

# 語彙の「広さ」を扱った語彙テスト研究の動向

— 日本語学習者用語彙テストの作成に向けた英語教育分野の研究概観 —

ジャ・ブルブル

(2016年10月6日受理)

“Breadth” of Vocabulary and Vocabulary Test Development:  
A Review of Research in the Field of English Education for the Development of Vocabulary  
Test for Japanese Language Learners

Bulbul Jha

**Abstract:** Vocabulary, being an important part of Second Language learning, has gained attention of many researchers and there have been many tests developed in English language for testing Vocabulary of learners. However, no such test exists in the case of Japanese language. Therefore, the purpose of this study is to review studies done in the field of English Language Education, in order to summarize the criteria needed in making vocabulary test. English vocabulary tests can be classified into two: vocabulary size or breadth tests and vocabulary depth tests, the current study focuses on the former. The pros and cons of the three different formats of vocabulary tests, namely, multiple-choice test, checklist test, and word levels test are discussed and it is clear that word levels test is the most suitable format for testing the vocabulary breadth of Japanese language learners. Furthermore, the characteristics specific to Japanese language need to be considered apart from sampling and context effect while developing the test in Japanese Language.

Key words: Vocabulary, Word levels test, Breadth of vocabulary, Multiple choice test, Checklist test

キーワード：語彙、語彙レベルテスト、語彙の広さ、多肢選択式テスト、チェックリスト式テスト

## 1. はじめに

Nation (2001)によると、語彙は言語の最小構成要素であり、言語研究や第二言語習得研究の研究対象として欠かせないものであるが、語彙が研究対象として注目され始めたのは近年になってからのことである。母語 (native language : first language とほぼ同義とし、以下 L1) 話者は膨大な語彙を有し、成人になっても毎年1000語族<sup>1</sup>を習得していくが (Nation, 2001)、第二言語 (second language : 以下、L2) 学習者では、語彙習得に意識的な努力が必要となる。それは、第二言語として日本語を学ぶ学習者においても同様である。本稿では、体系的な開発が遅れている日本語学習

者用語彙テストを作成する際に、どのような点を考慮すべきかを明らかにするため、語彙学習の重要性が指摘され、研究も盛んに行われている英語教育分野に焦点を当て、語彙テスト研究を概観しながら、その研究動向をまとめる。

語彙に関する研究では、語彙教授法、語彙学習方略、語彙の数と知識、受容語彙 (receptive vocabulary) と産出語彙 (productive vocabulary) の関係など、様々な側面が検討されてきた。特に、語彙の数と知識に関する研究が多く行われ、どのくらいの数の語彙を知っているかや、特定の語彙に対して、その意味を知ること、文法的役割を知ること、派生語を知ること、連語を知ること、などを語彙知識としてそれを測定するテ

ストが複数、開発されてきた。

語彙数を測定するテストは、語彙の広さのテスト、あるいは語彙サイズのテストであり、語彙知識を調べるテストは、語彙の深さのテスト、あるいは語彙の質のテストを指す。本稿では、このうちの前者に焦点を当て考察を進める。

## 2. 語彙の広さを扱ったテストの先行研究

語彙の広さを扱った研究を概観するにあたり、前半では、それに関わる3つの要因を述べる。3つの要因とは、語彙の数え方、テスト項目のサンプリングおよびサンプリングレートである。

### 2.1 語彙の数え方

語彙サイズのテストは通常、語彙サンプルのうち、学習者がどのくらいの語彙を知っているかにより、語彙サイズを予測できる、という考えに基づいている。例えば、3万語の中から無作為に抽出された300語のうち、30語を知っている学習者は、3万語の10%、つまり3000語を知っていると推測できる。ただし、語彙をどのように数えるかによって、総数は変わる。語彙を数える単位としては、延べ語数 (Tokens)、異なり語数 (Types)、語彙素 (Lemma)、語族 (Word family) などがあげられる。延べ語数は、文章中の語彙の総数であり、同じ語彙が何回出てても毎回数えられる。一方、異なり語数は、同じ語彙が文章中に何回出てても、1つとしか数えられない時の総数である。例えば、「山の後ろから山が見えていた」の延べ語数は8であり、異なり語数は7である。

