Syntax and semantics in OCS

Hanne Eckhoff

10 September
Early Slavic corpus linguistics

- The Tromsø Old Russian and OCS Treebank (TOROT)
- An expansion of the OCS part of the PROIEL corpus
- Many-layered annotation: lemmatisation, morphology, syntax (dependency grammar), information structure, customised tagging for various features (semantics, derivational morphology . . .)
- OCS Gospels (Mar. and Zogr.) are aligned with the Greek Gospels at word token level
- Very complex data that may at least to some extent compensate for the small size of the OCS canon
- Two case studies
  - What is the relationship between inflectional aspect marking and derivational aspect marking in OCS?
  - What factors regulate distribution of the nominative-accusative and the genitive-accusative? (The early history of the animacy category)
The aspect study

- Coauthored work with Dag Haug, to appear in *Diachronica* 32:2, data set and supplementary materials at http://hdl.handle.net/10037.1/10158
- A corpus study of the earliest attested situation, in OCS
- OCS has an inflectional expression of aspect (aorist, imperfect, participles), but also an affixation system reminiscent of the modern Slavic aspect systems
- No agreement as to which system is the primary exponent of aspect
- Old approach + new approach: Comparison of Greek and OCS aspect using rigorous statistical methods (Meillet 1902 with computers)
Method

- Data from the PROIEL/TOROT treebanks – annotation for both inflectional and derivational morphology, as well as syntax
- Codex Marianus/Codex Zographensis and the Gospels in their Greek original
- Advantages of using the Greek parallels
  - The Greek aspectual system is relatively well-understood
  - No formal similarities between e.g. the Greek and the OCS aorist
  - Correspondences must be based on semantic similarities
- Obvious disadvantage: We don’t have the exact Greek originals of the OCS texts – some noise in the data
- Gospels are the only text with sufficiently good Greek parallels
<table>
<thead>
<tr>
<th>Stem</th>
<th>Present</th>
<th>Aorist</th>
<th>Imperfect</th>
<th>Infinitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>tvori-&lt;sub&gt;ipfv&lt;/sub&gt;</td>
<td>tvoritъ</td>
<td>tvori</td>
<td>tvorjaаше</td>
<td>tvoriti</td>
</tr>
<tr>
<td>sъ-tvori-&lt;sub&gt;pfv&lt;/sub&gt;</td>
<td>sъ-tvoritъ</td>
<td>sъ-tvori</td>
<td>*sъ-tvorjaаше</td>
<td>sъ-tvoriti</td>
</tr>
<tr>
<td>sъ-tvarja-&lt;sub&gt;ipfv&lt;/sub&gt;</td>
<td>sъ-tvarjajetъ</td>
<td>*sъ-tvarja</td>
<td>sъ-tvarjaаше</td>
<td>*sъ-tvarjatи</td>
</tr>
</tbody>
</table>

- There is an inflectional exponent of aspect in the past tenses (and in the participles)
- The present and the infinitive do not express aspect inflectionally
Datasets

- We extracted all OCS verb forms and their Greek equivalents
- We eliminated *byti* (→ 23,538 verb tokens)
- Development set: those Slavic forms that have an inflectional exponent of ‘aspect’
  - aorists and imperfects
  - past and present participles
- all in all 13,707 tokens
### Tense correspondences in the development set

<table>
<thead>
<tr>
<th>Tense</th>
<th>Mar. aor</th>
<th>Mar. %</th>
<th>Mar. imperf</th>
<th>Zogr. aor</th>
<th>Zogr. %</th>
<th>Zogr. imperf</th>
<th>Zogr. %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aorist</td>
<td>2887</td>
<td>98.6</td>
<td>42</td>
<td>1.4</td>
<td>2604</td>
<td>98.2</td>
<td>47</td>
</tr>
<tr>
<td>Imperfect</td>
<td>79</td>
<td>11.1</td>
<td>631</td>
<td>88.9</td>
<td>73</td>
<td>11.0</td>
<td>592</td>
</tr>
</tbody>
</table>

**Table:** Slavic translations of Greek aorists and imperfects, Marianus: n=3639, Zographensis: n=3316
Participle correspondences in the development set

<table>
<thead>
<tr>
<th></th>
<th>Mar.</th>
<th></th>
<th>Zogr.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>past</td>
<td>%</td>
<td>pres</td>
<td>%</td>
</tr>
<tr>
<td>aor</td>
<td>1070</td>
<td>98.8</td>
<td>13</td>
<td>1.2</td>
</tr>
<tr>
<td>fut</td>
<td>0</td>
<td>0.0</td>
<td>2</td>
<td>100.0</td>
</tr>
<tr>
<td>pres</td>
<td>23</td>
<td>1.8</td>
<td>1225</td>
<td>98.2</td>
</tr>
<tr>
<td>perf</td>
<td>178</td>
<td>78.1</td>
<td>50</td>
<td>21.9</td>
</tr>
</tbody>
</table>

Table: OCS participles translating Greek participles, Marianus: n=2561, Zographensis: n=2286

Merged datasets: The OCS aorist/imperfect distinction and the present/past distinction in OCS participles follow the Greek and hence express aspect.
But didn’t the participles just express tense?

