

スペシャル・セッション〔パラ言語・非言語情報の知覚、分析、生成〕（招待講演） 1—10—11

パラ言語情報研究の課題

○ 前川 喜久雄

(独)国立国語研究所 研究開発部門

1. はじめに

近年、パラ言語情報、非言語情報に関連した研究が音声研究の本格的な対象として認知されるようになってきた。私がパラ言語情報の音声学的研究に踏み込む契機となったのは、本学会 1994 年度大会での藤崎博也氏の招待講演であったが[1]、この講演を聴くまで、私自身はパラ言語情報が真摯な研究の対象になるとは考えていなかったように思う。

今回のスペシャルセッションに際し、チュートリアル的な内容をという依頼をいただいたが、整然としたチュートリアルは不可能である。これは謙遜でも逃避でもない。基本的な問題が山積しているのが、パラ言語・非言語情報に関する研究の現状だからである。本稿では、パラ言語情報の研究に焦点をあてて、研究の基本的課題を検討してみたい。私見に対する皆様の建設的な批判を期待する。

2. 言語・非言語・パラ言語

2.1 非言語情報

音声がいわゆる言語学的情報（つまり音韻論や統語論が研究対象としている類の情報）だけを伝達しているのではないことは、日常の体験に照らしても十分に理解できる。以下藤崎による分類[1,2]を参考にしながら、音声に含まれる情報の種類を分類してみることにしよう。

まず音声信号には不可避免的にそれを発した個人に関する情報が付随している。話者の性差はその典型例であるが、その他にも、話者の年齢、感情ないし情動、気分(mood)等に関する情報もかなりの程度まで音声によって伝達されている。

性差、年齢差、感情、気分に通ずる特徴をいくつか指摘できる。第一に、これらの情報は話者によって意図的に制御される情報ではない(もちろん、意識的な演技ないし模倣の可能性は除外する)。第二にこれらは恒常的もしくは相当長時間にわたって持続する話者の身体的ないし心的状態に関わる情報である。第三に、これらの情報のなかには性差や年齢差のように、話者の音声器官の解剖学的構造によってはほぼ決定的に規定されてしまう情報が含まれている。最後に—これは主に上述した第三の特徴の帰結であるが—これらの情報にはかなりの程度まで、言語普遍性が認められる。

上に論じた情報を藤崎は「非言語情報」と呼んでいる。非言語情報の一部には、これまでの音声研究でも暗黙のうちに配慮がなされてきた。例えば合成音を作る際には、性別や年代の決定が必須である。性差の研究はもっとも進んでおり[3]、話者の個性については話者認識の研究があり[4]、感情や気分についても或る程度の研究は進められてきている[5,6]。

2.2 パラ言語情報

次に、音声で伝達する情報を言語情報と非言語情

報に分類するだけでは充分でないことを例によって示そう。今、「ナニヤッテンノ」というテキストが与えられたとする。テキストが与えられるとは、所与のテキストを構成する語(ないし形態素)の意味と音韻、および語の結合によって生じる文の構造の情報が与えられることを意味しており、これが藤崎の言語情報である。

本例の場合、それぞれ/na'ni/, /yaru/, /te/, /iru/, /no/ という音韻からなる五つの語が結合しており(疑問詞にはアポストロフで表記されるアクセント核が指定されていることに注意)、品詞はそれぞれ疑問詞、動詞、接続助詞、補助動詞、終助詞である。動詞と接続助詞の結合に際しては、音韻規則が適用されて、/yaru/+ /te/⇒/yaQte/ となる。これに補助動詞が結合する場合は、/yaQteiru/の他に/yaQteru/という補助動詞の縮約形が許容される。また、終助詞/no/が結合されるに際して、結合対象の動詞句がアクセント核をもたなければ助詞の直前(動詞句の末尾)の音節にアクセント核が挿入されること、補助動詞の縮約形の音韻にさらに変化が生じた/yaQte'Nno/の形も許容されることなども言語情報の一部である。

以上のような言語情報から音声信号を生成しようとする試みは、テキストからの音声合成(Text-To-Speech synthesis)の試みに他ならない。そこで市販の TTS システムに上記のテキストを与えてみよう。出力される音声信号は確かに「ナニヤッテンノ」と書きとることが可能な音声であり、いわゆる音声明瞭度は高い。しかし、その音声信号には何か重要な情報が欠落していると感じられることがある。

