

# Rapid #: -10732967

CROSS REF ID: **98691**

LENDER: **GZM :: Memorial Library**

BORROWER: **PIT :: Main Library**

TYPE: Book Chapter

BOOK TITLE: Studia Caroliensia : papers in linguistics and folklore in honor of Charles E. Gribble

USER BOOK TITLE: Studia Caroliensia: Papers in Linguistics and Folklore in Honor of Charles E. Gribble

CHAPTER TITLE: Macro-Level and Micro-Level Structural Analysis of Mixed Content Miscellanies

BOOK AUTHOR: David J. Birnbaum

EDITION:

VOLUME:

PUBLISHER: Slavica Publishers

YEAR: 2006

PAGES: 41-56

ISBN: 9780893573379

LCCN:

OCLC #:

Processed by RapidX: 6/21/2016 9:05:09 AM



This material may be protected by copyright law (Title 17 U.S. Code)

---

UW-Madison ILL Lending (GZM)  
728 State Street / Madison, WI 53706  
Email: gzmil@library.wisc.edu



GZM TN: 2844497

**Borrower:** RAPID:PIT

**Lending String:**

**Patron:**

**Journal Title:** Studia Caroliensia

**Volume:** Issue:

**Month/Year:** 2006 **Pages:** 41-56

**Article Author:** David J. Birnbaum

**Article Title:** Macro-Level and Micro-  
Level Structural Analysis of Mixed  
Content Miscellanies

**OCLC Number:** 85835051

**RAPID Number:** -10732967



**Location:** MEMORIAL stk

**Call #:** P121 S814 2006

**Request Date:** 6/20/2016 2:53:07 PM

**Odyssey:** 206.107.44.183

**Ariel:** 130.49.208.132



**Borrowing Notes:**

2m n  
**RAPID**

ODYSSEY

m 14

# Macro-Level and Micro-Level Structural Analysis of Mixed Content Miscellanies<sup>1</sup>

David J. Birnbaum

*Abstract.* The present report employs cluster analysis to examine structural correlations in the text of the series of apocryphal "Stories about Abraham" in twenty-two manuscripts. This analysis confirms many of the conclusions that scholars had drawn previously about the textual transmission of these stories using traditional methods; it also identifies new relationships among the manuscripts.

## Philological Background

The series of apocryphal "Stories about Abraham" (henceforth "Abraham") contains up to nine component texts, each of which may appear in up to three redactions (Miltenova 2005).<sup>2</sup> The fact that this work consists of discrete and identifiable subcomponents means that from a structural perspective it can be considered a sort of internal fixed-content miscellany, or miscellany within a miscellany, comparable to the many mixed-content or fixed-content miscellany manuscripts that are well known from the middle Bulgarian period (Miltenova 1982).<sup>3</sup> "Abraham" is closer in form to a fixed-content miscellany than to a mixed-content miscellany in that the contents are constrained by the need to refer to the Old Testament legend of the patriarch Abraham, and the order of the constituent parts is partially constrained by logical considerations (e.g., we expect that the "Story of Abraham" will precede "Ishmael" and "Isaac" as fathers precede sons chronologically; we expect that "Ishmael,"

---

<sup>1</sup> An earlier version of this paper was delivered at the March 2005 Midwest Slavic Conference in Columbus, OH. I am grateful to Sibelan E. S. Forrester, M. A. Johnson, Predrag Matejić, and Anisava Miltenova for comments and suggestions.

<sup>2</sup> "Story of Abraham" (two redactions), "Story of Sarah," "How Sarah Advised Abraham" (two redactions), "Ishmael," "Isaac" (two redactions), "Melchizedek," "Abraham and the Holy Trinity" (two redactions), "Death of Abraham" (three redactions), and "Samuel."

<sup>3</sup> Other micro-level miscellany-like work clusters include the "Erotapokriseis" (Miltenova 2004), the "Fiziolog," and the "Tales of Evil Women" ("Zlye ženy").

when present, will precede "Isaac" textually as Ishmael preceded Isaac chronologically; etc.).

A survey of twenty-two manuscripts that contain "Abraham" reveals patterns of structural agreement, which is to say that the appearance of certain "Abraham" stories in specific redactions in a particular manuscript is strongly correlated with the appearance of certain other "Abraham" stories in the same manuscript. An analysis of these correlations shows that these patterns of agreement can be used to divide the corpus into natural groups according to structural similarity.

### Mixed-Content Miscellanies

Miltenova (1982 and elsewhere), building on the work of Istrin (1922), Speranskij (1963), and others, has argued that coincidences in the presence or absence and in the order of texts within a miscellany may imply shared textual transmission, and her conclusions have recently been corroborated by computational analysis (Birnbaum 2003, Dubin and Birnbaum 2004). Until now, however, most arguments of this sort (and the supporting computational analysis) have been made on the "macro level," that is, on the level of examining the presence or absence of entire specific works within a mixed-content miscellany manuscript. In fact, however, if we view works like "Abraham" as micro-miscellanies (that is, as miscellanies embedded within miscellanies), it should be possible to apply the same type of analysis at the micro level, as Miltenova has done for "Abraham" (2005) and the *erotapokriseis* (2004), and to perform the same sort of computational validation.

