

日英コンパラブルコーパスにおける文類照応の指示詞表現の比較分析

竹井 光子
 広島修道大学・法学部
 takeim@shudo-u.ac.jp

吉田 悦子
 三重大学・人文学部
 tantan@human.mie-u.ac.jp

藤原 美保
 Willamette University
 mfujiiwar@willamette.edu

1. はじめに

語彙・統語レベルの厳格な規則と異なり、談話レベルの言語現象に関するルールは原則や傾向という色合いが強く、複合的な要因が影響するため複雑で知識として与えることが難しい。談話中の文間のつながりを実現する結束表現（照応表現など）は、その一例である。この知識として与えることが難しい談話レベルの言語現象への意識を効果的に活性化させ習得に導くための資源として日英データの相互比較が実現できるコンパラブルコーパスの構築と活用を目的とする研究が、日米大学における EFL/JFL¹教育環境の連携により進行中である（竹井ら, 2009; Yamura-Takei *et al.*, 2009）。

ある言語のテキストと文単位で別の言語に翻訳されたテキストの対をパラレルコーパスと呼ぶのに対して、同じサンプリング基準にしたがって収集された多言語のコーパスをコンパラブルコーパスと呼ぶ。後者は文単位の分析を超えた自然産出データを必要とする談話・語用論的、社会言語学的分析を目的とする場合に適している。また、表現形式を比較する対照言語学的視点からも有効である。

本稿は、談話レベルの言語現象である照応表現のうち文類照応の指示詞表現に焦点を当て、コンパラブルコーパスによる比較分析を行うことで、(A) 日本語・英語データの比較による言語別特徴、(B) 母語話者・学習者データの比較による学習者傾向を提示することを目的とする。さらに、分析結果から得られる外国語教育への示唆について述べる。

2. コーパス分析

2.1 コーパスの概要

日米の大学における EFL/JFL 教育環境の連携により、日米大学生が共通の題材および指示のもとに母語および外国語（学習言語）によって産出した談話サンプル（作文）を収集している。

これにより、英語学習者データ (EL), 日本語母語話者データ (JNS), 英語母語話者データ (ENS), 日本語学習者データ (JL) の 4 つのサブコーパスから成るコンパラブルコーパスが構築でき、2 言語間 (ENS-JNS, EL-JL) と母語・非母語間 (ENS-EL, JNS-JL) の比較が可能となる。談話モードは比較的産出が容易な物語文 (narrative text) に限定している。物語文の産出は、5 分程度のアニメストーリー映像 (PINGU) のあらすじの作文としている。

コーパスの基本データは以下の通りである。計 166 テキストで、総文数は 2,519 となる。

	ENS	EL	JNS	JL
テキスト数	48	36	34	48
文数	755	492	473	799
平均文/テキスト	15.73	13.67	13.91	16.65

表 1: コーパスの基本データ

2.2 文類照応の定義と分析

ある言語表現が文脈内の他の表現を指す時、それらの表現は照応関係にあるといい、指示する表現を照応詞、指示される表現を先行詞と呼ぶ。この照応関係は、先行詞の文法範疇によって名詞類照応、述語類照応、文類照応の3つに分類できるが、本研究の分析対象は、例 (1) のように先行詞が文または文章によって表現される事象である文類照応に限定する。

- (1) 母が一人で洗車をしていた。弟がそれを見て手伝いに行った。

文脈指示研究では名詞類照応が主流であり、文類照応のみに焦点をあてたコーパス研究は見当たらないが、竹井ら (2005) は、日本語読解教材コーパスに含まれる 1,382 個のゼロ代名詞のうち、21 個 (1.53%) が event を先行詞としているとし、Fais and Yamura-Takei (2003) は、電子メールコーパスに含まれる 295 個の (ゼロ) 代名詞のうち、27 個 (9%) が文類照応 (event reference) であると報告している。庵 (2002) は、新聞コーパスから抽出した (文脈) 指定指示用法の「この」と「その」500 個のうち 66 個 (13%) が文類

¹ English as a Foreign Language (EFL) / Japanese as a Foreign Language (JFL)

照応であったと報告している。また, Gundel *et al.* (2005) は, 会話コーパスに含まれる三人称代名詞 2,046個中110個 (5.4%) の *it* が文類照応であったと報告している。

