

係り受け解析における韻律情報の利用

The Use of Prosodic Information in Dependency Structure Analysis

電気通信大学 電気通信学部

Faculty of Electro-Communications, The University of Electro-Communications

尾関 和彦

Kazuhiko Ozeki

Natural language processing has traditionally relied solely on information extracted from written materials. Recently, as part of natural language processing and speech processing merge into spoken language processing, a new possibility is opening up to exploit prosodic information, which is lost when utterances are transcribed into letters or characters, for language processing. Such a possibility is worth pursuing from both scientific and technological points of view. This report focuses on syntactic information contained in prosodic features extracted from read Japanese sentences, and describes a method of exploiting it in dependency structure analysis. The basic idea here is to make a statistical model of feature distribution for each dependency distance. Then, by using the Bayes theorem, the probability of dependency distance for each phrase is calculated given a prosodic feature value associated with the phrase. The first point of this work is concerned with the use of pause information around a phrase in question. Three types of pauses are employed in this report: pre-phrase pause, post-phrase pause, and post-post-phrase pause. The second point is directed toward extraction and modeling of feature vectors related to the fundamental frequency (F_0) contours. It was shown that the use of multiple pause information is more effective than the use of single pause, say the post-phrase pause. Adding F_0 information to the multiple pause information was effective to improve the parsing accuracy, though the degree was not very large.

Key words: prosody, syntactic structure, dependency analysis, pause duration, fundamental frequency

1 まえがき

発話文の韻律とその統語構造の間には密接な関係のあることが知られている [1]。音声合成においては、統語構造に応じて文音声に適切な韻律を付与することが極めて重要であるため、発話文の統語構造と母音継続長、ポーズ長、基本周波数などとの関係が詳しく調べられている [2, 3]。一方、その逆問題、すなわち、韻律情報を文の統語解析に利用するための研究も行われている。英語に関しては、例えば、韻律情報を文の統語的曖昧性の解消に利用したり、統語解析結果の候補として得られるいくつかの統語構造に韻律的なスコアを付与する研究がある [4, 5]。日本語に関しても、韻律情報に基づいて句間の結合度を定義し、結合度の弱い句境界から順に文を分割して行くことにより、韻律的な句構造解析を行う方法が報告されている [6]。また、韻律情報に基づいて、ある句が直後の句に係るか否かを判定する研究も行われている [7]。日本語に対するこれらの研究は、もちろん、韻律情報から統語構造を知ることに関連している。しかし、韻律情報だけを用いて統語解析を行うことは現実には考えられない。従って、

種々の言語的な拘束条件や言語情報を用いた通常の統語解析において、更に韻律情報を追加したとき、解析精度がどの程度向上するかが問題であろう。しかし、統語解析器に韻律的知識を直接組み込んで利用する研究はこれまで殆んどなかった。

本研究では、統語解析の手法として「係り受け解析」と呼ばれている方法を採用し、その解析アルゴリズムに韻律的知識を組み込むことにより、どの程度解析精度が向上するかを調べた。韻律的特徴量としては、予備実験の結果から有効性が確認された文節間ポーズ長と基本周波数を取り上げた。

2 係り受け解析

2.1 係り受け構造

日本語は膠着語の特徴を有する言語であり、1つあるいは複数の自立語の連鎖の後に、0個以上の付属語の連鎖が接続して「文節」と呼ばれる単位を構成する。日本語文はこのような文節の列である。日本語文の統語構造を文節間の広義の修飾・被修飾関係に基づいて捉えようという考え方が古くからある

[8]. このとき修飾文節を「係り」文節といい、被修飾文節を「受け」文節という。そして、1つの文の中で、2つの文節が修飾・被修飾関係にあるとき、それらは「係り受け」関係にあるという。文が与えられたとき、その中の文節間の係り受け関係が分かればその文の統語構造が分かったと考える。これは「依存構造解析」とも呼ばれ、日本語以外の言語に対しても研究が行われている。日本語の統語構造を、このように2文節間の係り受け関係によって捉えることには異論もあるが、語順の自由度が大きい日本語文の構造を記述するには便利な考え方であり、よく用いられている。

日本語の平叙文においては、文節間の係り受けに次のような制約がある：

- 後方唯一性：文末以外の文節は、それより文末側にある唯一の文節に係る。
- 非交差性：係り受けは交差しない。すなわち、文節列 $\dots x \dots y \dots z \dots w \dots$ において、 x が z に係り、 y が w に係ることは許されない。

文節列に対して、文節間に上の2つの条件を満たす係り受け関係を指定したものを「係り受け構造」という。係り受け構造は「括弧付け」とも密接な関係があり、次のように再帰的に定義することができる：