そもそも、「ナニヤッテンノ」という発話はどういう状況で何を目的として発せられるかを考えてみてほしい。多くの方は単なる「質問」、つまり聞き手が今何をしているのかについての情報提供を要求する発話を想定されることだろう。しかし、なかには、「叱責」や「からかい」といった意図を想定された方があるかもしれない。実際、このテキストをそうした意図で発することは、日本語として完全に自然な言語行動である。

また、ここで大切なことは、「質問」「叱責」「からかい」等の話し手の意図が通常聞き手にも誤解されることなく伝わるという事実である。つまり、上記3種類の「ナニヤッテンノ」は、音声として伝達される場合、曖昧であるとみなすことができないのである。この事実は、上記三種の情報の識別に必要な音声上の手がかりが信号自体に内包されているためと考えるのが自然であり、実際、このことは文脈を排除した知覚実験によって確認できる。

音声信号を観察すると、「質問」では発話末で基本周波数が上昇し、他では下降する。「叱責」は発話時間長が顕著に短く、反対に「からかい」は長い。「叱責」は平均パワーが大きいなどの特徴を把握できる。さらに発話の韻律構造にも違いがあり、「叱責」は全体が1アクセント句として、「質問」は1ないし2アクセント句として、そして「からかい」は /na'ni/yaQte'Nno/ のように必ず2アクセント句で発音される(縦線がアクセント句境界)。さらに「質問」と「叱責」では第1アクセント句中のアクセントの方が基本周波数が高いが、「からかい」では第2アクセント句の方が高くなることもわかる。

Issues in the Study of Paralinguistic Information. By Kikuo Maekawa (Dept. Language Research, Natl. Inst. for Japanese Language.) Email: kikuo@kokken.go.jp

これらの音声特徴によって上記の意図の相異が伝達されているのである。

この例によって示される種類の情報を藤崎はパラ言語情報と呼んでいる。パラ言語情報の特徴を先ほどの非言語情報の説明と対比をとりながら列挙してみよう。

まず、パラ言語情報は話者が意図的に制御して生成する情報である。この点は言語情報と共通している。次にパラ言語情報は談話中で局所的に生じる情報であり、話者の恒常的身体／心理状態とは関係がない。この点も言語情報と共通である。パラ言語情報に言語普遍性が認められるかは興味深い研究課題であるが、おそらく非言語情報ほどに普遍的とは考えられない。

ここまでの説明で、パラ言語情報と非言語情報との相違はほぼ明らかになったが、言語情報との相違点はどこにあるのだろうか。藤崎は両者の相違点を、単一範疇の内部で連続的な変化が観察されるかどうかに求めている。つまり{sakura}や{tabe'ru}という形態素の意味(すなわち言語情報)が、本質的に離散的であり、「桜」「食べる」という範疇的な意味だけを示しているのに対して、「叱責」や「からかい」には範疇的な意味と同時に、さまざまな程度差を伴うという点である。同じ「叱責」にも強弱さまざまな「叱責」が存在し、一定限度内において、その程度差が音声によって伝達されることは経験的に確かであると言ってよい。

パラ言語情報にはさらに、基本的には文字によって伝達されないこと、音声の分節の特徴よりも韻律的特徴に強く依存することなどの特徴が認められる。これらには次節以下で触れることにする。

以上、言語情報、非言語情報、パラ言語情報について説明を加えたが、最後に、この分類にはグレーゾーンがあることも認めておかなければならない。しばしば問題となるのが、感情の帰属である。藤崎の分類で感情が非言語情報に分類されているのは、感情を生理学的な現象とみなし、意図的には制御できないとみなすからである。しかし、人間が感情の生起を自由に制御できないことは確かだとしても、一定限度内であれば、意図的に抑圧したり、増幅したりすることは可能であると感ぜられる。少なくとも性差や年齢差に比較すると、感情は明らかにパラ言語よりの現象であると言えるだろう。

