

国立国語研究所学術情報リポジトリ

調査の概要

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 石井, 正彦 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002913

調査の概要

言語体系研究部第二研究室
石井 正彦

1. 調査の特徴

国語研究所の語彙調査

- ① 新聞 (1949年, 1紙1ヵ月分)
- ② 婦人雑誌 (1950年, 2誌1年分)
- ③ 総合雑誌 (1953-54年, 13誌1年分)
- ④ 現代雑誌 (1956年, 90誌1年分)
- ⑤ 新聞 (1966年, 3紙1年分)
- ⑥ 高校教科書 (1974年, 理科・社会科9科目)
- ⑦ 中学校教科書 (1981年, 理科・社会科7科目)
- ⑧ 雑誌 (1906-76年, 1誌8年分)
- ⑨ テレビ放送 (1989年, 7チャンネル3ヵ月分)

- 1) 音声の語彙と画面文字の語彙
- 2) 言語の運用的な側面と語彙

2. 調査の目的

- 1) テレビ放送における語彙の実態
- 2) 現代日本語の語彙の実態
- 3) テレビ放送の語彙調査の方法論の確立

今回の報告書の目的

- ① いかにか調査したか
- ② どのような標本が得られたのか
- ③ 標本からテレビ放送の語彙について何を見出し得たのか

3. 調査の対象

- 1) 1989年4月2日(日)から同年7月1日(土)までの3ヵ月間に,
- 2) 全国放送網のキー局である6放送局の7つのチャンネルが,
- 3) 0時から24時までの一日をとおして放送した,
- 4) すべてのテレビ放送で,
- 5) 視聴可能であった,
- 6) 日本語の,
- 7) 音声および文字によって,
- 8) 表された語彙

NHK総合	1
NHK教育	3
日本テレビ	4
TBS	6
フジテレビ	8
テレビ朝日	10
テレビ東京	12

4. 調査の手順

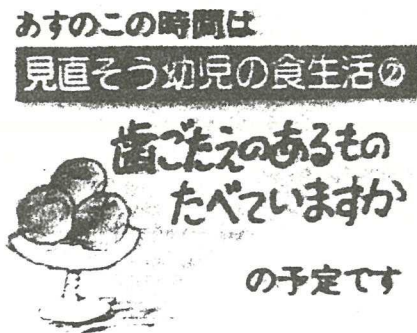
1) 標本抽出 (録画)

週・曜日・時間帯・チャンネルごとに標本数が等しくなるよう構成した集団からの無作為抽出 → 1標本5分, 1日4標本, 計364標本 (抽出比: 1/504)

2) 音声の文字化

3) 画面文字の採集

①テロップ, スーパーインポーズの類



S0294-080

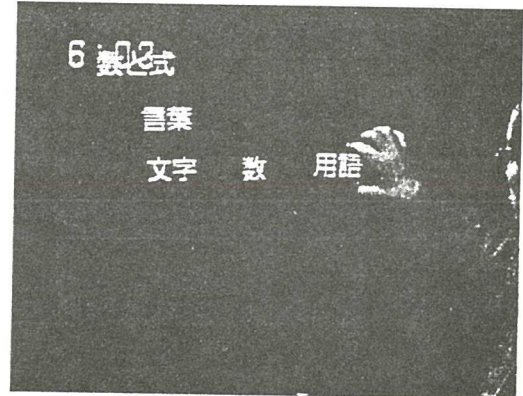


S0008-120

②フリップ, ボードの類



S0008-260



③セットの類



S0008-210

④実物の類



S0008-005

4) 単位切り (単位語の認定)

長い単位 (従来よりも、助詞・助動詞および固有名詞の範囲を拡張)

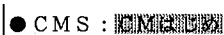
5) 同語異語判別 (見出し語の認定)

6) 情報付加 (語種・品詞, CM, 歌, 番組名, 視聴率, 話者など)

