

機能文節を導入した文節構造解析システム (ibukiC_v0.20)

池田尚志 脇田貴之 大口智也
(岐阜大学)

1. はじめに

文節は、日本語の構造にとって基本的で重要な文法要素である。しかし、いわゆる形態素解析システムの多くは、単語を切り出すこととその品詞などの属性を同定することを目的としており、文節あるいは文節構造に焦点を当ててはいない。文節については、構文解析(係り受け解析)を行う際に、形態素列の中に文節の区切りを入れる処理というところで簡単に扱われているようである。

我々は、文の意味内容を捉える方向で形態素解析を超えて文節構造解析を行うシステムとして ibukiC を開発してきた。形態素解析から構文解析(係り受け解析)に一步近づいたシステムともいえようか。07 年 3 月に開発途上版(v0.10)として公開したが、11 月末に v0.20 を公開予定である(<http://www.ikd.info.gifu-u.ac.jp/ibukiC/>)。V0.10 では、長単位の機能語約 2 万語を登録していたが、V0.20 ではこれを約 5,000 語に整理しなおした。そのうち約 3,500 語は、「V かもしれない」「V コトモノ(V の)」などの特殊な文節を構成するものとして通常の文節から分離し機能文節として分析した。鳥バンク(<http://unicorn.ike.tottori-u.ac.jp/toribank/>)の 15 万文を ibukiC_v0.20 で解析して、その中からランダムに抽出した 100 文について手作業で点検した結果では約 96% 程度の正確度であった。まだ辞書等の整備は不十分であるが、開発途上版 ibukiC_v0.20 として公開することとした。このシステムについて簡単に報告する。

2. ibukiC_v0.20 による解析例

ibukiC は、入力された日本語テキストを、日本語の基本的な文法的単位である“文節”に分割し、さらに文節の内部構造を分析するシステムである。図 1 に ibukiC_v0.20 による解析例を示す。

```
1;0;0;わたし/は;N;わたし;名/代/人; ; ; ;は; ; ; ; ;連用;  
1;1;0;紅白/歌/合戦/が;N;紅白;名/一般; ; ; ; ; ; ; ;複合語;  
1;1;1;特殊文節;TailN;歌;尾/名|名・一般/161.81; ; ; ; ; ; ; ;複合語;  
1;1;2;特殊文節;N;合戦;名/サ; ; ; ; ; ; ; ; ;連用;  
1;2;0;見/た/かつ/た/んです/が/、;P1;見る;動/1 段; ; ; ; ; ; ; ;直後;  
1;2;1;特殊文節;FuncP3;V/たい/、;機/動|形イ/117.37;た; ; ; ; ; ; ; ;直後;  
1;2;2;特殊文節;FuncP1;Y/ん/です/、;機/用|文末/134.164; ; ; ; ; ; ; ; ;連用、  
1;3;0;弟/は;N;弟;名/一般; ; ; ; ;は; ; ; ; ; ; ; ;連用;  
1;4;0;裏番組/の;N;裏番組;名/一般; ; ; ; ; ; ; ;の; ; ; ; ;連体;  
1;5;0;映画/を;N;映画;名/一般; ; ; ; ; ; ; ; ;を; ; ; ; ; ; ; ; ;連用;  
1;6;0;見/た/が/り/ました/。;P1;見る;動/1 段; ; ; ; ; ; ; ; ;直後;  
1;6;1;特殊文節;FuncP1;V/たい/がる/、;機/動|ラ行/117.28;ます/た; ; ; ; ; ; ; ; ;文末;。
```

図 1. 解析例(通常モード)

「わたしは紅白歌合戦が見たかったんですが、弟は裏番組の映画を見たがりました。」

図 1 の各行は 1 個の文節を表現している。(1;2;1)の文節「FuncP3;V/たい/,...」と(1;2;2)の「FuncP1;Y/ん/です/,...」は、(1;2;0)の「見たかったんですが」という文節から分離された“機能文節”である。「た」「が」は、これらの機能文節の“文節要素”として分析されている。

各行の内容は以下のとおりである。

文 id;文節 id;文節 subId;文節;文節カテゴリ;内容語;内容語の品詞;文節要素(1~6);係り先;句読点

“文節要素”は文節のカテゴリ毎に定義してあり(図 2)、否定(ない)/時制(た)や文節間の接続に関わる機能語(格助詞、接続助詞の類)の情報を中心に配置している。

ibukiC の解析には、通常モードと標準化モードがあって、先の例文を標準化モードで解析すると図 3 のようになる。標準化モードでは、「見たかったんですが」は「見たかったのだけれども」と、「見たがりました」は「見たがった」のように解析されている。標準化は辞書の記述に基づいて単純に語彙の変換をしているだけであるが、多くの表記のゆれや表現のゆれをある程度統一的な表現に変換してくれるので、応用の場面によっては有効に使えるものと考えている。

ibukiC はまた、通常の形態素解析のように形態素の列を表示することも出来る(図 4)。表示される内容は以下のとおりである。

文 id;文節 id;誤り可能性の有無;文節 subId;形態素;形態素 id;品詞;読み

“誤り可能性の有無”は、ヒューリスティックな規則によってその部分の解析が誤っている可能性がある判断された場合に“#”を、そうでない場合には” ”を表示する(図 6)。“誤り可能性”には、入力文は正しいが解析が誤っている可能性だけでなく、入力文が誤っている可能性も含む。要するに、解析結果中の何らかの乱れや稀な事象の検出を意図している。この“誤り可能性の有無”については、精度は未だ必ずしも十分ではないが、自動点訳システム ibukiTenC において、点訳結果の点検をする場合の点検箇所を誘導するための情報の一部として利用している。十分な精度と再現率が得られれば、利用範囲はいろいろ広がると考えている。

