

杉森直樹 (2010). 「V. 語彙力の測定・評価」. 石川祥一・西田正・斉田智里 (編著). 『テストイングと評価 4 技能の測定から大学入試まで』 (237–252 頁). 東京: 大修館書店.

第7章5節 語彙力の測定・評価

* 章構成は発表者が改変した。

はじめに

- 語彙力や語彙知識は英語運用能力の基盤を成す要素の一つであり、学習者の英語熟達度を判断する手段の一つになると考えられる。そのため、語彙力の構成要素を正しく分析し、それに基づいて学習者の語彙力を的確に測定・評価する語彙テストの開発は重要な課題であると言える。
- 本節では、語彙力の構成概念について述べた後、語彙知識の広さ・深さを評価する方法を述べる。最後に、語彙テストの今後の課題とテスト開発の留意点について述べる。

5. 語彙力の構成概念

- 語彙力とは多次的 (multi-dimensional) な構成概念から成り立っており、「意味を知っている」というような一面的な能力ではないと考えられている。
- 語彙知識の構成概念として、これまで (1) 広さ・深さ、(2) サイズ・構成、(3) 広さ・深さ・流暢さ、などが提案されてきた。最もよく知られている構成概念の定義は Nation (2001) のものである (p. 238)。
- これらすべての側面を測定・評価することは難しいが、語彙力の構成要素はお互いが密接に関連しているため、できる限り多くの側面をテストして全体的な語彙知識を評価することが必要になる。

5.1 受容語彙の広さ (サイズ) の測定・評価

- 一般的なテキストに含まれる語彙の 80% を認識するには、2,000 word families のサイズが必要になる (Milton, 2009)。このような語彙サイズを測定する際には、頻度情報に基づく語彙リストを作成し、その中から語彙を選定してテストに出題することが行われている。
- そのため、テスト作成の基準となる語彙頻度リストは信頼できるコーパスデータに基づいている必要があり、出題語彙の選定方法がテストの使用目的に対して適切かという妥当性の観点も考える必要がある。

5.1.1 The Vocabulary Levels Test

- Nation の The Vocabulary Levels Test (図 1, p. 240) は General Service List と University Word List を基に作成された語彙サイズテストで、2,000 語、3,000 語、5,000 語、academic、10,000 語レベルから成る。
- また、初級学習者 (バイリンガル版) 用に、2,000 語レベルまでのテストもある (Nation, 2004)。
- 各レベルは 30 問から成り、各レベルの正答率を使って簡易的に学習者の語彙サイズを推定することができる。また Nation (1990) には語彙レベルごとの学習ガイドが示されている。
- ただし、General Service List の古さや出題語彙のサンプリング数と選定方法に問題があること、および出題形式の問題から能力診断テスト (adaptive test) にするためには局所独立性が成り立たないという問題がある。

5.1.2 The Vocabulary Size Test

- Nation and Beglar (2007) は 1,000 word family を 1 レベルとして、全 14 レベル (14,000 word families まで測定可能) から成る The Vocabulary Size Test を開発した (British National Corpus に基づく)。

- ga それぞれのレベルは 10 問の多肢選択肢が設定されており、文内で目標語が提示される形式になっている (図 2, p. 242)。選択肢は GSL の 1,000 word family レベルの語で作成されている。正答数に 100 を掛けることで推定語彙サイズが得られる (バイリンガル版もあり)。

5.1.3 Check List

- Check List (Yes/No) 形式の語彙テストは、語彙項目をリストで提示し、それらの語彙を知っているかどうかをチェックさせて答えさせようとするものである。しかし、知らない単語でも偽りや錯誤により知っている単語であると解答することが可能なため、信頼性・妥当性に問題があるとされてきた。
- そのため、提示する項目に疑似語を含め、誤って疑似語を選択した場合には減点する方式を取ることでより正しく語彙サイズを推定できる (Anderson & Freebody, 1981)。
- ESL/EFL 用には、Eurocenter's Vocabulary Size Test 10KA (EVST) があり、コンピュータで受験できる (10 分程度)。出題される単語は 10,000 語の頻度リストから選定されており、疑似語が約 1/3 含まれている。すぐに結果が分かり、結果の解釈も容易なため学習者のプレイスメントとして機能する。

