

### 英語論文コーパスを利用した専門語リスト作成の実践

小屋, 多恵子 / KOYA, Taeko

---

(出版者 / Publisher)

法政大学小金井論集編集委員会

(雑誌名 / Journal or Publication Title)

法政大学小金井論集 / 法政大学小金井論集

(巻 / Volume)

12

(開始ページ / Start Page)

69

(終了ページ / End Page)

82

(発行年 / Year)

2016-03-31

(URL)

<https://doi.org/10.15002/00014104>

# 英語論文コーパスを利用した 専門語リスト作成の実践

小 屋 多恵子

## 1. はじめに

本論は、法政大学理工学部創生科学科の選択専門科目「コーパス言語分析」で実施した専門コーパス構築とそれを利用した専門語リストの作成を意図した、アクティブ・ラーニング形態の授業実践を報告することを目的とする。創生科学科は、物理学・数学をもとに、世の中のさまざまな事象や現象を解明・理解しようとする一連の教育体系を科学のみちすじと名付け、それを柱に理系・文系の科学領域を横断したフィールドで学び、理系ジェネラリストの育成をめざしている。「コーパス言語分析」の授業は、3年生の秋学期に開講されるが、それはゼミ配属が決定し、研究をスタートさせる時期と重なるため、言語研究の演習と各自の専門研究に関わる演習をリンクさせる授業を目指した。授業は、キーワードであるデータ駆動型学習（Data-Driven Learning：DDL）、アクティブ・ラーニング、気づきをベースに、各自の研究領域において英語で執筆された論文を効率的に読み進めるために基本となる専門語リストを作成し、効果的な研究活動を促進させることにつなげていくことにした。授業の設計から実践への過程と共に、受講者が英語専門論文コーパスから語彙リストを作成する過程を通して気づいたコメントを合わせて報告する。

## 2. 先行研究

コーパスを使用した英語教育を進める実践例とその効果が昨今多く報告されている（岩井 2009, 西垣他 2015, 中條 2012）。教員が学習者に提示するのは英語の知識ではなく、コーパスのKWIC（Key Word in Context）であり、文脈の中か

ら単語の使用法や共起語などを発見していくこの帰納的な方法は、データ駆動型学習（DDL）と呼ばれている。この学習形態が報告されるようになった背景として、1つにはコンピュータやデジタルデバイスが発展したことにより情報教育が活発化し、教育において自ら発見していくデジタル環境が整ったことがある。ICT教育は今や小学教育から推奨され、学習者の関心・意欲・態度に効果があると文科省により報告されている。また、別の背景としては、これも文科省が推奨しているアクティブ・ラーニングや気づきといった学習における重要概念との関連によるものである。アクティブ・ラーニングとは、「学修者の能動的な学修への参加を取り入れた教授・学習法の総称。学修者が能動的に学修することによって、認知的、倫理的、社会的能力、教養、知識、経験を含めた汎用的能力の育成を図る。」（文科省『用語集』から）と定義され、発見学習や調査学習を教室の中で協同で実践する学習法である。そこで気づきを積み重ね、発見していくことが重要であるため、これを体言した1つの方法であるDDLが注目を集めているのである。

コーパスを英語教育に取り入れた例は多いが、コーパス言語学といった専門の授業において、いかにアクティブ・ラーニングや気づきを体言するのかといった報告は少ない。学習者のニーズに応えるため、コーパスを使用してどのような活動を設計していけばよいのか。DDL、アクティブ・ラーニング、気づきからどのように専門分野の研究に応用・貢献していけばよいのかを考え、今回の「コーパス言語分析」の授業を実践していくことにした。

### 3. 「コーパス言語分析」とは

#### 3.1 目的

「コーパス言語分析」は、法政大学理工学部創生科学科が掲げる「俯瞰的な科学の視点で社会の諸問題を理解・判断し、マネジメントできる人材の育成」に基づいて作られた創生科学科専門科目の1つである。文系理系といった伝統的な枠組みを超えた様々な視点から問題を自ら解決できる人材を育てる意図がある。この授業を受講することによって、ことばを科学的に分析する手法を習得し、分析した結果を自らの分野に応用するための道筋を学ぶことができるように設計した。

