

感性の内省的言語化分析への第二外国語学習および翻訳分析の応用

福島宙輝^{*1}

Hiroki Fukushima

^{*1} 慶應義塾大学

Keio University

1. はじめに

味わいなどの感性にかかわる内省的な言語化は、一種の問題解決過程ととらえられる。問題解決は基本概念として初期状態、オペレータ、ゴール状態をもつ。感性の言語化においては初期状態は知覚情報(センサ情報)と言語知識、ゴール状態は適切な言語化ができた状態ととりあえず措定できる。オペレータには対象(例えばワイン)に関する知識や、どの感覚情報とどの言語表現が対応するかといった知識が含まれる。なお知覚については受動的な営為ではなく、能動性をもって対象との間に相互浸透が生じた結果として対象像が浮かび上がるという現象学的対象構成モデルを筆者は持っているが、とりあえず議論を前進させるために initial state として知覚情報が与えられているものとする。

表 1

		ゴール状態	
		明確	不明確
初期情報	あり	領域 A (良定義問題) ソムリエの言語表現など	領域 B 英作文自動採点 翻訳など
	不足	領域 C	領域 D 感性の言語表現

本研究の問題空間を示すために、ゴール状態と初期状態(初期情報)の有無からマトリックスを作成すると、表 1 のようになる。味わいの言語表現のなかでも、例えばソムリエの言語表現は、初期状態としての香りの知覚が明確で(知覚能力があり、知覚すべき香りが定義されている)、用いるべき言語表現のセットがあらかじめ定義されている。またゴール状態として香りが言葉によって理想的に表現された状態をあらかじめ想定出来る。領域 A に入り、良定義問題として考えることができる。一方で筆者の扱う感性の内省的な言語化は、知覚情報として何を知覚すべきかというあらかじめ規定されず、またどのような言語表現がゴール状態として求められるかも初めから明らかではないという問題、すなわち領域 D の問題となる。

2. 自由英作文の分析手法の応用

2.1 先行する研究

自由英作文の評価基準は、a. 全体的評価(テキスト全体の印象)、b. 分析的評価(語彙や文法などの項目別に評定)、c. 特定要因の評価(修辭的特徴など)の 3 つの基準に大別される

[Perkins, 1983]. このなかでも、自動採点システムの開発のように、定量的指標を用いて英作文を評価する試みは、主に全体的評価について研究が進められてきた。

こうした研究の目的は評価基準の設定、すなわち既存テストなどによる学習者の“英語力”を目的変数とした場合の説明変数を求めることである。

主な先行研究を幾つかあげると、石岡[2004; 2009]では語彙の重要性が強調され、単語の出現頻度ベクトル、文や句の長さ、論理構成のための語、修辭句などが重視されている。[小林・金丸, 2012]では以下の 12 の説明変数を採用する。すなわち(1)総語数、(2)異なり語数、(3)異語率、(4)平均単語長、(5)平均文長、(6)助動詞の頻度、(7)冠詞の頻度、(8)代名詞の頻度、(9)等位接続詞の頻度、(10)従属接続詞の頻度、(11)関係詞の頻度、(12)受動態の頻度、である。使用語彙の豊富さを重視した研究としては石川[2005]が標準化 TTR (Standardized Type-Token Ratio) が受容語彙力と中程度の相関を示すことを明らかにしており、水本[2008]ではこうした語彙指標の中で総語数(Tokens)以外では使用語彙の難易度を示す指標(J8 index, JACET8000)における難しい語の使用割合指標)や、語彙の豊富さを示す指標(D, [Malvern & Richards, 2002])などが英作文の総合評価値と相関が高い傾向にあることを示している。なお TTR (Type-Token Ratio, 「異語率」)は、延べ語数と異なり語数の比率を示す、汎用性の高い頻度指標である。ただし TTR はテキストや発話の全体の量が数値に過度の影響を与える(長く書くと語彙の豊富さを表す指標が低下する)という難点があることから、Carroll の修正 TTR (CTTR)¹や標準化 TTR などの改良版の使用が推奨されている[田島 et al., 2011].