音 声

標本#0008 : 標本番組
「ニュースステーション」 : 番組名
20.7/48.2 : 視聴率

話者記号

●CMS : 


a (歌) : ミテ<みる, みる> ミテ<みる, みる> スベスベ<すべすべ, すべすべ> キレーナ<きれいな, きれいな> テ<て, て> ファミリーフレッシュ<ふあかりいふれっしゅ, ファミリーフレッシュ>

a : オーイ<おい, おい>> デキタワヨ<できる, できる> ワレナガラ<われながら, ながら> キレーナ<きれいな, きれいな> サラダガ<さらだ, サラダ> デキマシタ<できる, できる> ワタシノ<わたし, 私> テダツテ<て, て> ホラ<ほら, ほら>> キレーナモンダ<きれいな, きれいな> ワタシワ<わたし, 私> ファミリーフレッシュ<ふあかりいふれっしゅ, ファミリーフレッシュ>

b : ライムノ<らいむ, ライム> カオリガ<かおり, 香り> トツテモ<とても, とても> フレッシュ<ふれっしゅ, フレッシュ> アブラヨゴレガ<あぶらよごれ, 油汚れ> サツト<さつと, さつと> オチテ<おちる, 落ちる> スバヤイ<すばやい, すばやい> アワギレ<あわぎれ, 荒ざれ>

a : マイニチノ<まいにち, 毎日> コトダカラ<こと, こと>

a (歌) : テニ<て, て> ヤサシ<やさしい, やさしい> ファミリーフレッシュ<ふあかりいふれっしゅ, ファミリーフレッシュ>

●CME : 




e : エ<え・ええ, え・ええ>> ネンノタメ<ねんのため, 念のため> モーシアゲテオキマスト<もうしあげる, 申し上げる> エイドトユーノワ<えいと, エイド> キュージョスルトカ<きゅうじよする, 数値する> キューサイスルトカユー<きゅうさいする, 数値する> イミデス<いみ, 意味> マイトシコーレイ<まいとしこうれい, 毎年恒例> ジャイアンツエイド<じやいあんつえいと, ジャイアンツエイド>

f : カレーライスオデスネ<かれえらいす, カレーライス>

e : ハイ<はい, はい>

f : +エ<え・ええ, え・ええ>> ジャイアンツガ<じやいあんつ, ジャイアンツ> ユーショースルマテ<ゆうしよする, 優勝する> タツト<たつ, 断つ>

画面文字

   出現形 見出し語

CS0008-005 4, ファミリーフレッシュ<ふあかりいふれっしゅ, ファミリーフレッシュ>
CS0008-010 4, ファミリーフレッシュ<ふあかりいふれっしゅ, ファミリーフレッシュ>
CS0008-010 1, 天然香料成分<てんぜんかうりようせいぶん, 天然香料成分> 配合<はいごう, 配合>
CS0008-020 1, ライムの<らいむ, ライム> 香りが<かおり, 香り> フレッシュ<ふれっしゅ, フレッシュ>
CS0008-030 1, すばやい<すばやい, すばやい> 泡ギレ<あわぎれ, 荒ざれ>
CS0008-040 1, 手に<て, て> やさしい<やさしい, やさしい> ファミリーフレッシュ<ふあかりいふれっしゅ, ファミリーフレッシュ> 花王<かおう, 花王>

S0008-090 1, 黒鉄ヒロシ<くろがねひろし, 黒鉄ヒロシ>
S0008-100 1, 川崎徹<かわさきとほる, 川崎徹>
S0008-110 1, 糸井重里<いとしいげさと, 糸井重里>
S0008-120 1, 1986年秋<せんきゅうかきはちじゅうろくねんあき, 1986年秋> ジャイアンツエイドは<じやいあんつえいと, ジャイアンツエイド> 始まった<はじまる, 始まる>

7) 集計

5. 標本一覧

標本 0008		番組名 「ニュースステーション」							
語数	延べ	異なり	語の密度	延べ	異なり	CM語数	延べ	異なり	
音声	288	167	音声	71.7	41.6	音声	57	45	
画面	138	60	画面	34.4	14.9	画面	23	19	
語種	和	漢	外	混	不明				
音声	208(72.2)	42(14.6)	21(7.3)	17(5.9)	0(0.0)	放送年月日	1989年4月3日(月)		
	105(62.9)	33(19.8)	13(7.8)	16(9.6)	0(0.0)	採録時刻	23:00~23:05		
画面	38(27.5)	30(21.7)	12(8.7)	21(15.2)	37(26.8)	チャンネル	10(テレビ朝日)		
	23(38.3)	18(30.0)	4(6.7)	10(16.7)	5(8.3)	ジャンル	報道系		
品詞	体	用	相	他	不明	本編秒数	241		
音声	123(42.7)	47(16.3)	56(19.4)	62(21.5)	0(0.0)	番組分数	75		
	83(49.7)	37(22.2)	31(18.6)	16(9.6)	0(0.0)	視聴率1	20.7		
画面	90(65.2)	8(5.8)	2(1.4)	1(0.7)	37(26.8)	視聴率2	48.2		
	44(73.3)	8(13.3)	2(3.3)	1(1.7)	5(8.3)	本編話者数	5		
画面媒体(延/異)	1: 31/27 2: 104/31 3: 3/3 4: 0/0					(男: 5 女: 0 不明: 0)			

