

Twitterを利用した新語・流行語研究の可能性

— アイドルグループ「Sexy Zone」の略語を例に —

Possibility of conducting research on neologisms / words in fashion using the Twitter
: A trial analysis of abbreviations of *Sexy Zone* (the name of Japanese idol group)

岡 田 祥 平

OKADA Shohei

1. はじめに

言語研究を行うにあたっては、研究対象の言語のデータを収集する必要がある。その手法としては、大別して、母語話者の内省を利用する場合¹⁾と、実際に使用された様々な言語資料から用例を収集する場合がある(松村一登2004での議論も参照)。それぞれの方法の長所と短所については岡田祥平(2013)でもまとめた²⁾が、日本語研究の世界においては、インターネットが登場して以降、実際に使用された様々な言語資料から用例を収集する方法の一つとして、インターネットを利用する可能性が模索され続けている³⁾。

日本語研究者が用例収集の対象としてインターネットに着目した理由はいくつも考えられるが、荻野綱男ほか(2005)は、その理由を、以下に示すように簡潔にまとめている。

- (1) 検索エンジンを経由すれば、簡単に大量の用例が入手できるので、WWWは用例データベースとしてきわめて有用である。また、ごく短時間に(秒のオーダーで)検索ができてしまうので、言語事実をすぐに確認できるということになる。

実際に使用された様々な言語資料から用例を収集する手法で言語研究を進める場合、「一般に、扱ったデータの量が大きければ大きいほど、そこから得られる情報に基づいた結論の信頼性が高くなる」(小野正弘2011)と考えられている。それゆえ、(1)に引用した荻野ほか(2005)が指摘する「大量の用例が入手できる」という点は、用例を収集する手法で言語研究を進める研究者にとっては非常に魅力的といえるのである⁴⁾。

また、インターネットの検索エンジンの「手軽さ」は、多くの人々が体感していると思われるので、このことをここで改めて強調するまでもないことだと思われる⁵⁾。

しかし、インターネットを利用して用例を収集する方法の問題点も、既にいくつも指摘されている(岡島昭浩1997, 田野村忠温2000, 田中ゆかり2003, 荻野ほか2005, 前川喜久雄2007, 荻野綱男2007・2008, 滝沢直宏2007, 荻野綱男ほか2007, 石川慎一郎2008など。本稿の5. 節の議論も参照)。そのひとつとして、前川(2007)は以下のような点を指摘している。

- (2) 所与の文書の出典情報。特にジャンル情報(新聞記事か小説か論文か, 論文なら領域は何か)をあらかじめすることができない。また文書の著者の社会的属性(性別, 年齢など)を知ることができない。そのため、言語学的な偏りを評価できない。

一方、荻野(2007)は以下のように述べ、前川(2007)が指摘する点は必ずしも問題にはならないとの立場に立つ。

- (3) 今までの言語研究だって、実は、テキストそのものだけでなく、その回りのさまざまな情報を集めてジャンルや書き手について調べてきたのである。WWWコーパスでそれができないはずはない。ただ、WWWが膨大な言語量があるために、全部の用例についてこのような調査をするのは大変だというだけのことである。

筆者は上に引用した萩野(2007)とは異なる立場から、すなわち、インターネットで検索を行う際、検索の対象を限定するという工夫を施すことによって、前川(2007)が指摘する懸念は、ある程度は解消できると考えている。そのような問題意識のもと、岡田(2013)では、インターネットを利用した用例収集の可能性を模索した先行研究を参照しつつ、以下のような手法を提案した。

- (4)・PDF文書に限定して検索をかける
- ・ドメインを限定して検索をかける
 - ・ニュースサイトに限った検索を行う
 - ・「国会会議録検索システム」(<http://kokkai.ndl.go.jp/>)を利用する
 - ・「青空文庫」(<http://www.aozora.gr.jp/>)を利用する
 - ・新聞社のデータベース⁶⁾を利用する
 - ・ブログを検索する
 - ・質問サイト(「Yahoo!知恵袋」,「教えてgoo」)を検索する
 - ・Twitter(ツイッター)のツイートを検索する
 - ・各種歌詞検索サイトを利用する
 - ・「少納言」(<http://www.kotonoha.gr.jp/shonagon/>)⁷⁾を利用する
 - ・「NINJAL-LWP for BCCWJ」(<http://nlb.ninjal.ac.jp>)⁸⁾を利用する
 - ・「国立国会図書館デジタル化資料」(<http://dl.ndl.go.jp/>)を利用する

本稿では、(4)であげた手法の中から、Twitterを利用して用例を探す日本語研究の実践を報告し、Twitterを利用した日本語研究の可能性を模索することにする。まず、2.節ではTwitterについて簡単な説明を行う。続く3.節ではTwitterを利用した用例収集の可能性を模索した先行研究を、4.節ではTwitterを利用した日本語研究の実践した先行研究を、それぞれ紹介する。そのうえで、5.節では筆者が利用を勧めるTwitterのツイート検索エンジンを紹介し、6.節ではそのツイート検索エンジンの使い方を概説する。その後、7.節では筆者が行ったTwitterを利用した新語・流行語研究の事例を紹介する。最後に8.節で本稿のまとめを簡単に行う。

本稿を大きく二分するならば、2.節から6.節までがTwitterを利用した日本語研究を行うための前提となる事柄を論じた部分であり、7.節が実際の分析結果を報告した部分となる。ご関心に応じて、該当箇所をお読みいただくということでも一向に構わない。

なお、あらかじめ、いくつかお断りしておく。

まず、Twitterを利用した日本語研究の可能性を検討する具体的な現象として、新語、流行語を対象にした理由については、7.1.節での議論もご参照いただきたい。また、分析対象として取り上げる現象がアイドルグループ名の略語という点に対しては、学術論文には「ふさわしくない」と思われる読者もいらっしゃるかもしれないが、今回は、あえて、このような、日本語研究とは無縁と思われる「卑近」な現象を取り上げることにした。一見すると「卑近」な現象であっても、実は、日本語研究(や言語研究)の世界に興味深い知見や視点を提供してくれる可能性があるということを示し、これまで、現代日本語を観察することには興味を持っていなかった方も現代日本語に関心を向けていただけるきっかけとなることを願い、本稿の研究、分析対象を設定した次第である⁹⁾。

また、教育学部の紀要という性格上、日本語研究を専門としない読者がいる場合のことも想定し、日本語研究を専門としない読者が必要と思われる事項の説明を註に加えた。註が多く、煩雑なであることを承知しているが、そのような意図があるため、どうかご容赦いただきたい。

2. Twitterとは

Twitterとは、「インターネットを通じて140字以内の『つぶやき』を不特定多数にリアルタイムに発信し、自分で選択した他人の『つぶやき』を受信するサービス」(津田大介2009)のことである。2006年7月から開始されたこのサービスは、2012年3月時点では、全世界で1億4,000万人、日本国内で1,400万人の利用者がいるという(総務省『情報通信白書平成24年版』<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/html/nc123220.html>。2013年6月8日閲覧)。

Twitterの主な利用者層であるが、株式会社電通パブリックリレーションズが2012年11月に15歳～69歳の男女1339人(男性674人、女性665人)を対象に行った調査によると、以下の図1に示したとおり、若年層ほどTwitterを「現在利用している【ほぼ毎日】」と「現在利用している【たまに】」とする回答が多かった(http://www.dentsu-pr.co.jp/wp-content/themes/dpr_themes/release/20121226_report.pdf。2013年6月8日閲覧。図1はリンク先に記載してあったデータを筆者でグラフ化したもの)。

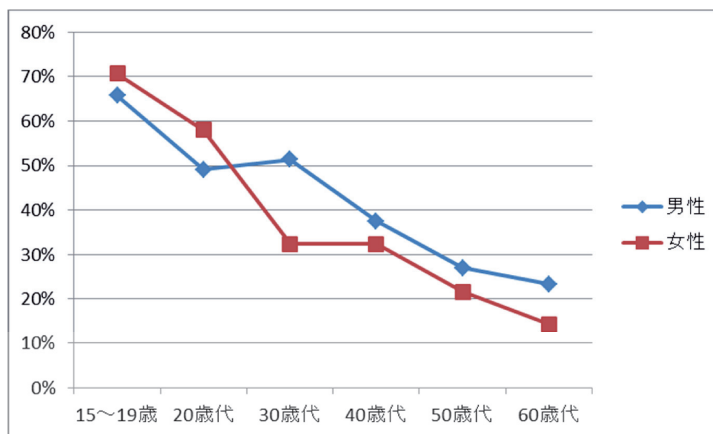


図1 性・年代別Twitter利用率(「ほぼ毎日利用」「たまに利用」の合計)

以上のような事実を踏まえると、Twitterは、若年層の言葉を中心とした言語データの集積体として捉えることが可能なのではないかと考えている。

なお、Twitter日本版(<https://twitter.com/>)では、当初は、140字以内の投稿のことを「つぶやき」と称していたが(そのため、上記に引用した津田2009も「140字以内の『つぶやき』」という表現を使用している)、2009年10月以降、投稿のことを「ツイート」と称するように変更された(<http://www.itmedia.co.jp/news/articles/0910/26/news084.html>。2013年6月8日閲覧)。そのため、本稿では、以降、Twitterに投稿された140字以内の投稿のことを「ツイート」と表記することにする。

3. Twitterを利用した用例収集の可能性について

Twitterを現代日本語研究の用例収集の材料として使用することの可能性について述べた文献としては、岡島昭浩(2011)があげられる。

岡島(2011)は、Twitterを利用した用例収集は、「細切れの発言の前後がたどりにくく、用語の意味を考察しようとする際に不便である」と述べている。しかし、逆に言うならば、「用語の意味を考察しようとする」研究(たとえば、新語や「目新しい」形態の使用状況を考察する研究)にTwitterを利用することの可能性は、検討する価値はあると思われる。

また、Twitterを利用した用例収集の利点として、岡島(2011)は、「ブログなどと違って、日付を換えて記事を作ったり、後日文言を修正したりすることが不可能なので、いつの発言であるのかが特徴できる」ことをあげている。つまり、Twitterを利用すれば、新語や「目新しい」形態の使用状況について、当該現象がいつごろから使用され始めたのかを明らかにしたり、当該現象の使用状況の変遷を計量的に記述したりと

いったことが可能になると考えられる。

ただし、Twitterは歴史が浅い（2. 節で触れたように、2006年7月サービス開始）ため、Twitterを言語変化研究に利用するとしても、現時点（2013年6月現在）では、ごくごく最近に生じた現象しか研究の対象にすることができない。さらに、「過去に遡ることが重視されていない」（岡島2011）ため、過去のデータをどのように検索するか、という大きな問題を解決する必要もある。岡島（2011）ではいくつかのツイート検索サービスを紹介されている¹⁰⁾が、現時点（2013年6月現在）では、それらのツイート検索サービスはいずれも日付を指定しての検索はできないようであり、Twitterを利用した言語変化の研究を行おうとする際の障壁になる。

しかし、岡島（2011）が指摘した「過去に遡ることが重視されていない」という問題点は、「TOPSY」（<http://topsy.com/>）というツイート検索エンジンを利用すれば、（ある程度は）解決できると筆者は考えている。そこで、本稿では、TOPSYを利用して、用例を収集した実践例を報告する。TOPSYの詳細は、5. 節で紹介する。

4. Twitterを利用した日本語研究の実践例

五味伸之ほか（2011）には、「Twitterについては、すでに様々な先行研究が行われて」いるという記述がある。同時に、五味ほか（2011）は、Twitterを利用した研究は、「情報処理側からの研究が主となっており、目的はIMEの開発や効率的な情報収集法などである」と指摘している。実際、筆者が言語学、あるいは日本語学の立場から、Twitterを利用した研究を行ったという例を探してみたものの、管見の範囲ではほとんど見当たらなかった（用例収集の対象の一つとしてTwitterを取り上げている研究はいくつか存在したが、用例収集の対象をTwitterのみに限定し、Twitterの特性を生かして、言語学的、あるいは日本語学的な考察を行った研究は、管見の範囲ではほとんど見当たらなかった）。

そのような意味で、五味ほか（2011）は、Twitterを利用した言語学的、あるいは日本語学的研究の実践の数少ない例の一つとしてあげられる。以下、簡単に五味ほか（2011）の概要を紹介し、Twitterを利用した日本語研究の可能性を考えたい。

