

係り受け解析における韻律情報の利用：多数話者による評価

The use of prosodic information in dependency analysis of Japanese sentences: evaluation by numerous speakers

電気通信大学 電気通信学部

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

高木 一幸

Kazuyuki Takagi

Prosody and syntax are significantly related with each other. In the field of speech synthesis, many efforts have been made to control prosody so that it reflects the syntactic structure of the sentence. However, the inverse problem, recovery of syntactic structure using prosodic information, has not been so much investigated. We have been working on a method that incorporates prosodic information into a dependency analysis parser as linguistic knowledge and exploit it in the search process. The basic idea is to make a statistical model of prosodic feature distribution for each dependency distance. Then, by using the Bayes theorem, the dependency distance of each phrase is predicted from a given feature value. Duration of pauses at phrase boundaries and relative fundamental frequency (F_0) contour features have been found to be effective for parsing in our previous works, in which evaluations were made in speaker-dependent conditions. In this paper, effectiveness of pause and F_0 information was examined in speaker-independent conditions by using prosodic features extracted from the spoken version of sentences uttered by 44 speakers. By simplifying the estimation of both pause and F_0 contour models, better performance was obtained. It was shown that the effectiveness of pause information was larger when pause models were estimated separately for zero-duration and non-zero-duration pauses. Linear combination of pause and F_0 information gave significant improvements in parsing accuracy. The optimum values of combination weight were different for different speakers, which may be caused by some speaker characteristic concerning the production of prosody.

Key words: prosody, syntax, pause duration, fundamental frequency, minimum penalty parser, dependency analysis

1 はじめに

韻律には様々な情報が含まれており、構文構造との関係に着目した研究がこれまでに多数行われている [1, 2, 3, 4, 5, 6, 7]。この中には日本語の韻律と構文構造を取り扱った研究 [1, 2, 3, 5, 7] もあるが、実際に韻律情報を構文解析アルゴリズムに組み込んで利用してはいない。我々は、以前から、総ペナルティ最小化パーザーを用いた係り受け解析 [8] において韻律情報を利用する方法を提案しており、それによって解析精度が向上することを見出している [13]。

さまざまな韻律特徴量とそれらの組合せの有効性について調査を行った結果 [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]、これまでに取り上げた韻律特徴量の中では、特に、着目文節直後のポーズ継続長、および着目文節と次文節の基本周波数 (F_0) 軌跡の概形の相対値が係り受け解析において有効であることがわかっている。

しかし、先行研究においては、韻律特徴量の分布関数の推定は話者毎に行われており、学習と評価に

用いた話者の数は最大 4 名と限られている。今後、音声理解に本手法を用いる場合には、分析対象話者の十分な量の韻律データを事前に利用できるとは限らないため、不特定話者に対応する必要がある。

したがって、最終年度では、多数の話者の音声データを用いて話者に依存しないモデルを作成し、不特定話者の条件下で実験を行った。また、利用できる文の種類はこれまでと同様、ATR 音素バランス 503 文に限られているが、学習用と評価用の文セットの組合せの種類を大幅に増やし、より信頼性の高い評価が出来るようにした [24, 25]。この実験条件の下で、ポーズ長分布のモデルとして文献 [17] で有効性を報告した方法についても再検証した。その際、不特定話者の音声データの解析を行う際に必要と考えられるポーズ長の正規化尺度について検討した。また、不特定話者の条件下でポーズ情報と F_0 情報を結合して利用する実験も行い、ポーズ情報と F_0 情報を単独で用いる場合に比べて良い性能が得られること、話者によってポーズ情報と F_0 情報の効果に差があることがわかった。

2 係り受け構造

係り受け解析とは、文中の各文節がどの文節に係る(どの文節を修飾する)かを定めることであり、日本語の構文構造は、係り受けを定めることにより記述することができる。

日本語の文は文節 w_i が連なったもの $w_1 w_2 \dots w_m$ とみることができる。文の長さは文中の文節数で表すことにする。例えば、文節数が m 個であるとき、文長は m であるという。文節 w_i が文節 w_j を修飾するとき、 w_i は w_j に係り、 w_j は w_i を受けるといい、そのときの係り受け距離 d は $d = j - i$ ($i < j$) で定義する。 w_i を係り文節、 w_j を受け文節と呼ぶ。このような修飾関係を係り受け関係という。

