

第二言語習得におけるインプット強化と自然言語処理技術

竹井 光子
広島市立大学
情報科学研究科

yamuram@nlp.its.hiroshima-cu.ac.jp

磯江 健史
広島市立大学
情報科学部

isoe@nlp.its.hiroshima-cu.ac.jp

相沢 輝昭
広島市立大学
情報科学部

aizawa@its.hiroshima-cu.ac.jp

1 はじめに

言語学習にコンピュータを活用しようとする試みは古くから盛んであり、CALL (Computer-Assisted Language Learning) として、認知心理学、人工知能、教育学、言語学など多彩な学問分野の研究成果を取り入れながら発展してきた。また、自然言語処理技術はその応用先の一つとしてCALLをとらえ、ICALL (Intelligent CALL) の設計に貢献している (Holland et al. 1995)。

一方、Chapelle (1998) は、CALL は第二言語習得研究からの知見を基に、理想的な習得環境を提供することを前提に開発すべきだと主張し、CALL が提供する習得条件として第二言語習得研究における7つの仮説を挙げている。

本稿では、この第二言語習得研究からの知見のうち、学習者が受け取る「インプット」の重要性に注目し、習得を促進するとされる「インプット強化」を取り上げる。そして、「インプット強化」を指導の中に取り入れる過程で、自然言語処理技術がいかに効率的に貢献できるかを論じる。

さらに、ICALL の教育理論的背景である第二言語習得理論と技術的土台である自然言語処理の融合を目指し構築した、談話理解レベルの「インプット強化」を提供するシステム例を紹介する。

2 第二言語習得理論

2.1 モデルの概要

Krashen(1985) は、第二言語習得を成功させるには、「理解可能な」インプットを十分に受けることが必須であると主張した。この「インプット仮説」(input hypothesis) を前提として、教育の効果を重視する流れの中、その効果を測定しようとする試みや、インプットだけでは不十分だとしてアウトプットの重要性を主張する

Swain (1985) の「アウトプット仮説」(output hypothesis) の提案など、これまで多くの理論的・実証的研究がなされてきた。これら様々な研究者によって提唱されてきた仮説等を統合し、Ellis (1995), Gass (1997) を参考に言語習得の流れを図式化すると、図1のようになる。

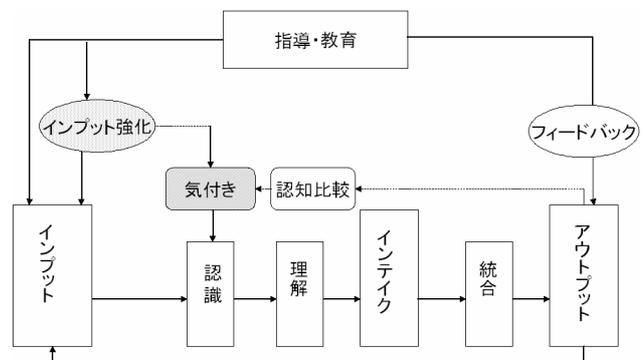


図1：第二言語習得モデル

「インプット」(input) は学習者が接する言語データであるが、このうち「認識」(appreciation) され、意味内容と言語形式の関係が「理解」(comprehension) されたものだけが「インテイク」(intake) として学習者の内部に残る。この「インテイク」が短期記憶の中に貯蔵され、学習者の言語システムとの「統合」(integration) がおきる。この「統合」によって発達する言語システムから「アウトプット」(output) が引き出されるのである。さらに、自己の「アウトプット」と与えられた「インプット」を認知比較することも「インテイク」に影響を与えている。

このうち、教育が関与できるのが、「インプット」の操作と「アウトプット」に対する「フィードバック」(feedback) の部分である。「インプット」から「アウトプット」に向けて、言語データの量は当然のことながら順に少なくなっていく。いかに、「インテイク」の量と質を保持させるかが、教育が担う課題である。