そこで授業は、基礎編と応用編の2部構成を取ることにした。基礎編の目的は、

(1) コーパスの定義・理念、コーパスの種類、コーパスの長所・短所、コーパスを使用してできることなどを理解すること、(2) コンコーダンスの使い方を学び、高頻度語・特徴語・コロケーション・クラスターなどを実践すること、の2つである。応用編の目的は、(1) 英語論文コーパスを構築すること、(2) 専門語リストを作成すること、の2つである。

### 3.2 受講者

「コーパス言語分析」は、法政大学創生科学科3年を対象に秋学期に開講される選択科目である。創生科学科の4つのフィールド（自然、物質、人間、知能）の中で、人間と知能フィールドを選択する学生に対する推奨科目であり、受講者数はおよそ50人である。受講生は、言語学に関する知識をほとんど持ち合わせておらず、言語が文理横断的な研究対象であることや長い歴史のある学問であることなど初めて知るものが多いが、毎年コーパスを用いた研究に興味を持ち、その技法を用いて卒論を書く学生が数人いる。また、「コーパス言語分析」の受講と同時期である3年の秋学期から各ゼミに所属し、研究をスタートさせていく。

### 3.3 使用したコーパスとコンコーダンス

基礎編で使用したのは、Brown family corporaである。これは、1960年代初頭の米語書き言葉をおよそ100万語集めたBrown Corpus, その英語版のLOB Corpus, さらに30年後の1990年代初頭の米語書きことばコーパスであるFROWN Corpus, その英語版のFLOB Corpusの4つのコーパスの総称である。このコーパスを使用すると、各時期の英米変種を比較する共時的的研究と、各変種で30年の間に起こった言語変化を調べる通時的的研究を行うことができる。これらのコーパスを分析するコンコーダンスは、AntConcを使用した。AntConcは、早稲田大学のLaurence Anthony氏が開発・発表しているフリーのソフトウェアである。豊富な機能の中で、KWIC表示, Clusters, Collocates, Word List, Keyword Listを使用した。

