Lexomics for Comparison
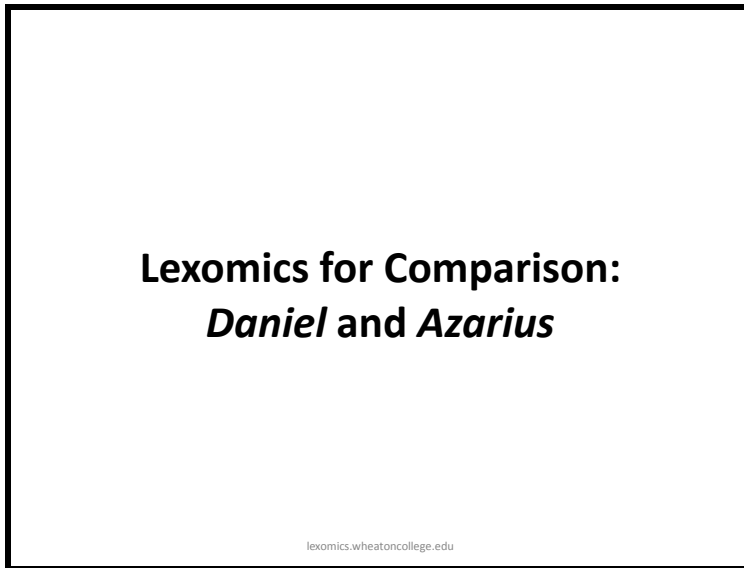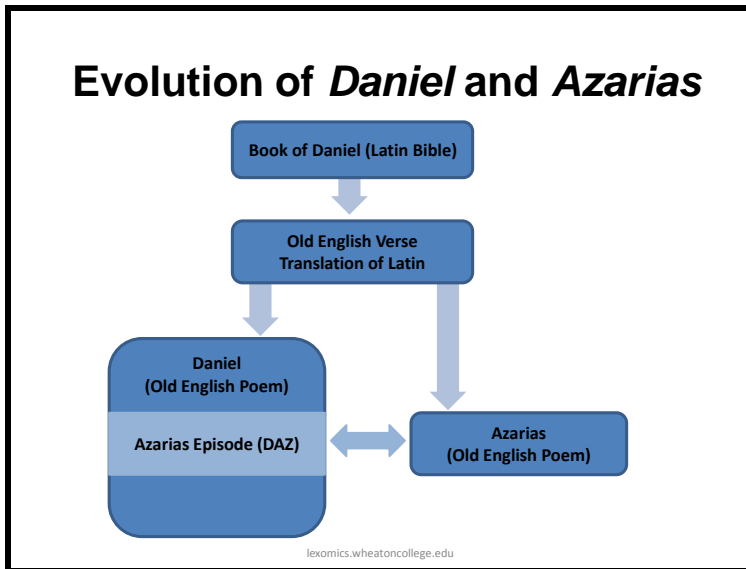
Daniel **and** *Azarias*

(Slide 1)

As our analysis of *Daniel* showed, Lexomics is a means of identifying shifts in vocabulary distribution. In *Daniel*, this difference in vocabulary is evidence of the poet's use of external sources. We can also use vocabulary analysis to aid in identifying similarities in vocabulary among multiple texts.

(Slide 2)

The Anglo-Saxon poem *Azarias*, copied into the Exeter Book in the mid to late 900s, recounts the story from the biblical Book of Daniel of the three youths in the fiery furnace. Because of its previously known similarities to a section of *Daniel*, *Azarias* serves as a "control" for checking the accuracy of our methods. Scholars hypothesize that both of the Anglo-Saxon poems share a common antecedent.

