

TR-IT-0153

ポーズ単位に基づく音声言語統合処理における
音声認識候補の言語解析実験

Analysis Experiments of Speech Recognition Candidates for
Integrating Speech and Language Processing Based on Pause Units

古谷 成章† 竹澤 寿幸

Narufumi KOYA† Toshiyuki TAKEZAWA

1996. 2

内容梗概

現在、(株) ATR 音声翻訳通信研究所の第四研究所で採用されている SSS-LR 連続音声認識プログラムを用いて得られた出力結果を、言語解析モジュールで解析を行なうための基礎調査を行なった。

今回の実験から音声認識部と言語解析部との部分木を用いた接続が不可能ではないことが分かると同時に、その実現にはまだ多くの課題が残されていることがあらためて確認された。また、依存構造解析を行なうだけでも音声認識部の入力の曖昧性をある程度解消できることも分かった。これに加え、「間投詞」を音声認識部の出力から削除するだけでさらに曖昧性の解消の効果が上がることも確認された。

ATR 音声翻訳通信研究所

ATR Interpreting Telecommunications Research Laboratories

†豊橋技術科学大学 情報工学課程

†Department of Information and Computer Sciences

Toyohashi University of Technology

© 株式会社 エイ・ティ・アール 音声翻訳通信研究所

© 1996 by ATR Interpreting Telecommunications Research Laboratories

目次

1	はじめに	1
1.1	目的	1
1.2	音声認識部、言語解析部概説	1
2	実験内容	2
2.1	文字列での接続	2
2.2	形態素列での接続	3
2.3	部分木での接続	4
3	実験結果	5
3.1	文字列で接続時の結果	5
3.2	形態素列で接続時の結果	6
3.3	部分木で接続時の結果	8
3.4	依存構造解析後の出力結果例	10
3.4.1	形態素列での接続	10
3.4.2	部分木での接続	11
3.5	追加実験結果	12
4	考察	15
4.1	今後の課題	15
4.2	まとめ	16
A	付録	18
A.1	使用対話文	18
A.2	音声認識部の出力例	19
A.3	文字列で接続時の入力例	19
A.4	形態素列で接続時の入力例	20
A.5	部分木で接続時の入力例	20
A.6	文番号14の形態素列での入力	21
A.7	音声認識部実行時間	22
B	今回作成したツール、実験結果	23

1 はじめに

現在、(株) ATR 音声翻訳通信研究所の第四研究所で採用されている SSS-LR 連続音声認識プログラム(音声認識部)を用いて得られた出力結果を、同様に第四研究室で採用されている言語解析モジュール(言語解析部)で解析を行なうための基礎調査を行なった。音声認識部の出力として「文字列」、「形態素列」、「部分木」の三つの状態それぞれで音声認識部と言語解析部をつなぐ場合の問題点の検討を行なった。

本報告書は、二章で今回の実験の内容についてふれ、三章で実験結果及びその検討を報告し、四章で今後の課題について述べている。また、今回の実験で用いたデータ形式などについての例は付録として収録している。

1.1 目的

(株) ATR 音声翻訳通信研究所の第四研究所では、音声言語統合処理研究が行なわれており、その一環として対話システムのために設計・開発された頑健かつ効率的で、しかも移植をする際に容易なものを目指した音声認識部、言語処理部がそれぞれ開発されている。現在、音声認識部で高い処理・認識性能が達成され、その認識結果がデータベースとして蓄積されつつある。また、言語解析部においても処理機構および知識がモジュール化された、さまざまなレベルで外部モジュールとリンク可能なものが準備されつつある。

今回、以上のように音声認識部と言語解析部との接続のための環境が整いつつあることをうけその予備調査として音声認識部と言語処理部との接続する際の接続レベルとして「文字列」、「形態素列」、「部分木」の三通りに関しての接続実験を行ないその際の問題点などに関する検討を行なうことを目的とする。

1.2 音声認識部、言語解析部概説

音声認識部の出力としては、前述のように第四研究室で使用している SSS-LR 連続音声認識プログラムを用いて得られた結果を用いている。今回は、基礎調査であることから文正解率の高いフレーム同期型でビーム幅が5000、L/A比が0.05のものを使用した。

言語解析部は、田代敏久氏等により開発された「音声言語処理のための構文解析ツール」を使用した。これは、「形態素解析」、「形態素調整」、「句構造解析」、「句構造調整」、「依存構造解析」、「格解析」、「意味解析」の七つの独立した基本的なモジュールで構成されたツール群であり、外部モジュールと容易にリンクできるように設計されている。この中の、前半部五つ、「依存構造解析」までのモジュールを利用して今回実験を行なった。

また、本実験で用いた会話は ATR 音声言語データベースの会話番号 TAS12009 (ホテルの部屋のキャンセル: 通訳者と申込者の会話) の申込者の発話である 24 ポーズ区間 (PAUSE UNIT) を用いた。

本実験では、実験、計測のすべてで HP-9000, Model 735/125 を用いた。

2 実験内容

今回、音声認識部と言語解析部をつなぐ場合にその両者間で受渡しされるデータの状態で「部分木」を用いた接続についての有用性などについて検討を行なうため、従来の接続で用いられていた「文字列」を用いた場合と「形態素列」を用いた場合を合わせて三通りについて接続実験を行なった。

「文字列」を用いた接続は、これまで多くの場合で用いられているように音声認識部と言語処理部の接続において最も安易な接続方法である。ただし、今回の実験で用いられている音声認識部ではすでに「形態素解析」、「句構造解析」までの処理が行なわれており、この部分において処理の重複が起こる。このため、今回行なった実験の中で最も処理効率の悪いものであると考えられる。

「形態素列」を用いた接続は、「文字列」を用いた接続に比べれば処理効率は良いと考えられる。しかし、「句構造解析」を重複して行なうためやはり効率的ではないと考えられる。また、形態素列であるため音声認識部と言語処理部との間で常に一意の形態素の対応がとれているとは限らないのでこれを調整する処理が別途必要になると考えらる。

「部分木」を用いた接続は、重複した処理を持たないことから今回行なった処理の中では最も効率的であると考えられる。ただし、形態素列の時と同様に音声認識部と言語処理部との間で常に一意の木の対応がとれているとは限らないのでこれを調整する処理が別途必要になると考えられる。