“読み”は、自動点訳への応用を考慮しており、「は」は「わ」、「おとうと」は「おとーと」など、点字としての表記で読みを表示している。

ibukiC は、辞書や規則の開発を進める上での有用な情報として、解析経過(解析途中の形態素/文節の候補のリスト、そのコスト)を表示することも出来る(図 5)。

○文節 カテ ゴリ コード	○文節カテ ゴリ	○要素1	○要素2	○要素3	○要素4	○要素5	○要素6	補足説明
0	NIL	Φ	Φ	Φ	Φ	Φ	Φ	初期カテゴリ
1	N	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	名詞
2	QC	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	」
3	KA	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	力系
4	SN	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	形式名詞
5	P1	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	動詞
6	P2	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	ダ系
7	P3	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	形容詞
8	P4	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	形容動詞
9	TO	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	引用の「と」
10	AN	Φ	Φ	Φ	Φ	Φ	Φ	連体詞
11	AV	Φ	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	副詞
12	I	Φ	Φ	Φ	Φ	Φ	Φ	感動詞
13	C	Φ	Φ	Φ	Φ	Φ	Φ	接続詞
14	UN	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	未知語
15	UK	Φ	Φ	Φ	Φ	Φ	Φ	空白
16	UP	Φ	Φ	Φ	Φ	Φ	Φ	句点のみ
17	QO	Φ	Φ	Φ	Φ	Φ	Φ	「
18	NUM	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	数詞
25	B?P	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	名詞
51	FuncN	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	機能辞文節
53	FuncKA	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	機能辞文節/力系
54	FuncSN	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	形式名詞
55	FuncP1	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	動詞
56	FuncP2	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	ダ系
57	FuncP3	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	形容詞
58	FuncP4	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	形容動詞
59	FuncTO	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	引用の「と」
61	FuncAV	Φ	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	副詞
101	TailN	取り立て(格の前)	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	名詞を作る接尾語
105	TailP1	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	動詞を作る接尾語
107	TailP3	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	形容詞を作る接尾語
108	TailP4	時制・極性	接続(→体言)	接続(→用言/節)	接続(→用言/格)	モード	終助詞系	形容動詞を作る接尾語
110	TailAN	Φ	Φ	Φ	Φ	Φ	Φ	連体詞
111	TailAV	Φ	接続(→用言)	取り立て(格の後)	取り立て(提題)	接続(→体言)	終助詞系	副詞

図 2. 文節構造の定義

1;1;0;0;わたし;は;N;わたし;名/代/人; ; ;は; ; ; ;連用;
 1;1;1;0;紅白/歌/合戦/が;N;紅白;名/一般; ; ; ; ; ;複合語;
 1;1;1;1;特殊文節;TailN;歌;尾/名|名・一般/161.81; ; ; ; ; ;複合語;
 1;1;1;2;特殊文節;N;合戦;名/サ; ;が; ; ; ; ;連用;
 1;1;2;0;見/た/かつ/た/んです/が/、;P1;見る;動/1段; ; ; ; ; ;直後;
 1;1;2;1;特殊文節;FuncP3;V たい;機/動|形イ/117.37;た; ; ; ; ; ;直後;
 1;1;2;2;特殊文節;FuncP1;Y のだ;機/用|文末/134.164; ; ; ; ; ;けれども; ; ; ;連用、
 1;1;3;0;弟/は;N;弟;名/一般; ; ;は; ; ; ; ;連用;
 1;1;4;0;裏番組/の;N;裏番組;名/一般; ; ; ; ; ;の; ; ;連体;
 1;1;5;0;映画/を;N;映画;名/一般; ;を; ; ; ; ; ;連用;
 1;1;6;0;見/た/が/り/ました/。;P1;見る;動/1段; ; ; ; ; ;直後;
 1;1;6;1;特殊文節;FuncP1;V たがる;機/動|ラ行/117.28;た; ; ; ; ; ;文末、

図 3 . 解析例 (標準化モード)

「わたしは紅白歌合戦が見たかったんですが、弟は裏番組の映画を見たがりました。」

1;0; ;0;わたし;530256;名/代/人;わたし
 1;0; ;1;は;70951;機/名|連用/101.141;わ
 1;1; ;0;紅白;556298;名/一般;こーはく
 1;1; ;1;歌;800147;尾/名|名・一般/161.81;か
 1;1; ;2;合戦;557575;名/サ;かつせん
 1;1; ;3;が;70893;機/名|連用/101.141;が
 1;2; ;0;見;551614;動/1段;み
 1;2; ;1;た;87940;機/動|形イ/117.37;た
 1;2; ;2;かつ;70427;機/形イ|活尾/70.122;かつ
 1;2; ;3;た;88916;機/用|文末/118.170;た
 1;2; ;4;んです;93040;機/用|文末/134.164;んです
 1;2; ;5;が;91598;機/用|連用c/131.147;が
 1;2; ;6;、;96591;記/読点;てん
 1;3; ;0;弟;593964;名/一般;おとーと
 1;3; ;1;は;70951;機/名|連用/101.141;わ
 1;4; ;0;裏番組;620154;名/一般;うらばんぐみ
 1;4; ;1;の;71399;機/名|連体/101.157;の
 1;5; ;0;映画;534607;名/一般;えいが
 1;5; ;1;を;70974;機/名|連用/101.141;を
 1;6; ;0;見;551614;動/1段;み
 1;6; ;1;たが;88577;機/動|ラ行/117.28;たが
 1;6; ;2;り;70338;機/動|活尾/60.111;り
 1;6; ;3;ました;88026;機/動|文末/117.170;ました
 1;6; ;4;。;96590;記/句点;まる

図 4 形態素列の表示例

105 100 1 わ[0 44 95 0] 未知語文節 独立未知語(0,0)

25 20 1 わた[530240 1 1 2] 名詞文節 独立体言(0,1)

15 10 1 わたし[530256 4 18 0] 名詞文節 独立体言(0,2)
25 20 1 わたし[530255 1 1 2] 名詞文節 独立体言(0,2)
29 24 1 わた[530241 9 23 1]し[96600 55 81 1] error 連用(0,2)
29 19 2 わた[530241 9 23 1]し[70288 55 116 0] 動詞文節 連用中止(0,2)

12 12 0 わたし[530256 4 18 0]は[70951 101 141 0 わ] 名詞文節 連用(0,3)
22 22 0 わたし[530255 1 1 2]は[70951 101 141 0 わ] 名詞文節 連用(0,3)
28 28 0 わた[530241 9 23 1]し[96600 55 81 1]は[70951 101 141 0 わ] error 連用(0,3)

