「情報処理学会論文誌」第33巻 第11号別刷　平成4年11月発行

音声対話文における助詞落ち・倒置の分析と解析手法

山本幹雄　小林聡　中川聖一

社団法人 情報処理学会

Vol. 33 No. 11 情報処理学会論文誌 Nov. 1992

音声対話文における助詞落ち・倒置の分析と解析手法†

山本幹雄†† 小林聡†† 中川聖一††

音声対話における発話文は，言い淀み，言い直し，間投詞，助詞の省略，倒置などの話し言葉特有の特徴を持つため，これまでの書き言葉に対する自然言語の解析手法をそのまま適用するには問題がある．本論文では解析において，まず問題となる名詞文節の助詞落ちと倒置について，実際の音声対話文約1,800文を分析し，その結果をもとに解析手法を提案する．音声対話文では，名詞文節の約4％の助詞が省略されていた．省略される助詞は「が，を，に，は」など述部に係る場合に必須格の機能を持つものが80％を占めていた．係り先の性質としては，述部に係る助詞落ち名詞文節の99％が最も近くの述部に係る．また，文頭にある助詞落ち名詞文節は「は」が省略される可能性が高く（68％），遠くに係る可能性を持っている．また，係り関係（格）については，述部の格構造の簡単な意味制約によって，90％が推定できることが分かった．倒置に関しては，述部に係る文節が倒置される場合が94％を占めており，倒置された句が1つ前の文節に係る場合が91％であった．また，倒置された句の直前の文節は必ず終止形で終わっていることが分かった．以上の分析を反映したヒューリスティックスを助詞落ちに関して5つ，倒置に関して2つ提案した．語彙が700の小規模な実験タスクで評価した結果，助詞落ち，倒置共に約90％の例を正しく解析できることが分かった．

1. はじめに

自然言語によるマン・マシンインタフェースとしての対話システムは盛んに研究されているが，まだ人間にとって自然な文を処理できるとは言いがたい．さらに，自然言語による対話は音声入出力によってはじめてその真価を発揮すると考えられるが，音声発話文に対する統語・意味・語用論的特徴の研究はまだ始まったばかりである[1)~3)]．これまでの自然言語処理の研究では，その多くが本や新聞・論文などの書き言葉を対象としているが，対話において人間が使用する発話文は一般に，言い淀み，言い直し，間投詞，省略の多用，話し言葉特有の言い回しなど，書き言葉と異なる点が多く，書き言葉の文法をそのまま利用するわけにはいかない[4)]．

本論文ではこれまでの書き言葉に対する文の解析手法を対話中の発話文に適用する際まず問題となる，名詞文節の助詞の脱落と文節の倒置の特徴を音声対話データベースのデータから分析し，解析手法を提案する．

以下，第2章で分析に使用した音声対話データベースについて，第3章で助詞落ちの分析結果，第4章で倒置の分析結果を述べる．これらの現象に関する言語学的・運用論的考察は，1つ1つ克明に説明する立場など種々考えられるが，本論文は工学的システムに応用する観点から大局的な観点に立ち，統計的に分析した．第5章でそれぞれに対する解析ヒューリスティックスを提案し，第6章で提案したヒューリスティックスを用いた実験システムと実験結果を示す．第7章で本論文の結論を述べる．

2. 使用した対話データベース

対話文の分析用として，日本音響学会連続音声データベースの書き起こしテキストを使用した[5)]．これには，音声によるシミュレーション対話の書き起こしテキストが含まれる．各対話は2人の人間がお互い見えない状態で電話（もしくはそれに類似した状況）で対話したものである．発声速度は各対話ごとにまちまちである．この書き起こしテキストの一部を今回の分析に使用した．助詞落ち・倒置共に，分析対象とした対話は以下のようなものである．

対話数：22人による20対話
総文数：1,818文
文節数：13,239文節（間投詞，相づち，言い直しを除いた文節数は8,520）
ドメイン：観光案内が半分で，残りは各種相談やその他の案内である．

今回の分析は名詞文節を対象としたものだが，その係り受け関係も対象としているため，すべての文の文節と係り受け関係の分析を行い基礎資料とした．文の認定（どこで切るか）は，1つ以上の述部を持っており，意味的にまとまりがあるものを文とした．述部が省略されているものについては述部がない文も認め

† An Analysis and Parsing Method of the Omission of Post-Position and Inversion on Japanese Spoken Sentence in Dialog by MIKIO YAMAMOTO, SATOSHI KOBAYASHI and SEIICHI NAKAGAWA (Department of Information and Computer Sciences, Toyohashi University of Technology).

