

MUC-7 Coreference タスク用システムの評価

福本 淳一 榊井 文人

沖電気工業(株)
研究開発本部 関西総合研究所
{fukumoto,masui}@kansai.oki.co.jp

あらまし

本稿では、7th Message Understanding Conference (MUC-7) 向けに開発した抽出システムのうち参照関係の解析精度を評価する Coreference タスク用のシステムについて報告する。まず、Coreference タスクの定義について述べた後、我々の評価用システムの参照関係の解析方法について述べる。抽出は Named Entity の認識において用いた表層レベルの情報を用いた認識と機械翻訳システムの解析モジュール得られた構文解析木の情報を用いた認識処理から構成されている。評価用に設定されたテキストにおいて f-measure で約 40% の精度を得ている。

キーワード

参照関係, 照応, 情報抽出, MUC, Tipster

Evaluation of Coreference Task System as used for MUC-7

Jun'ichi Fukumoto, Fumito Masui

Kansai Laboratories, R&D Group
Oki Electric Industry Co., Ltd.

Abstract

In this paper, we present the results of Coreference task evaluation of the 7th Message Understanding Conference (MUC-7) which is supported by DARPA in the States. We will show task definition of Coreference task and then report system overview of the CO task evaluation system. The system consists of surface level recognition module and pattern based recognition module using parsed tree. All the parsed trees of a text is converted into one tree in order to search possible candidates of anaphoric phrases. For the evaluation, the system attained about 40% f-measure score using formal run texts and training texts.

Keywords

Coreference analysis, Anaphora, Information Extraction, MUC, Tipster

1 はじめに

本稿では、Tipster Text Program Phrase III [2][3]の環で行われた情報抽出に関する国際会議 7th Message Understanding Conference (MUC-7) [1] 向けに開発した抽出システム [4] のうち参照関係の解析精度を評価する Coreference タスク用のシステムについて述べる。

MUC-7 では、設定されたタスクについて抽出精度を各参加システムについて評価を行う会議である。MUC-7 において設定されたタスクは、Named Entity (NE), Coreference (CO), Template Element (TE), Template Relation (TE), Scenario Template (ST) の5つがあり、Multi-lingual Entity Task (MET) として日本語と中国語の NE も設定された。TR は今回初めて設定されたタスクである。そして、それぞれのタスクについて各参加団体による抽出結果の評価が行われる。抽出対象テキストは、英語の新聞記事が設定されている。我々は、今回の評価で NE, CO, TE, TR および日本語 NE の5つのタスクの評価に参加し報告を行った [4][5][8][7][8][9]。

本稿では、Coreference タスクの概要及びその評価用抽出システムについて述べる。参照関係の抽出は NE の処理において用いた表層レベルの情報を用いた認識処理と機械翻訳システムの解析モジュール [10][11] から得られた構文解析木の情報を用いた認識処理から構成されている。機械翻訳の構文解析は1文単位で処理されているが、文にまたがる参照関係の解析を行うため1つの親ノードを与えることで、構文解析結果の解析木のすべてを小ノードとして持つような全体を一つの木構造に変換しそれに対して認識処理を行った。

以下では、まず、Coreference タスクの定義について述べた後、我々の評価用システムの参照関係の解析方法について述べる。最後に評価用のテキストを用いた評価結果について述べる。

2 CO タスク定義

CO タスクでは、テキスト中の代名詞、“the”+名詞、固有名詞などの名詞、名詞句のみについての参照関係を抽出する。動詞及び動詞を含むものについては対象外である。参照関係をもつ要素としては、テキスト中で意味的に同一のもので互いに参照しているものがある。抽出された参照関係の情報は NE と同様、SGML タグとしてテキスト中に埋め込まれる。

2.1 CO タグging例

図 1 にタグ付けされたテキストの例を示す。CO のタグを付与する対象のテキストは SGML タグ付けされている。“DOC”, “DOCID”, “STORYID”, “SLUG”, “DATE”, “NWORDS”, “PREAMBLE”, “TEXT” などがある。“<p>” はパラグラフの区切れを示している。

図に示すようにテキストの中で参照関係の存在する要素には“ID”として一意の要素番号が付与されており、その要素が他の要素を参照している場合は参照先の“ID”の値を“REF”として示すことで参照関係を示している。参照関係をもつ要素が連続して現われる場合、“REF”として示される値は、その参照関係の連鎖の要素のうちどの要素の“ID”を示しても構わない。“TYPE”は参照関係にある要素間の関係を示すためのものであるが、現在はすべての値が“IDENT” (for identity) となっている。