2.2 気づき仮説

認知心理学から影響を受けた第二言語習得理論の中での重要な課題の一つに「意識」(consciousness)がある。言語習得を起すためには、学習者がインプット中の特定の言語形式に意識的に注意を向けて取り込むことだとして、Schmidt (1990) は「気づき仮説」(noticing hypothesis)を提示している。学習者が意識的に注意を向けて気づいたものが、インテイクとして取り入れられるとして、インプットをインテイクに変えるのに「気づき」は必要かつ十分条件であると結論づけている。

2.3 インプット強化

学習者の内部に「気づき」を起こさせるためにはどのような工夫が可能であろうか。この「気づき」を促すための工夫として、学習者の「意識化」(consciousness-raising)を、教師が操作しうるインプットの操作の観点からとらえた「インプット強化」(input enhancement)がある (Sharwood Smith, 1991)。これは、ある言語形式(学習させたい項目)をハイライトしたり文字に色付けしたりするなど視覚的に強調してインプットの質を高める試みである。すなわち、学習者の注意を引くことにより「インプット」を「インテイク」として取り込む率を高めようとする目的の操作である。

例えば、Doughty (1991) は、英語の関係節の習得についてコンピュータの画面上での読解実験を行い、関係詞に「インプット強化」を施されたテキストを用いて読解をしたグループはテキストの理解度が高かったことを報告している。また、この研究から得られた示唆として、(1) 学習項目への高頻度の接触と、(2) 学習者の注意を引く工夫(インプット強化)の2つが習得に有効に働くという点がある。

Schmidt (1990) は、「気づき」の必要性はあらゆる段階における言語形式に当てはまるとしている。すなわち、「インプット強化」を施す対象は、語彙レベルから、統語レベル、談話レベルと様々である。どの学習項目に「インプット強化」を施すかの決定は教師の役割であり、学習の目的や学習言語の特徴などを考慮に入れて選択しなければならない。

2.4 インプット強化の自動化

ターゲットとする項目が決まれば、次はその項目を学習教材(コーパス)の中から抽出する作業である。これは、人手でももちろん可能だが、時間効率の点からはコンピュータによる自動処理に活躍の場があるはずである。すなわち、自然言語処理技術を用いた自動検出およびHTMLなどを用いて強調を施した出力の提供である。

自然言語処理を利用したCALLシステムの問題点として、解析誤りの存在がしばしば指摘される (Nerbonne, 2003 等)。現実として、CALLシステム的设计において言語分析は予め人手によって行う仕様である場合が多いし、語学教師の関心が必ずしも自動処理に向いているとは言えない。しかし、必要な自動言語分析を施し、その結果を人手で修正するという方法を取ることとも可能である。膨大な修正を要するようであれば意味がないが、土台となる形態素解析や構文解析の開発が進み、信頼性のある精度が得られている現在、自動化による分析時間短縮の恩恵は大きいと考える。この点は、2.3節で述べた「学習項目への高頻度の接触」を実現するため、多量のコーパスを分析して適切な教材を用意する必要があることから明らかである。

3 システム例

ここでは、読解という学習活動に焦点を当て、「インプット強化」による効果をねらいとする読解支援システムを紹介する。

3.1 読解過程モデル

「気づき」を促進させる言語形式を何に設定するかを考えるに当たり、まず読解過程のモデルについて触れる。

第二言語の読解研究においては、テキストに関する背景知識や読み手自身の経験に基づきながら推測しテキスト全体を理解していく「トップダウン処理」と、より小さな言語単位の認識からさらに大きな言語単位の理解へと段階的に処理が進んでいくという「ボトムアップ処理」が相互に影響し合いながら処理が進むと考えられている(図2)。

