Form-Meaning Regularities in Old English Thesaurus and Corpora

Drozhashchikh Nataliia, Efimova Elena and Meshcheryakova Evgenia
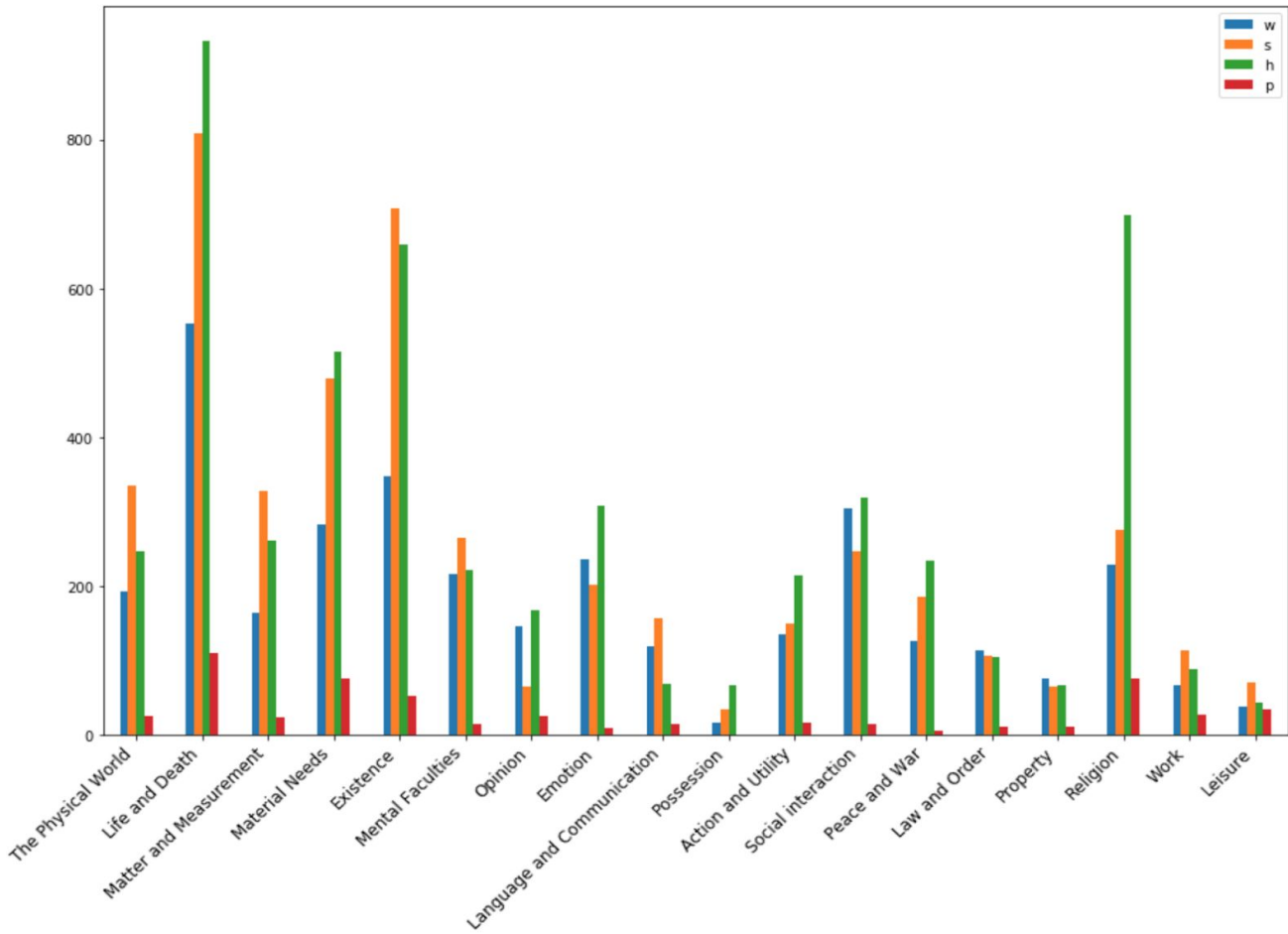
Supplementary materials

Form-meaning association at the lexical level

Table 1. The distribution of Old English initial graphemes <w>, <s>, <p>, <h> in 18 TOE semantic categories

	w	s	h	p
The Physical World	193	335	247	25
Life and Death	553	809	932	109
Matter and Measurement	164	327	261	23
Material Needs	283	478	514	76
Existence	347	707	658	53
Mental Faculties	216	264	222	15
Opinion	146	64	167	26
Emotion	236	202	307	9
Language and Communication	118	157	69	14
Possession	17	34	67	0
Action and Utility	135	150	214	16
Social interaction	304	246	318	14
Peace and War	126	185	234	5
Law and Order	114	106	105	10
Property	75	64	66	11
Religion	228	275	698	75
Work	66	113	88	27
Leisure	38	70	43	34

