russian corpus** and **linguistic case studies**.

#### **2.3.1 The Corpus and infrastructure**

The corpus will be based on the PROIEL corpus, but a separate installation of the application will be hosted by the University of Tromsø. The PROIEL corpus consists of two parts, the freely available open-source PROIEL annotation application and the PROIEL database. The already annotated OCS data are also publicly available, and will be imported into the application. Since the infrastructure already exists, we can start extending the Old Church Slavic corpus and building the Old and Middle Russian corpus from day one.

We plan to expand the Old Church Slavic corpus by 70 000 words and add 100 000 words of 11th–14th century Old Russian as well as 50 000 words of 16th–17th century Middle Russian. As far as possible we will use publicly available electronic resources and cooperate with other corpus resources to avoid duplication of effort, in particular we have text exchange agreements with the Russian National Corpus and the Bulgarian Academy of Sciences. Since the infrastructure is already in place, we can start annotating immediately at project start. Judging from experience from the PROIEL and Menotec (Old Norse) corpora, we expect annotators to be able to annotate at least 100 words per hour after a training period of 1-2 months. At this rate the full annotation should take 2400 hours, or 64 Norwegian standard working weeks. We propose to share the work between one permanent 100 % annotator assistant for one year supplemented by annotators working from abroad on an hourly basis. The annotation application is designed to be used from any computer with internet access without any special settings, which means that we can use highly qualified student assistants from any country, and thus become less vulnerable to periods of low activity if an annotator should quit or be less productive for some reason. With this arrangement we think it is realistic to complete the annotation within the limits of the first project year.

To ensure the quality of the corpus and remove errors and inconsistencies, we will recruit an assistant from the annotators to control and correct the analyses in the corpus. We estimate that the assistant can correct 250 words an hour, which means that the proofreading can be finished in 880 hours = 23.5 standard work weeks. We therefore propose to hire an assistant full time for half a year to do this work. Within the proposed frames, the corpus will be complete and ready for publishable research at the end of the project's third semester.

After the corpus is finished, we will develop corpus-based reading aids and other pedagogical tools in order for the corpus to be useful to a wider circle of students and scholars wishing to learn OCS and Old/Middle Russian.

### 2.3.2 The case studies

The second part of the plan for the Birds & Beasts Project is to carry out diachronic and contrastive linguistic case studies in order to test the Verb Classifier Hypothesis by means of corpus data and linguistic profiling. The prediction we pursue in the Birds & Beasts Project is that the development of verb classifiers has led to changes in the linguistic profiles of the relevant verbs. We focus on three types of linguistic profiles. A **grammatical profile** is the relative frequency distribution of the inflected forms of classes of words. Janda and Lyashevskaya (2011) have shown for Modern Russian that different classes of verbs show systematically different frequency distributions across the inflectional paradigm, and that these differences correlate with the aspectual properties of the verbs. Eckhoff and Janda (under preparation) have successfully applied grammatical profiling to Old Church Slavic data from the PROIEL corpus. However, in order to test the Verb Classifier Hypothesis, further work on grammatical profiling needs to be carried out with special focus on Old/Middle Russian data.

In the contrastive part of the project, we compare grammatical profiles of languages with different verb systems. The Barents Region provides an ideal testing ground, since two other important languages in the area, Norwegian and Sámi, have properties that contrast with Russian. Norwegian has an elaborate tense system, but no grammatical aspect. Sámi has no perfective/imperfective distinction, but has an elaborate system of procedurals. The differences and similarities between the grammatical profiles of Russian, Sámi and Norwegian verbs are relevant for the Verb Classifier Hypothesis, since they will contribute new knowledge about the role of procedurals in verb classifier systems.

A second type of linguistic profile relevant for the Verb Classifier Hypothesis is the **constructional profile**, i.e. the frequency distribution of the constructions a word occurs in (Janda and Solovyev 2009). Nessel, Janda and Eckhoff (in preparation) have tested out this method on the most frequent Old Church Slavic verb in the PROIEL corpus, *byti* ‘be’, and shown that constructional profiles represent a valuable tool in historical linguistics. With regard to the Verb Classifier Hypothesis we will compare the procedural verbs of Modern Russian with their cognates in Old/Middle Russian and Old Church Slavic. We surmise that changes in the constructional profiles will provide new knowledge about when and how the Russian verb classifier system emerged.

