Vol.2019-CH-119 No.11
2019/2/16

Body Language in Classic British Fiction: Words, *n*-grams, and topics

Tomoji Tabata,[a)]

Abstract: This pilot study takes a stylometric approach to investigate "body language" in classic British fiction. The specific research questions are how body-part words are distributed across registers, whether frequency patterns of body language make it possible to classify texts into meaningful sets, as well as what stylistic functions body-part expressions lend themselves to in fiction. To answer the questions, stylometric analysis is carried out in conjunction with topic modelling and qualitative interpretation of stylistics effects by paying close attention to individual words, *n*-grams, and topics.

Keywords: Body language, style, classic fiction, stylometry, topics

1. Introduction

This article was prepared as a handout for presentation given at the 119th SIG-CH (Computers and the Humanities) conference to be held at the Toyonaka Campus of the University of Osaka on the 16th February 2019.

A first step of text analysis often begins with identifying key words of a target author or a target corpus. Based on the assumption that key words encapsulate important lexical, semantic, thematic, or stylistic features of the author, we try to find words "that are evenly distributed across [her/his corpus] and display a higher frequency and a wider range than in a reference corpus of some kind" (Paquot and Bestgen, 2009). The most popular methods for detecting key words include application of statistical significance testings, such as chi-squared (χ^2), log-likelihood ratio (G^2)(Rayson and Garside, 2000), Mann-Whitney's U, and representativeness/distinctiveness (RD) index (Klaussner et al. 2015), among others. While χ^2 and G^2 can over-evaluate "bursty" words (those which occur extremely frequently in a very small number of texts, typically in a single text only), tending to give greater weights to high-frequency words due to their formulae, U and RD help to find lexical items "that are [more] evenly distributed across [the target set]". As a result of comparing a Charles Dickens corpus with a reference corpus of classic British fiction using RD, for instance, a list of following key words can be obtained:

> ***head****, down, old, corner, state,* given*,* ***legs****, fire, window, until, streets, night,* for*, outside, looking, air, dark, round, behind,* return*, shut, light, upon,* than*, hat, quiet,* without, would*, up, top,* future*, reference, wall, off, remarkable,* to, determined*, arranged, red, scarcely, not, but, heavy, bright, boy,* till*, slowly, street,* neither*, sky, expression,* desired*, stopped, through, with, dead, beside,* only*, dismal, sharp, leaning, windows, worn,* nevertheless*, holding,* ***face****,* nor, must*, bell, closed,* patience, therefore*, black, wasn't, stopping, times,* could*, anybody,* consciousness*, position,* ***faces****,* pleased*, various,* endeavour, use*, curious,* received*, shake, gradually, being,* feeling*, iron,* advice*, effort, wind, inside, here, back,* assured*,* ***hair***

Words in italics are the most representative and distinctive Dickens markers while words in Roman are consistently more overused in the reference set. What attracts attention in the Dickensian key words are body-part words such as *head, legs, face, faces*, and *hair*. Attentive readers of Dickens novels will associate such body-part expressions with description of actions, habits, emotions, and personalities of fictional characters. A team of researchers led by Michaela Mahlberg of the University of Birmingham are, in fact, working on characterization in the representation of body language from a corpus linguistic perspective (the CLiC Dickens project) (Mahlberg et al. 2016; Wiegand et al., 2017; Mahlberg and Wiegand, 2018).

The aim of this paper is to take a stylometric approach to investigate body language in classic British fiction with Dickens being the target author. The specific research questions are:

(1) how body-part words are distributed across registers
(2) whether frequency patterns of body language make it possible to classify texts into meaningful sets
(3) what stylistic functions body-part expressions lend themselves to in fiction

To answer the questions, stylometric analysis is carried out in conjunction with topic modelling and qualitative interpretation of stylistics effects by paying close attention to individual words, *n*-grams, and topics.

2. Body language across registers

As a first step to list up as many body-part words as pos-

[1] GSLC, University of Osaka, Toyonaka, Osaka 560–0043, Japan
[a)] tabata@lang.osaka-u.ac.jp

sible, all the word-types in a corpus of classic British fiction were examined. The corpus entitled **O**saka **R**eference **C**orpus for **Hi**storical/**D**iachronic **S**tylistics (ORCHiDS) is made up of 23 texts by Dickens, 24 major 18th Century novels, and 31 representative 19th Century fiction texts with a total tokens of 14,481,460 running words (Tables 2–4). Fig. 1 shows 150 lexical items found in the corpus.

abdomen, #abdominal, #abdominales, ankle, ankles, anus, arm, arm's, arms, arms', arse, artery, back, belly, bladder, blood, body, bone, brain, breast, breasts, buttocks, calf, cheek, cheeks, chest, chin, ear, ears, elbow, elbows, eye, eyes, eyebrow, eyebrows, eyelashes, eyelid, eyelids, face, faces, feet, finger, fingers, finger-nail, finger-nails, fingernail, fingernails, foot, forehead, foreheads, groin, gums, hair, hairs, hand, hands, head, heads, heart, hearts, heel, heels, hip, hips, instep, insteps, intestines, iris, jaw, jaws, kidney, knee, knees, leg, legs, ligament, ligaments, lip, lips, liver, lobe, lungs, mouth, mouths, muscle, muscles, nail, nails, navel, neck, necks, nerves, nose, noses, nostril, nostrils, organs, ovary, palm, palms, pinky, pituitary, #pore, pupil, ribs, #scalp, shin, shoulder, shoulders, skeleton, skin, skull, skulls, #sole, spinal, spine, spines, spleen, stomach, stomachs, teeth, tendon, tendons, thigh, thighs, #thorax, throat, throats, thumb, thumbs, tibia, #tissue, toe, toes, #toe-nails, tongue, tongues, #tonsils, tooth, torso, #uterus, #uvula, vein, veins, vertebrae, waist, waists, wrist, wrists

Fig. 1 150 body-pard words extracted from ORCHiDS

In an effort to answer the first research question, total frequencies (per 100,000 words) of body-part words were counted in each of fifteen registers in two 1,000,000-word balanced corpora: the FLOB corpus (British English) and FROWN corpus (American English). The two corpora were complied according to the same corpus design in terms of the size and number of samples that make up fifteen registers: nine informative prose registers and six imaginative (fictional) prose registers. Table 1 gives a descriptive summary of the corpus design.

