The University of Osaka
Institutional Knowledge Archive

Title	テクストマイニングとデジタルヒューマニティーズ 2021 （冊子）
Author(s)	
Citation	言語文化共同研究プロジェクト. 2022, 2021
Version Type	VoR
URL	https://hdl.handle.net/11094/88355
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

言語文化共同研究プロジェクト２０２１

テクストマイニングと
デジタルヒューマニティーズ
2021

黒　田　絢　香
藤　田　　郁
浅　野　元　子
福　本　広　光
田　畑　智　司

大阪大学大学院言語文化研究科

２０２２

言語文化共同研究プロジェクト２０２１

テクストマイニングと
デジタルヒューマニティーズ
2021

黒　田　絢　香
藤　田　　郁
浅　野　元　子
福　本　広　光
田　畑　智　司

大阪大学大学院言語文化研究科

２０２２

言語文化共同研究プロジェクト

テクストマイニングと
デジタルヒューマニティーズ
2021

目　次

田畑　智司　　プロジェクトの目的と活動 1

黒田　絢香　　トピックモデル可視化ツールの開発に向けて 5

藤田　郁　　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— 15

浅野　元子　　学術英語遠隔授業におけるジャンルの認識に関する調査
—日本の大学 1 年生を例に— 39

福本　広光　　米国一般教書演説における分離不定詞についての記述分析
— 1960 年代から 2020 年の用例を中心に— 57

「テクストマイニングとデジタルヒューマニティーズ」プロジェクトの目的と活動

本共同研究は，自然言語処理，コーパス言語学・計量言語学，数理統計学，データマイニング，機械学習など，諸分野の知見を有機的に統合した方法論を開発し，テクストマイニングを応用して人文学，言語文化学の諸問題にアプローチする，すなわち「デジタルヒューマニティーズ (Digital Humanities)」の実践と理論的精緻化の可能性を探る営みである。このプロジェクトは， 2001 年度に岩根 久教授，緒方 典裕助教授，および筆者の 3 名でスタートした「電子化言語資料分析の方法論」を基礎とするが， 2003 年度から名称を一部改め，言語文化研究科の大学院生もメンバーに加わった。2006 年度には三宅 真紀助教の加入を得て，対象言語も英・仏・ギリシャ語に拡がった。2011 年には言語文化教育論講座に着任した今尾 康裕講師が加入した。2014 年度後期から，さらに Hodošček Bor 講師が加わった。そして，2019 年度をもって退職された岩根 久教授の後任として，2020 年度に山田 彬尭講師の着任があり，現在の陣容となっている。(職位はいずれも当時)。2016 年度から，プロジェクトの名称を，当該リサーチコミュニティの名称としてより相応しい「テクストマニングとデジタルヒューマニティーズ」にアップデートしたが，研究の系統は創始時より常に一貫している。

「テクストマニングとデジタルヒューマニティーズ」プロジェクトは大きく分けて二つの層で構成されている。一つは研究基盤となるコーパス，テクストアーカイヴの開発・構築，もう一つは構築したコーパス，テクストアーカイヴからのデータ抽出法研究，並びに得られた高次元の言語データの計量分析である。前者には英・仏語の文学作品や，聖書（共観福音書）などの電子テクスト化，ロシア語政治演説コーパス，近代日本文学コーパスの編纂，マークアップ言語 XML による TEI（Text Encoding Initiative：デジタル化したテクストの国際互換規格の枠組）に準拠したタグ付けなど，人文学資料のデジタル化やマークアップ法，データ符号化方法論の開発などが含まれる。一方，高次元人文学データ分析の事例として，語彙・語法，コロケーション，意味構造，語用論などのレベルにおける言語使用の実態研究，高度な数理モデルや機械学習を応用した言語分析やテクストマイニング，文学作品の言語特徴の特定や，使用域間の言語変異や文体識別問題の考察，著者推定法の精密化研究を挙げることができる。

本プロジェクト班は言語文化研究科の専任教員 5 名と名誉教授 1 名 (今尾 康裕, 田畑 智司, Hodošček Bor，三宅 真紀，山田 彬尭，岩根 久名誉教授)，当研究科博士後期課程在学生 7 名（黒田 絢香，黄 晨雯，岡部 未希，徐 勤，福本 広光，藤田 郁，竹森 ありさ)，文学研究科博士後期課程在学生 1 名（涌井 萌子)，博士前期課程在学生 3 名（王 鈺，曹 芳慧，李 晨婕)，研究生 2 名（Vogatza Dimitra, 李 建蕾）に加え，OG の大阪医科薬科大学 浅野 元子氏（2020 年 3 月博士学位取得）・杉山 真央氏（2019 年 3 月博士学位取得)，本学非常勤講師の高橋 新氏，南澤 佑樹氏（本研究科博士課程修了)，摂南大学 後藤 一章氏（本研究科博士課程修了)，帝塚山学院大学 八野 幸子氏（本研究科博士課程修了)，数理・データ科学教育研究センターの上阪 彩香氏を主たる参加メンバーとしている。研究を遂行するために，コアメンバー以外も自由に参加できる月例の研究会・討論会などを通して，研究情報の交換，論文や開発ツール，構想段階のプロジェクトや進行中のパイロットスタディのプレビューなどを行っている。

2021 年度は，前年度に引き続き，パンデミック下全ての研究会をオンラインで開催した。ほぼ 2 年間，対面開催がなかったためか，今ではむしろ，オンライン開催の方がデファクト・スタンダードになってしまった感もある。インフォーマルな雰囲気の中で情報交換を行う過程で思いがけない新たな着想，セレンディピティを生むオフライン，対面でのやり取りができなかったのは残念であったが，その一方で，研究資料やデータの共有が効率化されるとともに，オンラインでのプレゼンテーションのノウハウが蓄積され，学術活動の引き出しが増えるという好循環も生じたことは確かであろう。

2021年度「テクストマイニングとデジタルヒューマニティーズ」研究会開催記録
およびメンバーによるDH関連学会での発表記録

第1回　2021年4月16日開催（Zoomを介したオンライン）
発表者・発表題目

全メンバー　2021年度の活動計画打合せ

第2回　2021年5月14日開催（Zoomを介したオンライン）
発表者・発表題目

涌井 萌子　「計量文体学を用いた匿名文書の帰属問題への挑戦
レ枢機卿のマザリナードを例に」

第3回　2021年6月11日開催（Zoomを介したオンライン）
発表者・発表題目

藤田 郁　“Topic modelling of Tennyson”
蒋　睿傑　「中国人日本語独習者がアニメに求める学習需要について
―人気アニメのセリフの特徴分析から―」

第4回　2021年7月2日開催（Zoomを介したオンライン）
発表者・発表題目

黄 晨雯　“What is popular in webnovels’ info?”
徐 勤　「中日大学生が書いた中国語作文における叙述視点
「外国語の勉強」と「留学」をライティングのテーマとして」

第5回　2021年7月7–9日開催（Zoomを介したオンライン）　Poetics and Linguistics Association International Conference (PALA) 2021, University of Notthingham, United Kingdom
発表者・発表題目

Tomoji Tabata　“Different paths to the same peak: Digital humanities and Spitzerian stylistics” (Keynote lecture)

第6回　2021年8月6日開催（Zoomを介したオンライン）
発表者・発表題目

浅野 元子　「ジャンルの理解度に関する考察
―サイエンスニュースを用いた学術英語への導入授業を例に―」
岩根 久　「Ronsardのソネの比較に向けて
Les Amours (1552) と *Les Amours* (1553)」

第7回　2021年9月3日開催（Zoomを介したオンライン）
発表者・発表題目

三野 貴志　「ThereAmalgam研究のこれから」
八野 幸子　「教科等横断的視点を取り入れた英語教育支援ツール作成に向けた語彙研究」
田畑 智司　「ワークショップ：近代日本文学作品コーパスのテクスト分析」

第 8 回　2021 年 10 月 1 日開催（Zoom を介したオンライン）
発表者・発表題目

福本 広光　「米国一般教書演説に出現する分離不定詞の用法について
1960 年代から 2020 年までの分析」
王 鈺　「分断・破壊事象を表す動詞の使い分けとその習得
―「切る」と類義語の関連性について―」

第 9 回　2021 年 10 月 2 日開催（Zoom を介したオンライン）　英語コーパス学会第 47 回大会
発表者・発表題目

黒田 絢香　「トピックモデル可視化ツールの開発に向けて」
藤田 郁　“LDA Topic modelling of Tennyson’s Poetry”（学生優秀発表賞受賞）

第 10 回　2021 年 11 月 5 日開催（Zoom を介したオンライン）
発表者・発表題目

三宅 真紀　「シナイ写本の字形分析<ふたたび>
―異常検知からのアプローチ―」
岡部 未希　「Dickinson の詩における宝石の比喩
―詩中で鮮やかに輝く自然―」

第 11 回　2021 年 12 月 3 日開催（Zoom を介したオンライン）
発表者・発表題目

竹森ありさ　「強意直喩表現 as white (or pale) as a sheet の意味変遷」

第 12 回　2021 年 12 月 16–17 日開催（Zoom を介したオンライン）　Workshop of Education on Digital Humanities: Innovative Application in the Big Data Era (WEDHIA) 2021, National Chengchi University, Taiwan
発表者・発表題目

藤田 郁　““Were not Tennyson’s words delicious?”: An Analysis Using LDA Topic Modelling”
黄 晨雯　“Topic analysis for the synopsis of web novels
―Using the Top2Vec Topic Model―”
黒田 絢香　“Applying topic modeling to full-length novels:
effective settings and visualizations”

第 13 回　2022 年 1 月 7 日開催（Zoom を介したオンライン）
発表者・発表題目

黒田 絢香　「トピックモデル可視化ツールの開発に向けて」
曹 芳慧　「Thomas Hardy, *Tess of the d’Urbervilles* における語彙的特徴」

第 14 回　2022 年 2 月 4 日開催（Zoom を介したオンライン）
発表者・発表題目

Hodošček Bor　「英語電子書籍テキスト処理
‘Standard Ebooks’ と spaCy を例に
高橋 新　“Study on the Application of Stylometric Methods for Analysing English Translations of the Bible: 17 + 3 Gospels of MARK and JOHN”
山田 彬尭　「イ形容詞文における丁寧語使用の歴史的変化
状態空間モデルを用いた時系列分析」

第 15 回　2022 年 3 月 4 日開催（Zoom を介したオンライン）
発表者・発表題目

李 晨婕　（発表予定）
今尾 康裕　（発表予定）

2022 年　2 月
研究代表者　田畑 智司

トピックモデル可視化ツールの開発に向けて

黒田　絢香

大阪大学大学院言語文化研究科

〒560-0043 豊中市待兼山町 1-8

E-mail:　kuroda22a@gmail.com

あらまし　本研究の目的は，トピックモデルの一つである LDA を用いた文学作品の分析をサポートするため，効果的なビジュアライゼーションツールを開発することである．数多くの文書とトピックから成る高次元で複雑なトピック構造から有益な情報を見つけ出すことは非常に労力を要するものであり，また最適なトピック数の選択やモデルの機械的な評価が難しい現状において，試行錯誤のプロセスは必要不可欠と言える．そこで本研究では，モデル出力結果から複雑な設定をすることなく直接グラフを描画することのできるトピック可視化ツールを目指す．複数のモデルを比較しながら探索的な分析を行うことのできるインタラクティブなツールを開発することは，LDA を用いる分析者に大いに寄与すると考えられる．

キーワード　LDA，トピックモデル，ビジュアライゼーション，ツール開発

Development of an Interactive LDA Visualization Tool

Ayaka Kuroda

Graduate School of Language and Culture, University of Osaka

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract　Textual analysis using LDA, a generative probabilistic topic model, is recently attracting attention among the field of literary analysis. However, we have had difficulty in choosing the number of topics when it comes to long literary texts, because each of them usually contains various topics. Thus trial and error is required in the analysis process.

We, therefore, attempt to develop an interactive LDA visualization tool which can compare multiple models we created by changing the number of topics. The aim of this tool is to make the trial-and-error process easier, and to help our understanding of complex but meaningful topical structures.

Keywords　LDA, topic model, visualization, tool development

1.　はじめに

1.1.　文学作品とトピックモデル

近年，機械学習アルゴリズムの一つであるトピックモデルをテクストマイニングに応用する研究が数多く見られる．トピックモデルには author-topic model (Rosen-Zvi et al. 2004) や dynamic topic model (Blei and Lafferty 2006) など様々な種類が存在するが，それらの基礎と

なった LDA (Latent Dirichlet Allocation) (Blei et al., 2003) は，最も有名かつ最も使用されるトピックモデルの一つであると言える．

LDA は文学作品の分析にも用いられており，従来の質的な研究にはなかった新たな観点から作品を見直すことができるとして注目されている．例えば，Jockers and Mimno (2013) では，19 世紀のフィクションを集めたコーパスに対してトピックモデルを適用することで，男性著者，あるいは女性著者が特徴的に使用する話題を発見した．また田畑 (2017) は，小説だけでなく様々なレジスターの文書が含まれた大規模コーパスである FLOB コーパスに対しても，そのコーパスの背後にある潜在的な意味構造を読み解く上でトピックモデルが有用であることを示した．

文学作品のように人文学の領分とされてきた研究対象に対して情報工学の手法を持ち込む，いわゆるデジタル・ヒューマニティーズと呼ばれるアプローチは，近年のテクノロジーの進化に伴いより活発となっており，質的研究が中心であった分野に新たな知見を数多くもたらしている．

1.2. 最適な設定を選ぶには

LDA を実行する際に，分析者は抽出するトピックの数を任意に設定することができる．分析対象の文書が少ない場合には 10 前後とすることが多いが，規模の大きなデータを扱う場合では 100 から 500 ものトピックを設定することもある．一般的には，この値が小さすぎれば抽出されるトピックは汎用的な語ばかりが含まれる解釈の難しいものとなり，一方で大きすぎれば過剰に細分化されるあまり，比較的出現頻度が低い各文書に特有の語が多く出現してしまうと言われている．そのため，実験対象コーパスの規模や分析目的に応じて適切な数値を都度判断する必要がある．

最適なトピック数の決定方法については数多くの議論がなされているが，どのような場合でも使用できる普遍的な最適解は未だ存在しない．現状ではトピックモデルの評価指標である Perplexity と Coherence の値を参考にして決定されることが多いが，この手法には問題点も存在する．

まず，Chang et al. (2009) が指摘した通り，機械学習モデルの精度を確率論に基づいて示す値である Perplexity は予測モデルの評価において有用であるものの，探索的な分析を行なうトピックモデルには適合しないことが多い．

一方 Coherence ではトピックの質，つまり生成されるトピックに含まれる単語の一貫性，解釈可能性を評価する．「人間にとってわかりやすいかどうか」という定量化しにくい評価を機械的に測定する必要があるため，その算出方法は物議を醸している．例えば Newman et al. (2010) のように WordNet や Google などの外部コーパスをベースとして単語間の類似度を計算する手法もあれば，Mimno et al. (2011) のように各トピックの頻出単語同士の共起頻度をもとに算出し，外部データを利用しないものもある．Lau et al. (2014) ではこれらを含めた複数の Coherence 算出方法を比較評価しているが，いずれも，特にトピックベースでの評価の場合，人間による評価とは必ずしも一致しないと結論づけられている．

また，Sbalchiero and Eder (2020) もトピック数の決定方法について検討を行っているが，学術論文のアブストラクトやニュース記事など単一の話題について書かれた文書を用いた実験が数多く行われている一方で，長編小説のように多様な話題で構成されている比較的長い文書については十分に検討されていないと指摘している．

以上のことから，トピック数を選択する際には，複数のモデルを構築した上で比較評価を行なう試行錯誤プロセスが必要不可欠であると言える．Perplexity や Coherence の値は参考になるものの，決定打とすることは難しい．実際に生成されたモデルを確認し，どのようなトピックが生成されているか，どれほどのトピックの粒度が分析に最適かを判断し，ときには複数のモデルによる結果を並行して表示しながら分析を進めていくことが重要であると考えられる．

