

小学校教科書の電子的再構造化

—外部情報参照のための構造化テキストの拡張—

中尾 桂子

神戸学院大学 人文学部

森下 淳也

神戸大学 国際文化学部

日本語 Non-Native 児童のための教科学習では、日本語能力や知識を補うために、読解、学習支援が行われる。本稿では、この教科学習支援の一環として利用する教材を想定し、教科書以外から画像やテキストを取り込んで元の教科書の補填を試みる。補填材料は、単語の読み方、イメージを膨らませるための画像や音声など、本来のテキストにはないものである。

Reconstruction to the schoolbook text by structure-ized text extension with external information

Keiko Nakao

Faculty of Humanities

Kobe Gakuin University

Jun-ya Morishita

Faculty of Cross Cultural Studies

Kobe University

In the subject study for a Japanese Non-Native child, the support is offered to reading comprehension or studying in order to compensate Japanese and knowledge. In this paper, we try to reconstruct the original textbook taking in pictures from other than textbook as a part of subject study support.

1. はじめに

本研究は、Non-Native 児童の教科学習支援を目指すものである。小学校教科書のパーソナライゼーションを行い、個人の能力を補う形式の読解力養成支援の一つの方法を提案するものである。

小学校在籍の外国人児童（以下 Non-Native 児童とする）に対して行われる教科学習を目指した日本語教育では、小学校の検定教科書が教材となることが多い。それは、教科書が、Non-Native 児童にとって、学習題材であると同時に、学習用日本語の模範例集ともなるためである[1]。

小学校の実際の教室活動の日本語は、口頭発話であっても、日常会話とは異なり、漢語や敬体が多用された学習用の日本語である。視点や立場の違いを明示するための言語的コードスイッチングが多く[2,3]、また、社会的、文化的に特化した文脈で自明の情報がたびたび省略される。

視覚的な文書構成は、共通部分が多いが、日本の小学校教科書の文章（discourse）の流れは、母国語の言語的特性とは異なるため、母国で培った読解ストラテジーをそのまま利用して日本語の教科書から必要な情報をスキ

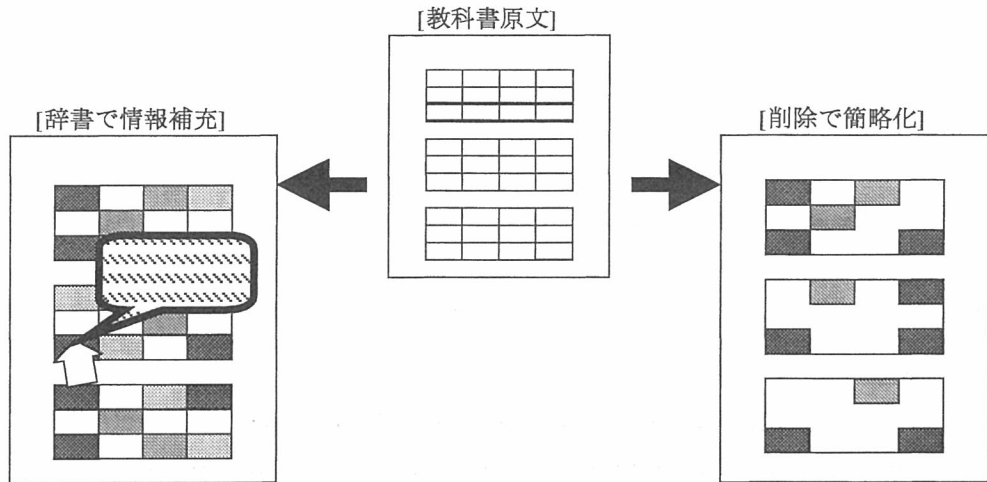
ミングすることは難しい。

したがって、Non-Native 児童は、教科書で概念や学習用の語彙を学ぶのみではなく、その他に、教科書の文章の流れ、省略の仕組み、それに応じて展開されている構成を把握し、教科書から、学習活動、思考活動に必要な情報を素早く的確に読み取る技能が必要になる[4-6]。

