

英語コーパス学会
大会予稿集2025

JAECS
CONFERENCE | **2025**
PROCEEDINGS

2025.9.2

JAECS
Japan Association for English Corpus Studies

ISSN: 2436 - 6647

目次

川越 涼太	
英語リスニングテキストにおける難易度決定要因の分析と自動推定の可能性	1
田畑 圭介／山縣 節子	
単純形副詞のタイプと容認性について	10
南 佑亮	
No point/sense/use の補文における形式的変異について	16
SHIBUYA Yoshikata／JENSEN Kim Ebensgaard	
<i>Bro</i> Constructions across America—A Twitter Corpus-Based Study—	22
安間 一雄	
高校英語教科書に見るテキストタイプの相違	28
MANABE Daisuke	
Exploring the Effectiveness of Data-driven Learning and Its Integration with Generative AI	37
宮崎 佳典／田中 省作	
<i>t</i> スコアと相互情報量における頻度効果の代数的考察	44
田中 悠介／瀬戸口 彩花／近 大志／神澤 克徳	
The KIT Speaking Test Corpus 専用検索インターフェースの開発	50
望月 圭子	
学習者コーパスに基づく英語対話能力の縦断的研究	57

英語リスニングテキストにおける 難易度決定要因の分析と自動推定の可能性

川越 涼太(九州大学 大学院生)
kawagoe.ryota.165@s.kyushu-u.ac.jp

An Analysis of Factors Determining the Difficulty of English Listening Texts and the Possibility of Automatic Estimation

KAWAGOE Ryota (Kyushu University, Graduate Student)

Abstract

This study aims to identify the factors that determine the difficulty of English listening texts and to explore the possibility of their automatic estimation. Listening texts from thirty-four past administrations of the Eiken Test were analyzed using indices from Coh-Metrix 3.0, with features extracted separately for dialogues and monologues. A random forest model was applied, treating these features as explanatory variables and Eiken grades as the target variable. Variable importance was evaluated using Gini Importance and SHAP values. The analysis identified the number of words per sentence (DESSL) as the most influential factor in dialogues, and lexical diversity (VOCD) in monologues. These findings suggest that processing load associated with long texts in dialogues and lexical recognition load in monologues are central to text difficulty.

Keywords

リスニング, 難易度推定, ランダムフォレスト, SHAP, 変数重要度

1. はじめに

リスニングの難易度に関する研究はこれまでに多数行われてきたが、リスニングに特化した妥当性の高い難易度推定ツールは現時点では存在していない。英語リスニングテキストを対象とした既存の難易度推定ツールとしては、Text Inspector (Bax, 2012)がある。しかしながら、無料版では1回につき250語までという使用制限があり、CEFRレベルの判定機能は有料版でのみ利用可能である。さらに、CEFRレベルの判定に用いられているアルゴリズムが非公開であるため、学術的な検証を十分に行うことができないという課題が指摘されている(内田・根岸, 2021)。同様に、CEFR-based Vocabulary Level Analyzer ver. 2.0 (Uchida & Negishi, 2018)は、テキストの難易度をCEFR-Jレベルで詳細に推定可能である点で有用であるが、ダイアログ形式のテキストには対応しておらず、妥当性の検証が今後の課題とされている。英語リスニング教材を対象とした難易度推定ツールとしては、Lexile Audio Analyzer (Koons & Van Moere, 2020)が挙げられる。

このツールは、語彙・文構造・音韻・音声的特徴など 100 を超える変数を解析し、最終的に選定された 13 の主要変数に基づいてリスニング教材の難易度を推定する。しかし、一般には公開されておらず個人での利用はできないことに加え、Lexile®値の算出方法も非公開であるため、Text Inspector と同様に学術的な検証が困難であるという課題がある。

これらの状況を踏まえ、本研究では、英語リスニングテキストにおける難易度決定要因を明らかにし、自動推定の可能性を探ることを目的とする。

2. 先行研究

リスニングの難易度には、様々な言語的および認知的要因が関与している。根岸(2008)は、英語リスニングにおける恒常的な難易度決定要因として、「言語的難易度」、「トピックの馴染みの度合い」、「理解の詳細さ」、「議論の複雑さ」、「発話速度」などを挙げている。また、Révész and Brunfaut (2013)は、語彙的特徴がリスニング教材の難易度に与える影響を詳細に分析し、(1) K1 機能語の出現率、(2) 学術語の出現率、(3) 語彙の多様性、(4) 語彙密度、(5) 因果関係を示す内容の 5 つの特徴が難易度の有意な予測因子であると報告している。さらに、高島(2010)は、英語リスニング問題において使用語彙の特性を調整することにより、問題の難易度を操作可能であることを示している。

英語リスニング教材の難易度推定に関する研究も進められている。根岸(2009)は、British Council が発行する英語教科書 *Listening* (Horváth & Pižorn, 2005) に収録された音声テキストを分析し、CEFR レベルの推定可能性を検討した。この研究では、語彙・文法・文構造といった言語的難易度の指標に加え、発話速度や音声の長さといった音声的特徴も分析に含められた。分析の結果、これらの客観的指標のみでは、CEFR レベルを正確に判定することは困難であることが明らかとなった。この課題に対処するため、根岸(2010)は、同じ教材を用いて「複雑さ」および「専門性」といった主観的評価指標を追加し、再度 CEFR レベルの推定を試みた。その結果、客観的な言語情報に加えて主観的な判断要素を取り入れることにより、リスニング教材の難易度を一定の精度で予測できることが示された。ただし、使用したデータは現実の生活から録った本物のテキストであり、異なるレベルの要素が混在している可能性が高いことが指摘されている(根岸, 2010)。そのため、試験問題のような精密に設計されたリスニング教材を対象とする場合には、客観的な指標のみでも難易度を推定できる可能性があると考えられる。

3. リサーチデザイン

3.1 研究目的と研究設問

本研究は、英語リスニングテキストにおける難易度決定要因を明らかにし、自動推定の可能性を探ることを目的としている。リスニングテキストは、話者構成や発話形式の違いにより、「ダイアログ(複数話者による会話形式)」と「モノローグ(単一話者による独話形式)」の 2 種類に分類される。そこで、以下のリサーチクエスチョン(RQ)を設定した。

RQ1: 英語リスニングテキストの言語的特徴から、教材の難易度をどの程度の精度で推定できるか。

RQ2: 英語リスニングテキストの難易度決定要因は、ダイアログとモノログでどのように異なるか。

3.2 データ

本研究では、実用英語技能検定(以下、英検)の難易度(CEFR レベル)の異なる 5 つの受験級(1 級, 準 1 級, 2 級, 準 2 級, 3 級)を分析対象とした。なお, 4 級・5 級は CEFR レベルが A1 未満であるため, 分析対象から除外した。また, 1 級の Part 4(インタビュー)は, 他のテキストとは種類が異なるため, 同様に除外した。データは 2013 年度第 3 回から 2024 年度第 3 回までの全 34 回である。分析対象は, 指示文や説明文を除いた問題に関連するスクリプトであり, ダイアログとモノログに分けて収集した。使用したデータの概要を表 1 に示す。

表 1 データの概要

	英検	<i>n</i>	総語数	平均語数	標準偏差
ダイアログ	1 級	340	39,712	116.80	47.68
	準 1 級	408	35,566	87.17	10.56
	2 級	510	34,572	67.79	8.54
	準 2 級	680	34,942	51.39	8.56
	3 級	680	20,527	30.19	4.78
総計		2618	165,319		
モノログ	1 級	340	61,269	180.20	37.65
	準 1 級	374	50,234	134.32	35.73
	2 級	510	36,591	71.75	6.29
	準 2 級	340	21,644	63.66	5.80
	3 級	340	12,922	38.01	3.77
総計		1904	182,660		

注: *n* = テキスト数

3.3 手法

3.3.1 テキスト分析に用いる指標の選定

本研究では, Coh-Metrix 3.0 (Graesser et al., 2011)の指標を用い, 各特徴量を抽出した。Coh-Metrix は, メンフィス大学の研究チームが開発したテキストの言語的特徴をコーパスに基づき分析できるツールであり, 言語的特徴を多角的かつ詳細に分析することが可能である。Coh-Metrix 3.0 では言語のさまざまな側面に関連する 108 の言語指標を数値化することができるが, 英検は語数やセンテンス数を厳格に統制しているため, パラグラフの数に関する特徴量 3 件, 語数・センテンス数に関する特徴量 2 件を除いた計 103 件の特徴量を抽出し, 分析に使用した。

3.3.2 分析の手順

分析には機械学習手法のランダムフォレスト (Breiman, 2001)を用いた。ランダムフォレスト(以下, RF)は, 分類や予測のルールに応じて条件分岐をしていくツリーの構造で, モデル化を行う決定木をいくつも組み合わせて多数決をとることによって精度の高い分類や予測ができる(小林, 2013)。本研究では, 英検の各受験級をそれぞれ 1~5 に置き換え(1 級=5, 準 1 級=4, 2 級=3, 準 2 級=2, 3 級=1), 各特微量を説明変数, 受験級を目的変数として RF を行い, 分類モデルを構築した。次に, RF により得られたモデルの分類精度を評価するために, 5 分割交差検証(5-fold cross-validation)を実施した。全データを重複なくランダムに 5 つのデータセットに分割し, そのうち 4 つを訓練データ, 残りの 1 つを検証データとしてモデルの性能を評価した。モデルの評価指標には, Accuracy に加え, クラスごとのデータ数の偏りに影響されにくい Macro-F1 スコアを採用した。

最後に, RF により構築した分類モデルから変数重要度を算出した。変数重要度の分析には, RF に組み込まれている指標である Gini Importance と, 各変数の予測への寄与を理論的に定量化する SHAP (SHapley Additive exPlanations)の 2 つの手法を用いた。Gini Importance は高速に計算可能である一方, 分割基準のバイアスをそのまま継承してしまうため, 変数の尺度やカテゴリ数が異なる状況では信頼性に欠けるという問題点が指摘されている (Strobl et al., 2007)。この問題に対する解決手段としては, SHAP (Lundberg & Lee, 2017)が挙げられる。SHAP は機械学習の予測を解釈するための枠組みであり, Lundberg and Lee (2017)は SHAP 値だけが局所的精度・欠損性・一貫性の 3 条件すべてを満たすと説明している。RF の実装および Gini Importance の算出には Python (ver. 3.11.13)の機械学習ライブラリ scikit-learn (ver. 1.6.1)を, SHAP 値の算出には SHAP ライブラリ (ver. 0.48.0)をそれぞれ使用した。

4. 結果と考察

4.1 RQ1 英語リスニングテキストの言語的特徴に基づく難易度推定

4.1.1 5 分割交差検証

英検の各受験級(5 クラス分類)における 5 分割交差検証の結果を表 2 に示す。5 分割交差検証の結果, ダイアログとモノログのいずれにおいても, 約 75%の精度が得られた。Accuracy と Macro-F1 の数値もほぼ同様であり, 今回使用した 103 の特微量から高い精度での分類が可能であることが示された。特に, Accuracy と Macro-F1 の双方において類似した傾向がみられることから, モデルは全体の分類精度のみならず, クラス間のバランスもある程度確保できていると考えられる。また, 両者における類似した精度結果は, 言語的特徴がテキストタイプに依存せず, 汎用的に難易度推定に寄与している可能性を示している。

英検の各受験級(2 クラス分類)における 5 分割交差検証の平均値と標準偏差を表 3 に示す。全体として, 5 クラス分類よりも高い分類精度が得られた。具体的には, ダイアログでは「準 1 級と 2 級」の分類で約 87%, 「準 2 級と 3 級」で約 92%の精度を示した。一方, モノログではそれぞれ約 97%および約 93%と, さらに高い分類精度が得られている。ただし, モノログにおける「2 級と

準 2 級」の分類では精度が約 70%とやや低く、このペアに関しては他の受験級と比較して分類が困難であると考えられる。

表 2 および表 3 に示した結果から、テキストの言語的特徴のみに着目した場合であっても、英検の受験級の難易度を一定の精度で推定可能であることが示唆された。

表 2 5 分割交差検証の結果 (5 クラス分類)

	指標	Fold1	Fold2	Fold3	Fold4	Fold5	平均	標準偏差
ダイアログ	Accuracy	0.761	0.796	0.756	0.725	0.769	0.761	0.026
	Macro-F1	0.751	0.782	0.748	0.706	0.750	0.747	0.027
モノローグ	Accuracy	0.785	0.759	0.698	0.772	0.755	0.754	0.033
	Macro-F1	0.764	0.749	0.667	0.757	0.732	0.734	0.039

表 3 5 分割交差検証の平均結果 (2 クラス分類)

	指標	1 級&準 1 級	準 1 級&2 級	2 級&準 2 級	準 2 級&3 級
ダイアログ	Accuracy	0.794(0.036)	0.871(0.033)	0.805(0.028)	0.921(0.036)
	Macro-F1	0.789(0.038)	0.868(0.034)	0.798(0.031)	0.921(0.036)
モノローグ	Accuracy	0.819(0.060)	0.974(0.016)	0.716(0.029)	0.943(0.028)
	Macro-F1	0.818(0.061)	0.973(0.017)	0.679(0.032)	0.943(0.028)

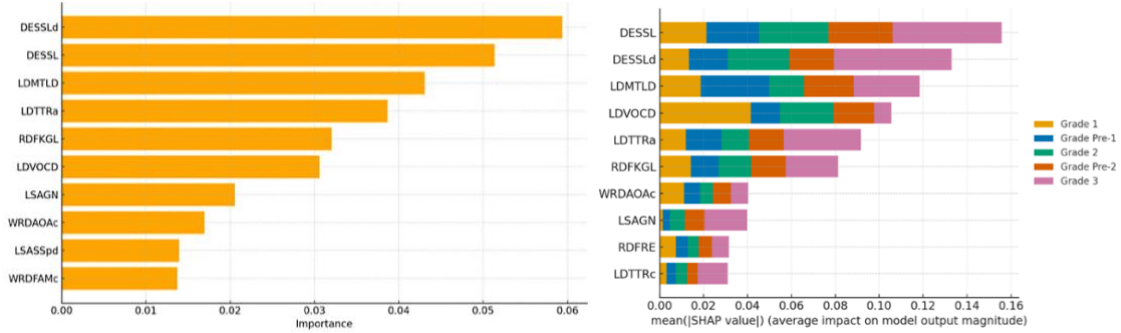
注:カッコ内の数字は標準偏差を示している。

4.2 RQ2 英語リスニングテキストの難易度決定要因

4.2.1 ダイアログ

ダイアログにおける Gini Importance と SHAP 値の上位 10 件の結果を可視化したものが図 1 である。両者の重要度を比較した結果、上位 10 件中 8 件が一致しており、重要度の順位も概ね一致していることが確認された。これは、テキストの言語的特徴を正確に捉えられていることを示している。特に、1 文あたりの単語数の平均を表す DESSL および標準偏差を表す DESSLd が最も高い重要度を示しているほか、MTLD, TTR, VOCD (語彙の多様性)、および FKGL (文章の難易度) も高い重要度であることが明らかとなった。さらに、各受験級における SHAP 値の比較では、1 級では VOCD, 準 1 級では MTLD, 2 級・準 2 級・3 級では DESSL(d) が最も重要な特徴量として挙げられた。このことから、上位の受験級においては語彙の多様性が難易度決定の主要因であり、下位の受験級では 1 文あたりの単語数が難易度を左右する要因である可能性が示唆された。

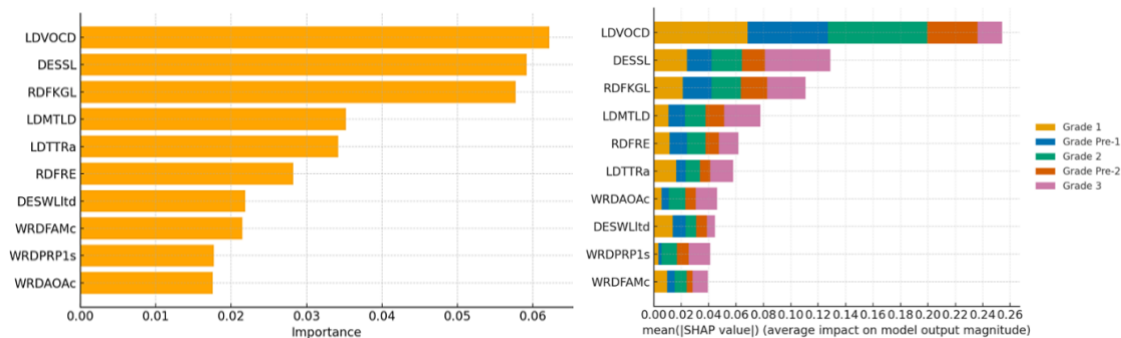
図1 ダイアログの変数重要度(上位10件)



4.2.2 モノログ

モノログにおける Gini Importance と SHAP 値の上位 10 件の結果を可視化したものが図 2 である。両者の重要度を比較した結果、上位 10 件の全ての特徴量が一致し、重要度の順位も概ね一致していることが確認された。これは、テキストの言語的特徴を正確に捉えられていることを示している。特に、語彙の多様性を表す VOCD が最も高い重要度を示しており、DESSL(1 文あたりの単語数)や FKGL(文章の難易度)も重要な特徴量として挙げられた。さらに、各受験級における SHAP 値の比較では、1 級から準 2 級までの各受験級で VOCD が最も重要な特徴量である一方、3 級においては DESSL が最も重要であった。ただし、VOCD は短いテキストでは測定することができず、2 級よりも下位の受験級(および一部の準 1 級)においては、数値がすべて 0 となっていた。SHAP 値は欠損値(0)も含めて重要度を算出するため、これらの影響により VOCD の重要度が過大に評価された可能性がある。

図2 モノログの変数重要度(上位10件)



4.2.3 ダイアログとモノログの比較

表 4 ダイアログとモノログの変数重要度の比較(上位 10 件)

