

## 英文和訳の評価:

# 分析的評価項目の策定と検証的因子分析による妥当性の検討

広島大学大学院 石原知英

This study tries to establish an analytic evaluation scale for translation so that the translation task will be revisited affirmatively in the field of English education in Japan. Firstly, three criteria of translation assessment (i.e. accuracy, readability, appropriateness) and subsequently nine items were determined, according to the literature review on the issue of translation quality assessment. Using this evaluation scale with nine items, seven judges scored twenty-three translated products by university students. By confirmatory factor analysis using structure equation modeling, the scale developed in this study was shown as valid enough to assess the translation quality. However, the correlation of the items and fit indexes showed that some items needed further revision.

### 1 問題と目的

外国語教育における翻訳の歴史は、古くはヨーロッパでのラテン語やギリシア語などの古典語の教育にまで遡ることができる。こうした古典語の教育の方法を現代の外国語教育に取り入れたのが、(文法)訳読式教授法と呼ばれるものである。日本におけるこうした教授法は、明治期の「直読直解」以来見られるが、現代にもなお根強く残っている。

近年、しばしば和訳<sup>1</sup>は批判の対象となってきた。静(2002)は和訳の問題点として、教室内で日本語を扱う時間が増え、英語を使う時間が減ることや、日本語を介して英語を理解する習慣がつくこと、日本語によって内容理解の評価が左右されうることなどを指摘している。

こうした議論において、和訳問題は、文法事項の確認や単語の意味理解の確認を含んだ、いわゆる英語の理解度のテストとして捉えられている。この前提は、英文和訳を擁護する立場の研究者の多くが共有しており、和訳によってどの程度内容理解が測れるのかという議論がなされている(例えば Buck, 1992 や Ito, 2004 など)。しかしこうした視点は、翻訳の産出の側面を考慮しておらず、また要約や英問英答などの方法によって、より効率よく成しえるという点で、先の否定的な指摘に対して効果的な反駁を行うことが難しい。

そこで本研究では、翻訳における産出の側面に焦点を当て、英文和訳を捉えなおすことを目指す。そのため、英語の理解のみではなく、日本語の分かりやすさや言葉の選択、文体などを含めて英文和訳を評価するための、新しい項目を策定する。そうすることで、翻訳の産出の側面にも目を向け、英語と日本語の比較に基づいた統合的な言語能力を問題にしたい<sup>2</sup>。具体的には、まず翻訳理論研究における議論を参考に、分析的評価の枠組みを

策定する。その後、策定した評価項目を用いて実際に評価を行い、その評価尺度の妥当性を検証的因子分析によって検討する。そうすることで、英文和訳の扱いを肯定的に捉えなおすための基盤整備を目指すのである。

## 2 評価尺度の策定

### 2.1 翻訳評価に関する議論の概観

そもそも、よい翻訳とは何か。この問いに答えるため、翻訳とは何か、という問いに立ち戻りながら、翻訳研究における評価に関する議論の変遷を概観する。

従来翻訳は、主に正確さ (Accuracy) の観点によって評価されてきた。この観点の背後には、翻訳とは原文を忠実に再現することである、という前提がある。したがって翻訳者に求められるのは、原文の構造や意味を模倣することであった。そういう意味で翻訳者は反逆者 (Traduttore; traditore) であり、また、翻訳は不可能である、といった極端な見方がなされてきたのである。時代を追うごとに、等価 (Equivalent) の考え方は、形式、内容、効果、テキストタイプと変化していくが、どの場合でも、原文との比較における等価に最大の価値が置かれてきた。

こうした考え方に対して大きな影響を与えたのは、その後の機能主義的な考え方の中で生まれた Vermeer (1989/2004) らのスコポス理論 (Skopos theory) である。この考え方は翻訳に、起点言語テキストからの脱却をもたらした。スコポス理論では、翻訳を一種の言語行為とみなし、その背後に行為の目的があり、訳す行為が持つ機能があると考えた。その翻訳の目的がすなわちスコポス (ギリシア語で目的を意味する) である。

