大規模コーパスを用いた日本人英語学習者用の語彙リスト構築と 教材分析システムの開発

Construction of a Vocabulary List for Japanese Learners of English and Development of a System for Analyzing Educational Materials Based on Large-scale Corpora

研究課題番号 16320076

平成 16-18 年度科学研究費補助金 基盤研究(B)研究成果報告書

平成 19 年 (2007 年) 3 月

研究代表者 相澤一美東京電機大学工学部教授

目 次

| 目次 | |
|---|-------------|
| はしがき | 3 |
| 0. 研究の背景と問題点 | |
| 0.1 研究の背景 | 19 |
| 0.2 研究計画 | 20 |
| 1. 学習語彙表の歴史 | |
| 1.1「JACET 基本語(第2次案)」(1983)刊行まで ············ | |
| 1.2 JACET 基本語リストの開発 ······ | 31 |
| 1.3 現在利用できる語彙表 | |
| 1.4「JACET 8000」(2003)の開発 ······ | 35 |
| 2. JACET 8000 の検証 | |
| 2.1 英語能力検定試験のテキストカバー率 | |
| 2.2 アカデミックテキストのカバー率 | |
| 2.3 学習者の受容語彙 | 45 |
| 2.4 学習者の発表語彙 | |
| 2.5 単語の難易度に関する英語教員の評価 | ····· 59 |
| 3. 既存コーパス資料の概観 | |
| 3.1 ICE-GB ····· | |
| 3.2 British National Corpus ····· | |
| 3.3 American National Corpus ····· | $\cdots 72$ |
| 3.4 Santa Barbara Corpus of Spoken American English | |
| 3.5 MICASE | 76 |
| 4. 新しいコーパス資料の整備と NEXT 1000 | |
| 4.1 米語コーパス | |
| 4.2 口語コーパス | 93 |
| 4.3 英米児童リーダー・コーパス | 99 |
| 4.4 教科書コーパス | 111 |
| 5. Supplement の開発 | |
| 5.1 Supplement の構築方法 | |
| 5.2 Supplement 636 の特徴 | 131 |
| 6. 語彙テスト開発の事例研究 | |
| 6.1 JACET 8000 語彙テストの開発(事例研究 1) | 134 |
| 6.2 JACET 8000 に基づく大学生受容語彙サイズテストの | |
| 開発と検証(事例研究 2) | 146 |
| 7. 研究の総括 | |
| 7.1 研究の限界 | |
| 7.2 今後の課題 | |
| 謝辞及び教科書発行者の略号と社名 | |
| 資料 1 Next 1000 Database ····· | 155 |
| 資料 2 JACET Supplement 636 Database ······ | 185 |

はしがき

本書は、平成 16-18 年度科学研究費補助金(基盤研究(B))による研究「大規模コーパスを用いた日本人英語学習者用の語彙リスト構築と教材分析システムの開発」の報告書である。

本プロジェクトは、大学英語教育学会(JACET)の英語語彙研究会の中心的メンバーによって計画立案された。本プロジェクトが最初に企画されたのは平成 14 年であったが、紆余曲折を経て平成 15 年に改訂され、平成 16 年から本格的に開始された。当初のメンバーは、相澤、上村、望月、投野、杉森、石川(敬称略)の 6 名であったが、望月氏がウエールズ大学での 1 年間の海外研修のため、5 名でのスタートとなった。この間、村田年、清水伸一の両氏にも、指導助言者としてプロジェクト会議に参加していただいた。

平成 17年度には、望月氏が海外研修から戻り 7名体制となった。当初は、JACET 8000 の修正 提案を行うことも念頭に置いていたが、プロジェクトの初期に行った JACET 8000 の妥当性検証により、現行のリストの高い有用性が確認された。そのため、 JACET 8000 を補完するための各種 の追加の語彙リストを作成することがプロジェクトの主たる目的となった。

平成 18 年度には、新たに村田氏を研究メンバーに迎えた。プロジェクトチームは、新たに整備した4つの大型サブコーパス(検定教科書、米語、米語口語、児童文学)から、それぞれ 1000 語の追加候補語を確定した。さらに、それらの中から、真に有用な約 600 語を精選し、 JACET 8000 Supplement として確定した。大規模コーパスに基づいた日本人上級英語学習者のための新たな語彙学習の指針として、これらの追加語彙資料は高い研究的・教育的価値を持つものと信じる。今後は、こうした追加リストに基づく教材や、教材分析システムの開発が望まれるところである。

この間、本プロジェクトが中心となって、以下の研究会やワークショップが企画・実施された。 これらは、幅広い内外の語彙研究者の集いの場となり、ひいては我が国の英語語彙研究の水準の向上に大きく資するものであった。

平成 16 年 8 月 科研プロジェクト公開講座(県立長崎シーボルト大学外国語教育センター) 平成 16 年 3 月 英語語彙研究会・英語辞書研究会合同ワークショップ(東京電機大学神田キャンパス) 平成 17 年 12 月 第 2 回英語語彙研究会研究大会(中央大学後楽園キャンパス) 平成 18 年 12 月 第 3 回英語語彙研究会研究大会(中央大学後楽園キャンパス)

3年間のプロジェクトでは、コーパスの整備やリストの作成など、高度に電子化された作業が必要とされた。その作業を中心になって進めてくれたのは、研究協力者の清水伸一氏である。また、大所高所からいつも適切な助言や励ましをいただいたのは、前回の JACET 8000 の編集時に JACET 基本語改訂委員会の委員長を経験された村田年氏である。また、研究代表者の手足となり、会議の雑用や資料の整備を手伝ってくれたのは磯達夫氏であった。深く謝意を表したい。

また、いくつかの教科書出版会社からは、今回のプロジェクトのために、有償無償で教科書の電子データの提供をいただいた。改めて深く御礼申し上げたい。

本報告書が、今後の日本の英語教育の実践や研究活動に、少しでも資することがあれば幸いである。

研究組織

研究代表者 相 澤 一 美 東京電機大学教授

(Aizawa, Kazumi) (Professor, Tokyo Denki University)

aizawa@cck.dendai.ac.jp

研究分担者 村 田 年 和洋女子大学教授

(Murata, Minoru) (Professor, Wayo Women's University)

murata.minoru@nifty.com

研究分担者 上 村 俊 彦 県立長崎シーボルト大学教授

(Uemura, Toshihiko) (Professor, Siebold University of Nagasaki)

yandi@sun.ac.jp

研究分担者 望 月 正 道 麗澤大学教授

(Mochizuki, Masamichi) (Professor, Reitaku University)

mmotizu@reitaku-u.ac.jp

研究分担者 投野 由紀夫 明海大学教授

(Tono, Yukio) (Professor, Meikai University)

y.tono@meikai.ac.jp

研究分担者 杉 森 直 樹 立命館大学助教授

(Sugimori, Naoki) (Associate Professor, Ritsumeikan University)

sugimori@mbox.kyoto-inet.or.jp

研究分担者 石川 慎一郎 神戸大学助教授

(Ishikawa, Shin'ichiro) (Associate Professor, Kobe University)

iskwshin@kobe-u.ac.jp

研究協力者 清 水 伸 一 安城学園高等学校教諭

(Shimizu, Shin'ichi) (Instructor, Anjo Gakuen High School)

shim@sk.aitai.ne.jp

研究協力者 磯 達夫 麗澤大学講師

(Iso, Tatsuo) (Assistant Professor, Reitaku University)

tiso@reitaku-u.ac.jp

(金額単位:円)

交付決定額 (配分額)

| | 直接経費 | 間接経費 | 合 計 |
|----------|--------------|------|--------------|
| 平成 16 年度 | 6, 600, 000 | 0 | 6, 600, 000 |
| 平成 17 年度 | 4, 800, 000 | 0 | 4, 800, 000 |
| 平成 18 年度 | 3, 100, 000 | 0 | 3, 100, 000 |
| 総計 | 14, 500, 000 | 0 | 14, 500, 000 |

研究発表

(1) 学会誌等

| | 十 五 恥 寸 | | <u> </u> | | | 1 |
|----|-------------------------|---------|--|---|-----|----------------|
| 番号 | 氏名 | 発表 年 | 論文名 | 雑誌名/書 籍名 | 号数 | ページ |
| 1 | Aizawa, K. & Iso, T. | 2007 | Estimating Word Difficulty: The Divergence from Frequency Levels | ARELE | 18 | 未定 |
| 2 | Iso, T. & Aizawa, K. | 2007 | The Role of Repetition and Context Richness in Inferring Word Meaning | Reitaku Journal of Interdiscipli nary Studies | 15 | pp.1-14 |
| 3 | 石川 慎一郎 | 2007 | L1/L2 コーパスの解析に基づく児童英語教育のための語彙マテリアル抽出システムの開発—小学校英語教育のための語彙選定の視点— | 紀要(中部 地区英語教 育学会) | 36 | pp.317- 324 |
| 4 | 石川 慎一郎 | 2007 | 日韓高等学校英語科教科書に見る使用語 彙の諸相:コーパス解析に基づく考察 | JACET 東 アジア英語 研究会研究 収録 | | 未定 |
| 5 | 石川 慎一郎 | 2007 | 多変量解析を用いた英語基本語抽出の試 み:主成分分析,因子分析,コレスポン デンス分析の応用可能性 | 統計数理研 究所共同研 究レポート | 199 | pp.1-22 |
| 6 | 石川 慎一郎 | 2006 | 日韓の中高英語教育における目標語彙水 準の経年的変化:1994~2005 年度大学 入試英語問題コーパスに基づく計量的調 査 | 神戸大学国際 コミュニケー ションセンタ 一論集 | 2 | pp.69-82 |
| 7 | 村田年・大崎さつき | 2006 | 「JACET8000」をスケールとして学習 指導要領の制限語彙及び中学校教科書語 彙を検討する | 和洋女子大 学紀要(人 文系編) | 46 | pp.83- 103 |

| 番号 | 氏名 | 発表 年 | 論文名 | 雑誌名/書 籍名 | 号数 | ページ |
|----|-------------------|---------|---|--|---------------|----------------|
| 8 | 上村 俊彦 | 2006 | JACET8000 と大規模英語コーパス | 県立長崎シ ーボルト大 学国際情報 学部紀要 | 7 | pp.309- 320 |
| 9 | Aizawa, Kazumi | 2005 | Assessing the Depth of Receptive Vocabulary Knowledge | JACET (Eds.), Facilitating Vocabulary Acquisition | | pp.18-28 |
| 10 | 石川 慎一郎 | 2005 | 大学生英語学習者の受容語彙力と発表語 彙力の関係:語彙サイズテストおよびエ ッセイ・コーパス分析に基づくアプロー チ | 紀要(中部 地区英語教 育学会) | 34 | pp.337-34 |
| 11 | 石川 慎一郎 | 2005 | 合衆国大統領選挙討論会における候補者 発話語彙の分析:コーパスに基づく批判 的談話分析の試み | 言語文化学会論集 | 25 | pp.3-15 |
| 12 | 石川 慎一郎 | 2005 | 日本人児童用英語基本語彙表開発における頻度と認知度の問題:母語コーパスと 対象語コーパスの頻度融合の手法 | 信学技報 (電子情報 通信学会) | TL200 5-25 | pp.43-48 |
| 13 | 石川 慎一郎 | 2005 | 米国司法 ESP 語彙表開発の試み: BROWN コーパスと米国司法文献コーパスの比較による特徴語の抽出 | 神戸大学国 際コミュニ ケーション センター論 集 | 1 | pp.13-28 |
| 14 | 石川 慎一郎 | 2005 | これからの英語語彙指導:単語の精選と 効果的提示-コーパス処理に基づく高校 生用英単語集の編纂をふまえて- | 明日の外国 語教育に同言 語文化教シン活動 (2004) (同志 社大学育研 究センター) | | pp.46-56 |

| 番 | | 発表 | m4 / 6 | 雑誌名/書 | H 184 | |
|----|----------------------|------|--|--|-------|------------|
| 号 | 氏名 | 年 | 論文名 | 籍名 | 号数 | ページ |
| 15 | 杉森 直樹 | 2005 | 日本人英語学習者が Speaking で使用する単語連鎖のレベル別比較-The NICT JLE Corpus を用いた比較 | The Proceedings of the NICT JLE Corpus Symposium | | pp.36-39 |
| 16 | 投野 由紀夫 | 2005 | 教材とコーパス | 立命館言語 文化研究 | 16(4) | pp.157-168 |
| 17 | 上村 俊彦 | 2005 | JACET8000 と電子版ニューヨーク・タ イムズ紙掲載の新刊書第 1 章の語彙 | 県立長崎シ ーボルト大 学国際情報 学部紀要 | 6 | pp.263-272 |
| 18 | Uemura, Toshihiko | 2005 | How Good Are Graded English Readers for ESL/EFL Students? | AsiaLex 2005 | | pp.338-344 |
| 19 | 石川 慎一郎 | 2004 | A Corpus-based Approach to the Synonymic Words of "Sorrow," "Grief," and "Sadness" | 言語と文学 (言語と文 学研究会) | 39 | pp.1-28 |
| 20 | 石川 慎一郎 | 2004 | 日韓の大学入試英語問題に見る構成語 彙の特徴:英文テキスト・コーパスの解 析に基づく考察 | アジアの英語と英語教育(大学英語教育学会中国・アジアの英語研究会) | 7 | pp.1-15 |
| 21 | 石川慎一郎 · 上村俊彦 | 2004 | JACET 8000 and Asian TEFL Initiative | The Journal of Asia TEFL | 1(1) | pp.333-347 |

(2)口頭発表

| | 中央儿女 | ı | | | |
|----|------------------------|---------|---|---|--------------------------------|
| 番号 | 氏名 | 発表 年 | 論文名 | 学会名 | 開催場所 |
| 1 | 相澤 一美 | 2007 | 英語読解力養成・テスト作成のためのリ ーダビリティの予測 | 慶應義塾大学 文学部英語読 解力テストフ ォーラム | 慶應義塾大学 |
| 2 | 石川 慎一郎 | 2007 | How to Estimate the Difficulty of Teaching Materials with JACET8000 Tools | Intensive Course in Corpus Linguistics 2007 | 明海大学 |
| 3 | 石川 慎一郎 | 2007 | 日韓高等学校英語科教科書に見る 語彙 統制の問題 | 第 64 回東アジ ア英語教育研 究会 | 西南学院大学 |
| 4 | 相澤 一美 | 2006 | 英語の語彙知識を科学する | 第9回流通科学 大学ドイツ語 教授法ワーク ショップ | 流通科学大学 |
| 5 | 相澤一美・磯達夫 | 2006 | 単語の難易度を予測できるか? - 単語の 頻度,難易度,教師の難易度評価の比較 - | 全国英語教育 学会第 32 回高 知研究大会 | 高知大学 |
| 6 | Aizawa, Kazumi | 2006 | Is Frequency of Words Related to Difficulty of Words? | BAAL/IRAAL 2006 | Cork University, Ireland |
| 7 | Aizawa, Kazumi | 2006 | Different Levels of Receptive Vocabulary Knowledge: How context and Multiple-choice Affect Vocabulary Test Score? | EuroSLA 2006 | Antalya, Turkey |
| 8 | Aizawa,K. & Iso, T. | 2006 | Reappraisal of JACET 8000 Based upon Web-Based VLT Scores | 大学英語教育 学会英語語彙 研究会第3回研 究大会 | 中央大学 |
| 9 | 石川 慎一郎 | 2006 | 言語研究における定量的視点と定性的視点:数字のその先へ—これからのコーパス言語学— | 大学英語教育 学会月例談話 会 | 大学英語教育学会本部 |

| 番号 | 氏名 | 発表年 | 論文名 | 学会名 | 開催場所 |
|----|--------|------|--|---|---------------|
| 10 | 石川 慎一郎 | 2006 | 発信語彙の選定手法―小学校英語教育用 語彙表 KUBEE の可能性― | 大学英語教育 学会英語語彙 研究会第3回研 究大会 | 中央大学 |
| 11 | 石川 慎一郎 | 2006 | 語彙サイズテストは英語総合力をどの程 度推定しうるか | 大学英語教育 学会英語語彙 研究会第3回研 究大会 | 中央大学 |
| 12 | 石川 慎一郎 | 2006 | An Introduction to Corpus Linguistics | 県立長崎シー ボルト大学外 国語センター 特別講義 | |
| 13 | 石川 慎一郎 | 2006 | Lexical Portfolio of Students' and Instructors' Presentations | 神戸大学国際 コミュニケー ションセンタ 一現代 GP シン ポジウム | 神戸大学 |
| 14 | 石川 慎一郎 | 2006 | 英語の語彙学習:経験から科学へ | Sel Hi 4 高等学校合同講演会 | 京都外国語大学 西高等学校 |
| 15 | 石川 慎一郎 | 2006 | An Introduction to Corpus Linguistics | 兵庫県教育委員会平成 18 年度英語教員集中研修 | 神戸大学 |
| 16 | 石川 慎一郎 | 2006 | L1/L2 コーパスの解析に基づく児童英語教育のための語彙マテリアル抽出システムの開発 —小学校英語教育のための語彙選定の視点— | 全国英語教育 学会第 32 回高 知研究大会 | 高知大学 |
| 17 | 石川 慎一郎 | 2006 | コーパスと学習英文法 | 平成 18 年度文 部科学省委 嘱・英語指導力 開発ワークシ ョップ | 関西学院大学 |
| 18 | 石川 慎一郎 | 2006 | 言語教育・研究のための統計:基礎の基 礎の基礎 | 立教大学英語 教育研究所講演会 | 立教大学 |

| 番号 | 氏名 | 発表年 | 論文名 | 学会名 | 開催場所 |
|----|---|------|--|--|-----------------------------|
| 19 | 石川 慎一郎 | 2006 | 英語研究・英文学研究とコーパス:新し い視点 | 日本仏語学会 シンポジウム 「文学・言語テ キストのコー パス分析・フラ ンス語・英語・ 日本語・」 | 慶應義塾大学 |
| 20 | 石川 慎一郎 | 2006 | 英語教育のためのコーパス<入門編>: 語彙指導と語法指導への視点 | 2005 年度外国 語教育メディ ア学会関東支 部講習会 | 大妻女子大学 |
| 21 | 石川 慎一郎 | 2006 | 日韓中等英語教育の達成水準について -試験コーパス分析の観点から- | 第 57 回大学英 語教育学会東 アジア英語研 究会 | 西南学院大学 |
| 22 | 石川慎一郎・倉 本充子・藪内 智・石川保茂・ 杉森直樹 | 2006 | 問題別討論会:語彙習得-新しい研究の 視点- Ⅲ 語彙指導とコーパス | 全国英語教育 学会第 32 回高 知研究大会 | 高知大学 |
| 23 | 石川慎一郎・村 田年・上村俊 彦・望月正道・ 清水伸一・杉森 直樹 | 2006 | 大規模コーパスを用いた JACET8000 の 拡張と精選 | 大学英語教育 学会第 45 回全 国大会 | 関西外国語大学 |
| 24 | Ishikawa, Shin'ichiro | 2006 | What Learners' Corpus Tells Us: A Quantitative Study on the Speeches by NS and NNS | CULI 2006 International Conference | Ambassador Hotel |
| 25 | Ishikawa, Shin'ichiro | 2006 | Vocabulary and Construction Seen in Japanese EFL Learners' Speech Presentation: A Corpus-based Approach | Asia TEFL International Conference 2006 | Seinan Gakuin University |
| 26 | 望月 正道 | 2006 | 語彙指導研究の視点から:語彙と読解は 相互依存 | 第 12 回日英・ 英語教育学会 | 成蹊大学 |
| 27 | 望月 正道 | 2006 | 前置詞の異なる語義の習得と辞書編集へ の示唆 | 第7回英語辞書 学ワークショ ップ | 和洋女子大学 |

| 番号 | 氏名 | 発表年 | 論文名 | 学会名 | 開催場所 |
|----|----------------------|------|---|--------------------------------------|--|
| 28 | 望月 正道 | 2006 | 新しい語彙構成テスト開発の試み | 第 32 回全国英 語教育学会高 知研究大会 | 高知大学 |
| 29 | 杉森 直樹 | 2006 | 大学生の英文エッセイライティング における使用語彙 | JACET 英語語 彙研究会第 3 回研究大会 | 中央大学 |
| 30 | 投野 由紀夫 | 2006 | コーパスと「100 語」が教えてくれた もの | 関西英語教育 学会 | 龍谷大学 |
| 31 | 投野 由紀夫 | 2006 | 人間の誤り、機械の誤り | アジア太平洋 機械翻訳協会 総会 | 総評会館 |
| 32 | 投野 由紀夫 | 2006 | 君にもできる「100 語でスタート!英 会話」 | 三重県高田中 学・高等学校 校内研修会 | 高田中学・高等学校 |
| 33 | 投野 由紀夫 | 2006 | 日本人の学習者コーパスから探る英 語習得項目の順序 | 長崎純心女子 大学 純心公開 講座 | 長崎純心女子大学 |
| 34 | 投野 由紀夫 | 2006 | コーパスの医学英語教育への貢献 | 日本医学英語 教育学会第 9 回総会 | 石川厚生年金会館 |
| 35 | 投野 由紀夫 | 2006 | コーパスを活用した効果的な語彙指導 | 福島県公立学 校教職員現職 教育 高等学校 経験者研修 | 福島県教育センター |
| 36 | 投野 由紀夫 | 2006 | 効果的な辞書指導 | 国際教育協議会 主催「英語科教育法集中講座」 | 三省堂本社 |
| 37 | Uemura, Toshihiko | 2006 | JACET8000 as an Evaluating Tool of English Vocabulary | 40th IATEFL | HarrogateInternational Conference Centre |
| 38 | 相澤 一美 | 2005 | テキスト読解と語彙学習 | JACET東北支 部 12 月例会 | 仙台市民会館 |

| 番号 | 氏名 | 発表年 | 論文名 | 学会名 | 開催場所 |
|----|-------------------|------|--|--|--------------------|
| 39 | Aizawa, Kazumi | 2005 | Mind the Gap: Discrepancies between Difficulties of Words and Their Frequencies | 大学英語教育 学会英語語彙 研究会第2回 研究大会 | 中央大学 |
| 40 | Aizawa, Kazumi | 2005 | Levels of Receptive Vocabulary Knowledge—The Role of Context and Multiple Choice in Vocabulary Test Scores | EuroSLA 2005 | Dubrovnik, Croatia |
| 41 | Aizawa, Kazumi | 2005 | Fostering Learner Autonomy of Learning Vocabulary through Mobile Phones | EuroCALL 2005 | Krakow, Poland |
| 42 | 石川 慎一郎 | 2005 | 日本人児童用英語基本語彙表開発における頻度と認知度の問題 | 電子情報通信 学会 思考と 言語研究会 | 早稲田大学 |
| 43 | 石川 慎一郎 | 2005 | ESP のためのコーパス分析入門: ESP の特徴語をどう科学的にとらえるか? | 大学英語教育 学会九州沖縄 支部 第 6 回 ESP研究会 | 宮崎大学医学部 |
| 44 | 石川 慎一郎 | 2005 | 英文科での学び: 見つめる目と見渡 す目 | 都留文科大学 英文科·英文学 会秋季講演会 | 都留文科大学 |
| 45 | 石川 慎一郎 | 2005 | 言語コミュニケーション文化学の新 しい展開:コーパスが拓く学際研究の 可能性について | 関西学院大学 大学院言語コ ミュニケーション研究会 | 関西学院大学 |
| 46 | 石川 慎一郎 | 2005 | 日加大学生作文コーパスの分析:語 彙,文法,文体の視点から | 大学英語教育 学会学習英文 法研究会 | 関西大学 |
| 47 | 石川 慎一郎 | 2005 | 2004 年米国大統領選における両候補 のディベート発話の比較: Corpus based Critical Discourse Analysis の視点から | 大学英語教育 学会英語辞書 研究会·語彙研 究会合同研究 発表会 | 東京電機大学 |
| 48 | 石川 慎一郎 | 2005 | 外国語教育・研究におけるコーパスの 有用性 | 島根大学法文 学部講演会 | 島根大学 |

| 番号 | 氏名 | 発表 年 | 論文名 | 学会名 | 開催場所 |
|----|--|---------|---|--|--------------------------------------|
| 49 | 石川慎一郎・相 澤一美・上村俊 彦・望月正道・ 清水伸一・杉森 直樹 | 2005 | 英語語彙研究の新展開 | 大学英語教育 学会第 44 回全 国大会 | 玉川大学 |
| 50 | Ishikawa, Shin'ichiro | 2005 | Corpus-Based Study on the Vocabulary Used by Japanese EFL Learners in Their Oral and Written Production | NICT JLE Corpus シン ポジウム | 京都セミナーハウス |
| 51 | Ishikawa, Shin'ichiro | 2005 | A Corpus-based Lexical Analysis of the English Section of the Daigaku Nyushi Center Test in Japan, the College Scholastic Ability Test in Korea, and the Scholastic Assessment Test in the USA | Foreign Language Education and Technology (FLEAT) 5th Conference | Brigham Young University (Utah, USA) |
| 52 | Ishikawa, Shin'ichiro | 2005 | Japanese EFL Learners' Use of the Vocabulary in Oral Production: A Study Based on The NICT JLE Corpus | 中部地区英語 教育学会第 35 回大会 | 山梨大学 |
| 53 | Ishikawa, Shin'ichiro | 2005 | Lexical Features of the English Composition by Japanese EFL Learners: A Quantitative Analysis of the Online Discussion Messages Posted by Japanese and Canadian College Students | ASIALEX 2005 Singapore International Conference | Hotel M |
| 54 | Mochizuki, Masamichi | 2005 | V_Links 2K: a new measure of lexical organization. | 大学英語教育 学会英語語彙 研究会第2回 研究大会 | 中央大学 |
| 55 | 杉森 直樹 | 2005 | N-Gram を用いた高頻度句動詞抽出 の試み | 第 6 回 JACET 英語辞書学ワ ークショップ | 東京電機大学 |
| 56 | 杉森 直樹 | 2005 | 大学生の英文エッセイに現れる語彙 フレーズの分析 | 第 31 回全国英 語教育学会札 幌研究大会 | 北海道教育大学 |

| 番号 | 氏名 | 発表 年 | 論文名 | 学会名 | 開催場所 |
|----|----------------------|---------|--|---|-----------|
| 57 | 杉森 直樹 | 2005 | 大学生の英作文における使用語彙の 分析 | 大学英語教育 学会第 44 回全 国大会 | 玉川大学 |
| 58 | 杉森直樹・望月 正道・相澤一美 | 2005 | JACET8000 に基づく CBT 方式の 語彙サイズテストの開発について | 大学英語教育 学会英語語彙 研究会第 2 回 研究大会 | 中央大学 |
| 59 | Uemura, Toshihiko | 2005 | How good are graded English readers for ESL/EFL students? | ASIALEX 2005 Singapore International Conference | M Hotel |
| 60 | 相澤一美・磯達夫 | 2004 | 繰り返しと手がかりが未知語推測に 及ぼす影響 | 全国英語教育 学会第30回長 野研究大会 | JA 長野県ビル |
| 61 | Aizawa, Kazumi | 2004 | Teaching and Assessing Receptive and Productive Vocabulary Knowledge | JACET Summer Seminar | 草津セミナーハウス |
| 62 | 石川 慎一郎 | 2004 | コーパスを利用した新しい英語語彙 指導 | 同志社大学言 語文化研究セ ンター講演会 | 同志社大学 |
| 63 | 石川 慎一郎 | 2004 | 法学系学生のための ESP 語彙表開発の試み:米国司法コーパスと FROWNコーパスの比較による特徴語の抽出およびESP語彙オンライン学修システムの構築に向けて | 言語文化学会 第 18 回大会 | 神戸大学 |
| 64 | 石川 慎一郎 | 2004 | 英語学習と語彙: コーパス処理技術を 応用した入試英単語集の編集につい て | 外国語教育メ ディア学会 (LET)関西支 部・大学授業研 究部会例会 | 流通科学大学 |
| 65 | 石川 慎一郎 | 2004 | 大学生英語学習者にみる受容語彙力 と発表語彙力の関係 | 全国英語教育 学会第30回長 野研究大会 | JA 長野県ビル |

| 番号 | 氏名 | 発表 年 | 論文名 | 学会名 | 開催場所 |
|----|--|---------|--|--|---|
| 66 | 石川 慎一郎 | 2004 | コーパスに基づく語彙表作成とその 授業への応用:各種ツールの活用法 | 外国語教育メ ディア学会 (LET)第 44 回 全国研究大会 | 西南学院大学 |
| 67 | 石川 慎一郎 | 2004 | コーパス研究の知見を英語教育にど う生かすか:コーパスを教える/コー パスで教える | JACET 関西支 部 2004 年度第 1 回談話会 | 京都外国語大学 |
| 68 | 石川 慎一郎 | 2004 | 日韓英語入試における使用語彙の計量的分析 | 第 45 回 JACET 東アジ ア英語教育研 究会 | 西南学院大学 |
| 69 | 石川慎一郎・村 田年・相澤一 美・上村俊彦・ 望月正道・清水 伸一・杉森直樹 | 2004 | JACET8000 に基づく語彙の応用研 究 | 大学英語教育 学会第 43 回全 国大会 | 中京大学 |
| 70 | Ishikawa, Shin'ichiro | 2004 | A Corpus-based Lexical Analysis of English Sections of College Entrance Exams in Japan and Korea | Asia TEFL International Conference 2003 | Olympic Parktel (Seoul, South Korea) |
| 71 | Ishikawa, Shin'ichiro | 2004 | Use of Computers for Analysis of English Usage: What Corpus Tells Us | 兵庫県教育委 員会平成 16 年 度英語教員集 中研修 | 神戸大学 |

(3) 出版物

| (0) | 山)及100 | | | | | | |
|-----|---|---------|---|---|---------|------------------|-----------|
| 番号 | 氏名 | 発表 年 | 論文名(書籍名) | 雑誌名/書 | 号数 | ページ | 出版社 |
| 1 | 石川 慎一郎 | 2007 | 英語教育のための基本語をどう選ぶか: コーパス言語各からの視点 | 英語教育 | 55(13) | pp. 10-12 | 大修館 書店 |
| 2 | 石川 慎一郎 | 2007 | 主要英語語彙表の概観 | 英語教育 | 55(13) | p.13 | 大修館 書店 |
| 3 | 望月 正道 | 2007 | テストで測れる語彙力,測れない語彙 力 | 英語教育 | 55(13) | pp. 31-33 | 大修館 書店 |
| 4 | 相澤 一美 | 2006 | 語彙習得をどう捉えるか | 言語 | 35(4) | pp.32 -37 | 大修館 書店 |
| 5 | Aizawa, Kazumi | 2006 | Rethinking Frequency Markers for English-Japanese Dictionaries | Murata, M., Minamide, K., Tono, Y., & Ishikawa, S. (Eds.) | | pp.82 -93 | 大修館書店 |
| 6 | 石川 慎一郎 | 2006 | コーパスと直観の融合: G4 への期待 | GCD 英語通 信 (大修館書 店), 40 | 40 | pp.12 -13 | 大修館書店 |
| 7 | Mochizuki, M. | 2006 | Acquisition of different senses of prepositions and its implications for lexicography | Murata, M., Minamide, K., Tono, Y., & Ishikawa, S. (Eds.) | | pp.26 2-273. | 大修館書店 |
| 8 | 村田年・望月正 道・相澤一美 (編) | 2006 | 『JACET8000 英単語チェックメイト』 | | | 全 215p | 桐原書店 |
| 9 | Murata, M., Minamide, K., Tono, Y. & Ishikawa, S. | 2006 | English Lexicography in Japan | | | 全 326p | 大修館書店 |
| 10 | 投野 由紀夫 | 2006 | Learner's Dictionary 概観 | 日本語学 | 2006(7) | pp.2 3-3 7 | 明治書院 |

| 番号 | 氏名 | 発表 年 | 論文名 | 雑誌名/書 籍名 | 号数 | ペー ジ | 出版社 |
|----|-----------------------------------|---------|---|---|-------|----------------|--------------------------------|
| 11 | Noguchi, J., T. Orr, & Y. Tono | 2006 | Using a Dedicated Corpus to Identify Features of Professional English Usage: What Do "we" Do in Science Journal Articles? | Wilson, A., Archer, D. and Rayson, P. (Eds.) Corpus Linguistics Around the World. | | pp.15 5-166 | Rodopi |
| 12 | Tono, Yukio | 2006 | English Bilingual Lexicography in Japan: Meeting Serious Challenges | Murata, M., Minamide, K., Tono, Y., & Ishikawa, S. (Eds.) | | pp.18 -25 | 大修館書店 |
| 13 | 相澤 一美 | 2005 | 日本人のための英単語学習法 | 日英英語文化 研究会(編)『日 英語の比較 - 発想・背景・文 化』 | | pp.14 9-151 | 三修社 |
| 14 | 相澤一美/石川慎一郎/村田年(編) | 2005 | 『「大学英語教育学会基本語リスト」 に基づく JACET 8000 英単語』 | | | 全 503p | 桐原書店 |
| 15 | 石川 慎一郎 | 2005 | 英和辞書と重要語ラベリング:科学と 教育の接点 | 英語教育 | 54(5) | pp.52 -53 | 大修館 書店 |
| 16 | Tono, Yukio | 2005 | Corpus-Based SLA Research:State of the Art of Learner Corpus Studies | Minami, M., Kobayashi, H., Nakayama, M. & Sirai, H. (Eds.) Studies in Language Sciences | 4 | pp.45 -77 | Kurosh io Publis hers |

| 番号 | 氏名 | 発表 年 | 論文名 | 雑誌名/書籍 | 号数 | ページ | 出版社 |
|----|-------------|---------|--|--|----|----------------|-------------------------------------|
| 17 | Tono, Yukio | 2005 | Exploring the potential of learner corpora for pedagogical lexicography | Ooi, V. et al. (Eds.). Words in Asian Cultural Contexts: Proceedings of the 4th Asialex Conference. | | pp.31 5-321 | Nation al Univer sity of Singap ore |
| 18 | Tono, Yukio | 2005 | Japanese Lexicography | K. Brown (Ed.) Encyclopedia of Language and Linguistics. | 6 | pp.10 5-109 | Elsevie r |
| 19 | 石川 慎一郎 | 2004 | 『VITAL 4500 英単語・熟語』 | | | 全 351p | 文英堂 |
| 20 | Tono, Yukio | 2004 | Research on the Use of Electronic Dictionaries for Language Learning: Methodological Considerations | Campoy, M.C. & Safont, P. (Eds.) Computer- mediated Lexicography in the Foreign Language Learning Context. | | pp.13 -28 | Univer -sitat Jaume -I |
| 21 | Tono, Yukio | 2004 | Multiple Comparisons of IL, L1 and TL Corpora: the Case of L2 Acquisition of verb Subcategorization Patterns by Japanese Learners of English | Aston, G., Bernardini, S. & Stewart, D. (Eds.) Corpora and Language Learners. | | pp.45 -66 | John Benja mins |
| 22 | Tono, Yukio | 2004 | Learner Corpus Research: State of the Art and Applications in SLA Research | Samad Arshad Abd. (Ed.) Perspectives on Grammar in Language Education | | pp.12 6-157 | Univer siti Putra Malays ia Press |

(東京電機大学・相澤一美)

0. 研究の背景と問題点

0.1 研究の背景

日本における従来の英語学習語彙表は、ほとんどが国内・海外の既存語彙表の比較と合成によるものであった。しかし、この 10 年間程で、海外の大規模な言語コーパスが利用可能になり、より正確な語彙頻度情報が得られるようになってきた。これによって、コーパス言語学の手法を応用した新しいタイプの学習語彙表を生成できる可能性が出てきた。本研究はその意味では、科学的学習語彙表作成の先駆的研究であり、単に語彙表作成だけにとどまらず、語彙表に基づきながらどのように教材を開発したり選択したりするかを、客観的に判断できるシステム開発を視野に入れている点で、包括的な語彙表作成と運用のための総合的な研究であるといえる。

これまで、ほとんどの海外の語彙表は、単一コーパスの単純集計を行ったものか、規模の小さなコーパス・データを教育目的で集約したもののどちらかであった。そのため、大規模かつ包括的なコーパス比較と分析に基づく学習語彙表作成は前例がほとんどない。さらに、その語彙表に学習的配慮を加えるため、独自の国内資料によるサブコーパス作成を実施したり、学習者の発信語彙を加味するための学習者に対する語彙知識の調査を予定するなど、独創的な手法を組み入れた研究計画は、その規模においても計画の独創性においても世界的に注目されるべきものである。

本研究の研究代表者や研究分担者の全員が大学英語教育学会(JACET)基本語改訂委員会の中心的な委員であり、2003 年 3 月には JACET List of 8000 Basic Words (通称『JACET 8000』)を発表した。しかし、海外のコーパス・データや国内の英語教育関連リソースの組織的収集がまだ十分ではなかったことや、時間的な制約があり選択語彙の吟味が完全でなかったなどの反省があった。そこで、2004 年 4 月から開始する 3 年間のプロジェクトを計画した。

本プロジェクトでは、学習語彙表の歴史を紐解くことから開始し、過去の JACET で開発した基本語リストの概観、それから現在利用できる学習語彙表の収集とその特徴についてのまとめを行い、問題点を整理する(第2章)。

次に、JACET 8000 の妥当性を再検討し、その問題点を克服するための追加の語彙表の作成も視野に入れる。特に、語彙頻度のデータと教育的な観点をいかに融合させるかについては、実際に学習者や英語教員からデータを収集することによって改善を図る計画を立てている。さらに、教材の語彙分布の分析などの分野では、JACET 8000 を基本語彙表として意欲的に検証する(第3章)。

JACET 8000 の検証が済んだ時点で、その検証結果を踏まえ、大規模コーパスに基づく新しい語彙表を開発することを最終目標とする。検証結果によっては、プロジェクトの最終目標が、JACET 8000 を完全に更新することになるかもしれないし、場合によってはJACET 8000 の部分的な修正や語彙の追加などになるかもしれない。その点ついては、検

証が済んでから再度検討することにして、研究プロジェクトのスタート時点では、もっと も困難な語彙表の刷新を想定することとした。

今回の学習語彙表の作成にあたっては、前回のJACET 8000の作成で開発し、会得したサブコーパス構築の経験を最大限に活用できると思われる。前回のJACET 8000の語彙表作成作業でも、検定教科書や学習用語彙表などの各種のサブコーパスを作成したが、今回もTOEIC、TOEFL、児童文学など大規模なサブコーパスを構築することが想定される。前回の経験を、今回の作業に大いに生かせると期待される。

さらに、今回のプロジェクトでは、JACET 8000 に基づく語彙テスト作成などを行い、 その結果を新しい語彙表に反映させるための環境を整備することもねらいとしている。

本プロジェクトは、日本の英語教育全般の指針ともなる英語学習語彙表を、従来とは全く異なる科学的方法で作成し、学習者の語彙知識に応じた教材を選択するための教材分析システムを実験的に開発することを目的とする。

0.2 研究計画

2004年から3年間の研究期間に、研究目的を遂行するため以下の計画を立案した。ただし、前項で説明したとおり、JACET 8000の妥当性の検証結果によっては、全面的な語彙表の改訂になるか、部分的な更新にとどまるのか、または語彙の追加になるかもしれない。そのため、現時点では、最も作業が困難な全面的な語彙表の改訂を前提に、研究計画を立案した。

1年目:

1) 英語母語話者の言語使用データ (コーパス) をベースとした基礎語彙表を作成する ため、利用可能な英語コーパス・データを可能な限り広範囲に収集する。

(コーパス・データの構築)

2)日本人にとっての英語語彙表とはどのようなものかを検討するため、日本で利用されている各種英語教材(教科書・副読本・会話教本)、検定試験(TOEFL/TOEIC/英検など)などの英文データを分野別コーパスとして構築する。また日本人英語学習者に望まれる最低限のauthentic English として、時事英語・映画のシナリオ・児童文学などのデータのコーパス化も行う。さらに、学習者の発信語彙のニーズを探るための調査も実施する。(サブ・コーパス構築・学習者の実態調査)

2年目:

1) 収集したコーパス・データをもとに、2003年に発表された最新の語彙表 JACET 8000 を視野に入れながら、1万語規模の語彙表を構築する。この際、最新のコンピュータ言語学の手法を導入し、複数コーパスの語彙リストから最適化した語彙リストを抽出する方法も開発、今後の語彙表作成の科学的手法として提案する。また発信語彙アン

ケートをもとに日本人学習者のニーズにあった語彙を補正して加える。

(新語彙表の完成)

3年目:

1) コーパス・データから基準となる語彙表を補填する活用語彙表(コロケーション 頻度表,句動詞頻度表,word familyによる表,動詞型頻度表など)を作成する。

(活用語彙表の作成)

2) 語彙表をもとにして(a)語彙テスト作成班,(b) 教材(リーディング教材・会話教材・他)分析パッケージ開発班に分かれてプロトタイプを開発し,新しい語彙表に基づいた教材選択の方法論や教材開発などを提案する。

(語彙テスト開発, 教材分析システムの開発)

なお、本研究の推進に当たっては、コーパス言語学のPaul Rayson氏 (Lancaster大学)、 語彙習得論のPaul Meara 氏(Wales 大学)の両名には、本研究のアドバイザーとして、必 要に応じて、研究全般についての助言を求める予定を立てている。

(東京電機大学・相澤一美)

1. 学習語彙表の歴史

英語教育において語彙制限という考えの誕生から「JACET 基本語 (第2次案)」(1983) 誕生までを中心に、語彙表と語彙指導の歴史を以下に簡単に記す。

- 1.1 「JACET 基本語 (第2次案)」(1983) 刊行まで
- 1.1.1 英語教育における語彙制限というアイディアの端緒

外国語教育の効率と学習者の負担軽減のために「語彙制限」というアイディアを持った2つのグループがあった。ひとつは、コロンビア大学の Edward Lee Thorndike を中心とし、語彙統計における頻度とレンジによって基本語彙表を作るという考えであった。もうひとつのグループは Harold E. Palmer を中心とし、研究者・英語教師の専門家としての経験から基本語を選定しようとのグループであった。両者は交流があり、会議も持っていた。

1. Thorndike の3万語表

Edward Lee Thorndike (1874-1949)

- 1921. *The Teacher's Word Book.* Bureau of Publications, Teacher's College, Columbia University. (1 万語) --- 41 種 400 万語のテキストに基づいて
- 1931. A Teacher's Word Book of Twenty Thousand Words: Found Most Frequently and Widely in General Reading for Children and Young People. (2 万語)
- 1944. The Teacher's Word Book of 30,000 Words. (3 万語) (Lorge と共著)
- Thorndike の客観的な量的接近法: Palmer の主観的な接近法——カーネギー会議では、選定した基本語については、大きな相違がないことを、お互いに確認する。

Thorndike の基本語選定基準は,

- 1. 頻度数
- 2. 構造語的価値(主として機能語)
- 3. 地理的広範囲
- 4. 分布值
- 5. 定義性(他の語の定義に利用し得ること)
- 6. 語形成上の価値(接辞をつけて語をふやせること)
- 2. Horn, E. 1926. A Basic Writing Vocabulary. Iowa City. [1 万語, 口語が多い.]
- 3. Ogden, C. K. 1930. The Basic Vocabulary. London.
 - _____. 1932. The ABC of Basic English. The Orthological Institute. [復刻版. 北星堂書店. 1980.]
 - _____. The General Basic English Dictionary. 1940. [復刻版. 『ベーシック英英辞典』 北星堂書店.]
 - _____. 1968. Basic English: International Second Language. Harcourt, Brace & World.
- 4. Faucett, L. & Maki, I. 1932. A Study of English Word-Values Statistically

Determined from the Latest Extensive Word-Counts. Matsumura Sanshodo.[5000語について]

5. Palmer の 3000 語表

Harold Edward Palmer

- 1912年に「機能語のリスト」を作る。
- 1921年に「中等学校における英語教育の調査と報告」の目的で日本に招聘され, 1936年に帰国。
- 1930. *(First) Interim Report on Vocabulary Selection*. Kaitakusha.[3000 word-units, さらに語義区分,屈折形,派生語などで分けられる]
- 1931. Second Interim Report on Vocabulary Selection. Kaitakusha. [3000 words]

 Interim Report on English Collocations. Kaitakusha. [斉藤秀三郎『熟語本位 英和中辞典』より再販]
- 1932. The first 600 English Words for a Classroom Vocabulary.

 The Grading and Simplifying of Literary Material. [選定語彙の教材簡易化への応用]
- 1933. Second Interim Report on English Collocations. Kaitakusha. [5747 連語] 1934. The IRET Standard English Vocabulary. IRET. (1000-Word Radius)

6. Michael West

Michael Philip West (1888–1973)

植民地インドのベンガルの Imperial Education Conference おいて「語彙制限」に基づいて教材の改良することになった。*New Method Primer* [新語の比率を 7.4 語に 1 語であったものを 44.7 語に 1 語に改良した.]

1934年のカーネギー財団の会議において、「語彙制限」が 16 議題のひとつになり、語彙選定小委員会が組織され(West, Palmer, Faucett (Thorndike, Sapir))、 報告書を出す。*Interim Report on Vocabulary Selection*. その結果 *General Service List* (1935、原案)が提案される。これに Irving Lorge による頻度統計数値を入れて 1953年版ができる。この 1953年版が LDOCE 初版の定義語彙の原型となった。

The New Method English Dictionary. Longmans, Green. 1935. —当初は 1799 語の定義語彙(のちに 1490 語)で、23,898 の見出し語を定義した。

Speaking Vocabulary in a Foreign Language. MLJ, 14 Apr. 1930. [1,000 語]

7. Carnegie Report

Carnegie Conference は 1934 年 10 月 15 日-20 日,"The Possibilities of English as a World-Language"をテーマに開かれた。この会議で「語彙選定小委員会」が組織され,討議した。この小委員会では,外国人に英語を教える場合にどういう単語か

ら教えるべきか, が話し合われた。Thorndike と Horn を統合した Faucet-Maki の最初の 1500 語と Palmer の 3000 語, West の定義語彙 1779 語を比較検討した。翌 1935 年 6 月にロンドンで会議が再開され, 選定作業を 11 月まで継続した。単語の選定基準を以下の 7 点とした。

- 1. 使用頻度:英語国民によく使われているか。
- 2. 構造上の価値:機能語はすべて入れる。
- 3. 普遍性:世界のどの地域でも使われているか。
- 4. 主題の範囲: 社会活動の多くの分野で使われているか。教室用語,実業用語,雇用関係,通信文,軍事用語,農漁業,機械工業,商取引,貿易・税関関係の9分野を網羅している。
- 5. 定義語:辞書を作る際,他の語を定義するのに役立つか。
- 6. 造語力:派生語がたくさんあるか。
- 7. 修辞語は選ばない。例えば、詩語、誇張語、荘重語、専門語、職業語、卑語、俗語、児童語などは不適当。

完成した報告書は,

Faucett, Palmer, West and Thorndike. 1936. *Interim Report on Vocabulary Selection for the teaching of English as a Foreign Language*. London, P. S. King and Son, Ltd.

この報告書の Part V に The General Service List (pp. 87-506)として 2060 語が収められ、それぞれに簡単な例文がついている。

1938年に Thorndike and Lorge, *The Semantic Count of English Words* が完成し、翌 1939年に West に Carnegie List の改定を委嘱し、1953年に完成した。ゆえにいわゆる General Service List は、単に West が作ったものではなく多くの研究者の合作である。

West, M. 1953. A General Service List of English Words with Semantic Frequencies and a Supplementary Word-list for the Writing of Popular Science and Technology. London, Longmans, Green.

8. Albert Sydney Hornby (1898—1978)

1924年に大分高商へ英文学の教師として着任するが、すぐに日本の学校でまず求められているのは文学教育ではなく、語学教育であるとわかった。パーマーを知り、1931年から「連語リスト」の編集を手伝い、英語教育における語彙制限に目を開かれていった。

1938. "Report on the year's work: dictionary problems." IRET Bulletin, 148.

1942. (et al) Idiomatic and Syntactic English Dictionary. Kaitakusha.

1947. "Editorial: vocabulary control and vocabulary layout." English Language Teaching 1.

- 1.1.2 学習指導要領における語彙制限
- 9. 学習指導要領の語彙規定(1958)
- 10. 全英連. 1963. 『中学校英単語活用集』南雲堂.
- 11. 全英連. 1964. 「高校英語の Minimum Vocabulary 試案」 全英連. 1981. 『高校基本英単語活用集』(初版 1963) 研究社.
 - 1) Thorndike, Faucett & Maki, Palmer, West 等を参照しながら、日本の中・高の教科書の語彙統計を主な拠り所として、現行の中・高で教えられている語彙を、これでよいと追認したものと判断される。
 - 2) Kučera, Carroll 等の新しい、客観的な語彙統計はほとんど参照されていないと 判断される。
 - 3) 英語教師の経験からの選語であると思われる。
 - 4) 4767 語という語数は「JACET 基本語 4000」よりも多く, またその当時の高校 英語 II B までの上限 2950 語と比べても多すぎると思われる。
- 12. 速川 浩. 1966. 『教科書に現われた英語単語の研究』大修館書店.速川 浩. 1972. 「使用単語統計研究(1)(2)」『英語教育』第21巻7,8号.
- 13. 垣田直巳ほか. 1977. 『電子計算機による英語教科書の使用語彙総覧 中学校編』渓 水社.

