

## 関西弁コーパスの紹介

# An Introduction to the Corpus of Kansai Spoken Japanese

ケビン・ヘフファナン

Kevin Heffernan

### 1. はじめに

本報告書の目的は、関西弁コーパスを紹介することである。この報告書ではまず、このコーパスを作成するに至った動機を述べ、その作成にあたって使用した方法論を説明する。そしてこのコーパスを用いた調査の例を提示していく。

### 2. 動機

このコーパスは、日本語における現在進行中の言語変化を研究するという動機から作成されている。新しい言葉、文法、発音などは、標準語として正式な形になる前に口語や地方の方言の特徴として現れることが多い(Chambers 2003)。従って、この研究の目標は日本語に起きている最新の日本語の特徴を発見するために、形式にとらわれていない日本語の話し言葉のサンプルを適量、記録することにある。

言語的な革命は都市の中心部から外に向かって流れていくという傾向が認められている(Chambers and Trudgill 1998)。したがって、都会に該当する大阪府をこのコーパスの焦点に当てることに決めた。大阪市とその周りの都市を含める大阪府の人口は約9億人であり(大阪府2012)、

東京に凌ぐほどである。大阪は日本で2番目に巨大な大都市であり、関西圏では経済、政治、文化の中心である。その上、大阪弁はよく方言研究の対象とされるだけではなく(平山 1997、陣内・友定 2005)、他の言語学分野の研究対象にもされている(例Fujimoto et al. 2002)。大阪弁は牧村(1984)、平山(1997)、山下(2004)、札埜(2006)など、多くの辞書や文法書及び資料においてかなり文章化されている。大阪弁は関西地方の方言の元になり、多くの関西弁の特徴は大阪周辺の地域で形成され、関西圏に広まっていったのである。

関西弁は日本国内で若い世代を中心に良いイメージを持たれている(陣内2005、2007)。日本の首都、東京周辺で使用されている標準語の圧力があるにも関わらず、それを押しのけるかのように関西弁は繁栄し続けている。つまり、このコーパスは新たな日本語の言語形式を発見することのできる理想的なツールある。

このプロジェクトの目的のひとつとして、他の研究者も使用することが出来るコーパスの構築がある。現在公表されている使用可能なコーパスのひとつに「日本語話し言葉コーパス」(前川他2000)というものがある。「日本語話し言葉コーパス」の主な目的は、音声認識の研究にある。それぞれのコーパスが目的とするものは大きく異な

るため、日本語話し言葉コーパスと本稿で紹介する関西弁コーパスについても、いくつかの重要な点で異なっている。日本語話し言葉コーパスのスピーチ標本には二つの種類がある。学会などにおける研究者の発表を録音したもの、そして、報酬を支払われた一般人によるシミュレーションの二つである。両者においてはともに、標準語という日本語の正式な形が使用されている。そのような正式なスピーチをする際に、話し手は方言を使用せず、当該言語の最も標準な言語形式を用いる(Labov 1972)。それに対して関西弁コーパスは、より砕けた表現をとりあげている。関西弁コーパス作成において使用した全てのインタビューは、お互いのことをよく知っている二者による自由な会話である。言い換えるならば、関西弁コーパスは、より砕けた関西弁での会話の標本から構成されているのである。

### 3. 研究手法

#### 3.1 データ収集

ここで言うデータ収集とは、社会言語学では一般的なインタビュー調査という研究手法を行うことであり、Tagliamonte (2006) や、小林・篠崎 (2007) の研究手法に従っている。インタビューアーは関西学院大学総合政策学部の大学生で、報酬を支払われた研究助手又はゼミ生である。インタビューの対象は、インタビューアーの家族の一員や知り合いであり、インタビューアーと親密な関係にある話者である。これによって、正式な話し言葉ではなく、くだけた表現が会話の中で使用されることが想定された。

それぞれのインタビューは約1時間で、内容が前もって決められてはいなく、自由な会話から構成されている。しかし、インタビューアーには、2つの条件を事前に伝えられた。1つ目は、表現の硬さについてで、2つ目は、話し手がたくさん話

すように促す工夫についてである。これらについて順に追って説明していく。

街角で取材をする際に、ニュース番組の記者が使うような習慣的なインタビュー様式をとらないということである。このような話し方は、非常に形式的であり、標準的な文法で使用されている。インタビューされる側、つまりインタビューにおけるメインの話し手は、インタビューアーの話し方に順応する傾向にある。つまり、インタビューアーが標準語で質問するとき、話し手は標準語で返事をしてしまう傾向が高いのである。従って、1つ目の条件はインタビューアーが常に日本語の話し言葉を使用することである。

