

## 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

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  
mfujiwar@willamette.edu

### 1. はじめに

本稿の目的は、談話レベルのメタ言語能力育成のための日英コンパラブルコーパスの構築計画とその外国語（日本語・英語）教育における活用案を理論的な背景をふまえて紹介することである。

母語に対する直感の意識化が外国語学習におけるメタ言語意識を活性化するという主張に示唆を得て、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メタ言語意識の相互活性化を実現する指導法の研究を着想した。日米の大学に在籍する日本語・英語の母語話者および学習者の産出データを共有できる環境を利用して、言語資源の構築とそれを利用した指導カリキュラムの開発を目指す。

指導項目としては、談話レベルの言語現象である照応関係に焦点を当てる。照応表現の種類・連続性の分析を行うとともに、その分析結果を談話能力の一側面（照応連鎖による結束性の理解）に関するメタ言語能力の育成をベースとする外国語教育に活用する案を提案する。

### 2. 理論的背景と先行研究

本研究の核となる 3 つのキーワード（メタ言語意識、談話能力、コーパス）について、その定義および理論的背景を外国語教育との関連を紹介しつつ説明する。

#### 2.1. メタ言語意識 (metalinguistic awareness)

「メタ言語意識」とは、意味とむすびつく言語形式や構造、機能に向けられる意識的な内省であり、言語を客体化して分析し運用に応用することができる能力を言う。ここで言う言語の構造とは、音韻、形態、語彙、統語、談話など様々なレベルを網羅する。例えば、メタ統語意識 (metasyntactic awareness) は、非文の判定や訂

正ができるかどうかの能力を示す<sup>1</sup>。

母語 (L1) の獲得にともないこのメタ言語能力も発達するが、外国語 (L2) に接することによりさらに意識化が促進すると言われている。直感的判断が支配しやすい母語と異なり外国語は客体化がしやすいこと、2 言語の比較対照により明らかになる相違点や類似点により言語の構造的特性に気づきやすくなること、その理由である。山田 (2006) は同様の概念を L1, L2 の言語形式につながる出入力チャンネルの成長によって強化される「共通基底能力」と定義している。

メタ言語意識への関心は、focus on form (Long, 1991) に具現化されている明示的文法指導の再評価とともに高まっており、認知的アプローチの第二言語習得理論において論じられる注意 (attention)、気づき (noticing)、意識高揚 (consciousness raising) などの概念と微細な違いはあるものの共通するものがある (参照: Robinson, 2001)。

メタ言語意識と第二言語能力との関係を調査し両者間に正の相関があることを報告した Renou (2001) や、言語形式への意識や気づきに注目した読解支援システムを提案した竹井ら (2004) など、この分野の先行研究は多岐に渡る。

#### 2.2. 談話能力 (discourse competence)

「談話能力」とは、自然な意味的なまとまりのある文章や会話を理解・生成する力である。Canale & Swain (1980) では、文法能力、社会言語学的能力、方略能力と並び、コミュニケーション能力の構成要素の一つとして取り上げられている。実際の言語教育の場においては、語彙・統語レベルの言語項目を使いこなす能力である文法能力に焦点をあてた指導が先行していることは否定できないが、談話レベルの指導への関心も、研究、実践の両面で高まりを見せている (参照: McCarthy, 1991; Thornbury, 2005)。

<sup>1</sup> メタ言語意識の体系的な定義や関連研究については、Mora (2008) に詳しい。

談話レベルの特性としてしばしば取り上げられる結束性 (cohesion) は、指示 (照応)、代用、省略、接続および語彙的結束に分類される (Halliday & Hasan, 1972)。特に、指示 (照応) 表現については言語間の相違が顕著で母語話者レベルの習得が難しい項目であると同時に、対照言語学、計算言語学などの分野で興味深い研究テーマとなっている。

比較的厳格な規則 (hard rule) である語彙規則・文法規則と異なり、原則 (principle) または傾向 (preference) という色合いが強い談話レベルの法則は、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く、教科書や参考書でも体系的に扱われることが少ない。

しかし、ライティング評価を決定する要因の調査結果から結束性の重要性を導き出した Chiang (2003) や、学習者が使用する照応表現の母語話者とのずれを指摘した Kang (2004, 2005) など談話レベルの教育の必要性を示唆する研究成果は多い。

### 2.3. コーパス (corpus)

「コーパス」とは、ある目的をもって収集される (多くの場合、付加情報を付与されて) 電子データとして保存されている自然発生言語の集合で、ある特定の言語のジャンルの特徴を保持する代表性 (representativeness) とそれにとりあう規模 (size) が必要となる (Baker et al., 2006)。その下位分類、用途は、実に多様である。

近年、コーパスはますます大規模化し自然言語処理の分野で自動処理の研究が積極的に進んでいるが、あえて分析過程における人の関与<sup>2</sup>を好む言語教育研究の分野では小規模コーパス (small corpus) の活用価値も高い (Ghadessy et al., 2001)。

2 語以上の言語データから成る多言語コーパス (multilingual corpus) のうち、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との対をパラレルコーパス (parallel corpus) と呼ぶ。一方、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 (comparable corpus) と呼ぶ<sup>3</sup>。前者の文単位の対応付けが容易であるという利点を活用した