ダイアログ		モノログ	
特徴量	SHAP 値	特徴量	SHAP 値
DESSL (1 文の単語数)	0.156	LDVOCD (語彙の多様性)	0.254
DESSLd (1 文の単語数)	0.133	DESSL (1 文の単語数)	0.129
LDMTLD (語彙の多様性)	0.118	RDFKGL (文章の難易度)	0.111
LDVOCD (語彙の多様性)	0.105	LDMTLD (語彙の多様性)	0.078
LDTTRa (語彙の多様性)	0.092	RDFRE (文章の難易度)	0.062
RDFKGL (文章の難易度)	0.081	LDTTRa (語彙の多様性)	0.059
WRDAOAc (内容語の習得年齢)	0.040	WRDAOAc (内容語の習得年齢)	0.046
LSAGN (既出情報の重なり)	0.040	DESWLltd (1 単語の文字数)	0.045
RDFRE (文章の難易度)	0.031	WRDPRP1s (一人称単数代名詞の頻度)	0.041
LDTTRc (語彙の多様性)	0.031	WRDFAMc (内容語の親密度)	0.039

ダイアログとモノログにおける上位 10 件の SHAP 値を表 4 に示す。両者の重要度を比較した結果、ダイアログでは 1 文あたりの単語数 (DESSL) や語彙の多様性 (MTLD, VOCD, TTR) を表す特徴量が上位を占めているのに対し、モノログでは語彙の多様性 (VOCD) や文章の難易度 (FKGL) を表す特徴量が上位に挙げられるなど、テキストタイプによって重要な特徴量に違いがあることが明らかになった。また、ダイアログにおいては、既出情報の重なりを表す LSAGN、モノログにおいては 1 単語あたりの文字数の標準偏差を表す DESWLltd や一人称単数代名詞の頻度を表す WRDPRP1s、内容語の親密度を表す WRDFAMc という異なる特徴量も確認された。これらの結果より、ダイアログでは長文処理負荷、モノログでは語彙認知負荷が難易度に直結している可能性が示唆された。

5. まとめ

本研究は、英語リスニングテキストにおける難易度決定要因を明らかにし、自動推定の可能性を探ることを目的として分析を行った。Coh-Metrix 3.0 を用いて抽出した 103 の特徴量に基づき RF による分類を行った結果、ダイアログとモノログともに約 75% の分類精度が得られ、テキストの言語的特徴から英検の受験級の難易度を一定の精度で推定できることが明らかになった。Gini Importance および SHAP 値による可視化の結果、ダイアログでは 1 文あたりの単語数 (DESSL)、モノログでは語彙の多様性 (VOCD) が特に重要な変数として検出された。また、ダイアログとモノログの変数重要度を比較した結果、テキストタイプによって重要な特徴量に違いがあることが明らかになり、ダイアログでは長文処理負荷、モノログでは語彙認知負荷がそれぞれ難易度に

強く影響している可能性が示唆された。

最後に、本研究の課題として、今後取り組むべき点を 4 つ示す。第一に、英検以外のリスニング教材への適用である。本研究では英検のリスニングテキストを対象としたが、今後は他のリスニング教材にも対象を広げ、得られた結果との比較を通じて、分析手法の汎用性や有効性を検証していく必要がある。第二に、音声的特徴を考慮した分析である。これまでの分析では、テキストの言語的特徴に関する指標のみを用いてきたが、今後は発話速度や音声の長さといった音声的特徴を加味することで、リスニング教材の難易度をより多角的かつ精緻に推定することを目指す。第三に、回帰モデルの構築とその妥当性の検証である。本研究において重要度が高いと判断された指標を用いて回帰モデルを構築し、リスニング教材の難易度推定を試みる。第四に、リーダビリティ指標との比較分析である。ARI, Gunning Fog, SMOG などの既存のリーダビリティ指標と本研究で用いた指標とを比較することで、リスニング教材に最も適した難易度評価指標の特定を目指す。

謝辞

本研究は JSPS 科研費 JP23K21949 の助成を受けたものです。

引用文献

- Bax, S. (2012). Text Inspector: Online text analysis tool. Available at: <https://textinspector.com/>.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223-234.
- Horváth, K. F., and Pižorn, K. (2005). *Listening*. British Council.
- 小林雄一郎(2013) 「機械学習と自然言語処理の技術を用いた習熟度推定—現状と課題—」 『メソドロジー研究部会 2013 年度報告論集』, 4, 12-23.
- Koons, H. H., & Van Moere, A. (2020). The Lexile Framework for Listening. Electronic publication (<https://metametricsinc.com/wp-content/uploads/2020/01/Lexile-Framework-for-Listening-White-Paper-1.pdf>; accessed June 30, 2025).
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.) *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- 根岸雅史(2008) 「オーセンティック・リスニング・テキストの CEFR リスニングのレベル判断における諸問題」 『ARCLE REVIEW』, 3, 100-109.
- 根岸雅史(2009) 「自然言語コーパスに基づく学習教材作成のための基礎的研究：英語リスニング・テキストの CEFR レベルの決定要因とそれに基づくレベル推定の可能性」 『コーパスに基づく言語学教育研究報告 3 フィールド調査, 言語コーパス, 言語情報学』,

3, 195-210.

根岸雅史(2010) 「CEFR に基づくリスニング・レベルの基準特性を探る」 『ARCLE REVIEW』, 4, 80-89.

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, 35(1), 31-65.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(25), 1-21.

高島裕臣(2010) 「TOEIC 形式リスニング問題の難易度決定要因を探る：題材文構成語彙特性の観点から」 『中国地区英語教育学会研究紀要』, 40, 21-29.

Uchida, S., & Negishi, M. (2018). Assigning CEFR-J levels to English texts based on textual features. In Y. Tono & H. Isahara (Eds.) *Proceedings of the 4th Asia Pacific Corpus Linguistics Conference*, 4, 463-467.

内田諭・根岸雅史(2021) 「英語読解教材の CEFR レベルの推定：CVLA の妥当性評価」 *Journal of Corpus-based Lexicology Studies*, 3, 1-14.

単純形副詞のタイプと容認性について

田畑 圭介(新潟県立大学)

tabata@unii.ac.jp

山縣 節子(京都外国語大学)

s_yamaga@kufs.ac.jp

Types and Acceptability of Flat Adverbs

TABATA Keisuke (University of Niigata Prefecture)

YAMAGATA Setsuko (Kyoto University of Foreign Studies)

Abstract

The flat adverb “deep” and the -ly adverb “deeply” are observed to waver in their acceptability, as in “He dug the hole deep/*deeply” and “The ground was deeply covered with snow.” This study will illustrate based on the Movie Corpus (works since 2000) that the fluctuation of acceptability can be explained through the feature of <+overlap in meaning>, which is merged into Nevalainen’s (1997) four categories of flat adverbs. We conclude that blurring of the judgments occur in the type of <+overlap in meaning>, which shows a high degree of semantic and pragmatic overlap with -ly adverbs.

Keywords

単純形副詞, The Movie Corpus, 容認性

1. はじめに

Biber et al. (2002, pp. 568-9)は、(1)が単純形副詞のすべてではないと述べつつ、単純形副詞として生起できる語を例示している。

(1) clean clear dear deep direct fine first flat free full high last light
loud low mighty plain right scare sharp slow sure tight wrong

Peters(2004, p.591)は単純形副詞のリストとして、次のものを示し、コンテキストや慣用化の影響を受け、単純形副詞が選択されるとしている。Biber et al.が挙げていなかったものを太字で表示する。

(2) **bad** **cheap** clean clear **close** deep direct **easy** **fair** flat
high loud **quick** right sharp **short** slow tight **wide** wrong

Peters(2004, p. 591)は、(2)のうち一般的な単純形副詞として *cheap, close, flat, high, right, wrong* を選定している。これらは *come close, going cheap, fall flat, fly high* のように慣用的な言い回しの中で用いられる。また *come quick* のように単純形副詞を用いると、通例、*-ly* 副詞を用いたときよりも、口語的に響くと述べている。Peters は(1)に見られる *low* を(2)で挙げていないが、Peters (2004, p.330)は *lowly* が通常 *humble* を意味する形容詞として解釈される点を指摘し、*-ly* 副詞が存在しないものは単純形副詞から除外している。

Eastwood(1994)の場合は、*cheap(ly), loud(ly), quick(ly), slow(ly)*が単純形副詞として最も一般的で、続いて *direct(ly), tight(ly), fair(ly)*がそれに続く高頻度のものだと述べている。Eastwood も、単純形副詞が日常的な言い回しの中で用いられる点について指摘し、比較的長い表現やそれほどの頻度では用いられない表現では *-ly* 副詞が選択されるとしている。

- (3) a. talk so loud d. play fair
 b. go slow e. Do you have to rustle that newspaper so loudly?
 c. fly direct f. We need to take action quickly. Eastwood (1994, p. 263)

久野・高見(2015, p. 2, p. 14)は、「*tight, slow, quick* などの副詞は、*tightly, slowly, quickly* の *-ly* 副詞より口語的で、会話や表示、広告などで多く用いられ、後者は文語的で、書き言葉で多く用いられる傾向があります」と説明し、「高さ」「広さ」「近さ」「深さ」を表す 4 組の副詞、*high/highly, wide/widely, close/closely, deep/deeply* を考察し、「形容詞形の副詞は、物理的な文字通りの意味を表すのに対し、*-ly* 副詞は、より抽象的で程度の高さや大きさ、深さを表す」とまとめている。久野・高見はこれらの形容詞形の副詞と *-ly* 副詞が相互に置き換えることができないとも述べており、文字通りの意味と比喩的な意味の相違を明らかにしている。

- (4) a. The stone sank deep / *deeply into the pond.
 b. Henry David Thoreau lived deep/*deeply in the woods.
 c. He dug the hole deep/*deeply.
 d. We love our children *deep/ deeply.
 e. His words wounded her feelings *deep/deeply.

しかしながら、ジーニアス英和辞典 6 版では、物理的な深さを示す *deeply* の記述が確認される。

- (5) The ground was deeply covered with snow.

本稿では、Nevalainen(1997)が示した(6)の4つの分類をもとに、単純形副詞を精緻化し、対応する *-ly* 副詞との意味・用法上の重なり <overlap in meaning>の指標を追加することで、(4a-c)と(5)といった容認性の揺れが導かれることを論じる。検証にあたっては、口語表現を集積している *The Movie Corpus* の 2000 年以降の作品に基づき考察を行い、*-ly* 副詞と意味・用法上のオーバーラップの度合いが高い、<+overlap in meaning>のタイプが容認性の揺れを生じさせると結論づける。

- (6) a. functional and semantic differentiation (HARD versus HARDLY)
- b. semantic differentiation within the same adverb category (LATE versus LATELY)
- c. vague, multiple conditioning of choice (CLOSE versus CLOSELY)
- d. stylistic and register differentiation (SLOW versus SLOWLY)

2. -ly 副詞と交換可能な単純形副詞: bad と loud

Longman Dictionary of Contemporary English (LDOCE) (<https://www.ldoceonline.com/>)では、(7)のように、副詞 bad は badly と同意で、a word used to mean ‘badly’ that many people think is incorrect と記述している。Oxford Advanced Learner's Dictionary (OALD) (<https://www.oxfordlearnersdictionaries.com/>)と Oxford Advanced American Dictionary (OAAD) (https://www.oxfordlearnersdictionaries.com/definition/american_english/)では、くだけたアメリカ英語の用法で、非標準だと読み取れる。

(7) LDOCE

bad³ *adverb* spoken

a word used to mean ‘badly’ that many people think is incorrect

I need that money bad.

Examples from the Corpus

bad

- He needed a drink pretty bad.
- The front of the shop had been blown away, and the roof was badly damaged.
- Like I hate it and I do bad at it.
- How bad do you want it?

(8) OALD

bad *adverb*

/bæd/

/bæd/

★ (North American English, informal) badly

- She wanted it real bad.
- Are you hurt bad?

(9) OAAD

Definition of **bad** *adverb* from the Oxford Advanced American Dictionary

bad *adverb*

/bæd/

(informal) (non-standard)

badly

- She wanted it real bad.
- Are you hurt bad?

単純形副詞の現代用法に着目するため、本稿では The Movie Corpus を用いて 2000 年以降の映画を調査対象とする。bad の副詞用法を抽出するために、動詞+代名詞+bad の形式で検索すると、3 例以上で bad が副詞として扱われるものは次のものとなる。

表1 2000 年以降の映画作品に見られる 3 例以上の VERB+ PRON bad

順位	VERB+ PRON bad	頻度	順位	VERB+ PRON bad	頻度
1	GOT IT BAD	51	10	WANT HIM BAD	5
2	WANT IT BAD	25	11	WANT YOU BAD	5
3	HAD IT BAD	14	12	WANTS IT BAD	4
4	HAVE IT BAD	9	13	WANTS YOU BAD	4
5	HURT HIM BAD	8	14	BEAT YOU BAD	3
6	TREAT YOU BAD	8	15	NEED IT BAD	3
7	WANTED IT BAD	7	16	TREAT ME BAD	3
8	HURT ME BAD	6	17	TREATS YOU BAD	3
9	HURT YOU BAD	5	18	WANT ME BAD	3

副詞 bad は非標準用法とされているなか、The Movie Corpus では上記のように、その使用が観察される。Eastwood は、単純形副詞の特徴として、表現の日常性を指摘しているが、表 1 の表現はこの特徴を満たすものとなっている。[get] it bad (54 例/badly 0 例) と[have] it bad (26 例/badly 0 例)は be in a bad situation or position と解釈される表現で、The Movie Corpus では [get] it badly および[have] it badly が 0 例となることから、got it bad や had it bad が定型表現として定着していることがわかる。動詞が want になると、[want] it bad (37 例/badly 9 例)、

[want] you bad (9 例/badly 3 例) , [want] him bad (8 例/badly 0 例)となり、最後の例を除き、badly の使用が確認できる。その他、目に付く動詞としては treat PRON bad (24 例/ badly 27 例), hurt PRON bad (27 例/ badly 10 例), need PRON bad (8 例/ badly 11 例), beat PRON bad (5 例/ badly 2 例)が挙げられる。need については、The Movie Corpus では、[need] a drink/bath/toilet bad のように代名詞以外の例も見られ、[need] ~ bad の生産性が確認できる。上記以外では、(7)-(9)の用例に示されている、pretty/real bad の使用が見られ、[hurt] real bad が 53 例、[hurt] pretty bad が 38 例観察され、be cut up real bad のような、be (...) up real bad が 39 例検出され、こちらも bad の慣用的表現となる。単純形副詞 bad は慣用表現を除き、badly と交換可能となることから、Nevalainen(1997)の分類では(6d) stylistic and register differentiation として扱われると帰結される。

副詞 loud は talk loud, sing loud, speak loud のように「大声で話す・歌う」表現を作り出す。loudly と交換可能で、他の語との共起性については loudly と同等となる。ウィズダム英和辞典第 4 版の副詞 loud の項目では、「loud enough, as loud as, too [very, so] loud の形でよく用いる」とあるが、The Movie Corpus ではいずれの用法も loudly の形態でも確認できる。「loud が文末にない場合は loudly を用いる」とする英和辞典も見られるが、The Movie Corpus では、文中での VERB loud enough to の使用が 5 例、VERB so loud for ...の使用が 3 例観察され、単純形副詞 loud は文末に限られるわけではない。loud も Nevalainen(1997)の分類では、-ly 副詞と交換可能となる(8d) stylistic and register differentiation のタイプとなる。

3. -ly 副詞と意味の棲み分けを持つ単純形副詞: clean と clear

副詞 clean の語義は小西(1989, pp. 398-9)では(10)のように解説されている。

(10) a. S V (O) clean M A<人[物]>がすっかり[完全に, 全<]>(Oを)Vする

<<M は方向や経路を表す副詞(句)>>

b. S clean V S<人>が完全に[すっかり](Oを)Vする<Vはおもに forget>

c. S V (O) clean S<人>が(Oを)正々堂々と[フェアに, あざやかに, 見事に]Vする

The Movie Corpus で VERB+ clean を検索し、上位 20 位の中で、clean が副詞のものは、go clean, get clean, come clean となる。go clean は、Why don't you go clean up?のように clean は動詞として機能しているものと、He said the bullet went clean through.のような単純形副詞のものが見られる。(10a)にあるように、直後に方向・経路を示す副詞が後続するときに clean が単純形副詞であると判別できる。get clean away「まんまと逃げきる」と come clean 「白状する」のように VERB+ 副詞 clean は定型的な表現であることが確認できる。副詞 clean と cleanly の交替は、基本的に「あざやかに/見事に...する」を意味するときのみとなるが、The Movie Corpus の例を見ると、There's a way we can do this cleanly. / We need to take him cleanly. / I want this to go quickly and cleanly.のように、cleanly が選択される環境は文が比較的長いとき、あるいはそれほど頻度では用いられない表現のときだと推察される。Nevalainen(1997)の分類では(6c)

vague, multiple conditioning of choice に分類され、副詞 *clean* と *cleanly* の共起語の違いは、副詞 *clean* の存在価値につながっている。

単純形副詞 *clean* の語義は小西(1989, pp. 407-8)では(11)のように示されている。

(11) a. S V (O) *clean* はっきりと V する

b. S V *clean* M S<人[物]>がすっかり[完全に]M へ[まで]V する

<<M は方向・経路などを表す副詞句>>

c. S V *clean* (of O) S<人など>が O(場所[人])>から離れて[に触れないで]V する

(11a)は *clearly* と意味が重なり交換可能であるが、*very* など程度を表す副詞が直前にくるときは *clearly* に限られる(ジーニアス英和辞典第 6 版)。小西(1989, p. 407)は「*Speak clear.*はくだけた口語的(あるいはときに非標準的)な響きを伴う」と述べ、*The Movie Corpus* では *Aren't I speaking clear enough?*が1例確認される。(11b,c)については、*get clear away*(逃げ切る)、*loud and clear*(はっきりと)、*Stand clear of the closing doors, please.*といった慣用的表現で単純形副詞 *clear* が使用されている。*clear* の使用にはある程度慣用性が見られ、(11b,c)では *clearly* とは交換可能でないと帰結される。*clear* および *clean* は Nevalainen(1997)の分類では(6c)のタイプとなるが、*clearly*、*cleanly* との意味・用法の重なりはごく一部となる。

