

工学系話し言葉コーパスにおける日本語の使用実態 —使用頻度の高いサ変動詞の共起名詞を中心とした分析—

猪狩美保・岩崎夕子・菅谷有子・単 娜・古市由美子・村田晶子・山口真紀・山崎佳子

1.はじめに

現在日本における工学系の留学生や外国人研究員は増加の傾向にあり、工学系に特化した日本語教材が必要となってきたが、重田(2007)が示唆しているように、現時点では彼らの研究生生活を支える日本語教材は十分とは言えない。また工学系の場合、多くの留学生が来日後に研究室に所属し、専門的な語彙を含む難解な日本語環境に置かれているにも関わらず日本語学習には十分に時間をかけることができないのが現状である¹。

本研究グループでは工学系の留学生や研究員が効率的に日本語学習を進めていくために、工学系に特化した教材の作成を目指し、工学系話し言葉コーパスの構築および分析を行っている²。工学系の研究室で実際に使用されている話し言葉を集めたコーパスの分析を通して、留学生の研究生生活に必要な言語運用のサンプルを収集し、教材作成につなげることを目的としている。今回の分析では使用頻度の高いサ変動詞と共起する名詞を中心に分析を行い、共起名詞の特徴を明らかにすることにより、工学系に特化した教材の開発に向けての一助としたい。

2.先行研究

単他(2008)では工学系話し言葉コーパスにおける日本語の使用実態を動詞を中心に調査し、動詞の出現頻度やそれと共起する名詞、形容動詞、副詞について分析している。その結果として、特にサ変動詞に関しては専門分野にかかわる語彙が多いことや、これらの動詞と共起する名詞は2級以上が多いことなどが指摘されている。ただし共起名詞に関してはまだその特徴が明らかにされておらず、課題として残されている。また、日本語教育においては動詞と名詞を共起表現として提示する必要性について複数の先行研究ですでに言及されている(三國・小森 2008、大曾 2006、小宮 2005)。

3.研究課題と分析方法

3.1 研究課題

以上の先行研究を踏まえ、本研究では工学系話し言葉コーパスに現れたサ変動詞(出現頻度上位10位語)と格関係にある名詞を分析する。具体的な研究課題は以下の3点である。

- (1)格関係別に名詞がどのような分布になっているのか。
- (2)格関係にある名詞の語種はどのようなものになっているのか。
- (3)格関係にある名詞の使用領域はどのようなものか。

上記の3点から、サ変動詞と格関係にある名詞はどのようなものかを明らかにする。

3.2 コーパスの概要

今回の分析対象となるデータは電気系工学専攻のある研究室ゼミでの発表・報告を含む話し言葉を音声データとして収録したものである。収録期間は2007年11月から2008年2月までの4ヶ月間で計11回、総時間数は約20時間である。収録した音声データの日本語発話部分を文字化し、話し言葉コーパス(延べ226,194語)を作成した³。

3.3 分析方法

はじめに作成した話し言葉コーパスを分析ツールKH coderを用いて形態素を同定し、品詞別の出現頻度の結果に基づいて出現頻度の高いサ変動詞を抽出した⁴。次にサ変動詞と格関係にあるすべての名詞について用例を取り出し、格関係別に分類し分析を行った。

4.分析結果

工学系話し言葉コーパスにおける出現頻度の高いサ変動詞の上位10位は次の動詞である。

- ①説明する ②確保する ③計算する ④変化する
- ⑤設定する ⑥評価する ⑦発表する ⑧設置する

⑨制御する ⑩比較する

これらのサ変動詞と共に起る名詞を格関係別に分類し、今回は特に①「を格」②「が格」③「に格」④「で格」の4つの格関係に絞り分析を行った。

①「を格」には「Nを」の他、「を格」が「Nは(を)」、「Nも(を)」、「Nも(を)」として現れたものも含め、②「が格」には「Nも(が)」も含めた。③「に格」には「Nに(動作主)」、「Nに(場所)」、「Nに(対象)」④「で格」には「Nで(場所)」、「Nで(手段)」がそれぞれ含まれている。

4.1 研究課題1：格関係別

図1は格関係別の名詞の分布をグラフで表したものである。今回の分析対象であるサ変動詞10例のうち9例は他動詞で、1例「変化する」のみが自動詞だったため、「を格」関係が多く観察された。自動詞の「変化する」で「を格」が出現したのは、「変化させる」という使役形として用いられていたことによる。

