

## 英語論文作成支援を目的とした日本人大学生の学習者コーパス構築

鈴木千鶴子, スーザン福島, 渡辺洋子, ○金城由美子, 吉原将太  
(長崎純心大学)

### 1 目的

本論文では、「英語卒業論文作成支援を目的とした学習者コーパス構築と教育システム開発」研究<sup>注)</sup>における以下の研究過程のうち、1)～3)に関する成果を報告する。

- 1) 英語卒業論文をデータとする**学習者コーパスの構築**
- 2) **学習者コーパスの母語話者コーパス等との比較対照による分析**
- 3) コーパス分析結果を基にしたレベル別習得項目の抽出と**教材コンテンツの精選・作成**
- 4) 教材コンテンツ・チュートリアルとコーパスデータ・レファレンスツールを組み込んだ教育システムの開発
- 5) 教育システム使用による学生の自立学習支援効果の評価

特に、Paul Nation の Basewords リスト(BWL)と、Averil Coxhead の Academic Word リスト(AWL)との比較対照により、①構築した日本人大学生の学習者コーパスのレファレンスツールとしての使用可能性の評価・検証、②Web チュートリアルに取り上げるべき Academic Words の教材コンテンツの項目抽出、を行う。

### 2 方法

#### 2.1 コーパスデータ

2001 年度～2005 年度入学 (2005 年 3 月～2009 年 3 月卒業) 学生 5 学年分の英語卒業論文提出分 212 編をデータ・ソースとし、各編ともタイトル・ページ、レファレンス、図・表を除いた本文 (Abstract, Introduction, Discussion, Conclusion の 4 セクションより構成) の全テキスト部分をコーパス化 (現段階で POS タグ無) した。

#### 2.2 対照コーパス

本報告において使用した対照データは、Paul Nation の Basewords リスト(BWL)および Averil Coxhead の Academic Word リスト(AWL)の 2 種類。

#### 2.3 分析ツール

コーパスデータのプロフィールを把握するために、i) Laurence Anthony の AntConc3.2.1w でレシマ処理による Word List, ならびに ii) WordSmith4.0 で、type/token ratios(TTR) を算出した。なお、主要目的の上掲①②の分析には、Paul Nation の Range32 を使用した。

### 3 結果

#### 3.1 コーパスデータ概要

上述の方法 2.1 ならびに 2.3 により、本学習者コーパスについて以下の概要が得られた。

表 1 学習者英語卒論コーパスのプロフィール

year	Papers no.	Word Type	Token	TTR	Stand. TTR
2001	42	7528	150991	6.49	34.97
2002	42	7031	183508	5.18	34.28
2003	41	8027	183724	5.74	36.15
2004	39	7861	180047	5.70	33.81
2005	48	9782	245386	5.19	34.89
Overall	212	19800	943656	2.70	34.82

注) 年度は入学年度, TTR (Standardized TTR 含む) は WordSmith4.0 により,  
 その他は, AntConc3.2.1w による。

### 3.2 BWL との比較

上述の方法 2.2 ならびに 2.3 により, 学習者コーパスにおける Basewords の割合について, 以下の結果が得られた。

表 2 学習者英語卒論コーパスの語彙リスト別分類

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
one	720078/77.49	3460/14.01	998
two	49580/ 5.34	2428/ 9.83	922
three	51184/ 5.51	2013/ 8.15	562
not in the lists	108441/11.67	16790/68.00	?????
Total	929283	24691	2482

Paul Nation により Basewords1 に挙げられる見出し語で 998 語類の全てが当該学習者卒論コーパスに含まれていた。同じく Basewords2 とされる見出し語類 988 の中で, 以下の語等 66 語は含まれていなかった。

aeroplane	arch	axe	baggage
bare	bribe	bunch	bundle
canal	carriage	centimeter	chalk
damp	ditch	donkey	drawer
envelope	fasten	fur	gallon
grey	hammer	knot	litre
loaf	lodging	log	metre
millilitre	millimetre	mud	nonsense
oar	paw	penny	pigeon
pint	procession	punctual	quart
rake	razor	repair	ripe
roar	roast	rug	saddle
sand	saws	scrape	shave

shield	shilling	soap	spade
spill	stove	tent	voyage
wax	whip	whistle, etc.	