- (1) x が文節ならば、 (x) は係り受け構造である。
- (2) X_1, X_2, \dots, X_r が係り受け構造で、 x が文節ならば、 $(X_1 X_2 \dots X_r x)$ は係り受け構造である。
- (3) 上の(1)と(2)を満たす記号列のみが係り受け構造である。

係り受け構造 $(X_1 X_2 \dots X_r x)$, $X_i = (\dots, x_i)$ においては、 x_i が x に係ると考える。文節列が与えられたとき、それに適切な括弧付けをして係り受け構造になるようにしたものを、その文節列上の係り受け構造という。例えば、

- ((この)(花は)(とても)美しい)
- (((この)花は)(とても)美しい)
- (((この)(花は)とても)美しい)
- ((この)((花は)とても)美しい)
- (((((この)花は)とても)美しい)

などは全て文節列 /この/花は/とても/美しい/ の上の係り受け構造である。長さ n の文節列上の可能な係り受け構造の総数 $D(n)$ はカタラン数

$$D(n) = \frac{1}{n} {}_{2(n-1)}C_{n-1}$$

で与えられることが知られている [9]。

2.2 係り受け解析アルゴリズム

1つの文の係り受け構造は、どの文節がどの文節に係るかを指定することによって定まる。従って、日本語文を文節の列として $w_1 w_2 \dots w_m$ と表すと、その係り受け構造は各文節の番号 i にそれが係る文節の番号 $S(i)$ を対応させる写像

$$S : \{1, 2, \dots, m-1\} \longrightarrow \{2, 3, \dots, m\}$$

を与えることによって決まる。このとき、後方唯一性と非交差性は、次のように表現することができる：

- $\forall i \in \{1, 2, \dots, m-1\} : i < S(i)$
- $\forall i, j \in \{1, 2, \dots, m-1\} :$
 $i < j \Rightarrow [S(i) \leq j \text{ or } S(j) \leq S(i)]$

このような写像 S は、 $m \geq 3$ ならば複数個存在する。従って、係り受け構造を決定するためには、更に何らかの拘束条件を導入し、候補を絞る必要がある。古典的な係り受け解析では、文節を構成する形態素情報に基づいて、文節 w_i が文節 w_j に係ることが許されるかどうかを判定する規則を設けることにより、候補を絞ることが行われていた [10]。しかし、この方法では多くの曖昧性が残り、候補を少数に絞ることが難しい。近年では、文節 w_i が文節 w_j に係ることが許される程度を「整合度」、「ペナルティ」、「選好度」などの実数値で表すことが行われるようになってきている [11]。この方法では、2文節間の「係り受けペナルティ」の総和が最小になる（あるいは「整合度」や「選好度」の総和が最大になる）ような係り受け構造を求める。すなわち、文節 w_i が文節 w_j に係るときの係りにくさをペナルティ関数 $F(w_i, w_j)$ で表すとき、その総和

$$\sum_{k=1}^{m-1} F(w_k, w_{S(k)}) \quad (1)$$

を考える。ペナルティ関数 $F(w_i, w_j)$ には、文節 w_i, w_j が持つ形態素情報のような記号で表される情報、 w_i, w_j の間の文節間距離のような整数値で表される情報、あるいは、そのような係り受け文節対

が出現する確率や韻律的特徴量のような実数で表される情報など、種々の異なるタイプの言語的知識を組み込むことができる。

係り受け解析は、式 (1) の最小値を与える係り受け構造 \hat{S} ，すなわち

$$\hat{S} = \operatorname{argmin}_S \sum_{k=1}^{m-1} F(w_k, w_{S(k)})$$

を求めることによって行う。この計算は、動的計画法の原理に基づいて効率よく実行することができる [12]。本研究では、この「総ペナルティ最小化法」を基本的な手法として用いている。

3 音声言語資料と韻律的特徴量

音声言語資料として ATR 音声データベース セット B [13] を用いた。このデータベースには、新聞、雑誌、小説、手紙、教科書等の出版物から抽出された 503 文のテキストと、それを 10 名の話者が読み上げた音声データが含まれている。テキストには、音素表記、文節境界などの言語ラベルが付けられ、また、音声には音素とその始点・終点、ポーズなどに関する詳細なラベルが付けられている。

まず、係り受け解析に有効な特徴量を見出すため、種々の韻律的特徴量を取り上げて予備的な係り受け解析実験を行った [14]。そこで取り上げた特徴量は、着目文節の直後のポーズ長、着目文節と直後の文節の基本周波数 (F_0) 曲線から得られる特徴量、着目文節と直後の文節のパワー曲線から得られる特徴量など多種類に上るが、その中で着目文節の直後のポーズ長が最も明確な有効性を示した。そこで本研究では、まず最初に着目文節の直後のポーズ長の利用を試み、それを出発点として、逐次取り上げる特徴量の種類を増加して行った。

