〔書　評〕

国立国語研究所編

『テレビ放送の語彙調査Ⅰ～Ⅲ』

石　野　博　史

1　はじめに

本書は日本で初めて行われた本格的テレビ語彙調査の報告書である。

これまで一定以上の規模を持つ日本語語彙調査を中心的に推進してきたのは国立国語研究所である。1950～1960年代にかけて行われた婦人誌と総合誌，それに雑誌九十種の語彙調査，60年代から70年代にかけての新聞の語彙調査，80年代の教科書の語彙調査がそうであるが，本書のテレビ語彙調査もこの系列に属する。

初期の語彙調査は手作業中心の，せいぜいで電動計算機を使う程度の面倒な調査であったが，今はもちろんコンピューターを使用する。そのコンピューターも最初の新聞語彙調査のころから見れば，信じられないほどに小型化し，かつ高性能のものとなった。したがって，一見調査は楽になったようにも思われるが，本書を見ると，実際は必ずしもそうでないことが分かる。ことに，以前の雑誌・新聞・教科書とは違い，テレビは言語的には音声主体のメディアである。方法論的にも従来はなかった問題が多い。担当者には依然骨の折れる調査であったと思われる。

さて調査は，1989年4月～1990年3月の1年間に，東京をキーステーションとする6放送局7チャンネルが放送した全放送番組中に出現した語彙を対象として行われた。そのうちの3か月分について分析処理が終わり，95年から99年にかけて本報告書が刊行された。初めに各冊の大体について述べる。

2　報告書の概要

（1）「報告Ⅰ」

最初の報告書（「報告Ⅰ」）は「方法・標本一覧・分析」というサブタイトルが付けられている。要するに担当者たちが「いかに調査したかを詳細に記す」ことを最大の目的として書かれたものである。そして実際，調査の目的から対象，調査ならびに分析の方法や結果の一部に至るまでが懇切丁寧に述べられている。一般に人は調査の手順よりも結果のほうに関心を抱く。それだけにこの「報告Ⅰ」は少々読みづらいのではないか。テレビで一番多く使われている語のトップ・テンはこれこれ，といった調子の歯切れのよさは，ここにはない。しかし，今回の語彙調査が結局どのようなものであったかを理解

するには，この報告書を通読するのが一番である。

（2）「報告II」

第2の報告書は，非常に大部であるが，中身は事実上すべて語彙表である。全部で20種類の語彙表が載っている。手早く結果を知りたい人のために，主な語彙表から，頻度数1〜10位を紹介しておこう。なお，「本編」とは「CM」以外の部分の意である。「本編」「CM」とも音声と画面文字（記号を含む）に分けて数字があがっている。

〈本編・音声度数順〉（数字は各項左が出現度数，右がパーミル＝千分率の値）

①「ええ」（「え」も）3268，31.7，②「はい」2346，22.8，③「うん」（「ん」も）1922，18.6，④「あの」1517，14.7，⑤「ああ」（「あ」も）1404，13.6，⑥「そう」1348，13.1，⑦「事」1325，12.9，⑧「する」1230，11.9，⑨「これ」1181，11.5，⑩「この」1120，10.9

〈本編・画面度数順〉（数字については同上）

①「0（零）」454，22.4，②「（」403，19.9，③「）」402，19.9，④「1」289，14.3，⑤「2」252，12.4，⑥「・」175，8.6，⑦（記号）「曇」167，8.2，⑧「5」165，8.2，⑨「3」163，8.2，⑩「、」162，8.0

〈CM・音声度数順〉（数字は度数）

①「よい」（「いい」も）110，②「なる」（成・為）91，③「新発売」70，④「この」69，⑤「おいしい」65，⑥「する」62，⑦「新しい」57，⑧「あなた」52，⑧「わたし」52，⑩「これ」50

〈CM・画面度数順〉（数字は度数）

①「。」368，②「・」209，③「、」191，④「（」180，④「）」180，⑥「新発売」107，⑦「○R」（丸にR）79，⑧「よい」（「いい」も）57，⑨「！」54，⑩「税」49

以上，4種の頻度上位語を掲げた。〈本編・音声〉の1〜3位が「ええ」「はい」「うん」という応答語で占められている点，〈CM・音声〉の10位内に「新発売」「おいしい」「新しい」といった語が顔を出している点，〈本編〉〈CM〉を通じて〈画面〉に記号類が多い点などに，テレビ語彙の特徴がうかがえるであろう。

（3）「報告III」

第3の報告書は，調査を担当した研究員による論文5編を収める。個人的な論文であり，報告臭の少ない分，読みやすい。それぞれの論文については，冒頭の「この報告書の目的」に要領よく述べられている。ここでは題名と内容の要点のみを簡単に記す。