2.2 応用の可能性

第二言語習得の場合には、目的変数として学習者の既存テストの点数などを設定することによって、説明変数の妥当性を検討することができる。しかし本研究であつかうような感性の言語表現は、どのような表現が「良い」表現なのかを事前に決定すること、あるいは他のテストなどとの相関を示すといったような、目的変数を設定することが難しい。日本語と英語の言語的な違いもあり、例えば助動詞など品詞についてはその差異が顕著である。さらに接続詞など感性の言語化表現に重要度が低いと思われるものは分析から除外できる。

¹ CTTR : $\text{Type} / \sqrt{(2 \times \text{Token})}$ により算出される。延べ語数の 2 倍の平方根における異なり語数の割合。同一語が繰り返されているのではなく多様な語彙が使われているかどうかを示す。

これらをふまえ、今回は英作文の分析において説明変数として一般に用いられている基準のうちで、感性の言語表現を分析する上で助けとなるような変数があるかを検討すべく、「ワイン雑誌コーパス」「日本酒雑誌コーパス」「日本酒自作言語化コーパス」の3つの言語データを用いて比較してみたい。2種類の雑誌コーパスは、日本国内において市販されている日本語のみ(外来語を含む)によって記述された日本酒・ワインの銘柄レビュー文である。「日本酒自作言語化コーパス」は、筆者を含む6名によって記述された、日本酒の内省的言語化コーパスである。

表 1:各コーパスの概要

	ワイン 雑誌	日本酒 雑誌	日本酒 自作
総抽出語数	201,294	50,147	34,930
異なり語数	9,449	3,047	3,082
総文数	11,421	2,552	1,344
総段落数	1,957	1,167	442
平均出現回数	10.50	8.47	5.81
出現回数標準偏差	64.55	34.92	22.45
TTR	0.05	0.06	0.09
CTTR	14.89	9.62	11.66
平均文長	17.62	19.65	25.99

表 1 に、各コーパスの概要及び TTR, CTTR, 平均文長を示す。TTR は 1 に近いほど総語数に対する異なり語数の比率が多いことを示す(例えば 1000 語の作文が 1000 の異なり語で記述された場合、TTR=1.00 となる)。なお本分析の目的は各コーパスの概略を知ることであり、Tokens もコーパス間で大きく開きがることから厳密な統計的分析を目的としないことを付記する。

表 1 に示された CTTR を比較すると、雑誌間では日本酒雑誌コーパスが語彙が豊富であり、日本酒雑誌コーパスと日本酒言語化コーパスの比較では日本酒雑誌コーパスの語彙が豊富であることが示された。

2.3 応用可能性の検討

筆者の直観としては日本酒言語化コーパスのほうが語彙が豊富という印象を持っていたが、数値としては逆を示すこととなった。おそらくこの原因は日本酒の雑誌では味わいの記述に加えて銘柄の紹介や製造法、土地柄などが記述されることがあり、異なり語数が多いことにある。この現象を「豊かな記述」とするには慎重さを要する。味わいの記述の豊かさ、詳細さを志向した場合には何度も同じ「酸味」という語について言及することが考えられ、「精緻な記述」と「何度も同じ言葉を使っている」ことを表面的な情報から判断できるかという課題が示され、TTR や CTTR をどのように扱うかはコーパスの指向性を含めてさらなる検討が必要であることが示された。

3. 翻訳ストラテジーの応用可能性

翻訳学領域のなかでも本研究で着目したいのは、翻訳の方略についてである。前項で示した第二言語習得分野が主に本研究のテキストの定量的分析手法への援用となることが期待されるのに対し、翻訳分野の知見は分析の背後にある理論面での応用が期待される。

翻訳と感性の言語化の相同点は、言語化(翻訳)の前後においてなにかしらの「等価性」を求めるという点にある。翻訳はソーステキストからターゲットテキストに言語をまたいで変換される際に、逐語の意味の等価性や著者の意図の等価性など、様々な等価性が求められ、求める等価性に応じて様々な方略が用いられる。感性の言語表現において等価性は、感覚と言語表現の間に生起する。表現者は知覚情報を言語という異なる系に翻訳するのである。

