

OPI データに見る「話題」と「語彙」の関係

—OPI データを用いた語彙研究—

橋本 直幸

首都大学東京

要旨

本稿は、これまで研究の蓄積の乏しかった OPI データを使った語彙(実質語)研究について、新たな方法を提案するものである。語彙は、個人差や話題差による違いが大きく影響し、語彙研究をするにも大量のデータを必要とする。本稿では、この課題に対処するために、OPI の特徴を生かし、まず「話題」という単位でデータを切り取り、それによって、語彙の使用実態を把握しやすいよう試みる。さらに、個々の語の使用状況を観察するのではなく、その語が所属する意味分野の出現状況を観察する方法を提案する。これにより、その話題における使用語彙の実態を把握することができる。

【キーワード】 話題、語彙、意味分野、『分類語彙表』

1. はじめに

本研究は、OPI データを利用した語彙研究の方法について、提案を行うものである。OPI データに基づく文法(機能語)研究はこれまでに数多く行われており、多くの成果を残している。それに比べ、語彙(実質語)研究は、ほとんどないと言ってよい。

本稿では、OPI データに基づく語彙研究のひとつのアプローチとして、まずデータを「話題」という単位で分類し、その後、国立国語研究所により作成された『分類語彙表』⁽¹⁾の意味分野を利用して使用語彙を把握する方法を提案する。

2. 語彙と話題の関係

前田(1982)は、「語彙」を「ある特定の条件にかなった語の集まり」と定義し、そのうえで、以下のように述べている。

どういうところを「特定の条件」と考えるかによって語彙の範囲は変わってくる。特定の条件というのは語について考えうる条件であればどのような条件でも良いのである。そして、どういう目的で語彙を考えるかによって「特定の条件」というものも変わってくることになる。(p.9)

では、日本語教育という目的の下では、どのような条件で語をグルーピングすることが望ましいのであろうか。語のグルーピングということに関して、橋本(2008)では、『日本語能力試験出題基準』の五十音順語彙表を「話題」という単位で配列し直した語彙表を提案している。この考え方の背景にあるのは、「言語活動・タスクと言語形式の融合」という考えである(橋本・山内(2009)参照)。日本語学習者が目的とするところは、言語活動が行えるようになることであり、実際の授業ではそのための手段として言語形式を指導する。従って、この両者を結びつけたかたちで教えることが重要であり、橋本(2008)では、とくに

「語彙(言語形式)」と関連が深い「話題(言語活動)」を結びつけて提案している。

つまり、日本語教育という目的の下では、語彙を「話題」という単位でグルーピングすることが最も自然で望ましいと考えられる。幸い、OPI の評価基準である「ACTFL 言語運用能力基準—話技能」では、「話題」は「テキストの型」「正確さ」などとともに、「タスク・機能」の遂行能力を支える重要な要素の一つとして位置付けられている。「話題」という概念は、言語研究ではあいまいであると言われている⁽²⁾。しかし、OPI においては、「話題」に関する明確な定義こそないものの、テスター間では共通した認識があると考えられる。OPI の試験官養成用マニュアルでも、「OPI ではいくつもの話題をとりあげることが重要である。」と述べられていたり、「話題の数は二つでは少なすぎ、最低三つは必要」といった記述もある。つまり、OPI では「話題」というものに一定の単位を与えており、データとして切り取ってくることも可能であると考えられる。

また、OPI における話題の選択はもちろん被験者によって異なるが、しかし、その中でも「町の説明」「料理の説明」「映画のストーリー」「スポーツのルール説明」など、「定番」とも言える話題があることも事実である。

以上のように、OPI データは、「話題」という単位でデータを切り取って来ることが比較的行きやすく、また定番の話題に限定すれば、同質のデータを多く集めることができる。このことから、OPI データは、日本語教育における語彙研究に適したデータと言えるのではないだろうか。

3. 研究方法

3-1 語彙習得研究の壁

語彙習得研究について概観した山内(2004)では、第二言語習得研究の分野で行なわれている「語彙習得研究」の多くが、語彙学習ストラテジーの研究や、特定の語群の習得研究に偏っていて、「語のあつまり」としての語彙習得研究がほとんど行なわれていないことを指摘している。理由は言うまでもなく、実質語が個人差や話題差に大きく左右されるからである。山内(2004)では、このことを証明するために、調査データ(OPI 被験者)の数が増えるごとに、語彙的形態素数の異なり数がどの程度増えていくかということを、具体的なデータとともに示している。以下、表1は、OPI の上級話者 12 人分のデータをランダムに1データずつ増やしていき、どのように異なり語彙的形態素数が増えていくかを示したものである。