2つ目の条件として、インタビューアーは話し手からできるだけ沢山の話を聞き出そうとするように努めることである。そのために、インタビューアーは相手の経験を聞いたり、フォローアップの質問をしたりするように指導される。なぜならば、社会言語学を専門とする研究者は話し手の話に集中し、インタビューアー自身の話にそれほど着目しないという傾向があるからである。

現在までに終えたインタビューのテーマは、地震の経験や仕事場での出来事、経済、あるいは恋愛、海外生活や学校生活などについて多岐に及んだ。またインタビュー内容は、デジタルボイスレコーダーを使用して録音された。音の質に関しては、静かな場所を選んでインタビューするようインタビューアーに依頼したが、それ以外には特に留意点はないと考え、提示しなかった。

#### 3.2 話し手の分布

この調査の最終目標は、約200人の話し手の話を記録することである。この200人のうち、約120人は大阪近辺で生まれ育った生粋の大阪弁話者であると言える。これら、120人の話し手は表1で見られるような性別と年齢層に分けることができる。

	中学生	高校生	大学生	25-30 歳	30-39 歳	40-49 歳	50-59 歳	60-69 歳	70-79 歳	80歳 以上
女性	6	6	6	6	6	6	6	6	6	6
男性	6	6	6	6	6	6	6	6	6	6

表1：大阪近辺で育った120人を対象にした分布

残りの80人の話し手は以下のような所属となる。

- 成人してから大阪に引っ越してきた者(30名)
- 人口2万3千人ほどの小さな町で、大阪から車で3時間ほどかかる兵庫県多可郡多可町というところに居住する者(30名)
- 日本語を第二言語として学んだ海外の留学生(20名)

上記のように、インタビューにおいて多様な話し手が加わることで、大阪弁調査の幅が広がり、それは大阪弁コーパス作成の上で重要な要素である。また、本インタビュー調査の主要な目的ではないが、成人になってから大阪に引っ越してきた話者と日本語を母国語としない話者にとって、本インタビュー調査は関西弁習得の機会という意味も含まれていた。また一方で、本調査は、新しい大阪弁が多可町の人々に伝播し、その周りの地域にも普及するという機会も含めている。

### 3.3 調査対象の個人情報

インタビューの終了後、話し手に個人情報の回答への協力を求めた。尋ねた個人情報は以下のようなものである。

- 性別
- 年齢
- 職業
- 出生地
- 出身地
- 現在地
- 最終学歴
- 英語力
- 両親の出身地
- 両親の最終学歴

### ●両親の職業

### 3.4 「見かけの時間」という研究手法

幅広い年代に渡った調査の際には「見かけの時間」(英語apparent-time construct)という手法を使用することが可能である。しかし、この研究手法は最近発見されたものである。1960年代に入ってから、William Labovが年代別のスピーチと言語変化の増量の関係を発見するまで、この手法は知られていなかった。それまでの言語学者たちは、言語変化というものはその変化が完了した後にしか見られないと信じていた。その際に用いられていた手法は、現代の文章と過去の古い歴史的な文書を比較した手法であった。しかし、実際には歴史的な文書の使用は、その時々起こっていた言語変化の経過観測を妨げていたのである。「見かけの時間」という手法を使用すると、観測された時代別の言語データを図として表すことができる。すると現在進行中の言語変化とその経過が明確に反映されるのである。この技術で大切なことは、人が成長するにつれ、言語システムは完成に近づくということである。つまり、思春期を越えると、発音および文法はあまり変化しない傾向にある。一方で、新しい語彙項目は生涯に渡って加えられる。さらに言うならば、人は、生涯に渡って自分が子どもの時に獲得した文法およびアクセントで話す傾向があるのだ。従って、青年時期は言語獲得において重大な時期であると考えられる。この時期が終わると、個人の文法システムは化石のように変化しにくくなる。この言語システムの化石化はコミュニティ内での言語変化を観測する上でも役に立つ。例えば、70代の高齢者の文法およびアクセントは、その人が子どもだった60年以上前に、そのコミュニティで話されていた言語と捉えることができる。つまり、高齢者の話し方によって過去を顧みることができるのだ。研究者は、高齢者の話し方を若者のものと比較する。

また、観察された系統的な違いは進行中の言語変化であると推測される。このように、実時間自体は経過していないが、研究者は時間が経つにつれて言語変化が生じていることに気づくことができる。時間が実際に経ったような錯覚から「見かけの時間」という名前が付けられた。