言語能力は個人差が大きく、一様には判断できないものである。にもかかわらず、その指導は、かならずしも個別に行われるとは限らない。また、言語的な能力の向上は、個人の能力と理解のスピードに合わせて繰り返し指導が必要なものであるが、十分ではないことも多い。個人のペースや目線に合わせた指導が難しく、教材開発等に関する教師の負担は大きい。

2. 一般的な読解支援と本研究の位置づけ

1節で述べたような背景をふまえ、我々は、Non-Native 児童のために情報スキミング技能養成を期待して、ディスコースの再構造化と再編纂の繰り返しによるテキストのパーソナライゼーションに取り組んでいる。



[図 1: 読解支援のあり方]

一般に、Non-Native に対して行われる個別対応の教材はコンピュータを用いるものが多い。これらは、テキスト読解支援のあり方から見ると、図 1 に示すように、大きく 2 つの立場に分かれる。1 つは、テキストを構成する文の構造を簡略化する立場であり、もう 1 つは、テキストに付加的に情報を補うアプローチ、または、不足している情報を補填するというものである。

テキストの文の簡略化とは、文中の、いわゆる文節を、述語が取り得る必要最低限の文節(必須項)に絞り、構造をシンプルにすることで、いわば、文を短くすることである。小学校在籍の Non-Native 児童の学習支援としては早くから認識されており、算数や理科、社会の教材で、実現方法が提案されている [7,8]。

もう一方のテキスト補充方式は、主に、Web 上の電子化されたテキストに対して、電子辞書で新出語の意味情報を補う形式のものである。こちらは、辞書を引くことができる年齢以上の学習者を対象として提案されるものが多く、主にインターネットやパソコンを介した支援である。

本研究では、これまで、読解支援のうちの「簡略化」という方向でテキストのパーソナライゼーションを検討してきた。ただし、従来型のテキスト簡略化の考え方には以下のような問題がある。

- 1) 元の教科書と見比べることができず、簡略化された部分が把握できないため、重要箇所と副次的な箇所の価値の違いが把握しづらい。
- 2) 簡略化されていても、情報は、常に受け身で与えられる。文章構造や構文を手がかりとして自力でスキミングできる能力がつきにくい。

いずれ自律的に読解していかなければならないことを念頭におけば、簡略時に、修飾部分を完全削除するのは不相当である。このことから、本研究では、学習者に削除の意図を十分把握させ、その意図を利用して、必要情報をスキミングする練習を行うべきであると考える。そこで、文構造上で必須部分と不要部分との違いを踏まえられるように簡略化して提示するという立場をとってきた [9,10]。つまり、「情報過多である」として過剰情報を削除するのではなく、テキストはそのままにして、必要箇所のみ焦点を当てるのである。これにより、テキストのスキミングの際の知識獲得の技能的な自立へと結びつくよう、自律を目指した読解支援を提案している。

3. これまでの問題点

本研究で言うテキストのパーソナライゼーションとは、使用者のニーズに応じてテキストの見え方を差し替えることである。すなわち、元のテキストデータに、そのテキストの一部分だけ重ねて表示するように見せること

であるが、実際は、教科書のテキスト構造の再編成を行い、視覚的にオリジナルを変化させて見せるというものである。手順は以下のとおりである。

- (1) Xml[11]を用いてテキストを構造化し、テキストに可逆性、可変性を持たせる。
- (2) Xslt[12]で検索した結果を別の xml テキストに返還する。
- (3) 一部が前景化して見えるように、テキストの一部の強調を html で表示する。

このような方法を用いた理由は、簡略化の履歴を情報として Non-Native 児童に示したいという希望と、小学校の教科書の構成が複雑であるという問題が影響する。

小学校の教科書は、指導者の立場や少し客観的な解説者の立場、身近なアドバイスを与える立場など、複数の立場で記述された文章や補足資料が散逸して並列するという構成をとっている。つまり、複数の文脈が同一ページに一度に展開される複雑な文書である。

このような書物のテキスト化に xml を利用することで、中尾・森下(2002)は、テキストの一部を前景化するといった表示切り替えが行えること、ならびに、同列に扱える文脈と別の次元の文脈とを区別して階層を視覚化することが可能であることを示した[12]。