Textual Similarities

Daniel (1. 279):   Ða Azarias ingeþancum
Azarius (1. 1):   Him þa Azarias ingeþoncum

Daniel (1. 280):   hleoðrade halig þurh hatne lig
Azarias (1. 2):   hleoþrede halig þurh hatne lig

Daniel (1. 281):   dreag dæda georn, drihten herede
Azarias (1. 3):   dreag dædum georn, dryhten herede

lexomics.wheatoncollege.edu

(Slide 3)

As you can see from these pairs of lines, *Azarias,* particularly the beginning of the poem, is very similar to one section of *Daniel,* with only minor differences in spelling and morphology separating the two texts.

***Daniel* and *Azarias* (900-word chunks)**

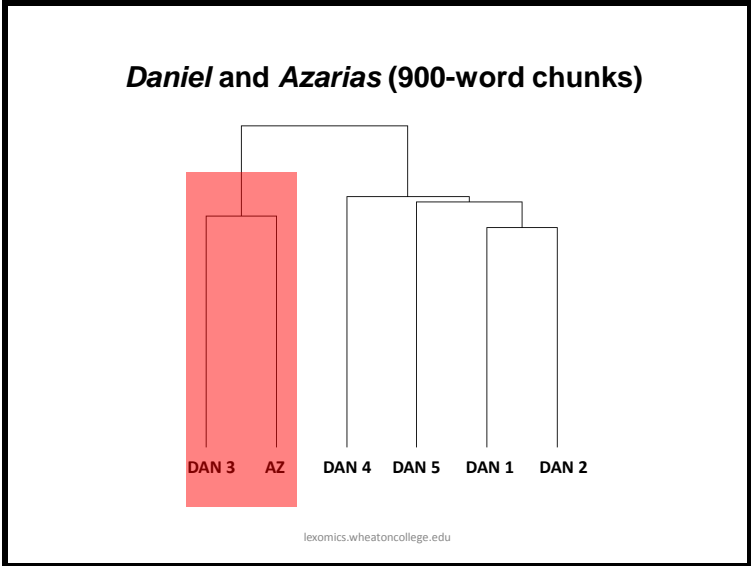DAN 3   AZ   DAN 4   DAN 5   DAN 1   DAN 2
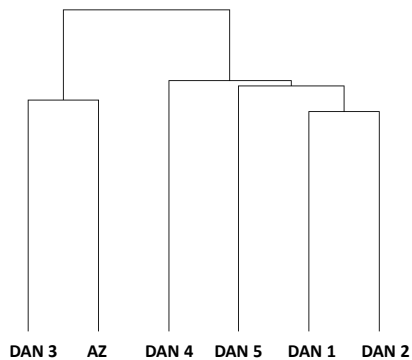
lexomics.wheatoncollege.edu

(Slide 4)

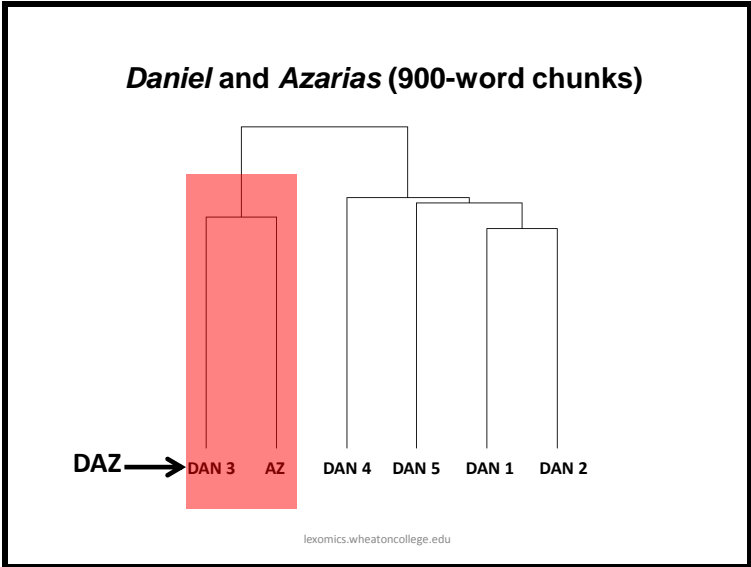When we produce a dendrogram of *Daniel* and *Azarius* together, we see that the section of *Daniel* containing the Azarias episode (labelled "DAZ") clusters together with *Azarias*.

