

「論文」

‘To spread the Word by which himself had thriven’: Analysis of Alfred Tennyson’s Use of Language Based on the LDA Topic Model

Iku FUJITA

Abstract

This study aims to provide pieces of newly discovered dimensions on the poetic style of the Victorian poet Alfred Tennyson, using a stylometric approach, the latent Dirichlet allocation (henceforth, LDA; Blei et al., 2003) topic model. Many studies have examined Tennyson’s poetry by focusing on similarities or differences between Tennyson’s and other poets’ styles, syntax or lexicons. However, most take a qualitative approach to specific poems based on close reading of texts. This study attempts to uncover new aspects of Tennyson’s style through a balanced combination of close reading and quantitative analysis. LDA is a method dedicated to analysing big data, making it possible to read text data from a ‘distance’. Moreover, ‘distant reading’ enables us ‘to focus on units that are much smaller or much larger than the text: devices, themes and tropes’ (Moretti, 2013: 48–49). Using probabilistic calculations, LDA identifies semantic connections hidden behind words in the target corpus. Although several studies have employed the topic model to investigate prose texts, very few have applied the technique to analyse poetry. However, Fujita (2022) showed that LDA is effective in examining Tennyson’s poetry. This study examines the lexicon of Tennyson’s 603 works, focusing particularly on the nouns he uses. Results of this analysis show that some topics (semantically classified word groups), such as ‘immortality’, one of Tennyson’s major themes, are the most prevalent in his body of work.

1. Introduction

Alfred Lord Tennyson (1809–1892) is one of Britain’s representative poets of the

Victorian period. He became a poet laureate in 1850, after the death and the resignation of the previous laureate, William Wordsworth (1770–1850). Tennyson wrote more than 600 poems, including unpublished ones, of which more than 80 percent are lyrical poems. However, his major works are narrative and epic poems, which tend to be longer than lyrical poetry. One of his famous lyrical poems, *In Memoriam A.H.H.* (1850; henceforth *In Memoriam*), consists of more than 2,800 lines. It can be said, then, that one feature of Tennyson's body of work is his long poetry.

Numerous past studies exist on Tennyson's life and his poetry, style, prosody and lexicon (e.g., Dixon, 1896; Tennyson, 1897; Nakamura, 1967; Ricks, 1969, 1987; Nishimae, 1979, 2000; Bloom, 1985; Hair, 1991; Kabata, 2001, 2007; Noguchi, 2011; Holmes, 2012; Thomas, 2019), and much of these studies adopt a qualitative approach to specific poems. Plamondon (2005) is the sole study that takes a quantitative approach and investigates the difference in some sounds' frequencies between Tennyson's and Robert Browning's poetry. However, other than Plamondon's research, few Tennysonian studies employ quantitative approaches amongst extant studies. The qualitative studies above point out or investigate topics such as 'Auditory Value in *Enoch Arden*' (Horiuchi, 1992), 'Tennyson and Death' (Bruce, 1917), and 'Tennyson and Spiritualism' (Elliott, 1979). Horiuchi (1992) offers detailed observations of sounds and prosody in *Enoch Arden* (1862), as reflected in the title of his paper. Bruce (1917) and Elliott (1979) discuss the themes of death and spiritualism respectively. Bruce (1917) focuses on poems such as 'Lover's Tale' (1833), 'Morte d'Arthur' (1842), 'St. Agnes' Eve' (1836) and 'Gareth and Lynette' (1872), while Elliott (1979) spotlights comments for Tennyson's works from Tennyson himself and his close people, such as Frederick, to explore his works like 'The Ancient Sage' (1885), 'The Silent Voices' (1892) and *In Memoriam*. Tennyson's 'Tithonus' (1862), *In Memoriam* and 'Crossing the Bar' (1892) often attract attention from literary critics in examinations of the subject of immortality in his poetry (e.g., Shaw, 1976; Elliot, 1979; Perrine, 1966). Ricks (1987) is a representative study that comprehensively annotated Tennyson's works with abundant materials. Various critical studies refer to Ricks (1987), or to the earlier edition, Ricks (1969), and Shaw (1973) and Hair (1991) are no exception. Shaw (1973) discusses the similarity between Tennyson's syntax and style and that of other poets and playwrights, for example, Homer, William Shakespeare, John Keats, Wordsworth and others, by identifying and counting the number of allusions to other

authors, as identified in Ricks (1969). Hair (1991) illustrates qualitative features by comparing words, concepts or motifs with those of other poets. A recent critical study by Thomas (2019) employs close reading to point out the *echoes* in Tennyson's works of his predecessor as poet laureate, Wordsworth.

What these earlier studies have in common is not only that they undertake detailed and in-depth close readings, but also that they focus on specific words, phrases or works. However, there is rarely critical research that presents an overall picture by observing an entire work(s). In other words, although there are much *other elements* we might wish to see outside of a specific/representative work, we are missing them by focusing on only a limited part of Tennyson's works.

Moretti (2013: 48) explains, 'the trouble with close reading [...] is that it necessarily depends on an extremely small canon' in the literature research. He continues, saying close reading leaves 'great unread' in works in question since it focuses on a limited part of work(s). Moretti adds, 'if you want to look beyond the canon' (2013: 48), 'distant reading' will work effectively for this. His 'great unread' corresponds to the abovementioned *other elements*. Though Moretti does not specify what the 'canon' encompasses, his remarks can be applied to research on individual poets/authors. Nonetheless, the intent of Moretti (2013) and the present paper is not to criticise studies based on close reading or the approach itself. *Other elements*/'great unread' can cover a wide range, but this study focuses on two divergences: the possibility of previous studies might have missed and not discussed poems, which contain the same themes as other Tennysonian canons; the prospects of unmentioned topics by earlier scholars, though written within his poems. More precisely, as mentioned in previous studies, even though immortality has been a general theme, the poems in question were supposedly limited to the works mentioned above, such as *In Memoriam*, 'Crossing the Bar' and 'Tithonus'. Likewise, some studies indicate relationship between seas and emotions of characters as well as what seas connote in his particular poems, for instance 'Mariana in the South' (1832), 'Ænone' (1832), 'The Mermaid' (1830) and 'Morte d'Arthur' (Fulweiler, 1965; Keirstead, 2019). Meanwhile, Okazawa (1969) points out Tennyson's preference of waters, namely seas, rivers and lakes; rivers have not been fully recognised by Tennysonian scholars compared to seas. *The Princess* (1847) and some other works have also been given earlier attention in the way Tennyson depicts female characters; however, male protagonists have not been

noticed as much as their female counterparts have. Because few studies exist that employ distant reading, including statistical methods and stylometric approaches, this paper aims to reveal the ‘great unread’ in Tennyson’s poems through a quantitative method, the latent Dirichlet allocation (LDA) topic model (Blei et al., 2003).