五味ほか（2011）では、動詞「違う」の形容詞化¹¹⁾現象について、Twitterを対象に用例を収集し、分析を行った結果、以下のような結論を導き出している。

(5)①Twitterにおける「違う」の形容詞化率は、わずか2.80%に留まっている。

②ただし、「違わない・違（う）くない」での形容詞化率34.87%であり、全体の形容詞化率2.80%という値と比較して有意に高率である。

③「違う」が動詞活用のときは漢字表記、形容詞活用するときにはかな表記になる傾向が見られた。これは、コンピュータ、携帯電話などの入力装置の漢字変換機能が、「違う」を形容詞型に活用させることに対応していないためだと考えられる（形容詞型の活用をさせた「違う」を漢字でTwitterに投稿する際には、「違う」「違い」と漢字変換させた上で、活用語尾を削除し、形容詞型の活用語尾を新たに入力しなければならない手間がかかる）。

④総ツイート数が多いTwitterをよく使用しているユーザの方が、「違う」を形容詞化しやすい傾向があると考えられる。

実は、動詞「違う」の形容詞化現象については、五味ほか（2011）とほぼ同時に発表された石井由希子（2011）でも取り扱われている。ただし、その両者は言語データの収集の方法が異なる。すなわち、五味ほか（2011）はTwitterを利用した言語運用の実態を記述した研究である一方、石井（2011）はアンケート調査¹²⁾による言語意識の実態を記述した研究と位置づけられる。したがって、五味ほか（2011）の結果と石井（2011）の結果を比較すると、言語運用の実態と言語意識の乖離という問題（これは、1. 節で述べた、言語研究における言語データ収集の二つの方法の長所と短所にもつながる問題である）を考える際には、非常に興味深い視点を提供してくれるため、ここでも簡単に紹介しておく（言語研究において、この問題を考えるのは非常に興味深く、また、非常に重要だと筆者は考えているが、紙幅の都合上、本稿では深く立ち入らない）。

五味ほか（2011）と石井（2011）の結果を比較、対照させると、以下の表1、表2のようになり、両者の結果には大きな相違が存在することが読みとれる。

表1 五味ほか（2011）における

「形容詞化率」 ¹⁴⁾	
チガ（ウ）クナイ	34.87%
チガ（ウ）カット	2.66%
チガ（ウ）クテ	1.91%
チガ（ウ）ケレバ	5.79%

表2 石井（2011）の調査結果¹³⁾

	違和感なく、 許容できる	聞いたことあり、 自分でも使うことがある
チガクナイ（疑問文での 使用の場合）	52.90%	94.0%
チガクナイ（平叙文での 使用の場合）	18.0%	54.0%
チガカット	46.0%	54.0%
チガクテ	32.4%	68.0%
チガクテモ	46.0%	84.0%
チガケレバ	53.1%	62.0%

五味ほか（2011）と石井（2011）とは、言語運用の実態調査と言語意識の実態調査という調査方法上の大きな違いがある上に、調査対象現象の認定方法の違い¹⁵⁾や、調査対象となったレジスター（言語使用域）¹⁶⁾が異なるなど、種々の違いが存在する。そのため、五味ほか（2011）の結果と石井（2011）の結果とを、同一レベルで比較、対照させることは問題もあろうが、それでも、両者の際は、言語運用の実態調査の結果と言語意識の実態調査の結果とのあいだには乖離が少なからず認められることを計量的に示すことができたという点で、意味がある試みだと考える。すなわち、ある言語形態を「使用する」という意識の比率が、実際の使用率と呼応しているとは限らない、ということである。このことは、その事実、言語意識の調査（内省を使った調査）を行うと同時に、言語運用の実態の調査（用例を収集する調査）を行うことの重要性を示唆していると考える。

なお、五味ほか（2011）は、1. 節で述べたインターネットを利用した用例収集の長所を利用した研究、すなわち、検索エンジンを利用して、大量の用例を手軽に収集した研究ではないことは、付記しておきたい。実は、五味ほか（2011）では、分析対象とする「サンプルは、Twitter, inc. が提供するAPIの内Streaming API::sample メソッドを利用し、全世界のパブリックツイート（公開ツイート）から大まかに1%を取得した」とあり、Twitter全体を分析の対象にしてはいないようである。また、データの収集、分析にあたっては以下のような、人文系の研究者には敷居が高い手法が取られている¹⁷⁾。

- (6) ・ツイートデータの受信には、短時間で多量のデータ保存処理を超並列に実行出来る、プログラミング言語 Erlang を用いた
- ・データベースには、JSON 形式で直接保存できる CouchDB を採用した
- ・稼働率を向上させる目的で、受信クライアントは3台の Linux サーバ上で独立に同時運用した
- ・収集完了後、データベース内のツイートに対し、正規化プログラムを適用したサンプルが収集された

さらに、五味ほか（2011）では、データ収集期間が2011年7月30日から2011年9月1日という約1ヶ月間に限定されており、通時的観点からの研究は意識されていない¹⁸⁾。

5. ツイート検索エンジン「TOPSY」について

3. 節で述べた通り、先行研究（岡島2011）で紹介されていたツイート検索エンジンは、日本語研究には必ずしも適していると言いがたいものであった。しかし、「TOPSY」(<http://topsy.com/>)というツイート検索エンジンは、日付を指定しての検索も可能であり、ツイート検索を利用した言語変化研究を含め、様々な観点からの日本語研究に利用できそうな感触を、筆者は抱いている。

TOPSYの利用方法の詳細については、6. 節で紹介するが、概要は以下のとおりである（『IT用語辞典バイナリ』の「Topsy」の項<http://www.sophia-it.com/content/Topsy>。2013年6月7日閲覧）。

- (7) Topsyとは、Topsy Labs, Inc.が提供しているリアルタイム検索のサービスである。Twitterを中心とするリアルタイム性の高いメディアに特化して検索を行うことができる。

Topsyは2006年に設立され、2009年にサービスの提供を開始した。Googleが2009年12月に、Web検索結果ページ上でTwitterを対象とした検索機能を提供し、その後、同機能を取り外した。その代替需要もあって、2011年7月現在、Topsyは手頃なTwitter検索のサービスとして注目を集めている。

インターネットで情報を収集すると、特に過去のツイートの検索を行う際には、TOPSYの利用を勧める人が多いようである。たとえば、テクニカルライターの原 如宏氏は、「過去のつぶやきも探したいなら、ツイッターの投稿を検索のデータベースに利用する「TOPSY」がお薦め」と述べている (<http://www.yomiuri.co.jp/net/qanda/20110711-OYT8T00181.htm>。2013年6月7日閲覧)。また、ブログ「Startup Reality」の2011年4月29日付の記事「Twitterの過去のツイートを検索するにはTopsyが最強」(http://d.hatena.ne.jp/a_kimura/20110429/1304078734#20110429f2。2013年6月7日閲覧)によれば、「過去のツイートを検索するならTopsyが最強」である理由として、「超高速」「抜けがない(なさそう)」「期間指定もできる」「時系列ソートもできる」「言語指定もできる」という五つの点をあげている¹⁹⁾。それらの利点に加え、ここでは、「検索の結果、示されるヒット件数が安定している」という点もTOPSYの利点であるということを描きつけておきたい。

一方で、インターネットの検索エンジンを利用して日本語の用例収集を行っている研究者によって、インターネットの検索エンジンの検索結果(件数)が安定していないということが報告されていることには留意したい。たとえば、田野村忠温(2008・2010・2012)は、Googleの検索結果には「時間的安定性」が欠けている²⁰⁾ことを指摘している。また、田野村(2008)では、Googleの検索結果には「論理的整合性」が欠けている²¹⁾ことも指摘されている。さらに、小野(2011)は、インターネットの検索では、「[ユーザーの地域]」「どのデータセンターにアクセスしているか」「ブラウザ」「OS」などが、検索結果を左右するという」と述べている。これらの問題点は、言語研究、中でも計量的な研究にとっては、深刻である。検索結果が安定しないということは、検索結果に基づいた計量的観点からの議論の妥当性を疑わせることにつながるからである。しかし、TOPSYについては、簡単であるが筆者が調査した結果、先行研究で指摘されたインターネット検索エンジンの問題点のような現象は確認できず、安定的な検索結果を得ることが確認できた²²⁾。それゆえ、TOPSYを利用した言語調査は、ある程度の信頼性があると言えそうである。

以上の議論をまとめるならば、ツイート検索サービスTOPSYを利用すれば、Twitterを利用した日本語研究、特に新語や「目新しい」形態の使用状況を考察する研究や、近年生じた言語変化の研究の新たな手法を展開する可能性がある、ということになる。

なお、TOPSYでは、2008年5月以降のツイートが検索可能のようであることも付記しておく。(<http://searchengineland.com/topsy-now-searching-tweets-back-to-may-2008-49162>。2013年6月7日閲覧)。

6. TOPSYの検索方法

6.1. ツイートされた日時の指定をしなくて検索する場合

TOPSYで、ある単語やある表現を、日時の指定をせずに検索する場合には、トップページ<http://topsy.com/>にアクセスし、そこに表示されている「検索窓」に検索したい単語や表現を入力し、「検索窓」の右端にある「虫眼鏡」のアイコンを押すだけでよい(図2参照)。ただし、ここで注意したいのは、「検索窓」の上部に記載している検索対象の候補の中から、「ツイート」を選択したほうがよさそうである、ということである²³⁾。

図2の画面から検索を実行すると、まず、図3のような画面が表示される。図3は、「日本語」を検索語としてTOPSYで「ツイート」を検索した結果、最初に表示される画面である(2013年6月8日20時に検索)。ここで注意しなければならないことがある。それは、検索窓の下に、「æ¥æ-èについて3,483件のツイート」と表示されている²⁴⁾。ここで注意を要するのは、この3,483件という結果は、ツイート全体を検索した結果ではなく、検索時点から遡ること2時間のあいだに発信されたツイートのみを対象に検索した結果の件数である、という点である。それは、図3の左端の「過去1時間」「過去2時間」…と表示されている欄の「過去2時間」のところの色が濃くなっていることからわかる。もし、ツイッター全体を検索した結果を知りたい場合は、左の欄の「過去すべて」をクリックする。そうすると、TOPSYが検索対象としているツイー

TOPSY

Search and Analyze the Social Web.

すべて LINKS ツイート 写真 動画 エキスパート トレンド

Try a social search



図2 TOPSYのトップページ

TOPSY

日本語

ツイートを検索

検索オプション

æʏæ-èʔについての3,483件のツイート中1-10件目の検索結果 [Eメールアラートを設定](#)

最新の検索結果

過去1時間	911
過去2時間	3,483
過去1日	26K
過去7日	200K
過去30日	866K
過去すべて	12M



mou_uknow 유노최고!!(mou)

【京セラ3日目】ユノ最後の言葉。いつが一番幸せか考えてみました。日本語変かもしれないけど、この、ステージが一番幸せです。皆さんがいるから僕たちがいる。最近わかってきました。#東方神起 #tohotimetour

11分前 Reply Retweet Favorite 90



akisameruu 【お絵描き】秋雨ルウ【大好き】

【なぜか炎上】人気漫画『進撃の巨人』作者ブログが荒らされる 韓国語と日本語が入り乱れる状態に <http://t.co/JYpFVSv7Pr> 素直に作品を楽しめないとは、なんと貧しい心の持ち主は。例え仮に全員日本人だったとしても、心から軽蔑する。

2時間前 Reply Retweet Favorite 14

図3 TOPSYによるツイートの検索結果画面①（ツイートされた日時を指定しない場合）

ト全体を検索した結果を表示することができる（図3の例であれば、11,821,597件であった）。

以上述べた通り、TOPSYで、ツイートされた日時の指定をせずに検索をした結果、直後に表示される画面は、図3に示したとおり、ツイート全体を検索した結果ではないことに注意すべきである。ツイート全体を検索した結果を知りたい場合は、検索を実行した直後に表示される画面上で、さらに画面左の欄において、「過去すべて」を選択する必要があることを繰り返しておく。

なお、図3を見ると、検索結果画面でツイートを確認できることがわかる（どのような文脈で検索語が使用されているのかが確認できる）。さらに詳細を知りたいければ（当該ツイートがTwitter公式ページではどのように表示されるか、など）、表示されているツイートの左下のTwitter公式ロゴ、もしくはツイートされた時間が表示されている箇所（たとえば「〇分前」と表示されている箇所）をクリックすればよい。

6. 2. ツイートされた日時を指定して検索する場合

TOPSYで、ある単語やある表現を、ツイートされた日時の指定を指定して検索する場合には、<http://topsy.com/advanced-search>にアクセスする（図3で示した検索結果の、「ツイートを検索」の右にある「検索オプション」をクリックしても、当該ページにアクセスできる）。当該ページにアクセスすると、以下の

