

第 1 部

言葉の使用と言語の仕組に関する研究と方法

多量の実例データの観察に基づく 言語現象の研究

藤村逸子

大規模コーパスから得た多量の実例をデータベース化し、探索的に観察することによって、これまでに気付かれなかった言語事実を発見し、発話時に発話者が行う言語使用の傾向を記述し、言語の規則的な特徴の解明に貢献することができる。本章では、データベースの作り方と観察の方法を説明したあと、日本語の品詞と格表示、フランス語の冠詞に関わる研究を紹介する。

キーワード：コーパス、帰納法、言語使用、ジャンル、プロトタイプ

1 はじめに

言語を研究する方法には大きくわけて帰納的・経験論的方法と演繹法とがある。帰納法は個々の言語現象の観察から出発して一般的原理を目指すボトムアップの方向である。演繹法はある一般的原理から出発して、その原理のもとで個々の現象の関係の解明を目指すトップダウンの方向である。実際の研究において、帰納法と演繹法は相反するわけではなく、2つの方法を組み合わせることによってよりよい結論を導き出すことができるが、研究者によってどちらを重視するかには違いがある。ここでは、帰納的方法を重視する立場をとる。

帰納的方法は新しい方法ではない。昔から辞書編纂や文献学や歴史言語学

は言語資料の観察を重視する学問であり、学者は文献を調査し、実例を書きとめて研究していた（たとえば見坊(1990)を参照）¹。しかしその後、言語学が言語（ラング）を研究する学問と位置付けられてからは、言語使用の痕跡である言語資料をいくら観察しても、言語を知ることにはならないと批判されることもあった。現在帰納的な研究方法が盛んになっている理由のひとつは、コンピューター技術の発展にともなって、大規模なコーパスの処理が容易になったからである。コーパスはサンプルである。帰納法は本質的に確率論的な研究方法である。コーパスから何らかの結論を引き出しても、どこまで一般化できるかについての保証はない。しかし、コーパスが大規模であれば信頼性は高まる。

コーパスから得た多量の実例をデータベース化し、それを観察し、記述すると、人間の言語行為の傾向を知ることができる。そこから、人間の言語のルールがいかにして作り上げられるかを推測し、説明できる場合がある（たとえば、Bybee and Hopper (2000)を参照）。規則性を記述することはできるが、説明は容易ではなく、後の研究に任せなければならない現象もある。さらに、言語現象の中には共時的原理によって説明のできない不規則性も存在する。特に語彙と定型表現に関わる部分は不可思議である。

本章では第2節において、データを観察する方法としてのデータベースの作成法を概説する。続く第3、第4、第5節では、多量の実例の観察に基づく研究の実際を紹介する。第3節では日本語の色彩語の文法形式の問題を扱い、母語の言語使用の記述の1例とする。第4節ではフランス語の冠詞をテーマに、外国語や過去の言語使用に関しても、実例の収集によって面白い新発見が可能であることを示す。第5節では日本語の「被動作主」の格表示をテーマに、発話における選択の傾向の観察が、人間の言語一般の傾向の解明に結びつきうることを述べる。

2 データベースの作り方

2.1 データベースとは

本節では、実例を自分の目で観察するためのデータベースの作り方を紹介する。実例は数万件の規模であっても、自分の目で観察することを勧める。数値だけに基づいた議論は避ける方がよい。

ソフトウェアは、表計算ソフトの Microsoft Excel を用いる。専用のデータベースソフトではなく Excel を使うのは、Excel には汎用性が高いという大きな利点があるからである。Excel には、集計、グラフ化、統計検定などの基本的な操作が1つのソフトウェアで行える、本格的な統計ソフトなどの他のソフトウェアへのエクスポートが容易である、単純な表の構成なので例を目で観察することが直観的に行えるなどの長所がある。また、本稿で勧める手作業で扱える範囲のデータ（数万件まで）であれば、分量の面でも Excel で扱うことの不都合はない。

この方法は、言語研究においてはるか昔から採用されている、紙のカードを使った実例データベース（カード型データベース）の電子版である。あるいは、1960年代に梅棹忠夫が『知的生産の技術』（1969）で紹介したカードによるデータ管理の現代版ともいえる。梅棹は次のように述べている。「カードの操作のなかで、いちばん重要なことはくみかえ操作である。知識と知識とを、いろいろにくみかえてみる。あるいはならべかえてみる。そうすると一見なんの関係もないようにみえるカードとカードのあいだに、おもいもかけぬ関連が存在することに気がつくのである。」（p. 58）電子版のデータベースを使うとこの操作がたやすくできる。

2.2 実例データベースを作る

以下では、図1に掲げた色彩表現データベースを例にして、リレーショナル型データベースの作り方を説明する。原則は、1件の例は Excel の1行（1レコード）に対応させること、1つの問題に関するデータは1つのワークシート上の1つの大きな表（テーブル）にまとめることである。色彩表現に関する1万件の例（データ）は、1つのワークシート上の1万行の表（テーブル）になる。

図1の第1行目は見出し行である。各列はフィールドであり、見出し行には、左から横に[No]、[(Keywordの)品詞]、[不要]、[(Keywordの)形態]、[ジャンル]、[前文脈]、[key(word)]、[後文脈]、[N(=被修飾名詞の意味カテゴリー)]、[N(大区分)(=被修飾名詞の意味の大カテゴリー)]の順にフィールド名が並んでいる。2行目以降には縦に実例が並んでいる。1番目のレコードは「茶色のソフトをかぶり、青のスプリングコート姿でハンドルを」という例である。「青の」が[Key(word)]のフィールドに収められ、左と右のセルに前後の文脈が表示されている。例は、キーワードを中心にして、左右のセルに前後の文脈が表示されるKWIC(Keyword in Context)形式に整形してある。Excelでは1つのセルに、最大32,767文字を記入できる。前後の文脈を収めるには十分な量である。1番目の例の場合、各フィールドには以下の情報が収められている。[No]:1、[(Keywordの)品詞]:名(詞)、[不要]:空白(=不要ではない)、[形態]:の、[ジャンル]:新聞、[前文脈]:「茶色のソフトをかぶり、」、[Key(word)]:「青の」、[後文脈]:「スプリングコート姿でハンドルを」、[N]:衣服、[N(大区分)]:人造物・衣服。このうち、[不要]のフィールドと、意味カテゴリーに関する最後の2つのフィールドへの記入は、一定の分析法に基づいて研究者が手作業で行う。

No	品詞	不要	形態	ジャンル	前文脈	Key	後文脈	N	N(大区分)
1	名		の	新聞	茶色のソフトをかぶり、	青の	スプリングコート姿でハンドルを	衣服	人造物・衣服
2	名		の	新聞	帯を裏返したものに	青の	フェルトペンで《中内功とダイ	着色剤	着色剤
3	名		の	新聞	＼T2＼	青の	着物姿で会見に応じた千代大	衣服	人造物・衣服
5	名	x	の	文学	、それと対照的に深い	青の	、海の上の空を三機編隊の	環境	自然
6	形		い	新聞	田さんの胸ポケットに、	青い	「ドラえもん」の携帯電話スト	人造物	人造物・衣服
7	形		い	新聞	を演じている。そそのの	青い	Tシャツ姿。日本のラジオ体操	衣服	人造物・衣服
8	形		い	新聞	＼T2＼「子供の	青い	あざなら、ワシらと同じだなも	身体	自然
9	形		い	新聞	が効果的だ。橋脚も、	青い	イーストリバーの波も見える。	環境	自然
10	形		い	新聞	う素朴な疑問も、手に	青い	インクをつけてティッシュでふい	着色剤	着色剤
11	形		い	文学	のにだまって、八津は	青い	カキの実をたべたのである。も	植物	自然
12	形		い	文学	ajiro.txt 園子は、その	青い	ガス灯がなにかおそろしいもの	光	人造物・衣服
14	形		い	新聞	が報道陣の見守る中、	青い	クーラーボックスを手に、病院内	人造物	人造物・衣服
15	形		い	文学	を焼土の上に立って、	青い	ケンボナシの実があったので、	植物	自然
16	形		い	文学	洗い終ると粗い布地の	青い	コートに身をつつんだ。ひきちぎ	衣服	人造物・衣服
17	形		い	新聞	一帽に白のTシャツ、	青い	ジーンズ姿で連日の特訓。編曲	衣服	人造物・衣服

図1：データベースの例（色彩表現）

データベースの作成の前には、コーパスの構築あるいは既存コーパスの選

択を行い、必要な例を検索して採取することが必要である。出来上がるデータベースの規模や質は、研究者のテキスト処理の能力に左右される。コーパス構築については第4章に紹介がある。コーパスの選択については本章の各節で紹介する。テキスト処理とExcelに関しては第4部に解説がある。

具体的な作業手順の概要は以下のとおりである。

- (a) コーパス中からキーワードの文字列検索を行うことによって例を収集し、テキスト形式で保存する。コーパスによって検索方法は異なる。
 - テキスト形式などの汎用コーパスの場合には、目的とテキスト処理の能力に応じて、様々な検索方法を選択することができる。
 - オンラインコーパスなどの場合、検索方法は一般に制約があり、不自由である。
- (b) テキストエディターを使うか、プログラムを作成するなどして、全ての例のキーワードの前後にタブを挿入する。その結果をExcelで読みこむと、KWIC形式の表示になる。
- (c) 1行目に見出し行を作り、[前文脈]、[Key]、[後文脈]のフィールド名を入力する。
- (d) 列(=フィールド)を新しく挿入し、見出し行に[不要]と記入する。不要例にはxの印をつける。
- (e) 列(=フィールド)を必要に応じて追加し、フィールド名をつけ、各行の該当箇所に情報を入力していく。

2.3 データベースを探索的に使う

データベースが一旦出来上がったら、Excelに付属した次のような機能を使ってデータを整理し、データベースを充実させる。そして、データを探索的に観察する。すなわち、「くみかえてみる、並べ替えてみる、そうすると思いがけない関連の存在に気がつく」という作業を行う。

- (a) 並べ替え (ソート)
- (b) オートフィルタ
- (c) ピボットテーブルとピボットグラフ
- (d) 付属の文字列関数や検索関数
- (e) ユーザー定義関数
- (f) グラフ

3 当たり前のことを発見し説明する

データベース化した多量のデータを観察すると、話者が無意識のうちに行っている言語使用の傾向を記述することができる。言語使用の研究には未開拓の領域がある。流行の理論において重要視されない現象や、内省や小規模なコーパスなどによる従来の方法では観察の容易でない現象の中には、基本的かつ重要な事実であっても、記述されないままのものが存在する。

以下では、日本語の連体修飾用法における色彩表現の文法形式の使い分けの問題を取り上げる。色彩形容詞と「色彩名詞+の」（「白い雪」と「白のウェディングドレス」）はどのように、またなぜそのように使い分けられるのかという問題である。

3.1 日本語の色彩語

まず、例文(1)、(2)を観察する。(1)では下線の引かれた7カ所の色彩表現はすべて形容詞である。(2)では同じ色彩表現が「名詞+の」に置き換えである。アスタリスクはその直後の表現が使用不可であることを示している。

- (1) 青い空と白い雲を背景に、黒い上着と茶色いTシャツを着た*緑い眼の少女が、赤いバラを手に、黄色い顔をして立っている。
- (2) *青の空と*白の雲を背景に、黒の上着と茶色のTシャツを着た緑の眼の少女が、赤のバラを手に、*黄色の顔をして立っている。

例(1)においては「緑い目」のみが不可である。「緑」には形容詞の用法がないので、形態論的に不可能である。一方、例(2)においては「青の空」「白の雲」「黄色の顔」が不可である。言われてみれば当たり前と感じられようが、この現象を指摘する研究は以前には存在しなかった。

観察対象のデータベースは、前節の図1で示したものである。コーパスは、『CD-毎日新聞'99データ集』の前半年分と『CD-ROM 新潮文庫の100冊』のうち1950年以降に出版された42冊の小説である。コーパスの形態素数は、形態素解析プログラム [\(茶釜](#)

[、unidic\)](#) を用いて解析すると、それぞれ合計約12,560万個と約2,350万個である。以下では、便宜的に色彩形容詞の語幹と色彩名詞とを色彩語とよぶ。データの収集の概要は以下のとおりである。まず、「白」、「黒」、「赤」、「青」、「黄色」、「茶色」、「緑」の7つの色彩語に「い」と「の」が後続する例を機械的に検索する。次に、データを均一化するために、「黄色い声」、「黒い笑い」、「緑の党」など実際の色を指示しない例と、「深い青の空」、「赤と白の旗」、「青白い顔」など「い」と「の」の交替が統語的に制約されている例とを、不要例として手作業によって除去する。残るのは3119例であり、これが観察対象のデータとなる（検索条件や検索文字列の選択に関する詳細は、藤村(2003)を参照）。

この条件における色彩語の生起数は、小説では新聞に比べ、割合として10倍以上多い。最も生起が多いのは「白」であり、以下の順序で減少していく。以下のうち、下段の数字は小説と新聞の合計の生起数である。順序は新聞も小説も全く同じである。この順序は、上記の条件下における日本語の色彩語の出現傾向を示しているといえよう。

白	>	黒	>	赤	>	青	>	黄色	>	緑	>	茶色
1230	>	771	>	538	>	254	>	178	>	85	>	63

3.2 名詞と形容詞の選択に関わる要因

色彩を表現するのに名詞が選択されるか形容詞が選択されるかの傾向を決

定する要因に関して、3.2.1 から 3.2.3 までの観察を行うことができる。

3.2.1 色彩語の種類

名詞が選択されるか、形容詞が選択されるかは色彩語の種類によって大きく異なる。図 2 は、色彩語ごとに、「い」と「の」との共起傾向を百分率で示したものである。数字は素の生起数を示す。「白」「黒」「赤」「青」では、約 90%が「い」と共起している。「黄色」は約 70%が「い」と共起しているが、「茶色」は約 80%が「の」と共起している。「緑」には「い」との共起は、もちろん存在しない。）

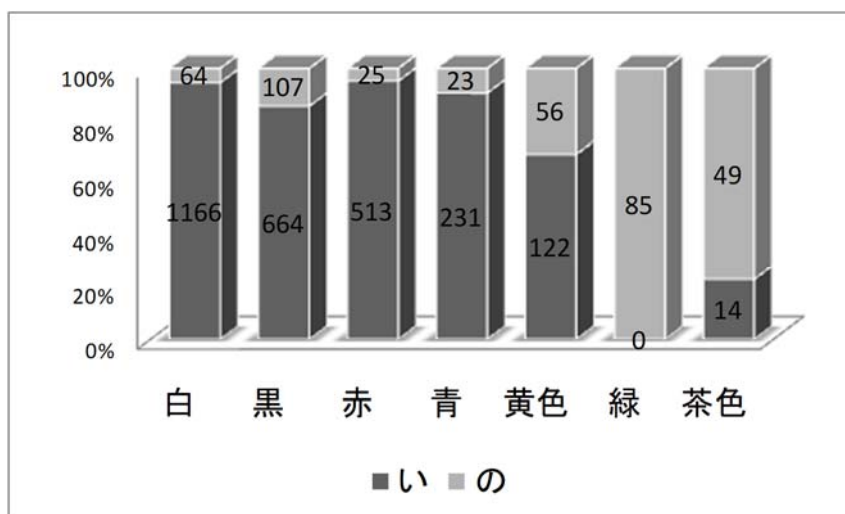


図 2：名詞と形容詞の出現傾向は色彩語によって異なる

色彩語によって選択傾向が異なるという事実を、意味や文法機能や音声などの共時的な要因によって説明するのは困難である。むしろ、通時的な要因を考慮に入れるほうがよい。語にはそれぞれ歴史があり、それが現在の使用に制限を与えていると考えられるからである。