Table 1 The structure of FLOB/FROWN corpora

Category	Register	No. of texts
A	Press: Reportage	44
B	Press: Editorial	27
C	Press: Review	17
D	Religion	17
E	Skills, Trades and Hobbies	38
F	Popular Lore	44
G	Belles Lettres, Biographies, Essays	77
H	Miscellaneous	30
J	Science	80
K	General Fiction	29
L	Mystery and Detective Fiction	24
M	Science Fiction	6
N	Adventure and Western	29
P	Romance and Love Story	29
R	Humour	9
	Total number of texts in the corpus:	500

The categories A–J are informative registers, and K–R fictional registers.

As Fig. 2 illustrates, body-part related words are predominantly more frequent in fictional registers in both British and American English. The overall distributional pattern in the two corpora is remarkably similar: the category most given to body language is Romance and Love story (P), with Humour (R) having the least recourse to parts of body in the fictional registers.

Of the 150 items given in Fig. 1, words with # were not found in the two corpora and 121 words are common in both corpora. Fig. 3 is a result of conducting Principle Component Analysis on a frequency matrix of the 121 body-part words across the 15 registers in the FLOB and FROWN corpora.[*1] Whereas informative registers form a closely-knit cluster, fictional registers are widely dispersed. Of further note is that fictional registers in the two national varieties of English have remarkably similar positions in the 3D diagram. To sum up, body language is a distinctive feature of fiction since depiction of body part plays significant roles in fictional settings.

Table 2 ORCHiDS Dickens Component

No.	Texts	Category	Date	Tokens
1	*Sketches by Boz*	Sketches	1833–1836	187,493
2	*The Pickwick Papers*	Serial Fiction	1836–1837	300,269
3	Other Early Papers	Sketches	1837–1840	67,121
4	*Oliver Twist*	Serial Fiction	1837–1839	157,707
5	*Nicholas Nickleby*	Serial Fiction	1838–1839	322,393
6	*Master Humphrey's Clock*	Miscellany	1840–1841	47,331
7	*The Old Curiosity Shop*	Serial Fiction	1840–1841	217,521
8	*Barnaby Rudge*	Serial Fiction	1841	254,141
9	*American Notes*	Sketches	1842	101,941
10	*Martin Chuzzlewit*	Serial Fiction	1843–1844	336,630
11	*Christmas Books*	Fiction	1843–1848	154,449
12	*Pictures from Italy*	Sketches	1846	72,553
13	*Dombey and Son*	Serial Fiction	1846–1848	342,386
14	*David Copperfield*	Serial Fiction	1849–50	356,110
15	*Bleak House*	Serial Fiction	1852–1853	354,369
16	*Hard Times*	Serial Fiction	1854	103,443
17	*Little Dorrit*	Serial Fiction	1855–1857	338,502
18	*Reprinted Pieces*	Sketches	1850–1856	91,574
19	*A Tale of Two Cities*	Serial Fiction	1859	136,259
20	*The Uncommercial Traveller*	Sketches	1860–1869	142,973
21	*The Great Expectations*	Serial Fiction	1860–1861	184,889
22	*Our Mutual Friend*	Serial Fiction	1864–1865	326,158
23	*The Mystery of Edwin Drood*	Serial Fiction	1870	94,143
			Sum of word-tokens in the set of Dickens texts:	**4,690,355**

Table 3 ORCHiDS 18th Century Subcorpus

No.	Author	Texts	Date	Tokens
1	Defoe	*Captain Singleton*	1720	110,843
2	Defoe	*Journal of Prague Year*	1722	94,695
3	Defoe	*The Military Memoirs of Captain George Carleton*	1728	80,612
4	Defoe	*Moll Flanders*	1724	136,241
5	Defoe	*Robinson Crusoe*	1719	232,324
6	Fielding	*A Journey from this World to the Next*	1749	45,003
7	Fielding	*Amelia*	1751	211,678
8	Fielding	*Jonathan Wild*	1743	69,938
9	Fielding	*Joseph Andrews*	1742	125,342
10	Fielding	*Tom Jones*	1749	346,256
11	Goldsmith	*The Vicar of Wakefield*	1766	62,976
12	Richardson	*Clarissa*	1748	935,894
13	Richardson	*Pamela*	1740	438,937
14	Smollet	*Peregrine Pickle*	1751	342,200
15	Smollett	*Ferdinand Count Fathom*	1753	157,641
16	Smollett	*Humphrey Clinker*	1771	150,395
17	Smollett	*Sir Launcelot Greaves*	1760	89,388
18	Smollett	*Roderick Random*	1748	191,602
19	Smollett	*Travels through France and Italy*	1766	121,074
20	Sterne	*A Sentimental Journey*	1768	40,783
21	Sterne	*Tristram Shandy*	1759–1767	186,426
22	Swift	*A Tale of a Tub*	1704	44,121
23	Swift	*Gulliver's Travels*	1726	104,047
24	Swift	*A Journal to Stella*	1710–1713	189,587
		Sum of word-tokens in the set of 18th Century texts:		**4,508,003**

[*1] The frequency matrix of the 121 body-part words across the 15 registers in the FLOB/FROWN corpora is omitted due to space constraints. The data is available upon request to the author.

