



日英バイリンガル文法テストの開発：多読とのかかわりにおける文法能力の発達を測定する

著者	吉澤 清美, 瀨 敦子, 大槻 きょう子
雑誌名	関西大学外国語学部紀要 = Journal of foreign language studies
巻	16
ページ	45-60
発行年	2017-03
その他のタイトル	Developing a Japanese-English Bilingual Grammar Test: Measuring Japanese EFL Learners' Grammatical Ability in Relation to Extensive Reading
URL	http://hdl.handle.net/10112/11035

日英バイリンガル文法テストの開発： 多読とのかかわりにおける文法能力の発達を測定する

Developing a Japanese-English Bilingual Grammar Test: Measuring Japanese EFL Learners' Grammatical Ability in Relation to Extensive Reading

吉澤清美 高瀬敦子 大槻きょう子
Kiyomi Yoshizawa Atsuko Takase Kyoko Otsuki

This paper focuses on the development of the Japanese-English Bilingual Grammar Test (JEBGT, henceforth), which is intended to measure Japanese EFL learners' grammatical ability in relation to extensive reading in English. The paper starts with the rationale for developing the JEBGT: the lack of previous studies of the effects of extensive reading on the development of grammatical ability and the lack of instruments which measure grammatical ability with independent tasks. The test specifications, the principles for item writing, specifications of the Overall Test Structure, and specifications for individual tasks were made. After a small-scale pilot study, minor revisions were made. Then, the JEBGT was administered to 450 Japanese EFL learners at the university level. The test data was Rasch analyzed and the findings were presented.

キーワード

grammatical ability, a Japanese-English bilingual grammar test, extensive reading

1. はじめに

第2言語習得研究 (Second Language Acquisition Studies) においても、目標言語の言語形式 (過去時制を表す形態素、-ed などの form を指す) に触れる頻度 (frequency) が言語形式の獲得に大きな影響を与えていると指摘されている (例、Ellis, 2002; Hamrick & Rebuschat, 2014)。更に、母語、外国語に関わらず、リーディング研究に関して、読みを難しくしている要因の一つに活字に触れる量 (print exposure) が十分でないことを挙げられる。これは読書量が少ないということである。とりわけ、日本のように目標とする外国語を使う機会がほぼ外国語の授業だけであり、それ以外のところでは当該の外国語を使わなくても生活できる環境、

目標言語が英語の場合は外国語として英語 (English as a Foreign Language, EFL) を学習する環境では、print exposure が極めて少ない。近年、外国語教育・研究において、多読が取り上げられている。特に、日本の英語教育において多読は教育機関で実践され、その効果についても発表されている。しかしながら、L2リーディングカリキュラムを俯瞰した場合、多読を実践している言語カリキュラムの割合は非常に小さいと Grabe (2009) は言う。その理由をいくつか挙げており、その一つが多読クラスの授業形態や内容が、教師中心の授業とは違い、教師がクラスで文法項目の説明などを行わない、また学習者が入学試験など高リスクの試験対策準備を授業で行わないような教授法であり、学校などの管理職者や教師は多読授業に違和感を覚えるとしている。Grabeによると、多読が更に普及するには、多読の効果を示す実証的な研究からくる証拠が必要であるとしている。本研究はこのニーズに対して、これまであまり実証研究が少ない多読が文法能力の発達にどのように効果があるのかを研究するための測定道具となる日英バイリンガルテストの開発を述べる。

2. 多読の効果に関する先行研究

多読が1980年代に英語教育の中で採用され始め、その後多読が英語学習におけるモチベーション向上・情意面に与える影響と同時に、英語力向上に非常に効果がある指導法であるとの認識が世界的に広まってきている。英語力向上にもたらす多読の効果は小学生から社会人まで様々な年齢層の様々な分野に現れ、世界各地で報告されている。例えば、英語力に関して述べれば、リーディング力・内容理解力・およびリーディングスピード向上 (Beglar et al., 2011; Bell, 2001; Elley & Mangubhai, 1981; Huffman, 2014; Iwahori, 2008; Takase, 2007)、リスニング力向上 (Elley & Mangubhai, 1981; 高瀬, 2010)、ライティング力向上 (Hafiz & Tudor, 1989; Irvine, 2011; Lai, 1993; Mermmelstein, 2015)、語彙とスペリング向上 (Krashen, 1989)、語彙力向上 (Cho & Krashen, 1994)、日本人中学・高校生の模擬試験の伸び (Furukawa, 2008; 鈴木, 1996; 渡辺, 2014)、クローズテストの伸び (赤尾, 2015; Mason & Krashen, 1997; Takase, 2008, 2009; Takase & Otsuki, 2012; Walker, 1997)、TOEIC の伸び (Nishizawa et al., 2010)、および様々な英語力の伸び (Yamashita, 2008) などがある。