3.2.2 色の性質：自然色か人工色か

どの色彩語も自然の色を示す時には形容詞で表されやすく、人工色を表す時には名詞で表されやすい傾向があり、その差は統計的に有意である。

図3は、色彩語が形容詞として使用される割合を、指示対象が自然色の場合と人工色の場合に分けて折れ線グラフで示している。

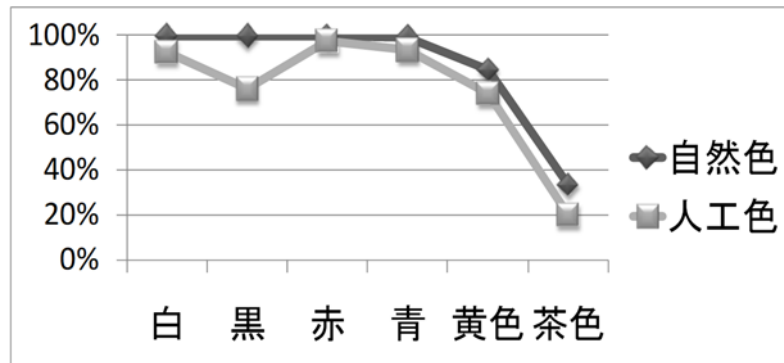


図3: 各色彩語の形容詞としての使用の割合 (自然色と人工色)

自然色を指示する場合、「白」、「黒」、「赤」、「青」はほぼ100%形容詞となる。例(1)、(2)で見たように、「の」の使用はほとんど常に不可である。一方、「黄色」や「茶色」は「の」の使用の制約が緩いので、自然色の場合にも、例(3)のように「の」を用いて問題がない場合が多い。

- (3) 茶色の葉、黄色の葉、*黒の葉、*赤の葉、*青の葉
茶色い葉、黄色い葉、黒い葉、赤い葉、青い葉

しかし「黄色」や「茶色」も、名詞と人工色、形容詞と自然色の共起傾向と無関係であるわけではない。頻度は図3のとおりであるが、使用の可否に関しても制約のある場合がある。たとえば、例(4)で「の」が使用されると、まるで黄色に塗った偽の歯をポケットから出して見せたかのようなのである。この文脈では、自然の歯の色を指して「黄色の歯」は変であり、実際、原文では「い」が使われている。

- (4) 頼圀は黄色 (い/の) 歯を出して笑った。(渡辺淳一『花埋み』)

3.2.3 被修飾名詞の意味

被修飾名詞の内包（意味の限定）が弱く、外延（指示範囲）が広い場合には形容詞が選択され、逆に、内包が強く、外延が狭い場合には名詞が選択されやすい。

「白」「黒」「赤」「青」が自然色を指示する場合にも、まれではあるが例(5)のように「の」が出現することがある。例(5)の被修飾名詞はバラの1品種である。自然色であるが「の」が不自然ではない。自然色が「の」で示されるのはこのような限定的な意味をもつ被修飾語にかかる場合に限られる。

- (5) 真っ白なスノーグーズ、黄色のゴールデンセレブレーション、赤のザ・ダークレディ、ピンクのオールドピンクモスなど色彩豊かなイングリッシュローズを数多く見ることができる。(毎日新聞 1999 年)

例(6)では被修飾名詞は形式名詞の「もの」であり、内包は乏しい。文脈から判断して、人工的に作られた商品が指示対象であり、色も人工色であろうと判断できるにも関わらず、「白のもの」は不可である。「スカーフ」や「スニーカー」のように名詞の内包が豊かになると、「白のスカーフ」「白のスニーカー」のように「の」が可能となる。

- (6) 本来は、バレンタインデーのお返しに1カ月後の3月14日、男性から女性に白いものを贈る日をホワイトデーというそうです。(毎日新聞 1999 年)

3.3 説明

データのソート、グループ分け、グラフ作成、頻度計算などを繰り返しながらデータを観察するにつれ、記述は正確になっていく。要因は多数ある。検討を繰り返すうちに、形容詞であること、名詞であることの意味はなにかという問題がそこにあることに気づく。以下では、自然色と人工色の対立の

問題のみに限定して考え方の道筋を示す。他の要因についても同じ作業を行ううちに、分析はより精密になる。

自然色と人工色の区別は言語学的な概念ではない。自然色と人工色の対立を形容詞と名詞の対立に関連づけるためには、さらに抽象的な考察が必要である。形容詞のプロトタイプの機能は、ある存在物の状態や属性を描写することであり、名詞のプロトタイプの機能は、ある存在物を指示することである(cf. Croft (2003: 183--188))。自然の色はそれ自体独立して存在するわけではなく、常にある別の存在物の特性である。「青い空」の「青」という特性は「空」の状態であり、「空」の存在なしに「青」だけが存在することはない。一方、人工の色はそれ自体としての存在が可能である。人工色は典型的には色見本帳の中に存在する。「黒のドレス」における「黒」は「ドレス」とは独立して、「黒色」を指示しうる。自然色と人工色の違いを形容詞と名詞の違いに対応させることが可能なのはこの理由によると考えられる。

4 誰も知らないことを発見する

大量のデータに基づく帰納的方法は、外国語の研究や歴史的な言語研究に適した方法である。ある言語の研究を母語話者以外が行うデメリットは、母語話者のもつ言語直観を働かすことができないことである。一方、メリットは母語話者の気付かない独創的な研究の可能性があるということである。また、学習者に役立つ実用的価値のある研究につながる可能性も高い。コーパスを用いた方法を採用すると、デメリットを補うことが可能である。本節では、フランス語の不定冠詞複数形の2つの形式間の選択の問題を取り上げる。

4.1 フランス語の不定冠詞複数形

フランス語の初級文法では不定冠詞の複数形に関して次のように教える。不定冠詞複数形の *des* (発音は[de])は、形容詞が名詞の前に置かれると、原則として *de* (発音は[də])に変わる(例(7)(8))。ただし、話し言葉では

des のままであることが多い。また、形容詞と名詞が複合名詞を構成する場合には例(9) (10)のように des のままである。なお、形容詞の位置は一般に名詞の後にであり、短い語形の、高頻度の形容詞だけが名詞の前に置かれる。フランス語の形容詞は男女単複の4種類に語形変化をする。以下、本文中で形容詞を挙げる際には男性単数形を代表形として用いる。

- | | | | |
|-----|------------------------------|------|----------------------|
| (7) | des conditions intéressantes | (8) | de bonnes conditions |
| | des 条件(N) 有利な(ADJ) | | de 良い(ADJ) 条件(N) |
| | 有利な条件 | | 良い条件 |
| (9) | des petits pois | (10) | des grands magasins |
| | des 小さい(ADJ) 豆(N) | | des 大きい(ADJ) 店(N) |
| | グリーンピース | | デパート |

この規則は明解であるとは言い難い。話し言葉とは何なのか、複合名詞とは何なのかなど問題が多い。実際話し言葉でも、de を使った(11)のような表現は存在する。また、ハイフンで結ばれて表面的には1語化した複合名詞でさえ、(12)のように de の使用されている例を見つけることができる。

- (11) Je te souhaite de bonnes vacances.
私 あなたに 願う de 良い(ADJ) 休暇(N)
楽しい休暇を過ごしてね。
- (12) Dieu aime la justice et le bon droit et qu'on ne le trompe pas par de faux-semblants. (Oldenbourg, Z., *Les cités charnelles ou l' Histoire de Roger de Montbrun* 1961)
よって(PREP) de 偽の(ADJ)-外見(N)
神は正義と正当性を愛し、人が見せかけで神を騙さないことを望んでいる。

フランス語教育の現場では、この問題を学習者にどのように教えるのが最適なかわからない。この問題に関する文献は、規範文法書以外に存在しないのが実情であり、調査したい気持ちに駆られる。

4.2 コーパスの選択とデータ

コーパスのジャンルとコーパスの規模、および採取したデータの数は表1に挙げたとおりである²。20世紀後半の資料としては、学術書、新聞、小説、一般雑誌、議会発言録、インターネット上の議論（ネットニュース）、私的な話し言葉の7つのジャンルを選んだ。17世紀から20世紀前半までの資料は学術書と小説を用いた。コーパスの規模は合計約1億8000万語である。機械的な検索と手作業による修正の結果、最終的に得た用例数は、合計約15,000例である。

表1：コーパスと de と des の頻度

時期	ジャンル	詳細	年	コーパス の規模 語数(千)	用例数		
					<i>de</i>	<i>des</i>	合計
20世紀 後半	議会 発言録	Hansard (カナダ議会)	1986-1988	3 300	214	77	291
	新聞	<i>Le Monde</i>	1997, 2001	13 000	930	145	1075
		<i>Libération</i>	1992-1993	1 500			
	一般雑誌	<i>Actuel</i>	1990-1992	2 000	417	120	537
		<i>Marie Claire</i>	1990-1993	3 100			
	ネット ニュース	個人登録に よって配信 されたもの (323種)	2000年10月	21 200	1113	530	1643
	小説	Frantext	1951-2000	15 600	902	228	1130
学術書			11 900	659	30	689	
私的な 話し言葉 (インタ ビューや語 りを含む)	Blanche- Benveniste		89	71	267	338	
	Corpus Allier		72				
	Onlineデータ (Delic, Elicop)		不明				
20世紀 前半	小説	Frantext	1901-1950	20 600	1477	215	1692
	学術書			13 500	821	38	859
19世紀 後半	小説	Frantext	1851-1900	14 500	1267	104	1371
	学術書			4 400	312	16	328
19世紀 前半	小説	Frantext	1801-1850	12 600	891	43	934
	学術書			8 100	661	10	671
18世紀	小説	Frantext	1701-1800	12 000	981	25	1006
	学術書			11 200	1172	44	1216
17世紀	小説	Frantext	1601-1700	3 800	300	29	329
	学術書			7 000	557	92	642
合計				179,500	12745	2013	14758

de と *des* は多機能の高頻度語である。*de* はフランス語の語彙中、もっとも頻度の高い語である。*de* には冠詞としての別の用い方も存在する上に、英語の *of*、*from* に相当する前置詞でもある。*des* には不定冠詞複数以外に、前置詞の *de* と定冠詞の連続 (=of the) が縮約した形としての用い方もある。不定冠詞の *de* と *des* を機械的に収集するためには、前置詞の直後や、特定の他動詞の直後など、冠詞以外は出現できない文脈をターゲットに定める必要がある。コーパスは品詞情報を含まないものを用い、検索のあと、不要例は手作業により除去した。品詞情報付きコーパスを用いなかった理由は、*de* と *des* の品詞の判定が極めて難しく、既存の品詞解析プログラムの性能では解析ミスが多発が明らかであったからである。形容詞は *ancien* (古い) , *beau* (美しい) , *bon* (良い) , *grand* (大きい) , *gros* (太い) , *joli* (か

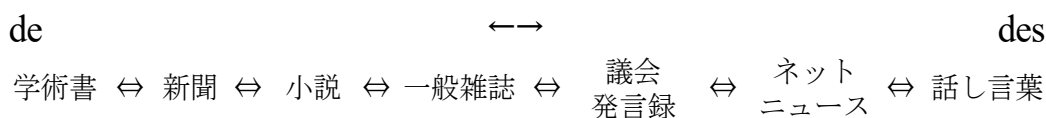
わいい) , mauvais (悪い) , nouveau (新しい) , petit (小さい) を対象にした。使用頻度が高く、被修飾名詞との共起制約が小さいものを選んだ。

4.3 de と des の選択に関わる要因

de と des の間の選択に関与する要因に関して、データから次の 4.3.1 から 4.3.4 までの観察ができる。

4.3.1 ジャンル

コーパスのジャンルによって傾向は大きく異なる (表 1)。予想通り、私的な話し言葉では des が多く de が少ない。学術書では des が少なく de が多い。しかし、話し言葉にも de はあるし、学術書にも des が存在する。私的な話し言葉と学術書の間は次のように連続している。



4.3.2 形容詞の種類

形容詞の種類によって傾向は大きく異なる。petit (小さい) は、des との共起傾向が特に強い特異な形容詞である。nouveau (新しい) は de との共起傾向の強い形容詞である。nombreux (多い) や excellent (素晴らしい) と des の共起はほぼ皆無である。

4.3.3 petit の特異性

petit (小さい) は、あらゆる時代あらゆるジャンルの資料を通じて、des との共起傾向が強い。図 4 と図 5 は、20 世紀後半の全てのジャンルの資料と 17 世紀から 20 世紀後半までの小説の資料における des の生起の割合を形容詞別に示したものである。petit を示す線は他の形容詞よりも常に上にあり、des との共起が常に最も頻繁であることがわかる。私的な話し言葉において、

petit と des との共起は 90%以上の高い割合であるが、他のジャンル、他の時代でも同じように共起傾向が強い。この現象は 1647 年に規範文法書 *Remarques sur la langue françoise* (フランス語に関する覚書) に取り上げられ、以降 de の使用が規範となったと言われる。des の使用が 17 世紀の資料で比較的多く、18 世紀になると減るのはこの規範化に呼応していると思われる。しかしいずれにせよ、petit と des の共起傾向は規範化以前の 17 世紀から今に至るまで常に強い。

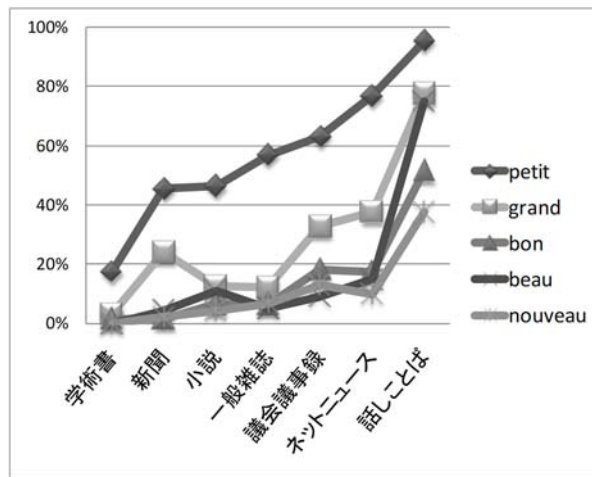


図 4 : 20 世紀後半の全てのジャンルにおける形容詞別の des の割合

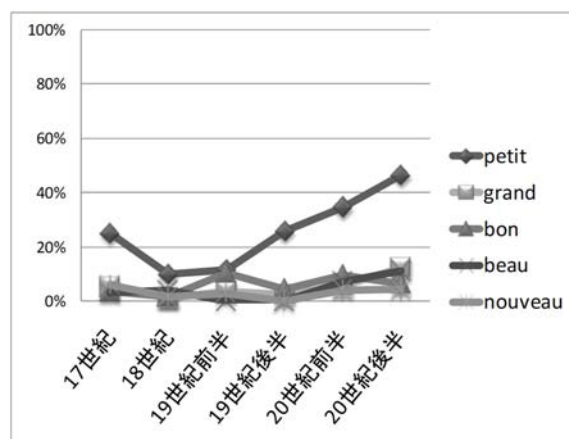


図 5 : 17 世紀から現代までの小説における形容詞別の des の割合

petit と des の共起の強さが何に由来するのかは難しい問題であるが、petit の特殊な用法のせいではないと言える。たとえば例(13)では同じ文中で、des と petit、de と grand (大きい) が共起している。petit と grand の意味は単純に「小」と「大」である。