4 ポーズ情報の利用

4.1 係り受け距離とポーズ長

日本語文を文節列 $w_1 w_2 \dots w_m$ で表すとき、 $j - i$ ($j > i$) を w_i と w_j の文節間距離という。また、 w_i が w_j に係るとき、 w_i の係り受け距離は $j - i$ であるという。係り受け解析は、どの文節がどの文節に係るかを決定することであるから、文末を除く各文節の係り受け距離が決まれば、係り受け構造が決定する。従って、各文節の係り受け距離と関連のある特徴量は全て係り受け解析に有効な情報を担って

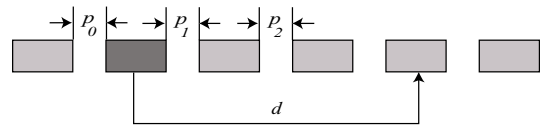


図 1: 第 0 ポーズ長 p_0 、第 1 ポーズ長 p_1 、第 2 ポーズ長 p_2 と着目文節の係り受け距離 d 。黒い箱は着目文節を示す。

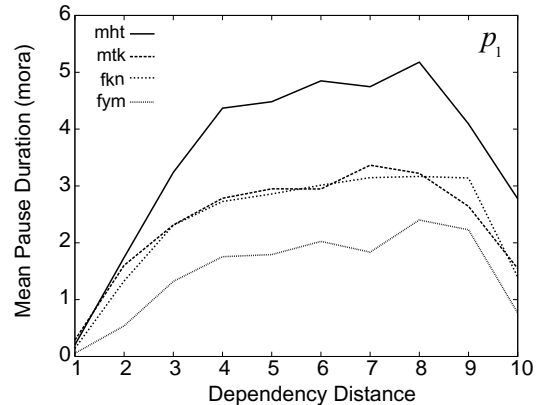


図 2: 4 名の話者に対する係り受け距離 d と第 1 ポーズ長平均値の関係。ポーズ長は当該発話の平均モーラ長を単位として表した。

いる可能性がある。本研究では、まずポーズ長がそのような情報を持っているかどうかを調査した。図 1 に、ここで取り上げたポーズの種類を示す。 p_1 は着目文節の直後のポーズ長であり、最初に取り上げたポーズである。これを、ここでは第 1 ポーズ長と呼ぶ。 p_2 は着目文節の直後の文節の直後のポーズ長であり、第 2 ポーズ長と呼んでいる。また、 p_0 は着目文節の直前のポーズ長であり、第 0 ポーズ長と呼んでいる。着目文節とは、その係り受け距離を問題にしている文節である。

図 2 はデータベース中の 4 名の話者 mht, mtk, fkn, fym について、係り受け距離に対する第 1 ポーズ長の平均値の変化を示したものである。ポーズ長は平均モーラ長を単位として表されている。話者によって傾斜は異なるが、 d が 4 程度までは第 1 ポーズ長平均値が d の増加と共にほぼ線形に増加することが分かる。係り受け距離と第 1 ポーズ長の間にはこのように密接な関係がある。係り受け距離 d が 4 より大きくなるとポーズ長平均値は飽和する。 d が 8 を越えるとポーズ長平均値は減少しているが、この部分はデータ数が少ないため結果の信頼性が低い。

図 3 は、同じ 4 名の話者について、係り受け距離に対する第 2 ポーズ長の平均値の変化を示したものである。 $d = 2$ のとき、第 2 ポーズ長平均値が極端に小さくなっているが、これは、そのとき、着目文

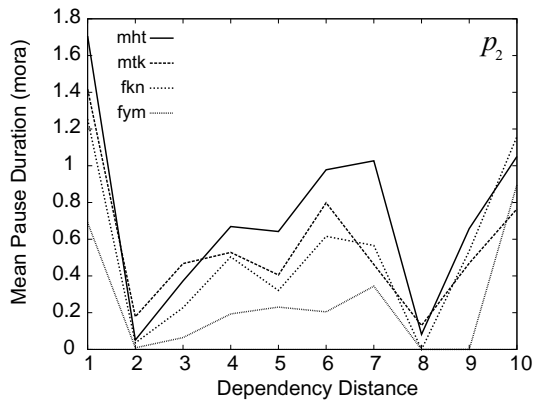


図 3: 4 名の話者に対する係り受け距離 d と第 2 ポーズ長平均値の関係。ポーズ長は当該発話の平均モーラ長を単位として表した。

節の直後の文節の係り受け距離が 1 となり、その直後のポーズ長が小さくなることによるものと説明できる。 d が 2 以上のところでは d が増加するに従って第 2 ポーズ長も増加する傾向が見られる。このように、第 2 ポーズ長も着目文節の係り受け距離 d と関係があるので、係り受け解析に有効な情報となる可能性がある。