4. 物理的な近さ、高さ、広さ、深さを表す単純形副詞: *close*, *high*, *wide*, *deep*

単純形副詞 *close*, *high*, *wide*, *deep* は、物理的な近さ、高さ、広さ、深さを表し、*closely*, *highly*, *widely*, *deeply* はより抽象的で、程度の密接さ、高さ、広さ、深さを表すと久野・高見(2015)が指摘している。*high* は「(物理的に)高く」、*highly* は「(比喩的に)非常に、大いに」を意味し、互いに交換可能ではない。*wide* についても「(物理的に)広く」の意味で、強意的に *apart*, *awake*, *open* などの語を修飾するのに対し、*widely* は抽象的に「(広範囲にわたって)広く」の意味を表す。*high-highly*, *wide-widely* は意味的に異なり、互いに交換可能でないことから、Nevalainen(1997)の分類では、*late-lately* と同じ(6b)のタイプと扱うべきである。

close は「(場所的に)近く」、*closely* は「注意深く」「念入りに」と抽象的な意味を表し、基本的に交換不可である。ただ、ジーニアス英和辞典 6 版が示すように、(距離・時間が)近くに[で]、すぐ: *I saw him riding his bike, followed closely by his dog.*の語義も示されており、ごく一部は交換可能となる。Nevalainen(1997)は *close-closely* を(6c)としているが、属性としては、(6b)と(6c)の間に位置するものと考えられる。

deep については(4a-c)のように物理的な深さを意味するのに対し、*deeply* が(4d,e)のように程度の深さ・大きさを意味することになるが、(5)のように *deeply* が物理的に深い意味を表す例も存在する。対照的に *deep* が程度の深さを意味する、*As you stand here in the darkness, look deep into your heart, into your soul, and ask yourself.*や程度の大きさを意味する *Let us drink deep of dod til morning.* (深酒をする)といった例が *The Movie Corpus* で確認でき、単純形副詞の *deep* と *deeply* が表す意味が重なりあう事例が見受けられる。その他、

The Movie Corpus では、dig deep が 166 例見つかり、物理的な例とともに、You've got to dig deep in your history.のような比喩的な用法も観察される。dig deeply については、3 例見つかり、いずれも「掘り下げる」の比喩的な意味で使用されているが、The shell doesn't flex, so we know that this turtle didn't dive deeply in the ocean.のような物理的距離の例や I always wanted to create a tablet that tells the seasons. Hm. I've always wanted to spend more time in the water. Swimming deeply where the world accepts me for all I am.のような物理的な深さと比喩的な深さが混交した例も The Movie Corpus に見られる。deep-deeply については、(6c) に分類はされるが、互いの意味のオーバーラップの度合いが大きく、(6c)と(6d)の間に位置するものと想定される。このことが、結果的に(4a-c)と(5)のような容認性の違いを生み出すことになると帰結される。(6c)については、-ly 副詞と意味的なオーバーラップの度合いが高い<+overlap in meaning>のタイプに属す単純形副詞が容認性で揺れる可能性がある結論づけられる。ただし(6d)と異なり、deep-deeply は完全に交換可能なわけではなく、deep の物理的距離の優位的特性は認められることから、会話体で物理的距離を指示するような状況では、Dig the hole deep.のように単純形副詞のほうがより自然だといえるだろう。

引用文献

- Davies, M. (2019). *The Movie Corpus*. Available online at <https://www.english-corpora.org/movies/>.
- Eastwood, J. (1994). *Oxford Guide to English Grammar*. Oxford University Press.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.
- 井上永幸・赤野 一郎 (編) (2018) 『ウィズダム英和辞典第 4 版』三省堂.
- 小西友七 (編) (1989) 『英語基本形容詞・副詞辞典』研究社出版.
- 久野暲・高見健一 (2015) 『謎解きの英文法 副詞と数量詞』くろしお出版.
- 南出康世・中邑光男 (編) (2022) 『ジーニアス英和辞典第 6 版』大修館書店.
- Nevalainen, T. (1997). The processes of adverb derivation in Late Middle and Early Modern English. In M. Rissanen, M. Kytö, & K. Heikkonen (Eds.), *Grammaticalization at Work: Studies of long-term developments in English* (pp. 145–189). Mouton de Gruyter.
- Peters, p. (2004). *The Cambridge Guide to English Usage*. Cambridge University Press.

No point/sense/use の補文における形式的変異について

南 佑亮(神戸大学)
y-minami@people.kobe-u.ac.jp

(Formal variation in the complement of *no point/sense/use*)

MINANMI Yusuke (Kobe University)

Abstract

This study investigates formal variation in an impersonal construction where a class of modal noun phrases (*no point/sense/use*) take a complement. The results of a corpus survey using COCA will reveal the following: (i) “there + be” form is preferred over the other options when the complement is in the *in*-gerund form; (ii) it is rare for the *to*-infinitival form to be used in lieu of the gerund form; (iii) the presence of the subject in the complement clause increases the relative frequency of the *to*-infinitive variant and that of the *in*-gerundive variant. It will then be argued that the skewed distributions that were observed are best explained from the perspective of Construction Grammar and the Complexity Principle.

Keywords

モーダル名詞, 補文の変異形, 構文文法, 構文スキーマ, 複雑性の原理

1. はじめに

英語には、「it/there + be 動詞 + 名詞句 (no NOUN) + 動名詞句補文」という形式を持ち、「～しても意味がない, 無駄である」という意味を表す構文現象(以下, 非人称名詞構文)がある。

- (1) a. There is no point in getting angry. (OALD, s.v. *point*)
- b. There’s no sense in worrying about it now. (OALD, s.v. *sense*)
- c. It’s no use running—the train has already gone. (OALD, s.v. *use*)

非人称名詞構文には形式上の変異がある。(1)を見るだけでも「主語が it か there か」「補文の動名詞句に in が伴うか否か」という 2 つの変項とそれを具現する 4 種類の変異形の存在が推定できるが, 実際にはさらに多くの変異形が存在し, その全てが(1)の 3 つの名詞句において観察される。本研究の目的は, コーパス調査によって非人称名詞構文の変異形のタイプとその分布傾向を明らかにし, 認知・機能主義的な説明を試みることである。リサーチクエスションは以下の通りである。

RQ1: no point/sense/use による非人称名詞構文の変異形の分布に共通する特徴は何か。

RQ2: そのような共通の特徴をもたらす認知的・機能的な動機は何か。

2. 先行研究

2.1 no use の補文における変異—神谷・高橋・神田 (2004)

神谷・高橋・神田 (2004) は、特に(1a)の構文パターン(以下「no use 構文」)の形式的変異を BNC で網羅的に調査し、no use 構文に関する辞書記述の改善案を提出している。存在が指摘されている no use 構文の種々の変異形のリストのうち本稿との関連で特に重要なのは、補文の形式が(動名詞句ではなく)to 不定詞句の事例(=(2)), および補文に「主語」が伴う事例(=(3))である。ただし、非人称名詞構文の変異形を幅広く検証しているものの、各変異形の事例数が極めて少ないため、挙げられている変異形の網羅性およびその分布特徴は明らかにされていない。

(2) It was no use to go in there. (神谷・高橋・神田 2004: 106)

(3) a. “There is no use me staying,” she said to Marie. (同上: 110)

b. It is no use for him to try to practise “walking” (...). (同上: 106)

2.2 貝殻名詞(shell nouns)の下位分類—Schmid (2000)

Schmid (2000) は、命題レベルの複雑な概念内容を単純な名詞句にまとめるという談話上の機能を果たす用法を持つ一群の名詞を貝殻名詞(shell nouns)と名付け、その分類と機能について認知・機能主義的な立場から包括的に分析している。貝殻名詞は命題レベルの情報を担う節表現と共に使用される特徴があり、用法は 6 種類に大別されている、本稿が扱う 3 つの名詞はこのうち Modal の用法(=(4), (5))に最も近いと考えられる。ただし、Schmid はモーダル性を持つ名詞(以下、「モーダル名詞」)に伴う補文節を that 節(定形節), wh-節, to 不定詞節のみに限定し、動名詞節は「動詞の補文節では重要だが名詞の補文節では非常に周辺の」(p.31)としてその考察対象から除外しているため、本研究が注目するタイプの現象は扱われていない。

(4) Perhaps both tasks will be worthless because there is **a good chance** that at least one of his rivals will beat him. (Schmid 2000: 240)

(5) There will be **no need** to change money to visit EC countries. (Schmid 2000: 250)

3. コーパス調査(RQ1)

3.1 データの収集とタグ付けと集計

RQ1 に対する答えを得るため、COCA(Davies 2008-)を用いて調査を実施した。データの収集は、名詞句 no point, no use, no sense それぞれについて抽出したテキストデータ全てを Excel ファイルに移したうえでデータを選別し、該当するすべての事例に手動でタグ付けをおこない、データベースを作成した。

データ抽出は、神谷・高橋・神田 (2004)で存在が指摘された変異形を想定し、二段階に分けて実施した。第一段階では、no point/sense/use の補文に(i) V-ing を伴う場合、(ii) in V-ing を伴う場合、(iii) to 不定詞を伴う場合の 3 通りを想定し、該当事例を検索して抽出した。第二段階では、それぞれの補文構造に主語が表現されている事例を、主語が代名詞 1 語の場合に限定して抽出した。次に、抽出したデータのうち、対象となるデータの “no N” の直前の形式を一件ずつ順に確

認し、分類(タグ付け)を行った。以上の方法で作成したデータベースから、(A)there + be, (B)it + be, (C) ∅(なし)の3つのパターンに該当する事例を対象データとしてカウントした。

データの集計結果の確認に移る前に、データ選別の過程で本研究の対象からは除外したものについて触れておく。まず、(6)のように **there** も **it** も伴わずに **ain't** から始まるタイプの事例が少数(%程度)見られた。二重否定を伴っておりこれ自体が興味深い現象だが、本研究の調査対象からは除外した。また、(7)のように当該名詞句が他動詞の目的語に該当する事例も少なからず抽出されたが、これは「非人称」ではないため、除外した。(ただし、3.3.1の議論も参照)

(6) **Ain't no point being shy about it now!** (COCA 2001 FIC)

(7) **Honestly, Lorelai, I see no point in continuing this conversation.** (COCA 2006 TV)

3.2 データの集計結果

3.2.1 補文に主語を伴わない場合

表1に示す通り、**in** を伴う動名詞(**V-ing**)と **in** を伴わない動名詞(**in V-ing**)の相対頻度は3つの名詞句が三者三様の特徴を示している。**no point** は明らかに **in V-ing** を好むが、逆に **no use** は **V-ing** を好み、**no sense** にはそのような偏りがない。

表1 補文が動名詞の場合(主語なし)

	V-ing	in V-ing	total
no point	512	1832	2344
no sense	407	367	774
no use	696	153	849

しかし、名詞句に先行する(A)-(C)の3つの形式パターンも考慮すると、**V-ing** の時よりも **in V-ing** の時に **there + be** 形式の占める割合が高くなっていることが判明する(表2)。これは、話者が **in V-ing** を使う時に、名詞句の種類を問わず **there + be** 形式を選好する傾向を示唆している。

表2 動名詞補文の2つの変異形において名詞句に先行する非人称形式の事例数

	V-ing				in V-ing			
	there	it	∅	there (%)	there	it	∅	there (%)
no point	378	0	129	74.6	1395	3	288	82.7
no sense	160	2	232	40.6	223	2	126	63.5
no use	300	150	239	43.5	106	4	27	77.3

この傾向は特に **no sense** と **no use** において顕著である。**no sense** の場合、**it + be** はほとんど存在せず **there + be** と **∅** の二者択一に近い状態だが、**V-ing** の時に **∅** の比率が大きいのに対し、**in V-ing** の時は **there + be** が多数派となる。**no use** は、**it + be** の生起頻度が **V-ing** の場合

に極端に偏っているため、結果として補文形式全体の中で **there + be** の事例が占める割合は **in V-ing** の場合の方が大きくなっている。

表 2 から分かるように、**no sense** と **no use** ほど明確ではないものの、**no point** についても同じ傾向が見られる。なお、**no point** には、**in V-ing** の場合に **no point** には動詞 **see** を伴う事例 (3.1 節の(7)参照) が相対的に高い頻度で観察されるという独自の特徴がある (パラダイムに **see** も加えた場合に全体に占める割合 = **no point** (7.6%), **no sense** (2.0%), **no use** (3.5%))。このパターンは、形式上は「非人称」とは言えないが、**there + be** を用いた表現と意味は限りなく近い関係にある上、**in V-ing** に偏って分布する点も共通している (**V-ing** の時、**see** の事例は 0.3%のみ)。

次に、当該構文の補文形式として、動名詞の代わりに **to** 不定詞が選択されている事例の割合を表 3 に示す。**to** 不定詞の使用は非常に限定的であり、最も頻度の高い **no use** の場合でさえ全体の 7%程度に留まっていることが分かる。

表 3 非人称名詞構文の補文における動名詞と **to** 不定詞の生起比率

	(in) V-ing	to V	total
no point	2344	21 (0.9%)	2365
no sense	774	11 (1.4%)	785
no use	849	63 (6.9%)	912
total	3967	95 (2.3%)	4062

3.2.2 補文に主語が伴う場合

表 4 に示すように、補文に主語 (人称代名詞 1 語) が表現されている場合、補文の 3 つの変異形 (**V-ing**, **in V-ing**, **to** 不定詞) の分布比率は、主語を伴わない場合 (表 1, 表 3) と比較すると、次の点で異なる。第一に、主語なしの場合よりも **to** 不定詞の使用率が高くなっている。第二に、2 つの動名詞パターンにおいて **in V-ing** が選択される割合が主語なしの場合よりも増加している。

表 4 非人称名詞構文の補文に主語 (人称代名詞 1 語) が伴う事例の内訳

	①Sbj V-ing	②in Sbj V-ing	(①+②)	for Sbj to V
no point	12	141	156	7 (4.3%)
no sense	8	25	33	1 (3.0%)
no use	11	14	25	8 (24.2%)

4. 考察 (RQ2)

4.1 **there be** と **in V-ing** の組み合わせが選好される動機

3.2.1 節で提示した現象の動機について考察する。構文文法 (Construction Grammar) において、話者の言語知識は構文スキーマと構文のネットワーク関係 (Diessel 2023) から成り立っていると考える。この言語モデルでは、あらゆるレベルの複雑性と抽象度を持つ構文 (constructions)

は様々な形で結びつけられており、構文同士が様々な条件下で互いに影響を及ぼし合っている。当該現象の場合、話者の知識において当該構文の補文形式については 2 つのスキーマ [no N in V-ing] と [no N V-ing] (N = point, sense, use) がパラダイムの関係にあり、話者は様々な条件に応じて 2 つのうちのいずれかの形式を選択する。このように考えると、no N の前に there + be が先行する形式を選択する際には、there + be を伴う構文として定着している抽象度の高い構文スキーマ ([there be no NP PP(in NP)]) が想起されるため、話者がこのスキーマにより近い形式にしようとして in を伴う動名詞句を含むスキーマの方を選択している、と分析することができる。確かに、there + be を伴う非人称名詞構文における PP は in を主要部とする前置詞句といっても実質的には補文であり、pivot 位置に来る名詞句の存在する場所についての描写ではなく pivot が表すモーダルの意味と結びつけられる命題を表しているため、非人称名詞構文は意味概念・統語構造の両面において一般的な there 存在文とはかなり遠い関係にある。だが、言語使用者が表現選択において常にそのような入念な解析を行っているとは限らない。当該現象の場合、表面上の形式的類似性のみ注目し、coda 位置の句は前置詞を伴うのが自然と判断する浅い解析 (shallow parsing) が優先されている可能性が十分にある。同様に、to 不定詞補文の事例が極めて稀である理由は、意味的な区別に基づく説明 (Wierzbicka 1988) よりも、異なる変異形間の関係という観点から考察すべきである。動名詞と to 不定詞は特に動詞の補文においてはパラダイムを成すケースもあるが、非人称名詞構文の場合は、in V-ing と V-ing という形式上の差異が最小限の変異形がパラダイムの競合状態にあるため、to 不定詞がそのパラダイムに加わりにくくなっていることが推定できるからである。このように、文法形式間のパラダイム関係を考慮すると、使用頻度の程度差という、意味概念の整合性からは説明がしづらい事実を捉えることが可能となる。

4.2 補文の主語を伴う場合に to 不定詞と in V-ing の確率が上がる動機

3.2 節で見たように、非人称名詞構文の補文形式は動名詞の 2 つの変異形が主流であって to 不定詞は周辺的な地位にあるものの、補文に主語が出現する環境では主語なしの場合に比べて to 不定詞句の比率が上がり、動名詞句の 2 つの変異形の中では in V-ing の方が優勢になる。本論では、この現象が、複雑性の原理 (Rohdenburg 1996, 2006) の帰結であると主張する。複雑性の原理は「同じ事柄を表すための選択肢としてより明示的な形式とそうでない形式がある場合、認知的により複雑な環境では、より明示的な表現形式が好まれる」というものである。これを非人称名詞構文に適用すると、(i) 主語の表現されている補文は、主語のない補文よりも処理が困難であり複雑な環境であり、(ii) 節(あるいは文)相当の内容を表現する手段として動名詞よりも to 不定詞の方が明示的である (Vosberg 2009: 219) ということから、主語が表現される時に to 不定詞の使用率が向上するという予測が導かれる。同様に、動名詞の 2 つの変異形において一貫して in V-ing が好まれる現象についても、V-ing よりも in V-ing の方が明示的であると仮定することで説明がつく。すなわち問題の頻度の偏りは、in を伴った場合に in が to 不定詞句における主語標識 for と同等の役割を果たすことの帰結であると分析できるのである。

5. 結論

本研究は、no point/sense/use の補文形式に見られる形式的変異現象に着目し、変異形にどのような種類と分布傾向があるか(RQ1)、その分布傾向をもたらす動機は何か(RQ2)、という2つの問題に取り組んだ。RQ1については、(i)there be と in V-ing という組み合わせが最も好まれ、(ii)補文での to 不定詞句の使用は周辺的であり、(iii)補文に主語が表現される場合は to 不定詞句と in V-ing の選好性が上がる、という事実を明らかにした。RQ2については、構文文法における構文ネットワークおよび複雑性の原理の観点からの説明を試みた。本研究は、今なお未開拓の部分が多く残るモーダル名詞(modal nouns)の研究に先鞭をつけるものと位置付けられる。

引用文献

- Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Diessel, Holger (2023) *The construction: Taxonomies and networks*. Cambridge: Cambridge University Press.
- 神谷昌明・高橋薫・神田和幸 (2004) 「動名詞を用いた慣用表現・コロケーション It is no use~ing 再考—British National Corpus を検索して」 『中京大学教養論叢』 45(1), 93-128.
- Rohdenburg, Günter (1996) 'Cognitive complexity and increased grammatical explicitness in English', *Cognitive Linguistics* 7, 149–82.
- Rohdenburg, Günter (2006) 'The role of functional constraints in the evolution of the English complementation system.' In Dalton-Puffer, Christiane, Kastovsky, Dieter, Ritt, Nikolaus and Schendl, Herbert (eds.), *Syntax, style and grammatical norms: English from 1500–2000*, 143–66, Bern: Peter Lang.
- Schmid, Hans-Jörg (2000) *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin: Mouton de Gruyter.
- Vosberg, Uwe (2009) 'Non-finite complements.' In Rohdenburg, Günter and Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*, 212–227, Cambridge: Cambridge University Press.
- Wierzbicka, Anna (1988) *The semantics of grammar*. Amsterdam: John Benjamins.