- (13) Or, cette relation entre la taille et le taux de mortalité est probablement un simple biais statistique, lié au fait qu'un taux calculé sur des petits effectifs est beaucoup plus fluctuant que quand il est calculé sur de grands effectifs. (*Le Monde*, 06/10/97)

des 小さい(ADJ) 数(N) de 大きい(ADJ) 数(N)

ところで、身長と死亡率とのこの相関はおそらく単なる標本の偏りによる。小さな数をもとに計算された割合は、大きな数で計算されるよりずっと変動が大きいことに関係している。

4.3.4 形容詞の強調

重複や副詞の添加などにより形容詞が強調されると de が生起しやすい。

(14)、(15)はともに私的な話し言葉からの例である。de が使用されにくいはずの私的な話し言葉において de が使用されるのはこの場合が多い。

- (14) il y a de grands grands immeubles hein. (Delic)

de 大きな(ADJ) 大きな(ADJ) 建物(N)

大きな大きな建物があるよね。

- (15) parce que euh /de, Ø/ toute façon depuis l'Antiquité il y a toujours eu des bijoux et de très beaux bijoux (Corpus Allier)

de とても 美しい(ADJ) 宝飾品(N)

だって、いずれにしたって太古の昔からずっと、ふつうの宝飾品ととてもきれいな宝飾品とがあったんだから。

図 6 は、形容詞の強調の有無と de と des の選択傾向をまとめたグラフである。形容詞の強調があると、全てのコーパスにおいて de の使用が増加している。話し言葉でも強調があると de の使用は 20 例中 10 例(50%)にみられるが、強調がないと 320 例中 63 例(20%)にすぎない。de の使用は、昔の文法家が気まぐれに定めた単なる書記法上の規則などではなく、発話の現実に直結した生きた現象であると考えられる。

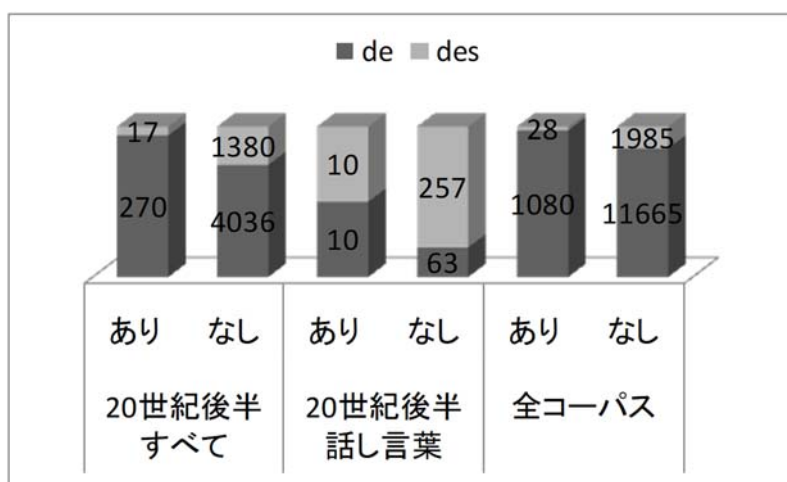


図 6 : 形容詞の強調の有無による de と des の選択傾向

5 新しい発見を先行研究に統合する

最後に、すでに記述が蓄積されている研究分野において、多量のデータの観察から得られる情報がなしうる貢献について述べる。以下では、2項文（他動詞文）における「被動作主」の格の表示の問題を取り上げる。なお、2項文の「被動作主」とは、いわゆる「直接目的語」とほぼ同じ概念を指すものとして読み進んでいただきたい。

5.1 「被動作主」の格表示の傾向

日本語では、典型的な 2 項文（他動詞文）において、「動作主（主語）」の格形は「が」であり、「被動作主」の格形は「を」である（「真理がピアノをひいた。」）。英語では、名詞の「動作主」と「被動作主」の間に違いは存在

しない(「Mary played the piano.」)が、人称代名詞では違いがある(「She met her.」)。「被動作主」の格形が「動作主」と同じか否かは言語ごとに定まっているが、潜在的には言語を超えて同じ条件が働いていると考えられ、「Animacy (有生性)」と「Definiteness(定性)」の要因が関与すると言われている (cf. コムリー(2001: 199--214), Croft(2003: 166--175))。これらの要因は、以下に示す通りの含意的階層関係を成しているとされ、この枠組によって日本語も英語も説明が可能である。

Animacy scale (有生性のスケール) :

Human (人間) > Animate (有生) > Inanimate (無生物)

Definiteness scale (定性のスケール) :

Personal pronoun (人称代名詞) > Proper name (固有名詞) >
 Definite NP (定名詞句) > Indefinite specific NP (不定・特定
 名詞句) > Non-specific NP (不特定名詞句)

「被動作主」の特徴がスケールの左端に近いほど、明示的に表示されやすく、右端に近いほど明示的には表示されなくなる。日本語の典型的な2項文の「被動作主」は、右端の無生の不特定名詞句であっても明示的に表示される(「智史が絵を描いた。」)。英語では、左端の人間を指示する人称代名詞のみが明示的に表示される。

この現象は次のように言い換えることが可能である。

1. 人間の言語には「被動作主」の卓立性 (prominence)が高いほど、「被動作主」を明示的に表示する傾向が存在する。
2. 個々の発話者は、個々の言語使用において、「被動作主」の卓立性 (prominence)が高いほど、「被動作主」を明示的に表示する傾向が存在する。

さて、日本語の2項文に戻る。日本語では、願望文(「一たい」)や「好き

だ」、「分かる」などを述語とする非典型的な2項文においては、「動作主」と「被動作主」を同じ格形で示すことができる。「被動作主」は、「動作主」と同じ「が」で表示されることも(「大己が英語が分かる。」、動作主」とは異なる「を」で表示されることもある(「大己が英語を分かる。」。この現象は上述の現象によく似ている。卓立性の高い「被動作主」を「を」で表示する傾向がコーパスで確認できたならば、他の言語で観察されている現象と同種の現象とみなすことが可能かもしれない。データをさらに細かく観察すると、人類の言語一般や、人間の発話傾向の一般に関する知識に貢献できるかもしれない。以下ではこの問題の一端を扱う。

5.2 卓立性：人間と人間の心

コーパスは『CD-毎日新聞データ集』の1991年から1999年(9年分)を原則として使ったが、述語によっては生起が多すぎるためにコーパスの規模を調整した。検索は、「被動作主+「が」/「を」」の直後に述語が後続するもののみを対象とした。機械的な検索のあと、手作業により不要例を除去した。検索条件やコーパスの詳細は、藤村(2009)を参照されたい。

次の表は、述語が「分かる」と「好きだ」の場合の「を」と「が」の生起頻度を、「被動作主」のカテゴリー別に整理したものである。「が」と「を」の間の選択は、「被動作主」のカテゴリーによって傾向が異なる。

表2：「被動作主」の特性と「が」vs「を」(「好きだ」と「分かる」)

述語	「被動作主」の特性	「を」	「が」	計	「を」の割合
好きだ	名詞節(「こと」「の」)	0	362	362	0%
	名詞句(人間以外)	83	1738	1821	4.6%
	名詞句(人間)	114	296	410	27.8%
分かる	名詞節(「こと」「の」)	60	1108	1168	5.1%
	名詞句(人間以外)	209	2956	3165	6.6%
	名詞句(人間)	29	67	96	30.2%
	名詞句(人間の心理)	62	176	238	26.1%

表 2 が示すように、「被動作主」が形式名詞を伴う名詞節(例(16)(17))の場合と、人間以外の名詞句(例(18)(19))の場合には「が」の使用が圧倒的であり、「を」は使用できない文が多い。一方、「被動作主」が人間を表す卓立性の高い名詞句の場合(例(20)(21))は、「を」が約3割の例において使用されている。すなわち、この問題は上記の問題と大いに関連していると考えられる。なお、以下の例文の出典は、全て、毎日新聞 1999 年である。

- (16) 足のけりが氷に伝わるのが分かった。
- (17) 手先を使うことが好きだよね。
- (18) 結末が分かっているのに読んでしまう。
- (19) 学生時代から旅と本が好きで、1年休学して東欧を放浪。
- (20) 初めておやじを分かったように思う。
- (21) ひとは、なぜ人を好きになるのだろう。
- (22) 他人の上に立つ人は、他人の痛みを分かる人でなきゃ。
- (23) 公害被害者の家族の思いを分かってもらえ、うれしい。

面白いのは、同じように抽象名詞であっても人間の心理を表す場合(例(22)(23))とそうでない場合(例(18))とで、「を」の出現傾向が全く異なることである(表 2)。述語が「分かる」の場合には、「被動作主」が人間の心理(「気持ち」、「苦しみ」、「苦労」、「思い」、「心」、「痛み」など)の例が多数出現するが、その場合には、人間の場合と同様の高頻度で「を」が選択される。人間と人間の心理はどちらも卓立性が高いので、同じ扱いを受けると考えるとうまく説明できる。「被動作主」が人間の場合に「動作主」とは異なる格形で表示されやすい理由として、「動作主」と「被動作主」との曖昧性の排除の機能が挙げられることがあるが、「被動作主」が人間の心理の場合に曖昧性は存在しない。すなわち、この事実は曖昧性による説明に対する反証となる。

6 おわりに

本章では、大規模コーパスから得た多量のデータをデータベース化する方法と、データベースを使ってデータを観察する方法を説明した。ケーススタディとして、日本語の色彩表現における形容詞と名詞の間の選択、フランス語の複数不定冠詞の2つの形式間の選択、日本語の「好きだ」と「分かる」の「被動作主」の2つの表示形式間の選択の問題を紹介した。

この方法を用いて発話者の行う言語使用の痕跡を詳しく観察することにより、言語使用の傾向を推測できる。また、言語使用の傾向と言語の規則は連続していることがわかる。言語の規則は、多数の要因が関与する多変量的現象であると考えられる。また、語彙に関わる部分には強い不規則性が存在し、通時的考察が必要である。1つの言語の局所的な文法現象の観察を、人間の言語活動の一般的傾向と関連付けることによって、人間の言語の一般的问题へとつなげていく面白さがある。

注

- ¹ 伝統的には電子化されているか否かに関わらず、特定の言語研究のための言語資料を corpus という。本稿では、言語研究のために用いる電子化された言語資料をコーパスと呼んでいる。
- ² コーパスの出典等の詳細は Fujimura et al. (2004) を参照のこと。オンラインコーパスの Delic と Elicop からの用例は本稿のために追加した。

参考文献

- Bybee, Joan L. and Paul Hopper. (eds.) (2000) *Frequency and the Emergence of Linguistic Structure* (Typological Studies in Language). John Benjamins.
- コムリー・バーナード 松本克己・山本秀樹訳(1992)『言語普遍性と言語類型論—統語論と形態論』ひつじ書房 (Comrie, Bernard. *Language Universals and Linguistic Typology: Syntax and Morphology*. (1981) Oxford: Blackwell.)
- Croft, William. (2003) *Typology and Universals*. Second Edition. Cambridge : Cambridge University Press.
- 藤村逸子 (2003) 「色彩名詞と色彩形容詞の対立— 新聞と文学のコーパスからわかること—」

『日本語学習辞書編纂に向けた電子化コーパス利用によるコロケーション研究 中間報告論文集』： pp. 25-48.

Fujimura, Itsuko, Mitsumi Uchida, and Hiroshi Nakao (2004) *De vs des* devant les noms précédés d' épithète en français: le problème de *petit*, In Gérald Purnelle et al. (eds.) *Le Poids des mots* 1: pp. 456--467 Louvain-la Neuve: Presse Universitaire de Louvain.

藤村逸子(2009)「他動性再考：「被動作主」を表示する「が」と「を」の交替」*Asian and African Studies* 13(1) : pp. 73--102 University of Ljubljana.

Fujimura, Itsuko (2010) Interaction entre la syntaxe des lexèmes et le sémantisme des parties du discours: Nom vs. Adjectif de couleur en japonais in Massimo. Moneglia et als. (eds) *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective* : pp73-96 Firenze : Firenze University Press.

見坊豪紀 (1990) 『日本語の用例採集法』南雲堂.

梅棹忠夫 (1969) 『知的生産の技術』岩波書店.

de Vaugelas, Claude Favre (1647) *De, & des, articles. Remarques sur la langue française : utiles à ceux qui veulent bien parler et bien écrire.* Gallica, Bibliothèque Nationale de France, pp330--331.

<<http://gallica.bnf.fr/ark:/12148/bpt6k84316s> > 2011. 6. 12

コーパス

Delic (Description Linguistique Informatisée sur Corpus)

<<http://sites.univ-provence.fr/delic/corpus/index.html> > 2011. 6. 12

Elicop (Etude Linguistique de la Communication Parlée)

<<http://bach.arts.kuleuven.be/elicop/>>2011. 6. 12

大量発話データに基づく 語を超えた構成単位の研究

滝沢直宏

本章では、言語使用における語を超えた構成単位、具体的には語と語の慣習的な結びつきであるコロケーションおよびそれをさらに拡張したレキシカルバンドルについて述べる。これらを認定するにあたっては、母語話者の直観的判断では不十分であり、電子化コーパスが不可欠な研究ツールである。

キーワード：

コロケーション、レキシカルバンドル、コーパス、MI-score

1. はじめに

コロケーションとは、ごく簡単に言うと語と語の慣習的な結び付き（共起関係）のことを指す。例えば、Hunston（2002）は、コロケーションをまず the statistical tendency of words to co-occur. (p. 12) と簡潔に定義し、後に the tendency of words to be biased in the way they co-occur (p. 68) と述べ直している。コロケーションは、各語の意味から全体の意味を導き出せる単位であり、この点で意味的に不透明なイディオム（例えば、「死ぬ」を意味する kick the bucket）とは区別される。田野村（2010）は、共起関係を語と語に限定せず、「・・・句や節、あるいは、構成素を成さない語連続もコロケーションの観点から観察すべき対象の範囲に含める。」(p. 2) と述べているが、筆者はさらに広げて品詞などの範疇と言語表現の共起関係などもコロケーションに含めて理解したい。

コロケーションに留意することは、「母語話者らしい」自然な言葉遣いのために必要不可欠であり、その重要性に関しては、多くの研究者が異口同音に指摘してきている。例えば、コロケーションの重要性を理論的、実践的観点から述べた Lewis (2000)は、次のように述べている。

We now recognise that much of our 'vocabulary' consists of prefabricated chunks of different kinds. The single most important kind of chunk is collocation. Self-evidently, then, teaching collocation should be a top priority in every language course.

