

語彙の重要度に基づく文構造自動書き換えシステム

- 「やさしい日本語」 への変換による読解支援 -

中尾 桂子

大妻女子大学短期大学部

森下 淳也

神戸大学大学院国際文化学研究科

文構造の書き換えによる文章の簡略化が、専門教育を受ける初級日本語学習者の読解支援として行われているが、人手による簡略化は量産できないため、読みたいものの入手が容易くなるよう、文の簡略化を自動化するシステムを制作する。本システムは、自然な日本語の文章を電子化してシステムに入れると、各文を構成する語のうちの重要語を残して余剰成分を削除、または、表示を変えて出力し、型どおりの「単純な」文に自動変換する。その際、文脈を把握するために不可欠な語が削除されてはならないため、文脈上重要度の高い語はそのまま保有しておくことで、文脈を維持するように気をつけている。一般に言う文章の簡略化では、通常、自然言語処理技術を利用した構文解析に基づくが、本稿のシステムではごく限られた日本語知識の初級日本語学習者を使用者として想定しているため、構文を正確に解析する方法で文の簡略化を目指すのではなく、いわゆる初級日本語学習者にとって「単純な」基本文型を明示する簡略化を優先している。本システムでは、文脈維持のために、述語に基づく文の必須要素とテキスト中の使用語彙を統計的に計算した結果から語の重要度を判断し、強調表示に利用しているが、テキスト量により重要度に信頼性が低くなることが予測されることから、本システムで簡略化された文章を利用し、初級から中級レベルの日本語学習者に対して読解テストとそれに対するアンケートを行って、本システムの可能性を検証した。結果、初級学習者にはある程度の支援が行えるものであると考えられることから、初級日本語学習者の専門教育における読解支援の一環として、文を簡単な構造のものに書き換える方法とシステムを提案する。

Automatic rewriting system of sentence structure

based on Vocabulary importance degree

Keiko Nakao

Japanese literature course
Otsuma Women's College

Jun-ya Morishita

Faculty of Cross-cultural Studies
Kobe University

Sentences have been simplified by rewriting the sentence-structure as comprehension support of the beginner's who receives the professional education. We produced the automating simplification system because the simplification by people cannot be mass-produced. Putting some electronic Japanese sentences in this system, it changes automatically the outputs depends on the user's settings, as leaving important words or deleting surplus elements. In general, it is usually based on parsing using the natural language processing technology in the simplification, but in this system it does not concerned to simplify it by the method of understanding the syntax accurately. This system can show the rough outline, because of giving priority to the simplification that specifies the basic sentence pattern and importance degree of vocabulary contained in the text. Asking 10 foreign students if it's useful or not, they said that they think it might do some support for beginners. So we propose this rough type system has also availability to the comprehension support for beginners.

1. はじめに

義務教育機関の日本語学習者や大学院への留学生など、専門教育を受けることが主たる目的である初級日本語学習者、並びに、その指導者には、できる限り早く専門知識の学習に移行したいという希望がある。

そのため、小中高の外国人児童・生徒や大学院の留学生には、初級の段階から専門教育につなげるための教育が、読解を中心に考慮されはじめ、多くの教材が作成されてきた[1-2]。

また、授業や課外における教員の指導の他にも、逐次読みや内容解釈の支援として、チュー

ター制度や読解支援システムが利用される[3-4]。

e-Learningに関連する読解システムの多くは、文章の構成要素である単語(形態素)レベルに対訳をつけるというものが多く、語彙力の補助や辞書で意味を確認する作業の軽減が利用の主な目的と想定されている。

ただし、初級日本語学習者が読む場合には、語彙レベルの意味が理解できたとしても、内容の論理的な理解は難しいときも多い[5]。それは、ネイティブのための自然な日本語の文章構造が「複雑」であることによる。

そこで、初級レベルの文法や構文知識でも、ある程度の因果関係などを論理的に把握するこ

とを目的とし、初級学習者の既習の文法構文知識の範囲で専門書などを書き直す、簡略化が行われ、簡易版テキストの使用や簡略文章を用いた指導が行われるようになっていく[6-7]。