1.1.3 コーパスに基づく語彙表

- 14. 竹蓋幸生. 1983. 「実用英語語彙リスト」別冊 The English Journal.
 - 1) Kučera の上位 11112 語と Carroll の上位 10175 語と竹蓋(1981)の全語彙 9714 語の和集合をとり、7500 語の頻度別実用語彙リストを作ったもの.
 - 2) 客観的なリストを寄せ集めたため、例えば、同じ州名でも 2000 語以内に入っているものと 8000 語あたりにやっと入るものがあるといった偶然性が見られる。教育経験からの手直しはしてないのが特徴である。
- 15. Kučera & Francis (1967) と Carrol, et al. (1971)
 - Kučera, H. and Francis, N. 1967. *Computational Analysis of Present-day American English*. Brown Univ. Press.
 - Carroll, J. B., Davies, P. and Richman, B. 1971. *The American Heritage Word Frequency Book*. American Heritage Pub. Co.
 - 1) Kučera(1967) と Carroll(1971) の約1万語(11112 語と 10658 語)の共通語は 7360 で 75%弱である。
 - 2) Kučera の特徴として新聞雑誌等の時事的用語の頻度が高いが,1962年以後話題 になった語,あるいは文語的な語は頻度が低いか,あるいは出ていない。
 - 3) Carroll の特徴としては時事的用語の頻度が低く、物語に現れるようなやや文語的な語の頻度が高い。また、理科の教科書に現れる程度の自然科学用語の頻度も高い。

- 16. Francis & Kučera. 1982. Frequency Analysis of English Usage: Lexicon and Grammar. Houghton Mifflin Co.
 - 1) Kučera (1967) と同一の Brown corpus を品詞よりもさらに細かい 76の Tags に分け、 それぞれの 100 万語に対する出現頻度数、15 ジャンルの中のいくつに現れているか、 500 サンプルのいくつに現れているかを示した。
 - 2) Alphabetical Frequency List と Rank List の2本立てになっている。
- 17. 淀縄光洋. 1983. 『高校英語語彙の実態と学習語彙の在り方』東京都立教育研究所.
 - 1) 当時使用の高校英語教科書のうち英語 I , 英語 I (26 種類 52 冊) に出現する全語 彙を調査し、アルファベット順に示したもの。延べ 8204 語。
 - 2) 単語の頻度数は出てないが、range (何種の教科書に出ているか) が示されている。
 - 3)以下の区分がなされている。
 - A(中学校 5 種に出現する語),B(中学で既習と想定された語),C(「英語 I」の新語),D(『英語 II の新語),E(「英語 I」「英語 II」に出現する教科書の種類数) 左の*(頻度 10 種類以上),中の*(中学必修語 490 語),右の*(JACET 基本語に出ている語)が参考に示されている。
 - 4) 教科書が出版されてすぐに作成された語彙リストで、当時としては画期的な速さであった。
 - 5)「速川リスト」「全英連リスト」と大ざっぱにつき合わせてみると、これら2リストよりも一段と Kučera (1967), Carroll(1971)のリストに近く、日本の高校教科書が新たな方向へ動いていることが感じられる。
- 1.1.4 辞書の定義語彙に基づく学習基本語
- 18. Longman Dictionary of Contemporary English の定義語彙 2000 語(1978)
 - 1) Longman 社の一連の辞書, *First English Dictionary*. 1968, *Junior English Dictionary*, 1965, *An International Reader's Dictionary*, 1965, 等は Michael West の definition vocabulary を用いてきたが、LDOCE もこの伝統を踏んで West を基にしている(序文)。この点で(もちろん英語テキスト出版資料等によって一部手直ししているが)古く、難点があると言わざるを得ない。
 - 2)「JACET 基本語第 1 次案」が少数の語を除いて LDOCE2000 語をすべて無条件に 採用したのは問題であった。この点「第 2 次案」では正されている。
 - 3) LDOCE の編集者が他の点では革命的とまで言われた辞書を作りながら,2000 語の 定義語彙の選定に客観的なコーパスよりもむしろ West に拠ったのは問題であった。

- 1.1.5 学習基本語彙選定の基準と考え方(参考文献)
- 19. Fries, C.C. (太田 朗訳). 1957. 『外国語としての英語の教授と学習』研究社.
 - 1)機能語・代用語・否定と肯定の分布を有する語は、英語学習上基本的であり、不可欠なものであるから、制限単語の語彙項目とは別枠にすべきこと。
 - 2) 発表のための語彙としては、現在のいずれの語表も口語体英語が不完全で、将来の語彙統計にこの面の資料が必須であること。
 - 3) 外国人が英語を学習する場合,英語の持つ語彙のすべてを英語国民のように習得することは絶対に不可能であるから、制限語彙の設定は不可欠であり、それらの語彙の持つ意味を厳格に制限し、同時に意味の範囲を明確に示さなければならないこと。すなわち、語彙リストとして具備すべきものは使用頻度にとどまらず、意味の統計(semantic count)が必要であること。
 - 4) 語数を可能な限り制限することに主眼がおかれている。特に初級レベルで、その外国語の音体系の習得を重視し、さらには文法構造における習慣形成を目的として語彙の導入を必要最小限に抑えることを薦めている。一これは構造言語学に基づく外国語教育観と言える。
 - 5) これに対して『英語教育』1983年10月号の「ノーショナル・シラバス特集」に見られるように、外国語教育が文法とともに「場面」「コンテクスト」「意味」等を重要視することになると、「語彙制限」に対する考え方も変わってこよう。
- 20. Fries, C.C., & Traver, A.A.(増山節夫訳). 1958. 『英語の制限単語表―教育語彙選定の方法と問題―』大修館書店. (1958) [上記 19 の考えを具体化したもの] 次の文献の分析と検討を行い、包括的な解説と批判をしている。
 - 1. Basic English
 - 2. West
 - 3. Palmer & Horn
 - 4. Thorndike
 - 5. Interim Report
 - 6. Faucett & Maki
 - 7. Janet Aiken: Little English

Fries はBasic English を高く評価している。

- 21. Fries, C.C., & Traver, A.A. 1950, 1958. English Word Lists: A Study of Their Adaptability for Instruction. Ann Arbor: George Wahr Publishing Co. / Tokyo: Taishukan. [文献について]
- 22. ミシガン大学英語研究所 (池永勝雅訳). 1958. 『英語の語彙』大修館書店. Rosenzweig, M. R. & McNeill, D. 1962. "Inaccuracies in the semantic count of Lorge and Thorndike," *American Journal of Psychology*, 75, 316-318 [意味統計の問題点]
- 23. 『基本英語百科辞典』研究社. (1964) [5600 語について]

- 24. Mackey, W.F. 1965. *Language Teaching Analysis*. Longmans. (伊藤・和田・池田共訳) 1979. 『言語教育分析』大修館書店. [学習語彙選択の基準について]
- **25.** NHK 通信教育部. 1966. 「テレビ英語会話初級使用単語研究」[34. 稲村松雄(1970)に解説あり]
- 26. Urdang, L. 1966. "An 'Unabridged' word-frequency count of American English: A proposal for an integrated system," *Word*, 22, 103, 294-302 [統計法のあり方]
- 27. 『英語の語彙』(現代英語教育講座 5) 研究社. (1966)
- 28. Mackey, W.F., & Savard, J.G. 1967. "The indices of coverage: A new dimension in lexicometrics," *IRAL*, 5, 2 & 3, Jul., 71-122 [coverage(適用度)の基準]
- 29. Preller, Arno G. 1967. "Some problems involved in compiling word frequency lists," *MLJ*, 51, 399-402 [語彙統計の問題点]
- 30. Fiks, A. I. & Corbino, J. P. 1967. "Course density and Student perception," *LL*,17, 1 & 2, Jul., 3-8 [言語材料の質の問題]
- 31. Shaw, J.R. 1970. The New Horizon Ladder Dictionary of the English Language, Signet. [5000 語が 1000 語ごとに 5 段階に分けられている]
- 32. Richards, Jack C. 1970. "A psycholinguistic measure of vocabulary selection," *IRAL*, 8, 2, May, 87-102 [familiarity の基準]
- 33. Siliakus, H. & Morris, K. 1970. "Some reflections on the lack of accuracy of word frequency lists," *Review of the Institute of Applied Licguistics*, 9, 11-18 [語彙統計のあり方]
- 34. 稲村松雄. 1970. 『語彙・連語の指導』(講座 英語教授法 7)研究社. 語彙選定の基本的態度
 - 1. 対象とする英語は標準的なものでなければならない。
 - 2. 素材は Written English だけでなく, Spoken English も含む。
 - 3. 頻度と分布一使われる範囲、分布が広い語。広い地域で使用されている語でなければならない。
 - 4. 種類一機能語と内容語。どのような内容語を選定するか。
 - 5. 特殊性―日本人が外国語として学習する特殊性。必ずしも英米の同年齢の語彙選択 とは一致しない。幼年期の語彙を学ばないこともある。
- 35. 速川 浩. 1971. 「内容語の制限が教材に及ぼす影響」 『英語教育』 第20巻3号.
- 36. 池田哲郎.1971.「教科書英語の変遷(1)幕末篇」『英語教育』第 20 巻 3 号.
- 37. 長谷川潔. 1971. 「Speaking Vocabulary と教科書の語彙」 『英語教育』 第 20 巻 3 号.
- 38. Ure, J. N. 1971. "Lexical density and register differentiation," Perren, G. E. and J. L. M. Trim(eds.) 1971 Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics, Cambridge 1969. London: Cambridge University Press. 443-452. [言語材料の問題点]

- 39. Richards, Jack C. 1971. "Coverage: What it is and what it isn't," Review of the Institute of Applied Linguistics, 13, 1-15 [coverage の基準]
- 40. Tarnoczi, L. 1971. "Wortbestand, Wortschatz, Wortfrequenz," IRAL, 9, 4. Nov., 297-318 [語彙統計]
- 41. 三浦順治. 1973. 「第二言語としての英語基本語彙選択の条件」『秋田大学教育学部研究紀要』 三浦(1973)の基本語彙選択の条件
 - 1) 連想の構造への配慮が必要
 - 2) 日本語の語彙の使用率も無視できない―日本人の生活の実態との関係で
 - 3)全体として語彙素統合論的であると同時に意味素統合論的でなければならない。これは部分的に参考になる、興味深い点があるが、語彙リスト作成についてはやや具体性を欠いていると思われる。
- 42. 「特集=語彙の指導」『英語教育』第 22 巻 7 号. 1973.
- 43. 清水貞助. 1973. 「重要語とは何か 語彙研究の歴史と考え方 —」『英語教育』第 22 巻 7 号.
- 44. Harris, A.J. & Jacobson, M.D. 1973-4. "Some comparisons between basic elementary reading vocabularies and other word lists," *RRQ* 9, 1, 87-109 [語彙表の比較]
- 45. O'Shea. 1974. "Vocabulary selection through word list comparisons," *Pasaa*, 4, 1, Mar., 91-102 [語彙表の比較]
- 46. O'Rourke, J.P. 1974. Toward a Science of Vocabulary Development. Mouton.[語彙学習の重要性と諸相について]
- 47. Richards, J. C. 1974. "Word lists: Problems and prospects," RELC J, 5, 2, Dec., 69-84. [文献目録]
- 48. Special English Word Book. VOA 1972, 国際コミュニケーションズ. (1976)
- 49. 古宮照雄ほか. 1976. 「英語教育における語彙の研究:中学校用英語教材語彙の統計分析」『広い旅空の下で』松柏社.

古宮照雄ほか. 1976.「中学校用英語教科書における語彙構成の分析とそのワードリスト」『千葉大学教養部研究報告 B-9』

- 50. 小笠原林樹. 1976. 「指定語彙の数とその構成の問題点」「慣用事実から見た「共通語彙表」」 『現代英語教育』第13巻6号.
- 51. 清川英男. 1975. 「英語教育のための5つの語彙統計の比較」『産業能率短大紀要』 第12号.
- 52. Kelly, E., & Stone, P. 1975. Computer Recognition of English Word Senses. North-Holland. [品詞・意味別の頻度が出ている.]
- 53. 清川英男. 1976. 「Spoken Word Lists に関する考察」『英語教育』第 25 巻 2 号. 清川英男. 1976. 「口語英語の基本語リスト作成の試み」『専修語学ラボラトリー論集』 第 5 号.

- 54. 小笠原林樹. 1977. 「軽視される語彙の研究と教育」 『現代英語教育』 第13巻10号.
- 55. 垣田直巳ほか. 1977. 『電子計算機による英語教科書の使用語彙総覧 ─ 中学校編』渓水社.
- 56. 小川 清. 1977. 「語彙・連語の指導」 『英語教育』 第 26 巻 7 号.
- 57. 「特集=語彙指導—何をどう教えるのか」 『英語教育』第26巻4号. 1977.
- 58. 中村嘉弘. 1978. 「アメリカ英語の統計的研究 (1)」『淑徳大学研究紀要』第13号.
- 59. 小西友七. 1980. 『英語基本動詞辞典』研究社.
- 60. 上田明子. 『英語基本語彙辞事典』中教出版. (1983) [3000 語について]
- 61. 山田雄一郎. 1980.「語彙選定:心理言語学的考察」『佐賀大学教育学部研究論文集』 Vol. 28. No. 2 (1)
 - 1) Frequency を利用して「第1基本語彙」を決定する。その際補助基準として、
 - 1. 機能語
 - 2. frequency は低くても必要度の高い語(数字,曜日,月名)
 - 3. category における上位語(連想が上位語 → 下位語と働くため)
 - 4.「場面的連想」を喚起しやすい語(hospital, school)を優先させる。
 - 2)「第1基本語彙」の中の内容語を "centers of interest" として連想によって「第2 基本語彙」を決定する。
 - 3) こうして決定された全体を、再び frequency, familiarity, coverage, needs などの基準によって整理する。
 - この山田(1980)は主観的選定方法として、かなり具体的で、注目に値すると思われる。
- 62. 升川 潔ほか. 1980. 「英語語彙力と英文読解能力 1, 2, 3」『英語教育』第 24 巻 1, 2, 3 号.
- 63. Hindmarsh, R. 1980. Cambridge English Lexicon. Cambridge University Press.
 - 1) 外国語としての英語学習用購読教材作成とテスト (FCE, CPE) 問題作成のための 資料作りという動機から, 17 種類の語彙資料・リストを参考にして作成された。
 - 2) 4500 語に 5 段階の頻度区分がなされ、語義項目はさらに 2 段階上の 7 段階までなされている。
 - 3) 出来上がったリストは、たいへん多くの資料を用いてはいるが、Michael West と Thorndike の影響が色濃く見られ、伝統的で、一部は古いと言わざるをえないであるう。
 - 4) 意味の頻度値を詳しく、広範に示した点で高く評価できる。
- 64. The Oxford-Duden Pictorial English Dictionary. (1981)
- 65. 竹蓋幸生. 1981. 『コンピュータの見た現代英語―ボキャブラリーの科学』エデュカ 出版部.
- 66. Rivers, W.M. 1981. Teaching Foreign-Language Skills. 2nd edition. Chicago U. P.
- 67. 中村嘉弘. 1981.「語彙指導における基本点」『英語教育』第30巻2号.

- 68. 島岡 丘. 1982. 「6. 語彙」『英語学と英語教育』大修館書店.
- 69. 城戸法政・村田年ほか. 1982. 「高等学校英語 B. 作文教科書における語彙調査とその分析」『木更津高専紀要』第15号.
- 70. 垣田直巳ほか. 1985. 「英語の語彙指導」 『英語教育』 第34巻7号.
- 1.2 JACET 基本語リストの開発
- 1.2.1 「JACET 基本語 第 1 次案」(1981)

「JACET 基本語第 1 次案」は、JACET 教材研究委員会が 2 年間にわたって購読用教科書に関する調査と研究を行ってきた結果の公刊である『大学一般教養課程における英語購読用教科書のあり方』(1981、発行所: 大学英語教育学会)の一部として発表されたものである。

本語彙表は、一般教養課程終了時までに修得する(少なくとも recognition のレベルで) ことを目指すべき語数として 4000 語ぐらいが適当と判断し、具体案を1つのたたき台とし て示したものである。その作成手順と基準はおおよそ次のようなものであった。

Kučera & Francis (1967) および Carroll et al. (1971) からそれぞれ頻度, U 値に基づき 6000 語を選び, さらに *Longman Dictionary of Contemporary English* (1978) (*LDOCE*) の 2000 語を加えて基本台帳を作った。

- 1) Kučera & Francis (1967) あるいは Carroll et al. (1971)のいずれにしか現れず、かつ頻度 5000-6000 位の単語、および *LDOCE* の中で、委員会が日本人学習者には必要でないと判断した語を除いた。
- 2) 形容詞に -ly を加えた副詞は原則として除いた。
- 3) 不規則動詞の活用形は、頻度の高いものだけを採用した。
- 4) 複合語のうち、容易に意味が類推でき、頻度の低いものは除いた。
- 5) 同綴異語の区別はしなかった。

(以下略)

従って、基本的には頻度を重視するが、上記の基準を設定し、委員会で協議の上、除いたり、新たに加える作業を行って、4064 語を確定した。

1.2.2 『JACET 基本語 第 2 次案』(1983)

「第1次案」発表と同時にJACET 全会員にアンケート調査をし、その結果を踏まえて基準の立て直しをし、2年後に『英語購読用教科書のあり方』についてのアンケート調査報告 —「JACET 基本語第2次案」を中心に—』として発表した。

「JACET 基本語第 2 次案」の改定基準

下記の5つの語彙表を第1次資料とし、原則として次のいずれかに該当すれば採用とした。

- 1) Kučera & Francis (1967)あるいは Carroll et al. (1971)の 2000 語レベルまで。
- 2) 上記のいずれにも 4000 語以内に入っている語。
- 3) Cambridge English Lexicon の Level 2 (1215 語)まで。
- 4) Francis & Kučera (1982) の frequency 100 以上。

5) Longman Dictionary of Contemporary English の defining Vocabulary 2000 語以内のもの。

さらに第2次資料として7つの語彙表を参考資料とし、9つの細目を立て、4つの具体的な作業手順に基づいて、慎重に3990語を選定した。

結果的に見ると、頻度・分布度優先、基本形ではなく、派生語・変化形の形のままで登録 されている語がかなり多く見られる語彙表であると判断される。

1.2.3 『JACET 基本語 4000』(1993)

「JACET 基本語第 2 次案」は広く英語教育界に受け入れられ、多くの教師によって教室で使用されたばかりか、研究論文の語彙関係の検証に用いられ、大学入試問題の語彙検定の指標としても使われようになった。これをさらに改訂してほしいとの要望もあったが、JACET 教材研究委員会の活動はリーディング研究の方へ移っていて、語彙表改訂に力を注げる状態にはなかった。

そこで,基本語彙表はそのままにして,各種教授場面においていろいろな使い方ができるような情報を付加し,またいくつかの別表を作成した。

1) 使用頻度順の5段階表示:主に使用頻度順によって5つの段階にわけた。

 $1:1\sim500$ 位, $2:501\sim1000$ 位, $3:1001\sim2000$ 位,

4:2001~3000位, 5:3001~5000位

使用頻度順の5つの段階に分ける際に典拠として参照したのは,『プロシード英和辞典』初版の「キーワード5000」(竹蓋幸生氏作成)である。

- 2) 品詞名の表示: Longman Lexicon of Contemporary English の品詞表示を基本としたが、基本 4000 語の観点から一部省いたものがある。
- 3) 意味・機能による 14 分野別の情報: それぞれの見出し語をその意味と働きによって 14 の範疇に分けた。 Longman Lexicon of Contemporary English の 14 の分類をそのまま採用し、この分野の分け方に従った。辞書に掲載されていない語形についてはその語の基本形や変化形、あるいは同意語などを参照して編集委員が決定した。

別表:

- 1) 使用頻度別一覧表:「アルファベット順一覧表」の全見出し語について,第1段階から第5段階まで,各段階別の一覧表を作成した。
- 2) 意味・機能の分野別一覧表:「アルファベット順一覧表」の全見出し語について、 A~N の各分野別にそれぞれ一覧表を作って示した。いろいろな活用法が可能であろう。

- 1.3 現在利用できる語彙表
- 71. 北海道大学言語文化部英語教育系. 1995. 『北大語彙表』(Hokkaido University English Vocabulary List) [独自の 1200 万語の分析資料と既発表の 12 種類の語彙表の検討結果から作られた,ほぼ基本語方式による 7416 語の語彙リスト.]
- 72. Zeno, S. M. et al. 1995. *The Educator's Word Frequency Guide*. Touchstone Applied Science Associates, Inc. [Carroll et al.(1971)の発展型と見ていいであろう. アメリカの教科書等 17,274,580 語の分析による 154,941 語の word-type リスト及びrank-order リスト.]
- 73. Dale, E. 1983. Readability Revisited: The New Dale-Chall Readability Formula. Brookline Books. [リーダビリティー公式用の 3000 語リスト.]
- 74. Cambridge International Dictionary of English. Cambridge Univ. Press. 1995. [定義 語彙 2000 語リスト. 頻度に他の6つの要素を加味して選定されている. 他の語を説明 するのに役立つ語, 特に文法用語を多く含めている.]
- 75. Collins COBUILD English Dictionary. 2nd edition. HarperCollins Publishers. 1995. [2 億語の通称 Bank of English より出現頻度数によって上位 14,700 語を選び、それを 5 段階に分けて示したリスト. 第 3 段階までで 3400 語.]
- 76. Longman Dictionary of Contemporary English. 3rd edition. Longman Group Ltd. 1995. [定義語彙 2000 語リスト,及び Longman Spoken 3000: Longman Spoken Language Corpus 等により基本語となる話し言葉用 3000 語を選び、1000 語、2000 語、3000 語分けて本文中に示した. Longman Written 3000: 1000 語、2000 語、3000 語に分けて本文中に示した. 同じ語でも話し言葉と書き言葉では頻度が異なるので貴重な資料である. 基本 150 語について頻度のグラフがついていて参考になる.]
- 77. Oxford Advanced Learner's Dictionary. 5th edition. Oxford Univ. Press. 1995. [定義 語彙 3500 語のリスト. Oxford の語彙資料に British National Corpus を加えて、出現 頻度にいくつかの条件を加味して 3500 語を選び、定義用語彙とすると共に重要基本語 として本文中の記述を詳しくしている. 2005 年の 7 版はこの「学習基本語彙」という 性格をさらに強めている. The Oxford 3000 との名称をつけ、特にコミュニケーション のための語として使えるように意味・解説と用例を示している.
- 78. 竹蓋幸生. 1988. 「キーワード 5000: SYSTEM」『プロシード英和辞典』(福武書店) [「白色語彙・マーク V」の頻度上位 15000 語の変化形等を原形に集め、整理して見出 し語 6500 語を得た。それから固有名詞、略語等を除き、教育的配慮から数語を加除し、 頻度上位 5000 語を選んだリスト。このリストはさらに、自前の資料、6 分野 72 編、 1,317,642 語の分析により「現代英語のキーワード」(5000 語)となり、さらに「プラス α 2000」を加えて合計 7000 語とし、いずれの分野においても高いカバー率を持つ語彙リストとした.]
- 79. Nation, P. (1996). Vocabulary Lists. English Language Institute Occasional

Publication No.17. Victoria University of Wellington. [このリストはビクトリア大学で教えるための 3000 語の英単語リストで、8つの原則に基づいて選出されている。原則とは、1)必要性、2)頻度、3) 範囲、4)経済性、5)規則性、6)定義性、7)教室や指導での必要性、8)借用語を含む原則である。最初の 1000 語は、受容と産出の両方に使うことを目的とし、残りの 2000 語は、主として受容語彙として使うことを目的にしている。]

- 80. Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238. [学術的文書を読むときに共通して使われる 570 語の語彙をリスト化したもの. さまざまな分野の学術的文献をコーパス化し、その頻度情報に基づき、最も頻度が高く、かつ、広範な分野で使われる 570 語を選出している. 570 語は、頻度に基づき 10 のレベルに分けられている.]
- 81. 岡田晃忠・葉島千歌. 1999. 『基本語彙 3000』日本大学理工学部 LL 教室. [日本大学理工学部生用の語彙表 3492 語の中から基本語 3000 語を選んだ語彙表. Longman Dictionary of Contemporary English(1995), Cambridge International Dictionary of English(1995), Collins Cobuild English Dictionary(1995) の 3 種の定義 語彙に、英文雑誌 Mini World の英文資料を加えて、3456 語を抽出し、そこから 3000 語を選んで Step 1—Step 5 及び Appendix に分けて表示したもの.]
- 82. Yoshioka et al. (1997), English Vocabulary for Academic Purposes: Analysis of Data from the Physical and Social Sciences. Liber Press. [文系から 6 分野, 自然理系から 4 分野を選び、それぞれの分野から概論書 1 冊ずつを集め、その英文(各 200 ページ分)をインプットし、統計処理と加除の調整を行い、さらに中学基本語を外し、頻度順に並べた 7962 語のリスト. そのリストを詳細に分析している. 全資料の量が 255,495 語と少ないが、始めて各専門に進む前に習得すべき専門基本語彙を示した意義は大きい.]
- 83. アルク語彙プロジェクト. 2000. 「SVL 標準語彙水準 12000」アルク. [Adam Kilgarriff の British National Corpus Frequency List を見出し語としたものをベースに、先行語彙リスト,各種英語コーパス、日本語コーパス、学習英和辞書の重要語表示,大学入試問題,日本語の使用頻度表等多くの文献資料を参考にし、さらに複数のネイティブ・スピーカーの直感による判断を求めて決めた 12000 語の語彙リスト. 特に、日本人の日本語使用の実態を参考にした点に特徴がある. この語彙水準を教材化したものに『SVL 標準語彙水準 12000 準拠 Power Words』等がある.]
- 84. 中華民國教育部編. 中華民國 84 年. 『国民中學過程標準』. [台湾の学習指導要領の学習語彙. 中学指定語彙約 1200 語の具体的な参考語が, 1. People, 2. Personal Characteristics, ... といった範疇別とアルファベット順の二通りの方法で提示されている. さらに選択科目の英語として上記必修語 1200 語を含めて約 2000 語のリストが同様の二通りの方法で示されている.]

1.4 「JACET 8000」(2003) の開発

1.4.1 プロジェクトの開始まで

毎年いくつかの学会において『JACET 基本語 4000』を基準語彙表にした発表がなされ、 また教育現場においても高く評価され、広く用いられてきたが、いっぽうでは語彙表の改訂 を求める声が高まっていた。

本部研究企画委員会ではそれだけの委員を集めることは困難と思われた。そのようなときに、連絡用に電子メールを用いた全国規模の改訂委員会のアイディアが出され、2000年1月の理事会で承認され、全国の JACET 役員と英語辞書研究会会員にメールとメーリングリストを使って委員の公募がなされた。その後、設立総会を経て、「8000語を選び、1000語ずつレベル分けして8段階にする」というおおまかな方向性が決定され、「常任委員会」「作業部会」「編集委員会」などの各組織を立ち上げて作業を行うこととなった。

1.4.2 プロジェクトの経緯

以下プロジェクトの活動経緯を簡単にまとめておく。

2000 年 1 月 JACET 理事会において「基本語改訂委員会」設置が承認される

2000年4月 公募により、約70名の委員が決まる

2000年4月 委員の互選により常任委員を選出

2000年5月 基本語改訂委員会設立総会を開催(於 早稲田大学)

2000 年 11 月 JACET 8000 第 1 回編集委員会(於 沖縄国際大学)

2001 年 7 月 JACET 8000 第 2 回編集委員会(於 JACET 事務所)

2001 年 9 月 JACET 第 40 回大会(於:藤女子大学)において「JACET 基本語(改 訂 4 版)について」の題でシンポジウムを開催し、進捗状況を報告

2002 年 8 月 JACET 8000 第 3 回編集委員会(於 早稲田大学)

2002 年 9 月 JACET 第 41 回大会(於:青山学院大学)において「JACET 基本語改訂 における科学的学習語彙リスト構築の試み」の題でシンポジウムを行い, 8000 語の仮案を発表

2002 年 12 月 JACET 8000 第 4 回編集委員会(於:早稲田大学)

2003年3月 『大学英語教育学会基本語リスト』を公刊

1.4.3 作業手順

1. 作業用スケール(作業のための基準となる語彙表)の作成

1 億語の BNC より「BNC 頻度順 100,000 語リスト」と、変化形を基本語に寄せて上位語をリストした「BNC 5516 語リスト」の 2 つを作成し、以後の作業の基礎データとした。後者は、Adam Kilgarriff 氏が BNC に基づいて作成した BNC Lemma List 6300 を加工したものである。

2. JACET 8000 サブコーパスに基づく頻度順位表の作成

委員個人が研究用に作成した各種コーパス(検定教科書,雑誌・新聞,映画,児童文学,BBC・CNNなどのスクリプト,センター試験・STEP・TOEFL・TOEICなどの資格試験など)から,語彙出現頻度データをデータ処理グループに提出してもらい,処理グループで頻度データの集計を行い,「サブコーパス頻度順位表」(8000語)を作成する。

3. 頻度統計処理による8000語の決定

BNC での頻度と JACET 8000 サブコーパスでの頻度を、対数尤度(log-likelihood)の 視点から比較して BNC の順位を調整し、上位 8000 語を選び、「JACET 候補語 8000」を 決定する。

4. 「JACET 候補語 8000」の順位の再調整

日本の教育現場で使用することを前提に、上記の手順で選び出された「JACET 候補語 8000」の順位を、「高校教科書コーパス頻度順意表」(4126 語)と照合して再調整する。同時に、大文字語や派生形などを一括収録し、別表「plus 250」を作成する。

結 語

学習語彙表は、今や教師・専門家の主観的選定から、統計処理による客観的データの活用と教師・専門家の判断の綯い交ぜから、純粋に統計処理のみによる「客観的選定」に進んできた。上で解説した「JACET 8000」はまさにそのような統計処理を駆使した客観的選定法を取っている。それゆえに、それなりの素養のある研究者であれば、選定作業のすべてを検証することが理論上は可能である。しかも日本における英語教育現場を十分に反映し、大学生・成人社会人の望ましい語彙力の指標のひとつを示したものとなっている。学習語彙表の作成は新たな時代に入ったと言えるであろう。

しかしながら、学習語彙表は、使用目的に応じているいろな作成方法が可能であり、主観 的選定から客観的な統計処理によるものの間の各種の表が存在すれば、それぞれがそれなり に有効な使い道があろう。相当数の学習用語彙表が存在し、教師がそれぞれの教育現場にお いて、適切なものを選択できる状態であることが望ましいであろう。

(和洋女子大学・村田 年)

- 2. JACET 8000 の検証
- 2.1 英語能力検定試験のテキストカバー率

JACET 8000 は、日本の英語学習者を念頭に置いた語彙リストであるが、その精選や確定の手法は客観性が高いと思われるが、英語圏で作成された英語テストで検証しても十分にその有用性が認められるかどうかを検証することにする。

調査は、清水伸一氏の語彙分析用 Perl スクリプト(V Level Marker)を分析用ソフトとして選定し、センター試験(英語)と英語能力試験(IELTS, TOEFL, TOEIC)の使用語彙が、JACET 8000 によるレベル別の語彙カバー率を検証した。

使用した試験データは以下の4種類である。

(1)大学入試センター試験(英語)の英文テキスト

1986 年から 2004 年の本試験と追試験, さらに 2005 年本試験, 2006 年本試験・リスニング問題スクリプトを含む。

- (2)IELTS(International English Language Testing System)データ
 - IELTS 2 University of Cambridge Local Examinations Syndicate 2000. Cambridge IELTS2: Examination papers from the University of Cambridge Local Examinations Syndicate. Cambridge: CUP
 - IELTS 3 University of Cambridge Local Examinations Syndicate 2002. Cambridge IELTS2: Examination papers from the University of Cambridge Local Examinations Syndicate. Cambridge: CUP
- (3)TOEFL (Test of English as a Foreign Language) データ
 - toefl01 ETS 1995 TOEFL Test Preparation Kit Workbook . NY: ETS
 - toefl02 ETS 1998 TOEFL Practice Tests Workbook Volume 1 Practice Tests A. NY: ETS
 - toefl03 ETS 1998 TOEFL Practice Tests Workbook Volume 1 Practice Tests B. NY: ETS.
 - toefl04 ETS 1998 TOEFL Test Preparation Kit Workbook, Practice Tests A. NY: ETS.
 - toefl05 ETS 1998 TOEFL Test Preparation Kit Workbook, Practice Tests B. NY: ETS.
 - iBTest1 & iBTest2 Phillips, D. 2005 Longman Preparation Course for the TOEFL

 TEST Next Generation iBT Complete Test 1. NY: Pearson Education, Inc.
- (4) TOEIC (Test of English for International Communication) データ
 - toeic1.txt ETS 2000 TOEIC Official Guide & Tests (Japanese Ed.) Tokyo: The Institute for International Communication
 - toeic2a.txt & toeic2b.txt ETS 2002 TOEIC Official Guide & Tests (Japanese Ed.)

Level Marker による「センター試験(英語)」と「3つの英語試験」の index と token は、それぞれ 5,247 と 119,328、 10,728 と 202,916 で、データ量から見ると「センター試験(英語)」は「3つの英語試験」の約半分となった。両データを Index ベースで見た場合、JACET 8000 によってカバーされたのは、「センター試験(英語)」74.1%、「3つの英語試験」50.37%であった。カバーされなかった 25.9%と 49.63%の内訳は、over 9(JACET 8000以外の語形)と others (固有名詞、数詞、縮約形、非語)で、それぞれ 12.08%と 13.82%、19.89%と 29.74%であった(図 2-1-1 参照)。

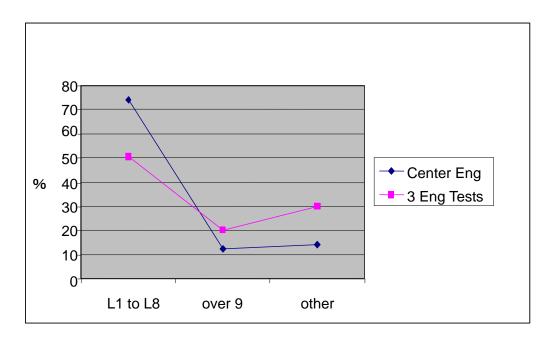


図 2-1-1 英語検定試験及びセンター試験「英語」の JACET 8000 でのカバー率

図 2-1-2 は、JACET 8000 でカバーされた両データの index について、それぞれレベルご との出現率を算出しグラフ化したものである。「センター試験(英語)」データと「3つの 英語試験」データは index のカバー率は、1%水準で r=.900 と強い相関を示した。両データは、Level 1、Level 2 で 90%前後、Level 3 で $75\%\sim80\%$ のカバー率で大きな違いは認められなかったが、Level 4 で大きな数値の開きが認められた。「センター試験(英語)」データでは Level 4 で 32.4%(1000 語中約 320 語が出現)、「3つの英語試験」データでは 77.7%(1000 語中約 770 語が出現)となり大きな差が生じた。この傾向は Level 8 まで継続した。

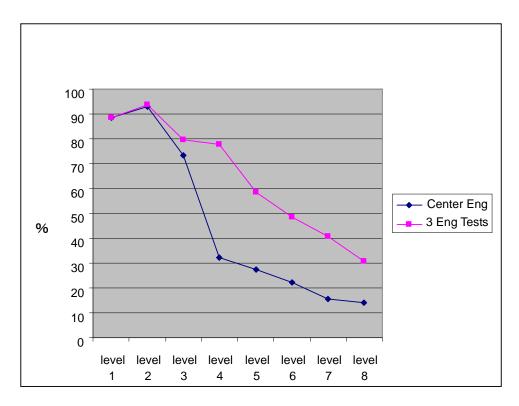


図 2-1-2 英語検定試験及びセンター試験「英語」の JACET 8000 でのレベル別カバー率

「センター試験(英語)」データは、試験 40 回分とリスニング問題1回分のファイルの合計 41 ファイル,「3つの英語試験」データは 16 ファイルから構成されている。両データの多くの語形は、限られたファイルにのみ出現し、多数のファイルで見られるものは限られている。

WordSmith Tools の Wordlist 機能を使い頻度順リストを出力すると、各語形の出現回数と出現したファイル数の情報を得ることができる。両データの語形(type)の内、全ファイルに出現したものは、それぞれ 69 語と 156 語であった。ちなみに、両データの総 type 数は 7039 語と 11940 語であり、総 type 数に占める「すべてのファイルに出た type 数」の割合は、それぞれ 0.98%、1.31%であった。同様に、ファイル比で約 25%以上、すなわち「センター試験(英語)」の 41 ファイル中 10 ファイル以上、「3 つの英語試験」16 ファイル中 4 ファイル以上に出現した type の数を調べると、それぞれ 925 語(13.14%)と 3812 語(31.93%)であった。(括弧内は総 type 数に対する比率) 換言すると、「センター試験(英語)」データの約 87%の type と「3 つの英語試験」データの約 69%の type とは、ファイル数でそれぞれ 10 ファイルと 4 ファイルよりも少ないファイルにのみ出現したことになる。

2.2 アカデミックテキストのカバー率

電子版 New York Times 紙に掲載された新刊図書 103 冊の第 1 章英文を, 2005 年 3 月 13 日から 8 月 14 日にかけて収集した。既存の英語語彙リストは, この英文テキスト・データをどの程度カバーできるか検証するために, 語彙表 JACET 8000 とアルク社の ALC Standard Vocabulary List (通称 ALC SVL) ¹⁾を用いたテキストのカバー率についての検証を試みる。

JACET 8000 に採用され ALC SVL には採用されていない語形, 逆に ALC SVL に採用され JACET 8000 には採用されていない語形を調べるために, ソフトウェア WordSmith Tools Version 4. (以下, WS4) の stop list に 2 つの語彙リストを登録して相互の出現頻度順語形リストを出力した。図 2-2-1 は, JACET 8000 に含まれず, ALC SVL (12000 語)中あるいは ALC SVL 上位 8000 語中に存在する語形, 逆に ALC SVL (12000 語)あるいは ALC SVL (上位 8000 語)に含まれず, JACET 8000 に存在する語形の数を示したものである。なお, 図 2-2-1 の ALC12K は ALC SVL 全体 (12000 語), ALC8K は ALC SVL 上位 8000 語を, J8K は JACET 8000 を示す。

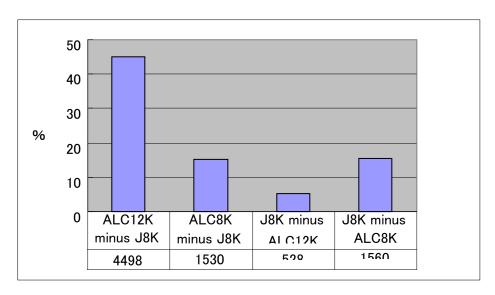


図 2-2-1 JACET 8000 と ALC SVL との比較

Kilgarriff のレマ・リストは、JACET 8000 と ALC SVL(12000 語)両語彙リストの語形を確定する際に利用されている。このため、JACET 8000 と ALC SVL(12000 語)に採用された多くの語形は共通である。グラフから明らかなように、ALC SVL(上位 8000 語)と JACET 8000 の約 8 割強の語形は共通である。JACET 8000 に存在せず、ALC SVL(12000 語)にある語形は約 45000。逆に、JACET 8000 のみにある語形は約 530 であった。

表 2-2-1 は、2つの語彙リストに採用された語形を比較するためにまとめたものである。 この表から推測できるように、Kilgarriffのレマ・リストの語形の多くが JACET 8000 と ALC SVL(12000 語)に認められる。しかし、名詞+s(outskirts, thanks, works),英語綴り(catalogue, enquire, whisky),短縮形(exam, math, photo)など, JACET 8000 のほうが Kilgarriff のレマ・リスト語形をより忠実に採用している。なお,Kilgarriff レマ・リストと JACET 8000 では,ともに"congratulation"(基本形)を採用しているのに対して,ALC SVL (12000 語)では"congratulations"(複数形)を採用している。ALC SVL(12000 語) 選定時に使用したコーパス・データの語形の出現頻度では,"outskirt, thank, work"については基本形,"congratulation"に関しては複数形が有意に高頻度であったため採用されたのであろうか。

表 2-2-1 Kilgarriff's Lemma リスト, JACET 8000 と ALC SVL における語形比較

| Kilgarriff's Lemma | JACET 8000 | ALC SVL |
|--------------------|-------------------|-----------------|
| catalogue | catalogue | catalog |
| congratulation | congratulation | congratulations |
| disc | disc | |
| enquire | enquire | inquire |
| Euro | Euro | |
| exam | exam | (examination) |
| OK | | OK |
| okay | okay | |
| outskirts | outskirts | outskirt |
| photo | photo | (photograph) |
| telecommunication | telecommunication | |
| thanks | thanks | (thank) |
| whisky | whisky | whiskey |
| works | works | (work) |
| | Apartheid | |
| | Internet | |
| | math | |

JACET 8000 では、語彙選定時に利用した英文コーパス・データの特徴が反映したためか、ALC SVL(12000 語)にはない語形 "Euro"、"Apartheid"、"Internet"が採用されている。なお、JACET 8000 の語形はすべて原形や基本形が採用されている。一方で、固有名詞(地名・人名、関連の形容詞)、曜日名・月名、数詞、頭字語、省略語、アポストロフィ(')付き語などは、出現頻度が高い場合であっても採録されていない。同音語については、ALC SVL(12000 語)では May と may、March と march、Miss と miss は別の語形としてそれぞれカウントされている。JACET 8000 では、このような語源・語義にもとづく

区別はおこなわれておらずひとつの語形とみなされている。

図 2-2-2 は、新刊書 103 冊の第1章テキスト・データの語彙を3つの語彙リストがそれぞれどの程度カバーするか次の計算式で求めた。

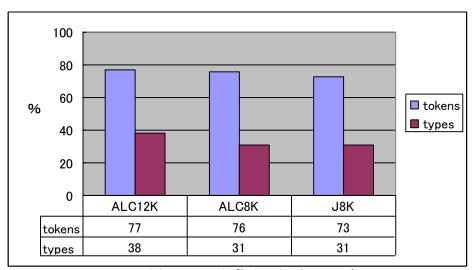


図 2-2-2 語彙リスト別カバー率

図 2-2-2 によると、各語彙リストは token レベルで新刊書 103 冊の第 1 章英文テキスト・データの語彙 73%から 77%をカバーしている。Laufer(1989)や Nation(2001)によると、外国語としての英文テキストを十分読みこなすためには、英文テキストの語彙の 95%以上が既知であることが好ましい。ニューヨーク・タイムズ紙で紹介された新刊図書を読むためには、大学生レベルの語彙表である JACET 8000 や ALC SVL(12000 語)でカバーされる出現頻度の高い語形とともに、さらに豊富なボキャブラリーが必要とされる。ちなみに、typeレベルのカバー率は 31%から 38%であった。

WS4 によって JACET 8000 未掲載語として出力された NYT103 頻度順リストを調べると、出現頻度が 10 回以上の語形は 828 語(全頻度 31,354 回)であった。その中で新刊書 103 冊中、10 冊以上に出現した語形(レンジ 10 以上)は 482 語(全頻度 24,785 回)、30 冊以上に出現した語形(レンジ 30 以上)は 87 語 2)(全頻度 15,866 回)であった。換言すると、103 冊中 30 冊以上の新刊書で 10 回以上使われている語形 87 語は、NYT103 テキスト全体で出現頻度 10 回以上の語形 828 語の 11%(87/828)にすぎないが、全出現頻度では 51%(15,866/31,354)を占める。これらの中には、多くの JACET 8000 基本語の変化形や派生形が存在している。ちなみに、レンジ、出現頻度ともに 10 の語形 482 語について見ると、それぞれの比率は 58%(482/828)、79%(24,785/31,354)となる。

NYT103 の JACET 8000 未掲載語であっても、実際にはその原形や基本形は既知のものが多い。JACET 8000 原形とその変化形、基本語とその派生形についての知識を高めることで、このような「未知の語形」の数を大幅に減少させることができる。JACET 8000 で

カバーされず、WS4 の出現頻度順リストに頻度 10 回以上の語形として現れた 828 語中、NYT103 の 10 冊以上(レンジ 10 以上)に出現した 482 語を,語形の意味や形態論上の特徴をもとに大別したものが表 2-2-2 である。

| | range ove | er 30 texts | range over 10 texts | | |
|---------------|-----------|-------------|---------------------|------------|--|
| | word form | total freq | word form | total freq | |
| 動詞の活用形 | 49 | 10922 | 254 | 15215 | |
| 名詞の複数形 | 16 | 1547 | 108 | 3454 | |
| 人名·地名等 | 2 | 299 | 36 | 1226 | |
| 数詞•曜日•月名 | 13 | 2250 | 39 | 2887 | |
| 所有格 's | 0 | 0 | 5 | 94 | |
| 縮約形 | 7 | 848 | 31 | 1737 | |
| 接尾辞 -er, -est | 0 | 0 | 9 | 172 | |
| 計 | 87 | 15866 | 482 | 24785 | |

表 2-2-2 JACET 8000 がカバーしなかった語形

すでに述べたように、JACET 8000 の語形は、動詞の原形、名詞・形容詞等の基本形である。 表 2-2-2 の「動詞の活用形」、「名詞の複数形」、「数詞・曜日・月名」、「所有格's」の付いた語形、縮約形(例 'd、't 、've など)、接尾辞('er, 'est, 'ness)の付いた派生語は、その出現頻度の高さにかかわらず JACET 8000 の「語の定義」によりリストから排除されている。同様に、高出現頻度の人名や地名、これらに由来する語形も JACET 8000 から排除されている。

未知の語彙が多い「大人の英文テキスト」を読んで概要を理解するためには、既知の語彙、縮約形、「所有格's」の付いた語形が文構造の中でどのような位置に置かれ、どのような統語上の役割を果たしているのかを認識できることが重要となる。同時に、人名または地名由来の語形が英文テキストの中でどのような文体上の役割を担っているか知ることも重要である。NYT103全体で出現頻度が30回以上の地名、人名または地名由来の語形としてWS4が出力した語形は、

(地名関連) YORK(38:122) AMERICA(27:45) ENGLISH(21:46) FRENCH(21:43) EUROPE(21:35) PARIS(20:37) BRITISH(15:39) ISRAEL(4:30)

(人名関連) JOHN(27:64) JAMES(16:32) ROBERT(15:61) WILLIAM(15:33)

VIRGINIA(13:36) JACK(10:32) ROOSEVELT(8:36) HARRY(7:54)

ROY(6:89) RUTH(5:57) DISNEY(3:71) EISNER(3:43) CARTER(3:32)

TAFT(2:35) KENNEDY(1:35) MATHEWSON(1:30) MCGRAW(1:30)

(WS4 出力リストから直接引用した語形の綴りはすべて大文字)

であった。ただし、括弧の中の数字は、語形が出現したテキスト数 (レンジ) と出現頻度を表す。 NYT103 の多くは、フィクションや人文科学または社会科学系分野のノン・フィクショ

ン新刊書の第1章である。ISRAELを除く地名や関連語形は、15冊以上に見られることから比較的に多くの新刊書で言及されている。上記の(地名関連)から、どの国家や地域がよく言及されているか推定できる。また、(人名関連)から、どのような人名がフィクションまたはノン・フィクションの登場人名として多用されているかわかる。NYT103にはテキスト数(レンジ)が10以下の(人名関連)も多く含まれている。出現頻度の割に限られた英文テキストのみに出現する人名は、特定の人物に関する言及や記述の文中に用いられていることが推測される。このような地名や人名によって形成される「英文テキスト」のコンテキストを把握することは、実際の英文テキストの正しい理解の前提となる。

Coaxhead の語彙リスト Academic Word List (以下, AWL) では、570 語の基本語と 形態論上、または語源上関連する「家族語」の存在を重視している。たとえば、analyse (米 語綴り analyze)を例に取ると、家族語 (analysed, analyser, analysers, analyses, analysing, analysis, analyst, analysts, analytic, analytical, analytically, analyze, analyzed, analyzes, analyzing)が存在する。英語学習者が体験的に学んだ語形成の規則を AWL の「家 族語」に当てはめてみると、一見未知と思われた語形が形態論上あるいは意味論上共通性 を持った既知の原形や基本形と「家族関係」にあることが推測される場合が多々ある。

文部科学省の学習指導要領から推測すると、日本の中学・高校の学生の英語語彙力 3 は必ずしも高くはない。日本の学生がその語彙力不足を補って英文テキストを読むためには、「検定済み学習用英文テキスト」の学習を通じて学んだ基本的な英文法や英語語彙の語義、動詞の変化形や名詞の複数形、形容詞・副詞の派生形などの体系的な語形成の規則や統語論上の文法規則を手がかりとすべきであろう。

(上村俊彦・県立長崎シーボルト大学)

(注)

- 1) ALC SVL 12000 http://www.alc.co.jp/goi/PW_top_all.htm
- 2) JACET 8000 の語彙をベースにレマ化し、再度 WS4 で頻度リストを出力すると、カバー されなかった 87 語のすべては JACET 8000 の「語の定義」でリストに採録されなかった語形であった。
- 3) 日本の学生の多くは、文部科学省検定済み英語教科書で英語を学び、大学入試センター試験を体験して大学生となっている。大学入試センター試験に出題される問題は、文部科学省の『学習指導要領』のもとで検定された高校教科書に基づき出題されている。『高等校学習指導要領』によると、文部科学省検定済み英語教科書で学ぶことのできる英語ボキャブラリーは 2,700 語程度である。

中学・高校の英語教科書として文部科学省の検定に合格するためには、中学英語教科書の語彙は 900 語程度、高校英語教科書の語彙は 2700 語程度(中学レベルの 900 語に加えて、「英語 I」 400 語、「英語 II」 500 語、「リーディング」 900 語)でなければいけない。平均的な日本人学生の高校終了時レベルの語彙力は、JACET 8000、ALC SVL(12000 語)はもとより、多様な語彙によって構成されている NYT103 英文テキストに遙かに届かない。

2.3 学習者の受容語彙

2.3.1 研究の背景

英語のコミュニケーション能力を推定するための指標として、語彙能力が使われることがある。この背景には、英語の運用能力で最も重要なのは語彙能力である、という主張が支持されている事実がある。上級の学習者ほどたくさんの単語を知っていて、そのレベルが高くなればなるほど、英語能力も高くなるのは、英語教師としての経験からも妥当であると思われる。

