The University of Osaka
Institutional Knowledge Archive

Title	LDAトピックモデルによるIPAテキスト分析の試み ：アルフレッド・テニスンの韻文を用いて
Author(s)	藤田, 郁
Citation	言語文化共同研究プロジェクト. 2022, 2021, p. 15-38
Version Type	VoR
URL	https://doi.org/10.18910/88358
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

LDA トピックモデルによる IPA テキスト分析の試み
—アルフレッド・テニスンの韻文を用いて—

藤田　郁

大阪大学言語文化研究科
〒560-0043 豊中市待兼山町 1–8
Email: u256780k@ecs.osaka-u.ac.jp

概要　本稿は，Latent Dirichlet Allocation (LDA) トピックモデリングが詩の韻律の分析にも応用できるかどうかを検証することを目的としている。LDA はテキストに隠された意味構造を検出する文体分析手法であり，現在では主に散文テキスト分析へ応用する研究が多い。LDA を用いた詩に関する研究は比較的少ない中で，Navarro-Colorado（2018）は，スペイン語で書かれた詩においても LDA で韻を検出できる可能性を示唆した。しかし，英語では，綴り字と発音・音は必ずしも一致せず，また音節や音素の単位で押韻される例も数多く見受けられる。そのため，綴り字のままの英文テキストに LDA を適用することは，テキスト中の韻を検出するための最適な方法とはいえないと考えた。本研究では，英語の綴り字と音の不一致を考慮し，国際音声記号 (IPA) で書かれたテキストと，IPA で書かれた語を母音と子音に分割したテキストの 2 種類のテキストを分析することにより，IPA テキストへの LDA 適用の可能性，及び IPA テキストや音単位のテキストを分析する際の問題点を提示する。

キーワード　トピックモデル，LDA，国際音声字母，アルフレッド・テニスン，詩

An Exploratory Investigation of IPA Texts with LDA:
Using Alfred Tennyson's Poems

Iku Fujita

Graduate School of Language and Culture, Osaka University
1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract　This paper aims to investigate whether Latent Dirichlet Allocation (LDA) topic modelling can also detect the rhymes in poetry. LDA is a stylometric method that detects hidden semantic structures in texts and is now used primarily to examine prose texts. Among a relatively small number of studies on poems using LDA, Navarro-Colorado (2018) suggested that LDA can also detect rhymes in Spanish-written poetry corpus. However, speaking of the English language, the spellings and the pronunciations/sounds do not always work correspond. Additionally, numerous rhymes can be found structured by the units of syllables or phonemes. Thus, applying LDA on raw English texts may not be the optimised method for detecting rhymes in texts. Considering the dissidence between spelling and sound of English, this study analyses two text-types, written in International Phonetic Alphabets (IPA) and a text of which all words in IPA are divided into vowels and consonants. Through the analysis and discussion

of the emerging results of LDA, this paper provides the possibility of employing LDA on IPA texts as well as issues of analysing IPA and more small-unit texts.

Keywords Topic Model, latent Dirichlet allocation, International Phonetic Alphabets, Alfred Tennyson, Poetry

1. はじめに

現在，デジタル人文学・計量的手法は人文学研究，文学研究及びイギリス文学研究に大きく貢献している。文学研究における研究対象はもっぱら散文作品が多いが，韻文作品も例外ではなく，この手法は現在さまざまな側面から詩の研究に新たな知見をもたらしている (Tartakovsky and Shen, 2019; Nakao and Jimura, 2016 等)。

計量的手法の中でも，トピックモデルは，テキストマイニングの分野において将来性のある手法であるとされている (Meeks & Weingart, 2012)。本論で用いるトピックモデリング手法 latent Dirichlet allocation (Blei et al.; 2003; 以下 LDA) は，特に多くの散文作品研究に用いられている (Tabata, 2017; Kuroda, 2017; Kiyama, 2018; Huang, 2020a, b)。しかし，トピックモデルの韻文作品への応用は，Rhody (2012)，Navarro-Colorado (2018)，Henrichs (2019)，Okabe (2019) 等を除き，散文作品と比べ未だ少なく，また LDA を国際音声字母 (International Phonetic Alphabets: 以下 IPA) の分析に応用した例は，筆者の知る限りではまだなされていない。

LDA とは，テキストコーパスにおける隠れた意味構造を明らかにする機械学習手法であり，テキスト内での（語の）共起傾向を基に，語をグループ分けする教師なし手法であり (Tabata, 2017: 52), 意味を持つ最小単位である語を分析するために通例用いられる。筆者もこれまでの研究では，引喩の中にある Tennyson 作品の特徴を明らかにすべく，語に着目した分析を行なっている。スペイン語で書かれたソネット[1]を LDA を用いて分析と考察を行なった Navarro-Colorado (2018) は，以下のように述べ，LDA がソネットに適用された場合に押韻する語も検出できることを示唆している。

> [T]he LDA-topic is used as a poetic motif, considering motif here as a set of words that usually appear together. Two kinds of motifs have been defined: phonetic motifs and semantic motifs. Phonetic or “rhyme” motifs include Topics whose keywords have similar soundS [sic], that is, they rhyme.
>
> Navarro-Colorado (2018: 11)

しかし，押韻は語単位で行われるよりもむしろ語よりも小さい音節あるいは音素単位で行われるものであり，加えて英語の発音は綴り字と一致しないものも多く，綴り字は同じでも同じ発音となるとは限らない (例: -oo-, c**oo**k /ʊ/; ch**oo**se /uː/)。また，「韻を踏む」という条件は，同じ語を使用する必要はなく，語やその意味，文法的機能などに関係なく，以下 (1) の例のように，母音や子音が同じ・似ている音で構成されている語を使用することも多い。英語詩において押韻が行

[1] 十四行詩とも呼ばれる。14 行から成る定型詩。

われる単位（音節や音素）と LDA が分析する単位（語）が異なっている以上， Navarro-Colorado (2018) の押韻する語も検出できるという指摘は，少なくとも英語詩には当てはまらない可能性があると考える。LDA の結果が押韻を検知したように見えたとしても，それは副産物としての結果であり，語単位での LDA を用いた分析は，押韻を検知させることを主目的にすることは妥当ではないと考える。

(1) () 内は IPA 表記，[] 内は押韻の種類を表示。（太字下線は筆者による）

THEY rose to where their sovran eagle **sails**,	(/s**eɪlz**/)	[1]
They kept their faith, their freedom, on the **height**,	(/h**aɪt**/)	[2]
Chaste, frugal, savage, arm'd by day and **night**	(/n**aɪt**/)	[2]
Against the Turk; whose inroad nowhere **scales**	(/sk**eɪlz**/)	[1]
Their headlong passes, but his footstep **fails**,	(/f**eɪlz**/)	[1]
And red with blood the Crescent reels from **fight**	(/f**aɪt**/)	[2]
Before their dauntless hundreds, in prone **flight**	(/fl**aɪt**/)	[2]
By thousands down the crags and thro' the **vales**.	(/v**eɪlz**/)	[1]
O smallest among peoples! rough rock-**throne**	(/θr**əʊn**/)	[3]
Of Freedom! warriors beating back the **swarm**	(/sw**ɔːm**/)	[4]
Of Turkish Islam for five hundred **years**,	(/j**ɪəz**/)	[5]
Great Tsernogora! never since thine **own**	(/**əʊn**/)	[3]
Black ridges drew the cloud and brake the **storm**	(/st**ɔːm**/)	[4]
Has breathed a race of mightier mountai**neers**.	(/n**ɪəz**/)	[5]

“Montenegro” (1877)

押韻している箇所をグルーピングする目的に特化するのであれば，国際音声字母 (International Phonetic Alphabets; 以下 IPA) に変換したものを用いることで，音を統一された基準で視覚的に捉えられる「形」で示すことができると同時に，homophone (異形同音異義語; read と reed, right と write, I と eye 等) のように音が同じにもかかわらず綴り字の違いにより別のグループに分類される問題を解決でき，トピックモデルの（特に）韻文作品への応用可能性が広がるのではないかと考えた。一方で，トピックモデルは先にも述べたように意味を持つ最小単位の語を分析の対象とし，コーパス内の文書に跨る意味的つながりを見つけることができるアルゴリズムである。ここでの「意味」とは，語の持つ辞書的な「意味」ではなく，またアルゴリズムの内部にセマンティックな辞書を有しているわけでもいため，IPA で書かれたテキストを LDA で分析することにより，見出す「意味」を「押韻（要素）」とすることは可能だと考える。しかし前述の通り，LDA を IPA テキストに応用した例はまだなく，意味を持たない単位である音節や音素に焦点を当てた場合にトピックモデルが適切な分析手法となり得るか，また音の背後にある要素＝韻を検知できるか否かを検討する必要がある。本稿では，まず raw テキストと IPA テキストを比較す

ることで，IPA に変換したのみのテキストに LDA を実行することによってどのような違いが得られるか検討し，更に IPA テキストを子音と母音とで分割した場合に得られる結果を検討する。

2. データ及び方法

2. 1. データ

Delphi Poet Series 20, *Complete Works of Alfred*, *Lord Tennyson* に収録されている韻文作品から，作品毎のトークン数が 100〜199 の作品 116 編を本稿の分析対象とする。トークン数が 100〜199 である作品群は，Tennyson の作品のおよそ 4 分の 1 を占め，最も作品数が多い群である。加えて，本作品群にはソネットをはじめ，(脚) 韻を踏んでいる作品が多く，IPA テキストの分析結果を LDA がどのように示すか, 押韻等の要素を検知できているか否かを検証するには, 押韻の (少) ない自由詩等を用いるよりも適切であると考えた。116 編の作品は，本文のみがテキストファイルに収められており，テキストファイルのタイトルに短縮した作品の題名と発表年が付してある。これらのファイルを一括して raw テキストとし，IPA 表記ファイル作成の元データとするとともに，分析結果の比較対象として用いる。

IPA 表記への変換は，オンラインで使用可能である toPhonetics (https://tophonetics.com) というコンバータを利用した。IPA の表記は，イギリス英語とアメリカ英語が選択できるほか，IPA に変換された文字列とともに変換前の英文を併記させることも可能であり，また，and の発音における /ænd/ と /ənd/ に見られるアッシュと (æ) シュワ (ə) の差のような強形と弱形表記を選択することができる。本稿で用いるデータは，全てイギリス英語，IPA テキストのみ表示，強形の設定で変換した。尚，コンバータの辞書にない単語は， IPA に変換されることなく入力されたままの文字列で出力される。この変換されなかった文字列は，本稿の分析では英語表記のままで分析にかけることとした。上記の手順を踏まえ，ファイルごとに変換作業を行い，出力された結果を新たなテキストファイル (raw ファイルと同様にファイル名を付与済み) に書き込んだものを IPA テキストとして分析対象データとした。

3 節で考察を行うデータは，先に用意した IPA テキストの内容を子音と母音で分割した IPA sep (arated) データである。テキストエディタに正規表現を入力し，連続する子音 (例：/sk/ (/ɑːsk/)) ／母音字 (例：/əʊ/ (/səʊbə/)) は分割せず，あくまで子音と母音が連続している箇所を分割するよう設定した。そのため，この分割は音節や音素と同等の音の単位を示していない。これら分割した IPA 表記は，本項では「音」あるいは「（音の）要素」として扱う。また，本稿では半母音 /j/ は子音として扱っている。分割に際して，IPA テキストに使用されていた強勢記号をはじめ記号類は，母音の長音記号のみを残して削除した。また，分割する際にその音 (群) が語頭にあったか，語中，語末にあったかを示すために，ギリシャ文字 ϕ (phi; ファイ) を用い，語頭の場合には文字の後方のみ，語中の場合には文字列の前後に，語末の場合は文字の前方にのみ ϕ を付与した。ϕ を用いたのは，このマーカーをハイフンなどの記号にしてしまうと，LDA を実行する際に語の一部として認識されず，自動的に半角スペースに置換されてしまうため，英語及び英語の IPA 発音表記では用いられず，また視覚的にも判別しやすい文字を付与するためである。以