(Slide 5)

(Slide 6)

This clustering demonstrates that Lexomic methods can be used to identify similarity even when there are many minor differences between the texts being compared.
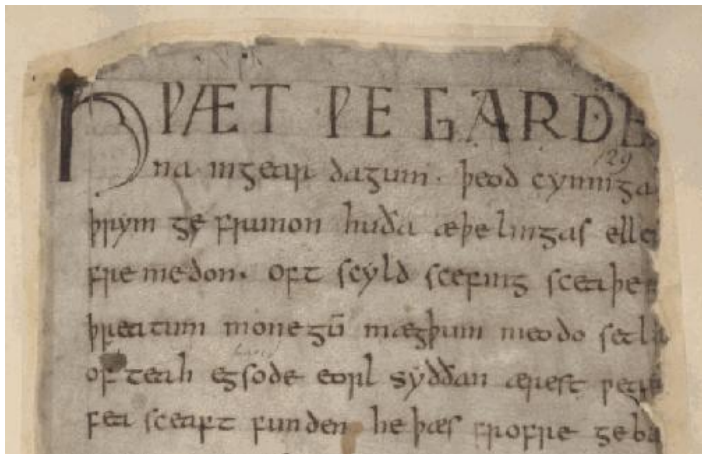
Beowulf **Translations**

Lexomics for Comparison:
Beowulf **Translations**

lexomics.wheatoncollege.edu

(Slide 7)

The DAZ section of *Daniel* and *Azarias* have very similar content and come from the same source, so it is perhaps not surprising that they would cluster together, but we cannot easily determine if the two poems are linked because they have the same content, or because they are stylistically similar. To separate out the influence of different kinds of similarity—in content or in style—we investigated multiple translations of the poem *Beowulf*. These translations are all the same in content, but of very different styles.
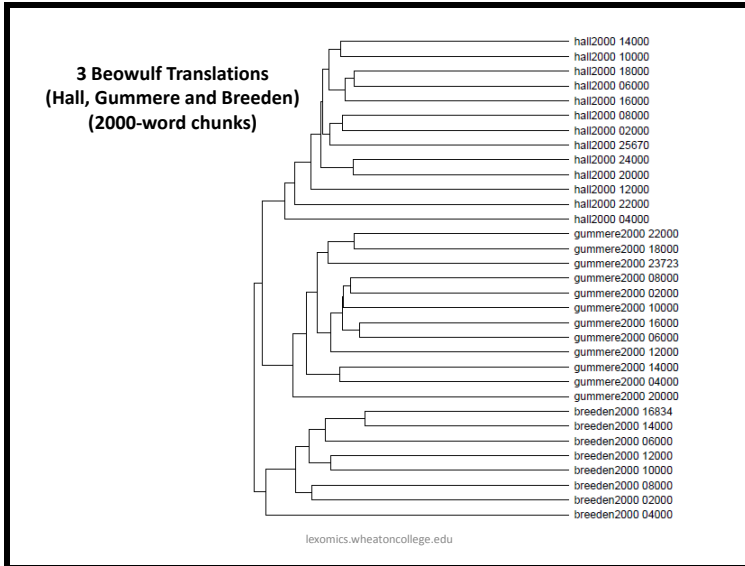

Beowulf

lexomics.wheatoncollege.edu

(Slide 8)

The poem *Beowulf* has been translated many times in the past two centuries. Because all translations are derived from the same original text, they are basically identical in content.