2.2 CO タグ付けの基準

CO タスクで抽出するものは、名詞、名詞句、代名詞で参照関係にあるものを抽出する。代名詞は名詞、名詞句、代名詞を参照する人称代名詞、指示代名詞を抽出し、疑問代名詞は抽出対象としない。日付 (“January 23”), 金額 (“\$1.2 billion”), 割合 (“17%”) は参照される場合、名詞句として抽出対象となる。

参照関係を示す要素としては、“president”, “problem” のように type として表現されるものと、“rate”, “temperature” のように function として表現されるものがあるが、どちらも抽出対象とする (the [per share value of [\$4.02]] ... [The stock price]). また、次のような文において

The temperature rose to 90 degrees before dropping to 70 degrees.

“temperature” は “90 degrees”, “70 degrees” の両方とも参照関係にあるものとして抽出する。ここで、“90 degrees” と “70 degrees” が同じであるため実際には誤りであるが、“temperature” はある時点で “90 degrees” であり、また、別の時点で “70 degrees” であり、どちらも “temperature” という type であると解釈し、両方とも参照関係としている。

以下に抽出対象に関するいくつかの注意点について述べる。例中では “*” で示されたもの同士が参照関係にあるものとする。

```

<DOC>
<DOCID> nyt960325.0578 </DOCID>
<STORYID cat=f pri=r> A7644 </STORYID>
<SLUG fv=tia-z> BC-AUP-BLOOM </SLUG>
<DATE> &LR; </DATE>
<NWORDS> <COREF ID="3"> 03-25 </COREF> </NWORDS>
<PREAMBLE>
BC-AUP-BLOOM
ARO
(For use by New York Times News Service clients)
By Tomi Clarke c. <COREF ID="1"> 1996 </COREF> <COREF ID="5"> Bloomberg Business News </COREF> <COREF
ID="7"> Coltec </COREF> Sees <COREF ID="0" TYPE="IDENT" REF="1"> 1996 </COREF> Net Profit <COREF
ID="13"> 2.6 Million </COREF> Lower on <COREF ID="9"> Fokker </COREF> Collapse
</PREAMBLE>
<TEXT>
<p>
<COREF ID="15"> New York </COREF> , <COREF ID="2" TYPE="IDENT" REF="3"> March 25 </COREF> <COREF
ID="4" TYPE="IDENT" REF="5"> (Bloomberg </COREF> ) - <COREF ID="6" TYPE="IDENT" REF="7"> Coltec In-
dustries Inc. </COREF> said loss of sales to the Dutch jet-maker <COREF ID="8" TYPE="IDENT" REF="9"> Fokker NV
</COREF> , which filed for <COREF ID="35"> bankruptcy </COREF> <COREF ID="32" MIN="earlier"> earlier this month
</COREF> , will reduce <COREF ID="10" TYPE="IDENT" REF="6"> its </COREF> <COREF ID="11" TYPE="IDENT"
REF="0"> 1996 </COREF> net income by <COREF ID="12" TYPE="IDENT" REF="13"> 2.6 million </COREF> , or 4
cents per share.
<p>
...

```

図 1: CO タギング例

名詞語句

名詞語句は参照関係の存在する場合にのみ抽出する。名詞語句の参照関係の例を以下にいくつか示す。

- *London* ... *London*-based ...
(2つの“London”は参照関係)
- *Reuters Holding PLC* ... *Reuters* announced that ...
(“Reuters Holding PLC”が“Reuters”で参照)
- Equitable of Iowa Cos ... located in Iowa.
(IowaはNamed Entityの一部であるため参照されない。)
- The price of *aluminum* siding has steadily increased, as the market for *aluminum* reacts to the strike in Chile.
- He was accused of money laundering and *drug* trafficking. However, the trade in *drugs* ...
- *The rate, which was *6 percent*,* was higher than that offered by the other bank. (“The rate, which was 6 percent,”と“6 percent”が参照関係)

動名詞

動名詞は抽出しない。例えば、“*Slowing the economy* is supported by some Fed officials; *it* is repudiated by others.”の文の場合、“Slowing the economy”, “it”は共に抽出しない。但し、ing形であっても“program trading”, “excessive spending”のように名詞等になっているものは抽出する。

代名詞

代名詞の抽出例を以下に行くつか示す。

- “There is no business reason for *my* departure”, *he* added.
- *He* shot *himself* with *his* revolver.
- *He* is, *himself*, unsure of the outcome.