表1 データ数と異なり語彙的形態素数の関係 (山内 2004、p.157 表 11)

データ数	1人	2人	3人	4人	5人	6人	7人	8人	9人	10人	11人	12人
異なり語彙的形態素数	137	202	281	318	336	364	382	411	422	446	467	478

表1を見ると、データを足していけばいくほど、使用される語の種類(異なり語彙的形態素数)も増えていくことがわかる。山内(2004)では、この結果について「いくらデータを足しても、もうそれ以上、異なり語彙的形態素数が増えることがないという状態、つまり、

上級話者が使用し得る語彙的形態素数がすべて出尽くしているような状態になるまでデータを増やすことができればいいのであるが、そのデータ量とは、一体どのぐらいのものなのであろうか」(p.158)と述べ、実質語の習得研究に大量のデータが必要であることを指摘している。

この山内(2004)の試みは、上級話者が使用する実質語全体を OPI データから示そうとしたものであるが、本研究では、アプローチを変え、ある特定の話題に限定した中で、語の使用状況を把握する方法を考えてみることにしたい。

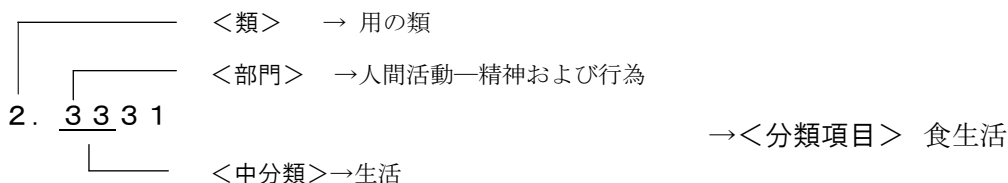
3-2 意味分野による使用語彙の把握

第2節では、学習者の語彙の使用実態を把握するのに、語彙の性質から、まず話題別に分ける必要があることを述べた。ここでは、もう一つの提案として、語自体の使用実態を把握するのではなく、その語が所属する意味分野を把握する方法を提案したい。例えば、「スポーツ」という話題に限定したとしても、サッカーのルールを説明するのか、相撲について説明するのか、あるいは、水泳について説明するのかで、使用される語は大きく変わってくる。話題を限定することである程度の絞り込みはできるかもしれないが、それでもやはり使用される語の増加は避けられない。そこで、本研究ではさらに絞り込むために、個々の語の使用実態を把握するということはせず、語が所属する意味分野の使用実態を把握することにする。

意味分野の決定は、『分類語彙表』の「分類項目」を利用する。『分類語彙表』では、<類><部門><中項目><分類項目>の4階層で分類され、分類番号が付されている。<類>は、分類番号の整数位にくる番号で、品詞論的に「1. 体の類(名詞の仲間)」「2. 用の類(動詞の仲間)」「3. 相の類(形容詞の仲間)」「4. その他の類」に分類されている。次の<部門>は小数点以下1桁目の番号で「1. 抽象的關係」「2. 人間活動の主体」「3. 人間活動—精神および行為」「4. 生産物および用具」「5. 自然物および自然現象」に分かれている。さらに、小数点以下1桁目と2桁目を合わせた<中項目>、その下の3桁目、4桁目を合わせた<分類項目>で細分化している。

例えば、「食べる」という語には、「2.3331」という番号が付いているが、以下のような構造を示しており、最終的に「食生活」という分類項目に分類されている。

「食べる」



本研究では、「話題」という単位で切り取った OPI データに、この分類項目を付与し、その使用状況を観察する。データは、ある程度の量が収集できることから、OPI で定番ともいえる「料理の説明」「スポーツの説明」について見ていく。

4. 分析

先に述べた方法で、OPI 中級から上級のタスクの中で定番とも言える「料理の作り方」「スポーツのルール説明」を例に、その使用語彙を把握する。

手順としては、OPI データ(KY コーパス^③)、『ACTFL-OPI 入門』収録データ^④、『インタビュー形式による日本語会話データベース』収録データ(非母語話者)^⑤、および個人的に採取した OPI データから該当する話題に関して話している部分各 10 データ(被験者 10 人分)ずつを抜き出し、そこで使用されている実質語に『分類語彙表』の分類項目の番号を付与する。『分類語彙表』に収録されていない語についても、該当すると考えられる番号を付ける。

まず、第 3 節で紹介した山内(2004)に倣い、異なり語の増加を見ていく。ただし、既に述べたとおり、語数の増加を見るのではなく、意味分野の数がどのように増えていくかを見る。以下、「料理の作り方」を表 2-a、「スポーツのルール説明」を表 2-b にそれぞれ示す。