Another methodology that is ideal for shedding light on the development of verb classifiers is **radial category profiling** (Nessel et al 2011). The Russian verb prefixes are polysemous, insofar as each prefix constitutes a network of closely related meanings (“radial categories”, Lakoff 1987). A radial category profile is the relative frequency distribution of the subcategories of the radial category network of a morpheme or word. While previous scholarship has suggested that prefixes change their meanings over time (Dickey 2007), no large-scale quantitative studies have been carried out (see however Dmitrieva 1991). The CLEAR group at the University of Tromsø has worked out radial category profiles for verb prefixes in Modern Russian, which provide a good point of departure for comparisons with Old/Middle Russian and Old Church Slavic. Corpus data from Old/Middle Russian will enable us to create radial category profiles for prefixes in Old/Middle Russian. Comparison with Modern Russian will reveal to what extent the profiles change over time, thus providing new insights about the historical development of verb classifiers.

Table 1 offers a **detailed project plan**, demonstrating how the Verb Classifier Hypothesis can be tested empirically through six realistic and feasible case studies based on linguistic profiling of corpus data. Each case study is assigned to one semester, thus setting up clear milestones for each semester of the project. The first three case studies will be carried out while the Old/Middle Russian corpus is being built; they are not directly dependent on Old/Middle Russian data, but will set up a context for the interpretation of the data from the Old/Middle Russian corpus. In order to guarantee

that the projected results are achieved, in accordance with the policy of the Norwegian Research Council we apply for a moderate amount of buy-out from teaching in the two final semesters, since the relevant case studies are the most labor-intensive.

Semester:	Case study:	Main author responsible:
1.	Old Church Slavic-Modern Russian comparative case study	Eckhoff
2.	Sámi-Russian comparative case study	Janda
3.	Norwegian-Russian comparative case study	Nesset
4.	Diachronic grammatical profiling (OCS, Old/Middle Russian)	Nesset, Janda & Eckhoff
5.	Diachronic constructional profiling (OCS, Old/Middle Russian)	Nesset, Janda & Eckhoff
6.	Diachronic radial category profiling (OCS, Old/Middle Russian)	Nesset, Janda & Eckhoff

**Table 1: Project plan for Birds & Beasts case studies**

### 2.3.3 Project management

The Birds & Beasts Project will be organized as part of the CLEAR research group at the University of Tromsø. The work in the group is organized according to two principles. According to the **“Divide and Conquer Principle”** we divide projects into feasible case studies of suitable size for one or two journal articles. Each case study departs from a clearly stated hypothesis, pursues a relevant methodological tool and analyzes a clearly defined set of data. According to the **“Teamwork Principle”** all case studies are first discussed by the whole group in a brainstorming session, before a subset of the group members is assigned responsibility for the study. All case studies should have at least two co-authors, since in our experience (small) groups of linguists work more efficiently and achieve better results than individual researchers. All articles resulting from the case studies are presented at least one or two times for the rest of the group before they are submitted to refereed journals. The “Divide and Conquer Principle” and the “Teamwork Principle” have proved very effective in organizing the work of the CLEAR group, as demonstrated by the large number of publications produced by the group members in recent years.

The CLEAR group at present consists of two full professors (Tore Nesset and Laura Janda), one assistant professor (Svetlana Sokolova) and four PhD students (Anna Endresen, Julia Kuznetsova, Anastasia Makarova and one to be hired in the fall of 2013). For the purposes of the Birds & Beasts Project we wish to extend the CLEAR group by hiring Dr. Hanne Eckhoff as a researcher, as well as one postdoctoral fellow and two corpus assistants.

The leader of the Birds & Beasts Project, **Tore Nesset**, is a professor of Russian linguistics and the co-leader of the CLEAR group at the University of Tromsø. He has extensive experience in project management. He recently completed the Exploring Emptiness project funded by the Norwegian Research Council that provided a number of new insights about aspectual prefixes and morphological variation in Russian verbs. He is currently completing a large historical linguistics research project at the prestigious Centre for Advanced Study (Norwegian Academy of Science and Letters) called “Time is Space: Unconscious Models and Conscious Acts”. Professor Nesset has published extensively on Russian verbs from the perspective of cognitive linguistics, including several large-scale studies involving linguistic profiling and statistical analysis of corpus data.