### Micro-Level Clustering

The present study is based on a micro-level cluster analysis of the "Abraham" texts in twenty-two manuscripts (see list in appendix A and detailed information in Miltenova 2005). Six clustering methods<sup>4</sup> were applied both with and without length-normalization, yielding twelve dendrogrammatical visualizations (appendix B). The small number of component texts in "Abraham" suggested that length-normalized similarity calculations were likely to be

<sup>4</sup> Cluster analysis builds family trees by grouping items according to degree of similarity. The present study used the following six methods: agglomerative single linkage, average linkage, complete linkage, Ward's linkage, weighted linkage, and divisive linkage. Similarities were calculated according to an algorithm first presented in Birnbaum 2003 and described from a clustering perspective in Dubin and Birnbaum 2004. Dendrograms were generated with the R software system.

most revealing, and agglomerative complete linkage was favored because of its greater conservatism, although it turned out that the differences between the analyses with and without length-normalization were relatively minor.

Micro-level analysis divided the corpus coarsely into four clusters of two or more manuscripts, as follows:

- 1) 29, 1938, 109, Belovo (and sometimes 794)
- 2) 241, 828, iv.a.24 (and sometimes 794)
- 3) 100, 149, 433, 738, 104, 326 (and sometimes 794; also possibly Svištov, about which see below)
- 4) 213, 677, 740, 1221, 13225 (and possibly Svištov, about which see below)

Some of these clusters were extremely stable across methods; e.g., the first emerged under eleven of the twelve methods. Correspondences of specific works within the clusters are easily visualized using plectogram renditions (appendix C).<sup>5</sup> 476 and iii.a.43 emerged as a smaller group under Ward's and weighted clustering without length-normalization and under complete linkage with length-normalization (where the group also included the promiscuous 794).

Most of the natural groups that emerge from the cluster analysis mirror the groupings that philologists had postulated previously on the basis of traditional non-formal visual comparisons, as follows (Miltenova, p.c.):

- 1) Set 1 corresponds to what has been called the Archaic Group, with a possible Athonite protograph.
- 2) Set 2 has not traditionally been identified as a coherent and discrete branch of the tradition.<sup>6</sup>
- 3) Part of Set 3 (100, 149, 433) corresponds to what has been called the Bulgarian Group, which has been considered an offshoot of the Archaic tradition. Another part of Set 3 (326, 738) corresponds to what has been called the Mixed Balkan Group, with multiple protographs. Clustering within Set 3 in most of the dendrograms

<sup>5</sup> See Birnbaum 2003 and, originally, Olmsted 1994 for a discussion of plectograms.

<sup>6</sup> The degree of similarity within this group is not as strong as within the others, but it is visible under agglomerative complete and Ward's linkage with and without length-normalization and under average and weighted linkage with length-normalization.

reflects this subdivision, as well as the relationship of both of these subgroups to Set 1.

- 4) Set 4 corresponds to what has been called the Wallachian-Moldavian Group (with a middle Bulgarian protograph).

Miltenova (2005) has emphasized that one stable feature that unites Set 3 internally and with Svištov is the presence of the archaic first redaction of "The Death of Abraham." The other text in Svištov, the second redaction of "Abraham and the Holy Trinity," unites that manuscript instead with Set 4. In the analysis presented here, Svištov does not cluster closely with Set 3 under any method; instead, some clustering methods associate it with Set 4, while others do not associate it closely with either Set 3 or Set 4.

### Conclusions

That the computational analysis confirms classification that had been determined previously by visual inspection supports the hypothesis that the computational clustering is producing meaningful results, that is, results that are congruent with expert philological judgment. In light of that confirmation, one might draw the following conclusions that go beyond traditional observations:

- 1) The Mixed Balkan group, previously identified as having multiple protographs, is related most closely to the Bulgarian Group.
- 2) Set 2, not previously identified as an integral cluster, shows at least modest internal coherence. Under some models it emerges as related most closely to the Archaic and Bulgarian and Mixed Balkan Groups, while under others it is not especially close to any of the other Sets. Specifically:
  - a. (3 methods) Sets 1 and 3 form a cluster. The members of Set 2 are related to them, but do not form a distinct subcluster by themselves (average, single, weighted linkage without length-normalization).
  - b. (2 methods) Sets 1 and 3 form a cluster. Set 2 forms a distinct cluster that is closer to Sets 1 and 3 than it is to 4 (Ward's linkage without length-normalization, Ward's linkage with length-normalization).

- c. (2 methods) Sets 1 and 3 form a cluster. Set 2 forms a distinct cluster that is no closer to 1 and 3 than it is to 4 (complete linkage without length-normalization, complete linkage with length-normalization).
- d. (1 method) Sets 1 and 3 form a cluster. Set 2 does not form a distinct cluster and its members are no closer to 1 and 3 than they are to 4 (divisive linkage without length-normalization).
- e. (4 methods) Sets 1 and 2 form a cluster. Set 3 forms a distinct cluster that is closer to 1 and 2 than it is to 4 (average, single, weighted, and divisive linkage with length-normalization).