表 2-a 話題「料理の作り方」におけるデータ数と異なり意味分野数の関係

データ数	1人	2人	3人	4人	5人	6人	7人	8人	9人	10人
異なり意味分野数	28	58	70	74	81	81	87	87	88	89
新しく追加される意味分野数		30	12	4	7	0	6	0	1	1

表 2-b 話題「スポーツのルール説明」におけるデータ数と異なり意味分野数の関係

データ数	1人	2人	3人	4人	5人	6人	7人	8人	9人	10人
異なり意味分野数	9	32	60	74	86	89	95	97	102	103
新しく追加される意味分野数		23	28	14	12	3	6	2	5	1

この表を見ると、わずか 10 のデータではあるが、異なり意味分野数の数はほぼ頭打ちとなり、新しく追加される意味分野の数が、後半あまり増えていないことが分かる。つまり、それぞれのタスクを達成するのに、必要な語の意味分野がある程度の段階で出尽くしていると考えられる。

では、具体的にそれぞれのタスクで、どのような意味分野の語が使用されているのだろうか。それぞれの話題で、延べ使用数の多かった意味分野を上位 20 位まで表 3-a、表 3-b に示す。表の中の、意味分野の番号は『分類語彙表』の分類番号を示している。品詞に関しては、意味分野と直接関係がないので、整数位は省略している(複数の品詞にまたがる場合は、語例のところを破線で区切って示す)。「延べ使用度数」は、その意味分野に所属する語が何回使用されていたかを示す。「データ数」は、調査対象となった 10 データ中、いくつのデータにその意味分野の語が使用されていたかを表している。

表 3-a 話題「料理の作り方」頻出意味分野と語例

意味分野		延べ 使用 度数	デー タ数	語例 (括弧内の語は、『分類語彙表』に収録されていない語)
.4330	調味料・こうじなど	25	9	調味料、塩、酢、砂糖、醤油、味噌、ソース、バター、コ チュジャン、(トマトソース、スパイスソース)
.4310	料理	17	8	中華料理、ロシア料理、スパゲッティ、(肉料理、四川料理、 インスタントラーメン、キムチインスタントラーメン、水 餃子、ガーリックブレッド)
.5050	味	17	8	味、下味 おいしい、甘い、辛い、ずっぱい
.3842	炊事・調理	15	8	料理、味付け 作る、炒める、炒る、焼く、煮る、沸かす
.5402	草木	10	5	にんじん、ピーマン、ジャガイモ、生姜、葱、カリフラワ ー、にら
.1532	入り・入れ	9	9	入れる
.1962	助数接辞	8	7	～cc、～分、～時間、～時
.4323	魚・肉	8	6	肉、牛肉、豚肉、鶏肉、手羽先、ソーセージ
.1920	程度・限度 程度	7	5	以外 適当、いちばん、とつても、(すごく)
.4520	食器・調理器具	7	5	鍋、フライパン、ナイフ、お皿
.2590	固有地名	7	4	日本、中国、韓国
.3020	好悪・愛憎	6	5	好み 好き、大好き
.1651	終始	6	5	最初、最後
.1800	形・型・姿・構え 形	6	3	形、丸、四角 丸い、四角い
.3331	食生活	5	5	夕ご飯 食べる
.5130	水・乾湿	5	5	水、お湯
.4320	米・ぬか・小麦粉な ど	5	5	野菜、小麦粉、(フラワー)
.1130	異動・類似	5	3	差 違う 同じ
.1220	成立	5	3	できあがり なる、できあがる
.1346	難易・安危	4	3	簡単、大丈夫

