

EDR 日本語コーパス参照支援ツール

鈴木 しとな † 渡部 憲二 ††
上原 徹三 † 石川 知雄 †

† 武蔵工業大学工学部 †† 大日本印刷情報システム (株)

EDR 日本語コーパスは、品詞タグなどの諸タグと、係り受け関係を含む文構造の添付により、日本語処理にとって有用なデータとして利用価値が高いが、一方で、データの格納形式が複雑となり、その理解と取扱いが容易でないという問題点を持つ。そこでわれわれは、同形式のコーパスの作成、編集、参照、照合などの機能を提供するツール群を作成中である。この中の参照支援ツールは、C または C++ 言語による応用プログラムから呼び出せる関数群として試作した。本ツールでは、EDR 形式のコーパスの保持する品詞タグと係り受け情報とから、文節情報と文節間係り受け関係を求めて応用プログラムに提供することで、利用者がコーパス形式の詳細に煩わされずに、用語や文法的情報の検索・照合などの調査を行なうことを支援する。その機能としては、コーパスファイル設定機能、形態素・文節係り受け関係・複合語等の文法情報およびその他の一般情報の取り出し機能などがある。EDR 日本語コーパスを上記の目的で参照している既発表論文を調べ、多くの場合に本ツールの機能が適用できることを確認した。他のツールについても概要と今後の課題を記した。

An Access Tool for EDR Japanese Corpus

SHITONA SUZUKI, † KENJI WATABE, ††
TETSUZOU UEHARA † and TOMOO ISHIKAWA †

† Faculty of Engineering, Musashi Institute of Technology

†† Dai Nippon Printing Information Systems Co., Ltd

EDR Japanese Corpus is valuable for Japanese language processing because it contains grammatical data including part-of-speech tags, dependency relationships and syntax trees. But the Corpus has so complicated inside-structure that its handling is difficult for us. So we propose some tools for handling the Corpus records. Access Tool for EDR Japanese Corpus, one of such tools, consists of a set of functions called from application programs in C/C++. These functions are used to fetch words or "bunsetsu" and their grammatical information, such as part-of-speech tags or dependency relationships by traversing the syntax tree. By investigating recent papers utilizing EDR Japanese Corpus, the functions of our tool are found to be applicable to the usage in these papers.

1. はじめに

自然言語の研究のためにコンピュータで参照できる形の辞書や索引の作成が盛んになっている。その中には、実際に使用された文例を集めたコーパスがある。コーパスには、単に文例を多数集めたもの²⁾から、文例に対して品詞情報や係り受け関係を含む構文情報を付加したタグ付きコーパス（解析コーパスとも言う）などの種類がある。特に、タ

グ付きコーパス¹⁾³⁾⁶⁾は形態素解析や構文解析などの自然言語処理の研究において、文法規則の学習や検証のため、また、解析結果の評価のためなどの目的で広く利用されている。

日本語に関するタグ付きコーパスの大規模なものとしては、日本電子化辞書作成の EDR 日本語コーパスが広く用いられている¹⁾⁸⁾。このコーパスは約 22 万文の文例に対し、各単語の品詞、活用形などの形態素情報と、係り受け関係を含む構文木

情報を保持しており、上のような研究対象として最も広く用いられている日本語コーパスであるといえる。その各文例に対するデータ形式は、文を構成する単語に関する読み、表記、品詞、活用形などの形態素情報と、単語間の合成関係を単語を葉に持つ構文木のリストで表している構文情報、文の中の意味的な部分構造を示した意味情報からなる。

EDR 日本語コーパスは、これらの情報をコード化せずに通常の文字表現によって示しているので、人がそのまま読むことも一応は可能であるが、複雑な階層構造を含む表現の意味内容を解釈することは容易でない。文法規則の学習と検証および解析結果の評価のためにこの EDR 日本語コーパスを全体的に調査する際には、調査者は計算機プログラムを作成し、調査項目に応じたプログラム機能を実現する必要がある。調査者は、EDR 日本語コーパスのデータ形式を把握し、調査プログラムの中にそれを解釈する機能を実現しておく必要がある。しかし、上記のようにこの処理は煩雑である上に、調査者の本来の目的にとっては余計な処理である。