省略語、関係代名詞

省略語、関係代名詞については参照関係を抽出しない

- Bill called John and spoke with him for an hour.
(“spoke”の省略された主語と“Bill”は参照関係を抽出しない。)

- the movie which I saw ...
("which" と "the movie")

並列語句

並列語句の場合は、全体を参照されているものは全体を、また、並列句の各要素が別々に参照されているものはそれぞれを別に抽出する

- *The sleepy boys and girls* enjoy *their* breakfast.

(但し、“The sleepy boys and girls” に対しては MIN=“boys” がタグの中に付与され全体の最小のものを示すことができる。)

- *Edna Fribble* and *Sam Morton* addressed the meeting yesterday. *Ms. Fribble* discussed coreference, and *Mr. Morton* discussed unnamed entities. *Ms. Fribble* was *president* and *CEO of Amalgated Text Processing Inc.*

(“Edna Fribble” は “Ms. Fribble”, “president”, “CEO of Amalgated Text Processing Inc.” に参照され, “Sam Morton” は “Mr Morton” に参照されている。)

同格

同格の関係にある語句はあるものに対する別の記述になっていることから参照関係として抽出する。

- Julias Caesar, the/a well-known emperor, ...
(以上の語句全体と “the/a well-known emperor” が参照関係)
- Peter Holland, 45, deputy general manager, ...
(以上の語句全体と “deputy general manager” が参照関係)

Predicate Nominals

- *Bill Clinton* is *the president of the United States*.
- Phinneas Flounder may be the dumbest man who ever lived.
(確定的でないため参照関係として抽出しない。)

- Phinneas Flounder was almost the first president of the company.

(確定的でないため参照関係として抽出しない。)

2.3 評価方法

評価は以上の定義に基づいて人間によってタグ付けされた正解と各サイトの出力結果をスコアラーププログラムを用いて比較することによって Recall, Precision, f-measure の値を得ることによって行う。

MUC-7 では、1997年7月中旬にタスク定義とトレーニング用のタグ付きテキスト 30 記事がリリースされ、11月中旬の Dry Run の後、1998年3月の始めに Formal Run として最終の評価が行われた。評価期間はそれぞれ 1 週間であった。

3 CO タスク用評価システム

ここでは、今回の MUC-7, CO タスクの評価に参加したシステムについて述べる。MUC-7 用の抽出システムとしては、表層レベルでの CO 認識処理の後、英日機械翻訳システム [10] の解析モジュールをそのままの形式で用い、解析モジュールの出力に対して CO 処理規則を適用することで抽出システムを実現した。CO 処理規則は機械翻訳システムの文法記述言語をそのまま利用することで記述されている。

CO タスク評価用システムの処理の流れを以下に示す。3.7 から 3.11 までの処理が実際の参照関係に認識処理であり、これらの規則は翻訳システム用の文法記述言語を用いて記述されている。

3.1 SGML タグ処理

入力テキストの SGML タグ情報を用いて抽出対象の部分を選択する。

3.2 NE 認識処理

表層レベルでの Named Entity の認識を行う [7][8]。

3.3 表層レベルの CO 認識処理

表層パターン処理における人名の認識では Mr. Bean の様に認識された要素の一部の要素である Bean が再びテキスト中に現れた時、それも人名として認識を行う処理を利用し、表層レベルでの CO の処理を行っている。(但し、この処理は実際には上の NE 認

識処理の一部として実現されている。) 表層レベルの CO 認識処理モジュールは Perl で記述されている。

3.4 表層レベル NE, CO タグ処理

表層レベルの処理で付与された NE タグ, CO タグの情報を次の処理である形態素・構文解析で利用するために構文解析の内部表現である木構造のノードの情報の形式に変換を行うためのフィルタリング処理を行う。

3.5 形態素・構文解析

英日機械翻訳システムの解析モジュールを用いて入力文の形態素・構文解析を行う。形態素・構文解析処理においては、入力テキスト中の各文が順に処理される。構文解析の際の内部表現としては、1文が1つの木構造として表現されており、木構造の各ノードにおいて「属性、属性値」のペアで情報が表現されている。