<sup>2</sup> Sinclair (2001) は、大規模コーパスにおける DHI (delayed human intervention) に対して、EHI (early human intervention) と呼んでいる。

<sup>3</sup> 例えば、同じ事件を扱った新聞記事の多言語版はその代表例である。

用途は多岐に渡るが、翻訳作業による影響 (いわゆる translationese) が避けられないという欠点がある (McEnery et al., 2006)。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は後者が適していることになる。

学習者コーパス (learner corpus) は、外国語習得過程にある学習者によって産出されたテキストを体系的に収集したものである<sup>4</sup>。誤用分析や特定の語 (句) や文法パターンの頻度調査や母語話者コーパスとの比較による過剰・過少使用の調査などで第二言語習得研究に大いに貢献している (Granger, 1998)。

頻度などの統計的分析には、比較対象とする参照コーパス (reference corpus) が必要となる。British National Corpus (BNC) など既存の大規模コーパスがしばしば用いられる。

コーパスは、言語使用の特徴や傾向性を明らかにすることから、言語資源としての価値が認められ外国語教育への応用が盛んである。中條ら (2005, 2007) は、学習者自身がデータの中から帰納的に規則を発見するデータ駆動型学習 (data-driven learning: DDL) に基づく語彙・文法指導でその成果を報告している。

### 2.4. まとめ

メタ言語意識の開発に着目した言語教育、談話能力の育成を目的とするカリキュラム、コーパスを活用した指導法の研究はそれぞれに先行例 (前述) があるが、この 3 つを有機的に統合する試みは見当たらない。

また、それぞれの外国語教育環境が恩恵を受けることが可能となる双方向の連携作業によるコーパスの構築と活用 (後述) は新規的な試みであると言える。

## 3. コーパスの構築計画

### 3.1 データの収集

日米の大学において EFL/JFL 教育環境にいるプロジェクトメンバーが互いの外国語教育環境を利用してデータ収集を行う。その概要を図 1 に示す。

<sup>4</sup> 代表例として、ICLE (International Corpus of Learner English) や日本人学習者のデータと母語話者データを体系的に整備した NICE (Nagoya Interlanguage Corpus of English) などがある。

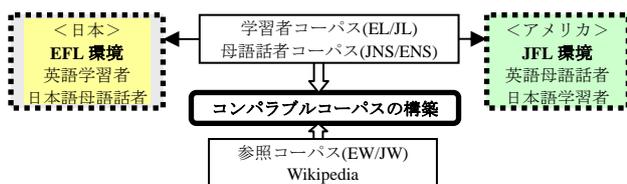


図 1：コーパス構築計画

まず、日米大学生が共通の指示のもとに母語および外国語によって産出した談話サンプル（作文）を収集する。すなわち、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つが収集できることになる。両環境における学習者のレベルをふまえ、談話モードは比較的産出が容易と思われる物語文(narrative text)に限定する。物語文の産出は、5分程度のストーリー映像のあらすじの作文とする<sup>5</sup>。これで、2言語間のコンパラブル(ENS-JNS, EL-JL)と、母語・非母語間のコンパラブル(ENS-EL, JNS-JL)が可能となる。

これに加えて、大学生コーパスの分析結果を統計的に比較するための参照コーパス(reference corpus)として Wikipedia の項目を利用する。Wikipedia は事典という性格上、叙述文(descriptive text)が大半であるが、大学生コーパスと同タイプの談話として、映画、ドラマ、アニメなどの項目に含まれるあらすじ(synopsis, plot)の部分を英語・日本語ページから収集することにする(EW, JW)。

これらのサブコーパスから以下の複合的な用途が考えられる。

- (1) JL - 母語(L1)の意識化(誤用分析)
- (2) EL - 母語(L1)の意識化(誤用分析)
- (3) JNS/JW - 学習者へのモデル提示
- (4) ENS/EW - 学習者へのモデル提示
- (5) JL+JNS - 母語話者・学習者の比較
- (6) EL+ENS - 母語話者・学習者の比較
- (7) JNS+JW - 母語(L1)能力の検証
- (8) ENS+EW - 母語(L1)能力の検証
- (9) JL+ENS - L1/L2間の比較
- (10) EL+JNS - L1/L2間の比較

<sup>5</sup> 一話5分程度、限定された登場キャラクター(=discourse entities)、セリフは実在しない言語を使用という、談話産出用の題材としての妥当性から、スイスのクレイアニメである Pingu を使用する。同一のエピソードを見て、母語と外国語の両方であらすじ文を作成することになる。

母語(L1)能力と外国語(L2)能力には相関があるとの指摘もある。自然な照応関係の生成能力に関してはどうだろうか。(7)(8)による検証結果をL2能力としてのJL, ELと比較することも可能であろう。