しかしながら、文法に関して言えば、文法能力向上への多読の効果を研究した実証的研究は存在するが数が少なく (Maruhashi, 2011; Rodrigo et al, 2004; Sheu, 2003)、Maruhashi 以外は研究対象者が日本人学習者ではない。Rodrigo et al. では研究対象者がアメリカ人のスペイン語学習者であり、Sheu では研究対象者が中国語母語話者である。つまり、現在の時点では、多読が日本人学習者の英語文法能力に与えた影響に関する研究は Maruhashi (2011) のみである。Maruhashi の研究は多読を主としたリーディング授業を受講した非英語専攻の137名の日本人大学生対象に、多読が文法力の向上に寄与するかどうかを検証したものである。参加学習者は

4月最初の授業と、7月最後の授業で文法テストを事前・事後テストとして受験し、その間約3ヵ月間（15回授業）は多読を実施し、授業内外でできる限り多くの英語の本を読んだ。文法テストには学習指導要領に記載のある以下の7つの文法項目で構成される合計37項目の問いが含まれていた。(1) 不定詞、(2) 関係代名詞、(3) 関係副詞、(4) itが名詞用法の句及び節を指すもの、(5) 時制（現在完了進行形・過去完了形・過去完了進行形・未来進行形及び未来完了形）、(6) 助動詞+受け身、(7) 仮定法。結果は、事後テストの平均が事前テストよりも有意に伸びていた。また、上記7つの文法項目の中では、関係代名詞と助動詞+受け身の2項目を除く5つの項目すべての難易度が下がり、伸びが見られた。この研究によると期間が僅か3ヵ月でも多読が日本人学生の文法力向上に寄与すると考えられる。しかしながら、Maruhashiの研究は多読が僅か3ヵ月であり、また統制群もなかった。

実際、大半の日本人学習者は中学・高校6年間に学校で或いは塾や予備校で高校・大学受験に向けて、大量の文法学習と訳（Grammar-translation）の授業を受け、大学に入学してくる。ところが、英語苦手・英語嫌いの大学生が多く、関西のある大学では、英語の単位を取得できずに再履修を余儀なくされた学生が毎年5～10%も存在していた（Takase & Otsuki, 2012）。彼らの調査研究によると、再履修学生81名に行ったアンケート調査（複数回答可）では、英語は苦手と答えた学生は52名（64.2%）、特に苦手な分野は文法であると回答した学生は過半数の28名（53.8%）で、リスニング（17名）、リーディング（16名）、語彙（13名）、ライティング（11名）と比較すると非常に多い。つまり、再履修の学生の大半が中学・高校から英語に躓き、その半数以上が文法に苦手意識を持っていた。一方、Takase（2009）によると、別の再履修生のグループが3ヵ月多読を行った後に行った調査では、42.1%が文法力の伸びを感じたと報告している。上記の様々な研究で示されているように、多読をすることにより、リーディング・スピードおよびリーディング力やリスニング力が伸びるのは明らかであるが、実際には多読で文法が伸びたと感じている再履修生もいるのである。

多読が文法能力の発達にどのように寄与するのかを研究する際、必要となるのが文法力向上を測定するテストであるが、日本人学習者に適した文法能力を独立させて測定し、テスト所要時間を多くとらないテストは殆どないと言える。本稿では文法能力測定のために作成した日英バイリンガル文法テストの開発プロセス並びに一回目のテスト実施のデータ分析の結果をまとめる。

3. 日英バイリンガル文法テスト開発プロセス

日英バイリンガル文法テストを開発するにあたり、テストの仕様書（specification）を作成した（Bachman, 1990；Bachman & Palmer, 1996；Carr, 2011）。次に、項目作成を行い、執筆者間での校正（review）、小規模なパイロットテスト、執筆者間での再校正（review）を行っ

た後、1回目のテストを実施した。Carr (2011) はテスト開発に関して仕様書を作成する必要があるが、それらはテスト作成の状況と目的 (Context and Purpose Specification)、テスト全体の構成に関する事柄 (specifications of the Overall Test Structure)、各テストタスクに関する事柄 (specifications for individual tasks) の3部からなると言う。以下では、日英バイリンガル文法テストの仕様書として、テスト作成の状況と目的、項目作成方針、テスト全体の構成、各テストタスクに関する事柄を説明する。

3.1 日英バイリンガル文法テストの仕様書

Carr (2011) は仕様書の「テスト作成の状況と目的」はテストの目的、測定する構成概念、テスト結果の解釈の枠組み (すなわち、集団基準準拠テスト、目標基準準拠テストとして取り扱うのか)、目標言語使用ドメイン (Target Language Use Domain) 並びにそのドメインでよく使用されるタスクタイプ、受験者の特性、許容可能な最低レベルのテストの有用性、リソースについて考える必要があると言い、Carr にもとづき、「テスト作成の状況と目的」に関する仕様書の作成を行った。下記ではその主な点について言及する。

