

Title	マルチメディア・コーパスの構築と活用 : 表現行動の計量的研究のために
Author(s)	孫, 栄爽
Citation	阪大日本語研究. 2010, 22, p. 65-90
Version Type	VoR
URL	https://hdl.handle.net/11094/10048
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

マルチメディア・コーパスの構築と活用 —表現行動の計量的研究のために—

Building and using of multimedia corpus:
For a quantitative study of expressive behavior

孫 栄爽
SON Youngsuk

キーワード：コーパス言語学、マルチメディア・コーパス、テレビ放送、映像・音声、表現行動、
非言語行動

要旨

言語と非言語行動（視線、表情、身振りなど）とを用いて情報を伝える行為を表現行動と呼び、表現行動に関するパターン化された傾向を見出すことを目的に、（主として）テレビの対談番組を資料に、言語のみならずその使用場面の映像・音声をも同時に参照できる「対談番組のマルチメディア・コーパス」を作成した。それによって、ある言語表現が、どの番組で、誰が誰に向かって話すときに、どのような視線・身振り・動作・表情とともに発せられたものか、またはそうした使用状況でどのような言語表現が発せられたかなどが、現実の大量のデータにもとづいて調べられるようになった。そして、実際にこのコーパスを使って、擬音語・擬態語と身振りとの関係や、終助詞「ネ」と視線との関係などを調べた結果、描写力が高いといわれる擬音語・擬態語の発話時ほど身振りが頻繁に伴うということや、「ネ」の用法によって視線行動に違いがみられるということなどが傾向として見出せた。これらは、従来の内省や実験観察的調査によって指摘されてきた内容に修正・補充すべきところがあることを指摘するものであり、同時に、表現行動研究におけるマルチメディア・コーパスの有用性を示すものである。ただし、表現行動の実態をより正確に知るためには、今後、本コーパスの改善のほか、他の資料によるマルチメディア・コーパスの作成が欠かせない。

1. 表現行動研究のためのコーパス

人は、言語とともに、さまざまな手段を用いて、情報の伝達を行う。対面的な言語使用場面の場合、そのさまざまな手段の中でとくに注目されるのが視線・表情・身振り・声の調子などの非言語行動であり、言語は非言語行動と関連し影響しあいながら共に現れるとあってよい。本研究では、このような行為全般を、林（1973）にならって「表現行動」と呼ぶ¹⁾。表現行動のさまざまな手段の中で、声の調子や大きさなどいわゆるパラ言語はそ

の位置づけに迷うところも多いが、本研究ではこれらも非言語行動に含めて考えることにする。

話し手の発した言語は、それだけが独立して存在するのではなく、非言語行動と共に表現行動を成し、その一部として現れる。日本では、林（1973）と南（1974）がこのことをいち早く指摘し、理論的モデルを提示した。その後、国語研（1984）では、人々の意識の側面から調査が行われている。

内省や意識調査とともに、実態調査も行われたが、国語研（1987）で代表される一連の研究は、その先駆的なものであったといえる。たとえば、言語と身振り（非言語行動）の関係についての分析項目では、発話量と身振りの出現量との関係や、発話位置による身振りの現れなどが精密に記述されており、二十年以上たった今からみてもその先見性は失われていない。これらの分析に国語研（1987）が用いた調査方法は、座談場面をあらかじめ設定し、座談メンバーの構成を実験的にコントロールした上で、その中で生じるすべての言動を観察・記録するという、実験観察的な調査であった。この調査方法は、調査の前段階で予想される関連要素を制御できるなど、目的に即したデータ収集に適している。

しかし一方で、実験観察的な調査は、実験的な側面が強いので調査結果が必ずしも現実の表現行動の実態を反映するものではないということ、データ収集の段階でかなりの時間と労力を要するので多数の被験者を対象とした大量のデータ収集が難しいということの、問題点も同時に持つ。そして、このことから、見出した調査結果が、表現行動に関するパターン化された傾向であるかもしくは（個人差などによる）偶発的なものであるかの判断が困難な場合がある。

表現行動に関する研究は、上記のほかにも、言語哲学、文化人類学、動作学、心理学などさまざまな分野で、さまざまな観点からなされてきている。ただし、その調査方法は上記と同様、内省や意識調査、実験観察的調査のいずれかであり、表現行動の実態が、現実の大量のデータにもとづき量的に論じられることは、これまでほとんどなかった。

こうした状況のなかで、表現行動に関わるパターン化された傾向を経験的に見出すには、内省や意識調査、実験観察的調査のみでは不十分で、現実の大量のデータにもとづいて、計量的な実態調査を行わなければならない。この、実際に使われたデータを大量に集め処理するという発想は、本来、コーパス言語学の考え方である。つまり、表現行動の研究においても、表現行動の全般が観察可能なコーパスがあれば、その実態を追及するための有用な手段になりうると考える。

そこで、本研究では、表現行動研究に利用できるコーパスを作り、それを使って表現行動に関する計量的な実態調査を行うことにした。表現行動の研究にコーパスを用いること

で、他の調査方法ではできなかった、あるいは難しかった表現行動に関するパターン化された傾向を見出すことを目的とする。そして、それは、表現行動の実態を知るために欠かせないものであると考える。

本稿では、表現行動研究に利用できるマルチメディア・コーパスの構築を中心に、実際にそれを使った調査についても論じる。コーパスの構築やそれに伴う方法論的な問題は、従来、コーパス言語学の中でも特に“corpus construction (corpus compilation)”で数多く議論されてきており、本稿もその枠組みの中で論じることにする (Meyer2002、Teubert & Krishnamurthy (eds.) 2007)。

2. マルチメディア・コーパス

コーパスとは一般的に「言語研究に使用しうる電子化された言語テキストの集合体」と定義される。この定義からもうかがえるように、従来のコーパスは主として、新聞や小説などの書きことばと、話しことばを書き起こした音声文字化テキストとを対象としてきた。実際の音声をも検索できるコーパスが登場したのは、比較的最近である。これらのコーパスは、主に、書きことばや話しことばの文字・語彙・文法・音声の研究のために作られてきたもので、本研究で試みるような、言語がどのような非言語行動とともに発せられたかに関する調査には利用できない。

そうした中で、石井 (2009) は、「マルチメディア・コーパス」という考え方を提案している。マルチメディア・コーパスとは、音声文字化テキストと映像・音声とを同期させ、音声文字化テキストから言語形式を検索すると同時に、その発話場面の映像・音声をも参照できるコーパスのことである。このコーパスを使えば、たとえば、ある単語を、誰が、誰に向かって、どのような表情で、どのような声の調子で発話したのか、あるいはそうした使用状況でどのような単語が発せられたのかなどを効率的に調べることができる。しかし、そのためには、あらかじめ、発話された言語、話し手、聞き手、表情、声の調子などに関する情報をコード化し、それらをコーパスに付与しなければならない。

理想的なマルチメディア・コーパスには、その映像・音声から発話場面の全般はもちろん発話参加者一人一人の細かい非言語行動も継続的に観察できること、全発話が精密に書き起こされた音声文字化テキストを成していること、その音声文字化テキストと映像・音声とが細かい単位 (たとえば、単語単位) で同期されていること、そして表現行動に関わるあらゆる情報が詳細かつ的確に搭載されていることが望まれる。その他、それらの情報が互いに自由に検索できるツール、KWIC (KeyWord In Context)、音声波形などが自動

的に作成・表示されるプログラムなどが備わっていればよいと考える。

3. テレビ放送の対談番組

マルチメディア・コーパスの主な特徴は、発話のみならず、その使用場面の映像・音声（とくに、映像）をも同時に参照できるところにある。したがって、発話とともに発話場面が観察できるならば、資料となりうるものに制限はない。だが、現実の大量のデータにもとづく計量的な調査を行うには、その表現行動が広範に、また容易に観察できる資料にもとづくほうが、目的に適すると考える。

そういった面で、テレビ放送（以下、テレビ）は、現時点でもっとも有用な資料となりうる。テレビは、我々の身近に存在するメディアで、そこからは、出演者たちが行った（表現行動研究のために実験的に制御されたものではない）実際の表現行動に関するデータが比較的容易に得られる。そして、その出演者たちによる表現行動のうち、言語によるところは主として音声から、非言語行動によるところは映像・音声から、その情報が得られる。このことから、テレビの音声を書き起こして音声文字化テキストを作成し、それとその時点での映像・音声とを同期させれば、両者が相互に検索可能になる。しかも、テレビでは、さまざまな発話参加者がさまざまな状況でさまざまな話題について発話しているので、それらを考慮に入れた形での表現行動の研究も可能である。

こういった利点から、筆者は、コーパスの作成にテレビを利用することとした。用いた資料には、「テレビ放送の語彙調査（NHK総合・教育）」「ABC放送」「対談番組（限定）」がある。

「テレビ放送の語彙調査（NHK総合・教育）」は、国語研究所の「テレビ放送の語彙調査」のうちのNHK総合・教育のデータで、1989年4～6月の三ヶ月間にNHK総合・教育で放送された番組の中から無作為に抽出した約8時間40分間の音声文字化テキストデータ（全発話対象）と映像・音声データを収めている（詳しくは、国立国語研究所1995・97・99、石井2009参照）。