Relationship (c) emerges under the most conservative models, complete linkage with and without length-normalization, but most other models entail some form of close relationship among Sets 1, 2, and 3.

### Further Research

Descriptions of the full contents of the twenty-two manuscripts studied here are not all available in electronic form, which makes it impossible to use the software tools described in this report to compare the clustering that emerges from the manuscripts as a whole to the clustering that emerges from "Abraham" alone. To the extent that the latter reflects genuine shared textual transmission, one would expect to find at least a modest correlation between the two levels, although verification will have to await the accumulation of additional data.

### References

- Birnbaum, David J. 2003. "Computer-assisted analysis and study of the structure of mixed-content miscellanies." *Scripta & e-Scripta* 1: 15–54. Preprint available at: <[http://clover.slavic.pitt.edu/~djb/2003\\_ljubljana/2003\\_ljubljana\\_paper.pdf](http://clover.slavic.pitt.edu/~djb/2003_ljubljana/2003_ljubljana_paper.pdf)>.
- Dubin, D. and D. Birnbaum. 2004. "Measuring similarity in the contents of medieval miscellany manuscripts." Paper presented at the ninth biennial meeting of the International Federation of Classification Societies, Chicago, IL, July 2004.

- Istrin, V. M. 1922. *Očerki istorii drevnerusskoj literatury domoskovskogo perioda, XI–XIII vv.* Petrograd: Nauka i škola. Reprint: Serija "Klassičeskaja učebnaja kniga." Moskva: Akademija, 2003.
- Miltanova, Anisava. 1982. "Kŭm karakteristikata na bŭlgarskata literatura prez XV–XVII v. (sbornici sŭs smeseno sŭdŭrŭžanie)." Diss. Sofija.
- . 2004. *Erotapokriseis. Sŭčinenijata ot kratki vŭprosi i otgovori v staro-bŭlgarskata literatura.* Sofija: Damjan Jakov.
- . 2005. "The Apocryphal Series about Abraham." In this volume.
- Olmsted, Hugh. 1994. "Modeling the genealogy of Maksim Grek's collection types: the 'plectogram' as visual aid in reconstruction." In *Medieval Russian Culture, vol. II*, edited by Michael S. Flier and Daniel Rowland, 107–33. *California Slavic Studies*, no. 19. Berkeley: University of California Press.
- Speranskij, M. N. 1963. *Rukopisnye sborniki XVIII veka; materialy dlja istorii rus-skoj literatury XVIII veka.* Predisl., podgotovka k pečati, red. i primečanija V. D. Kuz'min. Moskva: Izd-vo Akademii nauk SSSR.

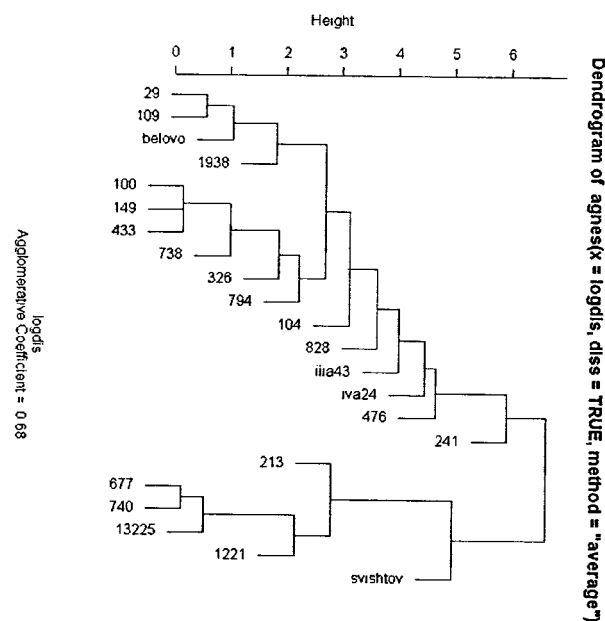
Dept. of Slavic Language & Literatures  
University of Pittsburgh  
1417 Cathedral of Learning  
Pittsburgh, PA 15260 USA  
djbpitt@pitt.edu