その一方で、単語の頻度と難易度は、反比例関係にあると漠然と思われてきた。具体的には、頻度が高いほど単語の難易度は易しいと思われてきた。つまり、頻度順位が高いほど、単語の難易度は低いということである。例えば、Read(1988)は、Nation(1983)の語彙レベルテスト(Vocabulary Levels Test)について、語彙レベルが上がるに従って学習者の得点が低くなることを、妥当性検証の根拠の1つとした。さらに、望月・相澤・投野(2003)によると、学習者の語彙と頻度の関係は、ピラミッドのようになっていて、頻度が高くなるほど知っている語彙が多くなり、逆に頻度が低くなれば知っている語彙が少なくなると示唆している。

この発想は、JACET 8000 (JACET 基本語改訂委員会, 2003) や Beyond 2000 (Laufer, 1995)にも見られる。前者は、British National Corpus (2000、以後 BNC)の頻度情報を中心に、日本人英語学習者が目標とすべき 8000 語を選んでいる。後者は、作文中に含まれる頻度 2000 語以内に含まれない単語の割合を算出して、エッセイのレベルを評価するものである。

しかし、単語の難易度と学習者の習得語彙の関係は、教師自身が漠然と感じてはいるが、 実証されてこなかった。本研究では、以上のような先行研究に基づき、習得されにくい語彙を難易度の高い語彙と仮定して、語彙の難易度と頻度の関係について調査し、語彙の難 易度が頻度によって予測できるかどうかを検証することにした。

2.3.2 先行研究の概観

JACET 8000 の難易度調査について、最近ではいくつか研究がなされてきた。 Aizawa(2006)は、Aizawa & Iso (2004)の開発した JACET 8000 に基づく語彙テスト (TDU VLT online)を用いて大学生の語彙力を測定した結果、頻度と単語テスト項目の正答率は、 4000 語レベルまでは関連が見られたが、5000 語を越えるレベルでは特に関連が認められなかった。この結果をレベル別に解釈すると、4000 語レベルまでの高頻度の語彙には、単語に出会う回数が増えれば増えるほど、自然と学習する機会が増えるので、それが学習を促進する要因となり、結果的に定着が早くなると結論づけた。他方で、5000 語レベルを越えると、単語の難易度は単語固有の特徴(品詞、文型、発音など)や外来語として日本語で使われているかどうかなどの要因が複雑に絡み合い、頻度と語彙の定着に直接的な関連はないものと解釈した。

Mizumoto (2006)は、8 レベルをそれぞれ 30 語 (計 240 語) 抽出し、4 選択肢でテスト する形式で調査を行った。その結果、(1)8000 語を 3 つのグループに分類できる (1000 から 2000 語レベル、4000 と 5000 語レベル、7000 語レベル以下、(2)語彙の頻度と親密度は一致しないこと、の 2 点を報告している。Aizawa の測定方法とは異なるが、難易度と頻度の関連がほぼ同じレベルを境にして見られなくなったことは、興味深い。しかし、Mizumotoが開発したテストでは、日本語訳を見てから単語を選ぶ形式のため、このテストで実際に測定している能力が受容語彙なのか発表語彙なのかの再検討が必要と思われる。

藤井(2006)は、英語力の異なるモニターに難易度のポイントを付けさせるタスクを実施した。英語力のレベルによって、モニターの評定に重み付けする方法を用いた。その結果、生起頻度から難易度順を導き出すのは困難であると結論づけた。この研究では、モニターの評価を英語力によって重み付けしているが、この方法の妥当性に検討の余地があるかもしれない。

以上の先行研究を集約すると、(1)単語の頻度レベルと難易度は、4000 語レベルまでは関連が見られること、(2)生起頻度は単語の難易度と一致しないこと、(3)単語の親密度は単語の頻度と一致しないこと、などがそれぞれ報告されている。しかし、それぞれの研究方法には再検討の余地も残り、単語の頻度と単語の難易度、英語教師による単語の難易度の評価について、最終的な結論を出すに至っていない。本研究では、JACET 8000 を語彙の頻度リストとして仮定し、語彙テストの正答率と語彙レベルとの関係を調査し、頻度によって難易度の推定が可能かどうかを調査する。

2.3.3 TDU VLT online

以下の手順で Web 版の語彙レベルテストを作成した。サンプル画面は、図 2-3-1 の通りである。

- (1) Schmitt, Schmitt, & Clapham (2001)のテスト形式を踏襲し、語義を日本語で与える形式。 各 30 語 (選択肢は 60 語) を 8 レベルでテスト。
- (2) テストは、Test 1 と Test 2 を準備。各レベル(1000 語)から計 60 語を抽出。錯乱 肢を入れれば 120 語をテストしたことになる。
- (3) 発表語彙を測定する要素を可能な限り排除した形式。
- (4) 電子辞書の使用を防ぐために、制限時間を設けた(各レベル3分、計24分)。



図 2-3-1 TDU VLT Online のサンプル画面

2.3.4 調査

(1) 目的

TDU VLT on Web を利用して、一般的大学生の英語語彙能力を測定し、頻度との関連を調査する。

(2) 被験者

首都圏の大学生 142名 (理工系 142名 + 外国語系 22名)

(3) 結果

TDU VLT Online による被験者の語彙サイズの平均は、4473.5 語(SD=944.4)となった。図 2-3-2 は、レベル別平均点と標準偏差を示している。レベル別に見ると、Level 1 が 54.42 で最も高く、Level 4 になるまでほぼ同じ割合で得点が減少した。しかし、Level 4 (27.99)から Level 5 (27.92)へは、0.07 点の得点減少にとどまった。さらに、Level 6 (25.12)で得点が減少するが、Level 7 では得点が上昇し 26.7 となった。1 要因被験者内計画の分散分析の結果、Level 4 と Level 5 の間には有意差が見られなかった。また、Level 6 と Level 7 の得点では、Level 7 の方が Level 6 よりも有意に得点が高かった(*p<.05)。結果的に、Aizawa の前回の調査と同じ傾向が認められた。

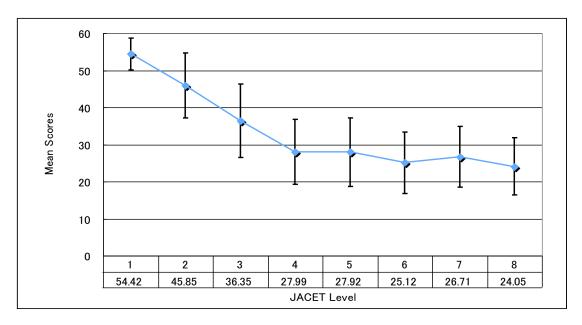


図 2-3-2 TDU VLT Online のレベル別平均点

さらに計 164 名の被験者を語彙サイズ別に 5 つのレベルに分けた。表 2-3-1 には,語彙サイズ別の各レベルの平均点と平均語彙サイズを示してある。各レベルの人数は,サイズ順に 2000 語から 6000 語に,それぞれ 12, 38, 63, 43, 8 名となった。(ただし,6000 語レベルには 7000 語レベル 1 名を含む)。

この結果からも明らかなように、6000 語レベルの被験者の Level 1 と Level 2 の得点差を除くと、Level 4 までの得点はほぼ同じ割合で減少しているが、Level 4 以降の得点には異なった傾向が見られた。レベル別の得点差を混合計画による分散分析で検定した結果、どの語彙レベルの被験者の得点でも、Level 4 と Level 5 には有意差は認められなかった。また、2000 語レベルの被験者の得点は、Level 4 から Level 8 まで統計的な違いはなかった。さらに、3000 語レベルと 6000 語レベルの被験者の得点は、Level 6 から Level 8 の間に有意差がなかった。以上の結果から、どのレベルであっても、Level 4 以降の得点間にはあまり違いがないことが明らかになった。

表 2-3-1 語彙サイズ別の各レベルの平均点及び平均語彙サイズ (/60)

| Vocabulary | N | | | | JACET | Γ Level | | | | Total |
|------------|----|------|------|------|-------|---------|------|------|------|-------|
| Size | IN | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Score |
| 2000 | 12 | 47.1 | 30.8 | 19.8 | 12.7 | 14.7 | 11.7 | 14.2 | 13.4 | 164.3 |
| 3000 | 38 | 51.3 | 38.9 | 28.0 | 20.5 | 19.6 | 18.7 | 19.6 | 17.4 | 214.1 |
| 4000 | 63 | 54.9 | 45.6 | 36.0 | 28.2 | 27.8 | 25.0 | 27.4 | 24.0 | 268.7 |
| 5000 | 43 | 57.8 | 54.2 | 45.5 | 35.2 | 35.9 | 31.8 | 33.0 | 30.4 | 323.8 |
| 6000 | 8 | 58.3 | 58.5 | 54.0 | 46.4 | 45.9 | 40.6 | 40.6 | 38.3 | 382.5 |

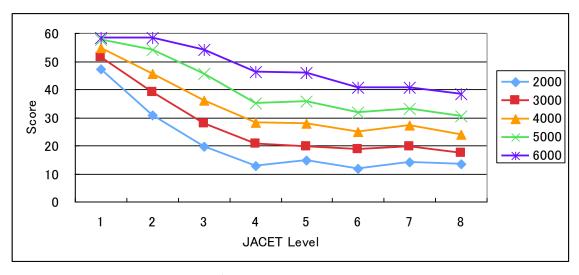


図 2-3-3 TDU VLT Online の語彙サイズ別平均点

2.3.5 考察

本研究の結果,いくつかの興味深い点が明らかになった。まず、学習者の語彙テストの結果では、先行研究の結果が裏付けられた。つまり、単語の難しさと頻度(本研究では便宜上 JACET 8000 とした)の関係は、4000 語レベルまでは関連が認められたが、それ以上のレベルでは頻度の得点の関連はほとんど見られなかった。また、学習者の語彙サイズを2000 語レベルから 6000 語レベルに分けて分析したが、ほぼ同じ傾向が認められた。

この結果は、2つの点で Aizawa (2006)の調査結果に対して示唆を与えてくれる。1つは、前回の調査の被験者が理工系学生のみであったことが結果に影響したのではないか、という疑問があった。本研究では、被験者数は限られるが、外国語系の被験者を確保して調査したが、前回と同じ傾向が顕著に現れた。この結果を見る限りでは、学習者の専門領域が語彙知識に大きな影響を与えているとは考えにくい。実際、工学的な専門英語を学習するのは卒業研究に進んでからで、被験者の英語学習はこの段階にまで進んでいない。

もう1つは、前回の被験者の語彙レベルが平均 4000 語台であることが、4000 語レベルに threshold level が認められた原因ではないか、という疑問もあった。今回は、比較的英語を得意とする被験者を増やしたが、4000 語に大きな境界が認められた。今回の結果は、英語学習に必要な core vocabulary とその後に続く ESP の語彙の境界が、4000 語レベル前後に存在している可能性を示唆するものである。

2.3.6 今後の課題

今回の調査では、被験者のレベルや専門性に限界があった。次回の調査では、他のレベルや専門領域の学生に協力を得て、今回の発見をより一般化できるようにすることが待たれる。

(相澤一美・東京電機大学, 磯達夫・麗澤大学)

引用文献

- Aizawa, K. (2006). "Rethinking frequency markers for English-Japanese dictionaries," In Murata, M., Minamide, K., Tono, Y., & Ishikawa, S. (Eds.) *English Lexicography in Japan* (pp.108-119), Taishukan-shoten.
- Aizawa, K. & Iso, T. (2004). TDU vocabulary levels test online. Available online: http://jacetvoc.kl.dendai.ac.jp/~aizawa/
- British National Corpus (2000). Oxford: Oxford University.
- 藤井哲郎(2006)「英語語彙のレベル分け基準-4800 語の親密度調査と JACET 8000 との違い」第7回英語辞書学ワークショップロ頭発表資料,和洋女子大学.
- JACET List of 8000 Basic Words (2003). Tokyo. Japan Association of College English Teachers.
- Laufer, B. (1995). Beyond 2000. A measure of productive lexicon in a second language. In L. Eubank, L. Selinker & M. Sharwood-Smith (Eds.), *The Current State of Interlanguage* (pp. 265-272). Amsterdam: John Benjamins.
- Mizumoto, A (2006). Developing another type of vocabulary size test. Paper presented at the 7th JACET English Lexicography SIG Workshop, Wayo Women's University.
- 望月・相澤・投野(2003). 『英語語彙の指導マニュアル』大修館書店.
- Nation, I.S.P. (1983). Testing and teaching vocabulary. Guideline, 5, 12-25.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19 (2), 12-25.
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55-88.

2.4 学習者の発表語彙

2.4.1 研究目的

英語学習者の自由英作文における語彙的な豊かさ(lexical richness)を測定する指標としては、Lexical Frequency Profile (Laufer & Nation, 1995)が知られているが、本研究では、日本人英語学習者(大学生)のエッセイライティングにおける使用語彙を JACET 8000 を用いて分析し、それを通じて学習者のエッセイにおける使用語彙のレベル分布についての考察を行う。

2.4.2 エッセイライティングの授業

今回分析の対象とした英文エッセイは,理工系大学の1年生を対象としたライティングの授業で課題として提出されたものである。年度当初に独自のプレースメントテスト(リスニング,リーディング,文法・語彙)を用いてクラス分けを行い,上位クラスと下位クラスに分けて授業を行った。クラスサイズは1クラス約35名となった。受講者は授業中にパラグラフライティングに関する指導を受け、250-300語の英文エッセイを書く課題を年間で8回与えられた。エッセイを書く際には辞書の使用は許可されており,授業時間外にも宿題として自宅でエッセイを書く作業を行わせた。エッセイの提出に関しては,課題をオンラインで電子的に提出するシステムがWeb上に構築され,学生は作成した英文エッセイをそれを用いて提出した。なお,授業ではテキストとして「First Steps in Academic Writing」(Addison Wesley Longman, 1996)を使用した。

2.4.3 エッセイ課題の収集とデータ処理

課題として提出されたエッセイの分析は、上位クラスと下位クラスの課題トピックの統一性を考慮して、収集したエッセイ課題のうち同じトピックで書かれた4回分(「How to Plan a Trip」「My Room」「The Best Pet」「My Opinion」)のみを分析の対象とした。提出された課題にはスペリングの誤りを含むものもあったが、今回の分析は使用語彙の分析であるので、明らかな誤りに対してはスペリングの修正を行った。分析に使用したソフトウェアは清水伸一氏の語彙分析用 Perl スクリプト(V8an.pl)及び Wordsmith である。エッセイにおける使用語彙の分析結果を以下に示す。

表 2-4-1 エッセイの使用語彙に対するレベル毎のカバー率

| JACET | level | level | level | level | level | level | level | level | others | total |
|---------|--------|-------|-------|-------|-------|-------|-------|-------|--------|--------|
| 8000 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | |
| indexes | 1064 | 837 | 634 | 387 | 307 | 243 | 207 | 172 | 2346 | 6197 |
| (types) | 1004 | 037 | 034 | 307 | 307 | 243 | 207 | 1/2 | 2340 | 0197 |
| % | 17.17 | 13.50 | 10.23 | 6.24 | 4.95 | 3.92 | 3.34 | 2.77 | 37.85 | 100 |
| tokens | 144378 | 9045 | 4598 | 2088 | 1018 | 1596 | 632 | 753 | 6241 | 170349 |
| % | 84.75 | 5.31 | 2.69 | 1.22 | 0.59 | 0.93 | 0.37 | 0.44 | 3.66 | 100 |

表 2-4-2 Wordsmith によるエッセイデータの統計量

| Standardized Type/Token | 33.62 |
|----------------------------|--------|
| Ave. Word Length | 4.1 |
| Sentences | 13,407 |
| Sentence length | 12 |
| Standardized. Sent. Length | 7.13 |

2.4.4 分析

2.4.4.1 全体データの分析

収集したエッセイ課題の全データサイズは、type 数が 6,197 語、token 数が 170,349 語となった。図 2-4-1 は、学習者のエッセイでは、どのレベルの語彙がどの程度の割合使用されているかを表 2-4-1 のデータを基に示したものである。グラフの横軸は JACET 8000 における語彙レベルを示しており、L1 は 1000 語レベル、L2 は 2000 語レベルというように表記されている。縦軸(対数表示)には各レベルの語彙がエッセイの中で何%を占めたかを Type(異なり語数)と Token(総語数)別に示している。

学習者がどれだけの異なり語数を用いているかに関しては、3000 語レベルまでの総計で40%を超えているが、6000 語~8000 語レベルの語彙もそれぞれ3%前後使用されている。これについては、辞書使用を許可したことが影響していると考えられる。一方、総語数で見ると、1000 語レベルの語彙で約85%を占め、2000 語レベルまでの語彙で90%を超えている。このことから考えると、今回調査した大学生のエッセイで使用された語彙の大多数は2000 語レベルまでであったことが示されている。これについては、エッセイの長さが250-300 語であったことや、トピック自体がカジュアルなものであったことも影響していると考えられるが、日本人英語学習者の発表語彙のサイズを間接的に示す一つの指標になると考えられる。

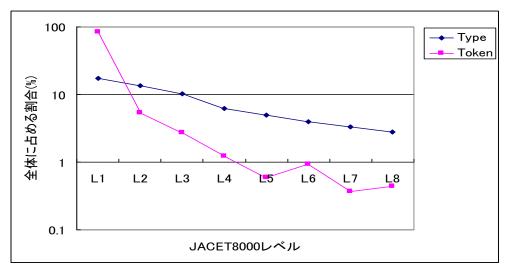


図 2-4-1 エッセイで使用された語彙のレベル別割合

次に、エッセイ中で使用された上位 300 語の頻度リストを以下に示す(レマ化済)。上位の語は多くが機能語で占められているが、最も使用頻度が高かった語が"I"であったのは、モノローグスタイルのエッセイであったためだと推測される。また、"dog"、"trip"、"travel"等の語彙が比較的上位にきているのは、トピックの影響であると考えられる。これら上位300 語は JACET 8000 では全てレベル 1 の語彙であった。

表 2-4-3 エッセイで使用された頻度上位 300 語リスト (同頻度の語はアルファベット順)

| 順位 | 語彙 | 頻度 | 順位 | 語彙 | <u>頻度</u> | 順位 | 語彙 | 頻度 |
|----|------|------|------------|-----------------------|-----------|----|--------|-------|
| 1 | I | 8326 | 28 | think | 922 | 55 | get | 526 |
| 2 | be | 7933 | 29 | they | 914 | 56 | second | 522 |
| 3 | to | 6074 | 30 | on | 905 | 57 | take | 515 |
| 4 | the | 5292 | 31 | good | 859 | 58 | which | 504 |
| 5 | a | 4220 | 32 | because | 838 | 59 | also | 474 |
| 6 | and | 4202 | 33 | when | 833 | 60 | friend | 470 |
| 7 | of | 3070 | 34 | there | 803 | 61 | them | 468 |
| 8 | in | 2701 | 35 | thing | 776 | 62 | three | 428 |
| 9 | have | 2470 | 36 | people | 760 | 63 | your | 412 |
| 10 | it | 2319 | 37 | this | 736 | 64 | these | 406 |
| 11 | you | 1999 | 38 | first | 723 | 65 | travel | 395 |
| 12 | for | 1937 | 39 | by | 668 | 66 | use | 389 |
| 13 | not | 1871 | 40 | dog | 668 | 67 | exampl | e 384 |
| 14 | my | 1757 | 41 | time | 657 | 68 | third | 382 |
| 15 | that | 1756 | 42 | trip | 637 | 69 | place | 381 |
| 16 | do | 1586 | 43 | make | 618 | 70 | decide | 380 |
| 17 | so | 1532 | 44 | play | 608 | 71 | day | 378 |
| 18 | can | 1496 | 45 | or | 607 | 72 | much | 370 |
| 19 | go | 1396 | 46 | at | 603 | 73 | must | 367 |
| 20 | like | 1288 | 47 | reason | 598 | 74 | game | 365 |
| 21 | if | 1277 | 48 | should | 594 | 75 | other | 360 |
| 22 | very | 1226 | 49 | as | 592 | 76 | from | 359 |
| 23 | want | 1119 | 50 | all | 587 | 77 | one | 347 |
| 24 | will | 1030 | 51 | many | 573 | 78 | money | 342 |
| 25 | but | 982 | 52 | he | 570 | 79 | some | 342 |
| 26 | with | 965 | 5 3 | about | 540 | 80 | who | 340 |
| 27 | we | 964 | 54 | me | 534 | 81 | more | 339 |

| 順位 | 語彙 | <u>頻度</u> | 順位 | <u>語彙</u> | 頻度 | 順位 | <u>語彙</u> | 頻度 |
|-----|---------|-----------|-----|-----------|--------|-----|-----------|-------|
| 82 | eat | 333 | 117 | how | 218 | 152 | fish | 156 |
| 83 | then | 332 | 118 | high | 212 | 153 | two | 156 |
| 84 | watch | 332 | 119 | up | 205 | 154 | live | 155 |
| 85 | their | 311 | 120 | would | 204 | 155 | sport | 153 |
| 86 | only | 303 | 121 | long | 200 | 156 | out | 151 |
| 87 | become | 297 | 122 | phone | 200 | 157 | home | 148 |
| 88 | what | 292 | 123 | Japanes | se 196 | 158 | way | 148 |
| 89 | buy | 291 | 124 | after | 195 | 159 | give | 147 |
| 90 | may | 289 | 125 | feel | 194 | 160 | opinion | 146 |
| 91 | music | 289 | 126 | too | 192 | 161 | world | 145 |
| 92 | cat | 285 | 127 | him | 189 | 162 | Japan | 144 |
| 93 | life | 278 | 128 | hotel | 189 | 163 | small | 141 |
| 94 | need | 276 | 129 | look | 188 | 164 | therefor | e 141 |
| 95 | student | 275 | 130 | person | 186 | 165 | family | 140 |
| 96 | book | 272 | 131 | often | 185 | 166 | new | 140 |
| 97 | see | 269 | 132 | say | 185 | 167 | why | 140 |
| 98 | than | 267 | 133 | know | 184 | 168 | part | 138 |
| 99 | food | 257 | 134 | listen | 183 | 169 | easy | 137 |
| 100 | lot | 257 | 135 | animal | 180 | 170 | happy | 136 |
| 101 | next | 256 | 136 | where | 177 | 171 | horse | 135 |
| 102 | now | 256 | 137 | English | 171 | 172 | country | 134 |
| 103 | finally | 252 | 138 | while | 171 | 173 | study | 134 |
| 104 | train | 250 | 139 | kind | 169 | 174 | every | 133 |
| 105 | plan | 246 | 140 | stay | 167 | 175 | able | 132 |
| 106 | child | 238 | 141 | such | 167 | 176 | talk | 132 |
| 107 | bad | 235 | 142 | especial | ly 165 | 177 | house | 131 |
| 108 | his | 235 | 143 | read | 165 | 178 | body | 129 |
| 109 | an | 233 | 144 | no | 163 | 179 | come | 127 |
| 110 | keep | 233 | 145 | movie | 162 | 180 | work | 124 |
| 111 | most | 233 | 146 | prepare | 162 | 181 | TV | 122 |
| 112 | importa | int 232 | 147 | big | 160 | 182 | before | 122 |
| 113 | year | 220 | 148 | our | 158 | 183 | cost | 122 |
| 114 | enjoy | 219 | 149 | point | 158 | 184 | car | 121 |
| 115 | school | 219 | 150 | since | 158 | 185 | however | r 121 |
| 116 | univers | ity 219 | 151 | always | 156 | 186 | room | 121 |

| 順位 | <u> 語彙</u> | <u>頻度</u> | 順位 | 語彙 | 頻度 | 順位 | <u> 語彙</u> | <u>頻度</u> |
|-----|------------|-----------|-----|-----------|---------|-----|------------|-----------|
| 187 | street | 121 | 222 | change | 91 | 257 | recently | 73 |
| 188 | interest | ing 120 | 223 | little | 91 | 258 | languag | ge 72 |
| 189 | call | 119 | 224 | underst | and 91 | 259 | same | 72 |
| 190 | problem | 118 | 225 | hard | 90 | 260 | week | 72 |
| 191 | human | 117 | 226 | help | 90 | 261 | action | 71 |
| 192 | walk | 116 | 227 | informa | tion 90 | 262 | could | 71 |
| 193 | class | 115 | 228 | trouble | 89 | 263 | feeling | 71 |
| 194 | us | 115 | 229 | away | 88 | 264 | mind | 71 |
| 195 | sleep | 114 | 230 | choose | 87 | 265 | race | 71 |
| 196 | she | 113 | 231 | eye | 87 | 266 | club | 70 |
| 197 | into | 112 | 232 | learn | 87 | 267 | those | 70 |
| 198 | dream | 111 | 233 | throw | 87 | 268 | few | 69 |
| 199 | well | 110 | 234 | summer | 86 | 269 | large | 69 |
| 200 | speak | 109 | 235 | fact | 85 | 270 | teacher | 69 |
| 201 | various | 108 | 236 | last | 83 | 271 | another | 68 |
| 202 | love | 106 | 237 | drink | 82 | 272 | bring | 68 |
| 203 | beautifu | ıl 105 | 238 | course | 81 | 273 | case | 68 |
| 204 | fire | 103 | 239 | forget | 80 | 274 | down | 68 |
| 205 | even | 102 | 240 | hour | 79 | 275 | once | 68 |
| 206 | health | 102 | 241 | night | 79 | 276 | ride | 68 |
| 207 | sea | 99 | 242 | stop | 79 | 277 | somethi | ng 68 |
| 208 | never | 97 | 243 | player | 78 | 278 | video | 68 |
| 209 | old | 97 | 244 | run | 78 | 279 | just | 67 |
| 210 | without | 97 | 245 | story | 78 | 280 | song | 67 |
| 211 | around | 96 | 246 | job | 77 | 281 | usually | 67 |
| 212 | check | 96 | 247 | number | 77 | 282 | write | 67 |
| 213 | fly | 96 | 248 | spend | 77 | 283 | meet | 66 |
| 214 | cause | 95 | 249 | hand | 76 | 284 | pretty | 66 |
| 215 | man | 95 | 250 | difficult | 75 | 285 | teach | 66 |
| 216 | sometin | nes 95 | 251 | find | 75 | 286 | visit | 66 |
| 217 | care | 94 | 252 | shop | 75 | 287 | die | 65 |
| 218 | its | 94 | 253 | free | 74 | 288 | order | 65 |
| 219 | any | 93 | 254 | future | 74 | 289 | seat | 65 |
| 220 | her | 92 | 255 | over | 74 | 290 | start | 65 |
| 221 | name | 92 | 256 | near | 73 | 291 | during | 64 |

| 順位 | 語彙 | 頻度 |
|-----|------------------------|--------|
| 292 | sing | 64 |
| 293 | experier | nce 63 |
| 294 | hope | 63 |
| 295 | public | 63 |
| 296 | meal | 62 |
| 297 | try | 62 |
| 298 | almost | 61 |
| 299 | color | 61 |
| 300 | put | 61 |

2.4.4.2 上位クラス・下位クラスの使用語彙比較

次に、上位クラスと下位クラスの間での使用語彙の差異についての分析を行った。上位クラス、下位クラスからそれぞれ 1 クラスを選び、エッセイ毎の使用語彙のレベル分析を行い比較した。それぞれのクラスのエッセイにおける使用語彙のレベル分布とその割合 (Token で算出)を以下に示す。図中の 4 本の折れ線はそれぞれのトピック毎の数値(1回目~4回目)を示している。なお、上位クラスと下位クラスの間には、プレースメントテストの成績において有意差があることが確認されている(F=763.1 df=1 p<0.01)。

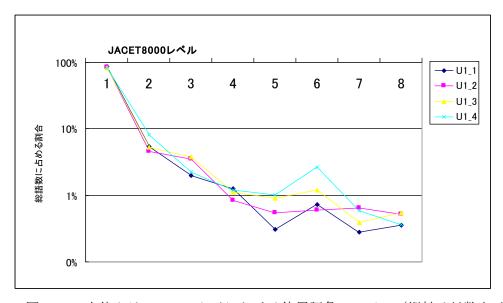


図 2-4-2 上位クラスのエッセイにおける使用語彙のレベル (縦軸は対数表示)

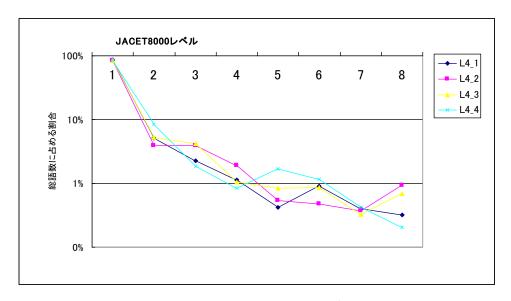


図 2-4-3 下位クラスのエッセイにおける使用語彙のレベル (縦軸は対数表示)

グラフのデータでは、上位学習者、下位学習者共に 1000 語レベルの語彙が 80 数パーセントを占めており、2000 語レベルの語を加えるとほぼ 90 パーセントを占める。また、4000 語レベルまではトピックによる使用語彙レベルのばらつきが比較的少なく、上位クラスと下位クラスの間での差もそれほど大きくないが、5000 語レベル以上になるとトピックによる使用割合のばらつきが増加する傾向が見られる。 語彙レベルが高く難しい語彙は使用頻度自体が低いということもあるが、これらのレベルの語彙の使用割合に学習者のレベルやトピックの影響が出てくると考えられる。

今回調査した4つのエッセイはグラフ中の数字の順のトピック($1\rightarrow 4$)で学生に課題として与えられたが,上位クラスの学習者のデータでは,後の課題になるに従って高い語彙レベルの語彙の使用割合が増えていく傾向が見られる。一方,下位クラスのエッセイでは 5000 語レベルを除いてこの様な傾向は見られていない。これには,各回のトピックの違いが影響を与えている可能性もあるが,上位学習者は,エッセイを何度か書いていくうちに,高いレベルの語彙をエッセイ中で使用する割合が増加する傾向があるのに対して,下位クラスの学習者は,例え辞書を使っていても,そのような力がまだ十分には養成されていない可能性があることが示されている。これについては,今後より多くのデータを分析して追加検証を行う必要がある。

2.4.5 結語

今回調査したエッセイにおいて使用された語彙は、全体として上位 2000 語までのものが大半であり、平易な語彙が多く使用される傾向が見られたが、これは、受講した学生の英語力の問題に加えて、エッセイの長さが 250-300 語という比較的短いものであったため、エッセイの中での話題の展開に制限があったことも一因と考えられる。また、今回のエッ

セイ課題では辞書使用を許可し、時間制限をせずに課題を書かせることを許可したので、個々の学習者が持つ発表語彙を必ずしも直接的に反映していない可能性もあるが、トピックの違いにかかわらず、エッセイライティングにおける使用語彙のレベル別割合は 4000 語レベルまでは一定の数値を取る傾向があることが示された。また、このレベルでは、上位学習者の方がトピックの違いによるレベル別割合のばらつきが少ない傾向を示している。しかしながら、5000 語レベル以上の語彙の使用に関しては、学習者の英語力やトピックの影響を受けてばらつきが発生しやすいという結果になった。

今後は、他大学の日本人英語学習者のエッセイとの比較、ICLE(International Corpus of Learner English)等を利用した他の EFL 学習者のエッセイとの比較、LOCNESS(Louvain Corpus of Native English Essays)等を利用したネイティブスピーカーのエッセイとの比較等の研究を行う予定である。また、より長いエッセイを書かせた場合との比較や、辞書使用を許可しない場合との比較、時間制限を加えた timed-writing をさせた場合との比較などを通して、英語学習者のエッセイライティングにおける使用語彙の特性を更に調査する必要があると考えられる。

(杉森 直樹・立命館大学)

参考文献

- Cushing-Weigle, S. (2002). Assessing Writing. Cambridge: Cambridge University Press.
- Hogue, A. (1996). First Steps in Academic Writing. New York: Addison-Wesley Longman.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

2.5 単語の難易度に関する英語教員の評価

2.5.1 研究の背景

JACET 8000 (JACET 基本語改訂委員会, 2003) は、British National Corpus(BNC) の頻度情報を中心に、日本人英語学習者が目標とすべき 8000 語を選んでいる。しかし、単語の難易度と学習者の習得語彙の関係は、教員自身が漠然と感じてはいるが、実証されてこなかった。特に、教員が単語の難易度をどのように感じながら指導しているのか、についての先行研究はほとんどない。

本研究では、習得されにくい語彙を難易度の低い語彙と仮定して、教員の難易度の推定 と語彙の頻度の関係について調査し、語彙の難易度が頻度や教員の直感によって予測でき るかどうかを検証することにした。

2.5.2 調査

(1) 目的

TDU VLT online で目標語として意味を尋ねた単語に対して、英語教員がどのように難 易度を判定するかを調査し、JACET 8000 のレベルと比較検証する。

(2) 被験者

日本の高校・大学レベルの英語教員 40 名(日本人教員 38 名, 母語教員 2 名) に協力を依頼した。

(3) アンケートの方法

以下の手順で教員アンケートを作成して、Web 上での回答の協力を依頼した。アンケートのサンプル画面は、図 2-5-1 に示すとおりである。

- 1) TDU VLT online で解答を求めた単語 480 語を 80 語ずつ6 セットに分け、1 セット ずつ日本語訳なしのリストを作成した。さらに、被験者ごとにセットの提示順を6 通 りに提示できるようにした。この配慮は、回答する順番が難易度の評定に影響を与え ないようにするためである。
- 2) 被験者へのアンケートのお願いの際に、「次の単語を『難しい単語』と『それ以外の単語』の 2 回に分けて提示することを想定した場合に、『難しい単語』として導入する単語にチェックを入れてください。ただし、各 1 セット(80 語)のうち、「難しい単語」と「それ以外の単語」の数が半数にならなくとも結構です。」という指示を出した。このように指示を単純化したのは、質問項目数が多いので、できるだけアンケート協力者の負担を軽減するためである。
- 3) 制限時間は特に設けなかった。

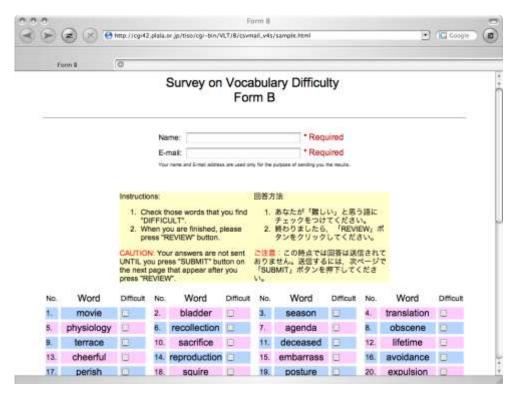


図 2-5-1 Web 版教員アンケートのサンプル画面

2.5.3 結果

アンケートの協力を得た 40 名の英語教員が「難しい語」として選んだ語(クリックした単語数)の基礎統計は、表 2-5-1 の通りである。延べ 4486 語が選択され、一人あたりの平均は、112.15 語(SD=73.1)となった。「難しい語」と「それ以外の語」に2回に分けて導入することを前提としたが、難しい語として選択された語は 25%程度に限られた。しかし、その一方で、最大語数は 289 語におよび、きわめて個人差が大きいことが明らかになった。

表 2-5-1 教員アンケートの基礎統計

| N | Total | Mean | SD | Morr | Min |
|----|--------|---------|-------|------|------|
| N | Clicks | (/480) | SD | Max | WIII |
| 40 | 4486 | 112.15 | 73.10 | 289 | 0 |

図 2-5-2 は英語教員が、レベル別に選択した延べ語数を示している。Level 8 から Level 1 に向けて期待値の線を引くと、ほぼこの線上に実測値が並んだ。1要因 8 水準の分散分析と LSD の多重比較によると、Level 1 と Level 2、Level 2 と Level 3、Level 4 と Level 5、Level 6 と Level 7、 Level 7 と level 8 の相互の間には、5%水準で有意差は認められなかった。つまり、Level 3 と Level 4、Level 5 と Level 6 の間にのみ、隣接するレベルでは、

差があったと結論づけられる。以上の結果から、教員は Level 4 と Level 6 の語彙を、相対的に難しいと判断していることを除けば、 JACET のレベルと教員の難易度の直感が全体的にほぼ一致していることが明らかになった。

次に、3名以上の教員(危険率 5% = 2人)によって難しいと判定された語を「難しい語」と仮定し、その総語数を算出した。危険率を考慮したのは、今回の調査が HP で該当する項目にクリックする形式で実施したため、操作の誤りで選ばれた語が含まれている可能性があったからである。図 2-5-3 は、難しい語とそれ以外の語の比較である。この図から明らかなように、4000 語までは右肩上がりで語数が増えていくが、4000 語を越えるレベルから「難しい語」が相対的に少なくなっていることが明らかになった。

さらに「難しい語」の基準を、「半数以上の教員が難しいと判定した語」と仮定して、その語数を算出した。図 2-5-4 は、レベル別の「難しい語」を示している。Level 3 までは、半数の教員が一致して選んだ語はほとんどなかったが、4000 語を越えるレベルから語数が増えている。興味深いことは、Level 6 の方が Level 7 よりも難しいと判定された語数が多かったことである。また Level 8 は語数が大きく上がるが、教員にとっても難しく感じられたのかもしれない。

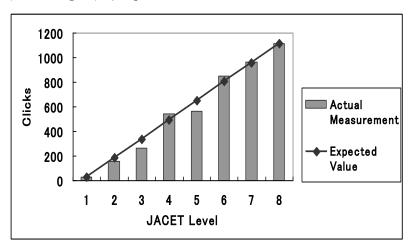


図 2-5-2 教員アンケートによるレベル別の難易度の延べ語数

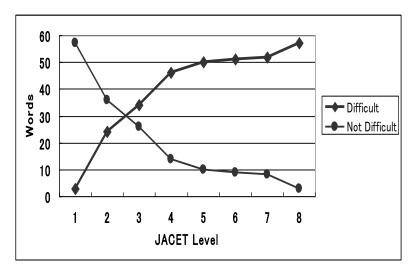


図 2-5-3 「難しい語」と「それ以外の語」の語数のレベル別比較 (60 語中)

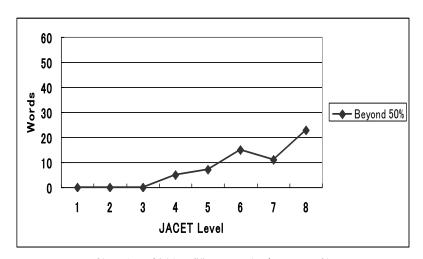


図 2-5-4 半数以上の教員が難しいと判定した語数

2.5.4 考察

アンケート結果は、分析方法によって異なった示唆が得られた。教員が「難しい語」として選択した延べ語数は、JACET 8000 のレベルが低頻度になるにしたがって、増える傾向が見られた。統計的には、隣接するレベルで違いが認められたのは Level 3 と Level 4、Level 5 と Level 6 の間のみであったが、延ベクリック数をグラフにすると、図 2-5-2 のようにほぼ一直線上になった。教員は、単語の頻度が低いとその単語は難しいと漠然と感じているが、JACET 8000 を使ってそれがほぼ証明されたことになる。この結果は、JACET 8000 のレベルが、単語の難易度を示していると解釈できるかもしれない。この解釈が可能であれば、教材作成や教材分析の際に、例えば JACET 8000 に基づくレベルの判定を行い、教材の難易度を明示することも可能になると考えられる。

次に、危険率5%を考慮して3名以上と、半数の20名以上の教員が選んだ語数を分析し

た。その結果,3名以上が難しい語として選択した単語の総数は,学習者の正答率とほぼ同じ傾向が見られた。つまり,4000 語レベル付近までは一定の割合で下降し,ほぼ一直線となった。この結果は、学習者にとって難しい単語とそれ以外の単語の別を,教員自身が比較的正確に理解していることを示唆していると解釈できる。

しかし、半数以上の教員が選択した語彙数は、3名以上が選んだ語彙数と比較すると、 上昇するレベルが 3000 語レベル以降となった。この結果は、全体的に低頻度になるほど 難しい単語の数が増えるという点においては教員の意見が一致しているが、どの単語が難 しいかという点においては、教員間に意見のばらつきがあることを示している。つまり、 難しい単語の数は 3000 語レベルを越えると増えてくるが、教員によって難しいと感じる 単語に差があることを示している。

以上の考察を総括すると、教員の単語の難易度の予測結果では、JACET 8000 の頻度と 難易度の関係は全体的にバランスが取れており、学習語彙表としての妥当性があると結論 づけられる。

(相澤一美・東京電機大学, 磯達夫・麗澤大学)

3. 既存コーパス資料の概観

今回のプロジェクトのために海外から購入した主なコーパスは、以下の通り1)である。

ANC American National Corpus First Release http://americannationalcorpus.org/FirstRelease/

BNC British National Corpus http://www.natcorp.ox.ac.uk/corpus/index.xml

ICE-GB British Component of the International Corpus of English http://www.ucl.ac.uk/english-usage/projects/ice-gb/index.htm

MICASE Michigan Corpus of Academic Spoken English http://micase.umdl.umich.edu/m/micase/

SBC Santa Barbara Corpus of Spoken American English Part-I http://projects.ldc.upenn.edu/SBCSAE/

清水伸一氏の V8 Level Marker により確認した各コーパスのデータ量 (index・token) は、BNC (68,596・11,328,146)、MICASE(28,848・1,794,311)、ANC(22,274・3,306,718)、ICE-GB (18,205・657,861)、SBC(8,737・222,973)であった。各コーパスについて、index 総数に対する JACET8000 カバーの語形(index)数の割合を算出すると、SBC(50.62%) > ICE-GB(39.65%) > ANC(32.87%) > MICASE(25.59%) > BNC (11.84%)となり、カバー率はデータ量と反比例の関係となった(図 3・1 参照)。

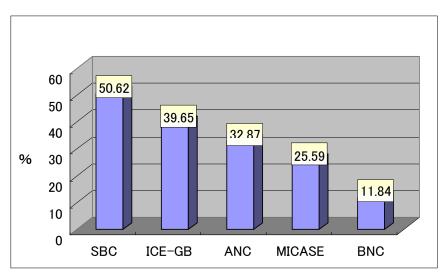


図 3-1 JACET 8000 による各コーパスのカバー率

図 3-2 は、JACET8000 でカバーされた各コーパスの index について、レベルごとの出現率を算出しグラフ化したものである。

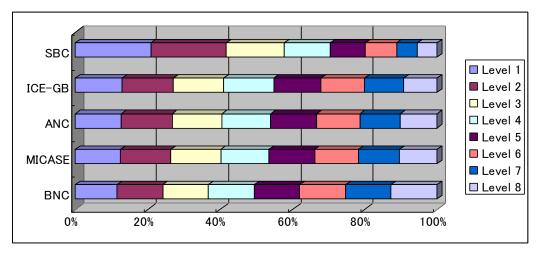


図 3-2 レベル別出現率

5つのコーパス中,最も小規模な SBC は,index 総数に占める JACET8000 カバー語の割合では最高率となったが,レベルごとの index の出現率では最低の割合となった。SBC を除く 4つのコーパスは,Level6でも 1000 語中 800 語を超える一致が認められたが,SBC では,Level4 で 1000 語中 540 語程度,Level5 では 1000 語中 400 語程度の一致にとどまった。Level8 では,SBC に次いで小規模なコーパスである ICE-GB が 640 語程度であったのに対して,SBC は大きく数値を下げて 220 語程度となった。

3.1 ICE-GB

3.1.1 ICE-GB の概要

The British Component of International Corpus of English (以下, ICE-GB)は、本プロジェクト予算を活用して購入した海外コーパスの1つである。 Greenbaum (1996:pp.5-6) によると、ICE-GB は 100 万語コーパス(500 テキスト×2000 語)で、「話し言葉」(300 テキスト)と「書き言葉」(200 テキスト)から構成されている。ただし、「話し言葉」テキストのうち、50 は会話スクリプトである。 1 テキスト・データは、1990年から 1994年に集められた教養のある、標準的な英語か構成されている。表 3-1-1 括弧内の数字はテキスト数を示す。

表 3-1-1 ICE テキストのカテゴリー

| Spoken | Dialogues | Private | Conversations (90) |
|---------|-------------|------------|----------------------------|
| (300) | (180) | (100) | Phonecalls (10) |
| | | Public | Class Lessons (20) |
| | | (80) | Broadcast Discussions (20) |
| | | | Broadcast Interviews (10) |
| | | | Parliamentary Debates (10) |
| | | | Cross-examinations (10) |
| | | | Business Transactions (10) |
| | Monologues | Unscripted | Commentaries (20) |
| | (120) | (70) | Unscripted Speeches (30) |
| | | | Demonstrations (10) |
| | | | Legal Presentations (10) |
| | | Scripted | Broadcast News (20) |
| | | (50) | Broadcast Talks (20) |
| | | | Non-broadcast Talks (10) |
| Written | Non-printed | Student | Student Essays (10) |
| (200) | (50) | Writing | Exam Scripts (10) |
| | | (20) | |

3.1.2 有用性

全体で 500 のテキストから成るコーパス・データは,付属の統語構造解析ソフトウェア International Corpus of English Corpus Utility Program (ICECUP) によって、個々の 語には詳細な POS タグ情報、clause には詳細な統語構造情報が付与されている。また、ICEC-GB と同様の仕様で、インド英語、シンガポール英語、ナイジェリア英語など,世界 各地域の英語コーパスが構築されており、ICEUP を用いて相互比較することができる。

ICE-GB は 1990 年代前半のイギリス英語である,最新のイギリス英語の研究を試みる場合,すでにコーパス構築から 10 年以上経過していることを念頭に置くべきであろう。

3.1.3 データフォーマット

コーパス・データには、テキストの構成要素(文、パラグラフ、表題等)、POS タグ、統語解析などの情報が付与されており、 ICECUP を使うことでコーパステキストに関するさまざまなパーシングが可能である。²⁾ なお、ICECUP のテキスト構造記述や統語構造記述の原型は、Svartvik & Quirk (1979) *Survey of English Usage and Survey of Spoken English* にある。

すべてのデータは、下記のように ICE 独自のフォームでタグ付けされているため、頻度順語彙リストを出力するには実際の発話部分のみを抽出することが必要である。

[<#2:1:B> <sent>]
PU,CL(main,inter,intr,pres)
INTOP,AUX(modal,pres) {Shall}
SU,NP()
NPHD,PRON(pers,sing) {I}
VB,VP(intr,infin,modal)
MVB,V(intr,infin) {go}
A,AVP(ge)
AVHD,ADV(ge) {first}
[<\$C>]

ICE-GB に関する資料には,以下のものがある。

Greenbaum, S. ed. (1996) Comparing English Worldwide: The International Corpus of English. Oxford: Clarendon Press Oxford.

Nelson, G.: Wallis, S. & Aarts, B. (2002) Exploring Natural Language: Working with the British Component of the International Corpus of English. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Svartvik, J. & Quirk, R. eds. (1979) *A Corpus of English Conversation*. Lund: C W K Gleerup Lund.