しかし、簡略化のための書き換えは、特に決まった方法があるわけではなく、書き換えに携わった個々人の文章表現力や主観に影響されることも多く、一貫性や均一性に疑問が残る場合もある。

また、書き換えの労力が大きいことから、希望する資料の簡易版を学習者の誰もが必要となきにすぐ手に入れられる環境にあるわけではない。

希望資料の自由な入手や多用な資料への対応を考えれば、ある程度の部分までの自動化が望まれる。

2. 研究目的

複雑な構造の文を含む文章を、単純な構造の文で記述された文章へと自動的に書き換える「文構造自動書き換えシステム」を開発し、その有効性を検討する。

このシステムは、専門教育を受ける初級日本語学習者の主な情報収集手段となる文章読解を支援するものであり、初級日本語での情報収集環境を提供するものである。

3. 構造簡略と自動化

自然な日本語の文章を電子化し、システムに入れると、各文を構成する語のうち、重要な語を残して余剰な語を削除、または、重要語を強調表示し、単純な構造の文に焦点を当てる形式で、簡単な構造に自動「変換」する。その際、文脈上、重要な意味の語が削除されてはならないため、重要度の高い語はそのまま保有しておくことで、文脈を維持するように気をつける。これは、中尾・森下(2007)で検討した文章の視覚的簡略化法を応用したものである[8]が、以下、簡略化の内容と方法について説明する。

3.1. 単純な構造の文

文が単純か否かについては、様々な捉え方があるが、本稿では、日本語能力検定試験の 3, 4 級を初級とし、そのレベルで扱われる基本文型をベースとした、いわゆる単文と、2 節からなる複文を単純な構造の文とする[9]。

『日本語能力検定試験出題基準』の 3, 4 級の基本文型文法項目とは、大別すると、①名詞(形容動詞、副詞)＋「だ(です)」, ②形容詞(相当語句を含), ③動詞(＋接尾辞, 補助動詞)④「てください」や「(に)なる, (に)する」「動詞ところだ(です)」等いくつかの語を慣用的に組

み合わせたもの、といった 4 種類程度を述語として利用するもので、肯定、否定、過去、現在の別はあるが、それら述語で文が終わるものとされている。

『日本語能力検定試験出題基準』に整理されている述語が、それぞれの初級日本語教育において提示される場合は、通常、いわゆる「基本文型」に準拠した語のセットで示される。それは、これらの述語が文末に来ると、省略、非省略の別はあっても、以下のような構造を基本的に持っていると考えられることによる。

[例]

- ・「～は～だ。」
- ・「～は(～が)～形容詞。」
- ・「～を～他動詞。」
- ・「～が～自動詞。」
- ・「～が～にある(いる)。」

語学教育における「基本文型」という考え方と、『日本語能力検定出題基準』の基本文型は少々異なっているが、ここでは、単語のセットで示される上記例のような文の型を基本文型と呼ぶことにする。

この、いわゆる基本文型は、英文法を参考に形成された考えで、戦前の日本語教育から脈々と受け継がれてきたものであるが、汎用性と一般性の高い文法要素の組み合わせを指すとともに、基本的にこれに還元されるという意味での規範性も含み、日本語教育ではシラバスインデックスとしても認識されている。

日本語教育では、この基本文型に基づいて初級学習者に文の組み立て方を紹介するが、その場合、それぞれの文型を構成する必須の要素を、可能な限り 1 語に絞り込んで用い、必要最低限の語(以下、必須要素)だけで構成された文を用いる。

この必須要素だけで構成された文で、単語の接続規則や文が成立するための必須の要素の重要性を学ぶため、初級学習者のコミュニケーションの練習は、基本の型に基づいて、ごく限定された必須語のみで組み立てられた文を用いて行われる。

したがって、初級の学習者は、日常のコミュニケーションでも、日本語らしくはないが、正確に意味が理解でき、情報を伝達することもできる基本文型に準拠して、教室で学んだ構造の文を用いて話す。

例えば、「私は朝ごはんを食べませんでした。(私は)とてもお腹がすいています。ですから、今日、(私は)大きいごはんを食べます」というような具合で、中間言語、すなわち、俗に言うフォリナートークである。