「テスト作成の状況と目的」に関する仕様書では日英バイリンガル文法テストが測定する構成概念はどのようなものであるのかをまとめた。はじめに、文法テストという名はよく使われるが、こういったテストは実際何を測っているのか改めて考察する。Grammar「文法」という語の意味は多岐にわたる概念を含む。まず意味を持つ単位という観点から grammar は phonology との対立概念として考えられる。また、チョムスキー言語学の立場では当該言語の話者が持つ文を生成・理解する能力、そしてそれを記述したものと考えられる。この意味での grammar は syntax であり、「語が整えられてさらに大きな単位を作り上げる際に、その機能や構成要素、さらに意味や用法を扱う分野」(永野, 1978, p.52) と簡潔に定義されるように、語を統合する(統語)知識と考えられ、ネイティブスピーカーが直感的に習得するものである。また一方では、grammar の指すものとして、言語の正しい使い方を述べることもある。いわゆる規範文法というものであり、この意味での grammar は規範となる用法を述べている規則の集合である。英語では to 不定詞における副詞の位置、文尾に前置詞を置かない、などがその例である。文法性の根拠は、権威ある文法書や辞書にあることが多い。規範文法は学校文法とも重なる部分が多い。

本稿で述べている日英バイリンガル文法テストは、上記の統語的観点と規範的な意味での文法の両方を問う内容とすべきであると考えられる。これは、日英バイリンガル文法テスト作成は多読が学習者の文法能力にどのような影響を及ぼすのかを検証する研究プロジェクトの一環として開発されたためである。多読によって、英語のネイティブスピーカーが直感的に習得する統語的知識と、英語の正しい使い方の知識、の二種類の知識の習得を測るという研究プロジェクトの趣旨を踏まえてのことである。

テスト結果の解釈の枠組みについては、各多読プログラム、研究の目的によって集団基準準拠テストあるいは目標基準準拠テストの枠組みで結果を解釈することができる。このため、結果の解釈の枠組みについては本稿では言及していない。（注1）

目標言語使用ドメイン（Target Language Use Domain）は学校あるいは塾などの教育現場のようなアカデミックなドメインを想定した。よく使用されるタスクタイプとしては、実際の多読教育の現場で使用されることを想定し、多読時間、授業時間に影響がでないように20～30分で実施を終了させることができ、中学、高校、大学などで多読を実践している教師が採点などに多くの時間を割けないことを鑑みた結果、実用性を重視し、選択肢形式を選んだ。受験者の特性に関しては、多読プロジェクトの参加者を念頭においており、日本語母語話者であり、日本の中学、高校教育の中で英語を学んできた大学生を想定した。許容可能な最低レベルのテストの有用性に関しては、実際の多読クラスでの実施を想定しているため、授業時間中に実施できる長さ、採点者の時間的負担などを考慮する必要があるため、テストの実用性に重きがおかれた。更に、測定道具としての要件である、信頼性、妥当性を確認する必要性を認識した。

表1 テスト作成の状況と目的に関する仕様書

Component	Explanation
「日英バイリンガル文法テスト」の目的	日本語母語話者が多読を行い、多読が文法能力にどのような影響を及ぼすかを測定する。
測定する構成概念	英語のネイティブスピーカーが直感的に習得する統語的知識と、英語の正しい使い方の知識
テスト結果の解釈の枠組み	（今回考慮しなかった。）
目標言語使用ドメイン（Target Language Use domain）並びによく使用されるタスクタイプ	学校、塾などの教育現場のようなアカデミックなドメインとし、よく使われるタスクタイプは選択肢形式とする。
受験者の特性	日本の中学、高校で英語教育を受け、大学で学ぶ日本語母語話者
許容可能な最低レベルのテストの有用性	信頼性（reliability） 妥当性（construct validity） 実用性（practicality）

3.2 日英バイリンガル文法テスト項目作成方針、全体構成並びに各テストタスクに関する事柄

本研究で開発した日英バイリンガル文法テストは EDiT Grammar に倣い、作成された。EDiT Grammar とは English Diagnostic Test のことであり、学習者の文法力を診断し、学習者が選択した回答を基に躓きの原因を特定し、よりよい言語学習プログラムを作るためのテストである(金谷他, 2006)。このように、EDiT Grammar の開発コンセプトは、学習者の苦手とする文法項目をあぶりだし、かつその誤答に至った原因を特定するというものである。多読によってどのような文法項目が影響を受けるか、または受けないかを調査するという目的と整合性が高いことから EDiT Grammar を文法テストの基本フォーマットとすることとした。