係り受け構造 S は、係り文節番号から受け文節番号への写像と考えることができる(式1)。

$$S: \{1, 2, \dots, m-1\} \rightarrow \{2, 3, \dots, m\}. \quad (1)$$

日本語の構文的性質により、係り先は後方の唯一の文節であり、係り受け関係は交差しないことから、写像 S は、以下の性質を持つ。

後方唯一性 $\forall i \in \{1, 2, \dots, m-1\} : i < S(i)$

非交差性 $\forall i, j \in \{1, 2, \dots, m-1\} :$

$$i < j \Rightarrow [(S(i) \leq j \text{ or } S(j) \leq S(i))].$$

これに加えて、係り受けが存在し得るためには、係り文節と受け文節を構成する形態素の品詞や意味などの素性の間に、整合性がなければならない。

本研究で用いている係り受け解析システム [8] においては、後方唯一性と非交差性の条件は、解析アルゴリズムに組み込まれている。整合性条件としては、文献 [11] を参考にして作成し、これまでの研究 [13, 15, 18] において改良された係り受け規則によって与えている。

一般に、ある文が与えられたとき、上記の条件を満たす係り受け構造は複数存在し、解析結果に曖昧性が生じる。この曖昧性を解消するために、整合性条件として音声の韻律情報も利用することが、この研究の目的である。

3 係り受け解析システム

3.1 総ペナルティ最小化法

我々が用いている総ペナルティ最小化法 [8] では、係り文節 w_i が受け文節 $w_{S(i)}$ に係るときの困難さ

を、式2のペナルティ関数 $F(w_i, w_{S(i)})$ で表す。

$$F(w_i, w_{S(i)}) = \begin{cases} -\log P(d | \mathbf{p}), & (w_i, w_{S(i)}) \in DR \\ \infty, & \text{それ以外の場合} \end{cases} \quad (2)$$

$(w_i, w_{S(i)}) \in DR$ は w_i と $w_{S(i)}$ が係り受け規則 DR を満たすことを表す。 $\mathbf{p} = (p_1, \dots, p_n)$ は w_i に対する n 個の韻律パラメータを成分とするベクトル、 $P(d | \mathbf{p})$ は \mathbf{p} が与えられたときに、 w_i と $w_{S(i)}$ の係り受け距離が d となる条件付き確率である。そして、唯一性、非交差性を満たす係り受け構造の中で、このペナルティ関数によって計算されたペナルティの総和

$$\sum_{i=1}^{m-1} F(w_i, w_{S(i)}) \quad (3)$$

が最小となる係り受け構造 S を解析結果とする。

3.2 ペナルティ関数

ペナルティ関数 $F(w_i, w_j)$ は、係り受け距離 d と着目文節(係り文節) w_i に関する韻律特徴量の間の統計的關係によって定義される [13]。 \mathbf{p} が与えられたときの d の条件付き確率 $P(d | \mathbf{p})$ は、直接求めることはできないが、ベイズの定理を使って式4により推定することが出来る。

$$P(d | \mathbf{p}) = \frac{P(\mathbf{p} | d)P(d)}{\sum_d P(\mathbf{p} | d)P(d)} \quad (4)$$

すなわち、 $P(d | \mathbf{p})$ は、学習用データから求めた $P(\mathbf{p} | d)$ と $P(d)$ を用いて計算することができる。ここで、 $P(d)$ は $P(d) = N_d / \sum_d N_d$ 、 N_d は学習データに出現する係り受け距離 d の文節の総数である。なお、本稿では、確率密度関数 $P(\cdot)$ として、Gauss 分布を用いている。

4 韻律情報

4.1 ポーズ継続長 x

4.1.1 係り受け距離に関する情報

係り文節直後のポーズ継続長 x 、すなわち文節 w_i と文節 w_{i+1} の間のポーズ長(以下、単にポーズ長と呼ぶ)は、係り受け距離と関係がある。図1は、ATR データベース B セット [9] に収録されている話者10名による音素バランス503文の朗読音声データから得た、ポーズ長の平均値と係り受け距離の関

係を示したグラフである．図 1 を見ると，係り先が遠くなるにつれて，ポーズ長が大きくなるのが分かる．特に，係り受け距離が 1 から 4 までの範囲では，ポーズ長の平均値は，距離に比例して増大している．