105 100 1 た[0 44 95 0] 未知語文節 独立未知語(1,1)

25 20 1 たし[514885 1 1 2] 名詞文節 独立体言(1,2)
29 24 1 た[514519 9 23 1]し[96600 55 81 1] error 連用(1,2)
29 19 2 た[514519 9 23 1]し[70288 55 116 0] 動詞文節 連用中止(1,2)

22 22 0 たし[514885 1 1 2]は[70951 101 141 0 わ] 名詞文節 連用(1,3)
28 28 0 た[514519 9 23 1]し[96600 55 81 1]は[70951 101 141 0 わ] error 連用(1,3)

15 10 1 し[510757 9 104 0] 動詞文節 連用中止(2,2)
15 15 0 し[901931 9 120 1] 動詞文節 文末(2,2)
25 15 2 し[510760 9 116 1] 動詞文節 連用中止(2,2)
35 30 1 し[510751 1 1 2] 名詞文節 独立体言(2,2)
55 30 5 し[59 48 99 4 4] 数詞文節 独立未知語(2,2)

32 32 0 し[510751 1 1 2]は[70951 101 141 0 わ] 名詞文節 連用(2,3)
34 34 0 し[59 48 99 4 4]は[70951 101 141 0 わ] 数詞文節 連用(2,3)

105 100 1 は[0 44 95 0] 未知語文節 独立未知語(3,3)

25 20 1 紅[556267 1 1 0 あか くない こー べに もみ] 名詞文節 独立体言(4,4)

15 10 1 紅白[556298 1 1 0 こーはく] 名詞文節 独立体言(4,5)

23 18 1 紅白[556298 1 1 0 こーはく]歌[800147 161 81 0 か] 名詞文節 連用(4,6)

23 18 0 紅白[556298 1 1 0 こーはく]歌[800147 161 81 0 か]合戦[557575 3 2 0 かつせん]が[70893 101 141 0] 名詞文節 連用(4,9)

25 20 1 白[602884 1 1 0 しろ しら はく] 名詞文節 独立体言(5,5)

29 24 1 白[630094 17 71 0 しら]歌[0 42 93 0] 未知語文節 独立未知語(5,6)
31 26 1 白[630094 17 71 0 しら]歌[537889 1 1 0 うた] 名詞文節 独立体言(5,6)
33 28 1 白[602884 1 1 0 しろ しら はく]歌[800147 161 81 0 か] 名詞文節 連用(5,6)

30 30 0 白[630094 17 71 0 しら]歌[0 42 93 0]合[802036 168 85 0 ごー] 未知語文節 連用(5,7)
31 26 1 白[630094 17 71 0 しら]歌合[537906 1 1 0 うたあわせ] 名詞文節 独立体言(5,7)

図 5. 解析過程の表示例

1;2; ;0;見;551614;動/1段;み
 1;2; ;1;た;88916;機/用|文末/118.170;た
 1;2; ;2;か;91802;機/用|文末/137.168;か
 1;3;#;0;っ;0;名/未知語/ひらがな;っ
 1;4;#;0;で;516985;動/1段;で
 1;5; ;0;すが;2000336;名/個/人名/姓;すが
 1;5; ;1;、;96591;記/読点;てん

図 6 誤り可能性の指摘例

3 . ibukiC の辞書構成

ibukiC では辞書や規則などのすべてのデータは、RDB (関係データベース) 上で管理している。図 7 に、ibukiC の辞書・規則類を管理している RDB のテーブル一覧を示す。

