

スペシャル・セッション〔自由発話の音声認識に向けて〕（招待講演）

1—3—10 言語研究における自発音声

○ 前川喜久雄(国立国語研究所) kikuo@kokken.go.jp

1. はじめに*

筆者の属する国立国語研究所では自発的なモノローグを対象とした『日本語話し言葉コーパス』(以下, CSJ: Corpus of Spontaneous Japanese と略称する)の構築を進めている。この研究は東京工業大学の古井貞熙教授を総括責任者とする文部科学省(科学技術庁)の開放的融合研究制度による総務省通信総合研究所との共同研究であり, 研究プロジェクト全体(1999年度より5年間)の目標は自発音声の音声言語情報処理研究の基盤を開拓することにおかれている[1]。

本稿の後半では現在までに構築されたCSJの資料を解析してその言語的特徴を概観するが, その前に自発音声とは何かという問題について若干の考察を加え, さらにCSJの特徴を紹介する。

2. 自発音声

従来実験音声学や音声言語処理で研究されてきた朗読音声(read/laboratory speech)には以下の特徴が認められる。

- A 発話の言語形式が予め決定されており, 結果的に内容も規定されている。
- B 通常, 練習を経た後に発話が収録され, 失敗とみなされた発話は収録の対象外となる。
- C 発話の場に話し手に影響を及ぼしうる聞き手が存在しないことが多い。

自発音声にはこの反対の特徴が認められる。つまり,

- A' 発話の言語形式が予め決定されていない。
- B' 収録が常にぶつつけ本番である。
- C' 発話の場に話し手に何らかの影響をおよぼしうる聞き手が存在している。

その結果, 自発音声と朗読音声との間には,

- 1) 朗読音声の文法形式は書き言葉に近いこと,
- 2) 朗読音声には, 言い間違いや言い淀みなど, いわゆる非流暢性が生じないか稀であり, 自発音声には多発しうること,
- 3) 自発音声には聞き手との相互作用に強く影響されること,

等の差異を指摘することができる。

このような自発音声を「自由発話」と呼ぶこともあるが, 筆者としては, 自由発話は拘束発話の対概念と規定したい。つまり, 発話の構成や継続時間を話し手が自由に制御できる発話が自由発話であり, そうでない

発話が拘束発話であるという区別を, 自発/朗読の区別とは別にたてたいと考える。いわゆるお喋りは自由発話かつ自発音声であるが, いわゆる講演音声の類は発表の内容や発表時間に一定の制約が存在するから, 自発的ではあっても自由ではない。

さらに自発/朗読という分類を二値的分类と捉えるべきでないことを強調しておきたい。例えば読みあげ原稿を用いた学会発表は上記Aの特徴に関しては朗読音声の特性をそなえているが, BやCに関しては自発音声である。また, A,B,Cの内部においても特徴の連続的な濃淡が観察される。連続性の存在は以下に報告するCSJの分析からも確認される。

3. CSJの特徴

CSJの設計と書き起こし基準の詳細については他の文献に譲ることとし[2,3], ここではCSJに格納される音声資料と書き起こしテキストの性格のあらましを解説する。

3.1 目的と規模

CSJは自発音声を対象とした音声言語処理研究と音声言語研究とのふたつの目的に資することを目的としている。前者については“自然”な話し言葉を対象とした音声認識のための言語モデルの構築が応用上の重要な目標である。

CSJの規模は700万語(時間にして700時間強)を予定しており, この全体について音声と書き起こしテキストを提供する。音声は講演全体をひとつのファイル(16khz, 16bit)とするが, 付加情報の形で, 200ms以上のポーズを基準とした転記基本単位(以下では基本単位と呼ぶ)への分割を施し, さらに「息」「笑い」「拍手」等の存在にもラベルをふる。

書き起こしテキストは, 語の検索に便利な漢字仮名混じり形式と音声の検索に便利な片仮名形式の双方を提供する。これらのテキストには言い間違いやフィルターなど自発音声を特徴づける諸現象, あるいは聞き取り不能箇所など20種程度のタグをふる。

テキストに対しては, 形態素解析を施し, 語境界と品詞の情報も提供する。形態論上の単位としては, 長短二種類の単位を採用する。例えば「国立国語研究所」は短単位では「国立/国語/研究/所」と分割されるが, 長単位では全体が一単位となる。

ところで, 必ずしも統計的なアプローチをとると限らない言語研究の立場からは, データの規模よりも, どちらかといえば付加情報の質が重要な問題である。CSJでは全体の一部50万語程度を「コア」と呼び, コアに

Spontaneous speech in the light of linguistics.
By Kikuo Maekawa (Nat'l. Lang. Res. Inst.)