低頻度ながら照応関係の一部を担う文類照応は, 照応解析において困難な課題の一つであることがしばしば指摘されているし (Fais, 2004等), 言語習得においても, 名詞類照応に比べ談話理解・運用の面で習得されにくいとの報告がある (伊藤ら, 1987等)。

英語の指示詞 (demonstratives) は, *this* と *that* (および *these* と *those*) であり, いずれも代名詞, 連体詞としての用法がある。一方, 日本語の指示詞は表2のようにコ・ソ・アの三系列に分かれ, さらに指示内容に応じて語形変化するため語彙的にも多様である。また, 代名詞, 連体詞, 副詞の3つの機能に分類できる。²

	コ系	ソ系	ア系
指示代名詞	これ ここ こちら こっち	それ そこ そちら そっち	あれ あそこ あちら あっち
指示連体詞	この こんな こういう こうした こういった このような	その そんな そういう そうした そういった そのような	あの あんな ああいう ああした ああいった あのような
指示副詞	こう このように こんなに こんなふうに	そう そのように そんなに そんなふうに	ああ あのように あんなに あんなふうに

表2: 日本語の指示詞

このうち文類照応表現として使用される可能性がある指示詞は, 英語では *this* と *that* である。加えて, 例 (2) のように, 人称代名詞の *it* も文類照応詞として機能する場合がある。

(2) John insulted the ambassador. That/this/it happened at noon. (Gundel *et al.*, 2005)

日本語ではソ系の指示代名詞の「それ」, 指示連体詞の「その」が一般的であるが, その他の指示連体詞および指示副詞やコ系の指示詞も候

² 外国語教育において, これらの指示詞は基本語彙であり, その現場指示用法は, 日英ともに比較的初期の学習段階で導入されるが, 初級・中級・上級を通して体系的に指導されることはなく, 文脈指示用法については習得に問題がある場合が多いことが指摘されている (新村, 1992)。

補となり得る。ちなみに, 本コーパスに現れた文類照応表現を下線で示している (表2)。

分析は, 人手によって (i) 文類照応の指示詞表現の抽出, (ii) 先行詞の判定の手順で行った。 (i) においては, まず *this*, *that*, *it* およびコ系・ソ系・ア系の指示詞すべての検出を行い, 文脈指示かつ文類照応の用法であるかどうかの判断を行った。ただし, ソ系の指示詞のうち, 文法化 (接続詞化) が進んでいる「それに」「それで」「それでも」「それから」などは抽出対象から除いた。接続詞的であるが, 指示性を残している「その時」「その後」などは対象に含んだ。

3. 分析結果と考察

3.1 照応詞

本分析で抽出した文類照応は, コーパス全体で 157 個であった。まず, 照応詞の言語形式および統語的な位置によって分別集計した結果を表3, 表4に示す。数値は, 各サブコーパスの文数が異なるため 100 文あたりの出現頻度の形で母語話者/学習者別に示している。

	主語	目的語	その他	計
<i>this</i>	2.1/0.0	1.2/0.0	0.5/0.0	3.8/0.0
<i>this</i> + N		0.1/0.2	0.8/0.0	0.9/0.2
<i>that</i>	0.5/0.0		0.8/0.2	1.3/0.2
<i>that</i> + N			0.0/0.8	0.0/0.8
<i>it</i>	0.1/0.6			0.1/0.6
計	2.8/0.6	1.3/0.2	2.1/1.0	6.2/1.8

表3: 英語データ分析結果(ENS/EL)

	主語	目的語	その他	計
これ		0.0/0.3	0.0/0.1	0.0/0.4
この+N			0.2/0.3	0.2/0.3
それ	0.4/0.1	4.4/0.5	0.6/0.1	5.5/0.8
その+N	0.0/0.1	0.4/0.1	4.0/4.4	4.4/4.6
そんな+N		0.2/0.0	0.6/0.0	0.8/0.0
そう			0.2/0.0	0.2/0.0
計	0.4/0.3	5.1/0.9	5.7/4.9	11.2/6.0

表4: 日本語データ分析結果(JNS/JL)

まず, 日英母語話者データ (ENS, JNS) を比較してみよう。照応詞の言語形式については, 英語では *this* が大半を占め, 日本語ではソ系の表現が圧倒的である。照応表現が現れる統語的な位置については, 英語は主語, 日本語は目的語³が多いという特徴が見られる。全体的な頻度

³ ここで言う目的語は, (動詞句内の) 前置詞の目的語や必ずしも必須項とは限らないものも含む広義の解釈をしている。

としては日本語の方が高い。⁴

次に、母語話者 (ENS, JNS) と学習者 (EL, JL) を比較してみよう。日英ともに母語話者に比べ学習者データでは使用頻度が低いことが分かる。特に、例 (3) のような、ENS データに顕著な主語の位置における this の用例が EL データでは全く見られない。

(3) His mother then went over to him and gave him a hug. **This** made Pingu really happy.

