

「とりあえず『日本語話し言葉コーパス』検索」 の可能性を検討する

—「雰囲気」という単語の発音を例に—

Validity of and Limit to “Searching the *Corpus of Spontaneous Japanese* for Now”

岡 田 祥 平
OKADA Shohei

キーワード

〔日本語話し言葉コーパス〕 the *Corpus of Spontaneous Japanese*

〔とりあえず検索〕 Searching for Now

〔日本語音声言語の実態〕 The Actual Condition of a Japanese Spoken Language

〔マクロ／ミクロな視点からの分析〕 Analysis from Narrow / Broad View

〔言語外的条件〕 Social Constraints

I. はじめに

近年の言語研究では、インターネット上の文書（以下、ネット文書と表記する）を巨大なコーパスと見なし、そこから用例を見出し、従来の研究では指摘されてこなかった新知見を提供するという潮流が生まれつつある^①。

その一方で、ネット文書を言語研究に安易に利用することに対し、注意を喚起する声もある（田中2003・荻野2004a・服部2004など）。その上で、田野村（2003）は、以下のように述べている。

インターネットを本格的な言語研究のための資料として捉えると、そこには種々の問題がある。しかし、同時に、昨今の急速なインターネットの普及によって初めて調査可能になったような現象が多いのも事実であり、その意味においてはインターネットの言語資料としての利用価値は過小評価できないものと思われる。

近年の言語研究において、「種々の問題がある」にもかかわらずネット文書を利用した研究が増えつつある背景は、結局は、上に引用した田野村（2003）のことばに尽きるだろう。

以上のように、近年、ネット文書を言語研究に利用しようとする様々な動きが見られる。しかし、注意しなければならないことは、ネット文書はあくまで書記言語であり、音声言語に生

起する音声・音韻をめぐる現象については、ネット文書を使用しても調査不可能だ、ということである²⁾。ネット文書を利用した言語研究が可能なのはあくまで書記言語に生起する事象に限定される。

それでは、音声言語に生起する事象の研究には、書記言語に生起する事象の研究にとってのネット文書のような存在、すなわち、比較的容易に、多量のデータに基づいて現象の実態を検証することができる「材料」は存在しないのであろうか。筆者は、国立国語研究所等が製作した『日本語話し言葉コーパス』(Corpus of Spontaneous Japanese。以下、英略称CSJと表記する。)が、そのような「材料」になり得ると考えている。

本稿では、CSJ検索、特に「とりあえずCSJ検索」の可能性と限界を考えていきたい。「とりあえずCSJ検索」とは、「あれっと思った」書記言語に生起する事象について「とりあえずネット検索」し、その結果を言語研究に利用することの可能性を検討した田中(2003)から着想を得た筆者による造語である。本稿では、「とりあえず検索」・「とりあえずCSJ検索」という類似した表現が出てくるが、次のように使い分けている。

- ・とりあえず検索：あれっと思った事象に出会ったとき、とりあえず「何か」を検索すること
- ・とりあえずCSJ検索：あれっと思った音声言語に生起する事象に出会ったとき、とりあえずCSJを検索すること

本稿では、まず、Ⅱ. ではネット文書と比較しながらCSJの概略を紹介する。そのうえで、Ⅲ. 以降では、「雰囲気」という単語の発音の実態を例にあげ、音声研究における「とりあえずCSJ検索」の可能性と限界を論ずることにする。

Ⅱ. 「とりあえず検索」の材料としての『日本語話し言葉コーパス(CSJ)』の概要・性格 — ネット文書と比較しながら —

『日本語話し言葉コーパス』とは、先にも触れたとおり、国立国語研究所等が製作した、日本語の自発発話の音声を格納した大規模コーパスである。日本語の音声言語のコーパスはCSJ以外にもいくつか存在している。しかし、CSJは、自発発話を音声学的な分析も可能なレベルの音質で録音していること、しかも、その録音音声ファイルが多様かつ精密な研究用の付加情報とあわせて公開されていること、などの点で、他の音声言語コーパスとは一線を画している。

CSJの詳細は前川(2004a)などで詳述されているので、本稿では、音声言語に生起する現象を「とりあえず検索」する際の対象としてのCSJの概要・性格を、書記言語に生起する現象を「とりあえず検索」する際の対象としてのネット文書と比較しながら、紹介していく。

表1 CSJとネット文書の性格

	CSJ	ネット文書
データの量・サイズ	有限 のべ話者数:1,417 人 発話時間数:662 時間 発話総語数 752 万語	有限であるが 膨大
言語資料のタイプ	大多数が独話	多種多様
データの流動性	ある	ない
産出者の属性情報	公開されている	不明
言語学的な 付加情報	ある 音声ファイルのほか、書 き起こしテキスト、形態論 情報、分節音情報、韻律 情報などが提供される	ない 生の文書のみ
表記のゆれ	ない	ある

表1は、CSJとネット文書の性格を対照表にしたものである。

まず、データの量・サイズについてである。CSJは大規模自発発話音声コーパスであるが、のべ話者数が1,417人、発話時間数が662時間、発話総語数が752万語とデータの量・サイズは有限である。一方、ネット文書は、データの量・サイズは有限とはいえ、非常に膨大である^③。したがって、データの量・サイズが限定されているCSJの場合には、生起頻度が低い現象の「とりあえず検索」は困難である可能性は否定できない。なお、CSJ製作の中心者の一人である前川喜久雄氏は、前川（2005b）で、CSJのデータの量・サイズは、「筆者の体験に基づいて音声学のための自発音声のデータマイニング的研究に必要と思われるデータ量の下限」を、ある程度満たしていると述べたうえで、CSJ以上の巨大な音声データベースの必要性は認めながらも、「従来音声学が取り扱ってきた問題が稀現象ばかりで構成されているわけではないから、研究に値する現象は」CSJのデータの量・サイズからも「数多く見つかるはずである」とも述べている。つまり、CSJのデータの量・サイズが有限であることは、大きな問題ではないという考え方も出来よう。そもそも、CSJのように何らかの情報が付与されたコーパス（CSJの付加情報については後述する）は、サイズが大きければ、それだけ付加情報の精度も下がる。それは、コーパスを利用した分析にとっては、時に致命的な問題になる（この問題については、伊藤2005などに詳しい）。したがって、コーパスのデータの量・サイズが大きければ大きいほどいいというわけでもない。本稿では、CSJのデータの量・サイズが有限であることが「とりあえずCSJ検索」にとっては有利なのか、それとも不利なのかについても検討したい。

第2は、言語資料のタイプについてである。CSJは音声言語の、ネット文書は書記言語の集合体と表現できるが、一言に音声言語、書記言語と言っても、それぞれのカテゴリの中には多様性が存在する。CSJの場合、格納している音声言語は、「学会講演」、「模擬講演」と呼ばれる独話（モノログ）が大半（CSJ全体の約90%・約605時間）を占める。その他、対話（ダイアログ）や朗読音声も格納しているが、独話と比較すれば少数である（合計約55時間分、詳細は前川2004aなどを参照）。また、発話のスタイルの観点からCSJに格納された音声言語の性格

を考えると、比較的フォーマルな場面で使用される言語と表現することができそうである。それゆえ、筆者は、CSJを真田信治氏が提唱する「標準語」(フォーマルな場面で話される言語、詳細は真田2000などを参照)を格納したコーパスと見なす立場をとっている(岡田・出野2006)。したがって、カジュアルな場面で使用される音声言語に生起する現象の分析には、不向きかもしれない⁽⁴⁾。一方、ネット文書は、フォーマルなものから個人のブログ・webといったインフォーマルなものまで、ありとあらゆるスタイルの書記言語が含まれている。それゆえ、ネット文書は「くだけた場面に表れがちな表現」の研究にも利用できるとされる(田中2003)。言語資料のタイプについての議論をまとめるならば、CSJは日本語の音声言語の一側面しか格納していないが、ネット文書は日本語の書記言語のかなりの範囲をカバーしている、と言えるのではないだろうか。CSJのそのような性格が「とりあえずCSJ検索」の際にどのような影響を与えるかについても、本稿では考えていく。

