

認知科学の新展開

3

乾 敏郎・安西祐一郎 編

運動と言語

今水 寛

銅谷賢治

二見亮弘

田邊敬貴

小椋たみ子

齊藤 智

山鳥 重

大津由紀雄



運動と言語

今水 寛

銅谷賢治

二見亮弘

田邊敬貴

小椋たみ子

齊藤 智

山鳥 重

大津由紀雄

〈編者紹介〉

乾 敏 郎(いぬい としお, 序)
京都大学大学院情報学研究所
安西祐一郎(あんざい ゆういちろう)
慶應義塾大学

〈執筆者紹介〉

今 水 寛(いまみず ひろし, 第1章)
科学技術振興事業団
銅 谷 賢 治(どうや けんじ, 第2章)
(株)国際電気通信基礎技術研究所(ATR)先端情報科学
研究部
二 見 亮 弘(ふたみ りょうこう, 第3章)
東北大学大学院工学研究科電子工学
田 邊 敬 貴(たなべ ひろたか, 第4章)
愛媛大学医学部神経精神医学
小 椋 たみ子(おぐら たみこ, 第5章)
神戸大学文学部心理学
齊 藤 智(さいとう さとる, 第6章)
大阪教育大学教育学部
山 鳥 重(やまどり あつし, 第7章)
東北大学大学院医学系研究科高次機能障害学
大津由紀雄(おおつ ゆきお, 第8章)
慶應義塾大学言語文化研究所

運動と言語

認知科学の新展開 3

2001年9月25日 第1刷発行

編 者 いぬい とし お あんざいゆういちろう
乾 敏 郎・安西祐一郎

発行者 大塚信一

発行所 株式会社 岩波書店
〒101-8002 東京都千代田区一ツ橋 2-5-5
電 話 案内 03-5210-4000
<http://www.iwanami.co.jp/>

印刷・三陽社 カバー・半七印刷 製本・牧製本

© 乾 敏郎ほか 2001
ISBN 4-00-006783-4

Printed in Japan

運動と言語

運動と言語への招待

「運動と言語」は、認知科学の新展開の中でもっとも急速に進歩しつつある分野である。認知科学の今後の展開を考えたときにも中心的な分野となることは、まずまちがいないだろう。読者の一部には、「言語と運動」は「高次と低次」あるいは「中枢と末梢(出力)」の関係にあると思う人がいるかもしれないが、じつはこれらの間には密接な関連がある。

1948年9月、数学者であり「情報科学の父」といわれる von Neumann, 生理学者 Lorente de No, 心理学者 Wolfgang Köhler, Karl Lashley らが集まり Hixon Symposium が開かれた。このときの論文集が1951年に『行動の脳内メカニズム』(Cerebral Mechanisms in Behavior)として出版されている。その中で Lashley は、自分のもっとも複雑な行動の一側面に関心を持っており、それは思考と行為を論理的に順序づけて計画することであると述べた。そして彼は次のように書いている。

言語の組織化に関する問題は、他のほとんどすべての大脳活動の特徴でもあると私は考えている。それは一連の階層的組織化である。すなわち、単語の発音における発声運動の順序、各文における単語の順序、各節における文の順序、文章における節の合理的な順序、といった階層的組織化で

ある。発話のみならず、すべての熟達した行為は、到達運動や把持運動のようなある運動における筋収縮の時間的協調にいたるまで、時系列的な秩序化という同一の問題を含んでいるようである。より基本的な行為における順序の基礎をなす神経機構を分析すれば、最終的には、論理というものを生理学的に解明することさえできるかもしれない。

確かに Lashley のこの言葉は、本書の目指す方向にかなり近いものである。しかし、このシンポジウムから半世紀以上が経過した現在、きわめて多くの事実が明らかになり、Lashley の時代とはまったく異なるレベルでこれらの問題が論じられるようになった。特にこの半世紀の間に、言語理論、運動の計算理論などはめざましい発展を遂げた。そこでここでは、運動と言語に関する主要な問題とそれらの間の関連性について、できるかぎり簡潔にまとめる。

運動制御と言語

上肢は、肩、肘、手首の3関節で、7自由度を持つ。全身の粗いモデル (gross body model) では31自由度だが、指の関節なども考慮に入れた精密なモデルでは244の自由度がある。手の運動は本文で詳しく紹介されるので、ここでは発声運動を中心に、運動制御の問題を整理しよう。