### 3.5 インタビュー内容の書き起こし

録音されたそれぞれのインタビュー内容は、生粋の大阪弁話者である研究助手によって書き起こされた。一回目の書き起こしが終わると、別の研究助手が録音されたインタビューを聞き、書き起こしの間違い箇所を訂正した。この作業を合計3回、それぞれ違う助手によって行われた。この経過については、インタビュー内容の正確な書写を確実にものにするために必要なことであると経験から言える。

インタビュー協力者の個人情報保護の観点から、インタビュー中のすべての個人名前は仮名にする。仮名が使えない場合、例えばインタビュー中にインタビューされた側が自分の名前に使われている漢字の意味を説明するなどの箇所があれば、その部分は削除された。

インタビューの書き起こしの際には適宜、平仮名、カタカナ、漢字、英語で書写された。書き起こしを行った研究助手には、方言を忠実に書写するように説明した。標準語ではない言葉、例えば「そやんな」、「なんてゆうか」や「よう食べる」などの話し言葉、また、省かれた音、例えば「そんな」、「食べてる」や「うん」の相槌などについても忠実に書き起こすように指導した。

## 4. 調査事例

ここでは言語接触とディスコースマーカという2つの調査事例について述べる。下で述べる調

査事例は仮説と結論のある完全な調査事例ではないが、これはむしろコーパス構築に向けた予備調査と言える。2つの調査事例共に、同じ調査対象から得られたデータを基に分析を行っている。28人の話し手から得られたデータは、本稿執筆時までに完全に書き起こされたものである。これは表2で見られるような性別と年齢層に分けられる。この段階で分析対象とする話し手はまだ少数で、「若者」、「中高年」、「高齢者」の3つの年齢層だけに分けられる。(中学生や高校生のグループからのデータについてはまだ書き起こしが完了していない。)すべての話し手は、大阪近郊の育ちである。

	若者	中高年	高齢者
女性	5	6	4
男性	5	4	4

表2：調査事例の28人の分布

### 4.1 関西弁コーパスを用いた言語接触の調査事例

コーパスは、コード切り替えや外来語の語法のような、言語接触の現象を研究するための理想的な道具である。第一の調査事例では、話し手の英語の能力と外来語の使用頻数との相関関係性を調査する。多くの外来語の表現は、その外来語表現と相対的に同じ意味を持つ日本語表現が併存し、使用されている。例えば、「オーナー」(英語 owner)は「店の支配人」としても表現される。この研究では、英語能力が比較的高い話し手が、純粋な日本語表現以上にカタカナ言葉を好みがちになるという仮説を検証する。

話し手には、それぞれ自分の英語能力を自己申告してもらった。この自己申告を基に、話し手は表3で見られるように、3つの層に分けられた。

自己申告の英語能力	人数
苦手(全く又はほとんど話せない)	11
平均(文法や発音をよく間違うが話せる)	7
得意(母国語なみに又は流暢に話せる)	10

表3：話し手の英語能力分布

それぞれの話し手のカタカナ外来語使用の割合は、以下の手順に沿って計算された。まず、インタビュー中のカタカナ外来語の使用数は、次の点を考慮しながら数えている。「ローソン」や「アメリカ」のように表現の仕方がカタカナのみの固有名詞は、外来語とせず、数えていない。そして、「キュウリ」や「イケメン」のような、カタカナに書き換えられても一般的に外来語として考えられないものも数に含めていない。しかし「サービス業」のようなカタカナと漢字の混合表現は数に含めた。

カタカナ外来語使用の割合は、以下のような計算式で算出した。

この値は、1000文字ごとに使われるカタカナ外来語の文字数である。計算の結果、話し手の値は9.2と73.7の間に分布することがわかった。英語能力によるカタカナ外来語使用の平均値分布は、図1のとおりである。

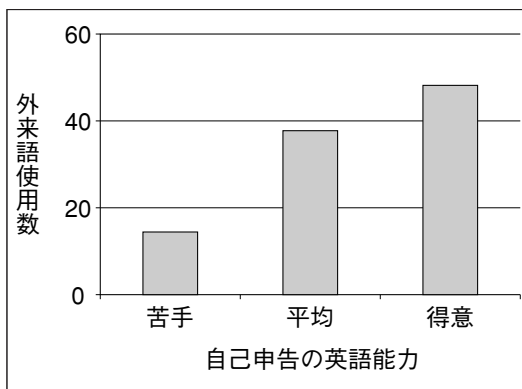


図1：英語能力別カタカナ外来語使用の平均値

これらの結果は、英語能力と外来語使用の間での明白な傾向を表している。つまり、英語能力の高い者が日本語を話す際、より多くの外来語を使用することを表している。この結果から、第二言語に接することは、第一言語の使用に影響を及ぼす可能性が考えられる。関西弁コーパスは、言語接触現象の調査を行う上で1つの理想的なツールと言えるであろう。