†† 豊橋技術科学大学情報工学系

た．また，間投詞，相づち，割り込み，言い直しについてはすべて無視し（1,818文中1,137文に存在した．これらの分析は今後の課題である），それらを取り除いたテキストを使用した．

名詞の助詞の脱落・誤りを分析する場合の問題点は，音声を録音テープから書き起こしてテキストにする段階で，無意識の内に助詞を入れてしまったり，誤りを正してしまう可能性である．また，ワープロの入力誤りの可能性もある．このため，データベースには音声を忠実に書き起こしたという信頼性が要求される．われわれは，テキストの元となった音声テープと，書き起こされたテキストと照らし合わせて20対話1,818文に対して誤りが見つからなくなるまでチェックした．

3. 名詞文節の助詞落ちの分析

係り受けに基づく構文解析および日本語の意味解析では名詞文節の助詞によって表現されている格情報が重要な役割を果たしている．助詞の省略は，格情報の欠落を意味するため，構文・意味解析にとって大きな問題となる．本章では，実際の音声対話文での名詞文節の助詞落ちを分析する．

3.1 脱落した助詞の種類

名詞文節全体と助詞の脱落した名詞文節を助詞によって分類したものを表1に示す．脱落した助詞は文脈を考慮にいれてわれわれが推定した．表層的にいくつかの可能性がある場合（例えば，格助詞と格助詞相当表現）は，すべての省略されている助詞が意味的に格助詞あるいは副助詞の「は」の機能を持っているため，格助詞あるいは副助詞の「は」を推定結果とした．しかし，実際に表層的に曖昧性を感じることはまれであった．また，人間にとって意味的に曖昧な例はなかった．表の左側は係り助詞「は」の省略をそのまま「は」の省略としたもの，右側は「は」が「が，を，に」などの格助詞としての機能を持つ場合にその格助詞のほうに分類したものである．表の右側で，「は」の省略と分類されているものは主題としての「は」である．また，「文節全体」の欄の数は助詞が脱落した文節の数を含んでいる．

表1 名詞文節の助詞の脱落の割合
Table 1 The omitted rate of post-positions.

助詞	「は」はそのまま			「は」を格助詞に分配		
	文節全体	脱落文節	比率%	文節全体	脱落文節	比率%
は	439	63	14	293	5	2
が	419	23	5	509	67	13
を	228	43	19	256	52	20
に	383	10	3	399	15	4
へ	10	—	—	10	0	0
で	277	2	1	277	2	1
まで・から	117	0	0	119	0	0
と(引用)	109	1	1	109	1	1
と(並列)	98	17	17	98	17	17
の(連体)	647	9	1	647	9	1
だ(助動詞)	501	3	1	501	3	1
その他	835	0	0	845	0	0
合計	4063	171	4	4063	171	4

以下に助詞の省略の例を示す．括弧の中の助詞が省略された助詞である．“ᵛ”はポーズを表し，[ ]内は発声速度を表す．これ以降の例文に対しても同じ表記を使用する．

- 高速(が)できてますので…[6.9モーラ/秒]
- 馬(を)飼ってますんで…[8.4モーラ/秒]
- 中央自動車道(に)乗っていただいて，[9.5モーラ/秒]
- バスᵛ(で)，ずーっと見てまわれるということなんですが，[7.1モーラ/秒]
- これ(は)もともとᵛ車の値段が安いもんで…[7.6モーラ/秒]

※この「は」は主題を表す「は」である．

- そこ(の)あたりᵛはᵛちょっと分からないんᵛですね．[7.6モーラ/秒]
- テニスᵛ(と)ゴルフᵛがあります．[7.9モーラ/秒]

省略に対する分類でその他に注意する点は以下のとおりである．

(1) 省略における「へ」は，「に」に置き換えることが可能なので「に」に統一した．

(2) 言い直しによる助詞のない名詞文節は省略と見なさなかった．例えば，「江戸時代の建物，建物はもうないんですけれども」のような文で，最初の「建物」という名詞文節は助詞が省略されているとは考えない．この現象は「言い直し」という枠組みで検討されるべきである．

(3) 「だ，です」は助動詞であり助詞ではないが，分類には入れた．しかし，数も少なく助詞の省略ではないので以下の分析では無視する．

(4) 1文節文（名詞1個からなる文節）は助動詞「だ，です」の省略と考えられるが，非常に特殊なので省略とは考えない．分類は「その他」とした．

（5） 名詞を並べただけで，「と」などで接続されていない並列句は一般に「と」の省略と考えないかもしれないが，機械によって他の助詞の省略と区別が付きにくく，機械処理を行う際に問題となると思われるため，省略の一種として扱った．

表1より，「は」の省略が最も多いが，比率から言えば「を」と「と（並列）」のほうが脱落しやすい．また，表の右側からは，格としては「が」と「を」が際だって脱落しやすい助詞と言える．また，「が，を，に」の必須格になりやすい格助詞と，潜在的に必須格の機能を持っている「は」を合わせると助詞落ち全体の約80%占める．一方，「まで，から」などの任意格になりやすい格助詞はほとんど脱落していない．

3.2 係り先までの距離

本節では名詞文節と係り先の文節との距離を分析する．ここで，距離とはいくつ先の文節あるいは述部に係っているかを意味している．表2は助詞落ち文節と名詞文節全体の係り先までの文節数をまとめたものである．表の列は係り先までの文節数を表す．「なし」の欄は述部の省略などによって係り先がない名詞文節の数である．また，斜線の左側が助詞落ち文節の数，右側が全体（助詞落ち文節を含む）の数である．助動詞「だ・です」を持つ名詞文節は表に含めなかった．

すべての助詞を合わせたものと「は，が，を」についての距離の分布グラフを図1に示す．実線が助詞の落ちていない文節を含めたものの距離で，点線が助詞落ち文節の係り距離である．「は，が」については，省略された場合のほうが，若干近くに係る傾向がある

表2 名詞文節の係り先までの文節数
Table 2 Number of phrases between modifying noun phrase and modified phrase.