5.1.4 望月テスト

- 日本人学習者を対象とした語彙サイズテストとして望月テストがある。出題される語彙項目は北大語彙表からで、1,000 語～7,000 語までのレベルがある (lemma 換算)。最新版のテストでは 1 レベル 26 問から成り、version も 3 つ用意されている。
- 望月テストは日本人英語学習者にとって意味が推測しやすい英単語が選択肢として出題されることや、日-英をマッチングさせる形式であるため、推定語彙サイズは多めに算出される傾向にある。

5.2 発表語彙の測定・評価

- 英語学習者の受容語彙と発表語彙の関係: (1) 受容語彙 > 発表語彙、(2) 受容語彙と発表語彙の比率は常に一定ではない、(3) 低頻度語になるほど受容語彙と発表語彙とのギャップが大きい、(4) 高頻度語になるほど両者のギャップは小さくなる、(5) テストで直接測定された語彙サイズの増加は、実際に使える語彙知識の増加と必ずしも一致しない。

5.2.1 The Productive Vocabulary Levels Test

- 学習者の発表語彙を測定するテストには Productive Vocabulary Test (Laufer & Nation, 1999) がある。このテストは文脈情報を与えて目標語のスペリングの一部をあらかじめ提示する形式をとっており (統制的発表語彙: controlled productive vocabulary)、Vocabulary Levels Test のオリジナル版に基づいて作成されている (p. 244)。
- 各レベルはそれぞれ 18 問で構成されており (計 90 問)、各レベルの正答率から推定発表語彙サイズを算出できる。このテストは、学習者がライティングにおいて、目標語が文脈に正しく合っているかを判断し、許容される形でその語を使うことができる能力があるかを測定している。
- オンライン版は <http://www.lexutor.ca/tests/levels/productive/>

5.2.2 語彙の豊かさの評価指標

- 学習者の発表語彙を客観的に測定する指標として、英作文における使用語彙を統計的に分析し、評価指標として使用するための研究が行われてきている。その中でも「語彙の豊かさ (Lexical richness)」を測定してライティングの評価指標とする方法がある (Read, 2000)。

(1) 語彙の多様性

- TTR (type-token ratio): 英作文に使用された語彙の総語数 (token) を異語数 (type) の割合を算出し、数値が高いほど多様な語彙が使用されていることになる。しかし、総語数が多くなるほど数値が低くなる傾向があるため、総語数の異なるライティングを比較する場合には注意が必要になる。
- Guiraud Index: TTR をベースに、総語数の平方根を分母とする指標。この操作によりライティングの長さによる影響を少なくしている。

(2) 語彙の洗練度

- 使用語彙の質的な豊かさについては、ライティングにおいてアカデミック語彙や低頻度語をどれだけの割合で使用できたかを指標とする。
- 洗練度は一般的に、特定の語彙リストとの照合を行って算出する。

(3) 語彙密度

- 語彙の情報密度は、総語数に占める内容語の割合を指標とする。内容語の比率が高いテキストは情報量が大きくなる傾向にあり、書き言葉の語彙密度は 40% 以上、話し言葉の場合は 40% 以下になるとされる。

→ 語彙頻度プロファイル (Lexical Frequency Profile: LFP)

- ライティングの使用語彙を 4 つの頻度レベルに分類し、それぞれのレベルに属する word family の割合を求め、それらの相対的使用割合により語彙の洗練度を評価する指標 (Laufer & Nation, 1995)。
- 判定の基準として使用される語彙リストは (1) 最高頻度 1,000 語、(2) 次に使用頻度の高い 1,000 語、(3) Academic Word List、(4) その他であり、100 語の英作文の 60 語が (1)、10 語が (2)、20 語が (3)、10 語が (4) に分類されたとすると、LFP は 60%-10%-20%-10% となる。
- この評価では、(1) と (2) の割合が低く、(3) と (4) の使用割合 (Beyond 2000) が大きいほど発表語彙の洗練度が高いと見なされる。
- 最近では Range や Ant Word Profiler というソフトウェアで、BNC の頻度データを基に 16 段階の分析が可能になっている。ただし、300 語以下の英作文では結果が不安定になり、トピックの影響を完全に排除できないという留意点もある。