辞書

Oxford Advanced Learner's Dictionary, 10th edition. [OALD]

Bro Constructions across America —A Twitter Corpus-Based Study—

SHIBUYA Yoshikata (Kanazawa University)

y.shibuya@staff.kanazawa-u.ac.jp

JENSEN Kim Ebensgaard (University of Copenhagen)

ebensgaard@hum.ku.dk

Abstract

This study explores how casual address terms—specifically *bro* phrases—function in American Twitter discourse. Using a geolocated corpus of tweets, the authors focus on five high-frequency expressions (e.g., *my bro*, *What’s good bro?*), analyzing their regional distribution through correspondence analysis and density heatmaps. The findings reveal systematic geographic variation: some phrases are used nationwide, while others are regionally concentrated (e.g., *What’s good bro?* in the Northeast). These expressions are treated as entrenched constructions in the sense of Usage-Based Construction Grammar (e.g., Goldberg, 2006), showing how social meaning, identity, and linguistic structure emerge from patterns of usage in digital contexts.

Keywords

bro constructions, Twitter discourse, regional variation, usage-based linguistics

1. Introduction

Address terms (e.g., *sir*, *ma’am*) are central to interaction, helping speakers identify and position interlocutors, manage relationships, and negotiate identity. Usage varies by formality: in English, formal terms like *sir* or *Ms.* convey respect, while informal ones like *bro* or *man* signal familiarity and group affiliation. Classic sociolinguistic work (e.g., Brown & Gilman, 1960; Braun, 1988) links address terms to social hierarchies, institutional norms, and politeness conventions, while pragmatic studies (e.g., Leech, 1999) highlight their discourse functions. Far less is known about address terms in digital discourse, where interaction unfolds under different social and technological conditions. Online, such terms initiate exchanges, build relationships, signal group membership, and convey stance. While studies such as Zappavigna (2012) and Page (2012) explore online affiliation and identity, they give less attention to the constructional and interactional properties of address terms. This study addresses that gap by analyzing casual address terms—especially *bro*—in Twitter discourse. Twitter’s

hybrid of spoken and written features, informal tone, and real-time environment make it ideal for observing how address terms enact stance, index group belonging, and reflect sociocultural affiliations. Focusing on recurrent constructional phenomena that we call *bro* phrases (i.e. recurring functional word strings featuring the word *bro*), the study examines their functions, geographic distribution, and regional variation in the U.S. within a cognitive sociolinguistic framework.

2. *Bro* in use: From kinship term to index of identity

According to the Oxford English Dictionary, *bro* originated in the Southern United States in 1832 as a shortened form of *brother*, first attested in Caribbean and African American Englishes. Early uses appeared as titles in folktales (e.g., *Bro Rabbi*) alongside variants such as *Brer*, *bra*, and *bredda*. Over time, *bro* shifted from a kinship term to an informal address among men, broadened to mean a male friend or peer, and more recently has evoked youth culture, male bonding, and sometimes irony or derision. Functionally, *bro* contrasts with *brother*: while *brother* typically appears in formal and kinship contexts, *bro* conveys informality, camaraderie, and group affiliation, particularly in youth speech, texting, and online discourse. Used vocatively (e.g., *Thanks for helping me out, bro*) or referentially (e.g., *He's my bro*), *bro* spans meanings from sincere to sarcastic, underscoring its pragmatic flexibility. From a Construction Grammar perspective (cf. Hilpert, 2019: 2), both *brother* and *bro* are entrenched lexical constructions with distinct forms and functions. This study focuses not on *bro* in isolation, but on *bro* phrases—recurrent expressions containing *bro* that function as constructions in their own right. Adopting a usage-based approach (e.g., Bybee, 2010), it examines their functions and geographic distribution in U.S. Twitter discourse, addressing the research question: How are *bro* phrases used, and how does their distribution vary regionally? The analysis draws on cognitive sociolinguistics (e.g., Kristiansen & Dirven, 2008) to investigate how such constructional patterns index identity and social relationships in digital communication.

3. Data and Methodology

To address the research question, this study draws on the test set of the Cheng-Caverlee-Lee Twitter Scrap (CCLTS; Cheng et al., 2010), which contains over 5 million tweets from 5,190 U.S.-based users with GPS-enabled location data. The training set was excluded due to its less reliable, self-reported locations. From the test set, all tweets containing *bro* were extracted, yielding 19,159 tweets from 2,973 users. Recurrent *bro* phrases were identified via bigram and trigram extraction, with inclusion criteria of (1)

forming a recognizable phrase-level construction (recurrent form-meaning pairing) and (2) occurring at least 150 times. Many bigrams were excluded for appearing mainly within longer trigrams (e.g., *lil bro* in *my lil bro*) or showing high internal variability (e.g., *bro I, up bro*). Five high-frequency phrases met the criteria: *my bro*, *thanks bro*, *my lil bro*, *my big bro*, and *what’s good bro*. These cover functions from expressions of camaraderie (*my bro*) and gratitude (*Thanks bro*) to kinship metaphors (*my lil bro*, *my big bro*) and informal greetings (*What’s good bro?*).

With the five target *bro* phrases identified, the next step was to examine whether their usage varies by U.S. region. Using correspondence analysis (CA)—a multivariate method for visualizing associations between categorical variables—each occurrence was linked to geolocation data and grouped into the U.S. Census Bureau’s four regions (Northeast, Midwest, South, West). Table 1 shows raw frequencies by region, with *my bro* most common overall and *what’s good bro* least frequent. While these counts suggest regional differences, CA allows both phrases and regions to be projected into the same low-dimensional space, revealing clustering patterns and cross-regional dispersion. This provides a nuanced, usage-based and cognitive sociolinguistic view of how *bro* phrases index regional variation in digital discourse.

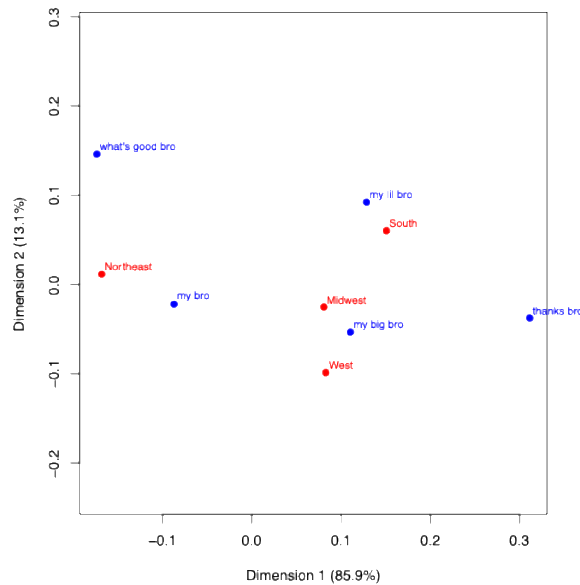
Table 1 *Frequency of bro phrases by U.S. region*

Phrase	Midwest	Northeast	South	West	Total
<i>my bro</i>	231	960	545	412	2,148
<i>my lil bro</i>	61	192	194	97	544
<i>thanks bro</i>	53	96	140	91	380
<i>my big bro</i>	29	95	84	63	271
<i>what’s good bro</i>	17	90	50	22	179
Total	391	1,433	1,013	685	3,522

4. Results

Figure 1 shows a correspondence analysis (CA) biplot of the five *bro* phrases across four U.S. regions. On Dim 1, *thanks bro*, *my lil bro*, and *my big bro* lie on the positive side, aligning with the South, West, and Midwest. *My bro* and *what’s good bro* are on the negative side, nearer to the Northeast. On Dim 2, *what’s good bro* and *my lil bro* score positively, associated with the South and Northeast, while *my big bro*, *thanks bro*, and *my bro* score negatively, closer to the West and Midwest. These results reveal clear phrase-region associations, forming the basis for the interpretation in the next section.

Figure 1 *Correspondence analysis biplot: bro phrases × U.S. regions*



5. Discussion

The CA shows clear regional patterns in the use of *bro* phrases across the United States, indicating that these expressions are socially meaningful linguistic units with varying degrees of regional entrenchment. On Dimension 1, *thanks bro* is most positive, clustering with the South but also linked to the Midwest and West; *what's good bro* is most negative, aligning with the Northeast; *my lil bro* is positive on both dimensions, near the South; *my bro* is negative on both, suggesting diffuse but Northeast-leaning use; and *my big bro* aligns with the Midwest and West.

To map concentration more precisely, normalized density heatmaps (Figure 2) were created. These reveal that *thanks bro* is the most evenly distributed; *my big bro* is broadly spread but especially strong in the South, East Coast, and California; *my bro* and *my lil bro* cluster in the Northeast; and *what's good bro* is sharply localized in New York.

Together, the CA and heatmaps show that while some phrases are geographically unmarked, others are strongly region-bound, supporting the usage-based view that linguistic knowledge is shaped by usage within specific social and geographic contexts. The findings answer the research question by showing that *bro* phrases vary in both function and regional concentration: localized forms reflect community-specific discourse norms, while widely dispersed forms follow general communicative routines. Theoretically, the results illustrate (1) in usage-based terms, how repeated community exposure fosters entrenchment; (2) in Construction Grammar (e.g., Goldberg, 2006), that

these high-frequency, coherent, and socially conventionalized expressions qualify as fully fledged *bro* constructions; and (3) in cognitive sociolinguistics, that regional and social identity is encoded in informal multi-word expressions as well as in vocabulary and pronunciation—underscoring the value of studying recurrent phraseological patterns as sites of social meaning-making.

Figure 2 *Regional density distribution of five bro phrases in the United States*



6. Conclusion

This study examined the regional distribution and constructional status of *bro* in American Twitter discourse, using a geolocated corpus and correspondence analysis. The results show that *bro* expressions are both socially meaningful and geographically patterned: forms like *thanks bro* are widely distributed and relatively unmarked, whereas expressions such as *What's good bro?* are tightly localized, indexing specific regional or cultural identities. The findings contribute to usage-based linguistics (e.g., Bybee, 2010) by demonstrating that highly informal expressions can display systematic regional variation, supporting the framework's emphasis on frequency, context, and

social interaction. They also show that multi-word expressions such as *my lil bro* and *What's good bro?* meet Construction Grammar criteria for constructions—being frequent, functionally coherent, and socially conventionalized—and reinforce a cognitive sociolinguistic perspective linking grammatical patterns and social meaning in regional contexts. Methodologically, the study highlights the value of combining correspondence analysis with spatial visualization to uncover patterns of variation and entrenchment that might be missed by purely qualitative or single-factor approaches. Future research could explore less frequent or longer *bro* expressions, incorporate sociodemographic variables, and use experimental methods to test the cognitive reality of these constructions. Overall, the regionally patterned use of *bro* constructions illustrates how they emerge, spread, and acquire meaning in context, revealing the dynamic interplay between grammar, identity, and geography.

Bibliography

- Braun, F. (1988). *Terms of address: Problems of patterns and usage in various languages and cultures*. Mouton de Gruyter.
- Brown, R., & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (Ed.), *Style in language* (pp. 253–276). MIT Press.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: A content-based approach to geo-locating Twitter users. *Proceedings of the 19th ACM Conference on Information and Knowledge Management*, 759–768.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Hilpert, M. (2019). *Construction grammar and its application to English*. Edinburgh University Press.
- Kristiansen, G., & Dirven, R. (2008). *Cognitive sociolinguistics: Language variation, cultural models, social systems*. Mouton de Gruyter.
- Leech, G. (1999). The distribution and function of vocatives in American and British English conversation. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 107–118). Rodopi.
- Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse and Communication*, 6(2), 181–201.
- Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. Bloomsbury Academic.

高校英語教科書に見るテキストタイプの相違

安間一雄(獨協大学)

ammakazuo@mac.com

Abstract

Managing text types is an integral part in reading comprehension and essay writing. Without good understanding and implementation of strategies for text types success in academic production as well as business handling would be hopeless. EFL textbooks — here we focus on MEXT-approved senior high school textbooks — do not appear to take into consideration a good balance of texts of various types. In this research 10 textbooks of relatively high market share were examined as to how varied their texts are in terms of four text types: expository, descriptive, persuasive, and narrative. The result was an immense degree of difference in text type orientations. One of our findings was the major contrast in expository versus narrative, which is often the focus of discussion in second language acquisition. With respect to readiness for university-level academic English education cultivating CALP as its basic element, selection of a textbook with a balanced repertoire in text types sounds rational and fruitful.

Keywords

高校英語教科書, テキストタイプ, expository, narrative, 教科書間格差

1. 背景と概要

高校英語教科書 10 冊(2015 年版, 3 年生用)を対象として, 収録されている本文テキストの各文を Bain (1866) による古典的分類に則り 4 種類 (expository, descriptive, persuasive, narrative) に分類し, さらに課内での位置づけも考慮した。その結果, 教科書によりテキストタイプの内訳は大きく異なることが判明した。例えば教科書 C では expository が 13%, narrative が 68%であったのに対し, 教科書 M ではそれぞれ 58%と 22%を占めていた。その上で, どの教科書がどのタイプを典型的に反映しているかを対応分析により明らかにした。本研究は高校での教科書の選定がその後の英語学習へのレディネス形成に大きな要因を占めている可能性を示唆している。

2. 研究課題

義務教育である中学校と異り高等学校においては教科書の採択が各学校に委ねられているため, 生徒の学習到達度並びにニーズに基づき様々な種類の教科書が存在する。一方で小中高大

に互る外国語教育の連繫という視点では、学習対象教材の円滑な推移が求められる。例えば物語文を多用する教材でその構成概念(どのように物語を読み解くか)に習熟した後で論説文を客観的・批判的に読むことを求めても多くの学習者は困惑を感じるであろう(奥村 2021; Yamashita, Shiotsu, and Kusanagi 2024)。逆もまたしかりである。更に、TOEIC, IELTS 等のアカデミックな英語能力を測定するテストに対しては相応の準備が必要である。本研究では高校英語教科書 10 冊に収録されたテキストタイプの変異を探索的に明らかにする。

3. 対象教材

本研究で対象としたのは次の 10 冊の教科書である。何れも 2015 年度に検定済み教科書として使用が開始されたものである。また、3 年生用教科書に限定した。これは最終学年用教科書に当該教科書の言語学習観が最もよく反映されていると考えたからである。

Crown English Communication III(三省堂)

Element English Communication III(啓林館)

Genius English Communication III(大修館書店)

LANDMARK English Communication III(啓林館)

MAINSTREAM English Communication III: Strategic Reading Focus Advanced(増進堂)

New One World Communication III(教育出版)

POLESTAR English Communication III(数研出版)

PRO-VISION English Communication III(桐原書店)

PROMINENCE Communication English III(東京書籍)

UNICORN English Communication III(文英堂)

4. テキストタイプ

読解に関する第 2 言語習得の分野では、談話の性質を表す要素として narrative と expository のテキストタイプがしばしば対比される。Berman and Katzenberger (2004) や Berman and Nir-sagiv (2007) は両者の語彙・文法・談話構造等の特徴を挙げた上で、母語話者に関しては narrative は低年齢層の児童に典型的なディスコースであり、expository は年齢増加に伴う知的成長に沿って獲得されるディスコースであると位置づけている。いわば BICS と CALP の習得と並行していると考えられる。一方、修辞学並びにその実践としての作文教育においては伝統的に Alexander Bain による 4 つのテキストタイプ、即ち descriptive, expository, narrative, persuasive が扱われてきた (DiPardo, 1990)。ただ、それぞれの中心的概念はあるものの境界線は曖昧である。そこで筆者は実際のテキスト分類に適した定義を下記のように行った。この定義に当たっては Perdue Online Writing Lab (n.d.) を参照した。

Descriptive text

- ・事実及び対象物に対する客観的記述。
- ・核となるテーマ構造が存在しない。例: 歴史的事実の列挙, 経歴, メニュー, 時刻表, 名簿

Expository text

- ・核となるテーマ構造が存在し、各テーマについて議論を詳細に評価し、その根拠を挙げ、得られた事実から新たな知見を提示する。典型的にはそれぞれのテーマが次のような談話構造を持つ。

具体的問題提起→解決／説明

- ・議論展開の談話マーカーを使用する。例: **therefore, in conclusion, nevertheless, to sum up**
- ・読者／聴衆の理解に譲歩した話法 (modification) スタイルを取らない。
- ・テーマに関する歴史的記述はここに含める。例: アメリカインディアンの迫害史

Narrative text

- ・個人を含む主体に関する事象の経時的展開を記述する。
- ・核となるテーマ構造は存在しないか、あったとしても非明示的である。典型的にはそれぞれの話題が次のような談話構造を持つ。これは主語の人称には依らない。
[1 人称である書き手]説明する[2 人称である聞き手／読み手][事象の記述]
- ・読者／聴衆と見解を共有するものとして筆者の主観を普遍化した表現が用いられる。これは典型的に 1 人称複数の視点で語られる。例: ‘harmful to our health’, ‘as much/many as’.
- ・具体的出典を挙げずに普通名詞あるいは代名詞で矮小化する。例: ‘Many studies have shown that ...’; ‘In 1953 two young scientists made ... discoveries’.
- ・客観的・普遍的価値のある事象についてではなく、特定の個人・集団を主題として取り上げる。例: ‘Finishing high school is difficult My cousin Nancy is a good example.’
- ・しばしば読者／聴衆の理解に譲歩した話法 (modification) スタイルを取る。
- ・しばしば談話のヘッジとして推測の助動詞 **may/might** や断定を和らげる副詞句・形容詞句 **probably, as it were** 等が使われる。
- ・議論の根拠を必要としない主張／記録はここに含める。例: 伝記・独白・日記

