

脳科学への期待

前川 喜久雄*

My Expectations about the Brain Sciences of Speech

Kikuo MAEKAWA*

SUMMARY: A layman's view about the brain science approach to speech was presented. The most obvious advantage of the brain science approach over the traditional behavioral approach consists in the possible clarification of speech encoding and decoding processes; the processes that remain largely untouched in the behavioral approach. Especially, construction of computational model of speech perception will open up a brand-new landscape of speech science. However, to achieve these goals effectively, brain science should pay more attention to the fruits obtained by the behavior approach. Experiments of brain science should be designed based upon sufficient understanding of the linguistic properties of experimental stimuli. It is especially important to pay enough attention for the probabilistic properties of linguistic signs. For example, the distinction between semantically normal and abnormal stimuli is not necessarily a dichotomy. It is possible to compute the occurrence probability of semantically deviant sentences using the power of powerful computer and large-scale annotated language corpora. In the same vein, retrieval of a speech corpus reveals that the long-believed distributional restriction of the /tyu/ mora in Japanese is not a strict one as long as spontaneous speech is concerned. Results of brain scientific experiments will enrich greatly our understanding of speech, if the experiments are designed correctly.

キーワード：脳科学 (brain science), 行動科学 (behavioral science), 音声研究, 確率 (probability), コーパス (corpus)

1. はじめに

筆者は音声研究を専門としているが脳科学については素人である。学生時代にペンフィールドとロバーツの本 (Penfield & Roberts, 1959) を読んで強い感銘をうけた。しかし大脳生理学に属する研究をおこなう機会が自分にあるとは思えなかった。1979年のことである。近年、脳科学として位置づけられる音声研究に接する機会が飛躍的にふえた。研究室のポストドクにも脳科学的な手法で音声を研究する人がいた。しかし私自身が「脳」をキーワードとする研究を立案したことはない。

一昨年、音声学会研究例会で脳と音声について何か話せと言われたとき、例会ならばと気楽にひきう

けた。専門的な発表に素人の見解がまじるのも悪くないと思ったし、何よりも正式な書き物として跡がこのこらないからである。ところが今般『音声研究』の特集にするという。一瞬、執筆を断わろうと思ったが、思いなおして引きうけることにした。やはり素人の見解を述べることに意味があると考えたからである。勿論、恥はかきたくない……が、慣れてもいる。

以下には筆者がこれまでに脳科学的音声研究に接してきた過程で感じた感想を記した。まず筆者が脳科学を音声研究の中でどう位置づけているかを述べ、次いで脳科学的研究に対する期待と不安を述べた。ただし、研究例会の席で述べた内容とはかなり異なる内容になっていることをお断わりしておかね

* 独立行政法人国立国語研究所研究開発部門 (Dept. Language Research, The National Institute for Japanese Language)

ばならない。素人の見解を述べるという趣旨に鑑みてご容赦いただきたい。

2. 行動科学と脳科学

脳科学的な研究は行動科学と対比されることが多い。両者の大きな相違点は人間の情報処理過程をブラックボックスと捉えるか否かにあるのだと思う。行動科学では、よく知られているように、刺激とそれに対する反応の組み合わせを解析する。ここでは情報処理装置としての脳はブラックボックスとして扱われている。中身が見える必要はない。勿論、行動科学的な研究のなかにも脳の内部構造を問題にするものはあった。両耳聴 (dichotic listening) の研究はその例である。しかし、その場合も左右の半球の内部はやはりブラックボックスである。

一方、脳科学ではブラックボックスの中身をのぞくことが目標である。PET や fMRI などの非侵襲的な脳活動計測技術がそれを可能にした。これは本当に大きな発明である。非侵襲的観測装置によって刺激に対する脳の反応が空間と時間の両面において把握できるようになったことは、20世紀の科学史のなかでも大書されるにちがいない。