対しては分節音ラベル、韻律ラベルその他の音声言語研究用情報を付加する。分節音ラベルは通常音声認識で利用される音素ラベルよりも音声よりのラベルセットを利用している。また、韻律ラベルとしては J_ToBI の拡張版を利用する予定であり、現在そのスキーム確定作業を進めている。

3.2 対象とする日本語変種

CSJ は日本語全国共通語のモノログを対象とした音声言語データベースである。CSJ では全国共通語を「多少なりとも公的な場面で発せられる母語話者による日本語で、文法と語彙が東京語に酷似した変種」と規定している。アクセント等の韻律特徴は共通語の要件としないが、コアに収録する音声に限っては、韻律まで東京語に酷似していることを要件としている。

3.3 発話のタイプ

CSJ には自発性が高低さまざまに相違する音声のパラエティを記録することを目指して、以下のような音声を収録する。

- 学会講演音声(300 時間程度)：日本音響学会を始めとする種々の専門学会における研究発表のライブ録音。
- 模擬講演音声(300 時間程度)：年代(20～60 代)と性別をバランスさせて人材派遣会社から派遣された話し手による少数の聞き手を前にした講演。講演の大まかなテーマは指定するが(Ex.「私の住んでいる町」「人生で一番悲しかったこと」)、具体的なテーマは話者に決定してもらう。講演時間は 10 分以上 15 分以内。
- その他(100 から 200 時間)：聴衆多数を前にした実際の講演、各種学校における講義、比較研究のための朗読音声等。

模擬講演を収録する目的は、学会講演では避け難い年齢や性別の偏りのないデータの確保、やはり学会講演では避け難い語彙の偏りの補正、またまった予稿が存在しないという点において学会講演よりも自発性の高い話し言葉の把握の三点である。

3.4 構築作業の現状

今年度末までに約 450 時間の音声を収録し、そのうち 320 時間程度の第一次書き起こし作業を終え、50 万短単位に対する形態論情報付与一次作業を終了する予定である。来年度からは通信総合研究所が形態素解析作業の自動化を開始する予定である。

4. 自発音声の音音的特徴

本節以下では CSJ に収録されたデータの予備的な分析結果を報告する。最初に音音的特徴を分析するが、現段階では分節音ないし韻律ラベリングをほとんど実施していないので、書き起こしテキストに記録された基本単位の開始/終了時間情報やテキストから計

算したモーラ数などの分析にとまらざるをえない。

分析では CSJ の講演データがどの程度日常の談話に近接しているか、また学会講演と模擬講演との間にどのような差異が観察されるかに注目する。

4.1 基本単位の長さ

CSJ では 200ms のポーズによって基本単位を認定している。図1はこの基本単位を構成する分節数の百分率である。音響学会(AS 約 30 時間)、国語学会(JL; 約 11 時間)、および模擬講演(AC 約 29 時間)のデータに加えて、国語研究所が 1950 年代に実施した『談話語の実態』研究のデータ(DZ)も示している[4]。

DZ データは東京の日常生活のなかで録音された対話を含む自由発話データ(約 30 時間)であり、CSJ のデータよりも自発性が高いと考えられる。CSJ に関しては模擬講演の方が学会講演よりも自発性が高いと想定される。また、同じ学会講演でも JL では AS よりも原稿を読みあげる形式の発表が多いので、AS よりも自発性が低いと想定される。

図1には、1 文節発話の比率が最も高く 2 文節以上は単調に減少してゆくパターンが観察される。文献[4]によれば、ニュースやニュース解説音声における 1 文節分の比率は 2% 程度であり、平均値は 16～21 文節に分布しているから、これは自発音声の顕著な特徴とみなしてよい。また 1 文節文の比率は、DZ > AC > AS > JL の順に低下しており、上述した自発性の予測と一致している(ただし DZ データと CSJ では基本単位の認定基準が異なっている点は注意が必要)。

模擬講演と学会講演を比較すると、3 文節までは AC の頻度が最も高いが、4 文節以上では AS、JL が逆転する。また基本単位の平均持続時間長ないしモーラ数を比較すると、模擬講演が学会講演よりも有意に

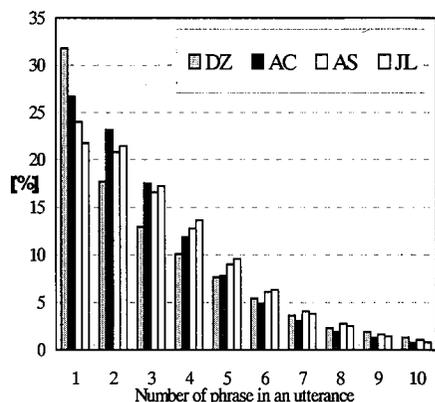


Fig.1 Percent distribution of utterance unit as a function of the number of syntactic phrase (*bunsetsu*) in an utterance.