係り受け距離が 4 を超える範囲では，この傾向ははっきりせず，係り受け距離が 9 と 10 の場合はポーズ長が減少しているように見える．これは，表 1 に示す通り，4 を超える係り受け距離のデータは非常に少なく，信頼出来る統計が得られないことによる．

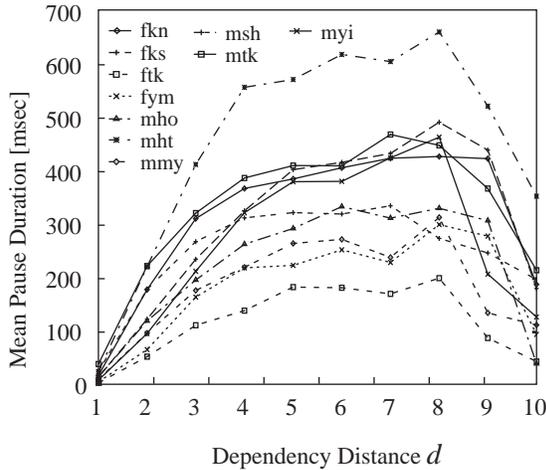


図 1: 文節直後のポーズ長 x の平均値の係り受け距離による変化 (ATR セット B の話者 10 名)

表 1: 係り受け距離頻度 (ATR503 文)

係り受け距離	1	2	3	4
頻度 N_d	1909	500	253	126
相対頻度 $P(d)$ (%)	65.3	17.1	8.7	4.3

5	6	7	8	9	10	計
73	35	13	9	3	1	2922
2.5	1.2	0.4	0.3	0.1	0.0	100.0

4.1.2 ポーズ長分布関数

図 2 は，ポーズ長のヒストグラムを模式化したものである．前述したように，ポーズ長は係り受け距離の情報を持つが，その分布はこの図のように，ポーズが無い場合の頻度が高く，ポーズが有る場合は連続的な分布をしている．

この分布を適切に近似する分布関数として，これまで，正規分布 [13, 14, 15]，ポアソン分布，指数分布，相対頻度分布 [18] を用い，係り受け解析におけ

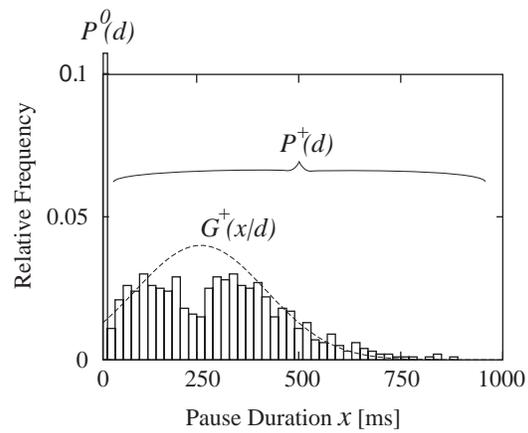


図 2: ポーズ長ヒストグラムの模式図

る有効性を検討してきた．しかし，先行研究 [18] によって，正規分布以外の分布関数を用いても，正規分布を用いた場合に比べて解析精度は同等であることが分かっている．

これらの関数を推定する際には，ポーズが無い場合はポーズ長が $x = 0$ のデータであるとして，ポーズが有る場合のデータと一括して取り扱った．しかし，ポーズ分布関数を，図 2 のように， $x = 0$ である確率 $P^0(d)$ と， $x > 0$ である確率 $P^+(d)$ および確率密度分布 $G^+(x | d)$ に分けて推定することにより，解析精度が向上することが確かめられている [17]．

4.1.3 分布関数の簡単化

図 3 は，実験 (後述) に用いた多数話者について，係り受け距離毎のポーズ長 x を単一 Gauss 分布とみなしてモデル化したときの，距離毎の分布の形である．これからわかるように，係り受け距離が 2 以上の場合は分散値が大きく，距離 3 以上の Gauss 分布はほとんど重なっている．われわれのこれまでの研究では，最大 10 までの全ての係り受け距離について別個の Gauss 分布を用いているが，再考の余地がある．そこで，ポーズ長分布の学習の際に，ある一定の距離以上のデータについては同一カテゴリとして扱うようにした場合に，係り受け解析性能がどうなるかを調べた．

表 2 は，ATR データベースのセット B の話者 10 名のデータを用い，話者については closed (特定話者)，文については open (後述) の条件で，同一カテゴリとして扱う係り受け距離の範囲を変えて求めた文正解率である．文正解率は，評価文のうち，文構造が正解ラベルと完全に一致したものの割合である．

