

報 告

# 多言語対応専門日本語読解学習支援システムの構想について

仁科喜久子

東京工業大学留学生センター 〒152東京都目黒区大岡山2-12-1

E-mail knishina@ryu.titech.ac.jp

## A Conception of a Multilingual Dictionary Linked to an Online Technical Japanese System

NISHINA, Kikuko

*International Student Center, Tokyo Institute of Technology  
2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152-8500*

本稿では多言語辞書対応専門日本語読解支援システムのコンセプトと内容の概略を述べる。このシステムの開発は東京工業大学留学生センターのプロジェクトチームにより、1998年度から始まり、5ヶ年計画で行われる。本センターで日本語研究を受ける留学生の殆どが理工系学生であり、その多くは英語を母語としない。一方、日本語で書かれた専門書読解のためには英語を介する以外には殆ど手段がない。特にアジア諸国の学生にとっては、母語以外の言語を通して、言葉を理解するのは困難だという声を聞く。このギャップを解消するための方策として、学生の母語から理解できる辞書を装備した専門書読解システムを開発するものである。本システムは、自然言語処理で用いられる形態素解析処理と構文解析処理の手法を用いる。語の意味に関しては、母語対応の意味辞書が複数存在する。さらにテキスト中の文法構造が理解できない場合、構文木を表示し、文中の構文関係が視覚的に分かるように出来ている。来年度からはマルチメディア対応して、音声システムの装備を行う予定であり、テキストを理解するための練習問題も搭載予定である。

キーワード: 理工系留学生 専門書読解 多言語対応辞書 構文構造 マルチメディア対応

### 1. はじめに - 計画に至った経緯

東京工業大学留学生センターでは、1994年創設以来理工系のための専門日本語理解に供する教材の開発を行ってきた。また研修コース、全学対象日本語コース（日本語補講）および大学院電気工学専攻において科学技術日本語の授業を開設している。これらのコースでの使用教材を留学生センターホームページ (<http://www.ryu.titech.ac.jp/ryu.html>) に掲載してきた。本計画は、さらに学習者の母語に対応した辞書を装備した本センター独自のオンライン学習支援システムを構築するものである。

### 2. システムの内容

#### 2.1 システム構築のコンセプト

大学院レベルの留学生にとっては論文読解も論文作成も英語である場合が多い。にも拘わらず、

研究室でのコミュニケーションは日本語が基本であることは過去の調査から明らかである。(参考文献7) 日本での留学期間が長い学生の中には、その隙間を埋めるために日本語学習を続けるという例がしばしば見られる。しかし多くの場合は研究に忙しく、独習に頼ることが多い。また日本に留学する学生の多くは非英語圏国からであるが、日本語学習の媒介言語は多くが英語である。留学生は英語が出来ることが前提とされるが、母語ではない言語を通して学習目標言語を正しく理解するのは障害が多い。この障害を解消する方法として、非英語圏からの理工系留学生のために母語から学べる日本語学習システム構築を企画した。

このシステムは最終的には読解・聴解・作文・発話の四技能を統合したものとするが、現時点では

読解部分の開発を行う。現在、科学技術日本語授業で用いる初中級以上の教材を留学生センターホームページにしている。また CASTEL-J が備えている学術用語集および北陸先端科学技術大学院大学で開発した Dictlinker も学習支援システム上においている。これらのツールやテキスト文章の使用許諾が学内のみという制限から、学内の日本語学習者にのみ利用を図っている。本年度の課題として、(1)システムに多言語辞書を加えること、(2)独習用文法学習システムの搭載することを中心に現状と展望を報告する。

2.2 システムの概要

自然言語処理の技術を日本語学習支援に導入することがこのプロジェクトの特色である。データベースサーバーには多言語辞書データベース、教材データベースが格納され、WWW サーバーにつながる。多言語辞書は、中国語、タイ語、インドネシア語、マレー語の各言語と日本語が対訳形式でそれぞれ格納される。また形態素解析プログラムおよび構文解析プログラムが WWW サーバーに搭載される。学習者はパソコン端末機を使用し、学習したい教材を教材データベースから選ぶか、学習者自身が読みたいと思う論文などの持ち込みテキストを入力し、学習する。