こうした機能主義的な考え方によって、翻訳テキストはその独自の価値を獲得し、同時に翻訳者はその独自の地位を勝ち得たといえる。翻訳者は、その原作者や原文テキストではなく、自らの翻訳とその読者に対して責任を持ち、ある程度の自由裁量によって、その読者と共同的に翻訳を成立させることを目指す、という見方が確立されたのである。

したがって翻訳は、「よい (Good)」「正しい (Correct)」「正確 (Accurate)」「誠実 (Faithful)」などの観点からでは測りえず、むしろ「語用論的に妥当である (Pragmatically adequate)」あるいは「語用論的に適切である (Pragmatically appropriate)」といった言葉による評価が妥当である (Schäffner, 1997)。なぜなら藤濤 (2007) が指摘するように、異言語文化を越える過程で起点テキストと目標テキストとの間にずれが生じるのはある程度必然的であり、そのずれは異文化コミュニケーションを成功させるためにとられた選択行為の結果であると考えられるからである。こうしてスコポス理論によって、原文との等価を目指す従来の翻訳評価のあり方が、翻訳されたテキストが、読み手が属している文化の中でどのように意図された役割を果たしているのか、という評価へと変化した。

このようにスコポス理論では、翻訳の質の評価に関する理論的な枠組みが提示された。本研究では、こうした理論を受け、どのように学習者の英文和訳を評価したらよいか、どういった観点によって学習者の翻訳の質を数値化することができるか、を問題にする。

そうした具体的な問題に取り組んだ研究には、例えば藤濤 (2004) がある。藤濤では、ヘッセの『車輪の下』とその日本語訳 2 作品からの引用をもとに、ある部分で採用された「翻訳の方法」(例えば脚注の使用、カタカナ外来語の使用、文構造の変更、削除、意識など) が、全体の「スコポス (目的)」にかなっているか、「結束性」<sup>3</sup>があるか、「内容」、「形式」、「効果」という 3 つの観点での原文への「忠実性」があるか、という点を 5 点満点 (5AB) で評価しようとした。しかしこれは、該当箇所における翻訳方法の選択が目的にかなって

いるかを評価するものであり、翻訳の質を点数化し、よしあしを判断するものではない。本研究では、教育環境を意図した評価尺度の開発を目指すため、スコボス理論の考え方をもとにして、具体的な尺度化に向けた検討が必要である。

より実用的な意味での評価の規準としては、The Institute of Linguist (IoL) の *Diploma in translation* がある。その観点は、(1) 内容理解、正確さとレジスター、(2) 文法、結束性と一貫性、全体の構成、(3) 技術的な観点：句読法、綴り、固有名詞、日付や図表など、の3つ (IoL, 2008, pp.11-12) であり、それぞれ、(1) どの程度正確に原文テキストを理解しているか、またその理解がどの程度正確に翻訳テキストにおいて再現されているか、(2) 翻訳テキストの言語の質はどうか、文法的に正確か、一貫性があり意味的にもまとまりのある文章になっているか、(3) 綴りや句読法、段落構成、図表などが正確に処理されているか、といった点が問題とされている (p.13)。

この評価尺度は翻訳の資格検定のためのものであり、試験内容は、それぞれの受験者が選んだ3種類のジャンルの課題文を訳すというものである。当然検定試験という社会的意味合いからも、プロ、あるいはプロ志望の翻訳者を対象としているため、それぞれの課題文はボリュームがある。そのため観点2や観点3の構成や段落の問題、あるいは観点1にあるような注の使用の問題なども考慮する必要が生じる。一方、本研究が意図している教室環境においては、課題文はたいてい1、2文、多くてもせいぜい1段落であることが多く、段落の再構成や注が必要であることは少ない。また、図表や日付などを訳すこともそれほど多くないだろう。そういう意味で、上記の観点を再構成しなおす必要がある。

