

アノテーションに基づく知的文書変換

東中 竜一郎

慶應義塾大学

政策・メディア研究科

長尾 確

日本アイ・ビー・エム (株)

東京基礎研究所

概要

アノテーションとは、コンテンツに対するメタコンテンツであり、XML形式のデータとして表現される。本稿ではアノテーションを持つテキストとアノテーションを持つ辞書を利用し、テキスト内の難解な言葉をわかりやすく提示する方法について述べる。具体的には、辞書の定義文のポップアップ表示や原文への挿入、原文の言い換えを提案し、その実現方法を説明する。特に、言い換えは多くの言語的操作を必要とする。本稿ではこの問題を比較的一般的な手法で扱う。

Intelligent Document Modification Based on External Annotations

Ryuichiro Higashinaka

Keio University

5322 Endou, Fujisawa,

Kanagawa 252-8520, Japan

rh@sfc.keio.ac.jp

Katashi Nagao

IBM Research, Tokyo Research Lab.

1623-14 Shimotsuruma, Yamato,

Kanagawa 242-8502, Japan

nagao@trl.ibm.co.jp

Abstract

The annotation is defined as the meta-level content of the original content, and is represented in XML-formatted data. In this paper, we propose the way in which difficult words are presented in more user-understandable form using annotated texts and annotated dictionaries. For its realization, we propose these methods: pop up and insertion of dictionary definitions and paraphrasing of the original text. Paraphrasing, among others, needs a great deal of linguistic knowledge and manipulations. We deal with this problem by using comparatively general methods.

1 はじめに

Web上にさまざまな文書が存在するがそれらは多様な内容と形式を持つために文書によっては背景知識を持たない人にとって分かりにくかったり、意味が取れなかったりすることがよく起こる。例えば専門用語であったり、一部でだけ通用するような単語を含む文書であると他の分野に携わっている人にとっては非常に難解なものになってしまうのは想像に難くない。

ここではWebの発達によって様々な文書にアクセスできる状況になったとはいえ、本当に文書が共有されているとは言えない。誰にとってもある程度理解可能な文書を提供することが可能になればWebの価値はより高まるのではないかと考える。難解な語を含む文書に出会った場合、人は通例辞書を引いたり、知人にその意味を尋ねたりして解決する。しかし、時間的制約などを考えると分からない単語に出会う度にそのようなことを行うのは効率が悪い。できることなら機械化できる場所は機械化したいと考えるのは当然のことである。

翻訳ソフトなどでよくマウス辞書が使用される。わからない単語の上にマウスポインタを置くとその単語の意味がポップアップして表示されるというものが、単語単位で区切りのある英語と違い日本語では単語の境界がどこかなのか分かりにくく実現は難しい。その上、従来のマウス辞書では語義が解決できておらず、複数語義のある単語を調べた場合、すべての語義を表示してしまう。これではどの意味にその単語が使われているのかが明快ではない。

では、ある単語についての語義を取得できた場合、次のステップはそのプレゼンテーションである。上に挙げた定義文のポップアップはその一つであるが、他に「挿入」と「言い換え」が考えられる。挿入は定義文を付加情報として文中に埋め込む方法で、言い換えは原文の表現を辞書に基づいて変換する方法である。言い換えは意味的な情報が必要となり一般に非常に困難である。

本稿ではこれらへの解決策として、アノテーションと呼ばれる付加情報を利用して上記問題を解決するシステムを提案する。2章でWebにおけるこのシステムの位置づけとアノテーションについて解説し、3章でアノテーションを持つ辞書を利用した定義文のポップアップや挿入、言い換えについて説明する。4章では言い換えを取り上げ、その手法を検討し、5章で具体的な言い換えルールを提示する。6章では実際に作成した言

い換えシステムの概要と実験結果を示し、7章で考察、8章でまとめを述べる。

2 アノテーション

アノテーションは、コンテンツに対するメタコンテンツであり、XML (Extensible Markup Language) 形式のデータとして表現される [11]。もちろん、アノテーションに対するアノテーションも考えられるが、これはまだ実現されていない。われわれは任意の HTML ドキュメントの任意のエレメントにアノテーションの XML データを関連付ける仕組みを開発した [1]。具体的には、XPointer と呼ばれる手法を用いて HTML エレメントを指定し、アノテーションデータのファイル名を関連付けるテーブルを用意する。