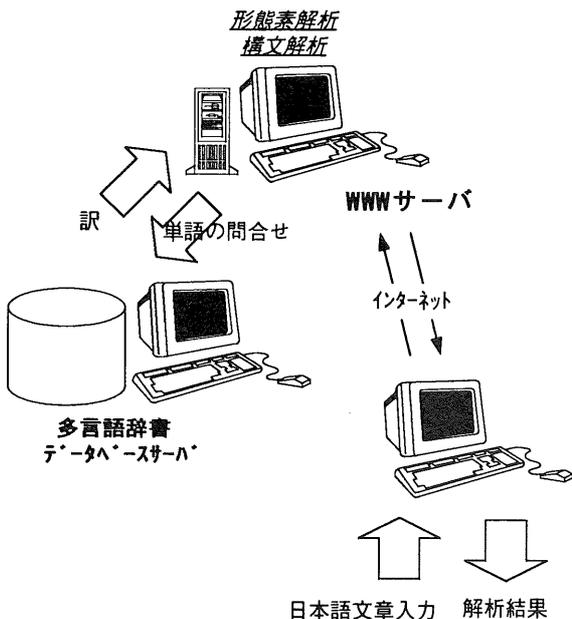


図1 学習支援システムの構造

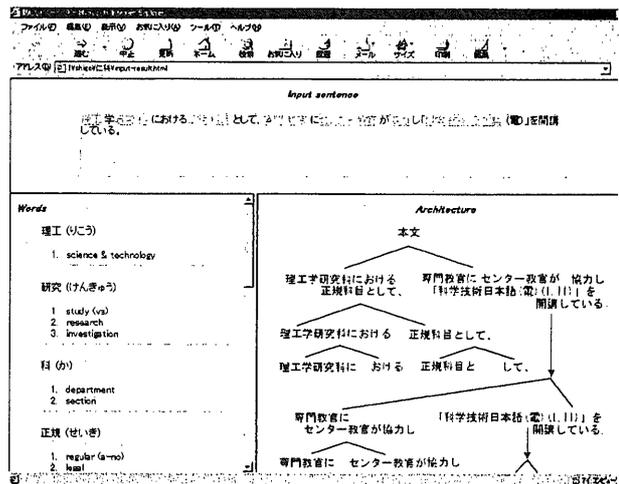


図2 学習システムの表示画面のイメージ図

(1) 母語対応辞書

システムに読み込まれたテキストは、図2の上部に表示される。選択したテキストはすべて形態素解析にかけられる。形態素解析とは、言語文における単語の下位単位をいう。「お-客-さん」のような単語中で、「客」は独立した単語であるが、「お」と「さん」は非独立の接辞である。このとき「お」「客」「さん」をそれぞれ形態素と呼ぶ。形態素解析プログラムは京都大学で開発された JUMAN を利用している。プログラム処理により解析された形態素が、学習者が選んだ多言語辞書データベースの中の一つの対訳辞書に行き、単語項目を問い合わせる。もし参照したい語の単語項目があれば、図2の左側に見られるように単語の意味が学習者の母語で表示される。図2の形態素解析結果の表示は、寺・北村らにより開発された DL の結果を用いたイメージ図であり、システム完成後の形とは多少異なる。(参考文献1)

(2) 構文構造の表示

図2の右側には、選択したテキスト中の任意のひとつの文について、構文解析を行った結果である。学習者がこのツールを利用することで、文法構造が学べる。解析プログラムは JUMAN と同じく京都大学で開発された KNP プログラムを使用する。現在の表示形式は、KNP から得られる木構造が学習者にとって、分かりやすく親しみの持てる形式かどうかについてアンケートを取った結果を

反映したものである。プログラムは JAVA で書かれており、<TEMP>というマーカーをクリックすると、更に木構造の詳細が見える。<TEMP>のあるところを繰り返しクリックすると、最終的には、すべての語の品詞が表示される。以上が本システムの現在までの開発状況である。

### 3. 開発の現状と検討点

今後実際に理工系専門日本語のための学習支援として実用化するまでには、諸々の問題が残されている。次の①から④の4項目は、解決すべき問題点である。

①フォント 英語を含めて5言語の対訳辞書を扱っているが、そのうち中国語とタイ語文字は特殊な文字コードであるため、WWW上に載せる方法について問題が残され、現在検討中である。

#### ②辞書の絞り込み

現在使用されている Dictlinker では、知りたい語に対して出力される辞書項目は、テキスト中の形態素の文字列と一致すれば、すべてを表示する。この場合は、学習者が候補の読み方と語義から正しいものを選ばなければならない。独習用とするためには、システムとしてかなり絞り込まれたものでなければならない。このためには、文あるいはディスコース中の共起関係を考慮したプログラム作成が今後の課題として残されている。

#### ③辞書項目の拡充（使用言語と分野別）

このシステムは、留学生の比率の高いアジアの4か国からの言語に焦点をあてている。そのため現在は CICC 辞書を利用している（参考文献8）。アジアの言語としては韓国語、ベトナム語、アジア以外のアラビア語、ロシア語、スペイン語などへの拡充も考えるべきである。また、今年度は情報系分野を対象としたが、他分野の専門辞書データを確保しなければならない。

#### ④構文解析結果の表示法

KNP の解析結果の表示は、JAVA によってプログラムが書かれている。木の枝の分岐点にある<TEMP>をクリックすると、下位の木の枝が出てくるが、それぞれのブロックに自動的なタグ付けが