※20位(延べ使用度数4)は、ほかに「時機」「終了・中止」「接近・切断」「熱」が同数

表 3-b 話題「スポーツのルール説明」頻出意味分野と語例

意味分野		延べ 使用 度数	デー タ数	語例 (括弧内の語は、『分類語彙表』に収録されていない語)
.3374	スポーツ	21	10	試合、ゲーム、サッカー、野球、プロ野球、相撲、バスケットボール、(シルム)、オフサイド、シュート、ストローク、(スポットキック) 押す、キックする、パスする
.1962	助数接辞	16	8	～人、～点、～カ国、～分、～分間、～マイル、～キロメートル、～回、～面
.3570	勝敗	10	9	勝ち、あいこ 勝つ、(勝ち抜ける)
.1940	一般・部分・全体	10	8	全部、すべて、一部、部分、～ずつ 一応
.4570	遊具・置物・像など	10	8	ボール、バスケットボール
.1920	程度・限度 程度	9	6	以上、以外 一番、たいてい、(すごく)
.2200	相手・仲間	8	7	相手、敵、チーム
.2010	われ・なれ・かれ	8	7	私、みんな、誰、君
.3392	手足の動作	7	5	持つ、打つ、投げる、叩く
.2020	自他	7	5	自分、向こう、個人
.1742	中・隅・端	7	5	中、真ん中
.1532	入り・入れ	7	5	入る、入れる
.2450	その他の仕事	7	4	選手、プレーヤー、ゴールキーパー、コックス、ディフェンダー
.3430	行為・活動	6	5	する、やる
.1740	左右・前後・たてよこ	6	3	左、右、前、後ろ
.1651	終始	6	5	最初、最後、始め
.3700	取得	6	5	取る
.1710	点	6	3	ゴール、ゴールイン
.1901	値・額	6	6	点、点数、スコア
.1730	方向・方角	6	6	方、方向、向こう側 向かう

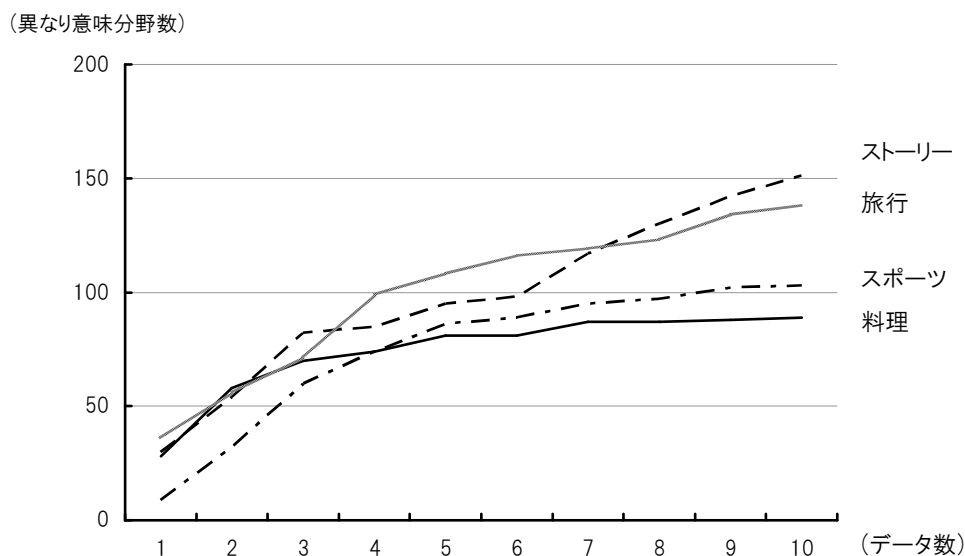
それぞれのタスク達成のためには、ここで示したような意味分野の語が必要となる。単純に語の出現頻度だけを数える方法では、大量のデータがなければ明らかな傾向は出にくい。意味分野を見る方法であれば、どのような語がタスク達成のために必要か見当をつけることができ、また個人差にも対応していくことができる。例えば、実際の授業で、クラス内にレベル差があった場合でも、意味分野を指導する方法をとれば、同じ意味でも「温める」「加熱する」「火を通す」など、難易度の違う語を学習者のレベルに応じ、柔軟に指導することもできる⁶⁾。

5. 課題 — 「語のまとめり」が見られない話題

前節までで述べた通り、30 分の OPI データをまとめて一つのデータとして扱うのではなく、まず話題で絞込み、その後、使用される語の意味分野の出現状況を見ることによって、少ないデータでもある程度の語のまとめりを観察することができるようになる。

しかし、話題によっては、この方法では語彙の使用状況を把握できないものもある。「料理の作り方」や「スポーツのルール説明」と同じく、中級から上級レベルの定番タスクでもある「ストーリー説明」「旅行の描写」などは、そのストーリーの内容や、旅行の内容によって、使用される語に大きな違いが出てくる。

「ストーリー説明」「旅行の描写」のデータ数と異なり意味分野数の増加傾向を、表 2-a 「料理の作り方」、表 2-b 「スポーツのルール説明」の結果と合わせ、グラフで示すと以下のようなになる。



グラフ1 話題別・データ数と異なり意味分野数の関係

この結果を見ると、第4節で述べたとおり、「料理の作り方」と「スポーツのルール説明」は、順にデータを増やして行っても、10人目のデータになるころには、新しく追加される意味分野というのはあまりなく、グラフが横ばいに近くなっていることがわかる。それに対して、「ストーリー説明」と「旅行の描写」は、10人目になっても異なり意味分野数は増え続けており、タスク達成に必要な意味分野の数が出尽くした状態だとは言いきれない。このように、話題によっては、必ずしも意味分野による使用語彙の把握という方法がうまくいかないものもあり、検討の余地がある。