図 3 は、同様に係り受け距離 d に対する第 0 ポーズ長の平均値の変化を示したものである。 d の増加とともに第 0 ポーズ長平均値は減少する傾向が見られる。ただし、 d が大きいところではデータ数が少ないので、結果は信頼できない。第 0 ポーズ長は着目文節の前に現れるポーズであるから、それが着目文節の係り受け距離と関係を持つことは一見不思議に思われる。しかし、第 0 ポーズ長は着目文節の直前の文節の係り受け距離と関係があり、また、係り受けの非交差性によって、着目文節の直前の文節の係り受け距離と着目文節の係り受け距離 d とは無関係ではない。これによって、 d と第 0 ポーズ長に関連が生ずるものと思われる。

以上のように、着目文節の周辺のいくつかのポーズ長はどれも着目文節の係り受け距離と関連を持っている。従って、どれもが係り受け解析に有効な情報となり得ることが分かった。

4.2 ポーズ情報に基づいたペナルティ関数

4.1 節で述べた係り受け距離 d とポーズ長の関係をどのように利用すればよいであろうか。 d とポーズ長の間に関係があると言っても、それは統計的な関係に過ぎない。そこで、本研究では着目文節に関わる韻律的特徴ベクトル p が与えられたとき、その

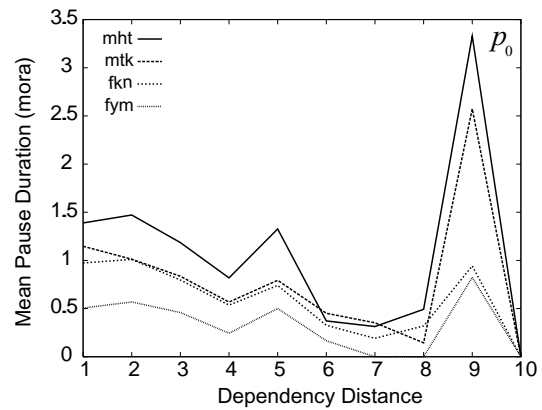


図 4: 4 名の話者に対する係り受け距離 d と第 0 ポーズ長平均値の関係。ポーズ長は当該発話の平均モーラ長を単位として表した。

文節の係り受け距離が d となる確率 $P(d | p)$ を推定し、これに基づいてペナルティ関数を定めることを考えた。 $P(d | p)$ を直接推定することは難しいが、 $P(p | d)$ は、係り受け距離毎に韻律的特徴量 p のデータを集め、何らかの分布関数を仮定することにより、推定することができる。そして、Bayes の定理

$$P(d | p) = \frac{P(p | d)P(d)}{\sum_a P(p | d)P(d)} \quad (2)$$

により $P(d | p)$ を求めることができる。ここで、 $P(d)$ は d の事前確率であり、学習データ中の係り受け距離が d である文節数の相対頻度から推定できる。本研究では $P(p | d)$ として正規確率密度関数を仮定し、学習データから平均と分散（あるいは共分散行列）を推定している [15]。係り受け距離の分布 $P(d)$ の情報だけでも係り受け解析に有効であることが知られている [16] が、ここでは、更に韻律情報が係り受け距離の分布に与える影響を利用する。

本研究は、韻律情報のみを用いて係り受け解析を行うことを目指しているわけではない。むしろ、通常の係り受け解析の中で、韻律情報がどの程度解析精度を向上させる働きがあるかを見ようとしている。通常の係り受け解析では、2 つの文節の形態素情報から、一方が他方に係り得るか否かを判定する規則（係り受け規則）を定め、それに基づいて係り受け構造を定める。ここでは、[17] の考え方に基づいて、人手で作成した規則 [16] を用いた。この規則では、係り受け文節の末尾の形態素を、連体詞、活用語の連用形、活用語の基本型・タ形、副詞、助詞“の”、助詞“が、に、より”、助詞“へ”、助詞“を”、その他の助詞、に分類し、また、受け文節の先頭の形態素を、名詞、名詞 + 判定詞、動詞、述語のイ・ナ形

表 1: 係り受け距離の頻度分布

距離	頻度	相対頻度 (%)	累積分布 (%)
1	1889	64.6	64.6
2	487	16.7	81.3
3	243	8.3	89.6
4	136	4.7	94.3

容詞, 連体形のイ形容詞, 連用形のイ・ナ形容詞, その他, に分類している. そして, それぞれの組合わせに対して係り受けが許されるか否かを記述している. 実際にはこの規則を受け文節の, 現実に現れる種々の変形に対処できるように補強している. 例えば「二時半ごろだった」(数詞 + 接尾辞 + 名詞 + 名詞 + 助動詞 + 助動詞)は, 名詞 + 判定詞として取り扱われるようになっている.