しかし、それだけで脳の機能が解明されるわけではない。見えるのは脳の反応パターンであり、その背後のメカニズムまでは見えないからである。現在の脳科学が目指しているのは、そのメカニズムに関する理論の構築でありモデル化である。

これを音響学的な音声研究に喩えるならば、音声のスペクトルを計算することができても、そのスペクトル形状がどのようにして形成されるかについての理論が確立されていない状態である。1930年代の音声研究がちょうどその状態にあった(前川, 2003)。音響音声学では千葉と梶山や Fant, Stevens らの活躍によって音響学的音声生成理論が確立され、1960年前後には音声研究の黄金時代が現出した。脳科学もやがてそのような時代を迎えるのではなかろうか。是非そうあってほしい。

3. 音声研究と脳科学

次に音声研究と脳科学の関係を考えてみよう。音声研究、特に伝統的な調音音声学は、上に述べた行動科学の構図で言えば、刺激の特性を研究する科学であると言ってよい。調音音声学における言語音の分類も、音響音声学における音響特徴量の抽出も、外在化された刺激の特徴に関する分析である。その成果は、音声を刺激として利用する脳科学の研究者にとってきわめて重要である。脳科学におけるデータ解析は、刺激の特性と脳の活動パタンの相関分析に他ならないからである。刺激に対する理解の深さが研究の死命を制する点において行動科学と脳科学にちがいはない。

一方、音声研究にとって脳科学はどのように貢献してくれるだろうか。もし音声研究が今後も外在的な刺激の研究に専念するのであれば、研究に脳科学的手法をもちいる必然性はほとんどないだろう。脳科学が必要となるのは、刺激が外在化する以前と以後の過程を研究する場合である。音声の encoding と decoding の過程である。

4. 音声受容研究における期待

筆者にとって脳科学への最大の期待は、それによって音声の受容機構の解明が進むのではないかという期待である。音声の研究は、生成、伝播、受容の三面からおこなえる。生成をあつかうのが調音音声学ないし音声生理学、伝播をあつかうのが音響音声学であるとする、受容をあつかう領域として知覚音声学なり聴覚音声学なりが存在すべきであるのだが、本誌の読者にはおわかりのとおり、これらの名称はかならずしも定着していない。その理由は音声の受容にかかわる人間の情報処理機構について手探りの状態が続いていることにある。

図1は W. Levelt が描く「話し手の設計図」(a blueprint for the speaker) である (Levelt, 1989)。図中の speech comprehension system には内部構造が全く示されていない。知覚研究がブラックボックス中のブラックボックスであることが端的に示されてい