ヒトは呼気流により咽頭にある声帯を振動させて原音を作る。そして、咽頭腔、喉頭腔、口腔、鼻腔からなる空間を声道と呼ぶが、声道の形状を変化させることにより共鳴特性が変わり、さまざまな音声を生成することができる。言い換えると、声道を通過することにより原音にフィルタがかけられて、さまざまな音声を作られるのである。ではどのように唇、顎、舌などの構音器官を動かして音声を作るのであろうか。

下顎および舌骨の運動においては図1のように、7つの筋肉群があり、4自由度で近似できる。舌は、舌骨につながる舌骨舌筋、下顎骨につながるオトガイ舌筋、軟口蓋につながる口蓋

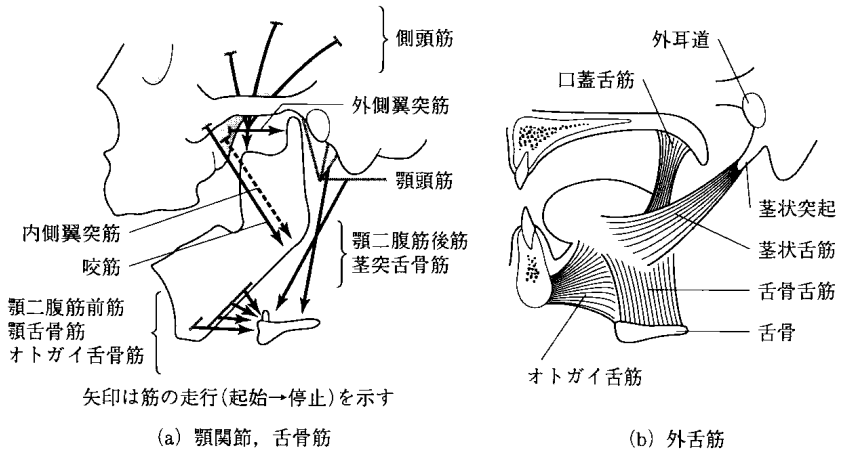


図1 下顎と舌骨の運動(澤島 2000 より)

舌筋, および茎状舌筋の4種によって前後および上下方向に移動する。舌の形状は舌の内部にある内舌筋によって制御される。まとめると, 構音器官の運動は以下のように9自由度ある。

- (1) 顎の上げ下げ
- (2) 顎に対して舌の上げ下げ
- (3) 舌体部に対する舌先端部の上げ下げ
- (4) 舌体部の前後方向の移動
- (5) 舌先端部の前後方向への移動
- (6) 上唇の上げ下げ
- (7) 顎に対して下唇の上げ下げ
- (8) 上下唇の前後移動
- (9) 口蓋の上げ下げ

しかし, 構音器官の簡易モデルとして使用される Maeda(1990)のモデルは,

下顎の位置, 舌の背面の位置, その形状, 舌尖の位置の4自由度と

唇の開き方(上下の幅と左右の幅), 咽頭の高さの3自由度のあわせて7自由度のモデルである。

運動学習と順逆問題

前項で述べたことから、構音器官の形状は7次元空間で表される。一方、聴覚音声空間は(ここでは紙面の関係で詳細は述べないが)3次元で表される。さて構音器官の形状が決定されると音韻が決まる。この変換を $f(\theta)$ で表すとする(図2)。発声の学習はこの逆変換、すなわち、 $x \rightarrow \theta$ の学習であると考えられる。ターゲットとなる音声 x を指定すると、それは聴覚音声空間では3次元の小さな領域 θ をなすと考えられる。このターゲット音声を出力するための構音器官の形状は7自由度なので、この変換は一意に決定できない(不良設定性)。このような関係は、**自由度の冗長性**と呼ばれている。そして脳は何らかの制約条件下でこの問題を解くことによって、構音器官の形状を決めているに違いない。しかし現実にはまだ冗長であり、一意に決定されていないらしい。実際、Hagiwara(1995)によると、英語の /r/ の発音において、その前後の音韻によって構音器官の形状が異なる場合がある(図3)。

以上をまとめると

- (1) 運動系は基本的に多自由度のシステムであり、したがって運動制御においては、逆変換がうまく計算されなければならない。
- (2) 運動制御には自由度の冗長性の問題があり、この不良設