この指摘よりもはるか以前、ロンドン学派の Firth (1957: 11)、さらには日本で英語教育に従事した Palmer らが、コロケーションの重要さを指摘している。しかし、半世紀以上前の時点では、語がどのような語と一緒に使われるのかを調べる手立ては、手作業によるデータ収集あるいは直観や内省による以外には存在しなかった。しかし今日では、電子化されたコーパス（コンピューターで処理可能な電子化された言語資料）が使えるようになったため、語と語の共起傾向についてコンピューターを用いて調べることが可能となったのである。

本章では、英語と日本語の語を超えた構成単位の様々なタイプを紹介しつつ、コーパスからの網羅的な抽出方法について言及する。

2. コロケーションの事例：英語

本節では、まず英語における事例を見ていくことにする。

2.1 語と語の共起：副詞＋形容詞の連鎖

語を超えた単位といっても様々な種類があるが、最初に最も単純な語と語の共起傾向の例を見てみよう。(以下、The British National Corpus (BNC

と略記) から例を引く。)

(1) ... the market will fail if transaction costs are prohibitively high. (BNC:ABE:3218)

(2) Although the idea of cooking with Parma ham sounds prohibitively expensive, it need not be so. (BNC:ABB:235)

(1) の例文の下線部分に、prohibitively という副詞が使われている。prohibitively の後ろには、「高い」を意味する high が来ている。ここでの「高い」は、fees を修飾していることから分かるように「値段が高い」という意味で使われていることに気付く。(2) では prohibitively と、やはり「値段が高い」ことを意味する expensive が一緒に使われている。prohibitively という語はあまり見かけない副詞であると思われるが、日本語に当てはめると、「法外に」という言葉に置き換えることができる。(例：「法外に高い料金」)

この prohibitively という語がどのような語を修飾するのが一般的であるかを知ろうと思えば、コーパスを用いてこの語を調べ、その右隣に来る語に注目すれば良い。すると、正に上の例のように、「値段が高い」という意味で使われている high と expensive がもっぱらであり、その他の語は少数であることが分かる。ここから、prohibitively と high および expensive の間には、慣習的な結び付きが存在していることになる。したがって、これは副詞とそれが修飾する形容詞のコロケーションの事例と見なすことができる。

ここで日本語の「法外に」がどのような語を修飾しているかを見てみると、以下に示すように、「値段が高い」という意味の形容詞を修飾している例あるいは、同様の意味を表す動詞との共起が目立つ。

(3) 法外に高い医師はもちろん、あまり安い医師にも問題がある。(毎日新聞 1991 年)

(4) . . . 禁止するから常習者は地下に潜り、やみ市場が生まれ、マフィアが介入して麻薬の値段が法外に跳ね上がる。(毎日新聞 1991 年)

このことから言語は違っても prohibitively と「法外に」という副詞は、かなり類似した共起傾向を示すことが確認できるのである。

「副詞＋形容詞」の例をさらに見てみよう。

(5) I was acutely aware of the urgent need to find some work to do, so I combed the national and local newspapers to see what was on offer. (BNC:A0F:683)

(6) In the political arena, too, American and British women demonstrate noticeably different qualities and values. (BNC:CBC:5085)

(5) の aware は「気付いている、意識にのぼっている」という意味の形容詞だが、その意識が非常に鋭いということを表すために acutely が使われている。この場合もまた、コーパスによって、acutely という副詞は aware のような形容詞と仲がいい傾向にあることが分かる。

(6) では different という形容詞に対して noticeably という副詞がかかっている。different の程度が高いことを強調し、全体として「著しい違いがある」という意味になっている。微妙な違いであると気付きにくいだが、その違いが大きければ誰でも notice することができるのである。この 2 つの語の連鎖も比較的仲の良い連鎖であると言える。

コロケーションは、形容詞＋名詞、副詞＋動詞など多くの品詞連鎖にまたがるが、中でも ly 副詞がかかわるコロケーションは、これまでの辞書記述において必ずしも十分に扱われてきていないので、とりわけ重要である。

2.2 whole と wholly : コロケーションと文法

では、次の文を見てみよう。

(7) For me, it started a whole new way of thinking about the photograph. (BNC:FBR:389)

この文の下線部は「全く新しい方法」という意味である。ここで注意すべきは、whole は形としては形容詞であるが、new を修飾する副詞となっているという点である。このような事例に出くわすと、whole という語がどのような形容詞を修飾しているのかを調べてみたくなる。「全く」を意味する副詞的な whole が修飾する形容詞にどのような傾向があるのかをコーパスで調べるのである。すると副詞用法の whole によって最も高頻度に修飾される語は new であることが分かる。(7) は、副詞的に使われている whole の典型的な用法を示していることになる。

new に次いで目立つ共起語は different である。そのほかの語は共起頻度も低く、重要な共起語と見なすことはできない。少なくとも私たち非母語話者が英語で whole を使う場合には、new や different を修飾する以外の使い方は避けた方が無難である。

このことから、whole という副詞は主に new もしくは different と共起すると考えることができるのである。逆に言えば、この new もしくは different を修飾する場合には全うな副詞形である wholly から ly を取って whole にしてもかまわないということである。副詞を作る語尾-ly を無闇に落とすわけにはいかず、どの語を修飾しているのかが関わっていることになる。見かけ上、形容詞の形をしていながら副詞的に別の形容詞を修飾しているということであるから、コロケーションが文法と関わる事例と見ることができる。

2.3 比較級・最上級:範疇が関わるコロケーション

次は比較級と最上級に関わるコロケーションの例を見る。

(8) Prince Albert and Queen Victoria, who was a martyr to seasickness, were also enthusiastic at a time when Anglo-French relations were noticeably warmer than in the previous 50 years. (BNC:BMJ:10)

(9) Bernini was arguably the most important architect and sculptor in my period ... (BNC:A0F:50)

(10) New York-based rare book dealer Bernard Breslauer has spent twenty-five years collecting these single illuminated leaves, amassing the finest private collection of its kind in America. (BNC:CKY:509)

(8) では noticeably という副詞が warmer という比較級を修飾しているが、この noticeably は warmer という特定の語ではなく、むしろ比較級と共起しやすい副詞であると言える。というのも、noticeably の後ろには、warmer に限らず多くの比較級が現れることがコーパスによって確認できるからである。先に (6) に関して、noticeably は different と共起しやすいことを述べたが、比較級とも同様の理由で共起しやすいと言える。つまり、比較されているものの違いが誰の目にも明らかである (notice できる程、大きい) ことを noticeably が表しているのである。したがって、比較級という範疇との共起傾向と different という特定の語の共起傾向は同じ理由に根ざしていると考えることができる。

同様に (9) では、arguably が the most important という特定の最上級と共起しているというよりも、最上級という範疇と共起していると考えた方が適当である。この副詞は他の多くの最上級形との共起がコーパスによって確認できるからである。

次の (10) は今までのものとは少し性格が異なっている。(10) では of its kind と the finest が不連続ながら関係を維持しているからである。この of its kind は「そのような種類の中で」という意味であるから、最上級

が成り立つ範囲を限定する役割を果たしていると言える。of its kind という 3 語がまとまってやや離れたところにある最上級という範疇と共起傾向を示していることになる。

3. 大規模コーパスによる網羅的抽出

これまで紹介してきた事例は、具体的な語句を定め、それをコーパスで確認するというものであった。関心をもった特定の語句の用法を知るには、コーパスに依拠しなくとも、多くの英文を読んでその語句に出会う度にノートに書き留めておき、集まった用例をよく観察するという方法をとることもできる。(既存の辞書などの文献から情報が得られることも多い。)

しかしどんなに頑張っても、コーパスを使わないことにはできないことがある。それは、語句を具体的に決めず品詞のみを頼りに行う調査である。例えば、「副詞＋形容詞」といった品詞の連鎖の調査がそれに該当する。読書中に出会うすべての「副詞＋形容詞」を書き留めていくことは、事実上、不可能である。一方、各語に対して品詞情報が付与されているコーパスを利用すれば、そういった調査も可能になる。品詞の連鎖を網羅的に抽出してリストを作ることによって、これまで全く意識にのぼっていなかった表現やどの辞書にも載っていない表現を見つけることができるし、その連鎖の細かな用法や頻度を知ることができる。

品詞情報が付与されている BNC の World Edition では、例えば、really good という「副詞＋形容詞」は<w AV0>really <w AJ0>good のようになっている。(w は word、AV0 は副詞、AJ0 は形容詞をそれぞれ表している。) こうした品詞情報を利用することで、コーパス中の特定の品詞の組み合わせを網羅的に抽出できる。網羅性の確保は、手作業では行い得ないコーパス利用の重要な利点と言える。(なお、あらかじめ品詞情報が付いていない電子テキストの場合には、品詞情報を自動的に付与してくれるソフトウェアを利用するという手立てがある。)

ここで BNC World Edition から「-ly 副詞+形容詞」を網羅的に抽出し、頻度順に並べてみよう。以下に、上位 5 位までを挙げる。

表 1 : 「-ly 副詞+形容詞」

847	really good
663	relatively small
574	particularly important
572	slightly different
506	completely different

この表から、どのような共起が多用されているのかというのは読み取れるが、どれもありふれており、英語表現として有益な情報が得られたという実感はないだろう。これは、共起頻度が高い 2 語であっても、両語の間に格別の相性の良さを認められる保証はないということである。例えば、really は good 以外の多くの形容詞を修飾することができ、同時に good も really 以外の多くの副詞によって修飾されうるからである。そこで、頻度ではなくある統計的な処理をすることによって語と語の結びつきの強さを測り、それによって頻度は低くても重要な連鎖を抽出する手立てを考えることになる。統計値には様々なものがあり、どの統計値を用いるとコロケーション抽出に有益であるかは、それ自体が重要な研究テーマであるが、本稿では MI-score (Mutual Information、相互情報量) を用いることにする。MI-score の算出には、コーパスから、共起頻度 (ここでは「ly 副詞+形容詞」)、個別頻度 (ここでは ly 副詞と形容詞の個別頻度)、そしてコーパスの総語数 (ここで用いている BNC は約 1 億語) が必要である。MI-score を求める式は、次の通りである (『応用言語学事典』 p. 637 より引用)。

$$\text{MI スコア} = \log_2 \frac{\text{ある単語 A と B の共起頻度} \times \text{総単語数}}{\dots}$$

単語 A の頻度 × 単語 B の頻度

実際に BNC から「副詞＋形容詞」を抽出し、MI-score を計算すると、その値が高い連鎖には以下の連鎖が含まれていることが分かる。

federally funded, grammatically well-formed, upwardly mobile, aesthetically pleasing, racially motivated, prelingually deaf, massively parallel, visually handicapped, deathly pale, diametrically opposed, ideologically unsound, racially discriminatory, fatally flawed, blissfully unaware

頻度はさほど高くなくとも、いずれも英語表現として重要な連鎖であることが分かる。このように、コーパスを用いることで、網羅的に重要な連鎖を抽出することが可能となるのである。

4. 日本語のコロケーション

これまでは英語のコロケーションの事例を見、前節では、副詞＋形容詞の連鎖の網羅的抽出について述べてきたが、ここでは日本語のコロケーションの網羅的抽出を試みる。

日本語の場合は、英語とは違って言語研究を目的に構築されたコーパス（狭い意味でのコーパス）は多くない。（但し、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」プロジェクトによる「現代日本語書き言葉均衡コーパス」(Balanced Corpus of Contemporary Written Japanese) が近々公開されれば、日本語のコーパスもかなり充実することになる。）現時点では、新聞記事の電子版など、広い意味でのコーパス（言語研究を目的として構築されたのではない電子化テキスト）を使ってコロケーションの抽出を行うことになる。本稿では、1991 年から 1999 年の「毎日新聞」の電子版（日外アソシエーツ）の本文を使用する。

新聞の電子テキストは言語研究を目的に構築されたわけではないので、前節で利用したBNCのような狭義のコーパスとは異なり、予め品詞情報が付与されているわけではない。そこで品詞タグを自動的に付与してくれる形態素解析器を利用し、品詞情報を付与する作業が必要となる。ここでは形態素解析器として、奈良先端科学技術大学院大学で開発された「茶釜」(ChaSen)と呼ばれる解析器(ver. 2.4.4)を利用する(辞書は、IPAdic(ipadic-2.7.0)を用いる)。

以下、英語と同様に「副詞+形容詞」の連鎖の抽出を試みる。例えば、

(11) ただ行動指針の中身となると、ビジョンから行動への第一歩というには、はなはだ心もとないものがある。(毎日新聞1995年)

という文には副詞「はなはだ」と形容詞「心もとない」の連鎖が含まれている。

新聞の電子テキストを「茶釜」で解析すると、(11)は次のようになる。「茶釜」では、必要に応じて様々な出力形式を指定することができる。ここでは集計の都合上、活用している用言をすべて辞書形に直して出力する。また、詳細な品詞情報を付与することも可能だが、ここでは最も簡素な品詞情報に留めている。)

(12) ただ_接続詞 行動_名詞 指針_名詞 の_助詞 中身_名詞 と_助詞 なる_動詞 と_助詞 、_記号 ビジョン_名詞 から_助詞 行動_名詞 へ_助詞 の_助詞 第一歩_名詞 と_助詞 いう_動詞/自立 に_助詞 は_助詞 、_記号 はなはだ_副詞 心もとない_形容詞 もの_名詞 が_助詞 ある_動詞 。_記号

1度このような品詞情報が与えられると、「副詞+形容詞」の連鎖を機械的に抽出することは容易である。

実際に、毎日新聞9年分の記事から「副詞+形容詞」を抽出し、出現して

いる連鎖を頻度順に示してみよう（以下では頻度 500 回以上に限定している）。

表 2：副詞と形容詞の共起（500 回以上）

4339	最も	多い
4095	ほとんど	ない
2055	全く	ない
1415	最も	高い
1311	かつて	ない
1251	できるだけ	早い
928	まったく	ない
923	あまり	ない
692	もっと	早い
667	極めて	高い
606	まだ	ない
557	とても	うれしい
540	本当に	うれしい
517	本当に	よい

ここから分かることは、少なくとも毎日新聞に関する限り「副詞＋形容詞」の連鎖で最も高頻度の連鎖は「最も・多い」であることが分かる。次いで「ほとんど・ない」（「ない」は「形容詞」として解析される）が続く。さらに先を見ていくと「できるだけ・早い」がある。この連鎖は、実際のテキストでは「できるだけ早く」という形で使われることが多いが、先に述べたように茶釜による解析を行う際に辞書形に戻しているため、「早く」も「早い」の形で集計されている。