3. 話し言葉と書き言葉

藤崎も指摘しているが、パラ言語情報は通常文字によって伝達されることがない。反対に言語情報は通常文字が伝達する情報である。これはわかりやすい特徴であるのだが、日本語のアクセントや中国語の声調のように、文字には間接的にしか表現されない言語情報が存在すること、感嘆符や疑問符などの書記記号(近年ではさらにいわゆるエモティコンの類)が専らパラ言語情報のために利用されることなどの点において、厳密な分類基準とするには問題がある。

しかし、私はパラ言語情報こそが話し言葉の本質にかかわる特徴なのだと考えている。後述するように、パラ言語情報の伝達には、イントネーション、リズム、発声様式など、音声の韻律的特徴が重要であるが、その多くは喉頭によって制御される特徴である[7]。一方、言語情報ないし文字情報は声道によって制御される音色の情報に多くを負っている。そして source-filter theory が仮定しているように、人間は声道と喉頭を基本的には独立に制御することが可能である。

その結果、話し言葉には、声道の制御によって実

現されるものと、喉頭の制御によって実現されるものと、ふたつの情報伝達チャンネルが並存することになり、この点が単一チャンネルしかもたない書き言葉との決定的な相違点となっている。いわば、書き言葉はモノラル放送、話し言葉はステレオ放送なのである[8]。

ちなみに非言語情報も話し言葉の重要な特徴ではあるが、手書き文字は或る程度まで性差や年齢差を伝達することから、パラ言語情報ほどには話し言葉の本質を規定しているとは考えられない。

このように考えると、パラ言語情報の研究が音声研究において有する価値も自ずと明らかになってくる。従来、音声情報処理の研究では、音声の文字に対する優位性として、情報伝達スピードの速さや、キーボードなどの入出力装置が不要であることが強調されてきている。しかしこれは言語情報という同じ土俵のなかでの比較である。同じ土俵のなかであれば、文字には文字固有の特長が認められる。保存性の良さや検索の容易性がそれである。これらが言語情報にとってきわめて重要であることには疑う余地がない。

パラ言語情報の研究は書き言葉とは別の土俵での研究であり、だからこそ比較を絶した価値を有している。書き言葉では伝達できない情報の処理を自家菜籠中のものとしたとき、音声処理技術の実用化にも、自ずと明るい展望が開けてくるだろう。

4. パラ言語情報研究の課題

本節では、音声によるパラ言語情報伝達過程の解明にかかわる基礎的課題を指摘する。

A. パラ言語情報の体系化

パラ言語情報の意味論的考察である。最重要であると同時に、目下実施が最も困難な課題である。①パラ言語情報カテゴリの包括的リストの作成、②基本的なカテゴリと派生的カテゴリの区別、③カテゴリ間の関係の解明、等の課題が考えられる。

B. パラ言語情報の伝達メカニズム

言語情報を対象とした音声科学研究のパラ言語版であり、多くの下位領域に区分できる。①特定言語でパラ言語情報の音声学の実現に利用される音声特徴の網羅(言語情報における音韻目録に該当)。それらの特徴の、②調音的、③音響的、④知覚的特性の解明、さらに⑤パラ言語に関わる音声特徴と言語にかかわる音声特徴との関係の解明。最後に、⑥パラ言語情報に関わる音声特徴の表記法の考案もこの領域に含めてよいだろう(6D 参照)。

C. 言語情報とパラ言語情報の交互作用

パラ言語的な意味のなかには言語情報とあわせて考察しなければならないケースがあり、その場合、交互作用の有無が重大な関心の対象となる。例えば日本語の終助詞(ネ、ヨ、カ...)と句末イントネーションの関係を内省すると、同じイントネーションが終助詞との組み合わせによって異なるパラ言語的意味を伝達することがわかる。これは明確な交互作用の例である。また、韻律と語彙が同じ意味の表現に寄与することもある。待遇の高い語彙(イラッシャル、メシアガル...)と丁寧な韻律との関係がその一例であるが、この場合にも、交互作用の有無が問題になる。さらに、言語情報とパラ言語情報とを意図的に矛盾させて皮肉を表現するようなことも我々はおこなっている。これらは元来言語学(語用論)の守備範囲に属する問題であるが、残念ながらあまり研究が進展していない。