第3はデータの流動性についてである。「データの流動性」というのは、現在存在するデータに新たなデータが加わったり、逆に現在存在するデータが削除されたりすることを指す。CSJは現在公開されているバージョンからの更新がない限り、データの流動性はないと言えよう。したがって、製作された時期以降⁽⁵⁾に表れた「きわめて新しい現象」については対応できない。一方、ネット文書は「常に変化している」(服部2004)ので、非常に新しい言語現象についても対応ができる。ただし、データに流動性があるということは「ある検索結果を時間が経った後に再度確認することは必ずしもできない」(服部2004)という弱点にもなり得る。つまり、CSJにはデータの「再現性」があるが、ネット文書には「再現性」は保証されないということになる。日比谷(2003)では、「公開されたコーパス」を利用した研究の利点のひとつとして、「追試の可能性」、すなわち「複数の視点から、同じデータを検討すること」ができることをあげている。データに流動性がないCSJでは追試も可能であるが、データに流動性があるネット文書では追試は不可能である。データに流動性がないことは、新語・新現象研究という立場からは短所であるが、「追試の可能性」という立場からは長所であるといえる。

第4は、産出者の属性情報についてである。CSJの場合、「話者のプライバシーを侵害しないと判断された範囲で話者の属性情報」が公開されている(前川2004a)。「話し言葉の多様性の一部は、性別、出生地、居住歴、学歴、講演経験の有無など、話者の社会的属性に起因している」(前川2004a)とされるが、CSJでは公開された話者の属性を利用した研究も可能である。一方、ネット文書の場合は「発信者属性がほとんど分からない」(田中2003)。そのため、ネット文書では、発信者属性を考慮に入れた研究は難しいとされる(田中2003)⁽⁶⁾。

第5に、言語学的な付加情報についてである。CSJの場合、収録された音声すべてについて、以下の言語学的情報が提供されている(前川2004a)。

- ・ 精密な転記テキスト
- ・ 形態論情報(テキストを語に区切って品詞をつけた情報)
- ・ 節単位情報(転記テキストにおける節(clause)境界の情報)

さらに、「コア(Core)」と呼ばれる一部のデータ(約50万語)には、上記の情報に加え、以下のような言語学的情報が提供されている(前川2004a)。

- ・分節音ラベル
- ・イントネーションラベル
- ・係り受け構造情報（節単位を領域として、その内部での文節間の修飾関係を示した情報）
- ・要約・重要文情報（重要文とは、講演を要約する目的で抽出された転記テキスト中の重要部分のこと）
- ・談話境界情報（談話内部における話題の階層構造を示す情報）

以上のように、CSJでは多種多様な言語学的情報が提供されている。これらを利用すれば、たとえば、「形容詞連用形のアクセント」という複数の条件が加わったものでも、比較的容易に検索することができる。これは、「とりあえず検索」の際の利点になると考えられる。

一方、ネット文書の場合、CSJに付加されているような言語学的情報は提供されない。そのため、まずは検索エンジンで自分が検証したい現象の候補を見出した後、さらに何らかの手法を採用し、自分が検証したい現象を絞り込む必要がある。それゆえ、ネット文書を利用した言語研究が、必ずしも簡便ではないことを指摘する声もある（たとえば、荻野2006c）。

最後に、表記のゆれについてである。これは、自分が求める現象の検索する際に直面する問題である。CSJの場合、「転記テキストには、発話を漢字仮名まじりで表記した基本形と片仮名だけで表記した発音形の2種類」があるが、「基本形は主として情報検索のための利用を想定しているので表記にゆれを生じさせないことを徹底して追及」されている（前川2004a）。そのため、CSJを利用して「とりあえず検索」する際、表記のゆれは考慮に入れる必要はない。

一方、ネット文書の場合は、表記の統一が図られていないわけではないので、当然表記のゆれを内包している。田中（2003）によると、ネット文書を対象に用例を探す際には、特にカタカナ表記のゆれが問題になるようである。これはネット文書の「検索方法が検索エンジンのしびりに従う」ためである。

以上、CSJの概要を、特に「とりあえず検索」の際に必要なと思われる観点に着目して、ネット文書と比較しながら紹介してきた。CSJは音声言語研究用に設計されたものであるため、基本的にはネット文書よりは「とりあえず検索」に適していると考えられる。ただ、上述した通り、データの量・サイズ、ならびに言語資料のタイプが限定されていることが、「とりあえずCSJ検索」の際に何らかの影響を与えることも考えられる。そこで、Ⅲ．以降では、「雰囲気」という単語の発音の実態を例に、「とりあえずCSJ検索」を実践していくことにする。

なお、「とりあえずCSJ検索」に類する研究として、前川（2004b・2005a）がある。前川（2004b）では「日本」という単語の、前川（2005a）では「NHK」という単語の発音のゆれの実態をCSJで検索した結果を報告し、CSJによる検索の有効性を示唆している。ただ、前川（2004b・2005a）では、主に変異形の生起確率に言及しただけであり、発音のゆれに関与する言語外的条件などについての分析まではしていない。本稿の（副次的な）目的は、前川（2004b・2005a）が示唆したCSJによる検索の有効性をさらにもう一步深め、「とりあえずCSJ検索」の可能性を検討することにもある。

Ⅲ. 「雰囲気」という単語の発音の実態をめぐる議論

現代日本語における「雰囲気」という単語の発音の実態をめぐる話題が、人口に膾炙されることは少なくない。しかし、従来の議論は、主に研究者や日本語に関心のある人々の日常の観察の結果や内省、意識調査などを基にしたものが中心で、音声言語における実態について記述・報告した文献は、管見の範囲では見当たらない。そこで、「雰囲気」という単語の発音の実態は、「とりあえずCSJ検索」の格好の材料だとも考えられる。

「雰囲気」という単語の発音のゆれの実態に関して言及した文献は、管見の範囲でも、井上編(1997)・近藤(1998)・「ことば会議室」(2000)・「日本語調査結果(99)」(2004)・鳥飼(2004)・窪蘭(2006)など、印刷媒体は勿論のこと、web上でのやり取りも存在し、枚挙に暇がない。それらの文献(ネット文書も含む)の多くで共通しているのは、日本語研究者などによる日々の観察や内省、あるいは意識調査に基づいた議論を展開している点、「雰囲気」を「フインキ」と発音する人が近年目立つようになってきたという点(「フンイキ」よりも「フインキ」の方が多数派だとする文献も存在する)、そして、「フインキ」と発音するのは若年層に多いという点である。また、どうして「雰囲気」という単語の発音にゆれが生じるのかということについて検討した文献もある(岡島1996・道浦2000・浅野2001・2004など。上に紹介した「ことば会議室」2000・鳥飼2004・窪蘭2006にも、この点についての言及がある)。

以上に紹介した文献は「雰囲気」を「フインキ」と発音している場合についての考察であったが、「おしえてねドットコム!」(作成年不明)には「フイーキ」という発音の存在を認める記述もある。

ただ、以上列举したように、「雰囲気」という単語の発音の多様性について言及する文献は数多く存在しているものの、自発的な音声言語における実態については詳らかではない。

Ⅳ. 分析対象・分析方法

今回の分析の対象資料は、繰り返し述べているように、国立国語研究所等が製作したCSJである。Ⅱ.でも触れたようにCSJは独話中心のコーパスであるが、今回は一定以上の用例数を確保するという観点から、CSJに格納された全てタイプの発話(独話・対話・朗読)を分析の対象にした。これは、Ⅱ.で詳述したとおりCSJはデータの量・サイズが限定されているが、そのようなCSJの中からさらに、たとえば「独話のみ」と発話のタイプに限定を加えると、分析に耐え得るだけの用例数を確保できなくなることを懸念したための処置である。

分析の方法は、以下のとおりである。既にⅡ.で紹介したように、CSJの転記テキストは、「発話を仮名まじりで表記した基本形と片仮名だけで表記した発音形の2種類がある」(前川2004a)。今回は、まず、「主として情報検索のための利用を想定しているので表記にゆれを生じさせないことを徹底して追及した」(前川2004a)基本形を利用し、CSJに出現する「雰囲気」という単語を抽出した。次に、抽出した「雰囲気」という単語について、「人間の耳で聞き分けられ仮名文字で表現できる範囲で、可能な限り正確に表記された」(前川2004a)発音形を利用し、発音の実態を調べた。