「ABC放送」は、筆者が、2006年10月30日～11月5日の一週間、大阪ABC放送で放送された全番組を録画したものである。「ABC放送」には、全発話ではなく、後述の活用例の分析項目（擬音語・擬態語と終助詞ネ）を含む前後の文脈のみが文字化済みである。

「テレビ放送の語彙調査（NHK総合・教育）」と「ABC放送」はいずれも、ある期間中に放送された全番組を対象としているので、中には、スポーツ中継番組など、出演者たち（この場合は、アナウンサーと解説者）の非言語行動がほとんど観察できない番組も多数

含まれている。両資料から、これらの番組は省き、話し手の非言語行動が比較的多く観察できる番組にかぎってサブコーパスを作成したところ、結果的に、対談番組やワイドショー番組などを主に扱うことになった。これは、特段の理由によって番組を限定したわけではない。

そこで、あらかじめ番組を対談番組にかぎって、新たにデータ収録を行うことにした。以下、このデータを「対談番組（限定）」と呼ぶ。対談番組にかぎったのは、映像に話し手が映っていることがもっとも多く、しかも継続的に観察できる場合が多かったからである。他にも、対話中に背景音などがほとんど挿入されないので、音声聞き取りやすいといった利点などもある。

本研究で作成したコーパスの大半は、この「対談番組（限定）」からなり、また「テレビ放送の語彙調査（NHK総合・教育）」と「ABC放送」も結果的には対談番組を多数含んでいることから、これら三資料を内容とする本研究のコーパスを、「対談番組のマルチメディア・コーパス」と呼ぶことにする。

4. 「対談番組のマルチメディア・コーパス」の構築

ここでは、「対談番組のマルチメディア・コーパス」の作成過程について、「対談番組（限定）」を中心に説明する。「テレビ放送の語彙調査（NHK総合・教育）」「ABC放送」に関しても、ほぼ同じ手順と方法を用いている²⁾。

4.1. 番組の選定

「対談番組（限定）」の番組選定は、以下の手順で行った。まず、大阪の6放送局7チャンネルにおける対談番組の数・種類・放送時間を知るために、各チャンネルで一週間に放送された全番組を録画し、対談番組の実態に関する事前調査を行った。次に、そこから、4人以下の発話参加者による23種類の対談番組を選び、各番組を対象に、2009年3月31日（月）から2009年8月9日（日）の間、録画作業を行った。最終的には、話題や発話参加者の属性などが特定のものに偏らないように注意しながら、106回分の放送を選定した（表1）。こうして集めた対談番組の放送時間数は、62時間42分である。そして、この中から、話し手の非言語行動が観察できる箇所などにかぎって音声を書き起した。その時間数は、39時間41分である。

表1：「対談番組（限定）」の各番組情報（文字化時間数の多い順）

番組名 (放送時間（一回分）、チャンネル)	発話参加者数	放送回数	放送時間	文字化時間数
こころの時代～宗教・人生（60分、NHK教育）	2人	8回	8時間00分	6時間10分
時事放談（40分、毎日放送）	4人	9回	6時間00分	4時間40分
生活ほっとモーニング（50分、NHK総合）	2～3人	6回	5時間00分	3時間58分
徹子の部屋（30分、ABCテレビ）	2人	7回	3時間30分	2時間58分
美女放談（30分、テレビ大阪）	2～3人	9回	4時間30分	2時間49分
ボクらの時代（30分、関西テレビ）	3人	5回	2時間30分	2時間05分
KYOTO塾（30分、関西テレビ）	2人	7回	3時間30分	1時間53分
ミュージズの晩餐（25分、テレビ大阪）	3人	9回	3時間45分	1時間50分
トップランナー（40分、NHK総合）	3人	3回	2時間00分	1時間42分
となりの子育て（30分、NHK教育）	4人	4回	2時間00分	1時間38分
グータンヌーボ（30分、関西テレビ）	3人	5回	2時間30分	1時間26分
知る楽（30分、NHK教育）	2人	5回	2時間30分	1時間01分
爆笑問題のニッポンの教養（30分、NHK総合）	3人	3回	1時間30分	58分
ホリデーインタビュー（23分、NHK総合）	2人	4回	1時間32分	56分
僕らの音楽（30分、関西テレビ）	2人	5回	2時間30分	52分
さんまのまんま（29分、関西テレビ）	2人	2回	0時間58分	48分
憲法記念日特集～今日は憲法記念日です（60分、NHK総合）	3人	1回	1時間00分	45分
日曜美術館（60分、NHK教育）	3人	2回	2時間00分	40分
アナザースカイ（30分、よみうりテレビ）	4人	3回	1時間30分	38分
ボランティア21（30分、テレビ大阪）	4人	3回	1時間30分	34分
プロフェッショナル・仕事の流儀（40分、NHK総合）	3人	3回	2時間00分	31分
カンブリア宮殿（50分、テレビ大阪）	3人	2回	1時間40分	29分
おしゃれ工房（30分、NHK教育）	3人	1回	0時間30分	20分
計		106回	62時間42分	39時間41分

これらの番組は、対談形式から、大きく、話し手役と聞き手役があらかじめ決まっている「インタビュー形式」（「徹子の部屋」や「トップランナー」など）と、対談の中で話し手役と聞き手役が自由に交替する「自由対話形式」（「ボクらの時代」や「グータンヌーボ」など）とに分けられる。そこで話された話題には、芸術・政治・経済・思想・宗教・教育・福祉・スポーツ・恋愛・旅行などがあり、比較的幅広い。

話し手の人数は、延べ310人（異なり人数192人）で、以下に、その性別・年代・職業・出身地についてまとめる³⁾。まず、性別による分布をみると（図1）、延べでは男性158人・女性152人、異なりでは男性101人・女性91人で、ほぼ均等に分布している。ただし、実際の発話時間を比べてみると、男性は24時間29分、女性は15時間11分で、男性の話し手によ

る発話時間数のほうが9時間20分ほど多い。

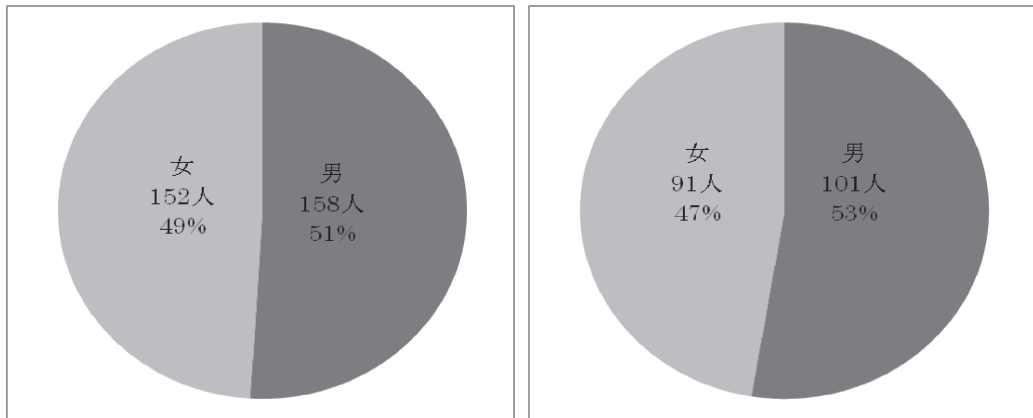


図1：話し手の性別の分布（左側：延べ人数、右側：異なり人数）

次に、年代の分布をみる（図2）。延べでも異なりでも、30代・40代・50代が、他の年代よりやや多いが、比較的均等に分布している。各年代の話し手による発話時間数を多い順に並べると、70代（7時間15分）、50代（7時間9分）、40代（6時間50分）、60代（5時間29分）、30代（4時間58分）（以下略）といった順番になる。その他、年代がわからない話し手による発話も1時間18分あった。