この規則によって, 文節 w_i が文節 w_j に係ることが許されるとき, 係り受けペナルティ関数 $F(w_i, w_j)$ を

$$F(w_i, w_j) = -\log P(j-i | \mathbf{p}) \quad (3)$$

と定義する. また, w_i が w_j に係ることが許されないときは $F(w_i, w_j) = \infty$ (実際には大きな実数値)とする. これが, 本研究で用いた基本的なペナルティ関数であるが, 複数のポーズ情報を用いるときは後で見るように修正を加える.

4.3 実験条件

3 節に述べたように, 本研究では ATR 音声データベースセット B を用いている. 総文数は 503 であり, 1 文当りの平均文節数は 6.8 である. また, 係り受け文節対の総数は 2923 である.

係り受け距離は表 1 に示すような分布を持っており, 距離が大きくなるに従って, 急激にその頻度は減少する. 実際, 距離 3 以下の係り受け文節対が全体の約 90% を占め, 距離 4 までに約 94% が存在することが分かる.

このデータベース中の文は, A から J までの 10 グループに分けられている. A から I までの 9 グループにはそれぞれ 50 文が含まれ, グループ J には 53 文が含まれている. 本研究では, これを学習データと評価データに分割する仕方を変えて, 表 2 のように 3 種類のデータセットを作成した.

係り受け解析の結果は, 文正解率で評価した. これは, 解析結果として得られる係り受け構造がデー

表 2: 学習データと評価データ

セット	学習データ	評価データ
I	D-J (353 文)	A-C (150 文)
II	A-G (350 文)	H-J (153 文)
III	A-C, G-J (353 文)	D-F (150 文)

タベースのラベルと一致する評価文数の, 全評価文数に対する割合である. また, 実験はすべて話者に関して閉じている. すなわち, 一人の話者の学習データを用いて種々のパラメータ推定を行い, その結果を用いて同一話者の評価データに対する文正解率を求めている. これを各話者に対して行う.

4.4 複数のポーズ情報を用いた係り受け解析

第 1 ポーズ長 p_1 のみを用いるときのペナルティ関数の定義は, 式 (3) において $\mathbf{p} = p_1$ としたもので:

$$F(w_i, w_j) = -\log P(j-i | p_1) \quad (4)$$

を用いた.

2 種類以上のポーズ長を併用するときは,

$$\mathbf{p} = (p_1, p_2), \quad \mathbf{p} = (p_0, p_1, p_2)$$

のような韻律的特徴ベクトルを考え, 多次元の同時分布を考えるのが基本的な方法と思われる. しかし, この方法を実際に試みたところ, 良い結果は得られなかった. そこで, p_1 と p_2 を併用するときは $P(d | p_1)$ と $P(d | p_2)$ を別々に求め, ペナルティ関数 $F(w_i, w_j)$ を

$$F(w_i, w_j) = -\{\alpha \log P(j-i | p_1) + (1-\alpha) \log P(j-i | p_2)\}$$

と定義した. ここで $0 \leq \alpha \leq 1$ は第 1 ポーズ情報と第 2 ポーズ情報の貢献度を調節する係数である. ただし, w_i と w_j が係り受け関係にあることが規則によって許されない場合は, $F(w_i, w_j) = \infty$ とする.

同様に, p_0, p_1, p_2 を併用するときは,

$$F(w_i, w_j) = -\{\alpha_0 \log P(j-i | p_0) + \alpha_1 \log P(j-i | p_1) + (1-\alpha_0-\alpha_1) \log P(j-i | p_2)\}$$

と定義した.

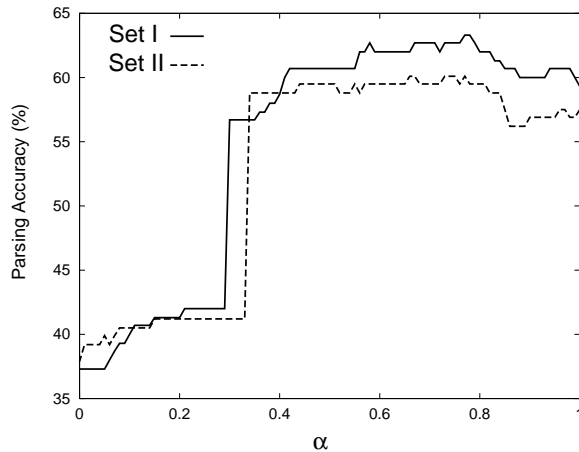


図 5: 係数 α と文正解率の関係．話者は mht. データは セット I, セット II.

