



Title	科学技術日本語は二つの場面においてどのように異なるのか？ : 科学技術系の学術論文とブログに対するMulti-Dimensional分析
Author(s)	Chen, Yukai
Citation	
Version Type	VoR
URL	https://doi.org/10.18910/100479
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

科学技術日本語は二つの場面においてどのように異なるのか？

—科学技術系の学術論文とブログに対する Multi-Dimensional 分析—

CHEN YUKAI (言語文化学・D1)

1. 研究目的

本研究の目的は、科学技術日本語は専門家コミュニティと一般向けという二つの場面においてどのように異なるのかを明らかにすることである。特に、学術雑誌上の論文とブログ上のテキストを例に、その言葉遣いの異なりを明らかにする。

人々が特定場面で特定の言語変種を使っており、このような特定場面と関わる言語変種はレジスターと呼ばれている (Biber & Conrad, 2019: 6)。また、異なる場面で異なるレジスターを使うことはレジスターバリエーションと言われている (Ferguson, 1994: 16)。このようなレジスターバリエーションの観点を、科学技術日本語の言葉遣いにも当てはめることができる。

科学技術日本語の範疇は、科学技術分野の内容を表す全ての日本語を含めている (畠田谷, 2001)。今まで、科学技術日本語は、理工系の留学生の勉強・研究活動をサポートするため、学術論文などを代表とした専門家コミュニティ内部の日本語にフォーカスする研究 (仁科 & 武田, 1993; 札野 et al., 2012) がメインであった。一方、科学技術分野の内容を専門外の人々に紹介するというサイエンスコミュニケーションの場面も存在している。

本研究はいままで扱ってきた専門家コミュニティ内部の場面から一歩踏み出し、サイエンスコミュニケーションの場面における科学技術日本語も射程範囲内に入れた。特に、この二つの場面での言葉遣いの異なりに着目している。

2. データ収集

本研究は、16 誌¹の科学技術系学術雑誌の日本語テキスト、および「note」²上で 20 の科学技術系ハッシュタグを含む無料記事 (2021～2023 年投稿のもの) を収集した³。さらに、学術雑誌からは、「論文」「研究」「報告」「総説」「論説」という文字が類別表示のメタデータに出現したドキュメントを抽出した。なお、メタデータがない場合はタイトルに「分析」「研究」を含むドキュメントを抽出した。note のデータから、専門家、専門組織の投稿テキストを抽出した⁴。テキストの前処理において、学術雑誌のテキストではタイトル、英文段落、テーブル、参考文献および引用元を示す括弧を削除し、数式を [MathExpr] という表記に置換した。note では英語段落を削除し、数式を「MathExpr」に置換した。

データ前処理の後、ginza (version: 5.2.0) と ja-ginza-bert-large (version: 5.2.0b1) でテキストデータのトークン化を行った。収集したデータのドキュメント数とトークン数の分布は表 1 の通りである。

¹ 収集した学術誌は、日本統計学会誌、日本惑星科学会誌遊星人、分析化学、地質学雑誌、哺乳類科学、育種学研究、日本公衆衛生雑誌、日本内科学会雑誌、日本口腔インプラント学会誌、薬学雑誌、自動車技術会論文集、材料、日本建築学会計画系論文集、日本機械学会論文集、電気学会論文誌C、人工知能学会論文誌である。電気学会論文誌Cは 2017-2019 年の文章であり、ほかの学術雑誌は 2021-2023 年の文章である。

² note は、「クリエイターが文章や画像、音声、動画を投稿して、ユーザーがそのコンテンツを楽しんで応援できるメディアプラットフォーム」である (<https://note.com/info/n/nea1b96233fbf> 2024 年 12 月 23 日閲覧)。

³ 学術雑誌は一般に学会の会員に向けて発行しており、閉ざされた専門家コミュニティの場面に使うものである。一方で、ブログは閲覧数が多いほど広告や有料記事などの収益が獲得でき、より一般向けであると考えている。

⁴ 下記いずれかの文字列を含むアカウントプロフィールまたはアカウント名の投稿記事を抽出した：「博士|修士|教授|講師|助教|教員|学者|研究者|研究員|ポストドク|科学者|サイエンティスト|専門家|エキスパート|技術者|開発者|プログラマー|医者|医師|薬剤師|建築家|建築士|職人|技師|エンジニア|コンサル|アナリスト|学系|実験室|ラボ|研究室|大学|研究所」

表 1：データ分布

	ドキュメント数	トークン数
学術論文	2,735	19,353,599
note	1,905	3,160,950
合計	4,640	22,514,549

3. 分析方法

本研究は、Biber(1988)が英語の言葉遣いの研究で使用した Multi-Dimensional 分析 (MDA) を採用し、それを日本語のテキストに適用している。MDA は、因子分析を通じて、テキスト内で頻繁に共起する（相関の高い）言語項目をいくつかの次元（潜在的なコミュニケーションの機能）に圧縮し、さらに各次元の状況に基づいてテキストの特徴を分析する方法である。