(1) sъgrĕšĭxъ prĕdavъ  krъvъ  ne
sъ.sin.i.aor.1sg  prĕ.betray.pstp.m.nom.sg  blood.acc.sg  not
povinъnъ
guilty.f.acc.sg
hēmarton  paradous  haima  athōion
‘I have sinned by betraying innocent blood.’ (Mar. Mt. 27.4)
Research questions

- Given that there is an inflectional exponent of aspect, what role for affixation?
  - What do the prefixes mean?
  - What do the suffixes mean?
  - How does affixation relate to inflectional aspect?
- First, we will study this issue based on the inflected forms
- Then we will test our hypotheses on the infinitives, futures and subjunctives
What role for the affixes?

<table>
<thead>
<tr>
<th>affix</th>
<th>aorist</th>
<th>imperfect</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>2156</td>
<td>437</td>
</tr>
<tr>
<td>prefix</td>
<td>2669</td>
<td>13</td>
</tr>
<tr>
<td>suffix</td>
<td>1643</td>
<td>2639</td>
</tr>
<tr>
<td>both</td>
<td>3154</td>
<td>996</td>
</tr>
</tbody>
</table>

Table: OCS affixation on participles and past tense verbs

- The meaning of prefixation (without suffixation) must be incompatible with imperfective aspect.
- A reasonable working hypothesis is that prefixation is already grammaticalised in the meaning of perfective aspect.
- Suffixation increases the frequency of the imperfect.
### What role for the suffixes?

<table>
<thead>
<tr>
<th>Suffix</th>
<th>All Verbs</th>
<th>Unprefixed Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aorist</td>
<td>Imperfect</td>
</tr>
<tr>
<td>–</td>
<td>4825</td>
<td>450</td>
</tr>
<tr>
<td>ě</td>
<td>563</td>
<td>555</td>
</tr>
<tr>
<td>ja</td>
<td>305</td>
<td>392</td>
</tr>
<tr>
<td>ěj</td>
<td>104</td>
<td>14</td>
</tr>
<tr>
<td>jaj</td>
<td>412</td>
<td>245</td>
</tr>
<tr>
<td><code>a</code></td>
<td>1185</td>
<td>1068</td>
</tr>
<tr>
<td>aj</td>
<td>98</td>
<td>251</td>
</tr>
<tr>
<td>a/aj</td>
<td>72</td>
<td>93</td>
</tr>
<tr>
<td>(V)va</td>
<td>2</td>
<td>63</td>
</tr>
<tr>
<td>i</td>
<td>1731</td>
<td>826</td>
</tr>
<tr>
<td>nó</td>
<td>197</td>
<td>0</td>
</tr>
<tr>
<td>ova</td>
<td>128</td>
<td>128</td>
</tr>
</tbody>
</table>

**Table:** Token frequency of suffixes on participles and past-tense verbs
What role for the suffixes?

- nq is a case apart
- no suffix is exclusively associated with the imperfect
- so we need to look for ‘paired verbs’, as in modern Slavic
- let us look at families of verbs, as defined by either a common root, or a common Greek correspondence
Aspect
Animacy and differential object marking
Reading

Hanne Eckhoff
Syntax and semantics in OCS
Aspect
Animacy and differential object marking
Reading

Distribution of aspect in verbs from da

Freq
0 20 40 60 80 100 120 140

Hanne Eckhoff
Syntax and semantics in OCS
Distribution of verbs and aspect in translations of eperōtaō

<table>
<thead>
<tr>
<th>Verb</th>
<th>Pres. (Gk. impf.)</th>
<th>Pres. (Gk. aor.)</th>
<th>Impf./pres.ptcp (Gk. impf.)</th>
<th>Impf./pres.ptcp. (Gk. aor.)</th>
<th>Other (Gk. impf.)</th>
<th>Other (Gk. aor.)</th>
<th>Aor./past (Gk. impf.)</th>
<th>Aor./past (Gk. aor.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>vъprašati</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vъprositi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glagolati</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prosili</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Aspect
Animacy and differential object marking
Reading
Categorising verbs