助詞	1	2	3	4以上	なし	合計
は	36/149	12/115	8/69	7/75	0/31	63/439
が	21/300	2/65	0/23	0/14	0/17	23/419
を	35/175	7/26	1/12	0/7	0/8	43/228
に	8/267	0/56	1/15	1/30	0/15	10/383
へ	−/7	−/2	−/1	−/0	−/0	−/10
で	1/148	1/49	0/19	0/28	0/33	2/277
から・まで	0/64	0/23	0/14	0/10	0/6	0/117
と（引用）	0/94	1/5	0/1	0/1	0/8	1/109
と（並列）	16/81	1/11	0/2	0/4	0/0	17/98
の（連体）	9/609	0/16	0/2	0/4	0/16	9/647
その他	0/433	0/98	0/78	0/69	0/157	0/835
合計	126/2327	24/466	10/236	8/242	0/291	168/3562

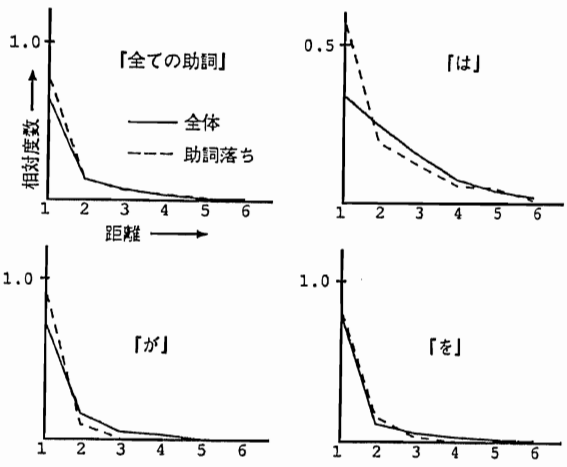

図1 係り先の距離と助詞の関係
Fig. 1 The relationship of number of phrases between modifying noun phrase without post-position and modified verb phrase for every omitted post-position.

が，助詞すべてで比較した場合，特に変わりはない．しかし，助詞落ち文節は対話文全体の特性を保存していると言えるから，この近くに係るという情報は助詞落ち文節の解析に利用することができる．

助詞落ちの文節は，直後にかかる可能性が最も大きく，全体の75%を占める．2つの文節までだと，89%になる．また，副詞を飛び越えて2つ先の文節に係る文節は「直後」に係ると言うことにすれば，直後に係る文節は83%を占める．

助詞が省略された述部に係る名詞文節を対象に，係り先の述部までの述部の数をまとめたものが表3である．例えば，「2」のところは，2つ目の述語に係っていることを示している．ここで，形式名詞[6]「こと」や「もの」につながる「いう」は独立した文節と数えず文節の一部とした．例えば，「研究センターというもの」は1つの文節として，この文節の「いう」は述部の数に入っていない．

この表より，最も近い述部に係る場合がほとんど(99%)であることがわかる．2つ目以降の述部に係る例は2例あるが，3つ先の述部に係っている例は「新富士まで^v値段^v(は)，新幹線(に)乗ろうと思って

表3 係り先の述部までの述部の数
Table 3 Number of verb phrases between modifying noun phrase and modified verb phrase.

述部の数	1	2	3	4以上	合計
文節数（全体）	2100	118	21	11	2250
文節数（助詞落）	140	1	1	0	142

いるんですが，ᵛどれくらいかかります？［6.2モーラ/秒］」の「値段」が「かかります」に係っている例である．「新幹線(に)乗ろうと思っているんですが」の部分は，途中で条件を言い忘れていることに気づいて，あわてて挿入した文と考えられる．この特殊な例を除くと1つの例だけが2つ先の述部に係っていることになる．このように2つ以降の述部に助詞落ち文節が係るのはきわめて希である．

3.3 文頭の名詞文節の助詞落ち

表2と図1より，係り助詞「は」は，もともと遠くに係る性質を持っており，「は」が省略された文節も遠くに係りやすいことが言える．また，「は」は格助詞の持たない主題提示という文法上の機能を持っているため「は」の省略に対する情報は重要である．この点に関して，省略された名詞文節が文頭である場合，「は」が省略されている可能性が高いという分析結果が得られた．

