

The lexical space of Old English poetry

Giuseppe Brunetti

Abbreviations

texts

and	=	Andreas
beo	=	Beowulf
best	=	Bestiary
dream	=	Dream of the Rood
eleg	=	Elegies
elen	=	Elene
genA	=	Genesis A
genB	=	Genesis B
gnomic	=	Gnomic poetry
guthA	=	Guthlac A
heroic	=	Minor heroic poems
jud	=	Judith
mald	=	Battle of Maldon
phoen	=	Phoenix
rid	=	Riddles

word classes

n	=	nouns
v	=	verbs
a	=	adjectives
av	=	adverbs
np	=	proper nouns
p	=	pronouns
pp	=	prepositions
c	=	conjunctions
d	=	determiners
nu	=	numerals
e	=	exclamations
cp	=	compounds
po	=	poetic words

Lexical style

Lexical style is here meant to be the distribution of word classes in a text, with the addition of compounds and poetic words (those used only in poetry). Forms are taken into consideration, that is the occurrences of lemmas.

Texts of very different lengths (*Maldon* 2074, *Phoenix* 3704, *Elene* 7284, e *Beowulf* 17219 forms) cannot be directly comparable through percentages: comparison has to be through a norm or average – that of the corpus to which they belong.

The corpus

The counts come from over a third of the entire Old English poetic corpus: 13,044 lines = 73,548 forms (= 7,548 lemmas). Collections have been made of the elegies, the minor heroic poems, the gnomic poems and the bestiary, so that the shortest text is at least 1000 words.

Theoretical and actual frequency

The statistic used is based on the difference between theoretical and actual frequency. *Beowulf*, for instance, with its 17,219 forms is little less than a fourth of the corpus, so the theoretical frequency of its nouns is little less than a fourth of the nouns of the corpus: the difference between actual and theoretical frequency shows whether the text has an excess or a deficiency of nouns – whether it is above, below or within the norm.

Z-scores

This difference is give in z-scores, which directly show whether it is significant or not: it is if it exceeds ± 2 (two standard deviations).

Tab. 1 Z-scores

	n	v	a	av	np	p	pp	c	d	nu	e	cp	po
and	2.50	-1.27	1.06	-2.04	-4.61	0.37	3.18	-0.86	-2.66	-1.25	0.34	7.17	5.08
beo	0.96	2.27	0.73	1.43	13.61	-3.90	-5.41	-1.67	-3.27	-3.67	-0.72	13.68	13.73
best	0.75	-2.58	3.47	0.10	-3.69	-3.05	1.89	-0.13	3.56	-0.62	-0.71	2.54	0.77
dream	-2.51	1.02	-0.07	1.09	-4.15	3.20	1.85	-1.78	1.04	-1.49	1.72	-3.89	-5.18
eleg	-1.25	3.25	0.60	2.09	-7.50	2.33	-2.45	-0.32	-1.07	-3.57	2.58	1.21	-0.55
elen	2.79	-2.08	-1.82	-0.18	-3.23	-5.84	6.34	-2.63	5.80	1.01	1.94	0.49	0.43
genA	3.34	-3.61	-3.35	-1.77	8.68	0.58	-2.08	1.54	-3.03	9.71	-1.74	0.54	-1.35
genB	-7.44	1.88	-5.16	1.25	-6.85	14.32	0.00	1.45	2.20	-0.94	1.54	-10.71	-11.16
gnomic	3.12	3.32	4.76	-3.01	-9.06	-1.22	-1.20	-1.70	-5.60	-0.79	-1.59	-2.15	-1.03
guthA	-1.63	1.10	-2.80	-2.56	-5.00	6.07	-0.53	1.55	3.07	-1.94	-0.52	-7.53	-10.70
heroic	-5.29	-2.98	-1.96	-0.09	31.71	-2.41	2.59	2.64	-2.03	-0.40	-0.07	-1.08	-1.25
jud	-1.36	-2.27	5.00	2.29	-2.98	-2.28	0.11	-1.51	6.03	-1.94	-1.10	3.46	5.81
mald	-4.74	3.44	-4.04	2.01	2.43	1.90	1.40	0.09	2.01	-0.73	-0.19	-6.04	1.87
phoe	5.75	-4.29	4.91	-0.19	-8.83	-9.76	2.48	2.41	3.42	-2.73	-1.50	-1.82	-2.22
rid	-1.88	2.91	0.23	-0.17	-9.46	4.59	-1.54	0.98	-2.18	4.08	-1.55	-5.44	-1.78