3.6 木構造変換処理

形態素・構文解析は一文単位の処理であるが、CO タスクの参照関係解析処理ではテキスト全体を処理対象とするため全文を各ノードとして持つ一段上の木構造への構造の変換を行う。例えば、図 2 に示すように、文 1 から文 6 までのすべての木構造が新たな文章ノードの子ノードとしてまとめられ、一つの木構造に変換される。この処理は Perl で記述している。

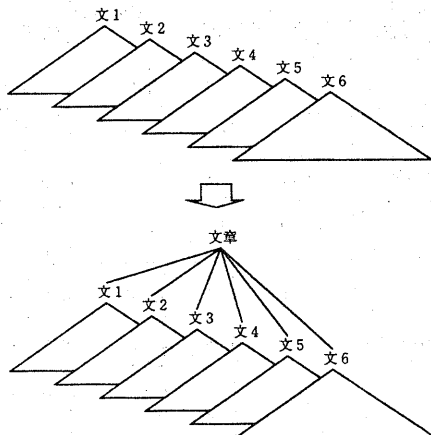


図 2: 木構造変換の概念図

3.7 Predicate Nominals, 同格パターンの参照関係解析

CO タスクの定義にあるような Predicate Nominals と同格表現の構文構造に当てはまる場合、そのパターンに相当する要素を参照関係として抽出する。

3.8 照応語マーキング処理

参照関係の解析候補となるテキスト中のすべての照応語をマークする。照応語としてマークするのは、代名詞と “the”+名詞語句がある。

3.9 代名詞 参照関係認識処理

代名詞の参照関係として、まず、直接話法表現の “I ...”, he said. のように話法の内部と外部の参照関係の解析を行う。その後、その他の代名詞の参照関係の解析を行う。

参照先の探索は、同一文内では文の右から左へ、その前の文では文の左から右へと順に探索を行う。代名詞が参照可能である名詞句に出会った時点で探索を終了し、その要素を参照先とする。参照先の探索順の概略を図 3 に示す。

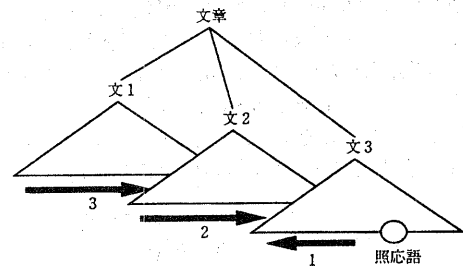


図 3: 参照先の探索順の概略

3.10 “the”+名詞 参照関係認識処理

“the”+名詞の参照関係認識は代名詞の参照解析と同様の探索順で行う。参照可能であるかどうかについては、まず、名詞語句の中心語(ここでは構文解析結果において中心語であるとされたものを選択する)が一致するかどうかについての判断を行う。次に同義となる言葉であるかどうかの判断を行う。

3.11 木構造のフラット化と参照属性の抽出

文章の最終出力結果を得るため木構造を平坦な構造に変換する。この際、句のヘッドに付与されている CO タグの情報をどの範囲の単語までが参照関係にあるのかを判断し、それらのノードの範囲を指定する。

3.12 CO タグの付与

フラット化された木構造から文章の表層情報を保持した属性値から参照関係要素の出力を行う。また、参照関係にあったものについては参照属性にカウンタによって同一の番号が付与されているため、CO タスクの定義に従い参照要素の ID および REF の各番号に付け替えを行う。最後に、参照関係の解析結果を基に CO タグ情報を原文にタグとして付与することにより CO タスク評価用の出力を得る。

この処理は解析結果の出力木から必要な CO タグの情報を原文に付与する処理として Perl で記述されている。

4 実験、評価

CO タスクの評価としては、実際には 11 月の Dry Run にシステム作成が間に合わず、評価としては CO タスク用評価システムで training 用にリリースされた 30 テキストと Formal Run 用に用いられた 20 テキストを用いた。評価は、Recall, Precision, f-measure (以下の表中では f) の 3 つの値で行われた。

$$\text{Recall} = \frac{\text{システムによる正解数}}{\text{全正解数}}$$

$$\text{Precision} = \frac{\text{システムによる正解数}}{\text{システムの全出力数}}$$

$$\text{f-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

表 1、表 2 にそれぞれ training テキスト、Formal Run 用テキストのスコアを示す。

以上の表から training 用 30 テキストについては f-measure 平均で 42.2、Formal Run 用 20 テキストについては f-measure 平均で 38.6 という結果を得ている。

