SlaVaComp

COMPutergestützte Untersuchung von VAriabilität im KirchenSLAvischen

Juliane Besters-Dilger, Gerhard Schneider Evgenij Filimonov, Susanne Mocken, Irina Podtergera, Pino Marco Pizzo

Projektlaufzeit: 01.10.2012 - 30.09.2015

Albert-Ludwigs-Universität Freiburg









SlaVaComp

Auf dem Weg zu einem kirchenslavischen Meta-Glossar

Dr. Irina Podtergera, Slavisches Seminar

Dr. Susanne Mocken, Rechenzentrum

Albert-Ludwigs-Universität Freiburg



Slavisches Seminar Universität Freiburg





Ziel Nr. 1: Ein zweisprachiges – griechisch-kirchenslavisches – Online-Wörterbuch mit möglichst präzisen Zeit- und Ortsangaben zu den einzelnen slavischen Belegen

- ➤ Im Bereich der Slavistik: Eine Datenbank mit umfassendem Überblick über die lexikalische Variabilität des Kirchenslavischen sowie mit Merkmalskatalogen über die regionalen Schreibweisen
- Im Bereich der Informatik: Entwicklung von TextGrid-Instrumenten für die Verwendung in der Korpuslinguistik, Editionsphilologie und digitalen Lexikographie für Sprachen mit nicht-lateinischen Schriftsystemen

Warum mit TextGrid?

- Die angewandten Verfahren
 - lassen sich auf vormodernes Schrifttum jeglicher Herkunft übertragen
 - sollen für die nicht-slavistischen Humanities offen zugänglich gemacht werden
- Stärkere Vernetzung zwischen den Philologien bzw. Geisteswissenschaften
- Neue Impulse für eine engere Zusammenarbeit zwischen Informatikern und Philologen

Zum Projektablauf

- 18 digitalisierte Glossare (die meisten mit Formenindizes) und 6 Editionen
- Arbeitsphase 1: Konvertierung der Quellen aus ursprünglichen .doc-Formaten mit individuell entwickelten idiosynkratischen Kodierungen nach Unicode Plain Text
- Arbeitsphase 2a: Zusammenführung der Glossare zu einem Meta-Glossar, Zuordnung der slavischen Übersetzungsäquivalente einem griechischen Lemma
- Arbeitsphase 2b: Überführung in TEI/XML-Format, Importieren in TextGrid

Konvertierung nach Unicode

καθυβρίζειν πορπτατη Triod καθυπαντᾶν εφετατη Triod καθυπομένειν πρετραπέτη Triod καθυποτάσσειν, /-ττειν ποκορητη Triod kaqubr...zein por4gati Triod kaqupant©n s7r5tati Triod kaqupomšnein pr5tr7p5ti Triod kaqupot£ssein, -/ttein pokoriti Triod

Links Textausschnitt aus Lex. Eccl. 244. Verwendete Fonts: Greek (Griechisch), Klim (Kirchenslavisch), Times New Roman (Belegstellen). Rechts derselbe Text in Roman Cyrillic Std

Konvertierung nach Unicode

александрианинъ сущ. м. мн. им. александрїане ἀπ' Άλεξανδρείας 442a 6

александринскый прил. ед. м. им. александрінскый τῶν ᾿Αλεξανδρέων 406в 3 τῆς ᾿Αλεξανδρείας 477г 36

àëåêñàíäðèàíèíú ñóù. ì. ìí. èì. àëåêñàí¨äð¿àíå a'p ' 'Alexandreiva' 442à 6

àëåêñàíäðèèñêûè iðèë. åä. ì. èì. àëåêñàí"äð¿èñêûè tw~n 'Alexandrevwn 406â 3 th~' 'Alexandreiva' 477ã 36

Bell.lud. 697, eingegeben mit dem Font Flaccidus. Derselbe Text in Roman Cyrillic Std.

Unterschiedliche Schreibstile

Unicode		Idiosynkratische Fonts	
Codepoints	Characters	Cyrillica Ochrid10	Cyrillica Bulgarian Std
00C0	À	λ	A
00E7	Ç	Z	3
00E8	è	Н	И
00EC	ì	М	М
00F9	ù	Щ	Ψ
00FD	ý	ቴ	ቴ

Open Type Fonts: Ladoga

```
      Stil 1: абвезинорт уфхч

      Stil 2: абвезинорт уфху

      Stil 3: абвезинорт уфху

      В пезинорт уфху

      В пезино
```



Entstehung des Meta-Glossars

2 zentrale Fragen:

Wie wird das zugrundeliegende Material in XML umgewandelt?

Wie erhält man eine einheitliche XML-Struktur, die sowohl den einfach als auch den komplizierter aufgebauten Glossaren gerecht wird?