こうした抽出は容易だが、このような情報を得ても有益な情報を抽出し得たと感じられることはないと思われる。その理由は、副詞「最も」は

「多い」以外の様々な形容詞をも修飾し得るからである（「最も・高い」「最も・小さい」「最も・暑い」のように）。つまり、日本語母語話者の頭の中で「最も・多い」が格別な結び付きを成しているのではない、ということである。

そこで、英語の場合と同様に MI-score を算出してみよう。すると、MI-score の値が高い連鎖として、以下のような連鎖が抽出される。

- | | |
|-----------------|----------------|
| (13) 何ら・やましい | (14) 誠に・喜ばしい |
| (15) くれぐれも・よろしい | (16) にわか・騒がしい |
| (17) 何とも・やり切れない | (18) はるか・遠い |
| (19) どうぞ・よろしい | (20) めっきり・涼しい |
| (21) 努めて・明るい | (22) なんとも・情けない |
| (23) 何とも・痛ましい | (24) ごく・浅い |
| (25) ずっしり・重い | (26) できるだけ・早い |
| (27) どんなに・心強い | |

これらの連鎖は、全てが高頻度に使われるわけではないが、母語話者であれば、2語が結び付いていると感じることができると思われる。

5. 「副詞＋形容詞」：2語を超える連鎖

前節で挙げた (13) から (27) の語連鎖を見ると、2つのことに気付く。

まず全てではないにせよ、生じる語形に制限があるものが結構見受けられるということである。例えば、「くれぐれも・よろしい」という連鎖は、このままの形で使われることはなく、「くれぐれも・よろしく」でのみ使われる。これほど明らかではないものの、「ずっしり・重い」は「ずっしり・重く」という形で使われるのが一般的だろう。前に挙げた「最も・多い」のように共起頻度が高い語連鎖の場合は、生じる語形に制限がないのが普通だが、MI-score 上位の幾つかの連鎖は、ある特定の語形でのみ使われる、

あるいは使われ易いと言える。(なお、これらの例から分かるように、品詞による抽出は、あくまで機械的に「副詞の直後に形容詞が出現している」ことを頼りに抽出作業を行っているので、副詞が形容詞を修飾しているか否かは問題にしていない。)

2点目は、「副詞+形容詞」の語連鎖の中には、その前後にどのような語が生じ易いかの見当がつくものがある、ということである。上記の例に関して言えば、以下のような連鎖が母語話者の想起しやすい典型ではないかと思われる。(以下では、形容詞を実際に使われる形に直して提示している。)

- (13') 何ら・やましい・ところはない。
- (14') 誠に・喜ばしい・限りである。
- (15') くれぐれも・よろしく・お願いします/お伝え下さい。
- (16') にわかに・騒がしく・なりました。
- (17') 何とも・やり切れない・思いがする。
- (18') はるか・遠く・に〇〇が見えます。
- (19') どうぞ・よろしく・お願いします/お伝え下さい。
- (20') めっきり・涼しく・なりました。
- (21') 努めて・明るく・振舞う。
- (22') なんとも・情けない・思いがする。
- (23') 何とも・痛ましい・限りである。
- (24') 震源の深さは・ごく・浅い。
- (25') ずっしり・重く・響く。
- (26') できるだけ・早く・〇〇して下さい。
- (27') どんなに・心強い・ことでしょう。

言うまでもなく、これら以外の形で生じ得ないということではないが、母語話者であれば、これらのパターンを容易に思い付き且つ自然な響きが

すると感じられるだろう。ということは、このような自然なパターンを日本語学習者が産出できるようになることは、言語の自然な使い方の習得において重要なことだと言える。

このような2語を超えた範囲において成り立つ語の連鎖は、レキシカルバンドル (lexical bundle) (Biber et al. (1999)の第13章) と呼ばれることがある。Biber et al. (1999: 989)は、

Lexical bundles can be regarded as extended collocations: bundles of words that show a statistical tendency to co-occur.

と述べ、例として in the case of the、there was no significant、it should be noted that など、数語から成るパターンを挙げている。また、この説明の通り、レキシカルバンドルをコロケーションが拡張されたもの、すなわち「拡大コロケーション」(extended collocation) と見ることもできると Biber et al. は言う。

レキシカルバンドルを抽出のためには、 n -gram を作成するのが有益である。 n -gram とは、 n 個の要素 (文字や単語あるいは形態素など) の連続を意味する。大量のデータを基にして n -gram を作成し、その頻度を見ることによって、レキシカルバンドルの抽出も可能となる。

レキシカルバンドルの抽出・認定には、以下のような手順が考えられる。

手順1: MI-score による有意義なコロケーション (2語の連鎖) の認定

手順2: 認定されたコロケーションを含む文の抽出

手順3: 形態素単位での n -gram の作成とその検討

このような方法を採用することで、2語に留まらず、より大きな単位での語の慣習的連鎖の抽出が可能となる。

6. Lemma の危険性：ある活用形で多く用いられる表現

本節では観点を変え、ある活用形で多く用いられやすい動詞句を取り上げる。例えば、次のうち (28) から (30) は、共に「名詞＋を＋動詞」というパターンをとっているが、このうち (29) と (30) は「バ形で使われることが自然な動詞句である。

- (28) 本を買う
- (29) 欲を言う
- (30) 元をただす

したがって、「*彼は欲を言った。」「*彼は決して欲を言いません。」のような文は、文意は明瞭であるが容認不可能である。「元をただす」も同様である。この点で、(29) (30) はごく一般的で活用形の制限のない (28) とは区別されねばならない。

同様に、その使用がほぼテ形に限定される「名詞＋を＋動詞」も多い。

- (31) 隙を見る (例：「店員の間を見て、・・・を盗んだ。」)
- (32) 身をもつ (例：「身をもって・・・を知った。」)
- (33) 暇を見つける (例：「暇を見つけては・・・している。」)
- (34) 合間を縫う (例：「仕事の合間を縫って・・・している。」)

これらもテ形以外で使われると、不自然な使用ということになる。

つまり、「名詞＋を＋動詞」というパターンの中には、実際に使われる動詞形がほぼ決まっているものもあるということである。

これまで英語研究の世界では、Lemma (辞書に見出し語として載る形) に一括して検討するのではなく活用形ごとに見ていくことの重要性が指摘されてきた。例えば、Stubbs (2002: 224) は次のように述べている。

... different forms of a lemma often have different collocational behaviour.

Hoey (2005: 5) も同様である。

Collocational analysis can be done on lemmas or words. Renouf (1986), Sinclair (1991), Stubbs (1996) and Tognini-Bonelli (2001) have all argued against conflating items sharing a common lemma ... on the grounds that each word has its own special collocational behaviour.

これらの指摘は語に関して述べられたものだが、語と語の結合（例えば、動詞句など）に関してより一層当てはまる。

以下では、「名詞+を+動詞（バ形）」で生じるのが一般的な動詞句を、毎日新聞（1991 年から 1999 年）から抽出した結果を示している（滝沢 (2007) 参照）。

表 3 : 「バ形で使われやすい「名詞+を+動詞」

動詞句の出現数	動詞句	「バ形」での出現数	「バ形」での出現割合(%)
72	元-を-ただす	70	97.2
204	裏-を-返す	196	96.1
99	欲-を-言う	85	85.9
28	元-を-たどる	22	78.6
44	表現-を-借りる	33	75.0
35	言葉-を-換える	25	71.4
29	結論-を-言う	20	69.0

この他、上記の例よりは頻度・割合が落ち、この表にあがっては来ないものの、「歴史を振り返れば」、「例外を除けば」、「力を合わせれば」、「理想

を言えば」、「目を見れば」、「対応を誤れば」などもバ形での使用が目立つ。

さらに、これらの「名詞+を+動詞（バ形）」は、これ自体が1つの固まり（prefabricated chunk）になって、より大きなパターン（レキシカルバンドル）の中で使われることがある。

- (35) 裏を返せば・・・とも言えるでしょう。
- (36) 欲を言えば、きりが無い。
- (37) 幾つかの例外を除けば・・・。
- (38) （みんなで）力を合わせれば、何とかなるでしょう。
- (39) 目を見れば、わかります。
- (40) 対応を誤れば、大変なことになる。

言うまでもなく、これらの動詞句が常にバ形で使われるわけではないことは、その百分率を見れば明らかである。しかし、はっきりとした傾向は指摘できるのであり、その傾向の存在は力説するに値する事実である。このような、例外はあるが明瞭な傾向があることについて、従来の研究ではあまり明示的に指摘されてきてはいないように思われるが、コーパスの出現によってこういった傾向の指摘が可能になったということである。

7. おわりに

本章では、以下のことを述べた。

- (i) 語と語の共起傾向（コロケーション）には様々なタイプがある。
- (ii) コロケーションに代表される言語の慣習的側面を捉えるには、コーパスの利用が必須である。

母語話者であっても、その言語の慣習性を意識しているわけではないので、母語話者の内制的言語知識が無意識に発動された結果を収集した大量の資料（コーパス）の中に慣用の姿を求めることが必要となってくる。

- (iii) コーパスを利用することで、MI-score などの統計値を利用するこ

とが可能となり、そのことによってその他の方法では発見が困難なコロケーションを捉えることが可能となる。

(iv) 2語の連鎖を超えたレキシカルバンドルの認定には、 n -gramの作成とその検討が有益である。

(v) 共起傾向を示す語の連鎖（コロケーション）の中には、全ての活用形で使われることがなく、一部の活用形に偏って用いられるものがある。その抽出においても、コーパスは有益である。

これまでの言語記述は、「ある特定の形で顕著に使われ易い」といった事実は無頓着であったように思われる。勿論、特定の形以外では使われることが決してない動詞句については指摘されてきた。しかし、特定の形以外での用法がありながら、ある形で使われることが顕著である、といったいわば「例外を許す傾向」を綿密に記述することは十分には行われてきていない。これが可能になったのは、大量の言語資料が電子化され、コンピューターによる処理が可能になったことによる。人間は、白黒が明瞭な事柄に関しては内省・直観で捉えることができるが、例外がある傾向の把握や、指摘されて初めて気付く慣習の把握となると困難である。

コーパスはまた、言語は我々が想像する以上にパターン化され、慣習に支配されていることを示している。そしてその慣習に従っているか否かが、言語使用の自然さ（naturalness）を保つ上で重要であると言える。こうした慣習的知識は、Hunston（2002: 20）が指摘するように「無意識的」なもので直観・内省で捉えることは難しい以上、コーパスを利用し、こうした傾向の把握をできるだけ網羅的に行うことが、言語記述において重要である。無意識に存在する知識を、コンピューターの利用によっていわば「あぶり出す」のである。それによって、従来の研究では必ずしも十分ではなかった「網羅的な言語の慣習性の記述」が可能になるのである。

* 毎日新聞の記事データは、毎日新聞社の許可を得て利用しているもので

ある。

参考文献

- Biber, Douglas, Stig. Johansson, Geoffrey. Leech, Susan. Conrad, and Edward. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Firth, John. R. 1957. "A synopsis of linguistic theory, 1930-1955," *Studies in Linguistic Analysis*. pp. 1--32. Oxford: Basil Blackwell.
- Hoey, Michael. 2005. *Lexical Priming*. New York: Routledge.
- Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- 国広哲弥. 1997. 『理想の国語辞典』大修館書店.
- 小池生夫他 (編) 『応用言語学事典』研究社.
- Lewis, Michael (ed.) 2000. *Teaching Collocation: Further Developments in the Lexical Approach*. Hove: LTP (Language Teaching Publications).
- Renouf, Antoinette. 1986. "Lexical resolution," in Meijs, W. (ed.), *Corpus Linguistics and Beyond: The Proceedings of the 7th International Conference on English Language Research on Computerised Corpora*, pp. 121--31. Amsterdam: Rodopi.
- Sinclair, John. M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, Michael. 1996. *Text and Corpus Analysis*. Oxford: Blackwell.
- _____. 2002. "Two quantitative methods of studying phraseology in English," *International Journal of Corpus Linguistics* 7,2: 215--244.
- 滝沢直宏. 2007. 「言語の慣習性とコーパス」『2007年日語教學國際會議論文集』 pp. 61-73.
- 田野村忠温. 2010. 「日本語コーパスとコロケーション」、『言語研究』138: 1--23. 日本言語学会.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

会話コーパスの構築による コミュニケーション研究

藤村逸子・大曾美恵子・大島デイヴィッド義和

大規模会話コーパスである『名大会話コーパス』の構築の方法を説明し、同コーパスで観察可能な、終助詞とコミュニケーション・ストラテジーの関わりを記述する。ウチの関係の会話では「よ」が多用され、直言法が多く、対立が忌避されないのに対し、ソトの関係の会話では「よね」が多用され、迂言法が多く、確認と同意が反復されることを、定量的および質的に示す。

キーワード：会話、ポライトネス、終助詞、コーパス構築、発話行為

1 はじめに

本章は、会話コーパスを構築してそれに基づく研究を行おうとする人々や、会話コーパスを使った研究がしたい人々のために、『名大会話コーパス』を紹介し、同コーパスを使ったコミュニケーション研究を提示することを目的とする。『名大会話コーパス』は、名古屋大学の教員と大学院生が中心となって構築した、日本語母語話者による約 100 時間の自然会話コーパスである。2009 年より文字化された資料が公開されており、研究のためであれば誰でも自由に利用できる。

書き言葉の電子化コーパスが数多く存在するのに対して、利用可能な話し言葉コーパスは数が少ない。自然会話を収録したコーパスはさらに限られて

いる。言語活動の基本が話し言葉であり、しかもそれが日常的なおしゃべりであるのなら、言語やコミュニケーションの研究のために自然会話コーパスの整備が進み、研究の進展が望まれることは言うまでもないが、汎用性の高い理想的な会話コーパスは現状では存在しない。本章では『名大会話コーパス』の特徴をいかした利用法として、定量的分析を取り入れたコミュニケーション・ストラテジーに関する研究を紹介する。

以下では 2 節で、『名大会話コーパス』の構築の方法を具体的に説明する。3 節と 4 節は、終助詞「よ」「ね」「よね」の使用を手掛かりとして、親疎関係によるコミュニケーション・ストラテジーの違いに関する分析を行う。3 節では定量的分析、4 節では質的分析を扱う。

2 『名大会話コーパス』の構築

『名大会話コーパス』は、2001～2003 年度に行われた科学研究費補助金による共同研究「日本語学習辞書編纂に向けた電子化コーパスによるコロケーションの研究」の一環として構築された¹。この構築の仕事には、研究代表者の大曾美恵子が、総責任者として細部にわたり実質的に関わった。『名大会話コーパス』は、2009 年 8 月より、国立国語研究所日本語教育ネットワークにおいて、文字化資料が「研究用データ」として公開されている（『名大会話コーパス』

<<http://dbms.ninjal.ac.jp/nuc/index.php?mode=viewnuc>>）。

近年、コーパスを利用した言語研究が盛んになっているが、利用できるのは、新聞、白書、書籍など、書き言葉コーパスが大半を占める²。そこで、話し言葉（会話）の文法、及びコロケーション研究のために、まず約 100 時間分の自然会話を収集し、コーパスを構築することとなった。

2.1 会話収集の方法と対象

会話の収集は主として、名古屋大学の日本語学・日本語教育学を専攻する大学院生に依頼して行った。その他、研究代表者、研究分担者が収集したもの、またそれ以外の人に依頼して収集してもらったものもある。こうして名