## 2.2 分析的評価尺度の観点と評価項目

ここまでの理論的背景と実際の評価尺度を、3名の大学教員の協力のもと総合的に検討し、本研究では以下のような3つの観点を策定した。

- A. 正確さ：起点テキスト（原文）の意味内容が正確に理解されているか
- B. 分かりやすさ：訳文テキスト（訳文）が分かりやすく自然な表現であるか
- C. 適切さ：原文の文脈や訳文の読者を考慮した適切な文体が選択されているか

さらに議論を重ね、実際に評価場面で用いることができるよう、各観点に基づく項目の策定を目指した。具体的には、それぞれの観点（上記 A, B, C）に対し、語や1文に関する指標（1, 4, 7）、2文以上の繋がりに関する指標（2, 5, 8）、文章のまとまりや全体に関する指標（3, 6, 9）の3つを設定し、計9項目を策定した。その際、この評価尺度が、ある程度どのようなテキストに対しても応用できるか、採点が簡単であるか、段階評価が可能であるか、合計点によって総合的な評価が可能であるか、様々な質の訳を弁別的に評価できるか、といった諸条件を考慮した。なお、1, 2, 3は「A 正確さ」を、4, 5, 6は「B 分かりやすさ」を、7, 8, 9は「C 適切さ」を、それぞれ代表する項目である。

1. 原文の語彙や文法を正しく理解して訳している
2. 原文の文の繋がりや段落の構成を正しく理解して訳している
3. 原文の伝える情報や作者の意図を正しく理解して訳している
4. 必要に応じた言い換えや省略、補足などがなされている
5. 訳文の各文が日本語として自然である
6. 訳文全体が読者にとって分かりやすい
7. 原文の文脈に即して訳文の語や表現を選択している
8. 訳文の文体が一貫している

### 9. テクストのジャンルや目的に応じて訳している

本研究において策定したこれらの観点および項目は、あくまで暫定的なものであるため、さらなる実証的な検討が必要である。そのため第3節以降では、本尺度を用いて得られたデータを分析し、妥当性の検討を行い、改善のための示唆を得ることを目指した。

## 3 評価尺度の検討

### 3.1 目的と方法

第2節で策定した分析的評価尺度の妥当性を、記述統計量と相関行列及び構造方程式モデリングを用いた検証的因子分析によって検討する。

### 3.2 評価対象

評価対象は、大学3年生23名（英語教育専攻22名、初等教育専攻1名、なお初等教育専攻の学生も英語を副専攻としている）による翻訳課題であった。課題はウェブ上のニュース記事の一部であった（付録1を参照）。

全体を踏まえて訳すことができるよう、はじめに全文を提示し、10分間で内容を理解するよう指示した。その後、課題範囲部分である最後の2段落を訳すように指示した。時間制限を設けなかったため、所要時間には個人差があったが、概ね15分程度（平均856.4秒、標準偏差275.1秒）であった。タスク中の辞書の使用は許可された。

### 3.3 評価者

評価者は英語教育を専攻する日本人大学院生7名であった。翻訳を専門としておらず、また翻訳を評価した経験がないため、採点に先立ち、評定のためのチュートリアルを実施した。具体的には、課題文の内容理解および評価尺度の説明などを20分程度で行った。そうして、課題テキストの十分な理解と評価に関する共通認識を確認した。

### 3.4 評価項目

評価項目は、先の第2節で策定した分析的評価尺度9項目に、総合的な評価尺度として「10. 総合的に判断して、あなたはこの訳に何点与えますか？」という項目を加えた、計10項目であった。項目1から項目9までは5段階による評定（「1 不満足」から「5 満足できる」）、項目10のみ10段階による評定であった。

### 3.5 調査および分析の手順

協力者による訳をすべてタイプし、A4用紙1枚に、上から課題の英文、訳文、評価尺度、の順になるよう印刷した。評定順序の影響を抑えるため、訳の並びが異なるセットを3種類用意し、それらのうちの1部を配布し、評定を依頼した。