語彙には、機能語と内容語という分け方もある。冠詞、前置詞、代名詞、接続詞、補助動詞など、単独では意味を持たず、内容語を修飾するなど、文中で他の語彙との関係において一定の役割を果たすものは総称して機能語と呼ばれる。他方、名詞、動詞、形容詞、副詞など、単独で意味を持つものの総称は内容語と呼ばれる。語彙テストによって語彙知識を調べるときは、内容語を扱うことが多い。

語彙研究において、語彙は語彙素と語族によって決められる。語彙素は、見出し語とそのいくつかの活用形と分解形の総体を意味する。また、語族は見出し語と意味が同じ活用形と密接な関係にある派生形を含む (Nation, 2001)。英語の例を見ると、'happy', 'happiness', 'happily', 'unhappy', 'unhappiness'などは、'happy' という見出し語の語族に属する。語彙素または語族を一つの語として扱うことは、学習者が活

用ルールと派生ルールを知っていれば、見出し語を知ることがその活用形、派生形を知ることにつながるという考えを基本にしている。両方の数え方は、ある程度の妥当性をもっているが、前者には過小評価が、また後者には過大評価の危険性が存在する (Ishii, 2005)。語彙サイズの研究では、特に、語彙素と語族の分け方が重要となる。L1話者を対象とした多くの研究において、語彙サイズの推測に別々の値を出しているのは、研究者によって語彙の数え方が違うからである (Read, 2000)。

### 2.2 語彙のサンプリング

語彙の数え方を決めた後に重要となるのは、語彙のサンプリングである。語彙のサンプリングは、辞書と語彙頻度リストの2つの情報源から可能となる。

#### 2.2.1 辞書からのサンプリング

L1話者の語彙サイズを扱った多くの研究は、辞書からのサンプリングを利用してきた。その方法として、各 m 番目のページから各 n 番目の語彙をとることがある (Nation, 1993)。しかし、この方法には高頻度語彙への偏りという問題が存在する。高頻度語彙は、辞書の中で多くのスペースをとり、選択の可能性が高いからである。Nation (1993) は、この問題の解決方法として、辞書のすべての語彙に番号をつけ、そこからページにかかわらず、各 n 番目の語彙を選択する方法を挙げている。

#### 2.2.2 語彙頻度リストからのサンプリング

L1話者でも、辞書にあるすべての語彙を知っているわけではない。学習者の場合は、辞書にあるすべての語彙を知る必要性がないので、語彙頻度リストがより妥当なサンプリング資源となり得る。L1話者の書き言葉コーパスと話し言葉コーパスに基づき、語彙の頻度を調べたリストが作成されている。例えば、West (1953) による General Service List がそれにあたる。

### 2.3 サンプリングレート

語彙サイズを扱うときにもう一つ重要な概念はサンプリングレートである。語彙サイズのテストでは、学習者がどの程度の語彙を知っているかの推測をサンプルによって行う。ただし、学習者に関しては、サンプル以外の語彙を知っているものの、テスト中の語彙を知らない可能性も考えられるため、サンプルが大きい方が望ましい (Meara & Buxton, 1987)。また、大量の語彙について尋ねるため、テスト形式は簡単な

なければならない。

テスト項目を選択した上で、次のステップはテストを作成することである。語彙知識を調べるためには、多肢選択式、マッチング式、翻訳式、チェックリスト式など、様々なテスト形式が利用できる。テストの目的によってテスト形式を選ぶ必要がある。次節では、英語の語彙サイズテストの各形式について、長所と短所を考察する。

## 2.4 多肢選択式テスト

これまでの研究で最も多く使われてきたのが多肢選択式のテストである。この形式では、目的語と選択肢を様々なフォーマットで提示できる。以下に例を示す。

表1 多肢選択式テストの例

1a	to <i>fade</i> means a. to make a loud sound b. to get something to eat c. to paint something a bright colour d. to become quieter
1b	When a light <i>fades</i> it a. gets weaker b. gets stronger c. flashes intermittently d. turns red
1c	The blue curtains began to _____ after they had been hanging in the sun for two months. a. fade b. die c. dissolve d. melt

(Meara & Buxton, 1987より引用)