- We observe that single formations select for specific aspectual preferences
- We can categorise verbs according to these preferences
  - \textit{pfv} verb only occurs in the aorist
  - \textit{pfv(<4)} verb only occurs in the aorist, but fewer than 4 times
  - \textit{ipfv} verb only occurs in the imperfect
  - \textit{ipfv(<4)} verb only occurs in the imperfect, but fewer than 4 times
  - \textit{\neg pfv} neutral verb, significantly more aorists
  - \textit{\neg ipfv} neutral verb, significantly more imperfects
  - \textit{\neg neut} neutral verb, no aspect dominates
- Significance level $p = 0.05$, binomial test
Testing the hypothesis

- We observed that the various verbal formations strongly select for one of the aspects.
- A likely hypothesis is that affixation is just another exponent of aspect.
- If so, affixation should act as the sole exponent of aspect in forms where there is no inflectional aspect.
- We can test this on infinitives (1317 tokens), which are uninflected in Slavic, but inflect for aspect in Greek.
- Other test set: OCS correspondents of the Greek Greek subjunctives (1684 tokens).
Infinitives and their Greek originals

Verb type
Freq
0 100 200 300 400 500
Aorist Present
- ipfv -ipfv(<4) - pfv - pfv(<4) - unkn. ∼ ipfv ∼ neut ∼ pfv

Hanne Eckhoff
Syntax and semantics in OCS
(2) \( \textit{vlastь} \quad \textit{imatь} \quad \textit{snь} \quad \textit{чčsky.} \)
\( \text{power.nom.sg} \quad \text{have.prs.3sg} \quad \text{son.nom.sg} \quad \text{of-man.m.nom.sg} \)
\( \text{na zemi} \quad \textit{otьpuшtati} \quad \textit{grěxy} \)
\( \text{on earth.loc.sg} \quad \text{otь.forgive.aj.inf} \quad \text{sin.acc.pl} \)
\( \text{ho huios tou anthrōpou exousian ekhei epi tēs gēs} \quad \textit{aphienai hamartias} \)
\( \text{‘(But that you may know that) the Son of Man has authority on earth to forgive sins’} \) (Zogr., Lk. 5.24)
Interim conclusions

- The OCS aorist/imperfect distinction expresses aspect (like Greek)
- Affixation creates derivations which robustly select for a single aspect
- These classes can express aspect where the morphology cannot
- So both inflection and affixation are exponents of aspect, and in the past tense and participle system, there is multiple exponence
- Is there a difference?
Greek aspect – single exponence

- Greek aspect is relatively independent of telicity

<table>
<thead>
<tr>
<th>Situation type</th>
<th>Imperfective</th>
<th>Perfective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement</td>
<td>habitual/progressive/conative</td>
<td>completive</td>
</tr>
<tr>
<td>Accomplishment</td>
<td>habitual/progressive</td>
<td>completive</td>
</tr>
<tr>
<td>Activity</td>
<td>habitual/ongoing</td>
<td>complexive/ingressive</td>
</tr>
<tr>
<td>State</td>
<td>ongoing</td>
<td>complexive/ingressive</td>
</tr>
</tbody>
</table>

Hanne Eckhoff

Syntax and semantics in OCS
Example
ôneeto - buy:PST.IMPF.3SG ‘often bought/was buying/trying to buy’
ônêsa - buy:PST.AOR.3SG ‘bought’
epoieeto stratêlasiên - do:PST.IMPF.3SG campaign:ACC ‘was campaigning/often campaigned’
epoiêsato stratêlasiên - do:PST.AOR.3SG campaign:ACC ‘campaigned’
edakrue - cry:PST.IMPF.3SG ‘was crying/usually cried’
edakruse - cry:PST.AOR.3SG ‘cried (for a bounded time)/burst in tears’
ebasileue - be king:PST.IMPF.3SG ‘was king’
ebasileuse - be king:PST.AOR.3SG ‘was king (for a bounded time)/became king’
Slavic-style aspect, ‘bounder perfectives’ (Bybee and Dahl 1989)

- Arise from adverbs such as ‘up’, ‘down’, ‘over’, ‘through’
- Other languages have bounder perfectives too (Lithuanian, Hungarian, Georgian – see areal studies by Arkadiev), but none have generalised them as far as Slavic
- Resulting perfectives are telic
- Prefix + atelic predicate yields only the shifted (telic) ingressive reading, not the complexive
- The complexive is later expressed by the prefix po, but this is not a productive pattern in OCS
- Telicity implies perfectivity, and atelicity imperfectivity
- Pragmatic strengthening: implication $\rightarrow$ entailment?
Our hypotheses