### 3.4 授業内容

全授業内容、授業の目的、学習形態は表1の通りである。

表1 スケジュールと授業内容

| 回数 |     | 内容  | 目的  | 学習形態      |
|----|-----|---|---|-----------|
| 1  |     | Introduction: 言語を科学するとは                     | ・身近なことばの現象について考え、言語研究とリンクさせる  | 全体授業・個別作業 |
| 2  | 基礎編 | コーパス言語学の基礎 (1)                              | ・コーパスの定義・理念、コーパスの種類、コーパスの長所・短所、コーパスを使用してできることなどを理解する                                      | 全体授業・個別作業 |
| 3  |     | コーパス言語学の基礎 (2)                              |   | 全体授業・個別作業 |
| 4  |     | 実践 (1) AntConc を使って分析してみよう：専門語リスト Lemma     | ・AntConc の Word List の使い方を学ぶ<br>・高頻度語・レマ化とはどういふものかを考察する                                   | 全体授業・個別作業 |
| 5  |     | 実践 (2) AntConc を使って分析してみよう：品詞タグ、BNC         | ・Go tagger <sup>1</sup> の使い方を学ぶ<br>・品詞タグのついた語彙リストを考察する<br>・BNC から話しことば高頻度語と書きことば高頻度語を考察する | 全体授業・個別作業 |
| 6  |     | 実践 (3) AntConc を使って分析してみよう：コロケーション          | ・AntConc の Collocates の使い方を学ぶ<br>・高頻度語とその共起語を考察する   | 全体授業・個別作業 |
| 7  |     | 実践 (4) AntConc を使って分析してみよう：クラスター            | ・AntConc の Clusters/N-Grams の使い方を学ぶ<br>・高頻度語のクラスターを考察する                                   | 全体授業・個別作業 |
| 8  |     | 実践 (5) AntConc を使って分析してみよう：特徴語              | ・AntConc の Keyword List の使い方を学ぶ<br>・特徴語を考察する  | 全体授業・個別作業 |
| 9  |     | まとめ   | これまでの復習   | 全体授業・個別作業 |
| 10 | 応用編 | Pilot study (1) :<br>英語論文コーパス構築             | ・10万語の専門コーパスを作成する   | グループワーク   |
| 11 |     | Pilot study (2) :<br>英語論文コーパス構築             | ・10万語の専門コーパスを作成する   | グループワーク   |
| 12 |     | Pilot study (3) : 専門語リスト作成                  | ・基礎編での実践と新たな指標 (JACET8000 <sup>2</sup> など) から独自の専門語リストを作成する                               | グループワーク   |
| 13 |     | Pilot study (4) : 専門語リスト作成                  | ・基礎編での実践と新たな指標 (JACET8000 など) から独自の専門語リストを作成する  | グループワーク   |
| 14 |     | Pilot study (5) : 専門語リスト作成                  | ・基礎編での実践方法と新たな指標 (JACET8000 など) から独自の専門語リストを作成する  | グループワーク   |
| 15 |     | Pilot study (6) : その他の分析 (コロケーション、クラスター)    | ・基礎編での実践方法から独自の専門語リストを作成する  | 個別作業      |
|    | 授業後 | レポート作成                                      | グループワーク・個別作業のデータを読み、自分なりに考察し、結論づける  | 個別作業      |
|    | 評価  | 提出レポートと基礎編の提出課題より評価をする。ポイントは、考察と気づき。テストはなし。 |   |           |

### 3.4.1 基礎編

まず、基礎編では、コーパス言語学の基礎を学習し、AntConcを使って実際にコーパスを分析し考察する実践練習を行った。コーパス言語学の基礎は、石川(2008, 2012)をベースにし、コーパスの定義・理念、コーパスの種類、コーパスの長所・短所、コーパスを使用してできることなどを解説した。この時間は、ともすれば、教師からの一方的な講義になってしまう傾向にあるため、受講者に課題を提示し、それを自分たちで調べる時間を設けた。例えば、代表的なコーパスとして、Brown corpusを紹介するときに、Brown Corpusの正式名称、代表編集者、コーパスサイズ、特長、問題点などを一定時間内でネットを使って調べる課題を入れた。続いて、実際にAntConcを使用して、コーパスから高頻度語・特徴語・コロケーション・クラスター分析を行った。これらの分析を通して、AntConcの使い方を学び、出てきたデータを考察した。高頻度語は、高頻度語にはどのような語があるのか、最も使用される動詞、名詞にはどのようなものがあるのか、レマ化すると結果にどのような違いがでるのか、品詞タグをつけた時に高頻度語はどの品詞で最も使用されるのか、などを分析結果から考察した。特徴語は、AntConcのKeyword List機能を使い、Brown family corporaの中から通時的比較か共時的比較を行い、考察を行った。コロケーションは、AntConcのCollocates機能を使い、Word Listを使用した分析結果の高頻度語である名詞がどのような語と共起するのかを調べた。クラスター分析は、コロケーションとの違いを確認した上で、Clusters機能の使用法を理解し、3-grams, 4-gramsで高頻度なフレーズを抽出した。以上が基礎編の内容であるが、受講者が、「調べる・分析する→考察する」を繰り返すことにより、データに基づいて自律的に思考できるように訓練した。