D. テキスト構造との関係

文章を効果的に朗読するためには、テキストの構造に即したパラ言語情報の制御が必要である。係り受け関係のような言語情報を適切に表現するだけでなく、新情報と旧情報の区別、強調の有無、適切な強調手段の選択、内容に即した速度や声質の制御などが要請される。この領域の研究を進展させるためには自然言語処理技術や意味理解研究の進展が必須である。

E. パラ言語情報の言語依存性

この問題は、言語依存性(あるいは普遍性)という観点から見たとき、パラ言語情報が非言語情報(普遍的)と言語情報(依存的)の間でどのような位置を占めるのかという問題である。外国語教育への応用を考える際には特に重要である。

5. パラ言語情報の研究手法

前節の基本的課題を達成するためには様々な学問領域に属する研究手法を動員する必要がある。

「体系化」は分析哲学を言語情報以外に拡張する試みと言ってよく、記号論・意味論の根幹に立ち戻った考察が必要である。「伝達メカニズム」と「交互作用」には従来の音声科学と同様、実験的手法が有効であり続けるだろう。

「テキスト構造」のためには、パラ言語情報に配慮して設計したコーパスの構築と、その統計的な分析が有益であろう。ただし、コーパスへの情報付与をおこなう前に「体系化」の研究を多少とも進める必要がある。付与するラベルの背景に客観的に理解可能な分類基準が存在していなければ、徒に大量のラベリングを実施しても、理解不能な情報になりかねない。

6. 私たちの研究

本節では、パラ言語情報という広漠な研究対象を巡る試行錯誤の記録として、私たちの研究グループが過去数年間におこなってきた研究の梗概を紹介する。

A. 「体系化」

この領域の研究成果は何も発表できていない。言語学の貢献が期待される問題であるだけに忸怩たる思いである。

B. 「伝達メカニズム」

相対的には研究が進展した領域である。これまでの成果の多くは文献[9]にまとめられている。

パラ言語情報の伝達には従来からの予想通り、音声の韻律的特徴が重要な役割を果たしているが、非言語情報と違って、パラ言語情報ではその変化が局所的に生じること、特に韻律的な音韻特徴が集中的に分布している発話の両端において顕著であることを確認した。

基本周波数に関しては、「疑い」や「感心」などのパラ言語情報の伝達において、発話冒頭のいわゆる句頭の上昇のタイミングに遅れが生じ、それと同時に上昇幅も拡大することが判明した(ただし、川上葵氏は主観的観察に基づいてこの事実を1950年代に報告しており[10]、またNTTの研究グループも音声合成用言語の研究で、この現象を制御するためのパラメータを導入している[11])。また、句頭の上昇に遅れが生じる種類の発話にアクセント核が含まれる場合、核による基本周波数の下降タイミングにもしばしば遅れが生じることが判明した。

発話末のいわゆる上昇イントネーションに数種類のタイプが存在することは、やはり川上によって夙に指

摘されているが[12]、私たちの実験でもこれが複数のパラ言語情報の識別に貢献していることが示唆されていた。

持続時間は、基本周波数とならんで最も変動が顕著な音声特徴であり、やはり発話の両端での変動が著しいことがわかっている。

伝達メカニズムに関する私たちの研究で真に独創的な成果と呼べるのは、或る種のパラ言語情報の表出においては、韻律的特徴と同時に分節的特徴にも組織的な変化が生じていることの確認である。

この現象は当初フォルマント領域において発見された。主として第2フォルマント周波数の高低に関係しているのが、舌の前後位置が変化しているものと推定したが、後にEMMAを用いて調音運動を測定したところ予想どおりの調音変化が観察された[13]。ただし両唇間の距離にも組織的な差異が観察されるので、実際には、舌と唇の相乗効果として、第2フォルマントに強い影響が生じているものと考えられる。

ところで、上述の舌や唇の変動は母音に限らず子音を含む発話全体に観察されることから、個々の分節音ではなく発話の全体に関わる音声特徴—すなわち声質(voice quality)—の制御であると考えられる。また声質の差異が発話の両端で殊に顕著となることから、韻律的特徴と共通の制御を受けていると想像される。