2.1 アノテーションの作成と管理

アノテーションの作成と管理のために、われわれはアノテーションエディターというオーサリングツールと、アノテーションサーバーを開発した。

アノテーション環境は図1のようにになっている。

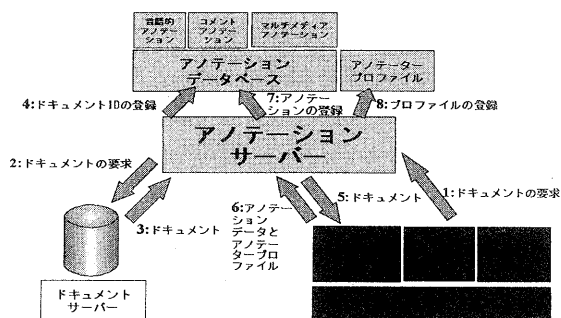


図 1: アノテーション環境

たとえば、HTML ファイルの場合、次のようなプロセスでアノテーションが作成され、管理される。

1. アノテーターと呼ばれるユーザーはエディターを起動して、対象となるドキュメントの URL を入力する。
2. アノテーションサーバーはエディターから URL を受け取ると、Web サーバーに問い合わせる。
3. アノテーションサーバーは Web サーバーからドキュメントを受け取る。

4. アノテーションサーバーはドキュメントのハッシュ値を計算すると、その値と URL をデータベースに登録する。
5. サーバーは、ドキュメントをエディターに送る。
6. ユーザーはエディターを使ってアノテーションを作成すると、それを自分のプロフィール情報(名前と専門分野など)と共にアノテーションサーバーに送信する。
7. サーバーは受け取ったアノテーションデータを URL と関連付けてデータベースに登録する。
8. サーバーは同時にアノテーターのプロファイル情報を更新する。

2.2 アノテーションエディター

アノテーションエディターは Java アプリケーションとして実装されており、アノテーションサーバーと通信できるようになっている。

アノテーションエディターは以下の機能を持っている。

1. URL を用いてアノテーションの対象となるドキュメントをサーバーに登録する。
2. Web ブラウザーと連動して、ドキュメントの任意のエレメントを選択できる。
3. XML 形式のアノテーションデータを生成し、サーバーに伝達する。
4. コンテンツが更新されたときに、以前に作成したアノテーションを再利用できる。

エディター(図2)の左側のウィンドウは、HTML ファイルの内部構造を示している。Web ブラウザー上で任意の HTML エレメントを選択すると、その部分のテキストがエディターに渡され、エディターの右側のウィンドウに表示される。選択された部分には自動的に XPointer と呼ばれるエレメントの ID が付与される。

このエディターを用いて、ユーザーは言語構造(構文や意味に関する構造)をテキストに関連付けたり、ドキュメント内の任意のエレメントにコメントを付けたりすることができる。

テキストの言語構造に関するアノテーションを言語的アノテーションと呼ぶ。言語構造は、まず自動的に生成されるが、その構造に曖昧さが含まれる場合は、それをインタラクティブに解消することができる。言語構造を修正するために、自動的に解析された構造をわかりやすく表示するための工夫を行なっている。

言語的アノテーションは図2の右に表示されている

画面上の操作によって容易に修正できる。

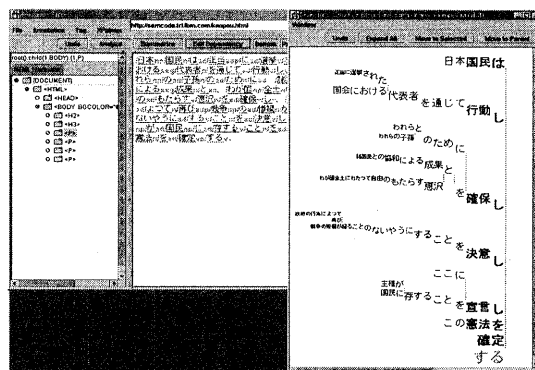


図 2: アノテーションエディターの画面

2.3 言語的アノテーション

言語的アノテーションは、ドキュメント内テキストエレメント(<H>, <P>, , , <DL>など)の文章の意味構造に関するアノテーションである。それは、語間の係り受け、代名詞の指示対象、多義語の意味など、かなり細かい情報を含む。このタイプのアノテーションは、ドキュメントの内容理解に大きく貢献し、文書変換以外にも、たとえば、内容検索や知識発見などに利用される。