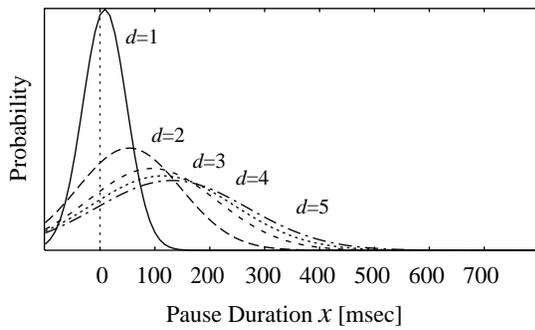


図 3: ポーズ分布確率密度関数の例 (多数話者)

$d \geq 3$ の場合, すなわち $P(d | p)$ を $d = 1, d = 2, d \geq 3$ の 3 カテゴリで計算した場合, 従来に比べて, 必要なパラメータは 3 分の 1 以下に減少でき, かつ, 正解率は僅かではあるが向上した. したがって, 本稿の実験ではこの条件を用いることにした.

表 2: ポーズ分布一括推定距離と文正解率 (%)

一括距離	$d \geq 2$	$d \geq 3$	$d \geq 4$	$d \geq 5$	従来
文正解率	53.67	53.86	52.95	52.91	52.91

4.1.4 ポーズ分布の分割モデル化

ポーズ分布を, ポーズ無しの確率とポーズ有りの場合に分割してモデル化する方法について, 文献 [17] で有効性を報告した. このモデルにおいて, $P(x | d)$ は次のように定義される.

$$P(x | d) = \begin{cases} P^0(d), & \text{if } x = 0 \\ P^+(d) G^+(x | d), & \text{if } x > 0 \end{cases} \quad (5)$$

ここで, $G^+(x | d)$ は学習データ中のポーズが有る場合のデータから推定したガウス確率密度関数である. ポーズが無い確率とポーズが有る確率は, 式 6 で推定する.

$$\begin{aligned} P^0(d) &= P(x = 0 | d) = N_d^0 / N_d \\ P^+(d) &= P(x > 0 | d) = N_d^+ / N_d \end{aligned} \quad (6)$$

ここで, N_d^0 と N_d^+ は, それぞれ学習データ中の係り受け距離が d である文節のうち, 直後にポーズが無い文節, 有る文節の数である.

このモデルの推定にあたって, 4.1.3 節のように, 係り受け距離のクラス分けを変えて行った予備実験において, $P(x | d = 1)$ と $P(x | d > 1)$ の 2 クラスで分布をモデル化したとき, 最も良い解析精度が得られた.

4.2 F_0 特徴量 f

4.2.1 特徴量抽出法

基本周波数 F_0 の特徴量もポーズ長と同様に, 時系列パターンを文節毎に区切って計算する. その利用方法として, これまでは, パターンの回帰直線を求め, その傾き, 当該文節の直線の終端の高さと後続文節の直線の始端の高さの差を特徴量とする [13, 14, 15], パターンを 2 本の折れ線によって近似し, その折れ線の端点と傾きについて後続文節との差を用いる [16, 18], パターンを 2 次曲線によって近似し, その曲線上の特徴点について, 後続文節との相対値を用いる [19], などが考案されている. 本稿では, このうち最も良い結果を与えた, 2 次曲線による近似法 [19] を用いた. 以下, この方法による F_0 特徴量の抽出手順を説明する.

まず, 対数 F_0 時系列を文節毎に 2 次曲線で近似し, 得られた曲線に対して, 文節内の有声区間の始端, 終端から 10% 内側の 2 点と中央, 計 3 点の対数 F_0 値を計測する. F_0 パターンに関しては, これら特徴点の相対値が重要であると考えられるので, 着目文節の中央値 f_2 を基準値とした 5 次元のベクトル $f = (f_1 - f_2, f_3 - f_2, f_4 - f_2, f_5 - f_2, f_6 - f_2)$ を着目文節の F_0 特徴量として用いる.