しかし、初級学習者のコミュニケーションは、

一方向であることが多い。多くの場合、日本語ネイティブからの返事は基本文型が明瞭に捉えられるとは限らないためである。したがって、十分、返事が理解できなかった初級学習者は、聞き取れた単語から、場面や状況といった情報を考え合わせ、内容を推測するということになる。

つまり、現行の日本語学習の流れで行くと、初級学習者が受け取る場合に「わかりやすい」または「やさしい」単純な構造の文とは、限られた語だけで構成され、基本文型が明確に判断できる文である。

初級学習で用いられる日本語文の多くは、述語を1つだけ持つ文、すなわち、単文と呼ばれる形で紹介されるが、そこには、『日本語能力検定試験出題基準』3, 4 級であげられている「なければなりません」「てもいいですか」「かもしれない」といったいくつかの語で組み合わせられた若干の表現的な文型も含まれる。

また、初級というのは、汎用性の高いものと構造の単純なものとを併せて紹介するレベルであるため、述語を複数持つ文、いわゆる複文も初級レベルで紹介される。

初級段階で言う複文とは、2つの短文を組み合わせたものであるが、初級段階の複文には、「〜て」や「たら」「れば」「とき」「なら」で接続した副詞節か、基本文型を2つ以上組み合わせる連体修飾節や、準体助詞「の」や形式名詞「こと、もの」による名詞節、他に、複合辞を利用して複文を形成するものがあるが、本稿では、述語を2つ持つ副詞節(いわゆる連用節)と、連体修飾節(名詞節を含)をやさしい構造のものとする。

ところで、教育用の基本文型は、従来、必須要素の中心的なものと述語のセットで提示されることが多いが、実際には、文末にどのような述語が来るかにより、文章構造を形成する上で重要な機能を担う語の数や性質が異なり、基本文型の構造が実際の文と一致しても、それ以外の要素でも意味や因果関係の理解の上で必要なものが存在する。

そこで、文章構造を形成する上で重要な機能を担う語として、日本語能力試験 3, 4 級の語彙を述語とした場合の各述語が持つべき必須の要素の組み合わせで作成したリストを用意し、本システムでの「基本文構造」とする。

3.2. 文構造の変換と文脈構造の保持

自然な文章の文は、装飾や意図の正確さのために、様々な修飾語や余剰な要素が盛り込まれ、組み立てられていることが多い。この文を、別

途用意する「基本文構造」リストの述語部分に合わせて必須要素だけに縮約するのであるが、しかし、日本語の自然な文章では、省略や文脈のつながりによる言い換えなどで、必須要素が必ずしも全て記載されているわけではない。

また、意味を把握するために文脈上必要な情報は、必須要素以外で継続して出されている場合もある。

この、文脈上、必要な語は消さずに残し、必須要素と併せて、長い文を縮約するために、文脈上必要な語として、意味上、重要な語を特定し、各枠組みによる基本文構造リストに合致した語とともに残しておくことで文脈形成上重要な意味の語を保持する。

3.3. 語の重要度と文脈上の一貫性

オリジナルのテキスト中の語彙の重要性を計量し、重要度評価の基準とする方法として、中尾・森下(2007)では、残渣 IDF 値を利用したが、テキストが小さい場合はまんべんなく出現することや特異性に対して、正確な状態を示す指標としては疑問である。

テキストが少量であった場合を考えると、語彙の基本統計量(頻度、標準化頻度、Type Token Ratio、中央値など)で、使用頻度を割合でみた方がよいと考えられる。

ただし、使用頻度だけでは重要かどうかはまだ不明であることから、文の中での機能を、重要度判定に組み込む。

以上のような使用頻度の高さと、それに加えて、3.1 で述べた必須の語の機能的な重要性、すなわち、主文の主語述語であることや、述語の必須語であることに基づき重要性を示すが、さらに、基本文型の意味などの補足説明を添付する。これは学習者に任せる形にはなるが、重要度については、学習者により何がどの程度必要かが異なるためである。必要ない学習者には不要であるが、初級レベルの文法機能を手がかりに読解する経験が不足な場合の学習者には、読解時の大まかな着眼点と概略把握のための補足支援として有益である。