図4のような画面が表示される。ツイートされた日時を指定して検索する場合は、「最も古い日付と時間」と「最新の日付と時間」を指定すればよい（時間は「PST」とあるとおり、日本標準時ではなく太平洋標準時で指定することになるようである）。また、この画面において、「検索対象のタイプ」を「ツイート」に設定することも忘れないことも大切である。さらに、「エキスパートを検索」の項目については、図4にある通り、「関連性」を選択する（その理由については、6.4.節を参照）。

すべてのキーワードを含む:

いずれかのキーワードを含む:

キーワードを含めない:

検索対象のサイトやドメイン:

検索対象のTwitterユーザ:

検索対象のタイプ: すべて ▾ ☐ トレンドだけを表示

言語を指定して検索: すべての言語 ▾

エキスパートを検索: 関連性 ▾

最も古い日付と時間: 12AM ▾ PST

最新の日付と時間: 12AM ▾ PST

検索

図4 TOPSYでツイートされた日時を指定して検索する場合の画面

図5は、図4の画面で検索対象期間を2013年5月1日AM12時から2013年5月31日AM12時までと指定し、「日本語」を検索語として「ツイート」を対象に「すべての言語」を選択して検索した結果、最初に表示される画面である（2013年6月8日20時40分に検索）。これを見ると、指定された日時の範囲で、「日本語」とい

TOPSY ツイートを検索 [検索オプション](#)

2013年5月04日12時05分から2013年5月31日12時の間の日本語のツイートについての8,557件のツイート中1-10件目の検索結果 [Eメールアラートを設定](#)

検索対象開始日
04/30/2013

検索対象終了日
05/30/2013

最新の検索結果

- 過去1時間 0
- 過去1日 150
- 過去5日 2,252
- 過去7日 2,872
- 過去30日 10K
- 過去すべて 135K

w7voa Steve Herman
Mainichi: 71% of #Japan voters consider as inappropriate Osaka mayor's comments on 'comfort women' - <http://t.co/WzPtXnyfGI> (日本語)
20日前 Reply Retweet Favorite 21

bot_furby ファービー
キス..カラノ重伝 日本語版 予約受付中! <http://t.co/fVGRelh2xs> <http://t.co/4b1r0ePCZb> <http://t.co/IRCr1e64Es>
16日前 Reply Retweet Favorite 15

yoshioterada Terada Yoshio
Java SE7API日本語版提供開始 <http://t.co/nrFqL02JFE>
25日前 Reply Retweet Favorite 76

teen_top_japan TEEN_TOP_JAPAN
<日本語訳> [TEENTOP No.1 Concert D-DAY] <https://t.co/ijdek0GBxy>
29日前 Reply Retweet Favorite 28

図5 TOPSYによるツイートの検索結果画面②（ツイートされた日時を指定した場合）

う単語を含むツイートが8,557件存在したことがわかる。なお、図5では、「検索対象開始日」が「04/30」、「検索対象終了日」が「05/30」と、それぞれ指定した日から1日早く表示されているが、なぜ、このようなことが生じるか、原因不明である。他の日時を指定して検索を掛けても同様の現象が生じるし、検索結果の画面で検索期間を修正しようとしても修正できない、ということを付記しておく。

6.3. 言語を指定した検索

TOPSYでは、「言語を指定して検索」することも可能である。「言語を指定して検索」する場合には、図



図6 TOPSYによるツイートの検索結果画面で
検索対象言語を指定する場合

4の画面の「言語を指定して検索」の欄にて検索対象言語を指定するか、検索結果が表示された画面(図3, 図5)の左端の下に、図6のような選択欄(「言語」という欄)が存在するので、そこで「言語を指定」してやればよい。

なお、TOPSYがどのような基準に基づいてツイートの言語を指定しているかについては、よくわからないことも付記しておく。

6.4. TOPSYを利用した用例収集の問題点

6. 節の最後に、TOPSYを利用した用例収集の問題のうち、筆者が気づいた点で、ここまでで触れることのできなかった事項をいくつかまとめておく。

一つは、TOPSYの検索結果は、検索語が含まれるツイートの数で、用例数ではないことである。つまり、一つのツイートに複数回検索語が含まれている場合でも、TOPSYの検索結果では「1件」とカウントされる、ということである。

また、TOPSYの検索システムが、判然としないところも問題点の一つであろう²⁵⁾。ここでは、筆者が感じた二つの問題点を記しておく。

まず、リツイートの扱いが不明である、という点も留意したい。リツイートとは、「ほかのユーザーの投稿を「転載」しつつ、再投稿する行為」(津田2009)のことである。TOPSYの検索結果には、検索語が含まれているツイートがリツイートされた件数が反映されるのか、それとも元のツイートしか反映されないのか、判然としない²⁶⁾。もし、前者であれば、リツイート数の多いツイートに検索語が含まれていた場合、実際の使用は1回とみなすべきであるにもかかわらず、検索結果は多くなってしまい、「歪んだ」結果を生ん

でしまうのだが、TOPSYの検索のシステムはこの点が判然としない。

次に、「エキスパートを検索」「ソート」についてである。図4の画面では「エキスパートを検索」という項目が、図6の画面では「ソート」という項目がある。この項目の選択肢は、図6に示されているとおり「関連性」と「日付」であるが、この項目は「関連性」を選択して検索をかけないと、検索結果表示画面において検索件数が表示されない(理由は不明。そもそも、「エキスパートを検索」と「ソート」が一体どのような機能なのか、詳細は不明である)。図7は、当該項目について「日付」を選択した以外は図3と同条件で検索をかけた結果であるが、図7を見て分かる通り、図3では表示されていた検索結果件数が表示されない(図4、もしくは図6の画面の当該項目について「関連性」を選択して検索を実行すれば、8,557件という

検索結果件数を得ることができる)。

以上のような事実を考えると、TOPSYを利用して得られたTwitterにおける用例数を計量的に検討することには、大きな危険性も含まれていると言わざるを得ない。また、しかし、本稿の目的は、Twitterを利用した言語研究、日本語研究の第一歩という意味で、この問題点については、今後の課題としたい。

TOPSY

[検索オプション](#)

2013年04月30日00時00分から2013年05月30日00時00分までの間のツイートについてのタイムライン [Eメールアラートを設定](#)

検索対象開始日

04/30/201



b2stv B2ST TV
[日本語字幕] BEAST(ビースト) - Crazy [#KPOPTV](http://t.co/sghXD0cPRE) #B2ST
2013年05月30日 [Reply](#) [Retweet](#) [Favorite](#) [4](#)

検索対象終了日

05/30/201



ohboy_dujun オボイ
DJ's Birthday support notice upload / 日本語 <http://t.co/5JP9bJYJPP>
2013年05月29日 [Reply](#) [Retweet](#) [Favorite](#) [18](#)

[最新の検索結果](#)

過去1時間 [1](#)
過去1日 [191](#)
過去7日 [1,473](#)
過去27日 [7,589](#)
過去30日 [8,510](#)
過去すべて [140K](#)



kui3 naoki
Amazing f(x) Special Episode(日本語字幕付)1/4[130524] (14:04) [#sm20942242](http://t.co/zyBhVNCaV3) <http://t.co/zyBhVNCaV3>
2013年05月29日 [Reply](#) [Retweet](#) [Favorite](#) [5](#)



fansofilwayne Lil Wayne Fans
Video: Tyga Speaks On Lil Wayne's Recent Health Scare【日本語字幕】 <http://t.co/lhgk3yd9OD>
2013年05月28日 [Reply](#) [Retweet](#) [Favorite](#)



sewoonggirl 성이(ソニア)
[日本語字幕] 2PM 'Grown' CD Delivery Event: <http://t.co/ZJQ4vYJkWd> (@YouTube)
2013年05月28日 [Reply](#) [Retweet](#) [Favorite](#) [4](#)

検索

[すべて](#)

図7 TOPSYによるツイートの検索結果画面③ (ツイートされた日時を指定した場合)

7. Twitterを利用した言語研究の可能性の実践—アイドルグループSexy Zoneの略語を例に

7.1. はじめに

Sexy Zone (セクシー・ゾーン。以下、本稿では「Sexy Zone」と表記する)とは、2011年9月29日に結成することが発表された²⁷⁾、ジャニーズ事務所に所属する5人組のアイドルグループである²⁸⁾。7.3. 節でも述べるように、芸能人の名前(グループ名)は省略されることが多いが、Sexy Zoneの略し方については、ゆれが観察される(一つの略語に統一されていない)ようである。たとえば、インターネットの質問サイト「Yahoo!知恵袋」(<http://chiebukuro.yahoo.co.jp/>)では、Sexy Zoneの略語が「セクゾン」なのか「セクゾ」なのかという質問が、複数件ヒットする。そのほとんどがSexy Zoneのファンから投稿されたものと推定できるため、Sexy Zoneのファンのあいだにおいても、Sexy Zoneの略語がどのようなものになるのか/どのようにすべきなのか(「セクゾン」なのか、「セクゾ」なのか)、迷いがあることが伺える。そして、Sexy Zoneがデビュー(2011年9月29日)から数ヵ月後の時点で、この種の質問が出始めたことも興味深い(この種の質問の「Yahoo!知恵袋」での初出は2012年1月26日)。なぜなら、この事実はSexy Zoneのデビューからきわめて短期間で、Sexy Zoneのファンの間でSexy Zoneの略語が「セクゾン」と「セクゾ」の双方が拮抗して使われることが意識されるようになったことを示唆しているからである。つまり、Sexy Zoneの略語をめぐる問題はきわめて短期間のあいだに形成された「ことばのゆれ」の現象であると考えられそうである。しかしながら、Sexy Zone「セクゾン」と「セクゾ」、それぞれの略語がどの程度使われているのか、また、このような「ゆれ」はどのように形成されたのか、詳らかではない。

3. 節で、Twitterは歴史が浅い(2006年7月サービス開始)ため、現時点(2013年6月現在)では、ごくごく最近に生じた現象しか研究の対象にすることができないと指摘したが、Sexy Zoneの略語をめぐる問題は2011年9月以降に生じた現象であり(Sexy Zoneデビュー以前は起こりえない問題だからである)、Twitterを使用した日本語研究の可能性を検討するには、格好の材料と考えた次第である。また、3. 節で

も紹介した通り、Twitterは「細切れの発言の前後がたどりにくく、用語の意味を考察しようとする際に不便である」(岡島2011)であるが、略語の使用状況の観察には意味を考慮する必要がないため、Twitterを使用した日本語研究の可能性を検討する際にも大きな問題にならないと思われる。

以上のような問題意識のもと、本節では、Twitterを利用し、Sexy Zoneの略語の使用実態を計量的に記述、分析することを通じ、Twitterを利用した日本語研究、特に新語・流行語研究の可能性を模索する。

以下、まず、7.2. 節で、「前部要素+後部要素」からなる複合語における略語形成規則を確認する。つづく7.3. 節では、7.2. 節での議論を踏まえ、(日本の)芸能人の名前の省略方法の傾向を概観する。そのうえで、7.4. 節では7.2. 節で確認した規則の例外について紹介する。以上の議論を踏まえ、7.5. 節ではSexy Zoneの略語として予想される形態を紹介し、7.6. 節から7.8. 節では、ツイート検索エンジンTOPSYを利用し、Sexy Zoneの略語の実態について調査、分析、考察を行う。最後に、7.9. 節では、本節で紹介した調査、分析で残された課題について述べる。

7.2. 「前部要素+後部要素」からなる複合語における略語形成規則

現代日本語において、「前部要素+後部要素」からなる複合語²⁹⁾ から略語が形成される規則の概略については、窪園晴夫(2002)にわかりやすくまとめられている。以下、「前部要素+後部要素」からなる複合語から略語が形成される規則について、本稿の議論に関係のある部分のみを、窪園(2002)から抜き出して紹介する。

現代日本語において、「前部要素+後部要素」からなる複合語から略語が形成されるパターンは、以下の2種類に大別できる(「<」の右側に記されたものが省略される前の複合語、左側が複合語から形成された略語、ゴシック体は略語の構成要素となった部分を表す。これ以降、同様)。

(8) 両方の要素を部分的に残すパターン

例： ポケモン<ポケット・モンスター
経済<経世済民
ワングル<ワンダーフォーゲル
ナベサダ<渡辺貞夫³⁰⁾

(9) いずれかの要素をそのまま残すパターン

例： ケータイ<携帯電話
スーパー<スーパーマーケット
ホーム<プラットホーム

なお、窪園(2002)は、(8)(9)、双方とも、「前半部分を残す」パターンの生産性が高い、と指摘している。つまり、(8)の場合であれば「ワングル」「ナベサダ」のようなパターンよりも「ポケモン」「経済」のようなパターンが、(9)の場合であれば「ホーム」のようなパターンよりも「ケータイ」「スーパー」のようなパターンが、それぞれ好まれる、ということである。また、(8)の場合、形成される略語は、前部要素、後部要素の、それぞれの語頭から2モーラ³¹⁾ ずつ採り、略語全体としては4モーラになるパターンが「最も基本的」と述べている。