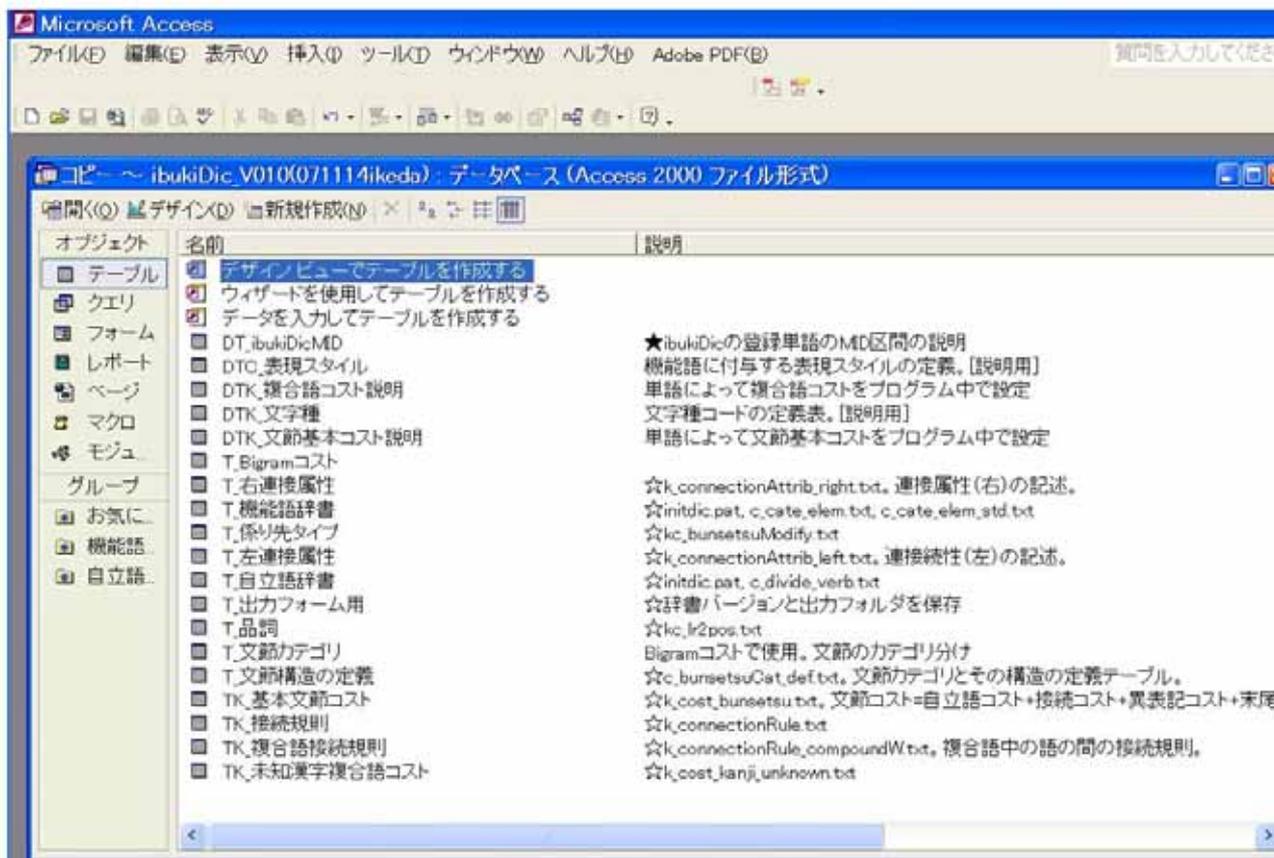


図 7 ibukiC の辞書・規則テーブルの一覧

ibukiC の辞書・規則などは、未だ十分整備された状態ではない。現在のところ機能語辞書には約 4,500 語が登録してあり、そのうち約 3,500 語は機能文節を分割している。内容語辞書には約 24 万語彙が登

録されている。図 8 に内容語辞書の一部を、図 9 に機能語辞書の一部を示す。図 10 に機能文節の例を示す。

ibukiC の解析は、基本的には左右の接続コードによる語と語の接続条件、文節の開始点・終了点の条件、をチェックして接続コストや単語コスト、部分的には bigram コスト、漢字複合語内部のコスト、なども利用して、コストを計算して、最小のコストの単語列/文節列を導き出すという方法で行われる。古典的な解析方法であるが、“形態素”を切り出すことに主眼を置くのではなく文節を切り出し、その構造を分析することに主眼を置いていること、機能語辞書に長めの単位の語を登録していること、“機能文節”を分割していること、などが特徴である。

接続コードは現在のところ、左接続属性が 125 個、右接続属性が 158 個が登録してある(図 11、12)。ibukiC では品詞は基本概念ではないが、左接続属性と右接続属性の組み合わせが ibukiC における品詞といていい。現在のところ、 $125 \times 158 = 19,750$ の組み合わせのうち、約 900 が実際の語彙として登録されており(つまり現在のところ ibukiC の品詞は、細分類すると約 900 個であるということになる)これらには利用しやすいように“品詞名”を与えている(図 13)。

接続規則は、語 A の後に語 B が隣り合って存在し得るかどうかを定めている。接続規則は接続のコストも記述している。コストは手作業で与えており、さほど微妙な調整はしていないがほぼ十分に機能している。

○MID	○登録	○左コード	○右コード	○見出し	終止形	○文字種コード	○単語コストアップレベル	○点訳表記	○読みコスト
501169	TRUE	13	66	いかんせん	如何せん	1	0	いかんせん	0
501171	TRUE	13	66	いかんなく	遺憾無く	1	0	いかんなく	0
501173	TRUE	1	1	いき	闕	1	2	いき	0
501174	TRUE	3	2	いき	遺棄	1	2	いき	0
501176	TRUE	9	30	いき	生きる	1	1	いき	0
501177	TRUE	11	41	いき	糝	1	0	いき	0
501179	TRUE	1	1	いきあたり	行き当り	1	2	いきあたり	0
501180	TRUE	9	31	いきいき	いきいきする	1	0	いきいき/	0
501181	TRUE	9	31	いきいき	生き生き	1	1	いきいき	0
501183	TRUE	13	62	いきいき	生き生き	1	0	いきいき	0
501185	TRUE	1	1	いきうつし	生き写し	1	2	いきうつし	0
501186	TRUE	1	1	いきうま	生き馬	1	2	いきうま	0
501187	TRUE	1	1	いきうめ	生理め	1	2	いきうめ	0
501188	TRUE	1	1	いきえ	生き餌	1	2	いきえ	0
501189	TRUE	1	1	いきおい	勢	1	2	いきおい	0
501190	TRUE	1	1	いきおい	勢い	1	1	いきおい	0
501191	TRUE	13	66	いきおい	勢	1	0	いきおい	0
501193	TRUE	9	27	いきおいこ	勢い込む	1	1	いきおいこ	0
501194	TRUE	1	1	いきがい	生甲斐	1	2	いきがい	0
501195	TRUE	9	28	いきかえ	生還る	1	1	いきかえ	0
501196	TRUE	1	1	いきがけ	行掛け	1	2	いきがけ	0
501197	TRUE	1	1	いきがみ	生き神	1	2	いきがみ	0
501198	TRUE	1	1	いきぎも	生きぎも	1	2	いきぎも	0
501199	TRUE	3	2	いきぎれ	息切れ	1	2	いきぎれ	0