- the lexical meaning of prefixes entailed telicity
- telicity was grammaticalised as perfective aspect by the time of OCS
- the original meaning of telicity *was not lost* (it was preserved all the way to modern Slavic)
- affixation is not yet obligatory in OCS: some (mostly unprefixed) verbal formations are neutral wrt. aspect
Predictions

1. Relatively few verbs will occur in both aorist and imperfect as affixation becomes more obligatory.
2. No imperfects from explicitly perfective verbs.
3. No aorists from explicitly imperfective verbs.
4. The complexive reading is possible for aorists of neutral, atelic verbs, but not those of perfective verbs.
5. Aorists of neutral, atelic verbs can also be ingressives.
6. The modern Bulgarian system does not preserve the OCS system, but has expanded it.
Prediction 1: ‘neutral’ verbs

- Only 74 verbs occur both in the imperfect and aorist, and only 38 have no predominant aspect (760 verbs have specialised).
- Some of the verbs have no partner (čuditi ‘wonder’), but appear to express aspect with inflection only.
- Some of the verbs have affixed aspectual partners, but sometimes express aspect with inflection only nonetheless (xotěti, vřsxotěti ‘want’).
- So OCS is relatively advanced in the grammaticalisation of prefixes.
Prediction 2: Imperfects from overtly perfective verbs

- No reliable examples in Mar./Zogr.
- A handful of examples in the literature
- Later evolution towards the modern Bulgarian system where this combination is used for iterated telic events?
- Note that iterated telic events are quite frequent in the NT, so if perfective imperfects were available for this meaning we would expect to find examples
Prediction 3: Aorists from overtly imperfective verbs

- No convincing examples
Prediction 4: Complexive readings

(3) \textit{sedmь bo ixь } \textit{iměšę } jo \textit{ženą}\hfill
\begin{center}
seven \ for \ of-them \ have-aor \ her \ wife
\end{center}

\begin{flushleft}
hoi \ gar \ hepta \ eskhon \ autēn \ gunaika
\end{flushleft}

‘For all seven were married to her.’ (Mar. MARK 12:23)

(4) \textit{vidita město ideže } \textit{ležа } xь

\begin{flushleft}
see \ \ place \ \ where \ \ lie-aor \ Christ
\end{flushleft}

\begin{flushleft}
deute \ idete \ ton \ topon \ hopou \ ekeito
\end{flushleft}

‘Come, see the place where He was lying.’ (Mar. MATT 28:6)

Prefixation is not an alternative here.
Statement of fact example

- Notice that the complexive reading borders on the statement of fact reading when the verb is atelic

(5) ĕstä prĕdъ tobojъ i ăxotъ i na rasptiixъ
eat-aor before you and drink-aor and on crossroads
učil esi
teach-l-ptcp are
ephagomen enōpion sou kai epiomen, kai en tais plateiais
hēmōn edidaxas
‘We ate and drank with you, and you taught in our streets.’
(LUKE 13:26)
Prediction 5: Neutral aorists as ingresses

(6) \textit{viděvĕ že narodi čjudišĕ se}
seen ptc people wonder-aor refl
idontes de hoi okhloi ephobĕthĕsan
‘When the crowd saw this, they were filled with awe’ (MATT 9.8)

(7) \textit{i vĕzide na gorŏ i prizva ježe samv xotĕ}
and went-up on mountain and called who self want-aor
Kai anabainei eis to oros kai proskaleitai hous ēthelen autos
‘Then Jesus went up on a hillside and called to himself those whom he had decided on’ (MARK 3.13)

(8) \textit{Vă utrĕi že denv vŏsxotĕ isv iziti v galilĕjŏ}
on morning ptc day want-pref.aor Jesus go-out to Galilee
Tĕi epaurion ēthelĕsen exelthein eis tĕn Galilaian
‘The next day Jesus decided to leave for Galilee’ (JOHN 1.44)
Conclusions

- The OCS systems attests an interesting stage of grammaticalisation where there are two exponents of aspect (in some forms)
- Later Slavic languages either lose the old exponent or (Bulgarian, Bosnian/Croatian/Serbian) kept both exponents, but changed the meanings
- The new aspect system includes telicity whereas the old one did not
- At the time of OCS, the old aspect system still has functions that the new system lacks (complexives of neutral verbs)
- The complexive reading later gets a separate exponent
- The new system takes over while the old system is still in force. Is this why the Slavic systems have the most ‘advanced’ bounder perfective systems?
Why this case study?