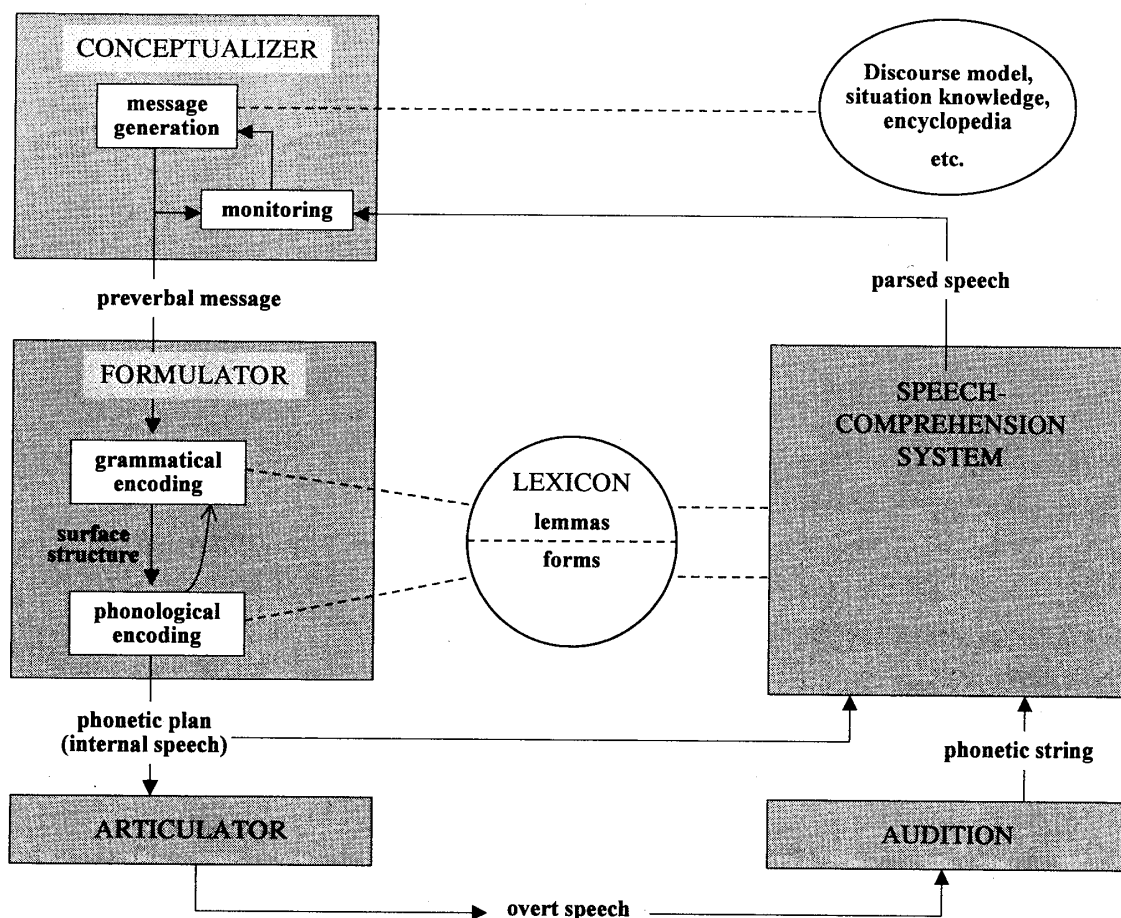


図1. W. Levelt による “a blue print of the speaker” (Levelt, 1989, p.9)

る。

音声器官に対する解剖学的知識は音声学者にとって必須の知識であるし、各器官の運動制御についても多くの知見が蓄積されている。そのため何らかの音声現象についてその生成プロセスを考える必要が生じた際には相当に具体的なイメージをうかべることができる（ただしすぐ後で述べるように、そのイメージが正確かどうかは別問題である）。

音響音声学の場合も同様である。音響学的音声生成理論（音源フィルター理論）は、やはり音声の音響特徴について考察するための強固な枠組みをあたえてくれる。

しかし、音声の受容を考える際にはよりどころとすべき枠組みが確立されていない。音声の知覚理論としてもっとも有名なのは音声知覚の運動理論（motor theory）であろうが、この理論については現在も盛んに議論がかわされている。運動理論でうまく

説明できる現象もあれば、そうでない現象もある。

音響的には明らかに異なった性質をもつ二つの音刺激が、或る言語の話者にとってはひとつの音素となり、別の言語の話者にとっては別々の音素となることはよく知られている。例えば日本語話者と英語話者における /r/ と /l/ がそれにあたる。従来の音声知覚研究の手法でこの現象を研究する場合に知ることができるのは刺激の音響的性質と心理学的な弁別閾の関係である。勿論これは貴重な知見であるのだが、言語が異なることによって何故受容のありかたが変化するのかという根本的な問いに対する回答とはなっていない。これと同じ問題は乳幼児による音韻獲得や方言識別の問題にも伏在している。

音声受容研究の最大の課題は、音が意味をもつ記号として成立する所以の解明であると筆者は考えているが、この問題の解明には脳機能の解明がどうしても必要である。実際、既にその方向の研究が盛ん

