

## 解説



## 頭の中に単語はいくつある？†

天野 成 昭††

## 1. はじめに

我々はいくつの単語を知っているのでしょうか？すなわち、我々の頭の中にはいくつの単語が詰まっているのでしょうか？試しに約22万語を収録した国語辞典をパラパラめくってみると、知らない単語があちこちに目につく。ということは、頭の中の単語数はきっと22万個よりも少ないはずだ。一方、小学校で習うことになっている教育漢字は1,006種類で、それぞれの漢字を含むような単語を最低1つは知っているだろう。だから、頭の中の単語数はきっと1,006個よりも多いはずだ。このように単語数の上限や下限を粗く推定することは可能である。しかし、頭の中の単語数は本当はいくつ？と聞かれると正しい答えを出すのはなかなか難しい。

さて頭の中の単語数を推定するときに、やっかいな問題点がいくつかある。問題点の1つは頭の中の単語数に個人差があることである。たとえば博学なAさんは“刮目（かつもく）”という難しい単語を知っているが、薄学なBさんはこの単語を知らないという状況は往々にしてある。これはつまり、Bさんの頭の中の単語数よりもAさんの頭の中の単語数のほうが、少なくとも1個多いことを示しているのである。このように頭の中の単語数には個人差があるということが、その数の推定を難しくしている。

もう1つの問題点は、頭の中の単語数が時間とともに変化することである。たとえばしばらく前までは、多くの人が“不借身命（ふしゃくしんみょう）”という単語を知らなかったはずである。しかし、ある相撲取りがこの単語を使ったことをきっかけに、今やほとんどの人がこの単語を知って

いる状態になってしまっている。つまり我々の頭の中の単語数が増えたのである。これとは逆に頭の中の単語数が減ることもある。たとえば、“国電”という単語は昔ならば皆がよく知っていた。しかし日本国有鉄道の民営化以後この単語は使われなくなり、現在では多くの人の頭の中から消えてしまっている。

このように頭の中の単語数の推定には、個人差や時間による変動の問題が存在する。しかし、これらの要因によって変動する単語数の割合は、頭の中の単語数の全体からみればそれほど大きくなく、変動しない部分、すなわち共通部分の方がずっと大きいはずである。なぜなら、もし個人差による変動が共通部分の単語数をしのぐほど大きいとすると、それは共通部分の単語数が少ないことを意味し、そのような少ない単語数によるコミュニケーションは、著しく不便になるはずである。ところが実際にはこのような不都合は起こっていない。これと同様に、もし時間変動が共通部分の単語数をしのぐほど大きいとすると、過去の言語的情報を現時点で理解し利用することが、極端に不便になるはずである。しかし実際にはこのような不便は生じていない。このように、たとえ個人差や時間による変動の問題が存在するにしても、その割合は大きくないと考えられるので、頭の中のおおまかな単語数の推定は、一応可能であると考えられる。

## 2. 頭の中の単語数の推定方法

## 2.1 文字単語の数

ではどのような方法で頭の中の単語数を調べることができるのでしょうか？よく用いられる方法は、まずある辞書から単語をいくつかランダムに選び出してテストセットを作り、そのセットの各単語を知っているかどうかを“はい”―“いいえ”で被験者に判断してもらう。そして式(1)で示す

† How Many Words Do We Have in a Brain? by Shigeaki AMANO (NTT Basic Research Laboratories).

†† NTT 基礎研究所

ように、テストセット内で“はい”と判断された単語の割合を辞書の全単語数にかけることによって、単語数を推定するという方法である。

$$\text{推定単語数} = \frac{\text{はいと答えた単語数}}{\text{テストセットの単語数}} \times \text{辞書の全単語数} \quad (1)$$

このようにして単語数を求めた主な研究結果を示そう。日本語を対象とした研究では森岡<sup>1)</sup>が、15人の義務教育修了者に37,970語の単語を知っているかどうかを判断させる実験を行い、頭の中の単語数を約30,000と推定している。他には阪本<sup>2)</sup>が、20歳の男性490人に500語を判断させる実験を行って、頭の中の単語数を約50,000と推定している。英語を対象とした研究では、Nusbaum, Pisoni, & Davis<sup>3)</sup>が、19,750単語を395単語ずつに分けたセットを、600人の大学生にランダムに割り振って、そのセットを判断させる実験を行っている。彼らは約14,000単語が頭の中にあることを示唆する結果をえている。このほかには、Aitchison<sup>4)</sup>が約50,000単語、Seashore & Eckerson<sup>5)</sup>が約150,000単語が頭の中にあると推定している。