言語的アノテーションは、具体的には XML 形式のタグファイルである。タグセットには、電総研の橋田らの提唱する GDA (Global Document Annotation) [10] のものを用いている。GDA は多言語間に共通な意味的・語用論的タグをドキュメントに付与することにより、その機械的な内容理解を可能にし、ドキュメントの検索・要約・翻訳を実用的なレベルで実現するとともに、ドキュメントの作成・公開(共有化)・再利用を考慮した統合的なプラットフォームを構築して、世界的に普及させようという、壮大なプロジェクトである。われわれのプロジェクトは GDA を現在の Web のアーキテクチャ上で利用可能にし、さまざまなサービスと連動させることによって、GDA の思想をより具体的な形で浸透させようとする試みの一つと位置付けられる。

一般に、GDA ドキュメントはネットワーク構造を成しており、そのリンクには、タグの入れ子構造によって定義される関係と参照関係の 2 種類がある。また、GDA のタグ集合は 10 項目以上からなるが、さしあたり、そのうちで自動タグ付け作業が比較的大変だと思われる、

統語構造、文法・意味関係、語義、照応、修辞関係という5項目だけを扱っている。GDA タグセットの詳細については、<http://www.etl.go.jp/etl/nl/gda/>を参照のこと。

文法機能(主語、目的語、間接目的語)、主題役割(動作主、被動作者、受益者など)、および修辞関係(理由、結果など)は関係属性によって表示する。関係属性は`rel="**"`という形で表される。主語、目的語、および間接目的語の主題役割の判断は難しいことが多いので、文法機能(sbj、obj、iob)を用いる。

このようなタグ付けは多くの労力を要すると思われるが、アノテーションエディターにいくつかの自然言語処理モジュール(統語・意味解析、照応解析など)を統合することによって、極力人間の負担を減らせるように工夫している。人間がインタラクティブに解析した部分は、事例として次の機会に再利用されるので、それによって解析の精度が少しずつ上がっていくことになる。解析の精度が上れば、それだけ人間の負担が減ると思われるので、将来的にはタグ付けのコストは十分に少なくなるだろう。

3 文書変換

アノテーションに基づく文書変換には、単語辞書を用いた定義文のポップアップ、原文への挿入、定義文に基づく原文の言い換えがある。

ここでは、機械的に言い換え等を実現するためにはどのような辞書が必要かを考察する。また、単語に対する定義文を発見した場合、その情報をどのようにユーザに提示するかは、非常に重要な課題である。

3.1 辞書の作成

一般的に国語辞典は以下のようなフォーマットで作られている。

(見出し語
(定義文① 用例)
(定義文② 用例)
(定義文③ 用例) ...
反義語
類義語
他の見出し語へのポインタ
その他注釈)

上で示すように一つの見出し語に対して定義文が複数個あることは多く、実際の例¹を示せば以下のようになるものである。

アイス	イギリス
	1 氷 アイススケート
	2 アイスコーヒー、アイスティー
	3 「アイスクリーム」の略

以上のように、機械的に辞書を引く場合見出し語だけでは、どの意味で使われているのかという問題が解消しないので単語に対しその語義をアノテーションとして付加することにする。辞書は語義と定義文の対で表わされており、語義として EDR 日本語単語辞書 [8] の概念識別子、定義文としては同辞書の定義文を用い作成した。

語義アノテーションの例
<code><np id="id0-0-1" sense="3bcdec"</code> 読み="ニッポン" 品詞="名詞-固有名詞-地域-国"> 日本</np> (3bcdec が概念識別子)
辞書 ² の例
概念識別子 定義文; 辞書見出し [読み]; ... 3bcdec 日本という国。; ジャパン [じゃぱん]; 日 [にち]; 日本 [にっぽん][にほん]; 日本国 [にほんこく];