分析に際しては、評定者と評定対象の組み合わせを1ケースとして捉え、すべての組み合わせ（評定者7名×評価対象23名）を別のケースとみなした。そのため計161ケースとして以下分析を行った。

まずMicrosoft Excel 2003およびSPSS11.5を用い、記述統計量、相関行列などから、収集されたデータの概要をまとめた(4.1)。その後Amos 4.0.2により、構造方程式モデリングを用いた検証的因子分析を行った(4.2)。最後にモデルの修正といくつかのモデルの比較検討を行い、尺度の改善に関する示唆を得た(4.3)。

#### 4 結果と考察

##### 4.1 記述統計量と相関行列

各項目の記述統計量および各観点内の信頼性係数を表 1 に、各項目間の相関行列を表 2 に示す。

表 1 各項目の記述統計量および各観点内の信頼性係数 (N= 161)

項目	平均	標準偏差	最小	最大	度数分布 (%)					α
					1	2	3	4	5	
1	3.69	1.07	1	5	1.9	15.5	19.9	37.3	25.5	
2	3.75	0.93	2	5	0.0	10.6	27.3	39.1	23.0	
3	3.73	0.99	2	5	0.0	14.3	23.6	37.3	24.8	
4	3.54	0.86	2	5	0.0	9.9	40.4	35.4	14.3	
5	3.61	1.01	2	5	0.0	18.0	23.6	37.3	21.1	
6	3.47	0.96	2	5	0.0	19.3	29.2	37.3	14.3	
7	3.71	0.91	2	5	0.0	9.3	31.7	37.9	21.1	
8	4.35	0.75	2	5	0.0	1.9	10.6	37.9	49.7	
9	3.97	0.92	2	5	0.0	6.2	24.8	34.8	34.2	
10 <sup>a</sup>	6.91	1.81	2	10	—	—	—	—	—	

a. 総合的評価の項目である項目 10 のみ 10 件法によって評定を得た。

表 2 評価項目間の相関行列 (N= 161)

項目	正確さ			分かりやすさ			適切さ		
	1	2	3	4	5	6	7	8	9
1	—								
2	<u>.59</u>	—							
3	<u>.68</u>	<u>.74</u>	—						
4	.47	.47	.51	—					
5	.31	.43	.46	<u>.43</u>	—				
6	.47	.56	.61	<u>.58</u>	<u>.72</u>	—			
7	.52	.56	.67	.50	.48	.56	—		
8	.35	.32	.27	.30	.15	.27	<u>.39</u>	—	
9	.34	.38	.43	.48	.38	.47	<u>.52</u>	<u>.50</u>	—
10	.74	.74	.83	.65	.61	.72	.72	.32	.47

注. 各観点内の項目間相関係数には下線を付した。

全体的な傾向として、項目 1 以外では 1 点が与えられることはなく、各項目とも 3 点から 4 点を中心に評定点が与えられているようであった。また項目 8 は他の項目に比べて点数の分布に偏りがあり、やや天井効果が見られたといえる。

各観点内の信頼性係数は、「正確さ」と「分かりやすさ」の観点が一般的な基準とさせている  $\alpha = .80$  を越え、「適切さ」の観点も  $\alpha = .72$  であった。また、表 2 中で下線を付した各観点内の相関係数においても、相対的に高い値が得られたことから、各観点の内的整合