下 (2) に，各テキスト内のテキストデータの一部を例として示す。

(2)

raw テキスト：　DEAR, near and true — no truer Time himself

IPA テキスト：　dɪə, nɪər ænd truː — nəʊ ˈtruːə taɪm hɪmˈsɛlf

IPA sep テキスト：　dφ φɪə nφ φɪəφ φr æφ φnd trφ φuː　nφ φəʊ trφ φuːə tφ φaɪφ φm hφ φɪφ φmsφ φεφ φlf

“A Dedication” (1862: l. 1)

各ファイルのトークン数は以下表1の通りである。尚，アポストロフィ (’) 及びハイフン (-) では語を分けないものとした。IPA sep テキストでは，子音と母音を分割した後のスペースで区切られている文字列 (例：φɪəφ) を一語としてカウントしている。Raw テキストと IPA テキストにおける異なり語数の差は30語となっているが，これは同音異綴りのタイプ違い (例：our, hour; /aʊə/) が90例，異音同綴り (例：after; /ɑːftə(r)/) が60例あり，その差分が異なり語数の差に反映されている。なお，同音異綴りには，いわゆる同音異義語 (例：made, maid; /meɪd/) のほか，所有格 –’s と複数形 –s (例：heart’s, hears; /hɑːts/)，過去形・過去分詞 –ed と –’d (例：looked, look’d; /lʊkt/) の接尾辞による違いがみられた。異音同綴りでは，冠詞の a/the の強形 (/eɪ /, /ði/) 弱形 (/ə/, /ðə/) の違い以外は全て語末の /r/ 音の有無によるものであった。

表 1　使用したデータ毎のトークン数と異なり語数

	raw テキスト	IPA テキスト	IPA sep テキスト
総トークン	15,528	15,528	45,065
異なり語数	3,651	3,621	718

2. 2.　データの実行及びトピック数の決定方法

LDA を実行するにあたり，MALLET 2.0.8 をインストールし，分析を行う。LDA 実行の際に任意のストップワードを指定することができるが，本分析では一切の語を削除せずに分析を行なった。また，LDA では，研究者（実行者）が予めトピックの数を設定する仕様となっている。本稿では，可能な限り恣意性を排除することを目的として，perplexity (モデルの予測精度の評価指標) と coherence (トピックの品質を測る評価指標) をテキスト種類毎に計算した結果から，適切なトピック数を 10 種類まで絞り込み，LDA を実行して得られる α 値（以下 3 節で詳述）と分析結果を比較検討し，トピック数を決定した。分析及び考察に用いる最適なトピックの数は，raw テキストと IPA テキストは 10，IPA sep テキストは 17 をそれぞれ最適なトピック数とした。

3.　分析及び考察

本節では，raw テキスト, IPA テキスト，IPA sep テキストの 3 テキストに対し，それぞれ LDA を実行し，得られた結果を提示する。

3. 1.　Raw テキストと IPA テキスト

3.1.1　Raw テキスト

本節ではまず，raw テキストの結果を提示する。以下表 2 に示すのは，raw テキストへの LDA 実行結果のうち，各トピックのα値と各トピックにおける最も顕著なキーワード (Keys) 上位 20 語である。トピックは表の左から右に向かいα値の降順で並んでおり, Keys は上から下に向かい，顕著度合いが高いものから低いものへ, 降順で並んでいる。例えば, 最もα値が高いトピックは，本出力結果では Topic 6 (α ＝1.88294) であり，最もα値が低いトピックは Topic 3 (α ＝0.07916) である。また， Topic 6 において，最も顕著な語の上位 1 位は the であり，20 位の語は as である。後の 3.1.2 節及び 3.2 節の表 4, 6 についても同様の表示をする。

表 2　LDA による各トピックのα値及び上位 20 語（raw テキスト）

Topics	Topic 6	Topic 4	Topic 5	Topic 2	Topic 1	Topic 7	Topic 0	Topic 8	Topic 9	Topic 3
α value	1.88294	0.21179	0.16916	0.16534	0.16431	0.14020	0.11504	0.09728	0.09695	0.07916
Keys (1–20)	the	i	thy	her	you	we	rosalind	our	riflemen	o'
	and	my	thou	she	your	our	valley	england	form	each
	of	me	thee	on	i	its	nothing	land	far	both
	to	thro'	oh	light	it	ye	falls	lord	storm	they
	in	brother	thine	sleeps	my	why	stream	their	your	together
	a	only	art	scorn	men	must	through	every	ready	side
	that	tho'	yet	form	again	ah	until	we	words	two
	with	eyes	heart	cold	me	gaze	along	freedom	he	other
	for	lilian	beware	saw	know	e'en	strange	farewell	nor	than
	is	hand	ere	ever	three	woe	full	god	city	beauty
	all	heart's	stay	nor	is	tear	living	through	sound	tears
	but	morning	wilt	had	room	tho'	aweary	part	against	look
	so	love	truth	with	hangs	us	bold	macready	do	an'
	from	with	year	rest	him	welaway	even	blood	trumpet	ralph
	not	thee	kiss	doth	little	grief	winter	thine	now	perfect
	his	mind	pass	year	should	bright	spring	thee	him	stars
	when	loved	canst	not	here	ecstasy	years	years	charm	over
	on	tis	sleep	down	read	soon	mourn	thought	thunders	a'
	be	dare	alas	hair	loved	joys	about	where'er	angel	luve
	as	sad	hope	after	leaf	glow	earth	french	speed	sighs

α値は，高ければ高いほど，そのトピックが該当コーパスに多く含まれていることを示す値である。一方でα値が非常に低い場合には，そのトピックがごく少数のテキストにしか現れず，場合によってはある 1 つのテキストにしか現れていないことが多い。現に，表 2 では，最もα値が高い Topic 6 の Keys は 20 語全て機能語であり，また表 4 に示す Raw テキストにおける高頻度語上位 20 位にそのほとんどが含まれており，テキスト内で広く使われている語であることが分かる。反対に，α値が最も低い Topic 3 には，特定の詩にしか使われていない固有名詞 (ralph) が

上位語に含まれており，このトピックが少数の詩にしか現れていないことが分かる。

3.1.2. IPA テキスト

次に，表 3 に示すのは IPA テキストに LDA を実行した出力結果，各トピックとそのα値，キーワード (Keys) 上位 20 語である。この IPA テキストのトピック数も，先の raw テキストと同じく 10 である。

表 3　Raw テキストにおける高頻度語上位 20 位

(下線は表 2 Topic6 の上位 20 語に含まれていることを示す)

Rank	Word	Freq	Rank	Word	Freq
1	the	993	11	her	134
2	and	615	12	s	127
3	of	451	13	is	120
4	to	256	14	for	110
5	in	247	15	my	109
6	a	233	16	all	103
7	that	196	17	thy	101
8	i	174	18	but	96
9	with	165	19	not	94
10	d	145	20	from	91
			20	so	91

表 4　LDA による各トピックのα値及び各トピック上位 20 語（IPA テキスト）

Topics	Topic 3	Topic 8	Topic 4	Topic 9	Topic 0	Topic 7	Topic 1	Topic 6	Topic 5	Topic 2
α value	1.94883	0.31232	0.19183	0.16167	0.14637	0.13617	0.13258	0.10989	0.10312	0.06632
Keys (1–20)	ðə	aɪ	ðaɪ	hɜ:	wi:	wɪl	fɑ:	ju:	lɔ:d	bi:
	ænd	maɪ	ðaʊ	ʃi:	ˈaʊə	rosalind	əʊ	jɔ:	gɒd	ðeɪ
	ɒv	mi:	ði:	hɜ:r	ɪts	wɪnd	ən	əʊld	ðeə	riflemen
	tu:	ɪt	ðaɪn	nɒt	waɪ	ˈi:vən	hi:	hɪm	bɪˈfɔ:	fɔ:m
	ɪn	səʊ	nɒt	ɒn	mʌst	stri:m	hɪm	hæd	baʊnd	ɑ:
	ə	ə	θru:	fɔ:m	wəʊ	daʊn	ðæn	gru:	stɑ:z	stɔ:m
	ðæt	əˈgɛn	ɑ:t	sli:ps	ɒft	ʧeɪndʒ	lʊk	jɔ:r	ænd	ðeər
	wɪð	əʊ	jɛt	lilian	geɪz	ˈvæli	saɪz	hɜ:d	blʌd	ˈrɛdi
	ɪz	lʌvd	wi:	nɔ:	deɪ	ˈfri:dəm	bɪˈni:θ	fli:t	rælf	laɪk
	ɔ:l	nju:	nɔ:	ɑ:	sti:l	əˈlɒŋ	əˈweɪ	pɑ:st	kɪŋ	hɑ:ts
	frɒm	ˈbrʌðə	ˈaʊə	hɑ:t	gləʊ	lænd	hæŋz	bɪˈfɔ:	tɔ:k	saɪd
	bʌt	wi:p	bɪˈweə	kæn	həʊp	ˈnʌθɪŋ	ˈhɛvɪli	bɪˈkɒz	streɪm	wi:
	wɛn	ˈlɪtl	lʌv	laɪt	haʊ	fɔ:lz	ˈswi:tə	trʌst	sɛd	deɪ
	ɒn	ju:	sɪns	əˈbʌv	bʌt	vɔɪs	əˈbaʊt	lʊkt	ðeəz	təˈgɛðə
	hɪz	ˈbju:ti	ˈfeəˈwɛl	ɪtˈsɛlf	ˈɛkstəsi	sæt	eɪ	mi:	bi:st	ˈɪŋgləɪnd
	fɔ:	mɔ:	steɪ	ˈvɜ:tju:	bɪˈfɔ:r	ʧaɪld	luve	hu:m	ðɛm	welaway
	æz	ˈəʊnli	hɑ:t	fi:t	gri:f	streɪndʒ	wɜ:dz	leɪt	greɪt	rɪˈvjuəz
	nəʊ	hænd	aɪ	ˈpɜ:fɪkt	teə	ˈmɪdl	teəz	gɛst	hu:	ˈɪndələnt
	ðeə	sʌʧ	macready	hɑ:f	ji:	əˈwɪəri	saʊnd	əˈfɑ:	ˈɑ:ftər	ˈpi:pl
	baɪ	fi:l	vɔɪs	hænd	reɪz	freʃ	hu:z	raɪmz	ˈʃʊəli	frɛnʧ

表 2 に挙げた raw テキストの結果と比較すると，トピックの番号やα値は異なるものの，各トピックの上位語に共通している点が多く見られる。上位 20 語に最も共通する語が多いのは，raw テキストの Topic 6 と IPA テキストの Topic 3 であり，20 語中 17 語が同じ語である。最も共通する語が少ないトピックは，Topic 1 (raw) と Topic 6 (IPA)，Topic 8 (raw) と Topic 5 (IPA) の 4 語である。全体では，各テキスト全 200 キーワードのうち，重複して現れた語は，その 6 割にあたる 120 語であった。

Raw テキストの Topic 4 と IPA テキストの Topic 8 は，共にキーワードの上位 3 語が一人称単数代名詞の主格 I，所有格 my，目的格 me となっているほか，Topic 5 (raw) と Topic 4 (IPA) では二人称単数代名詞及び二人称所有代名詞 thy, thou, thee, thine, Topic 2 (raw) と Topic 9 (IPA) では三人称単数女性代名詞 her, she, そして Topic1 (raw) と Topic6 (IPA) では 二人称代名詞の you, your がトピック毎に分類されている。Tennyson の作品を LDA で分析した際に，人称代名詞が各トピックに人称や単複，性差によって分かれ，作品毎の登場人物や語り手の違い，登場人物間の関係や作品のテーマ分類される点は，Fujita (2021) で得られた示唆と共通する点である。