ストラテジーという用語は、最適な方法で、在る特定の目標に到達するために用いられる一連の目的論的行動を意味する[ベイカー&サルダニャ, 2013]。Chesterman[1997]は翻訳のストラテジーを「理解ストラテジー」と「産出ストラテジー」に二分する。理解ストラテジーは減点の認知的分析と関連する方略であり、産出ストラテジーは翻訳テキストの産出に関するストラテジーである。このうちこれまでのところ翻訳学領域では理解ストラテジーが研究のメインピックとなっている。翻訳においてはこうしたストラテジーを使用しつつ、ソーステキストとターゲットテキストの間の等価性を担保が目指される。

等価については様々なレベルの分類があるが、知覚情報と言語記述の等価性という視点からみて重要なのは「指示的等価」であろう。これは ST と TT の語が現実世界にある同一物を指す場合であり、味覚的要素、香りの要素を名詞的に指摘する等価性に対応する。また言語は線条性をもつために、知覚情報の経時的流れを言語として再現すべきかという点からは「情報の流れ」の理論、及び結束性の理論も参照される必要がある。

4. まとめ

本稿では感性の内省的言語化資料の分析について、第二言語習得分野および翻訳分野からの知見応用を目指し、基本文献を紹介した。第二言語習得分野からは語意の抱負さの指標として TTR や CTTR による分析の有効性が示唆され、翻訳ストラテジーからはおもにソーステキストとの等価性の観点が重要であることが示唆された。等価性に関しては感性言語表現において重要な概念であるメタファ、類推などと同質の問題をはらんでいると思われる。ただし第二言語習得にとっての語彙力とは何かという問い、そして翻訳理論における等価性などは、どちらも未解決課題であり、分析手法として感性の内省的言語表現への応用の可能性は現状では工夫を要するものと思われる。

参考文献(論文誌と同じスタイルを推奨)

- Baker M, Saldanha G, 藤濤 文: 翻訳研究のキーワード. 研究社, 2013.9, (2013).
- Chesterman A: Communication and learning strategies for translators. *AILA Review*, Vol.12, pp.79-86 (1997).
- Higuchi K: Quantitative analysis of textual data : Differentiation and coordination of two approaches. *Sociological Theory and Methods*, Vol.19, No.1, pp.101-115 (2004).
- Malvern D, Richards B: Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, Vol.19, No.1, pp.85-104 (2002).
- Malvern D, Richards B: Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, Vol.19, No.1, pp.85-104 (2002).

-
- 小林雄一郎, 金丸敏幸: パターン認識を用いた課題英作文の自動評価の試み (テキスト・談話, 思考と言語一般). 電子情報通信学会技術研究報告 .TI, 思考と言語, Vol.112, No.103, pp.37-42 (2012).
- 水本篤: 自由英作文における語彙の統計指標と評定者の総合的評価の関係. 学習者コーパスの解析に基づく客観的作文評価指標の検討』(統計数理研究所共同研究レポート 215), pp.15-28 (2008).
- 田島ますみ, 佐藤尚子, 深田淳, 玉岡賀津雄: 記述式試験解答に対する答案としての評価と日本語力からの評価: 語彙力を中心に (< 特集> 日本語リメディアル教育とは). リメディアル教育研究, Vol.6, No.1, pp.39-46 (2011).
- 石岡恒憲: 小論文自動採点. 電子情報通信学会誌, Vol.92, No.12, pp.1036-1040 (2009).
- 石岡恒憲: 記述式テストにおける自動採点システムの最新動向. 行動計量学, Vol.31, No.2, pp.67-87 (2004).
- 石川慎一郎: 大学生英語学習者の受容語彙力と発表語彙力の関係: 語彙サイズテストおよびエッセイ・コーパス分析に基づくアプローチ. 中部地区英語教育学会紀要, Vol.34, pp.337-344 (2005).