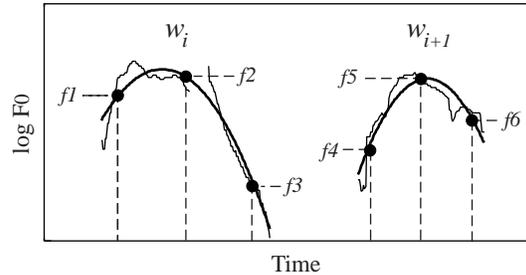


図 4: F_0 特徴量: 着目文節と次文節の F_0 時系列に当てはめた 2 次曲線上の 6 つの特徴点の相対値 $(f_1 - f_2, f_3 - f_2, f_4 - f_2, f_5 - f_2, f_6 - f_2)$ を用いる.

4.2.2 分布関数の簡単化

先行研究では, 5 次元の F_0 特徴ベクトル f を特徴量として用いている. しかし, 係り受け距離との関係が明らかな要素は, 着目文節と直後文節の中央点の差 $f_5 - f_2$ のみである. そこで, 本稿では, 5 次元全ての F_0 特徴量を用いる場合と, 中央点の差のみの 1 次元の特徴量を用いた場合での係り受け解

析精度の比較を行った。

図5は、係り受け距離が1~3までの場合の $f_5 - f_2$ の分布を示したものである。係り受け距離が $d = 1$ の場合は、 $d > 1$ の場合に比べて大きく異なるが、 $d > 1$ では同じような分布をしている。これに関しては、係り受け距離毎にではなく、 $d = 1$ と $d > 1$ の2クラスだけで分布関数の推定を行うことで、解析精度が向上することがわかっている [19]。

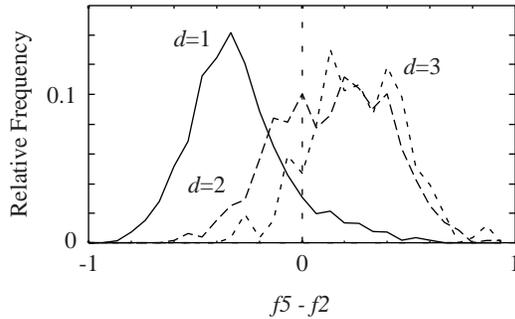


図 5: F_0 特徴量 $f_5 - f_2$ の係り受け距離別分布 (話者: mht)

表3は、5次元の F_0 特徴ベクトルを用いて係り受け距離毎に分布関数を求めた場合 (条件 C_1) と、1次元の F_0 特徴ベクトル ($f_5 - f_2$) を用いて $d = 1$ と $d > 1$ の2クラスだけで分布関数を求めた場合 (条件 C_2) の、係り受け解析性能の比較である。4.1.3節と同様、ATR データベースのセットBの話者10名のデータを用い、話者については closed (特定話者)、文については open (後述) の条件下での結果である。隣接係り受け正解率は、評価文中の文末を除く文節のうち、直後に係るか否か (すなわち $d = 1$ であるか $d > 1$ であるか) が正しく推定された文節の割合である。分布関数の簡単化により、文正解率は1.4ポイント、隣接係り受け正解率は5ポイント向上した。

表 3: F_0 モデルの条件と解析性能 (C_1 : 5次元の F_0 特徴ベクトル, C_2 : 1次元の F_0 特徴ベクトル (距離クラス数=2)).

条件	C_1	C_2
文正解率 [%]	48.8	50.2
隣接係り受け正解率 [%]	87.8	92.8

4.3 ポーズ情報と F_0 情報の線形結合

ポーズ長 x が与えられたときの距離 d の条件付き確率 $P(d | x)$ と、 F_0 特徴ベクトル f が与えられたときの距離 d の条件付き確率 $P(d | f)$ を式7のように線形結合し、ペナルティ関数の計算 (式2) に用いた [22]。

$$\begin{aligned} \log P(d | x, f) \\ \approx \alpha \log P(d | x) + (1 - \alpha) \log P(d | f) \quad (7) \end{aligned}$$

$\alpha = 0$ の場合は F_0 情報のみ、 $\alpha = 1$ の場合はポーズ情報のみを用いることに等しい。 α ($0 \leq \alpha \leq 1$) の最適値は実験的に求めた。

5 多数話者による評価実験

5.1 実験条件

学習および評価は、これまでの研究と同様 ATR 音素バランス 503 文 [9] を用いたが、最終年度においては話者数を大幅に増加し、多数話者のデータを用いた検証を行った。ATR DB セットBからは10名全ての話者 (男6名・女4名)、セットCからはセットBと重複していない34名 (男16名、女18名) の話者の音声データを用いた [25]。ただし、セットCのデータは音素バランス 503 文のうち、文セットA-Cの150文のみである。