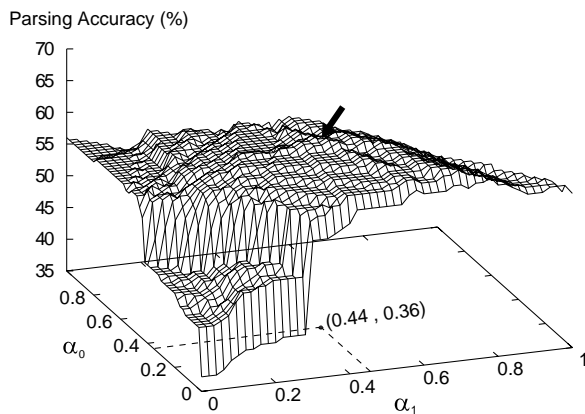


図 6: 係数 α_0, α_1 と文正解率の関係．話者は mht. データは セット I.

データベースには 10 名の話者が含まれているが、本研究ではその中の 4 名、mht, mtk, fkn, fym を使用した。図 5 は、第 1 ポーズ長 p_1 と第 2 ポーズ長 p_2 を併用したときの、係数 α に対する文正解率の変化の一例を示す。話者は mht である。 $\alpha = 0$ のときは第 2 ポーズのみを用いた場合になり、 $\alpha = 1$ のときは第 1 ポーズのみを用いた場合となる。この図が示すように、 α を 0 から増加させて行くと、あるところで文正解率が急激に上昇し、1 に近づくとまた少し下降する。このように、第 1 ポーズ情報と第 2 ポーズ情報がある割合で併用すると、それぞれを単独で使用する場合より文正解率が向上する。

図 6 は、第 0 ポーズ長 p_0 、第 1 ポーズ長 p_1 、第 2 ポーズ長 p_2 を併用したときの、係数 α_0, α_1 に対する文正解率の変化の一例を示す。話者は mht である。 $(\alpha_0, \alpha_1) = (0.0, 0.0)$ のときは第 2 ポーズのみを用いた場合になり、 $(\alpha_0, \alpha_1) = (1.0, 0.0)$ の

ときは第 0 ポーズのみを用いた場合になる。また、 $(\alpha_0, \alpha_1) = (0.0, 1.0)$ のときは第 1 ポーズのみを用いた場合になる。図では少し分かりにくいだが、 $\alpha_0 + \alpha_1 \leq 1$ の領域にしかグラフは存在しない。この例では、 $(\alpha_0, \alpha_1) = (0.44, 0.36)$ において、文正解率が最大になっている。このように、3 種類のポーズ情報を併用することの効果が見られる。

得られた文正解率を表 3 に示す [19]。これは、 α や α_0, α_1 などを各実験条件で最適値に定めた場合の結果で、セット I, II, III に対する文正解率の平均である。また、“Det” は決定論的解析法 [17] において 4.2 節に述べた係り受け規則を用いたときの文正解率であり、韻律情報は使用していない。この表から分かるように、どの話者に対しても、第 1 ポーズの情報に第 2 ポーズの情報を加えることにより、文正解率が向上する。また、それに第 0 ポーズの情報を加えることによって、さらに文正解率が向上することが分かる。第 0 ポーズの効果は大きいとは言えない。しかし、ここでの係り受け解析が着目文節の係り受け距離の確率分布を利用しており、着目文節は常にそれより文末側の文節に係ることを考えると、着目文節が発声される前のポーズである第 0 ポーズの情報が係り受け解析に有効であることは、興味深い事実と言えよう。

表 3 は係り受け距離毎の係り受け正解率をセット I の場合について示したものである。係り受け距離が 2 までは、ポーズ情報を追加するほど係り受け正解率が向上する。しかし、3 以上では、係り受け正解率は向上しないか、むしろ低下する傾向が見られる。表 1 に示したように、係り受け距離は短いところに大きく分布している。従って、係り受け距離 2 までの係り受け正解率の向上が文正解率の向上に寄与したと思われる。ただ、これまでの研究で、係り受け正解率と文正解率の間には必ずしも単調な関係があるとは限らない [14] ので、今後は、個々の文の個々の文節について、どのような場合にポーズ情報が有効に働いたかを調査する必要がある。

5 基本周波数情報の利用

5.1 基本周波数情報に基づいたペナルティ関数

音声の基本周波数 (F_0) は人にとって非常に重要な韻律情報であり、これが係り受け解析においてどの程度有効な情報となるかは興味ある問題である。 F_0 情報を取り扱う上で難しい点は、それが時間の

表 3: ポーズ情報を用いたときの文正解率 (%) . 数字は, セット I, II, III に対する結果の平均である . Det は決定論的解析法において, ここでの係り受け規則を用いたときの 503 文に対する文正解率を示す .