図 8 内容語辞書の一部

「言語・認識・表現」LACE 第 12 回年次研究会, 2007/12/1,2

OMD	登録	左コード	右コード	見出し	点訳表記	読みコード	分割文節カテゴリコード	単語コード	文節分割	文節分割置換	要素1置換	要素2置換	要素3置換	要素3置換	要素4置換	要素5置換	要素5置換	要素6置換	要素6置換	
86555	TRUE	120	168	だであらう	だで/あらう	0	55	0	V/で/ある/う/	Vだらう	だ φ	た φ	φ	φ	φ	φ	φ	φ	φ	φ
86711	TRUE	120	39	だんじやな	だんじや/な	0	55	0	V/ん/じや/ない/	Vのではない	φ た	φ た	φ	φ	φ	φ	φ	φ	φ	φ
86873	TRUE	120	168	だだらうか	だだらう-か	0	55	0	V/だ/う/か/	Vだらうか	φ た	φ た	φ	φ	φ	φ	φ	φ	φ	φ
86874	TRUE	120	168	だでしょう	だでしよー	0	55	0	V/です/う/	Vだらう	だ φ	た φ	φ	φ	φ	φ	φ	φ	φ	φ
86885	TRUE	120	168	だんです	だんです	0	55	0	V/ん/です/	Vのだ	た φ	た φ	φ	φ	φ	φ	φ	φ	φ	φ
87079	TRUE	120	170	だんじやなかつた	だんじや/なかつた	0	55	0	V/ん/じや/ない/た/	Vのではないかつた	φ た	φ た	φ	φ	φ	φ	φ	φ	φ	φ
87141	TRUE	120	190	だではないか	だでわ/ないか	0	55	0	V/では/ない/か/	Vではないか	た φ	た φ	φ	φ	φ	φ	φ	φ	φ	φ
87150	TRUE	117	30	にでかけ	に/でかけ	0	55	0	V/に/でかける/	VIにでかける	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87152	TRUE	117	34	にいい	に/いい	0	55	0	V/に/いく/	VIにいく	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87153	TRUE	117	34	に行	に/い	0	55	3	V/に/行く/	VIにいく	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87154	TRUE	117	103	にこ	に/こ	0	55	1	V/に/くる/	VIこくる	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87155	TRUE	117	103	に来	に/こ	0	55	0	V/に/来る/	VIこくる	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87156	TRUE	117	31	は	は	1	55	1	V/は/する/	VIはする	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87157	TRUE	117	31	も	も	1	55	1	V/も/する/	VIもする	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87158	TRUE	117	114	にき	に/き	0	55	0	V/に/くる/	VIこくる	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87159	TRUE	117	114	に来	に/き	0	55	0	V/に/来る/	VIこくる	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87166	TRUE	117	145	つつ	つつ	0	0	0	φ	φ	φ	φ	φ	つつ	ながら	φ	φ	φ	φ	φ
87169	TRUE	117	145	ながら	ながら	0	0	0	φ	φ	φ	φ	φ	ながら	ながら	φ	φ	φ	φ	φ
87191	TRUE	117	146	えなく	えなく	0	55	1	V/える/	Vできる	φ ない	φ ない	φ	φ	φ (連中)	φ (連中)	φ	φ	φ	φ
87214	TRUE	117	121	がたく	がたく	0	57	0	V/がたい/	Vがたい	φ	φ	φ	φ	φ	φ	φ	φ	φ	φ
87225	TRUE	117	146	がちに	がちに	0	58	8	V/がち/だ/	Vがちだ	φ	φ	φ	φ	φ に	φ に	φ	φ	φ	φ
87315	TRUE	117	146	たくて	たくて	0	57	0	V/たい/	Vたい	φ	φ	φ	φ	φ て	φ て	φ	φ	φ	φ
87316	TRUE	117	146	たくても	たくても	0	57	0	V/たい/	Vたい	φ	φ	φ	φ	φ ても	φ ても	φ	φ	φ	φ
87333	TRUE	117	146	たくな	たくな	0	57	1	V/たい/	Vたい	φ ない	φ ない	φ	φ	φ (連中)	φ (連中)	φ	φ	φ	φ
87334	TRUE	117	146	たくな	たくな	0	57	0	V/たい/ない/	Vたかない	φ	φ	φ	φ	φ て	φ て	φ	φ	φ	φ
87339	TRUE	117	146	たくな	たくな	0	55	0	V/たい/なる/	Vたくなる	φ	φ	φ	φ	φ て	φ て	φ	φ	φ	φ
87343	TRUE	117	146	たくな	たくな	0	55	1	V/たい/なる/	Vたくなる	φ ない	φ ない	φ	φ	φ て	φ て	φ	φ	φ	φ

図 9 機能語辞書の一部

○左 コード	○右 コード	○見出し	○点訳表記	○読 み コ ス ト	○分 割 文 節 カ テ ゴ リ コ ー ド	○単 語 コ ス ト ア ッ プ レ ベ ル	○文節分割	○文節分割_置換	○要素 1	○要素 1_置換	○要素2	○要素2_置換	○要素3	○要素3_置換	○要素4	○要素4_置換	○要素5	○要素5_置換	○要素6	○要素6_置換
101	142	からいうと	から/いうと	0	55	0	N/ノカンテンカラ/	N/ノカンテンカラ	Φ	Φ	Φ からいうと	Φ からいえば	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	142	からいえば	から/いえば	0	55	0	N/ノカンテンカラ/	N/ノカンテンカラ	Φ	Φ	Φ からいえば	Φ からいえば	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	142	からいって	から/いって	0	55	0	N/ノカンテンカラ/	N/ノカンテンカラ	Φ	Φ	Φ からいって	Φ からいえば	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	142	からして	から/して	0	55	0	N/ノカンテンカラ/	N/ノカンテンカラ	Φ	Φ	Φ からして	Φ からいえば	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	142	からすると	から/すると	0	55	0	N/ノカンテンカラ/	N/ノカンテンカラ	Φ	Φ	Φ からすると	Φ からいえば	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	142	からすれば	から/すれば	0	55	0	N/ノカンテンカラ/	N/ノカンテンカラ	Φ	Φ	Φ からすれば	Φ からいえば	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	37	ていった	て/いきた	0	57	1	V/て/いく/たい/	Vていきたい	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	37	ていだけた	て/いだきた	0	55	0	V/て/い/たく/たい/	Vてほしい	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	39	ていてはいけな	ていてわ/いけな	0	57	0	V/て/いる/て/は/いける/ない/	Vていてはいけな	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	39	ていてはならな	ていてわ/ならな	0	57	0	V/て/いる/て/は/いける/ない/	Vていてはいけな	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	38	ていてほし	て/いて/ほし	0	57	0	V/て/いる/て/ほしい/	Vていてほしい	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	39	ていなくもな	て/いなくも/な	0	57	0	V/て/いる/ない/も/ない/	Vていなくもない	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	39	ていなければいけな	て/いなければ/いけな	0	57	0	V/て/いる/ない/ば/いける/ない/	Vていなければならな	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	39	ていなければならな	て/いなければ/ならな	0	57	0	V/て/いる/ない/ば/なる/ない/	Vていなければならな	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
111	39	てきそうにな	て/きそーにな	0	58	0	V/て/くる/Rそう/に/ない/	Vてきそうにな	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	141	とが	とが	0	54	0	N/と/	Nと	Φ	Φ	Φ が	Φ が	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	141	とかが	とかが	0	54	0	N/と/か/	Nとか	Φ	Φ	Φ が	Φ が	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	141	とかに	とかに	0	54	0	N/と/か/	Nとか	Φ	Φ	Φ に	Φ に	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	141	とかを	とかを	0	54	0	N/と/か/	Nとか	Φ	Φ	Φ を	Φ を	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	141	とを	とを	0	54	0	N/と/	Nと	Φ	Φ	Φ を	Φ を	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
85	167	なのだ	なのだ	0	58	0	F/な/の/だ/	Fなのだ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
85	167	なんだ	なんだ	0	58	0	F/な/ん/だ/	Fなのだ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
85	170	なんだった	なんだった	0	58	0	F/な/ん/だ/	Fなのだ	Φ た	Φ た	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
104	181	の	の	0	54	1	N/の/	Nのモノ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
134	183	の	の	0	54	1	Y/の/	Yコトモノ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
102	48	のせい	の/せい	0	58	0	N/の/せい/だ/	Nのせいだ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
102	31	のせいに	の/せいに	0	55	0	N/の/せい/に/する/	Nのせいにする	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	142	のはずが	の/はずが	0	54	0	N/の/はず/	Nデアルはず	Φ	Φ	Φ が	Φ が	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ
101	142	のはずで	の/はずで	0	54	0	N/の/はず/	Nデアルはず	Φ	Φ	Φ で	Φ で	Φ	Φ	Φ	Φ	Φ	Φ	Φ	Φ