#### 4.2 関西弁コーパスを用いたディスコースマーカーの「なんか」の調査事例

コーパスは、言語変化の研究を行う上でも理想的なツールである。そのような言語変化の1つの例が、「なんか」という言葉である。日本国語大辞典(日本大辞典刊行会1975)によると、下の例文でみられるように「なんか」という言葉は古来、肯定の文脈で「指定されていない事物や事態」または「色々、あれやこれや」を意味する名詞として使用された。

[竹取物語] 今さへ何かと言ふべからず

19世紀の文献では、「なにか」の直前に来る語彙に類する名詞として使用されている例が見られる。

[滑稽本] お三絃(しゃむせん)や何角(なにか)も

それから1世紀後、「なんか」は、名詞に接してその名詞に対する消極的な意味を表現するようになり始めた。

川端康成[雪国] 私なんかまだ子供ですけど

近年の調査(内田2001、福原2009、李2006)報告によると、最近では「なんか」という言葉は節の最初に置かれて様々な意味を表したり、実用的な機

能を果たしたりする役目として使われたりしている。例えば、何かを言おうとするときに、どのように言おうかと考え巡らせている場合に用いられ、話す人と語彙の心理的な距離を作るために使われたりしている。

これらの例は、「なんか」という言葉が、名詞及び名詞類に接する類、そして節の前に来る類という3つの統語法に分かれることを示している。ここでは、関西弁コーパスを用いた「なんか」の調査事例を報告する。上述の28人から集められた会話について、3種類の「なんか」の分布状況を見るために、次のような測定を実施した。1人の話し手につき2000字のサンプルを、書き起こされたトランスクリプトの中間あたりから抜粋し、それぞれの発話を、節にするために品詞分析をした。それらの節の数を確認し、また、一つ一つの節に「なんか」という言葉があるか否かについても確認を行った。その結果、計4196の節数が認められた。これと同じ方法で名詞の数を数えると4213点見つかった。節の分析と同様に、それぞれの名詞の中に「なんか」という言葉があるか否かの確認も行った。

次に、それぞれの話し手のトランスクリプトについて、節や名詞の中の「なんか」の修飾する使用頻度を計算した。その結果、平均で4.2%の節が修飾されていた。節を修飾している例を(i)に挙げる。実際の使用頻度は、最も低いもので0%から高いもので19.3%と話し手の間で使用頻度に幅が見られた。この結果とは対照的に、修飾された名詞の割合はたったの0.8%であった。(ii)が名詞を修飾している「なんか」の例である。

- (i) なんかもちょっとしんどくなった部分があったけど。(CKSJ/005/F/3)<sup>1</sup>

- (ii) ほんで筆筒なんかは北向いてへんでしょ。  
(CKSJ/008/F/9)

表2は、節を修飾している「なんか」の使用割合について、年齢別に見たものである。表3は名詞を修飾した場合のものである。

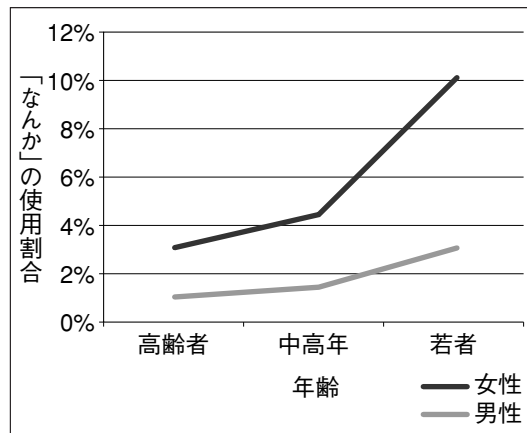


表2：年齢別にみた節を修飾しているなんかの使用頻度

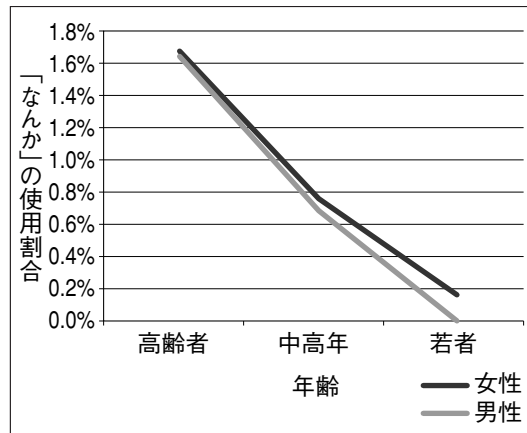


表3：年齢別にみた名詞を修飾しているなんかの使用頻度

これら2つの表は、際立って違うパターンを描いている。表2が示すように、節を修飾する「なんか」の使用頻度は世代が下るごとに上昇していることから、日本語の進行変化を示していると言え