短い($p < .001$). これらはいずれも、模擬講演が構文の単純さにおいて自然な話し言葉に近いこと、反対に学会講演にはある程度複雑な構文が用いられていることを示唆する結果である。

4.2 発話速度

基本単位に含まれるモーラ数を指標として発話速度を分析した。AS, AC, JL の3グループ関する分析結果を図2に示す。200ms以上のポーズを発話時間から除外したうえで、ファイルの冒頭、中間、末尾から抽出した各30基本単位の平均値をグループ毎に平均した値が示されている。

いずれのグループでも冒頭より末尾で発話速度が上昇している($p < .001$)。講演時間が不足すると上昇するのかそれとも一般に時間の経過とともに速度が上昇するのか、現時点では判断しがたいが、音声認識研究にとって重要な現象であろう。またASとJLの差は両学会の講演時間の差(12分と25分)を反映しているのかもしれない。

図2では測定位置を問わずに「AS>JL>AC」の順に発話速度が低下しており、学会講演は模擬講演よりも早口である。英語については朗読音声が発音よりも早口であることが報告されているので[5]、ここでもまた学会講演がより朗読音声に近いことが示唆されていると思われる。

4.3 (W) タグ

書き起こしテキストに付与されるタグのひとつに(W)がある。このタグは言い間違い・音声の転化、発音のなまけ等とみなされる現象に付与される。今回分析対象としたテキストに含まれた257893基本単位中に(W)は6412回出現していたが、そのうち2500例ほどは類型化することが可能な現象と思われる。以下に主要な

類型を例示する。

- **W** の脱落:「組合せ>クミアーセ」「川上>カーカミ」「ワ(副助詞)>ア」など(88例)。
- **W** の挿入:「場合>パワイ」「具合>グワイ」など(35例)。
- ナ行助(動)詞の撥音便化:「明らかになる>アキラカナル」「我々のところ>ワレワレントコロ」など(542例)。
- **R** の弱化:「それから>{ソレカー, ソエカラ}」「けれども>ケードモ」など(426例)。
- シとヒの混同:「ひとつ~シツ」「質問>ヒツモン」など(58例)。
- シとシュの混同:「試行>シュコー」「抽出>チューシツ」など(251例)。
- 狭母音の促音便化:「振幅項>シンブッコー」「建築家>ケンチックカ」「各グループ>カッグループ」など(68例)。
- 長母音の短縮:「傾向>ケーコ」「そう>ソ」「大域>タイキ」「言う(ユ) >ユ」など(927例)。
- テュ:「と言う>テュ(-)」(37例)。
- 撥音の長音化:「原因>ゲーン」「全員>ゼーイン」など(42例)。
- モーラ脱落:「ここ>コ」「~{です, ます}>~ス」「私>ダシ」など(94例)。

これらのなかには、音韻・文法環境からかなりの程度まで予測が可能な現象がある。例えば「狭母音の促音便化」は狭母音*/i/*ないし*/u/*が前後を無声子音に挟まれた環境、あるいは狭母音に後続する有声子音の調音位置が狭母音を含むモーラの子音と同一である場合に限って生じている。これなどは本来日本語話し言葉の文法の一部として記述されていてよい現象であろう。

「Rの弱化」と同一ないし類似の現象は、島根県出雲地方、東北の各地、沖縄などに存在することが知られている。日本語の*/r/*は音声学的には弾き音と言われるが、実際には調音上の自由度が高く、そのためには舌端がそり舌形状をとるだけでよい。そのために調音運動の怠けが生じやすく、弱化がおきやすい。上記諸方言と基本的に同一の現象が共通語にも生じていると考えられるが、まだ音韻現象として定着するに到っていない点は相違している。

「Wの脱落/挿入」「ナ行助(動)詞の撥音便化」「長母音の短縮」「モーラ脱落」などは、指摘されればすぐに存在に気づく身近な音声・文法現象である。しかし、これらの生起にかかわる言語的、社会的要因についての研究はほとんど進展していない。CSJの構築と解析が進めば、これら諸現象の生起確率を律する要因を研究する道がひらけるのではないかと思う。

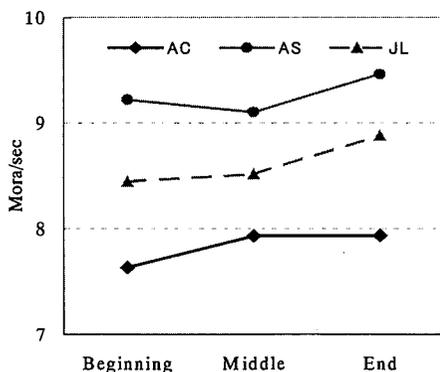


Figure 2. Comparison of speech rate between academic speech (AS, JL) and simulated public speech (AC) at three different utterance positions. Rate is measured as number of mora per second.