The basic idea behind LDA is that ‘documents are represented as random mixtures over latent topics, where each topic is characterised by a distribution over words’ (Blei et al., 2003: 996). LDA classifies words into groups, which are called topics. As Tabata (2018: 52) indicates, ‘LDA is a machine learning method for uncovering hidden semantic structures in a corpus of texts’. The main advantage of using LDA is that this approach allows to find semantic connections between words in texts, which other quantitative methods like keyword analysis, find it difficult to identify. LDA is especially dedicated to analysing big data, and it allows us to read text data from a ‘distance’. Using probabilistic calculations, LDA detects the semantic connections hidden behind words in the target corpus. Since LDA is more efficient algorithm on big data (Iwata, 2015), several studies have employed the topic model to examine prose texts (Onodera et al., 2016; Kuroda, 2017; Tabata, 2017, 2018, 2020; Kiyama, 2018; Matsukawa et al., 2018; Huang, 2020a, b), yet relatively few have applied the technique to analyse poetry (Rhody, 2012; Navarro-Colorado, 2018; Henrichs, 2019; Okabe, 2019). Although this might be perceived as a challenge for applying LDA on poetry, which is basically short-length compared to prose texts, Fujita (2022) shows that LDA is effective in considering Tennyson’s poetic works. Referring to Fujita (2022), this study investigates the lexicon of Tennyson’s 603 works, focusing in particular on the nouns he uses. The emerging LDA results show that some topics (semantically classified word groups) appear in multiple poems. For example, a topic represents the factors of ‘immortality’, one of the key themes in Tennyson. Further, this method reveals multiple associations with the word ‘man’ throughout his works. New ‘unread’ dimensions can be uncovered by a balanced combination of close reading and emerging results of quantitative analysis.

2. Methodology and data

2.1 LDA topic model

The LDA topic model calculates probabilities and classifies words into topics

based on the hypothesis that documents should be classified into several groups based on word co-occurrence (in the same document) trends. Here 'document(s)' is defined as consecutive segments of text data for analysis but that are not always relevant to a poetry/prose work (see Section 2.2 for more detailed explanation on slicing texts into consecutive segments of an equal size).

Figure 1 shows that, in running LDA, the passage at the left of the figure is analysed and the words are classified into groups called 'topics', and four topics as well as their constituent keywords are presented at the right of the figure.

LDA adopts a *bag of words* model, which means word order is ignored, unlike collocation and *n*-gram, which primarily take into account the order of words in a row. As mentioned previously in the first section, LDA uncovers a semantic connection between words; however, LDA does not draw on a semantic classification dictionary as Wmatrix (Rayson, 2009) uses. Therefore, it is the analyst who has to interpret outputs of LDA by combining their knowledge about the text data and assigns labels for each topic, for example, 'Arts', 'Budgets', 'Children' and 'Education' in Figure 1.

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.	"Arts"	"Budgets"	"Children"	"Education"
	NEW MILLION CHILDREN SCHOOL FILM TAX WOMEN STUDENTS SHOW PROGRAM PEOPLE SCHOOLS MUSIC BUDGET CHILD EDUCATION MOVIE BILLION YEARS TEACHERS PLAY FEDERAL FAMILIES HIGH MUSICAL YEAR WORK PUBLIC BEST SPENDING PARENTS TEACHER ACTOR NEW SAYS BENNETT FIRST STATE FAMILY MANIGAT YORK PLAN WELFARE NAMPHY OPERA MONEY MEN STATE THEATER PROGRAMS PERCENT PRESIDENT ACTRESS GOVERNMENT CARE ELEMENTARY LOVE CONGRESS LIFE HATTI			

Figure 1. Example article from the AP corpus (adapted from Blei et al., 2003: 1009, modified)

2.2 Data and settings for analysis

This study analyses 603 poetical works written by Tennyson. Each work was collected as it appeared in a first edition or, if not published, as first written. Of the 603 poems, 423 were compiled as part of the Delphi Poets Series 'Alfred, Lord Tennyson' (2013), and the other 180 were compiled by Ricks (1987). Minute but detailed emendation is given manually on digital documents converted by optical character reader/recognition. Table 1 shows statistical data of the Tennyson corpus.

Prose texts have been the subject of previous LDA studies. The number of studies

Table 1. Statistical data of Tennyson corpus

The number of poems	603
Total tokens	358,190
The shortest poem in number of words	12
The longest poem in number of words	26,383
Mean tokens per poem	594.01
Standard deviation	1,823.15

on poetry texts is extremely few compared with studies dealing with prose works. Poetry texts typically have fewer tokens per work than prose, which in part accounts for why there are not many studies of poetry using the LDA method. Further, LDA requires a significant number of documents for its application. Although the data size of the Tennyson corpus is not as large as those of some other prose text studies, the data for the topic model are considered sufficient, not only because the Tennyson corpus contains 603 poems but also because the total token of the corpus is over 350 thousand by reason of Tennyson’s several lengthy poems, such as *The Princess*, *In Memoriam* and *Maud* (1855).

In LDA, which ‘treats each document as a *bag of words*’, document size deviation is equivalent to bag size variation, and ‘the bigger the bag, the more words that will tend to be found together in the same bag’ (Jockers, 2014: 137). With regard to the study of prose text, Jockers (2014: 137) explains that LDA captures ‘some themes that run throughout and others that appear at specific points and then disappear in novels’ and that ‘it is useful to divide novels (and other large documents) into *chunks* or *segments* and then run the model’. Prose texts from earlier research that employed LDA, have been sliced into segments of equal size, such as 1,000 (Kuroda, 2017; Tabata, 2020), 2,000 or more words (Huang, 2020b). Splitting the poems in the Tennyson corpus into equally sized segments is preferable considering the lengthy poems in the corpus. Further, the large disparity in *bag* size, namely segment size, instantly affects the results since LDA uses raw word frequency to identify the word co-occurrence. In Table 1, the standard deviation value shows that Tennyson’s poems are of varying lengths. The high degree of variation in total word tokens in texts represents a potential methodological issue: a larger number of words may cause over-representing the topic since longer texts are more likely to contain a larger number of words belonging to a particular topic, compared with texts with an equal proportion of words for the same topic. Therefore, it is necessary to ensure texts to be fed into the

topic modelling of a reasonably comparable size, if not exactly the same. In regards to these concerns, Fujita (2022) suggests the relevant size of the segment for LDA practice in Tennyson's poems. This paper refers to Fujita (2022) and adopts a segment size of 594 words by the mean token value of the Tennyson corpus (Table 1). A segment is not exactly the same to one poem, but they correspond. For the poems that contain more than 594 words, for example, *The Princess*, of which total tokens are 26,526, is split into 45 consecutive segments, sequentially counting the number of tokens from the beginning (first token) to the end of the poem. When each text was divided into equal-sized consecutive segments, the two final parts were joined unless the final chunk was 12 word-long, the same length with the shortest poem in the Tennyson corpus (Table 1). Additionally, the poems with total tokens of fewer than 594 words will not be divided into duplicate segments but will be treated as a segment per work. Thus, the largest segment size is 606-word length, and the smallest segment size consists of 12 words.