3.2 プレゼンテーション

ある単語に対して語義が決定された場合、その定義文をユーザにどのように提示するかということは大きな問題である。プレゼンテーションの仕方によって最終的な理解を大きく左右するからである。本稿ではポップアップ、挿入、言い換えの手法を提案し、実装した。

3.2.1 ポップアップ

アノテーションによって文字列に境界を与えることが可能になるため、英語におけるマウス辞書のようにマウスポインタが指している単語を認識することが実現できる。指された単語に語義属性が含まれている場合はその語義をもとに辞書を引きその定義文をポップアップする。原文テキストは変化しないため、元のレイアウトを保持できる。定義文を参照できる単語に関

¹ 三省堂小学国語辞典 [9] による。

² ただし、辞書を用いた言い換えを行うときは定義文に言語的アノテーションを付ける。

してはテキストの背景色をわずかに変え、それとわかるようにした。分からない単語だけにユーザが問い合わせをするという形なので、ユーザにとって効率がよい。図3は画面例である。

日本国民は恒久の平和を企願し、人間相互の関係を支配する闘争いや心配事なく、るのであつて、平和を企図すること。信義に信頼して、われらの安全と生存を保持しようと決意した。われらは平和を維持し、専制と隷従、圧迫と偏狭を地上から永遠に除去しようと努めてある国際社会において名誉ある地位を占めたいと思ふ。われらは全世界の国民が、ひとしく恐怖と欠乏から免かれ、平和のうちに生存する権利を有することを確認する。

図 3: 語義のポップアップ

3.2.2 挿入

システムは語義の付与された単語を見つけるとその直後に括弧付きでその語義から引いた定義文を挿入する。この手法は一般的に出版物で多くなされる方法である。定義文が参照できる単語すべてについて括弧付きで説明を付け加えるため、ユーザにとって既知である単語も説明してしまう可能性がある。図4は画面例である。

日本(日本という国)国民(国家を構成する人々)は恒久(永久)の平和(争いや心配事なく、穏やかであること)を企願(念願)するし、人間(人間相互)関係(あつたのそれぞれの関係)人や物事の関わり合いを支配(自分の勢力下におき、治めること)する崇高(ただかくすぐれているさま)な理想(最も完全なものとして人が心に思いえがく姿)を深く自覚(物事を認識する)するのであつて、平和を愛する(愛してたいせつにする)諸国民の公正(公明正大であること)と信義(主人や国家に誠意をもって務めを果たすこと)に信頼(信頼)

図 4: 語義の挿入

3.2.3 言い換え

語義の付与された単語を定義文にそのまま置き換える(図5)。しかし、そのまま置き換えたのでは文との親和性がない。その点については次章で述べる。

日本という国国家を構成する人々(は)永久の争いや心配事なく、穏やかであることを企願するし、人間関係あつたのそれぞれの人や物事の関わり合いを物事を自分の勢力下におき、治めること(は)ただかくすぐれているさま(最も完全なものとして)人が心に思いえがく姿を深く物事を認識するするのであつて、平和を愛してたいせつにする諸国民の公明正大であることと主人や国家に誠意をもって務めを果たすことに信頼して拠り所とするして、われらの病氣や事故、危ないことがなく、きわめて安全なことと生きて存在するを保ち

図 5: 語義の言い換え

4 言い換えについて

ここでは前章のうち、言い換えを取り上げる。前の章でも述べたように単語をその定義文にそのまま置き換えたのでは全く文との親和性がない。これは当然で、辞書の定義文はその単語の説明であつて類義語などの「等価な表現」ではないからである。

ゆえに定義文を原文に埋め込む際、定義文を適切な形に変換する必要性が生じる。これは「言い換え」の一種であり、過去にいくつか研究がなされている。佐藤らは、複合名詞の言い換え[3]や、サ変名詞の言い換え[2]、格変換による言い換え[4]を提案し実装している。彼らは言い換えを以下の3つのクラスに分類している。