(Slide 9)

This dendrogram contains three different translations of *Beowulf*, by John Lesslie Hall, Francis B. Gummere and David Breeden, cut into 2000 word chunks.
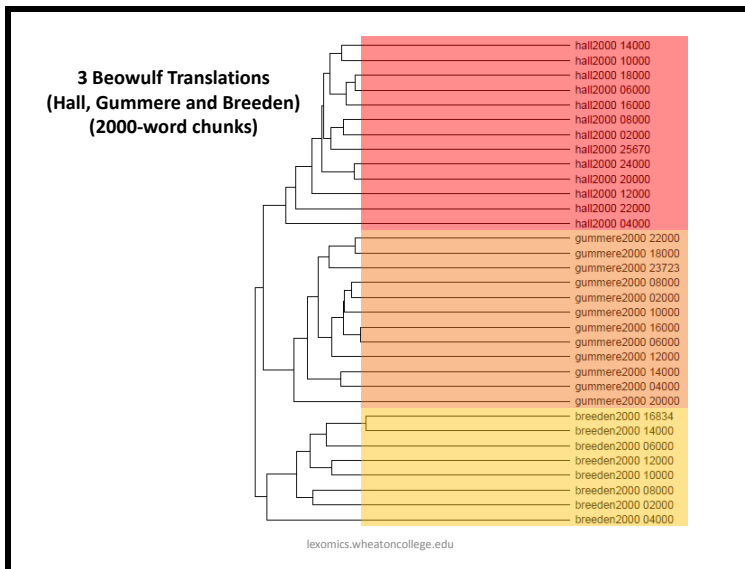
(Slide 10)

Even though each translation tells the same story in the same order from the same source, the dendrogram is organized by translator. Not a single chunk clusters with the wrong translator.

This unexpected result suggests that the vocabulary distributions measured by Lexomic methods are more closely correlated with style than they are with content.

Signed Poems of Cynewulf

Lexomics for Comparison:
Signed Poems of Cynewulf

lexomics.wheatoncollege.edu

(Slide 11)

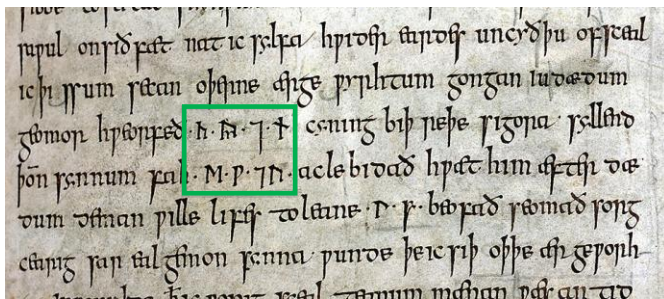Using Lexomic methods, we were able to group translations of *Beowulf* by translator. We next wanted to see if we could group texts by author.

Cynewulf Runic Signature

lexomics.wheatoncollege.edu

(Slide 12)

Because he "signed" his poems by including runes that spell out his name, Cynewulf is one of the only Anglo-Saxon poets whose name we know. There are four signed Cynewulfian poems: *The Fates of the Apostles, Elene, Juliana, and Christ II*. Many scholars think that Cynewulf also wrote *Guthlac B*, because the style of this poem is quite similar to that of the signed poems. The runic signature could have been in the concluding lines of the poem, but these are now lost.

We wanted to see if Lexomic methods could detect the same similarities of style that so many scholars have noted, so we produced a dendrogram of the signed poems of Cynewulf plus the poems *Guthlac A* and

Guthlac B. Although few critics think *Guthlac A* is by Cynewulf, it and *Guthlac B* are about the same Saint, so we included it for comparative purposes.