また、例 (4) のように JNS データに特徴的な目的語の位置におけるソ系指示詞の使用も JL では極めて少ない。

(4) ママは、ピングーを怒ってピングと遊び始めました。ピングーはそれを見て、焼きもちをやいてしまいました。

3.2 照応先

本分析によって抽出した 157 個の文類照応は全て前方照応であったが、そのうち 146 個 (93%) で直前の文が先行詞であった。残りは、直前の複数の文で表現された事象全体を指すもの、同文中の先行節を指すものであった。これは、日英間、母語・非母語間でほとんど差はない。

3.3 共起関係

文類照応表現を精査するため、母語話者データで文類照応表現が主語または目的語の位置にある場合 (ENS 31 例, JNS 26 例) について、共起している述語を調べた。

分類	述語例 [用例数]	
ENS	主語	(not) work [6], (not) go well [1], happen [1], capture/get/win one's attention [3], make one angry/happy/sad [4], cheer, frustrate, upset, worry [5], be a happy ending [1]
	目的語	(not) take [1], be upset about/by/with [3], know [1], see [5]
JNS	主語	～からだ, かなう [2]
	目的語	する [1] わかる/気づく [2] 見かねる/見はからう [2], 見る [12] やきもちを焼く/ 傷つく/怒る/反発する [6] 気に入る (否定) [1]

表 5 : 文類照応指示詞と共起する述語

日英ともに、感情表現 (upset, 傷つく等)、知覚表現 (see, 見る等) が目立っている。

指示連体詞 (この, その, そんな) と共起する名詞については表 6 の通りである。

分類	名詞例 [用例数]
ENS	point [2], time [2], way [1], situation [1], behavior [1]
JNS	ころ, 時, 中, 後 [14], 横 [2], こと [3] アピール, 結果, やり取り, 態度, 様子 [5]

表 6 : 文類照応指示連体詞と共起する名詞

まず日本語で顕著なのが「時間」を表す名詞であるが、これらは学習者データでも同様に多く見られた。⁵ その他の「その・この+抽象名詞 (態度など)」の表現については学習者データではほとんど見られない。

3.4 センタリングモデル

センタリング理論 (Grosz et al., 1995) は、隣接する文間の名詞類 (直接・間接) 照応に注目し、中心要素の連続性を 3 つの遷移タイプ CONTINUE (連続), RETAIN (保持), SHIFT (変移) および NULL (連続なし) にラベル付けすることで談話の一貫性との関係のモデル化を試みている。したがって、文類照応表現によって形成させる結束性は原則としてこの一貫性モデルに反映されない。

そこで、文類照応の出現環境をこの 4 つのラベル別に示してみた (図 1)。

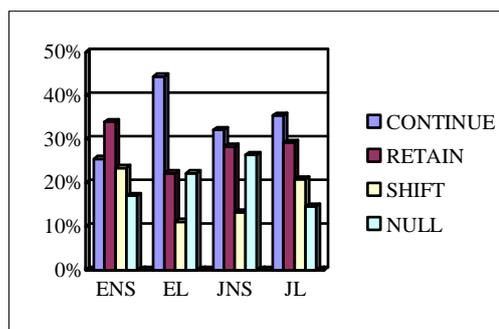


図 1 : 遷移タイプと文類照応

日本語・英語、母語話者・学習者によって多少の偏りはあるが、必ずしも名詞類照応関係が存在しない環境 (=NULL) で使用されているわけではないことが分かる。むしろ、例(5)のように名詞類照応 (破線) との併用 (すなわち

⁴ 西原 (1990) は、同様の内容の日本語と英文を比較した場合、日本語の方に指示詞の使用が多いとする調査結果を報告している。

⁵ 本橋 (2006) が指摘しているように、「その時、その後」などは初期の段階でフレーズとして学習していることが影響している可能性がある。

CONTINUE, RETAIN, SHIFT 環境での使用) が目立っている。

- (5) Kid Pingu's father got very angry, and scolded Kid Pingu. [RETAIN] / This made Kid Pingu even angrier, and he began rocking in his seat. [RETAIN]