今回の実験では、基本的に言語解析モジュールの変更をできる限り行なわずに実験を行なった。以下に、三通りの接続それぞれについての実験方法について示す。

2.1 文字列での接続

図 2.1 に音声認識部と言語解析部の両者間で受渡しされるデータを「文字列」とする接続を行なう際の今回使用した言語解析部の構成を示す。付録 A.2 に示されている形式である音声認識部からの出力を「文字列」に変換したものを形態素解析、形態素調整、句構造解析、依存構造解析の順に言語解析モジュールを通すことで最終的に依存構造の生成を行なった。

この接続方法は、言語解析部の基本的動作順序にのっとっているため基本的には接続可能であると考えられる。ただし、形態素調整部については調整規則を新たに作成する必要があった。

以下に今回の実験で変更、または追加を行なった部分について示す。(各モジュールに関する詳細については [1] を参照。)

- 形態素解析: 入出力方法に関する変更のみ行なった。
- 形態素調整: 調整規則が用意されていなかったのをこれを作成した。また、今回作成した変換規則のみでは数字の結合がうまくいかないためこの処理を代行するツールの作成を行ないこのツールと形態素調整モジュールを合わせて今回、形態素調整とした。
- 句構造解析: テキストベースの入力を想定した、構文木規則中に記号(句読点など)が

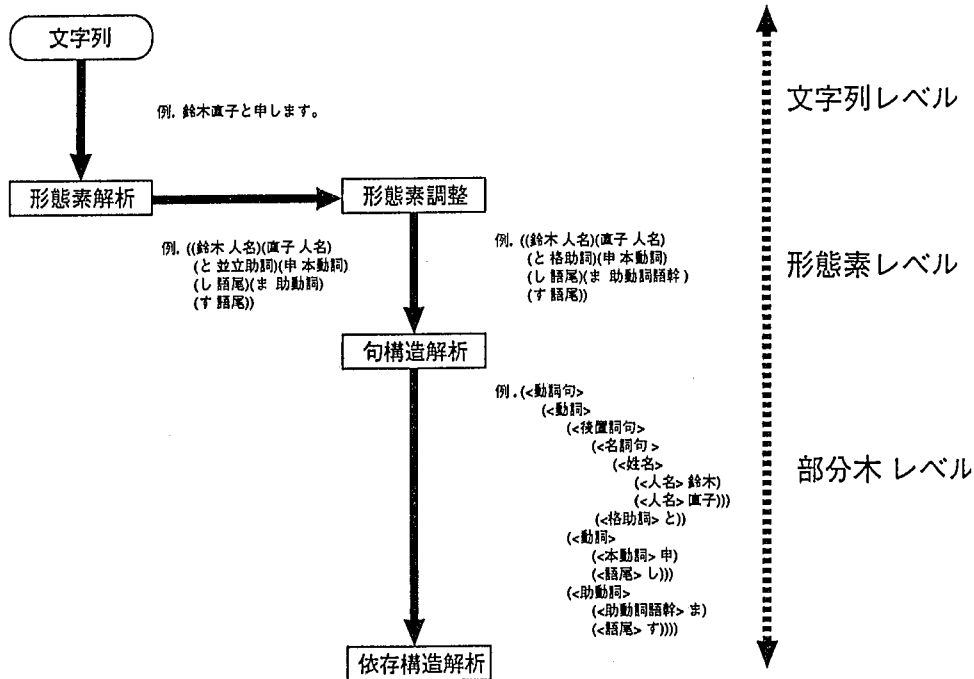


図 2.1: 「文字列」で接続時の言語解析部

含まれる構文木規則のみが用意されていたのでこれを記号の含まれないものに変更した。また、この構文木規則は文単位の入力を想定したものであったのでポーズ単位の入力を扱えるものに変更した。

- 依存構造解析：依存構造規則を得るための書き換え規則が用意されていなかったため、句構造解析で用いた構文木規則に基づいた書き換え規則を作成した。

2.2 形態素列での接続

図 2.2に音声認識部と言語解析部の両者間で受渡しされるデータを「形態素列」とする接続を行なう際の今回使用した言語解析部の構成を示す。

音声認識部の出力を「形態素列」に変換して形態素調整を行なう部分から以後は、「文字列」による接続と同様である。

今回、音声認識部の出力を実際に使用してみたことで音声認識部の出力の中には「間投詞」が多く含まれていることが分かった。「間投詞」は、「句構造解析」以後の処理であまり重要でないと考えられる。そこで今回、追加実験として「間投詞」を削除した結果での実験についても行なった。この際に、「間投詞」の削除処理が必要であり、この処理の容易性から「形態素列」での接続に関してのみ実験を行なった。また、「間投詞」の削除は先に述べた形態素調整として追加的に作成したツールの中で行なった。