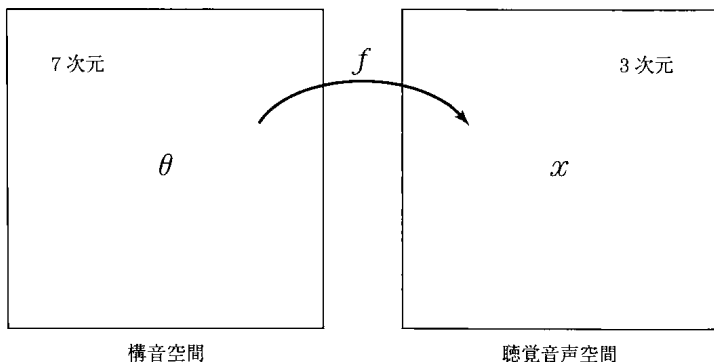


図2 構音空間から聴覚音声空間への変換

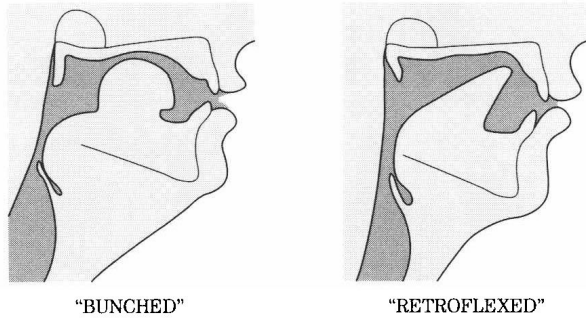


図3 /r/ の発音(Delattre and Freeman 1968 より)

定性をどのように解決しているのかを明らかにしなければならない。

言語の基本的特徴

N. Chomsky の有名な文「色のない緑のアイデアが猛然と眠る」を聞くと、我々は日本語として無意味であるにもかかわらず正しい文だと感じる。これは、統語範疇(品詞)の並び方が規則にあってからである。もし個々の単語の系列を覚えていたなら、このような配列に遭遇することは決してないから、正しい文だという判断はなされないであろう。我々が知っているのは語順ではなく統語範疇の配列の仕方なのである。さらに if-then, neither-nor など、ある程度はなれた語の関係(長距離依存)も正しく処理することができる。

If the student drinks a lot, then he goes to “Karaoke”.

Neither the woman who doesn't sing a song very well nor
the man who doesn't play instruments dislikes music.

そこで言語を処理するためには、

- (1) 個々の単語の属する統語範疇
- (2) 統語範疇の並び方
- (3) 長距離依存性

を知っていなければならない。

この点についてももう少し議論しよう。統語範疇(品詞)には日

本語の場合、名詞、助詞、形容詞、形容動詞、動詞、副詞、接統詞などがあり、英語の場合、名詞、動詞、前置詞、限定詞、形容詞、副詞などがある。このように言語によって統語範疇は異なる。さらに、語順も言語によって異なる。しかし、多くの言語では下記の2種類のうちのいずれかの語順になっている。

英語の語順	日本語の語順
VP: [_{VP} study music]	[_{VP} 音楽を学ぶ]
NP: [_{NP} professor of informatics]	[_{NP} 情報学の教授]
AP: [_{AP} fond of music]	[_{AP} 音楽が好きな]
PP: [_{PP} from Kyoto]	[_{PP} 京都から]

このように動詞句、名詞句、形容詞句、前置詞句は言語によっていずれかの順になる。句の中心をなす単語を X とし、句全体を **X** バーと呼ぶ。ここで X = V, N, A, P である。X は XP の主要部(head)と呼ばれる。したがって、それぞれの言語は、主要部先導型(head-initial)か主要部終端型(head-final)のいずれかになっている。

では、文はどのような構造を持っているのであろうか。以下のような生成規則によって英文を生成することができる(S は文, NP は名詞句, VP は動詞句, V は動詞, PP は前置詞句, Det は冠詞, N は名詞, PREP は前置詞をあらわす)。

- a. S → NP VP
- b. VP → V NP PP
- c. NP → Det N
- d. PP → PREP NP

ここで、矢印は(文脈とは無関係に)左の記号を右で書き換えよという意味である。実際の文を生成するには各統語範疇に属する単語と書き換えればよい。この規則は文脈に関係なく適用されるので、文脈自由文法(context free grammar, 略して CFG)と呼ばれている。ほとんどの自然言語の文は CFG で生成可能である(CFG については後述する)。

さて、生成規則(a)と(b)を適用すると

S → NP [V NP PP]

となる。たとえば,

[[_{NP} The man][_{VP} has [_{NP} nice PCs][_{PP} in his lab]]]

ここで、NP と VP は同じレベルであるが、最初の NP (主語) は次の NP (目的語) より高い階層にあると見ることができる。このとき、主語の NP は、目的語の NP を構成素統御 (c-command) するという。このように言語には階層構造がある。さらに、我々は以下のような二重、四重の埋め込み文を処理し、全体の意味を正しく理解することができる。

The man who likes the song which was sung by his friends
sings very well.