Figure 11. Dendrogram showing the results of cluster analysis using 22 approximately 1000-word chunks of *Guthlac A* and *B* and the signed poems of Cynewulf (*The Fates of the Apostles, Juliana, Elene* and *Christ II*). The exact word boundaries of each chunk are listed below. The * indicates those chunks that contain a runic signature.
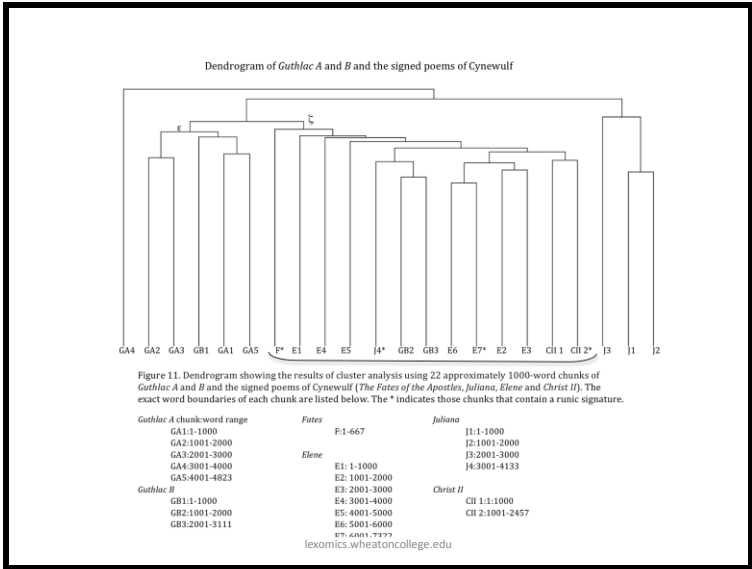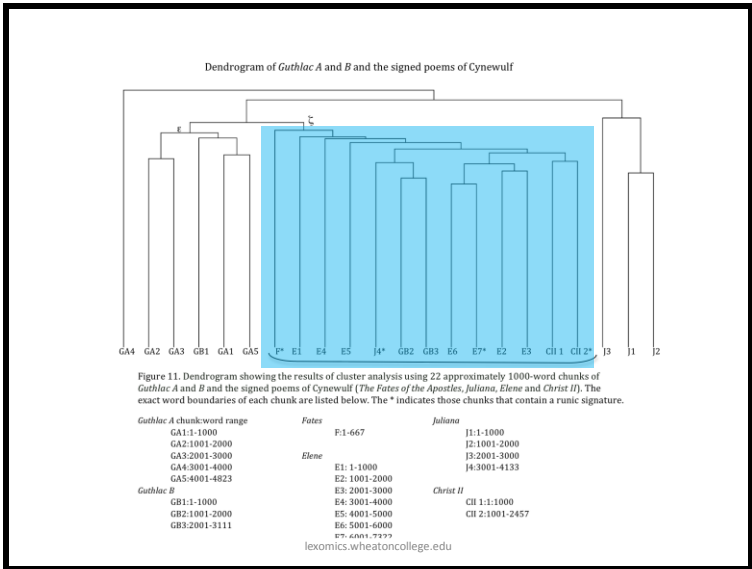
(Slide 13)

This complex dendrogram has four main clades.

Figure 11. Dendrogram showing the results of cluster analysis using 22 approximately 1000-word chunks of *Guthlac A* and *B* and the signed poems of Cynewulf (*The Fates of the Apostles, Juliana, Elene* and *Christ II*). The exact word boundaries of each chunk are listed below. The * indicates those chunks that contain a runic signature.
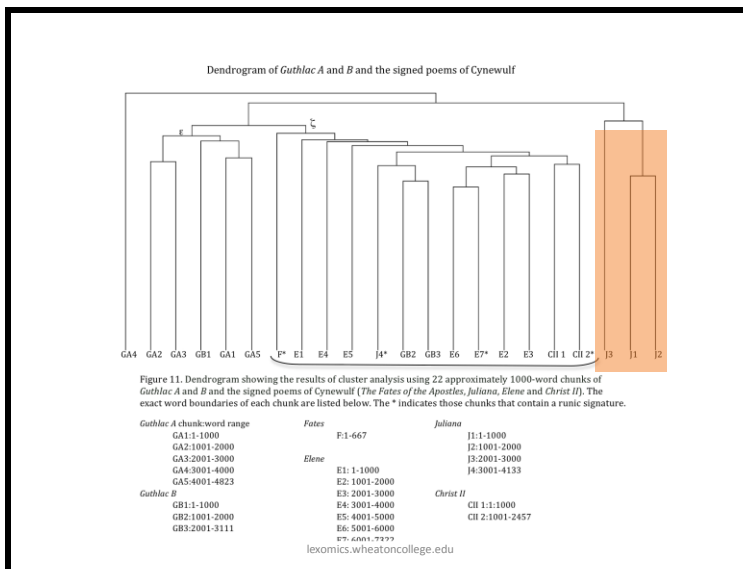
(Slide 14)