述部に係る文頭の名詞文節に関する分析結果を表4に示す．ここでいう文頭とは文の中で最初に出てきた名詞であることを意味する．距離は係り先までの文節の数である．この表から，文頭の名詞文節の約1割の助詞が省略されることが分かる．また，省略された助詞の内，2/3が「は」の省略であり，遠くに係りやすいことが言える．

3.4 係り先の述部の必須格と脱落した助詞の関係

述部に係っている文節の脱落した格助詞と述部（正しい係り先）の必須格との関係を調べたものが表5である．「は」についても格を持っているものはこの表の中に含まれている．「残り必須格」の欄の数字は述部の必須格の数から，他の文節の修飾によって埋まった必須格の数を引いたものである．脱落した助詞が必須格である場合とそうでない場合を区別してある．また，括弧の中は単語の単純な意味の制約によって文脈によらずに省略された格を推定できる場合の数である．

表4 文頭の名詞文節の係り距離（文節数）
Table 4 Number of phrases between the most left noun phrase and modified phrase.

助詞	1	2	3	4以上	合計
は	11/46	6/44	5/32	3/26	25/148
が	3/48	0/14	0/5	0/3	3/70
を	6/37	3/6	0/1	0/1	9/45
に	0/32	0/12	0/5	0/9	0/58
で	0/36	0/18	0/3	0/8	0/65
合計	20/199	9/94	5/46	3/47	37/386

表5 係り先の必須格と脱落した格の関係
Table 5 The relation between inner case in modified verb phrase and omitted post-position.

残り必須格		0	1	2	3	4
脱落格	必須格	—	70 —	52 (43)	7 (5)	0 (0)
	任意格	5 (5)	0 (0)	3 (0)	0 (0)	0 (0)

る．任意格の場合の意味による推定は表1から脱落する可能性のある「を，に，で」のうちのどれかを意味によって決定できる場合である．表5から，脱落するのは必須格がほとんどであり（94%），その約半数について1格1文節の制限によって，意味を使わずに格を推定できることを示している（残り必須格が1の場合）．さらに，残りの半数は単語の意味と選択制限を使うことによってほとんど（残りの81%）推定できる．任意格が省略されることは少ないが，省略された場合，残り必須格が0の場合は「が，を，に，で」の中から意味制限によって推定できる．係り先が正しければ，全体で結局90%の格を正しく推定できる．推定できない場合はいずれも，文脈処理などのより高度な処理が要求される．

3.5 名詞に係る助詞落ち名詞文節

述部としての名詞文節（「名詞＋です」など）に係る文節を除いて，その他の名詞に係る助詞落ちは「の」と「と」である．

「の」の省略の認定は，複合名詞をどう扱うかによって曖昧になる．例えば，「大阪近郊」は「大阪の近郊」の「の」が省略されているとも，複合した1つの名詞とも考えられる．このように判断が曖昧になりやすいため，明らかに「の」を入れたほうがよい場合にのみ「の」の省略と判断した．例えば，「そこ（の）あたりに」などである．明らかという制限を付けたため分析の結果は9例しかない．「の」の省略を処理する場合の問題は，「の」の省略を認めるために，助詞の省略された名詞が「の」が省略されたものなのか，次の述部に係る格助詞の省略なのかを一見して見分けられない点である．この点に関して，「の」の省略された文節が最も近くの述語に係ることができるかどうか分析した．結果は次のようになった．

述部の必須格で係ることができる：2

述部の必須格で係ることができない：7

必須格で係れないというのは，次のような例である．

「情報工学に関する科目(の)どちらかを受ける．

[9.1モーラ/秒]」

「の」が省略された文節は適当な助詞（上の例では「を」）を補って述部（「受けて」）に係ることができるが，その助詞は次の文節「どちらかを」が使用している．述部には1つの格に対して1つの文節しか係れないという原則によって，「を」を補って係ることはできない．また，もう1つの必須格である「が」は意味制約によって係ることができないので，結局「科目」は受けるに係ることができない．上記の必須格で係れない例はすべてこのような例である．

例が少ないので，確実なことは言えないが，「の」の省略された文節は次の述部に係れないことが多いので，解析が失敗してから次の名詞に係る可能性を検討すればよいと言える．

並列の「と」に関しては，「と」の省略された名詞文節と係り先の名詞文節がすべて，明らかに同じ種類の名詞であった．例えば，「囲碁部，合唱部が…」，「国語と社会，数学，理科…」などである．これは，名詞の概念階層を作成し，2つの名詞の1つ上の親概念が等しいとき「と」の省略の可能性を検討すればよいことを意味する．

係り先に関しては，「の」が省略された文節はすべて次の名詞文節に係っていた．「と」については，その88% (15/17) が次の名詞文節に係っていた．

4. 倒　置

倒置の分析の問題点は倒置の認定は文の認定基準からの影響を大きく受けることである．倒置された文（節）の前で文がいったん切れて，倒置された文節に続く文が省略されたものと考えると倒置でなくなる．判断基準は若干曖昧であるが，省略と考えた場合その省略部分が前の文と同じ内容になってしまう場合を倒置と認定した．例えば，「昔，ヾ遊廓だったところです．ここは．[7.0モーラ/秒]」の「ここは」以下を省略と考えると文脈から前の文と同じものが省略されていると考えられるので倒置と認定する．