採録時刻……………標本を採録した時刻 (24時制)

ジャンル……………番組のジャンル

- | | |
|-----------|------------|
| 1: 報道系 | 5: バラエティー系 |
| 2: 教育・教養系 | 6: ストーリー系 |
| 3: 一般実用系 | 7: スポーツ系 |
| 4: 音楽系 | 8: その他 |

本編秒数…………… 1 標本 5 分 (300 秒) のうち、本編部分の秒数

番組分数…………… 当該番組全体の分数

視聴率 1 …………… 局別世帯視聴率 } 標本が放映された 5 分間における視聴率

視聴率 2 …………… 総世帯視聴率

本編話者数…………… 本編部分に登場する話者の数

語数…………… 音声・画面それぞれの延べ語数・異なり語数

語の密度…………… 音声・画面それぞれの、1 分あたりの延べ・異なりの語数

CM 語数…………… 標本の中の、CM 部分の延べ語数・異なり語数 (音声・画面)

語種・品詞…………… 音声・画面それぞれ上段が延べ語数、下段が異なり語数

() 内は百分率。

語種: 和語 (和), 漢 (漢語), 外 (外来語), 混 (混種語)

品詞: 体の類, 用の類, 相の類, その他

画面媒体…………… 画面に現われた文字情報の、各媒体ごとの延べ語数・異なり語数

- 1: テロップ, スーパーインポーズの類
- 2: フリップ, ボードの類
- 3: セットの類
- 4: 実物の類

6. 調査の規模・主な数値

- | | | | | | |
|----------------------------|-------------|------|------------|------------|---------|
| 1) 母集団 | 時間数：2184時間 | 延べ語数 | 音声：約5700万語 | 画面：約1500万語 | |
| 2) 抽出比 | 504分の1 | | | | |
| 3) 標本 | 時間数：30時間20分 | 本編 | 96,832秒 | CM | 12,368秒 |
| 4) 延べ語数 | 全体：141975 | 本編音声 | 103081 | CM音声 | 9235 |
| | | 本編画面 | 20246 | CM画面 | 9413 |
| 5) 異なり語数 | 全体：26033 | 本編音声 | 17647 | CM音声 | 3455 |
| | | 本編画面 | 7970 | CM画面 | 3591 |
| 6) 語の密度（1分あたりの延べ語数，標本中央値） | | 本編音声 | 73.5 | CM音声 | 44.6 |
| | | 本編画面 | 8.2 | CM画面 | 42.0 |
| 7) 語の密度（1分あたりの異なり語数，標本中央値） | | 本編音声 | 41.2 | CM音声 | 36.3 |
| | | 本編画面 | 6.4 | CM画面 | 31.2 |

7. 標本の分析

- 1) 単位
364の標本（1標本5分）
- 2) 目的
 - ①テレビ放送の語彙の言語内的な側面の特徴
 - ②テレビ放送の語彙における言語内的な側面と言語外的な側面との関係

語の密度	本編とCM
語種構成	音声と画面
品詞構成	番組のジャンル
	チャンネル
	放送時間帯
	放送曜日
	番組の長さ
	視聴率
	話者
	視聴行動
	画面文字の媒体