All the segments of the central clade, indicated here in blue, are parts of the signed poems of Cynewulf, with the exception of *Guthlac B*, chunks two and three.

Figure 11. Dendrogram showing the results of cluster analysis using 22 approximately 1000-word chunks of *Guthlac A* and *B* and the signed poems of Cynewulf (*The Fates of the Apostles, Juliana, Elene* and *Christ II*). The exact word boundaries of each chunk are listed below. The * indicates those chunks that contain a runic signature.
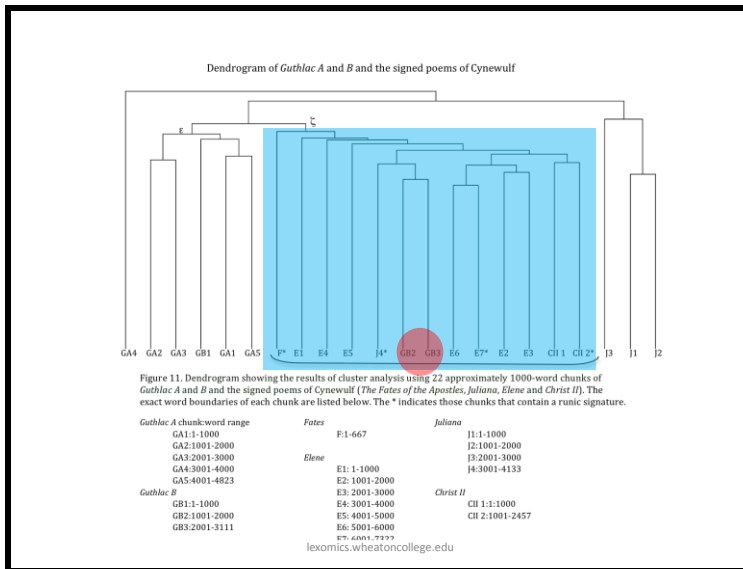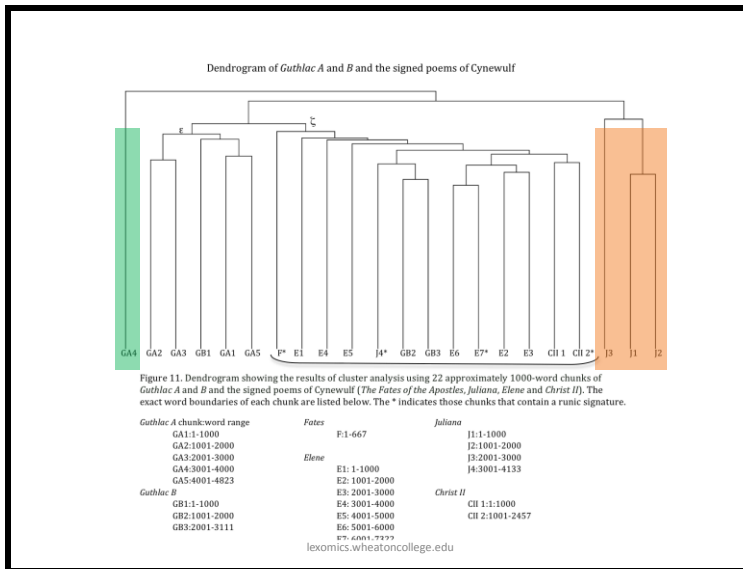
(Slide 15)

This placement strongly suggests that *Guthlac B*—at least the last two thirds of the poem—is by Cynewulf.