性は十分に高いと解釈できる。

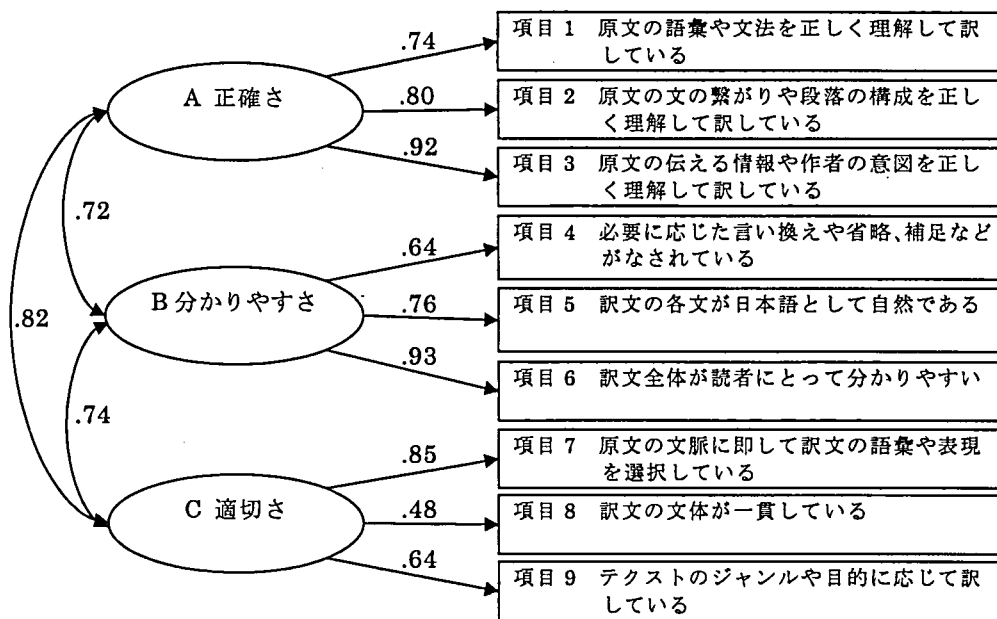
また表 2 中、各項目と総合的評価（項目 10）との相関係数は、項目 8 を除いてどの項目もかなり高い値であった。総合的評価は、様々な印象や要因を汲み取った評価であるとされており、総合的評価と相関の高い項目は、並存的妥当性が高いと捉えることが可能である（山西, 2004）が、本研究では、分析的評価の直後に総合的評価を行ったため、その解釈には慎重になる必要がある。

なお、9 項目全体の信頼性係数は  $\alpha = .89$ 、総合的評価（項目 10）と分析的評価項目の合計（項目 1 から 9 の合計点）の相関は  $r = .74$  であった。

#### 4.2 検証的因子分析

続いて検証的因子分析の結果を示す。

本研究では構造方程式モデリングを援用し、事前に想定した 3 つの観点と 9 項目による評価のモデルを検討した。具体的な手順としては、まず、事前の想定どおり、3 つの因子間にはそれぞれ共分散を仮定し、この 3 因子がそれぞれ 3 つの項目に影響を与えるというモデル（図 1 参照）に従い分析した。その後、モデル全体の適合度および各因子から項目へのパス係数を推定し、それらを妥当性の指標として検討した。なお図 1 中、潜在変数（観点）は楕円で、観測変数（各項目）は四角で表記し、測定誤差は省略した。また単方向の矢印は因果関係を、双方向の矢印は共変関係を示している。



$\chi^2(24) = 59.871, p < .001; GFI = .925; AGFI = .859; NFI = .921; CFI = .951; RMSEA = .097$

図 1. 3 因子モデルにおける各パスの標準化係数及び因子間相関と適合度指標

適合度指標のうち、GFI, AGFI, NFI, CFI の値は 1 に近いほど適合がよく、経験的に .90 以上を判定基準とすることが多いとされている（豊田, 1998）。本モデルのこれらの数値はどれも .90 を越えたため、概ね満足いく値であった。RMSEA は値が小さいほど適合がよ

く、.08 以下、あるいは .05 以下を判定基準とすることが多い（豊田, 1998）。ただし、サンプルサイズが小さい場合やモデルが簡素な場合には、比較的高い推定値となることが知られており（星野・岡田・前田, 2005）、本研究のように 9 つの観測変数と 3 つの潜在因子のモデルの場合には、ある程度満足のいく値であると捉えてもさしつかえないだろう。一方  $\chi^2$  値は、やや満足のいかない結果であった。ここでは帰無仮説を「モデルが適切である」とするため、5 %水準で棄却されるということは、モデルが適切であると言えない可能性がある、ということの意味するからである。ただし  $\chi^2$  値はサンプル数に影響を受けやすく、この数値だけでモデルを捨て去る必要はないとする見方もある（豊田, 1998）。