Professor **Laura A. Janda** is the co-leader of the CLEAR group at the University of Tromsø. Ever since her graduate work at UCLA, Janda has pursued cognitive linguistics, and she was president of the International Cognitive Linguistics Association 2007-2011. She is currently an associate editor of the journal *Cognitive Linguistics*. A well-known international authority on case and aspect in the Slavic languages, Janda has numerous publications in international refereed journals, many of which involve statistical analysis of large data sets from linguistic corpora. She has extensive experience in grant management both from the USA and Norway. At the University of Tromsø she is currently the project leader of the Neat Theories, Messy Realities research project funded by the Norwegian Research Council. She is also the co-leader (with Tore Nesset) of the Time is Space project at the Centre for Advanced Study.

Dr. **Hanne Eckhoff** is currently a postdoctoral fellow at the University of Oslo, where she is responsible for the Slavic part of the PROIEL corpus of Old Indo-European languages. She is also



project partner in the Old Norse corpus project Menotec (Bergen/Oslo) and a corpus consultant on the ISWOC project. Dr. Eckhoff is a specialist in historical and cognitive linguistics, with several publications on Old Russian, including her 2011 monograph *Old Russian Possessive Constructions: A Construction Grammar Approach* in the prestigious *Cognitive Linguistics Research* series by Mouton de Gruyter. Dr. Eckhoff's theoretical orientation and her interest in empirical statistical methods makes her a perfect fit for the CLEAR group of the University of Tromsø. As a member of Nessel and Janda's Time is Space project group, Dr. Eckhoff has collaborated with Nessel and Janda on two case studies of Old Church Slavic verbs. Dr. Eckhoff's expertise in building linguistic corpora as well as her competence in Old Church Slavic and Old Russian complement and strengthen Nessel and Janda's competence in aspectual semantics and linguistic profiling, thus guaranteeing a successful completion of the Birds & Beasts Project.

In addition to Nessel, Janda and Eckhoff, in order to successfully complete the Birds & Beasts Project, the CLEAR group needs to be extended by a postdoctoral fellow and two corpus assistants. The postdoctoral fellow will be expected to contribute to the practical corpus work, such as training and guiding annotators, preparing texts and overseeing quality control. The postdoctoral fellow will also be coauthor on two or more of the subprojects, and should also contribute a subproject in line with the main ideas of the project. The corpus assistants will annotate texts and correct annotations.

### 2.3.4 National and international cooperation

By joining the forces of Nessel, Janda and Eckhoff, the Birds & Beasts Project ensures national cooperation between the Universities of Tromsø and Oslo; while Eckhoff will be employed by the University of Tromsø, she will continue the close collaboration with the PROIEL/ISWOC corpus groups in Oslo. We will also collaborate with the aspectologist and corpus linguist Atle Grønn and the computational linguist Kjetil Rå Hauge (both in Oslo), who will visit Tromsø in semester 4 of the project period. As a result of wide-reaching national cooperation, the Birds & Beasts Project will have strong impact on the development of cognitive, Slavic, and corpus linguistics in Norway.

International cooperation represents an important aspect of the Birds & Beasts Project. In order to further strengthen our competence in statistical analysis of linguistic data, we will collaborate closely with professor R. Harald Baayen's group at the University of Tuebingen (Germany), who are the world leaders in this field. Professor Baayen is currently a member of Nessel and Janda's Time is Space group. He will hold two mini-workshops at the University of Tromsø (semesters 2 and 5), where he will offer "hands-on guidance" in the group members' work on concrete datasets from the case studies described in Table 1 above.

Another important international collaborator is professor Stephen M. Dickey at the University of Kansas. A leading expert on the history of aspect in Slavic, professor Dickey will be a co-author of one article on the development of procedural verbs relevant for the case study assigned to semester one in Table 1. In this connection, Janda will visit Kansas for two weeks in semester 1. Dickey is currently a participant in Nessel and Janda's Time is Space project; funding of the Birds & Beasts Project will make it possible to continue this fruitful collaboration.

Embarking on a corpus project without extensive collaboration with leading corpus linguists in Russia is unthinkable. We have collaborated with professors Ekaterina Rakhilina and Vladimir Plungian, two of the main authors of the Russian National Corpus for a number of years. Both participated in Nessel and Janda's Time is Space project. Professor Rakhilina has recently set up a group at the Higher School of Economics in Moscow, including some of the leading corpus linguists in Russia, such as professors Olga Lyashevskaya (formerly employed at the University of Tromsø and a member of the CLEAR group), Mikhail Daniel and Nina Dobrushina. Funding of the Birds & Beasts Project will facilitate extensive collaboration with Russian corpus linguists, from which the field of Slavic linguistics in Norway will benefit greatly. We also plan direct corpus collaboration with text and annotation exchanges.