## Appendix A: Manuscripts

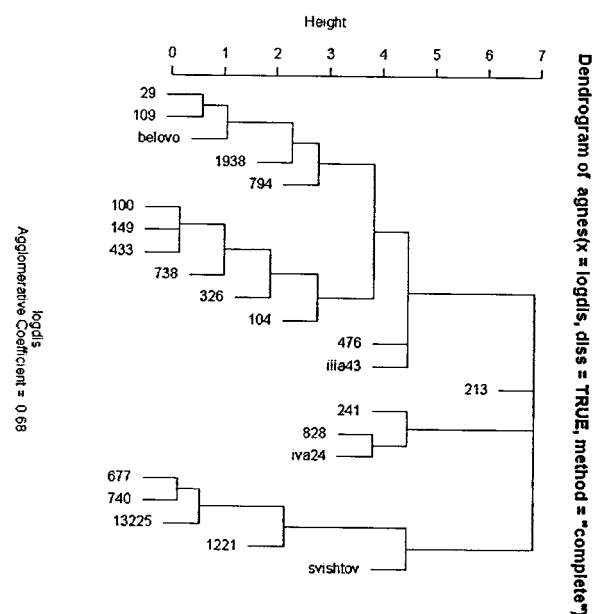
No. 104	National Library of Serbia, Belgrade (third quarter of the 14th century)
No. 29	Savina monastery, Montenegro (c. 1380)
No. 828	National Library of Serbia, Belgrade (1409)
No. 109	St. Paul monastery (15th century)
No. 677	Cyril and Methodius National Library, Sofia (Tikveš, end of the 15th century)
No. 740	National State Archive, Bucharest (16th century)
No. 13.2.25	Library of Russian Academy of Sciences, St. Petersburg (Moldavian origin, 16th century)
No. 149	National Library, Vienna (16th century)
No. 433	Cyril and Methodius National Library, Sofia (16th century)
No. 794	Russian State Library, Moscow (F. 304, 16th century)
No. 738	National Library of Serbia, Belgrade (17th century)
No. III.a.10	Library of the Croatian Academy of Sciences and Arts, Zagreb (16th century)
No. III.a.43	Library of the Croatian Academy of Sciences and Arts, Zagreb (16th century)
No. 241	Russian State Library, Moscow
No. IV.a.24	Library of the Croatian Academy of Sciences and Arts, Zagreb (1520)
No. 326	Cyril and Methodius National Library, Sofia (17th century)
No. 198 (1.221)	National Library of Serbia, Belgrade (1787)
No. 213	Petruševich collection, State Library, L'viv (18th century)
No. 1938	Russian National LibraryH St. Petersburg (17th century)
No. 476	Franko collection Institute of Literature, Kiev (written by Father Jaremeč'kyj-Bilahevič, 18th century)
Belovo	Belovo miscellany (17th–18th century)
Svištov	Amateur Art Club, Svištov (Bulgaria) (Svištov miscellany, 18th century), No. 145a

Note: The manuscripts from the National Library of Serbia (Nos. 104, 828, 738, and 198) were destroyed along with the library during the Second World War. Because they could not now be consulted directly in connection with this research, information about them is derived from earlier descriptions.

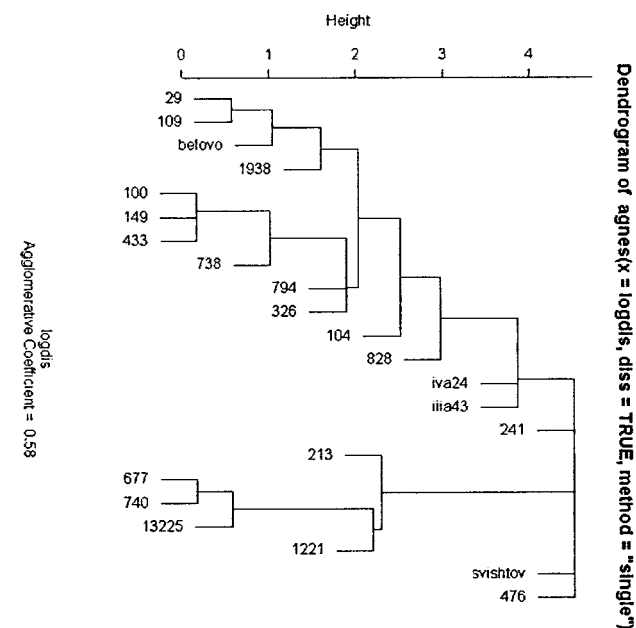
## Appendix B: Dendrograms



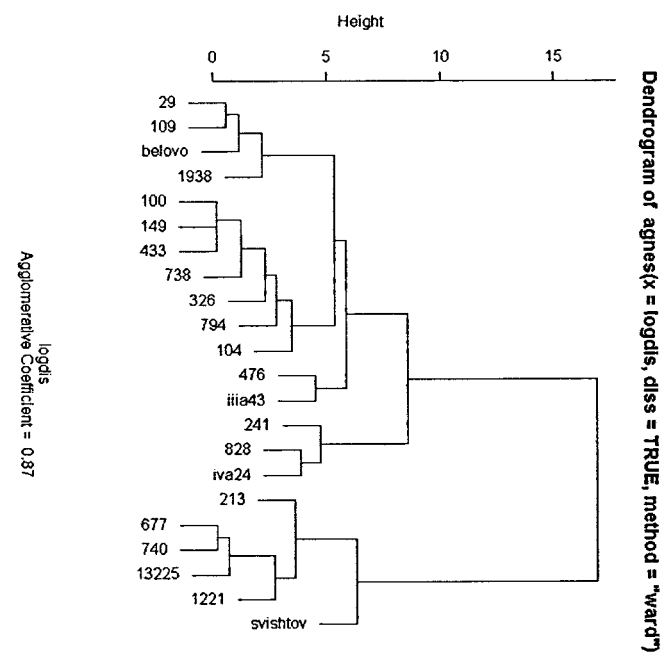
Agglomerative Average Linkage (no length normalization)