以下図 1，2 に示すのは，LDA の分析から得られる構成 (composition) データを用いて描画したヒートマップである。このヒートマップは，どのトピック（縦軸）がどの作品（横軸）に顕著に現れているかを色の濃淡で示すものであり，セルが濃い黒であればあるほど，そのトピックはセルが交差する軸に位置する作品に顕著なトピックであると言える。このヒートマップから，例えば共通して一人称単数代名詞がキーワードに含まれている Topic 4 (Raw) と Topic 8 (IPA) が顕著に現れる，セルが他のセルより濃い黒で示されている作品を比較する。Topic 4 (Raw) は，主に 24 作品に顕著に現れるトピックであり，Topic 8 (IPA) は 31 作品に顕著に現れていることが分かる。この二つのトピックが顕著である作品数は異なっているものの，それぞれに共通する作品は，以下表 5 に示す Topic 4 (Raw) と Topic 8 (IPA) に顕著な作品のリストにおいて網掛け太字で示している 19 作品である。

表 5　Topic 4 (Raw) と Topic 8 (IPA) に顕著な作品のリスト

Raw: Topic 4			
1	**Ancrntc31**	13	**MyRslnd32**
2	**ANDSKYW27**	14	PrftryP80
3	**APLLNSR27**	15	**Sonnt-Bu32**
4	**ASSTRSW27**	16	**SonntCl30**
5	**Bynvltn89**	17	**Sonnthb32**
6	Crssngt89	18	THBRDSF27
7	**DobtndP92**	19	**THDYNGC27**
8	**Herftnw50**	20	**Thrrthr32**
9	**Liln30**	21	**THSLGHT27**
10	LovndSr30	22	**ThSlrBy62**
11	Memywnf32	23	**TISTHVC27**
12	**MYBRTHR27**	24	**ToLnhst53**

IPA: Topic 8					
1	**Ancrntc31**	13	OHWRTHS27	25	ThSptfl68
2	**ANDSKYW27**	14	OLDG89	26	ThThrst89
3	**APLLNSR27**	15	Songvry30	27	**TISTHVC27**
4	**ASSTRSW27**	16	**Sonnt-Bu32**	28	To32
5	**Bynvltn89**	17	**SonntCl30**	29	**ToLnhst53**
6	**DobtndP92**	18	**Sonnthb32**	30	Toswhnw32
7	**Herftnw50**	19	ThBggrM42	31	YESTHRB27
8	InthGrd74	20	**THDYNGC27**		
9	**Liln30**	21	THGRVFS27		
10	**MYBRTHR27**	22	**Thrrthr32**		
11	**MyRslnd32**	23	**THSLGHT27**		
12	ODrlngR32	24	**ThSlrBy62**		

以上のように，raw テキスト及び IPA テキストへの LDA 実行結果は比較的類似している。ま

た，Topic 2 (IPA) のキーワードに含まれる /fɔːm/ (form) と /stɔːm/ (storm) は，押韻を LDA が検知した例のようにみえる。この 2 語は，raw テキストの Topic 9 の上位 2 語でもある。Topic 2 (IPA) が顕著に現れる 4 作品，Miltn-H62, Dulsms30, RiflmnF92, NatnlSn30 のうち，単語が使用されているのは以下 (3) に挙げる RiflmnF92 (“Riflemen Form”, 1892) のみである。作品原文をみると，確かに storm と form は脚韻である。しかし，storm と form 以外にも韻を踏んでいる語，音が複数確認できるが，これらの語は LDA の結果には反映されておらず，押韻要素が検出されていると断定できるものではない。これらのことから，IPA テキストにおいても raw テキストにおいても語の持つ意味とその共起傾向から，トピックが分類されていることが分かる。加えて，raw, IPA 各テキストの総語数 15,528 に対して，述べ異なり語数の差は 150 であり，その量，また種類を考慮すると，raw テキストと IPA テキストとでは，大きな差があるとは言えず，本稿で扱うデータにおいては LDA を実行するテキストを IPA 表記にしただけでは，音韻要素を LDA が検知できるか否かは不明である。そこで，IPA テキストを母音字と子音字に分割したデータ (IPA sep テキスト) を用いて再度 LDA を実行した結果を，次の 3.2 節で提示する。

(3) () 内は (行末 1 語の IPA 表記; 押韻している箇所の IPA 表記)

THERE is a sound of thunder afar,
Storm in the South that darkens the *day*! (/deɪ/; /eɪ/)
Storm of battle and thunder of <u>war</u>! (/wɔː/; /ɔː/)
Well if it do not roll our *way*. (/weɪ/; /eɪ/)
Storm, Storm, Riflemen **form**! (/fɔːm/; /ɔː(m)/)
Ready, be ready against the **storm**! (/stɔːm/; /ɔː(m)/)
Riflemen, Riflemen, Riflemen **form**! (/fɔːm/; /ɔː(m)/)
Be not deaf to the sound that <u>warns</u>, (/wɔːnz/; /ɔːnz/)
Be not gull’d by a despot’s *plea*! (/pliː/; /iː/)
Are figs of thistles? or grapes of <u>thorns</u>? (/θɔːnz/; /ɔːnz/)
How can a despot feel with the *Free*? (/friː/; /iː/)

“Riflemen Form” (1892; ll. 1–11; 飾り字は筆者による)

3.2. IPA sep テキスト

本節では，IPA sep テキストの結果を提示する。IPA sep テキストに LDA を実行する際のトピック数は，17 とした。以下表 6 に示すのは，各 17 トピックの α 値と各トピックにおける最も顕著なキーワード (Keys) 上位 20 語である。トピックは表の左上段から右下段へ α 値の降順で並んでおり，Keys は上から下に向かい，顕著度合いが高いものから低いものへ，降順で並んでいる。

表 6 において最も α 値が高いトピックは，Topic 3 (α = 6.10462) であり，最も α 値が低いトピックは Topic 6 (α = 0.09079) である。最も α 値が高く，複数の作品に現れる Topic 3 のキーワードをみると，raw テキストの高頻度語 (表 7) を構成する音が多く含まれていることが分かる。