発声様式(phonation type)にも、「落胆」ではbreathy、「疑い」ではcreakyのように、パラ言語情報と相関した差異が観察される。母音に限らず無声子音区間にも母音区間と類似した喉頭制御の差が観察されることから、これもまた声質の制御であろうと考えられる[14]。なお、同一のデータからARX法によって音源パラメータを推定する研究が宇都宮大学で実施されている[15]。

C. 「交互作用」

発話の丁寧さの伝達におけるイントネーションと語彙の貢献を実験的に検討した。この研究では熊本方言を利用した。無アクセント方言であるためにイントネーションの自由度が高く、また敬語を高度に発達させているからである。

6種類のイントネーションと、丁寧さの異なる3種類の終助詞との交互作用を検討するために、自然音声を刺激とし、一対比較法によって丁寧さを測定した。大学生と平均40歳代の成年被験者層を対象とした最初の実験ではイントネーションと終助詞の間には交互作用が存在しないことが確認された[16]。その後、被験者を中学生から70歳代まで拡大したが、丁寧さに関する限り、やはり交互作用は確認されていない[17]。また、熊本方言におけるイントネーションの丁寧さの序列を比較的簡単な規則によってその韻律構造から導くことが明らかになった。

D. 「テキスト構造」

この問題のために特別な研究は実施していない。しかし、現在私たちの研究グループが構築に取り組んでいる『日本語話し言葉コーパス』[18]の韻律ラベリングのために設計した韻律ラベリング体系X-JToBIでは、パラ言語情報に関連する音声特徴の表現に意を用いている。そのため上記コーパスのうち韻律ラベルが付与される予定の「コア」部分(43時間程度)は、パラ言語情報のテキスト構造の研究にも、或る程度利用できるものと期待している。

X-JToBIは朗読音声を対象としたJ-ToBIを自発音声用に拡張したラベリング体系である[19](5で論じたパラ言語情報そのもののラベルではなく、あくまで音

声特徴のラベルであることに注意)。

先に指摘した句頭の上昇の遅れや、発話末の上昇イントネーションの種別などはラベルによって検索できる。アクセント句中にポーズが挿入されたり(例:「こういう結果<pause>になりました」)、アクセント句末から2番目のモーラのピッチをアクセント核(?)とは無関係に上昇させること(例:「デイズニーランドで」)などによって生じる韻律境界の微小な差異も表現可能である。

E. 「言語依存性」

この問題は最近検討し始めたばかりである。文献[9]で日本人被験者に対して実施したものと同一の知覚実験を英語を母語とする日本語学習者名に実施したにとどまる。この実験からは、学習者の反応から導かれる知覚空間と日本人の知覚空間とは基本的によく類似していること、しかし、学習者はパラ言語情報に関わる音声特徴を識別できても情報のタイプを同定できないことがあること、パラ言語情報のタイプによって同定の困難に大きな差があることなどの知見が得られている[20]。

7. おわりに

パラ言語情報に関しては本稿で指摘した以外にも面白い研究課題がたくさん見つかるに違いない。視覚情報との統合や脳科学の観点からの研究などが思い浮かぶ。

パラ言語情報の音声科学的研究は、現在その揺籃期にあり研究テーマに困ることはない。創意と工夫でいくらでも面白い研究が実施できる。要するに開拓時代なのである。ただし、開拓時代であればこそ、基礎研究が重要であることも強調しておきたい。

言語情報に関する音声科学の発展は、音素や形態素といった言語学の概念に深く依存してきた。音響学的音声生成理論が確立されたことの重要性はいうまでもない。パラ言語情報の研究は、まだこれらの基礎概念、基礎理論が確立されていない段階にある。こういう課題にこそ果敢にとりくむべきであろう。私にも、開拓の面白さに惹かれてやや無反省に研究を拡張してきたという反省があり、「体系化」の重要性をひしひしと実感している。

付録:『パラ言語』の用法について

Paralanguage あるいは paralinguistic という術語は研究者によって随分異なる意味で用いられるので注意が必要である。本稿で取りあげた藤崎の分類は、情報学的な観点からの用法であったが、その他に、少なくとも、音声学的な用法と音韻論的な用法がある。代表例を示そう。