The man says that the woman who feeds the dogs who
chase the cat who I saw yesterday was looking for her CD.
したがって、単に階層構造を処理できるだけでなく、埋め込み
構造を正しく処理できる能力を持っている。こういった埋め込
み構造を処理できることが人間の持つ高い創造力の基礎になっ
ていると考えられる。

一方、動詞の意味を成立させるのに、最低必要な統語的意味
役割 (θ -role という) の集合を項構造という。主な θ 役割として、
行為者 (agent), 受動者 (patient), 主題 (theme, 動作によって
移動する物), 起点 (source), 着点 (goal), 場所 (location), 道
具 (instrument) などがある。例を挙げておく。

- 1 項必要 walk: (agent), die: (theme)
- 2 項必要 kill: (agent, theme), hit: (agent, theme),
chase: (agent, theme), come: (agent, source)
- 3 項必要 give: (agent, theme, goal),
send: (agent, theme, goal),
bring: (agent, theme, goal),
put: (agent, theme, location)

このように、動詞もしくは動作情報にはそれが意味を持つた
めに必要な名詞のカテゴリが決まっており、これらの知識も言語

システムではうまく獲得，処理しているのである．

まとめると，言語の重要な特性として，

- (1) 個々の単語の属する統語範疇
- (2) 統語範疇の並び方
- (3) 長距離依存性
- (4) 動詞が意味をなすために必要な統語的意味役割の集合
(項構造)
- (5) 埋め込み構造

があり，これら进行处理できるシステムが脳にある．

リカレントネット，CFG，チューリング機械

前述の文脈自由文法(CFG)は有限状態オートマトンでは識別できないが，プッシュダウンオートマトンによって文法に適合しているかどうかが判定できる．このことは，言語が有限個の状態からなる状態遷移図によっては記述できないということを意味しており，CFGを扱うには何らかのワーキングメモリが必要だということである．別のCFGの例をあげよう．いま (1) $S \rightarrow 0A1$ ，(2) $A \rightarrow 0A1$ ，(3) $A \rightarrow \epsilon$ という3つの規則を仮定する．ここで ϵ は空記号で，これは消去を意味する．規則(1)を1回，つづいて規則(2)を2回用いたのち規則(3)を用いると，系列0001111が得られる．この文法が導出する系列は0が n 回続き，その後1が n 回続く．このような系列が規則に合っているかを識別するには，プッシュダウンオートマトンが必要となる．

人工ニューラルネットワークの研究から，ある種の構文処理が，ワーキングメモリを持つ再帰型ネットワーク(recurrent network)で可能なことが示唆されている(図4)．図に示された長方形はニューロンの層で，左の数字はそれぞれの層にあるニューロンの個数を表している．このネットワークでは，実線の部分を除き，破線で示されている前向きの結合しかない．しかも，前段の各ニューロンは次段のすべてのニューロンと結合し