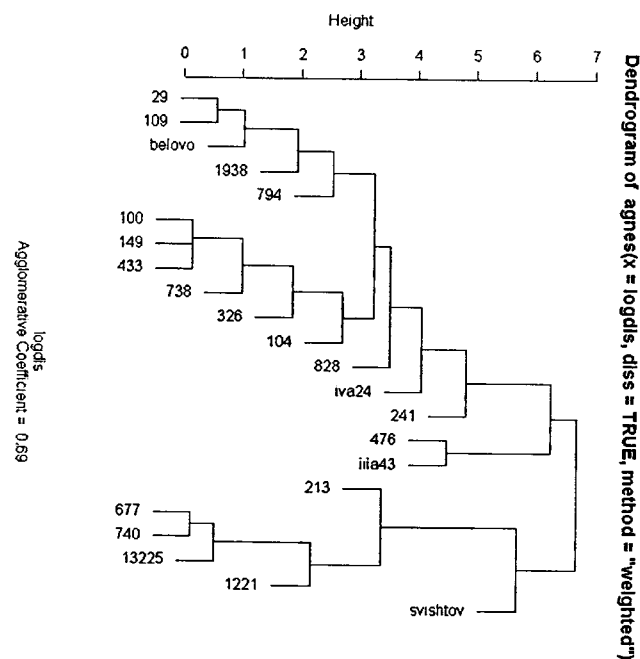
Agglomerative Complete linkage (no length normalization)



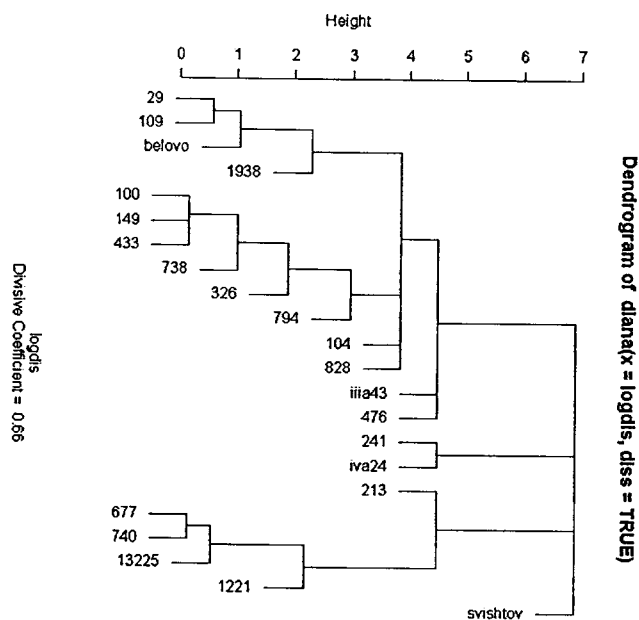
Agglomerative Single Linkage (no length normalization)



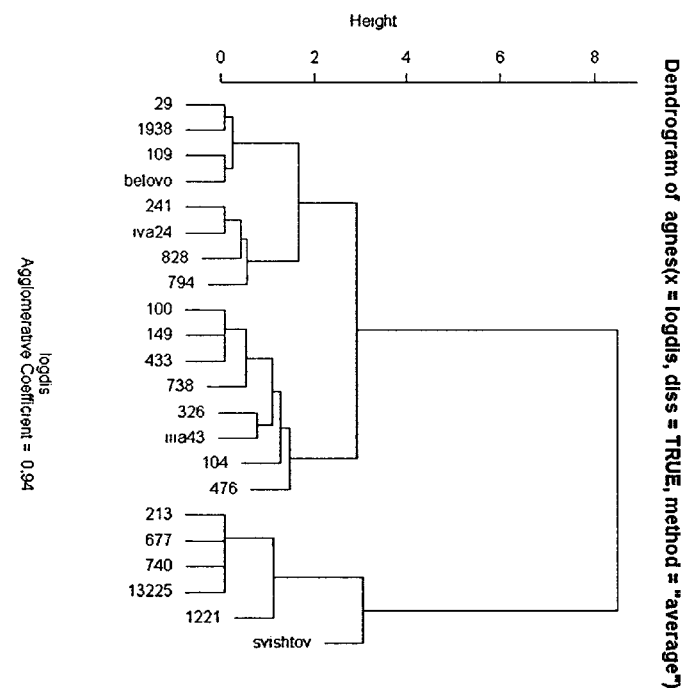
Agglomerative Ward's Linkage (no length normalization)