1. 構文的言い換え

言葉に関する知識によって実現可能な言い換え。単語を同義語や類義語に置き換える言い換えや、構造のマッピングに基づく言い換えをこのクラスに分類する。

2. 意味的言い換え

参照表現などを、それが指す内容で置き換える言い換えがこのクラスに含まれる。また、省略されているものを意味的に補う言い換えもこのクラスに含める。

3. 語用論的言い換え

(1)、(2)以外のより複雑な言い換え。ある状況において同じ効果を持つような文に言い換えるものがこれに含まれる。

本稿での言い換えはアノテーションを利用することにより以上のクラスの(1)と(2)を実現する。これはアノテーションによって指示詞が何を指すかというようなものを記述できることによる。

われわれはアノテーションを持つテキストとアノテーションを持つ辞書を利用することによって、より一般的な言い換えを実現するルールを作成を目指している。

今回試作したルールは日本語の文法依存であるが、適応するルールを変更することにより、特定の文法理論に依存しない言い換えのシステムが実現できると考える。

5 言い換えルール

言い換えは2つのエレメントの間で行われる。言い換えられる対象である原文のエレメントと、言い換える辞書の定義文全体を指すエレメントである。なお、使用するタグはJUMAN[7]とKNPによって得られた結果を手で係り受けを修正しGDAタグ化したものであり、品詞体系は茶筌[6]に合わせた。

言い換えられるエレメント
<np id="id0-0-9" sense="3bc9de" 読み="コッカイ" 品詞="名詞-一般">国会</np>
言い換えるエレメント
<su sense="3bc9de" id="id4-0"><vp><adp>
<np id="id4-0-1">立法権</np>
<ad id="id4-0-2" 読み="ヲ" 品詞="助詞-格助詞-一般">
を</ad></adp><v id="id4-0-3" 読み="モツ" 基本形="もつ" 品詞="動詞-自立" 活用="五段・タ行基本形">もつ</v></vp>
<np id="id4-0-4" 読み="コッカ" 品詞="名詞-一般">国家</np><n id="id4-0-5" 読み="キカン" 品詞="名詞-一般">機関</n></su>

具体的には原文テキストの各アノテーションのエレメントが語義属性を持っているとき、辞書引きを行いアノテーションの付いた定義文エレメントを取得しそれら2つのエレメントについて変換ルールを適応する。ルールには2種類考えられる。1つはどのようなエレメントの間の言い換えでも適用されるグローバルルールと、エレメントのタグ、属性によって適用するルールが違ふローカルルールである。

以下に今回試作した言い換えるルールを示す。なお、EとはEntryのことで置き換えられる対象のエレメントを指しDはDefinitionのことで定義文のエレメントを指す。

5.1 グローバルルール

1. DにEが含まれている場合は言い換えない

(理由) 語の説明に、説明する語が含まれてい

る場合明らかにより簡単な表現になっていないため。また、辞書としてそういった記述は不自然なため。

(例) 日本—日本という国。

2. 括弧付きで補足されているテキストエレメントは言い換えない。

(理由) 定義文に括弧付きで説明されている部分は補足的であり、そういった情報はすでに原文において説明されていることが多いため。

(例) 信託する—(責任や任務を)信用して委託する

3. 言い換え後2重否定になるようなものは言い換えない。

(理由) 置き換え後、原文よりも複雑な構造になると考えられるため。

(例) 無視してはならない—[無視 注意を向けない]→(×注意を向けないではない)

4. 基本単語の範疇ならば言い換えない。

(理由) 言い換えると逆に難しくなってしまうような非常に基本的な語はより分かりやすくする必要が無く言い換える必要性がないため。なお、それが基本単語でどれがそうでないかは三省堂小学国語辞典の基本語であるかどうかということを目指した。この辞書では「特に小学生にとっていいことば、これだけは意味も使い方も正しく知っていなければならないことば」というものを基本語として印を付けており、それをもとに基本語辞書を作成して利用している。

(例) 国家—一定の領土と住民からなり独立した統治機構を有する社会的集団の政府(これは言い換えない)