Persuasive/argumentative text

- ・トピックに関する発話者／筆者の見解・主張を確立させる。例: 意見, 要求, 問合せ, 意思確認, 賞賛, 叱責, 案内, 挨拶
- ・談話的効果を前提としてあるいは念頭に置いて表現される。例: 広告, 計画書, 演説

5. データ収集方法

一連のまとまりのある文章を「談話」とし、各談話内の段落毎にテキストタイプを上記基準に従って目視分類した。ただし独立した談話が 50 語未満である場合は省略したが、案内広告・旅程表のように個別の語句は短くても全体として 1 つの談話を構成するものは対象とした。

教科書の場合、通例 1 課の構成単位として本文テキスト以外に導入部と練習問題及び補充教材がある。導入部分は本文読解や練習問題指示と同様に教科書編集者のコメントであるので本研究の対象からは除外した。実際、導入部分は *Crown* のみにしかなく絶対数が少なく(約 18 万件中 1263 件)、かつこの部分には *expository* のテキストは皆無であった。一方、練習問題・

補充教材によっては新たな談話が提示されることがある。50 語以上であればこれも算入した。従って、ここではテキストセクションとして本文 (Main), 練習問題 (Skill), 補充教材 (Supplement) を設定した。

テキスト毎に教科書名 (Book)・セクション区分 (Section)・セクション相対位置(教科書内での当該談話の開始ページを 0・100% で相対化したもの, Page%)・テキストタイプ (Type)・テキスト長(語数, Length)を記録した。同一課の同一談話であるが引用が挿入されている部分は異なるテキストタイプと認定された。更にそれぞれのテキストについて Flesch reading ease score も追加記録した。

6. 分析ツール

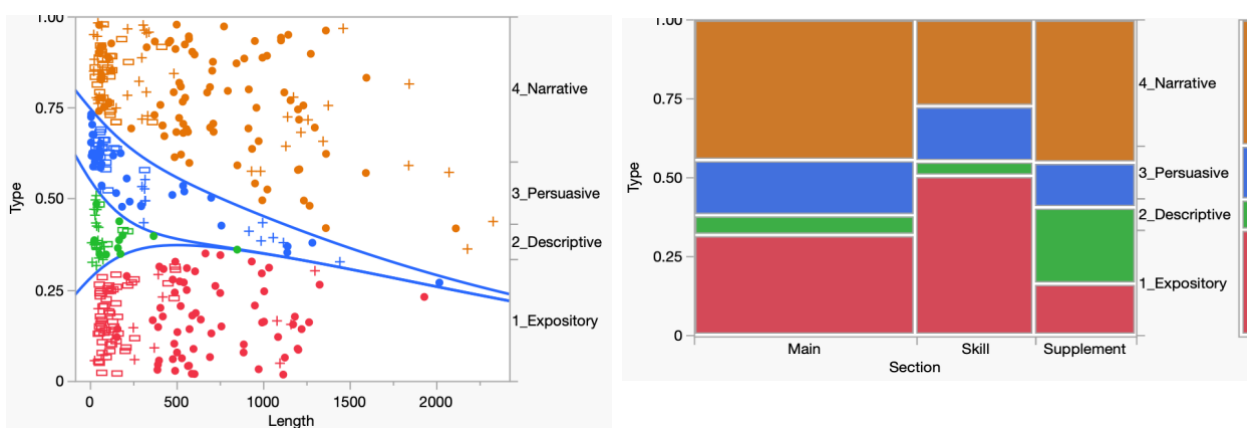
本研究では *JMP* version 18.2.0 (SAS Institute, 2024-2025) を用い, 要因計画法(二元配置の分散分析)・ロジスティック回帰分析・分割表分析・対応分析を行った。

7. 結果

7.1 テキストタイプの決定要因 (1)

要因計画法(二元配置の分散分析)により説明変数の効果を検証した。説明変数として教科書名・セクション区分・テキスト長(語数)を使用した。他の変数の影響はほぼ皆無であった。最大の要因はテキスト長である(データ省略)。ロジスティック回帰分析による次の図 1 が示すように、テキスト長が長くなればなるほど narrative である確率が大きくなる。

[図 1(下左) テキスト長によるテキストタイプの推移。図中丸い点はセクション= Main, 四角い点はセクション= Skill, 十字の点はセクション= Supplement をそれぞれ表す。各点の縦方向の位置は同一カテゴリー内でランダムに付与されている。]



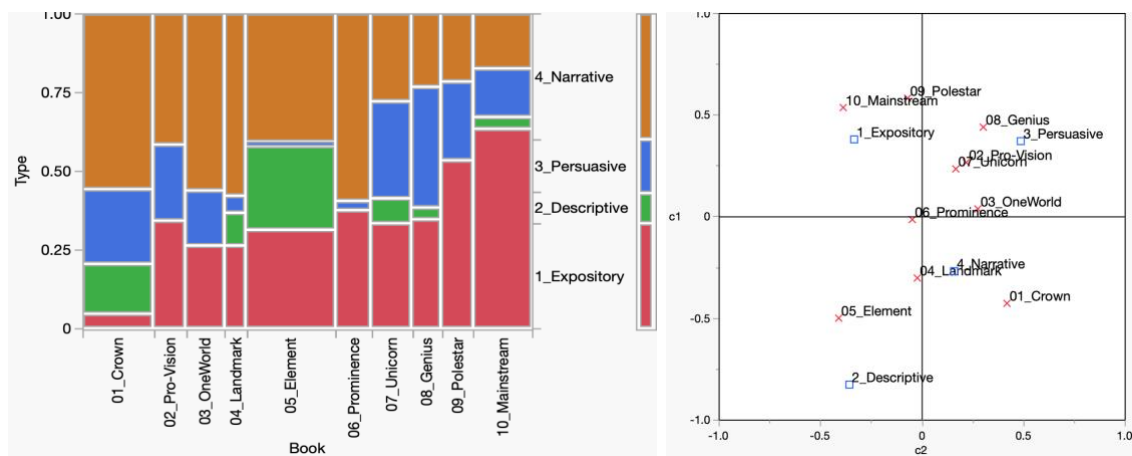
7.2 セクションによるテキストタイプの変異

[図 2 (上右) セクションによるテキストタイプのモザイク図]

Main と Supplement は expository / descriptive の配分が異なる。これは Supplement において資料的なテキストを提示して本文理解の応用を求めていると考えられる。Narrative の割合は殆ど変わらず、物語の扱い方に関しては両者で同様の発想をしていると考えられる。一方、Skill においては expository が Main の約 4 割増し、Supplement の 3 倍以上に達している。スペースが限られる読解・文法練習では事実の明示的説明である expository が効率的であると思われる。

7.3 教科書によるテキストタイプの変異

[図 3(下左) 教科書によるテキストタイプのモザイク図(全体)]



一見して明らかのように、教科書による差が大きい。特に顕著なのが *Crown* で、expository は全体の 5% しか収録されていない。一方 *Mainstream* は 63% である。Narrative を多用するのは *Prominence*, *Landmark*, *One World*, *Crown* で、何れも 1 冊内の半分以上を占めている。また、*Element* では descriptive を特に取り上げている。この教科書とテキストタイプの近接性を 2 次元にプロットしたもの(対応分析)が図 4 である。

[図 4(上右) 教科書とテキストタイプの対応分析(全体)]

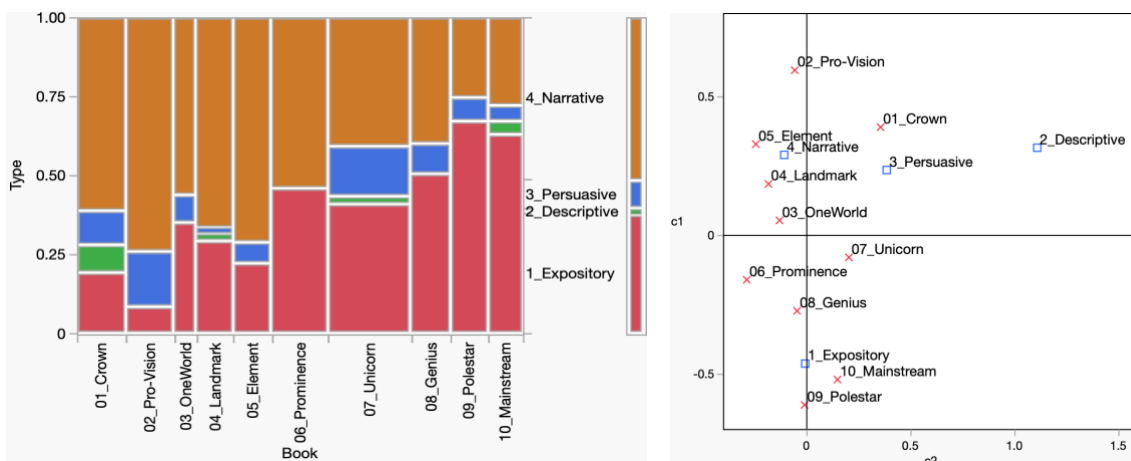
図はテキストタイプの変数の布置と教科書の変数の布置とを対応つけたもので、原点からの方向が近いものが近接性が高い。ここから次のように教科書の傾向を読み取ることができる。

- (1) expository の特徴は *Mainstream* に最もよく出ている。
- (2) descriptive の特徴はあまり明確ではないが強いて言えば *Element* に伺える。
- (3) persuasive の特徴はやや薄いながら *Pro-Vision* と *Unicorn* に見出される。
- (4) narrative の特徴は *Crown* に最もよく出ている。

7.4 テキストタイプの決定要因 (2)

本研究ではテキスト長を変数の1つに指定した。長時間学習者が触れているテキストタイプが学習の形成に大きく影響を与えると考えたからである。数量的に重要だと思われるのは Main の expository と narrative, Supplement の narrative の3つの場合である(数値省略)。まず Main に注目し、教科書毎のテキストタイプ(結果 7.3 の限定版)を見ると次のようになる。

[図 6(下左) 教科書によるテキストタイプのモザイク図 (Main)]



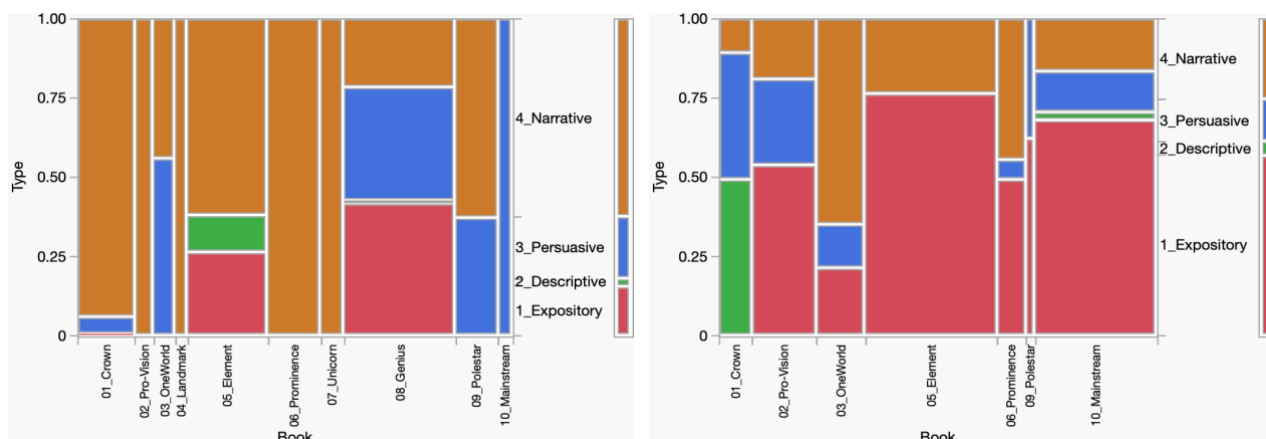
教科書での学習対象の主体とも言えるテキスト本体のみに限定すると上記結果 7.3 とはやや異なる特徴が見える。Narrative は依然多い(*Pro-Vision*, *Element*, *Landmark*, *Crown*, *One World*, *Prominence* で 50%以上)が, *Crown* の narrative 依存は薄れ, *Pro-Vision* と *Element* は *Crown* を上回る。Unicorn, Genius では persuasive が減り, その分 narrative が増えている。*Crown* でも persuasive はやや減ったが, 代わりに expository が増えている。更に, *Element* では descriptive がなくなった分, narrative が増えている。

[図 7(上右) 教科書とテキストタイプの対応分析 (Main)]

本体テキストのみの場合, 次の特徴が挙げられる。

- (1) expository の特徴は *Polestar* に最もよく出ている。
- (2) descriptive の特徴が出ている教科書はない。
- (3) persuasive の特徴が出ている教科書はない。
- (4) narrative の特徴は *Pro-Vision*, *Element*, *Landmark* に比較的好く出ている。

[図 8(下左) 教科書によるテキストタイプのモザイク図 (Supplement)]



補充学習教材である Supplement の部分では多くの教科書が narrative を使用している。気の置けない読み物といった感じなのか。Genius では全体とほぼ相似の構成になっており，expository, persuasive, narrative の 3 タイプである。Element にも若干その傾向があるが，ここでは descriptive を使うことが他の教科書では見られない特徴である。

[図 9(上右) 教科書によるテキストタイプのモザイク図 (Skill)]

本課での学習事項のまとめ・応用である Skill の部分では多くの教科書が expository を使用しており，論理的簡潔性あるいは明示性を重視していると考えられる。Crown は expository ではなく descriptive を使うことが他の教科書では見られない特徴である。

7.5 教科書による読みやすさの比較及び推移

各教科書について Flesch reading ease score (Readable, n.d.) を用いて読みやすさを数値化した。値が高いほど読みやすいことを表している(図省略)。全般的な読みやすさとしては Polestar(易)と One World(難), Polestar(易)と Crown(難)の間に大きな開きがあるが，これはそれぞれの教科書の対象者が異なるので仕方がない。年間を通しての読みやすさの推移も年度末を除いては大きな変動は少ない。年度末(100%近傍)に多くの教科書が難度を下げるのは，補充教材を組み入れているからであると思われる。ただし，実際の教室の現場でここまで扱うかどうかは定かでない。

7.6 テキストタイプによる読みやすさの比較及び推移

各テキストタイプについて Flesch reading ease score を用いて読みやすさを数値化した。値が高いほど読みやすいことを表している(図省略)。Narrative (8.305) と persuasive (8.512) が最も読みにくいとされ，何れも expository (9.763) 及び descriptive (11.563) との間で有意な差があった。これを時系列で見ると全体として読みやすさが僅かに上昇しているように思われる。特に descriptive は年度後半で明らかに他のテキストタイプを引き離して易くなっている。

8. 結論・まとめ・考察

まず、教科書によりそこに含まれるテキストのタイプが大きく異なることが判明した。大学においてアカデミックなテキストを読みこなすこと(即ち CALP の運用)が求められていることを考えると、それに適したテキストタイプへの準備が高校の英語学習において適切になされることが望ましい。英語運用能力のバランスという観点からは様々なテキストタイプに触れることが重要であるが、少なくとも本文テキストの半分以上を **narrative** が占める教材は避けた方がよいのではないかと考えられる。

次に、本研究ではテキスト長を説明変数を補強する要因として採用した。100 語のテキストも 1000 語のテキストも同一条件のテキストタイプとして取り上げるのは現実を歪曲している。後者に掛かる学習時間が前者の 10 倍であることが推測されるのであれば、当該テキストタイプへの接触量も 10 倍になると考えた。図 1 に示したように、テキストが長くなればなるほど **narrative** に接する機会が多くなる。これが全体として **narrative** の占める量を増やす要因になっていると考えられる。しかし長い読み物のために他のテキストタイプの学習時間が圧迫されているとも考えられる。

更に、テキスト本文と補充教材とで同一テキストタイプの傾向が見られる教科書がある一方、前者には少ないテキストタイプを後者で採用している教科書も見られた。このことが教科書全体としてのテキストタイプの傾向と本文のテキストタイプの傾向が若干異なっていることを説明していると思われる。特に本文テキストの内訳で見ると、**descriptive** と **persuasive** をその特色としている教材は見当たらなかった。第 2 言語習得分野で **narrative** 対 **expository** の対比に焦点が当てられているのは教材の実態を反映していると思われる。

最後に、テキストの読みやすさに関してはテキストタイプとの顕著な関連は見いだせなかった。**Descriptive** については他のテキストタイプより概して読みやすい傾向があるが、このテキストタイプの絶対数は多くないのでさほど考慮に値しない。ここで使用した **Flesch reading ease score** は単語長(単語に含まれる音節数)と文長(文に含まれる単語数)により読みやすさを判定するという単純な仕組みで、単語の難易度(頻度)や文構造の複雑さは考慮に入れていない。**Descriptive** なテキストには単語やフレーズのための資料が多いのでこのスコアが上がった可能性が高い。どのテキストタイプが学習者にどの程度の読解負荷を掛けるかは教科書編成あるいはカリキュラム施行上の大きな留意点である。読みやすさではなく談話を含めたテキストの難易度情報が得られることを今後の研究課題としたい。

引用文献

- Bain, A. (1866 = 2016). *English Composition and Rhetoric: A Manual*. Leopold Classic Library (originally New York: D. Appleton).
- Berman, R. A. & Katzenberger, I. (2004). Form and function in introducing narrative and expository Texts: A developmental perspective. *Discourse Processes*, 38(1), 57-94. https://www.tandfonline.com/doi/pdf/10.1207/s15326950dp3801_3

- Berman, R. A. & Nir-sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes*, 43(2), 79-120. <https://www.tandfonline.com/doi/pdf/10.1080/01638530709336894>
- DiPardo, A. (1990). Narrative knowers, expository knowledge: Discourse as a dialectic. *Written Communication*, 7(1), 59-95.
- JMP version 18.2.0. SAS Institute, 2024-2025.
- 奥村信彦 (2019) 「テキストタイプが第 2 言語 (L2) 読解力に及ぼす影響再考」『長野保健医療大学紀要』 Vol.5, 13-20.
- Perdue Online Writing Lab. (n.d.)
https://owl.purdue.edu/owl/general_writing/academic_writing/essay_writing/index.html
- Readable. (n.d.) About Readability. <https://readable.com/readability/>
- Yamashita, J., Shiotsu, T., & Kusanagi, K. (2022). Predictors of second language reading comprehension ability: A longitudinal study with learners from grade 9 to 11 in an English as a foreign language context. *Reading and Writing*, 37, 2733-2758. <http://hdl.handle.net/2237/0002011254>

Exploring the Effectiveness of Data-driven Learning and Its Integration with Generative AI

MANABE Daisuke (Kanazawa University, Graduate Student)

daisuke825@stu.kanazawa-u.ac.jp

Abstract

The present study aims to investigate the effectiveness of data-driven learning (DDL) and the integration of generative AI (GenAI) into DDL. This study reports the results of an experiment that examined the effectiveness of DDL and AI-integrated DDL, in which learners made use of GenAI during DDL, for learning an English construction. Furthermore, the results were compared with those obtained from two different teaching methods reported in Manabe (2025). Learners' attitudes toward DDL were also investigated. The results of pre- and post-tests showed that the DDL approaches were effective. A comparison of the four teaching methods demonstrated that each method produces varying degrees of learning effect. The post-questionnaire revealed learners' positive attitudes toward DDL and the use of GenAI for learning English.