- An example of how to use syntactic data
- An example of how to use information status data
- Published in *Russian Linguistics* 39:2, replication data at http://hdl.handle.net/10037.1/10190
The animacy category in OCS

- Modern Slavic case languages: the animacy category as a subgender – little variation with individual words or within semantic classes
- OCS provides the earliest attestation of the Slavic animacy category – genitive-accusative (GA) found in variation with nominative-accusative (NA)
- Masculine singular o-stem common nouns denoting male persons *may* get GA marking, as may masculine singular pronouns and nominalised adjectives/participles
- Found both with direct objects and after accusative-governing prepositions
- Extensive literature has argued that the GA-marking correlates with degree of social prominence (high-status adult men preferred) and pragmatic properties (definiteness, specificity)
Variation

(9) ašte žena pouštši mǫža posagnetə za ɨnə
if woman released man.ga marries after other.na
‘If a woman divorces her husband and marries another . . . ’
(Mar. Mk. 10.12)

(10) idi prizovi mǫžь tvoi
go summon man.na your.na
‘Go, summon your husband.’ (Mar. Jh. 4.16)
Not only social prominence: definiteness and specificity

- Meillet 1897: indications that the NA expresses indefiniteness and the GA definiteness; Greek definiteness an important factor for the realisation of *rabъ* ‘servant, slave’
- Comrie 1978: Coding of untypical objects drives the spread of the animacy category: not only highly animate, but also definite referents
- Huntley 1993: Strong tendency for definite referents to be coded with GA and indefinite ones with NA
- Timberlake (1997) on Old Russian: GA for highly individuated persons when participant focus (as opposed to event focus)
Differential object marking

- The animacy category in OCS seems a fairly typical case of differential object marking: “overt case-marking of direct objects to mark some objects, but not others, depending on semantic and pragmatic features of the object” (Aissen 2003)
- High-prominence NPs are object-marked, low-prominence ones not
- Semantic/pragmatic features vary: animacy, definiteness, specificity, topicality, referential persistence
DOM types

- Some languages have DOM of the coding/indexing type (only properties of object argument important)
- Some languages have discriminatory DOM – object marking depends on relative properties of subject and object
- Economy constraint: Overt marking of objects only if it is sufficiently similar to the subject
- Dalrymple and Nikolaeva 2011: DOM originates as (secondary) topic marking, but may grammaticalise in different ways
- What can we learn about OCS GA-marking using the DOM toolbox?
Information status annotation in the PROIEL corpus

<table>
<thead>
<tr>
<th>Context</th>
<th>Specific tag</th>
<th>Non-specific tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse</td>
<td>OLD</td>
<td>NONSPEC-OLD</td>
<td>previously mentioned</td>
</tr>
<tr>
<td>Scenario</td>
<td>ACC-INF</td>
<td>NONSPEC-INF</td>
<td>accessible by inference</td>
</tr>
<tr>
<td>Encyclopaedic</td>
<td>ACC-GEN</td>
<td>NONSPEC-INF</td>
<td>acc. from world knowledge</td>
</tr>
<tr>
<td>Situation</td>
<td>ACC-SIT</td>
<td>NONSPEC-INF</td>
<td>accessible by deixis</td>
</tr>
<tr>
<td>—</td>
<td>NEW</td>
<td>NONSPEC-OLD</td>
<td>not previously mentioned</td>
</tr>
<tr>
<td>KIND</td>
<td></td>
<td></td>
<td>kind-referring</td>
</tr>
<tr>
<td>QUANT</td>
<td></td>
<td></td>
<td>quantified</td>
</tr>
</tbody>
</table>

Table: Contexts and tags in the PROIEL corpus, adapted from Haug et al. 2014
Data extraction

- Extracted all animate masculine singular objects and complements of accusative-governing prepositions from the Codex Marianus + their parallels in Codex Zographensis if they were o- or u-stems (for nouns)
  - were not in an obligatory genitive position (in the scope of a negation, dependent on a supine)
  - were not governed by a verb that could possibly take genitive objects
  - had a Greek token alignment
- Proper nouns are all GA-marked
- Pronouns display NA–GA variation, but in a very different distribution
- We can say something about common nouns and nominalised adjectives and participles denoting humans
Many scholars say that the GA/NA distinction has ‘pragmatic properties’, but what does that mean? Principled IS data to the rescue.

Is the alternation sensitive to other discourse properties?