EDiT Grammar は名詞句を中心に問題を編成している。これは、名詞句を把握することは、句の境界を把握することであり、且つ内部構造を把握することである。そしてこのことは文の構造の理解に直結する(金谷他, 2006)。金谷他(2006)はこの名詞句への焦点化について具体的に次の5つの理由を挙げている。まず、主語と述語の把握は、基本的な文法力であり、文の主語の特定と名詞句の把握は強く関連していること。二つ目は、日本人学習者は文の初めに来る主語が極端に長い場合、これを主語と特定することが困難であるということ。三つめは、名詞句における日本語と英語の修飾関係の違いの理解が英語学習の躓きの原因であること。四つ目は名詞句が、the window near Ken's desk にみられるように階層構造をなしていることが多いため、この階層性や内包構造は学習が困難となる要因を診断するために利便性があること。そして最後に、名詞句を扱うことで、学習者の文法知識を広い視野で捉えられること。これは名詞句の中に動詞と目的語の関係、時制、相などの複雑な文法事象を見出すことができるからである(金谷他, 2006)。こうした理由で、今回作成した日英バイリンガル文法テストもまた、名詞句を扱う問題を中心に、他にも動詞句、仮定法、主語述語の一致、時制の一致、補語の外置といった文法事項を問うている。

日英バイリンガル文法テストは46個の質問からなる選択肢形式のテストである。表2は、文法テストで扱われている文法項目と問題数を示したものである。

日英バイリンガル文法テストの問題文は、日本でもよく使用されている *Cambridge Grammar of English* (Carter & McCarthy, 2006), *English Grammar in Use 4th edition* (Murphy, 2012), *Practical English Usage* (Swan, 2005) といった英語参考書に掲載されている文法項目に関する情報を基に作成した。問題文作成に際しては日本語訳との関連、当該テストの趣旨、日本人学習者の言語レベル、背景知識などを考慮した。

一問は、日本語訳、それに対応する英文、そして4つの選択肢という3つの要素からなる。誤答を導く選択肢の選定は、先述の統語的要素、規範文法、そして日本語の影響(transfer)という観点を基にして行った。統語的要素とは、上記で述べた語の配列、規範文法とは英語の正しい使い方、日本語の影響とは、主に日本語の語順を基にした英単語の配列、日本語訳とその

表2 文法能力テストの項目と問題数

		分類	問題数
名詞句	グループ1	前置修飾 + 主要部	6
	グループ2	名詞句 + 前置詞句	2
	グループ3	名詞句 + 現在分詞	2
	グループ4	名詞句 + 過去分詞	2
	グループ5	名詞句 + 関係詞	2
	グループ6	名詞句 + to 不定詞	4
動詞句	時制・相	単純現在	2
		単純過去	2
		単純未来	1
		現在進行相	1
		過去進行相	2
		未来進行相	2
		現在完了相	1
		過去進行相	1
		現在完了進行相	2
		過去完了進行相	2
	態	受動態（現在）	1
		受動態（過去）	1
助動詞 + 受動態		1	
仮定法			2
主語動詞の一致			2
時制の一致			2
補語の外置			3
			46

英文の対応部分における時制の不一致などを意味する。下の例1は、表1中の名詞句のグループ3、現在分詞を含む名詞句に関する問の例である。

例1 肌を太陽から守るそのクリームは意外に安かった。

_____ was unexpectedly cheap.

1. The protecting cream skin from the sun
2. The cream protecting skin from the sun
3. Skin from the sun protecting cream
4. The protecting skin from the sun cream

誤答を導く選択肢1、3、4に関して、選択肢1、4は英語の語順を問う統語的要素、選択肢3は

日本語の語順を再現した英単語の配置である。次に、表1における動詞句の時制・相グループの未来に関する問いの例として例2を挙げる。

例2 明日は誰がリサを幼稚園へ連れていくことになったの。

Who _____ Lisa to the kindergarten tomorrow?

1. took
2. is going to take
3. was going to take
4. has taken

誤答を導く選択肢1、3、4は日本語訳「連れていくことになった」の影響を考慮しての、過去、或いは完了を表す形式を含んだ選択肢である。翻って、次の例3は、主語動詞の一致に関して、三人称単数現在という文法規則を問う例である。

例3 全ての学生はその書類を次の月曜日までに提出しなければならない。

Every student _____ to submit the form by next Monday.

1. are needed
2. need
3. needs
4. is needed

このように、文法テストは語順にかかわるタイプ、すなわち母語話者が体得する統語知識と英語文法の規範的な面を問うタイプの双方の問題を含んでいる。

4 第一回日英バイリンガル文法テスト実施

4.1 実験参加者と実施手順

当実験の参加者は、関西のある私立大学の英語専攻以外の1・2回生学生450名であり、1年間のリーディングプログラムの受講者であり、多読グループ278名、精読グループの172名であった。両グループとも習熟度別ではないため英語力の差が大きく、初級レベルから上級レベル、つまりCEFRレベルで見るとA1～B2程度の様々な学生が混在していた。

多読グループ、精読グループ共に春学期及び秋学期、それぞれ15回、合計30回の授業があったが、春学期の最初2回は授業のオリエンテーションに次いで、第3週目に日英バイリンガル文法テストを実施した。試験時間は30分で行い、受験者はマークシートに正答をマークした。

