

Analysis of the Text

The edition of *Beowulf* on this website is supplemented by four appendices. Three of them illustrate spacing patterns within words, as these have been abstracted and sorted by computer programs; the fourth illustrates spacing patterns within word groups.

In place of similar appendices, this edition of *Alexander's Letter to Aristotle* provides a sampling of definitions and routines underlying these kinds of analysis. They can at once clarify how the *Beowulf* appendices were created and show how similar analyses can be carried out on this prose text. The fact that all of *Alexander's Letter* and well more than half of *Beowulf* were written by the same person makes their parallels all the more promising for study of poetic meter and linguistic prosody.

For computer-assisted analysis, the text is segmented in these ways.

Page is any string beginning with [and continuing up to the next occurrence of] (or the end of the text). Identification of the page is the folio notation, e.g., [fol. 126v]; see The Text, B(1).

Sentence is any string up to and including >; see The Text, B(2).

Word is any string ending with **blank numeral blank**.

Morpheme is any string ending with either **blank numeral blank** or **hyphen numeral hyphen**; see The Text B(5)(b). This will identify root and pre-base morphemes, but not inflectional morphemes; spacing does not occur at the onset of suffixed inflectional or most derivational morphemes.

Morph, when not a word or a morpheme, is any string terminated by **numeral** (and not hyphen or blank); see The Text, B(5)(b).

So, for example, to study the graphotactics of compound words, the procedure is this. (1) Select a domain—the whole text, or certain pages or sentences. (2) Scan for **word**. (3) Scan **word** for **morpheme**. (4) Discard any pre-base forms. (5) Re-scan **word** for **morpheme**. Then compile the list of words that have morpheme divisions within them.

With this list, proceed then to sort for any factors or set of factors that are of interest. The interest may be merely the range of measure for internal spacings. Of greater interest will be correlations of spacing and other features, whether the first element of a compound word is monosyllable or disyllabic, for example. Still further, if the first morpheme is monosyllable, is the syllable short or long? What is the phonological class of the terminal consonant(s)? And so on. All these kinds of correlations showed distinctive patterns in the *Beowulf* text.

For another example, to study the graphic clustering—the graphotactics—of connectives, set the domain as **sentence** and scan for the forms of interest, such as **ond** and its abbreviation 7 or **mid þȳ** or conjunctive **þā**; also check for the occurrences of relative **þæt** and its abbreviation **þ**.

Or in a restricted sense of connectives, scan **sentence** for various (stipulated) prepositions, such as **mid** or **on** or **tō** in their following context.

Or to test the scribal rules of mandatory word division at the end of a line of writing, compile a list of words and scan them for spacing notation **9** (represented by / in the printed text). Then proceed to sort for phonological features on either side of the division. Some of the data then pose such problems in phonology as syllable division involving the sequence **-st-**. Although initial consonant clusters **st-** are well established in Old English, what is the basis for the divisions **epis-tolan**, **fæs-tenu**, **swus-trum**, **magis-tre**, **wēs-tan**, **dyrs-tignesse**?

Relevant to more controversial questions of phonology may be line-end divisions such as **næd-rena** and **nic-res**, where the questions concern underlying forms and consonant lengthening.

A scan of separate words for either **īo** or **ēo** will assemble an array that includes **trīow-** and **trēow-** twice each on 127v and several more instances of both. With those will be found both **līod-** and **lēod-**. It will also pick up the specially interesting collocation 120r.2 **sēo fyrd 7 sī hīow**.

Or abstract a selection of pre-base morphemes by scanning a domain first for **word** and then scanning for **on-**, **be-**, **a-**, **wið-** and such. A separate—or comparative—analysis of **ge-** would proceed similarly.

Correlatives—typically **þā ... þā ...**—can be abstracted from sentence domains by scanning **sentence** for **þā** and then scanning the subsequent segment for the same form again. (As usual, the search would need to list **þā** and **ðā** as alternates.)

Beyond these kinds of analyses are those at phrasal level, such as the two described earlier, reflexive **wē ūs ge-restan** and the (quasi-) phrasal verb **on be-cwōm-**, in the last segment of ‘The Measure of Spacing.’

Of course there are potential analyses not limited to abstracting and sorting. A particularly interesting project is to adapt the text with its graphotactic notation for interpretation by a computer program that reads aloud (so to speak) the written text. The alphabetic string will require a system of transliteration for data the computer program is designed to process. The graphotactic notation will require a system of transliteration (transnumeration?) as well. For a start, relative length of spacing can be reprogrammed as relative length in the timing between segments. (Uncertain spacing ‘8’ (‘?’ in printed text) and line-end ‘9’ (‘/’ in printed text) would require special marking.) All this is straightforward programming. A program to correlate pitch patterns with syntactic and graphotactic patterns would be more difficult and tentative.