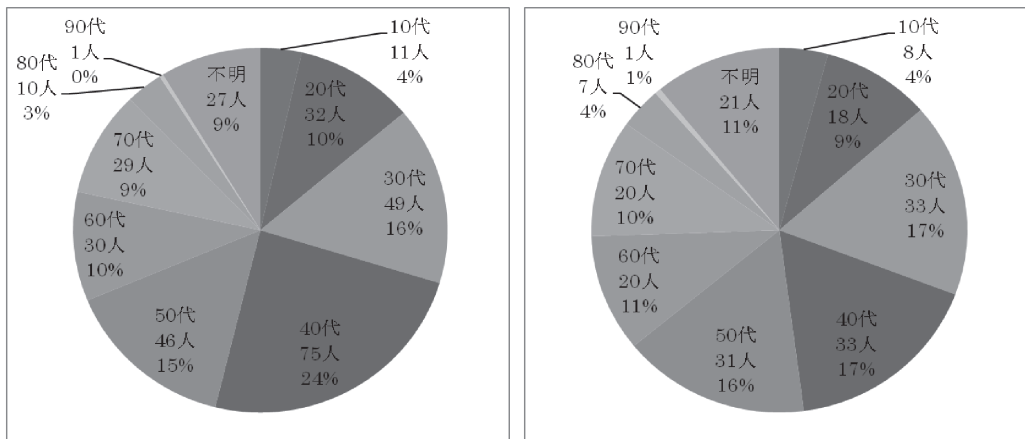


図2：話し手の年代の分布（左側：延べ人数、右側：異なり人数）

職業に関しても、特定のものに偏らないことを考慮したが、テレビメディア上の特徴もあって、俳優・アナウンサー・タレントなど、テレビに出演する機会が比較的多い職業の

話し手が多かった（図3）。「その他」は、異なり人数が10人以下の職業をまとめたもので、評論家・科学者・スポーツ選手・医者・僧侶・写真家・デザイナー・料理人など多種多様な職業がある。職業別に、その発話時間数をみると、俳優（6時間47分）、大学教授（4時間12分）、アナウンサー（3時間33分）、政治家（3時間29分）、作家（3時間27分）、タレント（3時間11分）、歌手（2時間22分）、実業家（2時間25分）の順であった。

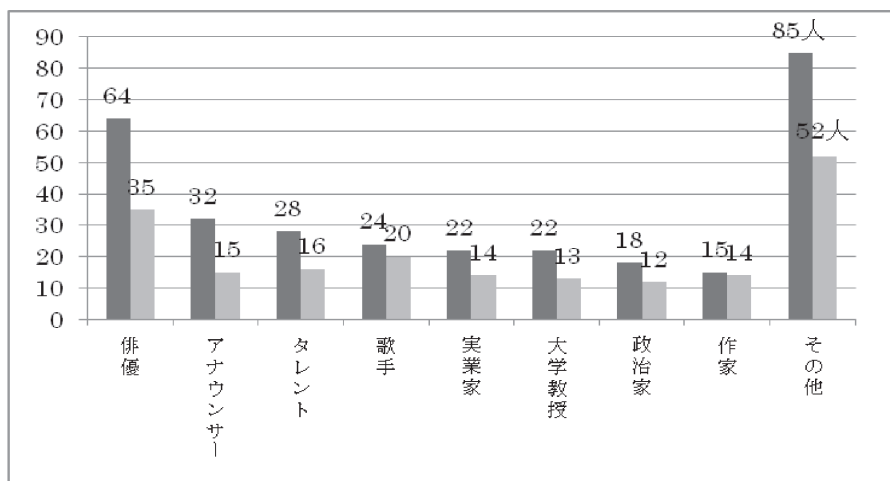


図3：話し手の職業の分布（左側：延べ人数、右側：異なり人数）

最後に、出身地を図4にまとめる。関東と近畿出身が、延べ・異なり人数で、全体の過半数を占めるが、そのうち約4割は関東出身であった。中でも東京都出身（11時間15分）の話し手による発話がもっとも多く、次いで京都府出身（3時間21分）、埼玉県出身（2時間47分）、大阪府出身（2時間29分）の話し手による発話が多かった。

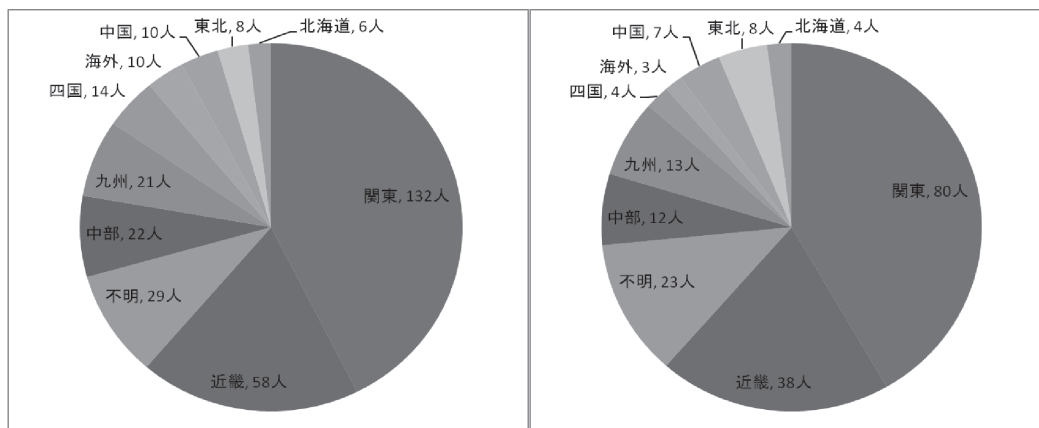


図4：話し手の出身地の分布（左側：延べ人数、右側：異なり人数）

4.2. 音声文字化テキストの作成

他番組に比べて、対談番組では、話し手が映像に多くかつ持続的に映ることは確かであるが、VTRなどが流れる場合などには、対談番組でも話し手の非言語行動が観察できない。これらの部分で発話された音声はそれを書き起こしたところで、話し手の表情や視線などとの関係についての調査には用いられない。したがって、あらかじめ映像の中から話し手の非言語行動が観察できる部分を調べて、その部分にかぎって書き起こし作業を行った。他にも、番組本編で話し手が同席者と話し合う場合以外（コマーシャル、明らかに視聴者に話しかける場合など）は省いた。

音声の精密な文字化作業には相当の時間と労力を要するので、作業を専門のテーブルライターに依頼し、それを筆者が点検するという方法をとった。入力には、マイクロソフト社製“Microsoft Office Excel”を使っている。以下、その概要をまとめる。

- ① 漢字仮名混じり文で一般的な表記を使用する。
- ② 発話どおりの書き起こしをする（フィラー、言い直し、いいよどもも含む）。
- ③ 文の認定基準には『話しことばの文型（1）』（国立国語研究所報告書18、1960）を参考にする。ただし、厳密な基準をもって文を認定したわけではない。
- ④ Excelの行ごとに一文（文相当）を入力し、文の終わりには文末句点「。」を施す（あきらかな疑問文には「？」をつける）。ただし、発話の途中で他者の発話が入った場合は途中までを一行とする。その場合「。」は付けない。
- ⑤ 一人の話者の発話を1行とする。話者が変わるたびに行を変え、話者記号（「A」「B」など）を付与する。
- ⑥ 文中はいわゆる文節相当で分かち書きする（ただし、補助動詞、形式名詞類は分かち書きしない）。
- ⑦ 笑い声も書き起こす。笑い声が発話と切り離せて独立している場合は一行とする。切り離せない場合には書き起こしの対象部分に「(笑)」というタグを前後に付与する。

図5は①から⑦の手順で作成した音声文字化テキストの一部である⁴⁾。

文番号	開始時間	終了時間	話者記号	発話文
286	0:20:48	0:20:50	A	もう こう ご家族も
287	0:20:50	0:20:50	B	はい。
288	0:20:51	0:20:52	A	娘さん 息子さん
289	0:20:52	0:20:52	B	うん。
290	0:20:53	0:20:53	A	いっしょって
291	0:20:53	0:20:56	A	その こう 成長していきますも この 秋田でね
292	0:20:53	0:20:54	B	はい。
293	0:20:56	0:20:57	B	うん。
294	0:20:56	0:20:59	A	柳葉さんと 同じように 成長していくわけですけど
295	0:20:59	0:20:59	B	はい。
296	0:21:00	0:21:02	A	子どもたちに こう 望むことなんて ありますか？
297	0:21:03	0:21:08	B	いや あのー (笑)まずは(笑) まあ 元気に いてくれば いいなっというのが いちばんですけど。
298	0:21:09	0:21:14	B	あのー せっかくね こういう いい 環境の 中 に いるんでね。
299	0:21:14	0:21:34	B	あのー いいものを見て えー いいものを 聞いて いいものを感じて えー 感受性の 豊かな 人間に なって で きっちり きちんと 人のことを こう んー 考えていけるようなね 人 に なってくれば 100% 十分ですね。
300	0:21:34	0:21:35	A	ううん。
301	0:21:35	0:21:35	B	はい。

図5：音声文字化テキストの一部（ホリデーインタビュー：NHK 総合_2009年5月5日放送）

4.3. 音声文字化テキストと映像・音声との同期

4.2の手順で作成した音声文字化テキストには、図6のように、文字化情報のほかにも、それが発話されたときの発話開始時点と終了時点（秒単位）が、文（および文相当）ごとに記録してある。

そして、その隣には“映像クリップ”という列が並んでおり、それをダブルクリックすると、発話開始時点から終了時点間の映像・音声再生されるようになっている。すなわち、文（相当）単位で、それに該当する発話場面の映像・音声瞬間に調べられるのである。