Keywords

Data-driven Learning, Generative AI, Applied Construction Grammar

1. Introduction

Data-driven learning (DDL) is one of the corpus-based methods to language teaching and learning, and it has gained increasing attention in foreign language education. In recent years, generative artificial intelligence (GenAI) also has made its way into the field of foreign language education. The potential synergy of DDL and GenAI has also been underscored in recent research (e.g., Mizumoto, 2023).

2. Research Background

DDL aligns well with usage-based theories of language because it emphasizes real language use (i.e., authentic corpus data). Along with a number of advantages, DDL has the potential to make L2 learning closer to L1 acquisition by exposing learners to large amounts of authentic data and encouraging them to learn inductively and autonomously. Learning a construction (a form-meaning pair in construction grammar; e.g., Goldberg, 2006) through DDL is referred to as "data-driven construction learning" (Gilquin, 2021). Manabe (2025) investigated the effectiveness of DDL with an explicit explanation of the form-meaning pairing of the target construction, and this approach was referred to as

“construction-centered DDL” (henceforth, *cxn-centered DDL*). However, there is still little research on the application of DDL from a constructionist perspective. Hence, the present study explores the effectiveness of DDL for learning a construction. This study also aims to explore the effectiveness of the integration of GenAI with DDL. Here, the approach that incorporates GenAI into DDL is referred to as “AI-integrated DDL.” The following research questions (RQs) were examined:

1. Are DDL and AI-integrated DDL effective for learning the *way* construction?
2. Are there differences in effectiveness both among DDL approaches?
3. Are there differences in effectiveness between DDL approaches and a traditional approach?
4. Do Japanese learners of English have positive attitudes toward DDL and the use of GenAI?

3. Method

3.1 Experimental design

The experiment consisted of educational intervention, pre- and post-tests, and pre- and post-questionnaires. The participants were divided into two groups: the DDL and AI-integrated DDL groups. Furthermore, to explore the differences in the effectiveness of teaching methods, this study incorporated data from Manabe’s (2025) experiment: the *cxn-centered* and traditional instruction groups. Since the experimental design in Manabe (2025) was the same as that in this study, the datasets are directly comparable.

The participants learned the *way* construction (Goldberg, 1995, p. 199) during an educational intervention. This study includes four types of instructional methods: DDL, AI-integrated DDL, *cxn-centered* DDL, and traditional instruction. In all three DDL groups, participants were provided with a paper-based concordance with 20 instances of the *way* construction and a worksheet. The instances of the *way* construction were extracted from the Corpus of Contemporary American English (COCA: Davis, 2008-). The DDL group received only the instances without any scaffoldings. The AI-integrated DDL group was allowed to use ChatGPT (OpenAI, 2025) during DDL. The *cxn-centered* DDL included an explicit explanation of the form and meaning of the *way* construction. The traditional instruction group received a teacher-centered lecture, mainly paying attention to the form of the target construction.

The participants were L1 Japanese speakers at a national university in Japan. This study included four groups of participants: DDL ($n = 12$; 10 undergraduates, 2 graduates; M age = 21.58 years; CEFR = B1–B2), AI-integrated DDL ($n = 12$; 8 undergraduates, 4 graduates; M age = 21.83 years; CEFR = B1–C1), *cxn-centered* DDL ($n = 20$; 18

undergraduates, 2 graduates; M age = 20.45 years; CEFR = B1–B2), and traditional instruction (n = 20; 19 undergraduates, 1 graduate; M age = 20.25 years; CEFR = A2–B2). Regarding proficiency levels, the scores of proficiency tests (e.g., TOEIC) were converted to CEFR (cf. Ministry of Education, Culture, Sports, Science and Technology, 2015). However, because of the random sampling procedure, there was substantial variation in participants' proficiency levels.

A vocabulary size test (VST; Hamada et al., 2021) was administered with a time limit of six minutes before pre-tests to examine participants' prior knowledge of English. The pre- and post-tests are composed of two types of tests: a sentence production task (SPT) and a translation task (TT). In the SPT, the participants were asked to generate as many sentences as possible containing the way construction within five minutes. The TT consists of five Japanese-to-English translation questions (5 min).

The pre-questionnaire collected information about speaker attributes (e.g., age). In the post-questionnaire, a Likert scale (five-point) question was carried out to investigate learners' attitudes toward DDL and the use of GenAI for learning English.

3.2 Analytical Procedure

3.2.1 Pre- and post-tests

The results of the SPT were analysed in three steps. In Step 1, the form and meaning were verified and errors were removed. In Step 2, remaining sentences were checked against COCA, and if a matching expression was found, the sentence was classified as correct usage. In Step 3, sentences remained after the first two steps were evaluated for appropriateness by L1 English speakers (three Americans and one Briton for the DDL and AI-integrated DDL groups, and three Americans and one Australian for the cxn-centered DDL and traditional instruction groups), using a four-point Likert scale (4 = appropriate). Sentences with an average rating of 3 or higher were considered correct. The results of the TT were rated by two L1 Japanese speakers (including the author), and answers receiving consistent evaluations from both raters were considered correct.

3.2.2 Statistical Analyses

R (R Core Team, 2024) was used for the statistical analyses. A Shapiro–Wilk test was conducted to assess the normality of the distributions for the VST, pre-test, and post-test scores. Due to space limitations, the detailed results of the Shapiro–Wilk test are not reported here. However, the normality assumption was violated in most cases (the majority with $p < .01$), and hence non-parametric tests were employed in this study.

4. Results and Discussions

4.1 Baseline comparisons

A Kruskal–Wallis test was conducted to first examine whether there were differences among the four groups in prior knowledge of English and the *way* construction. The Kruskal–Wallis test on the VST and pre-test scores of SPT and TT revealed a significant difference among the four groups only in the pre-test SPT ($\chi^2(3, N = 64) = 11.49, p = .009, \varepsilon^2 = .142$). The following post hoc test (Dunn’s tests with Bonferroni correction) showed a significant difference between the DDL and traditional instruction groups in the pre-test SPT ($Z = 2.94, p = .020, r = .368$), with the DDL group outperforming the traditional instruction group.

4.2 Within-group comparisons

As a result of Wilcoxon signed-rank tests on the SPT, significant gains were observed in all groups: DDL ($p = .002, r = .883$), AI-integrated DDL ($p = .002, r = .883$), cxn-centered DDL ($p < .001, r = .851$), and traditional instruction ($p < .001, r = .877$). In the TT, all groups also showed significant improvement: DDL ($p = .002, r = .883$), AI-integrated DDL ($p = .004, r = .883$), cxn-centered DDL ($p = .001, r = .877$), and traditional instruction ($p < .001, r = .877$). Based on the comparisons, all instructional methods in this study were found to be effective. Hence, the results supported RQ1, demonstrating that DDL and AI-integrated DDL were effective for learning the *way* construction.

4.3 Between-group comparisons

Since a significant difference was found between the DDL and traditional instruction groups in the pre-test SPT (Section 3.1), a residual analysis was conducted by regressing post-test scores on pre-test scores. A Kruskal–Wallis test on the adjusted residual scores indicated a significant difference among the four groups ($\chi^2(3, N = 64) = 22.10, p < .001, \varepsilon^2 = .318$). Following the Kruskal–Wallis test, the Dunn’s tests with Bonferroni correction showed that there was a significant difference between the cxn-centered DDL and traditional instruction groups ($Z = -4.683, p < .001, r = .567$), indicating that the traditional instruction was more effective than the cxn-centered DDL.

A Kruskal–Wallis test on the post-test TT also demonstrated a significant difference among the four groups ($\chi^2(3, N = 64) = 26.62, p < .001, \varepsilon^2 = .394$). Dunn’s tests with Bonferroni correction showed a significant difference: between the DDL and traditional instruction groups ($Z = -3.864, p < .001, r = .483$), between the cxn-centered DDL and traditional instruction groups ($Z = -4.747, p < .001, r = .593$), and between the AI-integrated DDL and traditional instruction groups ($Z = -2.689, p = .043, r = .336$). The

traditional instruction group outperformed the other three DDL groups. However, this result is not surprising, given the fact that only the traditional instruction group received a direct Japanese translation during the educational intervention.

RQ2 was not confirmed, as no statistically significant differences were found among the DDL groups. The traditional instruction group outperformed the cxn-centered DDL group in the SPT and all three DDL groups in the TT, thus RQ3 was confirmed.

4.4 Evaluation of DDL and the use of GenAI

RQ4 was confirmed, as the results of the post-questionnaire indicated learners' positive attitudes toward DDL (Figure 1; for the results of the cxn-centered DDL group, see Manabe, 2025, p. 54). Figure 2 shows that the participants in the AI-integrated DDL group also had a positive attitude toward using GenAI for learning English.

Figure 1 *Post-questionnaire results for the DDL group (n = 12; left) and the AI-integrated DDL group (n = 12; right)*

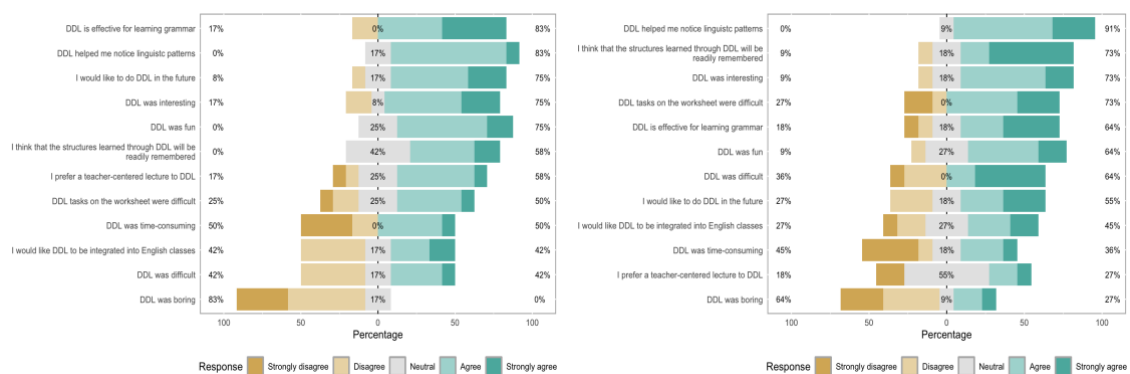
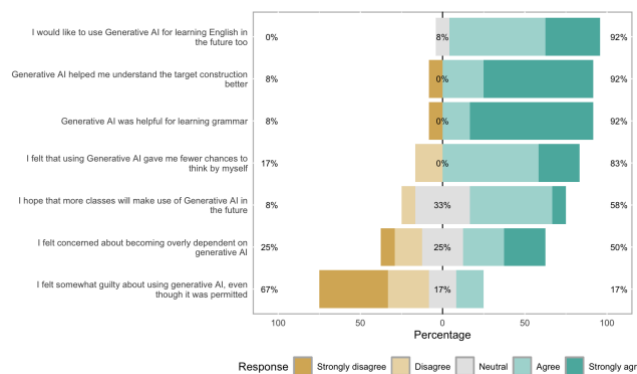


Figure 2 *Post-questionnaire results on GenAI (AI-integrated DDL group, n = 12)*



5. Conclusion

This study explored the effectiveness of DDL through an experiment, and all DDL approaches were found to be effective. However, there is room for improvement both in

the experimental methodology (e.g., considering learners' proficiency levels, conducting longitudinal experiments) and in DDL (e.g., exploring ways to integrate GenAI, revising explicit explanations, incorporating with other teaching methods and activities). To enhance the effectiveness of DDL, it is essential to explore the experimental and practical improvements such as the ones mentioned above, among others.

Acknowledgments

I wish to express my sincere gratitude to all the participants and informants who generously contributed to this research. This work was supported by JST SPRING, Grant Number JPMJSP2135.

Bibliography

- Davies, M. (2008-) The corpus of contemporary American English (COCA). Available online at <https://www.english-corpora.org/coca/>
- Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics*, 31(2), 229–247. <https://doi.org/10.1111/ijal.12317>
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.
- Hamada, A., Iso, T., Kojima, M., Aizawa, K., Hoshino, Y., Sato, K., Sato, R., Chujo, J., & Yamauchi, Y. (2021). Development of a vocabulary size test for Japanese EFL learners using the new JACET list of 8,000 basic words. *JACET Journal*, 65, 23–45.
- Manabe, D. (2025). Work your way through authentic data: Data-driven construction learning and its effectiveness explored through an experimental study. *English Corpus Studies*, 32, 41-64.
- Ministry of Education, Culture, Sports, Science and Technology. (2015). 各試験団体のデータによる CEFR との対照表 [Correspondence table between CEFR and test organization data]. https://www.mext.go.jp/b_menu/shingi/chousa/shotou/117/shiryo/_icsFiles/afieldfile/2015/11/04/1363335_2.pdf
- Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, 3(3), 100074. <https://doi.org/10.1016/j.acorp.2023.100074>
- OpenAI. (2025). *ChatGPT* [Large language model]. <https://chat.openai.com/>

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

t スコアと相互情報量における頻度効果の代数的考察

宮崎 佳典 (静岡大学)

yoshi@inf.shizuoka.ac.jp

田中 省作 (立命館大学 / 国立沖縄高専)

sho@lt.ritsumeimei.ac.jp

Algebraic Study on the Frequency Effect in t -score and Mutual Information

MIYAZAKI, Yoshinori (Shizuoka University)

TANAKA, Shosaku (Ritsumeikan University / NIT, Okinawa)

Abstract

This study examines the frequency effects in two major co-occurrence measures used in corpus research: t -score and pointwise mutual information. Prior studies often lacked explicit assumptions, hindering precise interpretation. By formally specifying these assumptions and algebraically analyzing both the aforementioned measures, this study provides a more detailed account of their behavior.

Keywords

共起指標, t スコア, 相互情報量, 頻度効果, 代数操作

1. はじめに

t スコアや相互情報量¹といった共起指標は共起性を測る, コーパス研究における基本的な道具立ての一つである。これらの共起指標は, それぞれの頻度, 共起した頻度 (共起頻度), そして総語数から算出される。そんな t スコアと相互情報量には, コーパス研究のテキストで, 次のようなことが付言されることがある。「 t スコアは, 共起の必然性の確信度で, 頻度が高ければ t スコアも高くなる」, その一方で, 「相互情報量は情報理論に基づいた予測可能性の一部で, 頻度が低い方が指標値は高く, 低頻度をより, ときには過剰に評価する」といった, それぞれで単純な相関めいた効果である。多くのテキストでは紙面の都合か, 各頻度や状況が曖昧で, 効果そのものも必ずしも明示的に記述されておらず, 精確な理解を難しくすることがある。そこで本発表は, このような t スコアと相互情報量の両指標における, 各頻度とその変化の関係を代数的に明らかにする。

2. 目的と諸定義

まず, 共起指標における「頻度効果」を確認しておく。共起指標における頻度効果とは, 語 x, y が共起したとき, 共起に関する頻度をのぞく他の状況は同等であるのに, x, y の共起頻度や x, y それ

それぞれの頻度，総語数に連動する変化である。ここでいくつかの疑問が浮かぶ。

- 「共起に関する頻度をのぞく他の状況は同等」とは，どういうものか
- 共起指標の値はどう変化するか
- 変化はどうして起こるのか，あるいは起こらないのか

本発表は，これらの疑問に，共起指標の定義式に対する高校数学程度の代数操作を通して答える。数式の見かけ上の複雑さはあるかもしれないが，代数操作で論が進むため，定性的な説明では生じがちな曖昧さがなく，記号的に明確である，という利点がある。

コーパスなどの特定の言語資料の下での，語 x と語 y の共起の状況を次のような分割表で表し，適宜，基準分割表と呼ぶ。

表 1 基準分割表と具体例

	y	\bar{y}			y	\bar{y}	
x	f_{xy}	$f_x - f_{xy}$	f_x	x	10	90	100
\bar{x}	$f_y - f_{xy}$	$N - f_x - f_y + f_{xy}$	$N - f_x$	\bar{x}	390	99,510	99,900
	f_y	$N - f_y$	N		400	99,600	100,000

表 1 で， $\bar{\alpha}$ は語 α 以外の語， $f_{\alpha\beta}$ は語 α と語 β の共起頻度， f_α は語 α の頻度， N は総語数を意味する。