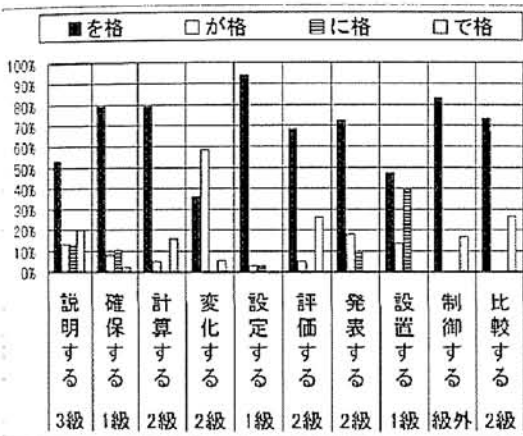


図1 格関係別の名詞の分布

4.2 研究課題2：サ変動詞と格関係を持つ名詞の語種

サ変動詞と格関係を持つ名詞を6つの語種別に分類した。語種は和語、漢語、カタカナ語、略語、複合語、その他(数字等)の6種類である。図2は名詞の語種別の分布を図で示したものであるが、これを見ると全体的に漢語名詞が比較的多く使われていることが分かる。また複合語の中にはカタカナ語と漢語の組み合わせが多く見られた。

4.3 研究課題3：名詞の使用領域

今回工学系留学生のための教材化への試みとして、サ変動詞と格関係にある名詞をその使用領域によって分類した。これは日本語学習にあまり時間をかけられない留学生や研究員が効率的に学んでいくため

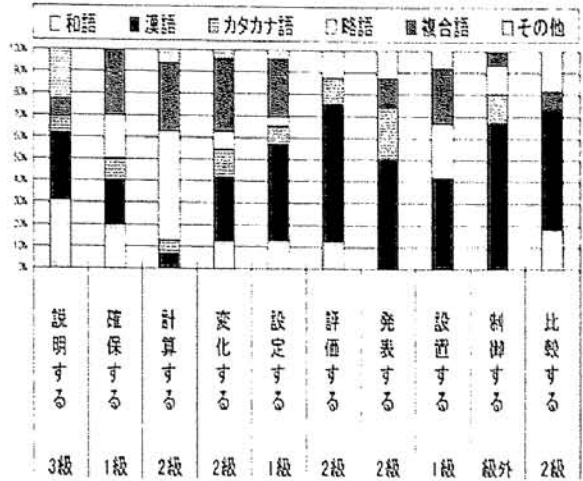


図2 名詞の語種別の分布

に、例えば研究室でよく用いられている語彙が日常的にも使われている語彙であれば日本語学習の早い段階から動詞と組み合わせて提示し学習していけるのではないかと考えたためである。

分類方法としてはまずサ変動詞と格関係にある名詞の使用領域を〈生活語彙〉、〈アカデミックな語彙〉、〈専門に特化した語彙〉と三つの分類基準を作り、その分類基準に基づいて日本語教師7名が出現した名詞の使用領域を判定した。その結果、使用領域の種類は以下の6種類となった。日常生活の中で使われることが多い①〈生活語彙〉(例：みんな/時間)、主にアカデミックな場面で使われる②〈アカデミックな語彙〉(例：相関性/有効数字)、特定の専門分野で使われる③〈専門に特化した語彙〉(例：制御系/発電機出力)の3つに加え、複数の使用領域にまたがって使用されると考えられる④〈生活語彙+アカデミックな語彙〉(例：結果/エリア)、⑤〈アカデミックな語彙+専門に特化した語彙〉(例：デルタ(H))、⑥〈専門に特化した語彙+生活語彙〉の計6種である⁵⁾。

なお、名詞の使用領域の判定及び検討の過程で日本語教師間で一致を見なかった名詞については、今回のサ変動詞と格関係にある名詞の使用領域分析からは除外した。その分析結果を図3に示す。

共起名詞の使用領域分布を見てみると動詞によってその分布傾向は異なっていることが分かる。例えば「確保する」、「設置する」、「計算する」と共起したのは〈専門に特化した語彙〉が多いのに対して、「説明する」、「設定する」、「制御する」は〈生活語彙〉との共起も比較的多くなっている。