このように推定された単語数は大きくばらついている。この原因の1つは式1で推定単語数を求めた場合、“はい”と判断された単語の割合が同じであっても、辞書の全単語数が大きいほど推定単語数が大きくなるという傾向があることである。推定された単語数が大きくばらついていることから分かるように、頭の中の単語数はまだ正確には把握されておらず、定説がないというのが実情である。

さらに問題点をあげるとすれば、これらの単語数が文字で書かれた単語（文字単語）を使って推定された結果だということである。すなわち、音声による単語（音声単語）を使って単語数を推定した場合、同じ結果が得られる保証がないのである。文字単語であろうと音声単語であろうと、推定結果は同じだろうと思う人がいるかもしれない。しかし、ごく最近の研究によれば頭の中の単語数と密接に関係すると考えられている単語のなじみの程度（親密度）が、文字単語と音声単語とで同一ではなく、両者の違いが少なからず存在することが分かっている<sup>6)</sup>。したがって、頭の中の音声単語の数は、文字単語とは異なった方法で求

める必要がある。

### 2.2 音声単語の数

音声は文字と違って時間の経過とともに消え去ってしまうという特徴を持っている。したがって、音声単語を聞き漏らさないためには、我々の頭の中ですばやく照合処理が行われなければならない。そのような音声単語の知覚処理を Marslen-Wilson ら<sup>7),8)</sup>は Cohort モデルとしてモデル化している。Cohort モデルの処理過程を以下に述べよう。まず最初に音声の開始部分（100～150ms）と語頭部分が一致する単語候補のグループ（cohort）ができる（図-1）。この処理段階は入力された音声、すなわち下位レベルからのボトムアップ情報のみに基づいて行われる。次に、時間軸に沿って音声と cohort 中の全単語候補との照合処理が、音韻を単位として逐次的に行われる。つまり、音韻が1つ入力されるごとに、それまでの音韻列と各単語候補の語頭からの音韻部分列との照合が行われる。この照合において不一致が判明した単語候補は cohort から削除され、cohort 中の単語候補数が徐々に絞り込まれてゆく。この処理段階では、上位レベルからのトップダウン情報も利用され、構文的、意味的に不適切な単語候補も cohort から削除される。このような処理を継続していくと、最終的に cohort 中の単語候補が1つになる。この時点ユニークネスポイントと呼び、このような時点において単語が知覚され

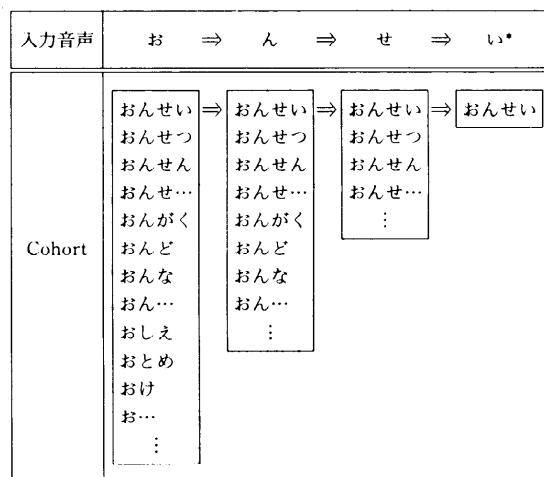


図-1 Cohort モデルの処理過程の概念図、音韻列を語頭部から順に照合し、cohort 内の単語候補が1つになる点(ユニークネスポイント：図中\*で示す)で単語が知覚される。

る。したがって、このモデルでは、時間軸に沿った left-to-right の音声処理がなされ、上位レベルの情報も利用した all-or-nothing の単語選択過程によって、最適効率で単語が知覚されることになる。