条件	mht	mtk	fkn	fym	Av.
p_1	58.0	57.8	55.4	54.5	56.4
$p_1 p_2$	60.7	59.2	57.6	56.5	58.5
$p_0 p_1 p_2$	61.4	59.6	58.9	57.1	59.3
Det					49.5

表 4: 係り受け距離毎の係り受け正解率 (%) . セット I の場合を示す .

d	Cond.	mht	mtk	fkn	fym
1	p_1	94.5	94.5	94.9	96.3
	$p_1 p_2$	94.3	94.9	95.1	96.5
	$p_0 p_1 p_2$	94.9	95.2	95.4	96.5
2	p_1	82.6	86.3	81.4	74.5
	$p_1 p_2$	87.0	88.2	85.7	83.2
	$p_0 p_1 p_2$	87.6	88.2	86.3	81.4
3	p_1	89.2	86.5	86.5	83.8
	$p_1 p_2$	89.2	83.8	86.5	81.1
	$p_0 p_1 p_2$	89.2	85.1	87.8	83.8
4	p_1	80.0	80.0	73.3	76.7
	$p_1 p_2$	76.7	73.3	76.7	70.0
	$p_0 p_1 p_2$	76.7	76.7	76.7	73.3

関数であり, そこからさらに何を抽出するかが問題となることである . ここでは, F_0 曲線の上昇・下降を表すために最低 3 点が必要と考え, 図 7 に示すように, 各文節の F_0 曲線に 2 次回帰曲線をあてはめて, そこから 3 点の対数 F_0 値を抽出した [18] . 基本的には文節の両端点と中点における対数 F_0 の値であるが, 本当の端点における値は誤差が大きい可能性があるため, 10% 内側の点を使っている . また, 重要なのは着目文節とその直後の文節の F_0 の相対関係であると考え, 着目文節と直後の文節から合計 6 点の対数 F_0 の値 $f_1, f_2, f_3, f_4, f_5, f_6$ を図のように抽出した . さらに, 全体の上昇・下降に依存しない特徴量を得るため, f_2 を基準点としてそこからの差のみを考え,

$$\mathbf{f} = (f_1 - f_2, f_3 - f_2, f_4 - f_2, f_5 - f_2, f_6 - f_2)$$

を F_0 に関わる韻律的特徴量とした . そして, ポーズ情報と併せ, 係り受けペナルティ関数を

$$F(w_i, w_j)$$

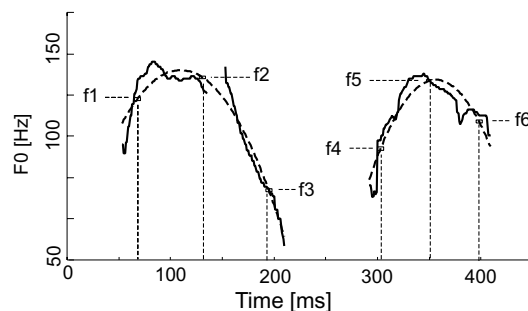


図 7: F_0 特徴ベクトルの抽出

$$= -\{\beta \log P_p(j-i | p_0, p_1, p_2) + (1-\beta) \log P_{F_0}(j-i | \mathbf{f})\}$$

と定義した [18, 20] . ここで,

$$\begin{aligned} \log P_p(j-i | p_0, p_1, p_2) &= -\{\bar{\alpha}_0 \log P(j-i | p_0) \\ &+ \bar{\alpha}_1 \log P(j-i | p_1) \\ &+ (1-\bar{\alpha}_0-\bar{\alpha}_1) \log P(j-i | p_2)\}. \end{aligned}$$

ただし, $\bar{\alpha}_i$ はポーズ情報のみを用いて係り受け解析を行ったときの最適な係数である . また, β は事後的に最適化した値を用いた .

w_i が w_j に係ることが許されないときは

$$F(w_i, w_j) = \infty \quad (5)$$

とした .

5.2 係り受け解析結果

表 5 に文正解率を示す [20] . 結果はデータセット I と II に対する文正解率の平均である . 話者 mtk については F_0 データがデータベースに含まれていなかったため, 実験を行っていない . このように, ポーズ情報に F_0 を併用することにより, 文正解率の向上が見られる . しかし, その程度は大きいとは言えない .

表 5: 3 種類のポーズ情報に F_0 情報を追加したときの文正解率 (%) . セット I とセット II に対する文正解率の平均を示す .

条件	mht	fkn	fym	平均
$p_0 p_1 p_2$	62.7	61.7	60.4	61.6
$p_0 p_1 p_2 F_0$	63.7	62.0	61.1	62.3

6 あとがき

係り受け解析における韻律情報の利用に関して、着目文節の周辺の3種類のポーズ長、および着目文節とその直後の文節の F_0 曲線から得られる特徴量の有効性について述べた。

本研究には多くの課題が残されている。例えば、ポーズ長が正規分布に従うことを仮定しているが、現実のポーズ長の分布は正規分布と大きく掛け離れている。従って、ポーズ長分布をよりよく近似する関数を見い出す必要がある。また、 F_0 曲線は複雑であり、そこから係り受け解析に有効な特徴量を抽出する方法について更に検討する必要がある。また、韻律情報の効果がより敏感に測定できるような解析結果の評価法についても考える必要がある。このような問題のいくつかに対しては、別の分担者が多数話者・話者非依存の条件下で検討している [21]。