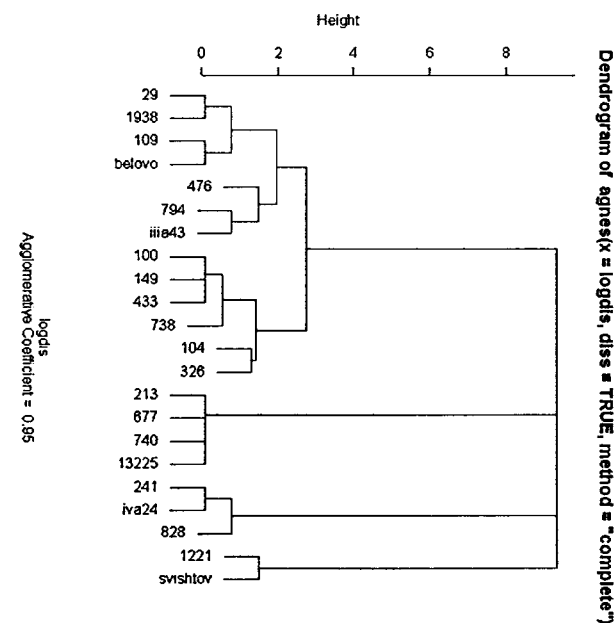
Agglomerative Weighted Linkage (no length normalization)



Divisive Linkage (no length normalization)

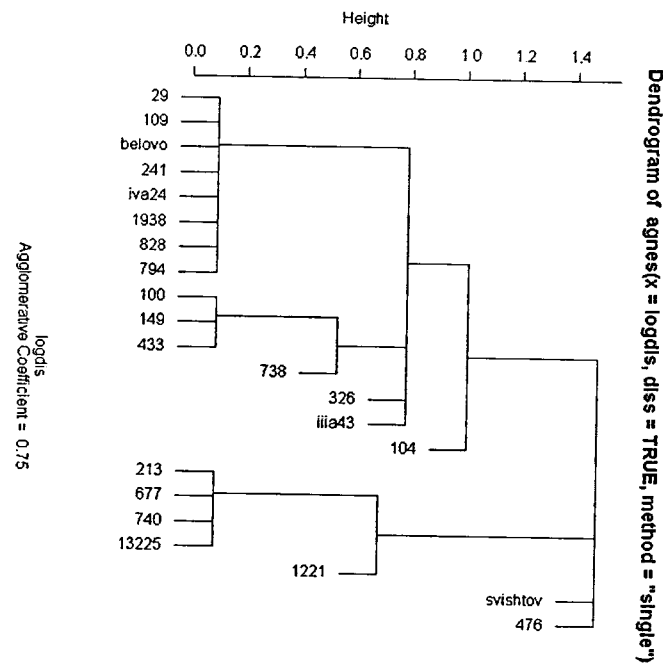


Agglomerative Average Linkage (with length normalization)

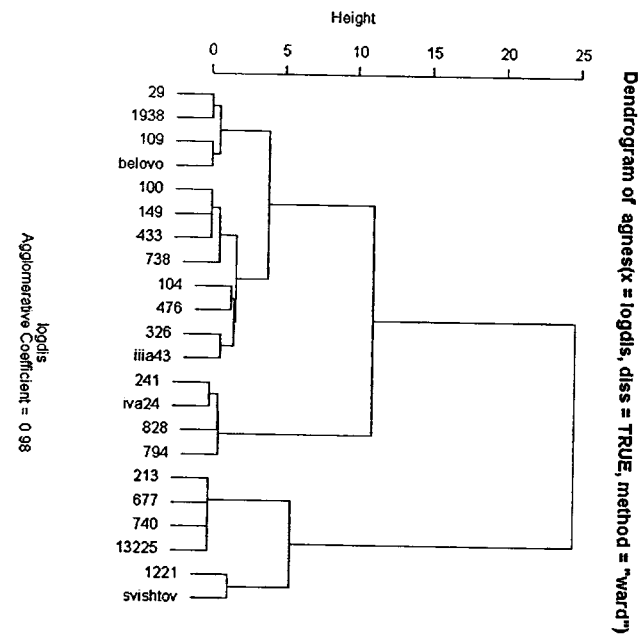


Agglomerative Complete Linkage (with length normalization)