におこなわれている。しかし、その成果は、少なくとも私の知る範囲では、満足できるものではない。異なる刺激に対する脳の反応部位や反応時間が異なっていることが示されても、それだけで脳機能が解明されたことにはならない。必要なのは音声受容の計算モデルである。現在までに得られている種々の音声受容現象を過不足なく説明できる計算モデル、いわば Neuro-cognitive theory of speech perception が確立されれば、音声受容の諸問題は、現在の音響音声学と同程度に明確なイメージをもって把握することができるようになる。そのような理論の出現を切望する。

5. 音声生成研究における期待

以上、音声受容の問題について述べたが、脳の計算モデルは音声生成の研究にも必要である。いわゆる調音結合の現象や条件異音の現象、ひいては音声変化の問題を根本的に理解するためには、記号が調音の運動制御に変換される過程のモデルが必要であり、それは脳機能の一部と考えるのが妥当であろう。生成研究の場合、身体運動全般の研究が進んでいることもあって、実際に具体的な成果が得られつつある。近年 NTT の菅田正彰氏らによって実施された調音運動プランニングの研究はその代表例といつてよいものである(菅田, 2002, 2004)。

また、狭い意味での脳研究とはいえないが、ATR の本多清志氏らのグループによる MRI を用いた調音運動観測技術の開発は音声研究にとって重要な貢献である(Honda et al., 2004; 正木, 2004)。筆者は X 線マイクロビームや EMMA (磁気センサーシステム) によって測定された自分自身の調音運動データを解析していたとき、舌や唇の上に配された観測点の軌跡から調音器官の全体像を想像するのに苦心した記憶がある(Maekawa & Kagomiya, 2000)。

そういう苦勞をしているときに、MRI による擬似動画をはじめて目にする機会があり驚いた。百聞は一見にしかずとはこのことだと思った。今後このようなデータが蓄積されてゆくことによって、古典的な調音音声学にもとづく調音運動の理解は大き

な修正を余儀なくされるのではないかと感じている。

6. 不安

不安も述べなければならない。一番大きな不安は上に述べた期待の裏返しである。つまり、脳科学的な音声研究が私の期待する方向にむかっているのかという不安である。勉強不足を棚にあげるようだが、脳科学研究のなかには、その研究のめざす目標が何であるのかが素人にはわかりにくいものが多いと感じる。

最近、NTT の柏野牧夫氏が日本音響学会でおこなった運動理論に関する講演は、音声知覚研究における計算モデルの重要性をはっきりと指摘しておられた点でこの不安を軽減してくれるものであった(柏野, 2004)。しかし、実際にどのような計算モデルを構築すればよいかは今後の課題と言うしかないのが現状のようでもあった。

生成面の研究については、先に菅田氏らの研究に言及した。このモデルは舌や唇などの調音体が、音素という時間区分に従って与えられた空間的な調音運動のターゲットをスムーズに通過するための軌跡を生成するものである。図1でいえば formulator 下部の phonological encoding と articulator をつなぐモデルである。

このような比較的表層レベルのプランニングの前段階として、語や文の外形(音形)を決定する過程が存在するはずであり、図1では grammatical encoding と conceptualizer がその機能を担っている。従来、この段階でのプランニングに関する研究は音韻論を含む文法理論の守備範囲とされてきたが、最近では脳科学の手法でこの領域の問題を検討する試みが始まっている。この種の試みについては知識が不足しすぎているのでコメントを避けることにするが、ひとつ重要と思われるのは、情報処理において上位のプロセスを検討するためには、下位のプロセスに関する知見がしっかりしていなければならないことである。下位プロセスを無視すれば、上位プロセスに関して引き出しうる結論の自由度が無制限

に増大するからである。

この点で、不安というよりむしろ危惧を感じるのは、そもそも音声や言語に対する理解が不足している上位プロセスの研究である。著名な研究者にもそのような研究をおこなう人がいる。文献の引用は避けるが、文法知識すなわち生成文法の脳内表現を探るために文字列の意味的正しさを判断させる課題をおこない、「ゆきをさわる」は正しく、「ゆきをしかる」は正しくないとしている研究があった。勿論「ゆき」=雪」という語の可能性を考えているのである。