5. 言い換えのときに、Dの前に読点をつける。

(理由) 言い換えた結果、だらだらと言葉が続いたり、係り受けがあいまいになることがあり、それを防ぐため。

(例) ここに主権が国民に—ここに、一国を統治する最高唯一の権力が国民に

6. Eにかかっている格助詞で、Dにかかっているものと同じ物があればその格助詞に係るものは削除する。

(理由) 定義文ではどういった文脈でその語が用いられているかを示すために「～などを」とか「～などが」のような格助詞句を持つ場合が多い。しかし、そのような句は原文の中にすでに存在する場合が多く言い換え後冗長になったり、意味が

わかりにくくなるため。

(例) 圧迫と偏狭を除去する一不要なものを取り除く [不要なもの] は言い換えに使わない。(× 圧迫と偏狭を、不要なものを取り除く)

7. D でシチュエーションの例示を行う句があればその部分は言い換えない。

(理由) 6 と似ているが、語の使われる文脈を示すために「において」や「において」などが使われることがある。これらも原文中にすでに自明である場合が多いため言い換えに使用しない方がよい。

(例) [成果一行為などの結果として生じたよい事柄] のうち「行為などの結果として」は言い換えに使わない。

5.2 ローカルルール

ルールに使われる記号は以下の意味で使用される。

N ::= n(名詞) np(名詞句)
V ::= v(動詞) vp(動詞句)
AD ::= ad(副詞) 形容動詞・名詞 adp(副詞句) 形容動詞・名詞句
AJ ::= aj(形容詞) ajp(形容詞句)
その他 助詞は ad
助詞を主辞に持つ句は adp

ルール表に示される N-N などは置きかえられるエレメント (E) が N で置きかえるエレメント (D) が N であるというような置き換えの対を示す。

N-N

(1) そのまま置き換える³。

(例) 惨禍 → わざわい、恵沢 → 恵み

N-V

(1) D を連体化し、「こと」をつけたもので置き換える。

(例) 協和 (心を合わせ仲良くする) → 心を合わせ仲良くすること

V-N

(1) D に係る vp があればその vp を E の活用で置き換える。

³ 複合名詞の場合は置き換えてしまうとつながりが悪くなり意味が取りづらくなる。その為に辞書引きによる置き換えは行わず、名詞間のつながりが分かるように助詞を挿入することが有効である [3]。
(1) N1-N2 → N1 の N2 といいかえる。
日本国民 → 日本の国民

(2) N1-N2-N3 N2 がサ変名詞の場合、N1 を N2 する N3 と言い換える。

V-V

(1) E の活用を D の活用に適応する⁴。

(例) 確認する → 確かに認める。

(2) D の表層格が、E に係る ADP の格を持たなければ置き換えない⁵。

AD-N

(1) D に係っている AD を活用させ置き換える。

(2) D に係っている AD がない場合は置き換えない。

AJ-N

(1) D に係っている AJ を活用させ置き換える。

(2) D に係っている AJ がない場合は置き換えない。

6 実装

以上のルールを実行するシステムを実装した。語義属性の付いているエレメントを見つけると、辞書を引き、アノテーションの付与された定義文を取得する。その上でアノテーションの構造により係り受け構造を理解し、上記のルールを適用した。日本国憲法前文の一文についての結果を示す。

6.1 結果

(変換前)

日本国民は正当に選挙された国会における代表者を通じて行動し、われらとわれらの子孫のために、諸国民との協和による成果と、わが国全土にわたつて自由のもたらす恵沢を確保し、政府の行為によつて再び戦争の惨禍が起ることのないやうにすることを決意し、ここに主権が国民に存することを宣言し、この憲法を確定する。

(変換後)

日本国民は正当に選挙された、立法権をもつ国家機関における代表者を通じて、物事を行い、われらとわれらの、血筋を受けついでいく人々のために、諸国民との、心を合わせ仲良くすることによる、生じたよい事柄と、わが国全土にわたつて自由のもたらす、恵みを、