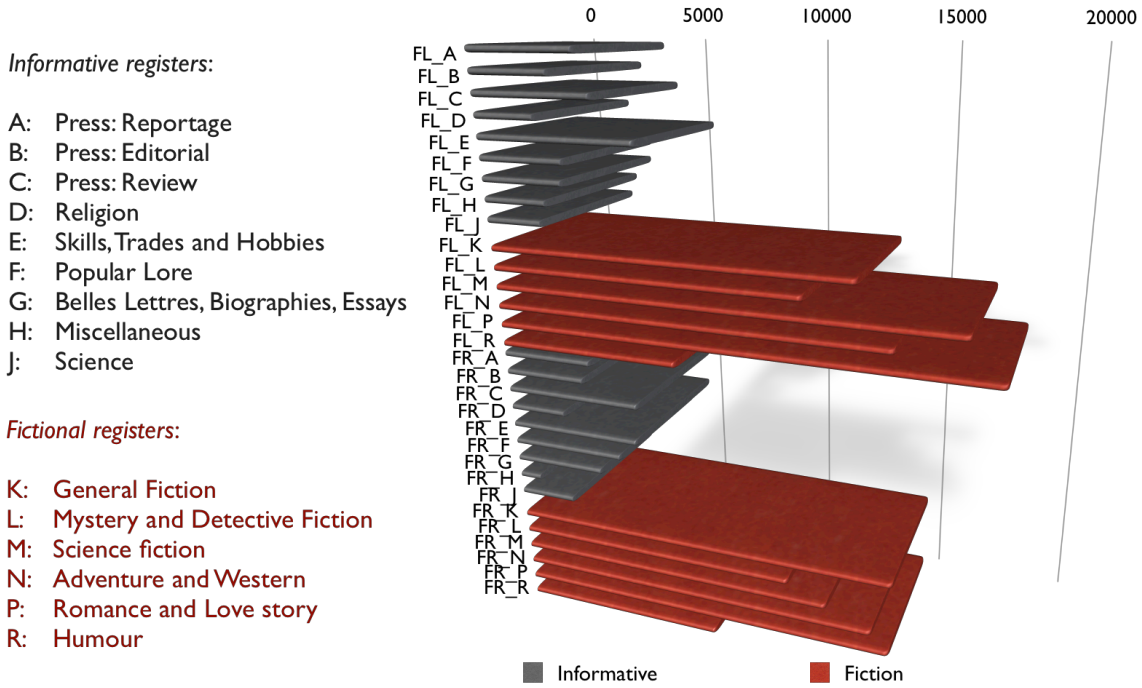

Fig. 2 Frequencies of body-part words (per 100,000 words) across 15 registers in the FROWN and FLOB corpora

Table 4 ORCHiDS 19th Century Subcorpus

No.	Author	Texts	Date	Tokens
1	Austen	*Emma*	1815	160,307
2	Austen	*Mansfield Park*	1814	159,792
3	Austen	*Northanger Abbey*	1803	77,250
4	Austen	*Persuasion*	1816	83,303
5	Austen	*Pride and Prejudice*	1813	122,253
6	Austen	*Sense and Sensibility*	1811	119,502
7	A.Brontë	*Agnes Grey*	1847	68,222
8	C.Brontë	*Jane Eyre*	1847	186,272
9	C.Brontë	*The Professor*	1857	88,081
10	C.Brontë	*Villette*	1853	194,479
11	E.Brontë	*Wuthering Heights*	1847	116,431
12	Collins	*After Dark*	1882	136,884
13	Collins	*The Moonstone*	1868	194,136
14	Collins	*The Woman in White*	1859	246,799
15	G.Eliot	*Adam Bede*	1859	215,101
16	G.Eliot	*Brother Jacob*	1864	16,683
17	G.Eliot	*Daniel Deronda*	1876	311,115
18	G.Eliot	*Middlemarch*	1871–1872	318,065
19	G.Eliot	*Silas Marner*	1861	71,389
20	G.Eliot	*The Mill on the Floss*	1860	207,341
21	Gaskell	*Cranford*	1851–1853	70,947
22	Gaskell	*Mary Barton*	1848	160,589
23	Gaskell	*Sylvia's Lovers*	1863	190,829
24	Thackeray	*Barry Lyndon*	1844	126,048
25	Thackeray	*Vanity Fair*	1848	303,935
26	Trollope	*Barchester Towers*	1857	198,369
27	Trollope	*Can You Forgive Her*	1865	316,151
28	Trollope	*Doctor Thorne*	1857	217,748
29	Trollope	*The Eustace Diamonds*	1873	269,818
30	Trollope	*Phineas Finn*	1869	263,231
31	Trollope	*The Warden*	1855	72,032
Sum of word-tokens in the set of 19th Century texts: 5,283,102				

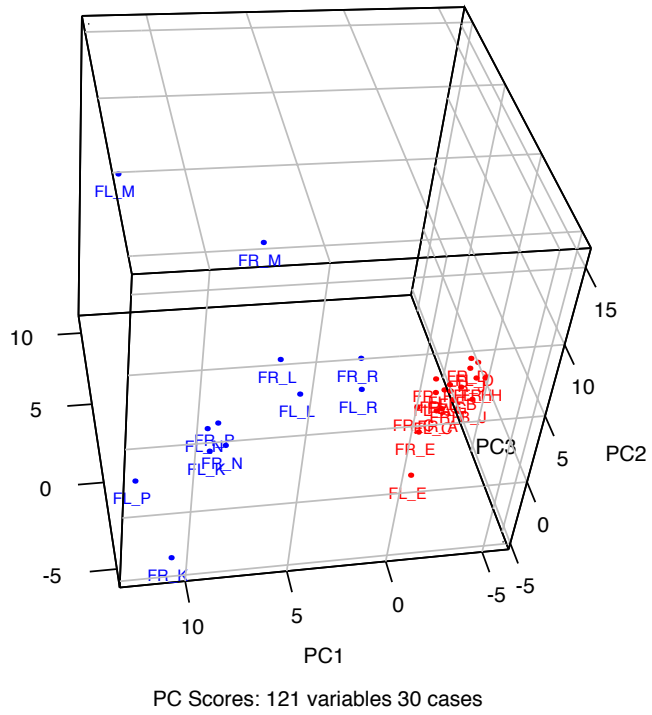

Fig. 3 Principal Component Analysis of 121 body-part words across 15 registers in the FROWN and FLOB corpora

3. Body-part words in classic British fiction

Normalized frequencies of the body-part words (per 100,000 running words) were counted in order to examine complex interrelationships between the terms, the texts, and association between the terms and the text. Word frequency profiles, of which part is shown as Table 5 were then fed to a series of multivariate analysis, including PCA and Correspondence Analysis, and machine-learning-based classification algorithm (Random Forests). The three subcorpora in the ORCHiDS are differentiated according to differing frequency patterns of body language.