また、それだけでなく、複雑なテキストの意味的な文脈構造明示にも応用できることを明らかにし、テキスト簡略化における1つの有効性の高い手段を示した[13]。

しかしながら、このアプローチを用いても解決できない根本的な問題がある。それは、「教科書の記述、補足資料の画像や説明に、そもそも不完全な部分がある」ということである。オリジナルのテキストにないものは、これまでの方法では対応できない問題であった。

読解により日本の小学校の文章の流れを把握するには、文章の流れの中で、省略、構成を把握し、教科書から、学習活動、思考活動に必要な情報を素早く的確に読み取るスキミング技能が必要である。これら、文章読解に備わるイメージスキーマの形成を考えれば、リソースを教科書から外に求めるという方法も検討する価値がある。

そこで、本稿では、簡略化、前景化といっ

た表示の切り替えに加え、補填というアプローチを検討してみる。補填に関しては次節4節から6節で述べる。

また、これまで、簡略化の1手段として前景化を提案していたが、実際には、名詞、動詞、学習指示を意図した数種類の文末表現を指標にしていた。これは、単語と指示表現を拾い読みする効果はあるが、単語の価値は同列に扱うものであった。

したがって、前景化の際、単語の文中での働きに応じて重みを変えて表示することも試みる。これは7節で述べる。

4. 補填

教科書のテキスト外に別の参考資料を求める場合、通常、辞書機能を追加するのが主流である。

しかし、辞書を引いて意味を理解する方法は、対象者が児童である場合、効果的ではない。それは、児童が発達過程にあり、以下にあげるような点が影響し、能力上、辞書の利用が無理なためである。

1. 母語でも語彙が少ない。
2. 同様のものが同様のものだと認識できない。つまり、事物のカテゴリ化能力が不完全で、おおよその意味が理解できてもそれが何に相当するものかわからない。
3. 過去の経験が少なく、眼前の新しいものと結びつけて意味を把握することが十分にできず、イメージの想起、認識が難しい。
4. 同じ事物を対象としても、母語と日本語とでは捉え方が異なり、発話かから同一物だと認識できない場合がある。
5. 単語が理解できても文全体で何を意味しているのかが理解できないことが多い。

児童は母語自体の意味の把握が不完全である。翻訳のための辞書が母語と与えられていたとしても、文単位で表される意味や概念が十分に理解できるわけではない。

全体像を構成する個々の事物が、それぞれどのように結びついているかといった関連性を理解していなければ、全体像を思い浮かべる際のイメージスキーマを得たことにはならない。辞書を見ても、全体像は把握しにくく、

それを土台に次の新知識を獲得していくことには至らない。

したがって、おおよその意味や全体像が把握できたとしても、次の概念に結び付けて学習を進めるうちに、いずれ、意味や概念がわからなくなってしまう。単語の意味や授業活動に対する理解は問題なく見えていても、テストで点数がとれないことや、学年が進むにつれて、進度に合わせていけなくなるという問題がよく指摘されているが、ここにその原因のひとつがあると考えられる。

以上から、本稿では、教科書の補填として、教科書テキストへ画像情報を追加することにより、イメージスキーマ形成をサポートする。教科書のテキストを補う目的で追加する情報は、主に具体例を示す画像であるが、以下の3つになる。

1. 文の構成要素である単語が示す具体例（画像・動画）
2. 同じ意味で使われている他の単語（テキストシソーラス）
3. 同じ言葉が別の状況で利用されている具体例（テキスト・画像）

これらの追加の画像や文字情報を教科書テキストに関連させて表示する。そのモデルを図2に示す。

具体例として画像を主に利用する理由は、「百聞は一見にしかず」の利点を考慮していることである。画像では視覚的に典型例を提示することになるが、その典型例は、ある意味で日本的なものごとの見方の方向を示すものであり、目線を統一し、何を見るべきかを画

一的にする。一般に、ステレオタイプで、没個性だと嫌われるものであるが、Non-Native 児童の最初のスキーマ習得の学習には効率的である。

また、複雑な作業や活動を通してはじめて把握できる抽象的な概念の場合、概念の下位構成要素である周辺関連物が、画像に現れるため、文字と同時に目で認識できる。イメージスキーマの形成に必要なものが複数提示できる利点から、教科書の日本語の概念イメージの速く、正確な認識が期待できる。