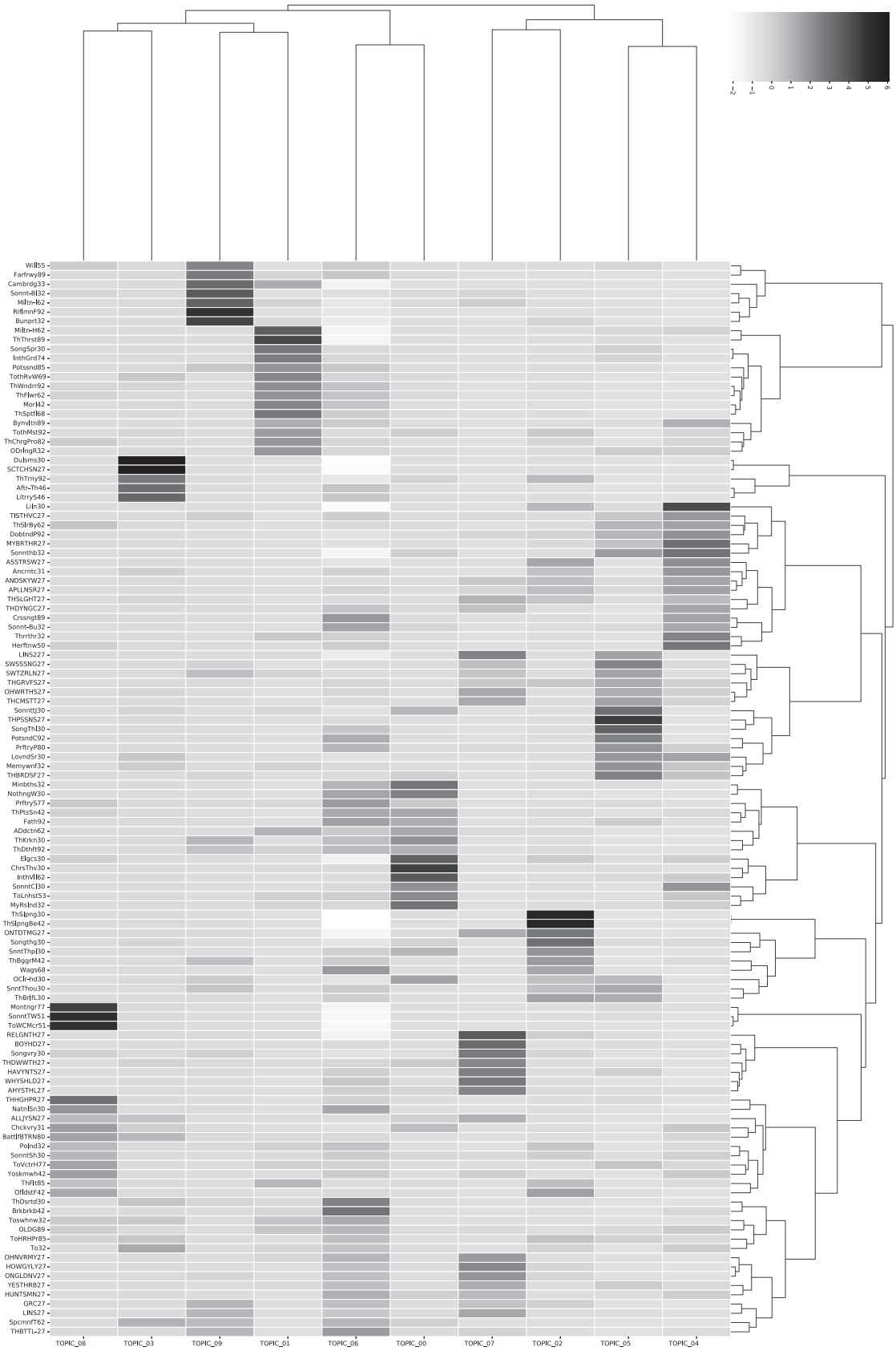


図 1 ヒートマップ（raw テキスト）

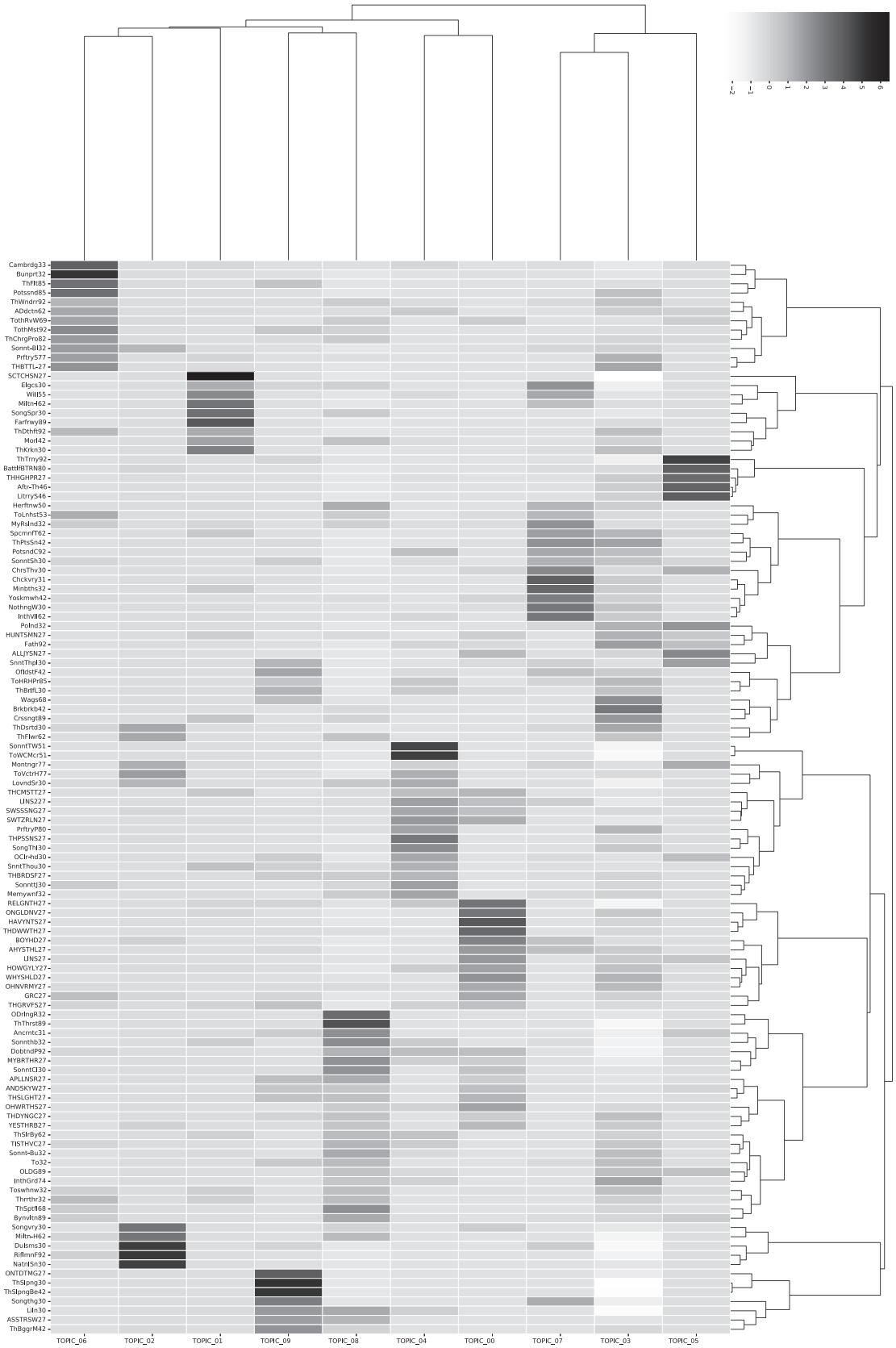


図 2 ヒートマップ（IPA テキスト）

しかし，高頻度語を構成する音の全てが Topic 3 に現れているわけではない。表 7 の頻度第 8 位にある i は，IPA 表記の場合は eye と同形となり，さらに頻度が増えるため，α値が高いトピックに分類されることを予想したが，前後に子音のつかない単独 aɪ は，α値が比較的低い Topic 11 にのみ現れている。このように，複数のトピックで，元の raw テキスト及び IPA テキストにおける語の頻度や，LDA の出力結果から推測できそうな分類が見られる一方で，IPA テキストに LDA を実行した際には見られなかった音のばらつきもある。

表 6　LDA による各トピックのα値及び各トピック上位 20 語（IPA sep テキスト）

Topics	Topic 3	Topic 10	Topic 14	Topic 12	Topic 4	Topic 13	Topic 2	Topic 5	Topic 15
Alpha	6.10462	3.61320	3.44713	2.93176	2.54886	1.86466	1.72640	1.42596	1.09778
Key (1–20)	φnd	φz	φɪφ	ðφ	φl	φəφ	ðφ	φɪφ	φɪφ
	æφ	ɪφ	wφ	φə	ðφ	φɪφ	φə	φlφ	wφ
	φn	lφ	φiːφ	φv	φaɪ	φrφ	bφ	φəʊφ	bφ
	φɛφ	φaɪφ	φz	ɒφ	φt	φs	φaɪφ	φi	φə
	φæφ	φd	sφ	φn	φæφ	φi	φeɪφ	φɒφ	φŋ
	φm	φəʊ	φt	φɒφ	φn	ɒφ	φeə	φfφ	φð
	φeɪφ	mφ	φəφ	φəʊφ	mφ	φv	φəʊφ	φntφ	nφ
	hφ	sφ	rφ	dφ	dφ	tφ	φk	φvφ	φlz
	φuː	φɑːφ	φŋ	frφ	kφ	φnφ	φæφ	ɔːφ	φwφ
	φɔːφ	hφ	φi	pφ	φɛφ	φuːφ	φɛφ	mφ	strφ
	fφ	φt	φʧ	φtφ	φs	vφ	φnz	spφ	φɜːφ
	φt	bφ	brφ	φld	φiː	glφ	φeɪ	φpφ	φvφ
	wφ	nφ	ɪφ	trφ	φd	φɪəφ	φaʊφ	flφ	θrφ
	φl	φʌφ	blφ	gφ	ʃφ	φtφ	φaɪə	ɑːφ	əφ
	φiː	φeɪ	φd	φmz	fφ	φzφ	φts	ʌφ	φɔːφ
	φʌφ	φiːφ	φnt	φmφ	φaʊ	φmφ	əʊə	φnd	φrφ
	φaʊφ	wφ	æφ	φr	φp	φns	klφ	φŋ	klφ
	φst	φf	φdφ	φɪφ	φʌφ	φkφ	plφ	φftφ	φɒφ
	φɔː	φs	φts	wφ	φuːφ	φɔɪφ	φmbφ	φaʊəφ	φɪəφ
	tφ	tφ	φð	ʤφ	nφ	φʃφ	eə	φtl	φndʒ

Topics	Topic 8	Topic 11	Topic 16	Topic 0	Topic 9	Topic 7	Topic 1	Topic 6
Alpha	0.80654	0.75588	0.63136	0.47514	0.45539	0.27874	0.18092	0.09079
Key (1–20)	hφ	aɪ	φʌφ	ðφ	jφ	fφ	φeφ	lφ
	φɜː	mφ	sφ	φə	φəφ	φɑː	φiφ	φiː
	φd	tφ	φɛφ	jφ	φɔː	φθ	φɔːφ	φeəφ
	ʃφ	ə	φðφ	φz	φr	əφ	φm	φk
	φɜːφ	φuː	φgφ	əʊ	əʊφ	ɑː	fφ	φr
	φɪφ	φaɪ	φəφ	φeəφ	φts	iːφ	rφ	φrφ
	lφ	ɛφ	φŋ	φaɪφ	φdφ	φɔɪ	φth	φŋglφ
	φɑːφ	φeəφ	əφ	φvφ	aɪφ	glφ	φi	ɑː
	φeɪφ	əʊ	ʌφ	φɛφ	φiːφ	φeɪ	bφ	φnʧ
	φps	φft	lφ	φŋk	φnz	φaɪəφ	φmφ	φiaφ
	φθ	φwφ	φuː	tφ	φʤφ	φsφ	φaφ	frφ
	φð	φvd	φr	φaʊə	φnʤφ	φɪəφ	φflφ	φlnφ
	φbφ	φsφ	φnlφ	φʊφ	eɪφ	φmbφ	stφ	eɪ
	φlf	φnd	φɪə	φuφ	φuəφ	φvnφ	φdφ	φmpl
	φld	lφ	njφ	smφ	φðd	welaway	gφ	φʊ
	grφ	kφ	φkt	φe	φæφ	φdnz	φnz	φgmφ
	aɪφ	φnsφ	ɑː	swφ	φtl	əʊφ	rosalind	φvl
	φeφ	φvlφ	φdn	φae	əʊəφ	ə	φspφ	φʃ
	drφ	aɪφ	φmφ	φskφ	φspφ	frφ	φshφ	φaɪφ
	φnfφ	φk	skφ	φltφ	catullus	φʊ	φtl	prythee

それぞれのトピックがどのような「意味」で分類されているのかを検討するために，各トピックに母音，子音の語頭音 (Xφ)，語中音 (φXφ)，語末音 (φX) がそれぞれ何音含まれているかを数え

たものを表 8 に示す。なお，語中音のみ，一語に複数音現れる可能性がある。表 8 をみると，いくつかのトピックの特徴が見て取れる。Topic 3, 10, 4 のキーワードには，語中子音 (C-Middle) が現れておらず，子音は語頭音 (C-Head) と語末音 (C-Tail) が多く現れており，語中子音が 0 である代わりに語中母音 (V-Middle) が複数現れている。このことから，この三つのトピックは CVC (子音-母音-子音)，VC，CV で構成される語の要素を多く含むトピックとみることができる。

表 7　raw テキストの単語頻度上位 20 位と単語を構成する音

Rank	Word	Sound	Word FREQ	Rank	Word	Sound	Word FREQ
1	the	ðφ φə (941) / ðφ φi (52)	993	12	s	φz (877) / φs (372)	127
2	and	æφ φnd	615	13	is	ɪφ φz	120
3	of	ɒφ φv	451	14	for	fφ φɔ: (79) / fφ φɔ:φ φr (31)	110
4	to	tφ φu: (289)	256	15	my	mφ φaɪ	109
5	in	ɪφ φn	247	16	all	ɔ:φ φl	103
6	a	ə (229) / eɪ (4)	233	17	thy	ðφ φaɪ	101
7	that	ðφ φæφ φt	196	18	but	bφ φʌφ φt	96
8	i	aɪ	174	19	not	nφ φɒφ φt	94
9	with	wφ φɪφ φð	165	20	from	fφ φrφ φɒφ φm	91
10	d	φd (632) / φt (1098)	145	20	so	sφ φəʊ	91
11	her	hφ φɜ: (115) / hφ φɜ:φ φr (19)	134				

* 下線は Topic 3 のキーワードに現れている音

** Sound 欄の () 内の数字は IPA sep テキストにおいて対応する音の頻度

表 8　IPA sep テキストの LDA によるトピックごとの音数比較（α値降順）

Topic	Alpha	V-Head	V-Middle	V-Tail	C-Head	C-Middle	C-Tail
3	6.10462	1	6	3	4	0	6
10	3.61320	2	3	2	8	0	5
14	3.44713	2	3	1	5	1	8
12	2.93176	1	3	1	8	2	5
4	2.54886	0	4	3	7	0	6
13	1.86466	1	5	1	3	7	3
2	1.72640	0	6	4	4	1	3
5	1.42596	3	4	1	3	6	3
15	1.09778	1	5	1	6	3	4
8	0.80654	1	5	1	5	2	6
11	0.75588	2	1	2	3	5	4
16	0.63136	2	3	2	4	4	4
0	0.47514	0	5	4	5	3	2
9	0.45539	4	4	1	1	4	5
7	0.27874	3	2	4	3	3	2
1	0.18092	0	4	1	5	5	3
6	0.09079	0	3	2	2	4	6

* V：母音，C：子音，-Head：語頭音，-Middle：語中音，-Tail：語末音

図 3　ヒートマップ（IPA sep テキスト）

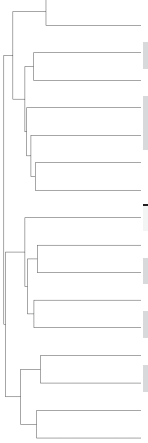
Topic 0,1,6 は語頭母音 (V-Head) が 0 であるものの，それ以外の音は全て現れているため，CVC(V) のように語頭が子音で始まり，後続する母音あるいは子音が繰り返される語を構成する音を多く含んでいるトピックと言える。また，17 トピックのうち 15 トピックにおいて，語中の母音は 3 つ以上出現しているが，Topic 11 と Topic 7 にはそれぞれ φeəφ と φaɪəφ, φɪəφ いずれ

も二重母音しか現れておらず，これらの二重母音と子音との間の関係によってトピックが分類されているとも考えられる。

以下に示す図 3 は，IPA sep テキストの LDA 実行結果のヒートマップである。これまでのヒートマップと同様，より黒いセルは，そのトピックが作品に顕著に現れることを示している。図 3 の上部にあるデンドログラムは，各トピックを分類している。このデンドログラムに従い，並べ替えた各トピックに語のどの位置にある音がいくつ含まれているかを数えたものが表 9 である。

図3及び表9のトピックを分類するデンドログラムでは，17トピックは表9上部の Topic 0 から Topic 8 までと，下部の Topic 3 から Topic 16 までに大きく分けられる。下部に分類される9つのトピックのうち，α値 ≥1.0 は6トピックであるのに対し，上部の8トピック中，α値 ≥1.0 は3トピックであり，下部のクラスターがよりα値が高く，複数の作品に顕著に現れるトピックを有していると言える。この表9より，上部，下部の各クラスター中，語頭音・語中音・語末音のそれぞれが最も多くキーワードに出現しているトピック（網掛け部）は6つあるが，以下で上部の3トピック (Topic 4, 1, 12) を raw テキストと比較しながら考察する。