そこで、この EDR 日本語コーパスの内部構造を解釈する機能を、調査者のプログラムから必要に応じて呼び出せるツールを用意することとした。これが本稿で主として述べる EDR 日本語コーパス参照支援ツールである。

さて、EDR 日本語コーパスは約 22 万文の文例を含むが、上のような目的のためにはこの量でも十分とは言えない。別種の文例についても、EDR 日本語コーパスと同じデータ形式で収容するならば（以下、この形式のコーパスを EDR 形式日本語コーパスと記すこととする）、上の参照支援ツールが同様に使用できる。このような新たな EDR 形式日本語コーパスの作成のために、形態素情報を含む文節情報の列と文節間の係り受け関係を受け取って、EDR 形式日本語コーパスの文例レコードを作成する EDR 形式日本語コーパス作成支援ツールを試作した。

また、この文例を検索し編集する編集システム⁴⁾を試作した。さらに参照支援ツールの関数群を内部で利用する応用プログラムの一種として、文構造照合支援ツール⁵⁾を試作した。

以下、EDR 日本語コーパスの含む情報の概要、EDR 形式日本語コーパスに対して試作中のツールの概要と関連について述べた後、EDR 形式コーパス参照支援ツールの試作機能を記し、最後に、ツール群に関する今後の進め方を述べる。

2. EDR 日本語コーパスについて

EDR 日本語コーパスは約 22 万文の文例を収めた大規模コーパスで、文例自身の他に品詞、意味概念などの形態素単位の情報や、係り受け関係を示す構文情報などのタグ情報を保持しているという特徴を持つ。このコーパスの含む文例を図 1 に示す。図のような文例一つに対する各情報のまとまりを、以下ではレコードと呼ぶこととする。

2.1 EDR 日本語コーパスの内部構造¹⁾

EDR 日本語コーパスの 1 レコード分の情報は大きく分けて、文情報、構成要素情報、形態素情報、構文情報、意味情報、および、管理情報によって構成されている。これらのうち、主な情報の内容を以下に示す。

2.1.1 文情報

文情報は文例毎に与えられる情報であり、テキスト番号と出典情報と文の内容とを含む。テキスト番号は文の管理番号であり、その文例を含む記事の種類とその記事内での文例の出現順序を示す。出典情報は出典テキスト名を示し、文の内容は文例を文字列表記として示す。

2.1.2 構成要素情報

構成要素情報は、文例内の各々の形態素に対して付加される情報であり、構成要素番号、表記、かな表記、概念選択、品詞、概念識別子、補足付き概念説明、複合語形態素番号によって構成される。構成要素番号は形態素番号、または複合語形態素番号を示す。概念選択は概念の同一性を示す番号で、EDR 単語辞書内の単語を一意的に識別するキーとなる。また、補足付き概念説明とは、単語辞書に登録された概念に当てはまらない場合に記述される概念の説明である。

2.1.3 構文情報

構文情報は文例の構文的な構造を示すために付加される下記の種々の記号を含む構文情報であり、文例をその構成要素のリスト表現として表現し、一つの文例を下記のような合成構造に基づく構文木として表す。この構文木の葉は形態素であり、葉識別子 'W' を先頭に置いて記述されている。より下位の節点のまとめ方を示す合成関係として、修飾合成（合成関係子 'M'）、統合合成（合成関係子 'S'）、数合成（合成関係子 'N'）複合語合成（合成関係子 'I'）の 4 種類を用いる。各合成の例を図 2 に示す。

修飾合成は連体修飾と連用修飾を示している。統合合成には、従属合成と等位合成がある。従属合成は、概念を持つ自立語と概念を持たない付属語

JCO0000666	0005000f4aa	朝日新聞	870208	10万円	はそのまま	にしている。	
1	10万	10000	数字	00010186a0	2	円	
接尾語	3bcc4	3	は	ハ	助詞	2621d5	
名詞	0fb1d8	5	に	ニ	助詞	2621d5	
動詞	"=Z 維持する"	7	て	テ	助詞	2621d5	
イ	動詞	"	る	ル	語尾	2621d0	
記号	2621d8	9	/	1:10万/2:円/3:は/4:そのまま/5:に/6:し/7:て/8:い/9:る/10:。/	10	。	
(S(t(M(S(t(S(W 1 "10万"))(t(W 2 "円")))))(W 3 "は"))(t(M(S(t(W 4 "そのまま")))(W 5 "に")))(t(S(t(W 6 "し"))(W 7 "て"))(W 8 "い"))(W 9 "る")))))(W 10 "。"))							
Z	維持する	[[attribute state][object [[main 2:円:3bcc44][number 1:10万:00010186a0]]]	[manner 4:そのまま:0fb1d8]]				DATE="95/7/12"