図 10 機能文節の例

○左コード	○左コード記号	○説明	○文節タイプ	機能語辞書中の語のカウン ト	自立語辞書中の語のカウン ト
121	l_funcY121	機能辞/形容詞/連用形 \forall する・なる	0	1446	
122	l_funcF122	機能辞/副詞 \forall する・なる	0	474	
123	l_funcY123	機能辞/動詞・形容詞/連用形・語幹	0	224	
124	l_funcAd124	機能辞/イ型/語幹	0	1	
125	l_funcAd125	機能辞/ダ型/語幹	0	279	
126	l_funcAd126	機能辞/形容詞/語幹	0		
127	l_funcAd127	機能辞/イ型・ダ型/終止形・語幹	0		
128	l_funcAd128	機能辞/形容詞/連体形	0	0	0
129	l_funcVK129	機能辞/イ型/終止形	0	0	0
130	l_funcY130	機能辞/動詞・形容詞/終止形(う○)・終止形 \forall と・などと	0	104	
131	l_funcY131	機能辞/動詞・形容詞/終止形(ウ×ダロウ○)・終止形	0	41	
132	l_funcY132	機能辞/動詞・イ型・ダ型/終止形・終止形・語幹	0	174	
133	l_funcY133	機能辞/形容詞/終止形	0	2	
134	l_funcY134	機能辞/動詞・形容詞/連体形	0	1159	
135	l_funcT135	機能辞/動詞/終止形・う	0	2	
136	l_funcY136	機能辞/動詞/終止形	0	334	
137	l_funcY137	機能辞/動詞・イ型・ダ型/終止形・or 語幹/→か	0	142	
138	l_funcT138	機能辞/動詞/終止形・ろう	0	2	
139	l_funcY139	機能辞/動詞・形容詞/仮定形 \forall こそ	0	6	

図 11 左接続コードの例

○右コード	○右コード記号	○説明	○係り先タイプ	○文節末尾コスト	機能語辞書中の語のカウン ト	自立語辞書中の語のカウン ト
157	r_funcTT157	機能辞/体言連体格文節	1	0	174	
158	r_funcTT158	機能辞/体言連体格文節/拡張格	1	0	124	
159	r_funcYT159	機能辞/用言連体格文節	1	0	622	
160	r_funcTJ160	機能辞/体言並列格文節	4	0	73	
161	r_funcT Y?J161	機能辞/体言 連用格?並列格 文節	5	0	22	
162	r_funcYQ162	機能辞/用言 引用格文節	2	0	1	
163	r_funcSe163	機能辞/用言文末文節/終止(だろう)	6	0		
164	r_funcSe164	機能辞/用言文末文節/終止(動)[3]	1	0	2424	3
165	r_funcSe165	機能辞/用言文末文節/終止(う)	1	0	56	
166	r_funcSe166	機能辞/用言文末文節/終止(形イ)	1	0	1183	
167	r_funcSe167	機能辞/用言文末文節/終止(形ダ)	1	0	899	
168	r_funcSe168	機能辞/用言文末文節/終端[1]	1	0	3186	
169	r_funcSI169	機能辞/用言文末文節/命令	6	1	191	4
170	r_funcSe170	機能辞/用言文末文節/終止(た)[4]	1	0	866	
171	r_endI71	機能辞/用言文末文節/終止(ん)[2]	6	0	0	0

図 12 右接続コードの例

id	○表示する品詞説明	○左コード	○右コード	見出しの先頭 (自立語辞書)	見出しの先頭 (機能語辞書)	語彙数 (自立語辞書)	語彙数 (機能語辞書)
1	名/一般	1	1	あばら		140940	
6	名/個/施設名	2	9	コロシウム		33	
8	名/個/自然名	2	11	エーゲ海		7952	
12	名/個/年号	2	15	安永		142	
20	名/時	7	4	いぜん		658	
21	動/力行	9	21	あば		898	
22	動/ガ行	9	22	いそ		149	
45	形/ナイ	10	39	くえな		23	
46	形/ダ/ニナノ	11	40	アブノーマル		934	
47	形/ダ/ニナ	11	41	アプリアリ		1939	
641	機/用文末/134.164	134	164		はずもありません		10
644	機/用文末/134.168	134	168		わけない		14
645	機/用文末/134.170	134	170		んじゃなかった		1

図 13 “品詞” の例

3 . Tori-Bank 1 5 万文の解析例にみる ibukiC_v0.20 の解析精度

Tori-Bank の約 1 5 万文を ibukiC_v0.20 で解析し、その中からランダムに抽出した 100 文を手作業で点検した。

その結果、100 文は 781 文節に解析されたが、そのうち筆者が解析誤り部分と判断したのは 35 文節であった (4.5%) (図 14)。文としては 14 文中に誤りが含まれていた (14%)。解析を誤った文節は次のとおりである。