1.3. LDA ビジュアライゼーションツール

前述のような試行錯誤プロセスにおいて，モデルビジュアライゼーションの効率化は重要な課題である．何百，何千もの文書と数多のトピックによって構成される高次元で複雑なトピック構造から分析に有用な情報を見つけ出すには的確な可視化が不可欠であり，バブルチャートやヒートマップ，ワードクラウド，ネットワークグラフなど様々なグラフが使用されてきた．しかし，そのグラフ描画プロセスが複雑で時間のかかるものであったり，大量，もしくは複雑なグラフによって表現されるものであるならば，特に複数のモデルを作成して比較する場合，分析には多大な労力を要する．

これまで幾多のトピックモデルのビジュアライゼーションシステムが開発されてきたが，中でも Sievert and Shirley (2014) によって提案された LDAvis[1]，Python 版である pyLDAvis[2]は代表的な LDA 可視化ツールであると言える．その開発の主眼は"both compact and thorough is interactivity" (Sievert and Shirley, 2014) とされており，インタラクティブなグラフを描画することによって，画面遷移を挟むことなく大量のトピック情報をコンパクトにまとめ上げることに成功している．

しかし一方で，多くとも 50 前後のトピック数を想定しているため抽出トピック数が大きすぎる場合にグラフが煩雑になりすぎる点や，各トピックがどの文書にどれほどの確率で出現しているかを示すことができない点など，限界も存在する．

1.4. 研究の目的

本研究の目的は，文学作品の分析を視野に入れた LDA のビジュアライゼーションツールを開発することである．前節までの内容を踏まえ，このツールの目標を以下の 4 つに設定する．

- 静的なグラフではなく，インタラクティブな出力であること
- トピック数を変化させて実験した複数のモデルの出力結果を比較できること
- 抽出するトピック数が 100 を超える場合でも対応できること

[1] https://github.com/cpsievert/LDAvis
[2] https://github.com/bmabey/pyLDAvis

- 各トピックの出現単語とその比率だけでなく，各トピックがどの文書に出現しているかを示すこと

このようなツールを開発することで，多大な労力と試行錯誤を要するトピックモデル分析をより効率的なものにし，文学作品の量的研究に寄与することができると考えている．

2. 手法

2.1. データと前処理

本稿では，サンプルコーパスとして Arthur Conan Doyle の長編小説コーパスを用いる．収録されているのは 29 作品，推理小説や歴史小説，ノンフィクションなど 5 種類のジャンルが含まれており，総語数は約 2,180,000 語である．

このコーパスに対し，Jockers and Mimno (2013) で用いられていた手法を参考に，以下の前処理を行った．

- チャプタータイトルや注など本文以外の要素を削除
- TreeTagger[3]によるタグ付けに基づき普通名詞のみを取り出す
- 1,000 語ごとのブロックに分割する

以上を行った結果生成された 409 ファイルが本稿での分析対象である．

2.2. MALLET の出力構成

モデリングには，Java ベースの自然言語処理ツールキットである MALLET[4]を用いた．このツールではモデルを様々な形式のファイルで出力できるが，今回必要となるのは主に (1) output-doc-topics コマンドによって生成される，各ドキュメントにおけるトピックの出現確率を示すデータ， (2) topic-word-weights-file コマンドによって生成される，各トピックにおける頻出単語リストとそれらの重みを示すデータ， (3) diagnostics-file コマンドによって生成される，複数の指標に基づいて各トピックの品質を示すデータの 3 種類である．

特に (3) の diagnostics データは，トピックに割り当てられた単語の数を示す tokens，トピックを指定された際の文書の予測確率からそのトピックの一般性を計る document_entropy，トピックの単語分布とコーパス全体の単語分布との距離を示す corpus_dist，Mimno et al. (2011) の手法に基づく coherence など，12 種類の指標を用いて各トピックの特徴やモデル全体における位置づけを表す有用なファイルである．

[3] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
[4] http://mallet.cs.umass.edu/

2.3. ツール開発

開発では，Python のグラフ描画ライブラリの一つである Plotly[5]，及びそれを組み込むことのできる Web アプリケーションフレームワークの Dash[6]を主に用いた．これらの特徴はインタラクティブなグラフを HTML 形式で出力できることにある．グラフの拡大や縮小，ドロップダウンメニューやスライダーなどによるグラフコントロール，クリックやホバーによるイベントの発生など，様々なカスタマイズが可能である．

3. 開発中ツールの概観と特徴

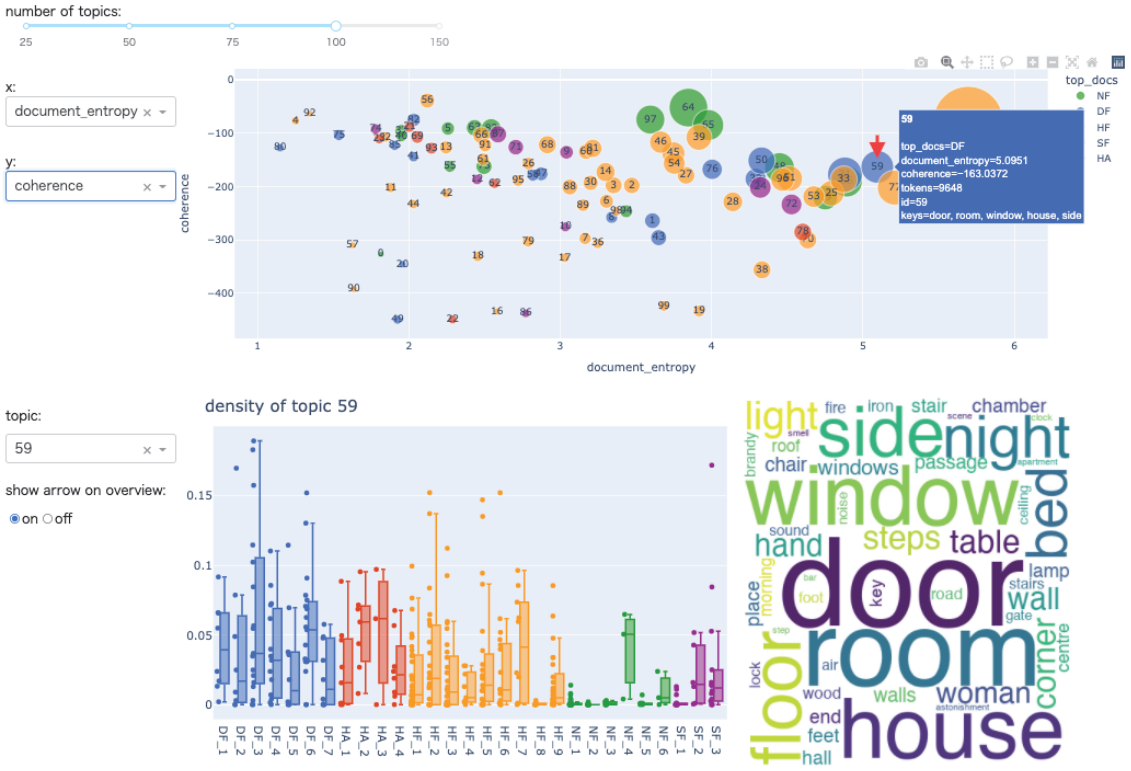


図 1: 開発中ツールの概観

開発中ツールの概観を図 1 に示す．大きく分けて 2 つの表示部があり，上部は前述の Diagnostics を参照してモデルの全体像を表す model view，下部は選択されたトピックの詳細をボックスプロット・散布図とワードクラウドで示す topic view である．model view でのマウスホバーによって各トピックのデータがツールチップの形で表示され，また下部の topic view に即座に反映されるため，どのようなトピックが存在するか，モデルの全体像を見ながら探索的に分析することが可能となっている．

5 https://plotly.com/python/
6 https://dash.plotly.com/

3.1. model view

前述の通り，上段の model view では Diagnostics に表示される 12 種類の指標に基づいてバブルチャートがプロットされる．バブルサイズは tokens に対応しており，大きなものほどトピックに含まれる単語数も多くなる．

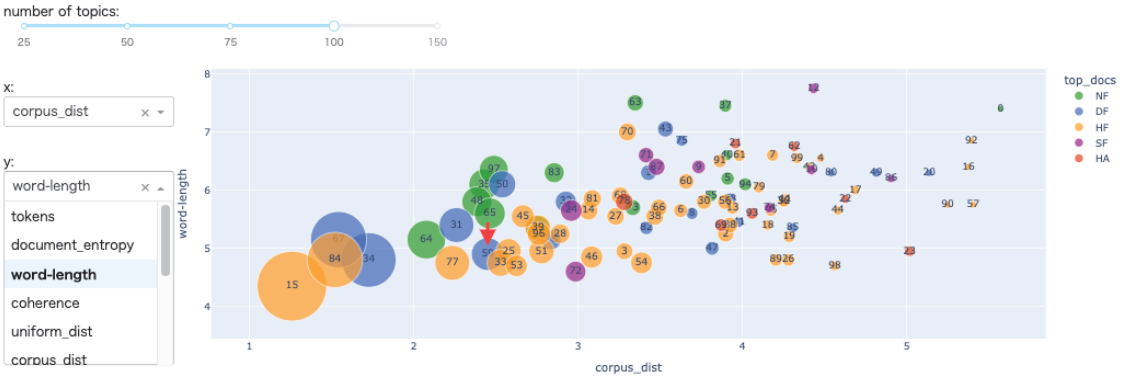


図 2: model view の x 軸と y 軸の変更

x 軸と y 軸に設定する指標はツール使用中に切り替えることが可能である．例えば図 1 では x 軸を document_entropy，y 軸を coherence に設定しているため，数多くの文書に出現する一般的なトピックは右に，比較的マイナなトピックは左に配置され，上部のものほど一貫性 (ここでの一貫性は単語の意味に基づく評価ではなく，構成要素の単語同士が同じ文書に出現する確率に基づく評価である) が高い．図 2 のように指定すれば，コーパス全体の単語分布に近い，つまり一般的なトピックほど左に配置されるようになる．y 軸で word-length を指定しているため，上部には平均単後長が長い高尚でアカデミックなトピックが配置されやすい．

ここに挙げた以外にも様々な指標が組み込まれているため，x 軸と y 軸の設定を切り替えることで，自身の分析目的に適うトピックを探し出すことが容易となる．

また，Plotly の特徴として，グラフの一部分をマウスドラッグで拡大することができる．そのため tokens の値が低い故に小さなバブルで表示されているトピックを選択する際，密集地帯であっても問題なく目的のトピックを探し出せる (図 3 を参照)．この機能は，トピック数の設定が 100 以上の大きな数値であった場合，特に真価を発揮すると考えられる．

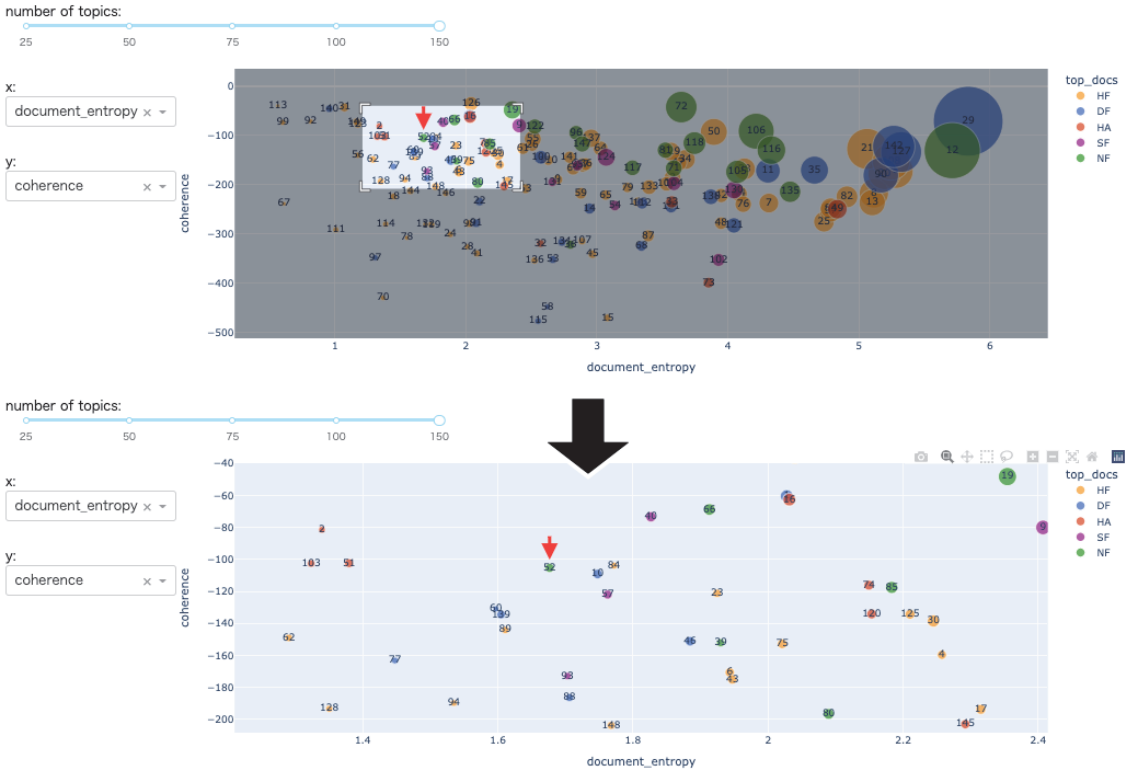


図 3: 拡大機能プレビュー

3.2. topic view

topic view では，当該トピックがどの文書にどの程度の確率で出現するかを示す散布図とボックスプロット，そして右側にトピックの構成単語とその重みが反映されたワードクラウドが描画される．この 2 つの図を読み解くことで，トピックや文書の特徴を解釈することができる．

例えば，図 1 に表示されている topic 59 は，ほとんどの文書において出現する一般的なトピックであることがわかるが，NF から始まるファイル，つまりノンフィクションでは NF_4 を除いてほとんど出現していない．そこでワードクラウドに注目すると，この topic 59 は，door, room, house, window, bed, floor など室内のものに関する語を中心に構成されていることがわかる．

NF_4，*The Case of Oscar Slater* は実際に発生した殺人事件について検証を重ねる形式の小説である．一方で，それ以外のノンフィクション小説は戦争を題材としたものがほとんどであった．この topic 59 の出現確率が高い作品は，推理小説を始め室内を中心に物語が展開するものが多く，逆に出現確率の低い作品は屋外で物語が展開する，あるいはより大局的な視点から出来事が描かれている作品が多いのではないかと考えられる．

3.3. トピック数の変更

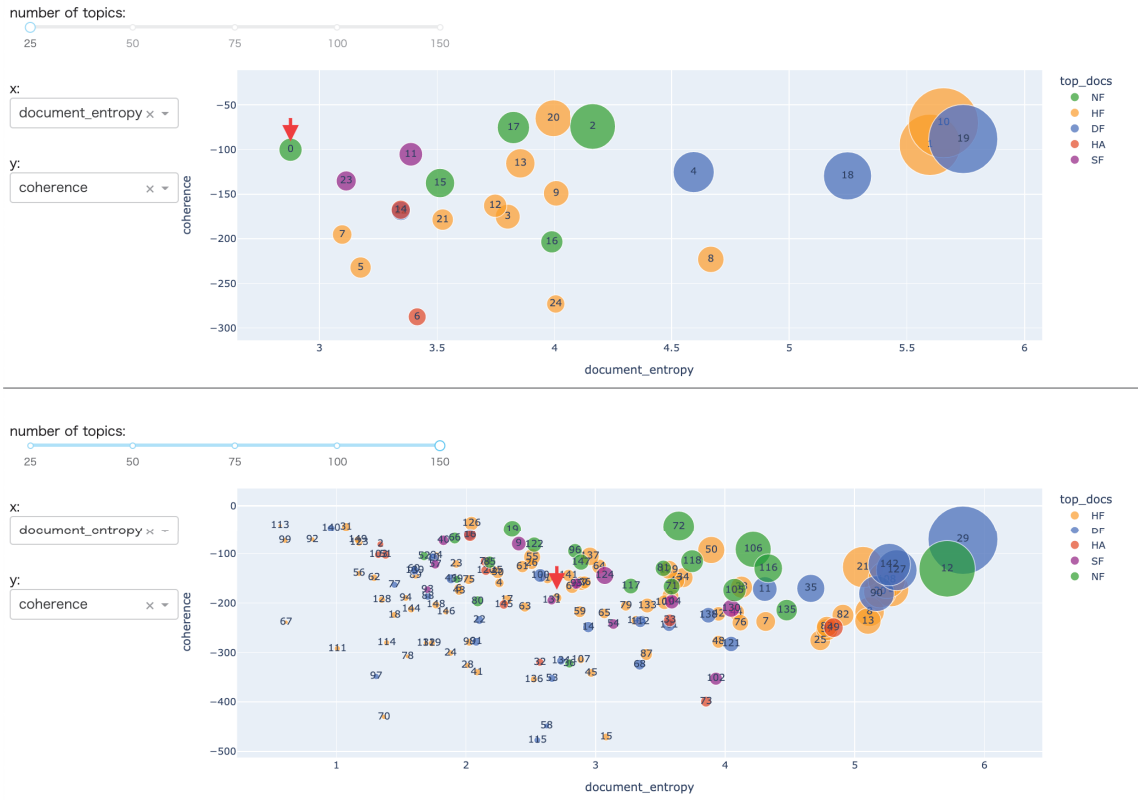


図 4: トピック数の変更イメージ

最上段のスライダーを操作することで，図 4 のように全トピック数を切り替えることができる．

これは，スライダーの入力に応じて設定を変更したモデルを生成するのではなく，事前にトピック数値を変更しながら複数のモデルを用意しておき，スライダーにその数値を登録することでアウトプットを切り替える形式となっている．モデルの生成は負荷が大きいため，都度生成する方式では細かな数値変更に対応できる一方，切り替えに時間がかかりすぎてツール全体の処理が遅くなることが考えられる．事前に様々なパターンのアウトプットを用意しておくことで，スムーズにトピック数を切り替えながら比較し，どのモデルが自身の分析目的に適っているか考察することができる．