倒置された文節の認定は，1文内で後ろから前に係っているものであるとした．**表6**に倒置された文節の分析結果を示す．表6の行は倒置された文節の種類，列はいくつ前の文節に係っているかを表している．丸括弧内の数値は句全体を1つの単位と見れば直前の文節に係っていると見なせる倒置の数である．例えば，「京都奈良が多いんですか，他の学校とかも．」は「学校とかも」が「多いんですか」に係っている倒置であるが，その間に「他の」があるため2つ前に係っていることになる．しかし，「他の学校とかも」全体で直前に係っていると考えたほうが自然である．

表6 倒置の分類
Table 6 The classification of inversion.

倒置された文節の種類	直前	2つ前	3つ前	合　計
「は」で係る文節	5	1 (1)	0	6
「が」で係る文節	0	0	0	0
「を」で係る文節	0	0	0	0
「に」で係る文節	1	1 (1)	1	3
「で」で係る文節	3	0	1	4
「まで・から」で係る文節	2	1〈 1〉	0	3
その他の助詞で係る文節	2	2 (2)	0	4
副詞文節	7	1 (1)	0	8
用　言	1	1 (1)	0	2
「の」で名詞に係る文節	0	2 (1)	0	2
合　計	21	9	2	32

また，かぎ括弧“〈 〉”の中の数字は複数の倒置があるため，直前でなくなった場合の数である．例えば，「1時間半くらいですか，ヾ新富士まで豊橋から．[9.5モーラ/秒]」の「豊橋から」は，「新富士まで」があるので2つ前の文節に係っていることになっている．

表6と倒置の例から言えることは次のとおりである．

(1) 述部に係る文節がほとんどである（「の」以外のすべて）(94%)．

(2) 1つ前の文節に係る文節が多い (66%)．また，表中の丸括弧とかぎ括弧の中の数値を直前に係るものとすれば，91% が直前の文節に係っていることになる．

(3) 倒置された文節が係る述部はかならず終止形で文法的に正しく文を終えている．

そのほかに特徴的な文として，倒置されてかつ，助詞が省略されたものが1文あった．それは，「何でしたっけ，ヾ名前．[12.3モーラ/秒]」である．この例は対話文として自然である．次章では，このような文も解析できる助詞落ち，倒置のためのヒューリスティックスを考察する．

5. 解析手法

5.1 解析の流れ

一般の構文解析システムによる名詞文節の解析は，名詞文節の助詞と係り先の格構造，概念構造などの組み合わせによって行われる．助詞が省略された時に可

能な助詞をすべて解析の候補とすると，係り先・係り関係が極端に曖昧になる．また，倒置についても，無制限に後ろから前への係り受けを認めると，曖昧性が爆発することは明らかである．このため，なんらかの方法によって候補を制限する必要がある．

この章では3，4章での分析を元に，曖昧さをあまり増やさず，助詞落ち名詞文節と倒置文節の係り先と係り関係を推定するヒューリスティックスを考察する．解析の大まかな流れは次のようになる．

（1） 標準的な係り受け解析システムに助詞落ちのヒュースティックスのみを入れて解析する．
（2） (1)で解析に成功したら終了．
（3） (1)で解析に失敗したら倒置の可能性があるとして解析をやり直す．

第1段階の解析が失敗してから倒置の解析ヒューリスティックスを入れるのは，効率のため（曖昧性が増加する）と，倒置には部分的な解析結果を参照するヒューリスティックスがあるためである．以下，それぞれのヒューリスティックスについて述べる．

5.2 助詞落ち文節解析のためのヒューリスティックス

助詞落ち文節解析のためのヒューリスティックスは係り先に関するものと，係り関係に関するものがある．以下に提案するヒューリスティックスを示す．

HJ1 助詞が省略された名詞は最も近くの述部に係る．
HJ2 述部に係る場合は，必須格を候補として考える．
HJ3 文頭の助詞落ち名詞文節には「は」を補った文節も文節切り出し結果の1つとして追加する．
HJ4 述部を飛び越さない次の名詞文節に「の」で係ることができる．ただし，これは述部へ係ることができない場合に限る．
HJ5 助詞の省略された名詞が，次の名詞文節と，概念階層上で同じ1つ上の親概念を持つ場合，並列の「と」の省略として係ることができる．