→ P_Lex

- ライティングを 10 語ごとのチャンクに区切り、その中の低頻度語の割合を計算してグラフ化するプログラム (Meara & Bell, 2001)。グラフは、横軸に 10 語のチャンクのうちの何語が低頻度語であるかを 0 語から 10 語まで取り、横軸にそのチャンクのパターン毎のパーセンテージが示されるようになっている。
- 自然言語では低頻度語の割合が低いパターンのチャンクが多く、相対的に低頻度語の割合が高いパターンは少なくなるという分布になり、グラフ化したカーブから算出される lambda の値が高いほど低頻度語の割合が高くライティングの洗練度が高いと解釈される。
- lambda の値は 150 語以上のライティングで安定し、テキストの長さの影響を受けにくくなる。プログラムは http://www.lognostics.co.uk/tools/P_Lex.zip から DL できる。

5.2.3 定量的評価指標の問題点

- 定量的指標は絶対的なものではなく、ジャンルやトピックなどの影響を受ける。そのため、ライティングの条件に一定の統制を加えた上で数値の解釈を行う必要がある。また、複数の指標を組み合わせたり、教師の評価との併用を考えたりする必要がある。

5.3 語彙知識の深さの測定・評価

- 日英対訳の語彙学習では、一定のサイズの語彙知識があっても、多義語やコロケーション、イディオムなどに対する応用力の低い語彙力になってしまう傾向がある。
- 一定以上の語彙サイズを持つ学習者が、より応用的な語彙力を身につけるために、語彙の深さを測定・評価する方法が必要になる。

5.3.1 The Vocabulary Knowledge Scale

- 与えられた語彙項目それぞれの知識を5つのカテゴリー (図4, p. 249) から選んで自己申告する。選んだカテゴリーによっては自分が本当にその単語についての知識を持っているかどうか記入する必要がある。
- このテストは受容・発表語彙知識を測定し、学習者の語彙力をより詳しく測定できる。

5.3.2 Word Association Test (WAT)

- 語連想知識やコロケーションの知識を測定するテストで (Read, 1993)、学習者の語彙知識を概観的に見ることができる (ただし、個々の語彙項目の知識を見ることはできない)。
- また、語彙サイズが十分でない学習者がこのテストを受けると、深さではなく広さを測るようなテストになってしまうため、対象となる学習者の語彙サイズを確認しておく必要がある。

おわりに

- 語彙テストの開発における今後の課題について、(1) 語彙力の構成概念・概念同士の関連性や英語熟達度との関係についての研究を進めること、(2) コンピュータ適応型テストの実用化 (項目のランダム提示 & 項目応答理論による信頼性の確保)、(3) 音声による語彙テストの開発 (従来のテストにはない) が挙げられる。
- また、テスト開発の際の留意点として以下の項目が挙げられる。
 1. テストの種類や目的は何か (到達度テスト / 熟達度テスト / 診断テスト)
 2. どのような語彙力を測定するテストにするか (サイズ / 深さ / 反応速度など)
 3. 「語」の単位を何にするのか (word family / lemma)
 4. 受容語彙 or 発表語彙を測定するテストか
 5. 書き言葉 or 話し言葉の語彙テストか
 6. 単独で単語を提示するのか、文脈の中で提示するのか
 7. 英語のみのテストにするのか、日本語の選択肢を含むバイリンガルテストにするか
 8. 記述式テストか多肢選択式テストか
 9. 一般的な語彙のテストか、ESP的な専門用語のテストか
 10. どのような語彙リストに基づいてどのレベルの問題を何題出題するのか
- 最後に Nation (2008) はよい語彙テストとしての基準を以下のように示している (詳細は p. 252)。
 - (a) 信頼性があること
 - (b) 妥当性があること
 - (c) 実用性があること

<コメント>

現在の語彙テストはすべて書き言葉の受容・発表語彙知識を評価しているので、話し言葉の受容・語彙知識の評価法を確立する重要性を議論した。学習者のスピーキングコーパスを収集し、個々人の語彙熟達度を評

価した研究に Crossley, Salsbury, McNamara, and Jarvis (2011) があり、この研究はライティングの評価で用いられている語彙の多様性などの指標を用いている。このような研究が進むことで、話し言葉における学習者の語彙力を評価することが可能になるとと思われる。

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, 42, 182-193.