To supplement findings from stylometric analysis, topic modelling was carried out using MALLET (a JAVA-based MAchine Learning LanguagE Toolkit developed at the University Massachusetts at Amherst)[*2]. What is of special interest is that a topic made up of body-language-related words was found to discriminate strongly in favour of the Dickens set (Fig. 4). Fig. 5 shows topic density distribution across the authors in the corpus. One Dickens sample is shown to have nearly as much as 40% accounted for by words belonging to the topic labelled facial/bodily gestures.

Full results will be provided at the 119th SIG-CH conference.

References

[1] Paquot, M. and Bestgen, Y. (2009). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction, in Jucker, A., Schreier D., and Hundt M. (eds.) *Corpora: Pragmatics and Discourse. Papers from the 29th International Conference on English Language Research on Computerized Corpora* (*ICAME 29*), pp. 247–269.

[2] Klaussner, C., Nerbonne, J., and Coltekin, C. (2015). Finding Characteristic Features in Stylometric Analysis, *Digital Scholarship in the Humanities*, Vol. 30, Supplement 1, i114–i129.

[3] Rayson, P. and Garside, R. (2000). Comparing corpora using frequency profiling, In proceedings of the *Workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics* (ACL 2000). 1–8 October 2000, Hong Kong, 1–6.

[4] Mahlberg, M., Stockwell, P., de Joode, J., Smith, C., and O'Donnell, M. B. (2016). CLiC Dickens: Novel uses of concordances for the integration of corpus stylistics and cognitive poetics. *Corpora*, 11 (3), 433–463.

[5] Mahlberg, M. and Wiegand V. (2018). Corpus stylistics, norms and comparisons: Studying speech in Great Expectations, in Page, R., Busse, B., and Norgaard, N. (eds.) *Rethinking Language, Text and Context. Interdisciplinary Research in Stylistics in Honour of Michael Toolan* New York & London: Routledge, pp. 123–143.

[6] Wiegand, V., Mahlberg, M., and Stockwell, P. (2017, September 18). Corpus linguistics in action: The fireplace pose in 19th century fiction. *The Programming Historian* blog post available from ⟨https://programminghistorian.org/posts/corpus-linguistics-in-action?ContensisTextOnly=true⟩ (accessed 2019-1-24).

Table 6 Words most strongly distinguishing between Dickens, 18th century fiction (ECF), and 19th century fiction (NCF)

Item	Gini Index	Accuracy	Dickens	ECF	NCF	Key Group
head	3.45934862	4.68480029	4.51573043	3.2804482	2.131026953	Dickens
belly	2.93615044	4.82666252	2.78455667	4.7231821	3.225332783	ECF
legs	2.85714671	3.87376684	3.92005145	2.426893	2.268022571	Dickens
breast	1.88716461	2.68483068	1.810141	-1.1644206	3.243361524	Dickens
heels	1.87774198	2.82987099	0.40838726	1.3367968	3.091538789	Dickens
cheeks	1.74496864	4.12310538	2.79298169	2.6308761	3.032714115	NCF
face	1.47790985	2.99578692	1.83803828	2.7015393	1.194675822	Dickens
hair	1.47287166	2.89868309	2.33714549	3.1574	-0.27317367	Dickens
back	1.34026863	3.51834501	2.47797086	2.1673312	2.045400912	Dickens
chin	1.31265667	3.21763246	2.85822505	0.9245071	3.195374236	Dickens
arm	1.239132	3.46846873	1.64456984	3.1226619	1.373849664	Dickens
hand	1.18311951	2.88112033	2.5254471	1.6153094	2.307059346	Dickens
feet	1.14666102	2.16540652	1.77208103	2.5939557	-0.14514389	Dickens
heart	1.07120984	2.26383803	1.35388105	1.8314778	1.146285049	NCF
throat	1.02456614	3.22698856	2.72740121	2.2293485	1.459072634	Dickens
eye	0.93785247	2.85880896	2.31836623	1.9874613	1.86526266	Dickens
finger	0.85558676	2.49350697	2.49097962	1.1607721	0.433995389	Dickens
hands	0.85394992	1.97852855	2.05698372	-0.1510919	1.753822422	Dickens
heads	0.82935414	2.27669403	1.46686413	-0.2009896	2.666760572	Dickens
shoulder	0.80448344	-0.1381218	1.79717065	-0.8040737	-0.1009469	Dickens
foot	0.79850267	2.50553778	1.89877841	0.3677838	1.763657861	Dickens
forehead	0.74678004	1.21123296	0.44854427	1.8473995	0.241281386	Dickens
body	0.71325697	2.04762125	1.37669456	1.5537086	1.080635767	ECF
nose	0.68183935	2.07042341	1.61407003	-0.1646997	1.876353453	Dickens
stomach	0.63990065	-0.56342138	-0.16153564	-1.0050378	-0.10344859	ECF
lips	0.63054557	0.52373244	0.68856853	2.0076963	-1.56256654	Dickens
elbows	0.60585681	0.99965539	0.48998943	1.8625435	-1.28048879	Dickens
faces	0.59524962	1.81912747	1.25259896	1.3573028	-0.73593074	Dickens
knees	0.59341116	2.05889911	2.032506	-1.0254725	2.034047374	Dickens
nerves	0.56793053	2.72545118	1.73799798	0.9454092	2.34839828	NCF
leg	0.54402506	0.96119038	1.23724356	-1.0050378	0.828239801	Dickens
fingers	0.5374956	1.80154455	-0.16632479	1.7123248	1.621181444	Dickens
knee	0.46714756	1.66255978	1.53421474	1.2017015	0.277456833	Dickens
tongue	0.45786288	1.57146577	1.54770367	0.582543	1.158450965	NCF
thumb	0.45440173	1.31711786	1.88495641	0	-0.83975465	Dickens
eyebrows	0.45310136	2.42132982	1.73399615	1.8711858	-0.14031677	Dickens
bone	0.44369894	-1.55017505	-0.12804738	-0.5984196	-1.6192606	ECF
tooth	0.43343815	-0.83254546	1.00503782	-1.7525586	-1.00503782	Dickens
spleen	0.43268168	1.8849915	1.41965726	1.5255594	0.646485551	ECF
neck	0.38464197	-1.52522174	-1.00503782	-1.7139779	-0.44766148	Dickens
eyes	0.36915844	2.02489705	1.00503782	1.9809722	1.731258439	Dickens
arms	0.35950636	1.89713505	1.40421357	0.2458438	2.017638471	Dickens
mouth	0.34897913	1.18871195	-0.09407625	1.716752	0.709218139	Dickens
arm's	0.34491626	1.43792127	1.68234392	1.0050378	0.751646028	Dickens