All words in texts were assigned part-of-speech tags using CasualConc 2.0.8 (Imao, 2022). This paper used a tagset called CLAWS5, as given in the British National Corpus. In order to consider prominent concepts, themes and subjects (all of these can be incorporated into topics discovered), the present study confines its scope to nouns used in the texts in the hope that nouns are more likely to encapsulate the ideational content of a text than other parts-of-speech. After the works were separated into 594-word consecutive segments, other words besides nouns were deleted using the part-of-speech tags; thus, LDA only analyses the target nouns.

LDA was applied to the segments using the MACHine Learning for Language Toolkit (McCallum, 2002). The number of topics was set at 20 based on consideration of emerging results of prior experimental trials, with the number of topics ranging from 10 to 200.

3. Results of LDA

This section discusses the results of LDA for the Tennyson corpus. Table 2 shows one of the output results of topic modelling, indicating the labels and alpha (α) value attached to each of the 20 topics, and the top 20 most salient keywords for each topic. The keywords are ranked in descending order of their weights from the top to the

bottom of the table. The α value represents the universality of each topic. The higher the value, the more prominently the topic appears in multiple works; the lower the value, the topic appears in a small number of works, sometimes only in a single work. In Table 2, the most prominent topic with the highest α value is Topic 17 ($\alpha = 0.88347$). In contrast to Topic 17, Topic 2's α value ($\alpha = 0.03632$) is the lowest among the 20 topics. Thus, it appears that the top 20 keywords of Topic 17 are the most significant elements, which appear in multiple works within the target corpus.

The labels in Table 2, are not the results that are automatically suggested by the LDA analysis. The author of this paper considered each label in terms of most relevant through a meticulous reading of the poems and the results. Table 2 shows that some topics are included with similar labels, but the keywords of these topics are somewhat different. Topics 17 and 7, the most significant topics, Topics 0 and 1 and Topics 6 and 18 have partial aspects in common. Topics 17 and 7 are both labelled as 'Life,' but Topic 17 is more about 'materials' in life, while Topic 7 relates to 'emotion'. Topics 0 and 1 are twin topics referring to the 'environment' depicted in the poems, but what separates the two topics is whether the 'environment' is near *sea* or *river*. While previous studies mainly focus on *sea* among waters as mentioned earlier, another watery element in nature, *river*, as well as *sea*, was detected from the LDA results. Some topics, Topics 11, 13 and 19, are about women, which, despite having α values that are not high, are associated with female protagonists and help support earlier studies that have highlighted women as constituting a dominant theme in Tennysonian poems.

Whereas Topics 11, 13 and 19 were found for females, Topics 6 and 18 both have males in their labels. Topic 6 refers to a *man* and his association with family members, while Topic 18 reflects relations between *man* and the society to which he belongs. The following section discusses Topics 17, 7, 0, 1, 6 and 18 as three pairs that correspond to each other. With the six topics, the next section further investigates whether other poems that have not been included in the discussion of previous studies have nonetheless been found to contain identical or similar themes as in canonical works, as well as other latent topics that have not gained much scholarly attention.

Table 2. A result of running LDA on the Tennyson corpus; 20 topics with their labels, alpha values and keywords

LABELS	Environment (sea)	Environment (river)	Lives in poems	Religious elements	Nature & beauty	Marriage	Man in family	Life (emotion)	War	Christianity
TOPICS	0	1	2	3	4	5	6	7	8	9
ALPHA VALUES	0.15124	0.07165	0.03632	0.05559	0.23016	0.05058	0.55437	0.77814	0.18978	0.09134
KEY WORDS (1-20)	sea	river	poet	gods	year	bride	man	love	war	fool
	isle	lake	goose	death	rose	moor	child	life	land	flesh
	sail	valley	reviewers	cannon	garden	woman	mother	heart	name	god
	boat	mountain	laurel	gleam	song	ruin	heart	death	battle	sin
	seas	pine	peacock	right	summer	gold	father	soul	glory	fire
	shore	palm	weather	children	flowers	fleet	hand	spirit	throne	mark
	ship	boughs	moment	league	leaves	hall	face	mind	men	priest
	water	brook	world	mountain	flower	cave	wife	eyes	hearts	ringlet
	sand	poplar	volume	roof	leaf	bridegroom	day	thought	voice	church
	melody	fern	stores	anger	birds	ward	life	hope	fame	bread
	ocean	dews	rhyme	valley	air	tide	men	light	freedom	wine
	rocks	fruit	drop	banner	roses	motion	house	tears	trumpet	beauty
	cliff	leaf	hammer	left	love	signs	head	pain	blood	hell
	world	water	chorus	master	morning	cap	boy	memory	kings	friend
	foam	bowers	poverty	brigade	spring	text	children	grief	sons	swine
	blast	grass	sow	charge	winter	fool	son	words	strength	brute
	hills	middle	throng	darkness	tree	beast	word	hour	power	soul
	bay	arches	eggs	people	wood	trifle	woman	things	shame	cross
	moon	summer	laughter	earthquake	woodland	wife	hands	brain	people	saints
	vessel	knees	pint	foe	lawn	water	night	blood	friends	shame
LABELS	Abstracts related to life	Lady & prince	Abstracts on human	Ladies	Object of love	Strength of nature	Nobles	Life (materials)	Man in society	Mother & queen
TOPICS	10	11	12	13	14	15	16	17	18	19
ALPHA VALUES	0.09013	0.0799	0.07364	0.05834	0.21145	0.19772	0.20564	0.88347	0.50529	0.04247
KEY WORDS (1-20)	time	lady	fame	princess	eyes	eye	king	night	man	mother
	sense	prince	charm	woman	love	soul	queen	light	men	queen
	prime	horse	music	prince	heart	earth	knight	day	world	farm
	throne	arms	woman	lady	beauty	woe	knights	sun	time	green
	spirit	knight	dream	women	tears	pride	hall	heaven	things	grave
	blue	wood	note	storm	cheek	air	eyes	voice	truth	dear
	angel	damsel	faith	form	lips	heaven	man	wind	earth	gate
	waters	knave	friend	riflemen	smile	sky	men	eyes	power	brook
	place	rest	deeds	college	morn	flame	maid	earth	peace	rave
	things	hall	birdie	girls	hair	ray	name	fire	days	cow
	depth	squire	others	boys	feet	hills	shield	star	people	maiden
	river	bridge	wits	honour	neck	gale	table	moon	years	new-year
	airs	court	devil	highness	rose	waves	quest	head	race	fellow
	eddies	walls	mood	men	arm	course	sword	time	nature	ill
	glooms	armour	lute	head	kiss	skies	realm	sound	age	cause
	eternity	charger	belt	ladies	brow	realms	court	shadow	life	dale
	raiment	shield	books	south	summer	view	field	stars	times	chestnut
	sides	<small>kitchen-knave</small>	boon	echo	sighs	light	face	death	hands	alleys
	impulse	fight	counsel	court	brows	might	hands	way	faith	daisies
	immortality	helmet	baby	lisette	kisses	glance	word	gold	land	bailiff

4. Discussion

This section discusses whether the LDA results detected any other poems that previous studies have not discussed in their analysis of specific themes as well as other latent topics that scholars have yet to focus on. Findings for the paired Topics 17 and 7 (section 4.1) show that immortality, one of Tennyson's themes, is depicted not just in his canonical works. In fact, Topics 0 and 1 (section 4.2) highlight *sea* as a significant motif in Tennyson's poems that have not been mentioned in previous research along with *river* as a prominent subject that scholars have rarely discussed. Topics 6 and 18 (section 4.3) illustrate how male protagonists are portrayed, whereas females have been more commonly discussed.