Entstehung des Meta-Glossars

Wie wird das zugrundeliegende Material in XML umgewandelt?

- In welchem Format liegen die Ausgangsdateien vor?
 - Microsoft Word-Format

ABER: Fonts sind nicht Unicode-konform!



Entstehung des Meta-Glossars

Wie wird das zugrundeliegende Material in XML umgewandelt?

- Limitierende Faktoren
 - Ausgangsformat .docx
 - strukturelle Heterogenität der Glossare



Entstehung des Meta-Glossars

Wie werden die zugrundeliegenden Word-Dateien in XML-Dateien überführt?

- keine Möglichkeit der vollautomatischen Überführung aller Dateien von .docx nach .xml
 - → Umwandlung erfolgt dateiweise per Skript



Entstehung des Meta-Glossars

Wie werden die zugrundeliegenden Word-Dateien in XML-Dateien überführt?

- Analyse der Input-Dateien nach verschiedenen Kriterien:
 - Codierung
 - Gliederungsmerkmale
- Zuweisung der passenden XML-Tags anhand einer durch den Philologen erstellten Strukturbeschreibung für jedes Glossars



Entstehung des Meta-Glossars

ἀγαθός: благь // τὸ ἀγαθόν: благо, благоє // τὸ ὄντως ἀγαθόν: соущоє благо // τὸ ὂν ἀγαθόν: соущоє благо // τὰ ἀγαθά: благам

Eintrag "ἀγαθός"

ἀγαθός (15)

□ благь *Sin23* f. 23r6, 23v24, 39r3, 309v21, *Sin24* f. 17r7, 26v19, 65r29-65v1, 251r12, 370r22

| ἀγαθόν τό

□ благо *Sin23* f. 39v15, 40r24, 51v16, 89v24, 309v29, *Sin24* f. 361r5

йγαθός adj. благословьнъ Parim благъ Evang Apost Psalt comPsalt Parim Triod блаженъ Triod добръ Evang Apost Psalt comPsalt Parim Triod идбъранъ Parim чистъ Evang щедръ Triod οὐκ ἀγαθός нεблагъ Psalt Parim άγαθόν n. благо Evang Apost Psalt comPsalt Parim Triod благодѣть Apost благость Psalt comPsalt Triod благостыни Psalt благъ Evang Apost добро Evang Apost Psalt comPsalt Parim άγαθὸν ποιεῖν благодѣнти Apost $\dot{\alpha}\gamma\alpha\theta\dot{\eta}$ $\pi\rho\hat{\alpha}\xi\iota\varsigma$ благодѣщниє comPsalt



Entstehung des Meta-Glossars

Eintrag ,ἀγαθός"

ἀγαθός благо 167: 8; 266: 3; 329: 28; 416: 8 благодать 1: 11 благодать (*благо) 270: 25; 331: 18 благость 173: 13; 335: 26; 346: 23 благочьстие (*благо) 333: 1 благъ 3: 20, 23; 15: 20; 20: 12; 144: 9; 298: 2; 335: 24; 336: 16; 337: 21; 346: 1, 16; 348: 14; 355: 22; 369: 3; 411: 17 благый 3: 12; 8: 17; 13: 7, 12; 14: 23; 15: 23; 20: 22; 25: 1; 57: 21; 145: 17; 154: 7; 157: 1; 162: 10; 173: 4; 178: 9; 201: 20; 229: 12; 231: 19; 233: 21; 235: 5 (2x); 253: 12; 269: 15; 272: 19; 275: 1; 284: 22; 287: 21; 299: 5; 301: 2; 303: 6; 304: 22; 312: 13; 327: 13; 328: 27; 331: 26; 339: 11; 346: 16, 18; 351: 4; 352: 21, 22, 23; 353: 8; 354: 18; 377: 3, 24 (2x); 383: 2; 388: 14; 390: 7; 417: 9; 436: 20 добро сущ. 5: 17; 225: 5; 226: 11; 335: 18; 362: 9; 367: 23; 378: 11; 391: 7; 411: 6; 416: 17 добро предик. 345: 13 добровазньство 334: 18 добронравьнь 115: 22 добръ 339: 19; 396: 23 добрый 70: 2; 74: 15; 86: 16; 88: 1; 149: 18; 179: 19; 276: 24; 390: 28; 391: 10; 392: 14; 393: 27 имѣниє 364: 11 мирьно 339: 29 пользьный 413: 3 сътъжжание 254: 23 оуне предик. 259: 21; 398: 7 τὰς ἀγαθὰς πράξεις добродѣюние 305: 9 ἐκπονεῖν τὰ ἀγαθά добродѣюние 272: 23 [ἀγαθῶν ἐλπίδων] добродѣюние 222: 22 φρένας ἀγαθάς доброоумие 146: 22 καλὰ καὶ ἀγαθά добрый 371: 6 τῶν καλῶν καὶ ἀγαθῶν добрый 371: 5 μέγα... ἀγαθόν οуне предик. 226: 25