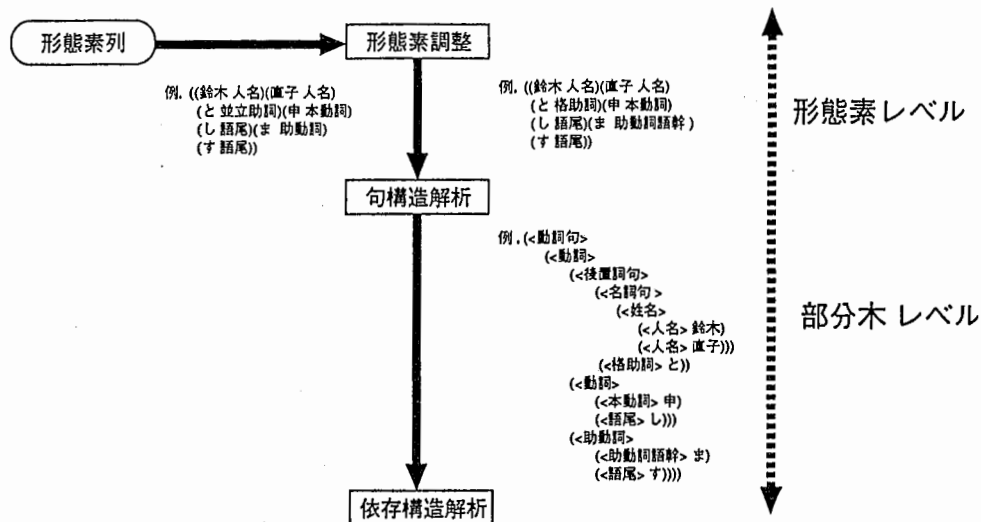


図 2.2: 「形態素列」で接続時の言語解析部

2.3 部分木での接続

図 2.3に音声認識部と言語解析部の両者間で受渡しされるデータを「部分木」とする接続を行なう際の今回使用した言語解析部の構成を示す。

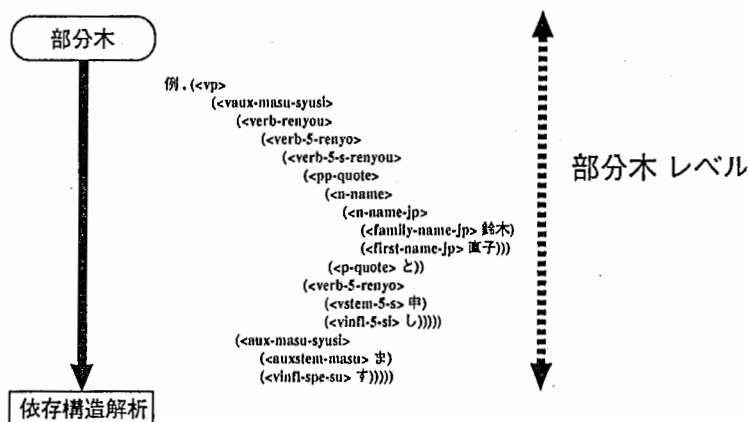


図 2.3: 「部分木」で接続時の言語解析部

- 依存構造解析：句構造調整部が利用できなかったため、SSS-LR連続音声認識プログラムで用いられている構文木規則に基づいた依存構造を得るための書き換え規則を作成した。

3 実験結果

3.1 文字列で接続時の結果

表 3.1: 実行時間

文番号	形態素解析			形態素調整			句構造解析			依存構造解析			合計時間		
	real	user	sys	real	user	sys	real	user	sys	real	user	sys	real	user	sys
1	13.2	12.7	0.4	2.1	1.3	0.1	2.0	0.9	0.1	4.1	2.7	1.3	21.4	17.6	1.9
2	10.1	9.7	0.3	1.8	1.5	0.1	1.6	0.6	0.1	4.0	2.6	1.3	17.5	14.4	1.8
3	25.7	25.3	0.3	3.2	2.5	0.1	3.5	1.6	0.1	4.4	3.0	1.3	36.8	32.4	1.8
4	13.1	12.7	0.2	1.7	1.4	0.1	1.9	0.8	0.2	4.1	2.9	1.2	20.8	17.8	1.7
5	9.5	9.0	0.4	2.3	2.0	0.1	1.5	0.6	0.1	3.9	2.5	1.3	17.2	14.1	1.9
6	6.1	5.8	0.2	1.7	1.1	0.1	1.0	0.5	0.0	3.8	2.5	1.2	12.6	9.9	1.5
7	3.0	2.7	0.3	0.9	0.6	0.1	0.5	0.2	0.1	3.6	2.2	1.4	8.0	5.7	1.9
8	35.5	34.3	0.4	3.3	2.5	0.2	3.7	1.8	0.1	4.8	3.4	1.3	47.3	42.0	2.0
9	6.5	6.2	0.2	2.0	1.7	0.1	0.9	0.4	0.1	3.8	2.4	1.3	13.2	10.7	1.7
10	9.4	9.0	0.3	1.9	1.5	0.2	1.3	0.6	0.1	4.3	2.6	1.3	16.9	13.7	1.9
11	21.3	20.8	0.4	2.7	2.4	0.1	3.0	1.4	0.2	4.3	2.9	1.3	31.3	27.5	2.0
12	10.2	9.7	0.3	2.4	1.6	0.1	1.1	0.5	0.1	4.0	2.6	1.2	17.7	14.4	1.7
13	6.8	6.4	0.3	2.5	1.5	0.1	0.9	0.3	0.1	3.8	2.5	1.2	14.0	10.7	1.7
14	5.0	4.5	0.3	2.0	1.0	0.1	0.8	0.2	0.2	3.7	2.3	1.3	11.5	8.0	1.9
15	5.2	5.0	0.1	2.4	1.5	0.1	1.0	0.3	0.1	3.7	2.4	1.2	12.3	9.2	1.7
16	4.6	4.3	0.2	1.3	1.0	0.1	0.7	0.2	0.2	3.6	2.3	1.3	10.2	7.8	1.8
17	9.0	8.7	0.3	3.1	2.8	0.1	1.6	0.3	0.2	3.8	2.2	1.4	17.5	14.0	2.0
18	10.3	9.9	0.3	3.1	2.1	0.2	1.3	0.6	0.1	3.8	2.6	1.2	18.5	15.2	1.8
19	16.2	15.6	0.3	1.9	1.6	0.1	1.7	0.8	0.1	4.1	2.7	1.3	23.9	20.7	1.8
20	11.2	10.6	0.3	1.4	1.1	0.1	1.2	0.6	0.0	3.7	2.6	1.1	17.5	14.3	1.5
21	11.9	11.5	0.3	2.1	1.8	0.1	1.3	0.5	0.1	3.9	2.6	1.2	19.2	16.4	1.7
22	27.0	26.5	0.4	2.3	2.0	0.1	4.8	2.3	0.1	3.7	2.4	1.3	37.8	33.2	1.9
23	24.8	24.4	0.3	3.8	3.5	0.1	3.2	1.4	0.2	4.3	3.1	1.1	36.1	32.4	1.7
24	16.9	15.6	0.4	2.4	1.1	0.1	0.9	0.4	0.1	3.9	2.3	1.4	24.1	19.4	2.0

表 3.1は、文字列で接続した場合の言語解析部の各モジュールでの処理時間を、コマンドラインの time 関数で測定した値である。そのため、この値は厳密な実行時間を表しているとは限らないが、今回の実験では、各モジュールの接続にはファイル入出力を用いているなどの理由で厳密な時間の測定は意味をなさないのここにとどめた。また、この値は画面やファイルへの出力などの処理も含めた値である。また、表 3.3、表 3.5についても同様である。

3.2 形態素列で接続時の結果

表 3.2: 入出力数変化

文番号	音声認識部	言語解析部		減少率(%)
	出力数	入力数	出力数	
1	37	25	32.4	
2	60	59	1.7	
3	84	82	2.4	
4	27	27	0	
5	66	64	3.1	
6	17	17	0	
7	3	3	0	
8	68	50	26.5	
9	45	45	0	
10	27	26	3.8	
11	24	24	0	
12	28	27	3.6	
13	77	76	1.3	
14	17	17	0	
15	24	24	0	
16	13	13	0	
17	64	64	0	
18	92	89	3.3	
19	24	24	0	
20	35	34	2.9	
21	30	30	0	
22	41	37	9.8	
23	79	72	9.7	
24	48	48	0	
平均	42.9	40.7	5.1	