Figure 1. The visualisation of the distribution of Old English initial graphemes <w>, <s>, <p>, <h> in 18 TOE semantic categories

The Chi Square statistics: $\chi^2 = 765.67$, $p = 3.637\text{e-}128$, $df = 51$ ($\chi^2 > 67.51$)

Table 2. An example of an alternative distribution contingency table

	religion	other semantic categories	total
p-	75	467	542
other initials	1201	11954	13155

total	1276	12421	13697

Table 3. Chi Square p-value and Phi coefficient data for Old English initial graphemes <w>, <s>, <p>, <h> in 18 TOE semantic categories

	w chi2, p-value, φ	s chi2, p-value, φ	h chi2, p-value, φ	p chi2, p-value, φ
The Physical World		26.87 2.17E-07 0.04		
Life and Death		0.045 0.83 0.002	0.69 0.4 0.007	2.57 0.1 0.01
Matter and Measurement		27.99 1.22E-07 0.05		
Material Needs		2.43 0.12 0.01	4.45E-05 0.99 5.70E-05	10.98 0.0009 0.03
Existence		39.33 3.59E-10 0.05		
Mental Faculties	12.83 0.0003 0.03	3.79 0.052 0.02		
Opinion	30.73 2.96E-08 0.05 / 5%		2.04 0.15 0.01	6.8 0.009 0.02
Emotion	19.79 8.63E-06 0.04		2.43 0.12 0.01	
Language and Communication	14.14 0.0001	17.76 2.51E-05		

	0.03	0.04		
Possession			17.74 2.53E-05 0.04	
Action and Utility	0.83 0.36 0.008		2.8 0.09 0.01	
Social interaction	50.36 1.28E-12 0.06			
Peace and War		0.006 0.94 0.0007	4.94 0.03 0.02	
Law and Order	16.76 4.23E-05 0.03			
Property	12.33 0.0004 0.03			0.74 0.39 0.007
Religion			165.79 6.13E-38 0.11	13.66 0.0002 0.03
Work		3.31 0.069 0.02		21.59 3.37E-06 0.04
Leisure		1.6 0.2 0.01		102.62 4.05E-24 0.08