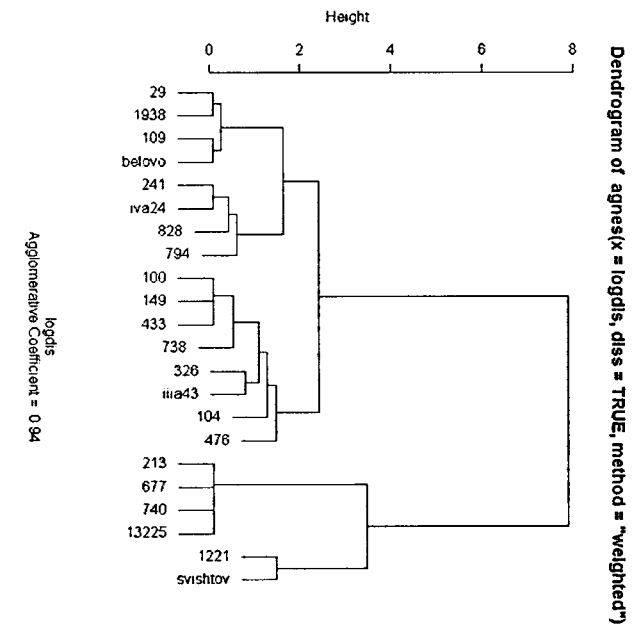




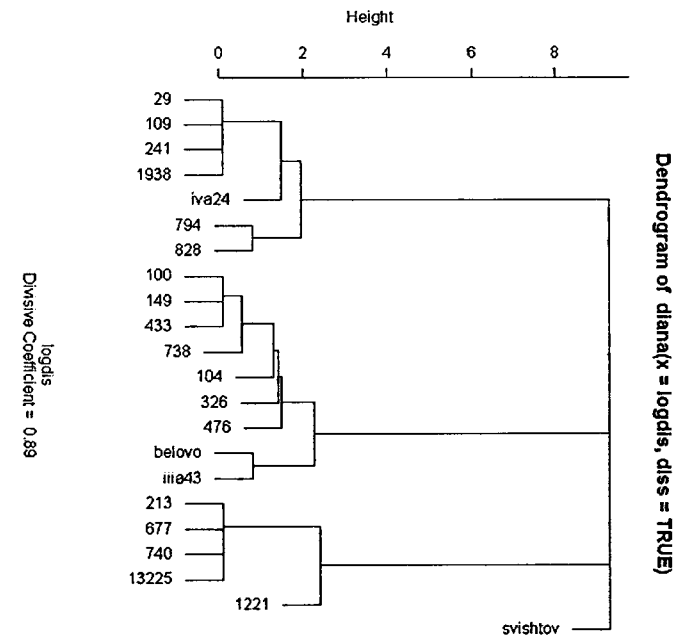
Agglomerative Single Linkage (with length normalization)



Agglomerative Ward's Linkage (with length normalization)

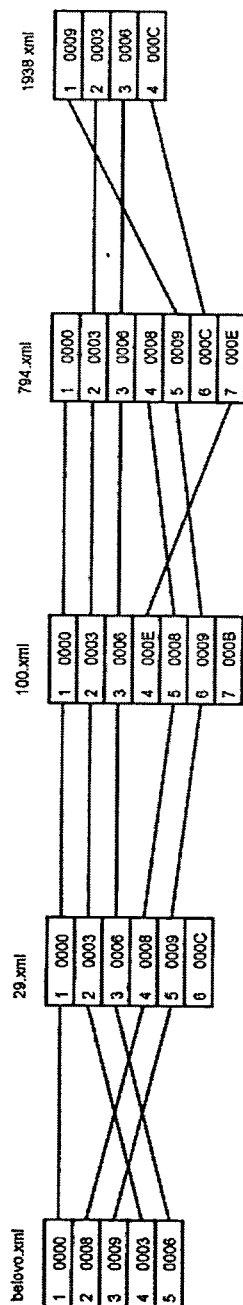


Agglomerative Weighted Linkage (with length normalization)

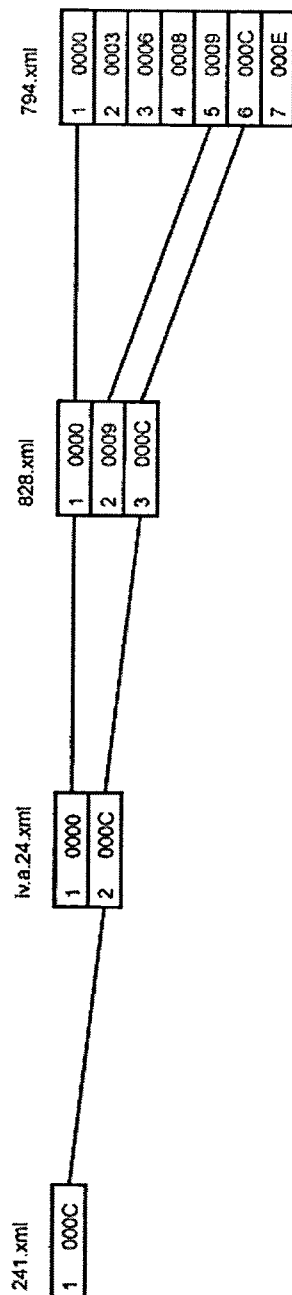


Divisive Linkage (with length normalization)