この研究がとある脳科学プロジェクトの成果発表会で発表されたとき、さすがに聴衆から人名（「ゆき」=由紀）の可能性を考えないのかという質問がでた。それに対する回答が「そういう周延的な用法は無視できると考える」旨であったのには一驚した。アクセントの問題を周延的と言われて腹が立つのは音声学者のヒガミかもしれないが、やはりそれだけの問題でもない。語彙アクセントが日本語文法の周延的要素であろうとなかろうと「雪」と「由紀」が別語であることにはかわりはないし、これらの語の生起確率に大きな差があるとも思われなからである。刺激の特性を正しく把握せずにおこなった知覚実験がどういう結果をまねくかは、行動科学の研究者ならば誰でも骨身に染みて理解している。

最後に下世話な不安もある。科学技術基本計画が施行された時期、「脳」は競争的研究資金獲得のキーワードと化したように見えた。最近ではやや沈静化しているようだが、それでも多額の研究費が脳研究に投入されている。そのこと自体の是非は問わないが、その影響として行動科学的な手法による研究を時代遅れとみなす風潮がうまれてはいないかと思われる。人文科学の領域ではそのような傾向はまだまだ生じていないと思われるが自然科学では杞憂とも言えないようである。

7. 脳科学の基礎としての行動科学

多くの読者には同意していただけるのではないかと思うが、そもそも脳科学は行動科学を置き換える

性格のものではない。ブラックボックスの中身をのぞけるのは脳科学の絶対的な強みであるが、その実験に用いる刺激の選択には行動科学の知見が欠かせない。そして、刺激としての音声や言語には、いまだ解明しつくされていないさまざまな特性がある。ここでは代表的な問題をひとつだけ指摘しておこう。それは言語刺激の生起確率に関する特性である。

先に「ゆきをしかる」という刺激に言及した。「ゆき」を「雪」に限定したとして、「雪を叱る」という文は本当に意味的に逸脱しているのだろうか。神様が雪に対して「あまり降り積もって人間を苦しめるんじゃないよ」とたしなめる童話であれば、「神様は雪を叱りました」という文は自然に用いることができる。このような文は有名な *Colorless green ideas sleep furiously* 同様、ほとんど生じないかもしれないが生じえない文ではない。問題はよく生じる刺激とほとんど生じない刺激とを意味的な逸脱に関して二分化してしまってよいかどうかという点である。「雪を触る」と「雪を叱る」との間に観察される脳活動の差異は、意味的逸脱の有無に関する binary な反応なのか、それとも逸脱の程度に関する gradual な反応の両極を観察しているのかという問題である。

実は、チョムスキーが *Colorless...* の例を挙げたのは一次言語資料だけに頼った構造主義言語学の欠点を指摘するためであった (Chomsky, 1957)。つまり一次言語資料をいくら探したところで、この文がみつかるわけがないから、明らかな非文である *Furiously sleep green ideas colorless* との差異をデータに基づいて判定することは不可能だという議論である。

いかにもチョムスキーらしい議論だが、言語資料の検索を人手に頼っていた当時の言語研究の状況では、この指摘に一定の説得力を認めなければならなかった。しかし計算機の能力が飛躍的に増大した現在では事情がかわりつつある。意味的逸脱を含んだ文の生起確率を推定することが不可能ではなくなりつつある。米国のペレイラは10億語規模のデータに種々の素性を付与したコーパスから潜在変数を含むクラスNグラムを計算することによって、*Furi-*

ously sleep... が生じる確率は *Colorless green ideas...* の確率の 20 万分の 1 以下であることを示している (Pereira, 2000; Liberman, 2003)。