4.2 データ分析

テストデータの分析にはラッシュ測定モデル（Rasch Measurement Models）の中のラッシュ2値的モデルを用いた。データ分析にはWINSTEPS（WINTEPS.com）、RUMM2030（RUMMLAB.com）を用いた。ラッシュ測定モデルでは、テストデータなどの真の意味では間隔尺度ではないデータを間隔尺度に変換し、受験者の能力とテスト項目の難易度を同一の尺度上に並べることが可能とする。受験者の能力と項目の難易度はロジット（logits）という単位であらわされ、理論的にはマイナス無限大からプラス無限大の値をとる。受験者の能力はロジットがプラスの方向に大きくなればなるほど高く、反対に、マイナスの方向に大きくなればなるほど受験者の能力が低くなることを示している。項目の難易度についても同様に、ロジットがプラスの方向に大きくなればなるほど難易度が高く、マイナスの方向に大きくなればなるほど難易度は低くなる。ラッシュモデルでは、ある受験者の能力が一つの項目の難易度と同じであれば、その受験者が当該の項目に正答する確率は0.5、受験者の能力と項目の難易度の差が0.5の場合、すなわち受験者の能力が項目の難易度より0.5大きい場合、当該の受験者がその項目に正答する確率は0.62、差が1の場合、確率が0.73、差が1.5の場合、0.82、差が2の場合、0.88のようになる。反対に、差が-0.5の場合、すなわち受験者の能力が項目の難易度より0.5小さい場合、確率は0.38、-1の場合、確率は0.27のように、受験者の能力が項目の難易度よりも大きければ、その受験者が当該の項目に正答する確率は0.5より上回り、受験者の能力が当該項目の難易度よりも小さければ、その受験者が当該項目に正答する確率は0.5より下回ると想定する。

5 第一回日英バイリンガル文法テスト実施結果

5.1 項目と受験者分布の比較（targeting）

ラッシュ分析では、項目と受験者の分布を比較し、テスト項目が受験者の能力に適したものであるのかどうかを判断する。表3は項目に関するラッシュ分析のまとめを示し、表4は受験生に関する分析をまとめたものである。項目ロジットは平均が0、標準偏差が1になるように設定されている。項目と受験生のロジットを比較すると、項目は平均が0、受験生は0.87となっており、受験生の方が項目よりもロジットが高く、日英バイリンガル文法テストの項目は受験生にとって易しいものであったことが分かる。この情報は図1の項目・受験者マップ（Item-person map）あるいはライトマップ（Wright map）からも読み取れる。図1は日英バイリンガル文法テストのデータをラッシュ分析にかけ、受験者の能力と項目の難易度を同じ尺度上に並べている。マップの左端の+3から-3の数字はロジットを示す。マップ中央の破線の左側が受験者の分布を示し、記号の#は6名分の人数を表し、・は1～5名分の人数を表している。マップの上に行けばいくほど受験者の能力が高く、下に行けば行くほど受験者の能力は低い。

他方、マップの右側は項目の分布を示し、上に行けばいくほど項目は難しく、下に行けば行くほど項目は易しい。この場合、Q42が一番難しい項目であり、Q30が一番易しい項目であることを示している。項目分布、受験者分布に共通して使われているMは平均値、Sは標準偏差1、Tは標準偏差2の値をそれぞれしめす。受験者と項目の分布を比較すると、多くの受験者が平均よりも上位に位置する。他方、項目に関しては、受験者の平均点（図1の中央の破線の左側のM）近くから標準偏差-2の値（図1の中央の破線の左側のT）の間に多く集まっているが、受験者の平均点（M）から上の範囲、特に受験者の平均点（M）近くから標準偏差+1（図1の中央の破線の左側のS）の間では項目が少ない傾向にある。この結果は、前述の結果と同様に日英バイリンガル文法テスト項目が受験者にとっては易しいことを意味しており、文法能力の高い学習者の能力を測定する項目が少ないことを示している。

表3 ラッシュ分析のまとめ：項目

	ロジッツ	INFIT		OUTFIT	
		MNSQ	ZSTD	MNSQ	ZSTD
平均	0.00	0.99	0.20	0.97	0.1
標準偏差	1.05	0.13	2.60	0.26	2.9
最大	2.49	1.36	9.30	1.81	9.9
最小	-2.07	0.70	-3.60	0.45	-4.5

表4 ラッシュ分析のまとめ：受験者

	素点	ロジッツ	INFIT		OUTFIT	
			MNSQ	ZSTD	MNSQ	ZSTD
平均	30.30	0.87	0.99	0.10	0.97	0.00
標準偏差	7.30	0.92	0.17	1.00	0.34	1.10
最大	42.00	2.81	1.70	4.90	2.93	5.20
最小	3.00	-3.07	0.58	-2.60	0.4	-2.40

5.2 信頼性 (reliability)

信頼性の指標として、WINSTEPSの分析結果では項目 (item reliability) は0.99、受験生 (person reliability) は0.82となっており、ともに測定道具として適する範囲である。通常、テストの信頼性を考える場合は後者の指標を参考にする。同様に、クロンバックアルファ信頼係数は0.86、標準測定誤差 (SEM) は2.77であった。(注2)

5.3 不適切項目の分析 (misfit items)

