

英語学習者の産出語彙を評価する 語彙の豊かさ指標の信頼性比較

小島ますみ (岐阜市立女子短期大学)

キーワード: 語彙の豊かさ、頻度表、LFP, P_Lex, Advanced Guiraud, S

1. はじめに

語彙の豊かさ指標とは、学習者の産出する語彙が、いかに多様で広範なものを量的に示す指標である (Laufer & Nation, 1995)。前提として、学習者の持つ語彙の広さが、産出される語彙に反映されると考えられている (Laufer & Nation, 1995; Meara & Bell, 2001)。語彙の豊かさを量的に表す指標は、母語習得、第二言語習得、医学臨床等の分野で、言語の発達指標として用いられてきた (Malvern & Richards, 1997)。近年、語彙の一般的な使用頻度に基づき、学習者の語彙の豊かさを測定する指標が提案されている (e.g. Lexical Frequency Profile, Beyond 2000, P_Lex, Advanced Guiraud, S)。語彙の頻度は、語彙の習得順序を決める重要な要因の 1 つであると考えられているため (Nation & Beglar, 2007)、一般的な頻度に基づく語彙の豊かさ指標は、タイプとトークンのみに基づく語彙の多様性指標よりも、学習者のメンタルレキシコンをより良く反映できることが期待されている。しかしながら、語彙の一般的な使用頻度に基づき学習者の語彙の豊かさを測定する試みはまだ新しく、スコアがテキストの長さの影響を受ける、明らかに熟達度の異なる学習者を区別しないなど、問題点も指摘されている (e.g. Meara & Bell, 2001; Meara, 2005)。

本研究では、語彙の一般的な使用頻度に基づき英語学習者の語彙の豊かさを測る 4 つの指標 (S, Beyond 2000, P_Lex, Advanced Guiraud) の信頼性を調査し、結果を比較・考察した。分析対象としたデータは、日本人大学生・大学院生が書いた 38 の英文エッセイであり、それぞれのエッセイを 5 つのセグメント (150 語×2, 200 語×2, 全体) に分割して、それぞれのセグメントにおける 4 つの指標のスコアを求めた。結果より、S (小島, 2010) が 150 語～400 語程度の短いテキストでも最も信頼性が高く、テキストの長さの影響も受けにくいことが示された。

2. 先行研究

まず, Laufer & Nation (1995) の提案する Lexical Frequency Profile (LFP) について説明する。この指標では、学習者の使用語彙を 4 つのレベル (1,000 語レベル, 2,000 語レベル, 学術語彙レベル, その他) に分類し、全体に対するそれぞれのカバー率をパーセントで表す。従って, LFP のスコアは 86%-7%-4%-3% のように表される。LFP は 4 つの数値がスコアとなり、統計処理等で扱いにくいことから、Laufer (1995) は、LFP の簡易バージョンである Beyond 2000 を提案している。これは、「学術語彙」と「その他」の語彙が、テキスト全体の語彙の何パーセントを占めるかで表される。上記の例で言えば、Beyond 2000 のスコアは 7% となる。

LFP や Beyond 2000 は、英語学習者のライティングやスピーキングにおける語彙の豊かさ指標として、様々な研究で用いられてきた (e.g. Laufer, 1998; Laufer & Paribakht, 1998)。しかし、スコアの信頼性についてはほとんど明らかになっていない。Nation (2008) は、LFP や Beyond 2000 を使用するためには、テキストの長さは 300 語以上で、すべてのテキストは同じ長さに揃えなければならないと述べている。しかし Nation は、どのようなデータをどのような方法で分析し、このようなテキストの長さに関する結論に至ったかを明らかにしていない。また、300 語のテキストで指標の信頼性がどのくらい高いのかも不明である。

Meara & Bell (2001) は、LFP・Beyond 2000 よりもテキストの長さの影響を受けにくい語彙の豊かさ指標を目指し、P_Lex を開発した。低頻度語の出現確率はポアソン分布に従うとし、ポアソ

ンカーブを表す λ 値を P_Lex のスコアとするものである。この指標の前提には、メンタルレキシコンのより発達した学習者は、より高い確率で低頻度語を産出するという想定がある。Meara & Bell (2001) は、P_Lex の λ は 150 語以上で安定すると述べているが、3 人の被験者の λ 値が語数の増加に伴いどのように変化するかグラフで表したのみで、他の被験者については明らかにしていない。また、150 語のテキストにおけるスコアの信頼性がどのくらい高いのかも不明である。