表 3.2は、形態素列で接続した場合の言語解析部の入出力数の変化、およびその入力に対する出力数の減少率を示している。

本結果から、一つのポーズ節に対する候補としての音声認識部の出力数が最大92個、平均すると43個もの出力があることが分かる。また、形態素列での接続の場合には依存構造解析までの処理を行なっても大きく曖昧性を解消することはないことが分かる。

ここで、出力される形態素列の中身について調べてみると、以下のような表現が多く含まれることが分かった。

- 間投詞が複数含まれるもの。

例、((お 接頭辞)(支払い サ変名詞)(が 格助詞)(どのように 副詞)(し 本動詞)(た 助動詞)(く 語尾)(い 本動詞)(い 語尾)(ま 助動詞)(す 語尾)(えー間投詞)(えー 間投詞)(し 本動詞)(お 接頭辞)(伺 本動詞)(っ 語尾))

- 数詞がポーズ節末に接続しているもの。
例、((分か 本動詞)(り 語尾)(ま 助動詞)(し 語尾)(た 助動詞)(八 数詞))
- 本動詞、語尾の順で複数無意味な接続が起こっているもの。
例、((はい 間投詞)(い 本動詞)(い 語尾)(い 本動詞)(う 語尾))
- 接頭辞、接尾辞が無意味な接続を起こすもの
例、((よろしく 副詞)(お 接頭辞)(お願いいたします 感動詞)(一 数詞))

それに加え例えば、((二 数詞)(四 数詞)(六 数詞)(九 数詞))のような数詞の連結したものが一つのポーズ節をなすときも出力数が多くなる傾向にあることが分かった。

表 3.3: 実行時間

文番号	形態素調整			句構造解析			依存構造解析			合計時間		
	real	user	sys	real	user	sys	real	user	sys	real	user	sys
1	3.1	2.1	0.1	3.3	1.5	0.2	4.5	3.3	1.2	10.9	6.9	1.5
2	3.5	3.1	0.1	3.8	1.7	0.2	4.6	3.4	1.2	11.9	8.2	1.5
3	5.9	5.5	0.1	14.7	7.3	0.2	6.6	5.3	1.3	27.2	11.1	1.6
4	2.3	1.8	0.1	2.1	1.0	0.1	4.2	2.8	1.3	8.6	5.6	1.5
5	3.5	3.2	0.1	2.0	0.8	0.1	4.0	2.8	1.1	9.5	6.8	1.4
6	1.5	1.2	0.1	1.1	0.4	0.1	3.8	2.5	1.3	6.4	4.1	1.5
7	0.8	0.6	0.1	0.5	0.2	0.1	3.6	2.4	1.2	4.9	3.2	1.4
8	4.0	3.7	0.1	7.1	3.2	0.2	6.2	4.7	1.3	17.2	11.6	1.6
9	3.6	2.5	0.0	3.1	0.4	0.1	3.9	2.5	1.2	10.6	5.4	1.3
10	1.9	1.5	0.2	1.2	0.4	0.1	4.0	2.2	1.1	7.1	4.1	1.4
11	2.0	1.2	0.0	4.3	2.1	0.2	4.4	3.0	1.3	10.7	6.3	1.5
12	2.0	1.7	0.1	1.4	0.5	0.1	4.0	2.5	1.3	7.4	4.7	1.5
13	4.1	3.8	0.1	1.9	0.5	0.2	4.2	2.7	1.4	10.2	7.0	1.7
14	1.4	1.1	0.2	0.6	0.3	0.1	3.8	2.5	1.2	5.8	3.9	1.5
15	1.7	1.5	0.1	0.7	0.3	0.1	3.6	2.4	1.2	6.0	4.2	1.4
16	1.2	1.0	0.1	0.6	0.2	0.1	3.6	2.4	1.2	5.4	3.6	1.4
17	3.5	3.2	0.1	1.4	0.3	0.1	3.7	2.6	1.1	8.6	6.1	1.3
18	4.6	4.4	0.1	2.2	0.9	0.2	4.1	2.8	1.2	10.9	8.1	1.5
19	1.9	1.6	0.1	2.6	1.3	0.1	4.2	2.8	1.3	8.7	5.7	1.5
20	2.4	2.1	0.1	1.7	0.9	0.1	4.1	2.6	1.4	8.2	5.6	1.6
21	2.0	1.8	0.1	1.2	0.6	0.1	3.9	2.6	1.2	7.1	5.0	1.4
22	2.6	2.3	0.1	4.5	2.3	0.1	4.8	3.4	1.4	11.9	8.0	1.6
23	4.4	4.1	0.1	4.9	2.4	0.2	5.0	3.7	1.3	14.3	10.2	1.6
24	2.8	2.6	0.1	1.9	0.9	0.1	4.2	2.8	1.4	8.9	6.3	1.6

表 3.3は、形態素列で接続した場合の言語解析部の各モジュールでの実行時間である。これも、表 3.1と同様に厳密な値ではない。

形態素列での接続は、形態素解析を行わないため表 3.1に比べ、総合した実行時間でかなり早くなることが分かる。また、句構造解析、依存構造解析の実行時間のほとんどは画面

やファイルへの出力の時間であることを考慮に入れると形態素調整部は全体の実行時間の半分以上を占めると考えられ形態素調整はかなり重い処理であることが分かる。

また、本結果の中で文番号8、11、17、23については、付録A.1で示されているようにこの音声認識部出力候補の中に正解が含まれていない。ここで、この誤認識の理由について考察すると、このそれぞれの文(ポーズ節)が比較的長い発話であることが挙げられる。長い発話では表3.2で分かるように一つのポーズ節の候補数が多くなる傾向にあり、総合的な処理時間が大きくなる傾向にある。これは、対話システムを構築した時に、聞き直しが起きる場合に認識時間がかかるというようなシステム利用者に不快感を与えかねない事態を引き起こしかねない。この観点から、早い段階での誤認識の発見の必要性もあるのではないかと考えられる。

3.3 部分木で接続時の結果

表 3.4: 入出力数変化

文番号	音声認識部	依存構造解析		減少率 (%)
	出力数	入力数	出力数	
1	122	60		50.8
2	80	50		37.5
3	112	94		16.1
4	29	27		6.9
5	141	85		39.7
6	55	23		58.2
7	3	3		0
8	269	255		5.2
9	116	55		52.6
10	114	47		58.8
11	312	312		0
12	52	30		42.3
13	121	96		20.7
14	39	39		0
15	40	40		0
16	28	28		0
17	128	128		0
18	216	121		44.0
19	207	73		64.7
20	53	27		49.1
21	70	50		28.6
22	51	48		5.9
23	157	108		31.2
24	102	54		47.1
平均	109.0	77.2		29.2