5. 自発音声の形態論的特長

続いて形態論的特長の分析に進む。CSJ が自発音声の特徴をうまくとらえているか、また現代日本語の社会的なユレを代表しているかに注目する。

5.1 品詞の分布

表1に学会講演(ACAD)と模擬講演(SIML)における品詞の比率を対照した。模擬講演の方が学会講演に比べて1)名詞が少なく、2)形容詞が多く、3)副詞が多いことがわかる。いずれもDZデータの分析結果とも一致しており[4]、再び模擬演の自発性の高さの傍証が得られたと考えられる。

Table 1. Percent distribution of POS in academic and simulated public speech.

	ACAD	SIML
フィラー	7.3	6.1
形状詞	1.3	1.5
形容詞	1.0	1.7
助詞	28.2	32.4
助動詞	10.5	12.5
接続詞	1.2	1.2
代名詞	1.5	2.4
動詞	12.5	12.9
副詞	1.9	3.8
名詞	27.1	19.8
連体詞	1.5	1.2
その他	6.0	4.3
総短単位数	166013	154776

5.2 語の融合

語の融合は話し言葉の顕著な特徴である。ここでは「デ+ワ」が「ジャ」に融合する現象に注目するが、この現象は文法的には助詞「デ」が融合するケース(Ex. 「日本では」>「日本じゃ」と助動詞「ダ」の連用形「デ」が融合するケース(Ex. 「この話題ではなく」>「この話題じゃなく」)に二分できる。表2, 3ではいずれのケースでも模擬講演に高い頻度で融合が生じており、やはり模擬講演の自発性が示唆されている。

Table 2. Fusion of case particle /de/ followed by topic particle /wa/. Frequency in CSJ

Type	/dewa/	/zya/	Total
ACAD	558	3	561
SIML	189	16	205
Total	747	19	766

Table 3. Fusion of auxiliary /de/ followed by topic particle /wa/. Frequency in CSJ.

Type	/dewa/	/zya/	Total
ACAD	299	72	371
SIML	114	85	199
Total	413	157	570

5.3 語形のユレと話者の出生地

最後に古くから東京において語形がユレていることが知られている語を表4にとりあげた。いずれの語も語形のユレに地域差が関係していると言われているので、話者の出生地との関係を分析してみた。

「ムツカシイ」の使用者は8名中7名が中部ないし西日本の出身、「ナクッテ」は11名全員が東京ないし神奈川の出身、「シンジク」の2名は東京の下町(葛飾)と北部近郊(埼玉県白岡)、「ジデンシヤ」は4名全員が東京都下の出身であった。いずれも従来報告されている地域差と一致しており、語形の地域差が現在も根強く残存していることをうかがわせる。

Table 4. Frequency distribution of selected lexical variation in Standard Japanese captured in CSJ.

語形	頻度	話者数	併用話者数
ムズカシイ	72	40	3
ムツカシイ	16	8	
ナクテ	135	69	6
ナクッテ	15	11	
シンジユク	18	7	1
シンジク	20	2	
ジデンシヤ	26	7	1
ジデンシヤ	26	4	

6. おわりに

CSJには講演のタイプによる自発性の高低が音声面でも文法面でも期待に添った姿で反映されていること、また言語研究で指摘されてきた日本語のユレの特性も反映されていることがわかってきた。まずは一安心である。今後は構築作業を急ぐと同時に、言語研究の観点からの解析も本格化させてゆきたい。

CSJは2004年3月の研究終了時には研究用に原則無償で公開する予定である[2]。

参考文献

- [1]古井他。「科学治術振興調整費開放的融合研究推進制度」音響学会誌, 56-11, 753-755, 2000.
- [2]前川他。「『日本語話し言葉コーパスの設計』。音声研究, 4-2, 51-61, 2000.
- [3]小磯他。「『日本語話し言葉コーパス』の書き起こし基準について」信学技報, NL2000-56, 55-60, 2000
- [4]国語研。「談話語の実態」国立国語研究所, 1955.
- [5]Hirschberg. In Holesmes ed. *Prosody: Theory and Experiment*. Kluwer Academic, 335-350, 2000.

謝辞: 古井代表をはじめとする融合研究メンバーに感謝します。なかでも品詞分布の相違を発見した小椋秀樹さん、短単位データのRDBを提供してくれた山口昌也さん、音声特徴データの解析を補助してくれた籠宮隆之さんの名前を記して特に感謝します。