7.3. (日本の)芸能人の名前の省略方法

窪園(2002)が指摘する「前部要素+後部要素」からなる複合語の略語の生産性の高いパターン、すなわち、「前部要素と後部要素の語頭の2モーラを結合し、全体として4モーラの略語を形成する」というパターンは、芸能人(個人、グループ双方)の名前の省略方法にも、顕著に認められる。たとえば、以下にあげる芸能人たちの名前の略語は、すべて、「前部要素と後部要素の語頭の2モーラを結合し、全体として4モーラの略語を形成する」というパターンである。

- (10) キムタク<木村拓哉(きむら・たくや)
タカミナ<高橋みなみ
ももクロ<ももいろクローバーZ

(10)にあげた芸能人は比較的最近の例であるが、「前部要素と後部要素の語頭の2モーラを結合し、全体と

して4モーラの略語を形成する」というパターンに従う芸能人の名前の省略方法は、近年に限ったことではない。(11)に示すように、「古い」芸能人にも当てはまる傾向なのである。

- (11) ドリカム<DREAMS COME TRUE (ドリームズ・カム・トゥルー)³²⁾

ブリブリ<プリンセス・プリンセス³³⁾

カツシン<勝新太郎 (かつ・しんたろう)³⁴⁾

エノケン<榎本健一 (えのもと・けんいち)³⁵⁾

アラカン<嵐寛寿郎 (あらし・かんじゅうろう)³⁶⁾

いずれの場合も、「前部要素と後部要素の語頭の2モーラを結合し、全体として4モーラの略語を形成する」というパターンに従っている。

以上をまとめると、日本の芸能人の名前の省略方法の大原則は「前部要素と後部要素の語頭の2モーラを結合し、全体として4モーラの略語を形成する」というものであり、しかも、その大原則は、近年に限ったものではない、と言えるのである。

7.4. 「前部要素+後部要素」からなる複合語における略語形成規則の例外—長母音が関与する場合

7.2. 節で、「前部要素+後部要素」からなる複合語の略語は「前部要素と後部要素の語頭の2モーラを結合し、全体として4モーラの略語を形成する」場合が生産性が高いとまとめたが、長母音が関与する場合、議論は若干複雑になる。

たとえば、(12)のように、後部要素の2モーラ目が長母音の場合、形成される略語では長母音が削除され、「2モーラ+1モーラ」の略語が形成されることがある。

- (12) テレカ<テレホンカード

カンペ<カンニングペーパー

フリマ<フリーマーケット

勿論、(13)のように、後部要素の2モーラ目が長母音の場合でも長母音がそのまま残り、「2モーラ+2モーラ」の略語が形成されることもある。

- (13) バトカー<バトロールカー

一方で、(14)のように、後部要素の2モーラ目の長母音が削除された上で、後部要素の3モーラ目が略語構成要素として採用され、「2モーラ+2モーラ」の略語が形成されることもある。

- (14) カラオケ<空オーケストラ

ヘビロテ<ヘビーローテーション

ヤフオク<ヤフーオークション

窪田 (2002) によると、後部要素の2モーラ目が長母音の場合、形成される略語は(12)のパターンを取ることが最も多いという。次いで(13)のパターンを取ることが多いが、このパターンを取るのは、(13)の具体例にあるとおり、長音を含む後部要素の冒頭2モーラそれ自体で意味を持っている場合に限られるという（「カー(car)」 「中^{ちゅう}」など）。そして、(14)のパターンを取る場合は珍しいという。

7.5. 節でも述べるが、本節の議論で重要なのは、後部要素の2モーラ目が長母音で、なおかつ3モーラ目が撥音の場合である。Sexy Zoneの後部要素「ゾーン」は、2モーラ目が長母音で3モーラ目が撥音だからである。このような場合は(12)のパターンをとらず、(14)のようなパターンに従うことが指摘されている（桑本裕二1998）。すなわち、後部要素の2モーラ目が長母音であって、なおかつ3モーラ目が撥音の場合には、後部要素の2モーラ目に存在する長音が消え、後部要素の3モーラ目に当たる撥音が略語構成要素として採用され、「2モーラ+2モーラ」の略語が形成されることが原則である、ということである。具体的には、(15)のようなパターンをとる。

- (15) ストコン<ストロベリーコーンズ

「前部要素+後部要素」からなる複合語の略語形成に長母音が関与する場合は、(12)から(15)に示したように複雑な様相を見せるが、桑本 (1998) は、この様相を最適性理論³⁷⁾ の枠組みを利用し、(16)のように一般化している^{38) 39)}。

- (16) ①形成される略語の前部要素部分が1モーラであってはいけない。

- ②形成される略語の音節数は、略語の構成要素となる元の語の前部要素、もしくは後部要素のいずれかの音節数と一致しなければならない。
- ③形成される略語に、長母音が含まれてはならない。
- ④形成される略語のモーラ数、もしくは分節音数は、略語の構成要素となる元の語の前部要素、もしくは後部要素のいずれかのモーラ数、もしくは分節音数と一致しなければならない。

桑本（1998）が提示した(16)の制約は、数字が小さいものほど、優先されるべきとされる。たとえば、「ゲームセンター」は「ゲーセン」と略されるが、形成された略語に長母音（「ゲー」）が含まれる。これは、(16)の③の制約に違反する。しかし、長母音を含まない「ゲセン」であると、形成された略語の前部要素が「ゲ」という1モーラになり、(16)の①の制約に違反する。最適性理論においては、上位にランキングされる制約ほど、守られなければいけないので、「ゲームセンター」の略語の場合、(16)の③の制約よりも(16)の①の制約が優先され、(16)の①の制約に違反しない「ゲーセン」という略語が形成されるのである⁴⁰⁾。その他、詳細は、桑本（1998）を参照にされたい。

7.5. アイドルグループSexy Zoneの略語はどのようなものになるのか

Sexy Zoneから略語が形成される場合、7.3. 節で説明した日本の芸能人の名前の省略方法である「前部要素と後部要素の語頭の2モーラを結合し、全体として4モーラの略語を形成する」に従い、「セクゾー」という形態が現れることが予想される。ただ、Sexy Zoneの後部要素である「Zone（ゾーン）」は、2モーラ目が長音であるため、7.4. 節で述べた(12)のように、「セクゾ」という長母音が削除された形態が現れることも予想される。さらには、Sexy Zoneの後部要素である「Zone（ゾーン）」は、2モーラ目が長母音、3モーラ目が発音であるため、7.4. 節で述べた(15)のように、長母音が削除され直後の撥音が残る「セクゾン」という形態が現れることも予想される。

それでは、Sexy Zoneの略語として、最も「適切」な形態は、「セクゾー」「セクゾ」「セクゾン」の一体どれなのであろうか。7.4. 節で述べた(16)の最適性理論の枠組みで考えると、以下の(17)ようになる。

(17) セクゾー：(16)の③に違反（「ゾー」という長母音の存在）

セクゾ ：制約違反なし

セクゾン：(16)の④に違反（元の語の前部要素「セクシー」の分節音数は6、後部要素「ゾーン」分節音数は3、略語「セクゾン」の分節音数は7）

実際には、既に7.1. 節で見たように、「セクゾー」は使用されず「セクゾ」と「セクゾン」とが併用されているようである。しかし、どちらが多数派なのか、どのようにこのような併用状況（「ゆれ」）が形成されたのかは、詳らかではない。そこで、次節以降、Twitterを対象に用例検索を行い、これらの問題点について明らかにしていく。

7.6. 検索方法

今回、利用するのはTwitterのツイート検索エンジン「TOPSY」である。詳細は既に6. 節で説明したが、今回は、6.2. 節で説明した手法、すなわち、ツイートされた日時を指定して検索を行った。

繰り返しになるが、Sexy Zoneは2011年9月29日にグループの結成が発表された。そこで、本稿では、グループ結成以降の2011年9月以降、1ヶ月毎に日時指定を行い、「セクゾ」「セクゾン」「セクゾー」について検索を実行して、ヒット数を集計した。検索の際には、検索対象を「ツイート」に限定した。一方、検索対象言語については限定を掛けなかった。

なお、「セクゾン」「セクゾー」の検索結果に「セクゾ」の使用例も含まれてしまうという懸念があったが、実際には「セクゾン」「セクゾー」の検索結果の実例の中に「セクゾ」が含まれていなかったことを、念のために付記しておく。

7.7. 結果

7.5. 節で述べたような方法で検索を掛けたと、以下の表3のような結果となった。

表3からは、以下のようなことがわかる。

- (18) ・やはり「セクゾー」はほとんど使用されていない。
 ・当初は「セクゾン」の使用が優勢であったが、現在では「セクゾ」の使用のほうが優勢である。
 ・「セクゾ」の使用率（「セクゾ率」）の変遷を観察すると、2011年9月から2012年7月までは上限を30%台の範囲として何回か増減を繰り返しているが、2012年7月以降は順調に「セクゾ」の使用が伸びる。

(18)の結果を踏まえ、「セクゾ」の使用率をグラフ化すると、以下の図8のようになる。

なお、「セクゾン」、ならびに「セクゾ」の初出は共に2011年9月29日（グループ結成が発表された日）であった。

表3 Sexy Zoneの略語の使用実態の推移

	2011年9月	2011年10月	2011年11月	2011年12月	2012年1月	2012年2月	2012年3月	2012年4月
セクゾ	2	22	81	188	312	713	614	566
セクゾン	47	129	1,293	1,172	597	1,538	4,216	3,983
セクゾー	0	1	0	0	0	0	0	1
セクゾ率	4.1%	14.5%	5.9%	13.8%	34.3%	31.7%	12.7%	12.4%

	2012年5月	2012年6月	2012年7月	2012年8月	2012年9月	2012年10月	2012年11月
セクゾ	1,201	571	650	2,155	5,671	9,187	10,409
セクゾン	2,573	1,368	2,673	2,605	4,585	7,135	7,351
セクゾー	0	0	0	0	2	14	9
セクゾ率	31.8%	29.4%	19.6%	45.3%	55.3%	56.2%	58.6%

	2012年12月	2013年1月	2013年2月	2013年3月	2013年4月	2013年5月	2013年6月
セクゾ	15,053	44,568	34,182	31,735	57,126	78,645	45,629
セクゾン	4,061	12,135	5,418	9,954	6,035	9,565	5,549
セクゾー	10	39	10	18	12	35	34
セクゾ率	78.7%	78.5%	86.3%	76.1%	90.4%	89.1%	89.1%

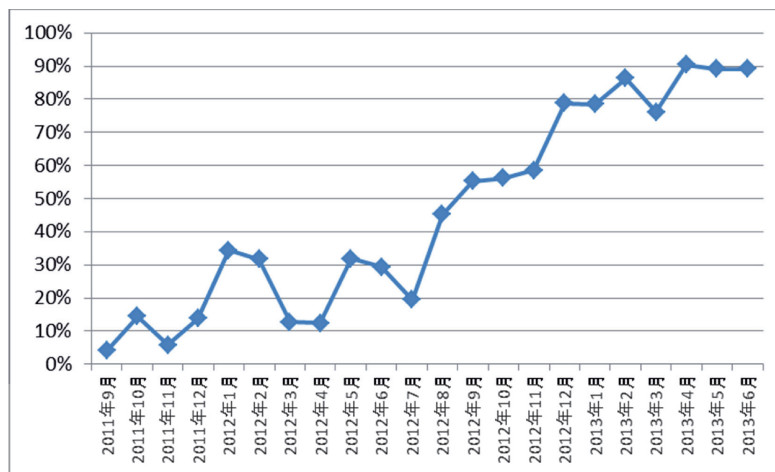


図8 「セクゾ」の使用率の推移

7. 8. 考察

表3、そして、図8の結果は、少なくとも筆者にとっては非常に興味深い。

まず、(17)で示したSexy Zoneの略語として予想される三つの形態のうち、最適性理論に基づく分析で、上位の制約に違反している「セクゾー」はほとんど使用されていないということが、計量的にも裏付けられたからである。また、当初は最適性理論に基づく分析で制約違反がある「セクゾン」という略語の使用主流だったものが、わずか2年弱の間に、最適性理論に基づく分析で制約違反のない「セクゾ」という略語の使用が主

流になっているからである。最適性理論に基づく分析で制約違反のある形式の使用から制約違反のない形式への使用への移行が、2年弱という短期間に、しかも明確な形で得られたのは貴重な例であるように思える。