4.1 Topics 7 and 17: Emotions and materials related to life

Based on the analysis of the outputs, it was found that Topics 7 and 17 are the most significant topics in the corpus. The highest α value ($\alpha = 0.88347$) was associated with Topic 17, while Topic 7 had the second highest α value ($\alpha = 0.77814$) among the 20 topics, slightly less than that of Topic 17. In terms of the top 20 keywords, Topic 17 mainly comprises physical nouns and nouns pertaining to tangible objects perceived through the senses, such as those related to the eyes, ears or skin. Moreover, these nouns represent entities closely associated with human lives (see Figure 3). On the other hand, the keywords of Topic 7 are relatively close to human lives but are mainly abstract nouns linked to people's emotions or thoughts (Figure 2). *Death* is the keyword of both Topics 7 and 17. Words connected to death, such as *grief*, *spirit*, *soul* and *memory*, are keywords of Topic 7, *heaven* is a keyword of Topic 17. Some of these



Figure 2. Word cloud of top 20 keywords
(Topic 7)



Figure 3. Word cloud of top 20 keywords
(Topic 17)

words may not be immediately linked to death, like *soul* or *memory*, but the connection becomes clear when considering notable poetic works related to this topic.

Figures 4 and 5 show the top 50 segments of density of Topics 7 and 17, respectively. Topic density represents the proportion that a topic accounts for in a segment. Segments with higher values have greater topic concentration. In Figures 4, 5 and the forthcoming density plot figures, the top 50 segments, in which the topic densely appears the most, are located horizontally, and the vertical axis shows the topic proportion. Poems containing over 594–606 words were sliced into multiple segments before LDA was implemented; therefore, if a topic significantly appears in multiple segments of one poem, the topic's proportions are located plurally on a vertical line of the poem. Additionally, when more than three segments of a poem are within the top 50, a box plot appears, and the median value is indicated by a black line inside the box. In these poems, to which the top 50 segments of Topics 7 and 17 belong to, the poet frequently writes about the death of people who are loved and are close¹. One of Tennyson's most representative works, *In Memoriam*, is not an exception and is often cited in discussions surrounding immortality in the context of death depicted in a work not only because Tennyson starts with the word 'immortal' right at the very beginning, as quoted below, but also because the poem expresses his lamentation over his best friend's death.

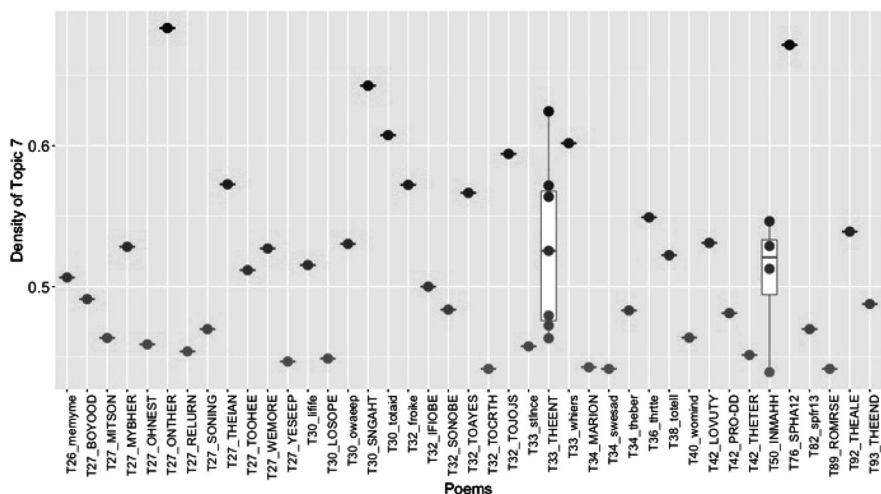


Figure 4. Density bar plot of top 50 segments of Topic 7¹

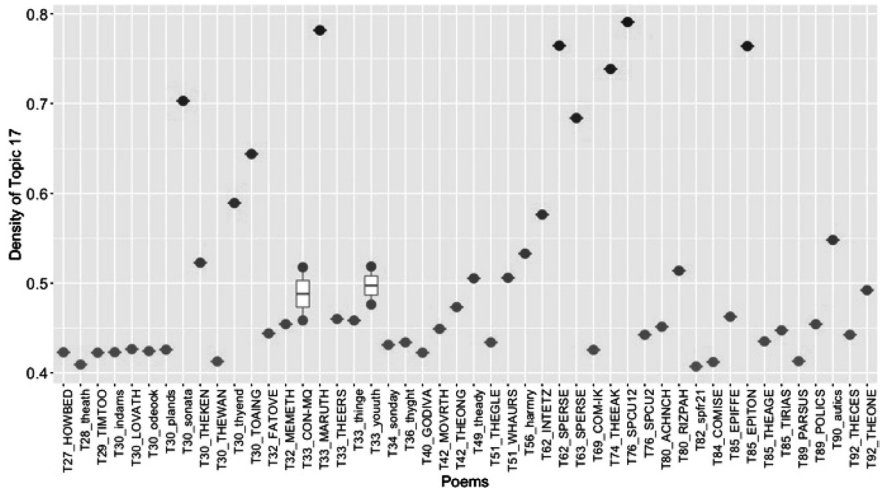


Figure 5. Density bar plot of top 50 segments of Topic 17¹

Strong Son of God, **immortal** Love,

Whom we, that have not seen thy face,

By faith, and faith alone, embrace,

Believing where we cannot prove;

(*In Memoriam A.H.H.*, 1850: ll. 1–4; bold added)

The meaning of immortality that is relevant to this poem is ‘[t]he condition of being celebrated through all time; enduring fame or remembrance’ (*Oxford English Dictionary*; *OED*, online, s.v. immortality, *n.* 2.); thus, it does not necessarily mean ‘the endless life or existence’ (*OED*, s.v. immortality, *n.* 1.). In *In Memoriam* and other significant poems under Topic 7 mention someone who has already passed away; therefore, the focus is not on a person who wishes to obtain eternal life or the depiction of everlasting life, unlike in Tennyson’s poem, ‘Tithonus’, which is often cited in discussions of immortality. Immortality in ‘Tithonus,’ or in its early original version ‘Tithon’ (1833), named after a figure from the Greek mythology, is more likely to be seen as expressing the desire for an endless existence. Further, it is not necessarily limited in its meaning of ‘enduring fame or remembrance’ after someone’s death. In fact, the majority of works in Topic 7 explore the death or loss of loved ones, but that is

a cue to thinking, remembering and recollecting people who are no longer with the other characters in the poems to some extent. These sensitivities seem to last forever for the people left behind, especially in the immediate moment of grief and despair over a loss. The feelings likely to be eternal are thus considered a form of immortality in the broad sense. While it is possible to identify two different definitions of immortality in Tennyson's poems, it is suggested, based on the results of LDA, that the poet's main concern is 'enduring fame or remembrance' for someone who has passed away or has simply left the other one's place.

