



# 生起頻度・親密度の観点からみた日本人英語学習者の自発的発話における使用語彙

横川, 博一

---

**(Citation)**

神戸大学国際コミュニケーションセンター論集, 3:59-77

**(Issue Date)**

2006

**(Resource Type)**

departmental bulletin paper

**(Version)**

Version of Record

**(JaLCD0I)**

<https://doi.org/10.24546/00519050>

**(URL)**

<https://hdl.handle.net/20.500.14094/00519050>



# 生起頻度・親密度の観点からみた日本人英語学習者の 自発的発話における使用語彙<sup>1</sup>

横川 博一<sup>2</sup>

## 1. はじめに

外国語運用能力の育成において、「語彙力」が決定的に重要であることは、経験的にも理論的にも明らかである。一般に、語彙力は、その広さ(width)と深さ(depth)といった比喻で、語彙数と語彙知識の量の観点から論じられることが多い(Nation, 2001 など参照)。また、言語の理解および産出のプロセスでは、メンタルレキシコンに格納されていると考えられている音韻、統語、意味などの語彙的情報を利用していることが明らかになっている(Levelt, 1989; こうした点について概説したものに、横川(2006)がある)。

横川・藪内・鈴木・森下(2006)では、約3週間にわたる短期海外英語研修に参加した日本人大学生を対象に実施した英語口頭能力自動測定試験(Harcourt/Ordinate Corporation 開発の Versant for English, 以前は PhonePass\_SET-10 と呼ばれていたもの)の発話データをもとに、使用語彙および発話潜時(発話開始までの時間)の観点から分析を行った。また、語彙に関しては、発話総語彙数(token)・語彙タイプ(type; 異なり語数)など使用語彙の観点から分析を行い、またリアルタイムでの言語処理能力の観点から、発話開始までの時間を計測した。その結果、(1) 海外英語研修の事前・事後で発話潜時は有意に速くなり、(2) 海外英語研修の事前・事後で発話総語彙数は有意に増加した。ただし、異なり語数およびワードファミリー(word family)には、事前と事後の間に有意差は見られなかった。このことから、3週間程度の英語研修であっても、発話潜時、つまり、相手の言っていることを理解し発話するまでの時間は短くなったことから、リスニングにおける理解がある程度高速に処理され、スピーキング・プロセスにおける語彙へのアクセスや検索などの自動化が進んだ可能性がある。そのことによって、「自分が何を言うか」という内容面の処理に注意を集中させることができ、研修後に発話総語彙数が増加したものと考えられる。

本稿では、横川・藪内・鈴木・森下(2006)で収集した発話データを、さらに、語彙の生起頻度(frequency)および親密度(familiarity)の観点から分析した結果を報告する。

## 2. 調査方法<sup>3</sup>

### 2.1 被験者

2006年2月18日～3月11日までの約3週間にわたってオーストラリア・メルボルン大学における海外短

<sup>1</sup> 本稿は、平成18年度神戸大学国際コミュニケーションセンター研究・教育プロジェクト「海外短期英語研修の口頭英語能力に関する調査」の研究成果の一部を成すものである。

<sup>2</sup> 神戸大学国際コミュニケーションセンター yokokawa@kobe-u.ac.jp

<sup>3</sup> 本稿は、横川・藪内・鈴木・森下(2006)と同一データを用いた分析であるため、本節の記述が横川ほか(2006)と一部重複することをお断りしておく。

期英語研修に参加した34名の日本人大学生(大学1~2年次生; 男子11名・女子23名)を調査対象とした。なお、被験者には聴覚障害はなかった。

## 2.2 海外語学研修の概要

学生らは3週間の英語研修プログラムを受講するが、原則として午前中は英語によるインターアクションを中心とした授業であった。主な授業内容のトピックは、Greetingsをはじめ、Exploring Melbourne / Cross-Cultural Communication / Families / Australian holidays / Weekends and recreation / Food in Australia / Sports in Australia / Australian history / Education in Australia / Work and career / Social issues / Movies and music / International travel / Multiculturalism などであり、単なる言語の研修にとどまらず、現代のオーストラリア社会をつぶさに理解する過程で、英語の力を伸ばす仕組みになっていた。