4. 文構造自動書換システム

本システムは、図 1 に示すように、3つの部分からなる。本システムの処理を概略すると、①読みたい文章をシステムに投入、②内部で下処理を行い、xml データに変換後、③投入時に設定した表示方法に基づいて表示返す、というものである。以下、それぞれの構成と各段階の処理について説明する。

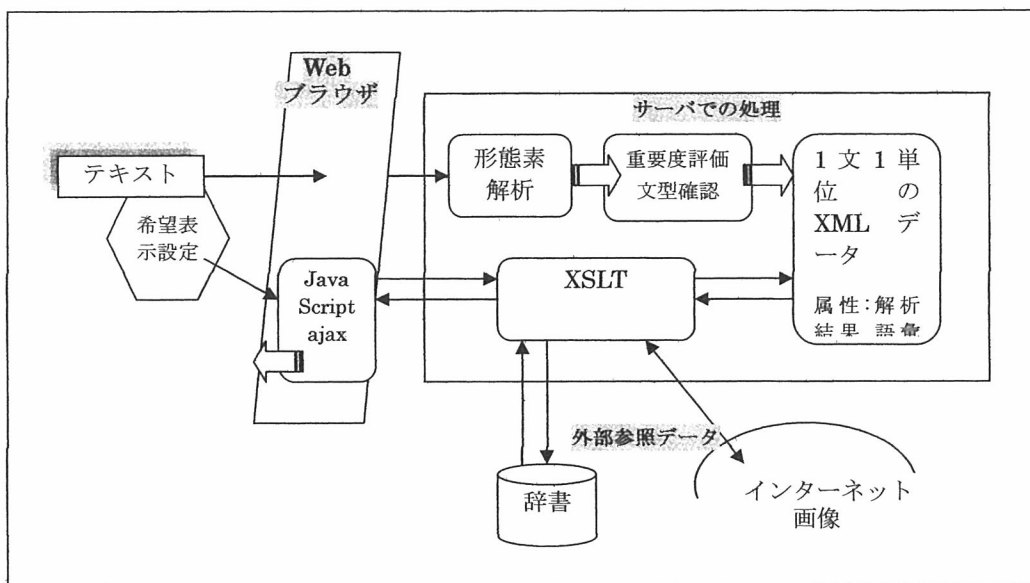


図1: 文構造自動書直システムの構成

4.1. テキストの投入

テキストの分量に応じてアップロードの方法を2つ用意する。1つは、ファイル単位で転送するもので、もう1つは上限を設けたフィールドに記入するものである。

いずれの場合でも、基本的に、1文は、句点を末尾に1つ持つもの、または、改行が1つあるものとする。句点のない文章は見出しやセリフ、引用などが考えられるため、文ではないと言えるのであるが、それらを後ろの文と区別する方が重要である。

また、段落が終わるまで改行が入っていない文章の場合は、句点が1文の手がかりとなるため、句点も1文の認定に利用する必要がある。

したがって、見出しや小タイトル等、改行が1つあるものも1文として扱い、それぞれの文どうしの関係は、配列以外、切り離して捉える立場である。

また、文字コードは、処理のそれぞれの段階で変換する必要があるが、アップロードされたテキストを受け取り、形態素解析する前に、認識して、文字コードを統一する。したがって、アップロードの際は、いずれの機械環境からでも利用が可能である。

なお、テキスト形式のファイルをアップする際に、テキスト内の記述形式をこちらの指定する形式で記述した場合は、元のテキスト文書の構造を反映させて表示することができる。これを利用する場合は、アップするテキスト内の章、節、段落といった階層間に、空行を、それぞれ、

章は3つ、節は2つ、段落は1つというように、反映させる階層の数だけ空行をあけることで、階層の区別を反映させた文書の構造を表示することができる。

また、ファイル単位のアップロードの場合、分量が多いと、表示準備が整うまでに時間がかかるという欠点がある。そこで、分量に応じておおよその処理時間をアップロード時に示すが、テキスト投入時に連絡先メールアドレスの指定を受けた場合には、表示準備が整い次第、結果出力のURLを指定先のメールアドレスへ送るというサービスを設ける。ユーザID、PSWを設定することもでき、URL送付時に送ることで結果やテキストの利用者を制限することができる。