In the prominent works of Topic 7, immortality is communicated through words like *memory*, *spirit* and *soul*, as well as through a contrast between alive and dead, in addition to using the exact words *immortality/immortal*. *In Memoriam*, 'Crossing the Bar' and 'Tithonus/Tithon' have often been the poetic works that are cited in discussions of immortality in various literary studies (Elliot, 1979; Perrine, 1966; Shaw, 1976; respectively). However, Topic 7 reveals that sense of immortality was a more intriguing concern for Tennyson from his early career until the end of his life, and the theme appears in different poems. Following are two extracts of poems in which Topic 7 significantly appears, 'To One Whose Hope Reposed on Thee' (1827) and 'The Death of the Duke of Clarence and Avondale' (1892):

To one whose hope reposed on thee,
Whose very life was in thine own,
How deep a wound thy death must be,
And the wild thought that thou art gone!

[...]

And **still** I hear the tolling bell,
For Memory makes each sense her own.
But **stay**, my soul! thy plaint forbear,
And be thy murmuring song forgiven!
Tread but the path of Virtue here,
And thou shalt meet with her in **heaven!**

('To One Whose Hope Reposed on Thee', 1827: ll. 1–4, 23–28; bold added)

THE bridal garland falls upon the bier,

The shadow of a crown, that o'er him hung,
Has vanish'd in the shadow cast by Death.

[...]

The face of Death is toward the Sun of Life,
His shadow darkens earth: his truer name
Is 'Onward,' no discordance in the roll
And march of that **Eternal** Harmony
Whereto the worlds beat time, tho' faintly heard
Until the great **Hereafter**. Mourn in hope!

(‘The Death of the Duke of Clarence and Avondale’, 1892: ll. 1–3, 12–17; bold added)

The high α value of Topic 7 suggests its universality in the corpus. Simply put, its elements generally appear in multiple poems. The top 50 segments of Topic 7, as shown in Figure 4, indicate diverse dates of publication, including a considerable number of works from Tennyson’s early career. As previously noted, *In Memoriam* and some other poems have been the main works cited in discussions surrounding immortality in Tennyson, but the present results suggest that the poet dealt with immortality in various other poems. Furthermore, it is one of the most familiar themes in his characters’ lives along with love and death, which are also always closely linked to human lives.

4.2 Topics 0 and 1: *Sea and river*

Topics 0 and 1 have keywords that describe entities related to water, such as *sea*, *river* and *lake* (Figures 6 and 7). The keywords also describe the surroundings of bodies of water in detail. These two topics specifically consist of words related to water in various natural settings, and one of the reasons these keywords to appear in the two topics can be explained by previous studies. Okazawa (1969) points out Tennyson’s particular love for and exploration of bodies of waters in nature, such as seas, lakes and rivers. Hair (1991: 42) argues Tennyson was ‘a landscape-painter in words, a colourist’, pointing out his skill at depicting sceneries through words. Tennyson succeeded in describing and conveying highly detailed scenes to readers, using words, which appear in Topics 0 and 1. Although the LDA results indicated that Tennyson illustrated the things around seas and rivers, previous studies have not frequently argued rivers in



Figure 6. Word cloud of top 20 keywords (Topic 0)

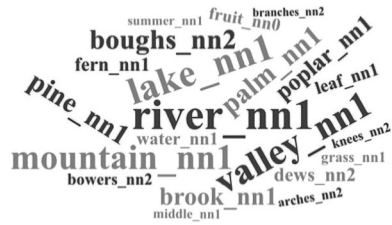
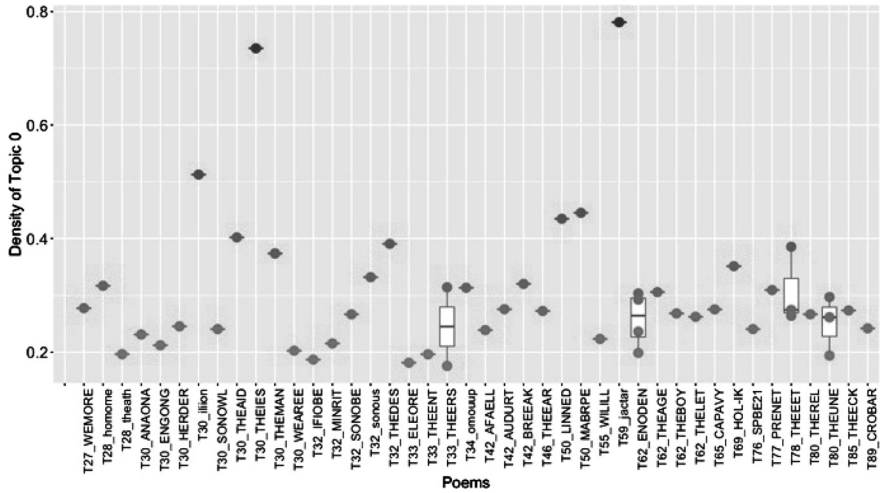
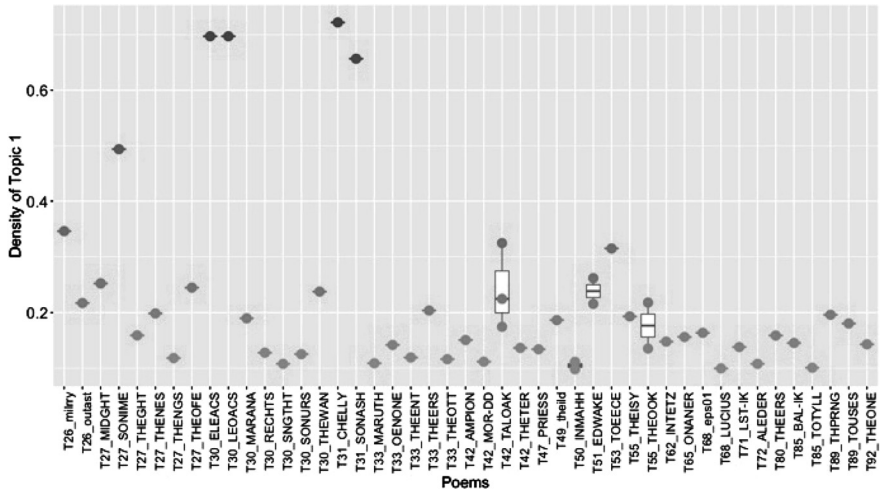