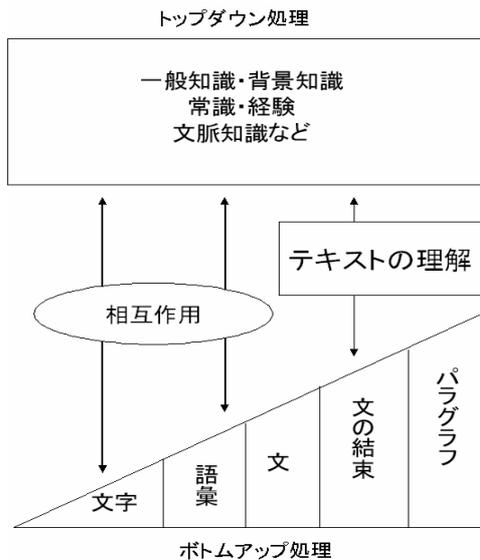


図2：読解過程モデル

自然言語処理技術利用によるインプット強化は、このボトムアップ処理の各段階内の項目を対象にすることになる。本研究では、「文の結末」に「注意」を向けさせる目的で「結束表現（文と文のつながりを保証する言語手段）」のインプット強化を試みた。結束関係に注目した理由は、横山(1999)が指摘しているように、談話レベルの法則は、統語レベルの規則に比べ複雑で知識として与えることが難しいからである。難しいがゆえに、教科書や参考書で体系的に扱われることが少ない。それよりもインプットから学ぶことを期待する方が効果的であると考えられる。

3.2 結束表現：指示詞

日本語の代表的な結束表現として、「これ、その、あそこ」などの指示詞表現がある。これらの表現は、形態素解析情報を用いて比較的容易に検出が可能である。指示詞に注意を向けさせることによって、照応関係を意識したり、「こ・そ・あ」の使い分けを認識させたりすることが期待できる。

3.3 結束表現：ゼロ代名詞

日本語の特徴的な結束表現として、ゼロ代名詞がある。ゼロ代名詞は表層上には現われないが、「現われない」ことによって結束関係を実

現している。母語話者はこの見えない情報を自然に補いながら談話理解を行っている。

「インプット強化」とは、注意を喚起する目的である言語形式を目立たせる手法であるが、ここでは、その操作を拡大解釈し「見えない存在」を「視覚化」することによって「強化」を実現しようと、ゼロ・ディテクター (ZD) を開発した。

3.4 ゼロ・ディテクター

ZD は、日本語談話テキストを入力として受け取り、節分割後、言語解析およびデータベースとの照合を経て、見えない存在であるゼロ代名詞を検出し、それを明示したテキストを出力として提供する (図3)。

検出手法は、用言 (動詞) や体言 (名詞) の項構造など理論言語学的な知見からの経験則によるアルゴリズムに基づいている。ゼロ連用格代名詞は、節に含まれる動詞がとる結合価パターンを顕在格と照合することにより検出を実現した。さらに、ゼロ連体格代名詞は、名詞の属性 (意味的に補完要素を要求する非自立的な名詞の特性) を特定することにより検出を実現している (詳細は、磯江ら, 2004)。