しかも、図8に見える「セクゾン」から「セクゾ」への移行の様相が、言語変化研究における「Sカーブ」的な様相を示しているのも、非常に興味深い。

言語変化研究における「Sカーブ」とは、図9のようなものである（井上史雄2008より引用）。言語変化の際、新しい形式の普及は、図9のように、「変化の萌芽期には進行速度が遅いが、しばらくたつと加速して急激に進み、やがて終息期には再び緩やかになる」（日比谷潤子2009）パターンを描くとされている⁴¹⁾。

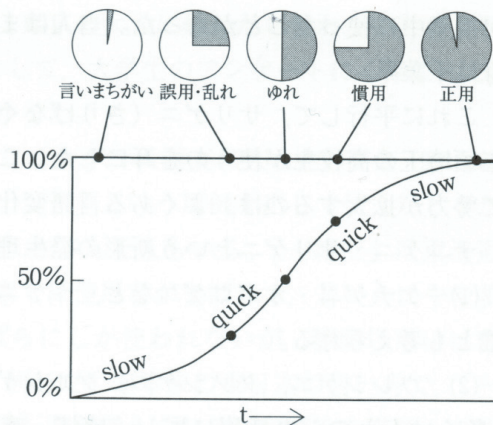


図9 言語変化におけるSカーブの模式図（井上2008より）

図8を見ると、「セクゾン」から「セクゾ」への変化も、緩やかながら「Sカーブ」を描いていることが読み取れる。すなわち、2012年7月までは増減を繰り返しつつも、「セクゾ」の使用率が爆発的に拡大することはなかった。しかし、2012年7月以降、「セクゾ」の使用率は爆発的に拡大し、半年強で「セクゾ」の使用率は90%弱に到達した（2013年2月）。その後は、2013年3月を除いて、「セクゾ」の使用率は90%程度で安定している。つまり、2012年7月までは変化の萌芽期で変化の速度は遅かったものの、2012年8月から2013年2月にかけては変化の加速期に入った後、2013年2月以降は変化の終息期に入り、変化の速度が緩やかになったと言える。

従来の言語変化における「Sカーブ」に言及した研究は数十年のスパンで分析したものが多い⁴²⁾が、わずか2年弱の期間でも「Sカーブ」を描く言語変化の例が見いだせたという点で、本稿の結果は、意義のあるものではないかと考えられる。

また、7.1. 節で述べたことも関連するが「Yahoo!知恵袋」におけるSexy Zoneの略語の問題の初出（2012年1月26日）が、「セクゾ」の使用率が初めて30%を超えた時期（2012年1月）であるというのも、非常に興味深い。というのも、言語変化の「Sカーブ」の議論に関連して、井上（2008）に、新しい形式の使用が「4分の1くらいの人に広がると「誤用・乱れ」として騒がれる」という指摘があるからである。Sexy Zoneの略語の問題も、まさに、新しい形式「セクゾ」の使用率が、Twitter上限定ではあるが25%を初めて超えた時点で、「セクゾン」か「セクゾ」かという議論が観察されるようになった。これは、井上（2008）の指摘とも呼応するものであり、非常に興味深い事象と考えられる。

以上の議論を踏まえると、「セクゾン」から「セクゾ」への移行は、最適性理論に基づく分析で制約違反のある形態から制約違反のない形態への言語変化と言うことができそうである。「セクゾン」から「セクゾ」への以降が言語変化と考えられるということは、Sexy Zoneというアイドルグループが、今後、芸能界に存在し続ける限りは、「セクゾ」という略語が使用されることが優勢であることが予想されるということでもある。

なお、今回の分析結果は、当該略語の使用者の感覚と照らしあわせても興味深い。7.1. 節で紹介した

「Yahoo!知恵袋」への質問（Sexy Zoneの略語は「セクゾン」か「セクゾ」か）の回答の中に、「私は最初「セクゾン」と言っていたが、今では「セクゾ」。後者のほうが言いやすい」というものが複数存在した。「言いやすさ」という主観的な感覚が最適性理論に基づく分析で制約違反の有無とも関わっているのかどうかは俄に判断できないが、最適性理論に基づく分析で制約違反がない形態のほうが当該言語において許容度が高いと考えられるため、最適性理論に基づく分析で制約違反がない「セクゾ」のほうが言いやすいとする声が複数あったことは、注目に値しよう⁴³⁾。

以上、本節では、Sexy Zoneの略語の使用の主流が「セクゾン」から「セクゾ」へと変化した理由について、主に最適性理論に基づく分析での制約違反の観点から考察したが、最後に他の観点からの考察についても簡単にやりたい。それは、「隠語」としての機能という観点である。

「Yahoo!知恵袋」への質問の回答の中に、「ファンは『セクゾ』と呼ぶ（一般人は『セクゾン』と呼ぶ）」というものが複数存在した。その回答があったのは2012年3月であり、Twitter上でもまだ「セクゾン」が優勢な時期であるが、そのような時期にあえて「多数派」とは異なる形態である「セクゾ」はファンが使うと意識されていたのは、「セクゾ」という形態に、ファンの間だけに通用するというある種の「隠語」のような性格を持ち合わせていた可能性がある。「隠語」を使うことの働きとして「仲間であることを確認し、結束を強める」（沖森ほか2011）ことが知られているが（柴田武1958の議論も参照）、一般人があまり使用していない「セクゾ」という略語をあえて使用することで、Sexy Zoneのファンとしてのアイデンティティを保持していた可能性もある（ただし、時間が経過するにつれ、当初は「隠語」的な使用であった「セクゾ」が多くの人々に流布し、「隠語」的性格を失った可能性が大きい）。

7.9. 残された課題

以上、本節では、Sexy Zoneの略語の問題について論じてきたが、最後に、残された課題についても、簡単に触れておく。

まず、なぜ、当初は「セクゾン」という、最適性理論に基づく分析で違反のある形式の使用が優勢だったのか、という問題である。この問題については、明確な答えを持ち合わせていない。そこで、現時点の筆者が抱いているいくつかの可能性を指摘しておく。

一つは、本分析で則った桑本（1998）の分析に改善の余地がある、という可能性である。註39で指摘した通り、桑本（1998）の分析では説明できない略語が複数存在する。ということは、桑本（1998）の分析には改善の余地があり、さらに妥当な分析を施した場合、Sexy Zoneの略語としては「セクゾ」より「セクゾン」のほうが適当だと導かれる可能性も否定はできない。ただし、現時点の筆者には、桑本（1998）の分析を再検討する能力はない。また、たとえ、「セクゾ」より「セクゾン」のほうが適切だという分析が可能であっても、その場合は、なぜ、「セクゾン」より「不適当」な「セクゾ」の使用が優勢になったのかという新たな問題が生じる。

第二の可能性は、「セクゾ」よりも「セクゾン」の方が元の語（Zone）が想起しやすい、といった要因があると考えられる。

窪菌（2002）は「「略語は元の単語を想起できる形でなくてはならない」という原理が働いて」といると指摘している。「Yahoo!知恵袋」には、「自分は『セクゾン』の方がわかりやすいので、そのように呼んでいる」という回答があった（http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1483482269。2013年6月30日閲覧）。その指摘を踏まえると、「ゾ」や「ゾー」から「ゾーン」という単語を想起するのは困難であるため、「セクゾ」ではなく、「セクゾン」が使用されていた、という可能性もあり得る。

すなわち、「セクゾン」は最適性理論に基づく分析では制約違反のある略語であるが、元の語が想起しやすいということもあり、Sexy Zoneがデビューした直後は、「セクゾン」という略語が使用される傾向にあった。そもそも、「セクゾン」は最適性理論に基づく分析では違反のある略語とはいえ、違反している制約のランキングは、(17)に示した通り、下位に属する。したがって、最適性理論に基づく分析では制約違反のある略語であるが、分かりやすさを優先し、「セクゾン」という略語が当初は好まれていた。しかし、Sexy Zoneが芸能界で活躍するに従ってグループとしての知名度が上がり、「セクゾ」であってもSexy Zoneというグループ名が復元できるようになったため、時間が経つにつれ、「セクゾ」の使用頻度が高くなったとい

う解釈である。この解釈は、7. 8. 節で紹介した、「セクゾはファンが使う、セクゾンはファン以外が使用する」という指摘にも矛盾しない。すなわち、グループのことをよく知っているファンであれば、「セクゾ」という略語あっても元のグループ名を簡単に復元できるため、「セクゾ」という略語を使用していた（厳密には、使用するというイメージが持たれていた）という可能性も考えられる⁴⁴⁾。

第三の可能性は、マスメディアの影響である。「Yahoo!知恵袋」には、「雑誌には『セクゾン』と書いてある」「Sexy Zoneのメンバーが『セクゾン』と言っていた」ため、自分も「セクゾン」と言うとの回答が、複数存在する。すなわち、メディアやSexy Zoneが「セクゾン」という略語を使用していたために、ファンや一般人のあいだでも、最適性理論に基づく分析では制約違反のある略語ではあるが、「セクゾン」という略語を使用していた、という可能性である。

この可能性は、言語学者の柴田武氏が提唱した「言語ボス」という考え方を援用すれば、うまく説明ができそうである。「言語ボス」とは、ある集団内のみに通用している語を「発明するにしろ、ほかの集団から移すにしろ、その集団で初めていい出して、それが広まるようなときに、そのいい出し手」（柴田武1956）のことを指す。そして、「あることばを口に出しても、その人が言語ボスでないと、それはその人だけで終わってしまいますが、その同じ言葉が、言語ボスの口から出ると、たちまち集団のなかに広まることもある」（柴田武1958）という側面を持つ。

柴田氏の言う「言語ボス」を本稿の分析に関連付けるならば、以下のようになろう。メディアやSexy ZoneのメンバーがSexy Zoneに関心を持つ人々の間で「言語ボス」的な役割を果たし、Sexy Zoneの略語は、当初、彼らが使用する「セクゾン」が一般にも流布した⁴⁵⁾。しかし、時が経るにつれ、メディアやSexy Zoneのメンバーの「言語ボス」的な役割が低下し、最適性理論に基づく分析では制約違反のある「セクゾン」が使用されなくなり、替わって最適性理論に基づく分析で制約違反のない「セクゾ」が使用されるようになった、という可能性も考えられる⁴⁶⁾。

以上、当初は「セクゾン」という、最適性理論に基づく分析で違反のある形式の使用が優勢だったのかという問題について、いくつかの可能性を指摘したが、どの可能性が正鵠を射ているのか（あるいは他の可能性があるのか）については、今後の課題にしたい。

今回の分析におけるその他の問題点としては、発言主の属性を考慮に入れた分析はできなかったということがあげられる。「Yahoo!知恵袋」には、既に繰り返し紹介した「セクゾはファンが使う、セクゾンはファン以外が使用する」という指摘の他に、「たぶんな情報で申し訳ないのですが、関西では「セクゾン」関東では「セクゾ」と言われています」と地域差を指摘する回答も会った（http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1297143867。平成25年6月30日閲覧）。それらの指摘が正鵠を射ているかを確認するためには、発言主の属性（Sexy Zoneのファンであるか否かや、発言主の出身地）を考慮した分析も行う必要もあると考えている。幸いなことにTwitterには「プロフィール欄」があり、そこには発言主がSexy Zoneのファンであることが明記されている場合や、発言主の出身地が記載されている場合もあるので、理論上は、発言主の属性を考慮に入れた考察も可能ではある。しかし、現時点のツイート検索エンジンでは、Twitterの発言と発言主のプロフィールを連携して検索することができないため（発言とプロフィールをそれぞれ別個に検索することは可能。岡島2011の議論も参照）、現在の筆者の技術では発言主の属性を考慮に入れた考察は不可能であった。この観点からの分析は今後の検索技術の進展に期待することにしたい⁴⁷⁾。

8. おわりに

以上、本稿では、Twitterを利用した日本語研究の可能性を模索してきた。7. 節で詳述した通り、Twitterを利用した日本語研究には、一定の可能性が見いだせたと信じている。

最後に、Twitterの言語資料的性格について、述べておきたい。

4. 節で紹介した五味ほか（2011）は、「Twitter で用いられている言葉は、従来の話し言葉とも、書き言葉とも異なる特徴を持った言葉であると考えられる」と指摘している。五味ほか（2011）は、その理由の一つとして、「何らかの「入力」を経ねばならないという事」をあげている。すなわち、「Twitter に投稿する時には、必ず何らかの入力装置キーボードもしくは携帯電話など）とかな漢字変換プログラムを用いねば

ならない」(五味ほか2011)が、漢字変換プログラムが対応していない／できない表記や現象については、当該現象の生起が抑制され、その結果、「Twitterにおいても、話し言葉において見られるような文法の揺らぎが見られるが、頻度は話し言葉の場合よりも低く、その現れ方は、話し言葉の場合とは様相を異にしている」(五味ほか2011)というわけである。