4.2. テキストの処理

テキストは、Web経由で受け取った後、まずは、文の認定と分かち書きによる単語認定と、品詞、読み方といった単語別の情報を得るために、形態素解析を行う。解析にはChasen[10]を利用する。

次に、形態素解析結果はオンメモリで保有しておき、語彙の統計データを計算した結果と、述語に応じた構文上の必須性の高さにより、重要度を決定して各単語にフラグを立てていく。

テキストは、最終的に、xmlで1文を1単位として構造化するが、その際、形態素解析の結果と、統計値、共起語で見た場合の頻度の高さを重要度判定値とし、xmlで構造化したテキストの属性として、各単語レベルの下位に組み込

んでいく。これがテキストの下準備である。

4.2.1. 共起語を含む語彙リストの作成

語彙統計値を計算する前に、形態素解析結果に対して、4-gram でテキスト中に出現する共起語を調べる。形態素解析では、通常、単語単位に分割されるが、意味の上で重要な表現は必ずしも 1 単語からなっているわけではない。単語どうしの組み合わせで出現率を判断するかしないかにより、文脈内の意味上の重要度計算に影響が出るため、どのような単語が共に接続して出現するかという共起セットで単語を捉える必要がある。そのため、共起語の使用頻度の高さを調べ、いわゆる慣用表現と言えるものを形態素解析結果の範囲を超えて確認し、意味の上で重要な語の特定の語に共起語という単語の組み合わせの情報と、その組み合わせでの頻度情報を属性に組み込んでおく。

なお、通常、コーパス言語学で共起語を特定する場合、「ミニテキスト中の語の観測頻度を使って、同テキストの理論頻度と比」べるなどの方法で、前後の語どうしとの組み合わせ確率を Z-score 値で求める[11]。この値が高ければ統計的有意性が高くなるとして、判定していくのであるが、本システムでは、新イディオム探索を目的としているわけではないため、ただ、共に出現する組み合わせも含めて、単語の出現頻度を多用性と仮定し、重要だと推測するのみであることから、4-gram で接続する組み合わせで捉えられたものの頻度が 2 回以上あれば、表現のセットとして語彙リスト作成の際に組み込むというものである。

4.2.2. 重要度評価と重要度値の属性化

語彙の統計量とは、標準化頻度に基づくものとする。頻度は、利用回数であり、文法用語以外で使用頻度の高いものは、意味上、重要性が高いはずだという予測のもと、使用頻度を計上することが多いが、テキストの大きさに左右されるものであるため、頻度実数を見るのではなく、頻度数を標準化する。

これを利用し、テキスト分量に応じて、頻度が分量の何割に相当するかを決定するのであるが、大まかに、3段階のレベルに分けて捉える。例えば、テキスト中の語の頻度が 1 回から最大 10 回までだとすると、頻度 1 回から 3 回までは少数派レベル、4 回から 6 回までは中程度、それ以上は多数派とするもので、機械的に約 30%、30%、40%と 3 分割し、そのうちのどの区分に位置する語であるかという 3 区分判定結果を表す値に置き換えてフラグを立てる。この方法は、先の 4.2.1 で調べた共起語セット属性と共に、

オンメモリで保有してある形態素結果と合わせて、xml データ属性に組み込んでいく。

また、頻度数に基づき、特徴語を計上する。特徴語とは、一般にコーパス言語学の分野では、「あるテキストの一定の範囲を他と比較し、有意に一定以上出現するもの」[11]を指すが、本システムの利用者は、投入するテキストが通常 1 種類だと考えられるため、現在のシステムでは、「使用頻度が多く、必須性の高いもの」という雑ばくな程度の定義にとどめておく。今後、参照データと比較する準備ができれば、参照情報内の語の出現状況と χ^2 乗検定等で比較した結果、比較用サンプルデータと比べて出現割合が多ければ、特徴語だとするように考え直したいが、本稿では今後の課題とする。