HJ1は，表3の結果（98%の述部に係る助詞落ち文節が最も近い述部に係る）から得られる．HJ2は3.4節の，HJ3は3.3節の議論から，HJ4とHJ5は3.5節の議論からそれぞれ得られる．HJ3によって，文頭の助詞落ち名詞文節は2つ以上離れた述部に係ったり，主題の「は」の省略として扱われることが可能となるが，HJ3が役立つ例は今回の分析対象には少なく，曖昧性も多くなるため次節の評価では使用しなかった．しかし，文頭の助詞落ち名詞文節は「は」の省略である可能性が高く，「は」は遠くに係る性質を持っているため，大規模なタスクでは役立つ可能性もあると予想される．

助詞が省略された場合の解析手法についてはすでにいくつかの研究があるが[7),8)]，いずれもメカニズムの提案であり，統計的な助詞落ちの特性までを考慮したものにはなっていない．このため，非現実的な文を扱うこともできるが，曖昧性などの点で問題がある．われわれのヒューリスティックスには助詞落ちに関する統計的な特徴が取り入れられているため，現実的な文に関しては，これまでの助詞落ちを解析するメカニズムと組み合わせることによって，より強力なシステムを構築できる．

5.3 倒置文節解析のためのヒューリスティックス

5.1節で述べたように，倒置の解析ヒューリスティックスは第1段階の解析が失敗してから導入される．このため，解析対象とされている文の部分的な解析結果が得られていると仮定する．倒置の解析はまず，どの部分解析木が倒置でないものかを決定し，残りの部分を倒置部分として後ろから前に係ることを許した係り受け解析を行うことによって実行される．倒置でないものを決定し，係り先を決定するヒューリスティックスは次のものである．

HT1 文の先頭を含み，終止形の述部で終わる最も長い部分解析木から順番に倒置でない部分の候補とする．
HT2 任意の部分解析木は直前（文の左隣）の部分解析木に係ることができる．

HT1は倒置部分は直前に係り，係り先は終止形の述部で終わっているという4章の分析による．係り先に関するヒューリスティックスがHT2である．4章の分析より，2つの文節が倒置されていて直前に係ることにならない倒置（第4章に例がある）も，直前の述部に係ることのできるほうの文節が直前の述部に係った解析結果を1つの述部（部分解析木）とすれば，HT2を2回適用することにより解析できる．また，連体修飾文節を伴う倒置も，連体修飾された部分解析木を1つの単位とすることによりHT2で解析できる．このようにすれば，第4章で分析したように91%の倒置が解析できると思われる．

6. 評価実験

6.1 実験システム

前節で述べたヒューリスティックスの有効性を調べるために構文解析の実験を行った．実験のために作成した構文解析システムは，交通規則文の解析のためにわれわれが以前作成したもの[9]を改良したものである．改良点，特徴，およびシステムの規模を以下に述べる．

（1） 図2に構文解析システムの構成を示す．改良点は文節解析規則を有限状態オートマトンで記述するようにした点と，係り受け解析をチャートパーザ[10]風に部分解析結果を保存しながら行う点である．係り受け解析はまずチャートデータベースから2つの部分解析木を選び，その2つの部分解析木が係り受けできるかどうかを調べ，できる場合には新しい部分解析木をチャートに加えるようになっている．解析に失敗したときにも，チャートに部分解析木が残っていることを利用して，倒置の解析対象を決定できるようにした．また，係り受け解析の各段階で助詞落ち・倒置の解析ヒューリスティックスが働くようになっている．

（2） ヒューリスティックスHJ1（最も近くの述部に係る）を適用するために，係り先との間に述部があるかどうかをチェックする必要がある．このため，文の部分解析結果を表現するデータ構造をそれが複数の述部を持つかどうかを表せるように拡張した．述部がある文節に係ると，結果としての部分解析結果には述部が係っていることを示すデータが付加される．さらに，このデータはこの部分解析結果が他の文節に係るときに伝播される．倒置解析のヒューリスティックスがない限り，文節は前から後ろにしか係らないので，このデータによって最も近くの述部であるかどうかを判別できる．

（3） 意味表現として平井の3つの視点からの意味マーカ[11]を使用している．

（4） 名詞が483，動詞が89，形容詞・形容動詞が27，その他が93の大きさのエントリを持つ辞書を使用した．

この構文解析システムに助詞落ち・倒置解析ヒューリスティックスを入れたもので実験を行った．

6.2 実験結果

実験対象としての文は分析に使用した例と評価用に新たに日本音響学会連続音声データベースの書き起こしテキストから抽出した文を用いた．分析に使用した例文から，助詞落ち文節を含む文は50文，倒置文は30文（2文は倒置を2か所持っている）すべてを使用した．新たな評価用文としては，助詞落ち文を50文，倒置文を27文使用した．合計，助詞落ちに関して100文，倒置に関して57文で実験を行った．助詞落ちの文については，解析が困難であると予測される文（2つ以上遠くの述部に係る文など）も，助詞落ち全体の割合と同じように分析の対象とした．また，対象が音声対話文であるため，助詞落ちや倒置以外の書き言葉と異なるところ，また非常に長い文が多く，これらに関しては本論文の主題ではないので，本システムの構文解析部で解析できない部分は解析できるように実験用データとして157文中132文について文を修正した．例えば次のように変更した．