図 4 に CO タスクに参加した全参加団体の結果 (Formal Run) をまとめたものを示す (この結果は

表 1: training テキストスコア

Doc. No.	Recall		Precision		f
9601260514	82/176	46.6	82/121	67.8	55.2
9601290937	23/75	30.7	23/38	60.5	40.7
9602070722	8/39	20.5	8/12	66.7	31.4
9603010607	24/68	35.3	24/53	45.3	39.7
9603030070	64/179	35.8	64/94	68.1	46.9
9603130722	54/127	42.5	54/74	73.0	53.7
9603250578	12/39	30.8	12/19	63.2	41.4
9604010706	18/66	27.3	18/28	64.3	38.3
9604040005	8/42	19.0	8/24	33.3	24.2
9604040662	19/66	28.8	19/28	67.9	40.4
9604090627	9/51	17.6	9/17	52.9	26.5
9604110457	16/70	22.9	16/24	66.7	34.0
9604280393	49/143	34.3	49/111	44.1	38.6
9605082226	8/56	14.3	8/20	40.0	21.1
9605140392	18/93	19.4	18/27	66.7	30.0
9605150480	38/105	36.2	38/54	70.4	47.8
9605150772	25/70	35.7	25/28	89.3	51.0
9605260530	33/102	32.4	33/40	82.5	46.5
9605280696	44/135	32.6	44/94	46.8	38.4
9606040592	25/102	24.5	25/50	50.0	32.9
9606270196	13/34	38.2	13/15	86.7	53.1
9607060278	20/81	24.7	20/39	51.3	33.3
9607080511	18/37	48.6	18/23	78.3	60.0
9607140269	9/30	30.0	9/19	47.4	36.7
9607180041	26/104	25.0	26/44	59.1	35.1
9607190250	24/67	35.8	24/58	41.4	38.4
9607200290	22/49	44.9	22/36	61.1	51.8
9607310536	24/78	30.8	24/31	77.4	44.0
9608310158	16/41	39.0	16/23	69.6	50.0
9609010014	47/140	33.6	47/61	77.0	46.8
TOTALS:	796/2465	32.3	796/1305	61.0	42.2

表 2: Formal Run テキストスコア

Doc. No.	Recall		Precision		f
9601080668	28/63	44.4	28/35	80.0	57.1
9601160264	23/128	18.0	23/84	27.4	21.7
9601260277	14/61	23.0	14/34	41.2	29.5
9602260421	14/44	31.8	14/18	77.8	45.2
9602290649	5/47	10.6	5/10	50.0	17.5
9603100074	43/87	49.4	43/52	82.7	61.9
9604040410	19/53	35.8	19/43	44.2	39.6
9604050312	8/45	17.8	8/18	44.4	25.4
9604300432	24/54	44.4	24/29	82.8	57.8
9605130706	6/54	11.1	6/15	40.0	17.4
9605200107	18/69	26.1	18/31	58.1	36.0
9605210576	32/70	45.7	32/37	86.5	59.8
9606040610	19/76	25.0	19/25	76.0	37.6
9606250230	28/88	31.8	28/29	96.6	47.9
9607160682	8/34	23.5	8/18	44.4	30.8
9607190256	27/77	35.1	27/42	64.3	45.4
9608080354	8/56	14.3	8/12	66.7	23.5
9608210097	6/58	10.3	6/13	46.2	16.9
9609100378	26/79	32.9	26/47	55.3	41.3
9609100495	7/37	18.9	7/7	100.0	31.8
TOTALS:	363/1280	28.4	363/599	60.6	38.6

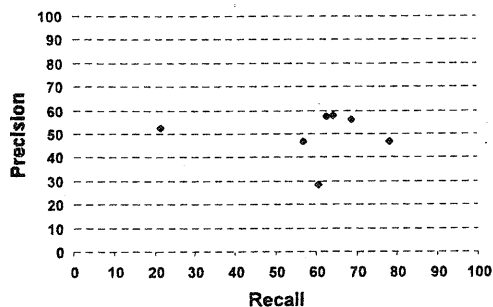


図 4: CO 評価結果 (全参加団体)