4.2.3. 基本文型情報の属性化

上記、2 種類の単語、または、表現といったフレーズレベルでの重要度評価の一方で、文構造の表示のためのマーカーも、xml テキストの単語以下に属性として組み込む。

基本文型リスト、単文・複文リストを用意し、それに合致する組み合わせの語があった場合も、基本文型情報とする。基本文型リストは、文字列と正規表現からなるパターンと当該基本文型で注意する点の記述からなるが、このリストの基本文型に合致する語には、それぞれに、文型セットであること、並びに、高必須度のものであることを示す情報を付与する。これらの情報は、希望表示形式に応じて利用する。

基本文型リストは、3 章で述べたように、日本語の特徴を加味して検討しているが、基本文型として着目する点を簡略すると、係り助詞、格助詞、述語であり、かつ、受け身形や使役形など構造に変化を伴う接尾辞が現れる場合ということになる。バリエーションが多く、定型外の例外処理が必要な動詞述語文以外は、定型パターンリストで判断する。動詞述語は文構造のバリエーションが定型パターンにはできないものが多いことから、本来、必須要素の決定には、述語動詞の格枠組みを用いて、正確な必須語を選択することが望ましく、市販の構文体系を含む語彙体系辞書[12]を参照したい。しかし、専門知識を紹介する学術的な文章の読解を想定するため、述語動詞のバリエーションはさほど多くないという傾向が見られる[13]ことから、今回のシステム作成においては、語彙体系辞書を利用せず、体系辞書の組み込みは今後の課題とする。

非常に簡雑ではあるが、必須格の代表となる助詞、「は」「を」「が」にのみ着目して重要性を判定する。語彙の出現率が高い場合は何ら

運動した直後は、**脈拍数や呼吸数が増える**が、しばらくすると**もとに戻る**。

運動すると**体温もあがる**が、脈拍数や呼吸数ほど**変化はしない**。

これは、汗をかいたりして**体温があがらないように調節している**からである。

ほかの動物も運動すると、人と同様に、体の**変化が見られる**が、**脈拍数や呼吸数や体温は**動物によって**違う**。

図 2: 複文の場合の表示例 (大木 1995)

かの強調表示がなされるが、受け身、使役などヴォイス関連で文構造が変化するもの以外の、例えば、対格を「に」で示す必要があるもので、かつ、対格の語の出現頻度が低い場合は、今回のシステム内での重要度が低くなり、強調表示の対象にならない。

以上のように、形態素解析処理の結果と、統計値からなる重要度や文型リストの情報を数値に置き換えた情報とを、文単位でシークエンスに xml タグを付与し、元データとして表示出力に備える。

4.3. 表示

投入時に設定した表示方法に基づき、xslt とブラウザローカルの ajax を利用して表示を返すが、表示は、第一段階のものと、表示後の利用中に追加で指定する第二段階のものに分けられる。

どちらの段階でも、表示自体は、サーバに問い合わせで行うのではなく、ブラウザ上で JavaScript を用いてローカルに行う。サーバに毎回問い合わせない方法を採用するのは、表示までの迅速さや処理コストの軽減を考慮することである。このために、4.2 でも述べたが、テキストデータを一度 xml データに変換するのである。元データの構造をそのままに、しかし、どのようにも表示を変化させられるため、さら

には、基本的な加工の結果から得られる文構造判定のための属性情報をそれぞれの構成要素の下位情報として添付したままデータを扱うことが、速度やコンパクトさにおいて簡便なことによる。

第一段階の表示は、重要度や「基本文型」に基づく必須語のみ表示や、重要性の高い語を強調させた表示、漢字の振り仮名をつけるといったもので、テキスト自体に備わる表示上の問題を扱っている。

また、第二段階の表示は、利用の際に必要となった参考資料を追加するもので、辞書、画像、テキスト中の語の利用例一覧 KWIC 表示である。

表示の例を図 2、図 3 に示す。図 2 は、複文が多い文章で、振り仮名非表示の例である。また、図 3 は、接続詞の多い文章の例である。いずれも主文の中の必須要素の文字表示を大きくし、下線をつけて主文必須要素であることを示している。複文の場合の接続表現と否定の「ない」が含まれる場合、また、慣用表現的なパターンの型を示す表現文型は字色を変え、図 3 の網掛け(黄色いマーカー)のように、文の論理性を読み取る手がかりとなる接続詞にも色表示する。