これと同質の問題が音韻にも存在することは確実である。「テュ」というモーラは日本語には存在しないか、存在しても外来語にしか用いられない「周辺の」モーラとみなされているだろう。しかし、筆者らの開発した『日本語話し言葉コーパス』(前川, 2004) を検索すると、752 万語のデータ中に「テュ」が 650 回以上出現しており、そのほとんどすべてが「～という」が「～テュ」ないし「～テュー」に転訛した例であることがわかる。或る CV ないし CSV 連鎖が日本語のモーラであるかないか、あるいは語種との関連において周辺のであるかないかもまた、いきなりは二分法で律することができない問題なのである。

8. まとめ

以上、脳研究に関する素人の管見を述べてきた。しいて結論を書くならば、脳科学と行動科学の成果はこれからも相補的に発展していく必要があるということである。現在は、脳科学が一方的に行動科学の成果を利用しているように見えるが、いずれ、脳科学の成果が行動科学にフィードバックされて音声や音韻の構造に新たな光をあてるにちがいない。そのような研究に数多く接することを願いつつ筆をおく。

【付記】一昨年の研究例会での私の発表は脳科学的音声研究に対する否定的見解と受けとられ易かったようである。二次会の後にも発表者のひとりであった辰巳先生からホテルの部屋に呼びだされて追及をくらった。酔余の議論であったので細部は記憶にないが(辰巳先生、覚えてますか?), 音素や語といった言語単位の確率的性格をどう考えるかが話題の中心であった。例会のコーディネータであった今泉先生もこの問題を重視しておられた。本稿を書くにあたっては、この深夜の議論を意識した。例会では言及しなかった生起確率の問題をとりあげたのはそのためである。

そのため今度は生成文法に拠る脳研究に対して否定的な見解を表明したと受けとられるかもしれないが、それも筆者の本意ではない。脳科学的であっても行動科学的であっても、人間の情報処理過程の実験研究には刺激の特性に対する深い理解が欠かせないこと、そしてその特性のなかには統計的な特性も含まれることを指摘したまでである。たとえ実験の対象が言語能力であっても、人間を相手に実験をおこなう以上、言語運用と記憶に対する配慮を欠いてはまずいと言い換えてもよい。

参考文献

- 柏野牧夫(2004). 「音声知覚の運動理論をめぐって」日本音響学会講演論文集 (2004 年秋季), pp.243-246.
- 菅田正彰(2002). 「発話機構の計算論的研究」 *Computer Today*, 103, pp.30-38, サイエンス社.
- 菅田正彰(2004). 「音を作り出す仕組みを探る」計測と制御, 43: 4, pp.301-305.
- 前川喜久雄(2003). 「解説」(杉等美代子・本多清志訳, 千葉勉・梶山正登著『母音—その性質と構造—』岩波書店).
- 前川喜久雄(2004). 「『日本語話し言葉コーパス』の概要」*日本語科学*, 15, pp.111-133, 国立国語研究所.
- 正木信夫(2004). 「MRI は音声生成研究に貢献し得るか?」日本音響学会講演論文集 (2004 年秋季), pp.247-250.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Honda, K., Takemoto, H., Kitamura, T., Fujita, S., and Takano, S. (2004). "Exploring human speech production mechanisms by MRI." *IEICE Trans., & Syst.*, E87-D, pp.1050-1058.
- Levelt, W. (1989). *Speaking: From Intention to Articulation*. The MIT Press.
- Liberman, M. (2004). "Large-scale Knowledge Resources in Speech and Language Research," Talk at the *International Symposium on Large-scale Knowledge Resources*, Tokyo Institute of Technology.
- Maekawa, K. and T. Kagomiya (2000). "Influence of paralinguistic information on segmental articulation," *Proc. International Congress on Spoken Language Processing (ICSLP 2000)*, Beijing, 2, pp.349-352.
- Penfield, W. and L. Roberts (1959). *Speech and Brain Mechanism*. Princeton Univ. Press (上村・前田訳『言語と大脳: 言語と脳のメカニズム』誠信書房, 1965).
- Pereira, F. (2000). "Formal grammar and information theory: Together again?" *Philosophical Transactions of the Royal Society*, 358 (1769), pp.1239-1253.