3) 方法

探索的データ解析

限られたデータ（数が少ない，分布型が仮定できない）で，
その構造・特徴をさぐる



データ（標本）の分布（形）を概観する

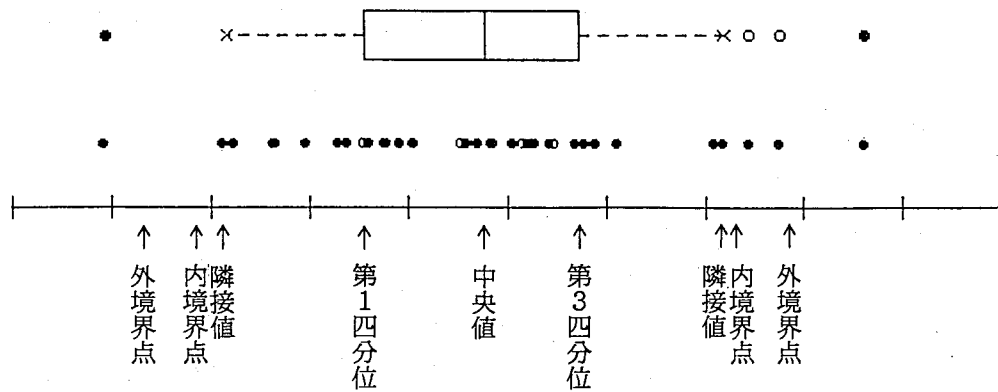
①ヒストグラム・幹葉表示による概観

②五数要約値と箱型図による概観

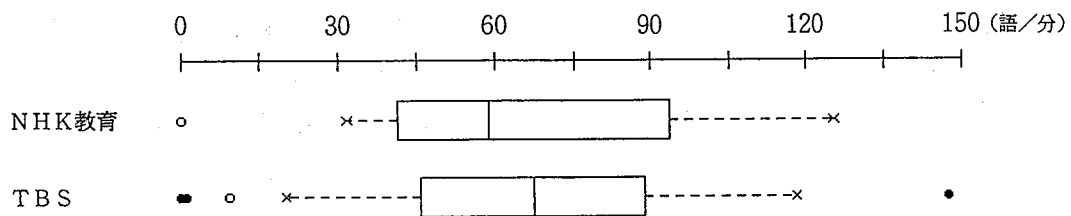
[表 1] 語の密度の五数要約値

	平均値	最小値	第1四分位	中央値	第3四分位	最大値	標本数
NHK教育	63.4	0.0	41.6	58.8	93.6	125.2	43
TBS	63.9	0.0	45.9	67.2	88.8	147.4	50

[図 1] 五数要約値と箱型図との関係



[図 2] 語の密度の平行箱型図



③外れ値の検討

N H K教育の外側値（括弧内は標本番号）

- (0100)「英語会話」
- (0112)「N響アワー」
- (0135)「レニングラード・バレエ・白鳥の湖」
- (0172)「芸術劇場・ミュージック・フランセーズ・フランス室内楽の精華」
- (0207)「ぼくの絵わたしの絵」
- (0340)「芸術劇場・サイモンフレストンオルガン演奏会」

T B Sの外側値・極外値

- (0037)「ノーカット世界名作劇場ポーランド映画傑作選・灰とダイヤモンド」
- (0125)「MTVジャパン・ヘッドバンガーズ」
- (0261)「ノーカット世界名作劇場・カサノバ」
- (0129)「東京音楽祭情報」
- (0169)「平成名物TV・トンガリ編」

8. 今後の予定

1) 語彙表の作成

見出し語データ（一部）

語種	品詞	読み	代表形	全体	音声	画面	本編音声	本編画面	CM音声	CM画面	
1	W	4	あ・ああ	あ・ああ<感>	1455	1451	4	1404	4	47	0
2	W	3	ああ	ああ<指>	3	3		3		0	0
3	H	1	ああ	青山円形劇場	1	1		1		0	0
4	W	3	ああいう	ああいう<達>	35	35		35		0	0
5	G	1	ああす	アース	4		4			0	4
6	H	1	ああすせいやく	アース製菓	2		2			0	2
7	G	1	ああすのおまっと	アースノーマット	4	1	3			1	3
8	G	1	ああすのおまっとりきつど	アースノーマット・リキッド	1		1			0	1
9	G	1	ああすれっどだふる	アースレッドW	3	1	2			1	2
10	G	1	ああち	アーチ	1	1		1		0	0
11	G	1	ああちすと	アーチスト	2	2		2		0	0
12	G	1	ああとねいちゃあ	アートネイチャー	25	10	15		2	10	13
13	G	1	ああとへらぼんど	アート・ヘラ・ボンド	1		1		1	0	0
14	G	1	ああとりぞおと	アート・リゾート	3		3			0	3
15	G	1	ああのるとしゅわるとつねがあ	アーノルドシュワルツネッカー	1	1		1		0	0
16	G	1	ああばん	アーバン	1	1		1		0	0
17	G	1	ああばんじおぐりつど	URBAN_GEO_GRID	1		1		1	0	0
18	H	1	ああばんほうゆん	アーバン方面	1	1		1		0	0
19	G	1	ああばんりぞおと	アーバンリゾート	1		1			0	1
20	G	1	ああまおおる	アーマオール	3	2	1			2	1
21	G	1	ああもんど	アーモンド	3	2	1			2	1
22	G	1	あある	R	1	1				1	0
23	G	1	あある	R<リコ->	2		2			0	2
24	G	1	あある	R<リョービ>	1		1		1	0	0
25	G	1	あある	R<リンア->	1		1			0	1

2) 語彙表の分析