Cohort モデルのユニークネスポイントは、辞書を定めると計算によって各単語ごとに一意にその値を求めることができる。そして値を求めるときに使う辞書が大きいとユニークネスポイントは単語の後ろの方になり、逆に辞書が小さいとユニークネスポイントは単語の前の方になるという性質がある。この性質を利用して天野<sup>9),10)</sup>は、音声知覚に使われる頭の中の音声単語の数を推定した。彼は、まず約6万個の音声単語について、その親密度を7段階で評定させるという実験を行った<sup>6)</sup>。この実験でえられた親密度の値の順に約6万個の単語を並べると、ある一定値  $x$  以上の親密度を持つ単語の数すなわち辞書の大きさ  $y$  は、親密度  $x$  の関数として表すことができる。この親密度  $x$  の関数として辞書の大きさ  $y$  を変化させ、Cohort モデルが予測する単語知覚時点すなわちユニークネスポイントと、心理実験で測定された単語知覚時点との相関が最大になるときの辞書の大きさ  $y$  を求め、それを頭の中の単語数の推定値とした (図-2)。それによれば、頭の中に約27,000個の音声単語が存在するという結果がえられている。

ただし、この推定のよりどころとなった相関の最大値は約0.5であり、それほど高くない。したがって約27,000個という推定値が本当に正しいかどうかについては、まだ検討の余地が残されている。さらに、この推定値は単語であるかどうかを音声で聞こえたらただちに反応するという実験状況、すなわち時間的拘束がかなり強い状況における推定値である。音声を聞いてからじっくり考える余裕がある場合のように時間的拘束が緩い場

合には、長期記憶に存在する情報を意識的に利用する割合が増えると予想されるので、これよりも大きめの推定値がおそらく得られるであろう。

### 3. ま と め

生まれたばかりのとき、我々の頭の中には単語は1つもなかったはずである。しかし大人になったとき、ここで示したように少なくとも数万の単語が頭の中に詰まっていると考えられる。単語をまったく知らない乳児がどのような仕組みで単語を覚えていくのかという発達の問題は、まだ完全には解明されておらず、現在多くの言語心理学者がこの問題に取り組んでいる。一方このような発達の対極にあるのが、頭の中の単語が徐々になくなっていく老化の現象である。その極端な例がお年寄りの痴呆症である。最近の高齢化にともなって、この問題がクローズアップされつつあるのは周知の事実である。老化にともなう単語の記憶の崩壊の仕組みを明らかにすることによって、それを未然に防ぐための対策も立てられるようになるであろう。

さて、あなたの頭の中にはいくつの単語が詰まっているのであろうか？4万個？それとも5万個？付録に単語の親密度<sup>6)</sup>を利用して作った簡単なテストを載せてあるので、それを使ってあなたの頭の中の単語数を調べてみてはいかがだろうか。

### 参 考 文 献

- 1) 森岡健二：義務教育終了者に対する語彙調査の試み，国立国語研究所年報，Vol. 2, pp. 95-107 (1951)。
- 2) 阪本一郎：読みと作文の心理，牧書店 (1955)。
- 3) Nusbaum, H. C., Pisoni, D. B. and Davis, C. K.: Sizing up the Hoosier Mental Lexicon: Measuring the Familiarity of 20,000 Words, Research on Speech Perception, Progress Report, Indiana University, Vol. 10, pp. 357-376 (1984)。
- 4) Aitchison, J.: Word in the Mind (2nd ed.), 290 p., Blackwell, Oxford (1994)。
- 5) Seashore, R. H. and Eckerson, L. D.: The Measurement of Individual Differences in General English Vocabularies, The Journal of Educational Psychology, Vol. 31, pp. 14-38 (1940)。
- 6) Amano, S., Kondo, T. and Kakehi, K.: Modality Dependency of Familiarity Ratings of Japanese Words, Perception & Psychophysics,

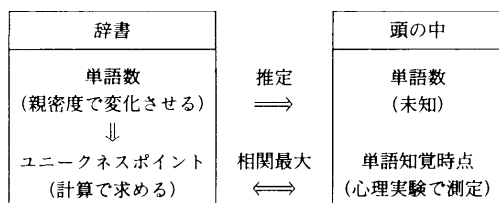


図-2 頭の中の音声単語数の推定方法の概念図

Vol. 57, pp. 598-603 (1995).