基本周波数 F_0 の時系列および音声区間の判定データは、ESPS/waves+ [12] を用いて推定したものを使用した。

学習・評価データの構成は、話者と文の両方に対してオープンな条件にするようにした。すなわち、文に関しては、評価データを連続する3つの文セットとし、残りの7セットを学習用とする。ATR DB セットBの場合、評価データをA-C, B-D, ..., J-Bの如くずらした10通りの条件全てにおいて解析を行ない、結果の平均値を求めた。

話者に関しては、ATR DB セットBの話者を評価する場合は、評価する話者1名以外の9名の話者を学習用とし、ATR DB セットCの話者を評価する場合は、ATR DB セットBの10名全ての話者を学習用とした。

表 4: 係り受け解析の文正解率 (%) (欄中上段は ATR セット B (10 名), 下段は ATR セット C (34 名))

評価文セット	A-C	B-D	C-E	D-F	E-G	F-H	G-I	H-J	I-A	J-B	平均
決定論的	50.0	48.7	46.7	43.3	44.0	38.7	42.0	49.0	54.2	56.2	47.3
ポーズ 正規化無し	56.6 51.5	52.4	51.5	47.2	46.4	42.7	47.3	55.8	62.6	62.6	52.5
ポーズ 平均モーラ長で正規化	56.5 52.7	52.4	50.9	47.1	45.7	42.7	47.1	55.6	62.8	63.0	52.4
ポーズ 平均 CV 音節長で正規化	56.1 51.8	52.7	51.7	47.1	45.6	42.9	47.1	55.8	62.9	63.1	52.5
ポーズ 平均音節長で正規化	57.0 52.9	53.3	51.5	47.1	46.4	43.1	46.9	55.4	62.3	63.3	52.6
ポーズ (分離推定 ・平均音節長で正規化)	58.1 57.4	55.5	52.8	47.1	46.5	42.4	47.4	55.6	62.0	63.9	53.1
F_0 (5 次元)	53.6 53.5	48.3	44.3	39.8	44.4	40.9	46.6	55.2	60.2	61.4	49.5
F_0 (簡単化)	54.0 54.8	52.3	48.6	44.6	44.7	41.4	46.4	53.8	58.1	60.3	50.5
ポーズ+ F_0	58.9 57.8	56.4	54.1	48.9	48.5	44.5	49.0	57.2	63.3	64.2	54.5
ポーズ (分離推定)+ F_0	59.3 58.6	56.6	53.9	48.7	48.0	44.1	49.1	56.6	62.3	64.7	54.3

5.2 結果

5.2.1 決定論的解析法

表 4 は、各条件での文正解率の一覧である。決定論的解析法 [11] では係り受け規則のみを用い、韻律情報は利用していない。これを韻律情報の有効性を検討するためのベースライン (基準値) とする。

係り受け解析性能のベースラインは、評価用文セットにより異なっていることがわかる。この原因は、音素バランス 503 文の各文セットは、構文的・韻律的観点からは片寄りがあるためである。例えば、文長の分布が異なること (長い文ほど解析精度は低い)、係り受け規則の被覆率が異なることなどである。

5.2.2 ポーズ情報の効果

ポーズ情報を用いた場合、全ての評価条件において、文正解率は決定論的解析法に比べて向上している。ATR DB セット C の話者に対する正解率はセット B の話者に比べて低い。ポーズ継続長をモーラ、CV 音節、音節のそれぞれの平均継続長で正規化した場合、僅かな差であるが、正規化しない場合に比

べて、音節数で正規化した場合が最も正解率が得られた。

ポーズの無い場合とある場合にデータを分割してモデル化した場合は、そうでない場合に比べ文正解率が、セット B の場合 (52.6%から 53.1%へ) 0.5 ポイント、セット C の場合 (52.9%から 57.4%へ) 4.5 ポイント向上した。これにより、不特定話者の条件においても、特定話者条件で行った先行研究 [17] の結果が確認された。