多肢選択式の語彙テストには、Wesche & Paribakht (1996) が述べるように、以下の問題点がある。

- (1) テストを作るのが難しく、フィールドテストおよび分析と改善に多くの時間と労力を要する。
- (2) 学習者が、求められている意味以外の意味を知っている可能性がある。
- (3) 学習者は、消去法によって正答を導くことができ、4択の場合、正答を推測できる可能性は25%ある。
- (4) 各項目は学習者のディストラクターの知識を検討している可能性もある。
- (5) 学習者が正答に至らないのは、語彙知識の不足だけでなくディストラクターのシンタックスがわからないから、という可能性もある。
- (6) 本テスト形式は、限定したサンプルしかテストできない。例えば、25項目の多肢選択式テストでは、1万語のうちの400語中、1語しかテストしないことになる。

しかし、これらの問題点があるにもかかわらず、実施の容易さ、採点のしやすさ、結果の統計分析のかけ

やすさ、項目分析による不適切な選択肢の見つけやすさといった利点があるので、現在でもこの形式は多用されている。

## 2.5 チェックリスト式テスト

語彙テストの中で回答形式が最も単純であるものは、チェックリスト式テストであり、yes/noテストとも呼ばれる。Zimmerman, Broder, Shaughnessy, & Underwood (1977) によって開発され、英語L1話者の語彙サイズの検討に使用された (Meara & Buxton, 1987)。Meara & Buxton (1987) は、この形式を利用し、L2学習者の語彙サイズの検討を行った。学習者は語彙リストを与えられ、各語彙を知っているかどうかについて答える。以下に例を示す。

表2 チェックリスト式テストの例

Tick the words you know the meaning of			
Gathering	Forecast	Descent	Revenge
Wodesome	Untamed	Awkward	Flane

(Meara & Buxton, 1987より引用)

チェックリスト式の語彙テストの問題点として、学習者の返答を確かめる方法がないことが挙げられる。その解決策として Meara & Buxton (1987) は、無意味語をリストに導入した。表2に示した 'Wodesome' と 'Flane' は無意味語である。無意味語を入れることによって、学習者が自分の語彙知識を過大評価しているかどうかについて確認できる。無意味語を選ばない学習者に問題はないが、無意味語を知っていると回答した学習者は、自分の語彙知識を過大評価しており、その学習者の得点を算出するときは Anderson & Freebody (1983) による以下の数式を利用する。

$$P(k) = P(h) \cdot P(fa) / 1 - P(fa)$$

$P(k)$  = 学習者が実際に知っている語彙数

$P(h)$  = 学習者が知っているとする本当の語彙の割合

$P(fa)$  = 学習者が知っているとする無意味語の割合

Meara & Buxton (1987) は、チェックリスト式語彙テストと多肢選択式語彙テストの相関関係を調べ、高い相関 ( $r = .703$ ) がみられたことを報告している。ただし、フランス語をL1とする英語学習者では、さらに高い相関 ( $r = .829$ ) がみられ、相関係数がこのような特定の言語話者で異なることは、無意味語に関する問題点を浮き彫りにした。つまり、特定の言語をL1とするL2学習者では、無意味語がL1の存在語に似ている場合があり、それを知らないと答えることは、別の言語をL1とする学習者と比べて難しくなる。Meara & Buxton (1987) は、チェックリスト式語彙

テストが、限定された時間内で他のテスト形式よりも多くのサンプル語を対象とすることができ、テストの作成時間および実施時間が短いという利点があったとした上で、無意味語へのL1の影響を検討することにより、このような問題が改善できることを主張した。

Meara & Buxton (1987) で使用されているテストのもう一つの問題点は、学習者の個人差である (Ishii, 2005)。学習者が自分の語彙知識について考えるとき、厳しい基準をとる学習者と緩やかな基準をとる学習者が存在する。前者では過小評価の可能性があり、後者では過大評価の可能性が否定できない。無意味語の採用は、過大評価の問題を解決できるが、過小評価の問題は解決できない。したがって、学習者の語彙サイズを推測しても、実際にそれと同じ語彙数をもつとは必ずしも言えない。

2.6 Eurocentres Vocabulary Size Test (EVST) :

Meara & Jones (1990a)