古屋周辺の地域だけでなく、その他の地域の居住者のサンプルも採集できるよう配慮した。

データ収集にあたっては、なるべく静かな場所で 30 分から 1 時間程度の自然な会話をしてもらい、カセットテープ、あるいは MD を使用して録音した。録音機器には外付けのマイクを装着した。録音されていることは会話参加者の全員に了解されている。映像はとっていない。

参加者は、日本語の母語話者であることという条件をつけただけで、会話内容などは全く制限しなかった。プライバシーを侵害するような話が出てきた場合、公開段階で削除する措置をとることにした。

文字化にあたって声の識別が容易にできるよう、参加者は原則 2 名とした。しかし、性別が異なる人の場合、識別の問題がないと考え、3 名～4 名の会話も収集対象とした。

収集者には、文字化の際の参考資料として会話の場面・状況に関するメモの作成を求めた。また、本章末に添付した収集記録票を作成し、収集者を介して、参加者に記入してもらった。

収集した会話の文字化は、まず、普段は講演などの文字化を行っている、書き起こしを専門とする会社に依頼して行った。次に文字化された会話のチェック、修正は会話収集にあたった者が担当した。

2.2 『名大会話コーパス』の概要

上記のような方法で収集され、文字化された約 100 時間分のデータは、129 のファイルで構成されている。収集の時期は 2001 年～2003 年の間である。会話参加者は、女性 162 名、男性 37 名で、14 歳から 92 歳まで、幅広い年齢層をカバーしている。大半が友人、同級生、家族など親しい者同士のインフォーマルな会話だが、初対面同士の会話や、仕事関係の知人同士など、丁寧体で話す間柄の会話もある。

参加者の出身地、現在の居住地とも様々で、方言も出てくるが、理解不可能な部分はないので、全員、共通語に近い話し方をしていると言っていいだろう。しかし、関西弁の「や (=だ)」「やろう (=だろう)」「あかん (=い

けない)」「ちゃう (=ちがう)」「やん (=じゃない)」「ねん (=の (だ))」「へん (=ない)」など、かなり頻繁に使われている。

また、主として名古屋大学の大学院生が収集にあたったため、名古屋近辺で収集された会話が多くのが、東京で録音された会話もかなりの数に上る。中には、北海道、新潟等で録音されたものもある。

2.3 文字化

2.3.1 取り決め

テープと MD の書き起こしを書き起こし担当会社に依頼する際に、次のような取り決めをした。

- (a) 漢字、仮名を使って表記する。
- (b) 読みが難しい漢字語には【 】の中に読みを入れる。
- (c) 発話の終わりには句点を打つ。読点は、文字化担当者に任せる。
- (d) 上昇調で終わる疑問文には最後に？を入れる。
- (e) 笑いは、<笑い>とし、相づちは（ ）の中に入れる。
- (f) 聞き取れないところには***を入れる。
- (g) 不確かな語は*XXX*と記す。
- (h) 言い間違いなど、訂正せず、そのまま記す。
- (i) 発話の重なりは表記しない。

2.3.2 文字化の問題点

文字化には次に述べるように様々な問題がある。

① 相づちの認定

「相づちは（ ）の中に入れる」としたが、何を相づちとするかは、研究者によっても考えが一定しない。相づちは聞き手側の行動であるが、どこから聞き手が話し手になるかの認定は難しい。同じ「うん」でも、質問に対する肯定的な返事の場合は、話し手として自らの答えを提示しているもので、相づちとは言えない。

大曾が修正した箇所もあるが、このコーパスの（ ）の中には、相づち以外のものも入っていると考えた方がよい。

② 聞き取りの難しさ

文字化の担当者は、当然であるが、会話参加者と面識がない。そこで会話の聞き取りの助けになるよう、テープまたは MD と共に、収集者が記した場面に関する情報を提供したが十分ではなかった。特に難しいのが固有名詞である。「やないさん」という名前が「ながいさん」となっていたりする。

同じく難しいのが聞き慣れない単語である。今回の会話参加者には大学関係者が多かったため、専門用語が飛び交うし、英語、フランス語など、外国語の単語も含まれている。次のような間違いが見られた。矢印の左は間違いで、右にあるのが、正しい。

- (a) 単語連結語が多くなる → 談話連結語...
- (b) そのときに*はんど*がつくとか... → ... ファンド...
- (c) 昔の*せんせい*文法だったら、 → 昔の生成文法だったら、

この他にも次のような間違いが見つかった。文字化担当者が聞き間違えたものである。

- (d) けっこう → かてきょう (=家庭教師)
- (e) 置くところに入れといても → ふところに入れといても
- (f) その方が良質ですね → 頭脳流出ですね
- (g) *いわや*みんな書いてるんですよ。 → 読みやあみんな...

不確かな語として、(b) (c) (g) のように*に囲まれているものは見直しの段階で必ず確認したが、問題は、(a) (d) (e) (f) などである。これらを見ると、我々が日頃、意味が理解できるように音声を聞き取ろうとしていることが分かって興味深いが、コーパス構築にあたっては音声データを丹念に聞き直し、訂正する必要がある。

日本語の場合、もう1つ、文字化に際して、入念な打ち合わせが必要な

は漢字の問題である。2 つ以上の読み方がある場合、仮名表記にするか、あるいは読みを示すべきであった。読みを示すためのマークの【 】の活用が期待されたが、【 】は難しい漢字語の読みを示すのに使うということで書き起こしを依頼したため、簡単な漢字の場合、【 】の中に実際の発音が表記されることはなかった。今回、問題になったのは次のような語である。

- (h) 私 (わたし／わたくし)
- (i) 何か (なにか／なんか)
- (j) 20 歳 (はたち／にじゅっさい)
- (k) 3 月末 (... まつ／... すえ)

この他にも話し言葉を文字にする時の表記の問題がある。たとえば、実生活では表記する必要のない相づちなど、平仮名で「うん」「えっ」と表記されたものが、正確に相づちの発音を表しているとは限らない。「そう、そう、そう」なのか「そ、そ、そ」の方が実際の音声に近いのか、表記にはどうしても聞く人の主観が入ってしまう。また、仮名は発音記号ではないので、すべての音声を表記することはできない。『名大会話コーパス』の利用にあたっては、こうした問題があることを念頭に置いておく必要がある。

今回は、会話の書き起こしを依頼した会社から納入されたファイルを少なくとも 1 度は、入念にチェックし、修正した。その際、収集にあたった者が会話に加わっている場合は、その人にチェックを依頼した。精度の高い文字化資料作成のためには、こうした地道な確認と修正が必要である。

2.4 ファイルの形式

文字化されたファイルは以下のような形式となっている。次に 1 つのファイルの最初の部分をサンプルとして示す。

@データ 38 (31分)

@収集年月日：2003年2月6日

@場所：F020 宅

@参加者 F025：女性 30 代前半、熊本県熊本市出身、東京都練馬区在住

@参加者 F020：女性 90 代前半、東京都渋谷区出身、東京都練馬区在住

@参加者の関係：祖母と孫

% c o m：F025 は 15 歳から東京に居住。

F020 は 35 歳から 68 歳まで熊本県に居住。

F025：和紙人形、また海外に送りたいんだけど、(はい) ここから選んでいい？

F020：はい、選んでちょうだい。

もうそこにあるのと、ここにあるのとだけですもん。

(うん、うん) うん。

F025：どうしよ。

どれでもいいの？

F020：うん、どれでもいいよ。

ここ、これもかわいいよ。

(ふーん) 今でも海外でそんなの、はやんないんじゃないの。

F025：うーん、はやってなくてもべつに、<笑い>べつに、いいんだ。

なんかきれいじゃない。

ヘッダーでは、@の後に、データ番号と会話の長さ、収集年月日に収集場所、参加者情報が記載されている。

参加者情報としては、性別、年齢 (5 歳刻み)、出身地、現在居住している場所が記されており、最後に参加者同士の関係が入っている。参加者名はすべてコード化され、女性の場合は F+数字、男性は M+数字で書かれている。その他、% c o m：のあとに追加情報が記されている。

2.5 公開に関する問題点

コーパスは公開して研究者・教育者が共有できるところに価値がある。し

かし、個人のプライバシーに関わる情報を公開することはできない。また、差別語の扱いに関する問題もある。大曾は、言語研究の資料としては、差別語もそのまま残しておいた方が良くと考えるが、国立国語研究所の事業としての公開にあたってはやはり考慮が必要である。『名大会話コーパス』には、大学院生が差別語について議論をしている場面があるのだが、こうした場面は公開版では削除した。

100 時間というのは、プライバシー侵害の恐れなど、問題のある部分を削除して残った会話の長さである。

2.6 今後の課題

『名大会話コーパス』は、文字化資料の公開のみで、音声データは公開していない。話し言葉では音声に含まれる情報が大きい。イントネーションで発話の意味が変わることもある。プロミネンス、アクセント、ポーズも重要な情報を伝える。音声も公開できるとよいが、プライバシー保護の壁が立ちはだかる。今回は文字化資料のみを公開するという条件で会話参加者の許可を得たので、音声の公開はできない。音声公開の許可もとったデータが待たれる。

100 時間の会話コーパスは決して十分ではない。『名大会話コーパス』の参加者は、年齢も背景も様々な人たちである。例えば、年代別に東京出身者の話し方を調べるなど、条件を揃えると、サンプル数はかなり少なくなる。今後は、多くの人々がデータを収集し、同じ形式のファイルを作り、なんらかの形でそれらを共有することが望まれる。

3 『名大会話コーパス』に基づくコミュニケーション研究

本節と次節では、『名大会話コーパス』をコミュニケーション研究のための資料として使う。相互行為論などのコミュニケーション研究では、綿密に注意を払って文字化を行う中で、会話参加者の意識的・無意識的な行為を発見するという手法がとられることが多いが、本コーパスの特徴はこれとは異

なる。『名大会話コーパス』は、会話コーパスとしては規模が大きく、内容はどれも雑談であり、話し手は 20 代の女性が多い比較的均一なデータであり、文字化の厳密性という面では欠けている点があるといった特徴がある。

研究対象の言語表現は終助詞の「よ」「ね」「よね」である。終助詞は使用頻度が高いので実例の収集は容易である。『名大会話コーパス』には、同一の話し手が、相手を変えていくつもの会話に参加しているものがある。これらの話し手の会話を終助詞の使用を手掛かりにして観察すると、話し手が相手との関係に応じてコミュニケーションのスタイルを変える様子を見ることができる。

「丁寧さ」に関しては、Brown and Levinson(1987) のポライトネス理論を借りて考察する。発話行為の特徴づけは、オースティン(1978)の「発語内行為(illocutionary act)」の考え方による。親疎関係の特徴づけのためには、三宅(1994)によるウチ、ソト、ヨソの枠組を用いる。

以下では、3.1 において終助詞のデータベースを作成する手順を説明し、3.2 では終助詞「よ」「ね」「よね」の基本的特徴を述べる。3.3 では研究対象とする話し手及び発話を説明し、3.4 では本章で用いるデータを特定する。3.5 では「よ」と「よね」の使用の傾向は親疎の程度と関連があることを示す。

3.1 文末表現データベースの構築

文末表現の使用を網羅的に研究することを目的とし、『名大会話コーパス』のうち早い段階で使用可能となった 78 個の会話を対象として、データベースを作った。ソフトウェアは表計算ソフトの Microsoft Excel を使った(データベース作成の詳細に関しては第 1 章を参照のこと)。収集したのは、次に挙げる文末表現が「。」「!」「?」の直前に現れる文である。

「よ」「ね」「の」「さ」「や」「かしら」「ぞ」「ぜ」「わ」「もん」「だ」「です」「ます」およびその長音形

まず、機械的な文字列検索によって約 23,000 例を抽出し、次いで、手作業によって不要例を除去した。不要例の中には、文末表現ではないものが多く含まれるが、文末表現であっても以下の 1 から 3 のタイプのものは除去した。

1. 終助詞が文に接続しないもの：「ねえ。」「だよね。」「あのね。」「でもねー。」「修学旅行もね。」「ちょっとね。」など
2. 統語的に完結しない文に接続するもの：「～からね。」「～てね。」「し～ね。」「～とかね。」「～たらね。」「～たりね。」など
3. 終助詞「か」が含まれるもの：「～かな。」「～かね。」

統語的に完結しない文を除いた主な理由は、文に接続しないものとの判別が、特に「ね」の場合、極めて難しかったからである。その結果、約 16,000 の例を得た。約 16,000 例のうちには、「だ」「です」「ます」による言いきりが約 2,000 例含まれている。これらは終助詞の使用と不使用を比較する目的で抽出したが、本稿では扱わない。これを除いた約 14,000 例のうち、最も高頻度なのは「ね」で約 4,000 例 (28%)、続いて「よ」が約 3,000 例 (21%)、「の」も約 3,000 例 (21%)、「よね」が約 2,600 例 (約 18%) である。この 4 つの合計は 88% を占め、極めて頻繁に使用される基本的な文末形式であることは明白である。その他は全てを合算しても約 12% にすぎず、偏りは大きい。以下は次の順で続く。「もん」「な」「わ」「もんね」「わね」「わよ」「かしら」「わよね」「もんな」「よな」「ぞ」。

3.2 「よ」「ね」「よね」の機能

「よ」「ね」「よね」の機能については、すでに多くの議論がなされているが (神尾 (1990)、宮崎他 (2002)、日本語記述文法研究会 (2003) など)、ここでは詳細な検討は行わず、大曾 (2005) の以下の特徴づけを採用する。

「よ」：話し手の「自己主張」を示す。

「ね」：聞き手との「一致志向」を示す。

「よね」：自らの情報、判断の確認を相手に求めていることや、自らの情報、判断と相手のそれが一致する（であろう）ことを示す。

終助詞の意味がイントネーションに従って変化するという問題についても数々の指摘がなされているが、本稿では検討しない。『名大会話コーパス』にはイントネーションの記載がないこと、また、イントネーションには発話者の属性による変動などの解決すべき課題があり、本稿の範囲を超えることがその理由である。

3.3 「話し手」と「相手」の関係

本章で観察するのは、以上の方法で作成したデータベースの内、表1に挙げる8つの会話に出現するデータである。これらはどれも2人の参加者による会話である。対象とするのは4人の「話し手」(F004, F011, F093, F098)の発話である。これらの「話し手」は全員が女性である。それぞれの「話し手」は、親しさの程度が異なる2人の「相手」と会話を行っており、親疎の違いによるコミュニケーション・ストラテジーの相違を比較することができる。

表1には、左からデータ番号、話し手、話し手の属性、親疎関係、2人の関係、会話の場所、相手、相手の属性、録音時間に関する情報が記載されている。データ番号はそれぞれの会話の番号であり、公開版の『名大会話コーパス』における各会話のファイル名と同一である。8つの会話の合計の録音時間は441分で、会話によって31分から89分までの幅がある。