変更前：「その下に名前を書くようになってるけど，ˇ誰も名前(を)かいてねえんじゃねえかなあ．[9.4モーラ/秒]」

変更後☆：「その下に名前を書くようになっているが，

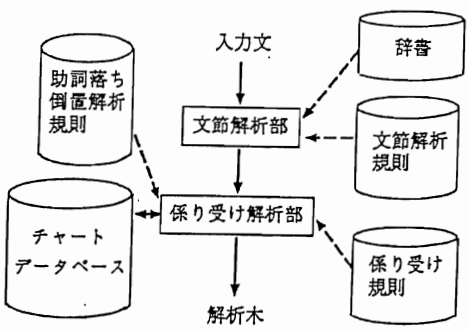

図2 実験用文解析システム
Fig. 2 The parsing system for the experiment.

表7 ヒューリスティックスによる助詞落ち・倒置文の解析結果
Table 7 The parsing results of spoken sentences by using our heuristics.

		解析成功						解析失敗	
曖昧さ		−2以下	−1	0	1	2	3以上	解析不能	誤解析
助詞落	分析文	0	3	36	4	1	1	2	3
	評価文	1	1	39	4	0	0	0	5
倒置	分析文	—	—	27	—	—	—	1	2
	評価文	—	—	24	—	—	—	2	1

☆ 変更はやや不自然かもしれないが，本論文のヒューリスティックスが文節内の解析から影響を受けることはほとんどないので，評価にはさしつかえない．

誰も名前(を)書いてないだろう.」

表7は実験結果である.解析に成功,または失敗したもので大きく分けた.解析結果は曖昧性によっていくつか出てくるが,その結果の中に正しい解析結果が含まれていれば成功とし,その助詞落ちに関係ある解析結果の曖昧さの数(係り先あるいは格の曖昧さの数)でさらに分類した.曖昧さの数は正しい助詞を補った文を解析し,そのときの曖昧さの数を基本とし,それより多くなったか少なくなったかで,プラスあるいはマイナスになっている.曖昧さの数が0の場合は助詞落ち文節の助詞が正しい係り先と共に一意に決定できたことを意味している.「は」については,格を持っているものはその格を正しく推定できれば成功とした.

失敗に関しては,解析結果が得られなかった場合と,結果は得られたがすべて誤っていた場合に分類した.また,「分析文」と書いてある行が分析に使用した文に対する,「評価文」と書いてある行が評価用に新たに抽出した文に対する実験結果である.分析文80文,評価文77文に対する実験結果に有意な差はなかった.

助詞落ちに関して,提案したヒューリスティックスが単純なものにも関わらず多くの助詞を一意に正しく決定できていることが分かる(75%).また,助詞が省略されたために,係り距離が短いというHJ1のヒューリスティックスによって,助詞があるときよりも曖昧さが減った例が5個あった.このように助詞落ちの情報を積極的に利用する使用法もあることが分かる.また,曖昧さが残ってしまうものに関しても解析結果に優先順位などをつけることによって,さらに改良できる.例えば,「が」と「に」の曖昧性がある場合には「が」のほうが落ちる可能性が高いので「が」を優先するなどである.例えば「いっぱいスキー場(が,に)あります.[12.1モーラ/秒]」の場合,「が」と「に」の曖昧性があるが優先順位によって「が」のほうを選択することができる.実際,文脈なしの状態で10人の人に聞いたところ全員が「が」の省略と答た.この現象に関しては,省略の優先順位や必須格などの議論との関連もあるだろうが,本論文の範囲を越えるので,詳しい研究は今後の課題としたい.

解析失敗に関しては,2つ先の述部に係るものと任意格の省略された例が解析不能で,「の」の省略と任意格の省略された例が誤解析であった.

倒置に関しては,4章の分析からも分かるように,倒置された文節を1つにまとめても,2つ以上前の述部に係る倒置がある.ヒューリスティックスはそれを無視しているため,そのような例は解析できなかった.しかし,89%の文は解析に成功している.

以上,正しく解析された例をいくつか挙げる.括弧内は助詞落ちに関して正しいと予想される助詞(すなわち推定された助詞),倒置に関しては,正しいと考えられる文の一例である.

助詞落ち

・先ほど挙げた3つのやつ(の)値段(を)1つずつ教えて下さい.

・他の人はどういう所(に)行くのですか?

助詞落ちと倒置の複合

・何ですか,名前(は).(名前はなんですか?)

倒置

・クラスの数はいくつ位あるのですか,一学年で.
(一学年で,クラスの数はいくつ位あるのですか.)

・一人3万5千円です,安くても.
(安くても,一人3万5千円です.)

次に解析に失敗した例を示す.[ ]の中は解析できなかった理由である.