EVSTは、コンピューター上で実施できるチェックリスト式テストであり、学習者の語彙サイズの検討に使用される。パソコンの画面には語彙が1つずつ呈示され、それとともに質問文 ‘Do you know this word?’ が出て、学習者は「はい・いいえ」で答える。テスト語彙には無意味語が含まれており、学習者にはそのことが告げられるが、もし無意味語を知っているとした場合は減点される。採点はコンピュータプログラムによって自動的になされ、短時間で多くの語彙をテストできるのが長所であるが、依然として無意味語の特質が問題となっている。Meara & Jones (1988) が指摘するように、いくつかの無意味語は簡単であり、他のものはL1話者でも混乱するものであった。さらに、英語では無意味語であっても学習者のL1で有意味語の場合は、その判断が難しくなる。すなわち、学習者のL1によってテスト成績が影響を受ける (Meara, 1996a ; Meara & Buxton, 1987)。さらに、無意味語を選択する場合は成績が低くなる。そのため、初級学習者の成績については、このテストの信頼性は低い。

2.7 語彙レベルテスト

語彙レベルテスト (word levels test) は、Nation によって1980年代に開発され、多くの国で利用されてきた (Read, 2000)。多くの研究者が注目し、新しい版も開発された (e.g., Beglar & Hunt, 1999; Schmitt, Schmitt, & Clapham, 2001)。語彙レベルテストでは、学習者が定義に合う語彙を選択し、その番号を空欄に記入するように求められる。例を以下に示す。

表3 語彙レベルテストの例

1. Business	
2. Clock	
3. Horse	_____ part of a house
4. Pencil	_____ animal with four legs
5. Shoe	_____ something used for writing
6. Wall	

(Nation, 1990より引用)

語彙レベルテストの目的は、学習者が各頻度帯で知っている語彙の数を調べることであり、5つのセクションに分けられている。そのうちの4つは、2000語、3000語、5000語、10000語のレベルであり、語彙の頻度に基づき作成されている。Read (2000) によると、語彙の頻度は、Thorndike & Lorge (1944) のリスト、West (1953) の General Service List、および Kučera & Francis (1967) のリストに基づいている。

残りの1つは、語彙の頻度ではなく University Word Level (大学語彙レベル) によるものである。University Word Level は学術的文章で多く使われ、普段の文章には使われない語彙が含まれる。University Word Level は Campion & Elley (1971) のリストに基づいている (Read, 2000)。Schmitt et al. (2001) は、Coxhead (2000) の Academic Word List に基づいて University Word Level の新版を作成し、Academic Word Level と名付けた。各レベルで、36個の語彙と18個の定義があり、6つのグループに分かれている。各グループ内の語彙は同じ品詞であり、意味が類似したものは1つのグループに入らないように構成されている。各レベルの語彙は、固有名詞と複合語を除いて無作為に選択され、各頻度レベルで学習者の語彙知識を計測できるようになっている。

テストの公開に向けては、その妥当性を検証する必要があり、語彙サイズのテストの妥当性を確認する方法の1つは、背景にある理論を検証することである。それは、学習者が低頻度の語彙よりも高頻度の語彙を知っている、ということである。Read (1988) は、学術的な英語を学習している81名を対象として、3ヶ月の間に2回の語彙レベルテストを実施し、その成績を分析した。その結果、1回目のテストでは、語彙頻度が下がるにつれて学習者の得点も減少する傾向がみられた。2回目のテストでは、1回目よりも成績が良かったことからコースの実績が評価されたが、語彙頻度が下がるにつれて成績が低くなる傾向がみられず、むしろ University Word Level の方が5000語レベルよりも成績が良かった。Read (1988) はその原因として、学習者が学術的な語彙を習得していることを挙げている。Schmitt et al. (2001) は、語彙レベルテスト

の新版について、妥当性の量的検証とあわせてインタビューによる質的検証も行った。そして、このテストが、実施方法の簡便性や採点の容易性を備えており、学習者の語彙に関する全体的なイメージを把握できる点で優れたテストであることを主張している。