本発表で対象とする共起指標の t スコアと相互情報量は，コーパス研究の文脈では，頻度ベースで示される場合が多く，次のように与えられる。ここでは x, y の共起の割合(確率) $p(x, y)$ ，それぞれの語の割合(確率) $p(x), p(y)$ を含む形も示しておく。

$$t(x, y) = \frac{1}{\sqrt{f_{xy}}} \left(f_{xy} - \frac{f_x f_y}{N} \right) = \frac{N}{\sqrt{f_{xy}}} \left(\frac{f_{xy}}{N} - \frac{f_x}{N} \cdot \frac{f_y}{N} \right) = \frac{N}{\sqrt{f_{xy}}} (p(x, y) - p(x)p(y))$$

$$\text{PMI}(x, y) = \log_2 \frac{N f_{xy}}{f_x f_y} = \log_2 \frac{f_{xy}/N}{f_x/N \cdot f_y/N} = \log_2 \frac{p(x, y)}{p(x)p(y)}$$

表 1 の具体例の x, y の t スコアは 3.04，相互情報量は 4.64 である。なお，コーパス研究では一般に，偶然の共起よりも多く出現する， $p(x, y) > p(x)p(y)$ が成立するような共起を対象とする。そのため，両指標とも正数の印象が強いかもしれないが，実際には，負の値にもなり，そのときは $p(x, y) < p(x)p(y)$ が成立している。

3. 出現状況と頻度効果

3.1 出現状況

共起の出現状況として，次の 3 つを考える。

- (1) $f_{xy}:f_x:f_y:N$ の比は一定で， f_{xy} が異なる場合
- (2) $f_{xy}:f_x:f_y$ の比と N は一定で， f_{xy} が異なる場合
- (3) f_{xy}, f_x, N は一定で， f_y が異なる場合

(1),(2)で変化した頻度は $u f_{xy}$ という具合に， $u > 0$ の定数倍で表現する。 $0 < u < 1$ のとき， $u f_{xy}$ は頻度が減少， $u > 1$ のとき $u f_{xy}$ は頻度が増加することを意味する。頻度は正の整数なので，

u は本来、不連続であるが、頻度効果を論じる際には、連続で考えても、特段、問題はない。(3)では uf_y となる。

基準分割表を(1),(2)では f_{xy} , (3)は f_y を u 倍したときの指標値の変化をみる。その際、変化した各値は f'_{xy}, f'_x, f'_y, N' のようにダッシュ付きで表す。基準分割表に対する具体的な指標値を $\psi(x, y)$ とすると、差 $\psi'(x, y) - \psi(x, y)$ や比 $\psi'(x, y)/\psi(x, y)$ を計算すれば、指標値の変化がわかる。差か比かは本質的ではなく、適宜、計算しやすい方を選択した。

3.2 (1) $f_{xy}:f_x:f_y:N$ の比は一定で、 f_{xy} が異なる場合

次のように基準分割表の各要素が u 倍となった場合である。表 2 の具体例は、表 1 の具体例を(1)の下で $u = 10$ としたもので、 t スコアは 9.60, 相互情報量は 4.64 となる。

表 2 (1)の分割表と具体例 ($u = 10$)

	y	\bar{y}			y	\bar{y}	
x	uf_{xy}	$u(f_x - f_{xy})$	uf_x	x	100	900	1,000
\bar{x}	$u(f_y - f_{xy})$	$u(N - f_x - f_y + f_{xy})$	$u(N - f_x)$	\bar{x}	3,900	995,100	999,000
	uf_y	$u(N - f_y)$	uN		4,000	996,000	1,000,000

指標値の変化をみていこう。冗長かもしれないが、最初なので、丁寧に展開してみる。

$$t'(x, y) = \frac{1}{\sqrt{f'_{xy}}} \left(f'_{xy} - \frac{f'_x f'_y}{N'} \right) = \frac{1}{\sqrt{uf_{xy}}} \left(uf_{xy} - \frac{uf_x \cdot uf_y}{uN} \right) = \frac{\sqrt{u}}{\sqrt{f_{xy}}} \left(f_{xy} - \frac{f_x f_y}{N} \right)$$

頻度の変化前後の指標値の比 $t'(x, y)/t(x, y)$ は、次のとおりである。

$$t'(x, y) / t(x, y) = \frac{\sqrt{u}}{\sqrt{f_{xy}}} \left(f_{xy} - \frac{f_x f_y}{N} \right) / \frac{1}{\sqrt{f_{xy}}} \left(f_{xy} - \frac{f_x f_y}{N} \right) = \sqrt{u}$$

ここから、(1)で頻度が u 倍に変化すると、 t スコアは \sqrt{u} 倍になることがわかる。平方根関数は単調増加関数なので、 $t(x, y) > 0$ ならば、 f_{xy} が増えれば t スコアは増加し、そして $t(x, y) < 0$ ならば f_{xy} が増えれば t スコアは負で減少するものの、値の大きさそのものは増加する、という単純な関係が見出される。実際、表 2 の具体例は表1の具体例を $u = 10$ 倍したもので、 t スコアは $\sqrt{10} \approx 3.16$ 倍になっている。

次に相互情報量をみてみよう。

$$PMI'(x, y) = \log_2 \frac{N' f'_{xy}}{f'_x f'_y} = \log_2 \frac{uN \cdot uf_{xy}}{uf_x \cdot uf_y} = \log_2 \frac{N f_{xy}}{f_x f_y}$$

$PMI'(x, y)$ は元の相互情報量とまったく同じ形になる。したがって、変化前後の比をとるまでもなく、この状況では相互情報量は頻度の影響をまったく受けないことがわかる。表 2 の具体例の相互情報量は 4.64 で変化していない。

この状況の要点は、 $f_{xy}:f_x:f_y:N$ の比が一定なことから、割合に直った $p(x, y)$ と $p(x)p(y)$ の関係

が変わっていないことである。 t スコアが検定統計量由来で、「頻度が高くなると共起の必然性の確証が高まる」といった説明が添えられることがある。この状況は、 $p(x, y)$ と $p(x)p(y)$ の関係が不変なまま f_{xy} が増加しているのだから、まさにその説明にあたる。一方、相互情報量の頻度効果は皆無で、「低頻度をより、ときには過剰に評価する」といった説明は成り立たない。つまり、 t スコアと相互情報量の頻度効果のよくある上記の説明は、(1)で同時に成立することはなく、前提としている状況が異なることがわかる。

なお、 $p(x, y)$ と $p(x)p(y)$ の関係が変わっていない分割表は、頻度の変化に対する条件を(1)からより緩めたものも考え得る。たとえば、 $f'_{xy} = 100, f'_x = 500, f'_y = 800, N = 100,000$ は、総語数は同じまま、 f_{xy} だけ 10 倍になったもので、 f'_x, f'_y の組み合わせは他にもある。したがって、(1)は $p(x, y)$ と $p(x)p(y)$ の関係が不変な状況の一例に過ぎない。 $p(x, y)$ と $p(x)p(y)$ の関係が変わっていない一般の状況の頻度効果も、ここまでと同様の展開で、結論も同じとなる。

3.3 (2) $f_{xy} : f_x : f_y$ の比と N は一定で、 f_{xy} が異なる場合

(1)に比べてイメージしにくいかもしれないが、 $p(y|x)$ が基準分割表から変わっていない、次の表 3 のような分割表となる。なお、 $p(y|x)$ は x が出現したときに y が共起する割合(条件付き確率)である。中心語 x に対する共起語を枚挙したなかに y が含まれる割合、と考えてもよい。 $p(x|y)$ も同様に基準分割表から変わっていない。

表 3 (2)での分割表と具体例($u = 10$)

	y	\bar{y}			y	\bar{y}	
x	uf_{xy}	$u(f_x - f_{xy})$	uf_x	x	100	900	1,000
\bar{x}	$u(f_y - f_{xy})$	$N - uf_x - uf_y + uf_{xy}$	$N - uf_x$	\bar{x}	3,900	95,100	99,000
	uf_y	$N - uf_y$	N		4,000	96,000	100,000

表 3 の具体例は、表 1 の具体例を(2)の条件下で $u = 10$ 倍としたものである。このとき、 t スコアは 6.00、相互情報量は 1.32 となる。(2)での指標値の変化は、とくに t スコアが記述の上でやや込み入っており、結論の概要をさきに示しておく。 t スコアの頻度効果は単調ではなく、 u の範囲によって振る舞いが異なり、一方、相互情報量では低頻度をより評価する効果が現れる。

$t'(x, y) - t(x, y)$ は次のようになる。

$$\begin{aligned}
 t'(x, y) - t(x, y) &= \frac{1}{\sqrt{uf_{xy}}} \left(uf_{xy} - \frac{uf_x \cdot uf_y}{N} \right) - \frac{1}{\sqrt{f_{xy}}} \left(f_{xy} - \frac{f_x f_y}{N} \right) \\
 &= \sqrt{f_{xy}} (-Du\sqrt{u} + \sqrt{u} + D - 1)
 \end{aligned}$$

ここで、 $D = f_x f_y / (N f_{xy})$ である。一見複雑に見えるが $v = \sqrt{u}$ とおくと v の 3 次関数として扱いやすくなり、 t スコアの増減は、次のように整理することができる。ただし、 $R = (1 - \sqrt{4/D - 3})^2 / 4$,

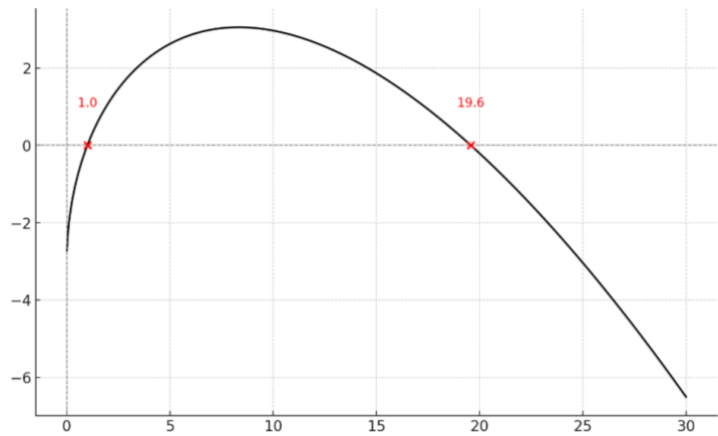
$u = 1$ は増減しないことは自明なので記述を省いている。

- i. $D < 1/3$ のとき, $1 < u < R$ では増加, それ以外では減少
- ii. $D = 1/3$ のとき, 全区間で減少
- iii. $1/3 < D < 1$ のとき, $R < u < 1$ では増加, それ以外では減少
- iv. $D \geq 1$ のとき, $u < 1$ では増加, それ以外では減少

実際の例を見たほうが理解しやすい。図1は表1の具体例に対して, 横軸を u , 縦軸を $t'(x, y) - t(x, y)$ で描いたものである。 $D = 1/25$ で, 上記では i の場合にあたる。 $R \approx 19.6$ で, $1 < u < R$ のときだけ t スコアは増加し, それ以外の範囲では減少する。つまり, $u > 1$ ではいったん増加し, そののち減少する。この状況での頻度の変化は, $p(x, y)$ と $p(x)p(y)$ の関係を変えるため, 結果, t スコアの振る舞いは単調ではない。

表3の具体例($u = 10$)では6.00と t スコアは増加しているが, $u = 20$ では2.83となり, 表1のときの3.04よりも減少している。 $u = 25$ で $p(x, y) = p(x)p(y)$ となり, t スコアは0となる。図1で $u = 25$, $t'(x, y) - t(x, y) \approx -3.04$ が確認できる。 $u > 25$ では, $p(x, y) < p(x)p(y)$ で, t スコアは負の値をとる。符号を考えない, t スコアの値の大きさだけをみても振る舞いが単調ではないことがわかる。なお, $p(x, y) = p(x)p(y)$ となるのは $u = 1/D$ である。

図1 表1例に対する u (横軸) \times $t'(x, y) - t(x, y)$ (縦軸)



一方, 相互情報量は次のようになる。

$$\begin{aligned} \text{PMI}'(x, y) - \text{PMI}(x, y) &= \log_2 \frac{N \cdot u f_{xy}}{u f_x \cdot u f_y} - \log_2 \frac{N f_{xy}}{f_x f_y} \\ &= (\log_2 N f_{xy} - \log_2 u - \log_2 f_x f_y) - (\log_2 N f_{xy} - \log_2 f_x f_y) = -\log_2 u \end{aligned}$$

u に対して相互情報量は単調に減少することがわかる。つまり, 頻度が低い方が高くなる, 低頻度をより評価する, という頻度効果が確認される。

3.3 (3) f_{xy}, f_x, N は一定で, f_y が異なる場合

この状況は, 今, 注目している語(中心語) x に対して共起する語(共起語) y との共起頻度 f_{xy} は

同じである。このとき、共起語の y が高頻度か低頻度かによる共起指標の変化をみる。次のような分割表で表すことができる。

表 4 (3)での分割表と具体例($u = 10$)

	y	\bar{y}	
x	f_{xy}	$f_x - f_{xy}$	f_x
\bar{x}	$uf_y - f_{xy}$	$N - f_x - uf_y + f_{xy}$	$N - f_x$
	uf_y	$N - uf_y$	N

	y	\bar{y}	
x	10	90	100
\bar{x}	3,990	95,910	99,900
	4,000	96,000	100,000

表 4 の具体例は、表 1 の具体例で y のみ $u = 10$ に変化した分割表で、 t スコアは 1.90、相互情報量は 1.32 となる。

t スコアの変化は次のようになる。

$$t'(x, y) - t(x, y) = \frac{1}{\sqrt{f_{xy}}} \left(f_{xy} - \frac{f_x \cdot uf_y}{N} \right) - \frac{1}{\sqrt{f_{xy}}} \left(f_{xy} - \frac{f_x f_y}{N} \right) = \frac{f_x f_y}{N \sqrt{f_{xy}}} (1 - u)$$

$$= -Eu + E$$

ここで、 $E = f_x f_y / (N \sqrt{f_{xy}})$ で常に正なので、頻度が増加すると t スコアは単調に減少する。

相互情報量は次の通りで、(2)と同じ形となり、 u に対して単調に減少することがわかる。

$$PMI'(x, y) - PMI(x, y) = \log_2 \frac{N f_{xy}}{f_x \cdot uf_y} - \log_2 \frac{N f_{xy}}{f_x f_y} = -\log_2 u$$

これらは、共起語が高頻度語か低頻度語かで、共起指標がどう変化するかを示すものであった。状況が曖昧なまま、「頻度が高いと、 t スコアも高くなる」といった先入観があると、高頻度の共起語でも t スコアは一般に高くなるように考えてしまうかもしれない。しかし、まったく逆の振る舞いをする。この状況下では両指標ともに、低頻度語との共起をより高く評価する。

4. おわりに

本発表は、代表的な共起指標である t スコアと相互情報量における頻度効果を、事前の出現状況を規定した上で、代数操作に基づいて明らかにした。管見の限りでは、現在、 t スコアと相互情報量に対する最も精確で、丁寧な資料は大名(2011)である。両指標をさまざまな角度からとらえ直すことで、より見通しが良くなり、理解が進むはずである。なお、対数尤度比など他の多くの共起指標でも同様に議論ができ、別稿にて報告予定である。

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.

大名力(2011)「MI-score, t-score 再考」英語コーパス学会第 37 回大会。

1 共起指標としての相互情報量は、正確には自己相互情報量と呼ぶべきものである。本稿ではコーパス研究での一般的な呼び方にならって、相互情報量で統一している。

The KIT Speaking Test Corpus 専用検索インターフェースの開発

田中 悠介(熊本学園大学)

瀬戸口 彩花(都城市立明和小学校)

近 大志(富山大学)

神澤 克徳(京都工芸繊維大学)

{yusuke.tanaka.07, guchi.a.chan7, nanou7614, rnkp43470}@gmail.com

Development of a Search Interface for the KIT Speaking Test Corpus

TANAKA Yusuke (Kumamoto Gakuen University)

SETOGUCHI Ayaka (Meiwa Elementary School, Miyakonojo)

CHIKA Taishi (University of Toyama)

KANZAWA Katsunori (Kyoto Institute of Technology)

Abstract

We developed a search interface for the KIT Speaking Test Corpus (KISTEC) to enhance its accessibility and practical value. While KISTEC represents a valuable resource for language research and education, it has not yet achieved widespread recognition or use. To promote its adoption among researchers, educators, and learners, a user-friendly and versatile search interface is essential. This paper introduces the design and key features of the KISTEC Search Interface, highlighting its five core functions: KWIC search, collocation search, wordlist generation, keyword extraction, and N-gram analysis. By integrating these tools, the interface not only facilitates access to the corpus but also encourages broader engagement and supports research in linguistics and language education.