How do the discourse properties of GA objects relate to subject properties?
Lexical variation

- Only 48 lemmata attested at all
- 26 occur only in the GA (most only once or a couple of times, but clear preference for отець ‘father’, 65 GA occurrences)
- Nine lemmata occur only in the NA
- Only 13 lemmata are attested in actual variation in the material
<table>
<thead>
<tr>
<th>Lemma</th>
<th>Meaning</th>
<th>NA</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>angelъ</td>
<td>‘angel’</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>běsъ</td>
<td>‘demon’</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>běşnojъje</td>
<td>‘being possessed by a demon’</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>dlъžnikъ</td>
<td>‘debtor’</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>drugъ</td>
<td>‘friend’</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>mladъпъьь</td>
<td>‘infant’</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>umъръ</td>
<td>‘having died’</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>pristavъникъ</td>
<td>‘steward’</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>pročъ</td>
<td>‘remaining’</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table**: Lemmata occurring only in the NA
## Variation lemmata

<table>
<thead>
<tr>
<th>Lemma</th>
<th>Meaning</th>
<th>NA</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>bogъ</td>
<td>'god'</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>gluhъ</td>
<td>'deaf'</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>gospodinъ</td>
<td>'lord'</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>gogodnyhъ</td>
<td>'having a speech impediment'</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>drugъ</td>
<td>'other'</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>možъ</td>
<td>'man'</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>oslablenъ</td>
<td>'paralysed'</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>pastyrъ</td>
<td>'shepherd'</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>rabъ</td>
<td>'servant, slave'</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>razboinikъ</td>
<td>'robber'</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>synъ</td>
<td>'son'</td>
<td>26</td>
<td>32</td>
</tr>
<tr>
<td>cěsarъ</td>
<td>'king, emperor'</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>člověkъ</td>
<td>'man'</td>
<td>3</td>
<td>28</td>
</tr>
</tbody>
</table>
How to deal with the lexical bias?

- Most of the NA and GA nouns have too few occurrences for us to be sure that they really occur with only one of the forms.
- Use variation nouns as a control set, to see if the conclusions that can be drawn from the full data set also hold in this limited set (bogросл ‘god’ excluded due to strong bias).
GA and Greek definiteness

<table>
<thead>
<tr>
<th>Greek</th>
<th>NA</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>no article</td>
<td>53</td>
<td>66</td>
</tr>
<tr>
<td>article</td>
<td>41</td>
<td>228</td>
</tr>
</tbody>
</table>

Table: OCS object marking by presence of Greek definite article, p-value < 0.00001, Fisher’s exact test, two-tailed

- Strong correlation between presence of Greek definite article and GA
- No strong tendency for indefinite Greek objects to be rendered as NA (cf. Huntley)
GA and Greek definiteness, variation set

<table>
<thead>
<tr>
<th>Greek</th>
<th>NA nouns</th>
<th>NA</th>
<th>GA</th>
<th>GA nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>no article</td>
<td>25</td>
<td>26</td>
<td>19</td>
<td>47</td>
</tr>
<tr>
<td>article</td>
<td>11</td>
<td>30</td>
<td>81</td>
<td>107</td>
</tr>
</tbody>
</table>

Table: OCS object marking by presence of Greek definite article in variation set vs. invariant lemmata, p-value = 0.00043 in the variation set, Fisher’s exact test, two-tailed
Examples

(11) *sъrěšteta vy člvkъ*
will-meet you.nom man.na
apantēsei humin anthrōpos
‘you will meet a man’ (Mar. Mk 14.13)

(12) *prizъvaše že vъtoricejо čka*
called ptc second-time man.ga
Ephōnēsan oun ton anthrōpon ek deuterou
‘They summoned the man once more’ (Mar. Jh 9.24)
## GA and information status, full set

<table>
<thead>
<tr>
<th></th>
<th>NA</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>anchored</td>
<td>24</td>
<td>32</td>
</tr>
<tr>
<td>accessible</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>old</td>
<td>19</td>
<td>132</td>
</tr>
<tr>
<td>non-specific new</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>non-specific inferred</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>non-specific old</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

**Table**: Information status by OCS case, all
GA and information status, variation set

<table>
<thead>
<tr>
<th></th>
<th>NA</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>anchored</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>accessible</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>old</td>
<td>13</td>
<td>57</td>
</tr>
<tr>
<td>non-specific new</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>non-specific inferred</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>non-specific old</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

*Table*: Information status by OCS case, variation set
Definiteness vs. information status

- Old and accessible objects have a strong tendency to be GA-marked
- New and anchored objects show variation
- No discernible difference between specific and non-specific objects
- This two-way separation is statistically significant (also in the variation set)
- So if ‘definite’ means old or accessible, and ‘indefinite’ means new (regardless of specificity or anchoring), OCS GA marking looks strongly like a limited kind of definiteness marking
Variation with new objects

(13) *privěšę člvkъ nětъ běsenъ*
brought man.na.new mute possessed-by-demon
prosēnegkan autōi anthrōpon kōphon daimonizomenon
‘they brought a man, mute and possessed by a demon’
(Mar. Mt 9:32)