助詞落ち

・そちらの方(を),先ほど申し込みました日程で申し込みたいんですけど.
[正しい係り先が2つ先の述部であるが,1つ目の述部「先ほど申し込みました」に係ってしまった(HJ1)]

・十万弱(で)往復は買える.
[十万弱を副詞としてとってしまい,「十万弱枚,往復を買える」という意味に解析してしまった]

倒置

・安定感があるとは言いますね,長い方が.
[正しくは2つ前の述部に係るべきだが,1つ前の述部「言います」に係ってしまった(HT2)]

7. おわりに

本論文では,音声から書き起こしたテキストを分析することによって,名詞文節の助詞の脱落の特徴を調べた.音声対話文では名詞文節の約4%において,助詞が脱落していた.構文・意味解析は助詞の情報に大きく依存しているため,大きな問題点であるといえる.述部に係る文節は最も近い述部に係る可能性が高いことと,必須格の助詞が落ちる場合が多いことが分かった.さらに,倒置文節についても調査し,倒置文

節は係り先の述部に近いところに存在していることが分かった.

これらの結果から，付属語の脱落・倒置を含む文の係り受け解析を行うための単純なヒューリスティックスを提案した．ヒューリスティックスはかなり荒い近似であるが，実際の対話でててくる文に対しては有効であることを実験により確認した.

今後の課題は，大規模なタスクに対しても助詞落ち・倒置を解析できるようにするために，分析対象を増やして一般性を上げ，より精密な助詞落ち・倒置のモデルを作成することと，今回，分析の対象から除いた言い淀み，言い直し，間投詞などの取扱いを検討することである．応用としては，音声認識のための情報としてこれらのデータが役立つかどうかを確認し，実際の音声対話システムに組み込むことによって，これまでよりもより自然な発話を扱える対話システムを構築することを計画している.

謝辞 日本音響学会連続音声データベースを使用させていただきました．作成された方々にお礼を申し上げます.

参考文献

1) 有田，小暮，野垣内，飯田：メディアに依存する会話の様式—電話会話とキーボード会話の比較—，情報処理学会研究報告，87-NL-61-5 (1987).
2) 井ノ上，江原，小倉：係り受け関係データから見たキーボード会話と電話会話の比較，第40回情報処理学会全国大会論文集（I），pp. 490-491 (1990).
3) 山本，小林，中川：対話における質問応答対の入れ子構造の分析，1991年電子情報通信学会秋季大会，D-54，pp. 55-56 (1991).
4) 保坂，竹沢，江原：対話データベースを利用した音声認識のための構文規則，情報処理学会研究報告，91-NL-83-13，pp. 97-104 (1991).
5) Itahashi, S.: Creating Speech Corpora for Speech Science and Technology, *Trans. IEICE*, Vol. E 74, No. 7, pp. 1906-1910 (1991).
6) 山口：体言，岩波講座 日本語 6-文法I，pp. 129-168, 岩波書店 (1976).
7) 中村，上原，豊田：非文法的な日本語文を取り扱う意味主導型解析メカニズム，情報処理学会研究報告，89-NL-70-8 (1989).
8) 柿ケ原，相沢：対話文における誤入力訂正処理，情報処理学会研究報告，87-NL-64-8，pp. 61-68 (1987).
9) 中川，竹本，田口：交通規則文に関する質問応答システム LICENCE における日本語文から一階述語論理への変換，情報処理学会論文誌，Vol. 32, No. 3, pp. 354-363 (1991).
10) Gazdar, G. and Mellish, C.: *Natural Language Processing in LISP*, pp. 181-214, Addison-Wesley Publishing Company (1989).
11) 平井，北橋：日本語解析システム MARION-IV における単文の構文および意味解析について，情報処理学会論文誌，Vol. 27, No. 9, pp. 892-899 (1986).

（平成4年2月24日受付）
（平成4年9月10日採録）

山本 幹雄（正会員）

昭和59年豊橋技術科学大学情報工学課程卒業．昭和61年同大学大学院修士課程修了．同年(株)沖テクノシステムズラボラトリ入社．昭和63年豊橋技術科学大学情報工学系教務職員．自然言語処理，人工知能に関する研究に従事．電子情報通信学会，人工知能学会，AAAI, ACL各会員.

小林 聡（学生会員）

1966年生．1991年豊橋技術科学大学情報工学課程卒業．現在，同大学研究科修士課程情報工学専攻在学中．自然言語処理の研究に従事．人工知能学会学生会員.

中川 聖一（正会員）

昭和51年京都大学大学院博士課程修了．同年京都大学情報助手．昭和55年豊橋技術科学大学情報工学系講師．昭和58年助教授．平成2年教授．工学博士．昭和60～61年カーネギー・メロン大学客員研究員．音声情報処理，自然言語処理，人工知能の研究に従事．昭和52年電子通信学会論文賞受賞．著書：「情報基礎学詳説」（分担執筆，コロナ社），「確率モデルによる音声認識」（電子情報通信学会），「音声・聴覚と神経回路網モデル」（共著，オーム社）など．電子情報通信学会，日本音響学会，人工知能学会，IEEE, INNS各会員.