Lexical space

The overall differences can be easily visualized. The word classes (plus compounds and poetic words) can be thought of as the coordinates of a multi-dimensional space in which the texts are embedded: the statistic of *multi-dimensional scaling* reduces the dimensions to two and graphs them.

Two complementary graphs represent the distances of the texts and the classes between each other and from the norm. The norm is the zero point of the axes – the hypothetical text whose z-scores equal 0.

Fig. 1 Lexical space: the texts

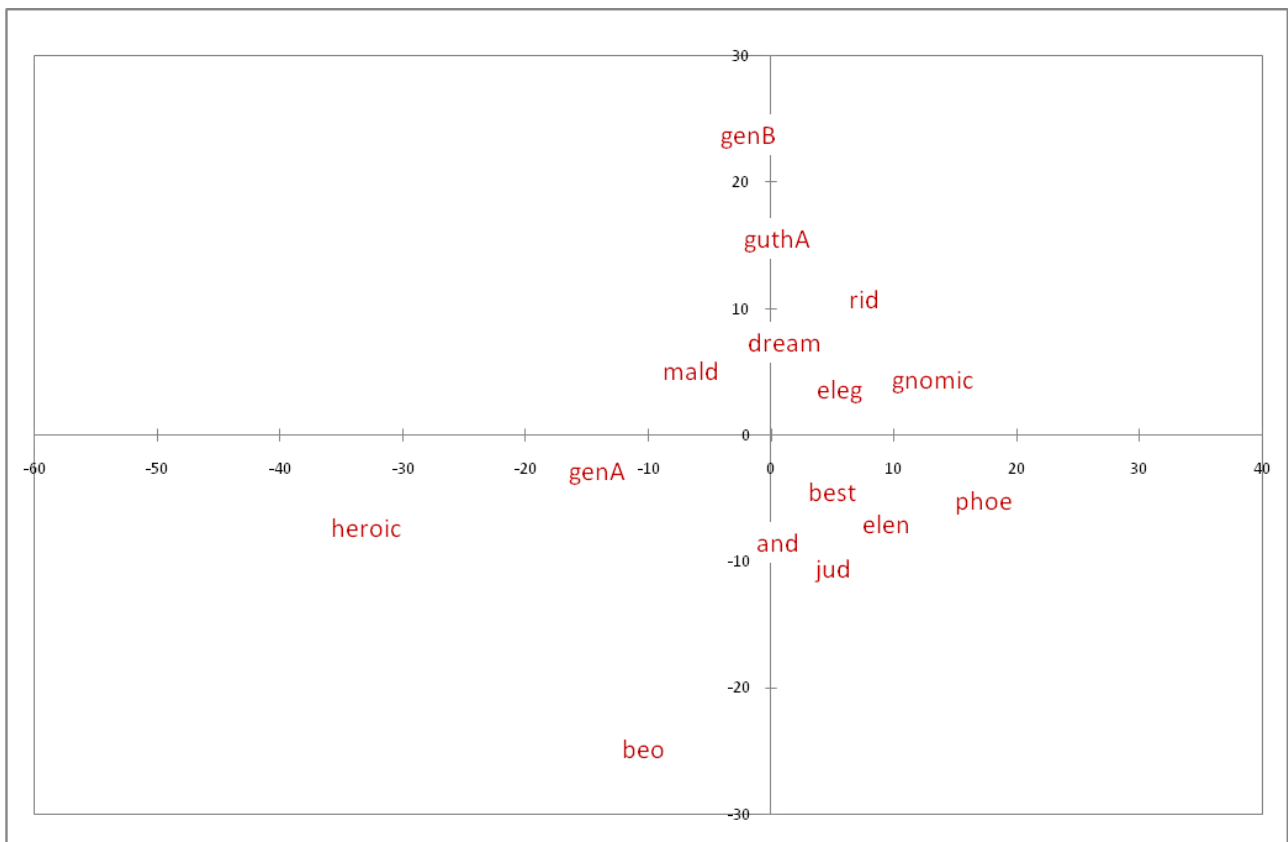
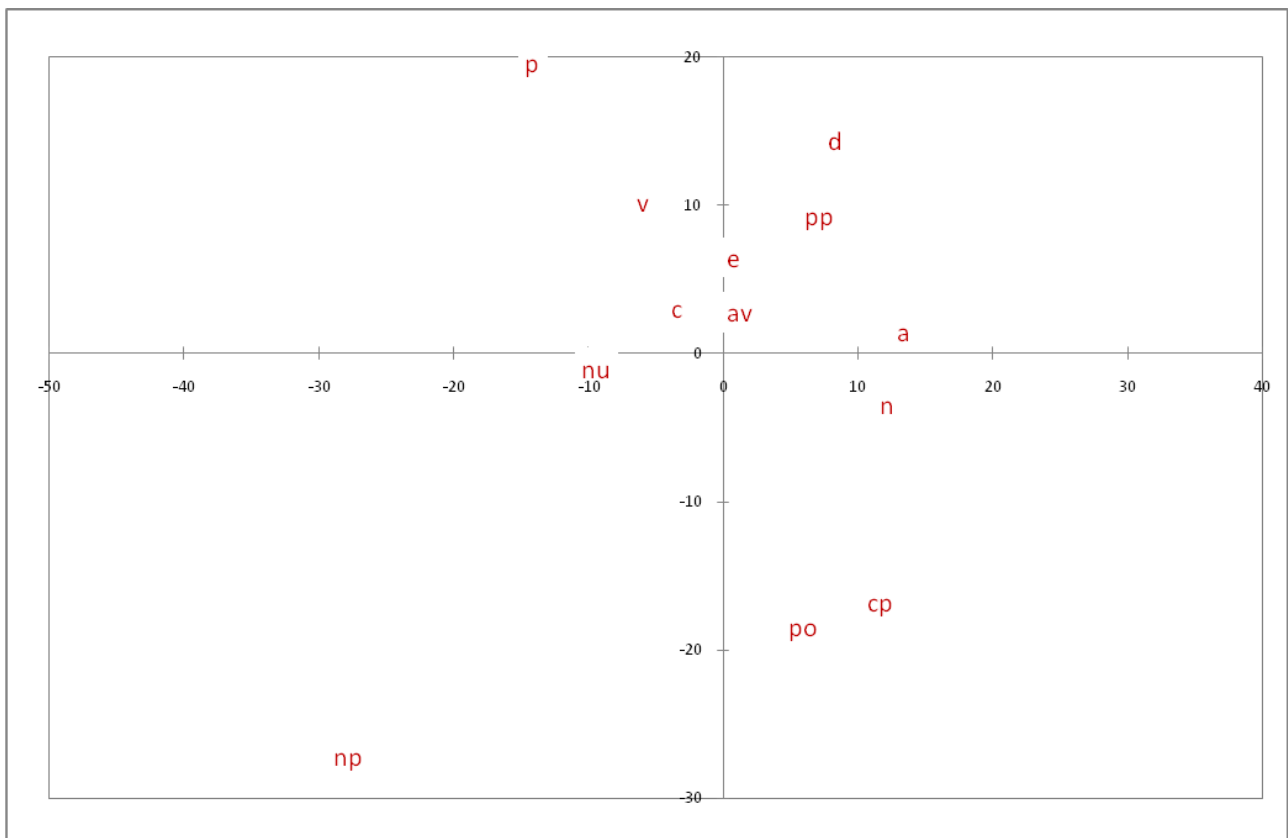


Fig. 2 Lexical space: the classes



The classes extend, vertically, from pronouns, determiners, verbs and prepositions to nouns, compounds, poetic words and proper nouns. What varies along this axis – from the minimum of *Genesis B* to the maximum of *Beowulf* – is nominality, also qualitative: three fourths of compounds and poetic words are nouns.