図1 EDR 日本語コーパスのレコードの例

を合成するもので、具体的には語幹と語尾、名詞と助詞などがある。等位合成は、概念を持つ自立語が対等の資格で結びつき一つの体言が形成される場合に用いられる。複合語合成は複合語の構成語を合成する。数合成は、数字列を一まとまりに合成する。

修飾合成の例
1. 連体修飾
(M(W1"主要"))
((W2"財界人"))
統合合成の例
1. 語幹と語尾
(S(t(W1"咲"))
(W2"く"))
複合語合成の例
(W 16 "デモ行進")
(I1 "デモ行進")
(W 16 "デモ")
(W 17 "行進")
数合成の例
(N"-3" (W 1"-"))
(W 2 "3"))

図2 合成関係の例

3. EDR 形式日本語コーパス関連ツール群

EDR 形式日本語コーパスは、1レコードの情報が多く、自然言語処理において有用なデータを含む。しかし、前述の通り、そのデータの格納形式は複雑であり、利用者はその格納形式を理解した上で、コーパスレコードから、形態素のタグ情報を取り出す処理や、係り受け関係にある要素を取り出す処理を応用プログラム上で実現する手間を必要とする。この処理は、利用者本来の目的である自然言語処理とは別のものである。そこで、われわれは、利用者が本質的な処理を行なうだけで済むような、EDR 形式日本語コーパスを軸にした作成、編集、参照、照合などのツール群を試作し、また、作成中である。

EDR 形式日本語コーパス関連ツール群のうち、EDR 形式コーパス作成ツールは、文例を構成単語と品詞情報、および、係り先文節情報の列として

受け取り、EDR 形式の合成関係子による構文木を表すリストを作成する。さらに、文番号などの情報を付加し、文情報と構成要素情報を作成する。

ただし、このツールの作成する係り受け関係情報は、本来の EDR 日本語コーパスの単語間の係り受け関係とは厳密には相異なり、橋本文法⁹⁾による文節間の係り受け関係を表現するものとした。これは、本来の EDR 日本語コーパスの係り受け関係の表現法が厳密には規定されていないこと（ただし、その形式から文節係り受け関係を取り出すことは可能である）、および、われわれの目的からは文節単位の係り受け関係が必要であったこと（例えば、文献⁷⁾では、EDR 日本語コーパスを文節単位の係り受け表現のコーパスに変換して用いている）による。

このツールを用いて、EDR 形式コーパスを作成することにより、EDR 日本語コーパス参照支援ツールが利用できる。実際にわれわれは、伊勢物語と竹取物語の文例を品詞情報と係り受け情報を合せて EDR 形式古文コーパスとして作成し、参照支援ツールを使用して、古文の係り受けに関する調査を行なっている。

また、EDR 日本語コーパス編集機能についても試作中である⁴⁾。このツールは、EDR 形式日本語コーパス内の文例および付加情報の誤りの訂正、および、用例の追加と削除を主な目的としている。

さらに、本稿の主題である EDR 形式コーパス参照支援ツールの応用プログラムの一つとして、EDR 形式日本語コーパスを対象とする文構造照合支援ツール⁵⁾を試作した。EDR 日本語コーパスの調査を行なうための応用プログラムを調査者が作成することなく、非手続き型の文構造照合記述言語 SSDL で照合条件を記述することにより対話形式でのコーパス内の文例の該当部分を照合する機能を実現するものである。