### 3.4.2 応用編

基礎編でコーパス言語学の基礎知識を理解し、実践を通してAntConcの使い方と分析・考察の仕方を学んだ受講生は、各自の研究テーマに沿った論文からコーパスを構築し、専門語リストを作成する課題に取り組んだ。ゼミ配属が決まり研究をスタートさせる時期でもあるため、自らが作成した専門語リストが英語で書かれた研究論文を精読していく際に、予め学習しておくことによって効率的な論文理解につながると同時に、リスト作成過程において、既知語であっても論文の

中で使用されている品詞が中学校・高等学校で学習したものとは異なっていたり、特殊な意味として使用されたりすることなどを気づくことが期待される。この専門語リスト作成はグループワークとし、同じゼミに所属をしている者または関連分野のゼミに所属している者同士で4,5名のグループを作り、作業を分担したり、話し合ったりしながら進めていくことにした。リスト作成は、(1)コーパス構築、(2)専門語の選定、(3)意味の付与の順で行われた。

### (1) コーパス構築

まず、グループ内でコーパスのソースを決定した。決定に際し、ゼミの指導教授に相談し、自分たちの研究に合った論文や本を紹介してもらった。また、コーパス作成のために、できる限り電子化された論文で、法政大学図書館の電子ジャーナルや電子ブックから論文が抽出できることを条件とした。さまざまなゼミに所属する学生でグループを作った場合には、すべての学生が取り組む研究の最大公約数である論文をソースにすることにした。

次に、コーパスを作成した。論文は過去3年間というように期間を限定し、そこからランダムに選択し、pdfファイルからtextファイル(ANSI)で保存した。語彙と構文を中心に分析調査する目的から、タイトル、著者の氏名・所属、小見出し、目次、謝辞、参考文献、別表(appendix)、写真、図、表、数式などは手作業で削除した。文字化けは元の論文を見ながら修正した。この作業によって、論文の本文のみをコーパス化することになる。専門コーパスの信頼性(Kennedy 1998, Nation 2011)をもとに、10万語以上のコーパスを作成させた。

### (2) 専門語の選定

コーパス構築後は、専門語リストに載せる専門語の選定を行った。そのもとになるのが、AntConcを使用した単語の頻度や特徴語指標とJACET8000、品詞タグ付き検索であった。手順は以下の通りである。

1. AntConcにより、高頻度語で特徴語指標の高いものを選抜する。特徴語検索には、参照コーパスとしてBrown family corporaを使用する。
2. JACET8000により、専門語が多く出てくる4000語以上の単語を中心に選抜する。ただし、高等学校までに学習した可能性がある4000語までの単語やJACET8000にカバーされない単語の中で、特に専門分野の論文を読むた

めに重要な語や、一般的な意味としては既習であるが専門的な意味として用いられている語、短縮形の語 (e.g. cp, svd)などはグループで話し合っけて削除しないようにする。

3. 多品詞語が多いため、品詞tagをつけてよく使われる品詞を特定する。

選定した専門語数は同じ作業を行ったとしても、専門分野やその論文によって異なることも想定されるが、150語以上をリストの目標として行った。

### (3) 意味の付与

選抜した単語の意味はコンコードダンスを見てどのような意味で使用されているかを吟味する。意味をつける際に、英辞郎、JST 科学技術用語日英対訳辞書、EFR 日英対訳辞書などを参照して付与した。

3つの行程から作成された2種類の専門語リスト例の一部を以下にあげる。

表2 専門語リスト例の一部 (テンソルの特異値分解に関する論文とEM アルゴリズムに関する論文から)

