

# Geschichten aus dem 'Kleinen Paris' Bulgariens

## Eine Digitale Edition des *Sbornik* von Pop Punčo (1796)

### 1. Projektvorhaben: Motivation, Relevanz, Ziel

Dass die Grundlage der Dynamik sprachlicher Entwicklungen ganz zentral auf Variation und diese ihrerseits zu einem grossen Teil auf Faktoren wie Sprach- und Kulturkontakt, Mehrsprachigkeit und L2-Erwerb beruht, lässt sich für die jüngere Zeit zunehmend empirisch anhand von grossen Datenmengen zeigen, die in linguistisch aufbereiteten Korpora zugänglich sind. Beachtung finden dabei insbesondere übergreifende konvergente Entwicklungen, da in einer globalisierten Welt Polyglosse und Kontakte zunehmend intensiver werden. Um jedoch auch Einblick in kontakt- und mehrsprachigkeitsbedingte Entwicklungen früherer Zeiten zu erhalten und so zu einem umfassenderen Bild der Dynamik sprachlicher Entwicklungen zu gelangen, sind Belege aus überlieferter älterer Literatur notwendig. Diese sind insbesondere dann von Interesse, wenn sie noch aus der Zeit vor der Kodifizierung sprachlicher Varianten als Standardsprachen und ihrer Verbreitung durch den Buchdruck stammen und somit eine Annäherung an die individuelle Sprachverwendung und die Herausbildung übergreifenderer Normen erlauben. Eine Aufbereitung dieser Daten für linguistische Zwecke und für ihre notwendige philologische Einordnung stellt gerade für Sprachen, deren vorstandardisierte Varietäten durch eine vergleichsweise geringe Zahl an Quellen belegt ist, vielfältige Herausforderungen dar. Dies trifft auch auf das süd- und balkanslavische Sprachgebiet zu, dessen literarische Diachronie aufgrund politischer Ereignisse durch eine diskontinuierliche Entwicklung und ein erst spät einsetzende vernakulare Literalität gekennzeichnet ist. Damit werden einzelne Quellen für das Nachzeichnen der sprachlichen Entwicklung umso wichtiger, zumal das Aufarbeiten der sprachlichen und literarischen Prozesse zudem Einblicke in die komplexe Kultur- und Ideengeschichte einer Region verspricht, die durch das Aufeinandertreffen von vielfältigen politischen, religiösen und kulturellen Einflüssen gekennzeichnet ist.

Ein Fenster zu den genannten Entwicklungen stellt der Sammelband des Popen Punčo (*Pop Punčov Sbornik*) dar, der Ende des 18. Jh. im Nordosten des heutigen Bulgariens entstand und eine Vielzahl an Texten unterschiedlicher Genres und unterschiedlicher Register, eingebettet in unterschiedliche Traditionen umfasst. Angesicht ihrer internen Vielfalt kann diese Quelle als ein Ausgangspunkt zum Verfolgen der Text- und Sprachgeschichte des balkanslavischen Raums dienen.

Um dies zu ermöglichen, besteht das Ziel des geplanten Projekts im Erstellen einer digitalen Edition des *Sbornik*, die für Linguisten (als morphosyntaktisch annotiertes Korpus) ebenso von Nutzen ist, wie für Philologen (durch übersichtliche Darstellung und Hypertextualisierung). Die digitale Edition wird nicht nur das Einbeziehen der Textdaten in quantitativen Studien zu areallinguistischen und dialektologischen Fragen ermöglichen, sondern auch das Sammeln von Daten zu den beinhalteten Texten und ihrer Sprache selbst. Zu diesem Zweck soll über 'philologisches Crowdsourcing' mithilfe interessierter philologischer Kolleg\*innen Information zur Textgeschichte der einzelnen Texte zusammengetragen werden, so dass daraus deren Verbreitungswege und damit die Transmissionswege, deren Zentren und Akteure, abgeleitet werden können.

### 2. Der *Pop Punčo Sbornik*

Der *Sbornik* ist aus diachroner und dialektaler Perspektive einzigartig. Die Sprache des Textes spiegelt den lokalen Dialekt des osmanischen Grenzgebiets an der unteren Donau spätes 18. Jahrhunderts wider, der sich in einer Region intensiven Sprachkontakts befindet und in den Literaturdenkmalen aus

dieser Region ansonsten kaum dokumentiert ist. Der *Sbornik* umfasst 763 Seiten Originaltext, plus 3 Seiten späterer Notizen (zusammen 383 Blätter). Er beinhaltet 80 Kapitel, meistens hagiographische und biblische Geschichten, inklusiv zweier Vorworte.

Die Quelle befindet sich in der Handschriftsammlung der Bulgarischen Nationalbibliothek in Sofia unter der Signatur 693. Die Handschrift wurde von der Bibliothek vollständig gescannt und ist als digitales Facsimile auf der [World Digital Library](#) und [Europeana](#) abrufbar. Der *Sbornik* wurde am Slavischen Seminar der Universität Zürich bereits in einen digitalen Text konvertiert.