映像・音声の再生には、石井（2009）と同様、オープンソースのメディアプレーヤー“MPlayer”（フリーソフトウェア）を使用した。ただし、石井（2009）では、Excel上でMPlayerを機動させるには、（発話時点をもとに）あらかじめバッチファイルを用意し、それをExcelのハイパーリンク（Hyperlink）機能で開かなければならなかった。本コーパスでは、そういった事前の用意がなくても、発話時点情報さえ入力してあれば、MPlayerが再生できるよう、Excelのマクロ（ソフトの操作などの実行手順を記述したもの）を使って、MPlayerが自動的に発話時点情報を読み込み、それに該当する部分の映像・音声のみを探し出し再生するようにした。他にも、Excel上でMPlayerの、再生時間の調整（たとえば、発話開始時点2秒前から終了後3秒までの再生など）、再生速度の調整、映像サイ

番号	映像	文番	開始時間	終了時間	映像クリップ	話者番号	発話文
2	T19150422_美女放談.mpg	1	0:02:44	0:02:45	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
3	T19150422_美女放談.mpg	2	0:02:45	0:02:47	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
4	T19150422_美女放談.mpg	3	0:02:45	0:02:46	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
5	T19150422_美女放談.mpg	4	0:02:47	0:02:50	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
6	T19150422_美女放談.mpg	5	0:02:49	0:02:50	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
7	T19150422_美女放談.mpg	6	0:02:50	0:02:54	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
8	T19150422_美女放談.mpg	7	0:02:52	0:02:54	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
9	T19150422_美女放談.mpg	8	0:02:54	0:02:54	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
10	T19150422_美女放談.mpg	9	0:02:54	0:02:56	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
11	T19150422_美女放談.mpg	10	0:02:56	0:02:56	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
12	T19150422_美女放談.mpg	11	0:02:57	0:03:00	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
13	T19150422_美女放談.mpg	12	0:03:00	0:03:00	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
14	T19150422_美女放談.mpg	13	0:03:00	0:03:04	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
15	T19150422_美女放談.mpg	14	0:03:03	0:03:09	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
16	T19150422_美女放談.mpg	15	0:03:09	0:03:10	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
17	T19150422_美女放談.mpg	16	0:03:23	0:03:28	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
18	T19150422_美女放談.mpg	17	0:03:29	0:03:29	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
19	T19150422_美女放談.mpg	18	0:03:30	0:03:41	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
20	T19150422_美女放談.mpg	19	0:03:35	0:03:35	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
21	T19150422_美女放談.mpg	20	0:03:36	0:03:40	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
22	T19150422_美女放談.mpg	21	0:03:37	0:03:38	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
23	T19150422_美女放談.mpg	22	0:03:39	0:03:39	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
24	T19150422_美女放談.mpg	23	0:03:40	0:03:45	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。
25	T19150422_美女放談.mpg	24	0:03:40	0:03:42	ダブルクリックで再生	B	《S》A[M]《S》ああ そんなこと [N]ないでしょう(笑)。
26	T19150422_美女放談.mpg	25	0:03:44	0:03:45	ダブルクリックで再生	A	《S》A[N]まあ なんかに 25よって 緊張します。

図 6：音声文字化テキストと映像・音声との同期

ズの調整、パソコン画面上の位置変更などを行うことができる⁵⁾。

4. 4. タグ情報の付与

4. 3 節で行った音声文字化テキストと映像・音声との同期により、(4. 5 節の Excel の検索機能を使えば) 音声文字化テキスト上のある単語から、その発話時刻に該当する発話場面の映像・音声を参照することが可能となった。しかし、4. 3 節の同期のみでは、視線や表情、声の調子などの映像・音声上の情報から、そこで用いられている単語を検索することはできない。このことを行うには、発話場面の映像・音声から表現行動に関わる情報を取り出し、その情報をコーパスに与えなければならない。映像・音声の情報にはさまざまなものがあるが、本研究は、表現行動の一部として言語と非言語行動とがどのように関連しあいながら現れるかに注目していることから、まずは、非言語行動に関する情報を取り出すことにした。

非言語行動の中で、とくに注目したのは、視線・身振り・動作・表情情報である。その理由は、林 (1973) も指摘するように、これらが服装の種類を選択やその着脱など他の非言語行動に比べて言語に伴って現れる傾向が強く、このことから言語との関係を調べるにあたって他の情報より優先できると考えたからである。声の調子や大きさなども、言語と

ともに随伴的に現れる非言語行動ではあるが、これらは、映像を考慮しなくても、従来の音声コーパスで取り出せる情報であるので優先させなかった。

そのほか、話者情報（性別、年齢、職業、番組内での役割、出身地）、番組情報（ジャンル、チャンネル、曜日、時間帯、長さ）も付与してある。ただし、これらの情報が、発話場面の映像を考慮しなくても付与できる性質の情報であるのに対し、視線・身振り・動作・表情は映像を観察することではじめてわかる情報である。これらの非言語行動情報をどのように分類・コード化し、コーパスに付与するかなどは、従来のコーパス言語学では、ほとんど論じられてこなかった課題でもある。

4.4.1. 非言語タグ情報の付与

テレビの映像から伝わる話し手の視線・身振り・動作・表情に関する情報を、特定の機器やプログラムなどを使って自動的に抽出・分類・コード化することは当面困難であり、現時点では手作業で行うしか方法はない。したがって、本研究でも、これらの作業を手作業で行っているが、できるかぎり個人差を少なくするため、それぞれに専門の担当者を二人ずつ割り当て、そのうち一人は分類・コード化を、もう一人はその内容を確認するといった方法で作業を進めた。表2に、付与したコードの一覧を示す。

表2：視線・身振り・動作・表情情報の分類・コード化

視線	身振り・動作	表情
{A}：聞き手を見る	{M}：身振りをする	{SS}：明らかな笑い
{K}：カメラを見る	{D}：動作をする	{S}：微妙な笑い
{A#}：聞き手以外を見る	{M#}：身振り・動作をしない	{S#}：笑っていない
{X}：視線が観察できない	{X}：身振り・動作が観察できない	{X}：表情が観察できない
{Y}：分類に迷う場合	{Y}：分類に迷う場合	{Y}：分類に迷う場合

まず、視線は、大きく、聞き手を見る場合とそうでない場合に分類している。たとえば、Kendon (1967) は、視線には「モニター機能 (Monitoring functions)」「調整機能 (Regulatory)」「感情表出機能 (Expressive functions)」の3機能があると指摘し、聞き手を見る場合とそうでない場合ではそれぞれの働きが異なることを論じているが、まだ言語との具体的な関係は明らかになっていないところが多い。「聞き手を見る {A}」とは話し手の視線が聞き手に向けられた場合である。「聞き手以外を見る {A#}」は聞き手以外の、たとえば、地面や天井、話し手の手元、発話場面にある物などに視線が向けられた場合を意味する。聞き手以外のものを見る場合でも、カメラを見るときは別の分類 {K} とした。

話し手がカメラを見ながら話すときには、他と違って、視聴者に直接話しかける場合が多いからである。

次に、身振りと動作について述べる。身振りと動作には、多種多様なものが含まれるので、まずはそれぞれの出現の有無を問題とすることにした。「身振りをする【M】」には、McNeill (1992) の「エンブレム (emblem)」と「表象的身振り (representational gesture)」に該当するものが含まれる。たとえば、「了解」の意味を表す「OKサイン」や、指差し、発話内容と類似したことを表現する身体の動き（「昨日、私、階段から落ちて」と言いながら人指し指を下の方へ動かす場合など）などがある。国語研(1987)の「指示(例示)的動作」「模倣(模写)的動作／象徴的動作」がそれに当たる。「動作をする【D】」には、話し手が自分の体や道具を使って他の事物や人に物理的な影響を与える行為（物を運ぶ、聞き手の手を掴むなど）、規則的に繰り返し現れる癖に近い行為、汗を拭くなどの偶発的・生理的の行為などが含まれる。また、物理的な影響を与えてはいないが、おじぎをする、うなづく、首をかしげるなども「動作をする【D】」に分類した。国語研(1987)の「偶発的・生理的動作」「首または頭部の前傾(後傾)動作」「表出的動作(の一部)」が、それに当たる。「身振り・動作をしない【M#】」は、身振りも動作もなかった場合である。表2にはないが、身振りと動作が並行して現れた場合は【MD】に分類・コード化した。

最後に、表情は、「笑っているかどうか」の観点から分類している。「笑い」を基準に分類したのは、あらかじめ調査目的が言語と表情「笑い」との関係性を調べること(後述の5.3節)にあったためであり、特段、他の表情よりも重要性が高いからではない。笑っているかないかは、基本的には、工藤(1999)の(1)(2)を基準に、顔の筋肉にそれらの動きが生じたかどうかで分類した。ただし、(1)(2)をみたま場合でも、「眉の内側(外側)がもちあげられる」「上脛がもちあげられる」とき(驚いた場合などの表情)は、「笑っていない《S#》」に含めている。

(1) 唇の両端が耳の方向へ水平に引かれること

(2) ほおがもちあげられること

「明らかな笑い《SS》」とは、(1)と(2)に明確な動きがあった場合である。それには笑い声が伴う場合が多いが、そうでない場合もある。「笑っていない《S#》」は、(1)と(2)に動きがなかった場合をさす。そして、「微妙な笑い《S》」とは、(1)と(2)のいずれかにわずかな動きがあった場合であるが、分類に迷った例の中でも、とくに《SS》と《S#》の間で迷った例がこれに当たる。