MDA を行うために、語彙、文法を含む言語項目の設定が必要である。本研究は Biber(1988)の英語言語項目を参考にしながら、抽出可能な日本語の基本的な語彙、文法の言語項目を広く網羅すると同時に、単語親密度、単語抽象度などの言語項目も取り入れた。最終的に設定した言語項目は表 2 の通りである。

表 2：言語項目

リーダビリティ	リーダビリティ	1	STTR	語彙	機能語	26	接尾辞_形容詞
		2	漢字比率			27	接尾辞_形状詞
		3	単語親密度			28	接続詞
		4	内容語含有率		指示詞	29	こ系
		5	語彙抽象度			30	そ系
		6	依存距離			31	あ系
語彙	内容語	7	第一人称代名詞	文法	中止形	32	ど系
		8	第二人称代名詞			33	第一連用形
		9	第三人称代名詞			34	第二連用形
		10	名詞		授受表現	35	てあげる
		11	動詞			36	てもらう
		12	形容詞			37	てくれる
	機能語	13	形状詞		常体敬体	38	です調
		14	副詞			39	である調
		15	格助詞			40	だ調
		16	係助詞		モダリティ	41	意志
		17	準体助詞			42	命令
		18	接続助詞			43	疑問
		19	副助詞			44	推量
		20	終助詞		その他	45	根拠
		21	助動詞			46	解釈
		22	感動詞			47	例示
		23	フィラー			48	曖昧表現（例：約、およそ、おむね、ほぼ、大体、だいたい）
		24	接尾辞_名詞			49	縮約形（例：ちゃう、じゃう）
		25	接尾辞_動詞				

設定した言語項目を正規化するため、言語項目がカウント数の場合は 1,000 トークンごとの出現頻度を算出した後、Z スコアに転換した；言語項目が平均値の場合はカウント数を直接に Z スコアに転換した。後の因子分析は、各言語項目の Z スコアで行っている。

4. データ分析と結果

具体的なデータ分析は factor_analyzer (version: 0.5.1) を使用している。相関が 0.8 以上の言語項目を削除した後で得た KMO 値が 0.9037 であり、因子分析に適している数値だと考えている。また、プロマックス回転を行った。図 1 のスクリープロットから判断し、第 3 因子の直前まで固有値が大幅に下がっているため、左側からの 2 つの因子が元のデータの分散を最も解釈できると考えている。今回は 2 因子を抽出した。

第 1 因子の因子負荷では、助動詞、接続動詞、動詞が高い正の負荷を占めており、一方で名詞、内容語が高い負の負荷を占めている。第 2 因子では、格助詞、語彙抽象度が高い正の負荷を占めており、終助詞と感嘆詞が比較的に高い負の負荷を占めている。

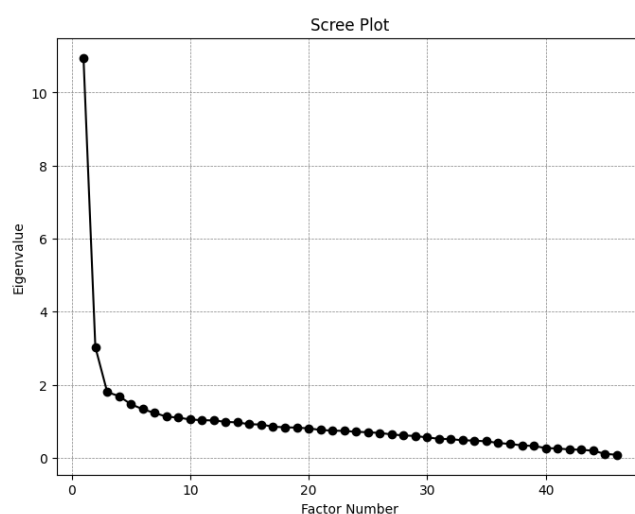


図 1：スクリープロット

そこで、第 1 因子を動詞化一名詞化の次元と名付け、第 2 因子をフォーマル—カジュアルの次元と名付けた。学术论文と note のテキスト上の因子スコアは表 3 の通りである。

表 3：因子スコア

	第 1 因子	第 2 因子
学术论文	-5.04	0.82
note	13.56	-1.91

第 1 因子では、学术论文は名詞化の表現に偏っており、note は動詞化の表現に偏っている；第 2 因子では、学术论文はフォーマルな表現に偏っている一方、note はカジュアルな表現に偏っていると考えている。

参考文献

- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge University Press.
- Biber, D., & Conrad, S. (2019). *Register, Genre, and Style* (2nd ed.). Cambridge University Press.
- Ferguson, C. A. (1994). Dialect, register, and genre: Working assumptions about conventionalization. *Sociolinguistic Perspectives on Register, 1994*, 15–30.
- 札野寛子, 藤井清美, 魚立康夫, & 松橋由佳. (2012). 「日本人学生とのインタラクションを通して学ぶ短期留学生のための

問題発見解決型科学技術基礎日本語プロジェクト」『専門日本語教育研究』 14, 35-40.

仁科喜久子, & 武田明子. (1993). 「助詞相当句を中心にみた科学技術日本語テキストの分析-理工系大学院留学生の専門書読解のために-」『東京工業大学人文論叢』 19, 139-152.

畝田谷桂子. (2001). 「科学技術日本語の提案」『留学生センター報告書 2001』 11-19.