We collaborate closely with Professor Anissava Miltenova and her colleagues at the Bulgarian Academy of Sciences, who run an UNESCO-funded project to develop an electronic



edition of the Codex Suprasliensis, a major manuscript in the Old Church Slavic canon. The Suprasliensis project provides high-quality text, which is annotated and published in the corpus application. The annotated output is then integrated in the electronic edition. The Suprasliensis project is also a great source of expertise on Old Church Slavic language and textology.

As part of the international cooperation that is essential to the goals of the Birds & Beasts Project, we will organize two workshops, one in the first semester and one in the sixth semester of the project period. In order to facilitate the collaboration with Russian corpus linguists, the workshops will be held at the Norwegian University Center in St. Petersburg, which specializes in organizing workshops and seminars as meeting places for researchers from Russia, Norway and other countries. All the international and national collaborators mentioned above will be invited to participate in the workshops in addition to a small number of international experts in the relevant fields. Table 2 provides an overview of the events planned as part of the Birds & Beasts Project.

Semester:	Event:	Invited:
1.	Workshop in St. Petersburg	All members of the CLEAR group and national and international collaborators
	Visit to Kansas University	L. A. Janda
2.	Mini-workshop on statistics	R. Harald Baayen
3.	Visit from Moscow	O. Lyashevskaya & E. Rakhilina
4.	Visit from Oslo	A. Grønn & K. R. Hauge
5.	Mini-workshop on statistics	R. H. Baayen
6.	Workshop in St. Petersburg	All members of the CLEAR group and national and international collaborators

Table 2: Planned Birds & Beasts events

## 2.4. Budget

In order to successfully complete the Birds & Beasts project, we apply for funding for one researcher (Hanne Eckhoff, 3 years), one postdoctoral fellow (3 years), one assistant (1 year) and one assistant (0.5 years). In compliance with NFR's current strategy, we apply for one semester buy-out from teaching for Nesset and Janda. Beyond the personnel costs (see budget in the electronic application), we apply for a total of 675 000 NOK for seminars, travel, corpus work, etc.

## 3. Key perspectives and compliance with strategic documents

### 3.1. Compliance with strategic documents

In its persistent focus on research and development in the Arctic areas, the University of Tromsø has always focused on Russian language, which is instrumental for future development and cooperation in the Barents Region. By contributing innovative research on the Russian language, the proposed research project raises the overall Norwegian expertise in Russian, and is thus in compliance with major strategic goals of the University of Tromsø and the Norwegian government. The focus on Sámi complies with the University's commitment to support this northern indigenous language.

### 3.2. Relevance and benefit to society

The Birds & Beasts project combines the theoretical study of language with practical applications that society will benefit from directly. Our **corpus** will provide a valuable resource for students, teachers and scholars in Slavic and general linguistics, and our **pedagogical applications** will offer new tools that will prompt a radical rethinking of the way Old Church Slavic and Old Russian is taught. Both corpus and pedagogical applications will be freely available on the Internet.

### 3.3. Environmental impact, ethical perspectives and gender issues

The Birds & Beasts project is organized so as to minimize travel and wasteful use of paper. We will use video conferencing extensively, and have purchased the SubEthaEdit software, that allows several people to work on a document simultaneously, thus minimizing the need for printouts.

The Birds & Beasts project does not involve ethically problematic research, and the CLEAR group will continue its persistent commitment to gender equality in our hiring of group members.

#### 4. Dissemination and communication of results

Each case study will lead to one or two **scholarly articles**, which will be submitted to “nivå 2” international peer-reviewed journals, e.g. *Cognitive Linguistics*, *Journal of Slavic Linguistics* and *Russian Linguistics* where the CLEAR group members are frequent contributors. We will organize two **workshops** (see Table 2), and our research will also be presented at **international conferences** such as the International Cognitive Linguistics Conference and the Slavic Cognitive Linguistics Conference, where the members of the CLEAR group are regular participants.