図 4 では 25, 50, 75, 100, 150 の 5 種類のモデルを登録しているが，25 以下，150 以上の数値でも問題なく設定できるほか，より多くのパターンを一度に登録することも可能である．

4. 結び

本稿では，LDA に基づくトピックモデルを効果的に可視化するツールの必要性について論じ，開発中のツールの概観と主要な機能を示した．静的なグラフではなくインタラクティブなグラフを出力することで，トピック数が多い場合でも快適に探索的分析を行うことが可

能となる．またスライダーを使ってトピック数を変化させることができる機能は，トピック数の決定に試行錯誤が必要不可欠である現状において，分析者に大いに寄与するのではないかと考えられる．

今後の課題として，インプットされたコーパスがサンプルコーパスよりも極めて多い文書数であった場合など，どのような種類のデータであっても topic view の表示が崩れないよう UI を改善していくことが挙げられる．

また，トピックリストから表示したいトピックのみを選択したり，特定の文書群に出現するトピックのみを選択したり，tokens など特定指標の値が一定の範囲を超えるものを除外するフィルタ機能を設けたい．分析目的に沿わないトピックを model view から取り除くオプションを追加することで，よりシンプルなプロットを描画できるようになると考えられる．

MALLET の diagnostics ファイルには，トピック間の類似性を評価する指標は組み込まれていない。一方で pyLDAvis では，多次元尺度法（MDS）に基づいてトピックを配置することで，トピック同士の類似性を距離に変換し可視化している。今後は類似性を評価する指標を組み込むことで，model view 部分のグラフ表示を切り替えたり，選択したトピックに類似したトピックをレコメンドする機能を追加することも検討している．

文　献

[1] Blei, M., Ng, A. and Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3: 993–1022.

[2] Blei, M, Lafferty, D. (2006). Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*, 113–120.

[3] Chang, J. Boyd-Graber, S. Gerrish, C. Wang, and D. Blei. (2009). Reading tea leaves: How humans interpret topic models. *Neural Information Processing Systems*.

[4] Jockers, M. and Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics* 41: 750–769.

[5] Kuroda, A. (2018). Topic Representation across Texts and Genres: Finding Key-words through Topic Models. (Unpublished master's thesis). Osaka University.

[6] Lau, J., Newman, D. and Baldwin, T. (2014). Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 530–539.

[7] Mimno, D., Wallach, H., Talley, E., Leenders, M. and McCallum, A. (2011). Optimizing semantic coherence in topic models. *EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272.

[8] Newman, D., Lau, J., Grieser, K. and Baldwin, T. (2010). Automatic Evaluation of Topic Coherence. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 100–108.

[9] Rosen-Zvi, M., Griffiths, T., Steyvers, M. and Smyth, P. (2004). The author-topic model for authors and documents. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 487–494.

[10] Sbalchiero, S. and Eder, M. (2020). Topic modeling, long texts and the best number of topics. Some Problems and solutions. *Quality & Quantity*, 54, 1095–1108.

[11] Sievert, C. and Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70.

[12] 田畑 智司 (2017)「FLOB コーパスの意味構造: 確率論的トピックモデルによる言語使用域の特徴付け」『統計数理研究所共同研究リポート』386: 1–17.

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。

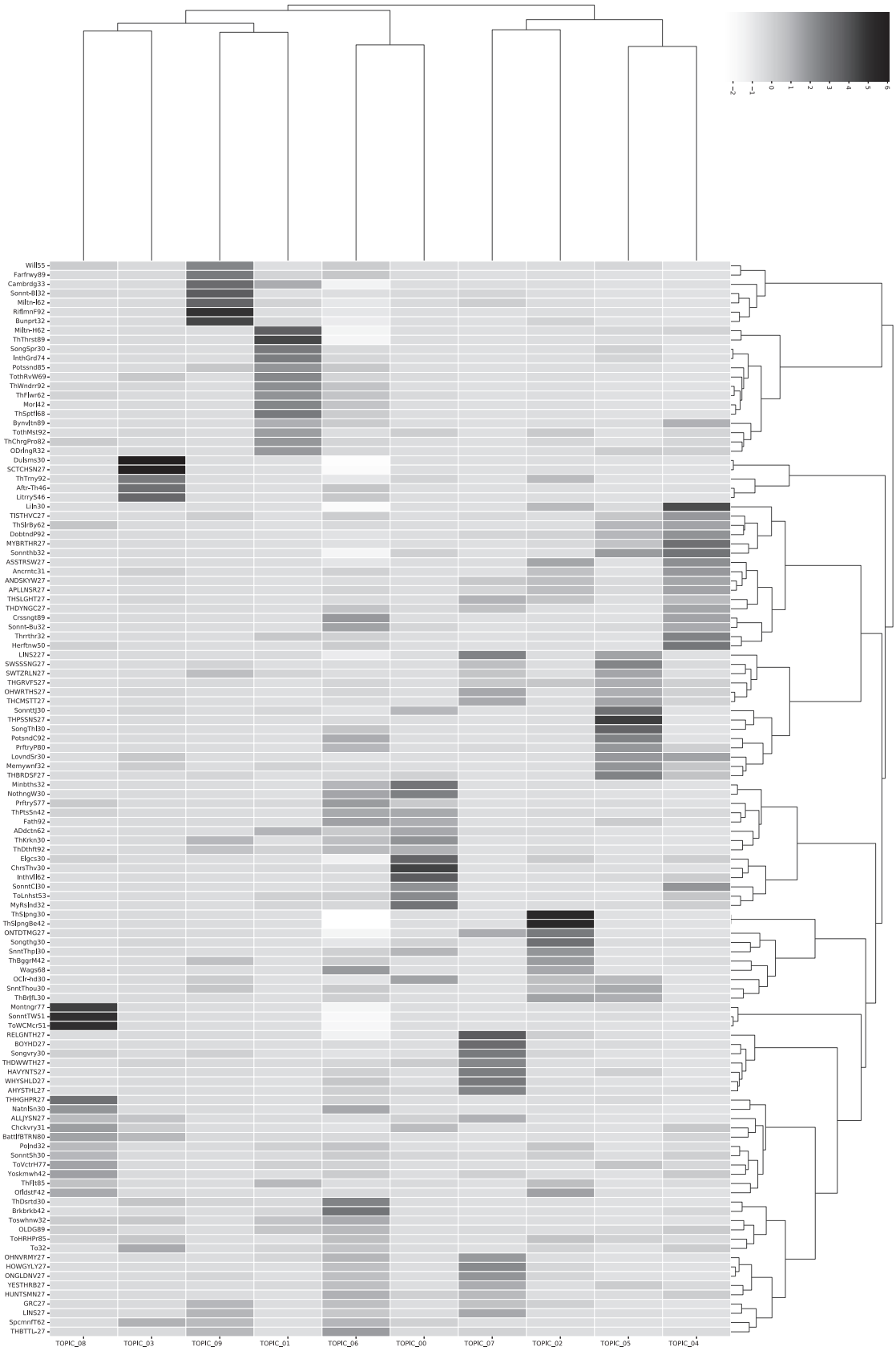


図 1 ヒートマップ（raw テキスト）

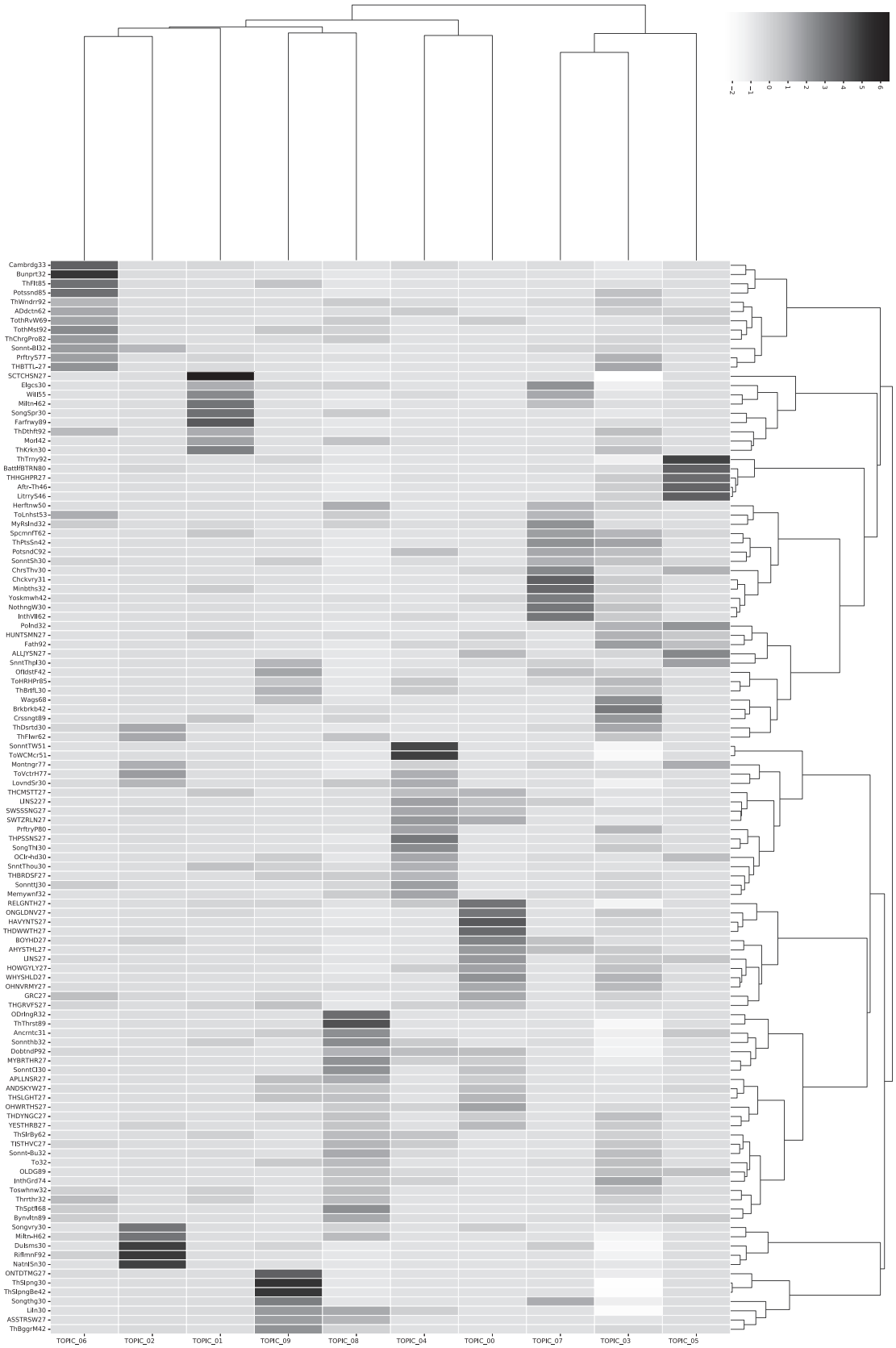


図 2 ヒートマップ（IPA テキスト）

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）

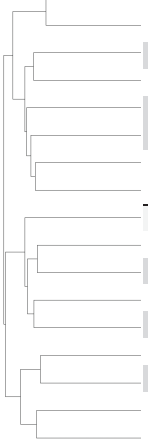
Topic 0,1,6 は語頭母音 (V-Head) が 0 であるものの，それ以外の音は全て現れているため，CVC(V) のように語頭が子音で始まり，後続する母音あるいは子音が繰り返される語を構成する音を多く含んでいるトピックと言える。また，17 トピックのうち 15 トピックにおいて，語中の母音は 3 つ以上出現しているが，Topic 11 と Topic 7 にはそれぞれ φeəφ と φaɪəφ, φɪəφ いずれ

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)

学術英語遠隔授業におけるジャンルの認識に関する調査
―日本の大学 1 年生を例に―*

浅野元子

大阪医科薬科大学医学部医学科語学教室
〒569-8686 高槻市大学町 2–7
Email: motoko.asano@ompu.ac.jp

概要　本研究は，コロナ禍における日本の大学 1 年生を対象に，前期 13 回の学術英語の遠隔オンデマンド授業で，ジャンルに基づくタスクを行い，ジャンルの認識について調査した。そしてジャンルの認識と学術英語全般の運用能力との関連性について考察した。20 名 (48.8%) の学生の期末試験での要約文において，授業課題の要約文と比較して，目標とするジャンルのテクスト構成の維持あるいは構成の改善が示されたが，約半数の学生は，目標とするジャンルではなく，もとの課題テクストのジャンルに特有の構成要素を再現していた。要約文の構成に基づくスコアと学術英語運用能力の指標として参照した *TOEFL ITP*®のスコアとの関連は示唆されなかった。遠隔での演習でもジャンルの認識が促進されることが示唆されたが，ジャンルの認識を深めるには，新しいアプローチを受け入れ，テクストの特徴や読み手などの状況を俯瞰的に捉えることなどを必要とするのではないかと考察された。本研究の結果は，授業にジャンルの枠組みを効果的に取り入れるために意義があると考えられた。

キーワード　ジャンル，コーパス，ムーブ，遠隔授業

Developing genre awareness in an academic English course:
A case of Japanese first-year undergraduate students

Motoko Asano

Osaka Medical and Pharmaceutical University
2-7 Daigaku-machi, Takatsuki, Osaka, 569-8686 Japan

Abstract　This study investigated whether an on-demand course of genre-based academic English for first-year undergraduates would help learners raise their genre awareness and whether the degree of genre awareness would be related to the students' English proficiency as expressed by the *TOEFL ITP*® scores. In the thirteen weekly classes, the participants performed various tasks such as extracting research information from science news (SN), analysing moves of SN texts that communicate research article information, and summarizing SN texts in a research article abstract style. Analyses of summaries from 41 students revealed that 20 participants (48.8%) improved or maintained their organizational performance in summarizing research information in a research article abstract style. The performance scores seemed to have

* 本稿は，大学英語教育学会 (JACET) 第 60 回国際大会の口頭発表 (「サイエンスニュースを用いた学術英語への導入授業―ジャンルの理解度に関する考察」浅野元子・藤枝美穂) のデータの一部を用いて新たに検討を行った。

no clear relationship with the participants' English proficiency based on their *TOEFL ITP®* scores. The on-demand genre-based academic English course successfully raised some students' genre awareness, focusing mainly on the rhetorical organization of genre texts. However, the degree of genre awareness may be related to whether or not learners accept new approaches to pay attention to genre-specific features of texts. The findings may be useful for effectively incorporating genre frameworks into classrooms.