Figure 7. Word cloud of top 20 keywords (Topic 1)

Tennysonian poems. Furthermore, in Topic 0, geographic features are expressed in words such as *shore*, *sand* and *rocks*, which are closely related to the sea, to represent landscapes exquisitely. Watercraft, for instance, *boat*, *ship* and *vessel* also appear, as well as the terms *world* and *seas*, which refer to regions/areas beyond the sea. One possible reason for the appearance of words like some different watercraft, *world* and *seas* in this topic is that Britain is an island country, and to go abroad means to ‘cross the sea’. In contrast, in Topic 1, besides *rivers* and *lakes*, words appear that communicate geographic features, such as *valley* and *mountain*, and also words related to plants, such as *pine*, *poplar*, *leaf* and *grass*. However, words associated with the world or with foreign regions/areas do not appear among the top 20 keywords in Topic 1. This suggests that one of the factors separating Topic 0 and Topic 1 from *rivers* and *seas* is the peculiar nature of Britain, which the sea separates from foreign nations.

The top 50 segments of density of Topics 0 and 1 are shown in Figures 8 and 9, respectively. Based on a close reading of the poems, which the segments appertain to, it is possible to deduce that Tennyson tends to represent feelings of awe, fear, loneliness and despair in the texts given to Topic 0. In scenes where the sea appears, the sound of the waves is almost the only sound that can be heard. The openness of the sea and the absence of any natural or artificial surroundings are also expressed. This nothingness around the person in the scene further amplifies the feelings of loneliness, despair or fear.

Moreover, war is most significant scenes in ‘Jack Tar’ (circa 1859)² and ‘The Revenge: A Ballad of the Fleet’ (1878). In these two poems, *sea* is depicted as the place where the war/battle occurs. Combined with images of shipwrecks and war, the sea is associated with the notion of inescapable, unwilling and lonely death. For years,

Figure 8. Density bar plot of top 50 segments of Topic 0¹Figure 9. Density bar plot of top 50 segments of Topic 1¹

scholars have considered *sea* in Tennyson as an attractive theme; however, significant poems in Topic 0, such as ‘Jack Tar’, ‘The Revenge: A Ballad of the Fleet’, ‘Ilion, Iliion’ (1830), ‘Lines [Here Often, When A Child, I Lay Reclined]’ (1850) and ‘Mablethorpe’ (1850), have seldom been cited in the discussion of *sea*. The LDA

results suggest a new aspect in some poems outside Tennyson's canonical or renowned works contain *sea* as their theme.

In contrast, in Topic 1, the anthropomorphic plants and animals talk, sing and surround the people in the poems. Unlike in Topic 0, living creatures and sounds are present in the poems. These works do not convey the impression of fear or despair that is felt in Topic 0's works. This therefore suggests not only the reflection of the characters' emotions but also Tennyson's intention to sketch and evoke these emotions not directly but in an indirect way.

4.3 Topics 6 and 18: *Men's various whereabouts*

The keywords of both Topics 6 and 18 designate *man/men* (Figures 10 and 11). This suggests that the topics are about *men* and some other people and objects that relate to/surround them. The keywords in Topic 6 are words related to *man* and to a man's family or household, such as *child*, *mother* and *wife*. The top keyword of Topic 18 is also *man*, but the plural form, *men* and *world* closely follow it. While Topic 6's keywords mainly have a close connection to family and familial life, Topic 18 contains keywords related to man and the world/society outside the home. Among the top 50 segments in which Topics 6 and 18 appear prominently (Figures 12 and 13), nine works in Topic 6 and three works in Topic 18 (within red lines in Figures 12 and 13) are narrative poems (Table 3).

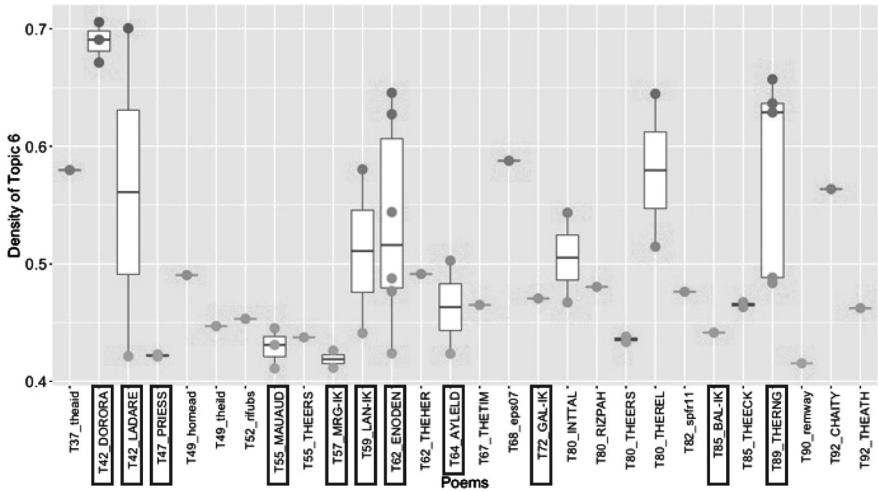
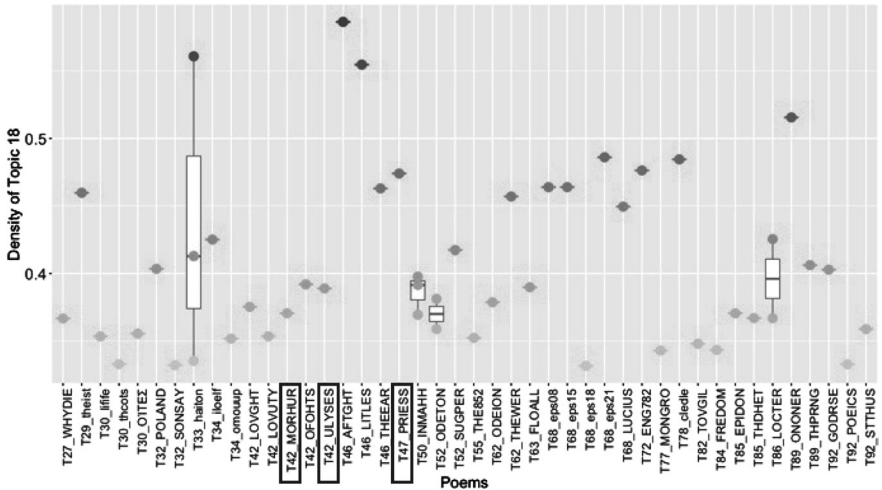
In the narrative poems of Topic 6, the main characters are female, and their relations with families are presented. In contrast, the main characters in 'Morte d'Arthur' and 'Ulysses' (1842) are powerful, sword-fighting males. The poem titled