表 3.4は、部分木で接続した場合の依存構造解析のモジュールの入出力数の変化と、その

入力数に対する出力数の減少率を示している。

部分木で接続した場合には、依存構造解析を行なうことで最大64%、平均で30%弱の一つのポーズ節の曖昧性のある候補の数を減らすことができることが分かる。ただし、部分木での接続では一つのポーズ節の曖昧性のある候補の数が最大312個、平均でも110個程度もあるので、依存構造解析後の出力数の平均では、形態素解析で接続した場合にくらべ悪い結果となっている。しかし、今回の実験では句構造調整は行なっていないので句構造調整後の結果では品詞の統合などが起こることを考慮に入れるとさらに出力数は減少すると考えられ、一概に部分木での接続の方が依存構造終了後の一つのポーズ節に関する曖昧性が高いとは言えない。

表 3.5: 実行時間

文番号	依存構造解析		
	real	user	sys
1	6.5	4.4	1.3
2	5.2	3.9	1.2
3	7.1	5.8	1.3
4	4.6	3.2	1.2
5	5.3	3.7	1.3
6	4.7	3.5	1.2
7	4.1	2.9	1.2
8	10.8	9.4	1.4
9	4.8	3.4	1.2
10	5.6	3.8	1.4
11	11.6	10.0	1.3
12	4.7	3.5	1.1
13	4.9	3.6	1.3
14	4.4	3.0	1.3
15	4.4	3.1	1.2
16	4.3	2.9	1.2
17	5.1	3.8	1.2
18	5.4	4.0	1.3
19	9.5	7.9	1.4
20	5.2	3.8	1.2
21	4.8	3.6	1.2
22	5.4	4.2	1.2
23	7.7	6.3	1.3
24	6.1	4.6	1.4

表3.5は、部分木で接続した場合の言語解析部の実行時間である。前の2つの実行時間と同様に厳密な依存構造解析の実行時間を表しているわけではない。

今回の実験では、句構造調整を行なっていない。句構造調整で行なわれるべき処理としては形態素調整、単語、熟語の接続にともなう木の結合、統合などが上げられる。今回行なった実験で形態素列での接続の場合の形態素調整部の実行時間は、言語解析部の中では重い処

理の部類に属することが分かった。句構造調整部も先に述べた処理内容を見ると形態素調整の実行時間と同等ないしそれ以上の処理時間がかかると考えられる。このことを考慮に入れると今回の実験の結果は、部分木で接続したものにかなり合理的な条件下で行なわれていることが分かる。しかし、この実験結果に形態素列での実験の結果表 3.3 の形態素調整部の時間を加えて比較してみるとそれほど大きな差がないことが分かる。ただし、前から述べているように二つの実験条件が違うためこの比較法の結果が意味があるかは分からない。

3.4 依存構造解析後の出力結果例

今回の実験で得られた、依存構造解析後の出力のうち特徴的なものについて以下に示す。ただし、今回の実験ではポーズ単位の依存構造生成規則が確立されていなかったため、出力結果が正確なものであるとはいえない。

3.4.1 形態素列での接続

1: 正解

```
((見出し た)
(カテゴリ <助動詞>)
(引数
((見出し ま#し)
(カテゴリ <助動詞語幹>)
(引数
((見出し てお#り)
(カテゴリ <助動詞語幹>)
(引数
((見出し し)
(カテゴリ <補助動詞>)
(引数
((見出し 予約)
(カテゴリ <サ変名詞>)
(引数
((見出し きょう)
(カテゴリ <副詞>))))))))))
```

2: 出力に明らかにおかしな見出しのものが含まれるもの

```
((見出し と)
(カテゴリ <並立助詞>)
(引数
((見出し 鈴木#直子)
(カテゴリ <人名>))))
((見出し お#ま#す#い#っ)
(カテゴリ <助動詞語幹>))
```

このような出力がかなり多く含まれていることが今回の実験で確認された。このような出力を削除することで、現時点で問題となっている一ポーズ節中に現れる複数の音声認識候補の数が多いことへの解決策の一つになるのではないかと考えられる。

3: 最後に数詞がついているもの

```
((見出し が)
```

```

(カテゴリ <接続助詞>)
(引数
  ((見出し で#す)
   (カテゴリ <助動詞語幹>)
   (引数
     ((見出し ん)
      (カテゴリ <準体助詞>)
      (引数
        ((見出し た#い)
         (カテゴリ <助動詞語幹>)
         (引数
           ((見出し し)
            (カテゴリ <補助動詞>)
            (引数
              ((見出し キャンセル)
               (カテゴリ <サ変名詞>))))))))))))))
  ))
((見出し 八)
 (カテゴリ <数詞>))

```

今回の実験分かった音声認識部の出力結果に現れる特徴の一つであるポーズ節末に数詞が接続される場合の依存構造解析後の結果がこの例である。こののような出力となつては、この時点で、この認識候補が誤っているかを判定するのは難しいと考えられる。このため、このような認識候補については依存構造解析以前に削除されるほうがよいと考えられる。

3.4.2 部分木での接続

1:

```

((見出し た)
 (カテゴリ <aux-ta>)
 (引数
  ((見出し ま#し)
   (カテゴリ <auxstem-masu>)
   (引数
     ((見出し てお#り)
      (カテゴリ <auxstem-te-oru>)
      (引数
        ((見出し し)
         (カテゴリ <aux-suru-si>)
         (引数
           ((見出し 予約)
            (カテゴリ <n-sahen>)
            (引数
              ((見出し きょう)
               (カテゴリ <n-time>))))))))))))))

```

2:

```

((見出し と)
 (カテゴリ <p-quote>)
 (引数
  ((見出し 直子)
   (カテゴリ <first-name-jap>)
   (引数

```

```

((見出し 鈴木)
 (カテゴリ <family-name-jap>))))))
((見出し お)
 (カテゴリ <prefix-o>))
((見出し ま#す)
 (カテゴリ <auxstem-masu>)
 (引数
 ((見出し し)
 (カテゴリ <verb-sahen-renyo>))))))
((見出し い#っ)
 (カテゴリ <vstem-5-w>))
3:
((見出し が)
 (カテゴリ <p-conj-syusi>)
 (引数
 ((見出し で#す)
 (カテゴリ <auxstem-desu>)
 (引数
 ((見出し ん)
 (カテゴリ <p-jun>)
 (引数
 ((見出し た#い)
 (カテゴリ <auxstem-optt>)
 (引数
 ((見出し し)
 (カテゴリ <aux-suru-si>)
 (引数
 ((見出し お#キャンセル)
 (カテゴリ <n-sahen>))))))))))
 (引数
 ((見出し あの)
 (カテゴリ <interj-hesit>))))))
((見出し 八)
 (カテゴリ <n-num-haq>))