以上の EDR 日本語コーパスに関するツール群とその関係図を図3に示す。

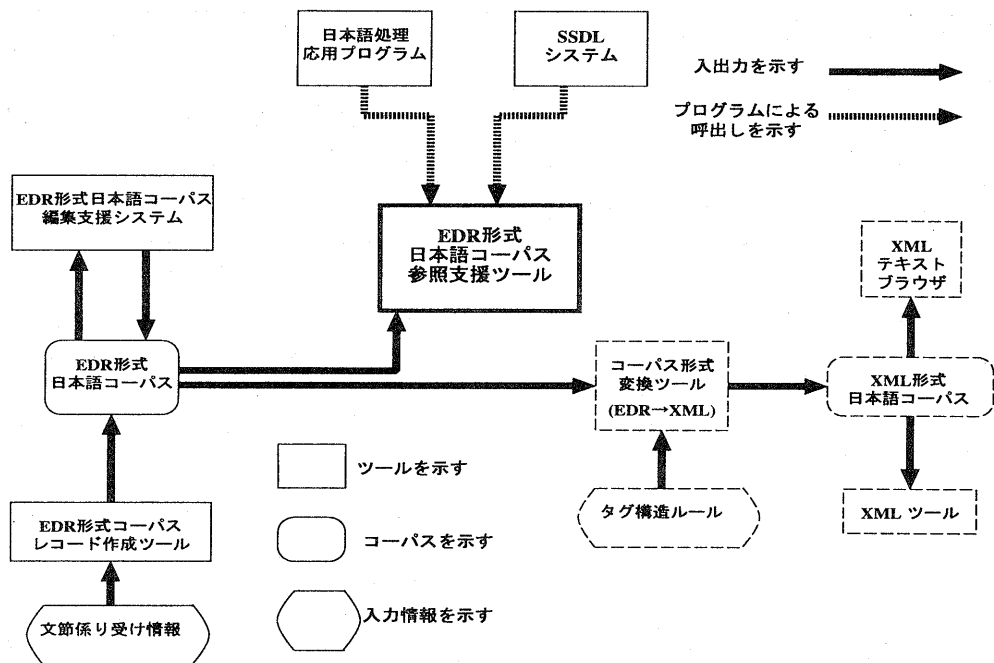


図3 ツール関係図

4. EDR 形式コーパス参照支援ツール

EDR 形式日本語コーパス参照支援ツールは、C 言語または C++ 言語で記述された応用プログラム内から呼び出される関数群の形で EDR 形式日本語コーパスの情報を提供する。

この章では、EDR 形式コーパス参照支援ツールに関して、ツールの機能を実現する関数群について示したあと、その適用性と効果について示す。

4.1 ツールの機能

提供する関数を機能ごとにグループにわけ、それぞれの機能を述べる。

4.1.1 ファイル設定用関数および設定用関数

ファイル設定用関数には、コーパスファイルの読み込み、インデックスファイルの読み込み、インデックスによる参照開始位置のセットなどを行なう関数がある。ここでのインデックスとはコーパスレコードを指定する文管理番号のことである。インデックスファイルは、インデックスファイル作成プログラムによって別途作成した。

- 関数 SetCorpusFile(char*)

引数 使用するコーパスの所在

戻り値 0(設定に成功) または -1(何らかの原因で失敗)

機能 コーパスファイルのオープンおよび先頭

レコードの読み込み。

4.1.2 一般関数

一般関数は、現在調査中のレコードに関してレコード番号、文管理番号、出典、文、日付、構文木を示すリスト、文の意味的な構造を示すリストをそれぞれ取り出す関数を含む。さらに調査対象のレコードを隣接する前後のレコードに移動する関数もある。

- 関数 AllBun()

引数 なし

戻り値 文

機能 現在調査中のコーパスレコードの文を返す。

4.1.3 文節要素用関数

文節は、参照支援ツールが、コーパスの構文木情報から作成するものである。本来の EDR 日本語コーパスでは文節という単位は用いられていないが、そこから文節という単位を取り出すことは可能である⁷⁾。参照支援ツールでは、文節要素用関数に対して、本来の EDR 日本語コーパスからも (EDR 形式日本語コーパスからも) 文節情報を認識して取り扱う。文節の作成方法については後で述べる。文節要素用関数には、現在の調査対象の文節 (カレント文節) を基準として、係り元の文節番号、係り先の文節番号、をそれぞれ戻り値に持