Figure 10. Word cloud of top 20 keywords (Topic 6)



Figure 11. Word cloud of top 20 keywords (Topic 18)

Figure 12. Density bar plot of top 50 segments of Topic 6¹Figure 13. Density bar plot of top 50 segments of Topic 18¹

The Princess is common to both Topic 6 and Topic 18. The protagonist of *The Princess* is a female, but this differs from the other works in these two topics. One of the poem's main themes is, to borrow the words of other scholars, 'women's role in society' (Clapp-Itnyre, 2000: 227) and 'literary and social inheritance as well as relations

Table 3. Narrative poems in top 50 segments of Topics 6 and 18

Titles of poems	Date of publication	Topic(s) appearing significantly
'Dora'	1842	Topic 6
'Lady Clare Vere de Vere'	1842	Topic 6
<i>The Princess</i>	1847	Topic 6 & Topic 18
<i>Maud</i>	1855	Topic 6
'The Marriage of Geraint'	1857	Topic 6
'Lancelot and Elaine'	1859	Topic 6
<i>Enoch Arden</i>	1862	Topic 6
'Aylmer's Field'	1864	Topic 6
'Balin and Balan'	1885	Topic 6
'The Ring'	1889	Topic 6
'Mort d' Arthur'	1842	Topic 18
'Ulysses'	1842	Topic 18

between the sexes' (Wright, 2015: 251). More directly, it can be characterised as gender equality. While the central theme of *The Princess* is equity between the genders, the poem at the same time reflects the Victorian attitudes toward the roles of men and women (Jimbo, 2009), as the lines below show.

Man for the field and woman for the hearth:
 Man for the sword and for the needle she:
 Man with the head and woman with the heart:
 Man to command and woman to obey;
 All else confusion.

(*The Princess*, 1847: ll. 2391–2396)

In the lines above, a contrast is presented between how men and women 'should be/should do', and men are described as strong and wise individuals who can control women. *Enoch Arden*, which is also a significant work of Topic 6, portrays the behaviour of the main female character, Annie, toward her husband. Annie, the wife of the protagonist, Enoch, has never disobeyed her husband in the seven years of their marriage, but in the scene presented below, she opposes her husband, who is going to leave her and their family behind to go to the Far East and make a fortune:

Then first since Enoch's golden ring had girt
 Her finger, Annie fought against his will:
 Yet not with brawling opposition she,
 But manifold entreaties, many a tear,
 Many a sad kiss by day by night renew'd
 (Sure that all evil would come out of it)
 Besought him, supplicating, if he cared
 For her or his dear children, not to go.

(*Enoch Arden*, 1862: ll. 157–164)

These lines in Tennyson's poems illustrate the position and roles of men and women in eighteenth- and nineteenth-century England: men were strong, went outside the home and existed at the centre of society, while women followed the men and stayed inside the home. The exact time period of the stories in *The Princess*, *Enoch Arden* and other poems listed in Table 3 are not necessarily mentioned directly; therefore, this paper does not assume that the historical setting is the Victorian era, but we can presume that it is safe to say these poems reflect the attitudes of society at that time and men's and women's roles during/prior to that age. Both Topic 6 and Topic 18 can be considered as suggesting the different positions of *men* and *women* and reflecting a male-dominated society in the poetic works. Additionally, Topic 6 includes the keyword *wife*, which refers to the spouse of a male, but does not include *husband*, which refers to the spouse of a female. These results indicate that Topic 6 is a male-centred topic about the home as well as Topic 18 is about male empowerment in society (especially in the context of sword-toting times/wars). Tennyson's poems tend to portray *men* as central to the home and society in/before Victorian times, and LDA detected different roles of *men* in the poems, thus Topics 6 and 18 are separated.

5. Conclusion

This study presented findings on the style of Tennyson's poetry using the LDA topic model. It showed that the latent topics behind the prominent elements of Tennyson's 603 poems can bring to light new dimensions through a balanced combination of close reading and quantitative analysis. The LDA results revealed

topics that reflect the notion of immortality, one of the major themes in his body of work as well as the poet's affinity for themes related to *sea* and *river* and to the Victorian era's social attitudes toward *men* and *women*. Through the stylometric approach, LDA enabled us to discover new aspects of Tennyson's poems by reading from a 'distance', which offers a different perspective than that of previous studies. However, this paper only investigated Tennyson's works. To examine whether the findings are peculiar to Tennyson or not, future research will expand the analysis to compare the results with various other poets' works, for example, William Wordsworth, John Keats and Robert Browning. All in all, through a quantitative approach, LDA revealed previously 'unread' parts of Tennyson's works, although there remain the other topics not addressed in this study. More research is needed to explore these topics lest we leave them 'unread'.

Notes

1. File titles appearing in figures comprise three parts: 'T' represents Tennyson; the numbers that follow reflect the last two digits of the year of publication (writing); the letters after the underscore denote the abbreviated poem title. The abbreviated poem titles in file titles can be distinguished in the following manner: works published during Tennyson's lifetime are inscribed in upper case; poems published posthumously or unpublished are inscribed in lowercase letters. For example, T26_memyme alludes to 'Memory [Ay me!]' written in 1826, whereas T27_BOYOOD refers to 'Boyhood', which was published in 1827 when Tennyson was alive. The first line of the poem and/or Roman numerals are encased in brackets to differentiate discrete poems with identical titles. Correspondence tables of abbreviated file titles and original poem titles presented in Figures 4, 5, 8, 9, 12 and 13 (Topics 7, 17, 0, 1, 6 and 18) are available online: <https://tennysondh.wordpress.com/figures-and-tables-with-their-titles/>
2. On 14 May 1859, 'Mr Peel took up Jack Tar to London; but A[lfred]. T[ennyson]. decided not to publish it' ([Hallam Lord Tennyson], *Mat[erials for a Life of A.T.]*. ii 217) (Ricks, 1987: 604).

Acknowledgement

This paper is a revised version of the author's presentation at PALA 2022 *Style and Sense(s)*, July 8, at Aix-Marseille Université. I would like to thank the chair and

audience members for useful discussion at the conference. This work was supported by Japan Science and Technology Agency Support for Pioneering Research Initiated by Next Generation [grant number JPMJSP2138]. I would also like to thank two anonymous reviewers for their insightful comments and suggestions. I own responsibility for other errors.