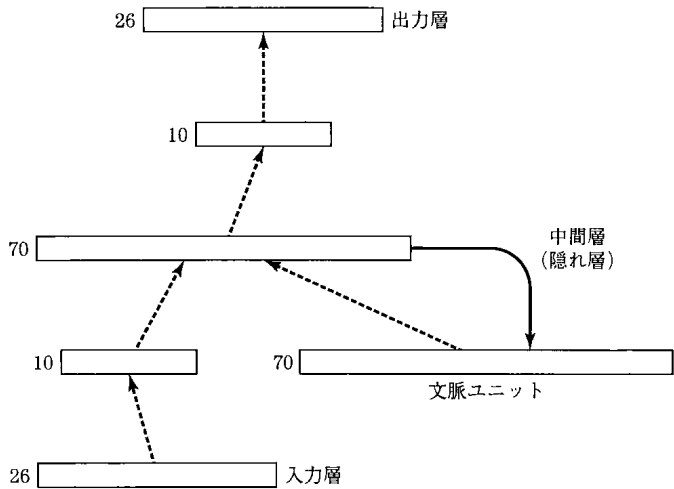


図4 エルマンネット

ている(全結合)。ここでこのネットワークでは入力層のニューロンの数と出力層のニューロンの数が同じであることに注目してほしい。入力層と出力層の各ニューロンはそれぞれ1つの単語を表現している。このネットワークに文を構成する各単語を時系列的に入力する。それぞれの単語が入力されると、出力層は次に入力されるであろう単語を予測して出力するのである。当然ながら、次の単語を正確に予測することは不可能である。しかし、次に入力される単語の統語範疇や長距離依存を含む単数・複数のかかりうけなどを正しく処理することができる。このような人工ニューラルネットワークを学習によって構成できることを Elman (1991, 1993) が初めて報告し、このニューラルネットワークはエルマンネットとよばれている。ここでは、学習アルゴリズムよりもこうしたネットワークによって構文構造を潜在的に理解、処理可能である点を強調したい。

すでに述べたように、文は単語系列では学習できず、長距離依存を含む統語範疇の並びを正しく認識しなければならない。さらに、関係節などの埋め込み構造を正しく処理しなければならない。このようなネットワークを用いると、埋め込み構造も

含めて自然言語の文法を処理できることが明らかにされた(乾 1997b, 1998b; 玉森, 乾 1999). さらに, 格関係も処理可能である(玉森, 乾 1998).

なぜエルマンネットではこのようなことが可能なのか. それは, 図中に示されている文脈ユニットが重要な役割を果たしているためである. この文脈ユニットは, 一時刻前の中間層(70個のニューロンがある層)の状態(各ニューロンの活動)をそっくりコピーし, 記憶している. 次の単語が入力されるとそれが処理されてこの中間層に入力されるが, 同時にコピーされ記憶されている一時刻前の中間層の状態もこの中間層に入力されることによって正しく次の単語の統語範疇を出力できるのである. この中間層のニューロン活動の状態(およびそのコピーである文脈層の状態)は, 文を構成する単語の長い系列の履歴を含んでいると考えられる. したがって, このような再帰構造をもつネットワークによって長距離依存の処理もできるようになるというわけである.

Rodriguez ら(1999)は, このネットワークがある種の文脈自由文法(CFG)を処理できることを示した. 現在, すべての文脈自由文法がこのような構造をもつネットワークで処理可能であるかどうかの検討が進められている. また, Siegelmann (1995)は, 再帰結合をもつより一般的なニューラルネットワークによってチューリングの限界を超える計算も可能であると主張している. 一般にニューラルネットワークの中に再帰結合があると, それによって活動のダイナミクスが生ずる. このような構造をもつことによって上述のような文脈依存的な処理などが行えるが, まだ一般の再帰結合をもつニューラルネットワークの能力に関する理論的な解析が進んでいないのが現状である.

動詞, 動作, 名詞

■ 動詞と動作

物体の絵に対してそれに関連する動詞を答えさせる(たとえ

ば、椅子に対して「座る」を答えさせると、ブローカ領野および右の小脳が強く活動した(Martin *et al.* 1995)。下前頭回44野および45野をブローカ野とよぶ(図5)。進化的には44野はサルのF4に対応し、45野はF5に対応すると言われている。サルでは、F4は主として到達運動(reaching)、F5は把持(grasping)の制御を行っていると考えられている。しばしばブローカ野に障害があると機能語(たとえば前置詞、格助詞)やある種の構文解析に欠陥が見られる(たとえば、乾 1998b; Elman *et al.* 1996)ことから、この領域が発話や文法に重要な役割を果たしているとされてきた。知らない外国語を聞いてもブローカ野は活動しないが、理解可能な外国語を聞くとブローカ野が活動する(Mazoyer *et al.* 1993; Nakai *et al.* 1999)。Martin ら(1995)の研究は、ブローカ野が動詞の文法的な構成に重要な役割を果たしていることを示唆している。

一方、被験者に道具名(たとえば、ハンマーや鉛筆、などの基本レベル)を答えさせると、下前頭回を中心に活動が見られた(Grabowski *et al.* 1998; Okada *et al.* 2000)。この活動のなかで、最も活発に活動したのは、下前頭溝および中心前溝の交点