適合度指標は様々なものが提案されている。しかし、それぞれの指標はモデルの複雑さやサンプル数などに影響されるため、どの指標のどの値で、どの程度適合度があると判断するのは、一概に決めることができない。先行研究においても、いくつかの指標を総合的に判断するという姿勢が一般的である。したがって本研究でも、図 1 中に列挙したように、先行研究で多く用いられている指標を中心に総合的に検討した。その結果、事前に想定したモデルがある程度得られたデータと適合していたと結論づけることにした。

#### 4.3 モデルの修正

本研究では本稿第 2 節で策定した分析的評価尺度の妥当性の検討を目的としたため、4.2 のモデルの適合度指標の検討によってひとまずの結論を得たといえる。しかし構造方程式モデリングの利点であるモデルの改善に関しても付記することで、データとモデルのずれの考察及び今後の項目改善への方向性を示すことが可能になる。

事前の想定によるモデル（図 1）を、便宜的にモデル A と呼ぶ。このモデル A に対して、「正確さ」の観点から項目 7 へパスを加えたものをモデル B（「正確さ」と「適切さ」から項目 7 へのパスがある）、モデル B の「適切さ」から項目 7 へのパスを削除したものをモデル C とした。モデル B 及び C に関して項目 7 を中心に修正を行ったのは、表 2 から明らかなように、項目 7 と項目 1, 2, 3 の相関が相対的に高くなっていたため、事前の想定では「適切さ」を測定するためのものであった項目 7 は、実際は「正確さ」の影響を大きく受けた可能性がある。またモデル A から項目 8 を削除したものをモデル D として検討した。これは、検証的因子分析の結果から、項目 8 へのパスが相対的に低かったため、項目 8 の構成概念妥当性が低いと考えられるためである。

表 3 に、事前に想定したモデル（モデル A）と、3 つの修正モデル（モデル B, C, D）の適合度指標を示した。モデルの比較に際しては AIC の指標も加えた。この数値は小さいほどよいモデルであると判断されるが、自由度の小さいモデルをよいモデルであると判定する性質がある。表 4 には、4 つのモデルの各パスの標準化係数及び観点間相関を示した。

表 3 から、どの修正案も、当初のモデルよりも適合度が高くなることが明らかとなった。ただし本研究のような検証的因子分析におけるモデルの修正に関しては、慎重な態度で臨む必要がある。それは、数値のみを判断基準としてモデルを採択するのではなく、あくまで解釈可能なモデルを採択する必要があるためである。本研究では事前に想定した 3 因子モデルがどの程度妥当であるかを検討することが主眼であったため、データをうまく説明するモデルを探索的に探ることは意図していない。そのため、こうした修正モデルを参考に、今後どのように項目を修正すればよいかを考える必要がある。

項目 7 に関しては、「適切さ」の観点からのパスと共に「正確さ」の観点からもパスがあるモデル B がより適合度が高い結果になった。本調査における項目 7 は、「原文の文脈

表 3 修正モデルの適合度指標

モデル	df	$\chi^2$	$\chi^2/df$	p	GFI	AGFI	NFI	CFI	RMSEA	AIC
A	24	59.871	2.49	<.001	.925	.859	.921	.951	.097	101.871
B	23	40.740	1.77	.013	.946	.893	.947	.976	.069	84.740
C	24	54.746	2.28	<.001	.929	.867	.928	.958	.089	96.746
D	17	29.712	1.75	.028	.957	.909	.957	.981	.068	67.712