The CLEAR group is committed to maximizing the availability of its output for users. Therefore, the **corpus** will be built in an entirely open-source format, and we will share the data freely under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 license. To offer users extended query possibilities, we propose to add the corpus to the treebank hosting facility of the INESS project at the University of Bergen, which offers a syntactic query interface already adapted to PROIEL-style data. Our **pedagogical applications** will be freely available on the Internet.

#### References

- Avanesov, R. et al 1975-. *Slovar' russkogo jazyka XI-XVII vv.* Moscow.
- Aikhenvald, A. Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford.
- Baayen, R. H., Milin, P., Filipovic Durdevic, D., Hendrix, P. and Marelli, M. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118, 438-482.
- Bermel, N. 1997. *Context and the Lexicon in the Development of Russian Aspect*. Berkeley.
- Dahl, Ö. 1985. *Tense and aspect systems*. Oxford.
- Dickey, S. M. (2007) A prototype account of the development of delimitative PO- in Russian. In D Divjak & A. Kochanska, eds. *Cognitive Paths into the Slavic Domain*. Berlin, 326–371.
- Divjak, D. & S. Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *CLLT* 2: 23–60.
- Dmitrieva, O. (1991) Formirovanie semantičeskoj struktury russkogo po-. *Aktivnye processy v jazyke i reči*, 68–74. Saratov.
- Eckhoff, H. M. and L. A. Janda. in preparation. “Grammatical Profiles and Aspect in Old Church Slavonic”.
- Gries, S. Th. & D. S. Divjak. 2009. Behavioral profiles: a corpus-based approach towards cognitive semantic analysis. In V. Evans & S. Pourcel *New directions in cognitive linguistics*, 57–75. Amsterdam.
- Jakobson, R. O. 1957/1971. Shifters. In R. Jakobson: *Selected Writings II*. The Hague, 130-147.
- Janda, L.A. 2004. A metaphor in search of a source domain, *Cognitive Linguistics*, 15:4, 471-427.
- Janda, L. A., A. Endresen, J. Kuznetsova, O. Lyashevskaya, A. Makarova, T. Nessel, S. Sokolova. To appear. *Why Russian aspectual prefixes aren't empty: prefixes as verb classifiers*. Bloomington, IN: Slavica.
- Janda, L. A. & O. Lyashevskaya. 2011. Grammatical profiles. *Cognitive Linguistics* 22:4, 719-763.
- Janda, L. A. & Olga Lyashevskaya. forthcoming. Semantic Profiles of Five Russian Prefixes. *Journal of Slavic Linguistics*.
- Janda, L. A. & V. Solovyev. 2009. What Constructional Profiles Reveal About Synonymy. *Cognitive Linguistics* 20:2, 367-393.
- Krifka, M. 1989. Nominal reference and quantification in event semantics. In R. Bartsch, J. van Benthem, and P. van Emde Boas *Semantics and Contextual Expression*. Dordrecht.
- Krifka, M. 1998. The origins of telicity. In S. Rothstein *Events and Grammar*. Dordrecht
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. Chicago.
- Leskien, A. (1871): *Handbuch der Altbulgarischen Sprache*. Heidelberg.
- Nessel, T., L. A. Janda & A. Endresen. 2011. Two ways to get out. *Zeitschrift für Slawistik* 56:4, 377-402.
- Nessel, T., L. A. Janda, H. M. Eckhoff. in preparation. The Many Faces of *byti*: What Was That Verb Doing in OCS?
- Nørgård-Sørensen, J. 1997. Tense, aspect and verbal derivation in the Novgorod birch bark letters. *Russian Linguistics* 21:1–21.
- Smith, C. 1997. *The Parameter of Aspect*. Dordrecht.
- Sokolova, S., L. A. Janda & O. Lyashevskaya. forthcoming. The Locative Alternation and the Russian ‘empty’ prefixes. In: D. Divjak & St. Th Gries (eds.). *Frequency effects in cognitive linguistics (Vol. 2)*. Berlin.
- Sreznevskij, I. 1893. *Materialy dlja slovarja drevne-russkago jazyka*. St. Petersburg.
- Strobl, C., J. Mally & G. Tutz. 2009. “An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests”. *Psychological Methods* 14:4, 323-348.
- Vendler, Z. 1957. Verbs and times. *Philosophical Review* 66, 143-60.
- Zaliznjak, A. 2008. “*Slovo o polku Igoreve*”: *vzgljad lingvista*. Moscow.