表 9　IPA sep テキストのトピック毎の音数比較（デンドログラム順）



Topic	Alpha	V-Head	V-Middle	V-Tail	C-Head	C-Middle	C-Tail	Head	Middle	Tail
0	0.47514	0	5	4	5	3	2	5	8	6
6	0.09079	0	3	2	2	4	6	2	7	8
4	2.54886	0	4	3	7	0	6	7	4	**9**
15	1.09778	1	5	1	6	3	4	7	8	5
1	0.18092	0	4	1	5	5	3	5	**9**	4
12	2.93176	1	3	1	8	2	5	**9**	5	6
7	0.27874	3	2	4	3	3	2	6	5	6
8	0.80654	1	5	1	5	2	6	6	7	7
3	6.10462	1	6	3	4	0	6	5	6	**9**
9	0.45539	4	4	1	1	4	5	5	8	6
10	3.61320	2	3	2	8	0	5	**10**	3	7
5	1.42596	3	4	1	3	6	3	6	10	4
13	1.86466	1	5	1	3	7	3	4	**12**	4
2	1.72640	0	6	4	4	1	3	4	7	7
14	3.44713	2	3	1	5	1	8	7	4	**9**
11	0.75588	2	1	2	3	5	4	5	6	6
16	0.63136	2	3	2	4	4	4	6	7	6

Topic 12 が最も顕著に現れる作品は，Sonnt-Bl32 (“Sonnet [Blow ye the trumpet gather from afar]”, 1832) である。以下 (4) には IPA sep テキストの Sonnt-Bl32 に，Topic 12 のキーワードを網掛けしたもの，(5) には raw テキストを左側に，IPA テキストを右側に示し，中央には行番号を付したものを示す。まず，語頭の /ð/ 音は，定冠詞 the や指示代名詞 that, this など用途が広く，他の複数のトピックのキーワードにもなっているが，Sonnt-Bl32では，先にあげた the, this のほか，those，そして than が3回使われている。この3回のうち2回は行頭で ‘Than when’ と，when とともに使われており，語末音 /n/ もまた，この Topic 12 のキーワードである。(5) の10, 12, 14行目では，行末でこの語末音 /n/ で押韻している。他にキーワードに含まれている脚韻の要素には /əʊ/ と /ld/ があり，2–3行目，6–7行目で使用されている。語中音 /əʊ/ は3行目の中だけで3回使われている。3行目冒頭2語の二重母音 /aɪ/ と /eɪ/ は Topic 12 のキーワードにはなっていないが，

非円唇低母音－非円唇高母音 /aɪ/ から非円唇中母音－非円唇高母音 /eɪ/ そして非円唇と円唇の中間で発音される弱母音－円唇高母音 /əʊ/ へと，二重母音をたたみかけると同時に非円唇を円唇に，低母音から高母音へと調音位置を移動させることにより，音を後続の /əʊ/ と呼応させるだけではなく，躍動感を持たせる効果があると考えられる。

(4)

blφ φəʊ jφ φi: ðφ φə trφ φʌφ φmpφ φɪφ φt gφ φæφ φðφ φə frφ φɒφ φm əφ φfφ φɑ:
ðφ φə hφ φəʊφ φsts tφ φu: bφ φæφ φtl bφ φi: nφ φɒφ φt bφ φɔ:φ φt æφ φnd sφ φəʊφ φld
əφ φrφ φaɪφ φz brφ φeɪφ φv pφ φəʊφ φlz ðφ φə bφ φəʊφ φldφ φɪφ φst ɒφ φv ðφ φə bφ φəʊφ φld
brφ φeɪφ φk θrφ φu: jφ φɔ:φ φr aɪəφ φn ʃφ φæφ φklz flφ φɪφ φŋ ðφ φεφ φm fφ φɑ:
əʊ fφ φɔ: ðφ φəʊφ φz dφ φeɪφ φz ɒφ φv Pφ φiaφ φst eə ðφ φə zφ φɑ:
grφ φu: tφ φu: ðφ φɪφ φs strφ φεφ φŋθ əφ φmφ φʌφ φŋ hφ φɪφ φz dφ φεφ φzφ φəφ φts kφ φəʊφ φld
wφ φεφ φn i:φ φvφ φəφ φn tφ φu: mφ φɒφ φskφ φəʊφ φz kjφ φu:φ φpφ φəφ φlφ φəφ φz wφ φɜ: rφ φəʊφ φld
ðφ φə grφ φəʊɪφ φŋ mφ φɜ:φ φmφ φəφ φz ɒφ φv ðφ φə pφ φɒφ φlφ φɪφ φʃ wφ φɔ:
nφ φaʊ mφ φʌφ φst jφ φɔ: nφ φəʊφ φbl æφ φŋgφ φə blφ φeɪφ φz aʊφ φt mφ φɔ:
ðφ φæφ φn wφ φεφ φn frφ φɒφ φm Sobieski klφ φæφ φn bφ φaɪ klφ φæφ φn
ðφ φə mφ φɒφ φzlφ φεφ φm mφ φɪφ φrφ φɪəφ φdz fφ φεφ φl æφ φnd flφ φεφ φd bφ φɪφ φfφ φɔ:
ðφ φæφ φn wφ φεφ φn Zamoysky smφ φəʊφ φt ðφ φə tφ φɑ:φ φtφ φə kφ φɑ:φ φn
ðφ φæφ φn ɜ:φ φlφ φɪə wφ φεφ φn ɒφ φn ðφ φə bφ φɔ:φ φltφ φɪφ φk ʃφ φɔ:
Boleslas drφ φəʊφ φv ðφ φə pφ φɒφ φmφ φəφ φrφ φeɪφ φnjφ φəφ φn

(5)

Blow ye the trumpet, gather from afar	1	bləʊ ji: **ðə** ˈ**tr**ʌmpɪt, ˈ**gæðə fr**ɒm əˈfɑ:
The hosts to battle: be not bought and *sold*.	2	**ðə hə**ʊsts tu: ˈbætl: bi: nɒt bɔ:t ænd s**əʊld**.
Arise, *brave Poles*, the *boldest* of the *bold*;	3	əˈraɪz, breɪv **pəʊ**lz, **ðə** ˈb**əʊ**ldɪst **ɒ**v **ðə bəʊld**;
Break through your iron shackles fling them far.	4	breɪk θru: jɔ:r ˈaɪ**ən** ˈʃæklz flɪŋ **ð**ɛm fɑ:.
O for those days of Piast, ere the Czar	5	əʊ fɔ: **ðəʊz d**eɪz **ɒ**v **P**iast, eə **ðə** zɑ:
Grew to this strength among his deserts *cold*;	6	**gr**u: tu: **ð**ɪs strɛŋθ əˈmʌŋ hɪz ˈ**d**ɛzəts **kəʊld**;
When even to Moscow's cupolas were *rolled*	7	**wɛn** ˈi:v**ən** tu: ˈmɒsk**əʊ**z ˈkju:pələz **w**ɜ: **rəʊld**
The growing murmurs of the Polish war!	8	**ðə** ˈ**grəʊ**ɪŋ ˈmɜ:məz **ɒ**v **ðə** ˈ**p**ɒlɪʃ **w**ɔ:!
Now must your noble anger blaze out more	9	naʊ mʌst jɔ: ˈ**nəʊ**bl ˈæŋgə bleɪz aʊt mɔ:
Than when from Sobieski, clan by *clan*,	10	**ðæn wɛn fr**ɒm Sobieski, klæ**n** baɪ klæ**n**,
The Moslem myriads fell, and fled before	11	**ð**ə ˈmɒzlɛm ˈmɪrɪədz fɛl, ænd flɛd bɪˈfɔ:
Than when Zamoysky smote the Tartar *Khan*,	12	**ðæn wɛn** Zamoysky sm**əʊt ðə** ˈtɑ:tə kɑ:**n**,
Than earlier, when on the Baltic shore	13	**ðæn** ˈɜ:lɪə, **wɛn ɒn ðə** ˈbɔ:ltɪk ʃɔ:
Boleslas drove the *Pomeranian*.	14	Boleslas dr**əʊ**v **ðə** ˌ**p**ɒməˈreɪnjə**n**.

"Sonnet [Blow ye the trumpet gather from afar]" (1832)

上記 (5) の中で押韻が確認できるものの，Topic 12 のキーワードに含まれていない要素に，1, 4, 5行目の行末の母音 /ɑ:/ (φɑ:) と，8, 9, 11, 13行目の行末音 /ɔ:/ (φɔ:) がある。φɑ: は Topic 7 (α = 0.27874) のキーワードであり，φɔ: は Topic 3 (α = 6.10462) と Topic 9 (α = 0.45539) のキーワードである。Topic 7 が顕著に現れる作品の上位5作品は，上位から順に Farfrwy89 ("Far far

away”, 1889), Songvry30 (“Song [Every day hath its night]”, 1830), ThSlpngBe42 (“The Sleeping Beauty”, 1842), ThSlpng30 (“The Sleeping Beauty”, 1830), SCTCHSN27 (“Scotch song”, 1827) だが，語末の /ɑː/ (φɑː) が最も多く用いられているのは，Farfrwy89 であり，以下 (6) に引用する raw/IPA テキストからも分かるように，この φɑː は押韻することを主目的として使われたというよりは，寧ろ far という語を反復 (repetition) させることによって得られる効果 (スウィングしているようなリズムを生み出すなど) を狙ったことにより同じ語が繰り返し使われ，結果として同じ音が多用されたと考えられる。この作品内で far という語が複数回用いられているだけではなく，φɑː の前にくる音は語頭音と語中音合わせて7種類45例であり，そのうち語頭音 fφ は28例，次に多いのは stφ の7例と大きく差が開いており，必然的に φɑː とともに far を構成する fφ も，このトピックのキーワードである。

(6)

WHAT SIGHT so lured him thro' the fields he knew	1	wɒt saɪt səʊ ljʊəd hɪm θruː ðə fiːldz hiː njuː
As where earth's green stole into heaven's own hue,	2	æz weər ɜːθs griːn stəʊl ˈɪntuː ˈhɛvnz əʊn hjuː,
Far — far — away?	3	*fɑː — fɑː — əˈweɪ?*
What sound was dearest in his native dells?	4	wɒt saʊnd wɒz ˈdɪərɪst ɪn hɪz ˈneɪtɪv dɛlz?
The mellow lin-lan-lone of evening bells	5	ðə ˈmɛləʊ lɪn-læn-ləʊn ɒv ˈiːvnɪŋ bɛlz
Far — far — away.	6	fɑː — fɑː — əˈweɪ.
What vague world-whisper, mystic pain or joy,	7	wɒt veɪg wɜːld-ˈwɪspə, ˈmɪstɪk peɪn ɔː ʤɔɪ,
Thro' those three words would haunt him when a boy,	8	θruː ðəʊz θriː wɜːdz wʊd hɔːnt hɪm wɛn ə bɔɪ,
Far — far — away?	9	fɑː — fɑː — əˈweɪ?
A whisper from his dawn of life? a breath	10	ə ˈwɪspə frɒm hɪz dɔːn ɒv laɪf? ə brɛθ
From some fair dawn beyond the doors of death	11	frɒm sʌm feə dɔːn bɪˈjɒnd ðə dɔːz ɒv dɛθ
Far — far — away?	12	fɑː — fɑː — əˈweɪ?
Far, far, how far? from o'er the gates of Birth,	13	fɑː, fɑː, haʊ fɑː? frɒm ˈəʊə ðə geɪts ɒv bɜːθ,
The faint horizons, all the bounds of earth,	14	ðə feɪnt həˈraɪznz, ɔːl ðə baʊndz ɒv ɜːθ,
Far — far — away?	15	fɑː — fɑː — əˈweɪ?
What charm in words, a charm no words could give?	16	wɒt ʧɑːm ɪn wɜːdz, ə ʧɑːm nəʊ wɜːdz kʊd gɪv?
O dying words, can Music make you live	17	əʊ ˈdaɪɪŋ wɜːdz, kæn ˈmjuːzɪk meɪk juː lɪv
Far — far — away?	18	fɑː — fɑː — əˈweɪ?