Table 5 Part of the frequency profiles for body-part words

texts	text.group	back	hand	head	face	eyes	hands	body	hair	heart	feet	blood	eye	arms	mouth	skin	arm	fingers	shoulder	foot	neck	faces	legs	nose
Dickens33_SB	Dickens	933.5	864.1	794.8	768.1	613.4	480.1	170.7	250.7	282.7	240.0	80.0	309.4	250.7	117.3	21.3	288.0	26.7	42.7	181.4	69.3	90.7	250.7	64.0
Dickens36_PP	Dickens	916.3	1216.2	1342.8	899.6	992.9	586.4	223.2	153.3	419.8	243.2	86.6	363.2	266.6	136.6	33.3	306.5	53.3	100.0	119.9	90.0	123.3	269.9	173.3
Dickens37a_OEP	Dickens	418.3	537.8	836.6	433.2	657.3	433.2	254.0	224.1	418.3	388.4	74.7	373.5	194.2	134.5	44.8	239.0	44.8	14.9	119.5	29.9	194.2	313.7	89.6
Dickens37b_OT	Dickens	1567.3	1300.8	1434.0	1110.4	1294.4	786.8	215.7	222.1	590.1	304.6	253.8	374.4	241.1	196.7	19.0	323.6	63.5	145.9	209.4	120.6	228.4	190.4	158.6
Dickens38_NN	Dickens	977.5	1114.0	1194.7	918.5	856.4	605.1	192.4	207.9	605.1	189.3	124.1	338.2	201.7	180.0	9.3	403.4	111.7	124.1	133.4	86.9	124.1	186.2	204.8
Dickens40a_MHC	Dickens	827.5	1328.4	1001.7	1023.5	892.9	479.1	413.8	392.0	871.1	326.7	152.4	196.0	65.3	65.3	0.0	370.2	43.6	87.1	196.0	65.3	196.0	239.5	130.7
Dickens40b_OCS	Dickens	998.3	1278.9	1237.5	883.3	1159.3	708.5	115.0	174.8	579.6	276.0	55.2	280.6	216.2	165.6	13.8	280.6	87.4	110.4	133.4	156.4	156.4	266.8	179.4
Dickens41_BR	Dickens	1224.5	1500.1	1248.1	1157.6	1071.0	893.8	311.0	192.9	645.7	259.9	189.0	330.7	303.2	149.6	31.5	370.1	70.9	196.9	275.6	122.1	177.2	185.1	78.7
Dickens42_AN	Dickens	706.6	726.2	814.5	549.6	628.1	677.1	245.3	147.2	235.5	235.5	58.9	402.4	147.2	117.8	49.1	245.3	206.1	39.3	147.2	108.0	245.3	235.5	88.3
Dickens43_MC	Dickens	891.5	1236.2	1200.5	948.0	817.2	674.6	104.0	208.0	573.5	133.7	121.8	291.2	175.3	139.7	14.9	365.5	101.0	112.9	160.5	53.5	68.3	181.3	112.9
Dickens43b_CB	Dickens	1211.0	1638.4	1528.3	1878.0	1282.2	1068.5	129.5	336.7	926.1	401.5	77.7	375.6	395.0	194.3	6.5	349.7	90.7	148.9	213.7	142.5	194.3	220.2	136.0
Dickens46a_PFI	Dickens	758.1	744.3	799.4	565.1	427.3	523.8	303.2	317.0	234.3	427.3	248.1	261.9	317.0	55.1	13.8	55.1	110.3	110.3	358.4	96.5	344.6	399.7	68.9
Dickens46b_DS	Dickens	1162.9	1618.7	1609.9	1758.9	1352.8	616.5	99.3	263.0	794.7	175.3	93.5	385.7	309.7	151.9	29.2	379.8	52.6	128.6	181.2	163.6	78.9	137.3	113.9
Dickens49_DC	Dickens	1134.9	1429.9	1393.4	1255.7	988.8	609.6	64.6	348.3	716.4	140.5	75.8	278.1	261.3	126.4	5.6	393.3	59.0	154.5	151.7	129.2	64.6	118.0	123.6
Dickens51_CHE	Dickens	772.3	545.5	1072.6	282.0	220.7	404.5	514.9	122.6	380.0	128.7	306.5	36.8	306.5	36.8	24.5	85.8	12.3	12.3	177.8	73.6	67.4	79.7	49.0
Dickens52_BH	Dickens	911.8	1239.3	1228.0	1058.6	827.1	592.8	87.5	313.4	527.9	135.5	70.6	367.0	338.8	124.2	14.1	352.9	90.3	149.6	138.3	110.1	76.2	141.2	48.0
Dickens54_HT	Dickens	967.6	1441.7	1315.9	1490.1	1074.0	706.3	135.5	329.0	532.2	106.4	38.7	425.7	222.5	145.1	29.0	319.3	29.0	116.1	261.2	135.5	116.1	203.2	164.5