ラッシュ分析では、ある受験者がある項目に正答するかどうかの確率を算出し、それが実際の受験生の能力・項目の困難度パターンとどの程度一致するのかが示す。モデルと実際のデー

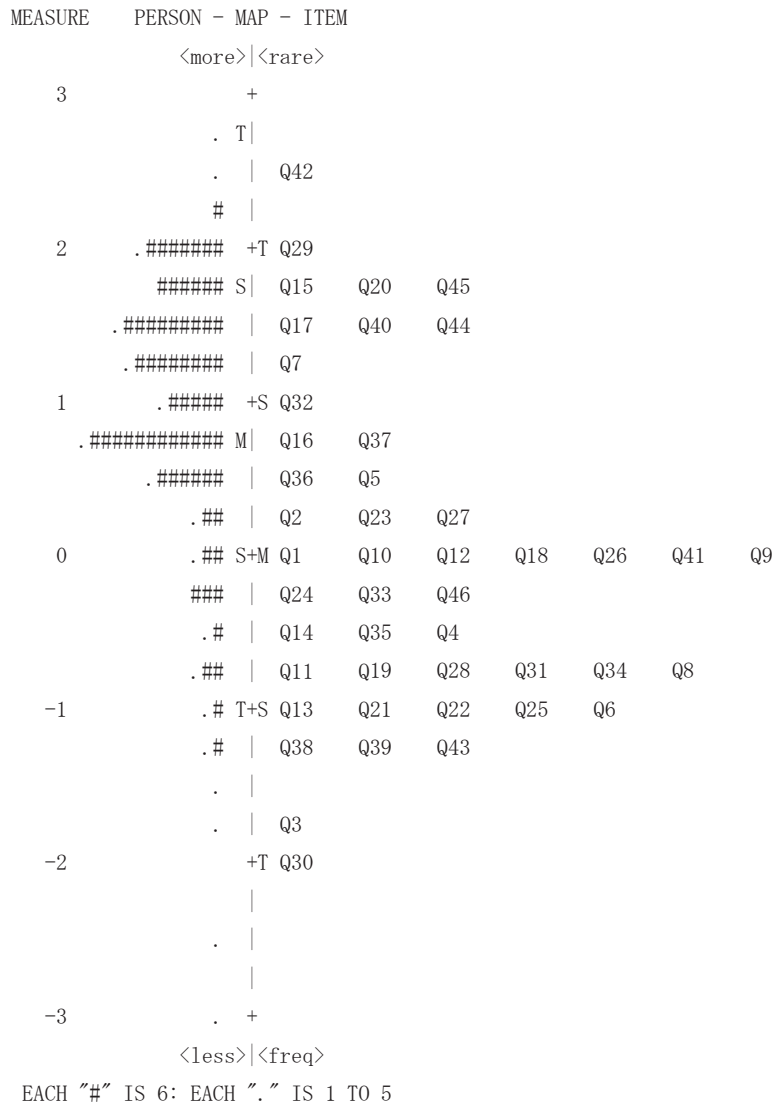


図1 日英バイリンガル文法テストのライトマップ

タの齟齬を指標として示すことにより、ラッシュ分析では測定に使われるテストや質問紙の項目が単一の構成概念（unidimensionality）を測定しているかどうかを確認することができ、テストの質を確保するメカニズムとなる（Bond & Fox, 2015, p. 42）。WINSTEPS では、モデルが算出した期待度（各項目に各受験者が正答する確率）と実際のデータ（各項目に各受験者が正答するかどうかを1、0で表す）の齟齬を残差平均（mean squares）として算出する。表3、4では INFIT MNSQ、OUTFIT MNSQ がそれに該当する。残差平均はその平均が1になるように設定してある。WINSTEPS では、残差平均を t 値、z 値などに変換して標準化しており、表3、

4のZSTDが標準化された残差平均を示し、平均が0、標準偏差が1になるように設定されている。Bond & Fox (2015)は、受験者数が500名未満の場合は、残差平均が0.7～1.3の範囲の項目、受験者が適切とされる。また、標準化された残差平均は-2から+2の範囲がよいとされるが、受験者数が多くなると解釈に注意が必要となる。表3の項目の結果を見る限り、標準化された残差平均の標準偏差が1よりもかなり大きいことから誤差が多いことが分かる。

残差平均と標準化された残差平均の指標から適切とされる範囲外の項目、受験者はミスフィット (misfit) と呼ばれる。第一回目のテスト実施の分析から不適切項目を抽出し、その原因を追究した。結果、項目7、5が残差平均、標準化された残差平均の4つの指標のすべてにおいて、ミスフィットであった。項目7、5はいずれも完了・進行相を問うものであり、類似した問題を例4として下記に述べる。

例4 Since Keisuke Kuwata released his first album, I _____ to his songs.