以上のように、本研究では、視線・身振り・動作・表情に関する情報を、まずは、表2のようにおおまかに分類・コード化した後、必要に応じて下位分類を行っている⁶⁾。ただ

し、こうした分類さえも判断に迷う例は多い。中でも、表情は、その変化が連続的で分類が難しかった。

4.4.2. 付与方法

分類・コード化した視線・身振り・動作・表情情報は、音声文字化テキスト上に盛りこんだ。その際に、音声文字化テキスト上の全ての発話を対象に付与したものと、特定の言語項目にのみ付与したのものがある。図7は、前者の方法で情報を与えた音声文字化テキストの一部である。

文番号	開始時間	終了時間	話者記	発話文
23	0:03:41	0:03:50	A	《S#》[A][M#]若い頃から 子育ては 地元で するって 思ってたっていうのは ちょっと 不思議にも 私は《SS》感じるんですが 《X#X#X#X#》何で そう 思ってたんですかね？
24	0:03:50	0:03:50	B	《S》[A#][M]あ [A][M#]そうですか？
25	0:03:50	0:03:51	A	《X#X#X#X#》ええ。
26	0:03:51	0:04:28	B	《S》[A#][M]何でって 言われてもなあ 《SS》《S#》[M#]まー 恐らく その 自分が 子どもの 時に 受けた 影響 お そりゃ もう お 【M】自然 ん 環境であり え 人間環境であり えー 社会環境って う 【M#】いわゆる 環境 全てが 自分に んー そのー 人として えー 何を 感じる[A]べきなのかとか[A#] あー 何を 考えていく う [A]べきなのかとか[A#] えー 何を 《SS》経験する[A]べきなのかとか[A#]っていうのを 《S#》あー 全部 教えてくれた所[A]なんですよな やっぱ。

(ホリデーインタビュー：NHK 総合_2009年5月5日放送)

図7：視線・身振り・動作・表情情報の付与（音声文字化テキスト上の全ての発話対象）

ここには、表2の各記号が細かく記されている。それぞれは、話し手の視線・身振り・動作・表情に変化があったところのみ付した。ただし、文（文相当）が始まるころには、変化の有無を問わず付与した。たとえば、図7の文番号23の表情情報を見ると、《S#》から始まり、「感じるんですが」の直前で《SS》へと変わっている。このことから、「若い頃から 子育ては 地元で するって 思ってたっていうのは ちょっと 不思議にも 私は」では、笑う表情を伴わなかった話し手が、「感じるんですが」のところで表情を明らかな笑いへと変えていたことがわかる。そして、それ以降の「何で そう 思ってたんですかね？」では、話し手の表情が観察できない状態であったことも判断できる。このように、全発話を対象に情報を与えれば、それぞれの非言語行動が、言語（発話）とどのように関連しながら現れたかを、その変動の中で動的にとらえることができる。この方法で情報付与できたのは、全番組の中で、19番組（いずれも発話参加者が二人の場合）である。そのうち、7番組には視線・身振り・動作・表情情報を、12番組には表情に関する

る情報のみを与えている。

次に、調査する言語項目をあらかじめ設定し、その部分にかぎって、非言語行動情報を付与したものもある。たとえば、図8はその一部である。

文番	話者記	発話文	視線情報
2	C	あー すごい●ね。	{A}聞き手を見る
10	C	国会 世界のためにね 働かな 駄目だ●な。	{A#}聞き手以外を見る
17	C	すごいね あれは●ね。	{A#}聞き手以外を見る
53	D	え 立派に守ってます●よ。	{A}聞き手を見る
57	D	これは まあ とても 想像できません●ね。	{A}聞き手を見る
145	D	体調が 十分でないっていうのは 分かります●ね。	{A}聞き手を見る
369	C	えー しっかりしてない●でしょ。	{A}聞き手を見る
371	C	だから 簡単に 言う と 大盤振る舞いしますと ああ いいなと 国民は 思う んだけれども 2年後から 消費税で 回収しますって 話●でしょ。	{A}聞き手を見る

(時事放談：毎日放送_2009年4月19日放送)

図8：視線情報の付与（音声文字化テキスト上の特定の言語項目対象）

図8の「視線情報」には、「{A}：聞き手を見る」と「{A#}：聞き手以外を見る」とが並んでいるが、これは、同じ列「発話文」の「●」の後にある文字列を発話したときの視線に関する情報である。それをみると、文番号2・53・57・145・369・371の「●」の後の、「ね」「よ」「でしょ」の発話時に話し手の視線は「聞き手を見る {A}」だったのに対し、文番号10・17の「な」「ね」では「聞き手以外を見る {A#}」であったことがわかる。こうして、特定の言語項目にかぎって情報を与えると、短時間で大量に付与できるが、前後の変動を考慮した調査などはできない。この方法で、54番組を対象に、文末モダリティ表現（「ネ」「ヨ」「ナ」「カ」「ダロウ」「デハナイカ」など）が現れた5103箇所視線情報を付与した。

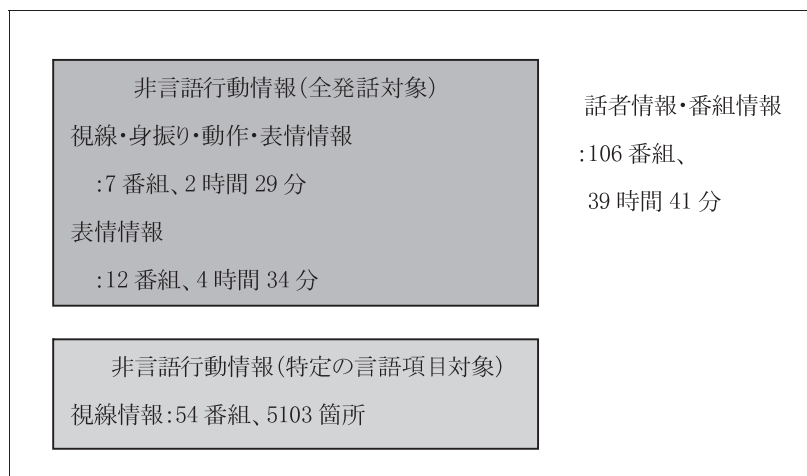


図9：「対談番組（限定）」におけるタグ情報

4.5. 検索

“Microsoft Office Excel”を使った上記のデータベースを本研究のコーパスとし、固有のインターフェイスは作成していない。したがって、文字列やタグ情報の検索・検索結果の絞り込みは、主としてExcelの機能を使って行うことになる。ここでは、その方法を、Excel（2007版）の「検索」と「オートフィルタ」機能を例に簡単に紹介する。後述の活用例は、実際にそれぞれの機能を使って行ったものである。

1) 「検索」機能

「検索」機能は、ある文字列やタグ情報を探し出すときに使う機能である。ここでは、「《SS》：明らかな笑い」がいつ現れたかを調べるときを例に説明する。

- ① 「ホーム」タブ→「検索」をクリック。すると、「検索と置換」というダイアログボックスが表示される。
- ② ダイアログボックスに、「《SS》」と入力し、「次を検索」をクリック。
- ③ 「次を検索」をクリックするたびに、「《SS》」を含む行が、次々と表示される。