基本文型リストに挙げられているものに関しては、その文法説明の注意が属性に埋め込まれていることから、必要な場合は、単語をクリッ

クすることで、その解説がフレームで参照でき、当該文型の場合の読み方の注意点を得ることができ、頻度と文型上の必須性ではまかなえない重要度判断の材料を学習者に与えるものである。

図 2.3 に示すような文章を読む途中で、単語の意味を知りたくなった場合、辞書機能を利用して、英語訳、日本語での言い換えや、日本語辞書を別ウインドウ(ブラウザによっては別タブ)に表示することができる。また、辞書ではよくわからない場合に備え、検索エンジン(google)を利用してインターネットのイメージ画像を検索し、1 つずつヒットした順に切り替えて表示させる機能を用意した。

その他、同文章内に複数出る単語の他の資料例を一覧表示にして使い方を比較することができる KWIC 表示を利用することができる。これらの機能は、数人の初級～中級の留学生に、どのような読解支援がほしいかというインタビューを行った結果を反映させている。インタビューについては 5 章で紹介する。

5. 成人留学生に対する印象調査

3 章で述べたように、1 文の構造上の必須要

素と、語彙間の共起や使用頻度に基づき計算された重要度を用いて特定する文脈上の重要語彙とを残して単純な構造に書き換えた文章が、初級日本語学習者にとって、元の文章よりも詠みやすくなっているか、また、内容の理解に齟齬が生じていないかについて、読解テストを行う。

テストは、トピック、事実関係、因果関係が理解できているかについての内容把握問題とし、テスト時の文章は、小学校程度の基礎知識について記述された文章と人文系論文を用いる。

事前調査として、日本語能力検定試験の 3,4 級に合格、または、その模擬試験等を用いて 2 級合格には至らない段階の非漢字圏日本語学習者 3 人に対し、任意の調査協力を依頼したテストでは、確かに読みやすいと思うという答えを得ることはできたものの、それは判断の根拠に疑問が生じるものであった。漢字が多い文章を見て冷静さを失い、教育的効果に着目するほどの余裕がなかったことで、システムの特徴や問題を考慮せずに、テストにのみとらわれていた。

そこで、本調査では、初中級段階の漢字圏の留学生 10 人にアンケートを協力することにした。自身の読解の問題や初級時からの課題を客

1. 動物埴輪の出現から終焉まで

古墳文化の波及に伴って、東北地方でも埴輪を樹立する古墳が現れた。

しかし、その後は、独自の継続的な発展を見せるのではなく、埴輪文化が北の地域まで広がり、そして短期間のうちに衰退していった。

たとえば、4世紀の築造と推定される宮城県名取市雷神山古墳、青塚古墳、山形川西町天神森古墳からは壺形と推定さ

れる埴輪の破片が出土している。

また、福島県会津坂下町鎮守森古墳からは川西編年のⅡ期にさかのぼる円筒埴輪も発見されている。

しかしながら、福島県会津盆地や山形県米沢盆地では

埴輪が発見されない空白地域になっている。

図 3：接続詞の多い人文系論文の例(今津 1990)

観的に振り返ることができ、かつ、余裕をもってテストとアンケートに臨めるとともに、自身の意見の十分な説明が期待できることによる。

まず、読解テストとして、運動直後の脈拍数、呼吸数、体温の変化といった基礎知識について記述された文章を読み、どのようなときに脈拍数の変化が生じるか、脈拍数と同様に変化するのは何か、どうして変化しないものがあつたのかについての質問に答えてもらい、その後、文章の視覚的な効果や、自身の読解の問題について回答してもらふ。

文章は、はじめに、主文の述部と連体修飾節内の動詞述語以外を消去した文章を読み、次に、全文表示の文章を読み比べる方法で比較してもらった。

文章読解テストが全問正解であつたのは 3 人であつたが、不正解の回答を含む留学生は、全て、全文表示で一部強調表示の文章の方が情報収集に便利であるとしていたが、おそらく、漢字圏であるため、語彙情報にのみ頼り、構文情報を考慮せずに語彙力だけで読んでいるため、内容の因果関係が十分読み取れていないことによると考えられる。