なお、今回は、公開された話者属性情報なども利用し、「雰囲気」という単語の発音の多様性に関与する言語外的条件についても、簡単ではあるが検討することにする。

V. 「雰囲気」という単語の発音のゆれ：CSJにおける実態⁽⁸⁾

前節で説明した手法に基づき、まず転記テキストの基本形を利用し、CSJに出現した「雰囲気」という単語を抽出した。その結果、異なり人数で193人の発話データから373回、「雰囲気」という単語が出現していた。次に、その373回について、転記テキストの発音形を利用し実際に生起した発音をまとめてみると、表2のようになった。

表2 CSJに観察された「雰囲気」という単語の発音

Wタグなし	フナイキ・(? フ)ンイキ・フ(? ン)イキ・フ(? ンイ)キ・フナイ<H>キ・フナイキ<H>
Wタグあり	フインキ・フイーキ・フイーンキ・フイーンキ・フイキ・フイン(? キ)・フイーキ・フーンキ・フンイク

*なお、表2中の各種タグは、以下のような場合に付与される(小磯ほか 2004)。

- ・Wタグ: 転訛, 発音の怠けなど, 一時的な発音エラー
- ・(?): 聞き取りや語彙の判断に自身がない場合
- ・<H>: 非語彙的な母音の引き伸ばし

表2上段・Wタグなしのものは、基本的には「フナイキ」という規範的な発音の枠内での変異の系統と言える。一方、表2下段・Wタグなしのものは、「フインキ」に代表されるように、非規範的な発音の変異の系統だといえる。

さて、表2の結果を見ると、「雰囲気」という単語には従来指摘されてきた「フインキ」(や「フイーキ」)以外にも多様な発音が存在することが分かる。異なり変異形数は15となる(規範的な「フナイキ」という発音も含む)。

ただし、上記にまとめた結果は、生起頻度が低く、偶発的現象・個人的な発音の癖によるものも含まれている可能性は否定できない。以下の分析では、そのような偶発的現象・個人的な発音の癖によるものは除外しなければならないであろう。ただ、では、生起頻度がどれくらい低ければ、偶発的現象・個人的な発音の癖によるものとみなせるのか、議論があろう。筆者は、岡田(2006a)で、生起頻度が極端に低いながら複数の人間に観察されるものは変異現象の可能性あると述べた(勿論、これは、言い間違い・偶発的現象が同時多発的に生じた蓋然性を否定するものではない)。そこで、今回も、岡田(2006a)にしたがって、ある特定個人にのみに観察された発音は言い間違い・偶発的現象の可能性が高いと判断し今回の分析からは除外する一方、生起頻度は(極端に)低いものの複数の人間に観察されるものは変異現象であると見なし、分析の対象に含めることにした。

以上のような観点から、CSJに現れた「雰囲気」という単語の発音(表2)のうち、(生起頻度は極端に低いものの)複数の人間に観察されるものを調べた結果、以下の表3のようになった。

表3 「雰囲気」の発音の変種・変異と生起度数・生起人数

実際の発音(発音形の表記)	生起度数(回)	生起人数(人)
フンイキ	272	160
フインキ	59	36
フ(?)ンイキ	12	11
フイキ	10	7
フ(?)ンイキ	5	4
フイーキ	4	4
フーイキ	2	2
フンイキ<H>	2	2
合計(生起度数)/異なり人数	366	193

*斜体字はWタグが付与された発音

**生起人数の最下段は異なり人数なので、生起人数の合計とは一致しない。(複数の変種を併用している人もいるため)

表3からまず読み取れることは、「雰囲気」という単語の発音でもっとも多く生起しているのは、規範的だとされる「フンイキ」だということである。生起度数で272回（総生起度数366回の約74.3%）、生起人数では160人（異なり人数193人の約82.9%）が、「雰囲気」という単語を規範的とされる「フンイキ」という発音で実現させている。これは、「フインキ」という非規範的な発音が目立つ（多数派だ）とする、Ⅲ.で紹介したいいくつかの先行文献との指摘とは一致しない。

この理由としては、主に以下の二つの可能性が考えられるであろう。

一つは、先行文献での指摘は主に意識調査の結果であり、意識・内省と実態との間に乖離があるという可能性である。近年、日本語母語話者の文法判断と実態との間に「ずれ」があることを指摘・検証する動きがある（杉村2002・川口2006など）。母語話者は自分が普段話している音声言語の音声実態、特に分節音については、文法以上に無自覚であるとも考えられる⁹⁾。したがって、今回の「とりあえずCSJ検索」の結果（表3）と先行研究の言及との乖離は、音声学の現象における母語話者の内省と実態との「ずれ」を意味している可能性がある。

もう一つの可能性は、CSJに格納されている音声の性質である。Ⅱ.でも指摘したように、CSJに格納された音声は基本的に独話で、その多くが日常会話と比較して相対的にフォーマルな場面で話されている「標準語」である。一方、日本語母語話者の中には、カジュアル場面では「フインキ」と発音しているが、フォーマル場面では「フンイキ」と発音しているという人も存在していると考えられる。そして、そのような人が意識調査で「雰囲気」という単語の発音を尋ねられたとき、フォーマル場面での発音ではなく、普段の日常生活での発音を意識し、「フインキ」と発音する、と答えている可能性もあろう。

以上のどちらの可能性が的を射ているかは、カジュアル場面の音声を収録し、分析しなければいけないだろう。表3の結果は、非規範的な発音を調べる場合、「とりあえずCSJ検索」だけでは明らかにならないこともあるということを意味しているかもしれない。ただ、「日本語調査結果(99)」(2004)の全ての世代を合わせた結果で、「雰囲気」という単語を「フインキ」

と「自分は読む」と答えた人が16.2%、「フインキ」と読む人が「周囲にいる」37.8%であり、「フンイキ」という規範的な発音のほうが優勢であった⁽¹⁰⁾。この傾向はCSJでの分析結果と極端に齟齬をきたすものではない。したがって、音声学的現象における母語話者の内省と実態との間に、必ず乖離が認められるというわけでもなさそうである。

また、表3からは、「雰囲気」という単語の代表的な変異は「フインキ」である、ということもわかる。「雰囲気」という単語について、規範的な発音である「フンイキ」と代表的な変異である「フインキ」の「カバー率」(「雰囲気」の総生起数に対して「フンイキ」・「フインキ」が占める割合)を表3から求めると90.4%となり、表3の上位2位までの発音が、「雰囲気」という単語の発音の大部分を占めていることがわかる。同時に、表2からは種類が減ったとはいえ、表3でも「雰囲気」という語の発音の多様性は多少なりとも維持されていることが分かる。表3によると、異なり変異形数は8である(「フン(?)イキ」・「フン(?)ンイキ」・「フンイキ<H>」を「フンイキ」に含めると考えても、異なり変異形数は5である)。「フンイキ」・「フインキ」が多数派とはいえ、のべ人数で33人は「雰囲気」という単語を多数派である「フンイキ」・「フインキ」以外の発音で実現させていることにも着目したい。

それでは、「雰囲気」という語において、どのような場合に規範的な「フンイキ」という発音以外のものが生起しやすいのであろうか。次節以降、用例数を確保すると言う観点から、表3で生起人数が多く、なおかつ「雰囲気」という単語の発音のカバー率が90%を超える、「フインキ」と「フンイキ」という発音に限定して、非規範的な「フインキ」という発音が生起する言語外的条件の分析を試みる。Ⅵ.では、個人差を捨象し、どのような属性を持つ人の集団、あるいはどのようなタイプの発話で「フインキ」という発音が生起しやすいかを分析する。続くⅦ.では、個人差に着目し、個人内での使い分け・ゆれの分析を行なう。いわば、Ⅵ.はマクロの視点からの分析であり、Ⅶ.はミクロの視点からの分析ということになる。

Ⅵ. 「フインキ」生起の言語外的条件の分析①：個人差を捨象し、話者属性・発話のタイプに着目した分析（マクロの視点からの分析）

先行文献のいくつかでは、「雰囲気」を「フインキ」と発音する人は若年層に多いと指摘するものがあった。そこで、話者の世代差と「雰囲気」という単語の発音(「フインキ」か「フンイキ」か)の関係を探ってみた。その結果が以下の表4である。また、あわせて表4で示した「フインキ」の生起確率をグラフ化したものが以下の図1である。