4.1. 補填箇所と材料

補填は、テキスト情報も追加するが、画像によって行うことを主としたい。しかし、いくつか考慮すべき問題がある。

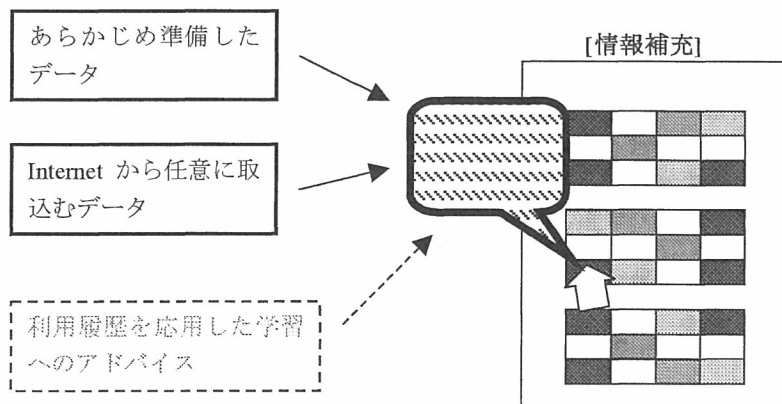
(1) 補填の単位の問題

個々の単語の意味が理解できるからといって、即、文、文章単位での理解につながるものではない。より大きな単位のイメージ補足情報も必要である。

しかし、一口に、文単位、文章単位と言っても、何をもって文とするか、一まとまりとするかの判断が難しい。また、常に、個々の要素を確認できるようにしておきたい。

(2) 量の問題

例えば、「もの」「こと」「空気」「でんぷん」といった抽象的な概念、または、学習活動の結果を抽象化されることでしかとらえられない概念を表す名詞がある。これらは、典型例として表示するイメージが1つに絞りきれない。抽象概念は、1つの画像



[図2:補充情報の種類]

だけから把握しづらく、同様のものをいくつも見比べ、把握する必要がある。したがって、画像は1枚に限ることができない。

(3) 質の問題

抽象概念を表すことばにイメージ画像を添付することで、逆に、物の見方を画一化してしまい、誤解を生じさせる恐れがある。

また、イメージ画像を参考程度に利用するか、確実な理解のために情報を利用するかにより、必要な画像の見せ方や性質が異なる。

(4) 補填箇所の問題

助詞、助動詞などの機能語については、画像イメージを提示するよりも、実際の使用例をテキストや音声で表示する方が状況を認識しやすい。

(5) 補填手段の区別の必要性

事物の捉え方により、視点や焦点を当てる部分が異なる。特に、抽象概念の場合は、

焦点を当てる軌跡を動画で示し、視線の推移を積極的に見せる方がよい。

また、感覚や情意、程度を表す語（形容詞や副詞）、助詞、助動詞などの機能語についても、動画を用いることや、実際の使用例をテキストや音声で表示することの方がイメージや知識が正確になる。

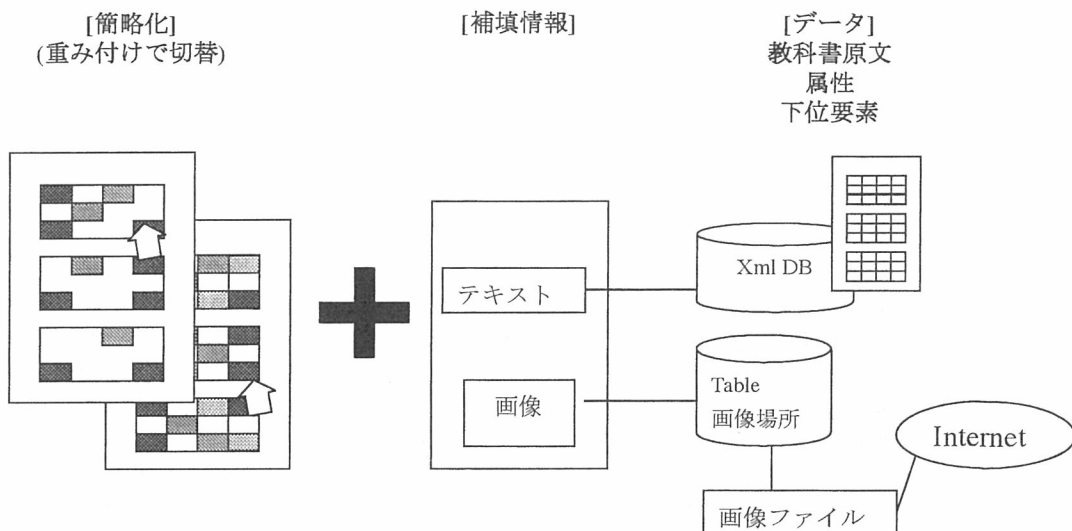
以上を考慮し、写真や絵などの画像イメージは、基本的に、普通名詞の単語に1枚、形容詞、副詞、動詞といった単語には、動画や複数の画像イメージを検討し、文、文章については、いくつかの画像、または、動画を1セット、提示する。