|    | 単語            | 頻度順位 | 特徴語順位 | POS   | 意味                        | JACET           |
|----|---------------|------|-------|-------|---------------------------|-----------------|
| 1  | tensor        | 18   | 2     | 名詞単数形 | テンソル                      | 欄外 <sup>3</sup> |
| 2  | decomposition | 45   | 8     | 名詞単数形 | 分解                        | 欄外              |
| 3  | tucker        | 68   | 13    | 名詞単数形 | タッカー                      | 欄外              |
| 4  | ontology      | 83   | 18    | 名詞単数形 | オントロジー                    | 欄外              |
| 5  | cp            | 150  | 35    | 名詞単数形 | 中央処理装置                    | 欄外              |
| 6  | svd           | 155  | 37    | 名詞単数形 | 特異値分解                     | 欄外              |
| 7  | hosvd         | 167  | 41    | 名詞単数形 | 高次元特異値分解                  | 欄外              |
| 8  | orthogonal    | 749  | 42    | 形容詞   | 直行の                       | 欄外              |
| 9  | nonnegative   | 182  | 48    | 形容詞   | 陽性の                       | 欄外              |
| 10 | symmetric     | 183  | 51    | 形容詞   | 対称な                       | 欄外              |
| 11 | xbf           | 226  | 53    | 名詞単数形 | 拡張可能なバイナリ形式               | 欄外              |
| 12 | nmf           | 241  | 57    | 名詞単数形 | 非負値行列を2つの非負値行列に分解するアルゴリズム | 欄外              |
| 13 | em            | 162  | 68    | 名詞単数形 | EM アルゴリズム                 | 欄外              |
| 14 | sparse        | 231  | 74    | 形容詞   | 希薄な                       | 欄外              |
| 15 | factorization | 275  | 77    | 名詞単数形 | 因数分解                      | 欄外              |
| 16 | multilinear   | 279  | 79    | 名詞単数形 | 可変多重線形                    | 欄外              |
| 17 | approximation | 227  | 82    | 名詞単数形 | 近似値                       | 欄外              |
| 18 | gaussian      | 285  | 88    | 名詞単数形 | ガウス分布                     | 欄外              |
| 19 | uniqueness    | 260  | 91    | 名詞単数形 | 唯一                        | 欄外              |
| 20 | taxonomic     | 315  | 96    | 形容詞   | 分類学上の                     | 欄外              |

表3 専門語リスト例の一部 (Progress of Theoretical & Experimental Physics = PTEP の宇宙背景放射についての論文)

|    | 単語           | 頻度 [回] | 特徴語順位 | POS   | 意味           | JACET |
|----|--------------|--------|-------|-------|--------------|-------|
| 1  | cmb          | 548    | 1     | 名詞    | cmb          | 欄外    |
| 2  | spectrum     | 485    | 2     | 名詞    | スペクトル        | 3     |
| 3  | model        | 634    | 3     | 名詞・動詞 | 模型 / 作る      | 1     |
| 4  | inflation    | 344    | 4     | 名詞    | 膨張           | 4     |
| 5  | parameter    | 372    | 5     | 名詞    | 媒介変数         | 4     |
| 6  | mode         | 336    | 6     | 名詞    | 様式           | 4     |
| 7  | field        | 351    | 7     | 名詞    | (電気・磁気などの) 場 | 1     |
| 8  | constraint   | 259    | 8     | 名詞    | 制御           | 4     |
| 9  | scale        | 371    | 9     | 名詞    | スケール         | 2     |
| 10 | scalar       | 230    | 10    | 名詞    | スカラー量        | 欄外    |
| 11 | fluctuation  | 242    | 11    | 名詞    | 波動           | 6     |
| 12 | polarization | 222    | 12    | 名詞    | 電極化 分極化      | 欄外    |
| 13 | lensing      | 208    | 13    | 名詞    | レンズ          | 欄外    |
| 14 | eq           | 228    | 14    | 名詞    | 方程式          | 欄外    |
| 15 | planck       | 202    | 15    | 名詞    | プランク         | 欄外    |
| 16 | density      | 228    | 16    | 名詞    | 比重           | 4     |
| 17 | distortion   | 199    | 17    | 名詞    | 歪曲           | 6     |
| 18 | temperature  | 294    | 18    | 名詞    | 温度           | 2     |
| 19 | emission     | 197    | 19    | 名詞    | 放出           | 3     |
| 20 | primordial   | 170    | 20    | 形容詞   | 始原の          | 欄外    |