中野洋「高頻度語から見たテレビ放送語彙の特徴」――過去に国語研究所が行ってきたいくつかの語彙調査における高頻度語と，今回の調査におけるそれとを比較し，どの調査においても，使用比率の最も高いクラスの語彙にはいわゆる基本語彙が多く含まれること，また同時に各調査に特徴的な語彙も含まれることを見いだしている。「報告III」のいわば総論的な位置を占める論文である。

山崎誠「用語類似度による標本（番組）分類」――〈本編・音声〉の300余りの標本（＝放送番組）間の語彙類似度を計算し，多変量解析を行ってそれらを分類した。つまり語彙

類似度という言語内的要素を用いた放送番組分類の試みである。結果は,「書きことば型」(代表例はニュース),「疑似話しことば型」(ドラマなど),「話しことば型」(対談・インタビューなど)の3グループが得られた。番組の内容や話題の類似性でなく,伝達形式の類似性によって分かれた点が注目される。

石井正彦「番組ジャンルの特徴語とジャンル間の特徴」――まず今回調査用に特に設けられた(言語外的基準による)7つの番組ジャンルのそれぞれから,そのジャンルに偏って使用される度合いの大きい語(特徴語)を各190語程度抜き出し,『分類語彙表』を用いて意味分野別に分けた。次にそれを多変量解析にかけてジャンル間の関係をとらえようと試みた。その結果,特徴的なものとして,「公的情報の伝達」のための語使用を特徴とするジャンル(報道・スポーツ)と,「私的日常世界の構成とその会話的表現」のための語使用を特徴とするジャンル(ドラマ・バラエティー),および両者の中間的なものの三者が見いだされた。山崎論文と同様の番組分類であるが,意味分類を利用した分,内容や話題面を重視した番組分類になっていると言える。

大島資生・小沼悦子「話者の属性から見た特徴語」――番組に登場する人物の性別・年代・職業の違いが,語使用にどのように反映されているかを見たものである。それによれば,例えば性別では,男性はスポーツ用語,女性は気象・食物・衣料用語が多い,年代では,20～30代は野球用語,40代は相撲用語が多いといったように,それぞれの話者が番組の中で演じている役割や個性を反映した語使用になっていることが明らかになった。テレビにおいては,個人としての話者でなく,役割としての話者の側面がより鮮明に前面に出てくるということである。

石井正彦・小沼悦子「語形の変異とその使用」――「報告II」の各語彙表は,同語と認定された語については,実際の出現語形でなく,統一語形のみを掲げている。本論文はこの点を補うため,自立語部分に変異形が見られた166語(異なり語形378)について,その使用状況を調べたものである(一例をあげると,接続語「では」は〈本編・音声〉での度数240であるが,その内訳は,「デワ」52,「ジャ」122,「ジャー」66となっている)。一部のものについては語形変異の原因をも探っている。その結果,ドラマやアニメなどのストーリー系の番組や,登場する話者の人数が多い番組は,話しことば的な変異形が多い(=話しことば度が大きい),女性のほうが男性よりも,またタレントのほうがアナウンサーよりも話しことば度が大きい,などのことが確かめられた。

以上が全体の概観である。次に「報告Ⅰ」から,重要と思われるポイントをいくつか選び,問題点の検討等を行う。

3　主な内容の検討

〈調査の目的〉

この語彙調査の目的としては,次の3点があげられている。(1)現代のテレビ放送に

おいて，どのような語が，どのような場合に，どのくらい使われるのか，その実態を調べる，(2)テレビ放送の語彙をとおして，現代日本語の語彙のありさまを追究する，(3)テレビ放送がその音声と画面をとおして伝える語を，どのように調査するか，その方法論を確立するための基礎を築く。これらの目的は至極もっともなものである。そして多分，当初はこの順序で重要度が考えられていたのであろう。しかし，結果的には第3の目的が最も重要視されるに至ったと言えるように思われる。というのは，初めの計画では1年間分のサンプル・データを分析する予定であった。それが作業上の理由で4分の1に縮小された。その結果，統計的に十分な分析をするには，データ量がいささか不足し，第1の目的には不満を残すこととなったのである。第2の目的は第1の目的に依存するから，これも当然不十分となる。