The segments that are not in the Cynewulfian central clade include all of those containing *Guthlac A* (which no one thinks is by Cynewulf), but also three sections of *Juliana*, which *is* thought to be by the poet—only the segment containing his runic signature is in the central clade. The rest of the poem is in this outlying clade, marked in orange.

Figure 11. Dendrogram showing the results of cluster analysis using 22 approximately 1000-word chunks of *Guthlac A* and *B* and the signed poems of Cynewulf (*The Fates of the Apostles, Juliana, Elene* and *Christ II*). The exact word boundaries of each chunk are listed below. The * indicates those chunks that contain a runic signature.
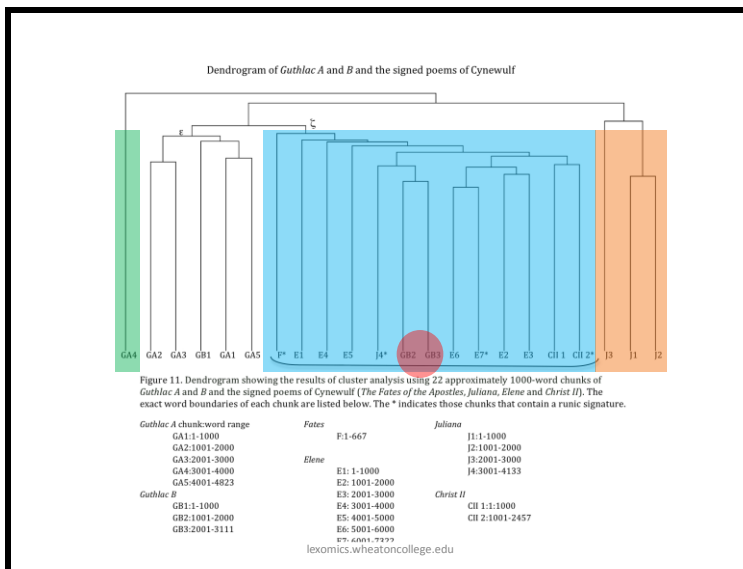
(Slide 16)

However, our previous research has indicated that the presence of multiple sources in a text can affect the shape of its dendrogram. In the case of *Juliana*, we know that the first three sections of the poem follow its Latin source extremely closely and that only the final section, which includes the runic signature, seems to be a composition entirely original to Cynewulf. The "outside" source actually makes up the great majority of the poem, and this is reflected in the dendrogram.

Figure 11. Dendrogram showing the results of cluster analysis using 22 approximately 1000-word chunks of *Guthlac A* and *B* and the signed poems of Cynewulf (*The Fates of the Apostles, Juliana, Elene* and *Christ II*). The exact word boundaries of each chunk are listed below. The * indicates those chunks that contain a runic signature.

(Slide 17)

Guthlac A is distinctly separated from the Cynewulfian material, but one section, *Guthlac A*, segment four (marked here in green) is even separated from the rest of *Guthlac A*. This segment, however, has a known external source: an Old English prose treatment of an episode in St. Guthlac's life where demons drag him to the mouth of hell and threaten to throw him in before they are stopped by St. Bartholomew.

Figure 11. Dendrogram showing the results of cluster analysis using 22 approximately 1000-word chunks of *Guthlac A* and *B* and the signed poems of Cynewulf (*The Fates of the Apostles, Juliana, Elene* and *Christ II*). The exact word boundaries of each chunk are listed below. The * indicates those chunks that contain a runic signature.

(Slide 18)

The complex geometry of this dendrogram can therefore be explained in terms of both authorship and the influence of external sources. We can therefore, use the identification of similarities enabled by the Lexomic methods to see how texts relate to each other, through authorship, sources and style. Lexomics does not replace close reading, philology, historical analysis or other traditional methods, but it *can* provide significant additional information about texts, their structures and relationships.

By: Michael Drout and Leah Smith

lexomics.wheatoncollege.edu