また、主として平日の午後を中心に行われた課外活動では、メルボルン大学のキャンパスツアーに始まり、市内見学、日本人会との交流、ノースメルボルンプールでの水泳、ウィリアムズタウンでの昼食会、ブッシュダンス講習、オーストラリア映画鑑賞、ブッシュ・トリップ(3日間)、フェアウェルパーティーへの参加など、体験を通じた英語研修が行われた。

なお、学生らは、すべてホームステイ(各ホームステイ先に1名)を行った。

## 2.3 英語口頭能力自動測定試験

海外短期英語研修に参加する学生の口頭英語能力を測定するため、研修の前後にHarcourt/Ordinate CorporationのVersant for Englishを実施した。研修前のプリテストは、出発3日前から前日までの間に、研修後のポストテストは帰国後3日以内に受験するよう指示した。

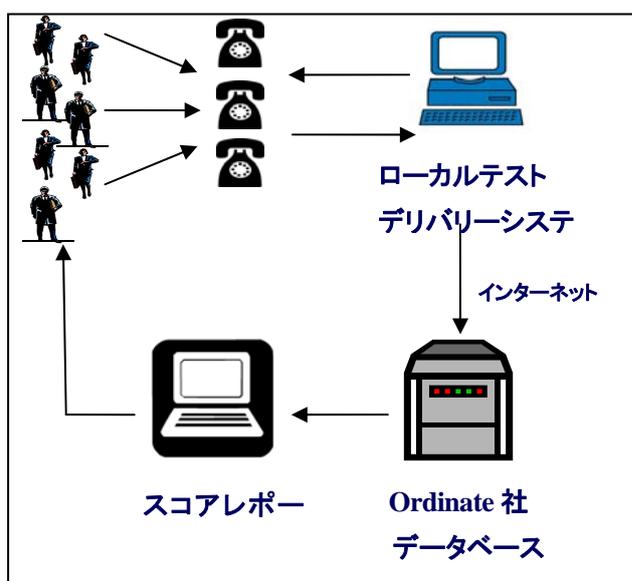


図1. SET-10の受験の流れ

本研究で使用したデータは英語口頭能力自動測定試験, Versant for English (Spoken English Test-10 minutes)を用いて収集されたものである。本試験は、米国カリフォルニア州に本社を置く Harcourt/Ordinate

Corporation ([www.ordinate.com](http://www.ordinate.com)) が、計算言語学・音声認識・音声処理技術・テスト理論の分野での研究成果を活用し、開発した自動試験である。本試験は電話上で実施され、24 時間受験が可能である。試験は音読 (Part A)、復唱 (Part B)、質問 (Part C)、文の構築 (Part D)、自由回答式質問 (Part E) の 5 つのパートから構成され、試験は約 15 分で終了する。試験終了後は音声認識システムを用いた自動採点システムにより採点される。試験終了後、数分以内でスコアレポートが Ordinate 社のホームページから閲覧可能となる。スコアレポートには、総合スコアのほか、文章構成、語彙、流暢さ、発音といったサブスコアが 20 点から 80 点の間で報告される。試験の信頼度係数は 0.97 である(この点については、Ordinate Corporation, *SET-10 Test Description and Validation Summary*, Menlo Park, CA, 20043 を参照)。

受験の際、各受験者は個別化された試験用紙と固定電話(携帯電話・ワイヤレス子機は不可)が必要である。試験用紙の表面には受験方法が説明されており、裏面が試験用紙となる。準備ができた段階で、試験用紙に印刷されてある電話番号に電話をかけ、プッシュボタンを使い、印刷されてある個別のテスト番号を入力するとテストが開始される。パート A の音読以外は、質問項目はすべて音声のみで提示される。受験の流れを図式化したものが図 1 である。

本研究での発話潜時を測定するにあたり使用した回答は、この Versant for English の Part E の自由回答式質問への音声回答である。Part E では 3 問質問され、受験者は各質問につき 20 秒の回答時間が与えられる。質問の内容は、個人の意見[例: Do you think television has had a positive or negative impact on family life? Please explain.]や好み[例: Do you like playing more in individual or in team sports? Please explain.]を問うものが多い。Versant for English では、電話でテスト番号が入力された段階で、項目バンクから問題をランダムに抽出するため、プリテスト・ポストテストともに、ある程度の重複の可能性はあるものの、各受験者間では基本的にランダムに別の質問が与えられた。