(“Far far away”, 1889)

Topic 3 (α = 6.10462) と Topic 9 (α = 0.45539) のキーワードである φɔː は，それぞれのトピックが顕著に現れる上位5作品を見ても，語末音の φɔː が脚韻として用いられている例は見つからなかった。語末音 φɔː が二つのトピックに共通して現れる位置の前後には，/iː/ や /uː/ などの長母音や，二重母音が現れていることが多いようにも見えるが，音韻的な理由で語末音の φɔː が

選択されたのか，または語末音 φɔ: を含む語に高頻度語が多いためにキーワードとなったのかを判断するには，更に詳細な考察が必要である。

Topic 1 が最も顕著に現れるのは，(3) にも引用した，RiflmnF92 (“Riflemen Form”, 1892) である。RiflmnF92 の1–12行までに，Topic 1のキーワード20位までの語を色付けしたものを以下 (8) に示す。Raw テキストの LDA の結果で同じトピックに分類された，storm と form を構成する音 (stφ, fφ, φɔ:φ, φm) は，本 IPA sep テキストの結果においても全てキーワードに入っている。

(7)

ðφ	φeəφ	φr	ɪφ	φz	ə	sφ	φaʊφ	φnd	ɒφ	φv	θφ	φʌφ	φndφ	φəφ	φr	əφ	φfφ	φɑ:				
stφ	φɔ:φ	φm	ɪφ	φn	ðφ	φə	sφ	φaʊφ	φθ	ðφ	φæφ	φt	dφ	φɑ:φ	φkφ	φəφ	φnz	ðφ	φə	dφ	φeɪ	
stφ	φɔ:φ	φm	ɒφ	φv	bφ	φæφ	φtl	æφ	φnd	θφ	φʌφ	φndφ	φəφ	φr	ɒφ	φv	wφ	φɔ:				
wφ	φεφ	φl	ɪφ	φf	ɪφ	φt	dφ	φu:	nφ	φɒφ	φt	rφ	φəʊφ	φl	aʊə	wφ	φeɪ					
stφ	φɔ:φ	φm	stφ	φɔ:φ	φm	Rφ	φiφ	φflφ	φeφ	φmφ	φeφ	φn	fφ	φɔ:φ	φm							
rφ	φεφ	φdφ	φi	bφ	φi:	rφ	φεφ	φdφ	φi	əφ	φgφ	φεφ	φnst	ðφ	φə	stφ	φɔ:φ	φm				
Rφ	φiφ	φflφ	φeφ	φmφ	φeφ	φn	Rφ	φiφ	φflφ	φeφ	φmφ	φeφ	φn	Rφ	φiφ	φflφ	φeφ	φmφ	φeφ	φn	fφ	φɔ:φ	φm
bφ	φi:	nφ	φɒφ	φt	dφ	φεφ	φf	tφ	φu:	ðφ	φə	sφ	φaʊφ	φnd	ðφ	φæφ	φt	wφ	φɔ:φ	φnz		
bφ	φi:	nφ	φɒφ	φt	gφ	φʌφ	φld	bφ	φaɪ	ə	dφ	φεφ	φspφ	φɒφ	φts	plφ	φi:					
ɑ:	fφ	φɪφ	φgz	ɒφ	φv	θφ	φɪφ	φslz	ɔ:	grφ	φeɪφ	φps	ɒφ	φv	θφ	φɔ:φ	φnz					
hφ	φaʊ	kφ	φæφ	φn	ə	dφ	φεφ	φspφ	φɒφ	φt	fφ	φi:φ	φl	wφ	φɪφ	φð	ðφ	φə	frφ	φi:		
fφ	φɔ:φ	φm	fφ	φɔ:φ	φm	Rφ	φiφ	φflφ	φeφ	φmφ	φeφ	φn	fφ	φɔ:φ	φm							

(8)

THERE is a sound of thunder afar,	1	ðeər ɪz ə saʊnd ɒv ˈθʌndər əˈfɑ:,
Storm in the South that darkens the day!	2	stɔ:m ɪn ðə saʊθ ðæt ˈdɑ:kənz ðə deɪ!
Storm of battle and thunder of war!	3	stɔ:m ɒv ˈbætl ænd ˈθʌndər ɒv wɔ:!
Well if it do not roll our way.	4	wɛl ɪf ɪt du: nɒt rəʊl ˈaʊə weɪ.
Storm, Storm, Riflemen form!	5	stɔ:m, stɔ:m, Riflemen fɔ:m!
Ready, be ready against the storm!	6	ˈrɛdi, bi: ˈrɛdi əˈgɛnst ðə stɔ:m!
Riflemen, Riflemen, Riflemen form!	7	Riflemen, Riflemen, Riflemen fɔ:m!
Be not deaf to the sound that warns,	8	bi: nɒt dɛf tu: ðə saʊnd ðæt wɔ:nz,
Be not gull'd by a despot's plea!	9	bi: nɒt gʌld baɪ ə ˈdɛspɒts pli:!
Are figs of thistles? or grapes of thorns?	10	ɑ: fɪgz ɒv ˈθɪslz? ɔ: greɪps ɒv θɔ:nz?
How can a despot feel with the Free?	11	haʊ kæn ə ˈdɛspɒt fi:l wɪð ðə fri:?
Form, Form, Riflemen Form!	12	fɔ:m, fɔ:m, Riflemen fɔ:m!

“Riflemen Form” (1892: ll. 1–12)

(8) の raw/IPA テキストの8, 10行目には，同じ母音を使った脚韻があり (/wɔ:nz/, / θɔ:nz /)，一見すると押韻を検知したようにも見える。しかし，raw テキストで頻度を調べると，form はコーパス内で全28例中14例，storm は全12例中10例がこのRiflmnF92に生起しており，特にこの2語に共通する，φɔ:φ, φm は，この結果のみでは押韻を検知したと断言することはできない。更に，この2語に共通する φɔ:φ を中心になる 3-gram の頻度 (頻度10以上) を以下表10に示す。

表 10　IPA sep テキストにおける φɔːφ を中心とする 3-gram (頻度 10 以上)

	Sound 1	NODE	Sound 2	FREQ
1	fφ	φɔːφ	φr	31
2	fφ	φɔːφ	φm	28
3	glφ	φɔːφ	φrφ	23
4	mφ	φɔːφ	φn	16
5	lφ	φɔːφ	φd	15
6	θφ	φɔːφ	φt	13
7	stφ	φɔːφ	φm	12
8	wφ	φɔːφ	φtφ	12
9	dφ	φɔːφ	φn	10
10	fφ	φɔːφ	φl	10
11	fφ	φɔːφ	φθ	10

Sound 1 の位置にくる音は，表10の11例中4例が fφ であり，fφ – φɔːφ の組み合わせののべ頻度は 79/180である。また，頻度順2番目には fφ φɔːφ, φm の28があり，φɔːφ 音の前に位置し，語頭音として語を形成する音の中では fφ が他の音よりも多く φɔːφ と結びつきやすいと言え，またこの語あるいは音同士の結びつきやすさや，組み合わせの頻度による影響を可能な限り避けた方が，押韻や音韻特徴に焦点を当てた分析ができると考える。この影響を避ける方法の一つとして，IPA 表記したテキストを更に分割する際に，本稿のように子音と母音で分けるのではなく，音節ごとに区切ることが挙げられるが，IPA テキストをどのように正確に音節ごとに分割するかなど課題も多く，稿を改めて議論したい。

最後に，Topic 4 を考察する。Topic 4 が最も顕著に現れる作品は，(9), (10) に挙げる MYBRTHR27 (“My brother”, 1827) である。この詩は四行連で書かれており，7つあるスタンザの

(9)

wφ	φɪφ	φð	fφ	φaφ	φltrφ	φiφ	φng	stφ	φɛφ	φp	aɪ	kφ	φeɪφ	φm	tφ	φuː	sφ	φiː	
ɪφ	φn	dφ	φɛφ	φθs	ʌφ	φnhφ	φiːφ	φdφ	φɪφ	φŋ	æφ	φpφ	φəφ	φθφ	φi				
δφ	φæφ	φt	frφ	φɛφ	φnd	sφ	φəʊ	dφ	φɪəφ	φr	ɪφ	φn	lφ	φaɪφ	φf	tφ	φuː	mφ	φiː
mφ	φaɪ	brφ	φʌφ	φðφ	φə														
mφ	φɪφ	φd	flφ	φaʊəφ	φz	ɒφ	φv	lφ	φʌφ	φvlφ	φɪɪφ	φst	sφ	φɛφ	φnt	æφ	φnd	hjφ	φuː
δφ	φæφ	φt	strφ	φuːφ	φd	δφ	φaɪ	fφ	φɔːφ	φm	twφ	φɒφ	φz	sφ	φæφ	φd	tφ	φuː	vjφ	φuː
δφ	φaɪ	lφ	φaɪφ	φflφ	φɪφ	φs	fφ	φeɪφ	φs	pφ	φiːφ	φp	wφ	φɒφ	φnlφ	φi	θrφ	φuː	
mφ	φaɪ	brφ	φʌφ	φðφ	φə														
wφ	φaɪ	dφ	φɪφ	φd	δφ	φeɪ	(δφ	φeə	δφ	φeɪ	dφ	φɪφ	φd	nφ	φɒφ	φt	fφ	φiːφ	φl)
wφ	φɪφ	φð	stjφ	φuːφ	φdφ	φiəφ	φs	kφ	φeəφ	φr	ɔːφ	φl	εφ	φls	kφ	φəφ	φnsφ	φiːφ	φl
bφ	φʌφ	φt	δφ	φaɪ	kφ	φəʊφ	φld	fφ	φeɪφ	φs	əφ	φlφ	φəʊφ	φn	rφ	φɪφ	φvφ	φiːφ	φl
mφ	φaɪ	brφ	φʌφ	φðφ	φə														
δφ	φeɪ	mφ	φaɪφ	φt	hφ	φæφ	φv	nφ	φəʊφ	φn	wφ	φɒφ	φt	jφ	φuːφ	φzd	tφ	φuː	glφ	φəʊ
wφ	φɪφ	φð	smφ	φaɪφ	φlz	æφ	φnd	ɒφ	φft	dφ	φɪφ	φspφ	φɛφ	φld	mφ	φaɪ	wφ	φəʊ	
wφ	φʊφ	φd	ʃfφ	φɪφ	φl	mφ	φiː	mφ	φəʊφ	φst	wφ	φɛφ	φn	fφ	φeɪφ	φdφ	φɪφ	φd	sφ	φəʊ
mφ	φaɪ	brφ	φʌφ	φðφ	φə														
δφ	φə	tφ	φəʊφ	φlφ	φɪφ	φŋ	ɒφ	φv	δφ	φaɪ	fjφ	φuːφ	φnφ	φəφ	φrφ	φəφ	φl	bφ	φɛφ	φl
δφ	φə	nφ	φaɪφ	φn	lφ	φəʊ	nφ	φəʊφ	φts	δφ	φæφ	φt	spφ	φəʊφ	φk	δφ	φaɪ	nφ	φɛφ	φl
aɪ	nφ	φəʊ	nφ	φɒφ	φt	hφ	φaʊ	aɪ	bφ	φɔː	sφ	φəʊ	wφ	φɛφ	φl					
mφ	φaɪ	brφ	φʌφ	φðφ	φə														
bφ	φʌφ	φt	əʊ	δφ	φə	ʃfφ	φɪφ	φl	dφ	φæφ	φŋk	mφ	φəʊφ	φld	δφ	φæφ	φt	slφ	φɪφ	φd
dφ	φʌφ	φl	sφ	φaʊφ	φndφ	φɪφ	φŋ	ɒφ	φn	δφ	φaɪ	kφ	φɒφ	φfφ	φɪφ	φn	lφ	φɪφ	φd	
δφ	φæφ	φt	drφ	φuː	mφ	φɔː	tφ	φeəφ	φz	δφ	φæφ	φn	ɔːφ	φl	bφ	φɪφ	φsφ	φaɪφ	φd	
mφ	φaɪ	brφ	φʌφ	φðφ	φə														
æφ	φnd	δφ	φɛφ	φn	aɪ	hφ	φʌφ	φrφ	φɪφ	φd	fφ	φɑːφ	φst	əφ	φwφ	φeɪ			
hφ	φaʊ	kφ	φʊφ	φd	aɪ	εə	hφ	φæφ	φv	bφ	φɔːφ	φn	tφ	φuː	stφ	φeɪ			
wφ	φeə	kφ	φeəφ	φlφ	φɪφ	φs	hφ	φæφ	φnd	ɪφ	φnhjφ	φuːφ	φmd	δφ	φaɪ	klφ	φeɪ		