文は「。」で括られるものを便宜上1つの文とし、階層ごとに一まとまりの文章であると捉える。追加情報は最小の単位である単語から文、階層へと範囲を拡大して準備する。

各単位毎に選べるよう準備された情報は、質的な問題に考慮すれば、3つに分けられる。

	データ種類	内容	状態	保管場所
教科書テキスト	Xml テキスト	漢字仮名まじり文	構文解析処理後分割	Xml DB
	属性	品詞情報/読み方	タグに属性として組込	Xml tag attribute
		構文情報	タグに属性として組込	Xml tag attribute
画像	ファイル指定ポインタ	タグに属性として組込	Xml tag attribute	
追加情報	テキストレベルの情報	単語の読み方	Xml テキストの下位要素	Xml テキスト検索の結果
		ローマ字での読み方	Xml テキストの下位要素	Xml テキスト検索の結果
		意味	Xml テキストの下位要素	Xml テキスト検索の結果
		教科書内での他の使用例	Xml リソース検索の結果	なし
		教科書内で同じ意図で利用されているものをリストアップした簡易ソース	Xml リソース検索の結果	なし
		教科書外の一般的な使用例	検索エンジンによる検索結果 (画像の場所情報情報)	Internet
	画像(動画)情報	選定情報	画像の場所情報	Internet
		教材準拠情報	画像の場所情報	Internet
		一般情報	検索エンジンによる検索結果 (画像の場所情報情報)	Internet

[表 1:データの状態と場所]



[図3:補填情報と表示の関係]

すなわち、①こちらで妥当性を検討した結果、個別の単語のイメージに準拠するものを選択し、独自に作成したもの、②教科書に準拠した教育用ホームページのいくつかを登録しておき、そこから取り込むもの、③全く任意にインターネットから自動検索で取り込むもの、の3つである。

このうち、②③は、インターネットから取り込む情報であるが、これらは、関連する単語とは必ずしも1対1の対応関係にない。しかし、テーマに共通点がみられるものを探すように検索条件を文中の単語から自動的に指定されるようにする。また、この他に「ごみ」処理問題を検討する必要がある。次節でこの手順を考える。

4.2. 補填方法とデータの扱い方

補填情報は、テキストレベルの情報と、画像や音声、動画である。

テキストレベルの情報、画像や動画は表1にあげたものをそれぞれ準備しておく。「選定情報」とは、手作業で選定した画像である。

「教材準拠情報」とは、教科書に準拠して作成されたホームページから引用する画像である。「一般情報」は検索エンジンで探してくる画像である。

「選定情報」のうち、単語レベルで付与される情報は、品詞の別により、その種類が異なるものの、基本的にこちらで集めた最適なもの

のである。教科書の内容からは外れないようにするが、基本的に単語自体に備わった概念や意味を示すものであり、一般的な日本語教育用教材を引用したものが多くなっている。

以上のような情報の性質とオリジナルのデータとの関係も概略を表1にまとめ、その関係を図3に示しているが、詳細は以下の通りである。

(1) テキストレベルの情報

オリジナルデータである教科書データは構文解析処理を行った結果や、その他、補填に必要な情報を構造化の際に盛り込んでいる。そして、そのデータは過去にある地域で4年間採用された5教科6学年分の教科書データであるが、Xmlで構造化された状態でXml DBに格納している。データの性質が1フィールドに複数の情報が入る構造であること、また、1つのキーでマルチバリューが検索可能であることの2点が、RDBを利用せず、Xml DBを利用する理由である。