Keywords genre, moves, on-demand classes, corpus

1. はじめに

学術英語教育の最終目標は，種々の専門領域特有の言語の使い方とコミュニケーションのスタイルを習得することとされる (飯島他, 2016) 。学術英語は「特定目的の英語」(田地野, 2009, p.130) (English for Specific Purposes, 以下ESP) の主要な位置を占める (Anthony, 2018; 寺内, 2010) 。ESPは，Swales (1981, 1990) によるジャンル分析を中心に発展してきた。

ジャンルの理論研究は，主にESP，北米を中心とするNew Rhetoric 学派，および Systemic Functional Linguistics に分類されてきた (Hyon, 1996) 。最近では，これらの学派は互いに独立したものではなく，融合して教育現場に応用されることが多い (Johns, 2008 & 2013) 。

近年，ジャンルとは「話し言葉や書き言葉の種類」(Hyon, 2018, p. 3; 以下，断りがなければ日本語訳は筆者による) であることに加えて「抽象的で，社会的に認知されたことばの使い方を指す」とされ (Hyland, 2007, p. 149)，「社会活動として理解されるもの」(Miller, 1984, p. 156) という概念が引用され, 支持されている (Swales, 1990, p. 18; Gentil, 2011, p. 7; Tardy et al., 2020, p. 288) 。

ESP のジャンルの考え方については，「テクストの構造を説明すると読む力がつくという有益な証拠が得られている」 (Swales, 1990, p.89) と説かれ，「第二言語ライティング教育に影響を及ぼした」 (Hyland, 2003, p18) とされる。しかし，ジャンルに基づく言語特徴に関する研究と比べて学習者に焦点を当てた研究は少ないとされてきた (Hyland, 2010) 。

ジャンル分析は，研究論文の緒言部に特徴的な「修辞の運び」"rhetorical movement" すなわち「ムーブ」 "moves" (Swales, 1990, p. 140) で構成されることが明らかにされた後，学術論文などの特定の目的のためのテクストを効果的に読み書きするために有効な方法として種々のテクストに応用されている (Biber et al., 2007; Dudley-Evans & St John, 1998; Flowerdew, 2012) 。ジャンル分析については，Swales (1981, 1990) による「ムーブの構成」 "move-structure" (Bhatia, 1993, p. 31) に関する解釈が進み「ジャンル分析におけるムーブ」"a "move" in genre analysis" (Swales, 2004, p. 228; 引用符の使用は原著による) を「コミュニケーション上の機能を担う談話または修辞の単位」(Swales, 2004, p. 228) とする考え方が受容され，「ムーブ分析」"move analysis" (Kanoksilapatham, 2005, p. 269; Moreno & Swales, 2018, p. 40) とも称される。

ジャンル教育研究の理論的枠組みを整理した Tardy らの研究 (2020) は，「ジャンルの知識」"genre knowledge" (Berkenkotter & Huckin, 1995, p. 13; Swales, 2004, p. 62; Tardy, 2009, p. 20) を「ジャンル特有の知識」"genre-specific knowledge" (p. 294) と換言して，以下の要素が「重なり合って層をなす動的な知識による多次元のもの」と定義している。

1. Formal [knowledge] (e.g., content, organization, lexicogrammatical features)

2. Process [knowledge] (e.g., composing, distributing)

3. Rhetorical [knowledge] (e.g., discourse community, social relations)

4. Subject-matter knowledge (e.g., disciplinary conventions) (Tardy et al., 2020, pp. 294–295)

Tardy らの研究 (2020) によると「ジャンルの認識」は「ジャンルの知識」と区別され，「ジャンルの役割を明示的に認識し，理解すること―ジャンルを分析し，学び，批評するための意識と過程」(p. 296) と定義される。

「ジャンルの考え方に基づく教育プログラム」“genre-based teaching programme” (Marshall, 1991, p. 9) は，ことばが社会的文脈の中で果たす役割について明示的かつ体系的な理解を促す (Hyland, 2007) 。その中でも「ジャンルの認識に関する教授法」 “genre awareness pedagogy” は，Tardy らの研究 (2020) によると「まず，よく知っているジャンルテクストの修辞の分析に焦点を当て，次にあまり知らないが関連のあるジャンルテクストに焦点を当てる中で，常に形式と文脈を結び付ける」とされ (p. 290)，主に「初期の学部生の教育」に向いている (p. 291) 。そして「ジャンルに基づくタスク」(Hyon, 2018, p. 130) は「修辞への意識向上のための活動」 “rhetorical consciousness-raising activities” を伴い，ジャンルの知識を「強化する」(Tardy et al., p. 291) 。

学部生の教育の例としては，Johns (1997) による北米の大学で英語を母語としない学生を対象とした授業実践がある。この授業 (Johns, 1997) では，生物学のレポートの「緒言―方法―結果―考察 (IMRD) 構造」 (Swales, 1990, p. 134) を理解させるために，生物学の教員と共同で，各自が書いたレポートを相互に評価させている。教員と 5 パラグラフ・エッセイの構成に慣れた学生との間でレポートの評価にずれが生じ，それを授業で伝えることが，テクスト構造に対する柔軟な考えとジャンルの認識を促すことにつながったとされる (Johns, 1997)。

日本の大学生を対象とした Yasuda (2011) の研究では，ジャンルに基づくライティングコースで，E メールを書くタスクを行う際に，“real-world context” (p. 117) を用いて，メールを書く目的，受信者，言語使用に注目し，明示的に指導を行なった結果，ジャンルについて学習した文脈から別の文脈への転移が可能であることが示唆されている。

異なる文脈を経験することについて，社会活動としてのジャンルを論じた Miller の研究 (1984, pp. 156–157) では「新規の文脈は，関連した類似性を再認することを通じて馴染みのあるものになる」“the new is made familiar through the recognition of relevant similarities” と記され，修辞的な文脈は同じものの繰り返しではなく，類似したものが幾度となく生起するため，人々がその類似性を認識して類似した反応を示すことを示唆している。この過程を Tardy らの研究 (2020, p. 301) では「再文脈化」“recontextualization” と称しており，その際に学習者は「修辞的文脈を鋭く認識して用いる能力を必要とする」(p. 300) と説かれる。

ジャンルの認識を調査した研究 (Negretti & Kuteeva, 2011) では，スウェーデンの大学の英語教員養成課程の学生に対して人文系研究論文のジャンル分析などを行うセミナーにおいて，参加者が「学術テクストの目的，聴衆，修辞的なムーブ，および構造などのジャンルに関連する要素の知識」(p. 104) をテクストの分析とエッセイのライティングに用いたとされる。この研究では，参加者全員が演習を通じて，テクスト構造などのジャンルに関することを理解するだけでなくラ

イティングに用いることができるようになった一方で，テクストに特徴があるのはそのテクストが状況に応じた目的を有するためであるという理解に至ったのは少数であったと報告される (Negretti & Kuteeva, 2011) 。ジャンルの認識に個人差がみられた理由については「学生間の違いを解釈するには注意が必要である」(p. 106) とされ，あまり明らかにされていない。

米国の大学で非母語話者を対象とする学術英語 (フォーマルなニュース記事，教科書，研究論文など) のジャンルに基づくリーディングコースの 1 年後に追跡調査を行った研究においても，ジャンルに関する明示的な指導による効果が持続したことが示される一方で，学生によって効果に差が認められたとされる (Hyon, 2001) 。

本研究では，先行研究を踏まえて，英語の遠隔オンデマンド授業において，研究成果を発信するサイエンスニュース　(以下 SN と記す) のテクストを中心に「ジャンルに基づくタスク」(Hyon, 2018, p. 130) を行った。異なるジャンルすなわち SN および研究レポートの目的，情報の受け手，テクストの構成などについて授業ビデオやタスク (課題) へのフィードバックなどを通じて明示的に伝えて，テクスト構成に焦点を当てることで，学生が SN ジャンルのテクストで伝えられる研究について，研究レポートのアブストラクトのテクスト構成を意識して要約することができるようになるかどうかを検討した。

TOEFL® (Test of English as a Foreign Language) は英語非母語話者の英語運用能力を測定するテストとして開発され，主に北米の大学などの教育機関で学ぶ際に必要とされる英語力を測る (Educational Testing Service, 2020) 。*TOEFL*®およびその教材の語彙を学術英語，標準英語，*TOEIC*®の語彙と比較した水本 (2006) の研究では，*TOEFL*®の使用語彙は特別に難しいわけではなく，学術英語の語彙に近いものであり，*TOEFL*®の特徴語が存在することを明らかにしている。*TOEFL ITP*® は，“Reading, Listening, and Structure and Written Expression” の 3 つのセクションで構成される団体向けの多肢選択式のテストである (ETS, 2022) 。期末試験での要約文 (以下 Writing F) のスコアを「ジャンル特有の知識」(Tardy et al., 2020, p. 294) の一部を用いる力を測るものと捉える場合，学術英語全般の運用能力を測る *TOEFL ITP*®との「相関関係を調査すること」に無理があることは自明であるが (Bachman, 1990, p. 248)，本研究では，ジャンルの認識と学術英語運用能力との関連性を考察するために，*TOEFL ITP*®のスコアを参照した。

2.　方法

2. 1. 目的とリサーチクエスチョン

本研究の目的は，学生が遠隔授業での「ジャンルに基づくタスク」(Hyon, 2018, p. 130) を通じて，ジャンルの認識が促されるかどうかを調査することである。もうひとつの目的は，ジャンルの認識と学術英語運用能力の指標として参照した *TOEFL ITP*®のスコアとの間に関連性がみられるかどうかを調査することである。この目的のために，以下のリサーチクエスチョンを設定した。

RQ1:　期末試験の要約文（Writing F）の構成には，授業課題の要約文（以下 Writing 1）の構成と比べて変化がみられるであろうか。(Writing F と Writing 1 の比較)

RQ2:　Writing F の構成に基づくスコアと *TOEFL ITP*® のスコアに相関がみられるであろうか。

(Writing F と *TOEFL ITP*®スコアの比較)

Writing 1 および Writing F のテクストを用いてコーパスを編纂し，構成要素を観察してコーディングを行った。構成要素の叙述量を計量し，研究レポートのアブストラクトに特徴的な構成要素の記述量の差を対数尤度比検定 (Dunning, 1993) を用いて調査した。コーパス間の頻度の比較には正規分布を前提としない対数尤度比検定が使用されることが多い (McEnery & Hardie, 2012) 。コーディングの結果をスコア化し，スコアの推移を調査した。

TOEFL ITP®のスコアとの関連性について, 要約テクストのスコアを点数とみなしてピアソンの相関係数を算出した。ピアソンの相関係数は 2 つの変数の直線関係の程度を記述する係数で，間隔尺度の場合に用いられる (竹内・水本, 2014) 。

2. 2. 対象者

本研究は, 2020 年度に実施された医学部医学科第 1 学年の学術英語授業への参加者を対象とした。必須科目 (国際言語文化 1 (英語)) の一部として，例年は 90 分の対面授業を 15 回実施するところ，当該年度は，コロナ禍で授業ビデオ，資料，課題などを学内の学習マネジメントシステム (LMS) により配信する 13 回の遠隔オンデマンド授業となり, 期末試験のみが対面で実施された。学生 113 名を 2 クラスに分けて 2 名の教員が 1 クラスずつ担当した。本研究では，筆者が担当した 56名のうちの，すべてのデータがそろった 41名 (男性 26名，女性 15名) を対象とした。41名の 2020 年 4 月 1 日時点の平均年齢は 19.5 歳 (標準偏差は 1.5)，中央値は 20 歳で，年齢が平均以上の学生は 16名，平均未満の学生は 25名であった。2021年 1月に実施された *TOEFL ITP*®スコアの平均は 475 (標準偏差は 42) であった。

2. 3. 授業概要

授業の概略を表 1 に示した。本授業では，学術英語の導入と第 2 学年以降の専門英語への橋渡しをねらいとして，研究成果を発信するSNを素材として用いている。2010年代半ばの*Scientific American 60-Second Science* は大半が 1 分程度のポッドキャストで，書き起こし (トランスクリプト) と引用元の研究論文などへのリンクが公開されている。2014 年に発行され研究論文を紹介した *Scientific American 60-Second Health* のポッドキャストの書き起こしテクストを調査した研究において，導入 (Lead)，本論 (News)，結論 (Conclusions)という情報提示の構成が観察され, 研究の背景, 方法，結果，結論，および話の落ち (punch line), 研究の成果物と著者の紹介 (References) などの要素が抽出されている (Asano, 2022) 。

1 回目の授業で ESP とジャンルの考え方を 4 本の授業ビデオで紹介した。その中で，学生が目指す英語使用者の例として，コロナ禍で逼迫する東京の病院の様子を伝える日本の医療従事者からの英語による発話 (BBC, 2020) の音声と教員による書き起こしを提示した。ジャンルとディスコースコミュニティの概念 (Swales, 1990) について，大学の授業における教員と学生などの身近な例を挙げて伝えた。ムーブについては, Gibbons (2007, pp. 280–281) による「ジャックと豆の木」を用いたムーブ分析の提案に従って，日本語で書かれた「桃太郎」(Hukumusume.com, 2022) を用

いた資料を用意した。第二言語の学習における第一言語での説明と直接教授法を比較した Cook (2010) は「第一言語による説明のほうが，効率がいい」(p. 23) と説いている。

表 1　授業の内容とタスク (課題)

授業 (回)	主な内容とタスク (課題)
1	ESPとジャンルの考え方の紹介，1分自己紹介の草案と録音
2	SN1 聴解，研究の要約表，抑揚などの書き取り
3	SN2 聴解，研究の要約表，ムーブ分析
4〜5	SN3〜4 前回の省察質問，聴解，研究の要約表作成，ムーブ分析
6	ESPとジャンルの考え方・SN1〜4の復習・要約課題提示 前回の省察質問，SN2の要約 (Writing 1)，録音
7	SN5 聴解，研究の要約表
8〜10	SN6〜8 聴解，研究の要約表，ムーブ分析
11	ESPとジャンルの考え方・SN5〜8の復習・要約課題提示 前回の省察質問，SN7の要約，録音
12	SN1〜8の復習 聴解，研究の要約表，ムーブ分析
13	課題 (要約文) の復習 省察質問

SN のサンプルテクストを用いた説明に際して，授業で扱うテクストが「より広い社会修辞学的文脈に埋め込まれたもの」(Swales, 1990, p. 44) であるという理解を促すことを目的として “OCHA” (Noguchi, 2003, p. 3) 思考法と“PAIL”分析 (Noguchi, 1997, p. 311) を紹介して使用した。これらの理論は「ジャンル別にスキーマ認識能力を用い，情報に選択的注意を払うことができ，それを言語の側面から分析できる能力」を支援する (野口・深山, 2010, p. 30) 。スキーマとは「個人が現在の読み書きの状況に持ち込む事前の知識」(Johns, 1997, p. 11) であり，テクストのテーマすなわち「内容スキーマ」“content schema” と慣例的な修辞などの「形式スキーマ」“formal schema” に大別され，双方に馴染みがあると理解が進むとされる (Widdowson, 1983, p. 105) 。SN テクストに関して，参加者の大半が双方に馴染みがないと推察されたため，OCHA 思考法と PAIL 分析を繰り返し勧めた。

OCHA 思考法は，テクストにおける言葉の使い方を「観察」(Observe) し，「観察結果を整理」して分類 (Classify) し，言語使用に関する「仮説」を導き出し (Hypothesize)，コミュニケーションに「応用」(Apply) するという思考過程を指す (野口, 2009, p. 11) 。PAIL はテクストの「目的」(Purpose)，「対象」 (Audience)，テクストに含まれる「情報」 (Information)，および「言語的特徴」 (Language) を考察することを指し，各項目の頭文字を取ったものである (野口・深山, 2010, p. 31) 。SN のサンプルテクストに関する PAIL の例として，学術誌に掲載された研究論文の内容を伝える目的で，一般の人々に向けて，研究のことを，比較的平易な言葉を用いて伝達している

ことを説明した。OCHA 思考法を用いて，SN のサンプルテクストにおいては研究者が行ったことをアナウンサーが三人称を用いて客観的に叙述していることなどに言及し，そのことを他の SN テクストに応用する方略を提示した。