Daller, van Hout & Treffers-Daller (2003) は、詳細な頻度表が無い言語でも基本語彙表さえあれば利用可能な指標として、Advanced Guiraud を提案した。これは Guiraud Index を応用したもので、低頻度語（基本語彙表以外の語）のタイプ数をトークンの平方根で割ることでスコアを算出する。Daller らは、スコアにおけるテキストの長さの影響や、スコアの信頼性については報告していないが、定義式がトークンの関数であるため、テキストの長さの影響は避けられないと思われる。

小島 (2010) は、より妥当性の高い語彙の豊かさ指標を目指し、S を開発した。S は、学習者の産出したテキストから得られる語彙の頻度別累積カバー率をデータとし、データに最も近似するモデルを求め、累積カバー率が 100% に達する時点での語彙頻度順位を推定する。この推定値が S のスコアとなる。

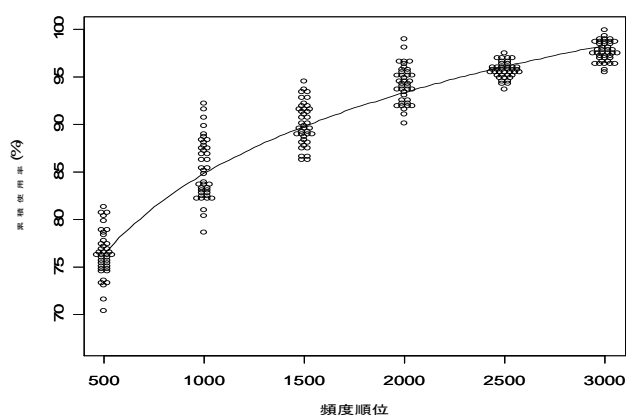


図 1 データに最も適合する頻度別累積カバー率グラフ例 (S=3,429)

小島(2010) は、S が熟達度の異なる学習者グループを区別し、200-600 語程度のデータにおいて、テキストの長さの影響を受けず、安定した結果を返すことを示した。しかし、200 語程度のテキストにおいて、S の信頼性がどのくらい高いのかという客観的な数値は示していない。

3. 本研究

3. 1. 研究目的とリサーチ・クエスチョン

本研究の目的は、一般的な頻度に基づく 4 つの語彙の豊かさの指標 (S, Beyond 2000, P_Lex, Advanced Guiraud) の信頼性を比較・検討することである。本研究では、以下のリサーチ・クエスチョンに取り組む。

1. 4 つの指標は、テキストの長さに依存しないか。
2. 150 語、200 語のテキストにおける 4 つの指標の信頼性はどの程度高いか。
3. 4 つの指標は、信頼性係数 .7 や .8 を得るために、どの程度のテキストの長さを必要とするか。

3. 2. 分析データ

Nagoya Interlanguage Corpus of English (NICE) (杉浦, 2008) 2.0 より抽出した 400 語以上の英文エッセイ 38 を対象とした。これらのエッセイは、名古屋市内の日本人大学生・大学院生が、辞書等の参照なしに 1 時間で書いたものである。参加者の専攻はさまざまであった。

3. 3. 分析方法

コンピュータプログラムで4つの指標のスコアを算出する前に、1つ1つのエッセイを実際読み、スペリングミスの修正を行い、固有名詞は除外した。また、英語ではない単語（例：*nenkin, arbiter*）や、明らかに学習者の意図に合致しないと思われる語（例：*bench* を *bank* と間違えているなど）については、語彙的エラーとして除外した。スペリングミスと語彙的エラーのチェックについては、調査で対象とする学習者のエッセイ 38 中 15 (39.5%) について、調査者ともう1人の評定者(英語教員・修士)により行った。評価者間信頼性を調べたところ、カッパ係数 0.85 であった。

38 のエッセイそれぞれについて、5つのセグメントにおける4つの指標のスコアを算出した。5つのセグメントとは、エッセイ最初の150語と2番目の150語、エッセイ最初の200語と2番目の200語、エッセイ全体である。語彙の頻度表として、BNC 話し言葉に基づく頻度表 (Nation & Beglar, 2007) を使用し、4つの指標のスコアを算出した。

4. 結果と考察

表1は、150語、200語、エッセイ全体におけるそれぞれの指標の平均スコアと標準偏差である。150語と200語のスコアは、2つのセグメントで求めたスコアの平均値を記載した。

表1 長さの異なるセグメントにおけるそれぞれの指標の平均スコアと標準偏差

セグメントの長さ	S (SD)		Beyond 2000 (SD)		P_Lex (SD)		AG (SD)	
150 語	2355.7	(996.5)	5.84	(2.72)	0.850	(0.380)	.392	(.206)
200 語	2331.0	(936.7)	5.67	(2.47)	0.854	(0.415)	.421	(.219)
全体	2285.7	(894.3)	6.77	(2.81)	0.853	(0.393)	.548	(.295)