表 4 修正モデルにおける各パスの標準化係数及び各観点間の相関

項目	観点		
	正確さ	分かりやすさ	適切さ
1	.74/.73/.73/.73	—	—
2	.80/.80/.80/.80	—	—
3	.92/.92/.91/.92	—	—
4	—	.64/.64/.64/.64	—
5	—	.76/.76/.76/.76	—
6	—	.93/.93/.93/.93	—
7	—1.56/.64/—	—	.85/.31/—1.85
8	—	—	.48/.58/.57/—
9	—	—	.64/.85/.88/.61
観点間相関			
正確さ	—		
分かりやすさ	.72/.73/.74/.72	—	
適切さ	.82/.55/.58/.84	.74/.60/.58/.76	—

注. 数値はすべて、モデル A / モデル B / モデル C / モデル D である。

モデルにおいて該当のパスがない場合は、係数を—で示した。

に即して訳文の語や表現を選択している」という項目であったが、原文の文脈が正しく理解されているかどうか、という点に関しての判断がより大きく影響を与えたと考えられる。この項目を「適切さ」の観点を測るために用いる場合、今後は多少の修正が必要であるといえるだろう。具体的には、より適切さに影響を受け、正確さからの影響が少なくなるような文言に修正する必要があると考えられる。例えば「訳文における語や表現の選択が適切である」とすると、ある程度原文との対応によらない評価が可能になるかもしれない。

また項目 8 を削除したモデル D は、比較したモデルの中で最も適合がよかった。そのため、項目 8 に関しても今後検討する必要があることが示唆された。ただし、単純な項目の削除は、信頼性にも影響を与える可能性があるため、今後は別の項目に変えて再検討する必要があるだろう。また 4.1 で指摘したように、項目 8 には天井効果が見られた。本調査の訳文は、どれもある程度の基準を満たしていたと捉えることができるが、別の協力者に対してこの項目で実施した場合には、また異なる結果になる可能性も残されている。

項目に修正を加えれば全体の適合度に差が生じることは容易に想像できる。そのため、項目の改善についてさらに検討し、そのモデルに対して再び検証を行うことが望まれる。



## 5 まとめ

本研究では、英文和訳のあり方を再考することを目指し、従来の英語の理解のみではなく、日本語の産出を含めた評価のあり方を提案した。具体的には、先行研究などをまとめ、3つの観点と9項目を策定し、その妥当性の検討を行った。その結果、いくつかの改善点は残されているものの、ある程度妥当性の高い尺度であることが結論づけられた。こうした尺度を用いて学習者の和訳を評価することによって、日本語の産出にも注意を向けた訳を目指すことが可能になり、ひいては教室内で英語と日本語をより意識的に、統合的に扱うことが可能になると考えられる。

最後に、今後の課題として以下に3点を挙げる。1点目は信頼性の問題である。今回は5件法によって評価を行ったが、それぞれの採点基準がやや曖昧であった。また、評定者や項目によってどの程度ばらついたのかという点も、同じデータを用いて検討する必要がある。2点目は実用性の問題である。本研究では7名の評定者が9項目で評価を行ったが、実際にはこうした手間をかけることが難しい状況が多い。したがって、何人の評定者による何項目の評価であれば、どの程度の信頼性を保つことができるのか、という実用的な問題も、今後さらに取り組んでいく必要がある。3点目は応用可能性の問題である。本尺度がどの程度他のテキスト、他の評定者、他の評価対象に応用可能であるか、という点も、今後の検討が必要である。たとえば詩の訳を評価する場合には、他の項目や観点が必要になるかもしれない。また内容理解が困難なテキストを用いた場合には、正確さの比重が重くなる可能性が高い。そういう意味では、今回の提案をもとに、今後、諸条件を変えた追調査を行いながら、少しずつ修正を加えていく必要があるだろう。

## 註

1. 本稿では「和訳」と「翻訳」を区別しない。Jakobson (1959/2004) の翻訳の分類に従えば、どちらも言語間翻訳 (Interlingual translation) であり、言語内翻訳 (Intralinguistic translation) や記号間翻訳 (Intersemiotic translation) とは区別されるためである。
2. 翻訳能力と英語教育の目的との関連や、そうした目的の妥当性についての論考は別の機会にゆずりたいが、母語である日本語と外国語である英語を含んだ言語能力の概念や、それを踏まえた英語教育の目的論については石原 (2007) で触れている。またその中で翻訳の役割については Zabalbeascoa (1997) などの考え方が参考になるだろう。
3. 藤濤 (2004) は「結束性」を、読者にとっての分かりやすさという意味で用いている。
4. 信頼性に関しては、同じデータをもとに Ishihara (in press) で論じている。