⁴ サ変名詞 + するの場合、ヴォイスやアスペクトなども考慮する必要がある [2]。

⁵ コロケーションの問題が生じるためである [2]。(例) 検討に着手する。→ 検討に始まっている。格レベルでの写像 [4] は今回実装していない。

しっかり保持し、政府の、動作や態度によつて再び、武器を使った戦いの、わざわざ、新たに生じることのないやうにすることを、心に決め、ここに、一国を統治する最高唯一の権力が国民に存在することを、はっきりと述べ、この、国の公機関が出す法令や命令を、確かに決める。

7 考察

ポップアップと挿入に関しては比較的分かりやすい文書に変換された。言い換え部に関しては、係り受けを明確にするために打った読点がかかなり多く挿入され、全体的に少し読みにくい感じを受けた。加えて、言い換えたことにより分かりやすくなった面もあるが、逆に冗長でわかりにくくなった面も多く見られる。やはり、辞書の定義文は見出し語とは等価ではなく、少し意味が広がってしまうからである。基本語以外の辞書引き可能な単語をすべて言い換えてしまうのも結果として意味が少し取りにくくなっている原因ではないかと考える。これは挿入にも言える。

8 まとめ

今回言い換えルールが比較的単純なため、場合によっては言い換え後のテキストが少し分かりにくくなるがあった。しかし、テキスト中の単語に語義属性を入れることによる機械的な辞書引きに関しては非常に有用だと感じ、専門用語等にアノテーションを付けることが一般化すれば Web のアクセシビリティは高まるだろう。

どの部分を言い換えてどの部分を言い換えてはいけないかというのは非常に難しい。文脈によって言い換えるかどうかを考慮したり、今回は言い換えの指標として三省堂小学国語辞典の基本語を用いたが、日本語語彙特性データベース [12] の利用も検討している。さらに、今後はプレファレンスを設定できるようにし、文書にアクセスする人にとっての分かりやすい言い換えを行うことを目指す。

最後に、アノテーションを付けることは現在比較的時間のかかる作業である。文脈情報を利用して、ある単語が辞書の中のどの語義として使われているか推測する機構をエディターに組み込むことを検討している。

謝辞

本研究に関して、有益な助言を頂いた慶應義塾大学の石崎俊教授に感謝します。

参考文献

- [1] Katashi Nagao et al. Semantic Transcoding: Making the World Wide Web more understandable and usable with external annotations. *TRL Research Report*. IBM Tokyo Research Laboratory, 2000.
- [2] 近藤恵子, 佐藤理史, 奥村学. 「サ変名詞+する」から動詞相当句への言い換え. *情報処理学会論文誌* vol.40 No.11, pp.4064-4074, 1999.
- [3] 佐藤理史, 論文表題を言い換える. *情報処理学会論文誌* vol.40 No.7, pp. 2937-2945, 1998.
- [4] 近藤恵子, 佐藤理史, 奥村学. 格変換による単文の言い換え. *情報処理学会研究報告* 00-NL-135 pp.119-126, 2000.
- [5] 益岡隆志, 田窪行則: 基礎日本語文法 (改訂版), くろしお出版 1992.
- [6] 奈良先端技術大学茶筌開発部 日本語形態素解析システム「茶筌」version 2.0 for Windows 1999. <http://cl.aist-nara.ac.jp/lab/nlt/chasen/>.
- [7] 松本裕治, 黒橋慎夫, 山地 治, 妙木 裕, 長尾 真. 日本語形態素解析システム JUMAN 使用説明書 version 3.3 1997. <http://www-nagao.kuee.kyoto-u.ac.jp/nl-resource/juman.html>.
- [8] 日本電子化辞書研究所. EDR 日本語単語辞書マニュアル 1996. http://www.iijnet.or.jp/edr/J_index.html.
- [9] 三省堂, 小学国語辞典 (第八版) 1989.
- [10] Koiti Hasida. Global Document Annotation. <http://www.etl.go.jp/etl/nl/gda/>.
- [11] Hiroshi Maruyama, Kent Tamura, and Naohiko Uramoto. XML and Java: Developing Web applications. Addison-Wesley, 1999.
- [12] 三省堂 日本語語彙特性データベース 1999. <http://www.brl.ntt.co.jp/cs/human/goi/index.html>