- 「大したことがなかったので」
 { 雨は大したことがなかったので試合は続行された }
 “たいした” は登録されていたが、“大した” は登録されていなかった。
- 「釣りざおに手ごたえが」
 { 釣り糸を垂れるとすぐ釣りざおに手ごたえがあった }
 “釣竿” は登録されていたが、“釣りざお” は登録されていなかった。
- 「目がけて」
 { 走者はゴール目がけてまっしぐらに走った。 }
 “目掛けて” は登録されていたが、“目がけて” は登録されていなかった。
- 「よぶことは」
 { この子の親は遠くに住んでいるので、すぐよぶことはできません }
 動詞 “呼ぶ” のかわりに、名詞の “よぶこ (呼子)” が採用されてしまった。
- 「くれとは」
 { 金をくれとは汚い }
 動詞 “くれる” のかわりに、名詞の “くれ (暮れ)” が採用されてしまった。
- 「何百ものワープロサブルーチンが」
 { 弊社のザッツテックスと呼ばれるソフトウェア開発ライブラリーには、何百ものワープロサブルーチンが入っています }

- 数量表現の後に続く機能語“もの”が登録されていなかった。
- 「方が」
{この企画はすぐ中止した方がよさそうだ}
“方(ほう)”ではなくて“方(かた)”が採用されてしまった。
- 「多数彼の」
{人々が多数彼のところにきて称賛の言葉を述べた}
“多数”と“彼”が、複合語として合成されてしまった。
- 「なければ」
{つづりに確信がなければ辞書で調べてください}
“無ければ”ではなく、“泣ければ”が採用されてしまった。
- 「あげさるぐつわを」
{強盗は銀行員を縛りあげさるぐつわをかませた}
連用中止の“縛りあげ”として解析しなかった。
「縛りあげ、さるぐつわを」なら正しく解析する。
- 「儀式だったものが」
{かつては意味深い儀式だったものが単なる法的手続きへと墮してしまった。}
“儀式だったものだ”という解釈が採用されてしまった。
- 「なんでも」
{あんなに歌えるならなんでも犠牲にする}
副詞の“なんでも”+“も”と解釈されてしまった。
- 「しゅーといって」
{風船がしゅーといってしぼんだ}
擬音の“しゅー”が登録されていなかった。
- 「燐オキシクロライドなどの」
{例えば、燐オキシクロライドなどの燐化合物は気化させて1,000℃に加熱されたスライスの上を通される。}
“燐”と“オキシクロライド”が、複合語として合成されなかった。

上記 14 件に対する解決策；

は単語を登録することで解決。

は、ひらがな名詞「よぶこ」と「くれ」、およびひらがな動詞「なける」の単語コストをあげることで解決可能。

は「なんでも」を副詞として登録して解決？

は複合語処理の調整が必要 { 複合せない / 複合せせる }

は、機能文節“N だったもの(文末)”および“N だったもの(中間)”を登録して解決？

は、ウム。

「言語・認識・表現」LACE 第 12 回年次研究会, 2007/12/1,2

ID	文節 I D	文節 Su bID	文節・形態素	文節のカテゴリ	自立語	自立語の品詞	機要素 1	機要素 2	機要素 3	機要素 4	機要素 5	機要素 6	係り先情報	句読点
7501	1	0	大	P4	大	形/ダ/ナノ	Φ	Φ	Φ	Φ	Φ	Φ	独 KD	Φ
7501	2	0	し/た/ことがな/かつ/た/ので	P1	N する	動/スル	た	Φ	Φ	Φ	Φ	Φ	直後	Φ
7501	2	1	特殊文節	FuncP3	V/こと/が/ない/	機/用 ナイ/134.39	た	Φ	ので	Φ	Φ	Φ	連用	Φ
9001	4	0	釣/り	P1	釣る	動/ラ行	Φ	Φ	(連中)	Φ	Φ	Φ	連用	Φ
9001	5	0	ざ/おに/手ごたえ/が	N	ざ	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
9001	5	1	特殊文節	N	おに	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
9001	5	2	特殊文節	N	手ごたえ	名/一般	Φ	が	Φ	Φ	Φ	Φ	連用	Φ
13501	1	1	特殊文節	N	目	名/形/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
13501	1	2	特殊文節	N	がけ	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
13501	1	3	特殊文節	UN	て	名/未知語/ひらがな	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
30001	5	0	よぶこ/とは	N	よぶこ	名/一般	Φ	と	は	Φ	Φ	Φ	連用	Φ
63001	1	0	くれ/とは	N	くれ	名/一般	Φ	と	は	Φ	Φ	Φ	連用	Φ
81001	4	0	何/百/もの/ワープロ/サブリーダー/が	N	何	頭/名 名・量/17.73	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
81001	4	1	特殊文節	NUM	百	名/数	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
81001	4	2	特殊文節	N	もの	名/形/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
81001	4	3	特殊文節	N	ワープロ	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
81001	4	4	特殊文節	N	サブリーダー	名/一般	Φ	が	Φ	Φ	Φ	Φ	連用	Φ
88501	3	2	特殊文節	FuncSN	Y/かた/	機/用 名/134.18	Φ	が	Φ	Φ	Φ	Φ	連用	Φ
94501	1	0	多数/彼/の	N	多数	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
94501	1	1	特殊文節	N	彼	名/代/人	Φ	Φ	Φ	Φ	の	Φ	連体	Φ
97501	2	0	なけ/れば	P1	なける	動/1段	Φ	Φ	ば	Φ	Φ	Φ	連用	Φ
102001	3	0	あげ/さるぐつわ/を	N	あげ	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
102001	3	1	特殊文節	N	さるぐつわ	名/一般	Φ	を	Φ	Φ	Φ	Φ	連用	Φ
111001	3	0	儀式/だった/もの/が	N	儀式	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	直後	Φ
111001	3	1	特殊文節	FuncP2	N/だ/	機/名 ダ文末/1021.170	た	Φ	Φ	Φ	Φ	Φ	直後	Φ
111001	3	2	特殊文節	FuncP4	Y/もの/だ/	機/用 形ダ/134.48	Φ	Φ	Φ	Φ	Φ	Φ	直後	Φ
111001	3	3	特殊文節	FuncSN	K/デアルコト/	機/形ダ 連用格/125.141	Φ	が	Φ	Φ	Φ	Φ	連用	Φ
124501	2	0	なんで/も	AV	なんで	副/ダ	Φ	Φ	Φ	Φ	Φ	Φ	直後	Φ
124501	2	1	特殊文節	FuncP3	F/も/	機/副 は・も/89.140	Φ	Φ	Φ	Φ	Φ	Φ	連用	Φ
141001	1	0	しゅ/ー/とい/って	N	しゅ	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
141001	1	1	特殊文節	UN	ー	名/記号	Φ	Φ	Φ	Φ	Φ	Φ	複合語	Φ
141001	1	2	特殊文節	N	とい	名/一般	Φ	Φ	って	Φ	Φ	Φ	連用	Φ
153001	1	0	燐	N	燐	名/一般	Φ	Φ	Φ	Φ	Φ	Φ	独体	Φ
153001	2	0	オキシクロライド/など/の	UN	オキシクロライド	名/未知語/カタカナ	Φ	Φ	Φ	Φ	Φ	Φ	直後	Φ
153001	2	1	特殊文節	FuncSN	N/など/	機/名 連用/101.143	Φ	Φ	Φ	Φ	の	Φ	連体	Φ