Entstehung des Meta-Glossars

```
ἀγαθός adj. благословинт. Parim
        благт Evang Apost Psalt
                comPsalt Parim Triod
       блаженъ Тпоп
      и добръ Evang Apost Psalt
      - - - comPsalt Parim Triod
       идбъранъ Parim
       чистъ Evang
       щедръ Triod
  οὐκ ἀγαθός нєблаг (Psalt Parim
  άγαθόν n. благо Evang Apost Psalt
      - - - comPsalt Parim Triod
      ■ благодѣть Apost
      благость Psalt comPsalt Triod
        благостыни Psalt
        благт Evang Apost
        добро Evang Apost Psalt
               comPsalt Parim
  ἀγαθὸν ποιεῖν благодѣнти Apost
  \dot{\alpha}\gamma\alpha\theta\dot{\eta} \pi\rho\hat{\alpha}\xi\iota\varsigma благодѣюниє comPsalt
```

```
Eintrag
"ἀγαθός"
```

```
άγαθός: σπαιτ<mark>//</mark> τὸ ἀγαθόν: σπαιο,
σπαιοις // τὸ ὄντως ἀγαθόν:
соущоιє σπαιο // τὸ ὂν ἀγαθόν:
соущоιє σπαιο // τὰ ἀγαθά:
σπαιαις
```

```
ὰγαθός (15)

□ благт Sin23 → 23r6, 23v24, 39r3, 309v21, Sin24 f. 17r7, 26v19, 65r29-65v1, 251r12, 370r22

□ ἀγαθόν τό
□ благо Sin23 f. 39v15, 40r24, 51v16, 89v24, 309v29, Sin24 f. 361r5
```



Entstehung des Meta-Glossars

- Strukturelle Heterogenität der Glossare erfordert ein Umwandlungsskript pro Glossar
 - → Zeitaufwand!
 - → Grad der Automatisierung bei der Umwandlung wirft die Frage nach "Kosten und Nutzen" auf



Entstehung des Meta-Glossars

Wie erhält man eine einheitliche XML-Struktur, die sowohl den einfach als auch den komplizierter aufgebauten Glossaren gerecht wird?



Entstehung des Meta-Glossars

•ἀγαθός adj. благословьнъ Parim благъ Evang Apost Psalt comPsalt Parim Triod блаженъ Triod добръ Evang Apost Psalt comPsalt Parim Triod идбъранъ Parim чистъ Evang щедръ Triod οὐκ ἀγαθός нεблагъ Psalt Parim άγαθόν n. благо Evang Apost Psalt comPsalt Parim Triod благодѣть Apost благость Psalt comPsalt Triod благостыни Psalt благъ Evang Apost добро Evang Apost Psalt comPsalt Parim *ἀγαθὸν ποιεῖν* благодѣнти Apost $\dot{\alpha}\gamma\alpha\theta\dot{\eta}$ $\pi\rho\hat{\alpha}\xi\iota\varsigma$ благодѣюниє comPsalt

Eintrag "ἀγαθός"

άγαθός: благь // τὸ ἀγαθόν: благо, благое // τὸ ὄντως ἀγαθόν: соущою благо // τὸ ὂν ἀγαθόν: соущою благо // τὰ ἀγαθά: благам

×ML

άγαθός (15)

□ благь *Sin23* f. 23r6, 23v24, 39r3, 309v21, *Sin24* f. 17r7, 26v19, 65r29-65v1, 251r12, 370r22

Ι ἀγαθόν τό

□ благо *Sin23* f. 39v15, 40r24, 51v16, 89v24, 309v29, *Sin24* f. 361r5



Entstehung des Meta-Glossars

Wie erhält man eine einheitliche XML-Struktur, die sowohl den einfach als auch den komplizierter aufgebauten Glossaren gerecht wird?