〈調査の対象〉

調査対象（母集団）は，NHK 2チャンネル（総合・教育）と民放5チャンネル（日本テレビ・TBS・フジテレビ・テレビ朝日・テレビ東京）が1年間に放送した，CMを含む全番組の語彙である。調査当時のテレビ放送状況からは当然の選択であった。ところが，当時と今とでは，テレビ放送のありように大きな変化が生じてきている。BSやCSの登場でチャンネル数が飛躍的に増え，特定ジャンルの番組だけを放送するチャンネルもできた。さらにデジタル化の本格的導入があり，インターネットとのからみもあって，テレビはかつての「送りっ放し」のメディアから，むしろデータ通信的に，視聴者1人1人が選択・構成して受容するメディアへと変貌を遂げつつある。したがって将来，新たにテレビ語彙調査を実施する際には，こうした点を考慮に入れて対象を選び直す必要があろう。その点，今回の調査には少々時期的な不具合があったと言わざるをえまい(ただし，地上波テレビ時代を総括する調査としての意義は残るし，方法論的な貢献はむしろ今後のものである)。

〈調査の方式〉

調査は1標本を5分とし，1日4標本，1週28標本を52週分，抽出比504分の1で母集団（＝放送番組）から抽出する方式で行われた。ただし，抽出は単純なランダム・サンプリングではなく，週・曜日・時間帯・チャンネルがほぼ等量となるよう工夫が加えられている。標本数1456，推定標本語彙数が音声約45万語，画面文字約12万語，この全部を分析する予定であったが，既述のように計画が縮小された結果，実際の分析対象は364標本，延べ時間数にして30時間20分となった。

1日を24時間としてサンプリングしたため，NHKの標本の一部には放送時間外のもの（つまり，音声・画面ともゼロのもの）が含まれたりもするが，結果への影響という点で特に重要と思われるのは，各標本の長さが一律5分であり，かつ開始時刻が毎時0分から，5分から，10分からというように5の倍数に設定されたことであろう。後者については，番組は5の倍数を持つ時刻を始まりや終わりとするものがほとんどであるから，

これにより集計される語彙の種類や量に偏りが出ることが懸念される（少なくとも画面文字の度数に多少の影響が出たことは担当者も認めている）。前者については，5分という長さが適当であったか否かが問題である。語彙調査は談話分析ではないから，文脈の影響はできるだけ排除すべきであると考えるなら，5分はあるいは長すぎるかもしれない。しかし，文脈を抜きにした言語研究はありえないという立場に立つなら，5分間から得られる文脈は多くの場合不十分であり，もっと長くすべきだという議論もありうるであろう。この問題は調査の目的や，資料の二次利用ということにかかわり，さらには標本採取（ビデオ録画等）に必要な人的および物理的手段の手当ということとも関係する。ここは問題を指摘するにとどめる。

〈調査の単位〉

語彙を計量するための単位をどうするか。分かち書きの習慣が確立している英語などとは違い，日本語の場合，単位の決め方にはいろいろな可能性があり，しかも，それが結果の数字を大きくも小さくもするので，たいへん重要かつ厄介な問題である。現に国語研究所のこれまでの調査でも，α単位・β単位，長単位・短単位，W単位・M単位など，調査ごとに多少とも異なった単位が用いられ，厳密な意味での比較を難しいものにしてきた。

同じことが今回の調査についても言える。これまでの調査と同様に，文節相当の比較的長い単位と形態素相当の比較的短い単位の両方を設定したとのことであるが，報告されているのは長い単位のみである。どのように切っていくかの具体的手順は省き，結果だけ述べると，今回の単位は，国語研究所でこれまでに案出された中でいちばん長いものになっている。固有名詞を人名・地名以外にも多数認定し，それらを全体で1単位としたこと，「～ている」「～てしまう」「～について」「～という（といった）」など助詞と動詞が複合して一つの補助動詞あるいは助詞のようになったものは分割せず，全体を一つの付属語と見なしたことなどが原因である。語彙表は自立語しか載せていないので，これら付属語扱いのものがどれくらい出現したかは分からないが，固有名詞だけ見ても，「桂三枝のニュース THIS EVENING」「小沢昭一のみちのくド真ん中」「小倉百人一首かるた名人位・クイーン位決定戦」など番組題名を中心に長いものが目立っていることは事実である。

これは普通の自立語つまり実質的な意味を担う語を重視したためという。番組題名などを一般の語と同じに扱うことには問題があろうから，この扱いでよかったとは思うが，得られた結果の数字はそのつもりで見る必要がある。人名を姓と名で分けるか分けないか，住居表示を都道府県名，市町村名，番地などまで全部含めて一つと扱うかいくつかに切って扱うか，「平成○年△月×日□曜日午前◇時」などをひと括りにするかしないかで，語彙量にどの程度の違いが生じるものなのか。マスメディアの場合，この種のものが出現する度合いは特に大きいと思われる。これらの点については，今後の検証に期待