## 2.8 まとめ

語彙テストの開発で最も重要な点は、サンプリングレートであり、サンプルが大きければ推測がよりの確になる。そのためには、テストの実施方法に簡便性が求められる。チェックリスト式の語彙テストは、作成時間とテスト時間を考慮すれば最も適切であると言えるが、受験者が正確に答えているかどうかを確認できない点が最大の短所である。その解決策として、無意味語を入れることがあるが、それ自体にも問題がある (Huibregtse, Admiraal, & Meara, 2002; Shillaw, 1996)。この問題は、多肢選択式の語彙テストであれば解決できるが、回答に際して推測の可能性が残る。そのため、各項目の選択肢の数を増やすか、項目自体の数を増やすことになるが、テストの所要時間が増えることが今度は問題となる。

上記のことをふまえるならば、語彙レベルテストは両方の問題点を解決できる。このテストでは、各項目セットに対して、3つの目標選択肢と3つのディストラクターがあり、3つの正答に至る確率は120分の1となる。したがって、語彙レベルテストは、推測の可能性が低く、テスト時間も短縮できるといえる (Ishii, 2005)。

## 3. 語彙の広さを扱ったテストの妥当性検証研究

前章では語彙の広さ、語彙サイズを扱ったテストについて先行研究を概観したが、テストの作成・開発では、その妥当性を調べた研究が、あわせて重要となる。本章では、それについての研究動向を述べていく。

### 3.1 語彙テストの妥当性検証の方法

語彙テストの妥当性を検討する際は、テスト自体の妥当性を検証する方法と、他のテストとの相関関係を調べる方法がある。Meara & Buxton (1987) は、チェックリスト式語彙テストと多肢選択式語彙テストの相関を調べたが、多肢選択式語彙テスト自体の妥当性は確認していない。通常、語彙サイズテストの妥当性を検討する際は、インタビューは実施されないが、Schmitt et al. (2001) が示した通り、項目分析では学習者の推測が確認できない点を、インタビューで見ることができる利点があり、質的な検証の必要性も指摘

されている (Ishii, 2005)。Schmitt et al. (2001) と同様に Read (1993) も、学習者の口頭による報告はテストへの反応について貴重な情報を与えるとしている。インタビューによって学習者の推測の問題が浮上し、以降のテスト開発でも重要な指摘点となっている。表4は、語彙サイズのテストの妥当性検証研究の一覧であるが、方法上は次の2点に注意する必要がある。

- (1) 2つのテスト間で相関関係を調べるときは、他方のテストが何を測定しているかを吟味し、明確にしなければならぬ。
- (2) 量的な分析だけではなく質的な分析も同時に行うことによって、テスト項目やテスト全体の妥当性がより適確に検証できる。

表4 語彙サイズの妥当性検証研究の一覧

	Quantitative examination	Qualitative examination
Checklist format		
Meara & Buxton (1987)	多肢選択テストとの相関	
Vocabulary Levels Test		
Read (1988)	スケーラビリティ	
Beglar & Hunt (1999)	TOEFL 得点との相関 項目分析 信頼性	
Schmitt et al. (2001)	項目分析 因子分析 スケーラビリティ 信頼性 実用分析	学習者とのインタビュー

### 3.2 語彙知識間の関係性を調べた研究

本節では、語彙知識間の関係性を調べた研究について述べていく。

#### 3.2.1 Schmitt & Meara (1997) の研究

Schmitt & Meara (1997) は、L2習得分野で種類の異なる語彙知識間の関係を調べた最初の研究である。関連語と接頭辞に着目し、これら2つの知識と、学習者の語彙サイズ (Nation, 1990の Vocabulary Levels Testで測定) と英語能力 (TOEFL得点) との相関を調べた。その結果、関連語の知識については、語彙サイズと英語能力との間で一定の相関がみられた。また、接頭辞の知識と語彙サイズの間にも弱い相関がみられた。加えて、関連語の知識と接頭辞の知識の間でも相関がみられた。しかし、Schmitt & Meara (1997) には、語彙サイズ以外では2種類の語彙知識しか扱っ

ていないという限界がある。さらに、テストの信頼性と妥当性が検証されていないことが問題点として挙げられる。Schmitt & Meara (1997) は、語彙サイズと、接頭辞の知識、関連語の知識とは関連があると主張しているが、語彙サイズの増加によってこれらの関係がどう変容するかは残された課題である (Ishii, 2005)。