表4 「雰囲気」の発音の生起度数・生起確率と生年代

生年代	10年代	30年代	40年代	50年代	60年代	70年代	80年代	全体
フンイキ	3(75.0%)	18(90.0%)	22(100%)	64(87.7%)	61(77.2%)	103(78.6%)	1(50.0%)	272(82.2%)
フインキ	1(25.0%)	2(10.0%)	0(0%)	9(12.3%)	18(22.8%)	28(21.4%)	1(50.0%)	59(17.8%)
人数	1人	13人	18人	40人	40人	69人	2人	183人

* 単位なしの数字が生起度数、()は各生年代別生起確率

** 生年代はすべて1900年代

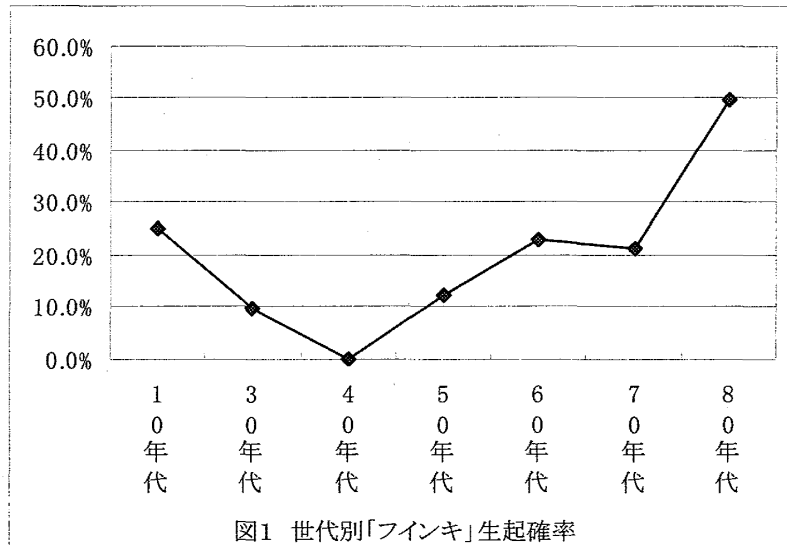


表4をみると、分析対象者のうち、1910年代生まれは1人、1980年代生まれは2人しかいないので、この二つの世代は世代差の分析に含めるのは不適當かもしれない。そのうえで表4と図1を眺めると、以下のような傾向を読み取ることができよう。

一つは、どの世代も、「フンイキ」という規範的な発音が多数派であるということである。1980年代生まれを除き、「フンイキ」という規範的な発音が生起する確率がほぼ7割を超えている。また、「フインキ」という非規範的な発音が生起する確率が若年層にかけて増加する傾向を読み取れる。ただし、1960年代生まれと1970年代生まれのあいだには差を認めることができない。1960年代生まれを若年層と見なすべきか否か、先行文献でいう「若年層」とは何を念頭に置いているのか、という議論はあるが（この件については後述する）、表4ならびに図1の結果は、「フインキ」という発音は若年層に多いという先行文献との指摘に、概ね一致していると言えよう。

次に、世代差以外の要因についても検討することにした。まず、表5は話者の性別と発音の関係を探ったものである。表5を見る限り、「フインキ」の生起確率は、女性で22.5%、男性で10.5%となり、女性のほうで「フインキ」という発音が生起しやすいと解釈できる。

表5 「雰囲気」の発音の生起度数・生起確率と性別

	女	男
フンイキ	161(77.8%)	111(89.5%)
フインキ	46(22.2%)	13(10.5%)
人数	110人	73人

*単位なしの数字が生起度数

**()は性別生起確率

さて、Ⅱ.でもふれたように、CSJは独話中心のコーパスであるものの、対話や朗読音声も格納されている。また、独話にも「学会講演」と「模擬講演」という2種類が設定されている。前者は各種学会での発表音声を実況録音したもの、後者は人材派遣会社から派遣された話者に

よる一般的なテーマについてのスピーチである（前川2004a。本稿の註4も参照のこと）。そして、統計的には、学会講演よりも模擬講演のほうでスタイルの低下が観察されるという（前川2004a）。また、CSJには学会講演・模擬講演といった独話以外に、量が少ないとはいえ対話（約12時間）や朗読音声（約20時間）も格納されている。それらを利用すれば、「雰囲気」という単語の発音と発話のタイプとの間に関係が認められるかどうか、調べることも可能である。そのような観点から、発話のタイプごとに規範的な発音とされる「フンイキ」と非規範的な発音とされる「フィンキ」の生起度数と生起確率を求めた結果が表6である。なお、表6中、独話の「その他」とは、「一般向けに開催された博物館の連続講演会、専門学校における日本語学の講義、国立国語研究所が開催した一般向け講演会の独話」を録音したもので、「いずれも学術的な講演であるが、話し手が専門家、聴き手が一般人という点で、学会講演とは相違している」という（前川2004）。また、表6での音声ファイル数と表4・表5での人数とが一致しないのは、1人で複数のスピーチを担当している場合もあるためである。

表6 「雰囲気」の発音の生起度数・生起確率と発話のタイプ

	独話			対話	朗読
	模擬	学会	その他		
フンイキ	228(80.3%)	21(91.3%)	9(100%)	5(71.4%)	9(100%)
フィンキ	56(19.7%)	2(8.7%)	0(0%)	1(16.7%)	0(0%)
音声ファイル数	182	18	4	4	6

*単位なしの数字が生起度数 **()は発話のタイプ別生起確率

表6からは、模擬講演に音声ファイル数が偏っていることにより用例数も模擬講演に偏っているため、ある種の傾向を見出すことは困難である。ただ、強いて言うなれば、模擬講演で「フィンキ」という非規範的な発音が生起しやすい傾向が認められるのではないか。また、数は少ないものの、朗読では規範的な「フンイキ」という発音しか生起していないことにも注意したい。同時に、朗読より用例数が少ないにもかかわらず、対話で1例であるが「フィンキ」という非規範的な発音が生起していることも注目に値しよう。少なくとも表6を見る限りでは、「フィンキ」という非規範的な発音の生起しやすさの序列は、概ね「朗読＝その他<学会講演<対話<模擬講演」となる。なお、この順番は前川（2004a）のいうCSJに格納された音声の自発性の序列（自発性が低いとされるほうから「再朗読ないし朗読<学会講演<模擬講演<インタビュー<課題指向ないし自由対話」⁽¹¹⁾）に概ね呼応している⁽¹²⁾。このことから、「フィンキ」という非規範的な発音の生起と発話の自発性には何らかの相関があると言えそうである。

さて、ここまでの分析では、前節では、「雰囲気」という単語の発音の多様性に関与していると思われるいくつかの要因について検討してきた。しかし、各項目ごとの分析のため、どの要因がもっとも大きな影響を与えているのか、判然としない。そこで、本節のまとめとして、ここまで取り上げた要因について「CHAIDによる決定木分析」を行ない⁽¹³⁾、「雰囲気」という単語の発音の多様性に最も大きな影響を与えている要因を浮き彫りにしたい。