したい。

〈調査結果の分析〉

まず，語彙の分析に先だって，標本（つまり番組）の分析を行っていることが注目される。標本の性質がよく分かっていないことに加えて，それが語彙の分布を左右する可能性が他の語彙調査の場合より大きいと予想されたこと，標本に見られる多様性が何によるものかを見ようとしたことなどが理由である。そして，そのための統計的手法として，探索的データ解析が用いられている。これはデータの分布状況を概観し，本格的分析への手がかりを得ることを目的として行う予備的分析であり，「幹葉表示」「五数要約」「箱ヒゲ図」といった視覚的手段を使用する。主にこの探索的データ解析の手法を用いて分析が行われているが，紙数の関係で，以下主な結果のいくつかに簡単に触れるのみである。

▽語の密度（1分あたりの語彙量）は，標本全体で，音声・本編は延べ語数73.5語，異なり語数41.2語，画面・本編は延べ8.2語，異なり6.4語，音声・CMは延べ46.6語，異なり36.3語，画面・CMは延べ42.0語，異なり31.2語（数字はいずれも中央値）である。

▽語の密度が大きい番組は，音声ではバラエティー番組，画面では情報番組である。

▽語の密度が小さい番組は，音声では音楽番組，画面ではドラマ番組である。

▽CMでは音声と画面の語彙量にほとんど差がない。

▽本編では音声の語彙量が画面のそれを大幅に上回る。

▽語種構成で見ると，音声では和語の割合が非常に高い。画面では語種の割合に差がない。

▽番組ジャンル別で見ると，スポーツ番組に画面文字が多い。

▽番組ジャンル間には，語の密度，語種構成，品詞構成にはっきりした差がある。

▽チャンネル別で見ると，NHK教育は，音声・画面の両方で語の密度が小さい。

▽四分位範囲（分布のちらばり幅）では，NHK総合の画面も語の密度が小さい。

▽話者の性別と語種および品詞分布の関係を見ると，語種では延べ・異なりとも女性は和語，男性は漢語が多い。品詞では延べ・異なりとも女性は用の類（動詞）と相の類（形容詞等）が多く，男性は体の類（名詞）が多い。

▽話者の職業別と語種および品詞分布の関係を見ると，語種ではアナウンサー・キャスター・司会者・政治家・大学教授は漢語が多く，俳優・声優・タレント・コメディアン・漫才師は和語が多い。品詞ではアナウンサーとキャスターに体の類が多い。また，アナウンサー・司会者・政治家・大学教授には相の類とその他の類が少ない。全体として，体の類と相・その他の類の分布は対照的である。

▽個人の話者と語種および品詞分布の関係を見ると，職業による類似性などよりもむしろ個人と出演する番組の性格との関連が強いようである。

▽国民1人あたりの平日1日の視聴語彙量(CMを除く)は，音声13950語，画面2640語（いずれも延べ語数）で，これはテレビ全体の送出量のそれぞれ13%，15%にあたる。

4　おわりに

テレビ放送の語彙調査と銘打った調査は，かつてNHK放送文化研究所でも行われたことがある（筆者もこれに携わった）が，これは放送用語の改善のためという実用上の目的に出るもので厳密に統計的なものではなかった。本格的な語彙調査としては本書のものが最初であり，それだけに試行錯誤的な部分が多かったのは当然である。それにしても番組を収録し，音声部分を文字化してコンピューター入力用のデータを作る作業は，エラーチェックなども含めて大変な時間と労力を要するものであり，報告書にまとめるまでの担当者諸氏の努力を多としたい。ただ残念なのは，何度も言うが分析データが予定の4分の1にとどまったことである。また，探索的分析が中心となって，量的にはともかく，質の面の分析に不満を残した。せっかく音声データがあるのだから，イントネーションやアクセントに多少触れるところがあってもよかったように思う。全体として第1段階の報告という感じは否めない。本来なら第2，第3段階の報告を期待したいところであるが，上述のように，現在は放送を取り巻く環境が急速に変化しつつある時代である。いずれは新しい環境に即した調査が行われなくてはなるまい。今回の調査で得られた経験や知見は，その際大いに役立つことであろう。

最後に，この調査のリーダーであり，「報告Ⅲ」に総論的な論文を載せている中野洋氏が1999年12月21日急逝された。ここに謹んで哀悼の意を捧げる。

(国立国語研究所報告112『テレビ放送の語彙調査Ⅰ――方法・標本一覧・分析――』1995年12月20日発行　秀英出版刊　B5判　412ページ　本体価格7767円／同114『テレビ放送の語彙調査Ⅱ――語彙表――』1997年5月20日発行　大日本図書刊　B5判　892ページ　本体価格9500円／同115『テレビ放送の語彙調査Ⅲ――計量的分析――』1999年4月26日発行　大日本図書刊　B5判　222ページ　本体価格5000円)

――城西国際大学教授――

(2000年10月11日　受理)