欲しい。たとえば、文中のある句 A と B の関係を A=述句 B =主部などのように表示したい。以上の点が現在開発している中で出てきた問題点である。

### 4. 今後の計画と展望

将来の開発計画として、以下のような拡充を検討している。学習上有用な機能として、次の⑤～⑪の7項目の搭載が考えられる。

#### ⑤履歴システムによる評価

履歴を取ることは、学習者・教師・システム作成者のいずれにとっても必要である。

#### ⑥練習問題の充実

留学生センターでは「やさしい科学技術日本語読解入門」（参考文献2）を発行した。本書は初級終了程度の学習者を対象とした科学技術日本語テキストであり、仁科の理工系日本語教育に関する調査結果に基づいて書かれたものである（参考文献6）。文法やディスコースに関わる練習問題が掲載されているが、教室で教師の指導の下で使用されるものとして作成されている。今後は独習に耐えられるように変更を加え、本書の内容をシステムに入れる予定である。

#### ⑦辞書から参照する例文コーパス

教科書（参考文献2）でも例文が入っているが、学習者のレベルと専門分野などの興味に合わせた例文が得られるような大量の例文コーパスの作成も検討する。このコーパスは学習者のみならず、日本語教師にも教材作成やシラバス作成上、必要なものである。

#### ⑧母語からの作文支援

多言語辞書、構文学習システム、文法練習、例文コーパスなどはそのまま作文支援に利用できる。

本システムでは母語から伝達したい概念内容を作文し、それがチェックできるはずである。このようなシステムは学習者に便宜をあたえるものと思われる。

その他、⑨マルチメディアとしての機能（音声・映像）⑩音声対話支援、⑪聞き取りなどの機能を付加することも現在検討中である。これらは2004年3月を目処に実装する予定である。

参考文献:

- 1.寺・北村・落水 (1996) 日本語読解支援システム Dictlinker 日本語教育学会秋期大会予稿集 pp.43-38
- 2.仁科 編集『やさしい科学技術日本語読解入門』(1999)東京工業大学留学生センター pp.1-120
- 3.仁科 (1999) 多言語対応専門日本語学習支援オンラインシステムの開発に向けて 第1回専門日本語教育研究討論会講演要旨集 pp.2-4
- 4.仁科・奥村 (1999) 科学技術日本語学習支援多言語対応辞書の項目設定に関する研究 日本語教育方法研究会誌 Vol.6 No.2 pp.32-33
- 5.仁科・小島 (1998) オンライン科学技術日本語学習システム構築のためのテキスト解析 (2), 日本語教育方法研究会誌 Vol.5 No.1 pp.34-35
- 6.仁科 (1997) 科学技術日本語学習システム開発のための基礎的研究 (学位論文) 東京工業大学 pp.1-127
- 7.仁科 (1991) 理工系留学生の日本語学習および能力に関する実態調査 (1990) 年度文部省科学研究費補助金試験研究 (B) 「科学技術日本語教材開発」中間報告書研究代表者仁科喜久子 pp.1-84
8. Machine Translation System Center of The International Cooperation for Computerization (1995) Technical Term Dictionary Part 1,2

なお本研究は留学生センターにおける総合的な研究プロジェクトであり、一部を文部省科学研究費

補助金基盤研究 (C) 「概念知識に基づく母語からの理工系専門日本語習得方法に関する研究」 (研究代表者 仁科喜久子) および (財) 電気通信進行事業団助成金 「多言語対応理工系留学生のための日本語読解学習支援システムの開発研究」 (研究代表者 仁科喜久子) により賄っている。

謝辞: 北陸先端科学技術大学院大学情報理工学研究科奥村学助教授は本プロジェクトでは科学研究費および電気進行事業団助成金の研究分担者であり、自然言語処理技術において、重要な役割を担っている。また、フーズラボの技術担当者傳亮氏、杉本茂樹氏、さらに本学条理工学研究科計算機工学専攻修士課程八木豊氏、阿辺川氏ら多数の協力により、ここまで立ち上げることができた。また本センター小島聡助教授はホームページ上の学習システムの立ち上げから管理運営を今日まで責任者として行ってきた。このことに感謝の意を表し、ここに特記する。その他、多くの研究分担者と協力者が関わり、協力をいただいたことも申し添える。

著者紹介

仁科喜久子: 東京工業大学留学生センター教授 【経歴】 学歴: 東京女子大学大学院文学研究科修士課程修了 博士 (学術) 東京工業大学 職歴: 埼玉大学教養部 助教授、東京工業大学留学生センター助教授を経て 現職。同大学大学院社会理工学研究科専攻併任教授 【専門】 日本語教育・日本語学
--