7 謝辞

本研究は長期に渡って多くの人との共同で行われたものである。電気通信大学助手 高木一幸氏はもとより、著者の研究室に在籍した学生諸氏、特に大学院生であった張玉潔、江口徳博、高坂和之、廣瀬幸由、久保田新、沖本真美子、呂美蓉の諸氏には、その協力に対して感謝の意を表したい。

参考文献

- [1] T. Uyeno, H. Hayashibe, K. Imai, H. Imagawa, and S. Kiritani, "Syntactic structure and prosody in Japanese: a study on pitch contours and the pauses at phrase boundaries," Annual Bulletin of Research Institute of Logopedics and Phoniatrics, University of Tokyo, Vol.15, pp.91-108, 1981.
- [2] 海木延佳, 匂坂芳典, "局所的な句構造によるポーズ挿入規則化の検討," 電子情報通信学会論文誌, Vol.J79-D-II, No.9, pp.1455-1463, 1996.
- [3] 海木延佳, 匂坂芳典, "局所的句構造に基づく F_0 制御," 電子情報通信学会論文誌, Vol.J83-D-II, No.9, pp.1853-1860, 2000.
- [4] P. J. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong, "The use of prosody in syntactic disambiguation," J. Acoust. Soc. Am., Vol.90, pp.2956-2970, 1991.
- [5] M. Ostendorf, C. W. Wightman, N. M. Veilleux, "Parse scoring with prosodic information: an analysis/synthesis approach," Computer Speech and Language, Vol.7, pp.193-210, 1993.
- [6] 小松昭男, 大平栄二, 市川薫, "韻律情報を利用した構文推定およびワードスポットによる会話音声理解方式," 電子情報通信学会論文誌, Vol.J71-D, No.7, pp.1218-1228, 1998.
- [7] 関口芳廣, 鈴木良弥, 菊川智之, 高橋安子, 重永実, "韻律情報を利用した連続音声の中の隣接句間の修飾関係の有無の判定," 電子情報通信学会論文誌, Vol.J78-D-II, No.11, pp.1581-1588, 1995.
- [8] 橋本進吉, 国語法要説, 明治書院, 1934.
- [9] 張玉潔, 日本語文の係り受け解析における統計情報の利用法に関する研究, 電気通信大学電気通信学研究科博士論文 (1999).
- [10] 吉田将, "二文節間の係り受けを基礎とした日本語の構文分析," 電子通信学会論文誌, Vol.55-D(4), pp.238-244, 1972.
- [11] 江原暉将, "最大エントロピー法を用いた日本語文節間係り受け整合度の計算," 言語処理学会第4回年次大会論文集, pp.382-384 (1998).
- [12] 尾関和彦, "係り受けの整合度に基づき最適文節列を選択する多段決定アルゴリズム," 電子情報通信学会論文誌, Vol.70-D, No.3, pp.601-609, 1987.
- [13] 阿部匡伸, 匂坂芳典, 梅田哲夫, 桑原尚夫, 研究用日本語音声データベース利用解説書 (連続音声データ編), ATR 自動翻訳電話研究所, 1990.
- [14] 廣瀬幸由, 尾関和彦, 高木一幸, "日本語読み上げ文の係り受け解析における韻律の特徴量の有効性," 自然言語処理, Vol.8, No.4, pp.71-89, 2001.
- [15] 江口徳博, 尾関和彦, "韻律情報を利用した係り受け解析," 日本音響学会誌, Vol.52, No.12, pp.973-978, 1996.
- [16] 張玉潔, 尾関和彦, "文節間係り受け距離の統計的性質を用いた日本語文の係り受け解析," 自然言語処理, Vol.4, No.2, pp.3-19, 1997.
- [17] 黒橋禎夫, 長尾眞, "並列構造の検出に基づく長い日本語文の構文解析," 自然言語処理, Vol.1, No.1, pp.35-57, 1994.
- [18] K. Takagi, H. Kubota, and K. Ozeki, "Combination of pause and F_0 information in dependency analysis of Japanese sentences," Proc. IC-SLP 2002, Vol.2, pp.1173-1176, 2002.
- [19] M. Lu, K. Takagi, and K. Ozeki, "Recovery of dependency structure using multiple pause information," Proc. of International Conference: Speech Prosody 2004, pp.513-516, 2004.
- [20] 呂美蓉, 日本語係り受け解析における複数ポーズ情報の利用, 電気通信大学大学院電気通信学研究科情報工学専攻修士論文, 2004.
- [21] K. Takagi, K. Ozeki, "Dependency analysis of read Japanese sentences using pause information: A speaker independent case," Proc. of International Conference: Speech Prosody 2004, pp.595-598, 2004.