5.2.3 F_0 情報の効果

F_0 情報を用いて解析を行った場合は、従来通り 5 次元の F_0 特徴量を用いる場合と、中央点の差のみの 1 次元の特徴量を用いて距離のクラス分けを 2 種類にして簡単化した場合の比較をした。文正解率の平均値は、決定論的解析法に比べ、5 次元の特徴量を用いた場合、セット B の場合は 2.2 ポイント、セット C の場合は 3.5 ポイント向上した。さらに分布関数の推定を簡単化した場合、セット B の場合は 1.0 ポイント、セット C の場合は 1.3 ポイント向上した。5 次元の特徴量を用いた場合、B-D, C-E, D-F の

評価文セットに対しては、決定論的解析法よりも正解率が下がったが、単純化した場合は、全ての評価文セットにおいて正解率が向上している。

5.2.4 ポーズ情報と F_0 情報結合の効果

ポーズと F_0 情報の結合は、式 7 のとおりに、結合係数 α によって $P(d|x)$ と $P(d|f)$ を線形結合することにより行った。係数 α の最適値は実験的に求めた。すなわち、話者毎に、 α の値を 0 から 1 まで 0.1 刻みで変化させて係り受け解析を行い、最良の文正解率を与えた α の値を最適値とした。表 4 の文正解率の数値は、他の実験と同様、 α が最適値のときの文正解率を評価文グループ毎に平均した値である。

ポーズの有無によってデータを分離して推定したポーズモデルを用いた場合、セット B とセット C のいずれに対しても、ポーズのみの場合に比べて 1.2 ポイント、 F_0 のみの場合に比べて 3.8 ポイント文正解率が向上した。ポーズと F_0 情報を結合することにより、それぞれ単独で使う場合よりも係り受け解析精度の向上に効果的であることが分かる。

また、ポーズの有無によってデータを分離して推定したポーズモデルを用いた場合には、そうでない場合に比べて、 α の最適値は 1 に近づく傾向が見られた。これは、ポーズ情報の効果が F_0 情報に比べて大きいことを意味しており、ポーズを分離推定したほうが良いことの傍証となっている。

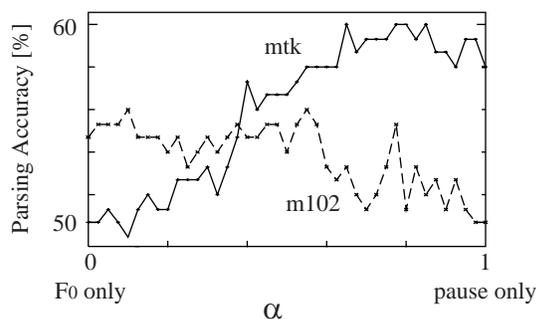


図 6: ポーズ情報と F_0 情報の混合係数 α と文正解率の関係の例

α の最適値は話者毎に異なっている。ある話者の α の最適値は 0 に近い一方、別の話者では 1 に近い結果となった。図 6 は、実験に用いた話者の中からこの傾向が対照的な 2 名を選び、 α の関数として文正解率を描いたものである。この違いは韻律の生成におけるある種の話者性を示していると思われるが、

この点についての考察は今後の課題である。

6 おわりに

我々は、日本語の係り受け解析において韻律情報を利用することにより、解析精度が向上することを報告してきたが、実験に用いてきた話者数・評価文の条件は限られており、韻律情報のモデルは話者依存であった。本研究の最終年度においては、44 名の話者による音声データを用い、評価用文セットの種類を増やし、話者独立に韻律情報モデルを作って実験を行った。

その結果、多数話者を用いた不特定話者対応の条件においては、ポーズ長と F_0 特徴量のモデルは従来より簡単なものを用いることで解析性能が向上すること、ポーズ長は平均音節継続長で正規化した方が良いことがわかった。また、ポーズ長のモデルは、ポーズがある場合と無い場合に分けてモデル化した方が良いことが分かった。ポーズ情報と F_0 情報を結合して韻律情報として用いれば、さらに係り受け解析性能は向上するが、ポーズ情報と F_0 情報の効果の程度は話者によって異なることが分かった。

今後は朗読文音声だけではなく、話し言葉の音声データを用いて実験を行いたい。

参考文献

- [1] 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,” 信学論, vol. J79-D-II, no. 9, pp. 1455–1463, 1996-9.
- [2] 海木延佳, 匂坂芳典, “局所的句構造に基づく F_0 制御,” 信学論, vol. J83-D-II, no. 9, pp. 1853–1860, 2000-9.
- [3] 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,” 信学論, vol. J71-D, no. 7, pp. 1218–1228, 1988-7.
- [4] N. M. Veilleux and M. Ostendorf, “Probabilistic parse scoring with prosodic information,” Proc. ICASSP’93, Minneapolis, MN, USA, vol. II, pp. 51–54, 1993.
- [5] 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中の隣接句間