(10)

WITH falt'ring step I came to *see*,	1	wɪð falt'ring stɛp aɪ keɪm tu: si:,
In Death's unheeding apathy,	2	ɪn dɛθs ʌnˈhi:dɪŋ ˈæpəθi,
That friend so dear in life to *me*,	3	ðæt frɛnd səʊ dɪər ɪn laɪf tu: mi:,
My brother!	4	maɪ ˈbrʌðə!
'Mid flowers of loveliest scent and hue	5	mɪd ˈflaʊəz ɒv ˈlʌvlɪɪst sɛnt ænd hju:
That strew'd thy form, 'twas sad to view	6	ðæt stru:d ðaɪ fɔ:m, twɒz sæd tu: vju:
Thy lifeless face peep wanly through,	7	ðaɪ ˈlaɪflɪs feɪs pi:p ˈwɒnli θru:,
My brother!	8	maɪ ˈbrʌðə!
Why did they (there they did not *feel*!)	9	waɪ dɪd ðeɪ (ðeə ðeɪ dɪd nɒt fi:l!)
With studious care all else *conceal*,	10	wɪð ˈstju:diəs keər ɔ:l ɛls kənˈsi:l,
But thy cold face alone *reveal*,	11	bʌt ðaɪ kəʊld feɪs əˈləʊn rɪˈvi:l,
My brother!	12	maɪ ˈbrʌðə!
They might have known, what used to glow	13	ðeɪ maɪt hæv nəʊn, wɒt ju:zd tu: gləʊ
With smiles, and oft dispell'd my woe,	14	wɪð smaɪlz, ænd ɒft dɪsˈpɛld maɪ wəʊ,
Would chill me most, when faded so,	15	wʊd ʧɪl mi: məʊst, wɛn ˈfeɪdɪd səʊ,
My brother!	16	maɪ ˈbrʌðə!
The tolling of thy funeral *bell*,	17	ðə ˈtəʊlɪŋ ɒv ðaɪ ˈfju:nərəl bɛl,
The nine low notes that spoke thy *knell*,	18	ðə naɪn ləʊ nəʊts ðæt spəʊk ðaɪ nɛl,
I know not how I bore so *well*,	19	aɪ nəʊ nɒt haʊ aɪ bɔ: səʊ wɛl,
My brother!	20	maɪ ˈbrʌðə!
But oh! the chill, dank mould that *slid*,	21	bʌt əʊ! ðə ʧɪl, dæŋk məʊld ðæt slɪd,
Dull-sounding, on thy coffin-*lid*,	22	dʌl-ˈsaʊndɪŋ, ɒn ðaɪ ˈkɒfɪn-lɪd,
That drew more tears than all *beside*,	23	ðæt dru: mɔ: teəz ðæn ɔ:l bɪˈsaɪd,
My brother!	24	maɪ ˈbrʌðə!
And then I hurried fast away;	25	ænd ðɛn aɪ ˈhʌrɪd fɑ:st əˈweɪ;
How could I e'er have borne to stay	26	haʊ kʊd aɪ ɛə hæv bɔ:n tu: steɪ
Where careless hand inhumed thy clay,	27	weə ˈkeəlɪs hænd ɪnˈhju:md ðaɪ kleɪ,
My brother!	28	maɪ ˈbrʌðə!

"My brother" (1827)

それぞれ1行目から3行目までが脚韻を踏んでいる。この7種の脚韻の音のうち，Topic 4 のキーワードに含まれているのは，部分的な (例えば，第3スタンザの脚韻は /i:l/ であるが，トピックのキーワードになっているのは子音の /l/ のみである) ものも含めて4種である。

この Topic 4 のα値は 2.54886 であり，17あるトピックの中で5番目に高い数値であり，他の作品にも当該トピックのキーワードが現れている可能性が高い。そこで，MYBRTHR27 の次，つ

まり2番目と3番目に顕著に Topic 4 が現れる作品，THBRDSF27 (“The bards farewell”, 1827) (以下(11), (12)) と Sonnthb32 (“Sonnet [Oh beauty passing beauty]”, 1832) (以下(13), (14)) の2作品も考察する。

(11)

snφ	φəʊφ	φdφ	φəφ	φn	δφ	φaɪ	klφ	φɪφ	φfs	ʃφ	φæφ	φl	hφ	φɪə	nφ	φəʊ	mφ	φɔ:	
δφ	φɪφ	φs	dφ	φi:φ	φp	tφ	φəʊφ	φnd	hφ	φɑ:φ	φp	əφ	φgφ	φεφ	φn				
bφ	φʌφ	φt	bφ	φæφ	φnφ	φə	krφ	φaɪ	æφ	φnd	bφ	φæφ	φtl	rφ	φɔ:				
ʃφ	φæφ	φl	fφ	φɔ:φ	φm	ə	fφ	φɪəφ	φsφ	φə	strφ	φeɪφ	φn						
əʊə	δφ	φaɪ	swφ	φi:φ	φt	kφ	φɔ:φ	φdz	mφ	φaɪ	mφ	φæφ	φdʒφ	φɪφ	φk	lφ	φaɪə		
wφ	φɒφ	φt	fjφ	φu:φ	φʧφ	φə	hφ	φæφ	φnd	ʃφ	φæφ	φl	strφ	φeɪ					
wφ	φɒφ	φt	brφ	φeɪφ	φn	ʃφ	φæφ	φl	fφ	φi:φ	φl	δφ	φaɪ	mφ	φɑ:φ	φstφ	φəφ	φz	fφ	φaɪə
ɔ:	frφ	φeɪφ	φm	hφ	φɪφ	φz	mφ	φæφ	φʧlφ	φɪφ	φs	lφ	φeɪ						
wφ	φεφ	φl	mφ	φaɪφ	φt	δφ	φə	krφ	φɑ:φ	φftφ	φi	εφ	φdwφ	φəφ	φd	fφ	φɪə		
ʃφ	φʊφ	φd	aɪ	bφ	φʌφ	φt	tφ	φʌφ	φʧ	δφ	φaɪ	kφ	φɔ:φ	φd					
ɪφ	φts	slφ	φaɪφ	φtφ	φɪφ	φst	sφ	φaʊφ	φnd	wφ	φʊφ	φd	kφ	φaʊφ	φʧ	δφ	φə	spφ	φɪə
æφ	φnd	bφ	φeə	δφ	φi	ɪφ	φndφ	φɪφ	φgnφ	φəφ	φnt	sφ	φɔ:φ	φd					
fφ	φʊφ	φl	wφ	φεφ	φl	hφ	φi:	njφ	φu:	δφ	φə	wφ	φɪφ	φzφ	φəφ	φd	spφ	φεφ	φl
δφ	φæφ	φt	dwφ	φεφ	φlt	əφ	φpφ	φɒφ	φn	δφ	φaɪ	strφ	φɪφ	φŋ					
æφ	φnd	trφ	φεφ	φmbld	wφ	φεφ	φn	hφ	φi:	hφ	φɜ:φ	φd	δφ	φaɪ	swφ	φεφ	φl		
θrφ	φu:	snφ	φəʊφ	φdφ	φəφ	φnz	kφ	φæφ	φvφ	φəφ	φnz	rφ	φɪφ	φŋ					
δφ	φi:φ	φz	aɪφ	φz	ʃφ	φæφ	φl	slφ	φi:φ	φp	ɪφ	φn	dφ	φεφ	φθs	dφ	φʌφ	φl	nφ	φaɪφ	φt
δφ	φɪφ	φs	hφ	φæφ	φnd	ɔ:φ	φl	nφ	φɜ:φ	φvlφ	φɪφ	φs	lφ	φaɪ					
eə	wφ	φʌφ	φns	əφ	φgφ	φεφ	φn	jφ	φɒφ	φn	ɔ:φ	φb	ɒφ	φv	lφ	φaɪφ	φt		
brφ	φeɪφ	φk	əʊə	δφ	φə	klφ	φɪə	blφ	φu:	skφ	φaɪ								
æφ	φnd	δφ	φaʊ	bφ	φaɪ	hφ	φεφ	φlz	əʊφ	φn	fjφ	φʊəφ	φrφ	φiφ	φz	nφ	φuφ	φrst	
ʌφ	φnfφ	φɜ:φ	φl	δφ	φaɪ	bφ	φæφ	φnφ	φəφ	φz	prφ	φaɪφ	φd						
bφ	φʌφ	φt	nφ	φəʊ	δφ	φæφ	φt	lφ	φɪφ	φvφ	φɪφ	φŋ	δφ	φi:	aɪ	kφ	φɜ:φ	φst	
æφ	φnd	kφ	φɜ:φ	φsφ	φɪφ	φŋ	δφ	φi:	aɪ	dφ	φaɪφ	φd							

(12)

SNOWDON! thy cliffs shall hear no more	1	ˈsnəʊdən! ðaɪ klɪfs ʃæl hɪə nəʊ mɔ:
This deep-toned harp again;	2	ðɪs di:p-təʊnd hɑ:p əˈgɛn;
But banner-cry and battle-roar	3	bʌt ˈbænə-kraɪ ænd ˈbætl-rɔ:
Shall form a fiercer strain!	4	ʃæl fɔ:m ə ˈfɪəsə streɪm!
O'er thy sweet chords, my magic lyre!	5	ˈəʊə ðaɪ swi:t kɔ:dz, maɪ ˈmæʤɪk ˈlaɪə!
What future hand shall stray?	6	wɒt ˈfju:ʧə hænd ʃæl streɪ?
What brain shall feel thy master's fire,	7	wɒt breɪn ʃæl fi:l ðaɪ ˈmɑ:stəz ˈfaɪə,
Or frame his matchless lay?	8	ɔ: freɪm hɪz ˈmæʧlɪs leɪ?
Well might the crafty Edward fear:	9	wɛl maɪt ðə ˈkrɑ:fti ˈɛdwəd fɪə:
Should I but touch thy chord,	10	ʃʊd aɪ bʌt tʌʧ ðaɪ kɔ:d,
Its slightest sound would couch the spear,	11	ɪts ˈslaɪtɪst saʊnd wʊd kaʊʧ ðə spɪə,
And bare the indignant sword!	12	ænd beə ði ɪnˈdɪgnənt sɔ:d!
Full well he knew the wizard-spell	13	fʊl wɛl hi: nju: ðə ˈwɪzəd-spɛl
That dwelt upon thy string;	14	ðæt dwɛlt əˈpɒn ðaɪ strɪŋ;
And trembled, when he heard thy swell	15	ænd ˈtrɛmbld, wɛn hi: hɜ:d ðaɪ swɛl
Thro' Snowdon's caverns ring!	16	θru: ˈsnəʊdənz ˈkævənz rɪŋ!
These eyes shall sleep in death's dull night,	17	ði:z aɪz ʃæl sli:p ɪn dɛθs dʌl naɪt,
This hand all nerveless lie,	18	ðɪs hænd ɔ:l ˈnɜ:vlɪs laɪ,
Ere once again yon orb of light	19	eə wʌns əˈgɛn jɒn ɔ:b ɒv laɪt
Break o'er the clear blue sky!	20	breɪk ˈəʊə ðə klɪə blu: skaɪ!