### 3.5 評価

応用編の専門語リスト作成を中心にレポートにまとめ、それを評価した。レポートには、リスト作成だけでなく、作成時に考慮したポイント (e.g. 特徴語指数や頻度) や専門語リストの中で特に意味や品詞から注意が必要な語、そして作成後の考察を入れるように指示した。また、リストに出てくる語の共起語やフレーズを AntConc の Collocates, Clusters を使用して分析することを個人作業とし、その作業は任意とした。個人作業の分析や基礎編での課題提出を含めて総合的に評価した。

## 4. 学生からのフィードバック

専門語リスト作成のレポートには、学生の気づきが多く書かれている。学生にとってこの一連の作業がどのような気づきを与え、自分の研究分野と関わってい

るのかを内省する機会であると考え。学生が考えた内容は主に以下の3つに集約される。

#### ●専門語リストの重要性

専門分野の論文を読む際に予め頻出専門語を理解しておけば、効率よく論文を読み進めることができると感じている学生が多かった。自らコーパスを構築し語彙を選定する過程を通して、専門語を知覚し理解することができたのではないかと考えられる。

#### ●専門語リストの応用

語の選定過程を自ら体験したことで、興味がある研究や分野にコーパス分析が応用できることを理解した学生がいた。例えば、TOEICや自分の苦手な分野の英文コーパスから自分だけの単語帳を作成したり、自分の研究分野の論文を年代ごとにコーパス化することによって、その研究対象の歴史的遷移を明らかにしたりすることがあげられる。

#### ●専門語リスト作成過程での気づき

専門語リスト作成過程で、日常的に使用しない語が多かったり、知っていると思っていた語が通常使用する語の意味とは異なる意味で使用されていたり、1つの単語がさまざまな品詞として使用されていたりすることに気付いた学生がいた。また、語彙の面から考えると、「長い文=難しい文であり読むのに抵抗あり」ではないことに気づいた学生もいた。一方で、構文の複雑性を考える必要がある点や、コーパスサイズだけでなくコーパス構築の際の手作業の正確さを重要視し、コーパスの質を高めることが必要である点を指摘する学生もいた。

## 5. おわりに：まとめと今後の課題

本論は、文理の枠を超えた人材の育成を目指す学科の学生に「コーパス言語分析」という選択専門科目の授業をどのように展開していけばよいのかを思考し実施した過程を報告したものである。授業は、現在の教育現場に広く根付いているインストラクショナルデザインのプロセスであるADDIE（Analyze「分析」→

Design「設計」→Develop「開発」→Implement「実施」→Evaluate「評価」に基づいている。まず、Analyze「分析」として、学生のニーズ分析から始めた。受講者は、コーパスや言語分析ということばを初めて耳にする学生であり、知能・人間フィールドで受講を推奨されている科目であるためになんとなく登録したといった学生である。学生のニーズは、単位のために必要が主であろう。このような学生が50名から60名受講するのであるが、何か受講してプラスになる内容を著者は考える必要があった。そこで、受講者に共通のニーズである「ゼミに所属し、専門研究をスタートさせる上で必要になる英語論文を読むための専門語リスト」を作成することを主眼にした。次のDesign「設計」では、語彙リスト作成にDDL、アクティブ・ラーニング、気づきをベースにした。コーパス構築から専門語リスト完成まで、グループワークを通して学生が自ら考え、気づき、実践していくプロセスを大事に授業を組んだ。その前段階を基礎編とし、そこでコーパス構築と専門語リスト作成に必要な知識とソフトの使い方などを学ぶことにした。Develop「開発」では、毎時間のプリントと授業支援システムの設定を準備した。毎時間のプリントは、毎回の授業の目標、方法、考察の仕方など、欠席した場合でも後から1人で取り組めるようになるべくわかりやすいプリント作成を目指した。授業支援システムも、毎時間課題を取り組むのに欠かせないツールであるため、課題の指示やプリントの添付、毎時間の小テストの作成やその他の資料の添付などを行った。学生は大学からパソコンを貸与されているため、同じ機種のパソコンを授業で使用することができた。Implement「実施」は授業であるが、そこでは毎時間授業の目的とゴールを明確にし、自分が何故この課題に取り組んでいるのか、何ができたらよいのかを重ねて説明した。毎時間の授業の目的とゴールに集中しすぎると15回の授業を通して設定した目的とゴールがわからなくなることも考え、全体の目的であるコーパスの知識を利用して英語の論文を読むのに必要なリストを作成することを明示した。最後のEvaluate「評価」は、学生からのコメントを読むと、こちらの意図した通り、専門語リストの重要性と効果を理解し、基礎編で学んだことを駆使して専門語リストを完成することができたとする一方で、やはり何のためにリストを作成するのかわからず、させられている感じが拭えない学生もいたようだ。