(14) *i se mōži na odrē nosěšte člvka*
and behold men on stretcher carrying man.ga.new
kai idou prosepheron autōi paralutikon epi klinēs
‘And behold, some men were carrying a man on a stretcher’
(Mar. Mt 9:2)
Variation with non-specific objects

(15) ėko na razboinika li izidete
as on robber.ga.non-spec interrog.ptc you-came-out
hōs epi lēistēn exēlthate
‘did you come out as though against a robber’ (Mar. Mk 14.48)

(16) ėko na razboinikъ li pridete
as on bandit.na.non-spec interrog.ptc you-arrived
hōs epi lēistēn exelēluthate
‘did you come here as though against a robber’ (Mar. Lk 22.52)
Variation with anchored objects

(A man was having a big dinner party and invited many)

(17) \textit{i posla rabь svoi vь godь veчerь reшti zьvanymь}
and sent servant.na his on day party.gen say invited.dat
kai apesteilen ton doulon autou tēi hōra tou deipnou eipein tois
keklēmenois
‘and he sent his servant on the day of the party to tell the
invited’ (Mar. Lk 14.17)

(18) \textit{edinь же оть stojгštixь [. . .] udari raba arxiereova}
one ptc of standing struck servant.ga archpriest’s
heis de tis tōn parestēkotōn [. . .] epaisen ton doulon tou
arkhiereōs
‘one of those standing there [. . .] struck the archpriest’s servant’
(Mar. Mk 14.47)
GA and other types of prominence

- Can other types of object discourse prominence help explain the remaining variation?
- GA objects are more salient (belong to longer/tighter anaphoric chains) than NA objects, but the difference is not statistically significant for old objects; reduces to information status
- Timberlake’s conception of participant focus might lead us to guess that GA objects would be more likely to be picked up again, but this does not hold
GAs and pickup rates

<table>
<thead>
<tr>
<th></th>
<th>full set</th>
<th>variation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>6.62</td>
<td>6.84</td>
</tr>
<tr>
<td>GA</td>
<td>6.17</td>
<td>9.24</td>
</tr>
</tbody>
</table>

Table: Mean number of pickups by object case, full set and variation set, p-value = 0.00240 in full set, 0.1212 in variation set, Wilcoxon test.

If anything, GAs are less likely to be picked up than NAs!
Pickup rates with first mentions only

<table>
<thead>
<tr>
<th></th>
<th>full set</th>
<th>variation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>6.30</td>
<td>7.29</td>
</tr>
<tr>
<td>GA</td>
<td>2.43</td>
<td>3.04</td>
</tr>
</tbody>
</table>

Table: Mean number of pickups by object case, new, non-specific new and anchored objects, p-value = 

- p-value < 0.00001 in the full set (66 NAs and 65 GAs),
- p-value = 0.00025 in the variation set (35 NAs, 23 GAs), Wilcoxon test
- The NA is especially likely to be used to introduce a referent which will be a recurrent participant in the following narrative.
Examples

(19) posledь že posъla kъ nîmъ synъ svoi
after ptc sent to them son.na his
husteron de apesteilen pros autous ton huion autou
‘Finally, he sent his son to them’ (Mar. Mt 21.37)

From the parable of the wicked tenants, where this is the first mention of the son who is subsequently sent to the tenants and brutally killed by them (eight further mentions).
Examples

(20)  

\[ i \quad se \ priněsę \ emu \ oslablenъ \ žilami \ na \ odrě \]

and lo brought him weakened. na sinews on stretcher

\[ ležěšťъ \]

lying

\[ kai \ idou \ prosepheron \ autōi \ paralutikon \ epi \ klinēs \]

beblēmenon

‘and, behold, they brought him a paralytic lying on a

stretcher’ (Mar. Mt. 9.2)

The paralytic goes on to be healed by Jesus after a lengthy
discussion with the Pharisees (eleven further mentions).
Topics and secondary topics

- Dalrymple and Nikolaeva (2011) propose that DOM (universally?) originates as (secondary) topic marking
- Aboutness topic: “a referent which the utterance is pragmatically presupposed to be about”
- How can we identify aboutness topics when native speaker intuitions are missing?
A topic guesser

- Place on the givenness hierarchy (old is best, new and acc-gen are excluded)
- Place on the hierarchy of syntactic relations (subjects are best)
- Place on the animacy hierarchy (humans are best)
- Word order (first is best; Greek in most cases)
- Realisation (prodrops, personal pronouns, personal names are favoured)
- Relative saliency: is the topic candidate a member of a longer and tighter anaphoric chain than the competition?
- Properties of the immediate antecedent: does it outrank the intervening referents on the relation, animacy and givenness hierarchies?
Topicworthyness rather than topicality