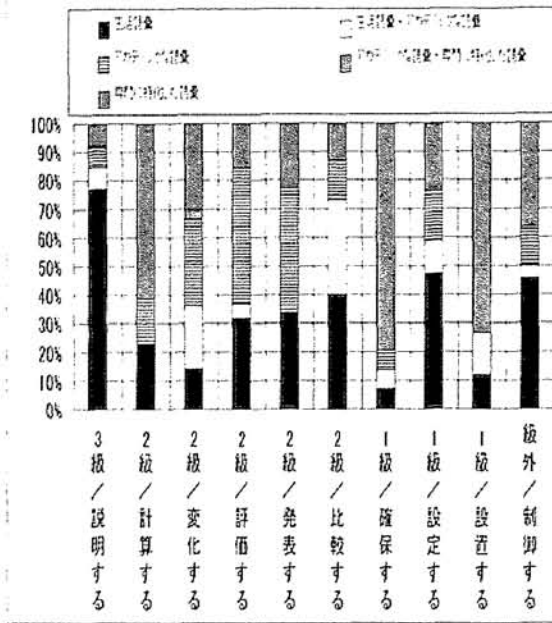


図3 共起名詞の使用領域分布

以下の共起名詞の使用領域分布図は、今回調査対象となった動詞の中から4つの動詞について共起名詞の使用領域の分布を具体的に示したものである。これを見ると、「設定する」、「計算する」は特に専門に特化した名詞との共起が多いことが分かる。また「計算する」は2級動詞だが、今回のデータでは格関係のある名詞は級外のものが多いという結果となっている。

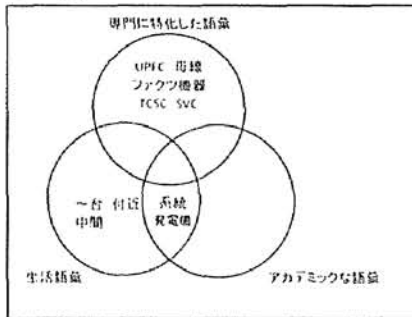


図4 共起名詞の使用領域別分布 (1級/設定する)

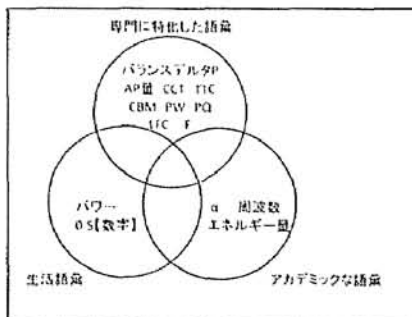


図5 共起名詞の使用領域別分布 (2級/計算する)

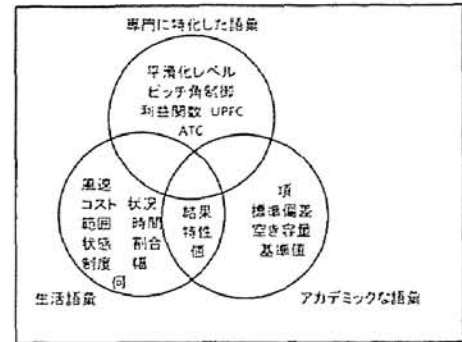


図6 共起名詞の使用領域別分布 (1級/設定する)

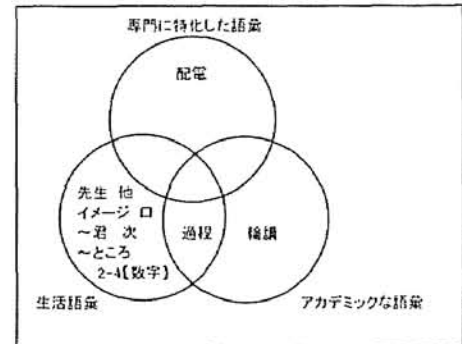


図7 共起名詞の使用領域別分布 (3級/説明する)

一方、1級動詞の「設定する」は〈専門に特化した語彙〉だけでなく、〈生活語彙〉との共起も多く見られた。また3級動詞の「説明する」も〈生活語彙〉の共起名詞が多く観察された。

共起名詞の使用領域と日本語能力試験でのレベルとの関係を見てみると、表1で示すような結果となった。共起名詞は全体で140例(延べ数)で、級外が55%、次いで2級が28%、1級を含めると約87%が2級以上の語彙となっていた。使用領域別に見てみると、〈生活語彙〉には3級、4級語彙が含まれているが、2級語彙が最も多く、他の4つの使用領域ではすべて2級以上であった。〈アカデミックな語彙+専門に特化した語彙〉、〈専門に特化した語彙〉では今回現れた名詞はすべて級外となっていた。

表1 共起名詞の使用領域と日本語能力試験でのレベル

	4級	3級	2級	1級	級外	計
生活語彙	11	5	20	1	8	45
生活語彙+アカデミックな語彙	0	0	8	2	4	14
アカデミックな語彙	0	0	11	3	19	33
アカデミックな語彙+専門に特化した語彙	0	0	0	0	1	1
専門に特化した語彙	0	0	0	0	45	45
計	11	5	39	6	77	140