ここまで取り上げた要因について「CHAIDによる決定木分析」を行なった結果が、図2である。

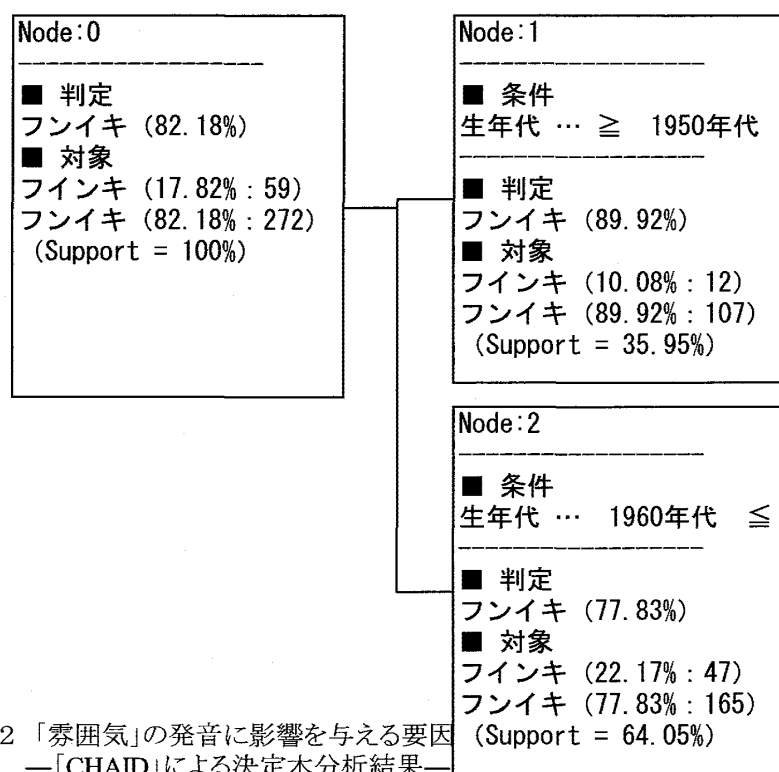


図2 「霧囲気」の発音に影響を与える要因
—「CHAID」による決定木分析結果—

図2を見ると、話者の生年代・話者の性別・発話タイプという三つの要因のうち、話者の生年代という要因が「霧囲気」という単語の発音の多様性に最も大きな影響を与えていることが推察できる。

図2によると、話者の生年代によって、1950年代生まれより年上のグループ（いわば中高年層）、1960年代生まれより年下のグループ（いわば若年層）と、二つのグループに分けることができる。そして、両グループとも規範的な「フンイキ」という発音が多数派であることがわかる。また、若年層で非規範的な「フインキ」という発音が増える傾向も認められる。これらの事実は、既に表4・図1の解釈でも述べたとおりである。ただ、ここでいう「若年層」とは、1960年代生まれが含まれていることに注意したい。註5で述べたことを踏まえると、音声収録は西暦2000年前後に行なわれたと推定できるので、1960年代生まれは概ね30歳代、1970年代生まれは概ね20歳代、1980年代生まれは概ね10歳代（後半）⁽¹⁴⁾であったと推測できる⁽¹⁵⁾。久野（2006）によると、「従来の方言研究では」「二〇代ないし三〇代は若年層とする」とあるが、Ⅲ．で紹介した先行文献の多くが念頭に置いている「若年層」とは10歳代・20歳代のような（註10の付表も参照）。したがって、今回の「とりあえずCSJ検索」の結果は、「フインキ」という非規範的な発音が、先行文献が指摘するよりも上の世代でも使用されている、ということの意味しているといえよう。

以上、本節では、どういう属性を持つ人の集団、あるいはこういった発話のタイプで「霧囲気」という単語が「フインキ」と発音されるか、個人差を捨象したマクロの視点から分析を試みた。その結果、若年層（1960年代・1970年代・1980年代生まれ）に「フインキ」という非規範的な発音が比較的多く観察される、という結果を得ることができた。このことから、少なく

ともマクロな視点からの分析には、「とりあえずCSJ検索」は、一定の有効性があると言えるのではないかと考えている。

VII. 「フインキ」生起の言語外的条件の分析②：個人差に着目した分析（ミクロの視点からの分析）

本章では、CSJに格納された各話者における「雰囲気」という単語の発音の変異・変種の使い分け・ゆれの実態について分析していく。筆者は、ある話者が同一現象について複数の変異を併用している場合、基本的には、話者が複数の変異を意図的に使い分けている場合を「使い分け」、話者が無意識に複数の変異を併用している場合を「ゆれ」と表現する立場をとる⁽¹⁶⁾。しかし、CSJを使用した分析の場合、話者に発話意図や意識をフォローアップインタビューなどで確認することは不可能である。したがって、「使い分け」と「ゆれ」を峻別することも不可能である。そこで、本稿では、CSJに格納された個人における「使い分け・ゆれ」の実態、と両者を区別しないことにする。

さて、CSJに格納された各話者における「雰囲気」という単語の発音の変異の使い分け・ゆれの実態について調べるにあたっては、各話者が相当数、「雰囲気」という単語を使用していなければならないだろう。たとえば、「雰囲気」という単語が1回しか出現せず、それを「フインキ」という非規範的な発音で実現している話者を「この話者は『フインキ』しか使わない話者である」と見なすことができないのは言うまでもない。というのも、たまたま限られたデータでは、その話者は「雰囲気」という単語を1回しか使用せず「フインキ」という非規範的な発音で実現させたが、もしさらに「雰囲気」という単語が生起する機会があれば「フンイキ」という規範的な発音が生起していた蓋然性も否定できないからである。しかし、では、一個人につき「雰囲気」という単語が何回以上使用されていれば、各話者の使い分け・ゆれの実態を分析できるのかという問いには、明確な回答を与えることはできないであろう。そこで、本稿では、便宜的に「雰囲気」という単語が3回以上使用した話者・42人について、発音の変異の使い分け・ゆれの実態を分析していくことにする。

CSJにおいて、「雰囲気」という単語を3回以上使用した話者の発音の変異の使い分け・ゆれの実態をまとめたものが、以下の表7である。表7中の「フンイキ」から「総計」までの数字は生起度数である。また、「フンイキ率」とは「フンイキ」という規範的な発音の生起率（「フンイキ」の数／総計×100）、「フインキ率」とは「フインキ」という非規範的な発音の生起率（「フインキ」の数／総計×100）のことである。また、表7では変異の使い分け・ゆれのタイプによって、話者を①から⑥までに分類している。なお、表2上段・Wタグなしの変異は規範的な発音「フンイキ」に準じるものとみなし、「フンイキ」とWタグなしの変異を併用している話者は、表7では「①準フンイキ専用話者」として分類した。各話者分類内での並び方は、「フンイキ率」が高い順番（＝「フインキ率」が低い順番）に、「フンイキ率」が同率の場合は生年代順、次いで男女順に並べている。なお、表中の話者IDは、筆者が便宜的に付与したもので、CSJにおける話者IDとは無関係である。

表7 「雰囲気」の出現回数3回以上の話者における発音の使い分け・ゆれの実態（生起度数）

	話者ID	生年代	性別	種別	フンイキ	フンイキ系変異形	フインキ	その他変異形	総計	フンイキ率	フインキ率
① フンイキ専用話者	1	70	男	模擬	7				7	100%	0%
	2	70	男	模擬	4				4	100%	0%
	3	70	男	模擬	3				3	100%	0%
	4	70	男	模擬	3				3	100%	0%
	5	70	女	模擬	5				5	100%	0%
	6	70	女	模擬	5				5	100%	0%
	7	70	女	模擬	4				4	100%	0%
	8	70	女	模擬	4				4	100%	0%
	9	70	女	模擬	3				3	100%	0%
	10	60	男	学会	3				3	100%	0%
	11	60	女	朗読	4				4	100%	0%
	12	60	女	模擬	5				5	100%	0%
	13	60	女	模擬	4				4	100%	0%
	14	60	女	模擬	4				4	100%	0%
	15	60	女	模擬	4				4	100%	0%
	16	50	男	模擬	5				5	100%	0%
	17	50	女	模擬	4				4	100%	0%
	18	50	女	模擬	3				3	100%	0%
	19	50	女	模擬	3				3	100%	0%
	20	50	女	その他	6				6	100%	0%
	21	30	男	模擬	3				3	100%	0%
①' 準フンイキ専用話者	22	50	男	模擬	7	フ(?)ンイキ×1			8	88%	0%
	23	70	女	模擬	3	フ(?)ンイキ×1			4	75%	0%
	24	70	女	模擬	3	(?)フンイキ×1			4	75%	0%
	25	50	女	学会	1	フ(?)ンイキ×2			3	33%	0%
② フインキ・フンイキ併用話者	26	60	男	模擬	5		1		6	83%	17%
	27	10	男	学会	3		1		4	75%	25%
	28	70	女	模擬	3		2		5	60%	40%
	29	60	女	模擬	1		2		3	33%	67%
	30	60	女	模擬	1		2		3	33%	67%
	31	60	女	模擬	1		4		5	20%	80%
③ フインキ専用話者	32	70	女	模擬			5		5	0%	100%
	33	70	女	模擬			4		4	0%	100%
	34	70	女	模擬			3		3	0%	100%
	35	60	男	模擬			5		5	0%	100%
	36	50	女	模擬			3		3	0%	100%
④ フンイキ・フイキ併用話者	37	50	女	模擬	4			フイキ×1	5	80%	0%
	38	70	女	対話	2			フイキ×1	3	67%	0%
⑤ フイキ専用話者	39	40	男	模擬				フイキ×4	4	0%	0%
⑥ その他	41	60	男	模擬	1	フ(?)ンイキ×1		フイキ×1	3	33%	0%
	40	50	女	模擬	1		1	フイーキ×1	3	33%	33%
	42	60	女	模擬	1	フ(?)ンイキ×1 フ(?)ンイキ×2		フイーキ×1	5	20%	0%