1 括弧内の情報はコーパス名、話し手番号、性別、年齢グループである。

る。さらに言えば、女性を筆頭にこの変化が表れている。表3でも年齢層によって顕著な違いが表れているが、名詞を修飾している「なんか」の使用頻度は、話し手が若くなればなるほどその数が減少している。最も若い年齢層ではほぼゼロに等しい。このことから、名詞の後に接する「なんか」の用法は日本語の中で廃語になりつつあるのではないかと予想できる。これらの結果から、関西弁コーパスが日本語の変化の調査のツールとして実用可能であることを証明していると考ええる。

## 5. 結論

本報告書では、関西弁コーパス作成に向けた基本的な調査材料の決定、及び、関西弁コーパスを用いた研究の方法論を紹介した。その例として、外来語の語法の研究と「なんか」における現在進行中の言語変化の調査事例を報告した。これら二つの研究結果については、本稿では十分な議論をし尽くせなかった。これは、本稿の第一の目的が、関西弁コーパスを用いて実施可能な研究事例の報告にあったからである。本報告書では、これら二つの調査課題の検討と報告のみにとどまったが、それらの調査結果から、関西弁コーパスを用いた今後の研究への展望を示すことができる。具体的に言うと、方言の獲得の検証や大都市地域から郊外コミュニティへの言語の最新特徴の伝播を検証していく上で、関西弁コーパスは確実に有用であると考ええる。また、関西弁コーパスの作成は、当該分野への学術的貢献のみならず、大阪弁の永久的な記録と将来世代のための言語の保存という意味でも重要な役割も果たしていくであろう。

## 【参考文献】

- 内田らら (2001)「会話に見られる〈なんか〉と文法化：〈前置き表現〉の〈なんか〉は単なる口ぐせか？」『東京工芸大学紀要』24-2, 東京工芸大学
- 大阪府 (2012)「大阪の毎月推計人口」  
(<http://www.pref.osaka.jp/toukei/jinkou/index.html>)  
(2012年4月アクセス)
- 小林隆・篠崎晃一編 (2007)『ガイドブック方言調査』ひつじ書房
- 陣内正敬・友定賢治編 (2005)『関西弁の広がりコミュニケーションの行方』和泉書院
- 陣内正敬 (2007)「若者世代の方言使用」『方言の機能』岩波書店
- 日本大辞典刊行会編 (1975)『日本国語大辞典』小学館
- 平山輝男編 (1997)『日本のことばシリーズ27：大阪府のことば』明治書院
- 福原裕一 (2009)「会話に見られる〈なんか〉の機能拡張—フェイス・ワークの観点から—」『国際文化研究』15, 東北大学
- 札埜和男 (2006)『大阪弁「ほんまもん」講座』新潮新書
- 前川喜久雄・籠宮隆之・小磯花絵・小椋秀樹・菊池英明 (2000)「日本語話し言葉コーパスの設計」『音声研究』4-2, 日本音声学会
- 牧村史陽編 (1984)『大阪ことば辞典』講談社
- 山下好孝 (2004)『関西弁講義』講談社
- 李秀賢 (2006)「若者ことばにおける〈なんか〉について」『日本文化學報』31, 台湾日本語文学会
- Chambers, J. K. (2003) *Sociolinguistic Theory*. 2<sup>nd</sup> ed. Oxford and Malden, MA: Blackwell.
- Chambers, J. K., and Peter Trudgill. (1998). *Dialectology*. 2<sup>nd</sup> ed. Cambridge: Cambridge University Press.
- Fujimoto, Masako, Emi Murano, Seiji Niimic, and Shigeru Kiritani. (2002) Differences in Glottal Opening Pattern between Tokyo and Osaka Dialect Speakers: Factors Contributing to Vowel Devoicing. *Folia Phoniatrica et Logopaedica* 54(3): 133-43.
- Labov, William. (1966/1996) *The Social Stratification of English in New York City*. 2<sup>nd</sup> ed. Cambridge: Cambridge University Press.
- Labov, William. (1972) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara. (2000) Spontaneous speech corpus of Japanese. In *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens, Greece, 31 May to 2 June 2000, pp. 947-952.
- Tagliamonte, Sali. (2006) *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