テスト後、主文の述部と連体修飾節内の動詞述語とそれらの述語が必須とする語を重要度に応じて強調表示した文章を見せ、先の 2 つの文章と比較して検討してもらったところ、10 人中 8 人が、このタイプの文の簡略化が望ましいと答えている。その理由としては、文法的な手がかりが示されていることと、それに応じてどこを見て読めばよいか分かるので助かるという印象を答えている学習者がいた。

簡単な印象アンケートではあるが、これらの結果から、トピック、事実関係、因果関係について、書き換え前と書き換え後では、読解の簡便さが違うという印象を持つ学習者が多くいることが明らかになった。

表示部分への希望や希望する読解支援については、漢字のよみ、ことばの意味の他に、ある程度のイメージがつかめないものもあるため、画像がほしいという希望や、単語の使用例、文法的な意味の説明、小さい文字や行間の狭い文章を 1 行表示するなど読みやすくすることがあげられていた。

今回協力を依頼した留学生は、概ね、文章の簡略化には好意的であつたが、インターフェースに関しては、わかりにくいとする人や、漢字圏特有の読解時の問題が理解できていない学習者には支援として利用しづらいのではないかという意見から、実用化にむけての課題も多く見られた。また、アンケート対象の人数が少なく、十分なアンケート結果かどうかは検討の余地が

あることから、改良と、アンケート調査を継続して、理解程度の変化を分析、検証することで、システムの有効性をさらに考察していきたい。

6. まとめ

本システムの文構造の書き換え方法は、一般に考えられている簡略化システムのように、自然言語処理技術を利用した方法とは異なるものであるが、ごく少数ながらも、留学生に対して行った読解テストとアンケートによれば、本システムの手法でも、初級の段階からの専門書読解にはある程度の支援として利用でき得るものであると考えられる。

本稿は、初級日本語学習者の専門教育における読解支援の一環として、文を簡単な構造のものに書き換えるシステムを提案したが、今後、より完成度の高いものになるよう再編纂しながら Web 上での公開を検討し、多くの初級日本語学習者に役立ててもらえるよう努力していきたい。

参考文献

- [1] 近松暢子(2003)「外国語としての日本語の読み・読解研究」『第二言語習得研究への招待』くろしお出版:67-85.
- [2] 岡崎敏雄(2004)「外国人年少者日本語読解研究方法論：原理論」『文藝言語研究 言語篇』45 号 筑波大学大学院人文社会科学研究科文芸・言語専攻 29-46.
- [3] http://www.mext.go.jp/a_menu/shotou/clarinet/003/001/004/002.pdf
- [4] <http://language.tiu.ac.jp/>
- [5] <http://www.jaist.ac.jp/~tera/>
- [6] 仁科喜久子・土井みつる・高野知子(2007)『初級文型で学ぶ科学技術の日本語 -An Introduction to Technical Japanese-』スリーエーネットワーク.
- [7] 光元 聡江・岡本 淑明(2006)『国語教科書対応—外国人児童・生徒を教えるためのリライト教材』ふくろう出版.
- [8] 中尾桂子・森下淳也(2007)「読解支援のための語彙統計量に基づくテキストの動的拡張—語彙の重要度に基づくテキストの簡略化—」情報処理学会シンポジウムシリーズ、人文科学とコンピュータシンポジウム 2007:ポスター発表.
- [9] 国際交流基金(1998)『日本語能力日本語能力試験出題基準』(財)日本国際教育協会.
- [10] <http://chasen-legacy.sourceforge.jp/>
- [11] シルヴィアン・グレンジャー(船城道雄・望月通子監訳)(2008)『英語学習者コーパス入門』研究社.
- [12] NTT コミュニケーション科学(1999)『日本語語彙体系』岩波書店.
- [13] 中尾桂子(2008.12 発表予定)『学習者コーパスの解析と評価(仮題)』大妻女子大学国文学会.