桑田佳祐が最初のアルバムを発表して以来、ずっと彼の歌を聴いている。

1. have been listening
2. have listened
3. am listening
4. listen

例4では1が正答であるが、2の解答も多くあり、項目5では34%、項目7では58%が誤答の2を選択した。更に、誤答2を選択した受験生の能力が正答1を選んだ受験生の能力よりも若干高かった。これは試験問題としては好ましくない。このような状況を引き起こしているのは、日本語で完了相と完了相・進行形の違いをはっきりと表現するのは難しく、英語の問題文だけで判断すると1、2いずれも正答となり得ると考えられる。対処法としては、両者とも正答にするか、2の選択肢を完了相以外のものに変える必要があると考えられる。

完了相に関連して問題となったのは項目42である。類似した問題を例5として下記に述べる。

例5 This is the first time that _____ Keisuke Kuwata's *Kimienotegami*.

今回初めて桑田佳祐の「君への手紙」を聴いた。

1. I hear
2. I've heard
3. I heard
4. I'd heard

例5の日本語文で動詞の活用だけを見れば、完了相と過去形はともに同じ形態素となる。このため、「今回初めて」の部分が完了相を使うことを示唆している。20%の受験生が正答2を選んだのに対して、44%の受験生が誤答3を選んでいる。正答を選んだ受験生の能力は誤答を選んだ受験生の能力よりも高く問題はないが、ラッシュモデルでは項目42に正答する確率が殆ど皆無、あるいは、非常に低い受験生がこの項目の正答を選んでいるという結果になっているため、残差平均、標準化された残差平均ともに高くなったと考えられる。

残差平均と標準化された残差平均の指標から適切とされる範囲外の項目として、項目15と20がある。これらは共に日本語の語順や日本語の表現をそのまま英語の語順、表現にあてはめたために生じた誤りと判断できる。項目15に類似した問題を例6として下記に述べる。

例6 _____ ? あなたの部屋の番号は何ですか。

1. What is your room's number?
2. What room is your number?
3. What is your room number?
4. Your room's number is what?

例6では、正答の3を選択した受験生は30%であるが、受験生の65%が誤答の1を選択した。これは日本語の「あなたの部屋の」を一字一句逐語訳していると思われる。Your room という名詞句の所有格の形を用いた誤答を選んだ受験生が正答を選んだ受験生の2倍以上にのぼった。誤答1を選んだ学習者の能力は正答を選んだ学習者の能力よりも低く問題はないが、項目42と同様に、ラッシュモデルでは項目15に正答する確率が殆ど皆無、あるいは、非常に低い受験生がこの項目の正答を選んでいるという結果になっているため、残差平均、標準化された残差平均ともに高くなったと考えられる。

5.4 錯乱肢分析

日英バイリンガル文法テストの各項目の錯乱肢が効果的に機能しているのかどうかを分析した。Bachman (2004) は少なくとも当該テストの受験者の10%が当該項目の錯乱肢を選択する場合、その錯乱肢は機能しているとする。Bachman (2004) は古典テスト論の枠組みの中で錯乱肢分析を行う場合に言及しているが(注3)、本研究では各項目の錯乱肢の選択率を調べた。その結果、三つの錯乱肢の選択率が10%を超えていた項目は2つだけであり、選択率が10%以下の錯乱肢が一つある項目は11 (23.9%)、二つある項目は22 (47.8%)、三つある項目は11 (23.9%)であった。このことは、約半分弱の項目では二つの錯乱肢は意図されたように機能しておらず、約4分の1弱の項目の錯乱肢はすべて機能していないと言える。

6. 第1回日英バイリンガル文法テスト実施結果のまとめと今後の展望

第1回日英バイリンガル文法テスト実施結果は次の点にまとめることができる。

- (1) 日英バイリンガル文法テストは文法能力の高い学習者の能力を測定する項目が少ない。
- (2) テストとしての信頼性は確保されている。
- (3) 不適切項目として、4項目が指摘されている。
- (4) 殆どすべての項目の錯乱肢が意図されたように機能していない。

上記(2)に関しては、文法能力の高い学習者の能力を測定する項目を加筆する必要がある。上記(3)に関しては、日本語で完了相と完了相・進行形の違い、また過去形と完了相の違いをはっきりと表現するのは難しく、この点に関しては、日本文の表現に工夫が必要とされる。上記(4)に関しては、テスト結果の解釈の枠組み、すなわち、当該テストを集団基準準拠テスト、目標基準準拠テストとして取り扱うのかどうかではテスト結果の解釈が異なってくる。集団基準準拠テストの枠組みでテスト結果を解釈するのであれば、錯乱肢は意図されたように機能していないと言える。錯乱肢の作成は、先述の統語的要素、規範文法、そして日本語の影響 (transfer) という観点を基にして行っており、そのいずれかの観点が機能していないとも考えられ、今後更なる錯乱肢分析を進める必要がある。

注

- (1) ただし、後述の信頼係数の算出する場合など枠組みが必要な場合は集団基準準拠テストの枠組みに従った。
- (2) 当該テスト結果を集団基準準拠テストの枠組みで解釈する場合、クロンバックアルファ係数は用いられる。
- (3) 当該テスト結果を集団基準準拠テストの枠組みで解釈する場合を想定している。