Beowulf, half-way between proper nouns and poetic words and compounds, has a marked scarcity of functional words (it is the only text, with the gnomic poetry, in which pronouns, prepositions, conjunctions and exclamations are all below the norm). Functional words, on the contrary, abound in *Genesis B* – and pronouns disproportionately.

The Battle of Maldon (mald), which has one of the highest shortages of nouns and the highest excess of verbs, is stationed near the verbs.

The Phoenix (phoe), on the other hand, with the highest excess of nouns and the highest shortage of verbs, is in the noun area.

Comparison between texts

It is possible to compare texts between them. Here are the two *Genesis* poems (the ± 2 band is the norm):

Fig. 3 *Genesis A*

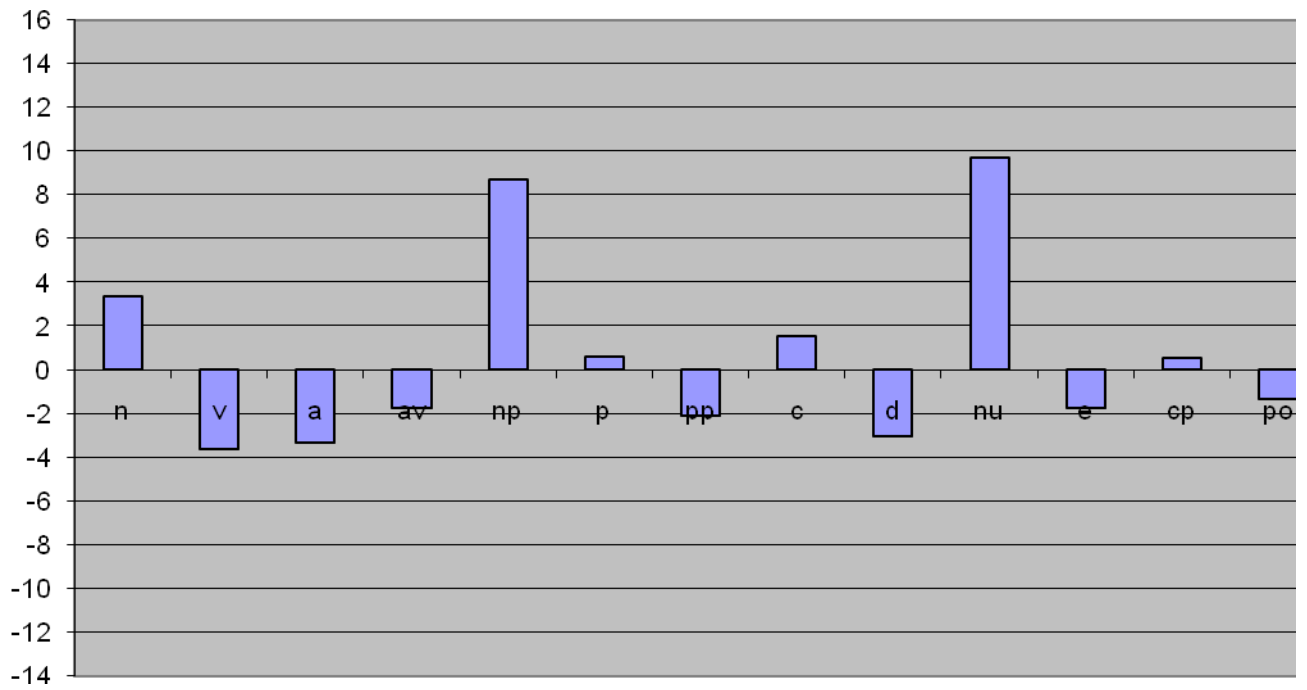
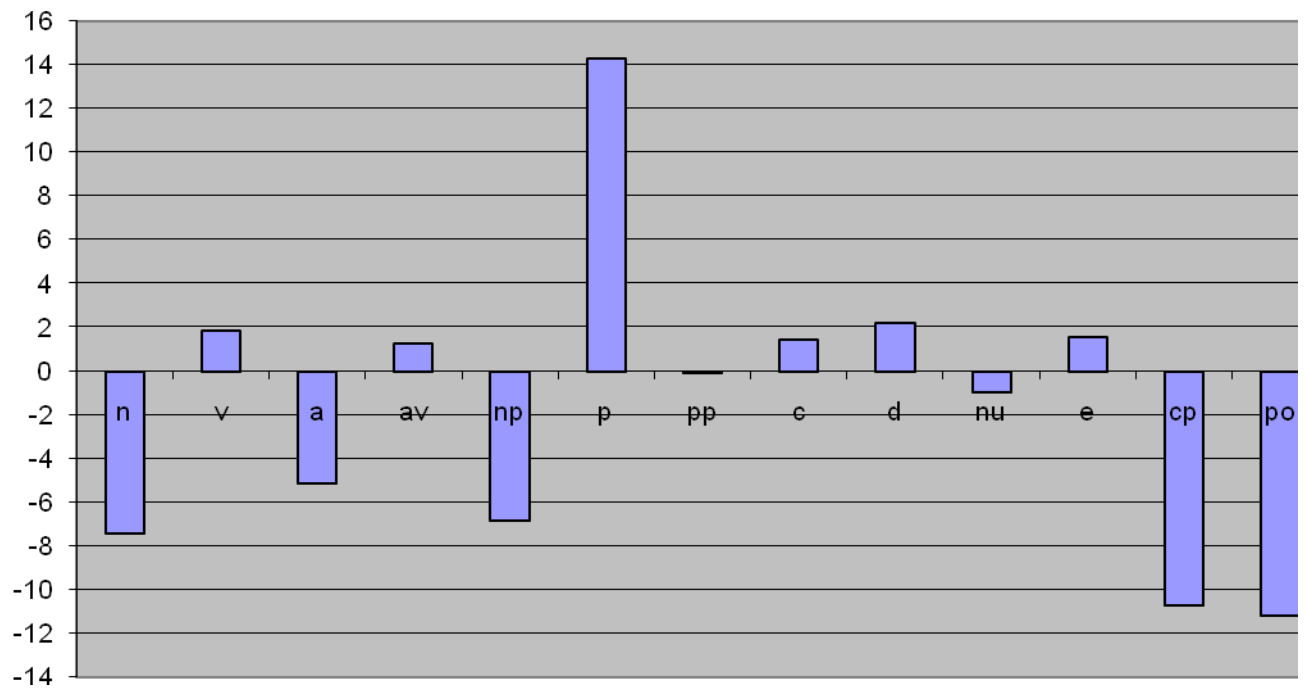
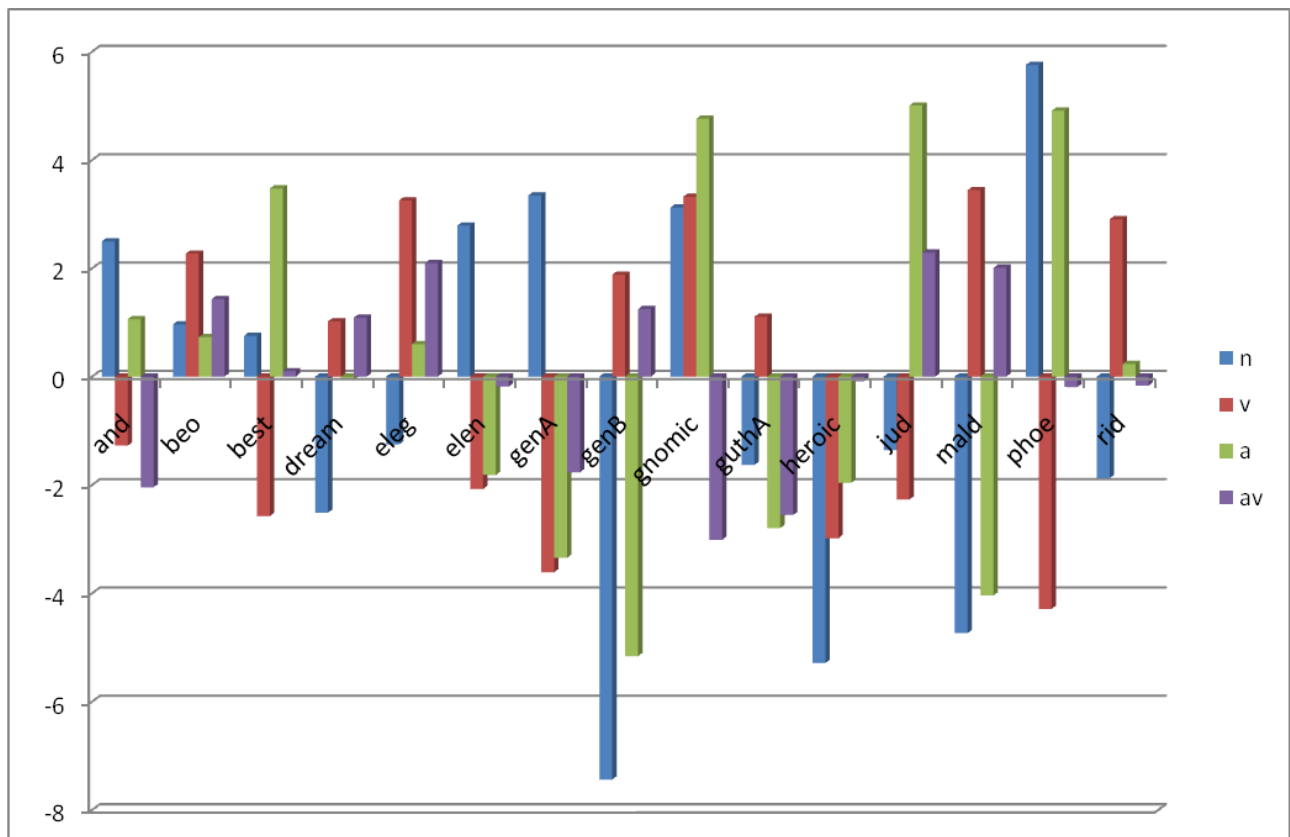