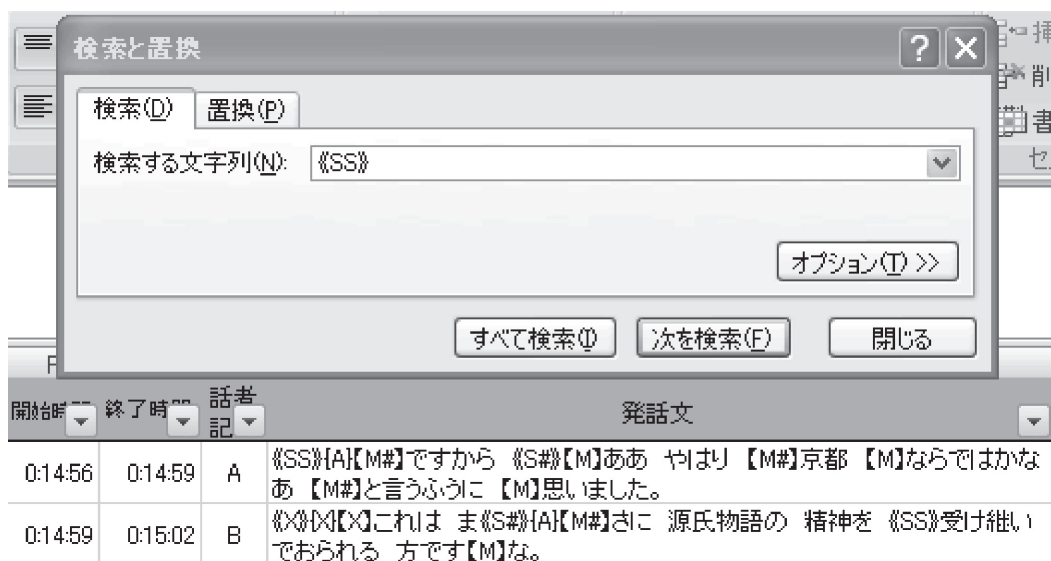


図 10：「検索」機能画面

- ④ それぞれの映像・音声は、映像クリップ（図6参照）をダブルクリックすると参照できる。

2) 「オートフィルタ」機能

「オートフィルタ」機能は、調べたい文字列やタグ情報を探し出すだけでなく、その

情報を含む行のみを抽出したいときにも使う機能である。その一例として、文末が「ネ」で終わった文にかぎってその際の視線を調べる場合を例に説明する。

- ① 先頭行全体を選択した状態で、「データ」タブ→「フィルタ」をクリック。すると、先頭行全体に「オートフィルタ矢印 (▼)」が表示される。

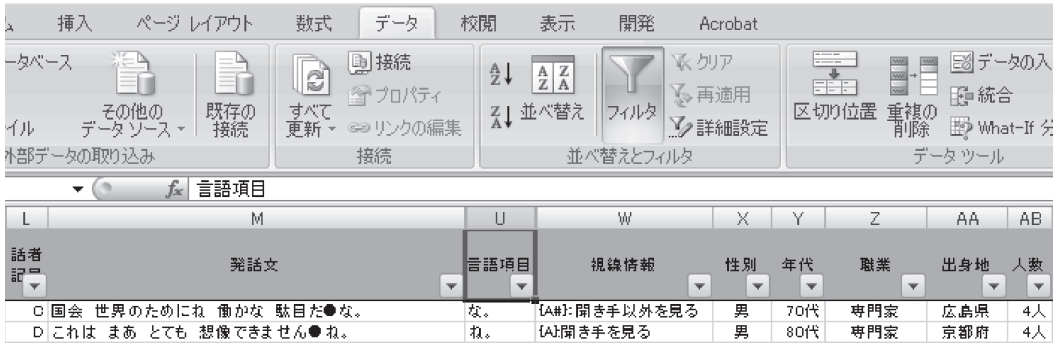


図 11: 「オートフィルタ」機能画面 I

- ② 「U.言語項目」には、「M.発話文」の文末モダリティ表現に関する情報が入力されている。その「オートフィルタ矢印 (▼)」をクリックすると、プルダウンメニューが現れる。メニューの中に、「ね」「よ」「な」などがあるが、その中で、「ね。」を選択する。
- ③ ②によって、「M.発話文」は「ね。」で終わる文のみが並ぶ。

L	M	U	W
話者記号	発話文	言語項目	視線情報
C	あー すごい●ね。	ね。	{A}聞き手を見る
C	すごいね あれは●ね。	ね。	{A#:聞き手以外を見る}
D	これは まあ とても 想像できません●ね。	ね。	{A}聞き手を見る
A	伝統を 守ってるんです●ね。	ね。	{X:視線が確認できない}
D	体調が 十分でないっていうのは 分かります●ね。	ね。	{A}聞き手を見る

図 12: 「オートフィルタ」機能画面 II

- ④ さらに、これと同じ方法で、たとえば、「X. (話し手の) 性別」「Y. (話し手の) 年代」にもフィルタをかければ、話し手の性別・年代からもデータを絞り込むことができる。
- ⑤ それぞれの映像・音声は、映像クリップ (図 6 参照) をダブルクリックすると参照

できる。

5. 活用例

上記の「対談番組のマルチメディア・コーパス」を使って試みた調査を三つ紹介する。いずれも、言語と非言語行動との関係についてのものである。それぞれにおける言語と非言語行動との関係については、すでに内省または実験観察的調査によっていくつかの指摘がなされてきたが、その実態が大量のデータにもとづいて論じられることはなかった。それには、いろいろな理由が考えられるが、その一つとして、調査方法の不在が挙げられる。その意味で、以下の調査は、マルチメディア・コーパスを用いることではじめて可能となったものとして位置づけられる。なお、各調査の詳細については、それぞれ、稿を改めて報告する予定である。

5.1. 活用例①：擬音語・擬態語と身振りとの関係

例えば「爆弾がドッカーンと爆発した」というとき、話し手は両手を素早く左右に広げようような身振りを伴う場合がある。Kita (1997) は、擬音語・擬態語の発話時には、このような身振り（映像的身振り iconic gesture）が、他の語類と違ってほぼ確実に現れるということ、「アニメ再生課題実験」という実験観察的な調査の結果から一般化して結論づけている。しかし、そこで調べられた擬音語・擬態語はわずか延べ83語であり、これらが何千語にも及ぶ日本語の擬音語・擬態語を必ずしも代表するとは思えない。

そこで、本コーパスの一部から、4.5節で紹介した方法で、擬音語・擬態語の使用例を検索し、その際の映像・音声から、発話時に身振りが現れたかどうかを調べた。分析に用いた擬音語・擬態語は延べ約1700語、異なり約600語である。擬音語・擬態語全般を代表させるほどの規模とはいえないが、調査の結果、身振りが現れる確率は、擬音語・擬態語の性質によって異なるということが明らかになった。中でも、身振りが頻繁に観察されたのは、A.文の中で副詞として働く際に「と」を義務的に伴うこと、B.臨時的または強調的な形をとること、C.音を表すこと、をより多くみだす擬音語・擬態語（たとえば、「ドッカーン」「ジャジャジャジャー」など）の発話時であった。田守・スコウラップ (1999) によれば、これらの語は描写力の高い語である。それに従えば、話し手は描写力の高い語を話すときほど身振りを頻繁に伴っていると考えられる。

5.2. 活用例②：終助詞ネと視線との関係

終助詞ネ（以下、「ネ」）には、さまざまな用法がある。ここでは、中でも、下の（1）（2）のように聞き手に同意や確認などを求めるような働きが感じられる場合と、（3）のようにそのような働きが感じられない場合について述べる。

- (1) a: 田中さんがいないとさびしいですね。 b: そうですね。 <同意要求>
 (2) a: 君はそのとき、たしか、中学生だったね。 b: はい。 <確認要求>
 (3) a: 何か不便な点は？ b: 学校が家から遠いことですね。 <自己確認>

蓮沼（1988）は、「ネ」の発話時における話し手の視線が、聞き手に同意や確認などを求めているかどうか積極的に関わるということを、内省にもとづいて指摘している⁷⁾。だが、その後、両者が実際にどのように関わりながら現れるかということが、大量の実例にもとづいて論じられることはなかった。

そこで、本コーパスの一部から、文末が「ね（え）」で終わる文（「ね。」「ね？」「ねえ。」「ねえ？」）を選定し、「ネ」と視線との関係を調べた。約1000例の使用例を集めることができたが、それらの用法を、視線を考慮せず音声のみであらかじめ分類した。分類に迷う例や、（1）（2）（3）以外の用法などは一旦除き、まずは用法の分類が可能だった約600例を対象に、映像からその際の視線を調べた。

その結果、（3）では話し手の視線が聞き手に向けられる場合もそうでない場合もあるのに対し、（1）（2）の発話時にはある状況以外ではほぼ確実に聞き手に向けられるということが確認できた。このことは、蓮沼（1988）の指摘とほぼ一致する。ただし、蓮沼（1988）は（3）のような場合に話し手は聞き手に目を向ける必要はないと述べるが、実際に調べた範囲では、（3）でも聞き手に目を向ける場合が過半数を超えていた。またある状況では、（1）（2）でも、視線が聞き手に向けられないこともあるなど、その実態は蓮沼（1988）の指摘よりも複雑であるように考えられる。

5.3. 活用例③：言語と表情「笑い」との関係

活用例①②では、あらかじめ調査する言語項目を「擬音語・擬態語」と「終助詞ネ」とにかぎっている。このことから、活用例①②は、言語の側面からみた言語と非言語行動との関係についての調査であるといえる。本コーパスを使えば、これとは逆に、非言語行動から両者の関係を調べることも可能である。ここで紹介する「言語と表情『笑い』との関係」はその一例である。

表情「笑い」に関しても、従来さまざまな分野で研究がなされてきたが、それと言語と

がどのように関わりながら現れるかを論じたものは少ない。志水ほか（1994）は、言語との関係の中で「笑い」を論じる数少ない先行研究の一つだが、そこでは笑いが「快の笑い」「社交上の笑い」「緊張緩和の笑い」などに分類されており、そこで基準とされているのは、言語というよりも言語使用場面に近い。本研究では、使用された言語にも注目しつつ、言語と表情「笑い」との関係を探ることを目的に、本コーパスの一部を使って調査を行った。ただし、まだ調査中で、計量的観点からの傾向などは見出せておらず、ここでは、方法論を紹介するに留める。

4.5節で述べた「検索」機能を使って、「明らかな笑い（《SS》）」で調べたところ、約3000例が見つかった。たとえば、図9の文番号23の《SS》はその一例である。