Dickens55_LD	Dickens	1040.3	1513.2	1129.0	969.4	1019.7	540.9	94.6	274.9	597.0	168.5	53.2	269.0	174.4	144.8	0.0	413.8	150.7	130.0	141.9	91.6	82.8	124.1	156.6
Dickens56_RP	Dickens	797.2	633.4	764.4	546.0	469.6	349.5	207.5	273.0	218.4	229.3	120.1	524.2	283.9	120.1	142.0	109.2	43.7	54.6	131.0	65.5	142.0	436.8	131.0
Dickens59_TTC	Dickens	991.5	1814.0	1292.6	1373.4	1211.8	822.6	154.2	462.7	543.5	323.1	257.1	249.7	359.9	154.2	0.0	477.4	242.4	183.6	176.3	95.5	396.6	73.4	117.5
Dickens60a_UT	Dickens	993.7	874.8	951.7	727.8	545.9	461.9	188.9	307.9	328.9	300.9	42.0	349.9	293.9	140.0	35.0	272.9	35.0	112.0	154.0	63.0	175.0	286.9	161.0
Dickens60b_GE	Dickens	1320.6	1455.9	1266.4	893.0	974.2	990.4	92.0	443.8	492.5	184.0	151.5	259.8	238.1	249.0	32.5	362.6	167.8	292.3	113.7	129.9	97.4	189.4	146.1
Dickens64_OMF	Dickens	1168.2	1416.5	1346.0	1162.0	1346.0	745.0	141.0	334.2	407.8	165.6	82.8	285.1	269.8	168.6	15.3	518.2	82.8	144.1	242.2	162.5	92.0	196.2	180.9
Dickens70_MED	Dickens	1190.3	1456.0	1041.5	1232.8	945.8	797.1	127.5	286.9	340.1	191.3	170.0	255.1	223.2	138.2	21.3	425.1	63.8	201.9	180.7	106.3	127.5	148.8	95.6
18_Defoe_Captain_Singleton	18thC	793.9	478.2	460.1	72.2	117.3	622.5	153.4	63.2	189.5	126.3	171.4	27.1	315.8	153.4	63.2	99.2	9.0	18.0	198.5	45.1	27.1	54.1	0.0
18_Defoe_Jnl_of_Prague_Year	18thC	551.5	683.4	359.7	131.9	119.9	275.8	383.7	36.0	179.8	119.9	203.8	48.0	155.9	59.9	0.0	0.0	0.0	24.0	71.9	48.0	71.9	12.0	12.0
18_Defoe_Military_Memoirs_of_Capt_George_Carleton	18thC	521.0	384.6	496.2	210.9	285.3	483.8	657.5	37.2	186.1	74.4	111.6	235.7	272.9	148.9	12.4	148.9	12.4	37.2	459.0	12.4	49.6	86.8	24.8
18_Defoe_Moll_Flanders	18thC	785.7	712.3	330.4	168.9	124.8	536.0	132.2	0.0	374.5	29.4	110.1	29.4	176.2	102.8	7.3	51.4	29.4	7.3	88.1	22.0	7.3	7.3	0.0
18_Defoe_Robinson_Crusoe	18thC	792.0	598.3	563.9	176.5	202.3	675.8	249.7	43.0	254.0	284.1	172.2	30.1	383.1	167.9	103.3	86.1	30.1	51.7	331.4	21.5	17.2	47.3	30.1
18_Fielding_A_Journey_from_This_to_Next_World	18thC	800.8	311.4	333.7	178.0	444.9	178.0	422.7	0.0	333.7	0.0	89.0	89.0	66.7	111.2	22.2	89.0	22.2	44.5	66.7	44.5	89.0	44.5	44.5
18_Fielding_Amelia	18thC	231.0	367.7	278.1	273.4	561.0	235.7	127.3	14.1	570.4	56.6	122.6	47.1	235.7	89.6	4.7	47.1	14.1	9.4	51.9	47.1	0.0	23.6	51.9
18_Fielding_Jonathan_Wild	18thC	171.4	242.9	585.8	71.4	700.1	700.1	171.4	57.1	314.3	85.7	185.7	171.4	228.6	85.7	42.9	57.1	42.9	14.3	42.9	71.4	28.6	28.6	28.6
18_Fielding_Joseph_Andrews	18thC	327.1	606.3	486.6	374.9	741.9	359.0	135.6	119.7	534.5	71.8	183.5	23.9	247.3	103.7	71.8	111.7	71.8	0.0	199.4	63.8	39.9	111.7	87.8
18_Fielding_Tom_Jones	18thC	256.9	421.4	430.1	329.0	505.1	294.4	138.5	43.3	663.8	69.3	239.6	95.2	184.7	129.9	28.9	103.9	31.7	14.4	80.8	46.2	37.5	57.7	60.6
18_Goldsmith_The_Vicar_of_Wakefield	18thC	476.0	460.1	222.1	412.5	126.9	174.5	158.7	190.4	920.3	63.5	79.3	222.1	301.5	0.0	47.6	190.4	31.7	15.9	126.9	63.5	111.1	15.9	0.0
18_Richardson_Clarissa	18thC	344.8	711.9	445.6	321.5	433.9	428.6	604.7	28.6	1354.8	115.6	59.4	304.5	118.8	108.2	14.9	53.0	65.8	21.2	90.2	61.5	39.3	6.4	20.2