### 3.2.2 Mochizuki & Aizawa (2000) の研究

Mochizuki & Aizawa (2000) も、学習者の語彙サイズと接頭辞の知識との関係を調べた。その結果、両者には Schmitt & Meara (1997) の相関係数よりも高い値 ( $r=0.65$ ) がみられたが、研究の問題点として、接辞のテストが実験前に十分に検討されていないことが挙げられる。

### 3.2.3 Ishii (2005) の研究

Ishii (2005) は、語彙サイズテスト、多義性テスト、派生語テスト、語彙選択テストを作成し、4つのテスト間の相関関係を調べている。各テストの例を表5～表8に示す。

いずれのテストも信頼性が高かったが、多義性テスト、派生語テスト、語彙選択テストは、語彙サイズの小さい学習者には不適であることが明らかとなった。4つのテスト間で一定の相関はみられたが、語彙サイズと多義性テストの間で最も高い相関がみられ、多義性テストと語彙選択テストの間の相関がそれほど高くはなかった。Ishii (2005) は、天上効果が表れたことを問題点として指摘し、天上効果を示した学習者を除外した場合、相関係数が低くなることを報告している。また、学習者の語彙サイズの相違によって4つのテスト得点の外形が異なることを述べている。しかし Ishii (2005) によると、テスト間の関係は直線的ではないため、テスト得点の解釈には注意が必要である。

## 4. おわりに

本稿では、英語教育の分野における語彙テストの研究に関して、語彙の広さを扱ったものを中心に概観した。より良い語彙テストを作成・開発するためには、語彙の広さと同時に語彙の深さを検討する必要があるが、語彙の広さに関する研究動向をふまえた現時点で、L2としての日本語語彙テストの開発に向けた重要な点を整理しておきたい。

学習者の語彙サイズのテストでは、多肢選択式テスト、チェックリスト式テスト、語彙レベルテストの3つについて概観した。多肢選択式テストでは、大きなサンプルでテストが出来ないという問題点が存在す

表5 Ishii (2005) の語彙サイズテストの例

例に従って、右側に与えられた意味に適する語を左側から選び、別紙の解答用紙に番号で答えなさい。	
例)	
1. business	
2. clock	a) 馬
3. horse	b) 壁
4. pencil	c) えんぴつ
5. shoe	
6. wall	

表6 Ishii (2005) の多義性テストの例

これは、多義語のテストである。例に従って、与えられた単語の意味を2つずつ選び、番号で答えなさい。	
例) run (b) (e)	
[a] 聞く [b] 経営する [c] 進む [d] 測る [e] 走る	

表7 Ishii (2005) の派生語テストの例

例に従って、与えられた単語を適切な形に変えて、表を完成させなさい。適切な形がないと判断した場合には、×印を書きなさい。	
例 : stimulate	
(名詞)	stimulation (動詞) stimulate
(形容詞)	stimulating (副詞) X

表8 Ishii (2005) の語彙選択テストの例

例に従って、空欄に適する語を選び、別紙の解答用紙に答えなさい。答えが分からない場合は推測せず、[c] の「分からない」を選びなさい。	
例)	
[a] empty [b] free [c] 分からない	
例1	He has a lot of _____ time.
例2	They drink fast, and the bottle is already _____.
例3	The family is away on holiday, and the house is _____.

る。それは、チェックリスト式テストによって解決できるが、チェックリスト式テストで調べた学習者の自己評価は、語彙知識を適確に反映しているかどうかの確認がとれないという短所がある。これらのことを総

合すると、3つの中では、語彙レベルテストが、語彙サイズの測定に最も適した方法であると言える。

また、次のような点にも注意する必要がある。

1つめは、サンプリングの問題である。辞書からサンプリングすることは、様々な問題点が存在するため、辞書ではなく使用頻度リストに基づいて語彙のサンプルを採取することが適切であろう。国立国語研究所から『現代日本語書き言葉均衡コーパス』の語彙表が公開されているため、この語彙表を活用して日本語の語彙テストを作成することができる。できる限り大きなサンプルで調べる方が学習者の語彙力を適確に把握できるため、可能な限り簡便な実施方法を採用することも大切である。