MUC-7 の会議において CO タスクの全体の報告において示されたものである。

以上の結果と比較すると我々のシステムは、Precision の値においてはほぼ同程度の性能を示しているが、Recall の値については大きな開きがある。これは、MUC-7 において 5 つのタスクに参加したため、実際の CO タスク評価用システムの開発に割くことのできた時間が 2 カ月弱と限られていたため多くのパターンを調査することができなかつたためである。また、図からもわかるように Coreference タスクはトップレベルの参加団体であっても Precision が 6 割程度であり、難度の高いタスクであることがわかる。

機械翻訳システムの解析モジュールを用いた点については、システムの開発期間が非常に短い間であったにも関わらず解析モジュールが同格表現や直接話法の処理などについて高精度の結果を得ることができたため、この部分については良い結果を得ることができた。しかし、新聞記事などの実際の文章を対象にしているため、固有名詞表現のいくつかが未知語となっているものもあり、この場合の構文解析結果は未知語でなかった場合と大きく構文構造が異なることなど、未知語が含まれていた場合の解析結果に対して予想できない部分があるなどの問題点があった¹。

5 おわりに

本稿では、MUC-7 の Coreference タスクの概要とその評価用に作成した CO タスク用評価システムについて述べた。参照関係の抽出は、表層パターンによる認識処理と、機械翻訳システムの解析モジュール

¹ 翻訳結果についてはなるべくそのままの状態での出力するため構文木の違いはそれほど大きな問題にはならなかった。

ルの解析結果を全体で 1 つの木構造に変換し、その木構造に対する抽出パターン規則を記述することで実現した。表層パターン処理は Perl で記述し、解析木を用いる抽出規則は、機械翻訳システム用の文法記述言語で記述した。MUC-7 で設定された training テキストと Formal Run 用のテキストを用いた評価においては、f-measure で約 40% の結果を得た。

今後は、評価結果の詳細な検討とその分析結果をフィードバックさせるなどのシステムの改良による評価の継続と日本語に対する参照関係の解析を行う予定である。

参考文献

- [1] Proceedings of 7th Message Understanding Conference (MUC-7), DARPA, 1998.
- [2] TIPSTER TEXT Phrase III 18-Month Workshop, DARPA, 1998.
- [3] 福本, 関根, 江里口: “MUC-7, Tipster 参加報告”, 情報処理学会研究報告 98-NL-127, pp.101-108, 1998.
- [4] Fukumoto, J., Masui, F., Shimohata, M. and Sasaki, M.: “Oki Electric Industry: Description of the Oki System as Used for MUC-7”, Seventh Message Understanding Conference (MUC-7), DARPA, 1998.
- [5] Fukumoto, J., Shimohata, M., Masui, F. and Sasaki, M.: “Oki Electric Industry: Description of the Oki System as Used for MET-2”, Seventh Message Understanding Conference (MUC-7), DARPA, 1998.
- [6] 福本, 下畑, 榎井, 佐々木, 杉尾: “パターン処理に基づく情報抽出システムの概要 - MUC7, MET2 参加システム -”, 言語処理学会第 4 回年次大会発表論文集, pp.230-233, 1998.
- [7] 下畑, 福本, 杉尾: “パターンと構文情報による固有名の情報抽出 - MUC7, MET2 参加システム -”, 「テキスト要約の現状と将来」言語処理学会第 4 回年次大会併設ワークショップ予稿集, pp.44-49, 1998.
- [8] 福本, 下畑, 榎井: “固有名詞抽出における日本語と英語の比較”, 信学技報 NLC98-21, pp.45-52, 1998.

- [9] 梶井, 福本 : “新聞記事からの要素間関連情報の抽出”, 信学技報 NLC98-22, pp.53-60, 1998.
- [10] Masui, F., Tsunashima, T., Sugio, T., Tazoe, T. and Shiino, T. : “Analysis of Lengthy Sentences Using an English Comparative Structure Model”, System and Computers in Japan, pp.40-48, SCRIPTA TECHNICA Inc., 1996.
- [11] 北村, 甲斐, 岡田, 永田 : “拡張性を重視した日英機械翻訳システム”, 電子情報通信学会技術報告 NLC, No.91-24 1991.