```

比較のため形態素列で接続を行なって依存構造解析を行なった結果の三つの例と同じものが入力された時の部分木で接続を行なって依存構造解析を行なった結果を示す。依存構造規則の違いから出力として違うものが出ている場合もあり、このままでは正確な比較はできないが、全体としてはカテゴリーが細かく設定されていることもあり、木が小さくなる傾向にあるように見える。

3.5 追加実験結果

今回の実験を通じ、「間投詞」を含む音声認識部の認識結果がかなり多く見られることが分かった。「間投詞」は、句構造解析においては木の結合を弱め、大きな木をできにくくする悪影響を与え、依存構造解析で係り受け関係を分かりづらくする。また、文の意味を考える際にはその意味を考える上であまり重要でない文成分であると考えられる。このことを考

考慮に入れ間投詞を含まない場合の依存構造解析までの実験を行なった。以下にその実験結果を示す。

表 3.6: 形態素接続時に間投詞を削除した場合の入出力変化

文番号	入力数	間投詞含む		間投詞削除	
		出力数	減少率 (%)	出力数	減少率 (%)
1	37	25	32.4	22	40.5
2	60	59	1.7	56	6.7
3	84	82	2.4	61	24.4
4	27	27	0	21	22.2
5	66	64	3.1	51	22.7
6	17	17	0	11	35.3
7	3	3	0	2	33.3
8	68	50	26.5	50	36.0
9	45	45	0	42	6.7
10	27	26	3.8	15	44.4
11	24	24	0	18	25.0
12	28	27	3.6	21	25.0
13	77	76	1.3	38	50.6
14	17	17	0	17	0
15	24	24	0	24	0
16	13	13	0	13	0
17	64	64	0	63	1.6
18	92	89	3.3	71	22.8
19	24	24	0	24	0
20	35	34	2.9	29	17.1
21	30	30	0	29	3.3
22	41	37	9.8	28	31.7
23	79	72	9.7	54	31.6
24	48	48	0	44	8.3
平均	42.9	40.7	5.1	33.5	21.9

表3.6は、形態素列による接続を行なう場合に間投詞を削除して言語解析部で処理を行なった場合の言語解析部の入出力数の変化を示している。表中「間投詞含む」の下の結果は表3.2のものと同じである。

この結果から、間投詞を含まない認識候補では多くの場合、依存構造解析後の結果に同じ結果が多く含まれていることが分かる。また、間投詞の削除が行なわれても候補の数が減少していない文番号14、15、16に関しては付録A.1を参照すると、正解文が数字のみのポーズ節であることが分かる。また、これらのポーズ節での認識候補は数字のみの表現となっている。(例を、付録A.6に示す。)それに加え、減少していない文番号14、15、16、19は音声認識部の出力候補が、平均の半分程度であることを考えると単に依存構造解析までを実行したのではこれ以上の曖昧性の解消は難しいのではないかと考えられる。

表 3.7: 間投詞の有無による正解文の音声順位変化

文番号	間投詞含む	間投詞削除
1	1	1
2	1	1
3	6	2
4	1	×
5	1	1
6	4	2
7	1	×
8	×	×
9	4	1
10	1	1
11	×	×
12	1	1
13	8	1
14	1	1
15	1	1
16	1	1
17	×	×
18	1	1
19	1	1
20	1	1
21	1	1
22	12	10
23	×	×
24	1	1

音声認識部からの出力には音声スコアも付加されているが今回の実験ではこのスコアの利用法に関する手段が提案されていないのでこの順位のみを利用している。表 3.7は、この順位が間投詞を削除することで起こる依存構造解析後の出力数変化によりどう変化するかを示したものである。また、表中の×印は、正解文が含まれていないことを示す。ここで、文番号 4、7 はこの処理を行なったために正解がなくなったことを示している。文番号 4 は、「と」という間投詞のみのポーズ節であり、この処理を施すことによりそのポーズ節自体が最終的になくなる例である。文番号 7 は、「あのキャンセルしたいんですが」という文中に間投詞が含まれていて、正解文としてはなくなってしまうが意味解析までを含めて考えた場合には正解が含まれている例である。この 2 つの例を除くと正解の順位が 1 位ないし 2 位にほとんどの場合がなっている。

4 考察

今回の実験では、部分木での接続と文字列、形態素列での接続では句構造調整部の準備ができていないために依存構造解析で用いた依存構造生成規則が違うためにそれぞれの厳密な比較が行なえなかった。しかし、今回の実験を通じ音声認識部と言語解析部が持つ今後の課題についての検討は行なえたので以下に示す。

4.1 今後の課題

● 音声認識部

- 音素表記であるアルファベットから日本語表記への変換
例、「しまったので」→「し待ったので」
- 言語解析部と音声認識部での形態素レベルでの品詞の整合性

- ポーズ節内の認識候補の数

今回の実験で、現在出力されている候補の中には三章二節で述べているような傾向があるのでこれを何らかの方法で削除できればかなりの数の候補数を減らすことができるのではないかと予想される。また、今回の追加実験を見ると間投詞がかなり候補数を増やす原因になっているようなのでこの点についても考慮に入れることで候補数の削減が行なえるのではないかと予想される。

● 形態素列で接続する場合の言語解析部

- 形態素調整部の確立

現時点での形態素調整部はアルゴリズム的に閉じた処理となっている。また、形態素の結合や品詞づけで確率を用いた処理を行なっているために常に正しい形態素調整が行なわれるわけではない。以上の点で、現時点での形態素調整部は不十分であるのでこれを解決するものを今後用意する必要がある。

- 一つのポーズ節内の認識候補の曖昧性の解消法の確立

- ポーズ単位の構文木生成規則の確立

● 部分木で接続する場合の言語解析部

- 句構造調整部の整備

音声認識部の品詞体系はかなり細分化されておりこれをそのまま言語解析をすることは困難であると考えられる。また、この細分化された品詞体系が原因で音声認識部の部分木での出力数は、形態素列で接続した場合に比べかなり多いので、曖昧性を小さくする上でも句構造調整部は重要であると考えられる。