つ関数とカレント文節に係り先があるかどうか調べる関数がある。

- 関数 CheckKakari()

引数 なし

戻り値 0(係り先がある) または -1(係り先がない)

機能 カレント文節に係り先があるかどうかを調べる。

4.1.4 形態素要素用関数

形態素要素用関数には、カレント形態素の構成要素番号、かな表記、品詞、概念、係り先の構成要素番号、かな表記、品詞、概念のそれぞれを戻り値に持つ関数がある。

- 関数 Kana()

引数 なし

戻り値 かな表記

機能 カレント形態素のかな表記を返す。

4.1.5 複合語情報用関数

複合語情報用関数は文内に複合語があった場合に使用する関数である。EDR 日本語コーパスでは単語と単語が結び付いて各々の概念とは異なった概念になるものを複合語と呼び、慣用句も複合語としている。情報の構成そのものは構成要素情報と変わらない。

複合語情報用関数には、複合語形態素番号、複合語の表記、かな表記、概念、をそれぞれ戻り値に持つ関数と、文内に複合語があるかどうか確認する関数、複数複合語があった場合、隣接する前後の複合語にそれぞれ移動できる関数がある。

- 関数 Fukugou()

引数 なし

戻り値 0(複合語がある) または -1(複合語がない)

機能 複合語があるかないかの判定。

4.1.6 ポインタ操作関数

ポインタ操作関数は、コーパスレコード内の調査点を移動する関数である。すなわち、カレント形態素およびカレント文節について前後に移動する。さらに、係り先文節内で形態素を移動する関数もある。

- 関数 NextKeitaiso()

引数 なし

戻り値 0(移動した) または -1(次の形態素はない)

機能 カレント形態素を次へ移動。(戻り値 0の時)

レコード内の情報を参照するための関数の一部を

付録に示す。

また、関数使用の例を図4に示す。

4.1.7 EDR 日本語コーパスの構文情報からの文節の作成

EDR 日本語コーパスの構文情報では、単語合成関係の一部として係り受けが示されており、結果として単語の係り受け関係が採用された形になっている。しかし、上記の理由により、EDR 日本語コーパス参照支援ツールでは、文節間係り受けを採用した。

この文節は本来の EDR 日本語コーパスの構文情報から作成することができる。具体的な例として、図5に簡単な構文木の例を示す。

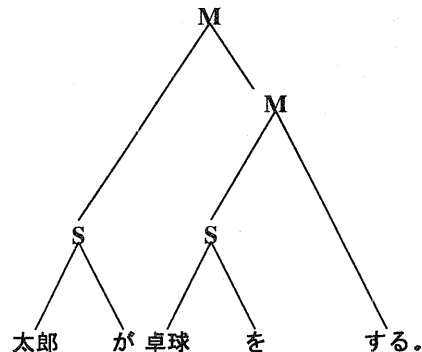


図5 構文木の例

このような構文木があった場合、文節の作成に必要となるのは修飾合成を示す合成関係子の‘M’である。まず、左部分木から順に構文木をたどっていく、葉レベルに最も近い頂点の合成関係子‘M’が現れた時点で、左の子を一つの文節とする。例の場合、‘太郎が’が一つの文節となる。次に、構文木を戻り、‘M’の頂点があれば左の子を一つの文節とする。このとき、左の子の中にすでに文節として確定しているものがあれば差し引く。構文木の根頂点まで戻れば、右部分木に移り同様の処理を進める。例の場合、‘卓球を’が一つの文節とみなされる。実際のコーパスのように、文が長い場合でも、この手順を繰り返すことで、文節が作成できる。また、最終の文節は最後の‘M’が現れた時の右の子、例の場合は‘する’となる。

図6のような場合でも同様に文節を作成することができる。つまり、構文木を下向きにたどって、葉レベルに近い合成関係子‘M’が現れたら、左の子(形態素1と形態素2の部分)を一つの文節とみなす。次に構文木を戻って、合成関係子‘M’が現れ