図 14 100 文中の解析誤り文節

4. おわりに

ibukiC_v0.20 について簡単に述べた。ibukiC は、単語の切り出しを基本的な任務とする形態素解析ではなく、文節を切り出し、その構造を分析する文節構造解析システムである。ibukiC_v0.20 では、より意味内容をとらえやすくする方向での分析を意図して、「(動詞+)かもしれない」「(動詞+)の」などに対して、「V かもしれない」「V コトモノ」などの機能文節（現在のところ約 3,500）を導入して解析している。

我々の研究室では、ibukiC を基礎部分に利用して、構文解析システム(ibukiS)、自動点訳システム(ibukiTenC)、機械翻訳システム(jaw)、手話テキストへの翻訳システム(jaw/SL)、さらにテキストマイニングなどについて研究・開発している。図 15 は ibukiS による解析例である。ibukiTenC は、ibukiTen と同様に研究室のホームページで公開している(<http://www.ikd.info.gifu-u.ac.jp/ibukiTen/>)。ibukiS、jaw も公開をめざしたい。

ibuikiC の辞書等は未だ整備が不十分な部分も多い。今後も整備を続けていく予定である。

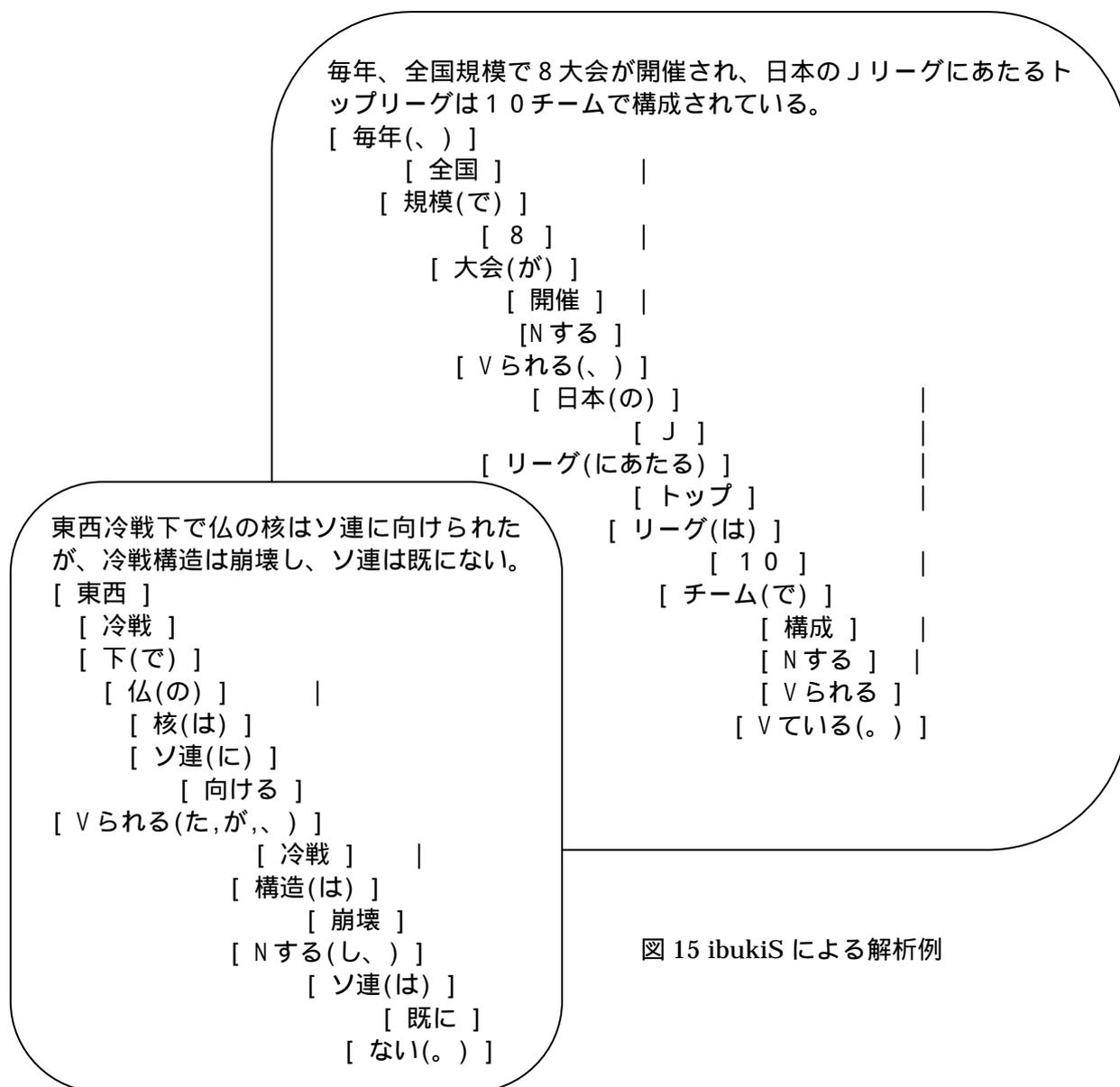


図 15 ibukiS による解析例