- 部分木での曖昧性の解消

今回の実験の結果を見ると部分木で接続した場合には一つのポーズ節での音声認識部が出力する候補の数が平均で100個以上あることが分かる。また、形態素列での接続であっても平均で40個程度あるので今後音声認識側の努力でその数が減少したとしても40個近くまでにしかならないであろうと予測される。また、依存構造解析後の処理でこの数がどこまで減らせるのかは分からない。以上の点からここでの何らかの方法による候補数の削減は重要であると考えられる。

- ポーズ単位の依存構造生成規則の確立

● その他

- 正解を含まないポーズ節の検出

前章でも述べたが、一発話長い場合に音声認識部が誤認識を行なう確率は高いと考えられる。また、一発話長い場合には、その音声認識候補の数は多いと考えられる。このことは、実際に対話システムを構築した時にシステムがユーザの入力を解析できない場合に処理時間がかかりユーザに不快感を与える原因になる可能性があることを意味している。この危険性を回避するためにも早い段階で正解を含まないポーズ節を検出する技術も必要になってくるのではないかと考えられる。

4.2 まとめ

今回の実験から音声認識部と言語解析部との部分木を用いた接続が不可能ではないことが分かると同時に、その実現にはまだ多くの課題が残されていることがあらためて確認された。

音声認識部で生成される部分木は、形態素の単位で生成された品詞の細分化された木である。言語解析部における依存構造解析はできる限り単語や品詞を統合化した部分木を扱うことを前提に開発されている。この音声認識部と言語解析部との間の矛盾を解消するためのツールとして句構造調整モジュールが提案されているが、これはまだ開発段階にあり実現されていない。このモジュールが実現されれば、音声認識部と言語解析部の接続は容易に可能になると予想される。

また、今回の実験で音声認識部からの一つのポーズ節に対する音声認識候補がかなり多く出力されることが確認された。しかし、今回の実験で依存構造解析を行なうだけでもその候補が30%近く削減できることが確認された。また、音声認識部の誤った出力にある程度の規則性があることや、間投詞の削除を行なうことで依存構造解析後の候補の数をかなり絞り込めることや、依存構造解析の結果にも認識候補のしほりこみのためのヒントが隠されていることも確認できた。このような、認識候補の絞り込みのためのいくつかの手法のもとになる可能性のある結果も得られたので認識候補の曖昧性解消もある程度は行なえるのではないかと考えられる。

参考文献

- [1] 田代敏久, 森元逞: “音声言語処理のための構文解析ツールキット”, 情報処理学会自然言語処理研究会, 106-12, (1995.3.10)
- [2] 田代敏久, 竹澤寿幸, 森元逞: “音声言語処理のための部分木併合手法”, 情報処理学会自然言語処理研究会, 109-4, (1995.9.14)
- [3] 田代敏久: “音声言語処理のための構文解析ツールキットユーザーズマニュアル”, ATR テクニカルレポート, TR-IT-0142 (1995.12)
- [4] 竹澤寿幸, 田代敏久, 後藤純司: “部分木を単位とする音声言語解析のための言語現象の調査”, ATR テクニカルレポート, TR-IT-0080 (1994.11)
- [5] 竹澤寿幸, 森元逞: “「ポーズの文法」基礎資料編”, ATR テクニカルレポート, TR-IT-0138 (1995.10)
- [6] 浦谷則好, 竹澤寿幸, 松尾秀彦, 森田千帆: “音声言語データベースの構成”, ATR テクニカルレポート, TR-IT-0056 (1994.5)

A 付録

A.1 使用対話文

文番号	正解文	正解の有無	間投詞の有無
1	きょう予約しておりました	有	無
2	鈴木直子と申します	有	無
3	都合で泊まれなくなってしまったので	有	無
4	あのキャンセルしたいんですが	有	有
5	はい	有	無
6	分かってます	有	無
7	と	有	有
8	キャンセル料はおいくらになりますでしょうか	無	無
9	はい	有	無
10	分かりました	有	無
11	お支払いはどのようにしたらよろしいでしょうか	無	無
12	分かりました	有	無
13	マスターカードで	有	無
14	五二七九	有	無
15	三九二零	有	無
16	二四六九	有	無
17	零零九八です	無	無
18	はい	有	無
19	九十六年四月までになっています	有	無
20	よろしく願いいたします	有	無
21	念のために	有	無
22	そちら様のお名前をお聞かせいただけますか	有	無
23	とジョンフィリップスさんですね	無	有
24	ではよろしく願いいたします	有	無

A.2 音声認識部の出力例

```
SS
Answer: - w a k a q t e m a s u - $
#r 0 0 1 掛かっています 79437089 41.371 -
#r 0 0 2 分かっています 79315032 41.308 @
#S 0 0 960 ms 167430 ms 168 s
PS
[1:1]
{
  <vaux-masu-syusi>
    (<vaux-te-renyo>
      (<verb-5-renyo-q>
        (<vstem-5-r> 掛か)
        (<vinfl-5-q> っ))
      (<aux-te-renyo>
        (<auxstem-te-iru> て)))
    (<aux-masu-syusi>
      (<auxstem-masu> ま)
      (<vinfl-spe-su> す)))
}
[1:2]
{
  <vp>
    (<vaux-masu-syusi>
      (<vaux-te-renyo>
        (<verb-5-renyo-q>
          (<vstem-5-r> 掛か)
          (<vinfl-5-q> っ))
        (<aux-te-renyo>
          (<auxstem-te-iru> て)))
      (<aux-masu-syusi>
        (<auxstem-masu> ま)
        (<vinfl-spe-su> す))))
}
[2:1]
{
  <vaux-masu-syusi>
    (<vaux-te-renyo>
      (<verb-5-renyo-q>
        (<vstem-5-r> 分か)
        (<vinfl-5-q> っ))
      (<aux-te-renyo>
        (<auxstem-te-iru> て)))
    (<aux-masu-syusi>
      (<auxstem-masu> ま)
      (<vinfl-spe-su> す)))
}
PE
SE
```

A.3 文字列で接続時の入力例

```
#START-SENTENCE
#START-PAUSE
[1]
掛かっています
[2]
分かっています
```

#END-PAUSE
#END-SENTENCE

A.4 形態素列で接続時の入力例

#START-SENTENCE
#START-PAUSE
[1]
((掛か 本動詞)(っ 語尾)(て 助動詞)(ま 助動詞)(す 語尾))
[2]
((分か 本動詞)(っ 語尾)(て 助動詞)(ま 助動詞)(す 語尾))
#END-PAUSE
#END-SENTENCE