ただ、「何らかの「入力」を経ねばならないという事」というのは、Twitterに限った話ではない。FacebookやLINEといったTwitter以外のSNS(ソーシャル・ネットワーキング・サービス)⁴⁸⁾は勿論、メール、ブログ、ホームページ、さらにはなども、全て「何らかの「入力」を経ねばならない」ものである。そして、それらのレジスターは、五味ほか(2011)も指摘するとおり、「インターネット上において、書き言葉の場とも話し言葉の場とも異なった、新しい言葉の場、仮に名付けるなら「打ち言葉」とでもいえる言葉の場」と言え、どのような言語が運用されているのか(あるいは、運用されようとしているのか)、言語研究、あるいは日本語研究の側面からも、注目に値する。

もっとも、「打ち言葉」という概念は、言語研究者、日本語研究者のあいだでは、決して新しいものではない^{49), 50)}。「打ち言葉」が言語研究者に使用されるようになったのは、インターネットや携帯電話の「黎明期」といえる1990年代後半まで遡ることができる。たとえば、読売新聞の連載「身のまわり言葉のまわり」の第4回「広がる「パソコン発」」(1998年11月13日・朝刊)には、「東海大学の朝尾幸次郎教授(応用言語学)は「書き言葉でも話し言葉でもない、『打ち言葉』とでもいうべきネットワーク上の文体、文化が育っていると見る」という一節がある。また、「打ち言葉」という単語は使用されていないが、既に1996年の時点で、朝尾幸次郎氏は、「パソコン通信、インターネットで交わされるメール、また会議室、ニュースグループに書き込まれる文章」は「文字を媒介とはしながら、それが表すのは話しことば」であり「究極の言文一致体」であると指摘している(朝尾幸次郎1996。朝尾1996は、この種の文体を「ネットワーク文体」と呼んでいる)。

その後、メール、特に携帯電話のメールについては、携帯電話が普及し始めた時期から日本語研究者の注目を集めてきた⁵¹⁾し、その他の「打ち言葉」についても、記述、分析が試みられて、日本語研究の資料としての性格の一端が明らかになっている⁵²⁾。

以上のような事情を踏まえると、「打ち言葉」という概念は、2013年時点では、もはや新しいものであるとは言い難いと思われる。しかし、Twitter(やその他のSNS)における言語についての詳細な記述、分析は、現在のところ、管見の範囲では見つけることができなかった。

2. 節で紹介した通り、Twitterには140字という字数制限が存在する。これは他の「打ち言葉」(メール、ブログ、ホームページ)には見られない大きな特徴である。そのような字数制限が存在するTwitterがその他の「打ち言葉」とどのように違うのか(あるいは共通点が見いだせるのか)といった、言語資料的性格の議論は、今後の課題としたい。

最後になるが、今回はTwitterを利用した言語研究の可能性を模索することが主目的であり、7. 節での分析、考察は、不十分な点が多々あることをお赦しいただきたい。特に、最適性理論についての理解が足りないため、7. 節で議論については、筆者の誤解によって、見当違いな結論を導き出しているかもしれない。また、7. 節の考察においては、種々の可能性を指摘するのみであって、その妥当性を検証するには至らなかった点も、今後の課題としたい。さらには、筆者自身はTwitterを利用していないため、6. 節で紹介したTOPSYの使い方にも、誤解がある可能性がある。本稿には様々な問題があることは承知しつつ、研究の可能性を模索するという意図で、あえて論文化した次第である。読者からの忌憚のないご叱正、ご批判、ご指導をいただければ幸甚である。

付記

本稿は、2013年2月2日に新潟大学教育学部にて開催された新潟大学教育学部国語国文学会平成24年度研究大会で行った講演「インターネットを利用して現代日本語を調べる―母語を見つめなおす手法としての提案―」で触れた内容の一部を発展させたものである。

付記2

本稿の校正作業中、6. 節で紹介したツイート検索エンジンTOPSYの仕様が変更されたことに気がつい

た（2013年9月13日確認）。詳細は別稿（本紀要の次号に投稿予定）で論じるが、本稿に関係する変更としては、日付を指定して検索した際に検索結果件数が表示されなくなったという点があげられる。これは、6.2.節で詳述した方法が使えなくなったということ、すなわち、本稿の7.節で行ったような分析が不可能になってしまったということを意味する。つまり、本稿で提案したTwitterを利用した新語・流行語研究の可能性が暗礁に乗り上げたということである。

今回のTOPSYの仕様の変更は、筆者にとっては大きな衝撃ではあると同時に痛手であるが、言語研究用に設計、提供されたサービスではないだけに、このような突然の変更（改悪）や、サービスの提供停止といった事態が生じることも覚悟しなければならぬと痛感した次第である（註10で言及した事例や田野村2012での議論も参照のこと）

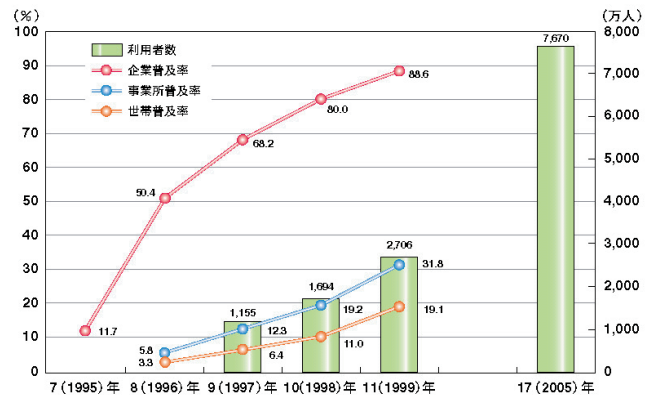
註

1) 「内省」とは、典型的には、「研究者自身が、自分のことばについて、どうか、どんな意味か、どんな場合に使うか、どんな感じがするかなどを内省して、その結果を研究資料とする」（徳川宗賢1988。傍点は筆者）方法である。しかし、筆者は、「内省」の概念をもう少し幅広く捉え、「ことばの使用者が研究者の質問などを通してその意図を推量しながら、いわば研究者にかかわって話者自身のことばについて内省を行う方法」（徳川宗賢1986。傍点は筆者）も、「内省」によるデータ収集の手法に含めたいと考えている。すなわち、筆者は、言語研究において母語話者の内省を利用する手法には、①研究者が自分の母語について内省を行う場合と、②研究者が母語としない言語について、その言語を母語とする話者に内省結果を尋ねる場合、とに下位分類できるという立場をとる。

2) 岡田（2013）については、別途、改めて公表することを考えている。

3) 日本でインターネットの商業利用が開始されたのは1993年のことである（『平成11年版 通信白書』<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h11/html/B1Z20000.htm>。2013年7月1日閲覧）。一方、日本語研究者が行った、インターネットを利用して用例を収集するという方法の提案は、岡島昭浩（1997）や田野村忠温（2000）が、比較的早い段階のものとしてあげることができる。また、言語学の啓蒙雑誌『月刊言語』（大修館書店。現在休刊中）では、1998年から2001年にかけて「インターネット言語学情報」という連載があった。以上のような事実を踏まえると、日本語・言語研究者がインターネットの存在を強く意識し始めたのは1990年代の後半からだと思われる。

図表 我が国におけるインターネットの普及状況



付図 日本のインターネット普及状況
（『平成12年度通信白書』より）

なお、『平成12年度通信白書』（<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h12/html/C1040000.htm>。2013年7月1日閲覧）によると、1990年代後半におけるインターネット普及率は以下の付図のとおりである。インターネットの企業普及率が50%を超えたのと軌を一にして、日本語研究者、言語研究者がインターネットに注目し始めたようで、興味深い。

4) たとえば、田野村忠温（2008）は、インターネット検索エンジンYahoo!が検索の対象としている「データ量は実に新聞記事のデータ3万年分以上に相当することになる」と推定している。

5) 既に2003年時点で、以下のような記述が見られる（田中ゆかり2003）。

あれっと思った事象に出会ったとき、または、何か調べなくてはならないことが発生した時、検索エンジンを使って、とりあえずネットで検索することが半ば習慣化してから、どのぐらい経つだろう

か？ 特殊な専門家を除けば、そのような習慣が芽生えてから、それほど時間が経っているわけではないのに、その手早さ感と一定の満足感から、基本動作化している人は、少なくないのではないだろうか。

- 6) 新聞社のデータベースはいろいろな形で利用ができるが、たとえば、新潟大学のネットワークから新潟大学附属図書館のホームページ (http://www.lib.niigata-u.ac.jp/searchworks/he_list.html) にアクセスすれば、朝日新聞 (「朝日新聞 聞蔵II ビジュアル for Libraries」)、読売新聞 (「ヨミダス文書館」)、新潟日報 (「新潟日報データベース」) のデータベースを、それぞれ無料で利用することができる。また、公立の図書館 (新潟県立図書館など) においても、各新聞社のデータベースを無料で利用することができる。
- 7) 大学共同利用機関法人人間文化研究機構国立国語研究所と文部科学省科学研究費特定領域研究「日本語コーパス」プロジェクトが共同で開発した『現代日本語書き言葉均衡コーパス』(BCCWJ:Balanced Corpus of Contemporary Written Japanese) のデータを検索できるサイトである。
- 8) 註7で説明した『現代日本語書き言葉均衡コーパス』を利用して、特に名詞や動詞などの内容語の共起関係 (コロケーション研究) や文法的振る舞いを網羅的に表示できるサイトである。
- 9) 荻野ほか (2005) では、インターネットを利用した日本語研究が持つ教育的効果が論じられている。本稿は、荻野ほか (2005) の影響を多分に受け、Twitterという学生にとっては身近なメディアを対象に、アイドルグループ名といった「卑近」な例であっても、十分に興味深い新知見を得ることができることを示し、日本語研究や言語研究に注目する読者が一人でも増えることを願って、執筆した次第である。
- 10) 岡島 (2011) で紹介されているツイート検索エンジンは、以下の四つである。なお、2013年6月現在、岡島 (2011) が記載しているURLから変更になっているものもあったため、参考までに2013年6月現在のURLも付記しておく。

- ・ Twitter Search (Twitter公式の検索サービス) : <https://twitter.com/search-home>
- ・ googletter : <http://www.googletter.com/>
- ・ tweet search : <http://twittell.net/search.php>
- ・ BIGLOBE つぶやき検索 : 2013年5月7日14時でサービス終了

- 11) 五段活用動詞「違う」を、形容詞型の活用をさせる現象を指す。具体的には、「違わない」(未然形) を「チガクナイ」、「違った」(連用形) を「チガカッタ」、「違う」(終止形・連体形) を「チゲー」という形で使用する現象である。なお、「チゲー」は、動詞「違う」の終止形・連体形「違う (ちがう)」から導き出される「チガイ」という形容詞型の終止形・連体形の俗語形と考えられる (「チゲー」については、井上史雄1988での議論も参照)。ただし、動詞「違う」の終止形・連体形「違う (ちがう)」から導き出される「チガイ」という形容詞型の終止形・連体形は、実際にはほとんど使われていないようである (北本洋子1995も参照)。それは、「終止形「違い」は動詞連用形およびそれから転じた名詞と同形であるために、形容詞終止形としての用法に違和感が強い」(北本1995) ためだと考えられる。したがって、動詞「違う」の終止形・連体形「違う (ちがう)」から導き出される「チガイ」という形容詞型の終止形・連体形はほとんど使用されず、その俗語形「チゲー」がもっぱら使用されていると考えられるのである。
- 12) 千葉大学の学生50名 (男子22名、女子28名) を対象に、2010年12月に実施された。
- 13) 石井 (2011) は、2種類のアンケート調査を行っている。一つは「「違う」の動詞型活用と形容詞型活用を含んだ口語文10例を提示し」違和感があるか/許容できるかを確認した「調査Ⅰ」、もう一つは「「違う」の形容詞型活用を含む、【調査Ⅰ】とは別の口語文8例を提示し」聞いたことがあるか/自分でも使うかを確認した「調査Ⅱ」である。
- 14) 五味ほか (2011) の「形容詞化率」とは、「形容詞活用例を動詞活用例と形容詞活用例の和で割った」値のことである。
- 15) たとえば、五味ほか (2011) では、「チガウカッタ」と「チガカッタ」双方とも形容詞型活用例と見なしている。一方、石井 (2011) では、形容詞型活用例としては「チガカッタ」しか認めていない。
- 16) レジスター (言語使用域) とは、社会言語学の用語で、「使用領域からみたことばのパラエティー」(真田信治2005) のことである。五味ほか (2011) はTwitterにおける言語運用を調査したのに対し、石井 (2011) は日常会話を念頭に置いた言語意識の調査 (口語文を提示した調査のため) と考えられ、双方の