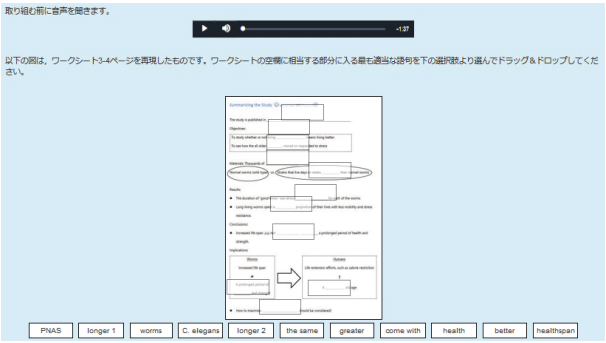
SN のサンプルテクストでは引用元の研究論文について紹介している。研究レポートの目的と読み手，種類などを紹介し，アブストラクトの構成については，SN の要約表を用いて明示的に伝えた。

2 回目以降の授業では，原則として 1 回の授業に 1 つのエピソードに関して，ポッドキャストとトランスクリプトを用いて SN を聴解し，SN で伝えられる研究の背景，方法，結果，結論を示す要約表を埋める形式の課題で内容の理解を促した。SN テクストのムーブ分析などの課題でテクスト構成に関する理解を図った (図 1) 。また，前回の省察質問として，以下のような問いについて記述式で LMS に回答してもらった。

1. サイエンスニュースで伝えられた研究でわかったことはなにか。
2. ムーブとはなにか。
3. サイエンスニュースのムーブのうち，Hook のムーブで話される内容はどんな役割（機能）を持っているか。

各々の課題に対しては，授業ビデオによるフィードバックを行った。

図 1　LMS 上の授業資料の例 (左：要約表，右：ムーブ分析)



SNの研究の内容に注目して要約する課題を授業期間中に 2 回実施した。1 つのエピソードに関して，研究レポートのアブストラクトのように，背景，方法，結果，結論という運びで記した要約の例を提示し，別のエピソードについて同様のスタイルで 70〜80 語に要約するように指示した課題シートを用意した。

LMS 上に，課題シートを用いた課題提示と復習などの具体的な手順をまとめたビデオを掲載した (図 2) 。手順の項目については，LMS 上にも掲示した。Writing 1 (1 回目) に関しては，1 つのエピソード (SN1) の要約の例を提示して別のエピソード (SN2) について要約する課題とし，要約文の案出に取り組む前に，ESP とジャンルの考え方について授業ビデオによる復習，SN1 の復習，要約の例の観察，および SN2 の復習を行うように促した。要約文は，草案を手書きまたは

電子的に行い，最終稿を LMS 上の所定の場所にタイプして提出するように LMS 上の文章とビデオで指示した。

図 2　要約課題 (Writing 1) に関する LMS 上のビデオと文章の掲示の例

Summary Writing

Orientation for Summary Writing ビデオ

Summary Writing課題の進め方を解説しています。**まずこのビデオを視聴して，PAILとOCHAの考え方を理解した上で**，以下の順番で取り組んでください。

1. このビデオの視聴
2. Study Guide (ビデオ3)の復習
3. Sunscreenの復習
4. Sunscreenのサマリーを観察
5. Breastfeedingの復習
6. Breastfeedingのサマリーライティング（下書き→タイプ入力）

要約文は LMS に記述してもらい，個々の提出物に対して，LMS 上で要約テクストの構成を中心に，テクストに含まれる情報，文法や技巧などを総合的に判断して採点し，フィードバックを行った (例 1, 2) 。学生によるフィードバックへのアクセスは任意であった。

例 1　ニュースの研究を理解する努力をしています。方法に相当する内容も入れるとなおよいでしょう。もう少し言い換えられます。 (ID 114)

例 2　内容を理解する努力をしていますが，サイエンスニュースの構造になっています。研究を報告する抄録の構造について授業ビデオの復習をお勧めします。 (ID 120)

期末試験は，対面で実施し，授業中の課題とは異なるエピソード (SN3) に関して 70〜100 語に要約してもらった (Writing F) 。

2. 4. コーパス

授業課題と期末試験の要約文 (Writing 1 と Writing F) を用いてコーパスを編纂した。Writing 1 は，LMS 上の個々の提出物をダウンロードしてテクスト化した。Writing F は，紙媒体の提出物を手作業で電子化して用いた。

2. 5. 要約文のコーディング

先の研究 (浅野・藤枝, 2021) で分析者間の協議により特定したものを本研究のテクストに照らして見直した。要約テクストにおける研究の背景 (Introduction, 例 3)，方法 (Methods, 例 4)，結果 (Results, 例 5)，結論 (Conclusions, 例 6) を伝達する記述をムーブとした。結論のムーブについては，研究論文のアブストラクトのように，テクストの最後に布置するものをコーディングの

対象とした。結論の内容が要約の第 1 文に提示される場合(例 7)，および要約の最後に SN の話の落ち (punch line) を再現している場合 (例 8)などは，SN に慣例的な要素を再構築したものと判断して一律にHookというコードを付与した。

例 3. Researchers investigated whether bre[a]st-feeding has a good effect on respiratory symptoms. (ID 109)

例 4. Firstly, researchers investigated the health records involving Facebook users and non-users, (ID 135)

例 5. They found that non-breast-fed infants were more inclined to develop respiratory symptoms than breast-fed infants by 27 percent. (ID 138)

例 6. The researchers point that they need to research another group in order to ensure whether the results are correct or not. (ID 113)

例 7. Having friends online is related to the better health. (ID 141)

例 8. [W]e all could use better ways to let affected individuals breathe easier. (ID 123)

各ムーブの出現について符号を用いた分類を筆者が２回行った。ジャンル分析における一致度や信頼性に関する Rau と Shih の研究 (2021)において，「一致率」“percent agreement” のみが有効であると結論付けられている。コーディングの一致率は，Writing 1 が 95.1%，Writing F が 95.6% であった。

コーディングの結果を 5　段階でスコア化した。アブストラクトの構成ですべての要素が出現する場合に 5，3つの構成要素が出現する場合に 4, 2 つの構成要素のみの場合に 3，1 つの構成要素のみの場合に 2とした。SNに慣例的な要素が出現する場合，その他の構成要素の有無に関わらず 1 とした。

2. 6. 分析

語彙プロファイルについて, 主に CasualConc (Version 2.1.7; Imao, 2019) と Langtest.jp (Mizumoto, 2022)，一部について R (R Core Team, 2021) を用いて調査した。要約文のコーディングと要約統計量の計算を，Microsoft Excel (Version 2111) を用いて行った。対数尤度比検定については，Langtest.jp (Mizumoto, 2022)を用いて検定を行った。また，対数尤度比を小林 (2019)に従い Microsoft Excel を用いて計算した。ピアソンの相関係数についても Microsoft Excel と Langtest.jp を用いて検討した。

3.　結果と考察

3. 1. RQ1 Writing F と Writing 1 の比較

授業課題の要約 (Writing 1) の総語数は 3,189語，異なり語数は 375語，要約文ごとの語数は☒65〜91語で, 平均 は 78 (標準偏差は 6) 語であった。期末試験での要約 (Writing F) の総語数は3,594語，異なり語数は 381語，要約文ごとの語数は 70〜110語で，平均 は 88 (標準偏差は 11) 語であった。総語数に統計学的に有意な差が認められたが (対応のある *t*検定により $t = -5.2$, $df = 40$, p

< .001, d = .82)，Writing 1 と Writing F の字数制限が各々70〜80 語と 70〜100 語であったためと推察される。

Writing 1 の頻度上位 10 語は *the, infants, of, breast, and, fed, to, asthma, that, symptoms* で，大半が SN2 の頻度上位 10 語 (*the, of, to, a, and, breast, in, infants, they, fed*) と重複していた。Writing F についても，頻度上位 10 語は *facebook, users, the, of, and, to, a, they, that, health* であり，SN3 の頻度上位 10 語 (*the, of, and, facebook, a, that, in, is, like, to*) に類似していた。

クラスター分析は，データの類似度を距離として示し，距離が近いものをまとめてクラスターを形成する (石田・小林, 2013) 。1,000 語あたりの相対頻度上位 50語を用いてユークリッド距離とウォード法によるクラスター分析を行うと，Writing 1 (右側のクラスター) と Writing F (左側のクラスター) に大別され，距離の小さなクラスターの形状より極度に酷似したテクストが存在しないことも示唆された (図 1) 。

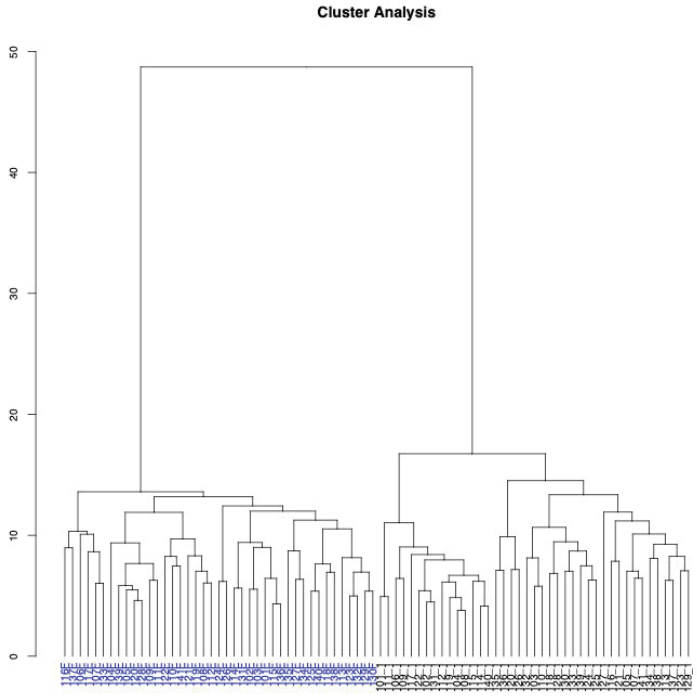


図 1　要約テクストの相対頻度上位 50 語によるクラスター分析

要約テクストにおけるムーブごとの叙述量の相対頻度を表 2 に示した。期末試験の要約文では，SN の構成要素を再現したムーブ (Hook) の叙述量に統計学的に有意な減少が認められた (対数尤度比検定による統計量 10.7, df = 1, p < .001) 。オッズ比は.53 であり，授業課題の要約文と比べてほぼ半減した。

表 2　Writing F と Writing 1 のムーブごとの叙述量 (1,000 語当たりの相対頻度)

	Writing F	Writing 1	対数尤度比	*p* 値	オッズ比	95%信頼区間
Introduction	95.4	25.7	45.3	<.01	3.99	[2.56, 6.24]
Methods	362.0	254.3	27.3	<.01	1.66	[1.37, 2.02]
Results	292.4	358.7	10.0	<.01	.74	[.61, .89]
Conclusions	191.4	256.8	12.3	<.01	.69	[.55, .85]
Hook	58.7	104.4	14.1	<.01	.53	[.38, .75]

要約テクストを背景，方法，結果，結論という運びで記述した学生は，授業課題 (Writing 1) では5名 (12.2%) であったが，期末試験での要約 (Writing F) では17名 (39.0%) であり，顕著な増加が認められた (表3) 。期末試験での要約文の例を以下に示した (例9，以下，文中に括弧で示したムーブの記載と下線による強調は筆者による) 。

例 9　[Introduction] The researchers investigated the connection between Facebook and the risk of death. [Methods] The first study population was 12 million Facebook users and non-users controlling age, race, and gender. The second study population was people having similar status. [Results] They found that Facebook users had a slightly lower risk of death than non-Facebook users by both studies. [Conclusions] They concluded that online world might reflect real world ties and interactions on Facebook might reflect actual human relationships.　(ID 122)

しかし，要約の第1文に結論を記述してSNのスタイルを再現した学生が Writing 1 に14名 (34.1%)，Writing F に16名 (39.0%) 認められた。

表 3　Writing 1 と Writing F のムーブ・スコア (N=41)

	Writing 1	(%)	Writing F	(%)
5 (4 つの構成要素が出現)	5	12.2	17	41.5
4 (3 つの構成要素が出現)	20	48.8	8	19.5
3 (2 つの要素が出現)	2	4.9	0	0.0
2 (1 つの要素が出現)	0	0.0	0	0.0
1 (SN の構成要素が出現)	14	34.1	16	39.0
平均 (標準偏差)	3.05 (1.55)	-	3.30 (1.84)	-

Writing 1とWriting Fの相関係数は $r = .40$で，中程度の相関が認められた。コーディングに基づく5段階のスコアを点数とみなすことで「相関係数の希薄化」(竹内・水本, p. 122) が生じた可能性を考慮すると相関があるといえる可能性がある。

個々の学生のWriting 1とWriting Fのスコアを比較すると，スコアが上昇した学生は17名 (41.5%)，変化がみられなかった学生は16名 (39.0%) であった。変化がみられなかった学生のスコアの内訳は，5が3名，4が4名，1が9名であった。9名 (22.0%) の学生においてスコアの減少がみられた。そのうち8名はWriting Fのスコアが1であった。残りの1名は5から4に減少しており，Writing FにIntroductionの記述がみられなかった。

期末試験での要約文において，スコアが上昇した17名とスコアが5で維持された3名の合計20名

(48.8%) の学生に，授業課題と同等あるいは改善した構成が認められた。

Writing Fのスコアの平均は，Writing 1より高かったが，ばらつきが大きく，統計学的に有意な差は認められなかった。しかし，Writing 1はLMSを通じた教室外でのタスクで，Writing Fは対面で持ち込み不可の期末試験の一部であったことを考慮すると，遠隔オンデマンド授業での演習でも，過半数の学生においてジャンルの認識が促進されたといえるであろう。

学生の背景を考慮すると，年齢が平均以上の学生において，期末試験での要約文 (Writing F) のスコアに比較的顕著な改善が認められたが，人数が16名と少なく，統計学的な有意差は認められなかった。年齢が平均未満の学生においては，期末試験の要約文のほうがスコアの平均値が低く，ばらつきも大きかった。また，Writing 1とWriting Fともに男子学生のほうが女子学生よりわずかに高かったが，スコアの伸びは女子学生のほうが顕著であった (表4) 。

表 4　Writing 1 と Writing F のムーブ・スコア (年齢別・男女別)

	Writing 1		Writing F	
背景 (人数)	平均	標準偏差	平均	標準偏差
平均年齢以上 (16)	2.94	1.57	3.75	1.69
平均年齢未満 (25)	3.12	1.56	2.92	1.91
男子学生 (26)	3.19	1.55	3.27	1.87
女子学生 (15)	2.80	1.57	3.20	1.90

Writing 1を行う際に前回の授業の省察として「サイエンスニュースで伝えられた研究でわかったことは何か」と「Moveとは何か」という問いに対して，2, 1, 0 の3段階のスコアで評価したところ，2が31名 (75.6%) と33名 (80.5%)，1が7名 (17.1%) と8名 (19.5%) ，0が3名 (7.3%) と該当者なし (0%) であり，遠隔授業で多くの学生がジャンルに関する要素を理解していたと推察される。しかし，その理解を対面での期末試験における要約文の構成に利用することができたのは半数の学生に留まった。

ジャンルの知識を戦略的に応用してライティングに取り組むことは，「書き手の習熟度に関係なく行われる」と説かれる (Tardy et al., 2020, p. 299) 。習熟度が高い言語使用者であっても，慣習などを熟知しない状況でルールに従って執筆を行うためにはジャンルの枠組みを援用することが必要となる。

また，テクスト構成に注目した Johns (1997) の演習によって，学生が研究レポートのスタイルに気づいたように，ジャンルについて理解するためには学生にとって新しいアプローチを必要とする。本研究の参加者と同じ医学部医学科第4学年を対象とした専門英語科目のジャンルに基づく授業の期末試験の成績と学術英語運用能力の指標として参照した*TOEFL ITP*®のスコアとの相関はほとんどみられなかったが (浅野，私信, 2022年1月13日) ，同学年の専門英語科目の成績と専門科目の共用試験の関連を調べた予備分析で専門英語科目の得点と共用試験の得点率とには中程度の相関があると判明している (栃澤，私信, 2022年1月19日) 。他学年の学生と比較することについては難しい面もあるが，これらの知見から，新しいアプローチを積極的に受容して学習することがジャンルの認識を深める鍵となる可能性が示唆される。