表7をみると、「雰囲気」という単語を3回以上使用した話者は、おおよそ、①フンイキ専用話者（21人）、②フンイキ・フインキ併用話者（6人）、③フインキ専用話者（5人）、と三つのタイプに大別できることが分かる。①フンイキ専用話者は、「雰囲気」という単語を3回以上使用した話者・42人の半数を占める。さらに、①'準フンイキ専用話者（4人）を加えると、「雰囲気」という単語を規範的な「フンイキ」という発音、あるいは「フンイキ」系統の変異のみを使用する話者は、日本語母語話者の中に相当数存在していると考えられる。これは、VI. 表3での分析とも関連するが、「フンイキ」という規範的な発音が、少なくとも「標準語」においては優勢である、と考えることができる。

それでは、①・②・③に属する話者の特徴とはいったいどのようなものであろうか。本稿では、石井（2002）の手法に倣い、特化係数（上田1981・1982などを参照）を使用して分析することにした。「特化係数とは、全体についての構成比を基準（期待値）にして」、各カテゴリに

についての「構成比を、その何倍にあたるかという形で、表現したものである」(石井2002)。

表7で①・②・③に分類された話者について、話者の性・生年代、そして発話のタイプ別に特化係数を求めたものが、以下の表8-1から表10-3である。特化係数を求める際の「全体の構成比」は、話者の生年代・性については、V. で述べた異なり人数・193人をもとに算出している。また、同一話者が異なる発話タイプに跨っている場合もあるので、発話のタイプにおける特化係数の「全体の構成比」は、発話タイプ別に異なり人数を求め、それを合計したものをもとに算出した。

表8-1 フンイキ専用話者・話者性別の特化係数

	特化 係数	フンイキ専用話者 の人数・構成比	全体の人 数・構成比
男	0.81	7(33.3%)	79(40.9%)
女	1.13	14(66.7%)	114(59.1%)

表8-2 フンイキ・フィンキ併用話者・話者性別の特化係数

	特化 係数	併用話者の 人数・構成比	全体の人 数・構成比
男	0.81	2(33.3%)	79(40.9%)
女	1.13	4(66.7%)	114(59.1%)

表8-3 フィンキ専用話者・話者性別の特化係数

	特化 係数	フィンキ専用話者 の人数・構成比	全体の人 数・構成比
男	0.49	1(20.0%)	79(40.9%)
女	1.35	4(80.0%)	114(59.1%)

表9-1 フンイキ専用話者・話者の生年代別の特化係数

	特化 係数	フンイキ専用話者 の人数・構成比	全体の人 数・構成比
10年代	0	0(0%)	1(0.5%)
20年代	0	0(0%)	0(0%)
30年代	0.66	1(4.8%)	14(7.3%)
40年代	0	0(0%)	23(11.9%)
50年代	1.15	5(23.8%)	40(20.7%)
60年代	1.28	6(28.6%)	43(22.3%)
70年代	1.18	9(42.9%)	70(36.3%)
80年代	0	0(0%)	2(1.0%)

表9-2 フンイキ・フィンキ併用話者・話者の生年代別の特化係数

	特化 係数	併用話者の 人数・構成比	全体の人 数・構成比
10年代	33.52	1(16.7%)	1(0.5%)
20年代	0	0(0%)	0(0%)
30年代	0	0(0%)	14(7.3%)
40年代	0	0(0%)	23(11.9%)
50年代	0	0(0%)	40(20.7%)
60年代	2.99	4(66.7%)	43(22.3%)
70年代	0.46	1(16.7%)	70(36.3%)
80年代	0	0(0%)	2(1.0%)

表 9 - 3 ファインキ専用話者・話者の生年代別の特化係数

	特化 係数	ファインキ専用話者 の人数・構成比	全体の人 数・構成比
10 年代	0	0(0%)	1(0.5%)
20 年代	0	0(0%)	0(0%)
30 年代	0	0(0%)	14(7.3%)
40 年代	0	0(0%)	23(11.9%)
50 年代	0.97	1(20.0%)	40(20.7%)
60 年代	0.90	1(20.0%)	43(22.3%)
70 年代	1.65	3(60.0%)	70(36.3%)
80 年代	0	0(0%)	2(1.0%)

表10- 1 ファインキ専用話者・発話のタイプ別の特化係数

	特化 係数	ファインキ専用話者 の人数・構成比	全体の人 数・構成比
学会	0.49	1(4.8%)	19(9.7%)
模擬	1.03	18(85.7%)	163(83.2%)
そのほか	2.4	1(4.8%)	4(2.0%)
対話	0	0(0%)	4(2.0%)
朗読	1.54	1(4.8%)	6(3.1%)

表10- 2 ファインキ・ファインキ併用話者・発話のタイプ別の特化係数

	特化 係数	併用話者の 人数・構成比	全体の人 数・構成比
学会	1.72	1(16.7%)	19(9.7%)
模擬	1.00	5(83.3%)	163(83.2%)
そのほか	0	0(0%)	4(2.0%)
対話	0	0(0%)	4(2.0%)
朗読	0	0(0%)	6(3.1%)

表10- 3 ファインキ専用話者・発話のタイプ別の特化係数

	特化 係数	ファインキ専用話者 の人数・構成比	全体の人 数・構成比
学会	0	0(0%)	19(9.7%)
模擬	1.20	5(100%)	163(83.2%)
そのほか	0	0(0%)	4(2.0%)
対話	0	0(0%)	4(2.0%)
朗読	0	0(0%)	6(3.1%)

石井（2002）には、「特化係数は、使用度数があまりに小さい単語では過大に出てしまう」という注意がある。今回の分析でも、各カテゴリに分類される話者数が少ないので、表 8 - 1 から表10- 3 の特化係数を「額面どおり」に受け止めて解釈してしまうのは、安易かもしれない。それでもなお、以下のような傾向を読み取ることができるのではないだろうか。

表 8 - 1 から表 8 - 3 は、話者の性別に着目した特化係数である。これらの表を見ると、全体的に男性の特化係数が低くなっているが、特に「ファインキ専用話者」の男性の特化係数が低いことが目につく（表 8 - 3 ・0.49）。つまり、「ファインキ専用話者」は男性が少ないことを意味している。

表9-1から表9-3は、話者の生年代に着目した特化係数である。これらの表で特徴的なのは、1960年代生まれの「フンイキ・フィンキ併用話者」の特化係数が特に高いこと（表9-2・2.99）、ならびに1970年代生まれの「フィンキ専用話者」の特化係数が高いこと（表9-3・1.65）であろう。これは、1960年代生まれでは2つの代表的な変異を併用している人が多いが、1970年代生まれになると「フィンキ」という非規範的な発音に収斂される人が多い、と解釈できる。

表10-1から表10-3は、発話のタイプに着目した特化係数である。「フンイキ専用話者」の学会講演で特化係数が低く（表10-1・0.49）、「フンイキ・フィンキ併用話者」の学会講演で特化係数が高くなっている（表10-2・1.72）。これは、学会講演では「フンイキ専用話者」が少なく、「フンイキ・フィンキ併用話者」が多いことを意味している。ただし、表10-1から表10-3は、各カテゴリが細分化され、各カテゴリに分類される数も少ないため、特化係数に信頼性が欠ける側面があることは否定できない。

「雰囲気」という単語の発音の変異・変種の使い分け・ゆれの実態の分析結果として表8-1から表10-3の解釈をまとめるならば、主に以下の二点に集約できるであろう。まず、規範的な「フンイキ」という発音と非規範的な「フィンキ」という発音を併用する人は1960年代生まれに多い傾向にあり、さらに1970年代生まれになると「フィンキ」という非規範的な発音だけを使用する人が多い傾向にある。また、「フィンキ」という発音だけを使用する男性は、女性に比較して少ない傾向にある。