Fig. 4 *Genesis B*

Nominal or verbal style

And all the texts can be compared in relation to the crucial classes (the ± 2 band is the norm).

Fig. 5 Nouns, verbs, adjectives and adverbs



With a marked excess of nouns and adjectives, *Phoenix* (*phoe*) has a nominal style.

A verbal style characterizes *The Battle of Maldon* (*mald*), whose verbs and adverbs on one side and nouns and adjectives on the other neatly straddle the norm.

Beowulf is within the norm, with a small excess of verbs.

Judith abounds in adjectives, and the gnostic poetry abounds in nouns, adjectives and verbs.

Genesis B is strongly denominalized.

Characteristic lexicon

Modal verbs can be part of the characteristic lexicon of a text. Gnostic **sculan** dominates gnostic poetry, followed at a distance by **cunnan**, which is the characteristic modal of *Elene*. *Maldon* has a positive difference of **motan** (“they asked that they might have passage”) and above all of **willan** (“those fled from battle who did not want to be there”), and the *Elegies* of **magan** (“I can tell about myself”). *Genesis A* and *Genesis B* have opposite differences, apart from **willan**, whereas *Judith* and the minor heroic poems have negative differences in all the modals. *Riddles* and *Dream of the Rood* do not vary significantly from the norm.

Fig. 6 Modal verbs