- Very few of the objects are aboutness topics by this (or any other) methods
- However, we can look at their ‘topicworthyness’ in terms of their topic score
- Can the score of the subject affect the encoding of the object?
Relative object and subject rank

<table>
<thead>
<tr>
<th>case</th>
<th>full set</th>
<th>variation set</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>33.80</td>
<td>32.64</td>
</tr>
<tr>
<td>GA</td>
<td>20.06</td>
<td>16.84</td>
</tr>
</tbody>
</table>

Table: Mean difference between subject and object topic score by OCS object case, p-value=0.00169 (full set) and 0.00752 (variation set), Wilcoxon test.

Is the difference due to the sheer high rank of GA objects? Uncertain: subjects are significantly higher ranked with NA objects than with GA objects in the full set, but the difference is not significant in the variation set. Must take this with a pinch of salt.
Could low-ranking subjects explain unexpected GAs?

(21) \[i \quad se \quad m"o\ži \ na \ odr\ \ nos\ste \ člvka\]
and behold men on stretcher carrying man.ga.new
kai idou prosepheron autoi [... ] epi klinês
‘And behold, some men were carrying a man on a stretcher’ (Mar. Mt 9:2)

(22) \[edin\ \že \ ot\ stoj\stix\ [... ] udari \ raba\]
one ptc of standing struck servant.ga
archpriest’s
heis de tis t\n parest\kot\n [... ] epaisen ton doulon tou
archhiere\s
‘one of those standing there [... ] struck the archpriest’s servant’ (Mar. Mk 14:47)
Conclusions

- In addition to animacy, what drives OCS GA marking?
- There is certainly lexical variation, but socially prominent referents are also discourse-prominent.
- Clear sensitivity to information status – old and accessible objects are usually GA; new (whether specific or non-specific, whether anchored or not) vary between NA and GA.
- There is a clear tendency for new objects to be encoded as NA if they are prominent participants in a story (have many further pickups).
- Discriminatory perspective: relative ranking of subjects and objects could possibly be relevant (but results are weak).
- General caveat: not much data!
- The GA is on its way to becoming the default choice: even in Codex Suprasliensis there are considerably fewer NA attestations.
Mt. 14.1

νῦν ὁ Ἱρόδης τῆς τέτραρχης ἤκουσεν τὰς δοξάσεις Ἰησοῦν, καὶ εἶπεν στοῖς ἵπποις τοῖς δικαστικοῖς αὐτοῦ.

‘At that time Herod the tetrarch heard the reports about Jesus, and he said to his servants:

- **vrěmē < *vermen-s** (neuter consonant stem)
- **isus-ov-্‌
- Syntax: participle coordinated with finite verb
Mt. 14.2

sъ estъ ioanъ krъstitelъ.
‘This is John the Baptist.’

- -tel’-: derives agentive nouns from verbs (krъst-i-ti ‘baptise’)
- Syntax: copular construction
Mt. 14.2

τῇ ψυχῇ ὁ τῇ μαρτυροῦσα.
‘He has been raised from the dead.’
Mt. 14.2

i sego radi sily dějótě sę o nemy.
‘And that is why powers are at work in him.’
Mt. 14.3

For Herod had seized John and bound him and put him in prison for the sake of Herodias, his brother Philip’s wife

jemъ: past active participle of jε-ti (*im-tei) ‘take’, also occurs as ımъ

َا-stems: accusative *ām > ʔ, genitive ās > y ?? a much-discussed mystery!

Syntax: Syntactic subordination but discourse coordination
Mt. 14.4

glaše bo emu ioanъ.
‘For John used to say to him’

- The imperfect: Slavic innovation
- Suffix -ěa- or -aa- + -x-, endings of the root aorist
- First palatalisation in 2nd and 3rd pers sg, x > š before -e-
Mt. 14.4

ne dostoĭ tĭ imēti eje.
‘It is not lawful for you to have her.’

- Syntax: Impersonal construction
- Genitive of negation
Mt. 14.5

i xotę i ubiti uboě sę naroda. zane ěko prka iměxǫ i.
‘And though he wanted to put him to death, he feared the people, because they held him to be a prophet.’

- Syntax: comparison construction
Mt 14.6

But when Herod’s birthday came, the daughter of Herodias danced before the company and pleased Herod

Syntax: Dative absolute
- irodijad-in-ь: possessive adjective
- *дъктĕр > дъштĕ (but gen. дъштер-е)
- irod-ovi: dat.sg. from the ŭ-stems