以上、本節では、「雰囲気」という単語の発音の変異・変種の使い分け・ゆれの実態という、いわばマクロの視点から分析を試みた。その結果、上述したような結論を得ることができた。このことから、ミクロな視点からの分析でも、「とりあえずCSJ検索」は、一定の有効性があると言えよう。ただし、ミクロな視点からの分析をするには用例数が少なく（これは、「雰囲気」という単語の発音、と言語変項をひとつに限定したためであるが）、マクロな視点からの分析と比較して、相対的に信頼性が落ちる側面も否定できない。

VIII. おわりに

本稿では、「雰囲気」という単語の発音の実態を例にあげ、「とりあえずCSJ検索」の妥当性を検証した。その結果、先行研究での指摘とは異なり「フンイキ」という規範的な発音が主流であること、非規範的な発音として代表的なものは「フィンキ」であること、「フンイキ」という規範的な発音と「フィンキ」という非規範的な発音が「雰囲気」という単語の発音の90%以上を占める一方で非規範的な発音の種類も多い、ということが明らかになった。また、個人差を捨象したマクロな視点から、ならびに個人差に着目したミクロな視点から、「雰囲気」という単語の発音の多様性に関与する言語外的条件の分析を行なった。その結果、マクロな視点からの分析では若年層（1960年代・1970年代・1980年代生まれ）に「フィンキ」という非規範的な発音が比較的多く観察されるという傾向を見出せた。またミクロな視点からの分析では、規範的な「フンイキ」という発音と非規範的な「フィンキ」という発音を併用する人は1960年代生まれに多い傾向にあり、さらに1970年代生まれになると「フィンキ」という非規範的な発音だけを使用する人が多い傾向にあること、「フィンキ」という発音だけを使用する男性は、女性に比較して少ない傾向にあること、を見出せた。マクロの視点からの分析、ミクロの視点

からの分析、双方の結果を総合して考えるならば、概ね、非規範的な「フィンキ」という発音は若年層に多い、という結論を導き出すことができよう。このように「とりあえずCSJ検索」で一定の結論を導き出すことができたということは、CSJはⅡ. で指摘したような問題点（主にデータ量の問題）があるにせよ、「とりあえずCSJ検索」には一定の有効性を認めることができるのではないだろうか。また同時に、Ⅴ. でも指摘したように、CSJは「標準語」のコーパスである以上、非規範的な発音については、日常会話よりも生起頻度が低くなってしまう蓋然性があることにも注意したい。

なお、非規範的な「フィンキ」という発音は若年層に多いという結論は、先行文献での指摘とも矛盾はしない。すなわち、少なくとも「雰囲気」という単語の発音の実態については、研究者の日々の観察の結果や母語話者の内省の結果は正鵠を射ていたことを、今回の「とりあえずCSJ検索」の結果は、実態のデータを伴った上で、実証したことになる。

今後の課題としては、ではどうして若年層に「フィンキ」という発音が多いのか、「フィンキ」という発音は言語変化の一端なのか、といった疑問点を解明することであろう。ただ、これらの点は、「とりあえずCSJ検索」の可能性と限界の検証という本稿の目的とは若干逸れる問題なので、本稿では言及しないことにする。

本稿の結論は、やはり用例数が不足している側面は否定できないものの、「とりあえずCSJ検索」は一定の有効性が認められる、ということになるが、最後に、「とりあえず検索」に限らず、CSJを使用した分析の際、(大きな)問題が存在していることを指摘しておきたい。本稿では「雰囲気」という単語の発音を表記する際、あえてカタカナ表記を採用してきた。それは、今回の分析に使用したCSJの転記テキストの発音形がカタカナで表記されているからである。しかしながら、実は、ここには大きな問題が潜んでいる。というのも、音声という物理的には連続的な情報が、CSJの転記テキスト作成の段階において、カタカナという離散的な情報に置き換えられているからである。今回はCSJの表記を全面的に信用して分析を行なったが、果たしてそれでいいのか、議論は分かれるであろう。特に「雰囲気」という単語の変異の一部には、物理的な情報としては前後の音素と連続している鼻母音を含んでおり、カタカナで離散的な情報として提示されたものをどの程度信頼できるものなのか、検討する必要があると考えている。

以上のように、筆者は、CSJの転記テキスト・ラベリングの妥当性については、様々な角度から検討すべき問題だと考えている（ただ、この問題は、音声学という学問のあり方の根本にもつながる問題だと考えている）。しかし、今回、このことについて論じる紙幅は尽きてしまった。本稿では、CSJ製作者側の見解として、前川（2005）を引用するにとどめておきたいと思う。

データベースの限界にも触れておこう。（中略）

次に、これも当然だが、データベースから直接抽出できる情報はほとんどすべて何等かの意味でデータベース構築作業者の解釈を経たデータである。7.3節で紹介した音声変異のデータ（発音形）には、転記作業者の耳というフィルタがかかっている。転記の統一には種々の工夫をこらしているものの、そこにバイアスが存在していることは否定できない。

付記

CSJの製作に関わられたすべてのみなさまに感謝と敬意を表します。

本文中の決定木の作成には、M道具工房製作のexcelアドインを使用させていただきました。

本稿は、2006年10月1日に開催された2006（平成18）年度第20回日本音声学会全国大会で発表した内容に修正を加えたものです。

本研究の一部は、2005年度第1回博報「ことばと文化・教育」助成（「標準語」の多様性）・代表：岡田祥平）を受けています。

参考文献

- 浅野 晃 (2001) 「ふいんき」『座観雑感 (2001年2月5日付)』
<http://zakkan.racco.mikeneko.jp/010205fuinki.html> (2006年7月20日閲覧)
- 浅野 晃 (2004) 「続・「ふいんき」」『座観雑感 (2004年12月1日付)』
<http://zakkan.racco.mikeneko.jp/041201fuinki2.html> (2006年7月20日閲覧)
- 石井正彦 (2002) 「テレビの単語使用—番組と話者からみた多様性—」佐藤喜代治編『国語論究 第9集 現代の位相研究』明治書院
- 伊藤雅光 (2005) 「計量言語学とコーパス言語学」『計量国語学』25 (2) pp.89-97
- 井上史雄編 (1997) 『社会方言学資料図集—全国中学校言語使用調査 (1993-1996)—』
- 上田尚一 (1981) 『統計データの見方・使い方』朝倉書店
- 上田尚一 (1982) 『データ解析の方法—質的データの解析—』朝倉書店
- 荻野綱男 (2004a) 「各種検索エンジンの実態と特徴」『日本語学』23 (2) pp.18-33
- 荻野綱男 (2004b) 「WWWによる方言語形の全国分布調査」『日本方言研究会第79回研究発表会発表原稿集』pp.61-68
- 荻野綱男 (2006a) 「検索エンジンGoogleの使い方とWWWコーパスによる日本語研究」城生佰太郎博士還暦記念論文集編集委員会編『実験音声学と一般言語学—城生佰太郎博士還暦記念論文集—』東京堂出版
- 荻野綱男 (2006b) 「wwwによる単語の文体差の研究」『日本語学会2006年度秋季大会号稿集』pp.139-146
- 荻野綱男 (2006c) 「形容動詞連体形における「な／の」選択について—田野村氏の結果をWWWで調べる—」『計量国語学』25 (7) pp.309-318
- 岡島昭浩 (1996) 「目についたことば (1996年10月14日付)」
<http://www.let.osaka-u.ac.jp/~okajima/menicuita/9610.htm#14> (2006年7月20日閲覧)
- 「おしえてねドットコム！」(作成年不明)「[No102]「雰囲気」は「ふいーき」って読んじゃいません？」
<http://homepage2.nifty.com/osiete/seito102.htm> (2006年7月20日閲覧)
- 川口 良 (2006) 「母語話者の「規範のゆれ」が非母語話者の日本語能力に及ぼす影響—動詞否定丁寧形「(書き) ません」と「(書か) ないです」の選択傾向を例として—」『日本語教育』129 pp.11-20
- 窪蘭晴夫 (2006) 「若者ことばの言語構造」『言語』35 (3) pp.52-59
- 久野 眞 (2006) 「現代の「若者」が使う方言」『日本語学』25 (1) pp.6-17