しかし、NULL 環境に注目してみると、ENS データ全体の NULL 93 例中 8 例 (8.6%), JNS データでは NULL 73 例中 14 例 (19.2%) に文類照応が存在していた。一例を(6)に示す。

- (6) Pingu walked over to Pinga's bed and she woke and started crying. [CONTINUE] / Mama hugged Pinga and began playing with her. [SHIFT] / Pingu saw this and became jealous. [NULL]

つまり、文類照応詞 *this* の存在が第 2, 3 文間の結束関係を形成し、基本センタリングモデルでは説明しきれない結束性(すなわち談話の一貫性を高める作用)を実現しているわけである。

4. まとめと今後の課題

日英母語話者データの比較分析において判明した言語別特徴として、まず「日本語の文類照応詞=目的語、英語の文類照応詞=主語」を挙げることができる。さらに、この対照的な結果が第二言語習得の困難点を示唆していることが、母語話者・学習者データの比較から明らかになった学習者傾向からも推測できる。

例(3)のような主語の *this* と感情動詞との組み合わせ文、例(4)のようなソ系指示詞と知覚動詞との組み合わせ文による結束表現の理解や運用は、母語データを手本としながら明示的に指導すべきポイントとなり得るであろう。

また、指示連体詞と共に文類照応表現となる抽象名詞についても、日本語・英語ともに、母語話者データから収集することで資源として活用することが期待できる。

同じ場面、状況、内容を描写することによって、異言語に共通する言語現象や概念があること、言語によって異なる表現形式があることをコンパラブルコーパスは気づかせてくれる。

本稿では、コーパス分析結果を外国語教育への示唆の視点からまとめたが、これらはすべて言語処理技術への応用の可能性を含んでいる。

今後は、コーパス規模を拡大しながら分析を続けることで、文類照応詞と共に起する動詞や抽

象名詞の収集を続けたい。また、文類照応がどのように談話の一貫性に貢献しているかについてもセンタリングモデルを利用して詳細な分析を行っていきたい。

参考文献

- Fais, Laurel (2004). Inferable centers, centering transitions and the notion of coherence. *Computational Linguistics* 30/2, 119-150.
- Fais, Laurel and Yamura-Takei, Mitsuko (2003). The nature of referent resolution in Japanese email. *Discourse Processes* 36/3, 167-204. Laurence Erlbaum Associates, NJ.
- Grosz, Barbara J., Weinstein, Scott and Joshi, Aravind K. (1995) Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 12/3, 175-204.
- Gundel, Jeanette K., Hedberg, Nancy and Zacharski, Ron. (2005). Pronouns without NP antecedents: How do we know when a pronoun is referential. In Antonio Branco, Tony McEnery and Ruslan Mitkov (eds.). *Anaphora Processing: Linguistic, Cognitive and Computational Modelling*. John Benjamins. 351-364.
- Yamura-Takei, Mitsuko, Fujiwara, Miho and Yoshita, Etsuko. (2009). Comparable local learner corpora: Raising discourse awareness. In *Proceedings of Pacific Associations for Computational Linguistics (PACLING 2009)*.
- 庵功雄 (2002). 「この」と「その」の文脈指示的用法再考. 一橋大学留学生センター紀要, 5, 5-16.
- 伊藤淳子・林部秀雄・小野博 (1987). 指示語を含む文脈の理解に関する発達的研究. 横浜国立大学教育紀要 27, 235-244.
- 竹井光子・相沢輝昭・藤原美保 (2005). コーパスとしての教材: 語学教師のための分析とツール. 言語処理学会第 11 回年次大会発表論文集.
- 竹井光子・吉田悦子・藤原美保 (2009). 談話レベルのメタ言語能力育成のためのコンパラブルコーパスの構築と活用. 言語処理学会第 15 回年次大会 発表論文集.
- 新村朋美 (1992). 指示詞の習得—日英語の指示詞習得の対照研究. 早稲田大学日本語研究教育センター紀要 4.
- 西原鈴子 (1990). 日英対照修辭法. 日本語教育 第 72 号, 25-41.
- 本橋美樹 (2006). 日本語学習者の指示形容詞の使用. 関西外国語大学留学生別科日本語教育論集 16, 61-72.