3. 2. RQ2 Writing F と *TOEFL ITP*®スコアの比較

Writing Fと*TOEFL ITP*®のスコアの相関係数は $r = .16$で，相関が認められなかった。学生の背景を考慮した際の相関係数は，男子学生において $r = .21$，女子学生において $r = .02$であり，明らかな相関はみられなかった。年齢が平均以上の学生において，$r = .29$，平均未満の学生において$r = .05$であり，年齢が平均以上の学生において，若干の相関が認められた。

フランス語を母語とする学生を対象としたCummingの研究 (1989) では，英語運用能力とライティングの技能に「有意な相互作用が認められなかった」(p. 118) とされる。また，第二言語学習と認知に関する研究方法を論じたJourdenais (2001) は「第二言語でのライティングの成果は，一般的な第二言語運用能力より母語でのライティングの経験に起因する可能性が高い」(p. 363) と述べている。Cumming (1989) は，第二言語学習者の「言語使用，談話の構成，意図などの文章[の意味以外]の面」(p. 127) に注目させることの重要性を説いており，PAIL分析 (Noguchi, 1997) の有用性が示唆される。

年齢が平均以上の学生において，Writing Fの要約文に比較的顕著な改善がみられ，Writing Fのスコアと*TOEFL ITP*®で測られる学術英語運用能力との間にわずかな相関が認められた。竹内 (2000, pp. 56–57) は「個人の莫大な知識や経験は，スキーマと呼ばれる数多くのプロトタイプ的な構造的知識として記憶されており，記憶の呼び出し，判断，内容理解など，あらゆる認知活動に関係する」と論じている。ジャンルに関する要素の知識も「読み書きの経験を積むことで既知のジャンルをより簡単に扱えるようになる」(John, 1997, p. 21) とされる。様々な経験に基づいて新しい考え方を受容し，テクストを俯瞰的に捉えることができるかどうかが鍵となる可能性がある。

なお，*TOEFL ITP*®のListeningとReading, ListeningとStructure and Written Expression, ReadingとStructure and Written Expression のスコアの相関係数は，各々 $r = .62$, $r = .52$, $r = .32$であり，相関が示唆された。

4. 終わりに

本研究では，大学1年生を対象とした学術英語の遠隔オンデマンド授業で「ジャンルに基づくタスク」(Hyon, 2018, p. 130) を行い，ジャンルの認識が促されるかどうかを調査した。研究を伝えるSNおよび研究レポートの目的，情報の受け手，テクストの構成などについて授業ビデオおよび課題へのフィードバックなどを介して伝えることで，学生がSNで伝えられる研究について，研究レポートのアブストラクトのスタイルを意識して要約することができるようになるかどうかを検討した。そして，ジャンルの認識と学術英語運用能力との関連性を考察するために，*TOEFL ITP*®スコアを参照して検討した。

得られた結果を以下に示す。

RQ1: 期末試験での要約文 (Writing F) において，授業課題の要約文 (Writing 1) と比べて，17名 (41.5%) の学生のスコアに上昇がみられ, 2回のスコアが5であった3名を加えると，20 名 (48.8%) の学生のスコアが上昇または保持された。年齢が平均以上においてWriting F の要約文に比較的顕著な改善がみられた。

RQ2: Writing F と学術英語運用能力の指標として参照した *TOEFL ITP*®のスコアに関連は示唆されなかった。年齢が平均以上の学生において，若干の相関が認められた ($r = .29$) 。

本研究における授業では，SN のテクストを理解するための方略として，聞き取り，伝えられる研究の要約表，テクスト分析，音読，要約など多面的なアプローチを行った。John (1997) は，テクストを理解するための方略には暗記，読み直し，要約などさまざまなものがあり，教員は，ひとりひとりの学生が各々の方略について「メタ認知的認識」“metacognitive awareness”を持ってタスクを完了するために使用する方略を考える力を身につけるように促すと説いている (p. 13) 。なお, 遠隔授業での要約課題を含む各種課題のフィードバックへのアクセスは任意であり，学生がフィードバックにアクセスしたかどうかは確認できていない。

Negretti と Kuteeva の研究 (2011) では，「ジャンルの認識はメタ認知的知識と密接に結びついている」と説かれる (p. 96) 。メタ認知とは，阿部と井田 (2010, p. 23) によると，メタ認知能力の評価尺度を提案した Schraw と Dennison の研究 (1994) より「人が何かを学ぶ場面での省察 (reflect)，理解 (understand) そしてコントロール (control) する能力」であると定義される。阿部と井田 (2010, pp. 24–25) は，メタ認知が，メタ認知に関する知識などの「知識的側面」と認知のプロセス，状況のモニタリング，コントロールなどの「活動的側面」に分類されると説き，「知識的側面」は，Schraw と Moshman (1995, p. 352) により「宣言的知識，手続き的知識，条件付き知識」“declarative, procedural, and conditional knowledge”に分類されていると説明している。

学習者のジャンルの知識は，Tardyらの研究 (2020) では，テクスト分析などを通じて表出するものを「ジャンルに関する宣言的知識」“a *declarative* knowledge of the genre”，ジャンルテクストの執筆などを通じて表出するものを「ジャンルに関する手続き的知識」 “a *procedural* knowledge of the genre” と分類される (p. 295; イタリック体は原著による) 。スウェーデンの大学の英語教員養成課程の学生を対象とした研究 (Negretti & Kuteeva, 2011) では「学術テクストのジャンル，談話，および修辞に関する知識に関するメタ認知の過程」を「メタ認知的なジャンルの認識」“metacognitive genre awareness” (p. 96) と定義して，すべての学生が「メタ認知的なジャンルの認識を手続き的に利用しようとした」 (p. 104) と述べている。NegrettiとKuteeva の研究 (2011) では，「文法に規則があることに気付いているという宣言的知識がその知識を用いることができるという手続き的知識になるかもしれない」 (DeKeyser, 1998, p. 48) という考えを支持して，ジャンルの特徴に関する知識が獲得され，その知識を用いるという手続き的知識に遷移させることが可能であると述べている。しかし，本研究では，「ジャンルに基づくタスク」(Hyon, 2018, p. 130) がジャンルの認識を高めるかどうかを検討するところまでとし，「ジャンル特有の知識」(Tardy et al., 2020 p. 294) 全般に関する調査には至っていない。また，NegrettiとKuteevaの研究 (2011)にみられるような, 学習で得た「宣言的知識」が演習などを通じて「手続き的知識」になる (DeKeyser, 2007, p. 9) という学習理論 (Lightbown & Speda, 2012 & 2021) を応用することについても今後の研究課題とした。

Writing FとWriting 1を精査した結果，約半数の学生の要約テクストに，論理構成の保持あるいは改善が示され，遠隔での演習でもジャンルの認識が促進されることが示唆された。しかし，残

りの半数の学生に，要約文におけるSNジャンルに特有の構成要素の再現やスコアの減少などがみられた。ジャンルの認識と学術英語運用能力の指標として参照した*TOEFL ITP®*のスコアとの明確な関連性は示唆されなかった。本研究の結果を踏まえて，ジャンルの認識には新しいアプローチを受け入れ，テクストの特徴や読み手などの状況を俯瞰的に捉えることなどを必要とする可能性について引き続き検討したい。

謝辞

本稿の執筆に際して，大阪大学大学院言語文化研究科 田畑智司教授，今尾康裕准教授，大阪医科薬科大学 藤枝美穂教授に貴重な助言をいただいた。大阪医科薬科大学IR室 栃澤健史副室長に予備分析結果の提供を受けた。記して謝意を表する。本稿の誤りはすべて筆者に帰する。本研究は，文部科学省科学研究費補助金 基盤研究 C (課題番号18K02966 研究代表者 藤枝美穂 教授)「英語医学論文投稿のための総合的支援モデルの構築」(2018〜2021年度) の助成を受けて行われた。

Appendix A: サイエンスニュース

SN1. Intagliata, C. (2017). Umbrellas plus sunscreen best bet to beat burns. Scientific American 60-second science. https://www.scientificamerican.com/podcast/episode/umbrellas-plus-sunscreen-best-bet-to-beat-burns/.

SN2. Beras, E. (2016). Breast-feeding benefits babies with genetic asthma risk. Scientific American 60-second science. https://www.scientificamerican.com/podcast/episode/breast-feeding-benefits-babies-with-genetic-asthma-risk/.

SN3 Intagliata, C. (2016). Online sociality linked to lower death risk. Scientific American 60-second science. https://www.scientificamerican.com/podcast/episode/online-sociality-linked-to-lower-death-risk/.

SN4 Intagliata, C. (2016). Ebola virus grew more infectious in the latest epidemic. Scientific American 60-second science. https://www.scientificamerican.com/podcast/episode/ebola-virus-grew-more-infectious-in-the-latest-epidemic/.

SN5 Intagliata, C. (2018). Catching flu also boosts heart risk. Scientific American 60-second science. https://www.scientificamerican.com/podcast/episode/catching-flu-also-boosts-heart-risk/.

SN6 Maron, D. F. (2015). Life extension may add just bad time. Scientific American 60-Second Science. https://www.scientificamerican.com/podcast/episode/life-extension-may-add-just-bad-time/.

SN7 Intagliata, C. (2016). Poor sleepers worse at recognizing unfamiliar faces. Scientific American 60-Second Science. https://www.scientificamerican.com/podcast/episode/poor-sleepers-worse-at-recognizing-unfamiliar-faces/

SN8 Intagliata, C. (2016). Viruses hijack the body's response to mosquito bites. Scientific American 60-Second Science. https://www.scientificamerican.com/podcast/episode/viruses-hijack-the-body-s-response-to-mosquito-bites/.

文献

Anthony, L. (2018). *Introducing English for specific purposes.* Routledge.

Asano, M. (2022). Science news on research findings in the mid-2010s: A corpus study envisaging the teaching of academic English. *Studies in the Humanities, Journal of Osaka Medical and Pharmaceutical University*, 53, 21-55.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

BBC. (2020). BBC global news podcast: China denies a cover up over coronavirus. Retrieved from https://www.bbc.co.uk/programmes/p08b9pvk on April 30, 2020.

Berkenkotter, C., & Huckin, T. (1995).*Genre knowledge in disciplinary communication* . Lawrence Erlbaum.

Bhatia, V. K. (1993). *Analyzing genre: Language use in professional settings.* Longman.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move; Using corpus analysis to describe discourse structure.* John Benjamins.

Cook, G. (2010). *Translation in language teaching.* Oxford University Press.

Cumming, A. (1989). Writing expertise and second-language proficiency.*Language Learning, 39(1),* 81-141.

DeKeyser, R. M. (1998). Beyond focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). Cambridge University Press.

DeKeyser, R. M. (2007). Introduction: Situating the concept of practice. In R. M. Dekeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1–18). Cambridge University Press.

Dudley-Evans, T. & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19(1),*61–74.

Educational Testing Service. (2020). *TOEFL®* program history. Volume 6. Available from https://www.ets.org/*TOEFL*/research/insight-series

ETS. (2022). ETS® *TOEFL ITP®* Available from https://www.ets.org/*TOEFL*_itp

Flowerdew, J. (2012). Reconciling contrasting approaches to genre analysis: the whole can equal more than the sum of the parts. In D. Belcher, A. M. Johns, & B. Paltridge (Eds.), *New directions in English for specific purposes research* (pp. 119–144). the University of Michigan Press.

Gentil, G. (2011). A biliteracy agenda for genre research. *Journal of Second Language Writing, 20,* 6–23.

Gibbons, P. (2007). Writing in a second language across the curriculum. In O. Garcia & C. Baker (Eds.), *Bilingual education: An introductory reader* (pp. 275-310). Multilingual Matters.

Hukumusume.com (2020). 桃太郎 Available from http://hukumusume.com/douwa/pc/jap/08/01.htm

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing, 12,* 17–29.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16,* 148–164.

Hyland, K. (2010). English for professional academic purposes: Writing for scholarly publication. In D. Belcher (Ed.), *English for Specific Purposes in theory and practice* (pp. 83–105). University of Michigan Press.

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly, 30,* 693–722.

Hyon, S. (2001). Long-term effects of genre-based instruction: a follow-up study of an EAP reading course. *English for Specific Purposes, 20*, 417–438.

Hyon, S. (2018). *Introducing genre and English for specific purposes.* Routledge.

Imao, Y. (2019). CasualConc (Version 2.1.7) [Computer Software]. Osaka University. Available from https://sites.google.com/site/casualconc/

Johns, A.M. (1997). *Text, role, and context: Developing academic literacies.* Cambridge University Press.

Johns, A. (2008). Genre awareness for the novice academic student: An ongoing quest. *Language Teaching, 41(2)*,

237–252.

Johns, A. (2013). The history of English for specific purposes research. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 5–30). Wiley Blackwell.

Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.),*Cognition and Second Language Instruction* (pp. 354-375). Cambridge University Press.

Kanoksilapatham, B.(2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes, 24,* 269–292.

Kuteeva, M., & Negretti, R. (2016). Graduate students' genre knowledge and perceived disciplinary practices: Creating a research space across disciplines.

Lightbown, P., & Spada, N. (2012). *How languages are learned. Fourth Edition.* Oxford University Press.

Lightbown, P., & Spada, N. (2021). *How languages are learned. Fifth Edition.* Oxford University Press.

Marshall, S. (1991). A genre-based approach to the teaching of report writing. *English for Specific Purposes, 10,* 3–13.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics.* Cambridge University Press.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech, 70,* 151–167.

Mizumoto, A. (2022). langtest.jp available from http://langtest.jp/

Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes, 50*, 40–63.

Negretti, R., & Kuteeva, M. (2011). Fostering metacognitive genre awareness in L2 academic reading and writing: A case study of pre-service English teachers. *Journal of Second Language Writing, 20*, 95–110.

Noguchi, J. (1997). Materials development for English for specific purposes: Applied genre analysis *TOEFL* pedagogy. *English Teaching, 52(3),* 303–318.

Noguchi, J. (2003). Teaching ESP writing: OCHA in a CALL class. *Cybermedia Forum, 4.*

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rau G., & Shih, Y. (2021). Evaluation of Cohen's kappa and other measures of inter-rater agreement for genre analysis and other nominal data. *English for Academic Purposes, 53*, 101026.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology, 19,* 460–475.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review, 7(4),* 351–371.

Swales, J. (1981). *Aspects of article introductions.* The Language Studies Unit, Aston University.

Swales, J. M. (1990). *Genre analysis.* Cambridge University Press.

Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge University Press.

Tardy, C. M. (2009). *Building genre knowledge*. Parlor Press.

Tardy, C. M., Sommer-Farias, B., & Gevers, J. (2020). Teaching and researching genre knowledge: Toward an enhanced theoretical framework. *Written Communication, 37(3),* 287–321.

Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writer's genre awareness, linguistic knowledge, and writing competence. *Journal of Second Language Writing, 20,* 111–133.

Widdowson, H. G. (1983). *Aspects of language teaching.* Oxford University Press.

浅野元子・藤枝美穂 (2021). サイエンスニュースを用いた学術英語への導入授業－ジャンルの理解度に関する考察 *The JACET International Convention Proceedings: The JACET 60th Commemorative International Convention,* (292).

赤塚麻里 (2013). 英語発音教育におけるプロソディー表記法の効果研究－日本人英語学習者を対象と

した実験授業から－学位論文　名古屋外国語大学

阿部真美子・井田政則 (2010).「成人用メタ認知尺度の作成の試み－Metacognitive Awareness Inventory を用いて－」立正大学心理学研究年報, 1, 23–34.

飯島優雅・渡辺敦子・マスワナ紗矢子・渡寛法・堀晋也・髙橋幸・金丸敏幸・田地野彰・寺内一 (2016).「日本の大学における学術英語カリキュラムの現状と課題－実態調査結果を踏まえて－」京都大学高等教育研究, 22, 95–98.