※47例に「その他」2例を含む。

5.まとめ

以上の分析結果をまとめると次のようになる。

- (1)格関係別：「を格」関係が最も多く観察された。
- (2)名詞の語種：全体的に漢語名詞が比較的多い。
複合語はカタカナ語と漢語の組み合わせが多い。
- (3)名詞の使用領域：動詞によって傾向が異なる。

「確保する」、「設置する」、「計算する」は〈専門に特化した語彙〉と共起して使われることが多い。一方、「説明する」、「設定する」、「制御する」は〈生活語彙〉との共起も多い。

共起名詞の約 87%が級外を含む 2 級以上の語彙であった。〈アカデミックな語彙+専門に特化した語彙〉、〈専門に特化した語彙〉では今回現れた名詞はすべて級外の語彙であることが分かった。

以上の結果をもとにして、工学系留学生や研究員に対しては動詞、名詞ともに個別にレベル別に提出していくのではなく、共起関係のあるかたまりとして提示できるのではないだろうか。例えば 3 級動詞の「説明する」と共起している名詞の中には「イメージ」、「過程」（ともに 2 級）があるが、これらは工学系の留学生に対しては「説明する」と合わせ、「イメージを説明する」、「過程を説明する」という形で初級の学習段階から提示していくことも可能ではないかと考えられる。

これと同様に「設定する」は 1 級動詞であるが共起例として「時間(4 級)を設定する」が見られる。「時間」は日常生活でもよく使われる語彙であるため、このようなパターンもあまり日本語能力試験のレベルにとらわれすぎずに、日本語学習のより早い時期から動詞と名詞の組み合わせとして提示していただける可能性がある⁶。コーパスデータの分析結果をもとに動詞と名詞とを組み合わせ提示していくことによって、工学系の留学生が実際に身を置く環境に合った日本語の学習を効率的に進めていけるのではないかと考える。

6.今後の課題

今後の課題として、今回の分析では 4 つの格関係の名詞のみに絞って分析を行ったが他の格関係も含めた詳細な分析を行っていききたい。また語彙の使用領域の分類についても分類方法などを含め再度検

討する必要がある。さらに今回の分析では分析の対象が電気系工学専攻の研究室のデータに限られているが、今後工学系話し言葉コーパスの構築を目指して他の工学系分野のデータも収集し、他分野に跨った分析を進めていきたい。

注

1. 単他 (2008) 参照。
2. 本研究における「話し言葉コーパス」は電気系工学専攻のある研究室で収集した音声データを文字化し、コーパスとしてまとめたものである。ここでいうコーパスとは「言語資料の総体」である。
3. このうち日本語母語話者の発話数は延べ 221,336 語で、非母語話者の発話数は延べ 4,858 語である。非母語話者による英語の発話は今回の分析から除外してある。
4. KH coder は日本語のテキスト型データを計量的に分析するために開発されたツールであり、日本語教育の分野においてもその有効性が主張されている(佐野・李 2007)。その詳細については <http://khc.sourceforge.net/> を参照されたい。
5. 今回はⒺは出現していないため図 3 及び表 1 に含めない。
6. 橋本・山内 (2008) 参照。

参考文献

- 大曾美恵子 (2006) 「日本語コーパスと日本語教育」『日本語教育』130,3-10
- 小宮千鶴子 (2005) 『理工系留学生のための「専門連語」集の作成』平成 14~15 年度科学研究費補助金基盤研究(c)(2)研究成果報告書
- 佐野香織・李在鎬 (2007) 「KH Coder で何ができるか—日本語習得・日本語教育研究利用への示唆—」『言語文化と日本語教育』33, 94-95
- 重田美咲 (2007) 「工学を専門とする大学院留学生の「正統的周辺参加」と日本語使用」『2007 年度日本語教育学会秋季大会予稿集』113-118
- 早瀬・猪狩美保・菅谷有子・吉市由美子・山口真紀・山崎佳子・岩崎夕子 (2008) 「工学系話し言葉コーパスにおける日本語の使用実態—動詞を中心とした調査—」『2008 清華大学日本語学国際シンポジウム予稿集』140-142
- 富井篤 (2002) 『科学技術日英英日コーパス辞典』丸海
- 橋本直幸・山内博之 (2008) 「日本語教育のための語彙リストの作成」『日本語学』vol.27-10,50-58
- 三國純子・小森和子 (2008) 「コーパスを用いた論文作成のための慣用的共起表現の抽出」『小田記念日本語教育研究会論文集』16,55-68

いがり みほ・すがや ゆうこ・たん な・ふるいち ゆみこ・むらた あきこ・やまぐち まき・やまざき よしこ/東京大学、いわさき ゆうこ/元東京大学 miho_igari@hotmail.com