親疎関係の特徴づけには、三宅(1994)によるウチ、ソト、ヨソの分類を用いる。すなわち、「ウチの人間は自己のまわりの家族やごく親しい人々、ソトの人間はごく親しくないが自己やウチと関連のある人々、ヨソの人間は

自己やウチとは関係がないがなにかのきっかけで関係をもちえる人々とする。」(三宅 1994 : 31) 『名大会話コーパス』には、ヨソの関係の会話は収録されていない。日本語では、相手がソトの人の場合とヨソの人の場合とで、言語行動に大きな違いがあると言われているので、親疎の差を、単に親疎の程度の違いと見るよりも、ウチとソトの関係の違いと見るほうがよいと考えた。

表 1 においてウチと呼ぶのは、恋人同士 (data092)、夫婦 (data090)、母と娘 (data042)、古くからの親しい友人 (data009) であり、絶対的な意味で親しい関係と言える。一方ソトは相対的である。ソトの程度が最も高いのは、data007 の初対面の仕事上の同僚同士のケースである。ソトの会話のうち、data007 と data035 では丁寧体が全面的に用いられている。この 2 つでは「話し手」は「相手」よりも年長で、かつ仕事上の先輩に当たる。一方、data041 と data101 では普通体が全面的に用いられている。この 2 つは学生同士の会話で、「話し手」と「相手」は社会的に対等であるが、年齢は「話し手」の方が下である。ウチの会話 (家族間) の data042 と data090 においては方言が全面的に使われている。

表 1 : 「話し手」と「相手」の属性と関係、会話の状況

データ番号	話し手	話し手の属性	親疎関係	2人の関係	会話の場所	相手	相手の属性	時間(分)
data092	F004	女性 20 代、兵庫県出身、愛知県在住	ウチ	恋人同士	車の中	M034	男性 20 代、東京都出身、韓国在住	59
data041			ソト	大学院の友人	相手の自宅	F091	女性 30 代、宮城県出身、愛知県在住	53
data090	F011	女性 40 代、大阪府出身、大阪府在住	ウチ	夫婦	自宅	M003	男性 40 代、大阪府出身、大阪府在住	60
data007			ソト	初対面の同僚	研究室	F089	女性 30 代、大阪府出身、愛知県在住	35
data042	F093	女性 20 代、岐阜県出身、	ウチ	母と娘	自宅	F097	女性 50 代、岐阜県出身、岐阜県在住	60

data101		岐阜県在住	ソト	大学院の同級生	大学院演習室	F057	女性 30 代、大阪府出身、名古屋市在住	31
data009	F098	女性 60 代、岡山県出身、愛知県在住	ウチ	学生時代からの友人	相手の自宅	F032	女性 60 代、東京都出身、東京都在住	84
data035			ソト	仕事上の知人	レストラン	M027	男性 40 代、熊本県出身、愛知県在住	59

本稿の 4 組のデータを、Brown and Levinson(1987:74)のフェイス侵害度の構成要素にあてはめると次のようになる。ウチとソトの違い以外に大きな差異はなく、ウチとソトの違いによるコミュニケーション・ストラテジーの違いの研究の資料として適切と考えられる。なおフェイスには、ポジティブ・フェイス（よいイメージを守りたい欲求）とネガティブ・フェイス（領域を守りたい欲求）の 2 面があるとされる。

- 「話し手」と「相手」の社会的距離 (Distance)には違いがあり、距離の小さいウチの関係と、距離の遠いソトの関係の間での比較が可能である。
- 「相手」が「話し手」に及ぼす力 (Power)には、大きな違いはない。
- 行為の負荷度 (Rating of imposition)に関しても、雑談のデータなので、依頼や謝罪があるわけではなく、大きな違いがあるとは言えない。しかし、発話行為の中にも行為の負荷度には差があり、たとえば「対立表明」は負荷度の高い行為といえる。

3.4 記述対象とする発話データ

本稿の考察は、上述の 4 人の「話し手」が使用し、以下の基準に当てはまる 277 の「よ」「ね」「よね」の例を対象とする。

(名詞、形容詞、形容動詞、動詞) + (「だ」、モダリティ、時制、否定などの表現) + (「よ」「ね」「よね」とその長音形) + 「。」

(例：「明日は浜松だよー。」「そんなこと言ってもしょうがないですね。」)

データは、文法的に「よ」「ね」、「よね」のどれとも置換可能な例に限られている。たとえば推測の形式の「だろう」には「よね」が接続しないので、「だろうね」はデータに含まれていない。また、異質の要因の介入を可能な限り防ぐために、以下も含まれていない。

1. 「のだ」を含む表現。「電話かけたのよ。」「試験だったのですね。」などは対象としない。「のだ」と終助詞の組み合わせは特別な意味を生むといわれているからである。
2. 同意を表す「そう（だ）」を含む表現。「そうね。」「そうですよね。」などは対象としない。定型句化の問題があり、特殊だからである。

3.5 親疎の程度と「よ」「ね」「よね」の使用

4人の「話し手」が使用する「よ」「ね」「よね」の生起頻度は図1の下部の表のとおりである。図1のグラフは、それを百分率に変換したものである。4人の「話し手」は全員が、ソトの人よりもウチの人に対して「よ」を多く使用している。「よね」は、逆にウチの人よりもソトの人に対して多く使用している。「ね」については、F093以外の「話し手」には、親疎による頻度の差を認められない。ウチの会話の中では、data042が特徴的である。F093はウチの人である母親に対して70%の高頻度で「よ」を用いている。一方ソトの会話の中では、data007が特徴的である。F011は、ソトの人である初対面の同僚に対して、「よね」を半数以上の発話で用い、「よ」の使用は少ない。

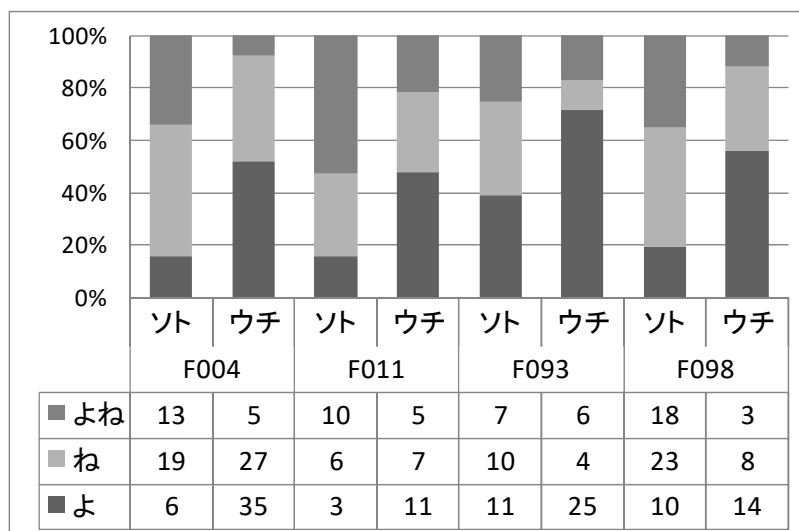


図1:親疎関係と「よ」「ね」「よね」の使用頻度

4 コミュニケーションの方略と終助詞の使用

4.1 発話行為、ポライトネス、表現

「よ」がウチの人間に対して多用され、「よね」がソトの人間に対して多用される理由については、次のような複数の解答を想定することができる。

1. 親疎の違いによって、会話の内容が異なる。
2. 親疎の違いによって、発話行為の種類が異なる。
3. 親疎の違いによって、ポライトネス・ストラテジーが異なる。

本稿では会話の内容に関する分析は行わない。また一見したところ、どの会話も内容に大きな違いはなさそうに思われる。

発話行為の種類は終助詞の選択と関係があると考えられる。発話行為の種類は、「一方的断定」「相手に関する断定」「対立表明」「同意表明」「同意要求」「確認要求」の6つにわけて観察する。明らかなのは、ウチの関係では「一方的断言」と「対立表明」が多く、ソトの関係では「同意表明」、「同意要求」、「確認要求」が多く「対立表明」が極端に少ないということである。

ポライトネス・ストラテジーも、終助詞の選択と関係があると考えられる。

以下は、滝浦(2005)による Brown and Levinson(1987)の紹介(p. 154)を一部改変したものである。直言とは、Grice(1975)が提示した 4 つの会話の公理に従った実践をいう。ウチの関係ではフェイス侵害度が高い行為も直言として表明されるが、ソトの関係では顧慮に満ちた方略が選択されると予想できる。

	顧慮の大きさ
(1) 直言 (軽減行為を伴わず、ありのままを明言する)	小
(2) ポジティブ・ポライトネス	↑
(3) ネガティブ・ポライトネス	
(4) ほのめかし	↓
(5) フェイス侵害行為をしない	大

なお本稿では、ポジティブ・ポライトネスを相手のフェイスを満足させる積極的かつ接近的な配慮、ネガティブ・ポライトネスをフェイス侵害行為の遂行の回避や迂言化などの消極的かつ離反的な配慮と定義して使う (Kerbrat-Orecchioni(1997))。

以下では発話行為ごとに、直言の場合に選択される終助詞について考察し、ポライトネス・ストラテジーに基づく表現法に関して、実例を挙げつつ記述していく。

4.2 一方的断言

「一方的断言」とは、話し手から聞き手へ向けて発せられる情報・主張の断定的な伝達を言う。直言の場合、この行為にもっともよく対応する終助詞は「よ」である。ウチの関係の会話では、「よ」による「一方的断言」が相手のフェイスへの顧慮なしに行われ、(1)から(3)のような例が数多くみつか

- (1) ウチ (data092) (明日の予定について)

M034 : 明日どうしてるかな。明日また家庭教だよ。

F004 : おー、なん、2時からか。

M034 : 2時から。うん。

F004 : ふーん、そっか。私明日浜松だよ。

M034 : あ、そうなんだ。

F004 : 忘れてたよー。

(2) ウチ(data009) (マンションの駐車場の様子について)

F09 : きょう、車と言え、お宅の駐車場で何か事故があったみたいよ、警察が来てたよ。なんか変な、前の、えーと、ライトがつぶ、つぶれてる大きな車があったよ。

F032 : あ、あの一、いたずらされたんじゃない？

(3) ウチ(data042) (ある狭い踏切の場所について)

F097 : (うん) そうすると、あの、踏切越えるところがあるやん。

F093 : うん、それはどこのことかわかるよ。私が言っとる踏切、違うところよ。もう、西可児駅のもうすぐ直前のところよ。

F097 : だからそこって、そこもせまいよっちゆうこと。

F093 : そこもせまいねえ。あっちは行かへん。

例(4)のように、「ね」を伴った「一方的断言」もある。ソトの「相手」に対する場合に比較的多い。「よ」のかわりに、一致志向の「ね」を使うこのような例は、ポジティブ・ポライトネスの方略によって、「相手」との情報の共有や共感を強調している。

(4) ソト(data041) (インターネット検索に必要な時間について)

F091 : そのそれぞれの、(うん) 情報を (あー) 3件なら3件とか。

F004 : うーん。絞り込むまで？

F091 : うーん。そう。

F004 : どうだろう。30分以上かかるねー。やっぱり。(うーん) か
かっちゃうねー。

例(5)はソトの関係の会話であるが、「よ」を伴う「一方的断言」が、ポジティブ・ポライトネスの方略に基づいて行われている。この発話では、称賛を目的とした一方的な強い主張がなされている。

(5) ソト(data035) (M027 が韓国で講演をすることについて)

M027 : 研究してる国の研究者がわざわざ、ね、海外に行って、話にきてくれたらうれしいもんですよ。

F098 : そうですね。韓国は、そりゃ韓国語までやってらっしゃるんだから。<笑い>それは珍しいですよ。

4.3 相手に関する断言

「相手に関する断言」は、相手に関わる情報の断定や評価である(例(6)、(7))。生起数は多くはない。「よ」ではなく「ね」が使われる。相手の領域内の情報についての断言は相手のネガティブ・フェイスを侵害するので、ソトの人に対してはなされにくい。ソトの関係の会話で、相手の領域内の情報について話す場合には、4.7 で見るように「よね」による「確認要求」の形が取られることが多い。

(6) ウチ(data009) (アメリカ大陸での待ち合わせについての昔話)

F098 : うん、じゃ、真ん中で会ったの？東から来て。

F032 : もちろん会った、うん。

F098 : は、すごいね。

F032 : デンバーで会った。

(7) ソト(data041) (赤ちゃんの性別についての話の後、「きましたねー」

は「お腹が大きくなってきましたねー」の意味)

F091 : ちょうど (うん) 検診に行ったときに (うん。向きが) 向きとかが (うん) 見えないところにいるみたいで (うんうん) わかんないみたい。うん。(ふーん、そっかー) そうなの。

F004 : うーん。結構あれだねー。きましたねー。へえー。うん。<笑い>へーえ。えーっ、でも変な感じじゃない？

4.4 対立表明

「対立表明」とは、直前の相手の発話に対して反対を表明する発話行為を言う。「対立表明」は相手のフェイスを脅かすため、ソトの人に向かって行うことは極めてまれである。一方、ウチの人に対しては、例(8)、(9)のように頻繁に行われる。この発話行為に伴われるのは「よ」である。

(8) ウチ (data092) (車の運転能力について)

M034 : <笑い>F004 だってペーパードライバーのくせに。

F004 : そんなことないよー。

M034 : ペーパードライバーじゃん？

F004 : いや、今ガンガン運転してるよ。

M034 : <笑い>ガンガン。

F004 : すごいよ。

M034 : どっちが上手かなー？

(9) ウチ (data042) (大学院の進級試験について)

F097 : 2年目に行けるかどうかちゅうのは、あるんでしょう、だから？

F093 : 進級試験ということ？ (うん) そんなやつあらへんよ。

普通に前期の試験あったらー、前期授業やって (うん) 前期の試験あったらー。

ソトの人に向けての「対立表明」は例(10)のような謙遜の場合のみに見られ、2例が確認できた。ネガティブ・ポライトネスの方略に分類できる。

(10) ソト (data035) (科学研究費補助金について)

M027 : F098 先生、毎年 [まいとし] 科研を。

F098 : いえいえ、去年取ってないですよ。

このほか 4.6 でみるように、ソトの関係において、「よね」による「同意要求」が、ポライトネスの方略による迂言的な「対立表明」として行われる場合もある。

4.5 同意表明

「同意表明」は、「対立表明」とは逆に、直前の相手の発話に対する同意を言う。「同意表明」の行為はフェイス侵害行為ではない。直言的に行うことによって、この行為は相手のポジティブ・フェイスを満足させる行為となる (cf. Kerbrat-Orecchioni (2011))。ソトの人に向けて頻繁に選択され、また何度も繰り返される傾向がある (例(11)から(14))。ウチの人に対しても「同意表明」は当然行われるが、「うん」などの一言で終わるのが普通である。「同意表明」には「ね」が伴われることが多いが、ソトの会話では「よね」も多い。場合によっては「よ」も選択されるが、そうするとフェイス侵害行為になる。

(11) ソト (data041) (インターネット検索について)

F091 : うん、すごく苦労して、(うん) 結局、(うん) 何にもわからなかったときもある。

F004 : あー、あるねー。よくある。(うーん) ん。それに何かやってみるうちに目的忘れちゃったりとかー。

(12) ソト(data007) (M大の学生がまじめであることについて)