(2) 教材準拠情報

あらかじめ評価しておいたホームページの画像やテキストを引用するものであるため、オリジナルの文の単位と関連づけて数種類ずつ用意する。

URLをファイル化し、xmlには、そのIDを属性としてタグに埋め込む。

(3) 一般情報

(1)、(2)の情報に飽き足らず、ほかの関

連情報も見分きたいという場合に行う検索である。検索エンジンで探す、条件指定は、関連させたい文中の単語のうち、述語の必須項を自動的に指定する。また、検索の結果から URL を抜き、ファイル化し、こちらのシステムに、そのページの画像を順に表示するようにする。

しかし、ヒットした結果は、すべて表示するわけにはいかない。特に、青少年に対して有害だと考えられるものは、幾重にもフィルタリングを行い、除外する必要がある。ただし、無関係だと考えられるもの、概念が捉えにくいものについては、今回は利用者にその選択をゆだねるものとしておく。

5. 表示の方法

これまで、読解支援を目指し、前景化という方法で簡略化の1つの手段を提案してきたが、今回、前景化する部分を見直し、その方法を変更した。

オリジナルテキストの視覚的可変性は、これまでの考え方では、名詞、動詞、文末表現のうち、学習指示を意図する10数種類の表

現に限っていた[12]。

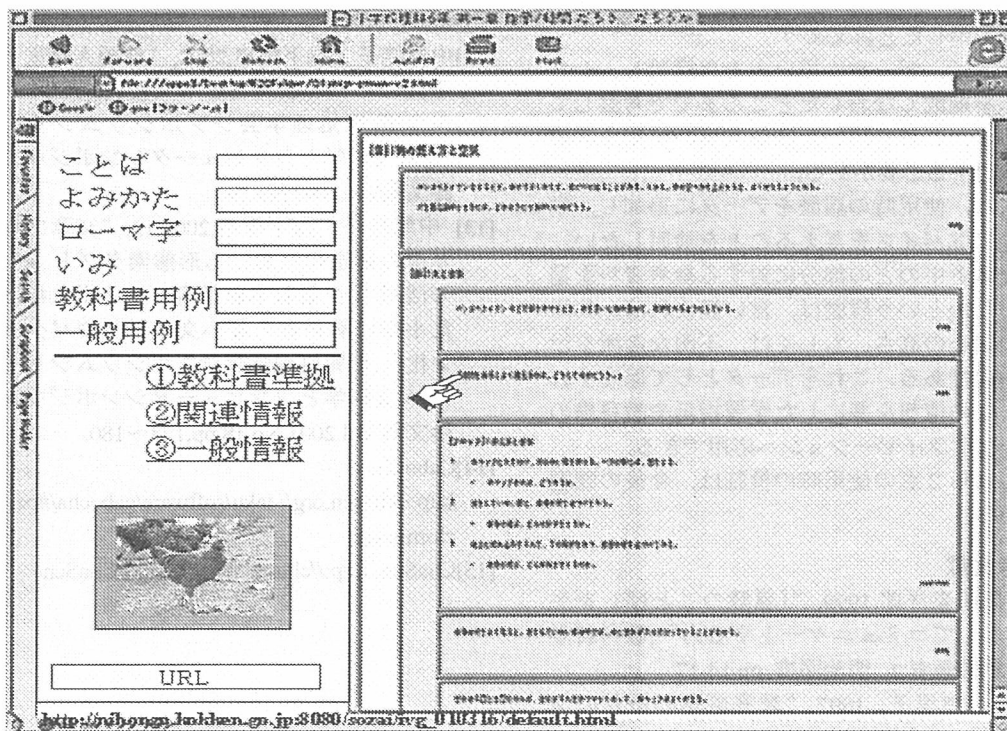
しかし、これでは、前景化する情報がすべて同列に扱われることになるため、実質的な意味のある単語や文字列を認識する練習にはなるが、結局は実質語を文頭から拾い読みすることと同じである。