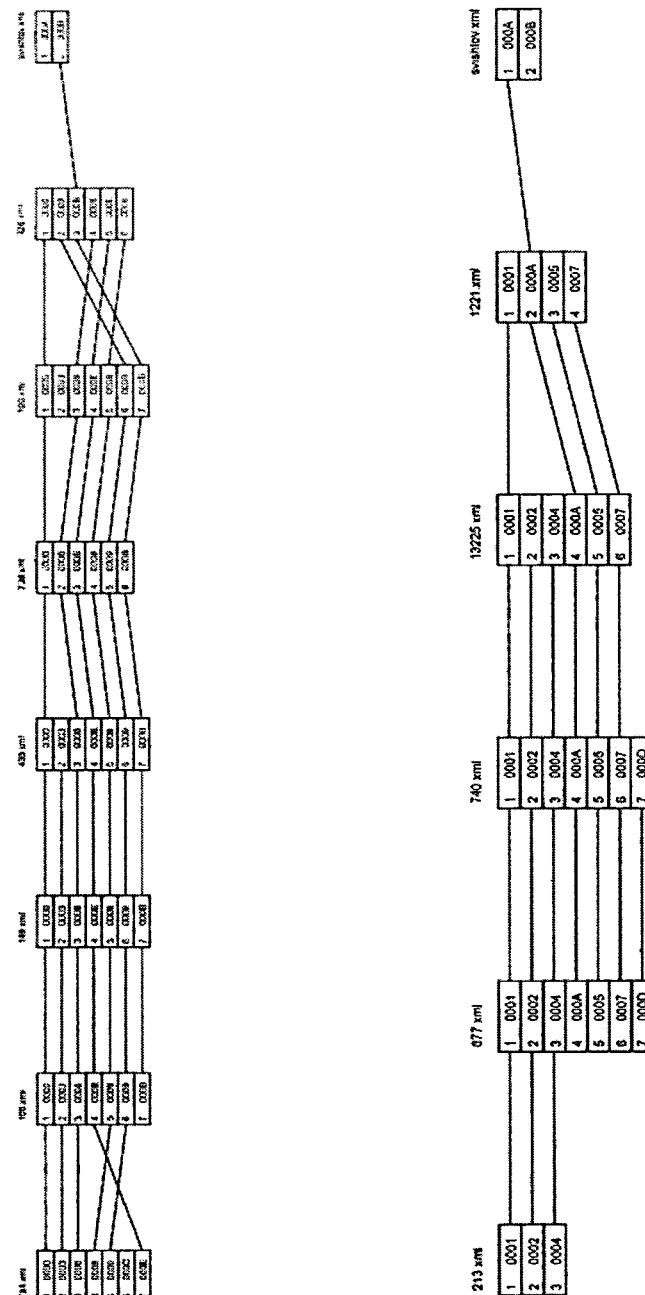
## Appendix C: Plectograms



Archaic Group (1) (with 794)



New Group (2) (with 794)

Bulgarian and Mixed Balkan  
Group (3) (with Svišov and 794)Wallachian-Moldavian Group (4)  
(with Svišov)

Key to Reference Numbers (arabic numerals after titles refer to redactions)

0000	Story of Abraham 1
0001	Story of Abraham 2
0002	Story of Sarah
0003	How Sarah Advised Abraham 1
0004	How Sarah Advised Abraham 2
0005	Ishmael
0006	Isaac 1
0007	Isaac 2
0008	Melchizedek
0009	Abraham and the Holy Trinity 1
000A	Abraham and the Holy Trinity 2
000B	Death of Abraham 1
000C	Death of Abraham 2
000D	Death of Abraham 3
000E	Samuel

## The Journal *Folia Slavica* and Charles E. Gribble as Its Angel

Catherine V. Chvany

In 1977, when Prof. Charles Gribble and a group of American Slavists published the first number of *Folia Slavica*, surely no one thought that Slavica Publishers' new journal would become one of the most valuable and desirable Slavic publications.

Wacław Osadnik (1983)<sup>1</sup>

*Folia Slavica* (1977–87) played an important role during lean years for American and international Slavistics. As inflation and exchange rates forced prices of European publications beyond the budgets of many libraries, let alone individuals, the *International Journal of Slavic Linguistics and Poetics* had ceased publication, while the U.S.-based journals of our field featured almost nothing in linguistics or poetics.<sup>2</sup> Launched with the generous support of Slavica's president, Charles E. Gribble (and subsidized by Slavica textbook profits), *Folia Slavica* more than filled the hiatus in *IJSLP* until the latter was resurrected—also at Slavica—in 1981. Initially, *Folia Slavica* was similar to *IJSLP* in its Jakobsonian focus on linguistics-philology-poetics and in its willingness to

<sup>1</sup> The epigraph is from the initial paragraph of Osadnik's review of the first three volumes of *FS* in *Sâpostavitelno ezikoznanie* (*SE*); the present English translation from Bulgarian is by Ernest A. Scatton.

<sup>2</sup> According to surveys in the 30th anniversary issue of the *Slavic and East European Journal* (*SEEJ* 31 [1987]), *SEEJ* contributions to linguistics and even to pedagogical applications were rare, except for reviews. Major exceptions were the two special issues devoted to selections from the 1974 SACRL (the first Soviet-American Conference on the Russian Language), *SEEJ* 19.1 and 19.2 (1975), guest-edited by Richard D. Brecht and Dan E. Davidson. Several more SACRL I papers appeared in the *Russian Language Journal* during the same span. Selected papers from 1981's SACRL II were published in *RLJ* XXXVI, No. 125 (1982). Otherwise, *RLJ* too showed a drop in language-linguistics contributions while gaining strength in literary and archival scholarship. As for the *AAASS Slavic Review*, its focus is mainly on social sciences, secondarily on literature, with only rare attention to language or linguistics in its book review section.