石田基弘・小林雄一郎 (2013). 『R で学ぶ日本語テキストマイニング』ひつじ書房

小林雄一郎 (2019). 『ことばのデータサイエンス』朝倉書店

竹内理 (2000). 『認知的アプローチによる外国語教育』松柏社

竹内理・水本篤 (2014).『外国語教育ハンドブック【改訂版】研究手法のより良い理解のために』松柏社

田地野彰 (2009). 「総合研究大学におけるEAPカリキュラム開発―専門教育との有機的連携に向けて―」福井希一・野口ジュディー・渡辺紀子 (編・著) 『ESP的バイリンガルを目指して』(pp. 130-142) 大阪大学出版会

寺内一 (2010). 「ESP の歴史と定義」 寺内一・山内ひさ子・野口ジュディー・笹島茂 (編)『21 世紀の ESP―新しい ESP 理論の構築と実践』 (pp. 3–16) 大修館書店

野口ジュディー (2009). 「ESP のススメ―応用言語学からみた ESP の概念と必要性―」福井希一・野口ジュディー・渡辺紀子 (編・著)『ESP 的バイリンガルを目指して』 (pp. 2–17) 大阪大学出版会

野口ジュディー・深山晶子 (2010). 「ジャンル分析と英語 ESP 教育への応用」寺内一・山内ひさ子・野口ジュディー・笹島茂 (編)『21 世紀 の ESP―新しい ESP 理論の構築と実践』 (pp. 26–36) 大修館書店

水本篤 (2006). 「*TOEFL* 語彙の語彙レベル、真正性、特徴の検証」言語文化学会論集, 27, 3–15.

米国一般教書演説における分離不定詞についての記述分析

—1960 年代から 2020 年の用例を中心に—

福本　広光
大阪大学言語文化研究科
〒560-0043 豊中市待兼山町 1-8
Email: u163938a@ecs.osaka-u.ac.jp

概要　本稿では，歴代米国大統領による一般教書演説 (State of the Union Address)を対象として，そこで用いられている分離不定詞の特徴を分析する。具体的には歴代大統領による演説データを集積して約 230 年分のコーパスを構築し，テクストデータにアノテーションを施して当該構造を抽出し，使用の実態を量的・質的観点から探ってゆく。分析においては，検索して得られた用例の通時的頻度推移を観察したうえで，比較的高頻度であった 1960 年代から 2020 年までに焦点を当て，その期間に使用された分離不定詞の用法について複数の観点から検討する。本試論は，一般教書演説を対象とした量的推移調査を基礎とした用例の文体論的分析を通して，特徴的な分離不定詞の用法の一端を炙り出すことが目的である。

キーワード 分離不定詞，一般教書演説，用法分析

A Descriptive Study on Split Infinitive in the SOTU:

With a Focus on the Examples from the 1960s to 2020

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

Abstract　This paper describes and analyzes some of the features of split infinitive appearing in the State of the Union Address (SOTU).　The investigation aims to shed light on some of the typical usages of split infinitive in political contexts. We construct a corpus by accumulating the data of speeches about 230 years from the end of the 18th century to the present, and annotate the text to extract the constructions' examples. In the discussion, we combine quantitative and qualitative perspectives. Specifically, we observe the transition in the frequency of split infinitive examples and focus on the period from the 1960s to 2020, when a relatively large number of contexts are available, to explore the characteristics of usage from the extracted examples. As a result of the analysis, we get several clues to uncover a part of the usage of the split infinitive structure, which is characteristic of the State of the Union Address.

Keywords split infinitive, State of the Union Address, usage analysis

1. はじめに

本稿で対象として取り扱う, (1) のような分離不定詞 (split infinitive) は中英語期から用例が確認されていたにもかかわらず, 規範文法の全盛期, つまり 18〜19 世紀にはその使用が厳しく非難されていた (Alford 1864; Fowler and Fowler 1906; Fowler 1926; Visser 1966) 。一方, 現代英語では比較的容認されるようになり(Curme 1931; Quirk et al. 1985; Huddleston and Pullum 2002 etc.), 使用される機会が増えつつある。

(1) Systematic investigations **to better understand** the likely mechanisms are urgently needed.
(COCA, 2019, ACAD)

しかし，文法書や場面状況によっては依然として批判を受けることもあり，現代英語における語法の中でも議論を巻き起こしているものの一つである。

分離不定詞はアメリカ英語において頻繁に見られる (Fischer 2007:268, Leech et al. 2009: 263, Mitrasca 2009: 118)と報告されており，筆者も特にアメリカ英語における分離不定詞の(特に通時的)な実態について大規模コーパスを用いて考察してきた (福本 2018, Fukumoto 2019) 。さらに近年では，米国大統領による演説 (スピーチ) において使用されている分離不定詞の用法についての分析も行っている (Fukumoto, forthcoming) 。

先行研究では，通時的・共時的コーパスを用いた量的分析は多く存在する (Fischer 2007, Mitrasca 2009, Calle-Martín and Miranda-García 2009, Perales-Escudero 2011, Calle-Martín and Romero-Barranco 2014, Calle-Martín 2015 etc.)ものの，大統領による演説をコーパスとして調査したものは管見の限りあまり存在しないため，考察の余地が多分に残されている。さらに，米国大統領による演説は，独立以降，時代を超えて世界的な影響力を持つ媒体の一つである。それらは内容もさることながら，語法や文体についても細心の注意が払われており，政治学は言うまでもなく，「聴衆に向けて，特定のメッセージをいかに効果的に伝達するか」という言語文化学の見地から論じられることがある。殊に就任演説など，歴史や伝統のある演説では明快で格調高い英語が採択されやすく，その演説がなされた時代の「正しい」英語の枠組みを十分に反映しているとされる[1] 。一見すると，格調高い英語を用いる傾向が強い大統領演説と，語法として批判を受けやすい分離不定詞は相容れないもののように思われるかもしれない。では，日常的なレジスターのみならず，政治的なテクストにおいてもそれらは実際に使用されているのだろうか？また，使用されているとすればどのような形においてか？ 本稿では，大統領演説の中でも長い伝統を持つ一般教書演説を対象として，これらの問いにアプローチしたい。

[1] 大統領演説テクストの持つ性質について，田畑智司教授に御教示いただいた。

2. 一般教書演説について

2.1. 概要 (Shogan (2015)を基に)

調査を行う前に，本節では今回の主たる媒体である一般教書演説について概観する。Shogan (2015)によれば，一般教書演説は米国行政府の長たる大統領が連邦議会議員を対象に行う施政方針演説であり，そこで大統領は過去一年の現状報告や政治課題について説明する。大統領の視点からは，政治要綱のすべてを一つの演説に盛り込んで伝える特別な機会であり，かつては憲法によって定められた義務から始まったものの，今や行政権や権威を行使するための原点にして，自らの権力について示す格好の場でもある。一方，連邦議会からも大統領の施政について詳細を知ることができる重要な演説の一つと目されている。

本演説は歴史も長く，初代大統領 George Washington によって 1790 年 1 月に最初になされて以来，原則として毎年行われている。時代の流れとともに形式や伝達方法が変化しているのも大きな特徴であり，1790 年から 1801 年までは口頭にて演説がなされていたが，Thomas Jefferson によって文書による伝達に形式が変更されたのを機に，1801 年から 1913 年までの約 110 年間にわたって大統領は年次報告を議会宛の公式文書として提出した。1913 年に Woodrow Wilson によって再び口頭での演説形式に戻されると，1934 年までハイブリッド，その後は 1981 年まではどちらの形式が取られることもあった。1981 年以降は口頭演説に一元化され，現在に至る。Shogan (2015) は，こうした形式や演説の長さの変化の背景にはテクノロジー，すなわちラジオやテレビ，インターネットの出現が重要な役割を果たしていると述べているが，これらのメディアの発達により，連邦議会や米国民のみならず全世界に注目される一大イベントとなっている。

2.2. 研究の動機

前節で見たように，一般教書演説は長い伝統を持ち，世界的な影響力もある。そのテクストの中で，使用の可否が問題となりうる分離不定詞は使われてきたのであろうか？既に述べたように，そのことに踏みこんだ先行研究は現時点では存在しない。そして一般教書演説は歴代大統領によってほぼ決まって行われており，形式，内容や意味合いに若干の変化はあれども，概ね均質かつ一定量のテクスト (18 世紀以降，現在までの約 230 年間のデータ) をまとめて得られることから，その時代の英語を十分に反映した歴史的言語資料として活用できると判断した。

3. データ・調査方法

本調査にて扱うデータを作成するべく，一般公開されたデータベース *The American Presidency Project* のサイトに所収の演説のスクリプトを集積する。本データベースは厳密にはコーパスではないが，そこから得られたスクリプトは電子テクストに変換することができる。そのためこの類の研究には有用であると考えられ，調査の前段階として，データのコーパス化を行った。 以下の表 1 は，本研究にて調査対象とする演説の一覧である。

表 1　一般教書演説コーパスの構成要素[2]

==========

Washington (1790-1796); Adams (1797-1800); Jefferson (1801-1808); Madison (1809-1816); Monroe (1817-1824); J.Q. Adams (1825-1828); Jackson (1829-1836); van Buren (1837-1840); Tyler (1841-1844); Polk (1845-1848); Taylor (1849); Fillmore (1850-1852); Pierce (1853-1856); Buchanan (1857-1860); Lincoln (1861-1864); Johnson (1865-1868); Grant (1869-1876); Hayes (1877-1880); Arthur (1881-1884); Cleveland1 (1885-1888); B. Harrison (1889-1892); Cleveland2 (1893-1896); Mckinley (1897-1900); Roosevelt (1901-1908); Taft (1909-1912); Wilson (1913-1920); Harding (1921-1922); Coolidge (1923-1928); Hoover (1929-1932); F.D. Roosevelt (1934-1945); Truman (1946-1952); Eisenhower (1953-1960); Kennedy (1961-1963); Lyndon Johnson (1964-1969); Nixon (1970-1974); Ford (1975-1977); Carter (1978-1981); Reagan (1982-1988); Bush (1989-1992); Clinton (1993-2000); G.W. Bush (2001-2008); Obama (2009-2016); Trump (2017-2020)

==========

分離不定詞は "*to* + 副詞 (相当句) + 原形動詞"という，特定の品詞語句の連結によって構成されているが，該当する構造を適切に抽出するためには，品詞情報の付与が不可欠となる。本研究では Brill Tagger を用いてアノテーションを行ったうえで，タグ付けされたテクストについて， コンコーダンサ AntConc を用いて当該構造を抽出する。[3] 次節以降で，本調査によって得られた結果を示し，分析を行う。[4]

4. 史的頻度推移の観察

以下の図 1 は，上記の手順で検索を行った結果，歴代大統領の一般教書演説において分離不定詞がどれほど使用されてきたかについて，その史的推移をあらわしたものである。(頻度は 1 万語あたりに調整済み)

---

[2] 各大統領には，演説を行うにあたりその原稿を執筆するスピーチライターが存在する。原稿を実際に書いているのは大統領自身ではなくスピーチライターであるため，演説に用いられている言語が必ずしも「大統領による英語」とは言い切れない可能性があることを研究会や学会にて指摘いただいている。筆者としては，確かに分析を行う際にはスピーチライターの存在を無視することはできないが，原稿を最終的にチェック，改訂した上で自分の言葉で発表するのは大統領本人であるということから，演説にて使用されている言語を「大統領による英語」とみなしたいと考えている。

[3] 分離不定詞を抽出する際の検索式は, "to_TO *_RB* *_VB"並びに"to_TO *_RB* *_VBP"である。

[4] 自動タグ付けは文脈を考慮していないため，品詞タグの付与が誤っている場合もある。今回の調査においても，動詞を修飾する副詞として使われているはずの *better* に形容詞 (JJ)のタグが付けられていたほか，副詞であるはずの *further* に動詞 (VB)のタグが誤って付与されていた。そのため，今回は CasualTreetagger および CasualConc も用例数の確認用として使用した。

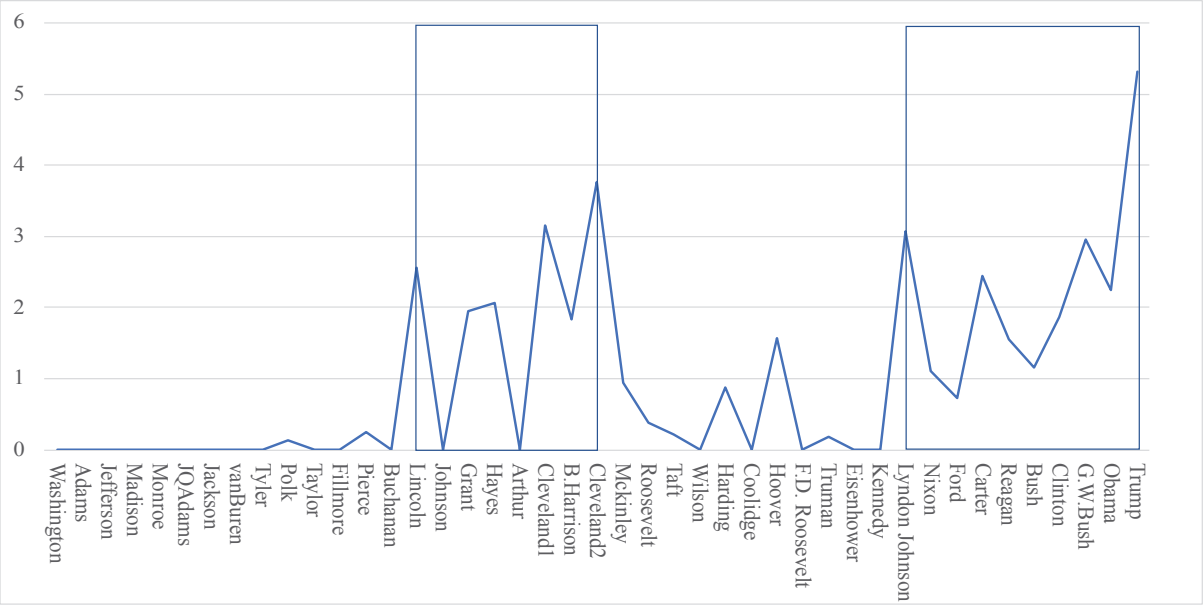


図 1 歴代米国大統領の一般教書演説に出現する分離不定詞の史的頻度推移

グラフから読み取れることとして， 18 世紀末以降, 分離不定詞は規範意識の影響によるものか, 19 世紀中頃になるまで一般教書演説中には出現することはなかった。本コーパスから得られた初例は第 11 代大統領 James Polk による, 1846 年のものであった。一般教書演説において分離不定詞を多用し始めたのは Abraham Lincoln である。彼は在任中の 4 回の全ての演説において最低一度は含めていた。彼を転換点として, 20 世紀初頭にかけては大統領ごとに許容度の差はあるものの, 演説中に分離不定詞が比較的多用されるようになった。一方で, 1910 年以降の大統領は使わないことが多くなり, この期間においては用例が一例も確認できなかった人物も少なくなかった。一方 20 世紀後半以降になると, 第 36 代大統領 Lyndon Johnson を皮切りとして（頻度的な揺れはあるものの）再び上昇傾向となり, 現在に至るまで任期期間中に一度も使用しなかった大統領は存在しない。とりわけ Donald Trump は, 1 万語当たりの相対頻度から, 歴代大統領の中で最も多くの分離不定詞を演説中に用いている。

これらのことから, すでに紹介したように一般教書演説はその形式に関して歴史的な変化があることに注意する必要はあるが, 頻度については概ね「19 世紀後半から 20 世紀初頭まで」、並びに 「20 世紀後半以降, 現在に至るまで」 (図 1 の枠で囲った部分) の期間は比較的多くの大統領が分離不定詞を用いており, それ以外の期間には用例数が極めて少ない」という量的傾向があると言ってよいだろう。この知見を基礎に, 次節では予備的分析として, 特に分離不定詞の使用頻度が高い大統領が多くいた 20 世紀後半の 1960 年代から 2020 年にかけて, その特徴的な用法について記述する。その際のリサーチクエスチョン (RQ) は以下の通りである。

(i) 一般教書演説にて使用されている分離不定詞の「構成要素」たる副詞および原形動詞にはどのようなものがあるか。

(ii) 分離不定詞と共起する表現や要素にはどのようなものがあるか。

5. 1960 年代以降の一般教書演説に出現する分離不定詞の特徴的用法についての分析

本節では, Lyndon Johnson の初年度 (1964 年) から Donald Trump の最終年度 (2020 年) までの期間に用いられた分離不定詞構造の特徴について, 複数の観点から考察してゆく。本調査では全部で 80 例が収集された。