F011 : 何か、世の中でよく人に聞くと、何か、もうそれこそお化粧しはじめたとかー。あの、いろんな大学にいろいろいるとね。

F089 : えー、そういうの全然。

F011 : そういうふうな話聞いたりもしたことあるけどー。

F089 : へえー、全然ないですよねー。

F011 : 全然ないですね。 (はい) うん。

(13) ソト(data101) (レポートを書くのに必要な日数について)

F057 : レポート 1 枚 2 日で書けるかな? (書けない。<笑い>) 書けないねー。前期わかったでしょ、って感じだね。<笑い> したらさー、1 週間ぐらいあったらさー、ぜんぶ。<笑い> 終わるよねー。

F093 : <笑い> 終わるよね。 <笑い>

(14) ソト(data035) (留学生が日本に来ないことについて)

M027 : 言語学で、生成文法なっちゃうと、ますますね、もうアメリカとかに行っちゃいますよね、みんなね。

F098 : 行っちゃいますよね。 なんかほかのこととか、そんなこと言ってもしょうがないですね。

「同意表明」に「一方的断言」の典型的マーカーの「よ」が使われるのは次の場合である。1 つには、(15)のように「話し手」と「相手」の情報量に差がある場合である。この例では自分の発話を繰り返しているともいえる。

(15) ソト(data035) (大学院の学生の専攻の偏りについて)

F098 : で、ま、それはいろいろ、その辺問題があつて、どうしてもや

っぱり日本語教育と現代日本語学が多いんです。

M027：多いですね。

F098：多いですよ。

ウチの関係の会話では、同意の要求に対して、まるで自分から持ち出した情報であるかのように「よ」で返答している(16)のような例もある。相手のフェイスを侵害する行為であり、特殊例と言える。

(16) ウチ(data090) (コンピューターに保存されている動画について)

M003：何でこんなところに入ってんねん。

F011：知らんよ、そんなもん。それがエラーになって、何かそれでも。

M003：ええやん、(うん?) 臨場感あるやん。

F011：臨場感あるよ。

4.6 同意要求

「同意要求」は、話し手が自分側の情報を聞き手に提示し、それに対する同意を求める発話行為である。「よね」、「ね」がこの発話行為に伴われる。例(17)から(20)は、直言的な「同意要求」の例である。

(17) ウチ(data042) (絵手紙について)

F093：Dのおばさん、やってたでしょう？

F097：そうそうそうそう。それでこないだEのばーばもそういうので送ってきたもんで、<笑い>絵手紙教室行きだしたなと思って。

F093：すぐわかるよね。(うん) おんなじような絵だもんね。(そうそう) ふーん。

(18) ソト(data101) (レポートについて)

F057：<笑い>だってもう1回あの一、(うん) 発表と同じぐらいの

時間をかけてさー、(うん) 調べてさー、レポート出すなんて
一、(うん) 大変だもん。

F093 : ねえ。 もう発表した人は、なしにしてほしいよね。 (ねー) そ
ういうのはだめなのかな。

(19) ソト(data035) (大学教員の研究休暇について)

M027 : アメリカみたいに自分の権利として取るという、そういう感じ
じゃないですよ。

F098 : ねえ。

M027 : だから取りたいけど、取れないってたくさんいる中で、こう、
半年とかずらして。

F098 : うん。 あれ、いいシステムですよ。 そうじゃなきゃ、もうほ
んとに、すり減るばかりですもんね。

(20) ソト(data041) (テレビの出演者について)

F004 : もろコメディだからね。 うん。 そうだねー。 うん。 うん。 なん
かさー、あのさ、アリー・マイラブのさー、(うん) あの女の
人、主人公の人。(うん) あの人さ、ちっちゃくて細いよね。
(うん) だからなんかさー、今までのイメージと違って、あれ
ー、ちっちゃくて細い人もうけるんだって思って、アメリカ人
の。(あー)

「同意要求」を行うためには、相手が同意に必要な情報をもっていることが前提となる。「同意要求」の行為は、「同意の行為を行う条件が相手に備わっている」と話し手が考えていることを相手に伝えることになるので、ポジティブ・ポライトネスの方略として使われることがある。以下の2つのソトの会話の例((21)と(22))では、「同意要求」の形式をとることで「私たちが共に知っているように」というメッセージを伝えて仲間であることを強調し

ている。しかし、「でも」の生起からわかるように反論も行っている。このような「よね」の使用は、「対立（の迂言的）表明」のためと考えることもできる。

(21) ソト(data035) (大学教員の研究休暇について)

M027 : H2さんも当たってないですよ。Bさんも。

F098 : でも、あそこはなんかうまく、半年ぐらい教えなくてもいいと
いうのをうまく回してらっしゃいますよね。

M027 : あ、そうですか。

(22) ソト(data007) (英語を教えることにした理由と職場について)

F089 : でもまあ英語もおもしろいかなって。うーん。そうしないと、
こっちに帰ってきてしまって、英語全然やってなかったら忘れて
しまうかなーと思ったんで。(うーん) ちょうどいいなっ
ていう感じですよ。

F011 : まあ、ここ、でも忙しいですよ。まあ、まだ来はったばかり
りやなので、まあそれほどすごく実感ないかもしれないけど、も
ういっそがしいところで。

4.7 確認要求

「確認要求」は、相手の領域の情報について、相手に確認を求める発話である。「よね」がこの発話行為にもっともよく適合する。直言的な「確認要求」は(23)から(25)のような場合である。

(23) ウチ(data042) (湯が沸いているかどうかについて)

F093 : ああ、危ない危ない。お湯わいとるよね。(うん) よし。ちょ
っとお母さん、ふた取って、そこ。ちょっと置いといた方がい
いか？

F097 : うん。

(24) ソト (data041) (安い航空券について)

F004 : そうなんだー。(うん) ふーん。そうかー。(うん) へー。ねえ、ねえ、ねえ、ねえ、あのさ、ANAだったよね。1万円のさ、チケット売り出してたんでしょ、(うーん) 25日に。あれどういうシステム？

(25) ソト (data007) (出身大学、大学院について)

F011 : F089 さんはたぶん関係ないでしょうね、そのCで勉強なさったことと今なさっていることと。

F089 : 全然。<笑い>

F011 : そうですね。だから大学院とかは関係ないですよね。その、Cとは。

F089 : Cは学部だけです。

「確認要求」はソトの関係の会話でよく見られる。相手の領域の内容を「ね」を使って直言的に断定することは、4.3 で見たようにフェイスを脅かすので、ソトの関係では断言を避けて確認の形をとる。このネガティブ・ポライトネスの方略は data007 の初対面の会話において特に頻繁に見られる。次例では、「話し手」は相手が英語の先生で忙しくて大変だと知っているにも関わらず、「大変ですね。」と断言せず、「よね」を用いて「確認要求」をしている。

(26) ソト (data007) (英語の先生の仕事について)

F089 : 毎日、ねー、続いてたらしんどいかなーと思ったりますんで。でも、来学期もなんか、月、水、金か、月、木、金か。

F011 : 英語はあれですよ。 (はい) 教えてらっしゃるんですよ。

(はい) 大変ですよ。

次も同様に、英語の先生に英語に関する話をするので断言を避けている。

「何となく英語って子音が多いですね。」よりも丁寧である。

(27) ソト (data007) (英語の特徴について)

F011 : それで一、まあ行くんですけど一、英語の、あの、えっと、テープとかをこうやって、一生懸命聞きながら。

F089 : あー、そうなんですか。

F011 : あの苦労してるんですけど。何となく英語って子音が多いですよ。

4.8 まとめ

表2は、ウチとソトの違いによる発話行為の出現傾向と、それぞれの発話行為を相手への顧慮なく直言的に行う際の典型的な終助詞の形式とを重ね合わせてまとめたものである。

表2 : 親疎の別による発話行為の出現傾向と典型的な終助詞の使用

	ウチ	ソト
1. 一方的断定	よ	
2. 相手に関する断定	ね	ね
3. 対立表明	よ	
4. 同意表明		ね、よね
5. 同意要求	よね	よね
6. 確認要求		よね

以下では、相手のフェイスへの顧慮を伴ったポライトネスの方略に基づく終助詞の使用も含めて解説する。

1. 「一方的断定」は、ウチの関係においては直言として「よ」を伴ってなされることが多い。ソトの関係では「よ」による断言が称賛の目的でな

されることもある（ポジティブ・ポライトネス）。

2. 「相手に関する断定」は「ね」を伴うが生起は少ない。ソトの関係では特に少ない。
3. 「対立表明」は「よ」を伴う。ウチの関係に多い。ソトの関係では、謙遜（ネガティブ・ポライトネス）の場合を除いて存在しない。「同意要求」の形をとったほのめかしとして表明される場合もある。
4. 「同意表明」はソトの関係で多く、ウチの関係で少ない。ソトの関係では、「同意表明」が「ね」「よね」を伴って何度も繰り返される。（ポジティブ・ポライトネス）
5. 「同意要求」の頻度に親疎の差は認められない。ソトの関係では、「一方的断定」のかわりに「よね」による「同意要求」が行われ、連帯が強調される（ポジティブ・ポライトネス）。
6. 「確認要求」はソトの関係に多い。ソトの関係では、「相手に関する断定」のかわりに「よね」による「確認要求」が行われ、断定を弱める（ネガティブ・ポライトネス）。

同一の話し手は、ウチの関係とソトの関係でコミュニケーションの方略を変えており、それが「よ」と「よね」の使用頻度に反映している。話し手がウチの関係において「よ」を数多く用いるのは、対立を恐れず、直言的なコミュニケーションを行うからである。ソトの関係で「よね」が頻繁なのは、同意表明を頻繁に繰り返し、対立や行き違いが生じないように確認要求や同意要求を行い、また、断言を避けるための方略としても確認要求や同意要求の形式が使われるからである。

以上、「よ」「ね」「よね」を伴う統語的に完結した文（「のだ」と「そうだ」を除く）とその文脈を観察し、親疎の違いによってコミュニケーションの方略がどのように異なるか、終助詞がそのためにどのように使用されているかを具体的に記述した。

5 おわりに

水谷 (1993)は「対話」と「共話」を対立させ、「共話」の特徴は、共通の理解を前提とし、いちいち相手の聞く意思をたしかめながら話すこと」であり、「対話」は「相手との共通の理解を前提とせず、相手の賛同や共感をとくに期待せず、しかも自分の意思や意見を相手に理解させることを目的として話すことである」(水谷 1993 : 9)という。本稿の結論はこの主張と矛盾している。なぜなら、共通の理解が存在するウチの会話においては、相手の賛同や共感を特に期待しない直言的断定が多いのに対し、共通の理解を前提としないソトの会話において同意や配慮のしるしが繰り返されるからである。

このような矛盾の理由は本稿のデータ分析の不備、特に、完結文に終助詞がついたもののみを対象データとしたこと、を含めて色々なことが考えられるが、1つには、水谷の「対話」はヨソの人をターゲットとし、「共話」はソトの人をターゲットにしている、ウチの人との会話が十分に観察されていないからかもしれない。この意味で、『名大会話コーパス』に含まれるウチの会話の資料はコミュニケーション研究に貢献できる可能性がある。また同時に、『名大会話コーパス』にはヨソの人、すなわち「通行人、電車などでまわりにいる人、サービス業の人など」(三宅 1994 : 31)との会話のデータが欠落していることも知っておく必要がある。

本稿においては、本章第2節で作成法を説明した文末表現データベースのデータのごく一部を観察した。『名大会話コーパス』からは文末表現以外にも種々のデータが採取できる。今後、研究がすすむことが期待される。

注

¹ 本書の執筆者のうちでは、藤村逸子、滝沢直宏、杉浦正利、大名力がこの共同研究に研究分担者として参加した。

² 現在、公開されている会話コーパスには、次のようなものがある。

現代日本語研究会編 (1997) 『女性のことば・職場編』 ひつじ書房 (約9時間)

現代日本語研究会編 (2002) 『男性のことば・職場編』 ひつじ書房 (約12時間)

Linguistic Data Consortium (1996) Callhome Japanese Speech (120件の電話会話)

国立国語研究所他編 (2004) 『日本語話し言葉コーパス』 (自由対話は 3.6 時間)

参考文献

- Brown, Penelope, and Stephen C. Levinson. (1987) *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Grice, Paul, (1975) Logic and Conversation in Cole, Peter, and Morgan, Jerry L. (eds.) *Syntax and semantics* 3. New York: Academic Press.
- Kerbrat-Orecchioni, Catherine (1997) A Multilevel Approach in the Study of Talk-in-Interaction, *Journal of Pragmatics* 7(1) pp. 1--20.
- Kerbrat-Orecchioni, Catherine (2011) From good manners to face-work: Politeness variations and constants in France, from the classic age to today in Bax, Marcel, and Dániel Z. Kádár (eds.) *Understanding Historical (Im)Politeness: Special issue of Journal of Historical Pragmatics* 12(1/2) pp. 133--155.
- 神尾昭雄(1990)『情報のなわ張り理論—言語の機能的分析』大修館書店.
- 水谷信子(1993)『『共話』から『対話』へ』『日本語学』12(4), pp4--10.
- 三宅和子(1994)「日本人の言語行動パターン—ウチ・ソト・ヨソ意識—」『筑波大学留学生センター 日本語教育論集』9, 筑波大学 pp. 29--39.
- 宮崎和人・安達太郎・野田春美・高梨信乃(2002)『モダリティ』くろしお出版.
- 日本語記述文法研究会(編)(2003)『現代日本語文法4 第8部 モダリティ』くろしお出版.
- 大曾美恵子(2005)「終助詞「よ」「ね」「よね」再考—雑談コーパスに基づく考察」鎌田修・畑佐由紀子・岡まゆみ・筒井通雄・ナズキアン富美子(編)『言語教育の新展開: 牧野成一教授古稀記念論集』ひつじ書房, pp. 3--15.
- 滝浦真人(2005)『日本の敬語論—ポライトネス理論からの再検討』大修館書店.

参考資料: 収集記録票

会話録音資料 収集記録票#

#

- 1. 収集者： _____#
- 2. 年月日： _____年 _____月 _____日 (_____曜) #
 時 間： _____時～ _____時#
- 3. 場 所： _____#

#

	例	A	B
氏名	山本太郎 (YT)		
性別	男		
年齢	23 歳		
職業	学生		
所属	名古屋大学 M2		
出身地	群馬県高崎市		
現住所	愛知県名古屋市		
現住所での居住期間	22 歳から現在まで 約 1 年 7 ヶ月		
これまでに最も長く 暮らした所、その期間	東京 4 歳から 22 歳まで 約 18 年間		
話し相手との関係	大学院の同級生		
連絡先電話番号	052-789-4787		

#

録音された会話を文字化し、プライバシーに関わる参加者の氏名、固有名詞等は伏せて電子化資料を作り、日本語・日本語教育研究の基礎資料として使用#
 することに同意します。#

署名 _____#

#

署名 _____#

#

ご協力、ありがとうございます。#

#

名古屋大学国際言語文化研究科教授：大曾 美恵子#