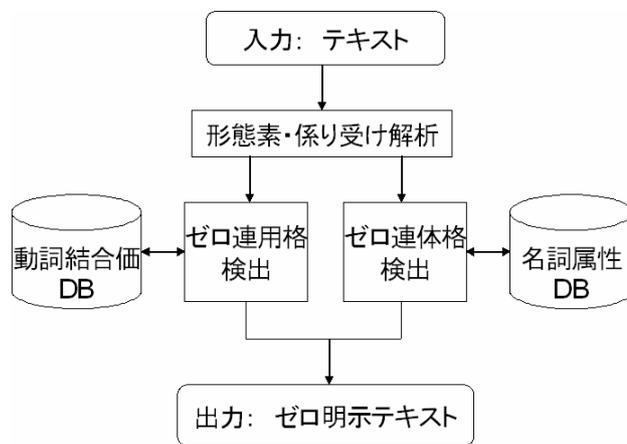


図3：システムの概要

3.5 ZD 出力の活用例

ZD の出力例を図4に示す。普通は表層上には姿を見せない「ゼロ代名詞」が、格情報とともにHTMLによるカラー指定で明示される。

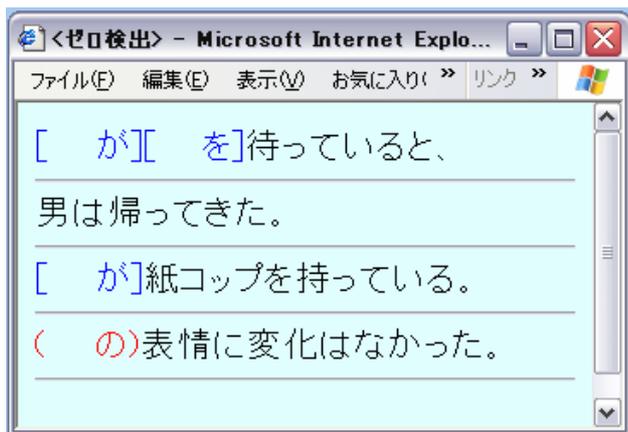


図4：ZD出力例

この出力を使うことによって、学習者の内部にどんな「気づき」が生じ、その「気づき」がどのように結束関係の理解に結びつくかについては、今後詳細な実証研究を待たなければならない。

しかし、普段ゼロ代名詞を意識することの少ない母語話者にとって、人手によりゼロ代名詞の存在を把握することは想像以上に難しく、ゆれも生じやすい。このことから、少なくとも、教師が本稿で取り上げた理論的背景を念頭において授業（あるいはCALL教材）を計画する際に、ZDの出力が補助的ツールとして機能する可能性は、自動化により拡大すると考える。

4 まとめと今後の課題

CALLは、第二言語習得理論からの知見を基盤として設計されるべきとする主張を受け、CALLで実現しうる習得条件の一仮説として「気づき仮説」を取り上げた。そして、その「気づき」を促進させる手法としての「インプット強化」に自然言語処理技術がいかに効率的に貢献できるかを論じた。

さらに、理論的背景である「インプット強化」と技術的土台である自然言語処理による自動解析を融合した事例として「ゼロ・ディテクター」を紹介した。「ゼロ代名詞」の指導／習得において、第二言語習得理論と自然言語処理の融合が効果的である点をまとめると、以下の2点がある。

(1)「インプット強化」は、知識として与えにくい学習項目、「気づき」が起こりにくい項目に、より効果があるはずである。

(2) 自然言語処理による自動検出は、人手による検出が困難な項目、人手による検出にゆれが生じやすい項目に、有効である。

例えば、結束表現である「指示詞」と比較した場合、(1)、(2)の両点において、「ゼロ代名詞」がこの融合の利点をより有効に活用した選択であると言えるだろう。

今後の課題としては、強化（ゼロ代名詞の明示）以外の「気づき」を促す手法として、結束関係の理解確認のための質問文（「誰が紙コップを持っているのですか？」など）の自動生成などが考えられる。

参考文献

- Chapelle, C.A. Multimedia CALL: lessons to be learned from research on instructed SLA. *Language Learning & Technology*, 2/1, 22-34.
- Doughty, C. 1991. Second language instruction does make a difference: evidence from an empirical study of relativization. *Studies in Second Language Acquisition*, 13, 431-470.
- Ellis, R. 1995. Interpretation tasks for grammar teaching. *TESOL Quarterly*, 29/1, pp. 87-105.
- Gass, S. 1997. *Input, Interaction, and the Second Language Learner*. Lawrence Erlbaum Associates.
- Holland, V.M., Kaplan, J.D. and Sams, M.R. (eds.) 1995. *Intelligent Language Tutors: Theory Shaping Technology*. Lawrence Erlbaum Associates.
- Krashen, S. 1985. *Input Hypothesis: Issues and Implications*. Longman.
- Nerbonne, John. 2003. Natural language processing in computer-assisted language learning. In Mitkov, R. (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford University Press.
- Schmidt, R.W. 1990. The role of consciousness in second language learning. *Applied Linguistics*, 11/2, 129-158.
- Sharwood Smith, M. 1991. Input enhancement in instructed SLA: theoretical bases. *Studies in Second Language Acquisition*, 15, 165-179.
- 磯江健史, 竹井光子, 相沢輝昭. 2004. ゼロ連体格代名詞の自動検出システム. 言語処理学会第10回年次大会発表論文集.
- 横山紀子. 1999. インプットの効果を高める教室活動: 日本語教育における実践. 日本語国際センター紀要第9号