18_Richardson_Pamela	18thC	334.7	847.0	334.7	330.1	350.6	398.4	582.9	15.9	1179.4	104.7	25.0	186.7	198.1	102.5	11.4	72.9	38.7	29.6	104.7	72.9	9.1	6.8	9.1
18_Smollet_PP	18thC	236.8	622.8	383.0	266.1	470.8	271.9	298.2	90.6	760.2	87.7	157.9	225.1	204.7	233.9	32.2	84.8	40.9	61.4	157.9	46.8	38.0	32.2	99.4
18_Smollett_FCF	18thC	203.0	539.2	323.5	228.4	672.4	234.7	279.1	82.5	1205.3	101.5	222.0	323.5	310.8	76.1	38.1	44.4	0.0	19.0	120.5	50.7	38.1	12.7	88.8
18_Smollett_HC	18thC	301.7	583.3	643.7	207.8	449.2	301.7	295.0	160.9	657.1	100.6	234.7	207.8	181.0	140.8	127.4	67.0	33.5	26.8	174.3	80.5	93.9	67.0	254.8
18_Smollett_LG	18thC	347.1	940.4	716.5	403.0	974.0	436.6	335.9	167.9	951.6	123.1	257.5	223.9	358.3	201.5	22.4	134.3	67.2	33.6	179.1	156.7	33.6	44.8	167.9
18_Smollett_RR	18thC	402.0	882.3	584.8	360.2	537.8	370.7	673.5	125.3	532.5	120.1	323.7	135.7	229.7	224.5	141.0	88.8	57.4	78.3	146.2	114.9	41.8	57.4	156.6
18_Smollett_TFI	18thC	90.9	396.4	553.4	190.0	223.0	123.9	396.4	214.7	90.9	396.4	173.4	214.7	99.1	115.6	115.6	57.8	41.3	16.5	272.6	49.6	90.9	132.1	24.8
18_Sterne_A_Sentimental_Journey	18thC	1127.8	1986.0	686.5	637.5	588.4	613.0	343.3	98.1	1422.1	98.1	245.2	294.2	122.6	98.1	0.0	318.7	73.6	24.5	294.2	122.6	122.6	98.1	49.0
18_Sterne_Tristram_Shandy	18thC	724.0	1464.2	1163.8	348.6	364.7	466.6	579.2	134.1	863.5	144.8	257.4	321.8	134.1	166.3	32.2	177.0	134.1	53.6	268.2	48.3	48.3	107.3	595.3
18_Swift_A_Tale_of_a_Tub	18thC	158.2	429.4	610.2	90.4	452.0	406.8	565.0	45.2	271.2	113.0	22.6	45.2	113.0	135.6	90.4	0.0	45.2	293.8	67.8	0.0	22.6	22.6	271.2
18_Swift_Gulliver's_Travels	18thC	482.1	636.3	626.7	424.2	433.9	539.9	617.0	173.5	96.4	1166.6	115.7	48.2	125.3	212.1	86.8	96.4	48.2	19.3	318.2	144.6	67.5	144.6	125.3
18_Swift_The_Journal_to_Stella	18thC	429.0	397.6	612.1	125.6	230.2	313.9	109.9	15.7	172.6	20.9	57.5	47.1	52.3	68.0	5.2	0.0	15.7	83.7	83.7	36.6	26.2	31.4	20.9
19_A.Bronte_Agnes_Grey	19thC	615.6	557.0	381.1	527.7	615.6	351.8	219.9	234.5	879.5	293.2	29.3	117.3	73.3	73.3	14.7	102.6	87.9	14.7	44.0	0.0	73.3	29.3	44.0
19_Austen_Emma	19thC	417.9	330.6	249.5	199.6	318.1	143.5	1203.8	87.3	411.7	6.2	31.2	193.4	12.5	31.2	18.7	99.8	0.0	6.2	49.9	0.0	18.7	0.0	0.0
19_Austen_Mansfield_Park	19thC	513.1	419.2	294.1	156.4	431.8	162.7	43.8	12.5	738.4	25.0	12.5	143.9	18.8	18.8	12.5	68.8	6.3	0.0	25.0	18.8	18.8	0.0	0.0
19_Austen_Northanger_Abbey	19thC	607.5	439.5	413.6	206.8	542.9	129.3	25.9	168.0	943.6	0.0	64.6	400.7	12.9	0.0	25.9	245.6	12.9	12.9	25.9	0.0	12.9	25.9	0.0
19_Austen_Persuasion	19thC	828.3	276.1	420.2	264.1	396.1	240.1	96.0	12.0	504.2	24.0	48.0	288.1	96.0	24.0	12.0	132.0	24.0	12.0	84.0	12.0	36.0	12.0	12.0
19_Austen_Pride_and_Prejudice	19thC	336.8	205.3	262.8	221.8	427.1	57.5	16.4	24.6	369.6	8.2	0.0	123.2	0.0	8.2	0.0	32.9	24.6	0.0	8.2	0.0	24.6	0.0	8.2
19_Austen_Sense_and_Sensibility	19thC	334.6	359.7	309.5	209.1	468.4	142.2	435.0	150.6	1062.3	25.1	16.7	158.9	33.5	25.1	25.1	25.1	25.1	16.7	16.7	33.5	8.4	0.0	0.0