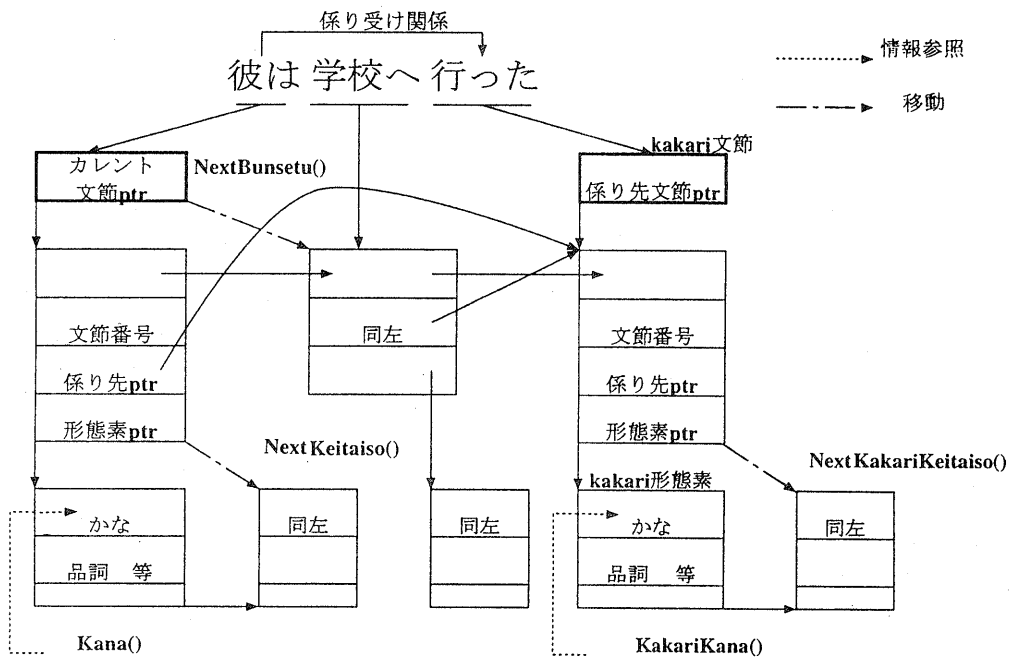


図4 関数使用の例

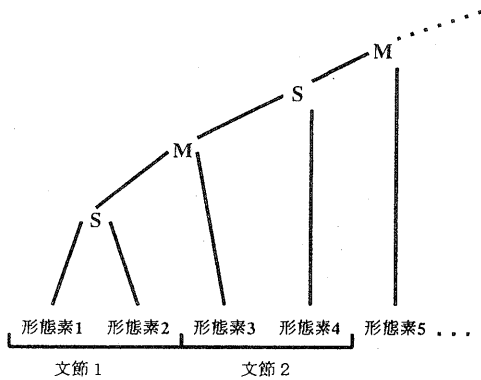


図6 構文木の例

たら左の子（形態素1、形態素2、形態素3、形態素4）を見る。この時、形態素1と形態素2はすでに一つの文節であることが確定しているので差し引く。よって、形態素3と形態素4が2番目の文節と判断できる。

4.2 適用性と効果

参照支援ツールの適用性について検討するために、まず、簡単な調査プログラムを、参照ツールを使用しない場合と使用する場合について作成し、両者を比較した。コーパスの文例から、副詞の係り先の候補数をカウントする調査例において、参

照支援ツールを使用しない場合、400行を越すプログラムを作成しているが、参照支援ツールを使用すれば20数行程度のプログラムで記述することができる。

例の場合は、自然言語処理の調査としてはごく簡単なものであるが、調査のためのプログラムが複雑になると、EDR日本語コーパス参照支援ツールを使用する効果がさらに強く現れると考えられる。

次に参照支援ツールの適用性の評価のため、実際にEDR日本語コーパスを調査・照合している既発表の論文を調べ、そこでのコーパス情報参照法について、本参照ツールの機能での実現性を検討した。

今回は次の各誌上に掲載された論文を対象とした（1995年から1998年に発表されたもの）：情報処理学会研究報告、EDR電子化辞書利用シンポジウム論文集、電子情報通信学会技術報告、言語処理学会発表論文集、情報処理学会論文誌、電子情報通信学会論文集。