Der *Sbornik* wurde im 1796 in Mokreš (in der Vergangenheit auch als 'Kleines Paris' bezeichnet), im Nordwesten des heutigen Bulgariens, von einem örtlichen Priester namens Punčo verfasst, dessen Sprache gut der lokalen gesprochenen Sprache entspricht. Der *Sbornik* ist ein prototypischer vorstandartisierte Text: er entsteht nicht nur in der Zeit vor der Vereinheitlichung der Literatursprache, sondern auch bevor die theoretische Debatte über die moderne bulgarische Orthographie (z.B. Rilski 1835, Keremidčiev 1943, Velčeva 1966) überhaupt begann. Der Autor gehört keinem zu dieser Zeit existierenden literaren Kreis (z.B. wie Rila Kloster, Vraca- oder Kotel-Schule) im Land an, und folgte den ansonsten üblichen orthographischen Normen nur eklektisch und inkonsequent. Er benutzt z.B. nur ein Jer-Zeichen (ѣ), die Akzentuierung schwankt, die Interpunktion ist unsystematisch, wenn überhaupt verwendet.

Die dialektale Vorlage für die Sprache des *Sbornik* entspricht der nordwestlichen Dialektgruppe des Bulgarischen. Dieser Dialekt liegt nicht nur in der Grenzzone zwischen dem Bulgarischen, Štokavisch-Serbischen und Rumänischen, sondern auch an der Peripherie des sprachlichen Balkanareals. Diese Zone ist für dialektologische und areallinguistische Untersuchungen bereits seit Ende des 19. Jh. von Interesse (z.B. Belić 1905, Sobolev 1998) und der *Sbornik* inzwischen als eine bedeutende Quelle diachroner Daten anerkannt (z.B. Petkanova-Toteva 1965, Šaur 1970, Sonnenhauser 2015). Nur wenige andere Texte aus dieser Zeit entsprechen den Dialekten dieser Zone (z.B. *Temski rukopis*, cf. Vasilev 1986); sie sind zudem in Umfang und inhaltlicher Vielfalt wesentlich kleiner.

Inhaltlich ist der *Sbornik* ein Beispiel für die spätere Entwicklung der *Damaskini*-Tradition. Diese literarische Tradition entwickelte sich ursprünglich aus Übersetzungen des neugriechischen *Thēsauros* von Damaskinos Studitis aus dem 16. Jahrhundert, der in späteren Handschriften um weitere Kapitel aus anderen Quellen ergänzt wurde. So kommen zunehmend zu den üblichen Hagiographien und Homilien auch säkulare Werke wie z.B. Anekdoten oder historische Chroniken hinzu. Als Autor gönnt sich Pop Punčo dabei stilistische Freiheit, indem er die meisten Kapitel in eigenen Worten neu erzählt und damit auch sprachlich von Vorlagen abweichen kann. Im Vergleich zu den anderen *Damaskini*-Werken ist der *Sbornik* auch aus dieser Sicht eigenartig. Es ist nicht nur eine gut erhaltene Quelle der historischen dialektalen Daten, sondern auch eine besondere Edition, die für die genuin balkanslavische Textgeschichte ebenso wichtig ist, wie für die Textgeschichte des Kirchenslavischen.

### 3. Edition und Korpus

Das Editionsprojekt beinhaltet zwei Teilbereiche: die Verbindung der Online-Edition mit den Scans der Handschrift und dem digitalisierten Text einerseits, und mit einem linguistisch annotierten und durchsuchbaren Korpus andererseits. Für beide Aspekte liegen bereits erste Vorarbeiten vor, auf die in der Projektarbeit aufgebaut werden kann.

Für die Verbindung der Online-Edition mit den Scans der Handschrift und dem digitalisierten Text wurde der Text durch die OCR-Software [Transkribus](#) im UTF-8 Format digitalisiert, mit erhaltenen Rand- und Fussnotizen, suprasegmentaler Markierung (Akzentuierung, Spiritus, Verkürzungen) und Ordnung nach Kapitel und Seiten. Für eine vollständige Edition sind noch letzte Korrekturen notwendig. Durch eine Konversion in das XML-basierte Format [TEITOK](#) (*Tokenized TEI Environment*; TEI

*Text Encoding Initiative*) werden weitere Textdaten einbezogen werden, wie z.B. die Farbe, Ornamente oder Illustrationen, sowie eine diplomatisierte Textfassung (in vereinfachter Latein-Schrift, ohne suprasegmentaler Markierung, für einfachere Volltext-Suche) und Übersetzungen. Die einzelnen Kapitel können als selbstständige *chunks* durch eindeutige Verweise abrufbar werden und dadurch als 'Prätext-Zitate' (Schweier 2012) interpretiert werden, die intertextuelle Bezüge zu ihren Vorlagen aufweisen.