文番号	開始時間	終了時間	話者記号	発話文
13	0:02:45	0:03:06	B	《S#》{A#}[M#]あ あえて {A}選んだんではなくて {A#}東京へ 出て行く 時にもう すでに あー 結婚したら あー 伴侶と共に え この 田舎で{A} 育ちたいと 自分の {A#}故郷で また あー {A}生活したいと 《X》{X}[X]んー いう思いを 持って 東京 出て行ったんですね。
14	0:03:06	0:03:07	A	《S#》{A}[M#]うん。
15	0:03:07	0:03:24	B	《S#》{A#}[M#]んで ずっと その 間 やっぱ んー 帰って えー 生活をおー 再開{A}したい という 思いは {A#}自分の {A}[M]青写真[M#]として 人生の [M]青写真として ずーっと 持って たんで {A#}[M#]あー ま 《SS》{A}自分の 夢といいますかね
16	0:03:24	0:03:24	A	《X》{X}[X]はい。
17	0:03:24	0:03:26	B	《S#》{A}[M#]ーっ 人生の 夢ですよ{A#}ね。
18	0:03:26	0:03:32	B	《S#》{A#}[M#]うん それを えー 叶えることが {A}できたということですかね うん。
19	0:03:33	0:03:36	A	《S#》{A}[M#]おー ま 私から すると 結婚したら 地元で
20	0:03:36	0:03:36	B	《X》{X}[X]うん。
21	0:03:37	0:03:40	A	《SS》{A}[M#]戻って 子育てだっていうふうなこと 全然 こう 浮かないんですけど
22	0:03:41	0:03:41	B	《X》{X}[X]うん。
23	0:03:41	0:03:50	A	《S#》{A}[M#]若い 頃から 子育ては 地元で するって 思ってたっていうのは ちょっと 不思議にも 私は《SS》感じるんですが 《X》{X}[X]何で そう 思ってたんですかね？

図13：音声文字化テキストの一部（ホリデーインタビュー：NHK総合_2009年5月5日放送）

話者A（アナウンサー）は聞き手役を担っており、対談中に、ゲスト（俳優）の話者Bにいろいろと質問をする。Bは近年、東京から地元（秋田県）へ引っ越しており、この対談もそこで行われている。対談の中でBは、地元で暮らすことが昔からの夢であったと話す。それに対し、Aはこれまでそのように考えたことがないと話しながら、なぜ地元で生活したいと思っていたのか、その理由をBに問う。文番号23で、そのときのAの表情の変化を追ってみると、Aの表情が、《S#》から《SS》へと途中から変わっている。表情の変化があった前後の発話内容を見ると、「ちょっと 不思議にも 私は 感じるんですが」で、ここでAは、Bの考えとは異なった意見を述べている。

以上と同様の方法で、《SS》を基準に、その際に発せられた発話内容を調べてみると、おもしろい話やおかしい話をしているとき以外にも、《SS》は頻繁に観察される。中には、文番号23のように相手の主張とは異なった意見を述べるときや、相手にとって都合の悪い話を持ち出すときも多い。このことから、これらの場合に話し手は、何らかの意図をもって《SS》を用いており、《SS》もまたなんらかの働きを担っていると推測される。だが、その意図や働きが何であるかは今後より詳しく検討しなければならない。

6. 「対談番組のマルチメディア・コーパス」の利点と問題点

活用例①②③のような調査には、発話場面の映像・音声を参照することが欠かせないが、従来そうした機能を備えたコーパスがまったくなかったわけではない。筆者の知る範囲では、他にも、「インタラクション・コーパス」や「DVD映画コーパス」などがある。

インタラクション・コーパスは、人と人との相互行為（インタラクション）に関する分析・モデル化を主な目的に作られたコーパスである（詳しくは、角ほか2003参照）。このコーパスでは、複数の人々を同時に録画・録音し、それをコーパスの資料とする。したがって、複数の発話参加者による複数の非言語行動が同時に観察でき、聞き手の反応なども考慮しながら話し手の表現行動に関する分析が行える。坊農（2008）は実際にインタラクション・コーパスを使って、それに関する興味深い指摘を行っている。だが、インタラクション・コーパスは、その開発が、映像上の視線や身振りなどに関する情報処理や、それらを統合的に扱う技術改善を中心に進められているので、音声から伝わる言語をどのように処理するかなどはそれほど問題とされておらず、基本的には、音声文字化テキストをもたない。したがって、たとえば、ある単語が発せられたときに、どのような視線行動や身振りが行われたのか、あるいは、ある視線行動や身振りが行われたときにどのような単語が発せられたのかなどの調査には適していない。また、現時点では、大量にデータを集めることも難しい。

DVD映画コーパスは、映画DVDを利用したコーパスである。このコーパスでは、映画の字幕データと、その字幕（に書いてある発話）が発せられたときの映像とが同期されていて、たとえば、字幕データ上のある単語で検索をかけると、それに該当する発話場面の映像・音声が再生される（詳しくは、佐藤・中川2006参照）。発話場面の映像・音声の検索には、すでに備わっている字幕データが利用できるため、それを用いれば、別段、音声を書き起さなくても、該当する映像・音声が検索できるという利点がある。だが、DVD映画コーパスは、その開発の主たる目的が外国語教育にあることから、画面に映っている

非言語行動の分類・コード化などは考慮されていない。したがって、ある単語からそれが発せられたときの映像・音声を検索することは可能だが、そのときにどのような視線行動や身振りなどが行われ、またはある視線行動や身振りが行われたときにどのような単語が発せられたかなどは調べられない。

こういった点から、従来の「インタラクション・コーパス」や「DVD映画コーパス」では、活用例①②③のような、言語と非言語行動との関係についての計量的な実態調査は難しく、現時点では本コーパスが、この種の調査にもっとも適していて、なおかつ規模の大きいコーパスであると考えられる。それに加え、本コーパスには、性別や職業などの話者情報、チャンネルや放送時間帯などの番組情報も付与してあるので、活用例①②③に、それらがどのように関わっているかなども調べられる。

しかし、本コーパスには問題も多い。ここでは、それについて、(テレビの)対談番組を資料とすることによる問題と、コーパスを作る際に本研究で選択した方法による問題とに分けて考えたい。

対談番組を資料とすることによる問題として、まず、対談番組で行われる表現行動が自然談話でのものと必ずしも一致するとはかぎらず、得られた結果がテレビ特有のものである可能性があるということが挙げられる。そして、対談番組の映像から一貫した発話場面が観察できないことや、話し手に比べて聞き手が画面に映ることが非常に少ないということなども問題である。したがって、たとえば、話し手が聞き手を見ながら「ネ」を発話するときでも、具体的に聞き手のどこを見ていて、その際に聞き手の視線はどこに向かっていたかなどはわからない。こうしたことから、対談番組では、話し手の持続的な表現行動の観察や、細かい非言語行動の記述、聞き手との相互行為を考慮した分析などはできず、非言語行動情報に関してもおおざっぱな情報しか得られない。

コーパスを作る際に本研究で選択した方法による問題としては、第一に、作成した音声文字化テキストの精度が、高度な音声研究などにも対応できるほどに達しておらず、そのテキストも漢字仮名混じり文で書かれていることから、検索も文字列による検索しか行えない。第二に、音声文字化テキストと映像・音声との同期が文(相当)単位であることから、文の長さが長い場合に、その中で使われた単語などが瞬時には検索できない。第三に、付与した情報の種類や量が少なく、とくに見出し語・語種・品詞など言語情報が付与できていないので、言語と非言語行動との関係を調べるにあたっていろいろな制限が生じる。第四に、すでに付与した非言語行動情報に関しても、本来ならば杉戸(1978)が身振りについて提示した試案のように、精密に分類・コード化しなければならないところがおおまかにしかできていない。第五に、その情報の与え方にも問題があり、国語研(1987)や坊農

(2008)のように、音声文字化テキストとは別の情報欄を設けて、そこに時間軸に沿って非言語行動情報を並記すべきだが、音声文字化テキストの文字列上に盛り込んでいることから、精密に情報を与えられず、その情報も必然的に文字列の影響を受けている。第六に、検索の際に正規表現などが使えないので、比較的複雑な検索になると手間がかかる⁸⁾。

今後、マルチメディア・コーパスを表現行動研究のための有用な手段とするには、これらの問題を解決しながら、対談番組以外の資料をも利用したマルチメディア・コーパスの構築が必要であろう。あわせて、表現行動に関わる要素（あるいは、表現行動というものを眺める際に必要な観点）と、それらのうちのどのような情報がマルチメディア・コーパスに付与できるかに関する考察もなされなければならない。表現行動に関わる要素にどういったものがあるかは必ずしも明らかになっていないが、ハイムズ (1974)、ネウストプニー (1979)、南 (1987)、杉戸 (1999) などが指摘するものはその候補となりうると考える。また、将来的には、近年 (笑) 顔認識などで広く知られている画像処理技術や、音声処理技術の発達により、映像・音声を取り込むだけで、そこから伝わる情報、たとえば、話し手が誰で、映像の中のどこに位置していて、誰に向かって、どのような表情をしながら、どのような音調で話したかなどが自動的に、しかも精密に分類・コード化することも可能になると予想される。こうした自動処理が可能になれば、マルチメディア・コーパスの構築、大規模化も、より容易になるものと考えられる。