And thou, by Hell's own furies nurst,	21	ænd ðaʊ, baɪ hɛlz əʊn ˈfjʊəriz nurst,
Unfurl thy banner's pride!	22	ˌʌnˈfɜːl ðaɪ ˈbænəz praɪd!
But know that, living, thee I cursed;	23	bʌt nəʊ ðæt, ˈlɪvɪŋ, ðiː aɪ kɜːst;
And, cursing thee, I died!	24	ænd, ˈkɜːsɪŋ ðiː, aɪ daɪd!

“The bards farewell” (13827)

(13)

əʊ	bjφ	φuːφ	φtφ	φi	pφ	φɑːφ	φsφ	φɪφ	φŋ	bjφ	φuːφ	φtφ	φi	swφ	φiːφ	φtφ	φɪφ	φst	swφ	φiːφ	φt							
hφ	φaʊ	canst	ðφ	φaʊ	lφ	φɛφ	φt	mφ	φiː	wφ	φɛɪφ	φst	mφ	φaɪ	jφ	φuːφ	φθ	ɪφ	φn	sφ	φaɪφ	φz						
aɪ	ɔʊφ	φnlφ	φi	ɑːφ	φsk	tφ	φuː	sφ	φɪφ	φt	bφ	φɪφ	φsφ	φaɪφ	φd	ðφ	φaɪ	fφ	φiːφ	φt								
ðφ	φaʊ	nφ	φɔʊɪφ	φst	aɪ	dφ	φeɔ	nφ	φɒφ	φt	lφ	φʊφ	φk	ɪφ	φntφ	φuː	ðφ	φaɪφ	φn	aɪφ	φz							
mφ	φaɪφ	φt	aɪ	bφ	φʌφ	φt	kφ	φɪφ	φs	ðφ	φaɪ	hφ	φæφ	φnd	aɪ	dφ	φeɔ	nφ	φɒφ	φt	fφ	φɔʊφ	φld					
mφ	φaɪ	ɑːφ	φmz	ɔφ	φbφ	φaʊφ	φt	ðφ	φiː	skφ	φɛɔφ	φslφ	φi	dφ	φeɔ	tφ	φuː	spφ	φiːφ	φk								
æφ	φnd	nφ	φʌφ	φθφ	φɪφ	φŋ	sφ	φiːφ	φmz	tφ	φuː	mφ	φiː	sφ	φɔʊ	wφ	φaɪφ	φld	æφ	φnd	bφ	φɔʊφ	φld					
æφ	φz	wφ	φɪφ	φð	wφ	φʌφ	φn	kφ	φɪφ	φs	tφ	φuː	tφ	φʌφ	φʃ	ðφ	φaɪ	blφ	φɛφ	φst	ʃφ	φiːφ	φk					
mφ	φɪφ	φθφ	φɪφ	φŋks	ɪφ	φf	aɪ	ʃφ	φʊφ	φd	kφ	φɪφ	φs	ðφ	φiː	nφ	φɔʊ	kφ	φɔφ	φntrφ	φɔʊφ	φl						
wφ	φɪφ	φðφ	φɪφ	φn	ðφ	φɔ	θrφ	φɪφ	φlφ	φɪφ	φŋ	brφ	φɛɪφ	φn	kφ	φʊφ	φd	kφ	φiːφ	φp	ɔφ	φflφ	φɔʊφ	φt				
ðφ	φɔ	sφ	φʌφ	φtl	spφ	φɪφ	φrφ	φɪφ	φt	iːφ	φvφ	φɔφ	φn	wφ	φaɪφ	φl	aɪ	spφ	φɔʊφ	φk								
ðφ	φɔ	bφ	φeɔ	wφ	φɜːφ	φd	kφ	φɪφ	φs	hφ	φæφ	φθ	mφ	φɛɪφ	φd	mφ	φaɪ	ɪφ	φnφ	φɔ	sφ	φɔʊφ	φl					
tφ	φuː	trφ	φɛφ	φmbl	lφ	φaɪφ	φk	ə	lφ	φuːφ	φtstrφ	φɪφ	φŋ	eɔ	ðφ	φɔ	nφ	φɔʊφ	φt									
hφ	φæφ	φθ	mφ	φɛφ	φltφ	φɪφ	φd	ɪφ	φn	ðφ	φɔ	sφ	φaɪφ	φlφ	φɔφ	φns	ðφ	φæφ	φt	ɪφ	φt	brφ	φɔʊφ	φk				

(14)

Blow ye the trumpet, gather from afar	1	bləʊ jiː ðə ˈtrʌmpɪt, ˈgæðə frɒm əˈfɑː
The hosts to battle: be not bought and sold.	2	ðə həʊsts tuː ˈbætl: biː nɒt bɔːt ænd səʊld.
Arise, brave Poles, the boldest of the bold;	3	əˈraɪz, breɪv pəʊlz, ðə ˈbəʊldɪst ɒv ðə bəʊld;
Break through your iron shackles fling them far.	4	breɪk θruː jɔːr ˈaɪən ˈʃæklz flɪŋ ðɛm fɑː.
O for those days of Piast, ere the Czar	5	əʊ fɔː ðəʊz deɪz ɒv Piast, eə ðə zɑː
Grew to this strength among his deserts cold;	6	gruː tuː ðɪs strɛŋθ əˈmʌŋ hɪz ˈdɛzəts kəʊld;
When even to Moscow's cupolas were rolled	7	wɛn ˈiːvən tuː ˈmɒskəʊz ˈkjuːpələz wɜː rəʊld
The growing murmurs of the Polish war!	8	ðə ˈgrəʊɪŋ ˈmɜːməz ɒv ðə ˈpɒlɪʃ wɔː!
Now must your noble anger blaze out more	9	naʊ mʌst jɔː ˈnəʊbl ˈæŋgə bleɪz aʊt mɔː
Than when from Sobieski, clan by clan,	10	ðæn wɛn frɒm Sobieski, klæn baɪ klæn,
The Moslem myriads fell, and fled before	11	ðə ˈmɒzlɛm ˈmɪrɪədz fɛl, ænd flɛd bɪˈfɔː
Than when Zamoysky smote the Tartar Khan,	12	ðæn wɛn Zamoysky sməʊt ðə ˈtɑːtə kɑːn,
Than earlier, when on the Baltic shore	13	ðæn ˈɜːlɪə, wɛn ɒn ðə ˈbɔːltɪk ʃɔː
Boleslas drove the Pomeranian.	14	Boleslas drəʊv ðə ˌpɒməˈreɪnjən.

“Sonnet [Oh beauty passing beauty]” (1832)

THBRDSF27 は，6スタンザからなる四行連の詩で abab のスタイル (1, 3 行目と2, 4 行目) で脚韻している。24行のうち，キーワードとしてマークされている部分は，先の MYBRTHR27 と比較すると少なく，3種であるが，そのうち，MYBRTHR27 と共通する語中音の母音 φεφ と，語末音の子音 φl (13, 15 行目)，及び語末音 φd (22, 24 行目) がみられる。もう一つの語末音に17, 19 行目の φt があるが，この脚韻は本来直前の母音 φaɪφ も含めて然るべきであり，また22, 24 行目の脚韻も同様に，正確には直前の母音 φaɪφ も語末音と合わせて扱うのが妥当であり，やは

り今後の研究では音節に分割して分析を進めることが課題であると言える。Sonnthb32 は， 4・4・3・3行からなる14行の詩である。本作品においても，語末音であり，かつ脚韻の一部である子音 φt と φl がマークされており，α値が示す通り，この Topic 4 に含まれるキーワードが複数の作品に跨って現れていることが分かった。

4. 終わりに

本研究では，韻文作品の韻律を，本来，語の持つ潜在的意味を探索する目的で用いられる LDA トピックモデルによって分析可能であるかを検証するために，原文のままのテキスト，IPA 表記に変換したテキスト，そして IPA 表記を母音と子音に分割したテキストを対象コーパスとして各テキストに LDA を実行し，その結果を比較考察した。

分析及び考察の結果から，本研究で使用したデータに関しては，原文のままのテキストと IPA に変換したテキストでは，明確に英語詩の押韻を LDA が検知しているという確証は得られなかったが，IPA 表記を母音と子音に分割したテキストでは，他のテキストよりも押韻を検知していると見られる例が複数見つかった。一方で，IPA 表記のテキストを分割する単位を音節とした方がより良い結果を得られる可能性も示唆された。また，本稿では主に押韻の例として脚韻を挙げるに留めてしまったが，韻文作品で用いられる韻は脚韻のみではなく複数存在するため，より網羅的な韻の種類にも目を向けることが今後の課題である。

Tennyson の作品や，Tennyson の用いる音に関する研究は数多く存在しているが，多くが質的研究に依拠しているものである一方で，LDA あるいは他の計量的研究手法の利点の一つに，より多くのデータを客観的かつ網羅的に分析できるという点が挙げられる。Tennyson，ひいては他の韻文作品や文学作品の音に関するこれまでの先行研究の知見と融合させ，新たな知見を得るための計量的研究手法の可能性を本研究では提示した。

文献

Blei, M. D., Ng, Y. A., and Jordan, I. M. (2003). “Latent Dirichlet Allocation.” Journal of Machine Learning Research 3, 2003, 993–1022.

Fujita, I. (2021). “‘Were not Tennyson’s words delicious?’: An Analysis Using LDA Topic Modelling.” WEDHIA 2021, National Chengchi University.

Henrichs, A. (2019). “Deforming Shakespeare’s Sonnets: Topic Models as Poems Author(s)” Criticism, 61 (3), 387–412.

黄 晨雯 (2020a). 「中国のミステリー小説におけるトピック解析の試み」，『大阪大学言語文化学』, pp. 1–17.

黄 晨雯 (2020b).「中国のミステリー小説をめぐる計量的分析—程小青と鬼馬星の作品を中心に—」，『テクストマイニンングとデジタルヒューマニティーズ 2019』（大阪大学大学院言語文化共同研究プロジェクト2019）, pp. 31–45.

Meeks, E. and Weingart, B. S. (2012). “The Digital Humanities Contribution to Topic Modeling.” Journal of Digital Humanities, 2 (1), 1–6.

黒田 絢香 (2017). 「小説テクストの計量的分析—アーサー・コナン・ドイルの作品から—」, 『テクストマイニンングとデジタルヒューマニティーズ 2016』（大阪大学大学院言語文化共同研究プロジェクト2016）, pp. 23–41.

Kiyama, N. (2018). “How Have Political Interests of U.S. Presidents Changed?: A Diachronic Investigation of the State of the Union Addresses through Topic Modeling.” English Corpus Studies, 25, 79–99.

中尾 佳行・地村 彰之 (2016). 「『カンタベリー物語』の写本と刊本における言語と文体について」堀 正広 (編), 『コーパスと英語文体』 (pp. 21–52). 東京: ひつじ書房.

Navarro-Colorado, B. (2018). “On Poetic Topic Modeling: Extracting Themes and Motifs From a Corpus of Spanish Poetry.” frontiers in Digital Humanities, (20), 5–15.

Okabe, M. (2019). “Thou and You in Emily Dickinson's Poems Using Topic Modeling: Reconsideration of Interjections.” Proceedings of Japanese Association for Digital Humanities Conference 2019, 125–131.

Rhody, M.L. (2012). “Topic Modeling and Figurative Language.” CUNY Academic Works, 2012, 19–35.

Tabata, T. (2017). “Mapping Dickens’s Novels in a Network of Words, Topics, and Texts: Topic Modelling a Corpus of Classic Fiction.” Japanese Association for Digital Humanities Conference 2017, Doshisha University.

Tartakovsky, R. and Yehsyahu, S. (2019). “Meek as milk and large as logic: A corpus study of the non-standard poetic simile.” Language and Literature, 2019, 28 (3). 203–220.

MALLET 2.0.8: https://mimno.github.io/Mallet/index (最終閲覧日：2022年1月20日)

toPhonetics: https://tophonetics.com/ (最終閲覧日：2022年1月20日)