研究では扱っているレジスターが異なると言える。

- 17) 一方で、検索エンジンを利用せずに独自でデータを収集する方法を採用したゆえに、五味ほか(2011)は、使用者の属性(「投稿者がこれまでに投稿したツイートの総数」「投稿者がTwitterに登録した年」「投稿者の総ツイート数をTwitterに登録してからの日数で割った数」)を考慮に入れた分析や、形態素解析を施したデータに対する分析が可能になっている。
- 18) 通時的な言語研究といえば、長期間(少なくとも数十年以上)にわたってデータを収集しなければならないように思われる。しかし、言語の運用の動態はダイナミックであり、現象によっては、現代に限られた期間(数年間という限定的な期間)であっても、通時的な観点からの言語研究は可能であると、筆者は考えている。そのような研究の実践として、本稿の7.節で述べたような試みを行った次第である。
- 19) paruparu氏のブログ「ありのごとくあつまりて」の2011年2月14日付の記事「Twitterのツイート検索なら断然 TOPSYが凄い!」(<http://arinogotokuatumarite.blog19.fc2.com/blog-entry-188.html>). 2013年6月7日閲覧)によると、TOPSYでは、以下のような検索も可能であるとのことである。
 1. そのページにリンクの張られたツイートを検索
 2. 特定のサイトにリンクの張られたツイートを検索
 3. 特定のユーザーのリンク入りツイートを検索
 4. 特定のユーザーがツイートした画像を検索
- 20) 同じ単語を検索しても、検索する日時によって検索結果が大きく異なることを指す。
- 21) インターネットの検索エンジンにおいては、「検索窓」に入力する単語の順番が異なっても、論理的に同等価であれば、同じ検索結果が出ることが期待される。しかし、田野村(2008)によると、Googleの検索では、論理的には同条件であっても、検索結果が異なることが報告されている。たとえば、Googleにおいて、「犬 猫」という条件で検索した場合と「猫 犬」という条件で検索した場合では(両者とも「犬」という単語と「猫」という単語の双方を含む文書を検索することを意味する)、検索結果が大きく異なるという(前者が1,790,000件。後者が758,000件)。したがって、Googleを使った検索は「論理的整合性」の側面で問題点があると言えるのである。
- 22) ただし、厳密な計画を立てて実施した調査の結果ではないので、このように断言することには、慎重にならなければならないことは承知している。この件については、別途、本格的な調査を行い、その結果を公表したい。
- 23) TOPSYによる公式の説明を発見できなかったので各検索対象の詳細は不明である。ただ、「LINKS」「写真」「動画」のみを選択して実際に検索を行うと、それぞれおおよそ、以下のようなツイートのみがヒットするため、日本語研究の用例収集の手段としてTOPSYを利用する際には、「ツイート」を選択するのがよさそうに思われる。

LINKS: ツイートの中に検索語とリンク先のURLの双方が含まれている場合のみ

写真: 写真をアップロードされたツイートで、なおかつ本文に検索語が含まれている場合のみ

動画: 動画がアップロードされたツイートで、なおかつ本文に検索語が含まれている場合のみ

なお、「エキスパート」と「トレンド」は、以下のとおり、検索語のヒット件数を表示するものではないため、日本語の用例収集の手段として使うことは難しいであろう。

エキスパート: 検索語について「最も言及(発言)」している人をエキスパートとしてピックアップする(<http://www.chrome-life.com/google-plus/3295/>). 2013年6月8日閲覧)。

トレンド: 「現在Twitter上を流れているツイートの中から多く話題に上っているキーワードをリアルタイムに抽出して表示する」(<http://www.sophia-it.com/content/Twitter%E3%83%88%E3%83%AC%E3%83%B3%E3%83%89>). 2013年6月8日閲覧)。

最後に、参考までに、TOPSYの検索では、「すべて」を選択して検索した結果の方が「ツイート」を選択した際の結果よりも著しく少なくなるという、「不思議」な現象が起きるということを付記しておく。

- 24) 検索語が日本語の場合、検索結果の画面では検索語が文字化けして表示される。ブラウザの文字コードを調整しても、この文字化けは改善しない。これは、ツイートされた日時を指定して検索した場合も同様である(図5.参照)。

- 25) インターネットの検索エンジンを利用して用例収集をすることの可能性を模索した種々の先行研究において、インターネット検索エンジンのシステムがよくわからないという点があげられているが(1. 節で言及したインターネットを利用して用例を収集する方法の問題点を指摘した諸文献も参照)、この問題点はツイート検索エンジンであるTOPSYにも当てはまりそうである。
- 26) いくつかの語を検索してこの問題を検討してみたが、リツイートまで検索結果に含まれている場合が確認された一方で、リツイートが検索結果に含まれていないと思われる例も存在した。
- 27) グループ結成発表日については、たとえば、「ジャニーズ新ユニット「Sexy Zone」最年少デビュー」(報知新聞社・<http://hochi.yomiuri.co.jp/entertainment/news/20110930-OHT1T00015.htm?from=yol>・2013年5月2日閲覧)を参照されたい。
- 28) グループの詳細については、以下のサイトを参照されたい(いずれも、2013年6月30日閲覧)。
 Sexy Zone オフィシャルウェブサイト(ポニーキャニオン運営): <http://sexyzone.ponycanyon.co.jp/>
 Sexy Zone オフィシャルウェブサイト(ジャニーズ事務所運営): <http://www.johnnys-net.jp/page?id=artistTop&artist=24>
- 29) 日本語における「複合語」とは、「二つの造語成分が、いずれも、語基であり、それらが複合的にむすびついて語および語の意味をくみだしているものである」(石井正彦1989)。なお、語基とは、「実質的・語彙的な意味を表し、語の中核的な部分になう造語成分」(石井1989)のことである。
 以上の定義を踏まえて、具体例をあげて説明すると、「新潟大学」「教育学部」「国際空港」は、いずれも、それぞれ語基である「新潟」「大学」「教育」「学部」「国際」「空港」から構成される複合語である、ということになる(ただし、「国際」は単独では語になることができないという点で、ここで例をあげたその他の語基と性質を異にする)。
 複合語と類似するものとして、「派生語」がある。派生語とは、語基に接辞(「形式的、補助的な意味を表し、語基に付いて、語のあり方(存在形式)を示すか、語基に形式的な限定を加える造語成分」のこと。石井1989)がついたものである(接辞には、語基に前接する「接頭辞」と、語基に後接する「接尾辞」がある)。たとえば、「無意識」「高さ」といった語が派生語である(「無意識」は語基「意識」に接辞「無」が、「高さ」は語基「高」に接辞「さ」が、それぞれ付いて形成された派生語である)。
 複合語、派生語は、ともに二つ(以上)の造語成分が結合して形成されたものであるため、双方をあわせて「合成語」と呼ぶ。ただし、複合語と派生語は、上述のような違いがあることに留意されたい。
 なお、石井(1989)は、合成語は、複合語、派生語、疊語の3種類から構成されるという立場をとっているが、本稿では、議論を簡略化するために、「疊語」は複合語の一種とみなす立場にたつことにする。疊語とは、「人々」「我々」「深々」といったように、「同一の語基がくりかえされたもの」(石井1989)であるが、これは石井(1989)も述べているように「複合語と同様に、語基二つのむすびつき」であるゆえ、疊語は複合語の一種と見なすことも可能であるからである(このような立場は、本稿独自のものではない。たとえば、近年出版された日本語語彙論の入門書である沖森卓也ほか(2011)では、合成語は複合語と派生語から構成されるという立場をとっている)。
- 30) サックス奏者。詳細は<http://www.sadao.com/>(オフィシャルウェブサイト。2013年6月30日閲覧)を参照。
- 31) 「モーラ」とは、音のまとまりの単位の一つである。また、直後に、「音節」と「分節音」という概念も出てくるが、隣接する概念であるので、ここで併せて説明しておく。詳細な議論は、たとえば斎藤純男(2007)などをご参照いただきたいが、基本的には、以下のような形にまとめられる。
 ・音節: 母音を中心とした音のまとまり。発音した際、あるいは知覚した際、切れ目が感じられない音の連続のこと。
 ・モーラ: 長さを基準とした時間的単位。日本語においては1文字が1モーラに相当する(ただし、拗音の場合は、かな2文字が1モーラとなる)。
 ・分節音: 音声の最小単位のこと。一つ一つの母音、子音を指す。
 各单位で以下の語をカウントした場合は、以下のようになる。

付表 音節数とモーラ数と分節音数の関係

漢字表記	読み	音節数	モーラ数	分節音数
新潟大学	ニーガタダイガク	5	8	13
小学校	ショーガッコウ	3	6	6
新大	シンダイ	2	4	6
切手	キッテ	2	3	4

- 32) 1988年結成, 1989年にメジャーデビューをしたバンド。以上, 「オリコン芸能人事典」の「DREAMS COME TRUE」の項より (<http://www.oricon.co.jp/prof/artist/7057/profile/history/>。2013年5月2日閲覧)。
- 33) 1983年から1996年まで活躍したガールズバンド。なお, 2012年に1年間限定で再結成された。以上, 「オリコン芸能人事典」の「プリンセス・プリンセス」の項より (<http://www.oricon.co.jp/prof/artist/19456/profile/full/>。2013年5月2日閲覧)。
- 34) 1931年生, 1997年没の俳優。以上, 「kotobank」の「勝新太郎」の項より (<http://kotobank.jp/word/%E5%8B%9D%E6%96%B0%E5%A4%AA%E9%83%8E>。2013年5月2日閲覧)。
- 35) 1904年生, 1970年没の喜劇俳優。以上, 「kotobank」の「榎本健一」の項より (<http://kotobank.jp/word/%E6%A6%8E%E6%9C%AC%E5%81%A5%E4%B8%80>。2013年5月2日閲覧)。
- 36) 1903年生, 1980年没の映画俳優。以上, 「kotobank」の「嵐寛寿郎」の項より (<http://kotobank.jp/word/%E5%B5%90%E5%AF%9B%E5%AF%BF%E9%83%8E>。2013年5月2日閲覧)。
- 37) 最適性理論 (Optimality Theory) は, 1990年代初頭にAlan PrinceとPaul Smolenskyによって提案されたもので, 「言語分析の常識であった「規則 (rule)」・「派生 (derivation)」という概念を否定し, 「制約 (constraint)」だけで言語現象を説明しようとする理論」(窪園晴夫1996) である。具体的には, 「人間の認知システムの中に普遍文法が存在し, それは全ての言語に共通した制約の集合から成る」が, 「各制約は違反可能であり, 各言語においてどの制約を優先的に遵守するかは異なる」と考える (深澤はるか2005)。すなわち, 最適性理論では, 各個別言語の違いは, 制約の優先順位 (ランキング) の違いに起因すると考える。
- 38) 桑本 (1998) では, 最適性理論の枠組みに則った方式で記述されているが, 本稿では一般の読者にも分かりやすいよう, 桑本 (1998) の記述を筆者が書き改めた。
- 39) もっとも, (16)の一般化でも説明できない場合は複数存在する。たとえば, (13)や(14)であげた例は, すべて説明できないようである。また, 桑本 (1998) では, 以下の例が説明できないとしている。

スケボー<スケートボード
 ストパー<ストレートパーマ
 テレコ<テープレコーダー
 レスカ<レモンスカッシュ

ただし, インターネットで検索してみると, 「スケボー」には「スケボ」, 「ストレートパーマ」は「ストバ」という語末の長母音が短呼した例がいくつもヒットする。語末の長母音が短呼した「スケボ」や「ストバ」は(16)の一般化には反しない。そもそも, 語末の長母音は音韻論的にも音声学的にも短呼する場合がしばしば観察される (助川泰彦ほか1999, 窪園晴夫2000, 前川喜久雄2000なども参照) ため, 「スケートボード」の略語は「スケボー」と「スケボ」が, 「ストレートパーマ」の略語は「ストパー」と「ストバ」が, 共時的にはそれぞれ, 「ゆれ」という形で併用されているのが現実ではないかと思われる。

- 40) なお, 「ゲムセン」という略語が形成される可能性があるが, 「ゲムセン」は以下に示す通り, (16)の②の制約に違反する ((16)の②の制約に違反しないためには, 形成される略語は2音節でなければならない)。

ゲムセン (3音節) < ゲーム (2音節) センター (2音節)

すなわち, 「ゲセン」という略語は(16)の①に, 「ゲムセン」という略語は(16)の②に, 「ゲーセン」という略語は(16)の③に違反するが, 最適性理論の枠組みでは, ランキングが低い制約に違反するものが選択されるので, 「ゲーセン」という略語が選択されるのである。