調査した論文のリストは、参考文献の後に適用性評価用調査論文として示した。

EDR日本語コーパスを他のEDR電子化辞書と併用して調査している論文については、EDR日本語コーパス参照支援ツールがコーパスだけを対象としているため、評価の対象とはしなかった。

その結果、13編の論文が評価の対象となった。論文中のEDR日本語コーパスの利用方法を例示する。

- EDRコーパス中の名詞の隣接文字の出現確率を集計する。¹⁰⁾
- コーパスの全文に対して動詞を発見し、表内に存在するかどうか確認する。¹³⁾
- 文字列のコーパス中の環境(ある文字列の前後の文字列)の抽出。¹⁴⁾
- 品詞の出現確率を求める。²⁰⁾

これらの調査の結果、対象となったすべての論文で参照支援ツールが利用できることがわかった。

5. 今後の課題と方針

われわれが作成した参照支援ツールは、文献調査の結果、一般的なEDR日本語コーパスの利用方法に対する適用性があることが確認できた。しかし、実際の自然言語処理の検討においてはコーパス以外のEDR辞書を併用する場合もある。これらの他辞書においても参照支援ツールと同様の機能を実現することを検討する必要がある。

また、今後の方針として、EDR日本語コーパスのXML文書化を検討している。XML文書とすることにより、今後のウェブブラウザなどの汎用的なツールを利用して、参照加工などの処理が自由に出来る可能性がある。さらに、現在のEDR日本語コーパスの構文情報では、合成関係は単語間に限定されているが、XML文書とすると、構文木の各頂点に構文要素タグ(例えば名詞句などを表示)を付加することが可能であるという効果もある。また、XML文書とすることにより、これらのタグのセットを固定させる必要がなく、必要に応じてタグセットの定義を変更することが可能である。

このように、試作ツールの改良とコーパス自体の表現力・汎用性の両面から将来の進め方を検討している。

6. おわりに

EDR形式日本語コーパスに関して、CまたはC++プログラムで呼び出すことのできる参照支援ツールについて述べた。調査の結果、本ツールはEDR日本語コーパスに関する過去の調査事例に対して適用性があることがわかった。本稿では、これに関連する他のいくつかのツールについてもその概要と今後の方向について記した。

なお、本研究の一部は文部省科学研究費基盤研究

No11680422(1999)によって行なったものである。

参考文献

- 1) 日本電子化辞書研究所.(1995).EDR電子化辞書仕様説明書, 日本電子化辞書研究所.
- 2) 毎日新聞テキストデータ.(1996). 毎日新聞社.
- 3) 黒橋禎夫他(1998). 京都大学テキストコーパス ver. 2.0. 京都大学長尾研究室
- 4) 潮 靖之, 本間 健一他.(1999秋).EDR日本語コーパス編集システム. 情報処理学会第59回全国大会. .
- 5) 渡部 憲二(1998)EDR日本語コーパスを対象とする文照合方法. 情報処理学会第125回自然言語処理研究会 98-NL-125
- 6) The Pen Treebank Project.
<http://www.cis.upenn.edu/treebank/>.
- 7) 植木正裕他3名(1998). 構造つきコーパスの共有化に関する一考察. 自然言語処理研究会. Vol.128-9,pp.61-66
- 8) 三吉 秀夫, 荻野 孝野(1999). 電子化辞書. 人文学と情報処理 No21,pp.43-53.
- 9) 橋本進吉(1959). 国文法体系論. 岩波書店