A.5 部分木で接続時の入力例

#START-SENTENCE
#START-PAUSE
[1:1]
(<DUMMY>
 (<vaux-masu-syusi>
 (<vaux-te-renyo>
 (<verb-5-renyo-q>
 (<vstem-5-r> 掛か)
 (<vinfl-5-q> っ))
 (<aux-te-renyo>
 (<auxstem-te-iru> て)))
 (<aux-masu-syusi>
 (<auxstem-masu> ま)
 (<vinfl-spe-su> す))))))
[1:2]
(<DUMMY>
 (<vp>
 (<vaux-masu-syusi>
 (<vaux-te-renyo>
 (<verb-5-renyo-q>
 (<vstem-5-r> 掛か)
 (<vinfl-5-q> っ))
 (<aux-te-renyo>
 (<auxstem-te-iru> て)))
 (<aux-masu-syusi>
 (<auxstem-masu> ま)
 (<vinfl-spe-su> す))))))
[2:1]
(<DUMMY>
 (<vaux-masu-syusi>
 (<vaux-te-renyo>
 (<verb-5-renyo-q>
 (<vstem-5-r> 分か)
 (<vinfl-5-q> っ))
 (<aux-te-renyo>
 (<auxstem-te-iru> て)))
 (<aux-masu-syusi>
 (<auxstem-masu> ま)
 (<vinfl-spe-su> す))))))
#END-PAUSE
#END-SENTENCE

A.6 文番号14の形態素列での入力

#START-PAUSE

[1]
 ((三 数詞)(九 数詞)(二 数詞)(〇 数詞))
 [2]
 ((三 数詞)(円 接尾辞)(九 数詞)(二 数詞)(〇 数詞)(の 連体助詞)(五 数詞))
 [3]
 ((三 数詞)(九 数詞)(二 数詞)(〇 数詞)(六 数詞))
 [4]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(〇 数詞)(五 数詞)(の 連体助詞)(五 数詞))
 [5]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(二 数詞)(〇 数詞)(の 連体助詞)(二 数詞))
 [6]
 ((三 数詞)(二 数詞)(九 数詞)(二 数詞)(二 数詞)(〇 数詞)(の 連体助詞)(五 数詞))
 [7]
 ((三 数詞)(五 数詞)(九 数詞)(二 数詞)(二 数詞)(〇 数詞)(の 連体助詞)(二 数詞))
 [8]
 ((三 数詞)(九 数詞)(二 数詞)(五 数詞)(〇 数詞)(五 数詞)(五 数詞))
 [9]
 ((三 数詞)(九 数詞)(二 数詞)(〇 数詞)(五 数詞)(五 数詞)(二 数詞))
 [10]
 ((三 数詞)(九 数詞)(二 数詞)(〇 数詞)(二 数詞)(二 数詞))
 [11]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(五 数詞)(〇 数詞)(の 連体助詞)(五 数詞))
 [12]
 ((三 数詞)(九 数詞)(五 数詞)(二 数詞)(二 数詞)(〇 数詞)(の 連体助詞)(五 数詞))
 [13]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(〇 数詞)(六 数詞))
 [14]
 ((三 数詞)(五 数詞)(九 数詞)(二 数詞)(〇 数詞)(五 数詞)(二 数詞))
 [15]
 ((三 数詞)(九 数詞)(二 数詞)(〇 数詞)(五 数詞)(六 数詞))
 [16]
 ((三 数詞)(五 数詞)(九 数詞)(二 数詞)(〇 数詞)(六 数詞))
 [17]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(〇 数詞)(五 数詞)(六 数詞))
 [18]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(二 数詞)(〇 数詞)(六 数詞))
 [19]
 ((三 数詞)(五 数詞)(九 数詞)(二 数詞)(二 数詞)(〇 数詞)(六 数詞))
 [20]
 ((三 数詞)(二 数詞)(九 数詞)(二 数詞)(〇 数詞)(六 数詞))
 [21]
 ((三 数詞)(九 数詞)(二 数詞)(三 数詞)(六 数詞))
 [22]
 ((三 数詞)(九 数詞)(二 数詞)(五 数詞)(〇 数詞)(六 数詞))
 [23]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(〇 数詞)(五 数詞)(五 数詞)(五 数詞))
 [24]
 ((三 数詞)(九 数詞)(二 数詞)(二 数詞)(二 数詞)(〇 数詞)(の 連体助詞)(六 数詞))
 #END-PAUSE

A.7 音声認識部実行時間

表 A.8: 音声認識部実行時間

文番号	real	user	sys
1	333	329.8	1.5
2	324	319.9	1.6
3	644	632.2	2.1
4	329	327.7	1.5
5	40	39.2	0.3
6	168	167.4	0.9
7	112	107.5	0.6
8	679	659.3	2.4
9	46	44.9	0.4
10	191	187.2	1.0
11	738	717.7	2.6
12	187	179.1	0.9
13	198	192.8	1.1
14	220	210.7	1.2
15	208	203.8	1.3
16	197	196.5	1.2
17	221	218.5	1.4
18	88	49.3	0.4
19	668	610.2	2.3
20	573	410.9	1.5
21	182	181.6	1.0
22	819	736.4	2.7
23	348	345.2	1.6
24	460	440.0	1.8

B 今回作成したツール、実験結果

今回作成したツールは、`~xkoya/Test` の以下のディレクトリにあります。各ディレクトリの README ファイルを参照して下さい。

DATA/ 今回実験で用いたデータ、規則生成ツールなどが収められている。
詳細は README ファイルを参照。

Score/ 音声認識部の出力を形態素列で取り出すためのツールが収められている。
詳細は README ファイルを参照。

aud2sem/ 音声認識部の出力を文字列で取り出すためのツールが収められている。
詳細は README ファイルを参照。

Tools/ 音声認識部の出力を部分木でとりだすためのツールと音声認識部の構文規則に基づいた依存構造変換規則を作成するためのツールが収められている。
詳細は README ファイルを参照。

tool/ 活用型活用形つき形態素列の作成を行なうツールが収められている。
詳細は README ファイルを参照。

以下のディレクトリは、田代敏久氏等によって作成された構文解析ツールの入出力部を一部変更したもの、および、今回追加したツールが収められている。構文解析ツールに関するものは、田代氏により詳しく説明されている(参考文献[3])のでそちらの方を参照して下さい。その他のものについては、README ファイルを参照して下さい。

morph/ 形態素解析プログラムが収められています。参考文献 [3] を参照して下さい。

madjust/ 形態素調整プログラムが収められています。参考文献 [3] 及び、README を参照して下さい。

cfg/ 構文解析プログラムが収められています。参考文献 [3] を参照して下さい。

Depend/ 依存構造解析プログラムが収められています。参考文献 [3] 及び、README を参照して下さい。