それぞれの指標に対して、表1の3つスコアをフリードマン検定で比較したところ、SとP_Lexでは有意差がなかったのに対し（それぞれ、 $p < .28$, $p < .73$ ）、Beyond 2000とAdvanced Guiraudでは有意差が確認された（いずれも、 $p < .001$ ）。多重比較の結果、Beyond 2000では、150語と全体、200語と全体の間に有意差が確認され、Advanced Guiraudでは、3つの長さの異なるセグメント間すべてに有意差が確認された（すべて $p < .001$ ）。SとP_Lexは、150～400語程度の学習者のテキストにおいて、テキストの長さの影響を受けにくいのに対し、Beyond 2000とAdvanced Guiraudはテキストの長さの影響を受け、テキストの長さが長くなるほどスコアが上昇する性質が示された。特に、Advanced Guiraudにおけるテキストの長さの影響が顕著であった。

表2は、それぞれの指標に対して、同じ長さのセグメントにおける2つのスコアのスピアマン順位相関係数を求めたものである。SとP_Lexでは有意な相関が確認されたのに対し、Advanced Guiraudでは、200語のセグメント間でのみ有意な相関がみられ、150語のセグメント間では有意な相関が見られなかった。さらにBeyond 2000では、2つのセグメント間ともに有意な相関は見られなかった。150語、200語程度の短いテキストにおいて、Sの信頼性が最も高く、続いてP_Lex, Advanced Guiraudであり、Beyond 2000の信頼性が最も低い結果となった。

表2 同じ長さのセグメントにおける2つのスコアのスピアマン相関係数

指標	150 語	200 語
S	.622**	.689**
Beyond 2000	.161	.305
P_Lex	.521**	.646**
Advanced Guiraud	.261	.474**

** $p < .01$

表3は、スピアマン・ブラウンの予言公式に上記150語と200語の相関係数を当てはめ、信頼性係数.70と.80を得るために必要なテキストの長さ（総語数）を推定したものである。

表3 信頼性係数.70 と.80 を得るために必要なテキストの長さ (総語数)

指標	信頼性係数	
	.70	.80
<i>150 語に基づき推定</i>		
S	212	365
Beyond 2000	1824	4168
P_Lex	323	552
Advanced Guiraud	992	1700
<i>200 語に基づき推定</i>		
S	210	362
Beyond 2000	1064	1822
P_Lex	256	438
Advanced Guiraud	518	888

表3より、同じレベルの信頼性を得るために、Sでは必要とするテキストの長さが最も短く、続いてP_Lex, Advanced Guiraud, Beyond 2000であることが分かる。

5. 結論

本研究は、語彙の一般的な使用頻度に基づき、学習者の語彙の豊かさを測定する4つの指標(S, Beyond 2000, P_Lex, Advanced Guiraud)の信頼性を比較した。結果より、本研究で調査した150語～400語の範囲で、SとP_Lexはテキストの長さの影響を受けにくいのに対し、Beyond 2000とAdvanced Guiraudはテキストの長さの影響を受け、テキストの長さが長くなるほど、スコアが上昇する傾向が見られた。150語～200語程度の短いテキストでもSとP_Lexは比較的安定していたのに対し、Beyond 2000とAdvanced Guiraudは信頼性が低い結果となった。同レベルの信頼性を得るために必要なテキストの長さは、Sで最も短い結果となり、続いてP_Lex, Advanced Guiraud, Beyond 2000であった。熟達度の低い英語学習者から長いテキストを引き出すことは大変困難であるため、Sはより幅広い熟達度の学習者に適用できる語彙の豊かさ指標であることが示された。

6. 主な引用文献

- Daller, H., van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in spontaneous speech of bilinguals. *Applied Linguistics*, 24 (2), 197-222.
- 小島ますみ(2010)新しい lexical richness 指標 S の提案:学習者の産出語彙頻度レベルの推定 『英語コーパス研究』第17号, 1-16.
- Laufer, B. (1995). Beyond 2000. A measure of productive lexicon in a second language. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage (Studies in Honor of William E. Rutherford)* (pp. 265-272). Amsterdam and Philadelphia, PA: John Benjamins.
- Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16 (3), 307-322.
- Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16 (3), 5-19.
- 杉浦正利 (2008). 『英語学習者のコロケーション知識に関する基礎的研究』平成17年度～19年度科学研究費補助金(基盤研究(B)) 研究成果報告書(課題番号 17320084)