- 1) 詳細は、http://www.ucl.ac.uk/english-usage/ice/design.htm#を参照。
- 2) ICE のタグ付けについては、http://www.ucl.ac.uk/english-usage/ice/annotate.htm を 参照。)

(県立長崎シーボルト大学・上村俊彦)

- 3.2 British National Corpus
- 3.2.1 BNC の概要

British National Corpus (以下 BNC)は、現代イギリス英語の1億語の均衡コーパスである。1991年に構築が開始され、1994年に完成した。2000年までは EU 圏内に使用が制限されていたが、BNC World Edition の公開により、全世界で利用できるようになった。BNC の構築は BNC Consortium という組織を作って、産学協同プロジェクトで行われた。これは Oxford University Press が中心となり、それに Addison-Wesley Longman、

Larousse Kingfisher Chambers といった辞典出版社, Oxford University Computing Services (OUCS), University Centre for Computer Corpus Research on Language (UCREL: ランカスター大学), British Library's Research and Innovation Centre といった大学の研究機関が参画した。さらに、財政的支援を Science and Engineering Council (当時:現在の EPSRC), DTI, British Library, British Academy などから受けて行われた。

BNC は 20 世紀後半,特に 1990 年代前半までのテキストを中心に収集されているが,プロジェクト完結後,データの増殖は行われていない。その代わり,BNC をもとにした2つの簡約版コーパス,BNC Sampler (書き言葉と話し言葉のコーパスを 100 万語ずつ収納)とBNC Baby (4つのジャンルをバランスよく 100 万語収納),を世に送り出している。

3.2.2 BNC の設計基準

BNC は全体が 9000 万語の書き言葉コーパスと 1000 万語の話し言葉コーパスに分れている。コーパスの構築に際して重要な設計基準 (design criteria) に関しては、検討の結果、以下のような基準を設けた:

① 選択基準 (selection criteria):

書き言葉(written text) 9000 万語

- -領域 (domain):
- a) 75%は情報提供型のテキスト (informative text) とし,以下の分野からほぼ均等 の分量を取るようにする:
 - applied sciences, arts, belief & thought, commerce & finance, leisure, natural & pure science, social science, world affairs.
- b) 残りの 25%は創作散文 (imaginative text) とし, literary and creative works などから採取する。

-媒体 (Medium)

- a) 60%の書き言葉テキストは書籍(books)とする
- b) 25%は定期刊行物(新聞・雑誌など)とする
- c) 5-10%をその他の印刷物 (パンフレット, 広告チラシ, など) とする
- d) 5-10%を未刊行物(個人の手紙, 日記, メモなど)とする
- e) 5%以下を政治演説, 演劇の台本, 放送台本などの「スピーチ原稿 (material written to be spoken)」とする

一時間 (Time)

原則として 1975 年以降のテキストを使用。ただし文学作品に関しては 1964 年からを 使用。

話し言葉 (spoken text) 1000 万語

- a) 大別して人口統計に準拠して標本抽出された自然発話データ (demographic part) と会議や催しなど特定の文脈に依存する発話データ (context-governed part) に分かれる
- b) すべての発話の音声データは National Sound Archives of the British Library に 保管される
- c) Demographic Spoken:
 - 124 名のボランティア (British Market Research Bureau の斡旋)
 - 4 つの社会階層 (AB, C1, C2, DE)
 - 英国の38の異なる地域に分布
 - 年齢、性別も同人数が分布するように調整
 - 被験者は2-3 日間のすべての会話を記録される
 - 会話に参加したすべての関連する人々からもコーパス収録の許諾を得る
 - 会話に参加した人たちの性別, 年齢, 方言, 職業なども記録

d) Context-governed:

- 次の4種類の社会的コンテクストからの会話を収録:
 - i) 教育または情報提供を目的とした催し(講義,ニュース放送,授業でのディスカッション,など)
 - ii) 職場での催し(セールス・デモ, 労働組合の会議, 面接, など)
 - iii) 団体または公共のイベント(説教,政治演説,議会会議,など)
 - iv) 余暇のイベント (スポーツ講評, 晩餐会スピーチ, クラブ会合など)
- すべての話者情報(性別、年齢など)がわかる範囲内で記述する

② 分類基準 (classification criteria)

分類基準も非常に細かく規定されている。これは採取するコーパスの元データの種類や特 徴に応じて柔軟に定義される。主に以下のような要素を含んでいる:

- 標本サイズ (テキスト総語数), 採取箇所 (初め, 終わりなど)
- トピック、主題
- 著者の名前,年齢,性別,出生地,居住地
- 対象となる読者の性別・年齢層
- 難易度(主観的な基準による)

3.2.3 BNC のテキスト構成

以下に BNC の下位分類のテキスト統計量を示す。

表 3-2-1 British National Corpus (World edition) の内訳とテキストの統計量

| Text type | Texts | Kbytes | W-units | S-units | percent |
|-------------------------------|-------|----------|---------|---------|---------|
| Spoken demographic | 153 | 4206058 | 4.30 | 610563 | 10.08 |
| Spoken context-governed | 757 | 6135671 | 6.28 | 428558 | 7.07 |
| All Spoken | 910 | 10341729 | 10.58 | 1039121 | 17.78 |
| Written books and periodicals | 2688 | 78580018 | 80.49 | 4403803 | 72.75 |
| Written-to-be-spoken | 35 | 1324480 | 1.35 | 120153 | 1.98 |
| Written miscellaneous | 421 | 7373707 | 7.55 | 490016 | 8.09 |
| All Written | 3144 | 87278205 | 89.39 | 5013972 | 82.82 |

BNC World Edition では、コーパスの総量が1億語をわずかに下回る。これはEU圏外にテキストの持ち出しを拒否した一部の著作権者がいるためである。

3.2.4 BNC の特徴

BNC の特徴は多岐にわたるが、本研究用のデータとしては以下のような点が強調される。

- (1) BNC は現代英語コーパスの中では、厳密な設計でテキスト間の標本抽出などが行われた、均衡コーパスとしては最も信頼のおけるものである。これによって、そこから得られる頻度・分布などの言語統計が、基準として現在入手できる最も信頼性の高いデータである、と言うことができる。
- (2) BNC はテキスト属性 (テキストに関連するさまざまな書誌情報や話者情報) が綿密に付されているため、それらを利用したテキスト間比較が柔軟に行えるという利点がある。
- (3) BNC は書き言葉、話し言葉に別れており、特に話し言葉の4割は自然発話であるので、それぞれのサブコーパスを中心とした語彙表の作成にも適している。
- (4) BNC にはテキスト内の構造化タグ (structural mark-up) や品詞タグ (POS tag) などが付与されているので、言語注釈を活用したさまざまな検索可能性が考えられる。

逆に BNC の弱点といえることを挙げれば以下のような点が指摘できる。

- (1) BNC は 1990 年代初めまでのテキストしか含まれておらず, 徐々に古さを感じさせるようになってきたが, データの更新は計画されていない。
- (2) 書き言葉 9000 万語, 話し言葉 1000 万語というサブコーパス構成も, 今日では話し言葉への比重をもっと多くすべきという要望が多い。
- (3) BNC のサイズは1億語であるが、辞書作成などには1億語でも不十分ということが言われてきており、現在の大規模コーパスは 10 億語クラスを基準としてきている。そういった意味では決して大きなコーパスではなくなりつつある。
- (4) BNC はイギリス英語のみを含んでいるので、アメリカ英語への志向の強い 日本の英語教育の資料としては偏りがある、と見なされる。

3.2.5 BNC の活用例

BNC は現在,世界で最も言語研究・教育に広く用いられているコーパスの1つである。いくつかの活用分野を簡潔に述べる。

3.2.5.1 辞書

BNC を公に使ったと言っている英語学習英英辞典は、コンソーシアム・メンバーでもあったロングマンの Longman Dictionary of Contemporary English (1995年の第3版からコーパス準拠)、およびオックスフォード大学出版局の Oxford Advanced Learner's Dictionary (これも1995年の改訂時にコーパス準拠と称した)の2冊である。

日本では BNC のライセンスが解禁になった 2001 年から, 小学館が BNC の検索サービスである『小学館コーパスネットワーク』を立ち上げ, 一般の英語教員などへの簡易インタフェースを提供すると同時に, 『ユース英和辞典』(小学館), 『コーパス英語類語使い分け 200』(小学館) などで, BNC からの頻度情報を活用した辞書作りを行っている。

3.2.5.2 文法書・語法書

文法書として BNC を部分的に使い自社コーパスを元に作成したものとして, Longman Grammar of Spoken and Written English (1999) が挙げられる。 Cambridge International Corpus を利用した複数の文法・語法書が出版されているが、その一部分がBNC である、という可能性が高いが、明確にはしていない。

3.2.5.3 語彙表

BNC の語彙表は 2000 年以前の World Edition 公開前から, Adam Kilgarriff が自身のホームページで公開していた (http://www.kilgarriff.co.uk/bnc-readme.html)。Kilgarriff の語彙表には, (1) Spoken/Written ごとの語彙表, (2) 基本形頻度リスト, (3) 5000 語毎テキスト内の単語頻度の分散データ,といった大別して 3 種類の資料が入手できた。こ

れらを用いて、アルクの Standard Vocabulary List 12000 が作成されたし、JACET8000 でも独自リストの他に、Kilgarriff のリストなども参考にしている。

World Edition 公開後(2001年)には、Lancaster 大学の Geoffrey Leech, Paul Rayson, Andrew Wilson が BNC を用いた詳細な語彙リストを出版した(Leech, G. P. Rayson, and A. Wilson (2001) Word Frequencies in Written and Spoken English: based on the British National Corpus. London: Longman)。彼らの web page には、全語彙リストのファイルがダウンロード可能になっている(http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html)。

3.2.5.4 その他の英語教材

BNC を用いた英語教材はまだそれほど多くはないが、日本では投野が NHK テレビ英会話『100 語でスタート! 英会話』(2003-2005 年度放映) において、BNC からの頻度データを用いた語彙シラバスの会話教材を提案している。なお、これらのデータに基づいた会話本(投野由紀夫(2003)『コーパス練習帳』その他)が日本(翻訳が台湾、韓国)で刊行されている。

(明海大学・投野由紀夫)

3.3 American National Corpus

American National Corpus (以下 ANC) は、大規模な現代米語の均衡コーパスを目指して構築が進められている。コーパス研究では以前から英米両方のコーパス作成が行われてきたが、イギリス英語の大規模コーパス BNC が完成し、コーパスの規模が 1 億語に及ぶに至って、同規模の比較可能な米語コーパスの必要性が真剣に議論されるようになってきたためである。

ANC は 1999年に BNC と同様のコンソーシアムが立ち上げられ, Catherine Macleod (当時ニューヨーク大学) が Project Manager を勤めたが,後に Randi Reppen と Nancy Ide に引き継がれている。コンソーシアムには Pearson Education の他,英国系の辞典出版社,および日本の辞典出版社が参画している。

ANC は現在のところ Second release が 2200 万語の規模で公開されている (本報告書の分析には時間の関係で、1st release のデータのみを使用)。当初は5年ほどで完成の予定であったが、出版社からのテキスト拠出や著作権許諾に手間取っており、未だ完成の目処が立っていない。データそのものは Linguistic Data Consortium (LDC) を通じて使用許可を得れば、研究用には安価で利用できる。

ANC 2^{nd} release の内容を簡単に説明する。BNC と同様、話し言葉と書き言葉の 2 つのセクションに別れている。現在のところ話し言葉が 400 万語弱、書き言葉が 1800 万語強集まっている。

| 表 3-3-1 American | National | Corpus | の内訳 |
|------------------|----------|--------|-----|
|------------------|----------|--------|-----|

| Spoken Totals | 2,474 files | 3,863,592 words |
|----------------|--------------|------------------|
| Written totals | 10,821 files | 18,530,112 words |
| Total | 13,295 files | 22,393,704 words |

ただし、現状では内容的にはかなりの偏りがある。話し言葉は7割以上が LDC が保有している Switchboard Corpus (Texas Instruments 社が実験で行った電話会話の記録で、まったくお互いに知らない実験参加者を機械的に結び付け、一定のトピックについて電話で会話させたもの)から成っており、非常に限定的なコンテクストでの話し言葉である。また書き言葉も New York Times, Slate Magazine, Biomed などの一部のコンテンツの分量が数百万語を超えるなど、テキスト・タイプのバランスが取れているとは到底言えない。

よって、現状ではこの ANC を BNC と比較する対象として独立して扱うのは極めて不適当であると思われる。少なくとも、ANC のデータでは足りない米語の他の分野のテキストをある程度集めて、コーパスの総量を BNC にできるだけ近づけること、またテキストタイプもなるべく BNC に近い分野のものを意識して採用するなどの配慮が必要である。このために我々の今回のプロジェクトでは ANC をコアとした「現代米語コーパス」として、ANC 以外のテキストも一定量集め合成する手法をとった。

(明海大学・投野由紀夫)

- 3.4 Santa Barbara Corpus of Spoken American English
- 3.4.1 Santa Barbara Corpus の概要

Santa Barbara Corpus (以下 SBC) は、現代口語米語のコーパスであり、ICE(International Corpus of English)の一部として Linguistics Department of the University of California, Santa Barbara で収集が続けられているものである。Part 1 は 2000 年にリリースされ、その後、Part 2 (2003), Part 3(2004), Part 4(2005)と続けてリリースされてきている。Director は、John W. Du Bois で、Wallace L. Chafe、Sandra A. Thompson、Charles Meyer、Robert Englebretson らがプロジェクトに参加している。

3.4.2 SBC の設計基準

SBC のデータの収集に当たっては、出身地域、年齢、職業、性別、人種的背景等を考慮してコーパスにバランスを持たせ、Face-to-face の会話、電話での会話、ゲームをしながらの会話、職場での会話、授業の会話、教会での説教等の多用な場面での自発的(spontaneous)会話を全米から集録している。各パートにはそれぞれ 15-30 分の会話が約

15 件収録されており、Part 1 から Part 4 までの合計で約 249,000 語のサイズとなっている。また、スクリプトのテキスト以外にも、録音された実際の会話音声を聞くことができる(個人情報は特定できないように処理されている)。収録されている会話のトピック一覧は以下の通りである。

Part 1

SBC001 Actual Blacksmithing

SBC002 Lambada

SBC003 Conceptual Pesticides

SBC004 Raging Bureaucracy

SBC005 A Book about Death

SBC006 Cuz

SBC007 A Tree's Life

SBC008 Tell the Jury That

SBC009 Zero Equals Zero

SBC010 Letter of Concern

SBC011 This Retirement Bit

SBC012 American Democracy is Dying

SBC013 Appease the Monster

SBC014 Bank Products

Part 2

SBC015 Deadly Diseases

SBC016 Tape Deck

SBC017 Wonderful Abstract Notions

SBC018 Vet Morning

SBC019 Doesn't Work in This Household

SBC020 God's Love

SBC021 Fear

SBC022 Runway Heading

SBC023 Howard's End

SBC024 Risk

SBC025 The Egg Which Luther Hatched

SBC026 Hundred Million Dollars

SBC027 Atoms Hanging Out

SBC028 Hey Cutie-Pie

SBC029 Ancient Furnace

SBC030 Vision

Part 3

SBC031 Tastes Very Special

SBC032 Handshakes All Around

SBC033 Guilt

SBC034 What Time is it Now?

SBC035 Hold My Breath

SBC036 Judgmental on People

SBC037 Very Good Tamales

SBC038 Good Strong Dam

SBC039 Pretty Busy Bird

SBC040 Beaten on a Regular Basis

SBC041 X Units of Insulin

SBC042 Stay Out of It

SBC043 Try a Couple Spoonfuls

SBC044 He Knows

SBC045 The Classic Hooker

SBC046 Flumpity-Bump Down the Hill

Part 4

SBC047 On the Lot

SBC048 Mickey Mouse Watch

SBC049 Noise Pollution

SBC050 Just Wanna Hang

SBC051 New Yorkers Anonymous

SBC052 Oh You Need a Breadbox

SBC053 I Will Appeal

SBC054 'That's Good', Said Tiger

SBC055 The Mama of Dada

SBC056 What is a Brand Inspection?

SBC057 Throw Me

SBC058 Swingin' Kid

SBC059 You Baked

SBC060 Shaggy Dog Story

口語英語コーパスであるので音声情報を含む情報付与形式が採用されており、フォーマットとしては、LDC Callhome format を使用する".trn"フォーマットと、Treebank の CLAN を使用する CHAT フォーマットの".ca"フォーマットの 2 種類で表記されている。

.trn format の例

2.660 2.805 JOANNE: But,

2.805 4.685 so these slides be real interesting.

6.140 6.325 KEN: ... Yeah.

6.325 7.710 I think it'll be real interesting

.ca format の例

JOANNE: But, %snd: "sbc015. wav"_2660_2805

so these slides be real interesting. %snd: "sbc015. wav"_2805_4685

KEN: ... Yeah. %snd: "sbc015. wav"_6140_6325

I think it'll be real interesting%snd:"sbc015.wav"_6325_7710

3.4.3 SBC の課題

SBC は、現代口語米語のコーパスとして貴重なものであると言えるが、音声英語コーパスということもあり、コーパスサイズがまだ約25万語と比較的少ないという問題がある。話し言葉英語のコーパスの整備は書き言葉に比べるとまだ遅れており、コーパスの規模も小さい。現代口語米語コーパスとして貴重なものであるので、今後も規模の拡大などの整備が期待される。

データは Linguistic Data Consortium (LDC) を通じて使用許可を得れば、研究用で利用できる。

(立命館大学・杉森直樹)

3.5 MICASE

3.5.1 MICASE の概要

MICASE は、The Michigan Corpus of Academic Spoken English の頭文字をとったものである。1997年ミシガン大学英語研究所は、学術的口頭発話の研究プロジェクトを開始した。大学の学術的口頭発話 200 時間分 (約 170 万語)を録音・転写する計画である。2001年に 190 時間以上の録音を終了し、2002年に転写と校正を終了した。全コーパスはwww.hti.umich.edu/m/micaseから入手可能である。

3.5.1.1 プロジェクトの目的

このプロジェクトは以下の4つを目的としている。

- 1.これまでにこのような学術的発話をデータベースにしたコーパスはない。
- 2.頻繁に用いられる文法・語句パターンを学術口語コーパスで調べれば、書きことばの コーパスに基づく現在の文法書・語法書との違いが見出せるだろう。
- 3.大学文化での経験が増すにつれ、発話パターンに一般的な変化が見出せるかもしれない。

4.学術口語コーパスに基づき、より適切な EAP プログラムや教材を開発できる。

3.5.2 学術発話の定義

MICASE の学術発話は19分から178分,2805語から30,328語に及ぶ。学術発話の定義は、学術的状況でなされた発話である。ジョークや譲歩、個人的な逸話、説明、知的な正当化などを含む発話行為である。「学術的議論」に限定されるものではない。

3.5.3 コーパスの構成

MICASEは、152の発話事例から成る。その内訳は次の通りである。

| 大小さまざまな講義 | 62 事例 |
|-----------------|-------|
| 学際的学会あるいは学部内の学会 | 13 事例 |
| 学生の研究発表 | 11 事例 |
| ディスカッション | 9 事例 |
| セミナー | 8 事例 |
| 学部生の実験室での授業 | 8 事例 |
| オフィスアワー | 8 事例 |
| 研究グループ | 8 事例 |
| 実験室などでの会合 | 8 事例 |
| 指導助言 | 5 事例 |
| 博論の口頭試験 | 4 事例 |
| チュートリアル | 3事例 |
| 面接 | 3事例 |
| キャンパス・博物館ツアー | 2 事例 |
| サービスに関する会話 | 2 事例 |

コーパスは、医学、歯学、経営、法律を除いた、生物健康科学、物理・工学、社会科学・教育、人文学部での学術発話をバランスよく取り入れている。話者はひとりの場合と複数の場合とがある。学部生、大学院生、教員、英語母語話者、非英語母語話者の発話が盛り込まれている。男女比はほぼ等分になっている。年齢は10代から50代以上を広く取り込んでいる。

3.5.4 データ収集

すべての学術発話はデジタルオーディオ・テープに録音し、2 名の研究者がその場で発話者の特定したり、フィールドノートをとったりした。しかし、小グループでの学術発話(指導、オフィスアワー、研究グループなど)では部外者の存在は障害になるので、録音機器のみ設定した。性別、年齢、大学での地位、母語などの情報は最後に質問紙で集められた。

表 3-5-1 は, 発話者の分類ごとの発話者数と語数を示している。男女比では, 46%対 54% で若干女性のほうが多い。 教員と学生では, 教員が 49%, 学生が 44%である。学部生と大学院生はちょうど半々である。言語地位は, 英語母語話者が 88%, 非英語母語話者が 12% である。

表 3-5-1 発話者の分類ごとの発話者数と語数

| 発話者の範疇 | | Total Speakers | Total Words | % of Total |
|--------|---------|----------------|-------------|------------|
| | | | | Corpus |
| 性 | 男性 | 729 | 786,487 | 46% |
| | 女性 | 842 | 909,053 | 54% |
| 学術的役割* | 教員 | 160 | 825, 829 | 49% |
| | [男性 | 84 | 446,925 | 26%] |
| | [女性 | 76 | 378,904 | 22%] |
| | 学生 | 1,039 | 742,348 | 44% |
| | 学部生 | 782 | 368,433 | 22% |
| | [男性 | 336 | 142,102 | 8%] |
| | [女性 | 446 | 226,331 | 13%] |
| | 大学院生 | 257 | 373,915 | 22% |
| | [男性 | 121 | 158,696 | 9%] |
| | [女性 | 136 | 215,219 | 13%] |
| 言語地位 | 英語母語話者 | 1,449 | 1,493,586 | 88% |
| | 非英語母語話者 | 122 | 201,954 | 12% |
| 総計 | | 1,571 | 1,695,540 | |

^{*}スタッフ, 訪問者などはこの表に含まれないために, 100%にならない。

表 3-5-2 は、学問領域ごとの発話者数と語数を示している。人文、社会科学、生物・健康科学、物理・工学の4つの領域の発話事例数はほぼ同じである。人文と生物・健康科学においては、教員の発話比率が学生を上回り、逆に、社会科学と物理・工学では学生の発話比率が教員を上回る傾向にある。

表 3-5-2 学問領域ごとの発話者数と語数

| | Speech | Speakers | Words | % of | % | % | % | % |
|----|--------|----------|-----------|--------|------|--------|---------|----------|
| | Events | | | Total | Male | Female | Faculty | Students |
| | | | | Corpus | | | | |
| 人文 | 36 | 349 | 434,669 | 26 | 56 | 44 | 63 | 29 |
| 社会 | 35 | 452 | 420,347 | 25 | 37 | 63 | 44 | 55 |
| 科学 | | | | | | | | |
| 生物 | 32 | 257 | 325,456 | 19 | 41 | 59 | 55 | 42 |
| 健康 | | | | | | | | |
| 物理 | 36 | 314 | 358,776 | 21 | 55 | 45 | 44 | 52 |
| 工学 | | | | | | | | |
| その | 13 | 199 | 156,292 | 9 | 37 | 63 | 20 | 41 |
| 他 | | | | | | | | |
| 総計 | 152 | 1,571 | 1,695,540 | | | | | |

表 3-5-3 は、独話、パネル、双方向的、混在という 4 つの談話形式ごとの発話者数と語数を示している。独話は 84%と圧倒的に教員によるものが多い。逆に、パネル形式では 76% と学生による発話が多い。

表 3-5-3 談話形式ごとの発話者数と語数

| | Speech | Speakers | Words | % of | % | % | % | % |
|-----|--------|----------|-----------|--------|------|--------|---------|----------|
| | Events | | | Total | Male | Female | Faculty | Students |
| | | | | Corpus | | | | |
| 独話 | 61 | 472 | 554,335 | 33 | 50 | 50 | 84 | 14 |
| パネル | 9 | 133 | 141,505 | 8 | 27 | 73 | 16 | 76 |
| 双方 | 57 | 643 | 715,333 | 42 | 46 | 54 | 26 | 63 |
| 向的 | | | | | | | | |
| 混在 | 25 | 323 | 284,367 | 17 | 51 | 49 | 54 | 39 |
| 総計 | 152 | 1,571 | 1,695,540 | | | | | |

3.5.5 MICASE の特徴

MICASE は学術的場面における口語コーパスであるために、口語的な語彙と学術的な語彙に特徴づけられる。口語的な語彙としては、mhm, oops, gee のような相づち、filler と呼ばれるつなぎ言葉、感嘆表現などが文字化されて、1 語として扱われている。また、書きことばでは、分かち書きされるが、口語では連続して1 語にように聞こえるため、gonna, gotta, hafta, lemme のように1 語として記述されたものが多く含まれる。または、cuz

のように発音に合わせて短縮して記述した語である。さらに、長音節語の第 1 音節あるいは最初の 2 音節のみに短縮した語である。例えば、intro は introduction、 grad は graduate、 stat は statistics を短縮した語である。

学術的な語彙としては, 次のような語が挙げられる。

academic, agent, basically, curve, decrease, engineering, error, physics, presentation, sample, solid, transfer, analysis, category, chemistry, definition, reference, review

このように MICASE コーパスに含まれる語彙は、学術的で口語的なものということができる.

(麗澤大学・望月正道)

4. 新しいコーパス資料の整備と NEXT 1000

4.1. 米語コーパス

4.1.1 研究の目的

JACET 8000(2003)は、BNC をベースに、新聞、CNNトランスクリプト、児童文学、アカデミックテキスト、日本の教科書コーパス、映画のスクリプト等のコーパスを加え、日本人の英語学習者に適するように合成して作られた語彙リストである。多くの研究者の手により、科学的な手法を用いて作られたこのリストの信頼性は高いと思われるが、いくつかの弱点も指摘されている。その中には、以下のようなものがある。

- ・BNC をベースにしていることから、米語データの比率が小さい。
- ・いわゆる「生活語」で、リスト内に含まれないものがある。

このような指摘を踏まえ、米語のコーパスを作成し、JACET 8000 と米語語彙リストを比較し、 その差異を検証することが本研究の目的である。

4.1.2 米語コーパス

今回の研究のために収集されたコーパスは,以下の通りである。

- (1) 新聞コーパス(USA Today Web 版; 2000 年 9 月~2006 年 7 月) 110,944,658 語
- (2) CNN トランスクリプト(Interview and Debate; 2000年1月~2006年6月) 25,150,021 語
- (3) American National Corpus (ANC 1st release; 2003) 12,469,119 語
- (4) Santa Barbara Corpus of Spoken American English (2000) 222,973 語
- (5) Michigan Corpus of Academic Spoken English (MICASE; 2003) 1,794,311 語
- (6) 子供向け米語コーパス1 KidsPost(Web 版;2000年1月~2006年5月) 1,291,461語
- (7) 子供向け米語コーパス2 Scholastics(Web 版; 2001年4月~2006年6月) 2,224,198語
- (8) 子供向け米語コーパス3 Nick.com(Web 版; 2001 年8月~2006 年5月) 1,896,654 語
- (9) 子供向け米語コーパス4 Time For Kids(Web 版; 2005 年9月~2006 年 4 月) 208,053 語
- (10) 小説(37冊;1949~2005) 4,567,257語
- ※米語コーパスサイズ合計 160,768,705 語

それぞれのコーパスの分析には、「v8an.pl」という perl スクリプトを用いた。また、後述する各リストの合成や log-likelihood の計算等にも perl スクリプトを用いた。また、ANC、Santa Barbara、MICASE は、独自のタグ付きコーパスだが、今回はタグを取り去って平テキストレベルで集計している。各コーパスの特徴は、表 4-1-1 のようになる。

表 4-1-1 米語コーパスの特徴

| コーパス | written/spoken | 大人向け/子供向 | 内容 |
|---------------|----------------|----------|-----------------|
| | | け | |
| USA Today | W | 大 | ニュース |
| CNN | S | 大 | ディベート・インタビュー |
| ANC | W/S | 大 | 整理されたコーパス |
| Santa Barbara | S | 大 | 整理されたコーパス |
| MICASE | S | 大 | 整理されたコーパス |
| KidsPost | W | 子 | 子供向けニュース |
| Scholastics | W | 子 | 子供向けニュース |
| Nick.com | W | 子 | 趣味・ジョーク等 |
| TFK | W | 子 | 子供向けニュース |
| 小説 | W | 大(一部子) | フィクション・ノンフィクション |

また,素データの JACET 8000 のカバー率は表 4-1-2 の通りであった。総トークン数におけるカバー率は約 86.5%であるが、縮約形、非単語、固有名詞を抜いて計算すると、そのカバー率は約 96.4%であった。

表 4-1-2 米語コーパスにおける JACET 8000 のカバー率

| | level 1 | level 2 | level 3 | level 4 | level 5 | level 6 | level 7 | level 8 |
|-----|-----------|----------|---------|---------|---------|---------|---------|---------|
| idx | 1136 | 1000 | 1000 | 1000 | 1000 | 1000 | 999 | 1000 |
| % | 0.159 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 |
| tkn | 112614570 | 10635495 | 4449862 | 5330128 | 2255775 | 1709302 | 1145318 | 884228 |
| % | 70.048 | 6.615 | 2.768 | 3.315 | 1.403 | 1.063 | 0.712 | 0.55 |

| over 9 | cont. | non | proper | total |
|---------|---------|---------|----------|-----------|
| 222012 | 5150 | 127061 | 351619 | 713977 |
| 31.095 | 0.721 | 17.796 | 49.248 | 100 |
| 5128556 | 2121308 | 3575277 | 10918886 | 160768705 |
| 3.19 | 1.319 | 2.224 | 6.792 | 100 |

※idx ... 見出し語数 tkn ... トークン数 cont ... 縮約形 non ... 非単語 proper ... 固有名詞

また、各コーパスの特徴を把握するため、以下の方法で log-likelihood(LL)を計算し、各コー

パスに特徴的に現れる語を特定した。

- (1) ターゲットリストを1つ決める。
- (2) 他の9個のリストの合計を求め、それをベースリストとする。
- (3) ターゲットリストとベースリストの LL を求める。
- (4) 上記(1)~(3)の作業を全てのリストに対して行う。

この結果, 今回収集された米語コーパス内での各コーパスの語彙的特徴を, LL の面から特定できる。以下は, 各コーパスの LL のうち, 1000 を超える値を示した語である。 ただし, 非単語や固有名詞等は省いてある。

(1) USA Today LL(897 語中, 上位 100 語)

say, game, season, million, its, official, the, company, in, team, coach, airline, year, last, after, Wednesday, player, for, Tuesday, win, Monday, Thursday, by, sale, state, billion, Friday, tech, spokesman, score, price, federal, flight, average, plan, Saturday, expected, consumer, from, analyst, play, including, lead, airport, quarter, Sunday, month, customer, industry, cost, university, rate, market, group, since, worker, second, during, city, according, report, league, share, playoff, manager, final, carrier, software, yard, third, nearly, passenger, three, loss, record, an, internet, stock, at, agency, his, hit, series, director, four, service, fund, than, county, offer, contract, revenue, injury, January, return, shot, include, technology, cent, September

(2) CNN LL(278 語中, 上位 100 語)

you, I, we, do, think, that, know, be, go, what, it, this, well, very, there, not, yes, so, they, right, mean, me, get, us, thing, really, people, my, thank, ph, all, now, lot, have, hello, just, your, here, about, want, happen, our, let's, back, and, why, how, because, hi, talk, OK, kind, absolutely, these, question, see, fact, like, he, wolf, actually, senator, talking, oh, tomorrow, tonight, look, come, him, Mr., something, sort, story, thanks, hear, if, somebody, certainly, obviously, president, but, welcome, course, wonderful, everybody, no, great, way, love, sure, believe, let, congressman, whole, Dr., then, thought, can, join, them

(3) ANC LL(106 語)

uh, yes, uh-huh, um-hum, um, I, oh, know, you, that, percent, it, like, well, guess, so, okay, really, slate, paper, they, just, kind, ha, story, do, client, article, or, right, my, twenty, but, time, news, and, hm, stuff, too, seem, nice, mean, thing, pretty, of, front-page, dear, bye-bye, art, not, real, thirty, eighty, pundit, century, one, hum, recycle, page, probably, piece, then, something, anyway, fifty, true, editorial, by, impeachment, little, think, old, b.c, journal, lot, editor, neat, column, km, them, spin, book, interesting, even, enjoy, seventy, globe, painting, read, because, huh-uh, sixty, actually, forty, museum, wow, modern, gosh, much, if, dole, newspaper, sort, click,

magazine, critic

(4) Santa Barbara LL(23 語)

I, you, yes, okay, gonna, oh, uh, em, know, like, do, um, it, so, well, mm, just, wanna, what, alright, not, that, cause

(5) MICASE LL(95 語)

um, okay, uh, you, so, mhm, I, gonna, yes, like, it, that, cuz, wanna, this, what, do, know, alright, oh, right, mean, just, here, then, we, be, if, there, uhuh, not, these, sort, thing, kinda, one, actually, can, think, really, kind, something, different, they, well, how, nineteen, look, see, or, basically, let's, probability, equal, class, minus, mm, maybe, equation, get, your, hm, species, stuff, gotta, example, very, receptor, node, lemme, about, why, but, because, way, zero, protein, talk, electron, hafta, cell, lotta, ha-, lecture, sorry, serotonin, guess, same, idea, fifty, kay, terms, photon, where, variable

(6) KidsPost LL(28 語)

kid, yesterday, school, book, quiz, Washington, birthday, animal, dog, today, your, reader, zoo, favorite, name, news, elementary, fun, girl, soccer, dinosaur, toy, sport, play, teacher, panda, scientist, pet

(7) Scholastics LL(38 語)

kid, age, your, read, learn, book, school, child, star, movie, space, earth, musher, fun, favorite, fave, student, animal, moon, reading, dog, soccer, birthplace, birthdate, statistics, mar, grade, California, planet, world, my, cool, zoo, skate, scientist, girl, character, calendar

(8) Nick.com LL(48 語)

your, today, you, birthday, day, kid, happy, chocolate, bear, if, star, creature, glue, cool, friend, dog, favorite, school, fun, sport, animal, mom, yourself, will, candy, cat, celebrate, dinosaur, homework, zoo, cookie, eat, girl, fave, wear, pet, oz, ice, butter, or, rock, spear, cream, good, get, color, can, like

(9) TFK LL(1 語)

kid

(10) novels LL (225 語中, 上位 100 語)

her, she, he, I, me, him, his, eye, my, hand, stare, you, then, smile, could, nod, thought, man, door, whisper, not, voice, look, them, see, dark, lord, herself, shake, head, room, lip, again, glance, arm, mouth, beside, ask, rose, face, stone, light, it, all, sigh, wagon, laugh, damn, hair, finger, turn, back, sit, down, hell, tell, breath, nothing, moment, hood, grin, away, vampire, seem, would, gaze, do, bed, slowly, feel, himself, god, beneath, stand, into, hear, frown, toward, cry, shoulder, walk, mother, old, softly, too, come, bean, around, as, mutter, suddenly, table, sword, mind, lark, if, no, let, sound,

sister

米語リストの合成

10種類の米語コーパスを合成するにあたり、単純な合計をとることは、1億語を超える新聞コーパスの影響が大きくなることと、いわゆる「生活語」は、子供向け米語のコーパスに多く含まれることから、以下のような操作を行い、10種類のコーパスを合成した。

- (1) 100 万語を超えるコーパスは、各語の頻度を 100 万語換算にして計算する。
- (2) 100 万語に満たないコーパスは、その頻度をそのまま用いる。
- (3) この結果から得られた頻度リストの上位に入っている固有名詞, 非単語, ハイフンで結ばれた複合語, レンジが低い(2以下)語, 一過性の高頻度だと推測される外来語(jihad 等)を除き, 8136番までの語を米語リストとした。8000語でないのは, JACET 8000の別表の語数を加えて, 同等の大きさにするためである。

上記(1)~(3)の操作により、巨大コーパスの影響を抑え、かつ子供向け米語コーパスの比率が高くなった。この操作リストと、単純な合計で算出した頻度リストとの差異は、8136 語中、873 語であったが、その差異にもそういった特徴が現れた。

この操作によって作成されたリストの JACET 8000 のカバー率は表 4-1-3 のようになった。総トークン数におけるカバー率は約 88.1%であるが、縮約形、非単語、固有名詞を抜いて計算すると、そのカバー率は約 96.5%であった。

表 4-1-3 操作後の米語コーパスにおける JACET 8000 のカバー率

| | level 1 | level 2 | level 3 | level 4 | level 5 | level 6 | level 7 | level 8 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|
| idx | 1136 | 1000 | 1000 | 1000 | 1000 | 1000 | 999 | 1000 |
| % | 0.159 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 |
| tkn | 6223230 | 495620 | 219154 | 191847 | 109852 | 91016 | 52988 | 46283 |
| % | 73.813 | 5.879 | 2.599 | 2.275 | 1.303 | 1.08 | 0.628 | 0.549 |

| over 9 | cont | non | proper | total |
|--------|--------|--------|--------|---------|
| 222012 | 5150 | 127061 | 351619 | 713977 |
| 31.095 | 0.721 | 17.796 | 49.248 | 100 |
| 272293 | 119574 | 137491 | 471679 | 8431026 |
| 3.23 | 1.418 | 1.631 | 5.595 | 100 |

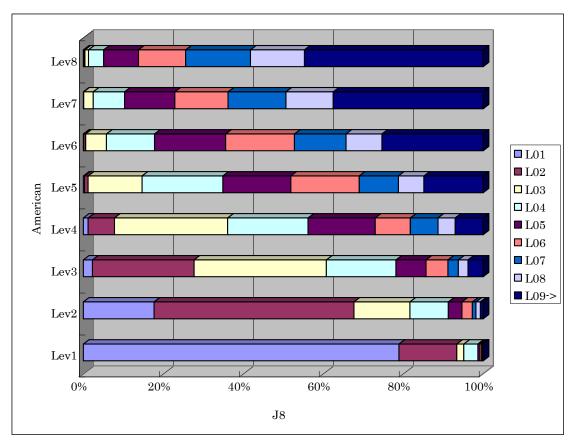


図 4-1-1 米語リストの Level 8 までと, JACET 8000 の Level 8 までのずれ

図 4-1-1 は、米語コーパスのレベル8までと、JACET 8000 のレベル8までのずれをグラフ化したものである。レベル3以降は、その重なり率が50%を下回り、レベル8では、JACET 8000 に含まれない語が半数近くになるなど、各レベル間の重なり度合いは低いと言える。

米語コーパスリストと JACET 8000 の誤差は 1340 語であった。この差異を品詞別に見ると、図 4-1-2 のようになる。ただし、英単語は多義語である可能性が高く、1語で複数の品詞を持つもの が多いため、実際にコーパス内で使用される場合の割合とは異なる。

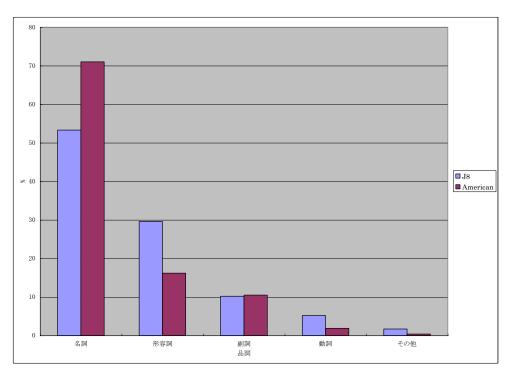


図 4-1-2 米語コーパスのみに含まれる語と JACET 8000 のみに含まれる語の品詞別割合

米語コーパス独自の語は、名詞の割合が多く、約7割を占める。JACET 8000 独自の語でも、 名詞の占める割合が一番大きいが、約5割ほどであり、その分、形容詞や動詞が多くなる。

この差異を意味的なカテゴリー別に見た場合,図 4-1-3 のようになる。また,表4-1-4 には,米語 コーパスリストに比較的多く現れる意味カテゴリー内の具体例が挙げられている。ただし,この意味的カテゴリーは,USA Today の新聞記事のカテゴリーを基に修正したものである。また,あるカテゴリーに多く含まれる語も、当然他のカテゴリーで用いられることがある。したがって、図中の割合は、実際のコーパス内での使用の割合とは異なる。

米語コーパス独自語には、生活に密着した場面で用いられるものが比較的多く、また、スポーツ、動物、芸能、食物に関する語の割合が高い。JACET 8000 独自語では、一般的に使われると考えられる語や、新聞等でよく見かけるような社会的な語、医学的な用語、学術論文等に用いられるような語の割合が高い。

また、JACET 8000 独自語には、書き言葉独特の形式張った語がやや目立ち、文語もわずかだが含まれている。

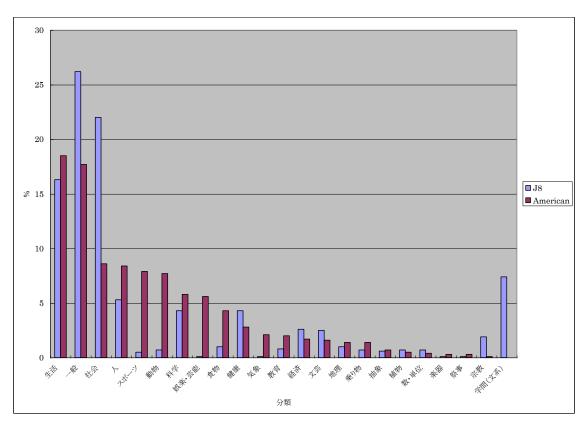


図 4-1-3 カテゴリー別に見る米語コーパス独自語と JACET 8000 独自語の割合

表 4-1-4 意味カテゴリー別に現れる語の例

| カテゴリー | 米語コーパスリスト独自語 | JACET 8000 独自語 |
|------------|----------------------------------|-----------------------------------|
| 生活の中で使用される | scary, trash, cooperate, closet, | gloomy, eyelid, apron, solemn, |
| と考えられる語 | sneaker, dodge, nap, bracelet, | plump, earnest, squeak, jug, |
| | pajama, itch etc. | stoop, cellar etc. |
| 一般的に使用され, | forth, hike, oversee, anytime, | honorary, deficiency, glittering, |
| 様々なレンジに現れる | longtime, rotate, trek, peek, | dubious, unwanted, simplicity, |
| と考えられる語 | skeptical, handout etc. | rightly, gradual, encompass, |
| | | broadly etc. |
| 新聞記事等によく見ら | capitol, patriot, lawmaker, | legitimacy, regression, decree, |
| れる社会的事象を表す | insurgent, recount, juror, | intruder, vicinity, hierarchy, |
| 語 | sniper, courtroom, artifact, | famine, electorate, |
| | obesity etc. | governmental, proliferation |
| | | etc. |
| 芸能,スポーツ以外で | astronaut, firefighter, | sovereign, theorist, chemist, |
| 「人」を表す語 | columnist, pundit, mentor, | occupant, trustee, nationalist, |
| | tycoon, moderator, camper, | archbishop, learner, chap, |

| | | _ |
|-------------|----------------------------------|----------------------------------|
| | paleontologist, magician etc. | elector etc. |
| スポーツに関連する語 | playoff, pitcher, skater, sled, | rugby, protector, lightweight, |
| | skateboard, surf, rebound, | striker, header, midfield, |
| | touchdown, volleyball, | sportsman |
| | gymnastics etc. | |
| 動物・鳥・魚・絶滅した | zoo, panda, alligator, buffalo, | watchdog, wasp, larva, gull, |
| 生き物等を表す語 | gorilla, ranch, mammoth, | herring, hedgehog, gall, |
| | squid, kitty, cheetah etc. | cuckoo, robin |
| 科学・工学的な分野で | email, spacecraft, download, | vector, thermal, plumbing, |
| 用いられる語 | website, clone, DNA, steroid, | modem, sodium, limestone, |
| | asteroid, blog, photon etc. | biotechnology, emulate, |
| | | interface, ethanol etc. |
| 娯楽・芸能関連の語 | premiere, superstar, audition, | acoustic, repertoire |
| | sequel, lineup, sitcom, gamer, | |
| | casino, soundtrack, rapper etc. | |
| 食物に関連する語 | cookie, pumpkin, burger, chili, | pastry, loaf, fig, barley, pear, |
| | lemonade, popcorn, doughnut, | tart, cider, whisky, |
| | pancake, beverage, | refreshment, gin |
| | mayonnaise etc. | |
| 健康・医療に関連する | allergy, psych, serotonin, | respiratory, leukemia, groin, |
| 語 | allergic, implant, polio, cramp, | hygiene, surgical, tablet, |
| | nurture, germ, pharmacy etc. | homosexuality, cardiac, |
| | | barren, traumatic etc. |

4.1.3 結果と考察

合成された米語リストの特徴は以下の通りであった。

- ・米語リストにのみ含まれ、JACET 8000 に含まれない語の全体を見渡すと、いわゆる「生活語」 が多く含まれる結果となった。子供向けテキストの比率を高くしたことが主な要因だと考えられる。 その分、学術的なテキストに含まれているような語彙は減少した。これは、学術的なテキストが ANCの一部にのみ含まれているためで、今回の米語コーパスの編成からこのような結果が生じたと考えられる。
- ・米語独特の語も増加しているが、全体としては、米語コーパスにはニュースが多く含まれるため、 時事英語が目につく。レベル1の部分にも、政治用語や経済用語が入ってきている。しかし、米語 コーパスリストと JACET 8000 との差異に注目すると、このような用語は JACET 8000 よりもやや 少なくなっている。このことは、他のコーパスと比較して米語コーパスが時事英語に偏っているわ けではないことを示している。

以上のことから、このリストは日常的に使用される米語の特徴を持っていると考えられるが、大学 またはそれ以上の専門性を必要とする学習者にとっては、より専門的な分野のテキストを収集し、 その特徴を反映させた方が良いかもしれない。

現在、いわゆる「生活語」をいかにして語彙リストに反映させるか、という議論がなされている。しかし、「生活語」という定義がはっきりしないことと、「生活語」重視で語彙リストを作成していったとき、「生活語」以外の語(時事英語、アカデミックな語等)がそのリストから削除されていくことになることから、どのようなバランスでコーパスを作成していくか、ということが今後の重要な課題であろう。

(安城学園高等学校•清水伸一)

参考文献

Leech, G., Rayson, P., and Wilson, A. (2001). Word Frequencies in Written and Spoken English: based on the British National Corpus. Longman, London.

Rayson, P. (2003). Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. Ph.D. thesis, Lancaster University.

清水伸一 (2005). 米国新聞(USA Today) に見るトピック別頻出単語. JACET 英語語彙研究会 第2回研究大会 ハンドアウト.