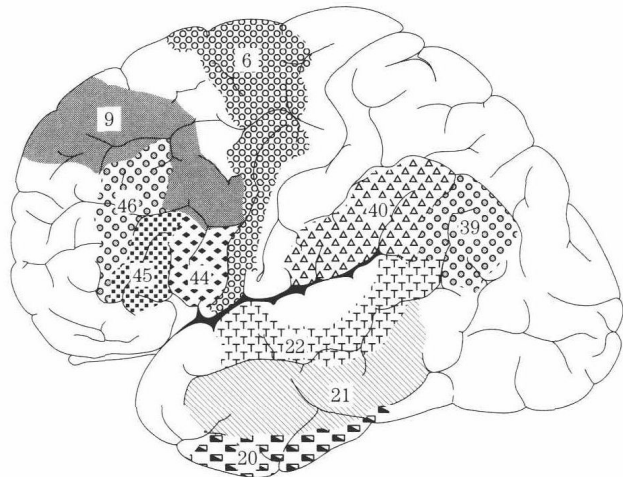


図5 言語に関わる活動部位

付近であり、9野と44野の境界付近であった(図5)。Grabowskiら(1998)は、この位置をLPREMと名付けた。またLIFGの活動も見られた。LIFGは45野におよそ対応する。この部分は言語課題、特に単語の選択と生成において活動することがこれまでも報告されてきた場所である。

■ 動詞と名詞

失語症研究から動詞の生成および検索には下前頭回が、また名詞の生成および検索には側頭葉が重要であると考えられている(たとえば、Damasio and Tranel 1993)。また、fMRIなどのイメージング研究などの結果を考慮すると、動詞(動作を表す単語)および名詞(物の名称)は、それぞれブローカ野を含む下前頭回、および側頭回に記憶されていることが推測される(乾1997a)。

解剖学的には45野は下側頭葉と連絡している。イメージング研究から下側頭葉は視覚情報のカテゴリ化を行っていると考えられている(Martin *et al.* 1996; Damasio *et al.* 1996)。このカテゴリ情報はさまざまな経路を経て、さまざまな属性情報と連合している。45野がこの領域と連絡しているということは、言語の統語解析(系列分析)と意味解析の統合の場でもあることが示唆される。次に統語解析とブローカ野の関係について考察する。

■ ブローカ野と頭頂葉との連関

ブローカ野も左の40野も、手の運動と音韻処理に関わっているようである。ブローカ野の活動が見られた課題を表1に記す。筆者(乾1998b, 2000a, b, c)は、44野は(手と構音器官の)運動系列の生成に関与しており、45野は運動系列の認識に関与していると考えた。ここでの「運動系列」とは、一次運動野などが出す個々の筋肉への指令信号を意味しているのではなく、より高次の「アクションユニット」とでも呼ぶべきものである。つまり、より複雑な指令信号をまとめたチャンクの符号であると考えられる。これはおそらく視覚的あるいは聴覚的に分節(seg-

表1 左の44野および45野で活動が見られた課題

左44野	左45野
音韻の短期記憶	単語のリストや文を聞く
内言(internal speech)	単語を構成する音韻の検出課題
黙読	単語リストの記銘時
手の運動イメージの生成	把持動作の観察
手のパターン認知	視覚刺激の移動ルールの顕在学習時
音楽のリズム判断課題	鏡映描写の学習の初期
	三段論法の推論

mentation)された単位に対応している。

左の40野は、音韻のワーキングメモリの働きがあると同時に、手の運動系列の生成・学習時にも応答する。これは何を意味しているのであろうか。筆者(乾1998a)は、以下のように考えた。ウェルニッケ野(22野の後部)は、単語音のパターン認知、より正確に言えば単語音の識別を行っており、この部位では、単語音の聴覚的表現が作られているものと思われる。左の40野はこの符号を入力として短期的に構音ループによって記憶保持されているものと考えられる。ここで重要な点は、表現の一貫性である。44野が運動系列の生成に関与しているとすると、40野との相互作用によって、それは聴覚情報に変換されねばならない。このような変換によって、44野で生成された内言が40野を経由して、ウェルニッケ野を活性化させ、なおかつ高次聴覚野も活性化させられるのである。これによって、内言でも自分の声が聞こえることになる。これは手の運動系列の生成においても同様で40野との相互作用によって運動-視覚変換がなされる。つまり、40野は44野との相互作用を行いながら、生成過程(メンタルシミュレーション)において系列の文脈情報を与えるのである。ここで45野には、直接側頭、後頭葉から長連合線維によって視覚情報や聴覚情報が入力される。45野ではこれらの情報を入力して、次の情報を予測する。予測された情報は45野と44野の相互作用によって運動情報(アクションユニット)へ変換される。