## 謝辞

評価の観点及び項目策定に際しては、広島大学の濱口脩先生、中村愛人先生、小野章先生にご協力いただき、議論を進めました。

## 引用文献

- Buck, G. (1992). Translation as a language testing procedure: Does it work? *Language testing*, 9, 123-148.
- Institute of Linguists. (2006). *Diploma in translation: Handbook and advice to candidates*. Retrieved March 13, 2008, from <http://www.iol.org.uk/qualifications/>

- Ishihara, T. (in press). Developing an analytic evaluation scale for English-Japanese translation: Considering its reliability using generalizability theory. *ARELE*.
- Ito, A. (2004). Two types of translation tests: Their reliability and validity. *System*, 32, 395-405.
- Jakobson, R. (1959/2004). On linguistic aspects of translation. In L. Venuti. (Ed.). *The translation studies reader* (2nd ed., pp. 138-143). New York: Routledge.
- Kubota, H. (2007, September 29). Colorless ink that produces a riot of color. *The ASAHI Shimbun*. Retrieved September 29, 2007, from <http://www.asahi.com/>
- Schäffner, C. (1997). From 'good' to 'functionally appropriate': Assessing translation quality. *Current issues in language & society*, 4, 1-5.
- Vermeer, (1989/2004). Skopos and commission in translational action. (A. Chesterman, Trans.). In L. Venuti. (Ed.). *The translation studies reader* (2nd ed., pp.227-238). New York: Routledge.
- Zabalbeascoa, P. (1997). Language awareness and translation. In L. van Lier & D. Corson (Eds.), *Encyclopedia of language and education volume 6: Knowledge about language* (pp.119-130). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- 石原知英. (2007). 「指導要領の変遷から見る学校英語教育の目的論の展開—豊かな言語感覚の育成に焦点を当てて—」『中国地区英語教育学会紀要』,37, 91-100.
- 静哲人. (2002). 『英語テスト作成の達人マニュアル』東京: 大修館書店
- 豊田秀樹. (1998). 『共分散構造分析<入門編>構造方程式モデリング』東京: 朝倉書店
- 藤濤文子. (2004). 「文学テキストの翻訳にみる異文化—評価分析のための方法—」『ドイツ文学』,115, 30-42.
- 藤濤文子. (2007). 『翻訳行為と異文化コミュニケーション—機能主義的翻訳理論の諸相—』京都: 松籟社
- 星野崇宏・岡田謙介・前田忠彦. (2005). 「構造方程式モデリングにおける適合度指標とモデル改善について: 展望とシミュレーション研究による新たな知見」『行動計量学』, 32, 209-235.
- 山西博之. (2004). 「高校生の自由英作文はどのように評価されているのか—分析的評価尺度と総合的評価尺度の比較を通しての検討—」*JALT Journal*, 26, 189-205.

#### 付録 翻訳テキスト (課題範囲のみ) と協力者の訳例

They encircled the element with organic compounds that can collect light from ultraviolet rays, and manipulated them at the molecular level. As a result, the rare-earth metal's ability to emit red light was heightened 100 to 1,000 times.

The colorless ink can be printed even on glass or vinyl. Three-dimensional objects that shine in red can be made from the ink mixed into transparent plastic.

研究者たちは、その成分を、紫外線から光を集めることのできる有機化合物でとりかみ、それを分子レベルに処理した。結果として、地球上にはめずらしい、金属が赤い光を発するという能力は100倍から1000倍に高まった。

この、無色のインクは、ガラスの上にもビニールの上にもうつる。赤色にひかる3Dの物体は、透明なプラスチックにまぜたインクから作ることができる。