資料 1 USA Today トピック別トークン数

| TOPIC | education | entertainment | life | money | nation | news | opinion |
|-------|-----------|---------------|---------|----------|----------|----------|---------|
| FREQ | 501401 | 8092218 | 4490680 | 12949028 | 10145981 | 15596057 | 1897112 |
| % | 0.5 | 7.3 | 4 | 11.7 | 9.1 | 14.1 | 1.7 |

| religion | science | sports | tech | travel | weather | unclassified | total |
|----------|---------|----------|---------|---------|---------|--------------|-----------|
| 195816 | 3521308 | 25204760 | 8724644 | 4950055 | 1902888 | 12772710 | 110944658 |
| 0.2 | 3.2 | 22.7 | 7.9 | 4.5 | 1.7 | 11.5 | 100 |

資料 2 CNN 番組別トークン数

| PROGRAM | Larry King Live | Late Edition | On the Story | Reliable Sources | Total |
|---------|-----------------|--------------|--------------|------------------|----------|
| FREQ | 16157067 | 5617109 | 1714398 | 1661447 | 25150021 |
| % | 64.2 | 22.3 | 6.8 | 6.6 | 100 |

資料3 novels コーパスの内容

Alan Dean Foster - Humanx 03 - Voyage to the City of the Dead (1984)

Alan F. Troop - The Dragon DelaSangre (2002)

Ann Rice - The Witching Hour - Mayfair Witches Book 1 (1990)

The Vampire Chronicles Volume 1 - INTERVIEW with the VAMPIRE (1973)

Ben Bova - Orion in the Dying Time (1990)

Bill Clinton - My Life (2004)

Carl Sagan - Pale Blue Dot (1994)

Clive Cussler - Golden Buddha (2003)

Dan Brown - The Da Vinci Code (2003)

Dave Pelzer - A Child Called 'It' (1995)

Dean Koontz - The Book of Counted Sorrows (2003)

David Weber and Eric Flint - 1633 (2002)

Frank Herbert - God Emperor of Dune (1983)

Greg Bear - The Infinity Concerto (Songs of Earth and Power Vol. 1) (1984)

J.D. Salinger - The Catcher in the Rye (1951)

James Patterson - London Bridges (2004)

Jennifer Weiner - In Her Shoes (2002)

Jonathan Dunn - The Forgotten King (2004)

Julie Garwood - The Wedding (1996)

Katherine Garbera - The Tycoon's Surprise (2003)

Larry Niven and Jerry Pournelle - The Burning City (2000)

Leigh Michaels - Dating Games (1993)

Lora Leigh - Kiss of Heat (2004)

Lori Foster - Indulge Me (2002)

Madeline Hunter - The Saint (2003)

Michael Crichton - The Andromeda Strain (1969)

Nora Roberts - Origin in Death (writing as J. D. Robb) (2005)

Orson Scott Card - Shadow Puppets (2002)

The Crystal City (2003)

Raymond Chandler - The Little Sister (1949)

Robert Jordan - Crossroads of Twilight (Wheel of Time 10) (2003)

Stephen King - The Girl Who Loved Tom Gordon (1999)

The Dark Tower VI: Song of Susannah (2004)

The Collective (2000)

Tamora Pierce - Circle of Magic 01 - Sandry's Book (1997)

Circle of Magic 04 - Briar's Book (1999)

Tom Clancy - Op Center 7 - Divide And Conquer (2000)

資料4 子供向けコーパス URL

KidsPost-http://www.washingtonpost.com/wp-srv/kidspost/orbit/kidspost.html

Scholastics - http://teacher.scholastic.com/scholasticnews/index.asp

Nick.com - http://www.nick.com/

TFK - http://www.timeforkids.com/TFK/

4.2 口語コーパス

4.2.1 目的

JACET 8000 策定のために用いたコーパスデータの多くは、印刷された「書き言葉」テキストデータから収集されたもので、「話し言葉」テキストデータは、データ収集や電子化の作業に時間を要したため十分なコーパスデータ量を確保することができなかった。「書き言葉」重視の JACET 8000 は、「話し言葉」コーパスの上位出現頻度語をどの程度カバーできるか検証するために、海外の主要な大規模コーパスの口語コーパス部を抽出し、その頻度順語彙リストと JACET 8000 の語形との比較検討をおこなう。

4.2.2 使用コーパスとデータ処理

海外大規模コーパスから,英口語データとしてBNC (「話し言葉」パートのみ),ICE-GB (「話し言葉」パートのみ),米口語データとしてANC (「話し言葉」パートのみ),MICASE (全体) と Santa Barbara Corpus (全体)を検証用口語コーパスデータとした。(個々のコーパスについては,第3章「新しいコーパス資料の整備」を参照。)

表 4-2-1 は、この口語コーパスデータに対する JACET 8000 (Plus 250 含む) のカバー率 を Perl スクリプト v8an.pl を用いて調べたものである。

| 表 4-2 | 2-1 | 口語コー | パスにおける | 3 JACET | 800 | 00 のカバー | -率 |
|-------|-----|------|--------|---------|-----|---------|----|
| | | | | | | | |

| | level 1 | level 2 | level 3 | level 4 | level 5 | level 6 | level 7 | level 8 |
|-----|------------|---------|---------|---------|---------|---------|---------|---------|
| idx | 1135 | 1000 | 1000 | 1000 | 999 | 998 | 998 | 998 |
| % | 1.224 | 1.078 | 1.078 | 1.078 | 1.077 | 1.076 | 1.076 | 1.076 |
| tkn | 14,102,161 | 761,870 | 288,274 | 350,121 | 225,303 | 175,467 | 54,015 | 42,419 |
| % | 81.468 | 4.401 | 1.665 | 2.023 | 1.302 | 1.014 | 0.312 | 0.245 |

| over 9 | cont | non- | proper | total |
|---------|---------|---------|---------|------------|
| 50,248 | 344 | 2,410 | 31,630 | 92,760 |
| 54.17 | 0.371 | 2.598 | 34.099 | 100 |
| 539,991 | 333,822 | 116,783 | 319,783 | 17,310,009 |
| 3.12 | 1.928 | 0.675 | 1.847 | 100 |

※idx... 見出し語数 tkn...トークン数 cont... 縮約形 non...非単語 proper... 固有名詞

4.2.3 追加候補語

表 4.2.1 の over 9 (index 50,248, token 539,991) は, 総語数 (index 92,760, token

17,310,009) から JACET 8000 の語形, 縮約形, 非語, 固有名詞を除いたものである。Index ベースで, 出現頻度上位 8000 語を目処に over 9 の語形を抽出し, JACET 8000 の「語の定義」を参考にして口語コーパスによる追加候補語のふるい分けをおこなったところ, 出現頻度 32 回までのものが 8015 語となった。内訳は, JACET 8000 採録の語形 (6,847), JACET 8000 以外の語 (1,168) であった。(8000 語を超過した主な原因は, 同頻度語の存在であった。ちなみに, 出現頻度 32 回の語形は 62 語であった。)

追加候補語の中で、着目すべき主要なタイプを以下に列記する。(BrE は British English, AmE は American English を示すものとする。)

(1) 間投詞 (exclamation)

aha, bingo, blimey (BrE), cheerio (BrE), darn, er, gee, golly, goody, heck, jeez, lo, oops, ouch, ow, poo, sh, ta (BrE), tut, ugh, uhuh, whew, whoa, yo, yuck

(2) 口語調 (colloquialism)またはくだけた表現 (informal)

dunno, gonna, gotta, hafta, lemme, lotta, oughta, sorta, wanna; babe, bollock (BrE), cos, dude, fella, feller, grad, iffy, kitty (<kitten), loo (BrE), macho, meself, nana, nappy (BrE), nay, nil, nutty, pee, potty (BrE), pup, spooky (esp. AmE), summat (BrE), telly (BrE), thingy, tot, tummy,

tuppence (BrE =twopence), wellie (BrE), whatnot, whatsit, yea, yucky, yummy

(3) 省略語 (Clipping)

advert (BrE), comp, demo, ex, hype, intro, max, mega, micro, mini, psych, prep, rep, semi, temp, tho, veg (BrE: cf. veggie (BrE))

(4) 擬声・擬態語 (onomatopeia)

beep, boo, ding, whack, whoop, zip

(5) 世俗語 (vulgar words, slang)

cunt, diddle, fart, pussy, wanker (BrE)

(6)外来語 (borrowing)

中期英語期までの借用語と、フランス語・ラテン語・ギリシャ語に由来するものを除いた外来語起源の語形。

buffalo, cafeteria, casino, chili, fiesta, lasagna, marijuana, mosque, posh, rabbi, ranch, shah, spaghetti, taco, totem, tornado, trek, vodka, yoga, yoghurt

(7) 学術語 (learned words)

口語コーパスデータ元となったオリジナルデータで頻出する語形の多くは、トピックと密接な関係が認められる。特に MICASE や Santa Barbara Corpus などは、米国大学の講義や会話などのアカデミックな環境下の発話データであり、やや専門的な学術語が多く認められる。以下から、多くの講義や会話のトピック

は,大学制度,生命科学や社会科学に関する語彙が多く見られる。

adrenaline, alkali, allele, android, anesthetic, annum, aquifer, benzodiazepine, cannabis, chloride, citric, clone, coefficient, colliery, coma, conductance, conductivity, contraception, cortex, cystoscopy, denominator, descriptor, détente, dissertation, dormitory, egalitarianism, endogenous, epistemology, feudalism, freshman, genetics, genome, germ, glucagon, glycolysis, hallucinatory, helium, hemoglobin, hernia, hydrochloric, hydroxide, inertia, kinase, libido, logistics, lupus, meiosis, metastasis, metastasize, mitosis, neutron, operon, phosphate, photon, plasmid, polymerase, proton, protectionism, redistribution, schizophrenia, semester, serotonin, sophomore, superego, tetanus, transfusion, transgene, transgenic, truss, tryptophan

(8) 新語 (new coinage)

英語の語形成のルールで、新たに形成された新語に由来すると思われる語形。 email (< electronic mail), plasticine (商品名由来: AmE = modelling clay), sitcom (< situation comedy), spokesperson (< spokesman)

なお, 語形・品詞・語源情報, BrE・AmEの判別については, 米語については *The American Heritage Dictionary of the English Language 4th Edition* (2000) Boston: Houghton Mifflin Company, 英語については *Concise English Dictionary 10th Revised* (2002) Oxford: Oxford University Press を用いた。

(県立長崎シーボルト大学・上村俊彦)

4.2.4 削除候補語

この節では、口語コーパスの頻度で上位 8000 語を選んだ場合、それから外れる JACET 8000 のうちの 1153 語について考察する。これらの語は、口語コーパス 8015 語と JACET 8000 の差分を求めることで、求められる。これらの 1153 語を口語コーパスにおける頻度順 8015 語から脱落した JACET 8000 の 1153 語とする。この 1153 語を以下、JACET 8000 独自語とする。

JACET 8000 独自語の特徴

それでは口語コーパスでもっともよく使われる 8015 語から抜け落ちた JACET 8000 独自語はどのような特徴をもつのだろうか。これについて考察する。

まず、形態的な特徴があるかどうかを考察する。JACET 8000 独自語の品詞分類は以下のとおりである。

| a | 264 |
|---------------|------|
| a/adv | 2 |
| a/ad/n | 1 |
| a/n | 50 |
| a/n/v | 1 |
| a/n/v/int | 1 |
| a/v | 4 |
| a/v/n | 1 |
| adv | 122 |
| adv/a | 1 |
| adv/a/n | 2 |
| conj | 1 |
| int | 1 |
| n | 504 |
| n/a | 26 |
| n/a/int | 1 |
| n/a/v | 4 |
| n/v | 67 |
| n/v/a | 1 |
| prep/conj/adv | 1 |
| pron | 1 |
| v | 70 |
| v/n | 27 |
| | 1153 |
| | |

ここで、JACET 8000 独自語の形態的特徴をよりわかりやすく見るために、主要な品詞である、名詞、動詞、形容詞、副詞の 4 品詞だけに分類される語に限定して考察する。この 4 品詞のみに分類される語は、JACET 8000 独自語で 1153 語中 837 語である。この 837 語の 4 品詞の配分が、JACET 8000 の配分と同じであるかどうか調べることにより、JACET 8000 独自語の形態的特徴が明らかになると想定される。4 品詞の JACET 8000 独自語と JACET 8000 での分布は以下のとおりである。

| | JACET 8000 | | JACET | |
|-----|------------|-----|-------|-----|
| | 独自語 | | 8000 | |
| 名詞 | 504 | 44% | 3619 | 57% |
| 動詞 | 70 | 6% | 869 | 14% |
| 形容詞 | 264 | 23% | 1361 | 21% |
| 副詞 | 122 | 11% | 495 | 8% |

名詞の比率は JACET 8000 独自語では JACET 8000 より 13%と大きく減少している。動詞も JACET 8000 独自語で JACET 8000 の 2 分の 1 以下に減少している。それに対して,形容詞は JACET 8000 独自語で JACET 8000 に対して 2%, 副詞は 3%上昇している。形容詞は 1153 語中 264 語で 23%を, 副詞は 122 語で 11%を占める。形容詞と副詞を合わせると JACET 8000 独自語の 34%, 約 3 分の 1 を占めることになる。

これは口語コーパスから形容詞と副詞がより多く脱落し、名詞と動詞は脱落しずらい傾向にあることを示唆している。

次に、JACET 8000 独自語の内容を考察する。JACET 8000 独自語の大半は、abrupt、accustom、comprehend、disapprove のように多くの分野で使われる抽象語である。分野が特定できるものは多くはないが、政治、経済、社会、法律、宗教、医学、軍事、生物、化学、地学、生活に分類すると以下の表のようになる。

| 分野 | JACET 8000 独自語 | | | |
|----|--|--|--|--|
| 政治 | aide, apartheid, bilateral, bureaucrat, censorship, congressional, | | | |
| | conservatism, czar, dictator, diplomat, dominion, duchy, | | | |
| | dynasty, elector, envoy, fascism, governmental, legislator, | | | |
| | municipal, patriotic, ratification, slavery, statesman, suffrage, | | | |
| | taxable, truce, etc. | | | |
| 経済 | buyout, competitiveness, conglomerate, consolidation, | | | |
| | devaluation, entrepreneur, fiscal, shipment, takeover, etc. | | | |
| 社会 | aristocracy, aristocratic, beggar, bourgeois, curator, peasantry, | | | |
| | etc. | | | |
| 法律 | acquit, creditor, culprit, debtor, heir, hereditary, indictment, | | | |
| | injunction, inquest, judiciary, libel, penal, policyholder, | | | |
| | proprietor, prosecutor, etc. | | | |
| 宗教 | altar, clergyman, convent, crypt, cult, diocese, evangelical, friar, | | | |
| | fundamentalist, goddess, ministerial, monastery, oracle, | | | |
| | orthodoxy, pagan, pilgrim, rector, shrine, vault, etc. | | | |
| 医学 | anal, anatomy, backbone, biopsy, cardiac, childbirth, cholera, | | | |

| | coronary, coroner, dementia, distal, duodenal, esophageal, fetal, |
|----|---|
| | gall, gastric, groin, hepatitis, infirmary, lymphocyte, neural, |
| | nostril, pharmaceutical, psychic, serum, surgical, ulcer, etc. |
| 軍事 | admiral, ambush, battalion, battlefield, cadet, cavalry, ceasefire, |
| | disarmament, enlist, fortress, garrison, grenade, marshal, |
| | militia, partisan, siege, etc. |
| 化学 | carbohydrate, catalyst, cellular, charcoal, ethanol, polymer, |
| | pylorus, etc. |
| 生物 | alga, ape, bacterial, falcon, fern, hedgehog, larva, otter, parasite, |
| | robin, toad, etc. |
| 地学 | crater, glacier, gorge, granite, lava, ore, volcanic, etc. |
| 生活 | briefcase, compass, cradle, cuisine, dine, eyelid, handkerchief, |
| | hearth, lavatory, luncheon, overcoat, quarrel, shopkeeper, wink, |
| | etc. |

(麗澤大学・望月正道)

4.3 英米児童リーダー・コーパス

4.3.1 データの収集

JACET 8000 を補完する目的でサブコーパスを構築するに当たり、その一つのポーションを児童用のリーダー(読み物)とすることにした。英語圏話者が広く親しんでいる児童用の読み物は、新聞や教材などのデータソースとは異なる重要語彙を浮かび上がらせる可能性があると考えたためである。

「赤毛のアン」「小公子」「魔法使いオズ」などの児童文学は、どれも英語圏の児童に広く膾炙したものであり、母語話者にとってはごく早い段階での語彙インプットになっているものである。こうしたデータを収集・分析することは、母語話者の語彙世界の根幹部分を探る一助になると思われる。

本プロジェクトにおいては、人口への膾炙という点を念頭に、下記のような二つのコーパスポーションを作成することとした。このうち、(1)は筆者が、(2)は望月正道氏が作成した。

- (1) オーセンティックな児童英米文学作品
- (2) グレイディッドリーダー化された児童用英米読み物

厳密には(1)と(2)は完全に同質のものではないが、本プロジェクトでは、「児童を主たる対象として書かれたフィクションの読み物」としてこれらを同等に扱い、広義の児童リーダーと呼ぶことにする。

本プロジェクトで収集された作品データを下記に示す。なお、有名なものについては邦 題名を付記している。また、グレイディッドリーダーのファイル名の冒頭の数字は、読解 難易度を示す。

(1) オーセンティックな児童英米文学作品(20作品)

| ファイル名 | 邦題作品名 | サイズ(KB) |
|----------------------|-------------------|---------|
| alice_looking glass | 鏡の国のアリス | 173 |
| alice_wl | 不思議の国のアリス | 152 |
| anne_avonlea | 赤毛のアン | 502 |
| anne_green gable | 赤毛のアン | 577 |
| anne_house of dreams | 赤毛のアン | 449 |
| beauty & the beast | 美女と野獣 | 497 |
| daddy long legs | 足長おじさん | 221 |
| dolittle | ドリトル先生 | 145 |
| golden road | 金の道(L.Montgomery) | 428 |
| jungle books | ジャングルブック | 281 |

| lame prince | 王子とマント(D. Craik) | 256 |
|-------------------------|------------------|-----|
| little prince | 小公子 | 324 |
| little_princess | 小公女 | 373 |
| oz, marvelous land of | オズの虹の国 | 250 |
| oz, wonderful wizard of | オズの魔法使い | 219 |
| peter pan | ピーターパン | 262 |
| peter rabbit etc | ピーターラビット | 165 |
| secret garden | 秘密の花園 | 444 |
| treasure island | 宝島 | 377 |
| wind in the willow | 柳の風(K. Grahame) | 334 |

(2) グレイディッドリーダー化された児童用英米読み物

| ファイル名 | 邦題作品名 | サイズ(KB) |
|----------------------------------|-------------|---------|
| (1) A Little Princess | 小公女 | 32 |
| (1) Aladdin | アラジン | 29 |
| (1) Mary Queen of Scots | メアリ・スチュワート | 34 |
| (1) Munity on the Bounty | | 32 |
| (1) The Coldest Place on Earth | | 31 |
| (1) The Elephant Man | エレファントマン | 28 |
| (1) The Phantom of the Opera | オペラ座の怪人 | 35 |
| (1) The President's Murderer | | 29 |
| (1) The Wizard of Oz.1 | オズの魔法使い | 29 |
| (1) Tom Sawyer.1 | トムソーヤ | 32 |
| (2) Anne of Green Grables | 赤毛のアン | 32 |
| (2) Dracula | ドラキュラ | 41 |
| (2) HenryVII Iand His Six Wives | ヘンリ7世と6人の后 | 34 |
| (2) Huckelberry Finn | ハックルベリーフィン | 31 |
| (2) Junglebook | ジャングルブック | 35 |
| (2) Robinson Crusoe | ロビンソンクルーソ | 34 |
| (2) Agatha_ChristieWoman_of1 | アガサクリスティ短編集 | 33 |
| (2) Alice's_Adventures_in_Wond.1 | アリスの不思議の国 | 34 |
| (2) Sherloch_Holmes.1 | シャーロックホームズ | 34 |
| (2) The_Death_of_Karen_Silkwoo.1 | | 31 |
| (3) Call of the Wild | 野生の声 | 59 |
| (3) Frankenstein | フランケンシュタイン | 50 |

| (3) Love Story | ラブストーリー | 48 |
|----------------------------------|----------------------------|-----|
| (3) Sky_Jack! | | 47 |
| (3) The_Bronte_Story | ブロンテ姉妹物語 | 56 |
| (3) The_Card | | 61 |
| (3) The_Last_Sherlock_Holmes_S | シャーロックホームズ | 54 |
| (3) The_Picture_Of_Dorian_Gray.1 | ドリアングレイの肖像 | 56 |
| (3) The_Secret_Garden | 秘密の花園 | 59 |
| (3) Through_The_Looking-glass | アリスの鏡の国 | 58 |
| (4) A Morbid Taste for Bones | | 92 |
| (4) Mr midshipman Hornblower | | 72 |
| (4) The Moonspinners | | 75 |
| (4) The Silver Sword | 銀の剣 | 80 |
| (4) The Songs of Distant Earth | | 83 |
| (4) The Thirty-Nine Steps | | 92 |
| (4) Cranford | クランフォード(E. Gaskell) | 83 |
| (4) Gulliver's Travel | ガリバー旅行記 | 82 |
| (4) Lord Jim | ロードジム | 104 |
| (4) Washington Square | | 87 |
| (5) David Copperfield | デイビッドコッパーフィールド(C. Dickens) | 136 |
| (5) Deadlock | | 109 |
| (5) Far From Maddening Crowd | | 138 |
| (5) Ghost Story | 怪談 | 112 |
| (5) Great Expectations | 大いなる遺産 | 131 |
| (5) Sense and Sensibility | 分別と多感(J. Austen) | 140 |
| (5) The Bride Price | | 109 |
| (5) The Dead of Jericho | | 135 |
| (5) This Rough Magic | | 119 |
| (5) Wuthering Heights | 嵐が丘 | 128 |

以上の合計は13039KB(約13MB)となる。

4.3.2 英米児童リーダー・コーパスの構成語彙プロフィール

以上 70 種類,合計 175 万語のデータを語彙解析し、まず、粗頻度表を作成した。 Wordsmith Version 4.0 によって基本的な語彙プロフィールを概観したところ、下記のようになった。

| 指標 | 数値 |
|--------------|-------------|
| 延べ語数 | 1751612 |
| 異なり語数 | 28663 |
| 標準化済 TTR 指数 | 39.91560745 |
| 平均語長 (文字数換算) | 4.128300667 |
| 文数 | 134362 |
| 平均文長(文字数換算) | 13.02518654 |

また、粗頻度表上位20語は下記のようになる。

| 順位 | 語 | 粗頻度 | 構成比 | PMW | レンジ数 | レンジ比 |
|----|------|-------|--------|--------|------|--------|
| 1 | THE | 88659 | 5.0616 | 50616 | 70 | 100 |
| 2 | AND | 62081 | 3.5442 | 35442 | 70 | 100 |
| 3 | ТО | 47103 | 2.6891 | 26891 | 70 | 100 |
| 4 | A | 37954 | 2.1668 | 21668 | 70 | 100 |
| 5 | Ι | 35378 | 2.0197 | 20197 | 70 | 100 |
| 6 | OF | 32321 | 1.8452 | 18452 | 70 | 100 |
| 7 | WAS | 27402 | 1.5644 | 15644 | 70 | 100 |
| 8 | HE | 26575 | 1.5172 | 15172 | 70 | 100 |
| 9 | IN | 24425 | 1.3944 | 13944 | 70 | 100 |
| 10 | IT | 22067 | 1.2598 | 12598 | 70 | 100 |
| 11 | YOU | 20570 | 1.1743 | 11743 | 70 | 100 |
| 12 | SHE | 18552 | 1.0591 | 10591 | 69 | 98.571 |
| 13 | THAT | 18478 | 1.0549 | 10549 | 70 | 100 |
| 14 | HER | 15546 | 0.8875 | 8875.3 | 70 | 100 |
| 15 | HIS | 14416 | 0.823 | 8230.1 | 70 | 100 |
| 16 | HAD | 13976 | 0.7979 | 7978.9 | 70 | 100 |
| 17 | SAID | 13661 | 0.7799 | 7799.1 | 70 | 100 |
| 18 | BUT | 13646 | 0.7791 | 7790.5 | 70 | 100 |
| 19 | FOR | 12389 | 0.7073 | 7072.9 | 70 | 100 |
| 20 | WITH | 12217 | 0.6975 | 6974.7 | 70 | 100 |

このうち、構成比は総延べ語数に占める個々の語の頻度を百分率で示したものであり、PMW (per million words) は百万語あたりの正規化頻度である。また、全体で作品ファ

イルは 70 種類あるが、このうちいくつのファイルに出現しているかを示すのがレンジ数であり、その割を百分率で表したのがレンジ比となる(上記では、SHE を含まないファイルが 70 のうち 1 つだけ存在することになる)。

次に上記のデータをレマ化した。レマ化にあたっては、JACET 8000 の本表作成時に使用したレマテーブルを利用した。

| 順位 | レマ | 頻度 |
|----|------|-------|
| 1 | the | 88659 |
| 2 | be | 67839 |
| 3 | and | 62081 |
| 4 | to | 47103 |
| 5 | I | 41641 |
| 6 | have | 40973 |
| 7 | a | 37956 |
| 8 | of | 32321 |
| 9 | he | 27952 |
| 10 | it | 24663 |
| 11 | in | 24425 |
| 12 | you | 22724 |
| 13 | not | 21511 |
| 14 | that | 19640 |
| 15 | she | 19399 |
| 16 | say | 16753 |
| 17 | do | 15638 |
| 18 | her | 15546 |
| 19 | his | 14416 |
| 20 | but | 13648 |

4.3.3 英米児童リーダー・コーパスの語彙レベル

以上のレマ化作業により、28663 語の異なり語数は 19885 語に集約された。なお、19885 語の中には、レマテーブルに存在しないため、自動的にレマ化されない派生活用形なども残っているが、これらはすべて極端な低頻度語となっている。

最初に 19885 語のうち、JACET 8000 に包含されている語の割合を見たところ、下記のようになった。

| 分類 | 語数 | 割合(%) |
|---------|-------|-------|
| J8 包含語彙 | 6531 | 32.8 |
| J8 外語彙 | 13354 | 67.2 |

以上より、19885 語中、約3割がJACET 8000 語彙であり、約7割がそのほかの語彙であることがわかる。JACET 外語彙の割合が高いように思えるが、ここには、固有名詞などが大量に含まれており、後でより詳細に分析することとしたい。

次に、JACET 8000 包含語彙 6531 語に対して、JACET 8000 に基づくレベル付けを行った。レベルは 8 段階とし、高頻度なものから順にレベル $1\sim$ レベル 8 となっている。なお、各レベルは原則として 1000 語であるが、レベル 1 に限っては、"plus 250"の基本語が含まれている為、1250 語となっている。各レベル別の語数は下図のとおりである。

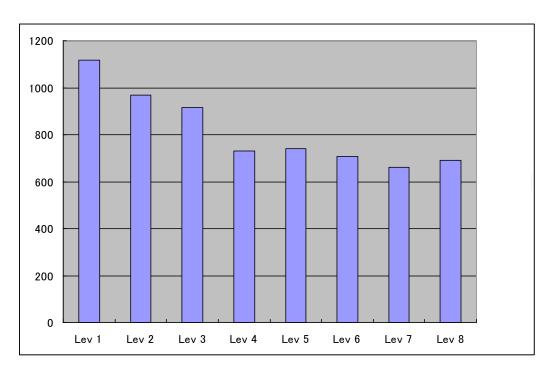


図 4-3-1 英米児童リーダー・コーパスの語彙レベル

上図より明らかなように、レベル1の語彙が多いが、レベル4からレベル8も一定数が使用されており、レベル4~8で語数の差はさほど見られない。児童リーダーの語彙とは、一般に予想されるように基本的な高頻度語だけで構成されているわけではなく、じゅうぶんな語彙的バラエティを持っていることが確認される。

4.3.4 英米児童リーダー・コーパス構成語彙の特徴:(1) 欠落語彙 ある語彙リストの特徴を見るには、当該リストと汎用性の高い参照リストを比較したう えで、1)参照リストにのみ含まれる語彙、および、2)当該リストにのみ含まれる語彙 の両方を見ていく必要がある。

ここでは、英米児童リーダー・コーパスを構成する 19885 語を当該リストとし、JACET 8000 に含まれる 8250 語を参照リストとして比較を行った。まず、参照リストに含まれているが、英米児童リーダー・コーパスリストには含まれていない「欠落語彙」について質的に検証しておく。

8250 語の参照リストに対して、当該リストは 19885 語とはるかにサイズが大きいわけであるが、にもかかわらず、参照リストに含まれ、当該リストにおいて欠落している語彙は 3200 語にのぼる。ここでは欠落語のうち、参照リストでレベル1に含まれる語の一部を見る。まず、plus 250 に含まれる語の中では、下記の 35 語が欠落している。

| zero | thirtieth | Korea | Ms. |
|------------|------------|----------|------|
| billion | fortieth | Chinese | CD |
| eleventh | fiftieth | Japanese | TV |
| twelfth | sixtieth | Korean | No. |
| thirteenth | seventieth | Asia | Ltd. |
| fourteenth | eightieth | Asian | |
| fifteenth | ninetieth | Sydney | |
| sixteenth | hundredth | Tokyo | |
| eighteenth | millionth | Seoul | |
| nineteenth | billionth | Beijing | |

これらは数詞や固有名詞などで、児童リーダーに出現しないことはある意味で当然といえよう。

次に、通常のレベル1語のうち、児童リーダーで欠落している22語を概観する。

| 順位 | 語 | 品詞 | 順位 | 語 | 品詞 |
|-----|---------------|---------|-----|-------------|-------|
| 127 | may | modal/v | 849 | stress | n/v |
| 337 | better | adv | 857 | traditional | a |
| 398 | percent | n | 873 | industry | n |
| 542 | international | a | 877 | total | a/n/v |
| 567 | best | adv | 878 | reduce | v |
| 583 | environment | n | 934 | involve | v |
| 615 | product | n | 935 | achieve | v |
| 722 | economic | a | 945 | basic | a |

| 731 | role | n | 955 | average | a/n |
|-----|------------|---|-----|----------|-----|
| 742 | movie | n | 985 | resource | n |
| 788 | technology | n | 989 | video | n |

上記のうち、may や better や best などは特定品詞の出現なかっただけで、児童リーダーリストにも表記形としては含まれている。しかし、そのほかの 19 語は欠落語である。これらの特徴を質的に見ていくと、movie/video などの現代文化に関する語、international/industry / technology / economic / resource などの現代の社会・技術・経済などに関する語などが目につく。また、reduce/involve / achieve などのように、概念自体は一般的であるが、同じ意味を表すより平易な語が存在する難易語もここに含まれる。

これらは、一般に「中学校程度」と理解されがちな JACET 8000 のレベル 1 語彙の中にも、一般の児童用語とはとうてい呼べない語彙が混入していることを示す。以上のような児童リーダーでの欠落語彙情報は、たとえばレベル 1 語彙を小学校や中学校の英語教育で活用しようとする場合、除去すべき語彙の手がかりを与えてくれるものである。

4.3.5 英米児童リーダー・コーパス構成語彙の特徴:(2)独自語彙

英米児童リーダー・コーパスに含まれ、JACET 8000 には含まれない独自語彙は 11128 語にのぼる。しかし、この中には、固有名詞や、自動レマ化でこぼれた派生形、または非語 (non-word) などが多く含まれている。独自語彙の中で、頻度順上位 20 語を下記に示す。

| 順位 | 語 | 頻度 | 順位 | 語 | 頻度 |
|----|---------|------|----|-----------|-----|
| 1 | ANNE | 3373 | 11 | DIANA | 683 |
| 2 | # | 1552 | 12 | JIM | 667 |
| 3 | 'D | 1397 | 13 | JOHN | 538 |
| 4 | ALICE | 1327 | 14 | GILBERT | 511 |
| 5 | MARILLA | 1304 | 15 | COLIN | 499 |
| 6 | CA | 1178 | 16 | DOROTHY | 490 |
| 7 | MARY | 1056 | 17 | О | 446 |
| 8 | SARA | 1035 | 18 | MATTHEW | 442 |
| 9 | WO | 811 | 19 | CATHERINE | 425 |
| 10 | PETER | 791 | 20 | TH | 376 |

20 語のうち、7割にあたる 14 語は人名である。残りは、誤混入した記号 (#) や、語要素の一部などである (would o'd、序数語尾の th、o'clock o o など)。このほか、低位の語彙には「視覚方言」に類するものも多くある。

本来、児童リーダー・コーパスの独自語彙は、JACET 8000 を拡張する上で貴重な資料となるはずであるが、上記のような語は考慮の対象とならない。そこで、11128 語のうち、分析対象として頻度 4 以上の 2669 語を選び、手作業でレマ化を行った。レマ化は、原則として s 形、ed 形、ing 形に対して機械的に行ったが、辞書なども参照しつつ、一部の語については分詞形を見出し語として残している。

レマ化によって当該語数は 2598 語になったが、ここにはまだ、人名・地名などの固有名詞や、非語などが大量に含まれている。そこで、コンピュータのスペルチェック機能を利用し、1) 大文字語、2) 非語、および、数詞・記号などをデータから除去することで、語数を約 1034 語にまで絞り込んだ。これらは、JACET 8000 を補完する"Supplement 1000"の候補語となるものである。

4.3.6 英米児童リーダー・コーパスから得られた追加候補語

以上の作業手順を経て、1034 語の追加候補語を得た。こここでは、品詞別にその一部を 概観しておくこととしたい。

4.3.6.1 名詞

名詞には古風な語や雅語が多い(morrow < tomorrow など)。そのほか、植物の名前などもある程度含まれている、また、児童向き読み物ではあるものの、意味的・形態的にかなり難易度の高い語も含まれている。下記は tion 型接辞を持つ名詞群である。

| consternation | 7 | apparition | 4 |
|---------------|---|----------------|---|
| dissipation | 7 | contortion | 4 |
| benediction | 6 | exertion | 4 |
| desertion | 6 | indiscretion | 4 |
| hallucination | 6 | inscription | 4 |
| desolation | 5 | predestination | 4 |
| elation | 5 | protestation | 4 |
| superstition | 5 | subtraction | 4 |
| adoration | 4 | supplication | 4 |
| affectation | 4 | | |

名詞の場合は、元になる動詞が JACET 8000 にすでにエントリーされている場合も多い。 上記の例について言うと、superstition、adoration、affectation、apparition などは一般的 に認知度の高い語であり、JACET 8000 の補完語として検討されるべきであろう。

4.3.6.2 動詞

動詞には物語の特性を反映して、伝達・発話に関するもの、感情に関するもの、擬音語や擬態語に由来する語などが散見される。

| daresay | 27 | nip | 9 |
|----------|----|---------|---|
| christen | 17 | revel | 9 |
| repent | 16 | clamber | 8 |
| chum | 14 | cower | 8 |
| scull | 14 | nourish | 8 |
| scuffle | 13 | snub | 8 |
| clatter | 12 | tryst | 8 |
| creak | 10 | blight | 7 |
| fidget | 9 | grope | 7 |
| flounce | 9 | keel | 7 |

上記についていえば、repent, clatter, creak, nourish, grope などは幅広いテキストでしば しば用いられる語彙であり、JACET 8000 補完語としてじゅうぶんに検討に値するもので ある。

4.3.6.3 形容詞

物語では、登場人物の感情描写が重要な位置を占める。そのため、感情や雰囲気に関する表現が多い。

| moody | 24 | suggestive | 8 |
|----------------|----|---------------|---|
| congenial | 8 | sunshiny | 8 |
| expressionless | 8 | surly | 8 |
| greenish | 8 | uneducated | 8 |
| innumerable | 8 | uninteresting | 8 |
| ladylike | 8 | unworthy | 8 |
| lithe | 8 | vile | 8 |
| pearly | 8 | airless | 7 |
| resentful | 8 | bonny | 7 |
| reverent | 8 | businesslike | 7 |

このうち、moody、pearly、resentful、suggestive などは一般的にも多く使う語彙である。 なお、形容詞の中には、uneducated など、過去分詞形が自立したものが多く含まれてい る。これらは、独立の語彙とみなすかどうかは微妙な判断となるが、こうしたタイプには、 ほかに aforesaid, undressed, beset, unafraid, unfastened, witted, ivied, singed な ども含まれる。

4.3.6.4 副詞

副詞は動作の様態をいきいきと描き出すものであるが、「物語」であることの必然的結果として、1034 語の中には各種の副詞が含まれている。下記は ly 副詞のうち高頻度の 20 語を示したものである。

| cowardly | 42 | impressively | 8 |
|----------------|----|---------------|---|
| cheerily | 19 | indifferently | 8 |
| blandly | 8 | luxuriously | 8 |
| blithely | 8 | naturedly | 8 |
| complacently | 8 | prettily | 8 |
| diligently | 8 | sagely | 8 |
| disconsolately | 8 | sleepily | 8 |
| exultantly | 8 | soberly | 8 |
| gruffly | 8 | steeply | 8 |
| hungrily | 8 | threateningly | 8 |

cowardly, **cheerily**, **indifferently** など, 一般的にもよく使う語が含まれており, これらは **JACET 8000** の補完語としても適切であると考えられる。

4.3.7 制約と課題

以上,英米児童リーダー・コーパスの語彙の特徴について概観し,JACET 8000 の補完 候補語について検討してきた。結果として,有用な候補語が多く見つかり,分析は一定の 効果をあげた。

しかしながら、今後のプロジェクトにおいては、児童リーダーの位置づけについて見直 しを行うことが必要となろう。というもの、今回のプロジェクトでは、児童文学作品を多 くデータに取り込んだが、分析の結果、その構成語彙は予想以上にレベルが高いことが明 らかになったからである。

一般に語彙表構築の基礎データとして児童文学を入れる場合、子供たちにとってなじみの深い日常語や生活用語をそこから抽出することを期待している場合が多い。だが、実際には、大半の児童文学作品は、語彙の観点においては、「児童」向けであるというよりも、あくまでも古典的な「文学作品」の一種であるように感じられる。この意味において、児童文学とグレイディッドリーダーは語彙的にはむしろ異質なデータでありうる。

もちろん、以上のことは語彙表構築のための基礎データとしての児童文学の価値を否定するものではない。実際、児童文学コーパスでなければ抽出が困難であった重要語彙も少なくない。たとえば、今回の分析で得られた JACET 8000 補完候補語の1つに nigh がある。これは near の同義語であるが、とくに詩的な文脈などで散見される語である。 COBUILD でも LDOCE でも"old-fashioned"というスピーチラベルを付与されているが、現代においても教養性の高い文章ではごく一般的に使用されるものである。おそらく、教材や新聞などのサブコーパスでは拾いにくいであろう、こうした語彙をうまく抽出するという目的のためには、児童文学は大いに有効なデータとなる。

今後は、児童文学の言語的特性をさらに詳細に検討しながら、全体のプロジェクトにおけるその位置づけについてもあわせて検討を重ねてゆくこととしたい。

(神戸大学・石川 慎一郎)

参考文献

- Carter, R., & Nash, W. (1990). Seeing through language: A guide to styles of English writing. Oxford, UK: Basil Blackwell.
- Dreyer, S. S. (1977). A guide to children's literature about the needs and problems of youth aged 2 to 15, London: Ags Publishing.
- Hunston, S. (2002). Corpora in applied linguistics. Cambridge, UK: Cambridge UP.
- 石川慎一郎. (2005). 「司法英語 ESP 語彙表構築の試み: FROWN コーパスと米国司法文献 コーパスの比較に基づく特徴語の抽出」『神戸大学国際コミュニケーションセンター 論集』 1,13-28.
- Leech, G., Rayson, P., & Wilson, A. (2001). Word frequencies in written and spoken English: based on the British National Corpus. Harlow, UK: Pearson Education Limited.
- 定松正. (1985). 『イギリス児童文学散歩』.中教出版.
- 瀬田貞二 et al. (1971). 『英米児童文学史』. 研究社.
- Stubbs, M. (2002). Words and phrases: Corpus studies of lexical semantics. Malden, MA: Blackwell Publishing.
- タウンゼント, J. R. (1982). 『子供の本の歴史』(上下). 岩波書店.
- Teubert, W., & Čermáková, A. (2004). Directions in corpus linguistics. In M.A.K. Halliday, W. Teubert, C. Yallop, & A. Čermáková. *Lexicology and corpus linguistics: an introduction* (pp. 113-165). London, UK: Continuum.

4.4 教科書コーパス

4.4.1 検定教科書コーパスの構築

日本人学習者が、その英語学習において最も大きな影響を受けるのは、中学校や高等学校での英語の授業であり、その中でも、学習教材として中心的地位を占める検定教科書の果たす役割は非常に大きいと言える。そのため、本研究で構築を目指す語彙リストにおいても、中学校や高等学校の文部科学省検定済教科書で使用されている語彙を反映させる必要があると考えられる。そこで、現行(平成 16~17 年度版)の中学校、高等学校英語検定教科書を可能な限り収集し、その本文及び練習問題の英文を電子化したサブコーパスを構築した。収集した教科書数は、高等学校「英語 I」が 29 冊、同「英語 II」が 21 冊、同Reading が 11 冊、同「オーラル・コミュニケーション I」が 9 冊、同「Writing」5 冊の計 75 冊、中学校が 1 年生用から 3 年生用までの計 21 冊である。サブコーパスのサイズは、1,029,810 語となった。収集した検定教科書のリストを以下に示す。

表 4-4-1 検定教科書サブコーパスに含まれる高等学校英語検定教科書

| 教科書名 | 種目(教科) | 発行者略称 |
|---------------------------------|--------|-------|
| ACORN English Course I | 英語 I | 啓林館 |
| All Aboard! English I | 英語 I | 東書 |
| Captain English Course I | 英語 I | 大修館 |
| Captain English Course II | 英語 II | 大修館 |
| CROWN English Series I | 英語 I | 三省堂 |
| CROWN English Series II | 英語 II | 三省堂 |
| Daily English Course I | 英語 I | 池田 |
| Daily English Course II | 英語 II | 池田 |
| Departure Oral Communication I | OCI | 大修館 |
| ENGLISH NOW I | 英語 I | 開隆堂 |
| EXCEED English Series I | 英語 I | 三省堂 |
| Genius English Course I | 英語 I | 大修館 |
| Genius English Course II | 英語 II | 大修館 |
| Genius English Readings | リーディング | 大修館 |
| Genius English Writing | ライティング | 大修館 |
| Interact Oral Communication I | OCI | 桐原 |
| Lingua-Land English Course I | 英語 I | 教出 |
| MAINSTREAM English Course I | 英語 I | 増進堂 |
| MAINSTREAM English Course II | 英語 II | 増進堂 |
| MAINSTREAM Oral Communication I | OCI | 増進堂 |

(表 4-4-1 続く)

| | (1 | |
|---|--------|-----|
| MAINSTREAM Reading Course | リーディング | 増進堂 |
| MILESTONE English Course I | 英語 I | 啓林館 |
| MILESTONE English Course II | 英語 II | 啓林館 |
| MILESTONE English Reading | リーディング | 啓林館 |
| NEW ACCESS to English Writing | ライティング | 開拓社 |
| NEW English PAL I | 英語 I | 桐原 |
| NEW English PAL II | 英語 II | 桐原 |
| NEW LEGEND ENGLISH I | 英語 [| 開拓社 |
| NEW LEGEND ENGLISH II | 英語 II | 開拓社 |
| NEW LEGEND ENGLISH READING | リーディング | 開拓社 |
| NEW STAGE English Course I | 英語 [| 池田 |
| NEW STAGE English Course II | 英語 II | 池田 |
| NEW STREAM I | 英語 I | 増進堂 |
| NEW STREAM II | 英語 II | 増進堂 |
| NEW STREAM Reading Course | リーディング | 増進堂 |
| NEW WORLD English Course I Revised Edition | 英語 I | 三友 |
| On Air Communication I | OCI | 開拓社 |
| ORAL COMMUNICATION EXPRESSWAYS I Advanced Edition | OCI | 開隆堂 |
| ORAL COMMUNICATION EXPRESSWAYS I Standard Edition | OCI | 開隆堂 |
| Planet Blue English I | 英語 I | 旺文社 |
| Planet Blue English II | 英語 II | 旺文社 |
| Planet Blue Reading Navigator | リーディング | 旺文社 |
| POLESTAR English Course I | 英語 I | 数研 |
| POLESTAR English Course II | 英語 II | 数研 |
| POLESTAR Reading Course | リーディング | 数研 |
| POLESTAR Writing Course | ライティング | 数研 |
| Power On English I | 英語 I | 東書 |
| POWWOW ENGLISH COURSE I | 英語 I | 文英堂 |
| PROMINENCE English I | 英語 I | 東書 |
| PROMINENCE English II | 英語 II | 東書 |
| PRO-VISION ENGLISH COURSE I | 英語 I | 桐原 |
| PRO-VISION ENGLISH COURSE II | 英語 II | 桐原 |
| PRO-VISION ENGLISH READING | リーディング | 桐原 |
| PRO-VISION ENGLISH WRITING | ライティング | 桐原 |
| | | |

(表 4-4-1 続く)

| Sailing Oral Communication I | OCI | 啓林館 |
|--------------------------------|--------|-----|
| SELECT Oral Communication I | OCI | 三省堂 |
| Surfing English Course I | 英語 I | 文英堂 |
| Surfing English Course II | 英語 II | 文英堂 |
| True Colors Oral Communication | OCI | 数研 |
| Unicorn English Course I | 英語 I | 文英堂 |
| UNICORN ENGLISH COURSE II | 英語 II | 文英堂 |
| VISTA English Series I | 英語 I | 三省堂 |
| Viva English! I | 英語 I | 第一 |
| Viva English! II(Unit 1) | 英語 II | 第一 |
| Viva English! II(Unit 2) | 英語 II | 第一 |
| Vivid English Course I | 英語 I | 第一 |
| Vivid English Course II | 英語 II | 第一 |
| Vivid Reading | リーディング | 第一 |
| Voyager English Course I | 英語 I | 第一 |
| Voyager English Course II | 英語 II | 第一 |
| Voyager Reading Course | リーディング | 第一 |
| WORLD TREK ENGLISH COURSE I | 英語 I | 桐原 |
| WORLD TREK ENGLISH COURSE II | 英語 II | 桐原 |
| WORLD TREK ENGLISH Reading | リーディング | 桐原 |
| WORLD TREK ENGLISH Writing | ライティング | 桐原 |

^{*}タイトルはアルファベット順に表記

表 4-4-2 検定教科書サブコーパスに含まれる中学校英語検定教科書

| 教科書名 | 出版社 |
|------------------------------|-----|
| COLUMBUS 21 ENGLISH COURSE 1 | 光村 |
| COLUMBUS 21 ENGLISH COURSE 2 | 光村 |
| COLUMBUS 21 ENGLISH COURSE 3 | 光村 |
| New Crown English Series 1 | 三省堂 |
| New Crown English Series 2 | 三省堂 |
| New Crown English Series 3 | 三省堂 |
| NEW HORIZON English Course 1 | 東書 |

^{*}OC=オーラル・コミュニケーション

^{*}発行者の略称は文部科学省の教科書目録の表記法に従った。

(表 4-4-2 続く)

| 東書 |
|-----|
| 東書 |
| 教出 |
| 教出 |
| 教出 |
| 開隆堂 |
| 開隆堂 |
| 開隆堂 |
| 秀文館 |
| 秀文館 |
| 秀文館 |
| 学図 |
| 学図 |
| 学図 |
| |

^{*}タイトルはアルファベット順に表記

4.4.2 JACET 8000 の教科書サブコーパスに対するカバー率

検定教科書サブコーパスの語彙に対する JACET 8000 のレベル別カバー率は以下の通りである。

表 4-4-3 JACET 8000 による教科書コーパスの分析結果

| | level 1 | level 2 | level 3 | level 4 | level 5 | level 6 |
|---------|---------|---------|---------|---------|---------|---------|
| indexes | 1128 | 998 | 988 | 758 | 703 | 632 |
| % | 6.38% | 5.65% | 5.59% | 4.29% | 3.98% | 3.58% |
| tokens | 842221 | 53367 | 23237 | 11195 | 5957 | 7006 |
| % | 81.78% | 5.18% | 2.26% | 1.09% | 0.58% | 0.68% |

| level 7 | level 8 | over 9 | cont. | non- words | proper nouns | total |
|---------|---------|--------|-------|---------------|-----------------|---------|
| 546 | 504 | 4642 | 75 | 1216 | 5487 | 17677 |
| 3.09% | 2.85% | 26.26% | 0.42% | 6.88% | 31.04% | 100% |
| 3533 | 3577 | 13179 | 8906 | 20528 | 37104 | 1029810 |
| 0.34% | 0.35% | 1.28% | 0.86% | 1.99% | 3.60% | 100% |

^{*}発行者の略称は文部科学省の教科書目録の表記法に従った。

Token で見た場合,Level 1 から 3 までで全体の約 90%(89.22%)をカバーしている。残りの約 10%のうち,電話番号や,数字の羅列等の non-words と固有名詞が 5.59%となっている。また,over 9 となっている語彙にも,tatami や shamisen 等の相当数の日本語が含まれる。従って,これらの語彙も含めると,カバーできる割合は約 95%となる(3000 語レベル+non-words+固有名詞=94.81%+ α)。これらの数値を考慮すると,検定教科書で使用されている語彙に対する JACET 8000 のカバー率は比較的高く,Level 3(上位 3000 語)を習得しておけば,多くの検定教科書に対応可能であると言える。

その他,以下の点が明らかになった。

- (1) 検定教科書は、Level 1 で 80%のカバー率を越える(81.78%)。
- (2) JACET 8000 ではカバーできない語が、1.28% (13,179 語:token) 存在する。
- (3) JACET 8000 の Level 8 までのカバー率は 92.26% となる。

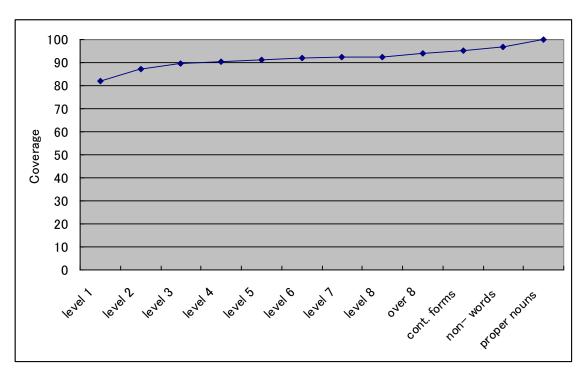


図 4-4-1 JACET 8000 による教科書コーパスのカバー率

4.4.3 教科書サブコーパスの語彙の特徴

JACET 8000 の検定教科書に対するカバー率は比較的高いといえるが、検定教科書で使用されている語彙の中には、JACET 8000 のリストには含まれていないものも存在する。そこで、これらの語彙にはどのようなものがあるかを分析した(短縮形、non-words、固有名詞は除く)。以下のリストは、教科書サブコーパスのみに出現した語彙(JACET 8000 には含まれていない語彙)の上位 500 語のリストである。

email homer octopus landmine albatross scold classmate chopstick shogi

carol cellphone well-known tricycle kilogram samurai spaghetti synthesizer

iack woof zebra subtitle graduation ramen left-handed hike tempura hamburger right-handed announcer rainy volleyball souvenir cloudy heartbeat natto kimono trash bakery karaoke panda lope T-shirt proverb prefecture haiku domo thirsty

somedaykyogensoftballjudocactuscafeteriavendnoodlekangaroopart-timesightseeingslurpsubwayclonesnowy

cf website max

DNA shampoo heavy-mud surf typhoon robopet hometown cooperation medaka sushi kimchi vase cookie koala kanji Valentine mammoth miso bamboo spaceship equator creek flea homepage sumo stomachache kendo toothache pacific moo

pacific moo toothache bookstore translator tuxedo zoo braille oink shin brooch alphabet

shamisen homeroom appetizer

dynamite snowboard angler humankind taco birthplace tortoise hot-air centigrade teammate gagaku dough dormitory dribble doggy geologist freshman ecosystem kiwi short-term miniskirt nonnative manga acronym swimmer alligator pheromone thank-you cartoonist politeness pet-robots cedar potluck hardworking prefix cleanup homesick refresh long-term humanistic er smog kabuki humid soy misunderstand mockingbird uh-huh omelet oxidation unburnable pavilion violinist photojournalist trestle woodcutter plankton muzukashii wrestler popcorn canyon reuse lastly closet sukiyaki powwow superman vanilla okonomiyaki symbolize forth warm-up tofu pitcher gonna burger backyard wetland doorbell badminton miller barrier-free barometer hatchling blackboard gasoline manly upside mural checkup burnable seafood cheeseburger eyesight squash drugstore

edelweiss

ketchup

oversleep

good-looking

steelhead

sneeze

scary

aloha

herbal

malaria

memorize

nonverbal

platypus wombat vinegar rectangle yoghurt waterway seaweed ouch eraser sunrise quack restroom unkind babe shawl brad juku wag chow kumara x-ray manmade loch malt homemade desalination rex

kindergarten sari shirataki
navigate tatami sumimasen
spike tsugaru tsugaru-jamisen

trendy antenna vice-director
bathtub newborn anytime
beginner nonstop catcher
bookshelf anime northeast
diligent barber old-fashioned

bedtime eel paralyze birch passerby grown-up birthrate guidebook sprain hairdresser carton touchdown handlebar bracelet copier brotherhood homestay grapefruit

humanism hitchhike chairperson islander insecticide cherish jellyfish itch coward nap knob detergent paste melon escalator

pumpkin pickle ethnocentrism salesperson proboscis euthanize self-image pula excursion skateboard retina fable

soybean sailboat fair-weather spicy stork fingernail sunglass tectonics hairstyle toothpaste urchin hiker

chili

kelpfast-foodpouchlongevityfirsthandrainwaternumeralhigh-techrecyclableobedienthitterreef

overpopulation juror riverside pajama left-hander scuba peel makeup slum

primrose one-on-one snowstorm proficiency pharmacy spinach reed ranch spoonful

salesclerk refrain stationmaster saltv roommate stewardess sidewalk tadpole sandal sneaker toothbrush schoolyard self-expression backpack ultraviolet self-reliance bestseller watermelon shaker blackberry zipper

supercontinentcacaosnowmansuspensecaribouricktablespooncarnivalfuton

braindead

speechless

yukata circumference gift-giving irrigation diaper kotatsu lullaby faraway low-ranking

rainforest fireman takoyaki
dandy frittata horsepower
windmill heron nonprofit
reggae judgmental right-hander

reggae judgmental right-har bean-throwing magnetron rocker

furoshiki mash superstar kendama microlight tattoo ling moa umpire ondol mustache von

shopowner origami anchovy yakisoba pancake aurora doping pedal barbaric bazaar nurture beep oppress bird-watching overcrowd boogieman overload botanical paella campfire pager cappuccino pastime carelessness pelican

checkout photosynthesis colonist psychologically

consulate relive

cucumber respiration cupcake schoolworkdreamer semicircle encyclopedia siesta faucet sundae fawn talkative fluently tombstone torchbearer forget-me-not globalization turquoise goldfish unforget tablegorilla walkway

grasshopper hospice
hatchery jazzy
hippie streetcar
humility bristle
internationalize snowball
jigsaw potassium

karate domos leek durian linguist fishburger

lotion mailbox motorbike multiplex

grader

world-famous

4.4.4 結果と考察

これらの検定教科書独自の語彙のうち、出現頻度が高いものについては、JACET 8000 に追加する候補語として今後の検討を要するものであると言える。特に、email、hamburger、subway、trash、zoo 等の日常語については、検定教科書において高い頻度で使用されており、英語学習用の語彙リストに含めるべきであると考えられる。一方で、日本人英語学習者用の教科書という特性を反映して、shogi、natto、miso、manga、tatami 等の相当数の日本語がローマ字表記で用いられている。これらの語は JACET 8000 への追加候補語にはならないが、日本の英語検定教科書の特徴であるといえるので、ここで触れておきたい。検定教科書における使用語彙のもう一つの特徴は、異文化に関する語彙である。検定教科書で扱われるトピックには異文化理解に関するものがしばしば用いられるので、kangaroo、kimchi、kiwi、taco、sari、paella 等の異文化に関する語彙が使用されている。