以下、適用性評価用調査論文

- 10) 下畑さより 杉尾俊之(1996). 隣接文字情報を用いたn-gram抽出文字列からの名詞句の自動抽出. 信学技報.NLC96-11,pp.13-18
- 11) 植木正裕他3名(1998). 構造つきコーパスの共有化に関する一考察. 自然言語処理研究会. Vol.128-9,pp.61-66
- 12) 西岡山滋之他2名(1998). コーパスからの日本語従属節係り受け選好情報の抽出. 信学技報.NLC98-11,pp.31-38
- 13) 大石亨 松本裕治(1997). コーパスからのアスペクト情報の獲得. 自然言語処理研究会.Vol.118,pp.15-20
- 14) 森信介 長尾眞(1995). nグラム統計によるコーパスからの未知語抽出. 信学技報.NLC95-8,pp.7-12
- 15) 出羽達也他4名(1997). コーパスからの言語知識抽出. EDR電子化辞書利用シンポジウム論文集,pp.17-23
- 16) 内元清貴他2名(1998). MEによる日本語係り受け解析. 自然言語処理研究会. Vol.128-5,pp.31-38
- 17) 白井清昭他2名(1995). EDR日本語コーパスからの確率文脈自由文法の自動抽出に関する研究. EDR電子化辞書利用シンポジウム論文集,pp.57-63.
- 18) 中山聡他4名(1995). EDRコーパスを利用した動詞の語義分類. EDR電子化辞書利用シンポジウム論文集,pp. 21-27
- 19) 荻野孝野他2名(1998). 電子化辞書を用いた, 文系研究者向き言語分析. 言語処理学会,pp. 418-428
- 20) 永田昌明(1995). EDRコーパスを用い

た確率的日本語形態素解析. EDR 電子化辞書利用シンポジウム論文集, pp.49-55.

- 21) 藤尾正和 松本 裕治 (1995). EDR 括弧付きコーパスを利用した, 統計的日本語係り受け解析. EDR 電子化辞書利用シンポジウム論文集, pp.49-55.
- 22) 宇津呂武仁 松本 裕治 (1996). コーパスからの下位範疇化優先度の学習. 信学技報. NLC-96-31, pp.31-38
- 23) 大倉清司 (1996). タグ付きコーパスからの統語・意味的知識の自動獲得. 信学技報. NLC-95-75, pp.9-13.

付録：参照支援ツールの主な関数の一覧

形態素要素用メンバ関数

- 関数 KeitaisoNum()
引数 なし
戻り値 構成要素番号
機能 カレント形態素の構成要素番号を返す.
- 関数 KakariKeitaisoNum()
引数 なし
戻り値 構成要素番号
機能 Kakari 形態素の構成要素番号を返す.
- 関数 Kana()
引数 なし
戻り値 かな表記
機能 カレント形態素のかな表記を返す.
- 関数 KakariKana()
引数 なし
戻り値 かな表記
機能 Kakari 形態素のかな表記を返す.
- 関数 Hinshi()
引数 なし
戻り値 品詞
機能 カレント形態素の品詞を返す.
- 関数 KakariHinshi()
引数 なし
戻り値 品詞
機能 Kakari 形態素の品詞を返す.
- 関数 Gainen()
引数 なし
戻り値 概念情報
機能 カレント形態素の概念情報を返す.
- 関数 KakariGainen()
引数 なし
戻り値 概念情報
機能 Kakari 形態素の概念情報を返す.

文節要素関数

- 関数 BunsetuNum()
引数 なし

戻り値 文節番号

機能 カレント文節番号を返す.

- 関数 CheckKakari()
引数 なし
戻り値 0(係り先がある) または -1(係り先がない)
機能 カレント文節に係り先があるかどうかを調べる.
 - 関数 KakariBunsetuNum()
引数 なし
戻り値 文節番号
機能 Kakari 文節番号を返す. (CheckKakari() の戻り値が 0 の時のみ)
- ポインタ操作関数
- 関数 NextKeitaiso()
引数 なし
戻り値 0(移動した) または -1(次の形態素はない)
機能 カレント形態素を次へ移動 (戻り値 0 の時)
 - 関数 NextKakariKeitaiso()
引数 なし
戻り値 0 または -1
機能 Kakari 形態素を次へ移動 (戻り値 0 の時)
 - 関数 NextBunsetu()
引数 なし
戻り値 0 または -1
機能 カレント文節を次へ移動. (戻り値 0 の時)
 - 関数 BackKeitaiso()
引数 なし
戻り値 0(移動した) または -1(次の形態素はない)
機能 カレント形態素を一つ前に移動 (戻り値 0 の時)
 - 関数 BackKakariKeitaiso()
引数 なし
戻り値 0 または -1
機能 Kakari 形態素を一つ前に移動 (戻り値 0 の時)
 - 関数 BackBunsetu()
引数 なし
戻り値 0 または -1
機能 カレント文節を一つ前に移動. (戻り値 0 の時)