- の修飾関係有無の判定,” 信学論, vol. J78-D-II, no. 11, pp. 1581–1588, 1995-11.
- [6] J. Venditti, S. Jun and M. Beckman, “Prosodic cues to syntactic and other linguistic structures in Japanese, Korean and English,” J. L. Morgan and K. Demuth (eds.), *Signal to syntax: bootstrapping from speech to grammar in early acquisition* (Hillsdale, NJ: Lawrence Erlbaum), pp. 287–311, 1996.
- [7] T. Ohsuga, Y. Horiuchi, T. Umeda, A. Ichikawa, “Estimating Syntactic Structure from Prosody in Japanese Speech,” *IEICE Trans. Inf.& Syst.*, vol. E86-D, no. 3, pp. 558–564, March 2003.
- [8] 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,” 信学論, vol. J70-D, no. 3, pp. 601–609, 1987-3.
- [9] 匂坂芳典, 浦谷則好, “ATR 音声・言語データベース,” 日本音響学会誌, vol. 48, no. 12, pp. 878–882, 1992-12.
- [10] “HTK: Hidden Markov Model Toolkit,” Version 3.1, <http://htk.eng.cam.ac.uk/>.
- [11] 黒橋貞夫, 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, 1994-1.
- [12] Entropic, Inc., “waves+ manual,” Product Release 5.3, 1998.
- [13] 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, 1996-12.
- [14] K. Ozeki, K. Kousaka, and Y. Zhang, “Syntactic information contained in prosodic features of Japanese utterances,” *Proc. EUROSPEECH’97*, Rhodes, Greece, vol. 3, pp. 1471–1474, Sept. 1997.
- [15] 高坂和之, 張玉潔, 尾関和彦, “係り受け解析における韻律情報の有効性,” 日本音響学会講演論文集, pp. 171–172, 1998-3.
- [16] Y. Hirose, K. Ozeki, and K. Takagi, “Effectiveness of prosodic features in syntactic analysis of read Japanese sentences,” *Proc. ICSLP2000*, Beijing, China, vol. 3, pp. 215–218, Oct. 2000.
- [17] K. Takagi and K. Ozeki, “Pause information for dependency analysis of read Japanese sentences,” *Proc. EUROSPEECH 2001 Scandinavia*, Aalborg, Denmark, vol. 2, pp. 1041–1044, Sept. 2001.
- [18] 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的特徴量の有効性,” vol. 8, no. 4, pp. 71–89, 自然言語処理, 2001-10.
- [19] 久保田新, 高木一幸, 尾関和彦, “係り受け解析におけるポーズ・ピッチの利用法の検討,” 日本音響学会講演論文集, 2-2-8, pp. 271-272, 2001-10.
- [20] K. Ozeki, K. Takagi, and H. Kubota, “The use of prosody in Japanese dependency structure analysis,” *Proc. ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*, Red Bank, NJ, USA, pp. 123–126, Oct. 2001.
- [21] K. Ozeki, K. Takagi, and H. Kubota, “Recovery of Japanese dependency structure using prosodic information,” *Proc. International Symposium on Prosody and Speech Processing*, Tokyo, Japan, pp. 169–174, Feb. 2002.
- [22] 久保田新, 高木一幸, 尾関和彦, “韻律を利用した係り受け解析におけるポーズ・基本周波数情報の結合法の検討,” 日本音響学会講演論文集, 1-P-22, pp. 395–396, 2002-3.
- [23] K. Takagi, H. Kubota, and K. Ozeki, “Combination of pause and F0 information in dependency analysis of Japanese sentences,” *Proc. ICSLP2002*, Denver, CO, USA, pp. 1173–1176, Sept. 2002.
- [24] K. Takagi, K. Ozeki, “Dependency analysis of read Japanese sentences using pause information: A speaker independent case,” *Proc. Speech Prosody 2004*, Nara, Japan, pp. 595–598, Mar. 2004.
- [25] 高木一幸, 尾関和彦, “係り受け解析における韻律情報有効性の多数話者による評価,” 日本音響学会講演論文集, pp. 239-240, 2004-3.