追加候補語を検討する目的で、教科書コーパスの語彙頻度と BNC の順位の比較も行った。JACET 8000 に含まれていない単語で、教科書コーパスで頻度の高い語(頻度 8 回以上)を絞り込み、これらの語を BNC Range で分析して、BNC でどのくらいの頻度になるかを調査した。以下のリストは、教科書コーパスで頻度 8 回以上の語を、BNC で 1000 語刻みに示している。アステリスクは、特に重要語と判断される語で、新たに語彙表に含む候補として考えられる語を示している。

List 1

copier, manly, *misunderstand, unkind, *upside

List 2

*announcer, *bakery, burnable, *cloudy, forth, humanism, humanistic, islander *rainy, *snowy, unburnable

List 3

*nonverbal, *politeness, refreshed, squash, *symbolize, *translator

List 4

*blackboard, *cooperation, *eyesight, *graduation, herbal

List 5

*alphabet, badminton, dough, *hamburger, *knob, *rectangle, *shampoo, *spaghetti, *spicy, *thirsty, *vase, *vinegar, *zoo

List 6

alligator, bedtime, cookie, *diligent, flea, malaria, pavilion, *proverb, *trendy, zebra

List 7

bracelet, brooch, *chairperson, doorbell, *equator, *homesick, kangaroo, karaoke, ketchup, mammoth, *proficiency, slurp, synthesizer, *toothache, *toothpaste, trestle

List 8

acronym, bamboo, cloning, detergent, dynamite, *geologist, *grapefruit, *humid, *kilograms, kindergarten, mural, noodle, numeral, *sightseeing, souvenir, spaceship

List 9

barometer, *cafeteria, handlebar, heartbeat, hiking, octopus, panda, popcorn, sari, smog, someday, subtitle, surfing, volleyball

List 10

*backyard, *birthplace, creek, dormitory, judo, kiwi, prefix, stork, *subway, *trash, typhoon

List 11

astronaut, cactus, *classmate, closet, *humankind

List 12

bathtub, kimono, oxidation

List 13

angler, cf, ecosystem, retina, samurai, scolded, tricycle

List 14

albatross, *freshman, pheromone

4.4.5 JACET 8000 との比較

最後に、JACET 8000 に含まれる語彙の中で、教科書サブコーパスでの頻度がゼロ又は 比較的低いものの分析を行った。中高の検定教科書ということを考慮すれば、Level 5 を 超えるような語彙についてはあまり多くが教科書サブコーパスには出現しないと考えられ るが、Level 4 までの語彙で、検定教科書に出現しないものに関しては、必要な場合には、 JACET 8000 における位置付けの再検討が必要となる可能性がある。 (1) JACET 8000 に含まれる語で教科書サブコーパスには出現しなかった語(Level 1 から 4 まで)

<u>Level 1(plus 250 を含む)</u>

該当無し

Level 2

accord

flatter

remark

Level 3

ally

compound

consultant

declaration

descend

fleet

flesh

frequency

grip

margin

strain

ye

Level 4

bastard \sin allocate allocation bid thinking amendment bind acceptance accommodation boost ass assault borough accountant accuracy assessment breach attendance builder administration adoption attribute capitalism allegation autonomy capitalist allege availability cater alleged aye certainty

certificate deficit hence chancellor devise hostility

chapel directive identification circuit disclose ideological

circular discourse implementation

clash discretion inadequate classification dismissal incidence clause distribute indication client dividend indicator closure dock induce coalition doctrine inherit collective domain initially duration commander input

complicate earl inspection eh installation comply conceive electoral intensity condemn elite interface intervention consistently empirical constituency endorse investigation enforce constitute investor

constitutionalenterprisejudicialconstraintenvisagejurisdictionconsultationequityjustification

contemplate exceptional lad contractor exclusively lease

conversion explicit legislation conviction legislative exposure coordinate facilitate legitimate corp liability fig correlation format limb correspondent formation locally

coup fraud magistrate
coverage fuck maintenance
creditor fucking marginal
database gay membership
defendant goddamn merger

9------

unemployment

minimum recession supervision mode recruit supplier mortgage referendum tenant motive refusal theft

namely regime theoretical naval relevance thereby nonetheless render thesis

obscure replacement transaction offence representation transformation

offender residential treasury opt resignation tribunal retail trustee output revenue overlap turnover parameter riot ultimately undermine parish screw undertake parliamentary segment unemployed payment sergeant

smash personnel utility socialism petition vat pit socialist venture plaintiff sociology venue portfolio specify verse vulnerable practitioner speculation

shareholder

predecessor squad whereby
pregnancy stance wholly
premise statistics withdrawal

premium statute
proposition striker

pensioner

prosecution subsequent
provincial subsequently
provision subsidiary
publicity subsidy

purchaser substantially rape successive ratio suite

上記の語彙には、学術、科学、政治、法律、時事等の所謂 ESP 語彙が多く含まれており、これらは専門用語であるため検定教科書では使用されていないことが分かる。また同時に、JACET 8000 が ESP 語彙を多く含み、各分野の専門語彙を収録していることが改めて示される結果となった。検定教科書には出現しなかった語の数は、JACET 8000 のレベルが上がるに従って増加しているが、レベル毎の語数に関しては、表3を参照されたい。

(2) JACET 8000 に含まれる語で教科書サブコーパスに 1 回しか出現しなかった語 BNC List 7

ale, alga, ammunition, apprehension, apt, aristocratic, arsenal, ascent, ashore, barge, battered, binocular, bitterness, boredom, brisk, buffet, bureaucrat, burglar, cannon, coarse, collectively, commercially, commonplace, comrade, confide, contamination, continual, cozy, crusade, dangerously, deception, descriptive, desperation, deterrent, diabetes, digit, diplomacy, disbelief, discontent, drip, dummy, dwell, dynasty, efficacy, emit, emotionally, enlightenment, epidemic, erode, examiner, expire, familiarity, fatigue, finely, flatten, folly, freezer, freshly, fungus, gamble, glossy, golfer, granite, groove, handwriting, harassment, herald, hierarchical, hymn, icy, inconsistent, indifference, inexperienced, infinitive, informative, inning, instability, instinctively, insulin, intensely, interestingly, internally, introductory, intuition, jay, jug, kidnap, latent, legendary, legion, lethal, lily, loaf, maritime, masculine, mast, maternity, meaningless, mend, moist, moonlight, morally, moth, multitude, mystical, novelty, nude, nylon, ordeal, ore, ornament, outrageous, outskirts, overtime, paperwork, parachute, parlor, pasta, pastel, persuasive, petroleum, plank, plantation, pointless, profoundly, prophet, proprietary, quarterback, queer, ramp, rationale, realization, reckless, reconsider, refreshment, relic, resistant, rewrite, rhyme, rim, ripe, ripple, robin, rumble, sandy, seaside, secrecy, shopkeeper, sparkling, spear, stalk, stark, starter, subscribe, tanker, terminology, testify, thinker, toad, trauma, trio, uniquely, universally, unload, unpredictable, uranium, utmost, versatile, vineyard, wavelength, whereabouts, woolen, yawn, youthful

BNC List 8

adolescence, afflict, affordable, agreeable, annex, anthropological, arrogance, automation, avoidance, awesome, backdrop, barbarian, barren, bestow, bewildered, blackness, bonnet, bravely, brittle, broadcaster, bun, cardinal, catastrophe, cerebral, challenger, chore, comprehend, crocodile, culturally, dazzling, delicately, dignified, dismal, distrust, diver, elegance, eligibility, emancipation, endlessly, enlist, erase,

evacuation, exposition, fern, forceful, forgiveness, friction, fright, gleaming, glittering, gosh, grandeur, grenade, grumble, guise, gull, hasten, heather, heavenly, herring, hideous, homelessness, humane, hush, illumination, improbable, improper, inconvenience, indispensable, infancy, insecurity, instinctive, intently, intercept, intimately, intruder, irregularity, lessen, lizard, luncheon, maximize, mischief, mouthful, mph, muse, mystic, nostalgia, objectively, objectivity, oppressive, oyster, passionately, paw, pebble, peep, persecution, pervasive, pious, pornography, postulate, protector, psychoanalysis, raider, ranger, realist, recess, recite, reformation, resonance, resurrection, retrospect, reverend, richly, routinely, rugged, sandstone, shotgun, simulate, solemn, soothing, spinal, stoop, strikingly, stubborn, supremacy, suspiciously, swirl, tempo, textual, thankfully, timely, tuberculosis, unavailable, undo, undue, unfinished, uniformity, unify, unnoticed, unseen, unthinkable, unused, uphill, vanity, vibrant, vulgar, walnut, ware, watchdog, whine, worthless

(立命館大学・杉森直樹, 東京電機大学・相澤一美)

5. Supplement の開発

- 5.1 Supplement の構築方法
 - 以下の5段階の手順に従って、JACET 8000 Supplement の第一次候補を抽出した。
- 1. 「教科書コーパス」「Spoken コーパス」「米語コーパス」「児童文学コーパス」それぞれの Supplement リスト (追加候補語リスト; 各約 1000 語)をマージする。マージしたリスト内に含まれる語の合計は、この時点で 3341 語であった。
- 2. 上記のマージリストをもとにして、頻度の合計とレンジを算出する。この際に、頻度 は各コーパスとも総語数を 100 万語として換算している。
- 3. 上記で算出されたマージリストを、レンジを第1キー、4つのコーパスの頻度の合計を第2キーとして降順でソートする。この作業によって得られた結果は、表 5-1-1 の通りである。
- 4. 上記3の作業の結果、レンジの最大値は3、最小値は1であった。つまり、4つのコーパスに共通して現れる語は見つからなかった。また、レンジ3の語は149 語、レンジ2の語は506 語であり、その合計は655 語であった。
- 5. レンジ1の語の3分の2以上は、上記4の655位の語の頻度よりも低く、また、それより高い頻度を持つ語でも、間投詞やスラング的な語などが多く含まれていた。また、もともとの4つのコーパス間でサイズのばらつきがあり、相対的に見てサイズの小さいコーパスである「教科書コーパス」と「児童文学コーパス」からの語が、同頻度で数十個並ぶというような現象も見られた。このため、上記4で得られた655語をSupplement 1000の第一次候補とした。

表 5-1-1 各コーパスの Supplement リストをマージした結果 (一部)

| WORD | 教科書 | 口語 | 米語 | 児童文学 | 頻度計 | レンジ |
|-------|--------|-------------|-------------|------|-------------|-----|
| email | 267.03 | 18.02425406 | 121.6118488 | | 406.6661028 | 3 |
| Z00 | 35.93 | 3.350662614 | 118.7835949 | | 158.0642575 | 3 |
| trash | 63.12 | 17.44655361 | 29.10061245 | | 109.6671661 | 3 |

次に、Supplement として候補に挙がった 655 語をさらに以下の手順で吟味した。

- (2) 候補語を,辞書検索ソフト PDIC 用に開発された大型辞書訳語データベース dic2001 (2001 年 8 月版; http://homepage1.nifty.com/yoshi_2000/PDEJ.HTM) と照合する。
- (3) 上記の操作によって意味のつかなかった語(52語)を抽出する
- (4) 抽出した語を研究グループで確認し、採用不適と判断したものは削除。複合語については機械的に solid の形で出力されており、最終的な採用決定には難しい判断が必要になったが、以下の複合語基準を採用した。

※複合語基準: 当該語の表記にあたり, open compound もしくは hyphenated

compound のほうが一般的と判断される場合であっても、solid compound の形が主要辞書のいずれかの見出しに含まれていれば採用とする。ただし snowborder など若干の教育的配慮による判断も許容する。

- (5) 最終段階ではレンジは無視し、合成頻度の上位で並べ替える
- (6) 最終的に 636 語を選定する

Dic2001 に含まれない語彙についての採否判断は下記による。

| er | 間投詞なのでカット |
|------------|--------------------------------------|
| email | 採用 |
| DNA | 略語なの d カット |
| Valentine | 固有名詞なのでカット |
| website | 採用 |
| homer | 採用 |
| max | 採用 |
| joey | 教科書に頻出の「カンガルーの子供」の意味だが一般性の低さを考慮し、カット |
| cellphone | 採用 |
| snowboard | 採用 |
| download | 採用 |
| nonnative | Solid 少なくカット |
| serotonin | ESP性高くカット |
| boo | 間投詞もしくは peek a boo の一部なのでカット |
| sunglass | 採用 |
| someplace | 略式語なのでカット |
| reuse | 採用 |
| rainforest | 採用 |
| finalist | 採用 |
| babysit | 採用 |
| restroom | 採用 |
| loo | 間投詞なのでカット |
| anime | 採用 |
| midterm | Solid 少なくカット |
| bestseller | Solid 少なくカット |
| ballpark | 採用 |
| surfer | 採用 |
| | |

| | T |
|--------------|---------------------------------------|
| schoolwork | 採用 |
| recyclable | 採用 |
| brandnew | 採用 |
| descriptor | 採用 |
| medalist | 採用 |
| goodnight | Solid 少なくカット |
| fielder | 採用 |
| paycheck | 採用 |
| skateboarder | 不採用,skateboard は採用したので |
| sitter | baby_sitter の一部のためカット |
| coworker | 採用 |
| allele | 「対立遺伝子」。ESP性高くカット |
| erotica | 教育語彙表性をふまえてカット |
| semi | 独立語と判断せずカット |
| cam | Video cam とか cam recorder などの一部なのでカット |
| feller | fellow の視覚方言なのでカット |
| checkout | 採用 |
| genome | 採用 |
| empowerment | 採用 |
| embedded | 形容詞として採用 |
| yummy | 採用 |
| unfairly | 採用 |
| veggie | vegetarian の略語のためカット |
| windowsill | 採用 |

特に判定が微妙な語に関しては、個別に検討した。個別の語の判定に関しては、以下の通りである。

(1) DNA

DNA のような略語については、JACET 8000 の策定作業の際は、本表でも付表である「Plus 250」でも、一定の範囲内(Mr, Mrs, Dr, Ltd, No)を許容し、それ以外は原則的に削除した。しかし、例えばcm(センチメートル)は入っているが、CM(コマーシャル)はないなど、略語の扱いは必ずしも首尾一貫していない面もあった。

DNA については本表においても「Plus250」の編集時に候補語となったが、「ただし同じ2500位以内のDNAに関しては、専門性が高いと判断し、掲載を見送った」(JACET, 2003,

p.116)という経緯があり、今回も採用とはしなかった。

(2) Valentine

Valentine は、検定教科書での頻度は高いが、固有名詞なので最終候補からは省いた。

(3) joey (カンガルーの子)

教科書コーパスでカンガルーを扱ったものに集中して現れた単語であるが、最終候補からは省いた。

以上の手順を踏んで、最終的な JACET Supplement 636 を決定した。リストは、巻末の資料 5 の通りである。

(安城学園高校・清水伸一/神戸大学・石川慎一郎)

5.2 Supplement 636 の特徴

JACET 8000 の追加語として提案された語彙の特徴をつかむため、Range32(Nation, 2005)を使用して分析した。Range 32 では、BNC を Word Family に換算して 14,000 語までの頻度で分析することが可能である。

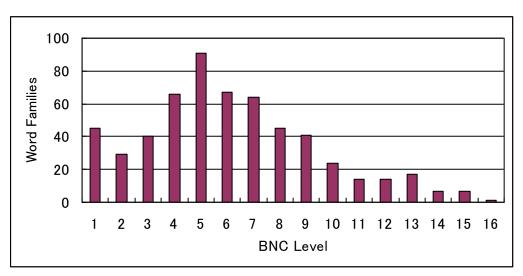


図 5-2-1 Range32 による Supplement 636 の分析結果

この図から明らかなように、Supplement 636 の候補となった語彙は、頻度が 9000 位までの語彙がほとんどで、最も多いのが頻度 3001 位から 7000 位まで語である。この分布を見る限り、この Supplement は JACET 8000 の語彙表の後半以降の順位を補っていると解釈することができる。

これらの分析結果の詳細は、表 5-2-1 の通りである。Type と Token の欄の数値が完全

に一致することから,異なり語数の率が 100%であることが読み取れる。また,この BNC の 14000 語(word families)のリストからも漏れる語が,50 語も存在することが明らかになり,非常に興味深い。

リストから漏れた50語は、以下の通りである。

anime, ballpark, bookstore, brandnew, cannonball, cellphone, cleanup, coworker, dative, diaper, email, envision, firsthand, graffito, grizzly, hometown, installment, lacrosse, lasagna, mailbox, meiosis, metastasize, mustache, nighttime, nonprofit, nonstop, orangutan, paralyze, paycheck, pediatric, playoff, psych, redo, rerun, restroom, roommate, schoolwork, snowboard, softball, sophomore, taco, takeoff, teeny, tine, tsunami, vend, watermelon, website, windshield, woof

表 5-2-1 Range 32 による Supplement 636 の分布

| WORD LIST | TOKENS/% | TYPES/% | FAMILIES |
|------------------|----------|----------|----------|
| one | 47/ 7.39 | 47/ 7.39 | 45 |
| two | 29/ 4.56 | 29/ 4.56 | 29 |
| three | 41/ 6.45 | 41/ 6.45 | 40 |
| four | 70/11.01 | 70/11.01 | 66 |
| five | 93/14.62 | 93/14.62 | 91 |
| six | 70/11.01 | 70/11.01 | 67 |
| seven | 64/10.06 | 64/10.06 | 64 |
| eight | 46/ 7.23 | 46/ 7.23 | 45 |
| nine | 42/ 6.60 | 42/ 6.60 | 41 |
| ten | 24/ 3.77 | 24/ 3.77 | 24 |
| 11 | 14/ 2.20 | 14/ 2.20 | 14 |
| 12 | 14/ 2.20 | 14/ 2.20 | 14 |
| 13 | 17/ 2.67 | 17/ 2.67 | 17 |
| 14 | 7/ 1.10 | 7/ 1.10 | 7 |
| 15 | 7/ 1.10 | 7/ 1.10 | 7 |
| 16 | 1/ 0.16 | 1/ 0.16 | 1 |
| not in the lists | 50/ 7.86 | 50/ 7.86 | ????? |
| Total | 636 | 636 | 572 |

リストから漏れた追加候補の単語には、以下のような特徴が見られる。 (1)日本語などの外国語から英語になった語 anime, lasagna, tsunami, taco

(2)時代を反映して使われるようになった語

科学技術 bookstore, brandnew, cellphone, email,

installment, mailbox, website, windshield

スポーツ ballpark, lacrosse, playoff, softball

社会経済 nonprofit

(東京電機大学・相澤一美)

- 6. 語彙テスト開発の事例研究
- 6.1 JACET 8000 語彙テストの開発 (事例研究 1)
- 6.1.1 はじめに

本論は、大学英語教育学会基本語改訂委員会が編纂した『大学英語教育学会基本語リスト(JACET 8000)』を元にした語彙テスト開発の意義と過程を論じるものである。

語彙の構成概念を構築する方法として、2 つのアプローチが提案されている。広がりと深さのアプローチ(Nagy & Herman, 1987)とサイズと構成のアプローチ(Meara, 1996)である。どちらも、広がりとサイズのように、学習者がどれくらい多くの語彙を知っているかという要素を取り入れている。このサイズという語彙の構成概念は、学習者の第二言語の熟達度と関係があることが明らかにされている。そして、サイズは、次のことから、語彙力を総合的に判断するもっとも信頼できる指標と考えられる。すなわち、語彙の構成を測定する妥当性・信頼性の高い方法が確立されていない、少数のサンプルから語彙の深さを測定するのは語彙全体を代表できない(Read, 2000)、それに対して、サイズテストは信頼できる語彙表を元に厳密な方法でサンプルを抽出した場合、学習者の得点から、その語彙全体を推察することができるからである(Meara & Wolter, 2004)。

語彙サイズを測定するテストは、チェックリスト・テスト (Meara & Buxton, 1987; Meara, 1992)がよく知られている。これは、頻度レベルごとに目標語と擬似語を混ぜて、1 語ずつ提示し、受験者は提示された語を知っている、または、知らないと回答する方式である。擬似語が含まれているため、知らない語に対して、知っていると回答する受験者は減点される。語彙レベルズ・テスト(Nation, 1990; 2001)は、語彙サイズ測定が目的で開発されたものではないが、語彙サイズ推定にも利用されている。2000 語、3000 語、5000 語、10000 語、学術語彙レベルの5つのレベルで、18 題ずつ英語の定義に合う語を選ぶ形式で、ワードファミリーによる語彙サイズを推定する。これを元にした発表語彙サイズテストも知られている(Laufer & Nation, 1999)。このテストは、文文脈で目標語の語頭の数文字が提示され、受験者は目標語の残りの部分を完成する形式で、受験者の発表的語彙サイズを推定するものである。

日本人英語学習者向けの語彙サイズテストとしては、日本人英語学習者のための語彙サイズテスト(望月,1998)、全英連語彙テスト(佐藤他,2005)、Mizumoto(2005)がある。望月(1998)は、日本語の単語に相当する英語を選択する形式で、北海道大学英語語彙表(園田,1996)に基づく 7000 語までの見出し語換算による語彙サイズを推定する。全英連語彙テストは、語彙レベルズ・テストの形式を採用し、日本語の訳語からそれに相当する英語を選ばせる 1000 語レベルを作り、最初級者の語彙サイズの測定を試みている。さらに、2000 語以上は語彙レベルズ・テストとリンクすることで、ワードファミリー換算による語彙サイズを推定する。Mizumoto(2005)は、標準語彙水準 12000(ALC、2001)に基づき、日本語に相当する英語を 4 択で選ばせる方式で、8000 語までの語彙サイズを推定する。

このように語彙サイズを測定するテストは数多く開発されているが、テスト方法や元に

する語彙表の違いから、推定語彙サイズに違いがあると考えられる。たとえば、Laufer(1998)は、語彙レベルズ・テストとその発表版の語彙サイズの違いを報告している。それによれば、イスラエルの 10 年生の受容的英語語彙サイズは 1900、発表的英語語彙サイズは 1700 ワードファミリーであるのに対して、11 年生はそれぞれ 3500、2550 ワードファミリーである。相澤(2005)は、語を知っているとすることの定義は語彙サイズテストごとに異なり、そのため、1 つのテストで算出される語彙サイズはそのテストに固有なもので、一般化できないと指摘している。

テスト方式の違いによる学習者の回答行動の違いを、Laufer & Goldstein (2004)と Laufer、Elder、Hill、& Congdon (2004)は語彙知識の強さの違いとしている。これらの研究は、テスト方式により学習者のもつ語彙知識の強さを測定できると仮定している。すなわち、受容と発表、再認と再生という2つの次元から、1)受容・再認、2)発表・再認または受容・再生、3)発表・再生という4つの方式のテストを作成し、この順に難易度が上昇するという仮説を実証している。そして、その難易度は学習者の習得した語彙知識の強さの違いであると主張している。これらの研究では、テスト形式による語彙サイズの違いには言及していないが、学習者の語彙知識の習得の強さに違いがあることを示している。すなわち、学習者がある語について発表・再生できる能力を持つならば、それより難易度が低い知識(発表・再認または受容・再生、受容・再認)も必ず持っているというものである。このサイズと強さのテストは、学習者の診断テストとしてだけでなく研究ツールとしても利用できる。

6.1.2 研究課題

本研究は、JACET 8000 を元語彙表にして、4つの異なる方式(和英記述、和英選択、英和選択)で語彙サイズを推定するテスト、JACET 8000 語彙テスト(J8VT)を開発することを目的とする。この4つの方式は、Laufer & Goldstein (2004)と同じもので、和英記述は発表・再生、和英選択は発表・再認、英和記述は受容・再生、英和選択は受容・再認に相当する。4つの方式で同じ学習者の語彙サイズに序列ができ、それが Laufer & Goldstein (2004)や Laufer et al (2004)の語彙知識の強さの違いを支持するものならば、学習者の語彙サイズの測定方法にある方向性を見出すことが可能になるかもしれない。

6.1.3 開発方法

6.1.3.1 和英・英和記述方式の開発

JACET 8000 の 5000 語までの 5 レベルから 25 語ずつを無作為に抽出した。抽出した 125 語に『JACET 8000 英単語』(相澤他,2005)から訳語を割り当てた。同じ方法で別の 125 語を抽出し、第 2 版とした。

版と和英・英和方式が選択でき、125 語を日本語または英語でコンピュータ画面上に提示し、その訳語を入力するコンピュータプログラムを開発した(図 6-1-1)。和英

記述方式では、日本語が提示され、与えられた文字で始まる、その意味を表す英単語を入力する。英和記述方式では、英語が提示され、その意味を日本語で入力する。

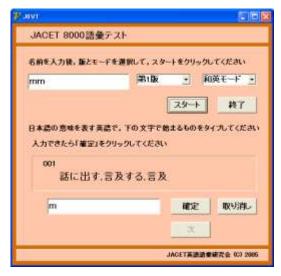


図 6-1-1 和英記述方式

図 6-1-2 英和記述方式

6.1.3.2 和英・英和選択方式の開発

和英・英和選択方式の開発には、和英記述方式、英和記述方式を用いて、パイロットスタディを行った。パイロットスタディの実施には、単語テストの選択肢を作成するためであると知らせ、英語を専攻とする大学生にテスト開発の協力を求めた。36名の学生が志願し、第1版あるいは第2版の和英方式、英和方式の順に制限時間を設けずに回答した。ほとんどの学生は和英・英和の2つの方式を90分で回答した。このパイロットスタディの回答結果は、2通りに利用した。1つは、誤答から正解の追加を決めるためである。たとえば、和英記述方式で、「認める、認知する」という日本語に対して、acknowledgeが正解としてプログラムされている。acknowledge 以外の回答は、誤答として処理される。パイロットスタディでは、以下のような回答がなされ、このうち admit は正解に近く、必ずしも誤りとはいえない。それゆえ、admit は正解としてプログラムに追加した。

55 認める,認知する accept(1) acknowledge(1) admit(5) allow(1) aquire(1) aquknowledge(1) attend(2)

パイロットスタディの回答結果の第 2 の利用法は、錯乱肢の作成である。Laufer & Goldstein (2004)は、受容・再認の錯乱肢の作成にあたり、目標語と同じ頻度レベルの同じ品詞の語で、意味的に関連しないものという基準を採用している。この基準によれば、目標語の意味を少しでも知っていれば正解を選択できることになるだろう。したがって、

わずかな語彙知識をも測定できることなる。しかしながら、この基準では、同じ意味場に 属する語が錯乱肢になった場合、目標語を正しく選択できるかどうかは定かではない。本 研究では、学習者の再認能力を意味的に近い選択肢の中から正解を選択できる能力と定義 し、パイロットスタディの被験者が誤答として回答したものを選択肢に利用した。たとえ ば、上に挙げた誤答例から、次のように選択肢を作成した。

55 認める,認知する acknowledge acquire allow attend

パイロットスタディの誤答の種類が選択肢を作成するに十分でない場合, 目標語と同じ文字で始まり, 同じ品詞の語を錯乱肢とした。

選択方式は、図 6-1-3、図 6-1-4 のように、和英であれば日本語、英和であれば英語が問題として表示され、その下に「わからない」を含めた5 つの選択肢が表示される。



図 6-1-3 和英選択方式



図 6-1-4 英和選択方式

6.1.3.3 回答方法

選択方式では、受験者は選択肢をクリックし、「次」ボタンを押すと回答が確定する。「次」を押さないかぎり、何度でも選択し直せる。記述方式では、受験者は回答を入力後、「確定」ボタンを押すと回答が確定し、「正解」または「不正解」が表示される。不正解の場合、1回のみ入力のし直しができる。

6.1.3.4 語彙サイズ推定

語彙サイズは、以下の式に基づいて推定した。

記述方式 正答数÷125×5000 選択方式 (正答数-(誤答数÷3))÷125×5000 「わからない」は、誤答には数えず、減点されない。

記述方式は、受験者自らが回答を入力するため、選択方式のようにまぐれ当たりする確率はきわめて小さいと考えられる。したがって、正答数を全問題数で割った値に、元となった語彙表の語数を掛けたものが、記述式の推定語彙サイズといえる。それに対して、選択方式では、「わからない」を除いた4つの選択肢の中の1つが正解なので、「わからない」以外を選択した場合、確率1/4で正解できる。選択方式の語彙サイズテストは、推定される語彙サイズが実際より大きいのではないかという懸念を持たれることがある。それゆえに、本研究では、あてずっぽうの修正式(Diamond & Evans,1973)

S = R - W/(k-1)

(Sは修正された得点, Rは正答数, Wは誤答数, kは選択肢数を表す)

に従って、選択方式の語彙サイズを推定する。上述のように「わからない」を選択した場合、あてずっぽうによる正答はありえないので、修正式には加えていない。

Hughes(2003)は、このような推測による正解を排除する方法は、2つの前提に基づいているとしている。1 つは、誤答はすべてあてずっぽうによるものであるという前提である。もう1つは、受験者は、等しい確率であてずっぽうにより正解するという前提である。この前提は必ずしも正しいとかかぎらない、したがって、あてずっぽうによる正解を除外した修正した得点は、真の得点を表しているわけではないと Hughes は述べている。

しかし、J8VTでは、誤答は減点されること、わからない場合は、「わからない」という 選択肢を選べば減点されないことをテストの最初に受験者に知らせている。したがって、 あてずっぽうによる正解を修正することで、真の語彙サイズに近いものを推定できるので はないかと考える。

6.1.4 J8VT の改良

杉森・相澤・望月(2005)は、J8VT の信頼性と 2 版の平行性の検証を目的に、関東圏の大学生と関西圏の大学生(43 名)を被験者にデータ収集を行ったが、J8VT に重大な欠陥が見つかったため、検証作業は行わず、テスト改良のデータとして利用した。具体的には、以下のような語が正解にもかかわらず、誤答として採点されてしまうので、正解となるようにプログラムを修正した。

英和記述

1. mention 言う

3. reply 返事する 14. boat 手漕ぎ船

15. forest 森

- 18. least 最も少なく
- 27. phrase 句
- 29. closely 密接に
- 37. formal 形式的な
- 38. media メディア
- 43. fairly かなり
- 44. forever 永遠
- 56. increasing 增加
- 58. deposit 頭金
- 66. empirical 実証的な
- 70. personnel 人事の
- 74. moving 移動している
- 80. supporting 支援
- 82. immense 莫大な
- 89. lesbian 女性の同性愛者
- 93 receptor 感覚器官
- 94. toxic 中毒の
- 95. dwelling 居住
- 97. damn 非難する
- 100. exploitation 開拓
- 102. coil 巻きつける
- 103. conspicuous 際だった
- 106. simplify 単純化する
- 118. disgusting 気分を悪くさせる
- 123. attempted 企てた
- 10. broad 幅の広い
- 11. wish 願望
- 70. personnel 社員の
- 79. hammer 金槌
- 92. cooperation 協同
- 94. toxic 有毒な
- 106. simplify 単純化する
- 118. disgusting いやになる

和英記述

- 76. 地主 landlord
- 92. 共同,協力 cooperation
- 93. 受容体, 受信機 receiver

これらの語が記述方式で入力された場合、正解となるようにプログラムの修正を行った。

6.1.5 J8VT の検証

6.1.5.1 検証の目的

改良した J8VT には、第1版と第2版という2つの版がある。さらに、それぞれの版に 和英記述、和英選択、英和記述、英和選択という4つの方式があるので、J8VT は総計8 つのテストから成る。J8VT に関して2つの検証を行う。最初は、8つの個々のテストが 信頼できるものであるという検証である。これは、8つのテストそれぞれで信頼性係数を 算出し、その値が高いものであれば、信頼できるテストであると考えることができる。

次に, 2 つの版が平行したテストになっているかという検証を行う。これは, 同一被験者グループが 2 つの版のテストを受験し, その平均点に有意の差がなければ, 2 つの版は平行であると考える。

6.1.5.2 被験者

被験者は、関東圏の 2 つの大学で志願者を募集した。総計 84 名の大学生・大学院生が被験者として J8VT を受験した。1 つの大学の被験者 19 名は、J8VT の 2 つの版を、和英記述、和英選択、英和記述、英和選択の 4 つの方式をこの順序で 8 つのテストを受験した。このグループを全受験グループとする。もう 1 つの大学の被験者 65 名は、無作為に 2 グループに分けられた。1 つのグループは、和英グループで、第 1 版和英記述、第 1 版和英選択、第 2 版和英記述、第 2 版和英選択の順番で 4 つのテストを受験した。もう 1 つのグループは、英和グループで、第 1 版英和記述、第 2 版英和記述、第 2 版 英和選択の順番で 4 つのテストを受験した。被験者には謝礼として図書カードを配布した。

6.1.5.3 方法

全受験グループには、受験のしかたを説明したファイルと J8VT のプログラムをメールで送付し、被験者が各自でテストを受け、結果のファイルをメールで戻すという方法でデータを収集した。したがって、8 つのテストを連続して受験したのか、数日に渡ってテストを受けたのか、1回にどれくらいの時間をかけたのかなどの詳細は明らかではない。

もう1つの大学では、被験者はコンピュータ教室に集まり、受験のしかたを説明した印刷物を受け取った。この印刷物は2種類あり、1つは和英グループ用で「第1版和英記述、第1版和英選択、第2版和英記述、第2版和英選択」の順番で受験するように指示している。もう1つは英和グループ用で、「第1版英和記述、第1版英和選択、第2版英和記述、第2版英和選択」の順番で受験するように指示している。どちらのグループも90分以内に4つのテストを終了した。

6.1.5.4 結果

結果は、8 つのテストごとに 3 つの受験者グループ(全受験、和英、英和)のデータを一緒にして表 6-1-1 に提示する。

表 6-1-1 第1版第2版の4つの方式の平均・標準偏差

| | 平均 | 標準偏差 | 平均語彙数 | 標準偏差 | 信頼性係数アルファ |
|---------|--------|-------|-------|------|-----------|
| 第1版英和記述 | 72.88 | 29.28 | 2915 | 1171 | 0.98 |
| 第1版英和選択 | 105.96 | 16.88 | 4238 | 675 | 0.96 |
| 第1版和英記述 | 64.51 | 28.12 | 2580 | 1125 | 0.98 |
| 第1版和英選択 | 106.45 | 13.13 | 4258 | 525 | 0.94 |
| | | | | | |
| 第2版英和記述 | 66.36 | 29.20 | 2654 | 1168 | 0.98 |
| 第2版英和選択 | 102.94 | 18.03 | 4118 | 721 | 0.97 |
| 第2版和英記述 | 61.20 | 27.25 | 2448 | 1070 | 0.98 |
| 第2版和英選択 | 104.63 | 14.34 | 4185 | 574 | 0.95 |

表 6-1-1 が示すように、2 つの版の 4 つの方式はいずれも 0.94 以上の高い信頼性係数を得ており、信頼性の高いテストであると言える。特に、記述方式のテストは、版や和英・英和という方式を問わず、いずれも 0.98 であり、きわめて信頼性が高いことがわかる。

次に、2つの版の平行性を検証するために、同一方式どうしで 2 つの版の t 検定を行った。その結果、和英記述、和英選択、英和記述、英和選択、いずれにおいても 2 つの版の平均には有意の差は見られなかった。このことは、2 つの版が平行であることを示していると考えることができる。

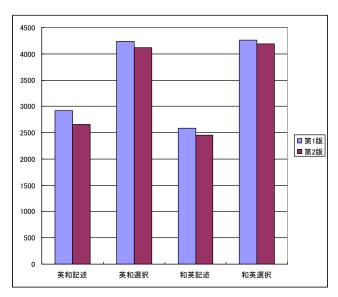


図 6-1-5 方式ごとの平均語彙サイズ

第 3 に、方式ごとの平均語彙サイズを見ると、第 1 版は英和記述 2915 語、和英記述 2580 語、英和選択 4238 語、和英選択 4258 語である。第 2 版は英和記述 2654 語、和英記述 2448 語、英和選択 4118 語、和英選択 4185 語である。どちらの版も英和・和英という産出する言語の違いよりも、記述か選択かというテスト方式の違いのほうが推定語彙サイズに大きな差をもたらしている。この違いが有意であるか調べてみるために、2 つの版の方式ごとに t 検定を行った(表 6-1-2、6-1-3)。

表 6-1-2 方式間の平均の差の有意性

| 第1版 | 英和選択 | 和英記述 | 和英選択 |
|------|--------|--------|---------|
| 英和記述 | p<.000 | n.s. | p <.000 |
| 英和選択 | | p<.000 | n.s. |
| 和英記述 | | | p <.000 |

表 6-1-3 方式間の平均の差の有意性

| 第2版 | 英和選択 | 和英記述 | 和英選択 |
|------|--------|--------|---------|
| 英和記述 | p<.000 | n.s. | p <.000 |
| 英和選択 | | p<.000 | n.s. |
| 和英記述 | | | p <.000 |

表 6-1-2, 表 6-1-3 が示すように, 2 つの版ともに産出させる言語の違いによる語彙サイズの差はないが, 記述か選択か, あるいは, Laufer & Goldstein (2004)と Laufer, Elder, Hill, & Congdon (2004)の言う再生か再認の違いによる語彙サイズには有意の差があることになる。

6.1.6 考察

結果で示したとおり、J8VT の 2 つの版は平行テストと考えることができる。このことは、2 つの版を同じテストとして結果を比較することができることを意味する。ある指導を行ったことの効果を語彙サイズの変化でみようとする研究を考えるならば、J8VT の第 1 版を事前テストとして、第 2 版を事後テストとして使用することが可能となる。また、学年始めに第 1 版、学年の終わりに第 2 版を実施して、1 年間の生徒の語彙サイズの変化を測定することもできる。このように J8VT の 2 つの版が平行テストであることの利用範囲は広い。

第2に、この研究は英和と和英の語彙サイズに違いがないという結果をもたらしたこと について考察する。従来、受容語彙は発表語彙の2倍と考えられているが、記述・選択と いう方式を統一すると、受容・発表語彙の差はないことになる。記述方式では、第1版は 和英(2580 語), 英和(2915 語), 第 2 版では和英(2448 語), 英和(2654 語)である。それに 対して, 選択方式では, 第1版は和英(4258語), 英和(4238語), 第2版では和英(4185語), 英和(4118語)である。今回の結果についてみると、英和選択による語彙サイズは、和英記 述による語彙サイズの 1.64 倍(第 1 版),1.68 倍(第 2 版)である。この数値は,従来考えら れている受容語彙と発表語彙の量的相違に近いものである。このように考えると、受容語 彙と発表語彙の相違とは、英和と和英の相違ではなく、再認と再生の違いと考えるほうが 妥当かもしれない。すなわち、選択方式による語彙を受容語彙、記述方式による語彙を発 表語彙とするならば,和英選択は英和記述の語彙サイズの 1.46 倍(第 1 版),1.58 倍(第 2 版)となり、英和選択と和英記述の比率と近いものになる。しかしながら、本研究は被験者 が84名と少ないものであり、しかも2つの大学の大学生・大学院生を対象としたもので あり,一般化できるものであるとは考えられない。受容語彙と発表語彙の違いについては, さまざまなレベルの学習者からなる多数の被験者が受験した結果にもとづいて議論すべき であろう。

第3に、本研究の結果は、J8VTが語彙サイズテストとして、2つの版、4つの方式で実施するテストが信頼できるものであることを示している。これは語彙サイズの測定が安定して行われることを意味するもので、語彙研究者にとって利用可能で役立つ道具が増えたことになる。しかしながら、本研究はJ8VTの妥当性検証は行っていない。他の語彙サイズテストとの併存的妥当性検証がまず考えられる。その他に、語彙サイズが異なると思われる中学生、高校生、大学生など英語熟達度の異なるレベルの学習者を被験者として、J8VTの結果がその違いを反映するかを調べるような妥当性の検証を行う必要がある。

6.1.7 まとめ

本研究は、JACET 8000 を元にした語彙サイズテスト(J8VT)を開発した過程を記述し、 テストの検証を行った。2 つの版と 4 つのテスト方式をもつ J8VT は信頼性の高いテスト であること、2 つの版は平行であることが実証された。妥当性については、今後の検証を 待たなければならない。

(麗澤大学・望月正道)

参考文献

- Diamond, J. & Evans, W. (1973). The correction for guessing. Review of Educational Research, 43, 181-191.
- Hughes, A. (2003). Testing for language teachers second edition. Cambridge: Cambridge University Press.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19, 255-271.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. Language Testing, 16, 33-51.
- Laufer, B., Elder, C., Hill, K, & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, 21, 202-226.
- Laufer, B. & Goldstein, A. (2004). Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning*, 54, 399-436
- Meara, P. (1992). *EFL vocabulary tests*. Centre for Applied Language Studies. University College Swansea.
- Meara, P. (1996). The dimension of lexical competence. In G. Brown, K. Malmkjaer,
 & J. Williams (eds), Performance and competence in second language
 Acquisition. Cambridge: Cambridge University Press.
- Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. Language Testing, 4, 142-151.
- Meara, P. & Wolter, B. (2004). V_Links: beyond vocabulary depth. *Angles on the English-Speaking World*, 4, 85-96.
- Mizumoto, A. (2005). Developing another type of vocabulary size test. JACET English Vocabulary Group Second Conference on Researching, Learning and Teaching Second Language Vocabulary, December 10th, Chuo University Korakuen Campus.
- Nagy, W.E., & Herman, P.A. (1987). Breadth and depth of vocabulary knowledge: implications for acquisition and instruction. In McKeown, M.G. & Curtis, M.E. (eds), *The Nature of vocabulary acquisition.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, 19-35.

- Nation, P. (1990). Teaching and learning vocabulary. Heinle and Heinle.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Read, J. (2000). Assessing vocabulary. Cambridge: Cambridge University Press.
- 相澤一美 (2005). 英語語彙習得研究の疑問点と今後の課題 学習者の受容語彙知識はどうなっているか . 第 44 回 JACET 全国大会シンポジウム英語語彙研究会企画「英語語彙研究の新展開」, 2005 年 9 月 9 日, 玉川大学.
- 相澤一美・石川慎一郎・村田年(編著) (2005). 『JACET 8000 英単語』東京:桐原書店. アルク (2001). 『標準語彙水準 12000』東京:アルク.
- 佐藤留美・石井 亨・篠原洋二・佐野義晴・山内 豊 (2005).リスニング語彙とリーディング語彙を測定・比較できるオンライン語彙レベル判別テストの開発. JACET 英語辞書研究会・英語語彙研究会合同ワークショップ「英語の辞書・語彙」研究発表, 2005年3月26日, 東京電機大学.
- 杉森直樹・望月正道・相澤一美 (2005). JACET 8000 に基づく CBT 方式の語彙サイズテスト開発について. JACET 英語語彙研究会第 2 回研究大会, 2005 年 12 月 10 日, 中央大学後楽園キャンパス.
- 園田勝英 (1996). 『大学生用英語語彙表のための基礎的研究』言語文化部研究報告叢書 7、 北海道大学言語文化部.
- 大学英語教育学会基本語改訂委員会 (2003). 『大学英語教育学会基本語リスト JACET 8000)』大学英語教育学会(JACET).
- 望月正道 (1998). 日本人英語学習者のための語彙サイズテスト. 『語学教育研究所紀要』 第 12 号, 27-53.

6.2 JACET 8000 に基づく大学生受容語彙サイズテストの開発と検証 (事例研究 2)

6.2.1 はじめに

JACET 8000 は、実データに基づいて切り出されたレベル別語彙表であり、8000 が頻度ベースで並んでいるという点で、学生の受容語彙力を評価する基準としても活用しやすい。本稿では、JACET 8000 をベースとして論者が作成した「大学生受容語彙サイズテスト」の概要を紹介し、被験者となった大学生の語彙サイズの一端を報告する。また、同テストから予測される学生の語彙サイズと、TOEIC®スコアの相関を調査し、受容語彙サイズテストの有効性について考えたい。

6.2.2 語彙テストの開発

語彙テストには、さまざまな形式がある。石川(2003)でも概観したように、代表的なものとしては、英語を見て該当する日本語を選ばせる $L2\rightarrow L1$ 型(英和型)、日本語を見て該当する英語を選ばせる $L1\rightarrow L2$ 型(和英型)、日本語を介在させずに、英単語の意味を英文定義から選ばせたり、英文の空所を埋めさせたりする $L2\rightarrow L2$ 型(英英型)などがある。

このうち, 日本において広く使われているのは, L1→L2 型の望月 (1998) である。 いわゆる望月 テストは, 下記のような形態をとる。

1. 小麦粉を焼いた菓子

2. 集まり、会

(1) birthday

(2) cookie

(3) fork (4) party

(5) star (6) sweater

望月テストは、Nation(2001)の形式にも似て、2 間ずつが一緒に出題されるため、錯乱肢の数が少なくできる。また、日本語から英語を探すというスタイルをとることで、ある程度、発信的要素も加味されている。これらの点から、同テストは、日本人対象の語彙テスト形式としてすぐれた点が多い。

ただし、本節では、狭義の「受容語彙知識」を測定しようとする狙いから、望月テストを一部変更し、英語を提示して日本語を答えさせるスタイルを採用した。加えて、実施の容易性を考え、問題数の圧縮を行った(石川 2003、2005)。テストは JACET 8000 の 8 レベルに対してそれぞれ作成され、1 レベルの問題数は 20 題とした。なお、以後の研究で、1 レベル 10 題のバージョンも作成したが、本節では 20 題バージョンについて報告する。実際のテストの一部 (レベル 3)を下記に示す。

1. brief 2. grand

(1)言語の (2)生の (3)生きた (4)簡潔な (5)神聖な (6)大きい

出題された語は、元リストからランダムに選択したうえで、人口に膾炙したカタカナ語になっている ものや、より平易な語成分の合成語になっているものについては差し替えを行った。錯乱肢は、 当該問題の前後の同一品詞語から採取した。

レベル1から8まで,8つのテストを受けた場合,テストされる語数は160語となる.これらは元リストの8,000語を代表する標本であり,1問の正答は50語の受容語彙力の保持を示すと考えられる。ゆえに,素点に50倍を乗することで,およその語彙サイズを測定することが出来ると考えられる。

6.2.3 調査方法

開発したテストを使用して、国立大学の1回生(文系・理系各学部ほぼ均等)100名を対象に試行テストを行った。その結果、レベル1の平均正答率は99.2%、レベル2の平均正答率は97.6%であることが分かった。レベル1~2の問題は点双列が0.1以下となり、これらのデータは総得点との関連度が低く、テストアイテムとしては弁別力が非常に低いことがわかった。

つまり、レベル 1~2 については、時間をかけてテストを行っても、そこから得られる情報量は多くないことになる。そこで、全体のテスト時間を圧縮するため、実際のテストはレベル 3~8 の合計120 問に対して行うこととした。なお、レベル 1~2 については、20 点満点に対して、それぞれ 20 点、19 点をみなし点として付与し、これらとレベル3以上の実際の得点の合計を 50 倍した数値をもって、受容語彙サイズを算出した。

正式なテストは、2005 年 10 月に行われた学内 TOEIC-IP テストの直前に実施された。被験者は、事前に受容語彙サイズテストの受験および得点データの研究目的での分析に応諾した 72 名である。全員が上記と同じ国立大学の 1 回生(文系・理系学部比は約4:6)である。

6.2.4 結果と考察

◆学習者の学習状況とJACET 8000 レベルは相関しているか?

頻度準拠のJACET 8000 による語彙レベルが、平均的な日本人大学生の語彙学習状況から見る「難易度」とどの程度関連しているかは興味深い問題である。下記の表は、レベルごとの平均正答数を出したものである。

表 6-2-1 レベルごとの得点状況

| レベル | Lev3 | Lev4 | Lev5 | Lev6 | Lev7 | Lev8 |
|-------|------|------|------|------|------|------|
| 平均正答数 | 16.8 | 10.3 | 9.3 | 8.8 | 7.3 | 5.8 |
| 標準偏差 | 1.82 | 3.05 | 3.69 | 3.14 | 3.35 | 3.19 |

上記をグラフ化したのが図 6-2-1 である。図に明らかなように、平均正答数はレベルの上昇にしたがって下降し、JACET 8000 のレベルが調査対象大学生にとっての実際的な難易度とおよそ関連していることが明らかになった。なお、ここで注目すべきはレベル 3 とレベル 4 の間に大きな差

が生じていることである。レベル 1~3 は平均正答率が 90%以上で安定しているが、レベル 4 になると平均正答率は 50%程度に落ち込み、以下、レベル 8 にかけて 30%程度まで低下する。

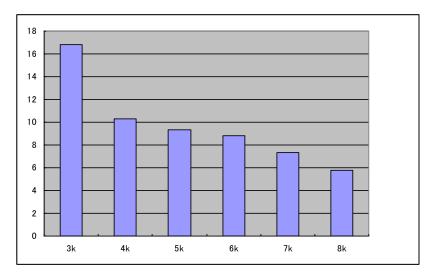


図 6-2-1 レベル別正答数

標準偏差に注目すると、レベル3に比べ、レベル4以上の標準偏差が総じて高く、とくに、レベル5の標準偏差が顕著に高いことがわかる。標準偏差の高さは、被験者の弁別力が高いというポジティブな可能性と、問題としての精錬度が低く、確信に基づかないランダムな回答行動が見られたというネガティブな可能性が存在する。この点については今後の検証が必要であろう。

次に、隣接するレベルごとの平均値に差が存在するかどうかを確かめるため、対応あり t 検定を行ったところ、基本的に隣接レベル間に有意差はあるが、レベル 5 と 6 の間には有意差が検出されなかった。このことは、対象被験者に限って言えば、6 つのレベルが、3>4>5&6>7>8 という状況にあることを示す。

| 表 | 6- | -2-2 | レベル差 |
|---|----|------|------|
| | | | |

| 比較レベル | Lev 3/4 | Lev 4/5 | Lev 5/6 | Lev 6/7 | Lev 7/8 |
|-----------|---------|---------|---------|---------|---------|
| 対応あり t 検定 | t=18.7 | t=2.1 | t=1.05 | t=3.96 | t=4.00 |
| p(両側) | p<0.001 | p<0.05 | n.s. | p<0.001 | p<0.001 |

◆総語彙数は TOEIC スコアと相関しているか?