5.1. 分離不定詞を構成する副詞および原形動詞について

まず本節では特に, RQ1 に答えるため, 構成要素となっている副詞および原形動詞に着目する。

5.1.1. 対象となる時代全体の傾向分析

対象となるレジスター (ここでは演説) の性質のためか, 分離不定詞の splitter (*to* 不定詞を「分割」する副詞相当句) となり得る副詞には, 「程度」や「時」の意味を表す副詞がその多くを占めており, 先行研究 (Albakry 2005: 146; Mitrasca 2009: 122 ほか) では小説などのテクストを調査対象とするものが多かったこともあり, 分離不定詞の構成要素として「様態」の副詞が中心的とされていたが, 今回の調査ではあまり多く見られなかった。以下の表 2 は, 今回抽出された分離不定詞の中で, 複数回, 分離不定詞の一部として使用されていた副詞の上位 13 項目とそれらに後続する動詞についての対応関係を示したものである。紙幅の都合で全てを示すことはできず, 早計な結論は出せないものの, 大まかな様相は見えてくる。

表 2 分離不定詞の構成要素として複数回使用されていた副詞および後続する動詞一覧[5]

splitter副詞	後続する動詞							
fully	control	deregulate	participate	fund	fund	restore	rebuild	fund
finally	end	put to rest	approve	slash	repeal	pass	close	end
better	protect	deliver	identify	secure	meet			
further	improve	protect	develop	protect	expand			
ultimately	have	eliminate	control	make clean				
just	review	mention	stop	talk				
successfully	participate	combine	deal with					
always	debate	be	remember					
quickly	enact	make available						
really	improve	solve						
carefully	explore	assess						
permanently	fix	extend						
greatly	increase	expand						

[5] 表中の後続動詞の並びは, 用例の使用年度の古い順である。

表 2 から読み取れることとして，やはり「程度」や「時・頻度」を表す副詞が構成要素として使われているパターンが多くを占めており，後続する動詞としてはそのほとんどが動作動詞であるというだけでなく，他動詞であるということも読み取れよう。このことから，分離不定詞を用いることにより，大統領が「特定の動作」に対してどのような「評価」を下しているのかについてはっきりと示すことができる。これは表 2 に現れていない構成要素についても同様のことが言えるため，大統領演説，少なくとも一般教書演説においては「*to* + 程度/時を表す副詞 + 他動詞の動作動詞」という語順が最も典型的なものであると考えられる。

5.1.2. 同じ型の分離不定詞を使用した用例

抽出された分離不定詞に使用されている副詞・動詞はきわめて多様であったけれども，全く同じ型の分離不定詞が別の大統領によって使用されていた事例がいくつか存在することは特筆すべき点かもしれない。[6] それらの用例を以下にて紹介する。

● ***to further protect***

(2) Now we intend to maintain the momentum we have developed by taking additional steps to further improve law enforcement **and <u>to further protect</u> the people of the United States.** **(Nixon 1973)**

(3) And so as we continue to diversify our fuel supply, we must step up domestic oil production in environment sensitive ways. **And <u>to further protect</u> America** against severe disruptions to our oil supply, I ask Congress to double the current capacity of the Strategic Petroleum Reserve. **(G.W. Bush 2007)**

上記 2 例では, Nixon と G.W. Bush が全く同じ分離不定詞の *to further protect* というフレーズを使用している。フレーズ自体が全く同じというだけでなく，その前後の接続詞，目的語も全く同じ (網掛け部分) ということも目を引くポイントである。この 2 人はいずれも共和党出身の大統領であることから, G.W. Bush が分離不定詞を使用するにあたり，過去の大統領による事例を参照して，敢えて同じ用法を使用している可能性を示唆するものである。

● ***to fully fund***

(4) Everyone in this chamber knows that Social Security is not prepared **<u>to fully fund</u>** their retirement. **(G.W. Bush 2001)**

(5) This program strengthens democracy, transparency, and the rule of law in developing nations, and I ask you **<u>to fully fund</u>** this important initiative. **(G.W. Bush 2008)**

(6) I am asking Congress **<u>to fully fund</u>** the Artemis program to ensure that the next man and the first woman on the Moon will be American astronauts, using this as a launching pad to ensure that America is the first nation to plant its flag on Mars. **(Trump 2020)**

---

[6] 大統領演説における定型表現などのフレーズには，過去の大統領の演説で用いられた言い回しを参考にする可能性があることについて，渡辺秀樹教授に御教示いただいた。

このパターンは，使用されている目的語は異なるけれども，類似した構文のもとで使用されていることが注目すべき点かと思われる。(5.2.1節も参照されたい)

● ***to finally end***

(7) With new resources, NIH will now become the most powerful discovery engine for an AIDS vaccine, working with other scientists **to finally end** the threat of AIDS. **(Clinton 1997)**

(8) We are working **to finally end** America's longest war and bring our troops back home. **(Trump 2020)**

この事例では，使用している大統領の政党は異なるが，これも互いに酷似した表現であると言えるだろう。

「全く同じ」形の分離不定詞が異なる大統領によって繰り返し使用されたのは本データからはこの 3 種類のみである。そのうちの一部はフレーズのみならず，前後の言い回しも類似したものを使用していたことから，大統領は語法についてもある程度それ以前の大統領による言い回しを踏襲しているということが示唆された。

以上から，ここまでの成果をまとめて RQ1 の答えとしたいが，少なくとも 1960 年代以降の大統領による分離不定詞としては，（それら以外のものもあるとはいえ）「不定詞マーカー*to*＋ 程度あるいは時の副詞 ＋ 他動詞の動作動詞」というパターンが大半を占めており，特定の動作に対する大統領による評価を明確に示すということを表していると考えられる。また，分離不定詞の構成要素として使用される副詞や動詞の種類は多様であるが，中には全く同じ組み合わせのもとで使用されていることもあり，過去の大統領による用例を参照している可能性が示された。

5.2. 共起する表現や要素について

本節では，RQ2 に答えるため，分離不定詞の用例に共起して使用されている表現や要素について検討する。

5.2.1. 命令や依頼の対象としての分離不定詞

この期間の周辺要素として特に目を引くのが，「～することを求める」という意味を表す構文と共起して，分離不定詞がその対象として扱われている事例がかなり多いことである。全部で 16 例確認され，以下の用例はその一部である。

(9) Tonight, I plead with the House **to immediately act** upon these measures and I hope take favorable action upon all of them. **(Lyndon Johnson, 1968)**

(10) I am therefore calling upon the Congress **to promptly enact** a new Heroin Trafficking Act. **(Nixon 1973)**

(11) I urge the Senate to move quickly and decisively in confirming Judge Anthony Kennedy to the highest Court in the land and **to also confirm** 27 nominees now waiting to fill vacancies in the federal judiciary. **(Reagan 1988)**

(12) Last year, I asked you **to voluntarily cut** the number and cost of earmarks in half. **(G.W. Bush 2008)**

上記の網掛け部分が命令や依頼に該当するが，ここの表現や時制なども大統領によって異なっており，目的語として *Congress* や *Senate* といった「議会」， *you* 「議員, 国民」などを取る。
すなわち，離不定詞を効果的に使うことによって，大統領の議会や他者に対する要望について明確に示すことができるという働きがあることがわかる。

尤も，2.1 節で見たように，一般教書演説は時代とともに大統領による議会への権力誇示の場としての側面が強くなっていったという背景があり，そもそも命令や依頼について表明する文脈は他のテクストに比べても多くなっている。従って，この分析によって得られる事柄も，ある意味では当然の結果と言えるかもしれない。さらに，これらの分離不定詞に先行する動詞は *ask* や *urge* といったものであるが，これらは *to* 不定詞を後続させるので，分離不定詞が生起しやすい環境にあるということも事実である。とはいえ，この形式，構文のもとで使用される分離不定詞は，管見では他のテクストジャンルにおいて見ることはほとんどなかった。一方，この時期に大統領を務めたほとんど全ての人物が最低一度はこの形式において分離不定詞を用いている。[7] 従って，これはその演説に出現する分離不定詞に特徴的な用法の一つとして捉えても良いと思われる。今後の課題としては，「20 世紀後半の一般教書演説に特有の用法と言えるか」については疑問が残る。もし他の時代においてもこのタイプの分離不定詞が頻繁に確認されるということであれば，それは時代によるものではなく，ジャンルの性質によるものである，と考えられるためである。

5.2.2. 先行接続詞 *and* との共起

また，接続詞 *and* と共起し，列挙された事柄 (主に *to* 不定詞)の一つとして分離不定詞が用いられている用例も散見された。

(13) A second principle of policy is the effort to control, *and* to reduce, ***and*** **to ultimately eliminate** the modern engines of education. **(Lyndon Johnson 1966)**

(13)のように, to 不定詞が複数回並置されている場合，その最後の箇所に分離不定詞が用いられる，というパターンが確認された。さらに特徴的な用法として, *and* が分離不定詞同士の平行構造を形成するために使用されるという例もあった。[8]ここでは「*to* + *further* + 動詞」という一連の形式が *and* を用いて並列されている。以下に，先ほど(2)にて検討した文を再掲する。

(2)' Now we intend to maintain the momentum we have developed by taking additional steps **to further improve** law enforcement ***and*** **to further protect** the people of the United States. **(Nixon 1973)**

---

[7] Lyndon Johnson 2 回, Nixon 1 回, Carter 1 回, Reagan 1 回, Clinton 3 回, G.W. Bush 4 回, Obama 1 回, Trump 3 回
[8] 本用例が平行構造をなしていることについて，渡辺秀樹教授より御示唆をいただいた。

さらに, 5.1.2 節にて取り上げた, (2) の *to further protect* と同じ構造の分離不定詞については, 文頭に *and* が置かれているが, これはいわゆる接続詞 *and* の「拡張的」用法 [9] であると考えられる。

(3)' ***And*** **<u>to further protect</u> America** against severe disruptions to our oil supply, I ask Congress to double the current capacity of the Strategic Petroleum Reserve. **(G.W. Bush 2007)**

このように, 分離不定詞は *and* と様々な用法のもとで共起していることがわかった。

それでは, RQ2 の回答をすることによって本節のまとめとしたいが, 一般教書演説で大統領が使用している分離不定詞に先行して共起する要素から用例を観察した結果, 命令や依頼の対象として機能させている用法や, 列挙, 並列構造の並置, 拡張的用法など多様な用法を持つ接続詞 *and* との共起を確認することができ, ディスコースの中で分離不定詞が強調される用法が特徴付けられた。

6. おわりに

本研究では, 分離不定詞構造が現代アメリカ英語においてどのような使用のされ方をしているかについて検討するために, 歴代米国大統領による演説, 特に歴史ある一般教書演説テクストをコーパスとして予備的な調査分析を行った。

量的変化としては, 分離不定詞は 19 世紀後半から 20 世紀初頭にかけて使用され始め, 20 世紀前半に一時的に減少したが, 後半以降再び増加傾向にある。すなわち, 一般教書演説のような伝統あるジャンルにおいても, 特に近年ではごく自然に発見できる構造であることがわかる。20 世紀後半以降, 特に 1960 年代以降の増加傾向が顕著であったことから, この期間において使用されている分離不定詞の用法について考察した。今回抽出された用例からは「*to*+ 程度/時をあらわす副詞 ＋他動詞の動作動詞」というパターンが最も多く, 特定の動作に対する大統領による評価を明確に示すために本構造が用いられているというのが自然であると考えられる。さらに, 命令や依頼, あるいは列挙や並列の *and* と共起して, ディスコースの中で強調されながら用いられるということも明らかにした。

ただし, 本試論はあくまで 1960 年代以降の特徴についての分析であり,「分離不定詞の特徴的用法」と指摘した点については他の時代における用例と比較することが今後必要となる。さらに, 今回指摘した事柄が演説に「特徴的」と言えるかについては, 演説以外のテクストとも比較しつつ, より高度な統計的手法も用いながら処理・分析される必要があるだろう。それらについては今後の課題とし, 別稿にて議論する予定である。

---

[9] ここでの文頭の *and* が「拡張的」用法であることについて, 渡辺秀樹教授より御示唆をいただいた。

注記・謝辞

本稿は， 2021 年 10 月 28 日に行われた大阪大学言語文化学会第 59 回大会での，筆者の口頭発表原稿に加筆修正を施したものである。学会発表に先立つ研究会にて発表の機会と有益な御助言を下さった，田畑智司先生をはじめとする先生方ならびに院生メンバーの皆様に感謝申し上げる。草稿作成にあたり貴重な御教示を下さった渡辺秀樹先生，および学会にて的確な御質問・御指摘を下さった大森文子先生，田村幸誠先生にもこの場をお借りして深く感謝申し上げたい。

文献

Albakry, M.A. (2005) *Style in American newspaper language: Use and usage.* Unpublished Doctoral Dissertation Thesis, Northern Arizona University.

Alford, H. (1864) *The Queen's English; Stray notes on speaking and spelling.* London: Strahan.

Calle-Martín, J. (2015) "The Split Infinitive in Middle English." *North-Western European Language* 68(2): 227–259.

Calle-Martín, J. and Antonio, Miranda-García. (2009) "On the Use of Split Infinitives in English." In Antoinette, R and Andrew, K (eds.) *Corpus Linguistics: Refinements and Reassessments*. 347–364. Amsterdam: Rodopi.

Calle-Martín, J. and Romeo-Barranco, J. (2014) "On the use of the split infinitive in the Asian varieties of English". *Nordic Journal of English Studies*, 13 (1) 130–147.

Curme, G.O. (1931) *A Grammar of the English language. Volume3: Syntax.* Boston; London: DC Heath.

Fischer, Roswitha. (2007) To Boldly Split the Infinitive - or Not? Prescriptive Traditions and Current English Usage. In Stephan E., Nils L, Joachim S. and Wim V.(eds.) *Germanic Language Histories "from Below" (1700–2000)*, 259–273. Berlin: Walter de Gruyter.

Fowler, H.W. (1926) *A dictionary of modern English language.* Oxford: Clarendon Press.

Fowler, H.W. and Fowler, F.G. (1906) *The King's English.* Oxford: Clarendon Press.

福本広光 (2018) 「アメリカ英語での分離不定詞使用実態調査と考察：使用されたsplitterを中心に」渡辺秀樹 (編) 『言語文化共同研究プロジェクト2017 レトリック、メタファー、ディスコース』75–87. 大阪大学言語文化研究科.

Fukumoto, H. (2019) "A Corpus-Based Analysis of Split Infinitives in American English: With Special Reference to "Splitters" from Diachronic and Rhetorical Viewpoints." 大阪大学言語文化研究科 (修士論文).

Fukumoto, H. (forthcoming) "Split Infinitive in American Presidential Speeches: Diachronic and Stylistic Approach Focusing on the Cases of Abraham Lincoln." 『文体論研究』第 68 号. 日本文体論学会.

Huddleston, R. and Pullum, G.K. (2002) *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Mitrasca, M. (2009) "The split infinitive in electronic corpora: Should there be a rule?" *Concordia Working Papers in Applied Linguistics,* 2: 99–131.

Perales-Escudero, M. D. (2011) “To Split or to Not Split: The Split Infinitive Past and Present.” *Journal of English Linguistics* 39 (4): 313–334.

Quirk, R., Greenbaum, S., Leech, J. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language.* London: Longman.

Shogan, C.J. (2015) “The President’s State of the Union Address: Tradition, Function, and Policy Implications.” (https://fas.org/sgp/crs/misc/R40132.pdf)

Visser, F, Th. (1966) *An Historical Syntax of the English Language. Part II: Syntactical Units with One Verb (Continued)*. Leiden: E. Brill.

データベース

The American Presidency Project. (https://www.presidency.ucsb.edu/) (2022年1月20日最終閲覧)

コンコーダンサ等

Brill Tagger (https://cst.dk/online/pos_tagger/uk/index.html)

AntConc (https://www.laurenceanthony.net/software/antconc/)

執筆者紹介

田畑　智司（たばた　ともじ）	大阪大学大学院言語文化研究科教授
浅野　元子（あさの　もとこ）	大阪医科薬科大学助教
黒田　絢香（くろだ　あやか）	大阪大学大学院言語文化研究科 博士後期課程
藤田　　郁（ふじた　いく）	大阪大学大学院言語文化研究科 博士後期課程
福本　広光（ふくもと　ひろみつ）	大阪大学大学院言語文化研究科 博士後期課程

言語文化共同研究プロジェクト 2021

テクストマイニングと
デジタルヒューマニティーズ
2021

2022 年 3 月 31 日発行

編集発行者　大阪大学大学院言語文化研究科

この用紙は人類環境浄化のため再生紙を使用しています。