[*2] http://mallet.cs.umass.edu/index.php

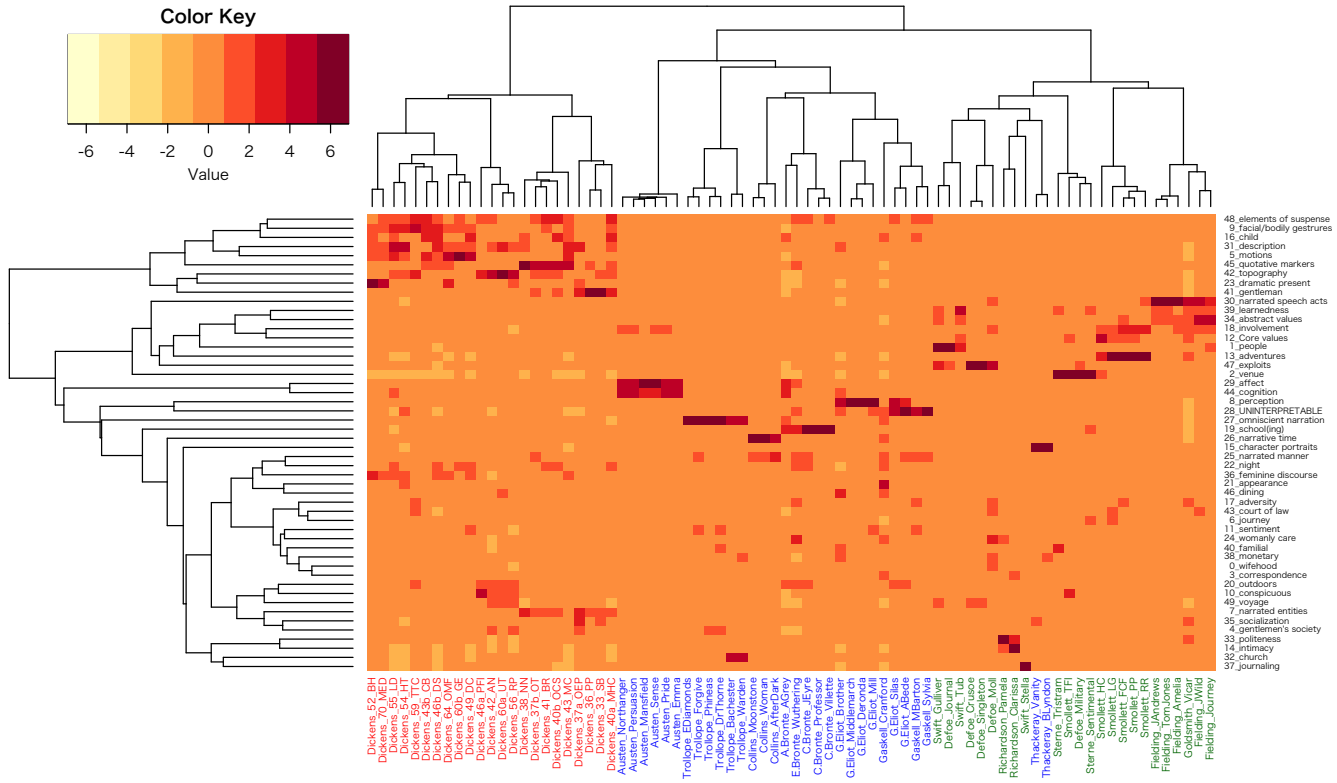

Fig. 4 Heatmap visualization of 50 topics across 78 texts (with mean weights scaled)

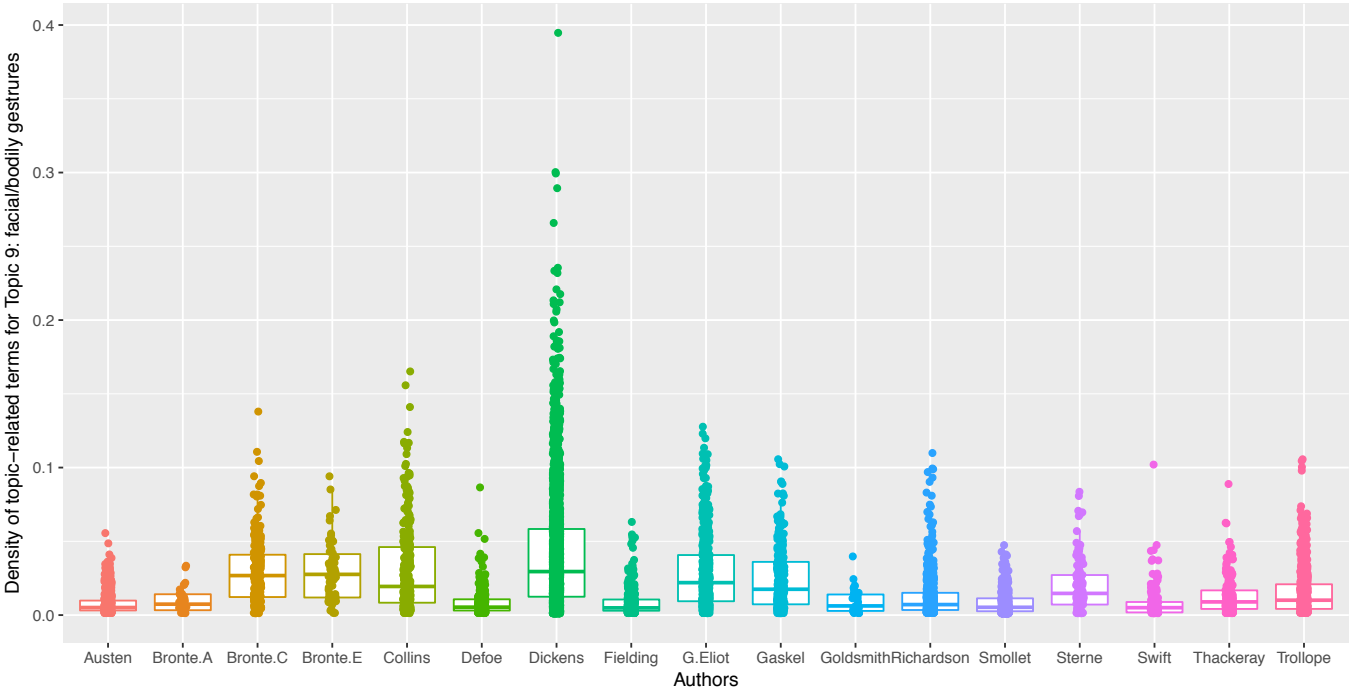

Fig. 5 Topic density distribution across the authors: Topic 9 facial/bodily gestures)