### 3.2 データの分析

コーパスの収集後にはコーパスの整備・分析が必要となる。コーパスに含まれる参照関係をモデル化する土台として、センタリング理論(Grosz et al., 1996)において談話要素の移り変わりを示すTRANSITION(Continue, Retain, Shift)とその連続パターンを利用する。さらに、各TRANSITIONと中心談話要素(center)の形式(すなわち照応表現:定名詞,指示詞表現,(ゼロ)代名詞,裸名詞など)との関係をまとめる。また、間接照応,同一参照関係の名詞表現(同義語,上位語,下位語)などの収集も可能であろうと予想している。また,(ゼロ)代名詞の連続使用や特定の所有表現が好まれる環境なども統計的に説明したい。母語話者が多用しながら学習者が習得できていない照応表現やTRANSITIONのパターンの抽出も行う。

### 4. コーパスの活用計画

構築したコンパラブルコーパスの活用法は、教授者用資源/学習者用資源,母語の直感の意識化支援/非母語の談話規則の顕在化支援の2つの視点(およびその組み合わせ)から検討し、英語教育,日本語教育の現場への還元を試みる予定である。

まずは、データの分析によって得られた資源を、教育的活用を意図したデータベースとして整備する必要がある。データ駆動型学習の知見も取り入れつつ、何をどのように提示するか設計から始めることになるであろう。さらに、データを使った学習法,教授法の工夫を学習者からのフィードバックも参考としつつ進める予定である。

試行的に行った結束性理解を目的とする授業では、学習者の産出データに含まれる語彙・統語・談話・語用論レベルの誤りや不自然さに接することで生起される母語的直感による判断をディスカッションなどによって顕在化させることで母語に対するメタ言語意識を活性化するというシナリオの実現に手ごたえを感じるフィードバックを得ることが出来ている。

## 5. 今後の課題

現在は、データ収集基準の検討を兼ねて大学生データの試行的収集を行うと同時に、参照コーパスとして、Wikipedia 日英データの人手による抽出を行っている。コンパラビリティの観点から、同項目の日英版についてバランスを考慮しながら収集を試みている。

Wikipedia は、その編集方法から多言語コンパラブルコーパスであるはずであるが、中にはパラレル（多くの場合、英語から日本語への翻訳）の色合を濃く残す例があるために注意が必要である。

本稿は、コーパス構築の計画と理論的背景の紹介が主目的であり、具体的な実施と検討はまだまだこれからである。今後の課題は、EHI (early human intervention) の前提としつつも、(前述のパラレルデータ除外を含む) データ収集、分析の過程における言語処理技術による部分的自動化の可能性を探ることである。

## 参考文献

- Baker, P., Hardie, A. & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Canale, M. & Swain, M. (1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1:1-47.
- Chiang, S. (2003). The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System* 31/4: 471-484.
- Ghadessy, M., Henry, A. & Roseberry, R. L. , eds. (2001) *Small Corpus Studies and ELT: Theory and practice*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Granger, S. (1998). *Learner English on Computer*. London: Longman.
- Grosz et al. (1995) Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 12/3, 175-204.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Kang, J. Y. (2004). Telling a coherent story in a foreign language: Analysis of Korean EFL learners' referential strategies in oral narrative discourse. *Journal of Pragmatics* 36: 1975-1990.
- Kang, J. Y. (2005). Written narratives as an index of L2 competence in Korean EFL learners. *Journal of Second Language Writing* 14: 259-279.
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, D. Coste, R Ginsberg, & C Kramsh (eds.), *Foreign Language Research in Cross-cultural perspectives*, pp. 39-52. Amsterdam: Benjamins.
- McCarthy, M. (1991). *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-Based Language Studies: An advanced resource book*. London/New York: Routledge.
- Mora, J. K. (2008). Metalinguistic awareness as defined through research. CLAD Website Resources(<http://coe.sdsu.edu/people/jmora/MoraModules/MetaLingResearch.htm>).
- Renou, J. (2001). An examination of the relationship between metalinguistic awareness and second-language proficiency of adult learners of French. *Language Awareness* 10/4: 248-267.
- Robinson, P. (2001). *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.
- Sinclair, J. M. (2001). Preface. In Ghadessy et al. (eds.)
- Thornbury, S. (2005). *Beyond The Sentence: Introducing discourse analysis*. Oxford: Macmillan.
- 坂上辰也・杉浦正利・成田真澄 (2008). 学習者コーパス「NICE」の構築. 杉浦正利 (代表) (2008)「英語学習者のコロケーション知識に関する基礎的研究」平成 17~19 年度 科学研究費補助金 (基盤研究(B)) 研究成果報告書 (課題番号 17320084) 名古屋大学大学院国際開発研究科
- 竹井光子・相沢輝昭・藤原美保 (2004)「読解支援システムへの認知的・第二言語習得理論的アプローチ:ゼロ代名詞による結束性. 教育システム情報学会誌 21/03, 205-213.
- 中條清美・西垣知佳子・内山将夫・原田康也・山崎淳史 (2005)「日英パラレルコーパスを活用した英語語彙指導の試み」『日本大学生産工学部研究報告B』 38, 17-31.
- 中條清美・西垣知佳子・内堀朝子 (2007)「パラレルコーパスを利用した文法発見学習の試み」『日本大学生産工学部研究報告B』 40, 33-40.
- 山田雄一郎 (2006)「英語力とは何か」東京: 大修館書店.