- 小磯花絵・間淵洋子・西川賢哉・斎藤美紀・前川喜久雄（2004）「転記テキストの仕様 Version.1.0」『日本語話し言葉コーパス』附属電子文書
- 「ことば会議室」（2000）「雰囲気。」<http://kotobakai.seesaa.net/article/8186933.html>（2006年7月20日閲覧）
- 近藤 明（1998）「『フインキ』と『フンイキ』」小矢野哲夫編「けとば珍聞」1998年9月号
<http://homewww.osaka-gaidai.ac.jp/~koyano/ketoba9809.htm>（2006年7月20日閲覧）
- 岡田祥平（2006a）「『縮約形』再考」『阪大日本語研究』18 pp.49-78
- 岡田祥平（2006b）「音声変異の重層性—音韻論的対立の有無の観点から—」『龍谷大学国際センター研究年報』15 pp.65-78
- 岡田祥平・出野晃子（2006）「『日本語話し言葉コーパス』に見る「標準語」の多様性」『東洋語言文化論集』第4集（華南理工大学出版社・中華人民共和国）pp.151-167
- 真田信治（2000）『脱・標準語の時代』小学館文庫
- 杉村 泰（2002）「コーパス調査による文法性判断の有効性—「～てならない」を例にして—」『日本語教育』114 pp.60-69
- 田中ゆかり（2003）「ネット検索は言語の研究に有用か」『コーパス言語学』（『日本語学』22（5））pp.111-123
- 田野村忠温（2003）「コーパスによる文法の研究」『コーパス言語学』（『日本語学』22（5））pp.174-186
- 鳥飼浩二（2004）「ふんいき／ふいんき」北原保雄編『問題な日本語—どこがおかしい？何がおかしい？』大修館書店
- 「日本語調査結果（99）」（2004）「『雰囲気』を『ふいんき』と読みますか？自分はどうですか？ 周囲にそう読む人がいますか？」
<http://pro.tok2.com/~nhg/research-2/research1-99.htm>（2006年7月20日閲覧）
- 服部 匡（2004）「WWW検索と日本語研究への応用」『日本語学』23（2）pp.6-16
- 日比谷潤子（2003）「変異理論のこれから」『英語青年』148（11）pp.44-45
- 前川喜久雄（2004a）「『日本語話し言葉コーパス』の概要」『日本語科学』15 pp.111-133
- 前川喜久雄（2004b）「ニホンかニッポンか」『文化庁月報』431 p.30
- 前川喜久雄（2005a）「NHKの発音」『情報通信ジャーナル』23（5）p.40
- 前川喜久雄（2005b）「自発発話とデータベース」『日本音響学会誌』61（9）pp.544-549
- 前川喜久雄（2006）「『日本語話し言葉コーパス』を利用した語形変異研究」『国際シンポジウム比較語彙研究X』pp.151-158
- 松井くにお（2004）「検索ロボット技術を活かしたWWW検索技術」『日本語学』23（2）pp.34-42
- 道浦俊彦（2000）「ことばの話197「フインキ」」「平成ことば事情（2000年11月12日付）」
<http://www.ytv.co.jp/announce/kotoba/back/0101-0200/0196.html#2>（2006年7月20日閲覧）

後註

- (1) ネット文書を利用した言語研究で、筆者が目についた最新のものとしては、荻野（2006c）がある。なお、荻野氏はネット文書を方言研究（荻野2004b）や文体差研究（荻野2006b）など、言語研究の様々な分野で利用しようとする試みを実践している。
- (2) 田中（2003）でも、ネット文書を利用した言語研究にはふさわしくないものとして、「ネット文書に含まれていない情報・検索不能なもの」と指摘し、その具体例としては「音声・アクセント・韻律の実態」をあげている。
- (3) 荻野（2004a）や松井（2004）では、ネット文書の言語量を新聞一千年分と推定している。
- (4) ただし、前川（2004a）によると、CSJ製作段階では、学会講演と模擬講演でスタイル差が生じるように工夫はされているようである。
- (5) 前川（2004a）によると、CSJの研究期間は1999年から2003年度ということから、音声収録（CSJのデータ収集は、研究期間の初期（西暦2000年前後）に行なわれたと推測できる。
- (6) 一方で、註1でも触れたように、方言研究にネット文書を利用することを模索する動きもある（荻野2004b）。
- (7) 田中（2003）は、ネット文書での検索には利用者のコンピュータリテラシーでは対処できない「ブラックボックス」が存在している、と指摘している。荻野（2004a・2006a）は、その「ブラックボックスの中身」を解明しようとした試みといえよう。
- (8) 本節の分析のうち、「異なり変異形数」・「カバー率」については、前川（2006）から着想を得た。
- (9) これはあくまで筆者の感覚に過ぎないが、日本語についての社会での議論は、文法や語彙に関することが中心であり、音声についてはそれほど議論がなされていないように感じる。中には音声についての議論もあるが、アクセント（平板化現象など）やイントネーション（いわゆる「尻上がりイントネーション」など）といった超分節的要素が主であり、分節的要素についての議論は少ないように思う（筆者は、ガ行鼻濁音をめぐる話題は、数少ない分節的要素についての議論のひとつだととらえている）。その理由のひとつとしては、日本語母語話者は自分たちが使用している音声言語を「日本語の音韻体系の耳」で聞いているため、分節音をめぐる多様性を看過していることも関係しているのではないかと考えている（この件については、岡田2006bでの議論も参照）。
- (10) なお、この調査は世代別の分析結果もまとめられている。調査結果を付表として提示するが、これを見る限り、あくまで意識調査であるものの（しかも自分の属性を偽ることが可能なインターネット上の調査なので、信頼性は落ちるが）、若年層以外は規範的な発音である「フンイキ」が多数派である。若年層に限っても、自分で読むか否かの問いに対しては、「自分は読まない」、すなわち規範的な発音である「フンイキ」が多数派になる。

付表 「日本語調査結果 (99)」 (2004) より

Q:あなたは「雰囲気」を「フインキ」と読みますか？		Q:あなたの周囲に「雰囲気」を「フインキ」と読む人はいますか？	
TOTAL:111		TOTAL:111	
自分は読まない	自分は読む	周囲にいない	周囲にいる
93 83.8%	18 16.2%	69 62.2%	42 37.8%
10代/20代:25		10代/20代:25	
自分は読まない	自分は読む	周囲にいない	周囲にいる
16 64.0%	9 36.0%	10 40.0%	15 60.0%
30代/40代:54		30代/40代:54	
自分は読まない	自分は読む	周囲にいない	周囲にいる
50 92.6%	4 7.4%	36 66.7%	18 33.3%
50代/60代:21		50代/60代:21	
自分は読まない	自分は読む	周囲にいない	周囲にいる
19 90.5%	2 9.5%	18 85.7%	3 14.3%

* 単位のない数字は人数, %は各解答の構成比

- (11) 「記号 ‘<’ はその左側よりも右側の方で自発性が高いことを示し、‘<<’ はその差が顕著であることを示す」 (前川2004a)。
- (12) なお、今回の分析結果の序列と前川 (2004a) のいう発話の自発性の序列は、対話と模擬講演で逆転があるが、これは今回の分析は模擬講演に用例が偏っていたことが影響したとも考えられる。
- (13) 決定木分析は大規模データの存在を前提としたデータマイニングの手法である。そのため、今回のように用例数が数百程度しか存在しない場合に適用することについての是非については、改めて検討しなければいけない問題だと考えている。
- (14) CSJでは、話者の生年代が5年刻みで公開されている。今回のデータに出てきた1980年代生まれ2人は、正確には「1980年から1984年生まれ」である。
- (15) このようにCSJでは話者の生年代は公開されていても、音声収録当時の話者の年齢が判然としない (推測するしかない)。この点は、CSJを使用して言語使用実体の世代差の分析をするときの弱点かもしれない。
- (16) 「使い分け」・「ゆれ」には、さらに、複数の変異が全くアトランダムに併用されるという自由変異と、何らかの言語内的要因が満たされた場合にのみある特定の変異が使用されるという条件変異という二つの場合が存在していると考えられる。