このように、授業はある一定の成果をあげることはできたが、授業参加が最後まで受け身の学生がいる点を含めていくつか今後の課題も残った。授業に消極的

な参加をしている学生に対しては、TAと共に課題を行っている時に巡回しながら個人的な指導をしていく必要があると考える。50人以上の受講者に個人的な指導をするのはかなり困難ではあるが、積極的な参加をしている学生に対してはTAを中心に対応し、教員は消極的な参加をしている学生中心に対応することによって、よりきめ細かい指導を行っていくことができるであろう。また、提出されたレポートのリストを見ると、選択した語に明らかに誤っている訳をつけているものもあった。レポートは提出後に返却をしないので、誤っているままの専門語リストを受講生は持っていることになる。これでは、リストを十分に活用できない。また、リスト完成後にこれを利用しているかは調査できていない。作成した専門語リストを使用しなければ、作成時の気づきは続かず、研究への効果も薄い。これらをいかに改善していくかは今後の課題とし、再び稿を変えて報告することにした。

#### [注]

1. 後藤一章氏によって作成された英文テキストに自動で品詞タグを付与する無償のプログラム。
2. JACET8000は、日本人英語学習者のための学習語リストであり、8000語までの単語を1000語ごとにレベルを分けたものである。大学英語教育学会基本語改訂委員会が編纂し、British National Corpusと日本人英語学習者の環境を踏まえて独自に作成されたサブコーパスに準拠し、中高の教育現場の状況にも配慮して作成された。数詞、助数詞、大文字で始まる語（国名、地名、人名、月、曜日）はリストから除外されている。
3. 8000語のリストに入らなかった語を表す。

#### [参考文献]

- 中條清美他（2012）「米国Reading教科書と英語Graded Readersの英語初級者向けコーパスデータとしての適性に関する研究」『日本大学生産工学部研究報告B』45, pp. 29-42.
- 石川慎一郎（2008）『英語コーパスと言語教育』東京：大修館書店。
- 石川慎一郎（2012）『ベーシックコーパス言語学』東京：ひつじ書房。
- 岩井千春（2009）「コーパス言語学を応用したESP教育：発見学習に基づいた学習者自

律』『大阪府立大学紀要（人文・社会科学）』57, pp. 1-12.

Kennedy, Graeme (1998) *An introduction to corpus linguistics*. London/New York: Routledge.

文部科学省 (2012) 「アクティブ・ラーニング」『新たな未来を築くための大学教育の質的転換に向けて～生涯学び続け、主体的に考える力を育成する大学へ — 用語集』 Retrieved February 12, 2016, from [http://www.mext.go.jp/component/b\\_menu/shingi/toushin/\\_icsFiles/afieldfile/2012/10/04/1325048\\_3.pdf](http://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/afieldfile/2012/10/04/1325048_3.pdf)

Nation, I. S. P. (2011). *Researching an analyzing vocabulary*. Boston, MA: Heline

西垣知佳子他 (2015) 「中学校英語授業における言葉を観察する眼を育てるデータ駆動型学習の実践 — ペーパー版DDLからタブレット版DDLへの発展 —」『千葉大学教育学部研究紀要』63, pp. 287-294.