Form-meaning association at the textual level

Figure 2. Topics visualisation (pyLDAvis, Intertopic Distance Map via t-SNE)
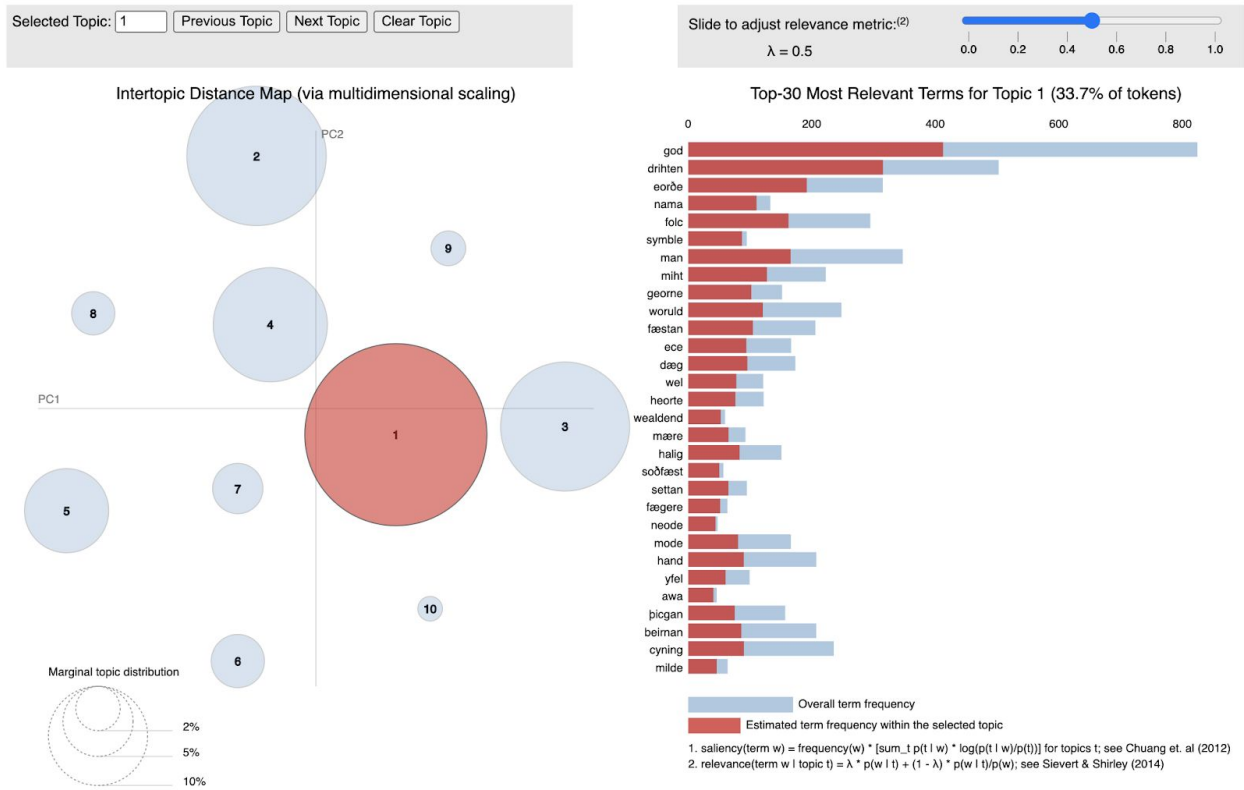
Figure 3. Umass topic coherence score (model trained on Calgary corpus, 3-fold cross-validation)
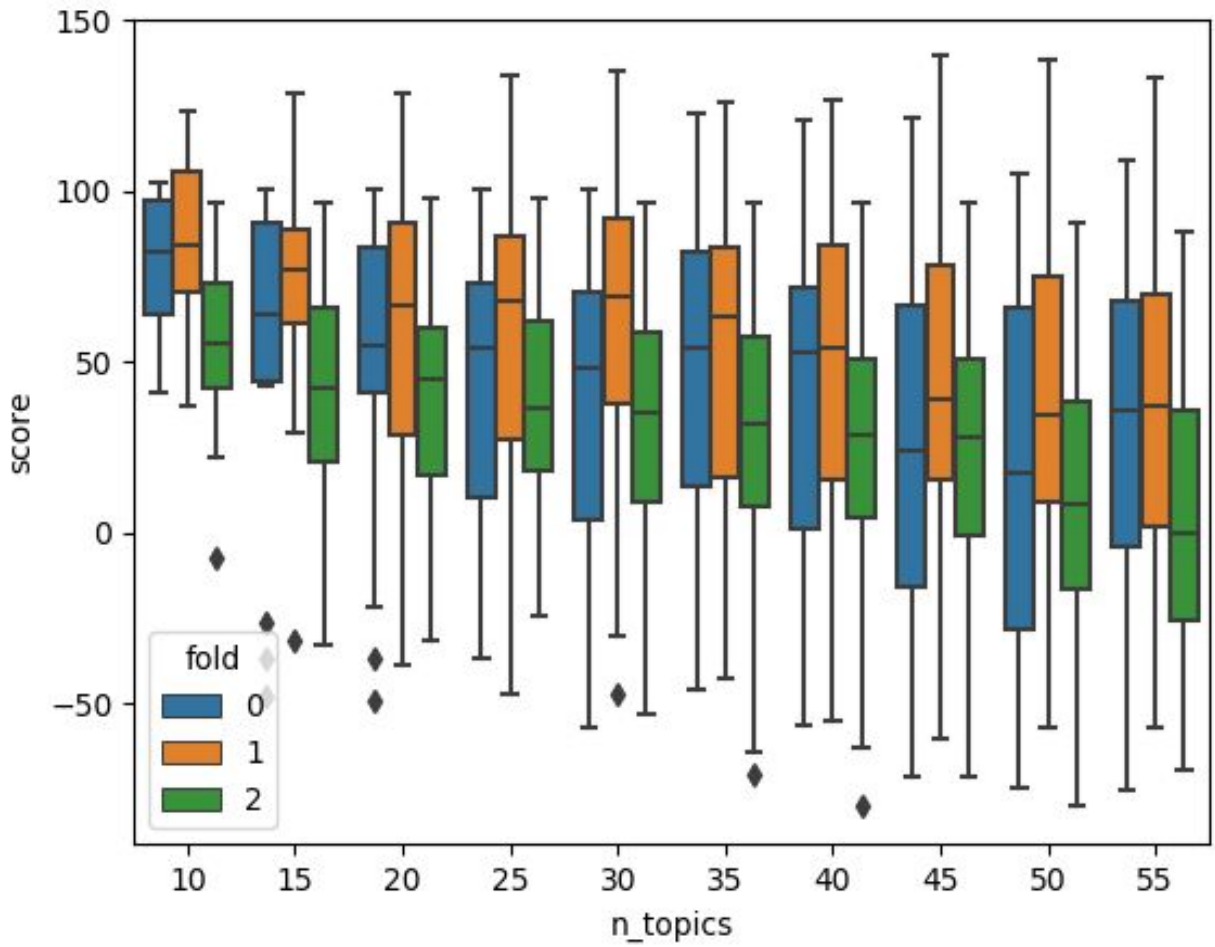
Figure 4. Umass topic coherence score (model trained on Calgary corpus, tested on YCOE corpus)