References

- Blei, M. D., Y. A. Ng., and I. M. Jordan (2003) “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993–1022.
- Bloom, H. (Ed.) (1985) *Modern Critical Views: Alfred, Lord Tennyson*. New York: Chelsea House.
- Bruce, H. (1917) “Tennyson and Death.” *The Sewanee Review* 25, 4: 443–456.
- Clapp-Itynre, A. (2000) “Marginalized Musical Interludes: Tennyson’s Critique of Conventionality in The Princess.” *Victorian Poetry* 38, 3: 227–248.
- Delphi Poets Series (2013) *Alfred, Lord Tennyson*. East Sussex: Delphi.
- Dixon, M. W. (1896) *A Primer of Tennyson*. London: Methuen & Co.
- Elliot, P. (1979) “Tennyson and Spiritualism.” *Tennyson Research Bulletin* 3, 2: 89–100.
- Fujita, I. (2022) “On Segment Size in Poetry Analysis Using the Latent Dirichlet Allocation Method.” *The Japanese Journal of Digital Humanities* 3, 1: 3–15.
- Hair, S. D. (1991) *Tennyson’s Language*. Toronto: University of Toronto Press.
- Henrichs, A. (2019) “Deforming Shakespeare’s Sonnets: Topic Models as Poems Author(s).” *Criticism* 61, 3: 387–412.
- Holmes, J. (2012) “‘The Poet of Science’: How Scientists Read Their Tennyson.” *Victorian Studies* 54, 4: 655–678.
- Horiuchi, T. (1992) “Auditory Value in *Enoch Arden*.” *Sendai Sirayuri Junior College Bulletin* 21: 47–62.
- Huang, C. (2020a) “Chugoku-no Misuterii Shousetu-ni Okeru Topikku Kaiseki-no Kokoromi (Experimental Topic Analysis on Chinese Mystery Prose Texts).” *Studies in Language and Culture Osaka University 2020*: 1–17.
- Huang, C. (2020b) “Chugoku-no Misuterii Shousetu-wo Meguru Keiryouteki Bunseki: Tei Shou Sei to Ki Ba Sei-no Sakuhin-wo Chushin-ni (Quantitative Analysis of Chinese Mystery Novels: Focusing on the Works of Cheng Xiao Qing and Gui Ma Xing).” *Text mining and Digital Humanities, Graduate School of Language and Culture, Osaka University 2019*: 31–45.
- Imao, Y. (2022) CasualConc (Version 3.0.2).
Available at: <https://sites.google.com/site/casualconccj/casualconc/CasualConc>
- Iwata, T. (2015) *Topic Model*. Tokyo: Kodan-sha.
- Jockers, L. M. (2014) *Text Analysis with R for Students of Literature*. Heidelberg, New York,

Dordrecht, London: Springer Cham.

- Jimbo, A. (2009) "Victoria-chou Shakai-no Daibensha-toshiteno Tenisun: *In Memoriam* Shuppan-no Kunou-wo Hete Eta Shousan (Tennyson, as an Opinion Leader of the Victorian Age: The Way He Won General Applause through the Publication of *In Memoriam*)." *The Bulletin of the Graduate School of Education of Waseda University* 17, 1: 355–364.
- Kabata, T. (2001) "'Nothing' in Tennyson's Poems." *Bulletin of Yonezawa Women's College* 36: 21–31.
- Kabata, T. (2007) "'Sun' in Tennyson's Poems." *Bulletin of Yonezawa Women's College* 42: 9–20.
- 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.
- Kuroda, A. (2017) "Quantitative Analysis of Literary Works: Novels of Sir Arthur Conan Doyle." *Text mining and Digital Humanities, Graduate School of Language and Culture, Osaka University 2016*: 23–41.
- McCallum, A. K. (2002) MALLET: A MACHINE Learning for Language Toolkit. Available at: <http://mallet.cs.umass.edu>.
- Matsukawa, H., M. Oyama, C. Negishi, and Y. Arai (2018) "Analysis of the Free Descriptions Obtained through Course Evaluation Questionnaires Using Topic Modeling." *Japan Journal of Educational Technology* 41, 3: 233–244.
- Moretti, F. (2013) *Distant Reading*. London: Verso.
- Nakamura, M. (1967) "Tennyson and his Christianity." *The Journal of Otemae Women's University* 1: 76–87.
- 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.
- Nishimae, Y. (1979) *Tenisun Kenkyu (Tennyson Research)*. Tokyo: Chuo-Shuppan.
- Nishimae, Y. (2000) *Tenisun-no Gengo Geijutsu (Tennyson's Linguistic Art)*. Tokyo: Kaibun-sya.
- Noguchi, T. (2011) "Tennyson's Ulysses in the Dramatic Monologue: The Dramatization of Visions Through the Wreck." *Hokusei Review, the School of Humanities* 48, 2: 1–11.
- 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.
- Okazawa, T. (1969) "A Study of Tennyson's Predilection for Waters." *Bulletin of Chofu Gakuen Women's Junior College* 2: 1–12.
- Onodera, D., L. Huang, and M. Yoshioka (2016) "Classification of New Article by Using Facet-Biased Topic Model and Distance Metric Learning." *The 30th Annual Conference of the Japanese Society for Artificial Intelligence 2016*: 1–4.

- Oxford English Dictionary* (online) <https://www.oed.com> Last accessed date: October 13th, 2022.
- Perrine, L. (1966) “When Does Hope Mean Doubt?: The Tone of ‘Crossing the Bar’.” *Victorian Poetry* 4, 2: 127–131.
- Plamondon, M. R. (2005) “Computer-Assisted Phonetic Analysis of English Poetry: A Preliminary Case Study of Browning and Tennyson.” *TEXT Technology*: 153–175.
- Rayson, P. (2009) Wmatrix: A Web-Based Corpus Processing Environment, Computing Department, Lancaster University. <http://ucrel.lancs.ac.uk/wmatrix/>
- Rhody, M. L. (2012) “Topic Modeling and Figurative Language.” *CUNY Academic Works 2012*: 19–35.
- Ricks, C. (Ed.) (1969) *The Poems of Tennyson*. London: Longman.
- Ricks, C. (Ed.) (1987) *The Poems of Tennyson*. vol. I–III. (Second edition) London: Longman.
- Shaw, W. D. (1973) *Tennyson’s Style*. Ithaca and London: Cornell University Press.
- Shaw, W. D. (1976) “Tennyson’s ‘Tithonus’ and the Problem of Mortality.” *Philological Quarterly* 52, 2: 274–285.
- 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.
- Tabata, T. (2018) “Mapping Dickens’s Novels in a Network of Words, Topics, and Texts.” *Text mining and Digital Humanities, Graduate School of Language and Culture, Osaka University 2016*: 51–60.
- Tabata, T. (2020) “Latent Topics in British Classic Fiction: Using LDA to Classify Texts into Meaningful Groups.” *Text mining and Digital Humanities, Graduate School of Language and Culture, Osaka University 2019*: 47–58.
- Tennyson, H. (1897) *Alfred Lord Tennyson: A Memoir by his Son*. vol. I–III. London: Macmillan.
- Thomas, J. (2019) *Tennyson Echoing Wordsworth*. Edinburgh: Edinburgh University Press.
- Wright, J. (2015) “The Princess and the Bee.” *The Cambridge Quarterly* 44, 3: 251–273.

(藤田 郁 大阪大学院)