そこで、より重要なものだけを前景化するように変更し、前景化する単語に重みを付ける。

単語の重み付けは、CaboCha[14]による構文解析結果を利用し、述語の必須項であるか否かにより決定する。

今までは、ChaSen[15]で形態素解析した情報をテキストの属性として各単語の下位に埋め込んでいたが、テキストをCaboChaで解析しなおし、再度オリジナルに埋め込みなおしたものを用意する。このデータを用いることにより、前景化する部分をより細かく選別することが可能になる。

図5は実際の表示で、階層別に区分けされたテキストと教科書の画像を表示したページにフレームをつけて、指定した単語に関する情報が確かめられるようにした状態である。



[図 5:教科書 Viewerでの表示例]

文や段落については、その範囲を示すマークを表示し、それを選択することでセットの一連の画像や用例テキストが表示されるようにした。

6. 結果の考察と今後の課題

本稿では、Non-Native 児童に対して行われる教科学習のための日本語教育における読解支援を目指し、先の読解支援システムを再検討し、変更を加えた。

今回の再考の結果に基づき、変更を加えたことにより、オリジナルのテキストと変化後の状態とを比較しながら、簡略化と補填が同時にできるものとなった。

テキストに追加した外部情報は、いずれの情報も、教科書のオリジナルテキストをより豊かにするものであるが、複数の異なる性質の情報である。

しかし、XML を用いてデータを扱っているため、追加部分とオリジナルとを明確に分割して保有している。したがって、自在に元のテキストにもどすことができ、また、変化の結果を別のものに応用しやすい。テキストや画像といったマルチメディア情報を扱う上での Mark-up Language の教育への応用方法の一例を示したと言えよう。

今回は、データの扱い方を再検討し、システムを編成しなおしたところまでで考察しているが、これを用いた読解支援の使用を実際に検討する必要がある。

また、使用時の履歴をデータに追加し、学習のアドバイスを与えることを検討したい。テキスト中のどの部分に対する参考資料を追加したかという履歴は、言い換えれば、利用者の関心の高さ、もしくは、不明な箇所を示すものである。これをデータとして保有すれば、この情報を基にした学習指示や教科書のパーソナライゼーションへ応用できる。

これら2点の使用時の検証は、今後の課題である。

参考文献

- [1] 黒崎東洋郎,1999,“「算数のことば」を生かしてコミュニケーションする力”,教科科学算数教育2, 明治図書, pp.14-17.
- [2] 石井恵里子, 1997,“教室談話の複数の文脈”,日本語学, 明治書院.
- [3] 岡本能里子,1997,“「教室談話における文体

シフトの指標的機能-丁寧体と普通体の使い分け-”,日本語学, 明治書院.

- [4] 松井嘉和,“読解教育”94号,『日本語教育』,1997.10,日本語教育学会.
- [5] 山田みな子,1995.7,“読解過程に見られる既有知識の影響と文法能力の関係について”,日本語教育86号,日本語教育学会.
- [6] 小林葉月,1999,“小学校における語彙教育と単語指導”,国文学解釈と鑑賞第64巻1号,至文社.
- [7] 岩沢正子・高石久美子,1994,“『算数』の教科学習を助ける日本語テキスト試案”,日本語教育83号,pp.73-84.
- [8] 池上摩希子,1998,“教科に結びつく初期日本語指導の試み-教材『文型算数』を用いた実践例報告-”,日本語教育97号,日本語教育学会,pp.118-129.
- [9] 中尾桂子・森下淳也,2003.5.30,“小学校教科書の学習者に応じた電子的再構造化について”,第58回人文科学とコンピュータ研究発表会.
- [10] XML: <http://www.w3c.org/XML/>
- [11] XSLT:
<http://www.alphaworks.ibm.com/tech/LotusXSL/>
- [12] 中尾桂子・森下淳也,2002,“外国人児童・生徒のための教科書読解支援Viewシステム”,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論文集Vol.2002,No.19.
- [13] 中尾桂子・森下淳也,2001.12,“文章ブロックの構造化における形態素タグとXMの活用-年少者日本語教育への応用にむけた小学校教科書の最小文章単位のパターン化”,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論文集 Vol.2001,No.18,pp.173~180.
- [14] CaboCha:
<http://chasen.org/~taku/software/cabocho/#performance>
- [15] ChaSen: <http://chasen.naist.jp/hiki/ChaSen/>