- 41) これは, 言語変化に限った話ではない。井上史雄ほか (2009) によると, 「普及伝播の過程でS字カー

ブが観察されることについては、地理学・生物学などで先駆的な研究があり、普及学という形で一般化された」とのことである。

- 42) 実際、数十年のスパンにわたってデータを収集し、言語変化の様相を描き出すのは、なかなか困難である。そのため、言語変化の研究、特に社会言語学的な研究においては、「現時点での年齢における分布を、そのまま進行中の言語変化の反映と解釈する」ことも多い（松田謙次郎2004）。そのような中で、国立国語研究所が約20年おきに山形県鶴岡市で実施している調査は実時間に基づいた言語変化研究の貴重な実践例であるが、いわゆる「鶴岡調査」でも言語変化における「Sカーブ」が見いだせることが報告されている（たとえば、井上ほか2009参照）。
- 43) 筆者は、以前、「発音のしやすさ」「言いやすさ」の問題は単純に生理的な要因で説明できないのはいかと指摘したが（岡田祥平2006）、「発音のしやすさ」「言いやすさ」という意識と最適性理論に基づく分析での制約違反との間に、何らかの関係性が見いだせるのではないかと考えている。
- 44) ただし、この可能性の妥当性を証明するのはなかなか困難である。新潟大学教育学部において岡田が担当している授業「国語学講義Ⅱ」にて、「ゾ」／「ゾン」から想起できる外来語を記述してもらう簡単なアンケート調査を実施したが、「ゾ」よりも「ゾン」のほうで「ゾーン」が連想される確率が高いという結果は出なかった。なお、回答者からは「ゾ」もしくは「ゾン」が、どのような語において、もしくはどのような文脈において使われるのかによって、連想されるものが変わってくるので、回答しづらい」という感想があった。このコメントを踏まえると、「元の単語の想起しやすさ」を厳密に調査するには種々の要素を考慮しなければならないと考えられる。今後の課題としたい。
- 45) もっとも、この可能性を採用するならば、なぜ、メディアやSexy Zoneのメンバーが「セクゾン」という略語を使用したのかという問題を考える必要が出てこよう。現時点の筆者には明確な回答を持ち合わせていないが、Sexy Zoneが所属するジャニーズ事務所に1986年から公演を行なっているミュージカル「PLAYZONE（プレイゾーン）」が「プレゾン」と略されている（<http://ja.wikipedia.org/wiki/PLAYZONE>。2013年7月1日閲覧）ことも関係しているのではないかと考えている。すなわち、ジャニーズ事務所関係者（ファンも含む）では「ZONE」という語を「ゾン」と略するのが慣習となっており、それをSexy Zoneの略語にも適用した、ということである。
- なお、TOPSYを利用してツイートを日時の指定をせずに検索した結果、「プレゾン」は多数ヒットしたが（16,778件）、「プレゾ」はほとんどヒットしなかった（30件。いずれも2013年7月1日午前10時40分時点での検索結果）。したがって、PLAYZONEの「ZONE」の略し方はSexy Zoneの場合とは異なり、「ゾン」が「ゾ」への移行はなかったと考えられる（その理由は不明）。参考までに、最適性理論に基づく分析では「プレゾン」は(16)の②と④（モーラ数、分節音数）に違反している一方、「プレゾ」は(16)の②と④（ただし、分節音数のみ）に違反しているということを付記しておく。
- 46) メディアにおいて、Sexy Zoneがどのように略されるかについては、彼らを取り上げられる頻度が高いアイドル雑誌を中心に調べる必要があると思われるが、今回はそこまで手が回らなかった。この件については、今後の課題としたい。
- 47) ツイート検索エンジン独自に何らかのプログラムを構築できる能力があれば、現時点でも、発言主の属性を考慮に入れた分析が可能なのかもしれない。ただし、筆者はそのような能力を持ち合わせていない。
- 48) SNSの詳細は、以下の石田富士夫氏の説明を参照にされたい（<http://www.tensuite.jp/ts/column/sns/001/>。2013年7月1日閲覧）。

SNSとはソーシャルネットワークサービスの略称で、同じ趣味・嗜好を持つ人と人のつながりを促進・サポートしたり、居住地、出身校、「友達の友達」といった直接関係のない他人との繋がりを通じて人間関係を構築する場を提供するコミュニティ型の会員制サービスを指します。

また2010年から急速に会員数を伸ばしているミニブログサービスのtwitterもSNSの一種とみなして良いでしょう。

SNSに参加するためには、個人情報登録して個人IDを取得します。（中略）

代表的なソーシャルネットワークサービスとして、日本最大の会員数を誇るmixi、モバイル向けのGREE、モバゲータウン、海外では世界最大の会員数を持つFacebook、それに次ぐMyspaceが有

名です。

2011年9月現在、Facebookは世界で約7億5,000万人、国内で約472万人、mixiは国内で約2,700万人の会員数を有します。

なお、石田氏はTwitterをSNSに含めているようであるが、含めない立場もあるようである。

- 49) なお、「打ち言葉」という言葉自体は、既にワードプロセッサが普及し始めた時期（1980年代前半）の使用例が認められる。たとえば、倉谷直臣（1982）には、「ワードプロセッサによる「打ち言葉？」」という表現がある（ただし、引用したとおり、鍵括弧、クエスチョンマーク付きで使用されているので、当時はまだ人口に膾炙した単語であったとは言えないと思われる）。
- 50) 「打ち言葉」の諸側面を論じた書籍としては、橋元義明（編）（2005）があげられよう。また、「打ち言葉」の定義に関する議論は、佐藤栄作（2009）も参照のこと。
- 51) 携帯電話のメールに観察される諸現象については、日本語学、国語教育の啓蒙雑誌『日本語学』（明治書院）の2000年10月号（第19巻第12号）で「ケータイ・コミュニケーション」という特集が、2001年9月号（第20巻第10号）で「ケータイ・メール」という特集が、それぞれ組まれている。他に、田中ゆかり（2010）も参考になる。
- 52) 次に示す雑誌『日本語学』の特集号などを参照。「ネット社会の集団語」（2006年9月号、第25巻第10号）。「ブログのことはば」（2007年4月号、第26巻第4号）。「ネット広告と日本語」（2008年5月号、第27巻第6号）。

参考文献

- 朝尾幸次郎（1996）「パソコン通信の文体論」『言語』Vol.25No.9
- 石井正彦（1989）「語構成」玉村文郎（編）『講座日本語と日本語教育6 日本語の語彙・意味（上）』明治書院
- 石井由希子（2011）「五段活用動詞「違う」の形容詞型活用」『千葉大学日本文化論叢』12
- 石川慎一郎（2008）「言語コーパスとしてのWWW—広がる可能性—」『日本語学』Vol.27No.2
- 井上史雄（1988）『日本語ウォッチング』岩波書店
- 井上史雄（2008）『社会方言学論考』明治書院
- 井上史雄・江川 清・佐藤亮一・米田正人（2009）「音韻共通語化のS字カーブ—鶴岡・山添6回の調査から」『計量国語学』26巻8号
- 岡島昭浩（1997）「インターネットで調べる」『日本語学』Vol.16No.2
- 岡島昭浩（2011）「ウェブ検索の応用」荻野綱男・田野村忠温（編）『講座ITと日本語研究6 コーパスとしてのウェブ』明治書院
- 岡田祥平（2006）「「縮約形」再考」『阪大日本語研究』18
- 岡田祥平（2013）「インターネットを利用して現代日本語を調べる—母語を見つめなおす—手法としての提案」新潟大学教育学部国語国文学会平成24年度研究大会配布資料
- 荻野綱男（2007）「コーパスとしてのWWW検索の活用」『言語』Vol.36No.7
- 荻野綱男（2008）「WWWをコーパスとして利用する研究—文系と理系の観点から—」『日本語学』第27巻第2号
- 荻野綱男・加藤 彩・本多さやか・谷口香織（2005）「WWWの検索による日本語研究」『東京女子大学日本文学』第101号
- 荻野綱男・末永絵梨・下重秋弓・三好亜萌（2007）「WWWの検索による日本語研究(2)」『東京女子大学日本文学』第103号
- 小野正弘（2011）「ウェブ検索概論」荻野綱男・田野村忠温（編）『講座ITと日本語研究6 コーパスとしてのウェブ』明治書院
- 沖森卓也・木村義之・田中牧郎・陳 力衛・前田直子（2011）『図解 日本語の語彙』三省堂
- 北本洋子（1995）「日本語動詞「違う」の形容詞型活用の実態」『横浜女子短期大学研究紀要』10
- 窪蘭晴夫（1996）「派生か制約か—最適性理論入門 [上] 制約理論の台頭」『言語』Vol.25No.5

- 窪蘭晴夫 (2000) 「日本語の語彙と音節構造—「女王」は「じょう」か「じょうおう」か?—」『日本語研究』第20号
- 窪蘭晴夫 (2002) 『くもっと知りたい! 日本語』新語はこうして作られる』岩波書店
- 倉谷直臣 (1982) 「ワードプロセッサは日本語を変えるか」『国文学 解釈と教材の研究』第27巻第16号
- 桑本裕二 (1998) 「日本語における複合語略語の音韻構造」『音韻研究』第1号
- 五味伸之・辰巳 暢・新田優喜 (2011) 「Twitter を利用した言語形態の変化についての研究」『福井工業高等専門学校研究紀要 人文・社会科学』第45号
- 斎藤純男 (2007) 『日本語音声学入門 改訂版』三省堂
- 佐藤栄作 (2009) 「文字言語の「書く」をめぐる」『論集V アクセント史資料研究会
- 真田信治 (2005) 「レジスター」真田信治・庄司博史 (編) 『事典 日本の多言語社会』岩波書店
- 柴田 武 (1956) 「集団生活が生むことば」石黒 修ほか (編) 『ことばの講座5 現代社会とことば』東京創元社 (本稿では、柴田 武1978収載のものに拠った)
- 柴田 武 (1958) 「集団語とは」日本放送協会 (編) 『NHK国語講座 日本の常識』宝文館 (本稿では、柴田 武1978収載のものに拠った)
- 柴田 武 (1978) 『社会言語学の課題』三省堂
- 助川泰彦・前川喜久雄・上原 聡 (1999) 「日本語長母音の短母音化現象をめぐる諸要因の実験音声学的研究と音声教育への示唆」アラム佐々木幸子 (編) 『言語学と日本語教育 実用的言語理論の構築を目指して』くろしお出版
- 滝沢直宏 (2007) 「巨大データの必要性—言語の周縁的・慣習的側面を探るために」『言語』Vol.36No.7
- 田中ゆかり (2003) 「ネット検索は言語の研究に有用か」『日本語学』Vol.22No.5
- 田中ゆかり (2010) 『首都圏における言語動態の研究』笠間書院
- 田野村忠温 (2000) 「電子メディアで用例を探す—インターネットの場合—」『日本語学』Vol.19No.6
- 田野村忠温 (2008) 「日本語研究の観点からのサーチエンジンの比較評価—Yahoo!とGoogleの比較を中心に—」『計量国語学』26巻5号
- 田野村忠温 (2010) 「日本語研究の観点からのサーチエンジンの評価・統一検索ヒット件数の時間変動のその後とWeb文書量の推計の修正—」『計量国語学』26巻8号
- 田野村忠温 (2012) 「日本語研究の観点から見た昨今のサーチエンジン事情—GoogleとYahoo!の技術提携の結果—」『計量国語学』28巻5号
- 津田大介 (2009) 『Twitter社会論 新たなリアルタイム・ウェブの潮流』洋泉社
- 徳川宗賢 (1986) 「ことばの調査・序説」『言語生活』第411号
- 徳川宗賢 (1988) 「方言の研究」金田一春彦・林 大・柴田 武 (編) 『日本語百科大事典』大修館書店
- 橋元良明 (編) (2005) 『講座社会言語科学【第2巻】 メディア』ひつじ書房
- 日比谷潤子 (2009) 「言語の変異」池内正幸 (編集) 『シリーズ朝倉〈言語の可能性〉3 言語と進化・変化』朝倉書店
- 深澤はるか (2005) 「最適性理論」西原哲雄・那須川訓也 (共編) 『音韻理論ハンドブック』英宝社
- 前川喜久雄 (2000) 「話し言葉における長母音の短呼—『日本語話し言葉コーパス』を用いた音声変異の分析—」『国語学会2002年度春季大会要旨集』
- 前川喜久雄 (2007) 「コーパス日本語学の可能性—大規模均衡コーパスがもたらすもの—」『日本語科学』22
- 松田謙次郎 (2004) 「ことばのバリエーション」郡司隆男・西垣内泰介 (編著) 『ことばの科学ハンドブック』研究社