謝辞

本研究はJSPS 科研費 24520678、15K02807 の助成を受けたものです。

参考文献

- 赤尾美和 (2015). 多読と英語力伸びの関連性—大学再履修クラスにおける多読授業. 『日本多読学会紀要』 8, 39-50.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

- Bachman, L. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bejar, D., Hunt, A., & Kite, Y. (2011). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning*, 61(4), 1-39.
- Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*, 1(1).
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New York: Routledge.
- Carr, N. (2011). *Designing and Analyzing Language Tests*. Oxford: Oxford University Press.
- Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English: A Comprehensive Guide. Spoken and Written English. Grammar and Usage*. Cambridge: Cambridge University Press.
- Cho, K. S., & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, 37(8), 662-667.
- Edinburgh Project on Extensive Reading. (1992). *The EPER guide to Organizing Programmes of Extensive Reading*. Institute for Applied Language Studies, University of Edinburgh.
- Elley, W. B., & Mangbhai, F. (1981). *The impact of a book flood in Fiji primary schools*. New Zealand Council for Educational Research/Institute of Education: University of South Pacific.
- Ellis, N. (2002). Frequency effects on language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188.
- Furukawa, A. (2008). Extensive reading program from the first day of English learning. *Extensive Reading in Japan*, 1(2), 11-15.
- Grabe, W. (2009). *Reading in a Second Language*. Cambridge, U.K.: Cambridge University Press.
- Hafiz, F. M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT Journal*, 43(1), 4-13.
- Hamrick, P., & Rebuschat, P. (2014) Frequency effects, learning conditions, and the development of implicit and explicit lexical knowledge. In J. Connor-Linton & L. Amoroso (Eds.), *Measured language: Quantitative studies of acquisition, assessment, and variation* (pp. 125-139). Washington, DC: Georgetown University Press.
- Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a foreign language*, 26(2), 17-33.
- Irvine, A. (2011). *Extensive Reading and the development of L2 writing*. Paper presented at the First Extensive Reading World Congress, Kyoto.
- Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language*, 20(1), 70-91.
- 金谷 憲／英語診断テスト開発グループ（2006）. 『英語診断テスト開発への道』 東京：英語運用能力評価協会
- Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the Input Hypothesis. *The Modern Language Journal*, 73 (iv), 440-464.
- Lai, F. K. (1993). The effects of summer reading course on reading and writing skills. *System*, 21(1), 87-100.
- Maruhashi, K. (2011). *The Effect of Extensive Reading on Japanese EFL Learners' Grammatical*

- Competence*. Paper presented at the American Association for Applied Linguistics Annual Conference in Chicago.
- Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, 25(1), 99-102.
- Mermelstein, A. D. (2015). Improving EFL learners' writing through enhanced extensive reading. *Reading in a Foreign Language*, 27(2), 182-198.
- Murphy, R. (2012). *English grammar in use* (4th ed.). Cambridge: Cambridge University Press.
- 永野芳郎 (1978). 『英語学概説』 東京: 英宝社
- Nishizawa, H., Yoshioka, T., & Fukada, M. (2010). The impact of a 4-year extensive reading program. In A. M. Stoke (Ed.), *JALT 2009 Conference Proceedings*, 632-640.
- Rodrigo, V., Krashen, S., & Gribbons, B. (2004). The effectiveness of two comprehensible-input approaches to foreign language instruction at the intermediate level. *System*, 32(1), 53-60.
- Sheu, S. P-H. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter*, 36(2), 8-26.
- 鈴木寿一 (1996). 読書の楽しさを経験させるためのリーディング指導『新しい読みの指導』(渡辺編) 116-123. 東京: 三省堂
- Swan, M. (2005). *Practical English usage* (3rd ed.). Oxford: Oxford University Press.
- 高瀬敦子 (2010). 『英語多読・多聴指導マニュアル』 東京: 大修館書店
- Takase, A. (2007). Japanese high school students' motivation for extensive L2 reading. *Reading in a Foreign Language*, 19(1), 1-18.
- Takase, A. (2008). The two most critical tips for a successful extensive reading. *Kinki University English Journal*, 1, 119-136.
- Takase, A. (2009). The effects of SSR on learners' reading attitudes, motivation, and achievement: A quantitative study. In A. Cirocki (ed.), *Extensive Reading in English Language Teaching* (pp. 547-560). Munich: Lincom.
- Takase, A., & Otsuki, K. (2012). New challenges to motivate remedial EFL students to read extensively. *Apples - Journal of Applied Language Studies*, 6(2), 75-94.
- Walker, C. (1997). A self-access extensive reading project using graded readers. *Reading in a Foreign Language*, 11(1), 121-149.
- 渡辺政寿 (2014). 公立中等教育における多読指導の成果—GTEC, 進研模試のデータから『日本多読学会紀要』7, 27-28.
- Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, 36, 661-672.