## Appendix: 「コーパス言語分析」受講後の学生の声

「コーパス言語分析」の最終レポートから気づきに関するコメントを抜粋した。コメントは以下学生が記した内容をランダムにそのまま記しているため、誤字脱字等がある場合もある。

### ●専門語リストの重要性

- ・自分がこの分野の英語論文を読むとき専門語リストを使えば効率よく読むことが可能だと考えられる。
- ・理系論文は専門単語を多く使用し、その単語の意味を理解している人にとってはスムーズに読めることができると考えられる。
- ・これから英語で書かれた論文を読む機会が増えてくると思うが、今回の課題を通して専門的な語彙を含む高頻度の語彙を学習することができてよかった。
- ・今回の作業を経て、これから扱う心理学の論文で出てくる単語の予習をすることができた。特に、特徴語や頻度順の高く、かつJACET8000のレベルが高いものをきちんと押さえておけば、これからは活かせると感じた。
- ・この専門語リストを作成することによって、専門分野の理解時のメリットがかなりあることがわかり、コーパス言語分析は有用性が高いことがわかった。
- ・一般的に使われている単語も宇宙論の知識を持っていないと間違えて理解して

しまうものも存在していた。ゆえに、天文学の知識が浅い人でも本実験の専門語リストをみることで、かなりすらすら読めるようになるといった意味では非常に意義のあるものとなった。

### ●専門語リストの応用

- ・論文を読むために専門語リストを作成したが、TOEICや自分の苦手分野の英文を用いて自分だけの単語帳が作れることが分かったので、ぜひ行いたい。
- ・n-gramを見ることで特殊な意味を持つ専門語リストの作成だけでなく、特殊な使われ方をしている単語の専門語リストの作成、ほかの統計手法による特殊語と比較などを行うことでより効果的な専門語リストの作成が期待できる。また、年代ごとにコーパスを作成することで、クラスタリング手法の歴史の遷移を理解できることが期待できる。

### ●専門語リスト作成過程での気づき

- ・単語レベルの分布をみると今回のコーパスは比較的簡単な単語から成り立っていることがわかるので、心理系の論文は読みやすい傾向になるのではないか。
- ・今回の分析結果として、専門語リストにでてきた語には日常的に使用しない語が多く、意味を調べないとわからないものが多く見られた。また、知っているような語も、通常使用する語の意味とは違う意味で使用されていることがあった。理系の論文を読む際にはこの専門語リストに出てきたような語を理解して読むと、意味・内容が捉えやすくなると考えられる。
- ・1つの単語から様々な品詞に変形していることがわかり、単語の覚え方にもある種参考になる部分もあった。
- ・この授業を受講する前までは、英語の長文が苦手だった。長い英語の文章は難しいという印象がなかなか抜けなかった。しかし、実際に英語の論文を分析してみると、文章に使われている単語は、親しみやすい語が多かったことがわかり、今後は英語の文章と向き合っていきたいと思う。
- ・今回題材にしたCMB（宇宙背景放射）、すなわち天文学の論文を読む場合、一定の宇宙論の知識が必要になる。たとえば、inflationという単語であると、一般的には金融などの急激な上昇を意味しており、日本ではインフレと用いられることが多い。しかし、天文学の論文となるとインフレーション理論に起因

して使われるものがほとんどである。

- ・コーパスの特性上、データの規模は大きいほど正確なデータが得られると考えられる。しかし量が多いだけで質が悪くなってしまうのも問題だと考えられる。少量のものであっても仲間同士で協力し、質を向上させることが良い結果につながるということがわかった。
- ・基礎レベルの単語が頻出しているとはいっても、複雑な文法で構成されている可能性があるので、安易に読みやすいと決めつけるのはよくないといえる。