ただ、このことから、同じ中級から上級へのタスクといっても、「料理の作り方」や「スポーツのルール説明」のように実質語の使用に一定の限定があるものと、「ストーリー説明」や「旅行の描写」のように「語のまとめり」を示せないものもあり、実質語がタスク達成に果たす役割が違うということがわかる。同じレベルのタスクであっても、その性質の違いについて考えていく必要があるだろう。

6. まとめ

本稿では、これまで研究の蓄積が少なかった OPI データを使用した語彙研究について、新たな方法の提案を行った。語彙研究は、最初に示した前田(1982)の指摘にあるように、

どの範囲で語のグルーピングをするかが重要になってくる。ただやみくもに大量のデータを収集するだけでは、本当に目的にかなった語彙研究はできない。OPI データは、話題をはじめとして、様々な面から統制のとれたデータであり、語彙研究にも広く利用できるものと考えられる。

また、本稿で行った意味分野による使用語彙の実態把握という方法も、本来、個人によって使用語彙、理解語彙が異なるという語彙の性質を反映したものと言える。

同じ OPI データを使った研究でも、語彙研究の場合、機能語研究とは根本的に違った方法を考えるべきである。

注

(1) 本研究では、『分類語彙表〈増補改訂版〉』(2004 年)の意味分類を用いる。本文中では、単に『分類語彙表』とする。

(2) 『応用言語学事典』(2003 年、研究社)「話題シラバス」の項には、以下のような記述がある。

このシラバスは一般の学習者にはわかりやすいことから、多くテキストやコースブックのシラバス・デザインに採用されているが、言語学者や言語教育者からは問題点が指摘されている。White(1988)は、多くの言語学者がこのシラバスは理論以前(pretheoretical)であると批判していると述べた上で、「話題」という概念があいまいであり、科学的検討の対象とはなりえないと述べている。(p.95)

(3) 「KY コーパス」は、平成 8～10 年度文部省科学研究費補助金による研究「第 2 言語としての日本語の習得に関する総合的研究」(研究代表者：カッケンブッシュ寛子)の成果の一部として作成された OPI の文字化資料。

(4) 『ACTFL-OPI 入門—日本語学習者の「話す力」を客観的に測る—』解説編では、実際のインタビューに即しながら各レベルの解説を行っており、本研究では、この解説用のインタビューデータを資料として使用した。

(5) 北九州市立大学情報メディア工学科上村隆一研究室によるプロジェクト。詳細は、<http://www.env.kitakyu-u.ac.jp/corpus/> 参照。

(6) 山内(2007)では、実質語の難易度について、paradigmatic に対立する語を集めることで、相対的に語の難易度を決定することができるとし、具体的に『分類語彙表』の同じ意味分類に所属する語どうしを比べ、語の難易度を決める方法を提案している。本稿の意味分野による使用語彙の実態把握の方法は、この指摘に示唆を受けたものである。

参考文献

国立国語研究所(2004)『分類語彙表〈増補改訂版〉』大日本図書。

橋本直幸(2008)「日本語教育版分類語彙表作成の試み」山内博之(編)『日本語教育スタンダード試案 語彙』9-91. ひつじ書房。

橋本直幸・山内博之(2009)「日本語教育のための語彙リストの作成」『日本語学』27-10, 50-58. 明治書院。

前田富祺(1982)「語彙」佐藤喜代治(編)『講座日本語の語彙 1 語彙原論』1-24. 明治書

院.

牧野成一監修・日本語 OPI 研究会翻訳プロジェクトチーム翻訳(1999)『ACTFL-OPI 試験官養成用マニュアル(1999 年改訂版)』アルク.

牧野成一他(2001)『ACTFL-OPI 入門 ―日本語学習者の「話す力」を客観的に測る―』アルク.

山内博之(2004)「語彙習得研究の方法 ―茶釜と N グラム統計―」『第二言語としての日本語の習得研究』7, 141-162. 第二言語習得研究会.

山内博之(2007)「日本語教育における語彙シラバスの作成について」『特定領域研究「日本語コーパス」平成 18 年度公開ワークショップ(研究成果報告会)予稿集』161-164.

White, R. V. (1988) *The ELT curriculum: Design, innovation, and management*. Blackwell.

付記

本稿は、日本語 OPI 研究会 研究発表会(2009 年 8 月 1 日、於(株)アルク)における口頭発表の内容を修正・加筆したものである。