- 7) Marslen-Wilson, W. D. and Welsh, A.: Processing Interactions and Lexical Access During Word Recognition in Continuous Speech, *Cognitive Psychology*, Vol. 10, pp. 29-63 (1978).
- 8) Marslen-Wilson, W. and Tyler, L. K.: The Temporal Structure of Spoken Language Understanding, *Cognition*, Vol. 8, pp. 1-71 (1980).
- 9) 天野成昭: 単語知覚に使われる心的辞書の大きさの推定, 日本音響学会秋季研究発表会講演論文集 I, 1-9-12 (1994).
- 10) 天野成昭: 心的辞書に含まれる単語数の推定, 日本音響学会春季研究発表会講演論文集 I, 3-9-7 (1995).

(平成7年5月25日受付)



### 天野 成昭

1959年生。1983年東京大学文学部卒業。1985年同大学院人文科学研究科修士課程修了。同年日本電信電話(株)電気通信研究所入所。同基礎研究所勤務。音声知覚研究に従事。日本心理学会、日本音響学会、日本神経回路学会、米国内音響学会各会員。

### 付 録

#### はじめに

頭の中の単語数のことを一般に語彙数と呼ぶ。このテストは語彙数を測定するためのテストである。

#### 実施方法

表-1のテスト項目の単語を知っているかどうかを“はい” - “いいえ”で番号順に判断する。“いいえ”の判断が3回連続して続いたらテストを中止する。そして“はい”と判断された単語のうち最も番号が大きい単語における推定語彙数(表の右端に書かれた数)をテスト結果として採用する。

#### 注意事項

この語彙テストは文字で表した単語の親密度<sup>6)</sup>を利用して作成してある。したがって、この語彙テストの結果は、あくまでも文字単語の語彙数を表し、音声単語の語彙数を表してはいない。つまり、この語彙テストを音声で行っても無意味であり、その結果は信頼できないことに注意してほしい。

表-1 単語親密度を利用した語彙テスト

番号	テスト項目	推定単語数
1	卒業 (ソツギョウ)	194
2	社長 (シャチャウ)	674
3	旅行 (リョコウ)	1,101
4	迷惑 (メイワク)	1,606
5	立派 (リップパ)	2,991
6	林檎 (リンゴ)	3,926
7	腕相撲 (ウデズモウ)	4,950
8	例年 (レイネン)	6,314
9	列車 (レッシュヤ)	7,653
10	論理 (ロンリ)	9,093
11	恋心 (コイゴコロ)	10,541
12	流域 (リュウイキ)	11,922
13	冷酒 (レイシュ)	13,425
14	老眼 (ロウガン)	14,887
15	免税 (メンゼイ)	16,222
16	恋仲 (コイナカ)	17,566
17	旅芸人 (タビゲイニン)	18,916
18	論評 (ロンビョウ)	20,153
19	牢獄 (ロウゴク)	22,694
20	浪花節 (ナニワブシ)	23,933
21	利潤 (リジュン)	25,282
22	満場 (マンジョウ)	26,635
23	変奏曲 (ヘンソウキョク)	29,209
24	隼的 (リョウテキ)	30,652
25	風窓 (カザマド)	31,991
26	有理数 (ユウリスウ)	33,341
27	撥音便 (ハツオンビン)	36,129
28	眉尻 (マユジリ)	37,606
29	来征 (ライセイ)	39,076
30	遊歴 (ユウレキ)	40,513
31	料飲 (リョウイン)	41,990
32	略解 (リョウカイ)	43,457
33	烈烈 (レツレツ)	44,933
34	流俗 (リュウゾク)	46,372
35	老健 (ロウケン)	47,766
36	曆象 (レキシヨウ)	49,141
37	靈剣 (レイケン)	50,482
38	訛音 (カオン)	51,770
39	疼痛 (トウツウ)	52,956
40	目迎 (モクゲイ)	54,114
41	袖 (ソマ)	55,123
42	雄竹 (オダケ)	56,070
43	鬱然 (ウツゼン)	56,916
44	結縮緬 (ロチリメン)	58,404
45	鷗 (アギト)	59,079
46	摩利支天 (マリシテン)	59,662
47	洒洒落落 (シャシャラクラク)	60,659
48	蓮卒 (ラソツ)	61,036
49	寸善尺魔 (スンゼンシャクマ)	61,360
50	拈華微笑 (ネンゲミシヨウ)	61,799