本テストから求められた推定語彙サイズは、近年、大学英語教育の外部指標として多用されている TOEIC スコアとどの程度相関しているであろうか?この点を検証するため、推定語彙サイズと TOEIC 総得点、および、リスニングセクション得点、 リーディングセクション得点との間でピアソンの積率相関係数を算出した。

なお,被験者 72 名の平均語彙サイズは 4862 語(標準偏差 629.2), TOEIC 平均総得点は

544.4(標準偏差 113.5), リスニングセクション平均は 283(標準偏差 58.0), リーディングセクション平均は 261(標準偏差 65.1)である。下図は、総得点と語彙サイズの関係を示した散布図である。

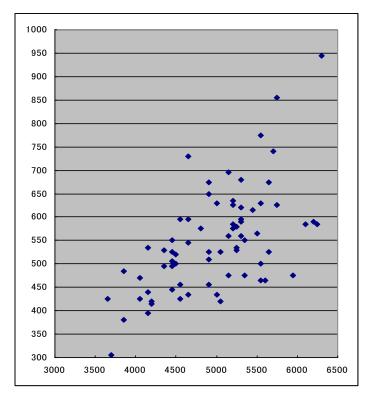


図 6-2-2 語彙サイズと TOEIC 総得点の散布図

次に、TOEIC の各得点と語彙サイズの相関値を示す。

表 6-2-3 語彙サイズと TOEIC 各セクション得点の相関

| | VOC/TOTAL | VOC/Listening | VOC/Reading |
|---|-----------|---------------|-------------|
| R | 0.600 | 0.422 | 0.667 |

語彙力が,英語力全体を要約する指標として有益であることはしばしば指摘されているが,本研究においても,語彙サイズは,TOEIC の総得点と.060 の相関を見せており,かなり強い相関が確認された。

とくにリーディングセクションとは.667 の相関が得られており、本研究で使用した語彙サイズテストは、TOEIC のリーディングセクション得点を予測する一定の目安として使用できるかもしれない。下記は、x軸に語彙サイズ、y軸にTOEICリーディングセクションスコアを散布図としてプロットし、線形近似曲線を追加したものである。

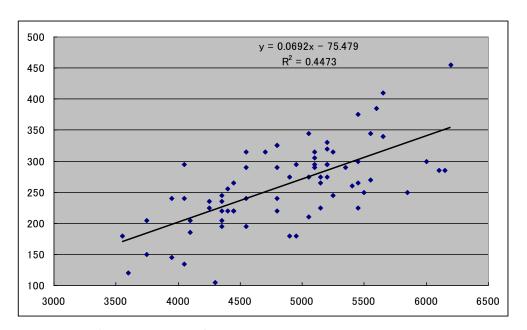


図 6-2-3 語彙サイズとリーディングセクション得点

単回帰分析を行うと,下記の式が得られる。

y=0.0692x-75.479 (変数 x: 語彙サイズ, y:リーディングセクションスコア)

もっとも、両変数間の.667という相関係数は、目的変数に対する説明力が44%であることを示しており(重相関係数 R2=0.4473)、このままでは、TOEIC スコアを予測する指標としての信頼度には限界がある。語彙テスト問題の内容の精選、あるいは、問題量の再検討によって、少なくとも.80程度の相関値を確保することが今後の課題となる。

◆テスト改良のための視点:より実施が容易で、より信頼性の高いテストとは?

上で触れたように、今後のテストの改良に当たっては、問題の精選が不可欠となる。そのためには、説明変数としての各レベル得点と、目標変数としてのTOEIC リーディング得点との相関値をとり、相関の高いものと低いものに峻別する作業が効果的である。

そこで、レベル $3\sim8$ の各得点、および、レベル $3\sim4$ 、レベル $3\sim5$ 、レベル $3\sim6$ 、レベル $3\sim7$ の組み合わせを用意し、目標変数との相関値の変動を調査した。

表 6-2-4 各レベルとの相関値

| レベル | 3 | 4 | 5 | 6 | 7 | 8 | 3-4 | 3-5 | 3-6 | 3-7 |
|-----|------|------|------|------|------|------|------|------|------|------|
| R | 0.49 | 0.45 | 0.49 | 0.48 | 0.46 | 0.41 | 0.56 | 0.61 | 0.66 | 0.66 |

以上より、レベル別に見ると、レベル3,5,6の相関値が高く、レベル4,7の相関値がそれに次ぎ、レベル8の相関値はかなり低いことがわかった。このことは、テストアイテムを減らす場合には、レベル8を減らすのが妥当であることを示唆している。

また、対象レベル中、もっとも基礎的で、かつ、目標変数との相関値が高いレベル3を機軸として、複数レベルを組み合わせたところ、相関値は.56 から.66 まで変動することが観察された。とくに、レベル 3~6 の4レベル計 80 題に限定した場合の相関値(.66)は、3~8 まで総計 120 題を課した場合の相関値(.67)とほとんど違わず、テストとして非常に効率性が高いことが明らかになった。

以上より、当該被験者を対象にした語彙サイズテストの改善の方向としては、レベル 3~6 に限定した上で、それぞれの問題数を増やすことが考えられる。これにより、現在以上の相関を確保することが期待される。

6.2.5 まとめ

以上,本節では,JACET 8000 に基づく受容語彙サイズテストの開発・実施結果について報告してきたが,本研究によって明らかになった知見を下記にまとめる。

- ・JACET 8000 の語彙レベルが上がるにつれて、被験者の平均得点が下がるという傾向が確認され、JACET 8000 の語彙レベルは、一定程度の学力を持った大学生の語彙サイズを測定する基礎資料として妥当性が高い
- ・レベル3と、レベル4以降に大きな得点差が観察され、当該被験者が今後体系的な語彙 学習を行う場合はレベル4を起点にすることが望ましいと判断される。
- ・レベル 5~6 には平均得点差が認められず、このレベルの語彙の定着は一様ではないと推定される
- ・語彙サイズテスト結果は、TOEIC 総得点と.60、TOEIC リーディングセクション得点と.67の相関を示し、かなり強い相関が確認できる
- ・語彙サイズテスト結果と TOEIC リーディングセクション得点間に一定の回帰式を求めることが可能である
- TOEIC リーディングセクションとの相関の点で各レベルを比較すると、レベル3,5,7の相関は高く、レベル8の相関は低い
- ・レベル 3~8 のすべてを測定した場合と、レベル 3~6 に限って問題数を 2/3 にして測定した場合の相関はほとんど変わらない
- ・レベル3~6に特化した語彙テストを開発することでより高い相関の確保が期待される

(神戸大学・石川慎一郎)

7. 研究の総括

7.1 研究の限界

本プロジェクトでは、3 年間の研究期間内に、学習語彙表の作成と教材分析システムの構築を最終的なミッションに掲げ、研究に邁進してきた。まず、前回作成した JACET 8000 を学習語彙表として検証し、その欠点を探ってきた。しかし、当初の予想以上に JACET 8000 の完成度が高いことがわかった。いくつか行った JACET 8000 の妥当性の検証の研究では、特に大きな問題点を発見することはできなかった。

その結果を踏まえて、2 年目途中から路線を転換し、検定教科書、口語、米語、児童英語に基づくサブコーパスをそれぞれ構築し、その中から JACET 8000 に漏れてしまった語を supplement として提案することになった。この軌道修正のために、語彙表作成までの研究期間は1年半余りの猶予しかなくなってしまった。

しかし、時間的な制約に加えて supplement の作成は、予想以上に困難な作業であった。新しく構築したサブコーパス(4 種類)のそれぞれについて、Perl スクリプト v8an.pl を用いて、JACET 8000 をベースとしたタグ付きの出現頻度順リストを出力した。非語、固有名詞等を排除した各リストを 100 万語換算で頻度補正した上で、4 リストをマージし、相対頻度とサブコーパス間のレンジ値の基準を 2 以上として JACET 8000 の supplement 候補語 636 語を提案するに至った。

以上のような軌道修正のため、教材分析システムの開発という本プロジェクトの研究課題が、未解決のまま残ってしまった。もう少し時間があれば、清水伸一氏に協力を依頼し、JACET 8000、Plus 250、Supplement 636 の 3 つのリストを合わせた形で、教材分析システムを構築することが可能であったと思われる。当面の間は、JACET 8000 と Plus 250 にもとづいた清水氏の Level Maker を代用して、教材の分析をするしかない。

また、今回発表した Supplement 636 も、十分な検証が済んだわけではない。場合によっては、今回の候補語と JACET 8000 の語を統合して、リストを再構築することも視野に入れることが考えられる。

7.2 今後の課題

本プロジェクトでは、検定教科書、口語、米語、児童英語の4分野のサブコーパスを構築することができた。これらのサブコーパスは、今後の語彙指導の研究や実践のための資料として大いに活用できると期待される。そこで今回の研究成果を、大学英語教育学会英語語彙研究会のサーバーから発信し、今後の研究のために役立ててもらえるような環境を整える予定である。

さらに、研究プロジェクトで購入した大型サーバーや大規模コーパスなどの資源などは、 今後の継続研究にも活用できる。公務や他の研究活動などで、当初の予定通りのエフォートが十分にできなかったことも反省としてある。 (東京電機大学・相澤一美)

謝辞

本プロジェクト推進にあたり,以下の教科書出版社からテキストデータをご提供いただきました。深く感謝申し上げます。

株式会社池田書店

株式会社旺文社

株式会社開拓社

学校図書株式会社

株式会社桐原書店

新興出版社啓林館

数研出版株式会社

株式会社増進堂

株式会社第一学習社

株式会社大修館書店

光村図書出版株式会社

(敬称略, 五十音順)

教科書発行者の略号と社名(本稿中のリストにおける掲載順)

啓林館 株式会社新興出版社啓林館

東 書 東京書籍株式会社

大修館 株式会社大修館書店

三省堂 株式会社三省堂

池 田 有限会社池田書店

開隆堂 開隆堂出版株式会社

桐 原 株式会社桐原書店

教 出 教育出版株式会社

增進堂 株式会社増進堂

開 拓 株式会社開拓社

三 友 三友社出版株式会社

開隆堂 開隆堂出版株式会社

旺文社 株式会社旺文社

数 研 数研出版株式会社

文英堂 株式会社文英堂

第 一 株式会社第一学習社

光 村 光村図書出版株式会社

秀文館 株式会社秀文館

学 図 学校図書株式会社

文部科学省の検定教科書に関する情報

http://www.mext.go.jp/a menu/shotou/kyoukasho/mokuroku/14/hakkou.htm

資料 1. NEXT 1000 Database

American English 1000 語

Spoken English 1168 語

英米児童リーダー 1134 語

検定教科書 944 語

(掲載は頻度順)

| | WORD |
|----|-------------------|
| 1 | email |
| 2 | Z00 |
| 3 | astronaut |
| 4 | cookie |
| | playoff |
| 6 | panda |
| 7 | forth |
| 8 | scary |
| 9 | pitcher |
| 10 | teammate |
| 11 | premiere |
| 12 | fave |
| 13 | pacific |
| 14 | skater |
| 15 | upcoming |
| 16 | capitol |
| 17 | pumpkin |
| 18 | trash |
| 19 | spacecraft |
| 20 | patriot |
| 21 | tornado |
| 22 | sled |
| 23 | skateboard |
| | phoenix |
| 25 | download |
| 26 | firefighter |
| 27 | surf |
| 28 | alligator |
| 29 | rebound |
| 30 | website |
| 31 | backpack |
| 32 | touchdown |
| 33 | clone |
| 34 | superstar |
| 35 | Valentine |
| 36 | rookie |
| 37 | DNA |
| 38 | lawmaker |
| 39 | max |
| 40 | hometown |
| 41 | tsunami |
| 42 | someday |
| 43 | |
| 44 | hike |
| | steroid musher |
| 46 | |
| | creek |
| 47 | cooperate |
| 48 | homer |
| 49 | classmate |
| 50 | birthplace |
| | |

| Rank | WORD |
|----------|-------------|
| 51 | closet |
| 52 | babe |
| | audition |
| 54 | burger |
| 55 | buffalo |
| 56 | backyard |
| 57 | cooperation |
| 58 | carol |
| 59 | makeup |
| 60 | snowboard |
| 61 | freshman |
| 62 | hoop |
| 63 | semester |
| 64 | gorilla |
| 65 | ranch |
| 66 | sidewalk |
| | mammoth |
| 68 | quake |
| 69 | wrestling |
| 70 | titan |
| 71 | canyon |
| 72 | squid |
| 73 | contestant |
| 74 | skier |
| 75 | cafeteria |
| 76 | insurgent |
| 77 | recount |
| 78 | subway |
| 79 | J |
| 80 | softball |
| 81 | sniper |
| 82 | courtroom |
| 83 | anthrax |
| 84 | artifact |
| 85 | sequel |
| 86 | volleyball |
| 87 | lineup |
| 88 | oversee |
| 89 | columnist |
| 90 | sitcom |
| 91 | videotape |
| 92 | flier |
| 93 | dough |
| 94 | chili |
| 95 | gamer |
| 96 | obesity |
| 97 98 | duo |
| | slay |
| 100 | sneaker |
| 100 | derby |

| Rank | WORD |
|------|----------------|
| 101 | dodge |
| 102 | lemonade |
| 103 | alphabet |
| 104 | anytime |
| 105 | gymnastics |
| 106 | accomplishment |
| 107 | hitter |
| | autograph |
| 100 | lacrosse |
| | birthdate |
| 111 | longtime |
| 112 | baseman |
| 113 | gasoline |
| 113 | spike |
| | |
| | semifinal |
| | trillion |
| | checkpoint |
| | kitty |
| 119 | |
| 120 | thunderstorm |
| 121 | bookstore |
| 122 | nickel |
| 123 | kindergarten |
| 124 | wildfire |
| | racer |
| 126 | |
| 127 | casino |
| 128 | asteroid |
| 129 | soundtrack |
| 130 | oops |
| 131 | |
| 132 | teamwork |
| 133 | allergy |
| 134 | catcher |
| 135 | pundit |
| 136 | blockbuster |
| 137 | rapper |
| 138 | crayon |
| | whoa |
| 140 | mosque |
| 141 | swimmer |
| 142 | popcorn |
| 143 | plaza |
| 144 | rotate |
| | psych |
| | mentor |
| | fundraising |
| | trek |
| | blog |
| 150 | finalist |
| 100 | THATISE |

| Rank | WORD |
|------------|-------------|
| 151 152 | fin |
| 152 | peek |
| 153 | atop |
| 154 | slugger |
| 155 | heck |
| 156 | doughnut |
| 157 | freestyle |
| 158 | dunk |
| 159 | caucus |
| 160 | filmmaker |
| 161 | grader |
| 162 | downhill |
| 163 | peel |
| 164 | photon |
| 165 | reed |
| 166 | televise |
| 167 | mascot |
| 168 | vanilla |
| 169 | nap |
| 170 | zip |
| 171 | shortstop |
| 172 173 | serotonin |
| 173 | educator |
| 174 | boo |
| 175 | scooter |
| 176 | mini |
| 177 | maneuver |
| 178 | dude |
| 179 | cheetah |
| 180 | lark |
| 181 | scorer |
| 182 | creepy |
| 183 | tug |
| 184 | hijack |
| 185 | nonprofit |
| 186 | awhile |
| 187 | slime |
| 188 | precipitate |
| 189 | moose |
| 190 | tyrannosaur |
| 191 | blackout |
| 192 | mule |
| 193 | enchant |
| 194 | tycoon |
| 195 | foil |
| 196 | роор |
| 197 | excerpt |
| 198 | rocker |
| 199 | disarm |
| 200 | neutron |
| | |

| Rank | WORD |
|------------|---------------------|
| 201 | dribble |
| 202 | pancake |
| 203 | tattoo |
| 204 | lunar |
| 205 | bamboo |
| 206 | slalom |
| 207 | reef |
| 208 | boulder |
| 209 | forecaster |
| 210 | rainy |
| 211 | gadget |
| 212 | fundraiser |
| 213 | ups i de |
| 214 | bracelet |
| 215 | teaspoon |
| 216 | skeptical |
| 217 | reunite |
| 218 | handler |
| 219 | tar |
| 220 | canine |
| 221 222 | spokesperson |
| 222 | hilarious |
| 223 | coyote |
| 224 | preschooler |
| 225 | antiwar |
| 226 | tester |
| 227 | inaugural |
| 228 | beverage |
| 229 | sideline |
| 230 | memorize |
| 231 | pickle |
| 232 | beaver |
| 233 | skinny |
| 234 | cleanup |
| 235 | kinless · |
| 236 | pajama |
| 237 | panther |
| 238 239 | affirmative |
| 239 | goo |
| 240 | hamster |
| 241 | cicada |
| 242 | pooch |
| 243 | syndicate |
| 244 | moderator |
| 245 246 | camper |
| | hippopotamus |
| 247 | hamburger |
| 248 | handout |
| 249 250 | diva + rodomović |
| 200 | trademark |

| Rank | WORD |
|------------|----------------|
| 251 | paleontologist |
| 251 252 | superhero |
| 253 | loot |
| 254 | indict |
| 255 | meteor |
| 256 | shooter |
| 257 | sophomore |
| 258 | commemorate |
| 259 | hubby |
| 260 | |
| 261 | footage |
| 262 | nonstop |
| 263 | grizzly |
| 264 | postseason |
| 265 | halftime |
| 266 | |
| 267 | |
| 268 | itch |
| 269 | bash |
| 270 | outfielder |
| 271 | jumper |
| 272 | carnival |
| 273 | toddler |
| 274 | inauguration |
| 275 | sprout |
| 276 | bipartisan |
| 277 | allergic |
| 278 | homemade |
| 279 | sleepover |
| 280 | |
| 281 | veterinarian |
| 282 | flop |
| 283 | offseason |
| 284 | |
| 285 | hack |
| 286 | infinity |
| 287 | someplace |
| 288 | roster |
| | magnet |
| 290 | auditorium |
| 291 | replay |
| 292 | sunglass |
| 293 | quilt |
| 294 | |
| 295 | marvel |
| 296 | tidbit |
| 297 | giraffe |
| 298 | crunch |
| 299 | |
| 300 | filibuster |

| Rank | WORD |
|------------|--------------|
| 301 | installment |
| 302 | prank |
| 303 | goody |
| 304 | rabbi |
| 305 | baton |
| 306 | portrayal |
| 307 | snowball |
| 308 | |
| 309 | leftover |
| 310 | skunk |
| 311 | endurance |
| 312 | Linebacker |
| 313 | bristle |
| 314 | limousine |
| 315 | whoop |
| 316 | peacekeeping |
| 317 | frenzy |
| 318 | assassinate |
| 319 | envision |
| 320 | makeover |
| 321 | rodent |
| 322 | mayonnaise |
| 323 | kangaroo |
| 324 | showcase |
| 325 | workout |
| 326 | acclaim |
| 327 | vend |
| 328 | mogul |
| 329 | dormitory |
| 330 | juggle |
| 331 | shred |
| 332 | courthouse |
| 333 | replica |
| 334 | furry |
| 335 | goofy |
| 336 337 | wacky |
| 337 | shovel |
| 338 | interception |
| 339 | ecosystem |
| 340 | graduation |
| 341 | handhe I d |
| 342 | aspire |
| 343 | |
| 344 | symbolize |
| 345 | alumnus |
| 346 | mailbox |
| 347 | ballpark |
| 348 | slug |
| 349 | hub |
| 350 | trivia |
| | |

| Rank | WORD |
|------|--------------|
| 351 | hijacker |
| 352 | ketchup |
| 353 | rescuer |
| 354 | horoscope |
| 355 | countdown |
| 356 | preschool |
| 357 | marshmallow |
| 358 | interstate |
| 359 | unleash |
| 360 | pastor |
| 361 | sparrow |
| 362 | vomit |
| 363 | roommate |
| 364 | goldfish |
| 365 | operative |
| 366 | matchup |
| 367 | starlet |
| 368 | flashlight |
| 369 | diaper |
| 370 | darn |
| 371 | sneeze |
| 372 | remake |
| 373 | pharaoh |
| 374 | festivity |
| 375 | hoof |
| 376 | crave |
| 377 | critter |
| 378 | bathtub |
| 379 | flicker |
| 380 | blizzard |
| 381 | nugget |
| 382 | midterm |
| 383 | syrup |
| 384 | medalist |
| 385 | craze |
| 386 | billionaire |
| 387 | prep |
| 388 | felony |
| 389 | helm |
| 390 | zebra |
| 391 | entertainer |
| 392 | barber |
| 393 | levee |
| 394 | surfer |
| 395 | cheer leader |
| 396 | driveway |
| 397 | brag |
| 398 | iceberg |
| 399 | healthcare |
| 400 | |
| | • |

| Rank | WORD |
|------|---------------|
| | homicide |
| 402 | blab |
| 403 | surpass |
| 404 | snag |
| 404 | pedal |
| 400 | yoghurt |
| 400 | yognur t |
| 407 | bedtime |
| 408 | husky |
| 409 | breakup |
| 410 | ZOOM |
| | gee |
| 412 | |
| 413 | implant |
| 414 | hottie |
| | hind |
| 416 | detainee |
| 417 | a. k. a |
| 418 | meteorologist |
| | mash |
| 420 | robotic |
| 421 | heavyweight |
| 422 | pup |
| 423 | kenne l |
| 424 | crank |
| 425 | navigate |
| | funky |
| 427 | weekday |
| | intern |
| 429 | anonymity |
| 430 | canister |
| 431 | rouge |
| 432 | revolve |
| 433 | tout |
| 434 | automaker |
| 435 | phenomenal |
| 436 | sweetie |
| 437 | observatory |
| 438 | overly |
| 430 | newlywed |
| 440 | vent |
| | pavilion |
| | tablespoon |
| 1/12 | polymerase |
| 111 | i neur genev |
| 444 | insurgency |
| 445 | terrapin |
| 440 | prairie |
| 44/ | cocoa |
| 448 | axiom |
| 449 | doggy |
| 450 | underdog |

| Rank WORD 451 harden 452 lobster 453 whopping 454 bobsled 455 lame 456 polio 457 tenet 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising 480 reindeer | |
|--|----------|
| 452 lobster 453 whopping 454 bobsled 455 lame 456 polio 457 tenet 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 453 whopping 454 bobsled 455 lame 456 polio 457 tenet 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 454 bobsled 455 lame 456 polio 457 tenet 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 455 lame 456 polio 457 tenet 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 456 polio 457 tenet 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 457 tenet 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 458 gag 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 459 crumble 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 460 mania 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 461 ferret 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 462 coefficient 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 463 strategist 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 464 staffer 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 465 runaway 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 466 showdown 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 467 yoga 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 468 twister 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 469 believer 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 470 oust 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 471 paparazzo 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | |
| 472 smelly 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | 4 |
| 473 meiosis 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | - 1 |
| 474 weaver 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | ┨ |
| 475 subtract 476 duel 477 parole 478 gymnast 479 uprising | ┨ |
| 476 duel 477 parole 478 gymnast 479 uprising | + |
| 477 parole 478 gymnast 479 uprising | 4 |
| 478 gymnast 479 uprising | + |
| 479 uprising | 1 |
| | \dashv |
| | \dashv |
| | 4 |
| 481 orangutan 482 cooperative | \dashv |
| 483 rink | ┨ |
| | 4 |
| | 4 |
| 485 dime 486 mow | ┨ |
| | \dashv |
| | \dashv |
| | 4 |
| | - |
| 490 brandnew | + |
| 491 intrigue | \dashv |
| 492 takeoff | 4 |
| 493 primetime | 4 |
| 494 carp | \dashv |
| 495 investigative | 4 |
| 496 merchandise | 4 |
| 497 oversight | 4 |
| 498 empower | |
| 499 witchcraft | 4 |
| 500 snarl | |

| Rank | WORD |
|------------|-------------|
| 501 | escalate |
| 502 | subpoena |
| 503 | stabilize |
| 504 | horseback |
| 505 | burp |
| 506 | fuzzy |
| 507 | animator |
| 508 | summertime |
| 509 | buckle |
| 510 | sloppy |
| 511 | woe |
| 512 | tomboy |
| 513 | newborn |
| 514 | luge |
| 515 | enchantment |
| 516 | zookeeper |
| 517 | crooner |
| 518 | umpire |
| 519 | cramp |
| 520 | backstage |
| 521 522 | standpoint |
| 522 | illustrator |
| 523 | apparel |
| 524 | beagle |
| 525 | badminton |
| 526 | lobbyist |
| 527 | preseason |
| 528 | terrier |
| 529 | squirt |
| 530 | nonnat i ve |
| 531 | utilize |
| 532 | spam |
| 533 | descriptor |
| 534 | retriever |
| 535 | autopsy |
| 536 | odyssey |
| 537 | snowy |
| 538 | excel |
| 539 | lumber |
| 540 | stomp |
| 541 | scavenger |
| 542 | humidity |
| 543 | babysit |
| 544 | buff |
| 545 | statewide |
| 546 | haircut |
| 547 | cosmonaut |
| 548 | lama |
| 549 | watermelon |
| | duplicate |
| | • |

| Rank | WORD |
|------|--------------|
| 551 | leash |
| 552 | tusk |
| 553 | nurture |
| 554 | firsthand |
| 555 | |
| 556 | tally |
| 557 | fanatic |
| 558 | onstage |
| 559 | taunt |
| 560 | |
| 561 | rectangle |
| | patriotism |
| | phony |
| | |
| 565 | snowboarder |
| 566 | |
| | wrestler |
| 567 | sculpt |
| 568 | bestseller |
| 569 | checker |
| 570 | diner |
| 571 | germ |
| 572 | trilogy |
| 573 | zipper |
| 574 | sesame |
| 575 | schoolwork |
| 576 | paycheck |
| 577 | songwriter |
| 578 | brownie |
| 579 | firefly |
| 580 | saxophone |
| 581 | obese |
| 582 | tidal |
| 583 | wrestle |
| 584 | milestone |
| 585 | browser |
| 586 | scroll |
| 587 | stallion |
| 588 | invader |
| 589 | frisbee |
| 590 | heighten |
| 591 | sunrise |
| 592 | flake |
| 593 | squash |
| 594 | timeline |
| 595 | reproducible |
| 596 | inflate |
| 597 | dismantle |
| 598 | spooky |
| 599 | bulge |
| 600 | encyclopedia |

| Rank | WORD |
|------------|---------------|
| 601 | primate |
| 602 | antitrust |
| 603 | feud |
| 604 | carnivore |
| 605 | cellphone |
| 606 | souvenir |
| 607 | coworker |
| 608 | booster |
| 609 | taco |
| 610 | bolster |
| 611 | wrongdoing |
| 612 613 | antenna |
| 613 | collaborate |
| 614 | fudge |
| 615 | trooper |
| 616 | spree |
| 617 | cam |
| 618 | constellation |
| 619 | standardize |
| 620 | boomer |
| 621 | sling |
| 622 | cartoonist |
| 623 | ivy |
| 624 625 | bison |
| 625 | bumper |
| 626 | undercover |
| 627 | piggy |
| 628 | soprano |
| 629 630 | commute |
| 630 | dissertation |
| 631 | karate |
| 632 | whiz |
| 633 | edible |
| 634 | lousy |
| 635 | setup |
| 636 | jellyfish |
| 637 638 | proton |
| 638 | nasal |
| 639 | spaceship |
| 640 | keystone |
| 641 | octopus |
| 642 | knack |
| 643 | bodyguard |
| 644 | rehearse |
| 645 | covet |
| 646 | relocate |
| 647 | paste |
| 648 | collegiate |
| 649 | smallpox |
| 650 | homeschool |
| | • |

| Donle | WODD |
|-------|--------------|
| Rank | WORD |
| 651 | whack |
| 652 | sharpen |
| 653 | raptor |
| 654 | jingle |
| 655 | cheerlead |
| 656 | cloudy |
| 657 | gamma |
| 658 | multiplayer |
| 659 | almond |
| 660 | prom |
| 661 | churn |
| 662 | sweaty |
| 663 | revere |
| 664 | stinky |
| 665 | flea |
| 666 | devastation |
| 667 | fastball |
| 668 | wrapper |
| 669 | pod |
| 670 | restart |
| 671 | headphone |
| 672 | skyscraper |
| 673 | hitch |
| 674 | dative |
| 675 | sweepstakes |
| 676 | rodeo |
| 677 | snapshot |
| 678 | cinnamon |
| 679 | notch |
| 680 | highland |
| 681 | woolly |
| 682 | snakehead |
| 683 | swan |
| 684 | silicon |
| 685 | cleric |
| 686 | inexpensive |
| 687 | ignite |
| 688 | reliever |
| 689 | mustache |
| 690 | genome |
| 691 | pharmacy |
| 692 | pediatric |
| 693 | fiance |
| 694 | aggressively |
| 695 | bribe |
| 696 | fiddle |
| 697 | bummer |
| 698 | pouch |
| 699 | derivative |
| 700 | candidacy |
| 700 | canardacy |

| Rank | WORD |
|------|--------------|
| 701 | metastasize |
| 702 | comma |
| 703 | plummet |
| 703 | |
| | elk |
| 705 | lookout |
| 706 | pretzel |
| 707 | smuggle |
| 708 | embedded |
| 709 | cello |
| 710 | devour |
| 711 | lighten |
| 712 | avalanche |
| 713 | infamous |
| 714 | geek |
| 715 | onscreen |
| 716 | broadband |
| 717 | groundhog |
| 718 | shimmer |
| 719 | intersection |
| 720 | sunflower |
| 721 | ceramic |
| 722 | ох |
| 723 | eater |
| 724 | |
| 725 | brute |
| 726 | vegetarian |
| 727 | tortoise |
| 728 | underscore |
| 729 | napkin |
| 730 | |
| 731 | yummy |
| 732 | freaky |
| 733 | tremendously |
| 734 | nab |
| 735 | |
| 736 | engrave |
| 737 | slayer |
| 738 | watcher |
| 739 | daydream |
| | teller |
| 740 | homeowner |
| 741 | fang |
| 742 | babysitter |
| 743 | roundup |
| 744 | |
| 745 | playful |
| 746 | cutie |
| 747 | carton |
| 748 | bullpen |
| 749 | • |
| 750 | beacon |

| 751 nightly 752 mural 753 skateboarder 754 songstress 755 funnel 756 fugitive 757 demo 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion 800 armpit | Rank | WORD |
|--|------|-------------|
| 753 skateboarder 754 songstress 755 funnel 756 fugitive 757 demo 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 751 | nightly |
| 753 skateboarder 754 songstress 755 funnel 756 fugitive 757 demo 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 752 | |
| 754 songstress 755 funnel 756 fugitive 757 demo 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 753 | |
| 755 funnel 756 fugitive 757 demo 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 754 | songstress |
| 756 fugitive 757 demo 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 757 demo 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 756 | |
| 758 lovable 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 759 matchbox 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 758 | |
| 760 ballroom 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 759 | |
| 761 saber 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 762 magnify 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 grandaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 763 mermaid 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 grandaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 764 fireman 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 763 | marmaid |
| 765 clench 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 766 motel 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 767 vie 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 768 empowerment 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 769 upbeat 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 770 beep 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 771 sandal 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 772 intentionally 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 773 fingernail 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 774 combo 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 772 | |
| 775 vase 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 774 | _ |
| 776 hustle 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 777 bagel 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 776 | |
| 778 amphibian 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 779 poker 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | Dage I |
| 780 aurora 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 781 gust 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | 700 | |
| 782 hospitalize 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 783 evaporate 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 784 dungeon 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 785 millimeter 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 786 chariot 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | uungeon |
| 787 varsity 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 788 rerun 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 789 granddaughter 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 790 curfew 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 791 broccoli 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 792 precipitation 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 793 synagogue 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | Droccoll |
| 794 chime 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 795 unhealthy 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 796 scurry 797 iguana 798 zombie 799 inversion | | |
| 797 iguana 798 zombie 799 inversion | | unhealthy |
| 798 zombie 799 inversion | | |
| 799 inversion | | |
| | | zombie. |
| 800 armpit | | |
| | 800 | armpit |

| Rank | WORD |
|------------|---------------------|
| 801 | jaguar |
| 802 | evacuee |
| 803 | fluffy |
| 804 | undec i ded |
| 805 | cherish |
| 806 | feline |
| 807 | powerhouse |
| 808 | journalistic |
| 809 | postwar |
| 810 | mimic |
| 811 | capitalize |
| 812 | safeguard |
| 813 | gruel |
| 814 | over load |
| 815 | popsicle |
| 816 | kayak |
| 817 | python |
| 818 | restroom |
| 819 | hacker |
| 820 | sunscreen |
| 821 | pesky |
| 822 | clubhouse |
| 823 | splinter |
| 824 | customize |
| 825 | pinpoint |
| 826 | infrared |
| 827 | retiree |
| 828 | tickle |
| 829 | subcommittee |
| 830 | ligament |
| 831 | quarantine |
| 832 | pore |
| 833 | leopard |
| 834 | chow |
| 835 | segregate |
| 836 | spicy |
| 837 | blister |
| 838 | eavesdrop |
| 839 | obstruction |
| 840 | shaggy |
| 841 842 | rhinoceros |
| 842 | doodle |
| | allele |
| 844 | burrow |
| 845 846 | tentacle |
| 847 | 00Ze |
| 848 | needy cornerback |
| | abide |
| 950 | proponent |
| 000 | իւ միույբուր |

| Rank | WORD |
|------------|------------------------|
| 851 | freeway |
| 852 | rainforest |
| 853 | gal |
| 854 | goof |
| 855 | gobble |
| 856 | feeder |
| 857 | erotica |
| 858 | fossilize |
| 859 | fray |
| 860 | equator |
| 861 | quarterfinal |
| 862 | bogey |
| 863 | lapse |
| 864 | computerize |
| 865 | torment |
| 866 | plume |
| 867 | goblin |
| 868 | meteorite |
| 869 | cannonball |
| 870 | |
| 871 | hog |
| 872 | repeal |
| 873 | stellar |
| 874 | lemur |
| 875 | mountainous |
| 876 | |
| 877 | layoff blogger |
| 878 | |
| 879 | rooftop parenthesis |
| 880 | inflatable |
| 881 | redesign |
| 882 | thirsty |
| | |
| 883 884 | throb |
| 885 | munch skepticism |
| 886 | |
| | graze |
| 887 | stint |
| 888 | handcuff |
| 889 | eerie |
| 890 | disqualify |
| 891 | hazel |
| 892 | greatness |
| 893 | spaghetti |
| 894 | anime |
| 895 | heartbeat |
| 896 | carousel |
| 897 | catapult |
| 898 | cuddle |
| 899 | slimy |
| 900 | sleek |

| Rank | WORD |
|------|---------------|
| 901 | fetus |
| 902 | shawl |
| 903 | cyber |
| 904 | nonfiction |
| 905 | brood |
| 906 | affirm |
| 907 | gill |
| 908 | salty |
| 909 | renovate |
| 910 | turnaround |
| 911 | clamp |
| 912 | nightclub |
| 913 | savvy |
| 914 | seafood |
| 915 | sweatshirt |
| 916 | guru |
| 917 | dent |
| 918 | speller |
| 919 | improvise |
| 920 | bulldog |
| 921 | toothbrush |
| 922 | slither |
| 923 | overflow |
| 924 | publicist |
| 925 | mastermind |
| 926 | screenplay |
| 927 | landfill |
| 928 | shoreline |
| 929 | redo |
| 930 | outspoken |
| 931 | seeker |
| 932 | tarantula |
| 933 | dwindle |
| 934 | coauthor |
| 935 | conspire |
| 936 | sunburn |
| 937 | derail |
| 938 | twirl |
| 939 | interrogation |
| 940 | eyewitness |
| 941 | abduct |
| 942 | pageant |
| 943 | slit |
| 944 | stardom |
| 945 | celery |
| 946 | unfairly |
| 947 | screener |
| 948 | spa |
| 949 | landslide |
| 950 | clam |
| 330 | O T UIII |

| Rank | WORD |
|------|----------------------|
| 951 | citation |
| 952 | boogie |
| 953 | cheesy |
| 05/ | yank |
| | spacewalk |
| | estrogen |
| 957 | |
| 958 | wand goaltender |
| 900 | cohost |
| 959 | COLIOS F |
| | standoff |
| 961 | veggie |
| | windshield |
| 963 | stench |
| 964 | embargo |
| 965 | upstate |
| 966 | seep |
| 967 | |
| 968 | elastic |
| | hardcore |
| 970 | strangle |
| 971 | barefoot |
| 972 | rearrange unearth |
| 973 | unearth |
| 974 | compassionate |
| 975 | measurable |
| 976 | accuser |
| 977 | eyeball |
| 978 | airwave |
| 979 | perk |
| 980 | modernity |
| 981 | tinker |
| 982 | tyrant |
| | wonder land |
| | clog |
| 985 | recede |
| 986 | swimsuit |
| 987 | midfielder |
| | diversify |
| 989 | |
| | arthritis |
| 991 | |
| 992 | rename |
| 993 | ferocious |
| 994 | graffito |
| | nighttime |
| 996 | bustle |
| 997 | waterfront |
| 998 | alot |
| 999 | perimeter |
| 1000 | ovation |
| | |

Rank WORD

資料 2. JACET Supplement 636

略号一覧

TXT 検定教科書

SP 口語英語

US 米語

CL 英米児童リーダー

FREQ 頻度

RANGE 4コーパスでのレンジ

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|-------------|---------|------------|-----------|----|-------------|---|
| | email | | 18. 024254 | | | 406. 666103 | |
| | Z00 | | 3. 3506626 | | | 158. 064258 | 3 |
| | astronaut | 91. 27 | | 57. 90494 | | 149. 17494 | 2 |
| | classmate | 116. 52 | | 20. 38462 | | 136. 904618 | 2 |
| | carol | 101. 96 | | 18. 55903 | | 120. 519028 | 2 |
| | panda | 63. 12 | | 50. 84729 | | 113. 967287 | 2 2 3 |
| | trash | | 17. 446554 | 29. 10061 | | 109. 667166 | 3 |
| | forth | | 48. 064677 | | | 103. 022486 | 3 |
| | cookie | | 4. 1594432 | | | 100. 991531 | 3 3 3 |
| | rainy | | 4. 0439032 | | | 83. 5064561 | 3 |
| | hamburger | 70. 88 | 2. 8307322 | 8. 758796 | | 82. 4695281 | 3 |
| | someday | 54. 38 | 3. 9861331 | 21. 62247 | | 79. 9885988 | 3 |
| | scary | 11. 65 | 23. 339098 | 44. 00491 | | 78. 9940089 | 3 |
| 14 | pacific | 36. 9 | 5. 3148441 | 33. 43954 | | 75. 6543812 | 3 |
| 15 | cloudy | | 3. 9283631 | | | 72. 2434732 | 3 |
| 16 | surf | 42. 72 | | 26. 40323 | | 69. 1232261 | 3 3 3 2 |
| 17 | creek | 37. 87 | 6. 470245 | 21. 11853 | | 65. 4587798 | 3 |
| 18 | hometown | 41. 75 | | 22. 33146 | | 64. 0814591 | 2 |
| 19 | cooperation | 28. 16 | 16. 291153 | 18. 64035 | | 63. 091504 | 3 |
| 20 | subway | 47. 58 | | 15. 48049 | | 63.0604863 | 2 |
| 21 | hike | 32. 04 | 4. 5060635 | 21. 54244 | | 58. 0885077 | 3 2 3 2 3 2 3 3 3 3 3 |
| 22 | vend | 50. 49 | | 7. 575636 | | 58. 065636 | 2 |
| 23 | website | 29. 13 | 2. 4841119 | 25. 93052 | | 57. 5446331 | 3 |
| 24 | clone | | 3. 0040423 | | | 57. 3020598 | 3 |
| 25 | pitcher | 10. 68 | 2. 3108018 | 43. 71597 | | 56. 7067723 | 3 |
| 26 | homer | 34. 96 | | 20. 53537 | | 55. 4953726 | 2 |
| 27 | playoff | | 2. 4841119 | 52. 62562 | | 55. 1097278 | 2 |
| 28 | teammate | 17. 48 | | 36. 29696 | | 53.7769582 | 2 |
| 29 | max | 22. 33 | 8. 2033464 | 22. 84921 | | 53. 3825572 | 2 3 3 3 2 2 |
| 30 | bookstore | 35. 93 | 3. 1773525 | 12. 49064 | | 51. 5979933 | 3 |
| 31 | volleyball | 32. 04 | 2. 7151921 | 14. 87139 | | 49.6265786 | 3 |
| 32 | bamboo | 37. 87 | | 9. 602637 | | 47. 4726366 | 2 |
| 33 | mammoth | 28. 16 | | 16. 35471 | | 44. 5147119 | 2 |
| 34 | spaghetti | 33. 01 | 5. 6614644 | 4. 3032 | | 42. 9746642 | 3 |
| 35 | cafeteria | 22. 33 | 3. 7550529 | 15. 64713 | | 41. 732181 | 3 |
| 36 | graduation | 32. 04 | 1. 8486414 | 7. 440684 | | 41. 3293254 | 3 |
| 37 | kilogram | 33. 01 | 7. 7989561 | | | 40.8089561 | 2 |
| 38 | closet | | 5. 2570741 | | | 40. 6704312 | 2 3 3 2 |
| 39 | softball | 22. 33 | 2. 6574221 | 15. 44911 | | 40. 4365359 | 3 |
| | alligator | 13. 59 | | 26. 22365 | | 39. 8136476 | 2 |
| 41 | pumpkin | 9. 71 | | 29. 49978 | | 39. 2097766 | 2 |
| | burger | | 4. 3327534 | | | 38. 7409023 | 3 |
| | semester | | 16. 348923 | 17. 26533 | | 38. 4742528 | 3 |
| | woof | 33. 01 | 5. 4303842 | | | 38. 4403842 | 2 |
| | cellphone | 33. 01 | | 5. 409182 | | 38. 4191822 | 2 |
| 46 | skateboard | 9. 71 | | 28. 09463 | | 37. 8046314 | 2 |
| | tornado | | 6. 1236248 | | | 37. 572521 | 3 |
| | backyard | | 6. 1236248 | | | 37. 4375317 | 2 2 3 3 2 2 |
| | skater | 3. 88 | | 32. 48729 | | 36. 3672913 | 2 |
| 50 | heartbeat | 32. 04 | | 4. 299036 | | 36. 3390358 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|--------------|--------|------------|-----------|----|-------------|---|
| | flea | 27. 19 | 3. 9861331 | 5. 142752 | | 36. 318885 | |
| 52 | freshman | 13. 59 | 4. 6216036 | 18. 09637 | | 36. 3079739 | 3 |
| 53 | backpack | 6.8 | 3. 2351225 | 25. 88291 | | 35. 9180329 | 3 |
| 54 | babe | 9. 71 | 5. 7192345 | 19. 34371 | | 34. 7729438 | 3 |
| 55 | canyon | 15. 54 | 2. 8885023 | 15. 89152 | | 34. 3200223 | 3 |
| 56 | alphabet | 18. 45 | 2. 4841119 | 13. 23554 | | 34. 169647 | 3 3 3 3 3 2 |
| 57 | octopus | 26. 22 | 2. 4263419 | 5. 265831 | | 33. 9121726 | 3 |
| 58 | touchdown | 7. 77 | | 25. 87736 | | 33. 6473575 | 2 |
| | noodle | 30. 1 | 3. 5239727 | | | 33. 6239727 | 2 |
| 60 | spaceship | 28. 16 | | 5. 281648 | | 33. 4416477 | 2 |
| 61 | snowboard | 14. 57 | | 18. 40397 | | 32. 9739653 | |
| 62 | shampoo | 29. 13 | 3. 6972829 | | | 32. 8272829 | 2 |
| | zebra | 25. 25 | | 6. 977939 | | 32. 227939 | 2 |
| | announcer | 24. 28 | | 7. 879378 | | 32. 1593782 | 2 |
| | firefighter | 4. 86 | | 27. 15117 | | 32. 0111717 | 2 |
| | birthplace | 11. 65 | | 19. 88903 | | 31. 5390254 | 2 |
| | makeup | | 6. 0080847 | | | 31. 2832365 | 2 2 2 2 2 2 2 3 3 3 2 |
| | gasoline | | 3. 812823 | | | 31. 1485641 | 3 |
| | telly | 4. 86 | | | | 31. 0876005 | 2 |
| | phoenix | | 3. 2928926 | | | 31. 019524 | 2 |
| | anytime | 7. 77 | 9. 8209077 | | | 30. 6713708 | |
| | superstar | 5. 83 | | 24. 74529 | | 30. 575291 | 2 |
| | thirsty | | 2. 8307322 | | | 30. 4827686 | 3 |
| | ups i de | | 6. 2391649 | | | 30. 3064348 | 3 |
| | cooperate | | 4. 6216036 | | | 30. 3027054 | 3 |
| | download | 2. 91 | | 27. 20698 | | 30. 1169764 | 2 |
| | kindergarten | | 7. 9144962 | | | 30. 0625868 | 3 |
| | kangaroo | 22. 33 | | 7. 629746 | | 29. 9597462 | 2 |
| | fro | | 2. 4263419 | | | 29. 8296709 | 2 |
| | souvenir | 24. 28 | | 5. 403926 | | 29. 6839259 | 2 |
| | oops | | 13. 864811 | | | 29. 4763632 | 3 |
| | doggy | | 5. 141534 | | | 29. 0958138 | 3 |
| | swimmer | 17. 48 | | 11. 50402 | | 28. 9840183 | 2 |
| | rookie | 4. 86 | | 24. 0027 | | 28. 8627031 | 2 |
| | vase | | 3. 5817428 | | | 28. 6661174 | 2 |
| | dough | | 2. 7729622 | | | 28. 6119531 | 3 |
| | popcorn | | 3. 4662027 | | | 28. 5474279 | |
| | snowy | 22. 33 | | 5. 838171 | | 28. 1681709 | 2 |
| | squash | | 9. 5898275 | | | 27. 7380075 | 3 |
| 1 1 | ranch | | 3. 9861331 | | | 27. 658841 | 2 3 3 3 3 3 |
| | memorize | 14. 57 | | | | 27. 6459782 | 3 |
| | peel | | 8. 7232768 | | | 27. 3829835 | .3 |
| | bakery | 23. 3 | | | | 27. 112823 | |
| | chili | | 6. 0080847 | | | 26. 9481263 | |
| | reed | | 8. 3188865 | | | 26. 7720246 | .3 |
| | gee | 1.11 | | | | 26. 7484087 | 3 3 3 3 3 3 |
| | spike | 9 71 | 4. 2749833 | | | 26. 7032184 | 3 |
| | buffalo | | 5. 0259939 | | | 26. 6664485 | 2 |
| | juror | 6.8 | | | | 26. 0896269 | 2 |
| | pavilion | | 3. 0618124 | | | 26. 0889607 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|-------------|--------|------------|-----------|-----------|-------------|----------------------------|
| 101 | audition | 4. 86 | 2. 1952617 | 19. 0244 | | 26. 0796635 | 3 |
| 102 | dormitory | 13. 59 | 4. 5638336 | 7. 554478 | | 25. 7083118 | 3 |
| 103 | tsunami | 3. 88 | | 21. 70145 | | 25. 5814453 | 2 |
| 104 | cleanup | 16. 51 | | 9. 067445 | | 25. 5774445 | 2 |
| | heck | | 14. 384741 | 11. 04186 | | 25. 426604 | |
| | wrestling | 4. 86 | 4. 2172133 | | | 25. 14522 | 2 3 2 |
| | vanilla | 13. 59 | | 10. 56395 | | 24. 1539459 | 2 |
| | dribble | | 2. 4841119 | 9. 732901 | | 23. 8670131 | 3 |
| | equator | 19. 42 | | 4. 418848 | | 23. 8388482 | 2 |
| 110 | tortoise | 18. 45 | | 4. 871959 | | 23. 321959 | 2 |
| 111 | sidewalk | 6.8 | | 16. 3787 | | 23. 1786952 | 2 |
| | derby | | 9. 4742874 | | | 23. 088397 | 2 |
| | yoghurt | 9. 71 | 6. 470245 | | | 23. 0259263 | 3 |
| | birch | 8. 74 | | | 14. 27257 | 23. 0125672 | 2 |
| | gorilla | 5. 83 | | 17. 14186 | | 22. 9718582 | 2 |
| | symbolize | 15. 54 | | 7. 389795 | | 22. 9297949 | 2 |
| | hoop | | 5. 141534 | 17. 67013 | | 22. 8116657 | 2 |
| | handout | | 14. 038121 | | | 22. 7895314 | 2 |
| | sneeze | 12. 62 | 2. 9462723 | | | 22. 6848088 | 3 |
| | bristle | | 9. 1276671 | | | 22. 6260435 | 3 |
| 121 | taco | | 2. 4841119 | | | 22. 4500403 | 3 3 3 3 2 3 |
| | pickle | | 4. 5060635 | | | 22. 4115872 | 3 |
| | nap | 9. 71 | 2. 1374917 | 10. 53691 | | 22. 3844044 | 3 |
| | jumper | | 13. 864811 | | | 22. 2287365 | 2 |
| | rotate | 2. 91 | 7. 741186 | | | 22. 1373314 | 3 |
| | whoa | | 10. 283068 | | | 21. 8332142 | |
| | seafood | 12. 62 | 4. 7949137 | | | 21. 6515811 | 2 |
| | pajama | | 4. 7371437 | | | 21. 5661125 | |
| | fin | 3. 88 | | | | 21. 4764519 | 3 |
| 130 | don | 8. 74 | | | | 21. 3916399 | |
| | trek | 2. 91 | 7. 0479455 | | | 21. 2748508 | 2 |
| | bracelet | 7. 77 | | | | 21.0115121 | 3 |
| | homemade | 9. 71 | 3. 0040423 | 8. 227555 | | 20. 9415971 | 3 |
| 134 | mini | | 10. 687458 | 10. 23284 | | 20. 9202971 | 2 |
| 135 | whoop | | 13. 1138 | 7. 662206 | | 20. 7760066 | 2 2 |
| 136 | photon | | 9. 8209077 | 10. 86711 | | 20. 6880191 | 2 |
| | sneaker | 6.8 | | 13. 65261 | | 20. 4526102 | 2 |
| | hum i d | 16. 51 | 3. 870593 | | | 20. 380593 | |
| | hitter | 6.8 | | 12. 95728 | | 19. 7572797 | 2 |
| | reef | 6.8 | 3. 2928926 | | | 19. 663999 | 3 |
| | pancake | 6.8 | 3. 1195824 | | | 19. 6493262 | 3 |
| | catcher | 7. 77 | | 11. 64538 | | 19. 4153793 | 2 |
| | grader | 5. 83 | 2. 541882 | 10. 94004 | | 19. 3119182 | 3 |
| | mash | 6.8 | | | | 19. 2249644 | 3 |
| | itch | 8. 74 | 1. 9641815 | | | 19. 0964081 | 3 |
| | ecosystem | 11. 65 | | 7. 443475 | | 19. 0934745 | 2 |
| | downhill | 3. 88 | 4. 2749833 | | | 19. 0455263 | 3 |
| | precipitate | | | 10. 06088 | | 19. 0152342 | 2 |
| | cartoonist | 13. 59 | | 5. 348375 | | 18. 9383753 | 2 2 |
| | artifact | | 3. 870593 | 15. 06463 | | 18. 9352217 | 2 |