2つめは、文脈の扱い方である。テストで文脈を使うことにより、学習者のテスト得点が変わる可能性がある。文脈がヒントとなり未知の語彙でも正答に至る可能性と、文脈が難しく正答に至らない可能性の両方が存在する。テストで文脈を与えるときは、学習者が構文や語彙を理解できるかどうかについて、十分な予備調査を行うべきであろう。

3つめは、多肢選択式テストの採点基準である。英語の語彙テストでは多肢選択式テストが多用されている。その場合は、正答と誤答が混在していたときの採点基準が明確でないため、日本語の語彙テストを作成する際は、多肢選択式テストの採点基準を吟味しなければならない。改善案としては、Ishii (2005) と同様に、2つの答えが正しい場合のみ点数を与える方法が考えられる。

最後に、日本語の語彙テストを開発する際は、日本語という言語に特有の注意点も存在する。

たとえば、外来語の取り扱いである。外来語は日本語の大部分になっているが、学習者はその語彙を日本語として理解できなくても、L1や日本語以外のL2の知識を使って意味が推測できる場合がある。

文字表記についても慎重を期す必要がある。英語はアルファベット言語であり表記は1種類であるが、日本語には平仮名、カタカナ、漢字という3つの表記がある。語彙の視覚呈示においては、通常、親近性の高い表記が使用されるが、語彙の一部または全部が漢字で表記される場合は、学習者のL1が漢字圏の言語か非漢字圏の言語かによって個人差が生じる。特に、中国語を母語とする学習者は、漢字を見ただけで、あるいは音を読めなくても、意味が分かる（推測できる）語彙がある。それを日本語の語彙力と評価してよいかどうか問題となる。同形または類似形での同義語と、同形または類似形での異義語の存在を考慮しつつ、テスト項目を作成する必要があるだろう。

## 【注】

<sup>1</sup> 語族は普段 language family と訳されるが、本研究では word family を意味する。

## 【引用文献】

- Anderson, R., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. in B. Huston (ed.) *Advances in Reading/Language Research*, 2 (pp. 221-250), Greenwich, CT: JAI Press.
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and the University Word Level vocabulary tests. *Language Testing* 16(2), 131-162.
- Campion, M. E., & Elley, W. B. (1971). *An Academic Vocabulary List*. Wellington: New Zealand Council for Educational Research.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing* 19(3), 227-245.
- Ishii, T. (2005). *Diagnostic tests of vocabulary knowledge for Japanese learners of English*. Unpublished PhD thesis, University of Nottingham.
- Kučera, H., & Francis, W. M. (1967). *A Computational Analysis of Present Day American English*. Providence, RI: Brown University Press.
- Meara, P. (1996a). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Willams (Eds.) *Performance and Competence in Second Language Acquisition* (35-53). Cambridge: Cambridge University Press.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing* 4(2), 142-151.
- Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (ed.) *Applied Linguistics in Society* (pp. 80-87). London: Centre for Information on Language Teaching and Research.
- Meara, P., & Jones, G. (1990a). *Eurocentres Vocabulary Size Test 10ka*. Zurich: Eurocentres Learning Service.
- Mochizuki, M., & Aizawa, K. (2000). An affix

- acquisition order for EFL learners: An exploratory study. *System* 28(2), 291-304.
- Nation, I.S.P. (1993). Using dictionaries to estimate vocabulary size: Essential, but rarely followed procedures. *Language Testing* 10(1), 27-40.
- Nation, I.S.P. (1990). *Teaching and Learning Vocabulary*. New York: Heinle and Heinle Publisher.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal* 19(2), 12-25.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-357.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition* 19(1), 17-36.
- Schmitt, N., Schmitt, D., & Clapman C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing* 18(1), 55-88.
- Shillaw, J. (1996). The application of Rasch modelling to Yes/No vocabulary tests. University of Wales, Vocabulary Research Group discussion paper (available at <http://www.swan.ac.uk/cals/calsres/vlibrary/1s96a.htm>).
- Thorndike, E. L., & Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review* 53(1), 13-40.
- West, M. (1953). *A General Service List of English words*. London: Longman.
- Zimmerman, J., Broder, P., Shaughnessy, K., & Underwood, B. (1977). A recognition test of vocabulary using signal detection measures and some correlates of word and non-word recognition. *Intelligence* 1, 5-31.

(主任指導教員 松見法男)