- Grundstruktur nach den Vorgaben der TEI entwickeln
- einfaches Glossar als Ausgangspunkt
- Struktur kontinuierlich erweitern



Entstehung des Meta-Glossars

```
<entry xml:id="xxx-xx">
    <note>
        <orig>Originaleintrag aus der Druckvorlage</orig>
    </note>
    <form type="hyperlemma" xml:lang="xxx">
       <orth>HYPERLEMMA</orth>
    <form type="lemma" xml:lang="xxx">
       <orth>LEMMA</orth>
        <cit type="counterpart" xml:lang="xxx">
            <form type="hyperlemma" xml:lang="xxx">
                <orth>Hyperlemma der Zielsprache</orth>
            </form>
            <form type="lemma" xml:lang="xxx">
                <orth>Lemma der Zielsprache</orth>
            <quote>Counterpart aus der Druckvorlage</quote>
        </cit>
    </form>
</entry>
```



Entstehung des Meta-Glossars

```
άγαθός adj. благословыть Parim
       олагъ Evang Apost Psan
             comPsalt Parim Triod
       олаженъ Triod
       добръ Evang Apost Psalt
             comPsalt Parim Triod
       идбъранъ Parim
       чистъ Evang
       щедръ Triod
  οὖκ ἀγαθός нεблагъ Psalt Parim
  άγαθόν n. благо Evang Apost Psalt
             comPsalt Parim Triod
       благодѣть Apost
       благость Psalt comPsalt Triod
       благостыни Psalt
       благъ Evang Apost
      добро Evang Apost Psalt
             comPsalt Parim
  άγαθὸν ποιεῖν благодѣнти Apost
  άγαθὴ πρᾶξις δηατομѣκιниє comPsalt
```

```
<entry xml:id="arg-46">
    <form type="hyperlemma" xml:lang="grc">
         <orth>άγαθός</orth>
                "lemma" xml:lang="grc">
         <orth>άγαθός:/orth>
             <gram type="pos">adj</gram>
         </gramGrp>
         <cit type="counterpart" xml:lang="cu">
             <form type="hyperlemma" xml:lang="cu">
                 <orth>благословьнь</orth>
             <form type="lemma" xml:lang="cu">
                 <orth>благословьнъ</orth>
             <quote><u>благословьнъ</u></quote>
             <bibl><biblScope unit="manuscript" {Parim /biblScope></bibl>
         <cit type="counterpart" xml:lang="cu">
             <form type="hyperlemma" xml:lang="cu">
                 <orth>благъ</orth>
             </form>
             <form type="lemma" xml:lang="cu">
                 <orth>благъ</orth>
             <quote>благъ</quote>
             <bibl><biblScope unit="manuscript"
                                                         /biblScope></bibl>
             <bibl><bibl>cope unit="manuscript"
                                                 -Apost<
                                                         biblScope></bibl>
             <bibl><biblScope unit="manuscript"</pre>
                                                 >Psalt</biblScope></bibl>
             <bibl><bibl>cope unit="manuscript"
                                                 >comPsal
                                                          t</biblScope></bibl>■
             <bibl><biblScope unit="manuscript"</pre>
                                                  Parim< biblScope></bibl>
             <bibl><biblScope unit="manuscript"</pre>
                                                         /biblScope></bibl>
```



Entstehung des Meta-Glossars

κυλισμός m. *κυλισμὸς βορβόρου* калъ тимѣнию Apost



Entstehung des Meta-Glossars

```
<entry xml:id="arg-5050">
    <form type="hyperlemma" xml:lang="grc">
        <orth>κυλισμός</orth>
    <form type="lemma" xml:lang="grc">
        <orth>κυλισμός</orth>
        <gramGrp>
            <gram type="pos">noun</gram>
            <gram type="gen">m</gram>
        </gramGrp>
    </form>
    <form type="variant">
        <orth>κυλισμὸς βορβόρου</orth>
        <form type="hyperlemma" xml:lang="grc">
            <orth>κυλισμός</orth>
            <orth>βόρβορος</orth>
        </form>
        <form type="lemma" xml:lang="grc">
            <orth>κυλισμὸς βορβόρου</orth>
        </form>
        <form type="colloc" xml:lang="grc">
            <colloc>κυλισμὸς βορβόρου</colloc>
            <orth>κυλισμὸς</orth>
            <orth>βορβόρου</orth>
        </form>
```

```
<cit type="translation" xml:lang="cu">
            <form type="hyperlemma" xml:lang="cu">
                <orth>каль</orth>
                <orth>тимЪниє</orth>
            </form>
            <form type="lemma" xml:lang="cu">
                <orth>калъ тимѣнин</orth>
            <form type="colloc" xml:lang="cu">
                <colloc>калъ тимѣнин</colloc>
                <orth>калъ</orth>
                <orth>тимѣнин</orth>
            </form>
            <quote>калъ тимѣнина</quote>
            <bibl>Apost</bibl>
        </cit>
    </form>
</entry>
```

Kirchenslavisches Meta-Glossar



Was passiert mit den in XML vorliegenden Glossaren?

- Inhaltliche, d.h. philologische Kontrolle
- Eingang in die Datenbank