| Rank | | | | US | CL | FREQ | RANGE |
|------|---------------|--------|------------|-----------|-----------|-------------|--|
| 151 | skeptical | 2. 91 | 6. 5280151 | | | 18. 903087 | 3 2 |
| 152 | squid | 2. 91 | | 15. 88435 | | 18. 7943463 | 2 |
| 153 | allergy | 3. 88 | 3. 1195824 | 11. 69289 | | 18. 6924696 | 3 |
| 154 | badminton | 10. 68 | 1. 9641815 | 5. 94103 | | 18. 5852118 | 3 |
| 155 | skier | 2. 91 | | 15. 6516 | | 18. 5616018 | 2 |
| 156 | roommate | 6.8 | 4. 5060635 | 7. 192617 | | 18. 4986807 | 3 |
| 157 | awhile | | 8. 3188865 | 10. 11335 | | 18. 4322361 | 2 |
| 158 | sprout | 4. 86 | 5. 2570741 | 8. 269472 | | 18. 3865458 | 3 |
| 159 | boo | | 7. 9144962 | 10. 44921 | | 18. 3637076 | 2 |
| 160 | bedtime | 8. 74 | 2. 7729622 | 6. 844696 | | 18. 3576584 | 3 |
| 161 | flier | | 3.812823 | 14. 52358 | | 18. 3364038 | 3 2 3 2 3 2 3 2 2 3 2 2 2 2 2 2 2 2 2 2 |
| 162 | dodge | | 4. 5060635 | 13. 49174 | | 17. 9978002 | 2 |
| 163 | ketchup | 10. 68 | | 7. 303892 | | 17. 9838923 | 2 |
| 164 | lemonade | | 4. 5638336 | 13. 38254 | | 17. 9463724 | 2 2 |
| 165 | cooperative | 3. 88 | 7. 7989561 | 6. 246823 | | 17. 9257787 | 3 |
| 166 | comma | | 12. 94049 | 4. 964659 | | 17. 9051488 | 2 |
| 167 | accomplishmen | 4. 86 | | 13. 03239 | | 17. 8923903 | 2 |
| 168 | paralyze | 7. 77 | 4. 3327534 | 5. 628989 | | 17. 7317423 | 3 |
| 169 | fireman | 6.8 | 6. 1813948 | 4. 737018 | | 17. 7184127 | 3 |
| 170 | sunglass | 9. 71 | | 7. 951543 | | 17. 6615433 | 2 |
| 171 | chow | 9. 71 | 3. 2928926 | 4. 478702 | | 17. 481595 | 3 |
| 172 | zip | | 6. 8746354 | 10. 53415 | | 17. 4087877 | 2 |
| 173 | mural | 12. 62 | | 4. 768697 | | 17. 3886965 | 2 |
| 174 | videotape | | 2. 7729622 | 14. 60951 | | 17. 3824716 | 2 |
| 175 | allergic | 3. 88 | 5. 2570741 | 8. 232519 | | 17. 3695932 | 3 |
| | paste | 9. 71 | 2. 3685718 | 5. 244296 | | 17. 3228682 | 2 2 3 3 2 2 2 2 2 3 3 3 3 3 2 2 2 2 2 2 |
| 177 | carton | 8. 74 | 3. 7550529 | 4. 807059 | | 17. 3021121 | 3 |
| 178 | nonstop | 8. 74 | | 8. 541235 | | 17. 2812354 | 2 |
| 179 | wrestler | 11. 65 | | 5. 622144 | | 17. 2721439 | |
| 180 | peda l | 6.8 | 3. 4084327 | 6. 85641 | | 17. 0648424 | 3 2 |
| 181 | infinity | | 9. 012127 | 8. 039456 | | 17. 0515829 | 2 |
| 182 | goldfish | 5. 83 | 3.870593 | 7. 18941 | | 16. 8900031 | 3 |
| 183 | kitty | | 4. 3327534 | 12. 55547 | | 16. 8882193 | 2 |
| | eyesight | | 2. 2530318 | | | 16. 8230318 | 3 2 2 3 |
| | diaper | 6. 8 | 2. 8885023 | 7. 123023 | | 16. 8115254 | 3 |
| | lacrosse | 3. 88 | | 12. 92236 | | 16. 8023632 | 2 |
| 187 | spicy | 9. 71 | 2. 599652 | 4. 477371 | | 16. 7870233 | |
| 188 | beep | 5. 83 | 6. 2391649 | 4. 716555 | | 16. 7857195 | 3 |
| 189 | bathtub | 9. 71 | | 7. 067448 | | 16. 7774483 | 2 |
| 190 | sitcom | | 1. 9641815 | 14. 71583 | | 16. 6800077 | 2 |
| 191 | nay | | 4. 1016732 | | 12. 55986 | 16. 6615323 | |
| 192 | moderator | | 7. 741186 | 8. 900185 | | 16. 6413711 | 2 |
| 193 | mosque | 4. 86 | | 11. 54585 | | 16. 4058532 | 2 |
| 194 | navigate | 9. 71 | | 6. 65419 | | 16. 36419 | 2 |
| 195 | rectangle | 10. 68 | | 5. 638676 | | 16. 3186763 | 2 |
| 196 | auditorium | 4. 86 | 3. 4084327 | 8.012703 | | 16. 281136 | |
| 197 | tablespoon | 7. 77 | 1. 9641815 | 6. 515624 | | 16. 2498052 | |
| 198 | sunrise | 10. 68 | | 5. 537551 | | 16. 2175513 | 2 |
| | cedar | 13. 59 | 2. 4841119 | | | 16. 0741119 | |
| 200 | coefficient | | 9. 6475975 | 6.368136 | | 16. 0157332 | 2 |

| Rank | WORD | TXT | SP | US | CL F | FREQ | RANGE |
|------|--------------|--------|------------|-----------|------|---------------------|--|
| 201 | crunch | 3. 88 | 4. 2172133 | 7. 896353 | | 15. 9935663 | 3 |
| 202 | lastly | 11. 65 | 4. 3327534 | | | 15. 9827534 | 3 2 |
| | nonprofit | 5. 83 | | 10. 12239 | | 15. 9523866 | 2 |
| | reuse | 13. 59 | 2. 2530318 | | | 15. 8430318 | 2 |
| | autograph | 2. 91 | | 12. 93005 | | 15. 8400523 | 2 |
| | snowball | 5. 83 | 2. 1952617 | | | 15. 7512724 | 2 3 2 |
| | psych | | 4. 2172133 | | | 15. 6901308 | 2 |
| | barber | 8. 74 | | 6. 941327 | | 15. 6813268 | 2 |
| | hamster | | 2. 7729622 | | | 15. 6458343 | |
| 210 | rocker | 5. 83 | | 9. 795218 | | 15. 6252184 | 3 2 |
| 211 | lark | 2. 91 | 2. 4841119 | 10. 17462 | | 15. 5687322 | 3 |
| 212 | tattoo | 5. 83 | | 9. 702439 | | 15. 5324389 | 2 |
| 213 | humidity | 4. 86 | 4. 8526838 | 5. 802873 | | 15. 5155568 | 3 |
| 214 | nickel | | 3. 0040423 | 12. 47773 | | 15. 4817744 | 2 |
| 215 | mailbox | 5. 83 | 2. 2530318 | 7. 363629 | | 15. 4466612 | 3 |
| 216 | teaspoon | 3. 88 | 2. 0797216 | 9. 478289 | | 15. 4380107 | 3 |
| 217 | thunderstorm | 2. 91 | | 12. 49777 | | 15. 4077747 | 2 |
| 218 | plaza | 3. 88 | | 11. 49098 | | 15. 3709816 | 2 |
| 219 | fiddle | | 10. 283068 | 4. 992577 | | 15. 2756446 | 2 |
| 220 | fundraising | 3. 88 | | 11. 38242 | | 15. 2624242 | 2 |
| 221 | vegetarian | 2. 91 | 7. 4523358 | 4. 877263 | | 15. 239599 | 2 3 3 3 2 2 2 2 2 3 3 2 2 3 3 2 2 3 2 2 3 2 2 3 2 2 3 2 2 2 2 3 2 |
| 222 | carnival | 6.8 | | 8. 362561 | | 15. 1625605 | 2 |
| 223 | mow | | 8. 8965869 | 6. 225706 | | 15. 1222926 | 2 |
| 224 | rainforest | 7. 77 | 2. 8885023 | 4. 440964 | | 15. 0994666 | 3 |
| 225 | skinny | | 5. 9503146 | 9. 132921 | | 15. 0832355 | |
| 226 | finalist | 3. 88 | | 11. 1932 | | 15. 0731992 | 2 |
| 227 | mentor | | 3. 5817428 | 11. 44357 | | 15. 0253176 | 2 |
| 228 | jellyfish | 9. 71 | | 5. 297833 | | 15. 0078331 | 2 |
| 229 | mustache | | 3. 1773525 | 5. 009053 | | 14. 9864051 | 3 |
| 230 | hairdresser | 9. 71 | 5. 2570741 | | | 14. 9670741 | 2 3 3 2 |
| 231 | mayonnaise | 4. 86 | 2. 3685718 | 7. 631288 | | 14. 8598599 | 3 |
| | snowman | 6.8 | 2. 3685718 | 5. 684981 | | 14. 8535533 | 3 |
| | bash | | 6. 412475 | 8. 386616 | | 14. 7990913 | 2 |
| 234 | newborn | 8. 74 | | 6. 011807 | | 14. 7518066 | 2 2 |
| 235 | hilarious | | 5. 3726142 | 9. 327562 | | 14. 7001763 | |
| 236 | zoom | 3. 88 | 3. 9861331 | 6. 79967 | | 14. 665803 | 3 |
| 237 | folder | | 7. 8567261 | 6. 791258 | | 14. 6479838 | 2 |
| 238 | neutron | | 4. 7949137 | | | 14. 575623 <u>1</u> | 2 |
| | crayon | | 2. 8885023 | 11. 56483 | | 14. 4533344 | 2 |
| | smog | 11. 65 | 2. 7729622 | | | 14. 4229622 | 2 |
| | antenna | 8. 74 | | 5. 374488 | | <u>14. 1144877</u> | 2 |
| | magnet | | 3. 1773525 | | | 14. 1141986 | 3 |
| | peek | 2. 91 | | 11. 14628 | | 14. 0562762 | 2 |
| | pouch | | 2. 1374917 | | | 13. 9220997 | 3 |
| | sophomore | 2. 91 | 2. 3108018 | | | 13. 919611 | 3 |
| | dude | | 3. 6972829 | | | 13. 9179164 | 2 3 3 2 |
| | cocoa | | 3. 5239727 | | | 13. 892513 | 3 |
| | pharmacy | 6.8 | 2. 0797216 | | | 13. 8868975 | 3 |
| | darn | | 6. 7590953 | | | 13. 878308 | 2 |
| 250 | enchant | 3. 88 | | 9. 950607 | | 13. 8306072 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|--------------|--------|------------|-----------|----|-------------|----------------------------|
| 251 | iceberg | 4. 86 | 2. 0219516 | 6. 904996 | | 13. 7869479 | 3 |
| | pastime | 5. 83 | | 7. 923691 | | 13. 7536913 | 2 |
| | nurture | 5. 83 | 2. 1374917 | 5. 700264 | | 13. 6677557 | 3 2 3 |
| | setup | | 5. 3726142 | | | 13. 5818441 | 3 |
| | knob | | 4. 7949137 | | | 13. 5349137 | 2 |
| 256 | goody | | 5. 7192345 | 7. 797193 | | 13. 5164273 | 2 |
| | hind | 2. 91 | 3. 812823 | | | 13. 4801544 | 3 |
| | wag | 10. 68 | 2. 7729622 | | | 13. 4529622 | 3 2 2 3 2 |
| | quilt | | 5. 4303842 | | | 13. 3817984 | |
| | babysit | | 7. 5678759 | | | 13. 3617306 | 2 |
| 261 | landfill | 4. 86 | 4. 2172133 | 4. 202913 | | 13. 2801267 | 3 |
| | restroom | 8. 74 | | 4. 524875 | | 13. 2648747 | 2 |
| | encyclopedia | | 1. 9064115 | 5. 465008 | | 13. 2014194 | 3 |
| | doughnut | | 2. 1374917 | | | 13. 1777505 | 2 |
| | haircut | 4. 86 | 2. 541882 | | | 13. 1729549 | 3 |
| | blackboard | | 2. 4263419 | | | 13. 1063419 | 2 3 2 3 2 |
| | anime | 8. 74 | | 4. 303186 | | 13. 0431858 | 2 |
| | shawl | 8. 74 | | 4. 277497 | | 13. 0174974 | 2 |
| | camper | | 4. 1594432 | | | 12. 9535772 | 2 |
| | sparrow | | 5. 7192345 | | | 12. 9305213 | 2 |
| | tug | | 2. 7729622 | | | 12. 9186701 | 2 2 2 2 2 |
| | cheetah | | 2. 7151921 | | | 12. 9057453 | 2 |
| 273 | fisher | 2. 91 | 2. 2530318 | | | 12. 8747642 | 2 3 2 3 |
| | subcommittee | | 8. 3766565 | | | 12. 8603282 | 2 |
| | tickle | 3. 88 | 4. 3905234 | | | 12. 7586847 | 3 |
| | giraffe | 4. 86 | | 7. 896585 | | 12. 7565848 | 2 |
| | quack | | 3. 0040423 | | | 12. 7140423 | 2 |
| | preschool | | 5. 4303842 | | | 12. 6866535 | 2 |
| | beginner | 9. 71 | 2. 9462723 | | | 12. 6562723 | 2 |
| | shred | | 2. 1952617 | | | 12. 6241596 | 2 2 3 3 2 2 |
| 281 | over load | | 2. 2530318 | | | 12. 6150733 | 3 |
| 282 | vinegar | 8. 74 | 3. 870593 | | | 12. 610593 | 2 |
| | loch | | 2. 8885023 | | | 12. 5985023 | 2 |
| 284 | hack | | 4. 5060635 | 8. 053163 | | 12. 5592265 | 2 2 |
| 285 | watermelon | 6.8 | | 5. 752402 | | 12. 5524015 | 2 |
| 286 | foil | | 2. 7151921 | 9.819432 | | 12.5346246 | 2 |
| 287 | burrow | 4. 86 | 3. 1773525 | 4. 450172 | | 12. 4875244 | 3 |
| 288 | firsthand | 6.8 | | 5. 68622 | | 12. 4862204 | 2 2 |
| 289 | millimeter | | 7. 8567261 | 4. 627966 | | 12. 4846919 | 2 |
| 290 | sandal | 7. 77 | | 4. 712189 | | 12. 4821895 | 2 |
| 291 | scorer | | 2. 3108018 | 10. 16245 | | 12. 473254 | 2 |
| 292 | fingernail | 7.77 | | 4. 702467 | | 12. 4724672 | 2 2 |
| 293 | shovel | | 4. 9682239 | 7. 481267 | | 12. 4494913 | 2 |
| 294 | zipper | 6.8 | | 5. 602746 | | 12. 4027464 | 2 |
| 295 | scooter | | 1. 9641815 | 10. 42527 | | 12. 3894533 | 2 |
| 296 | inexpensive | 4. 86 | 2. 4841119 | 5. 03938 | | 12. 3834916 | 2 3 2 2 2 2 |
| 297 | cherish | 7.77 | | 4. 557256 | | 12. 3272556 | 2 |
| 298 | utilize | | 6. 354705 | 5. 903284 | | 12. 257989 | 2 |
| 299 | ballpark | 4. 86 | | 7. 34124 | | 12. 2012399 | 2 |
| 300 | mule | | 2. 1952617 | 9. 989569 | | 12. 1848303 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|--------------|-------|------------|-----------|-----------|-------------|---------------------------------|
| 301 | polymerase | | 5. 6614644 | 6. 4951 | | 12. 1565645 | 2 |
| 302 | retriever | 2. 91 | 3. 2928926 | 5. 872129 | | 12. 0750212 | 3 |
| 303 | salty | 7. 77 | | 4. 254994 | | 12. 0249943 | 2 |
| 304 | flashlight | 4. 86 | | 7. 124809 | | 11. 9848086 | 2 2 |
| | replay | | 3. 9861331 | 7. 978001 | | 11. 9641345 | 2 |
| 306 | panther | 2. 91 | | 9. 033081 | | 11.9430807 | 2 2 2 |
| | axiom | | 5. 4303842 | 6. 487103 | | 11. 9174876 | 2 |
| 308 | | 9. 71 | 2. 1952617 | | | 11. 9052617 | 2 |
| | segregate | | 2. 541882 | | | 11.8797242 | |
| | boulder | | 2. 3108018 | | | 11. 8751819 | 3 2 |
| | ouch | 9. 71 | 2. 1374917 | | | 11.8474917 | 2 |
| | saxophone | | 2. 3685718 | | | 11.8430232 | 3 |
| | umpire | 5. 83 | | 5. 986142 | | 11. 8161421 | 2 |
| | syndicate | | 2. 8885023 | | | 11. 7964009 | 2 |
| | meiosis | | 5. 4881543 | | | 11. 7879627 | 2 |
| | surfer | 4. 86 | | 6. 923827 | | 11. 7838274 | 2 2 2 2 |
| | toddler | | 3. 4084327 | | | 11. 7632731 | 2 |
| | surpass | 4. 86 | | 6. 884077 | | 11. 7440773 | 2 |
| | portrayal | | 3. 9861331 | | | 11. 73856 | 2 |
| | summertime | | 5. 6614644 | | | 11. 7161502 | 2 2 2 2 2 |
| | subtract | | 5. 4303842 | | | 11. 6992223 | 2 |
| 322 | | | 2. 1952617 | | | 11. 5819474 | 2 |
| | endurance | 3. 88 | | 7. 683841 | | 11. 5638415 | 2 2 3 2 |
| | arthritis | | 4. 6216036 | | | 11. 5587994 | 3 |
| | phenomenal | | 4. 9682239 | | | 11. 5232254 | 2 |
| | tinker | 3. 88 | 3. 5817428 | | | 11. 5053076 | |
| | motorbike | | 5. 6614644 | | | 11. 4914644 | 3 2 |
| | parole | | 5. 1993041 | | | 11. 4648214 | 2 |
| | grizzly | 2. 91 | | 8. 534463 | | 11. 4444629 | 2 |
| | schoolwork | 5. 83 | | 5. 599368 | | 11. 4293683 | 2 |
| | proton | | 6. 1236248 | | | 11. 415502 | 2 |
| | lobster | 2. 91 | 2. 0219516 | | | 11. 3778303 | 3 |
| | recyclable | | 4. 5638336 | | | 11. 3638336 | 2 2 3 2 |
| | gallop | | 2. 1952617 | | 9. 134443 | 11. 3297047 | 2 |
| | redo | | 7. 1057155 | 4. 198792 | | 11. 3045076 | 2 2 |
| 336 | magician | 2. 91 | | 8. 394345 | | 11. 3043447 | 2 |
| | spokesperson | | 1. 9641815 | 9. 330176 | | 11. 2943573 | 2 |
| | swan | | 6. 2391649 | | | 11. 2855618 | 2 |
| | aspire | 3. 88 | | 7. 402959 | | 11. 2829592 | 2 |
| | canine | | 1. 8486414 | | | 11. 2271127 | 2 |
| | commute | | 5. 8925446 | | | 11. 2172382 | 2 |
| | northeast | 7.77 | 3. 4084327 | | | 11. 1784327 | |
| | beaver | | 1. 9641815 | 9. 162818 | | 11. 127 | 2 |
| | sesame | 2. 91 | | 5. 602291 | | 11. 1119429 | 3 |
| | orangutan | 4. 86 | | 6. 248212 | | 11. 1082122 | 2 |
| | pastor | 3. 88 | | 7. 211453 | | 11. 0914533 | 2 2 2 2 2 2 2 |
| | sling | | 5. 7192345 | 5. 35 | | 11. 0692341 | 2 |
| | ivy | | 5. 7192345 | | | 11. 0671172 | 2 |
| | adorable | 4. 86 | | 6. 206477 | | 11. 0664772 | 2 |
| | brandnew | 4. 86 | | 6. 178138 | | 11. 0381375 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|--------------|-------|------------|-----------|-----------|-------------|-----------------------|
| 351 | parenthesis | 2. 91 | 3. 7550529 | 4. 366486 | | 11.0315385 | 3 |
| 352 | marijuana | | 2. 4263419 | 8. 598888 | | 11. 0252302 | 3 2 |
| | toothbrush | 6.8 | | 4. 221 | | 11. 0210005 | 2 |
| | cello | 3. 88 | 2. 1952617 | | | 11. 0116804 | 3 |
| | takeoff | 4. 86 | | 6. 151345 | | 11. 011345 | 2 |
| | juggle | | 3. 4662027 | | | 11. 0046838 | 2 3 2 2 2 |
| | descriptor | | 5. 083764 | | | 10. 9622965 | 2 |
| | horseback | 4. 86 | | 6. 066815 | | 10. 9268154 | 2 |
| | mower | | 6. 0658547 | | | 10. 9258547 | |
| | medalist | 3. 88 | | 7. 030404 | | 10. 9104039 | 2 |
| 361 | installment | | 3. 0618124 | 7. 840388 | | 10. 9022006 | 2 |
| | grapefruit | 8. 74 | 2. 1374917 | | | 10.8774917 | 2 |
| | flop | | 2. 7151921 | | | 10. 8610402 | 2 |
| | affirmative | | 1. 8486414 | | | 10.8509432 | 2 |
| | broccoli | 2. 91 | 3. 3506626 | | | 10. 8494137 | 3 |
| | loot | | 2. 0797216 | | | 10. 8150437 | 2 2 3 2 |
| | dissertation | | 5. 4303842 | 5. 324074 | | 10. 7544584 | 2 |
| 368 | prank | 2. 91 | | 7. 830852 | | 10. 7408518 | 2 2 2 |
| | driveway | | 3. 812823 | 6. 91894 | | 10. 731763 | 2 |
| 370 | flicker | | 3. 6395128 | 7. 064911 | | 10. 7044236 | 2 2 |
| 371 | operative | | 3. 5239727 | | | 10. 7005468 | 2 |
| | fingerprint | | 2. 2530318 | 8. 398937 | | 10.6519683 | 2 |
| | heavyweight | 3. 88 | | 6. 697511 | | 10. 5775107 | 2 2 2 2 |
| 374 | standpoint | | 4. 6216036 | | | 10. 5761032 | 2 |
| 375 | potassium | 5. 83 | 4. 7371437 | | | 10. 5671437 | 2 |
| 376 | smelly | | 4. 2172133 | 6. 299871 | | 10. 5170839 | 2 |
| | prep | | 3. 5239727 | 6. 98953 | | 10. 5135027 | 2 |
| | tremendously | | 5. 6614644 | 4. 850553 | | 10. 5120177 | 2 |
| | rabbi | | 2. 7151921 | 7. 783713 | | 10. 4989052 | 2 |
| 380 | aurora | 5. 83 | | 4. 666266 | | 10. 4962664 | 2 |
| 381 | rouge | 3. 88 | | 6. 594498 | | 10. 4744981 | 2 |
| 382 | furry | 2. 91 | | 7. 494489 | | 10. 4044894 | 2 2 2 2 |
| 383 | revolve | | 3. 7550529 | 6. 591742 | | 10.3467952 | 2 |
| | fielder | 2. 91 | | 7. 402046 | | 10. 3120459 | 2 2 |
| 385 | hype | | 2. 1374917 | 8. 087046 | | 10. 2245375 | 2 |
| 386 | wrapper | 2. 91 | 2. 1952617 | 5. 104241 | | 10. 2095028 | 3 |
| 387 | paycheck | | 4. 5638336 | 5. 598814 | | 10. 1626478 | 2 |
| 388 | harden | | 3. 6972829 | 6. 44967 | | 10. 1469529 | 2 |
| 389 | specialty | | 3. 2351225 | 6. 901668 | | 10. 1367908 | 2 |
| 390 | escalate | | 4. 0439032 | 6. 079913 | | 10. 123816 | 2 |
| 391 | sweetie | | 3. 5817428 | 6. 537294 | | 10. 119037 | 2 |
| 392 | nip | | 4. 9682239 | | 5. 138124 | 10. 1063481 | 2 |
| 393 | cramp | | 4. 1016732 | 5. 962258 | | 10.0639316 | 2 |
| 394 | whiz | | 4. 7371437 | | | 10.0601434 | 2 |
| 395 | eater | 2. 91 | 2. 2530318 | 4. 883121 | | 10.0461524 | 3 |
| 396 | carp | 3. 88 | | 6. 14289 | | 10. 0228901 | 2 |
| 397 | workout | | 2. 4263419 | 7. 592154 | | 10. 0184954 | 2 |
| 398 | blizzard | 2. 91 | | 7. 062558 | | 9. 97255831 | 2 2 2 2 2 |
| 399 | hoof | | 2. 8885023 | 7. 08058 | | 9. 96908215 | 2 |
| 400 | cinnamon | 4. 86 | | 5.064312 | | 9. 92431248 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|--------------|-------|------------|-----------|-----------|-------------|---------------------------------|
| | cuddle | | 5. 6036944 | | | 9. 88992958 | |
| | suspense | 7. 77 | 2. 0797216 | | | 9. 84972162 | 2 |
| | slug | | 2. 4841119 | | | 9. 8200886 | 2 |
| | vent | | 3. 2928926 | | | 9. 81158899 | 2 |
| | dolly | | 4. 1016732 | | 5. 709027 | 9. 81070008 | 2 2 2 2 2 |
| | slum | 6.8 | 3. 0040423 | | | 9. 80404234 | 2 |
| | windmill | | 2. 0219516 | | | 9. 79195158 | 2 |
| 408 | intrigue | | 3. 5817428 | | | 9. 73602882 | 2 |
| | headphone | | 4. 6216036 | | | 9. 72225069 | |
| 410 | stomp | 3. 88 | | 5. 823292 | | 9. 70329194 | 2 2 |
| 411 | believer | | 3. 3506626 | 6. 319376 | | 9. 67003852 | 2 |
| 412 | tine | 2. 91 | 6. 7590953 | | | 9. 66909527 | 2 |
| 413 | fluffy | 2. 91 | 2. 1952617 | 4. 559251 | | 9. 66451285 | 3 |
| 414 | derivative | | 4. 6793737 | | | 9. 65816429 | 2 |
| 415 | whack | | 4. 3905234 | 5. 230554 | | 9. 62107767 | 2 3 2 2 2 |
| 416 | heron | 6.8 | 2. 7729622 | | | 9. 57296216 | 2 |
| 417 | envision | | 1. 9064115 | 7. 653045 | | 9. 55945617 | 2 2 2 2 2 2 |
| 418 | weekday | 2. 91 | | 6. 629771 | | 9. 53977063 | 2 |
| 419 | dative | | 4. 4482935 | 5. 089939 | | 9. 53823246 | 2 |
| 420 | skunk | | 1. 8486414 | 7. 684857 | | 9. 53349827 | 2 |
| 421 | crumble | | 3. 1195824 | 6. 391195 | | 9. 51077757 | 2 |
| 422 | bumper | | 4. 0439032 | 5. 338392 | | 9. 38229492 | 2 3 2 2 |
| 423 | chariot | 2. 91 | 1. 8486414 | 4. 617783 | | 9. 37642438 | 3 |
| 424 | overcrowd | 5. 83 | 3. 5239727 | | | 9. 35397275 | 2 |
| 425 | alumnus | | 1. 9641815 | | | 9. 34404329 | 2 |
| 426 | gill | | 5. 083764 | 4. 257714 | | 9. 34147845 | 2 |
| 427 | paddy | 2. 91 | 6. 412475 | | | 9. 322475 | 2 2 |
| 428 | coworker | 3. 88 | | 5. 401367 | | 9. 28136706 | 2 2 |
| 429 | mountainous | 4. 86 | | 4. 371562 | | 9. 23156242 | 2 |
| 430 | metastasize | | 4. 2172133 | 4. 977661 | | 9. 19487478 | 2 2 2 2 |
| 431 | keystone | 3. 88 | | 5. 265871 | | 9. 14587079 | 2 |
| 432 | inversion | | 4. 5638336 | 4. 569524 | | 9. 13335806 | 2 |
| 433 | jigsaw | 5. 83 | 3. 2928926 | | | 9. 12289257 | 2 |
| 434 | rehearse | | 3. 812823 | 5. 2523 | | 9. 06512343 | 2 2 |
| 435 | linguist | 5. 83 | 3. 2351225 | | | 9. 06512252 | 2 |
| 436 | sweaty | 3. 88 | | 5. 169537 | | 9. 04953737 | 2 |
| 437 | vomit | | 1. 8486414 | 7. 200377 | | 9. 04901842 | 2 |
| 438 | stabilize | | 2. 9462723 | 6.067036 | | 9. 01330867 | 2 2 |
| 439 | witchcraft | 2. 91 | | 6. 092265 | | 9. 00226492 | 2 |
| 440 | ferret | | 2. 541882 | 6. 384636 | | 8. 92651827 | 2 |
| 441 | leech | 4. 86 | | 4. 025453 | | 8. 88545256 | 2 |
| | pup | | 2. 1952617 | | | 8. 88348463 | 2 2 |
| | beagle | 2. 91 | | 5. 944641 | | 8. 85464146 | 2 |
| | ceramic | | 3. 9283631 | | | 8. 81258552 | 2 |
| | churn | | 3. 6395128 | | | 8. 81048948 | 2 |
| | crank | | 2. 1374917 | | | 8. 7974061 | 2 2 2 2 2 2 2 |
| | snag | | 1. 9064115 | | | 8. 77129479 | 2 |
| | intersection | 3. 88 | | 4. 889963 | | 8. 76996291 | 2 |
| | postwar | | 4. 2172133 | | | 8. 76065115 | 2 |
| 450 | duplicate | | 3. 0040423 | 5. 731353 | | 8. 73539544 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|-------------|-------|------------|-----------|-----------|-------------|---------------------------------|
| 451 | rivet | 4. 86 | 3. 870593 | | | 8. 73059302 | 2 |
| 452 | engrave | 3. 88 | | 4. 846879 | | 8. 72687926 | 2 |
| | implant | | 1. 9641815 | | | 8. 72635068 | 2 |
| | cucumber | 5. 83 | 2. 8885023 | | | 8. 71850225 | 2 2 |
| | brownie | | 3. 1195824 | | | 8. 7161523 | 2 |
| | sloppy | | 2. 6574221 | | | 8. 70070529 | 2 2 2 |
| | mince | | 5. 8347746 | | | 8. 68928799 | 2 |
| | overly | | 2. 1374917 | | | 8. 67087823 | 2 |
| | collaborate | | 3. 2928926 | 5. 374211 | | 8. 66710402 | |
| 460 | kenne l | | 1. 9641815 | 6. 68491 | | 8. 64909108 | 2 |
| 461 | skepticism | | 4. 2749833 | 4. 340556 | | 8. 61553918 | 2 |
| 462 | cole | 4. 86 | 3. 7550529 | | | 8. 61505293 | 2 |
| 463 | merchandise | | 2. 4841119 | 6. 127658 | | 8. 61176967 | 2 |
| 464 | lousy | | 3. 2928926 | 5. 312042 | | 8. 60493493 | 2 |
| 465 | flake | | 3. 0618124 | | | 8. 59619367 | 2 2 2 2 |
| 466 | highland | | 3. 5239727 | 5. 059141 | | 8. 5831141 | 2 |
| 467 | babysitter | | 3. 7550529 | 4. 815333 | | 8. 57038613 | 2 2 2 2 2 |
| 468 | exce l | | 2. 7151921 | 5. 832258 | | 8. 54745024 | 2 |
| 469 | varsity | 3. 88 | | 4. 612659 | | 8. 49265875 | 2 |
| 470 | fuzzy | | 2. 4263419 | 6.063928 | | 8. 49027009 | 2 |
| 471 | scroll | 2. 91 | | 5. 570486 | | 8. 48048635 | 2 |
| 472 | buff | | 2. 6574221 | 5. 792118 | | 8. 44954025 | 2 2 2 2 |
| 473 | germ | | 2. 8307322 | 5. 603085 | | 8. 43381732 | 2 |
| 474 | hippie | 5. 83 | 2. 599652 | | | 8. 42965203 | 2 |
| 475 | pee | | 3. 5239727 | 4. 857943 | | 8. 38191612 | |
| 476 | weaver | | 2. 0797216 | 6. 292503 | | 8. 37222413 | 2 |
| | leopard | 3. 88 | | 4. 478719 | | 8. 35871938 | 2 |
| | bulge | | 2. 8885023 | 5. 470095 | | 8. 35859746 | 2 |
| | empower | | 2. 2530318 | | | 8. 3532803 | 2 |
| | clamp | | 4. 1016732 | 4. 248321 | | 8. 34999371 | 2 |
| | lapse | | 3. 9283631 | | | 8. 34345688 | 2 |
| | drier | 3. 88 | 4. 4482935 | | | 8. 32829347 | 2 2 2 2 |
| | graze | | 3. 9283631 | | | 8. 25935415 | 2 |
| 484 | bison | 2. 91 | | 5. 342736 | | 8. 25273557 | 2 2 |
| | stallion | | 2. 6574221 | | | 8. 22689487 | |
| | dent | | 3. 9861331 | | | 8. 21259521 | 2 |
| | oppress | 5. 83 | | | | 8. 19857185 | 2 |
| | safeguard | | 3. 6395128 | | | 8. 17987997 | 2 |
| | terrier | | 2. 2530318 | | | 8. 17833747 | 2 |
| | sharpen | | 2. 9462723 | | | 8. 16751979 | 2 |
| | burp | | 2. 0797216 | | | 8. 14432838 | 2 |
| | screenplay | 3. 88 | | 4. 203271 | | 8. 08327063 | 2 |
| | computerize | | 3. 6395128 | 4. 407123 | | 8. 04663583 | 2 |
| | waffle | 4. 86 | 3. 1773525 | = | | 8. 03735248 | 2 |
| | twirl | 3. 88 | | 4. 150999 | | 8. 03099896 | 2 2 2 2 2 2 2 |
| | marmalade | | 5. 141534 | | 2. 854513 | | 2 |
| | fanatic | | 2. 3108018 | | | 7. 9890519 | 2 |
| | haze l | | 3. 6395128 | 4. 310269 | | 7. 94978206 | 2 |
| | warden | 2. 91 | | | | 7. 93599392 | 2 |
| 500 | sinner | | 4. 5060635 | | 3. 425416 | 7. 93147965 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|-------------|-------|------------|-----------|-----------|-------------|---------------------------------|
| 501 | pineapple | 4. 86 | 3. 0618124 | | | 7. 92181239 | 2 |
| 502 | squirt | | 1. 9641815 | 5. 923162 | | 7. 88734367 | 2 |
| 503 | pod | | 2. 7729622 | 5. 103539 | | 7. 87650074 | 2 |
| 504 | watcher | | 3. 0040423 | 4. 836558 | | 7. 84059992 | 2 2 |
| 505 | fudge | | 2. 4263419 | 5. 372282 | | 7. 79862349 | 2 |
| 506 | leek | 5. 83 | 1. 9641815 | | | 7. 79418153 | 2 2 2 |
| 507 | rearrange | | 3. 6972829 | 4. 058754 | | 7. 75603694 | 2 |
| 508 | daydream | 2. 91 | | 4. 830974 | | 7. 74097354 | 2 |
| 509 | abide | | 3. 2928926 | 4. 445889 | | 7. 73878172 | 2 2 |
| 510 | dismantle | | 2. 2530318 | 5. 485189 | | 7. 73822084 | |
| 511 | pelican | 5. 83 | 1. 9064115 | | | 7. 73641149 | 2 |
| 512 | lumber | | 1. 9064115 | 5. 82893 | | 7. 73534134 | 2 |
| 513 | playful | 2. 91 | | 4. 808331 | | 7. 71833132 | 2 |
| 514 | ligament | | 3. 2351225 | 4. 483135 | | 7. 71825718 | 2 |
| 515 | woolly | | 2. 6574221 | 5. 056542 | | 7. 71396447 | 2 2 2 2 |
| 516 | puddle | 2. 91 | | 4. 792994 | | 7. 70299359 | 2 |
| 517 | disposable | 4. 86 | 2. 8307322 | | | 7. 69073221 | 2 |
| 518 | checkout | 5. 83 | 1. 8486414 | | | 7. 67864144 | 2 |
| 519 | genome | | 2. 6574221 | 5. 008226 | | 7. 66564852 | 2 |
| 520 | bribe | | 2. 6574221 | 4. 997148 | | 7. 65456967 | 2 2 2 2 2 |
| 521 | mermaid | 2. 91 | | 4. 741921 | | 7. 65192141 | 2 |
| | clench | 2. 91 | | 4. 733646 | | 7. 64364588 | 2 2 2 2 |
| | gutter | | 3. 7550529 | | | 7. 63505293 | 2 |
| 524 | rinse | 4. 86 | 2. 7729622 | | | 7. 63296216 | 2 |
| 525 | clog | | 3. 5817428 | 4. 041103 | | 7. 6228456 | |
| | fertilize | | 3. 6972829 | | | 7. 57728288 | 2 |
| | hospitalize | 2. 91 | | 4. 651647 | | 7. 56164729 | 2 |
| | standardize | | 2. 1952617 | | | 7. 56137383 | 2 |
| | evaporate | | 2. 8885023 | | | 7. 52324927 | 2 |
| | sweatshirt | | 3. 2928926 | | | 7. 52028493 | 2 |
| | jingle | | 2. 3108018 | | | 7. 5191864 | 2 |
| | spooky | | 2. 0219516 | | | 7. 49803854 | 2 2 2 2 |
| | gal | | 3. 0618124 | | | 7. 4959175 | 2 |
| | jaguar | | 2. 8885023 | | | 7. 45125792 | 2 2 |
| | vowe l | | 3. 5239727 | | | 7. 40397275 | |
| | sympathize | 4. 86 | | | | 7. 40188198 | 2 |
| | waterproof | 4. 86 | | | | 7. 40188198 | 2 2 2 |
| | modernity | | 3. 3506626 | | | 7. 39426271 | 2 |
| | empowerment | | 2. 6574221 | | | 7. 37506428 | 2 |
| | needy | 2. 91 | | 4. 44809 | | 7. 35809009 | 2 |
| | stationery | 4. 86 | 2. 4841119 | | | 7. 34411194 | 2 |
| | affirm | | 3. 0618124 | | | 7. 32079958 | 2 |
| | perimeter | | 3. 2928926 | 3. 999985 | | 7. 2928771 | 2 |
| | punctuation | 3. 88 | | 4 0075:- | | 7. 28843266 | 2 |
| | embedded | 2.5. | 2. 3108018 | | | 7. 24861372 | 2 2 2 2 2 2 2 |
| | greatness | 2. 91 | 0.0010515 | 4. 306176 | E 40010: | 7. 21617638 | 2 |
| | converse | 0.00 | 2. 0219516 | | 5. 138124 | 7. 16007577 | 2 |
| | lasagna | 3. 88 | | 4 045000 | | 7. 11512252 | 2 |
| | overflow | | 2. 8885023 | | | 7. 10448778 | 2 |
| 550 | teller | | 2. 2530318 | 4. 828855 | | 7. 08188658 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|---------------|-------|------------|-----------|-----------|-------------|--------------------------------------|
| 551 | clarinet | 2. 91 | 4. 1594432 | | | 7. 06944324 | 2 |
| 552 | hearer | 2. 91 | 4. 1594432 | | | 7. 06944324 | |
| 553 | northwest | 4. 86 | 2. 1952617 | | | 7. 05526171 | 2 |
| 554 | yummy | | 2. 1952617 | | | 7. 05056362 | 2 |
| | unfairly | 2. 91 | | 4. 12223 | | 7. 03223032 | |
| | clam | 2. 91 | | 4. 115544 | | 7. 02554435 | 2 |
| 557 | | | 2. 1374917 | | | 7. 02120614 | |
| | elastic | | 2. 9462723 | | | 7. 0162021 | 2 |
| | citation | | 2. 8885023 | | | 7. 00402443 | 2 |
| | windshield | 2. 91 | | 4. 090807 | | 7. 00080694 | 2 |
| | cheesecake | | 2. 1374917 | | | 6. 99749167 | 2 2 2 2 |
| | freeway | | | 4. 443899 | | 6. 98578109 | |
| | wonderland | 2. 91 | 2.011002 | 4. 04305 | | 6. 95305017 | |
| | swimsuit | 2. 91 | | 4. 033524 | | 6. 94352371 | 2 |
| | warbler | | 3. 0618124 | | | 6. 94181239 | 2 |
| | genetics | | 3. 0618124 | | | 6. 94181239 | 2 |
| | dishwasher | | 3. 0618124 | | | 6. 94181239 | |
| | stairway | 2. 91 | | 4. 03023 | | 6. 94023018 | 2 |
| | inconvenient | | 2. 0797216 | | | 6. 93972162 | 2 |
| | graffito | 2. 91 | 2. 0707210 | 4. 01825 | | 6. 92824955 | |
| | pediatric | 2.01 | 1. 9064115 | | | 6. 91022578 | |
| | demo | | 2. 1374917 | | | 6. 89344633 | |
| | mimic | | 2. 3108018 | | | 6. 85261081 | 2 |
| | pore | | 2. 3685718 | | | 6. 84904545 | 2 |
| | granddaughter | | 2. 1952617 | | | 6. 7911814 | 2 |
| | intentionally | | 2. 0797216 | | | 6. 78738584 | 2 2 2 2 2 2 2 2 |
| | conditioner | 3 88 | 2. 8885023 | | | 6. 76850225 | 2 |
| | pinpoint | 0.00 | 2. 2530318 | | | 6. 75369473 | 2 |
| | bogey | | 2. 3108018 | | | 6. 72605241 | |
| | magnify | | 1. 9641815 | | | 6. 71221698 | 2 |
| | unreal | | 2. 7151921 | | 3 996319 | 6. 71151094 | |
| | perk | | 2. 6574221 | | 0. 000010 | 6. 70872588 | 2 |
| | rerun | | 2. 0797216 | | | 6. 68747558 | 2 |
| | funnel | | 1. 9064115 | | | 6. 66817127 | 2 |
| | teapot | 3 88 | 2. 7729622 | | | 6. 65296216 | |
| | handbook | | 3. 6972829 | | | 6. 60728288 | |
| | obstruction | ۷. ا | 2. 1374917 | | | 6. 60682406 | |
| | repeal | | 2. 1952617 | | | 6. 58144131 | 2 |
| | handcuff | | 2. 1952617 | | | 6. 52185775 | 2 |
| | measurable | | 2. 4263419 | | | 6. 48022684 | |
| | tiller | | 1. 9064115 | | 4. 567222 | | 2 |
| | renovate | | 2. 1952617 | | 7. 001222 | 6. 44850083 | |
| | capitalize | | 1. 9064115 | | | 6. 44737773 | 2 |
| | celery | | 2. 3108018 | | | 6. 43778288 | 2 |
| | foggy | ე 01 | 3. 5239727 | | | 6. 43778288 | 2 |
| | | 2. 91 | | | | | 2 |
| | migrant | 2. 91 | 3. 5239727 | | | 6. 43397275 | 2 |
| | torment | 0.00 | 2. 0219516 | | | 6. 42659577 | |
| | visualize | 3. 88 | | | | 6. 42188198 | 2 |
| | ooze | | 1. 9641815 | | | 6. 41387747 | 2 |
| 600 | cannonball | | 2. 0219516 | 4. 390652 | | 6. 41260375 | 2 |

| Rank | WORD | TXT | SP | US | CL | FREQ | RANGE |
|------|--------------|-------|------------|-----------|-----------|-------------|-------------|
| 601 | embargo | | 2. 3108018 | 4. 075537 | | 6. 38633861 | 2 |
| 602 | diversify | | 2. 3108018 | 4. 03257 | | 6. 34337196 | 2 |
| 603 | jumble | 2. 91 | 3. 3506626 | | | 6. 26066261 | 2 |
| 604 | tyrant | | 2. 1952617 | 4. 043337 | | 6. 23859884 | 2 |
| 605 | redesign | | 1.8486414 | 4. 362294 | | 6. 2109354 | 2 |
| 606 | amaze | 2. 91 | 3. 1773525 | | | 6. 08735248 | 2 |
| 607 | strangle | | 1. 9641815 | 4. 060412 | | 6. 02459332 | 2 |
| 608 | slit | | 1.8486414 | 4. 128155 | | 5. 97679656 | 2 |
| 609 | triangular | 3. 88 | 2. 0797216 | | | 5. 95972162 | 2 |
| 610 | nighttime | | 1. 9064115 | 4. 014894 | | 5. 92130596 | 2 |
| 611 | grocer | 2. 91 | 3. 0040423 | | | 5. 91404234 | 2 |
| 612 | coma | 3. 88 | 2. 0219516 | | | 5. 90195158 | 2 |
| 613 | earphone | 2. 91 | 2. 9462723 | | | 5. 8562723 | 2 2 2 |
| 614 | helium | 3. 88 | 1. 9641815 | | | 5. 84418153 | 2 |
| 615 | washer | | 3. 5239727 | | 2. 283611 | 5. 8075835 | 2 |
| 616 | nightingale | 3. 88 | 1. 9064115 | | | 5. 78641149 | 2 |
| 617 | sensibly | 2. 91 | 2. 8307322 | | | 5. 74073221 | 2 |
| 618 | spank | 3. 88 | 1.8486414 | | | 5. 72864144 | 2 |
| 619 | petal | 2. 91 | 2. 6574221 | | | 5. 56742207 | 2 |
| 620 | aerobic | 2. 91 | 2. 599652 | | | 5. 50965203 | 2 |
| 621 | termite | 2. 91 | 2. 599652 | | | 5. 50965203 | 2 |
| 622 | rehabilitate | 2. 91 | 2. 541882 | | | 5. 45188198 | 2 |
| 623 | cod | | 2. 3685718 | | 2. 854513 | 5. 22308529 | 2 |
| 624 | undu l y | 2. 91 | 2. 2530318 | | | 5. 16303176 | 2 |
| 625 | snip | 2. 91 | 2. 1952617 | | | 5. 10526171 | 2 |
| 626 | wheeler | 2. 91 | 2. 1952617 | | | 5. 10526171 | 2 |
| 627 | warp | | 2. 1374917 | | 2. 854513 | 4. 99200511 | 2 |
| 628 | joiner | | 2. 0797216 | | 2. 854513 | 4. 93423506 | 2 |
| 629 | burner | 2. 91 | 1. 9641815 | | | 4. 87418153 | 2 |
| 630 | campsite | 2. 91 | 1. 9064115 | | | 4. 81641149 | 2 |
| 631 | muffin | 2. 91 | 1. 9064115 | | | 4. 81641149 | 2 |
| 632 | typist | 2. 91 | 1. 8486414 | | | 4. 75864144 | 2 2 |
| | righteous | 2. 91 | 1. 8486414 | | | 4. 75864144 | |
| | quay | | 2. 1374917 | | 2. 283611 | 4. 42110242 | 2 |
| | windowsill | | 2. 0219516 | | | 4. 30556233 | 2 |
| | teeny | | 2. 0219516 | | | 4. 30556233 | 2 |

大規模コーパスを用いた日本人英語学習者用の語彙リスト構築と 教材分析システムの開発

Construction of a Vocabulary List for Japanese Learners of English and Development of a System for Analyzing Educational Materials Based on Large-scale Corpora

平成 16-18 年度文部科学省科学研究補助金 基盤研究(B) 研究課題番号 16320076

平成 19 年 3 月 10 日発行

発行者 文部科学省科学研究費補助金 基盤研究(B)

研究代表者 相澤 一美 (東京電機大学)

〒101-8457 東京都千代田区神田錦町 2-2 東京電機大学英語系列 電話 03-5280-3463

印刷所 有限会社 サンプロセス

〒207-0012 東京都東大和市新堀 1-1435-29 Tel 042-561-8810 Fax 042-561-8813

本書の一部あるいは全部を引用または複写製版する場合には、 本書より引用したことをお断りください。