Die Online-Edition wird weiterhin mit einer wiki-basierten Maske ausgestattet, wodurch weitere Kommentare zu den einzelnen Kapitel/*chunks* eingeführt werden können. Auf dieser Weise kann die Edition selbst als eine Plattform dienen, so dass die Analyse des *Sbornik* Verbindungen zu ähnlichen Projekten zur kirchenslavischen und balkanslavischen Literatur (z.B. [TOROT](#), [TITUS](#), [Obdurodon.org](#), [Cyrillomethodiana](#)) und Dialektologie (z.B. wie die Datenbank für das [Makedonische](#)), sowie auch zu historische balkan-orthodoxen und osmanischen Studien herstellen kann.

Der zweite Bereich des Editionsprojektes umfasst ein Korpus für quantitativen Studien. Der digitale Text, bzw. die *chunks* basierend auf einzelnen Kapiteln, wird vollständig tokenisiert und lemmatisiert. Für die Lemmatisierung werden Wörterbücher der für die Literatur der relevante Periode (z.B. Demina et al. 2012) sowie etymologische Wörterbücher der Standardsprachen aus dem Gebiet (z.B. Skok 1971, Georgiev et al. 1972–2006) verwendet

Die morphologische Annotation basiert auf [MultextEast v5](#), einem weitverbreiteten Modell für die Annotation von Wortarten und morphologischer Information. Die Unterschiede zwischen der Sprache von Pop Punčo und dem modernen Bulgarischen und Serbischen erfordern allerdings eine Anpassung des Tagsets. Diese Unterschiede zeigen sich z.B. hinsichtlich des postponierten Artikels, der in der Literatur aus dem Ende des 18. Jahrhundert üblich war. Die Beispiele in (1)–(3) weisen drei Varianten eines grammatischen Markers auf: zwei gekürzten Formen des Artikels in (1) und (2), eine volle Form in (3). Solche Formen waren (bzw. sind) charakteristisch für bestimmte Dialekte, unbekannt weder in früherer, noch späterer, standardisierter Literatur. Sie können somit helfen, den Text dialektal einzuordnen.

(1) *móžete da go izvadite o<sup>t</sup> stl<sup>b</sup>pa*  
 can.2PL.PRS to it.ACC remove.2PL.PRS from pillar.DEF  
 'Könnt ihr ihn aus der Säule herausnehmen?'  
*Pop Punčov Sbornik* (1796: 76r)

(2) *povele da iskopajut<sup>b</sup> dupka pri stl<sup>b</sup>pa*  
 command.3SG.AOR to dig.3PL.PRS hole.NOM.SG next to pillar.GEN.SG.DEF  
 'Er befahl ihnen, ein Loch bei der Säule zu graben.'  
*Pop Ioannov Damaskin* (1788; zit. nach Angelov 1958: 102)

(3) *podkàni se da slézi ot stól'pot<sup>b</sup>*  
 decide.3SG.AOR REFL to descend.3SG.PRS from pillar.DEF  
 'Er hat sich entschieden, von der Säule herabzusteigen.'  
*Berlin Damaskin* (1803: 182r)

Das Tagset beinhaltet spezifische Markierungsmodalitäten für die oben illustrierten und weitere, ähnlich gelagerte Fälle. Es kann damit sowohl innovative, als auch dialektalspezifische und archaische (z.B. verbaler Infinitivformen) Phänomene erfassen. Dadurch kann das *Sbornik*-Korpus zusammen mit bestehenden Korpora des Kirchenslavischen oder mit Daten aus benachbarten Varietäten für mehrsprachige Big-Data Analysen nutzbar gemacht werden, die für diese Varietäten bisher noch nicht möglich sind.

Die syntaktische Annotation basiert auf dem Modell [Universal Dependencies v2](#), das sich die Anwendbarkeit auf jede Sprache zum Ziel setzt. Gerade bei der vorstandardisierten, eklektischen

Sprache des *Sbornik* ist eine derartige universale Anwendbarkeit entscheidend. Einzelne Teile des *Sbornik* wurden in Vorarbeiten bereits mit dieser syntaktischer Annotation ausgestattet.

Die beiden Bestandteile unseres Projektes werden durch die Anwendung von XML-basierten Formaten online gestellt und durch Hyperlinks mit den Datenbanken bzw. Quellen verknüpft. Für das Hochladen werden [Webhosting-Services](#) der ZI-Abteilung der Universität Zürich benutzt. Als Schnittstelle für den doppelten Zugriff auf die Daten dient das XML-Format, das sowohl der Datenbankstruktur des Korpus als auch der TEI-Kodierung der digitalen Edition zugrunde liegt. In dieser Verknüpfung von Korpustechnologie und Editions Handwerk besteht der grosse Mehrwert dieses Vorhabens, nicht nur in Bezug auf die konkreten Daten, sondern auch als Vorlage für ähnliche künftige Projekte.