7. まとめと今後の課題

以上、本研究では、石井 (2009) にならって、言語のみならず、その使用場面の映像・音声をも同時に参照できる「マルチメディア・コーパス」の構築を、主としてテレビの対談番組を資料に試みた。まだ問題点や改善すべきところも多いが、本コーパスを使うことで、たとえば、ある単語が、どの番組で、誰が誰に向かって話すときに、どのような視線・身振り・動作・表情とともに発せられたものか、またはそうした使用状況でどのような単語が発せられたかなどが、現実の大量のデータにもとづいて、計量的観点から調べることが可能になった。

そして、実際にこのコーパスを使って、表現行動に関する分析、具体的には、「擬音語・擬態語と身振りとの関係」「終助詞ネと視線との関係」「言語と表情『笑い』との関係」の三つを調査したところ、擬音語・擬態語によって身振りが現れる確率は異なり、描写力の高いといわれる語の発話時ほど身振りが頻繁に現れるということ、また、「ネ」の用法によって視線の方向が異なるということなどが、計量的観点から傾向として見出せた。これ

らは、従来、内省や実験観察的調査によって指摘されてきた内容に修正・補充すべきところがあることを示すものであり、同時に、表現行動研究におけるマルチメディア・コーパスの有用性をも示すものである。

ただし、表現行動には、言語と非言語行動以外にもさまざまな要素が関わっている。たとえば、南（1987）が「談話の要素」として取り上げている参加者、話題、コミュニケーションの機能、表現態度、媒体、状況、ネットワーク、文脈などは、その候補となりうるもので、本稿で得られた傾向も、これらから影響を受けながら現れたものと推測される。したがって、表現行動の様相をより正確に捉えるためには、今後、これらの要素からも分析を行う必要がある。

そのためには、まず、各要素に関する情報を分類・コード化し、コーパスに搭載しなければならないが、これらの中には、発話場面の映像・音声から取り出せる情報も多い。このことから、6節で取りあげた問題点を改善しつつ、本コーパスをより高いレベルにしていくことで、今後、考慮できる情報の数や種類は増えていくと考える。一方で、対談番組の映像・音声からは取り出せる情報に制限があり、また、そこから得られる傾向が必ずしも（常識的な意味での）日常の談話場面に敷衍できるとはかぎらないことから、今後は、対談番組以外の資料を使ったマルチメディア・コーパスの構築も並行して行わなければならない。

注

- 1) ただし、ここでいう「表現行動」は、林（1973）が本来意図したものより、狭い範囲のものを指す。林（1973）は、話しことばのみならず書きことばも対象としており、視線や表情などはもちろん、人が作り上げた彫刻や建築なども表現行動の手段に含めて考えている。
- 2) 後述の「番組の選定」と「音声文字化テキストの作成」、付与したタグ情報（4.4節）の種類と量については、下表のように、資料間で異なる。

【付与したタグ情報（資料別）】

	「テレビ放送の語彙調査 (NHK 総合・教育)」	「ABC 放送」	「対談番組 (限定)」
付与した言語情報	見出し語 (読み・代表形)、語種、品詞	なし	なし
付与した非言語行動情報	視線・身振り・動作 (特定の言語項目対象)	視線・身振り・動作 (特定の言語項目対象)	視線・身振り・動作・表情 (全発話対象。ただし一部番組のみ) 視線 (特定の言語項目対象)
付与した発話参加者情報	性別、年齢、職業	性別、年齢、職業、 出身地	性別、年齢、職業、(番組内での) 役割、出身地
付与した番組情報	ジャンル、チャンネル、 曜日、時間帯、長さ、 視聴率	ジャンル、チャンネル、 曜日、時間帯、長さ	ジャンル、チャンネル、曜日、 時間帯、長さ

- 3) 話し手の年代・職業・出身地情報は『TVスター名鑑2009年版』(2008)による。そこに情報が記載されていない場合は、インターネット百科事典「Wikipedia」、あの一と検索「Spysee」を参照した。データベースには、情報源別に印を付してある。その他、肩書が複数ある場合は、一番前のものを選択した。
- 4) 入力ミスに関しては、映像・音声を確認しながら二度訂正したが、まだ完全とはいえない。分かち書きの区切り位置にもゆれがみられるところがあるが、まだ修正できていない。
- 5) EXCEL上でのMPlayer操作は、ソフト開発専門業者にマクロ開発を依頼して行った。
- 6) たとえば、5.1節の「擬音語・擬態語と身振りとの関係」では、「【M】:身振りをする」に分類したものを、「エンブレム (emblem)」、「映像的身振り (iconic gesture)」、「暗喩的身振り (metaphoric gesture)」、「直示的身振り (deictic gesture)」、「その他」に下位分類し、調査している。
- 7) 蓮沼 (1988) は、同意や確認などを求めるかどうかには視線が積極的にかかわることを主張し、視線によって、「ネ」の用法が変わりうることも示唆している。今後、そのことを確かめるには、(1) (2) と (3) の間で分類に迷った例も含めて分析しなければならない。
- 8) 正規表現とは、ある種の記号を組み合わせることで目的とする複数の文字列を簡略化して表す方法のことである。

引用文献

- 石井正彦 (2009) 「テレビ放送のマルチメディア・コーパス—映像・音声を利用した計量的言語使用研究の可能性—」『阪大日本語研究』21.
- 工藤力 (1999) 『しぐさと表情の心理分析』福村出版.
- 国立国語研究所 (1984) 『国立国語研究所報告80 言語行動における日独比較』三省堂.
- 国立国語研究所 (1987) 『国立国語研究所報告92 談話行動の諸相—座談資料の分析—』三省堂.
- 国立国語研究所 (1995・97・99) 『テレビ放送の語彙調査 I ~ III』大日本図書.
- 佐藤弘明・中川裕之 (2006) 「Web版多言語マルチメディアDVD映画コーパスの開発」『ことばの普遍と変容』
1 専修大学社会知性開発研究センター言語・文化研究センター.
- 志水彰・角辻豊・中村真 (1994) 『人はなぜ笑うのか』講談社.
- 杉戸清樹 (1978) 「身振りを記録する—「変位」の記録表試案—」『研究報告集1』.
- 杉戸清樹 (1999) 「表現行動の基盤」中村明ほか編『日本語表現』明治書院.
- 角康之・伊藤禎宣・松口哲也・シドニーフェルス・間瀬健二 (2003) 「協調的なインタラクションの記録と解釈」『情報処理学会論文誌』44 (11).
- 孫栄爽 (2009) 「終助詞ネの用法と視線行動—テレビ放送のマルチメディア・コーパスによる計量的分析—」『日本語学会 2009年度春季大会予稿集』.
- 孫栄爽 (印刷中) 「擬音語・擬態語と身振り—テレビ放送のマルチメディア・コーパスによる計量的分析—」『計量国語学』27-4.
- ネウストプニー J.V. (1979) 「言語行動のモデル」南不二男編『言語と行動』大修館書店.
- 田守育啓・ローレンススコウラップ (1999) 『オノマトペ—形態と意味 日英語対照研究シリーズ6』くろしお出版.
- 蓮沼昭子 (1988) 「続日本語ワンポイントレッスン」『言語』17-6.
- 林四郎 (1973) 「表現行動のモデル」『国語学』92 (『言語表現の構造』に再収).
- 坊塚真弓 (2008) 『日本語会話における言語・非言語表現の動的構造に関する研究』ひつじ書房.

南不二男 (1974) 『現代日本語の構造』 大修館書店.

南不二男 (1987) 「談話行動論」 国立国語研究所 (1987) 所収.

Hymes, D. (1974) *Foundations in Sociolinguistics : An ethnographic approach*. University of Pennsylvania Press. 唐須教光訳 (1979) 『ことばの民族誌—社会言語学の基礎—』 紀伊国屋書店.

Kendon, A. (1967) Some functions of gaze-direction in social interaction. *Acta Psychologica* 26.

Kita, S. (1997) Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35.

McNeill, D. (1992) *Hand and Mind*. Chicago: The University of Chicago Press.

Meyer, Charles F. (2002) *English Corpus Linguistics*. Cambridge: Cambridge University Press.

Teubert, W. & Krishnamurthy, R. (eds.) (2007) *Corpus Linguistics*. London: Routledge.

付記

「対談番組 (限定)」を資料とした「マルチメディア・コーパス」の作成にあたっては、(財) 博報児童教育振興会「第4回ことばと教育研究助成」の交付を受けた。また、本稿で用いた「テレビ放送の語彙調査 (NHK総合・教育)」の映像・音声データと音声文字化テキストデータは、日本学術振興会平成20年度科学研究費補助金 (基盤研究 (C)) 「計量的言語使用研究のためのマルチメディア・コーパスに関する基礎的研究」 (研究代表者: 石井正彦) の研究協力者として使用させていただいた。

(博士後期課程学生)

(2009年8月20日受付)

(2009年10月1日修正版受付)

(2009年11月2日再修正版受付)

(2009年11月19日再々修正版受付)

(2009年11月28日掲載決定)