音声学的用法の代表は、D. Crystal の著作である [21]。Crystal は分節音を中心に据えた伝統的な調音音声学をイントネーションの領域に拡張する作業の一環として、英語音声構成する特徴を segmental と non-segmental に分類し、さらに後者を prosodic system, paralinguistic system, non-linguistic features に分類している。このうち prosodic system にはピッチやラウドネスなど、英語のストレスやイントネーションの構造に関連づけられる諸特徴が属し、paralinguistic system には発声様式の諸特徴や「叫び」「忍び泣き」などの非言語行動が属する。英語音声は藤崎流に分類したならば、パラ言語情報の伝達に関与する音声特徴はむしろ prosodic system に属することになるだろう。

Crystal の用法には独自性が強いが、英国の言語学者の間では広く普及しているようである。

音韻論的(言語学的)用法の代表としては D.R. Ladd の著作が挙げられる[22]。そこでは paralinguistic という術語が contrastive の対義語として用いられており、音韻論的分析の基本となる二項対立とは異なる連続的な変化(gradience)を示すことが paralanguage の本質的特徴とされている。この点、Ladd の用法は藤崎のパラ言語情報に重なる。ただし、Ladd は、パラ言語情報と非言語情報の区別についてはほとんど議論していない。

謝辞: 日頃、多方面の議論に応じていただくことの多い藤崎博也、藤村靖、相澤正夫の各氏に感謝いたします。

参考文献

- [1] 藤崎. 「韻律研究の諸側面とその課題」音講論集(1994.6).
- [2] Fujisaki. Prosody, models, and Spontaneous In Speech. In Sagisaka et al (eds.) *Computing Prosody*. New York: Springer (1997).
- [3] Klatt & Klatt. "Analysis, synthesis, and perception of voice quality variation among female and male talkers," *JASA*, 87 (1990).
- [4] Furui. "Research on individuality features in speech waves and automatic speaker recognition techniques," *Speech Communication*, 5 (1986).
- [5] 北原・東倉. 「音声の韻律情報と感情表現」信学技報 SP88-158 (1988).
- [6] Murry & Arnott. "Toward the simulation of emotion in synthetic speech," *JASA*, 93-2 (1993).
- [7] 前川. 「韻律とコミュニケーション」音響学会誌, 55-2 (1999).
- [8] 前川. 「音声学」(『講座言語の科学』第2巻), 岩波書店 (1998).
- [9] 前川・北川. 「音声はパラ言語情報をいかに伝えるか」認知科学, 9-1 (2002).
- [10] 川上. 「文頭のエントネーション」国語学, 24 (1956).
- [11] 中蔦・水野. 「音声合成を制御する記述言語」音響学会誌, 55-2 (1999).
- [12] 川上. 「文末などの上昇調について」国語研究, 16 (1963).
- [13] Maekawa & Kagamiya. "Influence of paralinguistic information on segmental articulation," *Proc. ICSLP2000*, Beijing, 2, (2000).
- [14] 藤本・前川. 「パラ言語情報による発声様式の変化」信学技報 SP2002-62 (2002).
- [15] Kasuya, Yoshizawa, & Maekawa. "Roles of voice source dynamics as a conveyer of paralinguistic features," *Proc. ICSLP2000*, Beijing, 2, (2000).
- [16] 前川・吉岡. 「発話の丁寧さに対する語彙的要因と韻律的要因の寄与」国語学, 190 (1997).
- [17] 前川・榎・吉岡. 「発話の丁寧さにおよぼす語彙的要因と韻律的要因の関係とその社会差」信学技報, TL99-19 (1999).
- [18] 前川・籠宮・小磯・小椋・菊池. 『日本語話し言葉コーパス』の設計」音声研究, 4-2 (2000).
- [19] Maekawa, Kikuchi, Igarashi & Venditti. "X-JtoBI: An extended J_ToBI for spontaneous speech," *Proc. ICSLP 2002*, Denver (2002).
- [20] Maekawa & Erickson. "Perception of the paralinguistic information by non-native speakers," *Paper presented at LP2002*, Chiba (2002).
- [21] Crystal. *Prosodic systems and intonation of English*. Cambridge Univ. Press (1969).
- [22] Ladd. *The structure of intonational meaning*. Bloomington: Indiana Univ. Press (1980).