### 3.3 AWL との比較

Basewords 3 の 570 語類中、以下の 8 語は学習者コーパスで使用されていなかった。その 570 語類は AWL として、Coxhead により更に 10 のサブリストに分類されている。そこで、使用されていなかった 8 語を、以下に、それぞれ AWL サブリスト上 Academic Word として頻度の高いとされるランクより順に挙げる。

	AWL Sublist ランク
parameter	4
infer	7
offset	8
practitioner	8
commence	9
albeit	10
notwithstanding	10
straightforward	10

## 4 考察

以上の結果から、特に上述 1. 目的の研究過程 4) の教育システム開発に向けた、レファレンスツールと、チュートリアルコンテンツに関する考察を行う。

### 4.1 レファレンスとしての使用可能性

本学習者コーパスは、先ずデータ量的に十分であると言える。且つ語彙の、特に単語単独出現の観察点からは、Academic Writing として、基本語彙の段階別包含率も妥当であり、更に Academic Words も 98.6% を包含しており、学習用モデルデータとしてレファレンスコーパス化するに値すると判断される。

### 4.2 チュートリアルの構成要素

チュートリアルとして、初めに BWL と AWL の内で学習者コーパスに含まれていなかった語彙について組み込むべきかを検討する必要がある。

BWL において欠けていた単語 66 語類は、殆どが数量単位や英語母語話者の日常生活環境に特異的に依存する（例えば海運、牧畜、農業、動物等）語彙であり、論文作成上特に指導の対象とすべきものとは思われない。

また、AWL で使用されていなかった単語 8 語類は、外来語や、基本語の類義語がほとんどであり、AWL サブリストにおいて上位語であるため緊急性は高くはないが、学術分野によっては parameter や infer など基本的な概念用語である故、テーマ、ジャンルによっては使用指導の有用性は否定できない。

特に、表 2 においていずれの Basewords List にも分類されない、token で 11% 強、type で 68%

を占める単語は、日本語のアルファベット音読表記を含め、いわゆるテーマ・トピック固有の用語である可能性が高く、現段階では個々の論文作成において学ばれるべきものであると判断されるが、ジャンル別分析に基づくレファレンスコーパスとチュートリアルを作成する場合には、より詳細な考察が必要であろう。

## 5 今後の課題

今回は、単語を単独に対象とした分析に止めたため、更に Word Clusters と Collocation の観点からの研究が必要である。また、チュートリアル項目として本課題の先行研究において明らかとなった接続語 (connectors) の貧弱性と否定表現 (negations) 用語の未発達を、チュートリアルで取り上げる際の具体的方法の検討が必要である。これらの課題の具体的方向性を以下に記し、本研究の結語とする。

### 5.1 頻出 Word Clusters の特定

特に英語母語話者コーパスで発表予定の Michigan Corpus of Upper-Level Student Papers (MICUSP)を対照コーパスとして、本学習者コーパスにおける Word Clusters を検討し、その観点からレファレンスコーパスとしての妥当性を再評価し、併せてチュートリアルにおける Word Clusters 項目を抽出する。

### 5.2 チュートリアルの例文作成

上記同様に MICUSP 等の母語話者コーパスを対照コーパスとして、本学習者コーパスの Collocation を検討し、適切な例文を作成する。

## 謝辞

Averil Coxhead に本研究を進めるにあたり、研究協力者として技術的、英語教育学上のアドバイスをいただいた。ここに謝意を表す。

注) 平成 21 年～23 年度文部科学省科学研究費補助金 (基盤研究 (C) 課題番号 21520621) による。

## 参考文献

- SUZUKI Chizuko, FUKUSHIMA Susan, WATANABE Yoko, KINJO Yumiko, & YOSHIHARA Shota (2009). Developing an independent learning system by compiling a corpus of English graduation papers written by Japanese university students. *The 7<sup>th</sup> Asia TEFL International Conference Abstracts*. (Bangkok, Thailand)
- Coxhead, A. (2008). Phraseology and English for Academic Purposes: Challenges and Opportunities. In F. Meunier and S. Granger (Eds.), *Phraseology in language learning and teaching* (pp. 149-161). Amsterdam. John Benjamins.
- Coxhead, A. & Byrd, P. (2007).; Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of second language writing*, 16: 129-147.