Keywords

KIT Speaking Test Corpus; Learner Corpus; Spoken English; Search Interface

1. はじめに

著者らは、京都工芸繊維大学で開発・実施された英語スピーキングテストである KIT Speaking Test の解答音声を用いて The KIT Speaking Test Corpus (以下, KISTEC) を構築し, Web 上で公開している (Kanzawa et al., n.d.). 日本人が受験した英語スピーキングテストの解答音声を書き起こしたコーパスは他にも複数存在するが (Abe & Kondo, 2019; Ishikawa, 2023; 和泉他, 2004), KISTEC は各回答音声に 2 種類のスコアが付与されているといった独自の特徴を有して

いる。こうした特徴により、特定の要素が評価に及ぼす影響を詳細に検討するなど、従来は困難であった分析が可能となる。実際、KISTEC はすでに、学習者の非流暢性現象を自動検出する BERT モデルの性能評価 (Skidmore & Moore, 2023) といった研究に活用されている。

このように KISTEC は有用であるものの、一般への普及は十分ではない。KISTEC を研究者のみならず教育関係者や学習者にも広く活用してもらうためには、ユーザーフレンドリーな検索インターフェースの開発が不可欠である。本稿では、著者らが KISTEC のために開発した検索インターフェース (Tanaka et al., n.d.) の機能について紹介する。

2. KISTEC の概要

検索インターフェースの紹介に先立ち、KISTEC の概要を述べる。KISTEC は、京都工芸繊維大学で開発・実施された英語スピーキングテストである KIT Speaking Test の解答音声を書き起こしたものである。2018 年に、当時京都工芸繊維大学の 1 年生だった 575 人が受験しており、合計で約 75 時間分の回答音声の書き起こしが収録されている。

KIT Speaking Test は 3 つのバージョンがあり、それぞれ 9 問から構成されている。採点はタスク達成度とタスク伝達度という 2 つの観点から行われた。前者は問題が要求するタスクをどの程度達成できたか (Task Achievement: TA)、後者は回答をどの程度効果的に伝えられたか (Task Delivery: TD) を評価する。また、これらのスコアに加えて、受験者のジェンダーや国籍、TOEIC スコアなど、多様な情報が付与されている。

書き起こしには、フィルタータグ <F></F> や繰り返しタグ <R></R> を含む 16 種類のタグが付与されている。これにより、例えばフィルターや繰り返しといった非流暢現象とスピーキングテストのスコアとの関係を分析することができる。実際に、Tanaka et al. (2025) はフィルターの使用頻度が流暢性の評価 (TD スコア) に与える影響を詳細に分析し、両者の関係は線形的なものではなく、適度なフィルターの使用がむしろ流暢性の評価を高めることを明らかにしている。

KISTEC は現在、Web サイト (<https://kitstcorpus.jp>) から誰でもテキストファイル形式で入手可能である。しかし、この配布形式では利用者に一定のデータ処理スキルが求められる。そうしたスキルを持たない研究者や教育関係者、学習者でも容易に活用できるように、著者らは KISTEC 専用の検索インターフェースを開発した。次節では、その概要を述べる。

3. 検索インターフェースの機能

本節では、KISTEC 専用検索インターフェースの基本機能 (KWIC 検索, コロケーション検索, Wordlist, 特徴語抽出, N-gram) と、これらに付随するフィルタリング機能について紹介する。

3.1 KWIC 検索

KWIC (Key Word in Context) 検索は、検索語を含む行を前後の文脈とともに表示する機能である。表示結果を並べ替えることで、当該語の直前・直後に現れる語や表現パターンを直感的に把握できる。例えば、図 1 は at の検索結果である。Left Context には at の直前に出現する内

容, Right Context には at の直後に続く内容が表示され, 学習者が at をどのような文脈や構造で使用しているかを容易に確認できる。

図 1 KWIC の検索結果の例

KWIC Search Results (404 hits)

TA	TD	Left Context	Keyword	Right Context
1.5	1.0	I can't understand this conversation, but I eh will I eh support eh Bill. Eh because Bill is man. Eh I now not good	at	. Ah no I'm I'm eh I'm dislike woman so oh ou wakarane. I'm sorry.
1.0	1.5	Kate cough Kate cough loss lose it homework data yesterday. Kate is Kate report submit a report	at	tomorrow. Kate cough cough .
2.0	1.0	If I were Susan, I want to uh do I want study architecture. And, eh	at	the same time, I I do dance, and study very hard. And.
2.0	1.5	The man who has ah the bycicle he ah left by ah shopping mall as by	at	the shopping mall. So he so he left this that so he left that bike. uh

タグの表示有無は選択可能であり(図 1 はタグを非表示にした例), 表示順を TA スコアまたは TD スコアに基づいて並べ替えることも可能である。これにより, スコアの高い学習者と低い学習者とで, 検索語の使用傾向がどのように異なるかを直感的に把握できる。

3.2 コロケーション検索

コロケーション検索は, 特定の語がどのような語と共起しやすいかを把握するのに有用である。例えば, 図 2 は at の検索結果である。

図 2 コロケーションの検索結果の例

Collocation Analysis Results for 'at' - Frequency

Left 3	Left 2	Left 1	Keyword	Right 1	Right 2	Right 3
is (21)	is (27)	good (44)	at	the (84)	i (41)	i (25)
he (18)	not (18)	look (11)	at	first (63)	and (26)	so (22)
i (13)	the (15)	eh (11)	at	a (14)	so (12)	and (21)

図 2 では, Left 1 の最上段に good (44) と表示されている。これは, at の直前に最も高頻度で現れる語が good であり, その正規化頻度が 44 件であることを示している。この結果から, 学習者

は **good at** というコロケーションを頻繁に用いていることがわかる。

また、表示指標として **t-score** を選択することも可能である。これは、特定の語と語が偶然ではなく統計的に有意な結びつきを持つかどうかを判断するための指標であり、頻度だけでなく分布の偏りも反映する。これにより、生起頻度が低くても強い結びつきを持つ語を抽出できるため、高頻度語に偏らない共起分析が可能となる。特に、頻度の高い語だけでは見落としがちな学習者特有の表現や誤用傾向を把握する際に有用である。

3.3 語彙リスト

語彙リスト機能では、KISTECに含まれる語彙とその出現頻度を頻度順に一覧表示できる。図3からは、KISTECにおいて最も頻出する語は **I**、次いで **is** であることが確認できる。

図3 語彙リストの一覧表示の例

Word Frequency List

Word	Raw Frequency
i	13109
is	10908
to	10048
and	9538
the	8649

3.4 特徴語抽出

特徴語抽出とは、複数の文書を比較し、特定の文書群で顕著に使用される語彙を識別する機能である。この分析により、特定のスコア帯に共通する言語的特徴や言語使用の傾向を明らかにできる。本インターフェースでは、特徴語を抽出するために **TF-IDF** スコアを算出する機能を備えている。**TF-IDF** スコアは、単語の重要度を評価する統計的指標であり、単語が文書に現れる頻度 (**Term Frequency: TF**) と、その単語がどの程度希少または一般的であるかを示す逆文書頻度 (**Inverse Document Frequency: IDF**) とを組み合わせて算出される。**TF** が高い場合は、その語が特定の文書内で頻繁に出現していることを意味し、**IDF** が高い場合は、その語が全体的に希少であることを意味する。したがって、**TF** と **IDF** の双方が高い語は、希少でありながら特定の文書群で頻繁に使用されるため、特徴語として識別される。

図4は、得点群を「低得点群(0-39点)」「中得点群(40-70点)」「高得点群(71-100点)」と定義した場合における、低得点群の特徴語抽出結果を示している。低得点群の特徴として、フィルターの多用が確認される。得点群の区分は任意に設定可能である。

図 4 特徴語の検索結果の例

Characteristic Words Results for 'low' Score Category

Word	TF-IDF Score
eh	99.12
uh	76.4
ah	60.59
want	46.61

3.5 N-gram

N-gram とは、連続する n 個の単語で構成される語の並び(シーケンス)を指す。例えば、図 5 は 3-gram の検索結果であり、これを見ると最も使用頻度の高い 3 語のシーケンスは **I want to** であることがわかる。N-gram 分析は、学習者が頻繁に用いる定型表現や文型パターンを特定するのに有効である。

図 5 N-gram の検索結果の例

N-gram Analysis Results

N-gram	Frequency
i want to	1352
a lot of	386
take care of	373
if i were	318

3.6 フィルタリング機能

本インターフェースでは、問題、ジェンダー、各種スコアに基づくフィルタリングが可能である。例えば、検索対象を「女性」「TA スコアが 2.5 以上」「TOEIC スコアが 400 以上」といった条件で絞り込むことができる(図 6)。この機能を活用すれば、例えば語彙リスト機能において高得点群と低得点群の語彙使用傾向を比較し、高得点群が頻繁に用いる語彙や表現を抽出できる。

図 6 フィルタリング機能

TOTAL	TA	TD	TOEIC	TOEIC R	TOEIC L
Min	Min	Min	Min	Min	Min
Max	Max	Max	Max	Max	Max

Gender: Male Female Questions: Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9

4. おわりに

本稿では、著者らが KISTEC のために開発した検索インターフェースの機能を紹介した。本インターフェースには、KWIC 検索、コロケーション検索、Wordlist、特徴語抽出、N-gram の 5 つの基本機能があり、各種スコアやジェンダー、問題によるフィルタリングも可能である。本インターフェースの活用により、KISTEC のデータをより容易に利用できると期待される。

最後に、今後の展望について述べる。現状の機能は有用である一方、効果的に活用するには一定の知識や経験が依然として求められる。例えば、KWIC 検索の結果から共起パターンを直感的に理解するには、ある程度の慣れや背景知識が必要である。さらに利用しやすいインターフェースにするには、検索結果をグラフやチャートで視覚的に提示するといった工夫や拡張が必要である。

謝辞

本研究は JSPS 科研費 22K00736 の助成を受けたものです。

引用文献

- Abe, M., & Kondo, Y. (2019). Constructing a longitudinal learner corpus to track L2 spoken English. *Journal of Modern Languages*, 29, 23–44.
- Ishikawa, S. (2023). *The ICNALE Guide: An Introduction to a Learner Corpus Study on Asian Learners' L2 English*. Routledge.
- Kanzawa, K., Kobayashi, Y., Lee, J., Mitsunaga, H., Mori, M., Tanaka, Y., & Chika, T. (n.d.). The KIT Speaking Test Corpus (KISTEC). <https://kitstcorpus.jp>
- Skidmore, L., & Moore, R. (2023). BERT models for disfluency detection of spoken learner English. *Proceedings of 9th Workshop on Speech and Language Technology in Education (SLaTE)*, 91–92.
- Tanaka, Y., Setoguchi, A., Chika, T., & Kanzawa, K. (n.d.). Search Interface for the KIT Speaking Test Corpus. <https://www.kistecsearch.org/>
- Tanaka, Y., Setoguchi, A., Chika, T., & Kanzawa, K. (2025). Moderate use of fillers can enhance fluency assessments. In Lacy, B., Lege, R. P., & Swanson, M. (eds.), *Moving JALT into the Future: Opportunity, Diversity, and Excellence*. JALT.

和泉絵美・内元清貴・井佐原均 (2004). 『日本人 1200 人の英語スピーキングコーパス』 ア
ルク.

学習者コーパスに基づく英語対話能力の縦断的研究

望月 圭子 (東京外国語大学)

mkeiko@tufs.ac.jp

A longitudinal study of English conversational ability based on a learner corpus

Mochizuki Keiko (Tokyo University of Foreign Studies University)

Abstract

This presentation discusses the growth of English conversation skills among two English learners from their first year of high school to their fourth year of university.

The two English learners engaged in one-on-one English conversation with a native English speaker once a month for one hour from their first year of university to their fourth year. The analysis results showed that both learners exhibited an increase in vocabulary usage at the CEFR A2–B2 level and growth in complexity from the second year of university onwards. However, a decline in accuracy was observed. This decline in accuracy may be attributed to the use of more difficult vocabulary, which may have compromised accuracy. Fluency showed no growth, while a negative correlation was observed between accuracy and complexity, indicating that as complexity increased, accuracy was sacrificed.

Keywords

英語対話力・縦断的研究・複雑性・正確性・流暢性・CEFR-J Wordlist

1. 研究目的

本発表の目的は、2名の英語学習者を対象に、高校1年次時の2018年11月から大学2年次の2023年1月までの英語対話において、使用語彙の産出が、どのように成長するかを縦断的に分析することにある。分析対象の2名の学習者は、J003、J029と表記する。

2名の学習者の対話に産出された語彙を、CEFR-J Wordlistを用いて語彙のレベルをA1～B2に分類すると、以下の表1に示すとおりとなる。

表1 分析対象の英語対話のトピック

トピック	高校1年-3年次	大学1年-3年次
Lesson 1(自己紹介)	2018年11月	2021年11月
Lesson 10(ホームステイの経験)	2019年8月	2022年3月
Lesson 14(食と健康)	2019年12月	2022年7月
Lesson 20(AIとの共存)	2020年5月	2023年1月

2. 結果

学習者 J003 は、以下の表2に示すように、高校時代と比較して、大学生時代は A1 レベル語彙の割合が減少し、A2～B2 レベル語彙の割合が増加していることが観察される。

表2

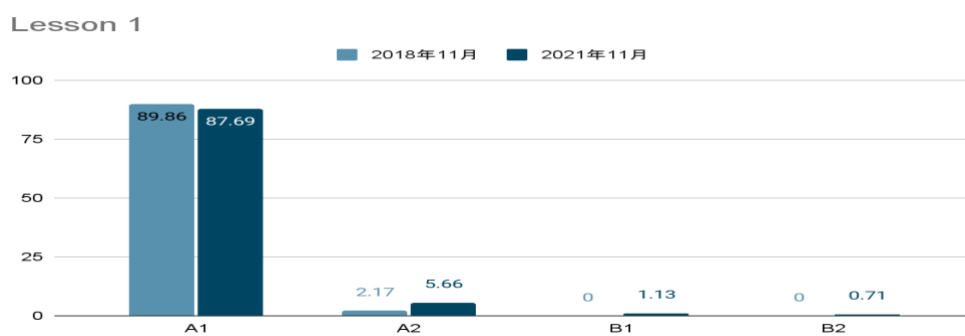


表3

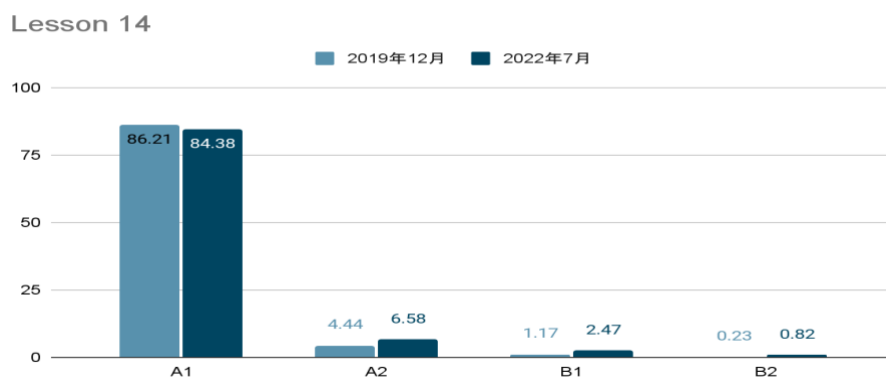
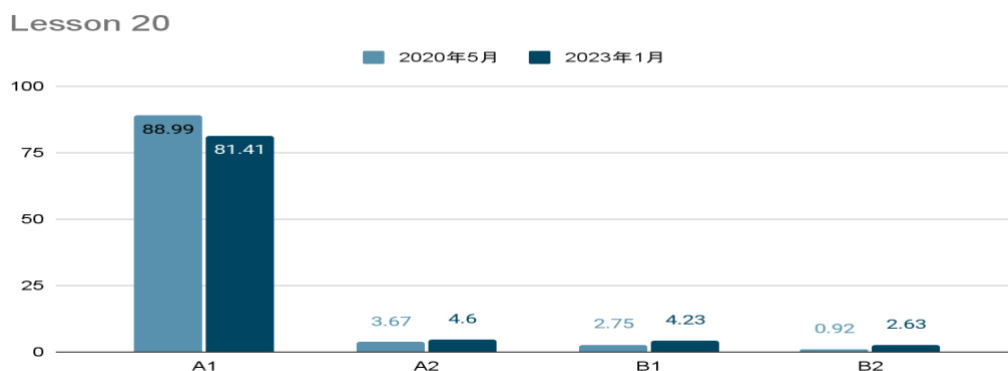


表4



次に、学習者 J029 も、対話レッスンを継続するにつれて、A2～B2 レベルの語彙割合の増加が観察される。以下の表 5～8 は結果を高校生時代と大学生時代別に整理したものである。

表 5 J003 高校1年 2018 年 11 月から 2020 年 5 月 高校3年

時期	AS-Unit あたりの節数	AS-unit あたりの token 数	従属節の数
2018 年 11 月	0.45	2.74	2
2019 年 8 月	0.5	3.26	4
2019 年 12 月	0.41	3.67	4
2020 年 5 月	0.6	4.48	3

表 6 J003 大学生時代

時期	AS-Unit あたりの節数	AS-unit あたりの token 数	従属節の数
2021 年 11 月	0.50	4.53	5
2022 年 3 月	0.46	6.17	6
2022 年 7 月	0.62	6.44	14
2023 年 1 月	0.77	6.90	14

表 7 J029 高校生時代

時期	AS-Unit あたりの節数	AS-unit あたりの token 数	従属節の数
2018 年 11 月	0.40	3.11	3
2019 年 8 月	0.17	2.12	2
2019 年 12 月	0.40	3.87	0
2020 年 5 月	0.45	3.89	6

表 8 J029 大学生時代

時期	AS-Unit あたりの節数	AS-unit あたりの token 数	従属節の数
2021 年 11 月	0.32	2.49	3
2022 年 3 月	0.42	4.10	6
2022 年 7 月	0.63	4.77	5
2023 年 1 月	0.63	5.17	13

J003、J029 とも全ての指標で成長が見られ、特に AS-Unit あたりの token 数と従属節の数を分析すると、複雑性が成長している。正確性の成長については、紙幅の関係でデータは割愛するが、J003・J029 ともに、誤りのある語数の割合は増加と減少を繰り返し、スピーキングレッスンの継続年数と正確性の発達の間には相関性が観察されない。

流暢性については、J003 の流暢性を示す項目の数値の増減は一定ではなく、流暢性が発達したと一概には言い難い。J029 に関しても、流暢性の発達は確認できない。

3. 結論

J003、J029 ともに、大学生時代の英語対話産出において、A2～B2 レベルの語彙使用が増加し、複雑性に成長が見られたが、正確性は低下傾向にあった。これは、難易度の高い語彙を使用することで正確性が損なわれた可能性がある。第三に、流暢性を示す指標は一貫した変化が観察できなかったことから、流暢性は成長しなかったと言える。正確性と複雑性には負の相関関係が観察され、難易度の高い語を使用することで、複雑性が成長した一方、正確性が犠牲になったと考えられる。

参考資料

『CEFR-J Wordlist Version 1.6』東京外国語大学投野由紀夫研究室. (URL: <https://cefr-j.org/download.html>)
https://www.arcle.jp/research/edu_english/data/pdf/pdf1_202208.pdf#:~:text=%E6%97%A5%E6%9C%AC%E4%BA%BA%E9%AB%98%E6%A0%A1%E7%94%9F

謝辞

本研究は JSPS 科研費 22H00076 の助成を受けたものです。

『英語コーパス学会大会予稿集 2025』(ISSN 2436-6647)

Proceedings of the JA ECS Conference 2025

刊行日 2025年9月2日

発行所 英語コーパス学会

事務局 〒464-8601 名古屋市千種区不老町 名古屋大学言語教育センター 小島ますみ研究室気付

e-mail: jaecs.hq@gmail.com URL: <https://jaecs.com/>