## 2.4 データ集計

今回、分析対象とした発話は、Versant for English の自由回答式質問 (Part E) に対する回答である。この問題は、1 回のテストにつき 3 問出題されるが、プリ・ポストテストのいずれかを受験しなかったテスト問題を除いて (4 テスト)、延べ 65 名分の合計 195 発話データを分析対象とした。

## 2.5 発話データの書き起こし

Versant for English の Part E の自由回答式質問を書き起こすにあたり、Ordinate 社独自の書き起こしインターフェースを利用した。この書き起こしインターフェースはウェブ上で運営されているため、パソコンとインターネット環境があればどこからでも利用可能である。書き起こしをした各個人にはユーザーネームとパスワードが与えられ、それを入力することで各個人のアカウントにログインが可能となる。ログイン後、各個人のアカウントに事前配布されている書き起こしセットを選択すると、最初の回答ファイルが自動的に再生される。書き起こし画面上には受験者が聞いた質問も表示される。書き起こしは画面上の特定のボックス内に記入していく。一音声ファイル終了ごとに、提出ボタンをクリックし、ウェブ上からインターネット経由で Ordinate 社のデータベースに送信する仕組みとなっている。

何をどのように書き起こしていくかは研究の目的により異なるが、本研究では Ordinate 社が普段使用している書き起こし基準を用いて行われた。書き起こしの基本は、回答者が各単語をどのように発音したか (Phonetic Transcription) ではなく、どの単語を発話したか (Orthographic Transcription) である。英語に存在す

る単語のみを用いて書き起こされ、英語にある単語で書き表せないものは発音の誤り、もしくは理解不能な単語 (Unintelligible) として、それぞれ別の記号を用いて処理された。また、言いよどみ (Stutter) や言い直し (Self-Correction), 「チェッ」というような単語や音素以外の口からの音 (Mouth Noise), 単語の一部のみ発話された音 (Partial Sound), 「あー」というような躊躇音 (Hesitation) などもそれぞれ別表記された。さらに背景ノイズ (Background Noise) や受験者以外の声 (Background Speech) なども注記された。

以上のような規則に基づき、筆者を含む 3 人が音声回答ファイルを書き起こした。全作業終了後、一人が全体の書き起こしデータをチェックし、必要に応じ修正を加えた後、Ordinate 社のデータベースから分析用として表計算ソフト EXCEL に一覧として抽出した。

### 3. 結果と考察

表 1, 2 は、プリテストおよびポストテストにおける発話語彙を生起頻度・親密度の観点から分析した結果を要約したものである。分析にあたっては、発話に現れた語をレマ化 (lemmatize) した。下の表には、それぞれ総発話語数に占める割合 (語彙占有率) ごとに、語数、British National Corpus (BNC) を指標とした頻度順位および平均頻度数の平均値、視覚 (文字) および聴覚 (音声) 親密度評定値の平均値が示されている。なお、親密度評定値には、日本人英語学習者を対象として調査した英単語親密度データベース (横川編, 2006; 準備中) を用いた。

表 1. プリテストにおける席頻度・親密度の観点から見た使用語彙

語彙占有率	語数	頻度順位	頻度数	親密度 (文字)	親密度 (音声)
0~50%	20	40.58	1348321	6.54	5.30
51~60%	13	84.92	481448	6.53	4.94
61~70%	24	267.54	210696	6.50	5.71
71~80%	41	343.20	206940	6.32	5.55
81~90%	84	649.89	72430	6.04	5.39
91~100%	177	819.78	41779	5.94	5.43

表 2. ポストテストにおける生起頻度・親密度の観点から見た使用語彙

語彙占有率	語数	頻度順位	頻度数	親密度 (文字)	親密度 (音声)
0~50%	17	38.63	1317839	6.55	5.27
51~60%	12	147.00	791429	6.31	5.09
61~70%	23	149.78	194271	6.41	5.21
71~80%	42	506.98	135388	6.35	5.67
81~90%	78	681.10	79741	6.00	5.35
91~100%	174	780.08	59048	5.97	5.54

まず、総発話語数を 100% としたときの発話占有率ごとの発話語数および発話語彙について見てみよう。プリテストでは 17 語、ポストテストでは 20 語で総発話語数の 50% を占め、きわめて少数の語の種類 (variety)

しか発話に用いていないことが明らかになった。具体的には、プリテストでは, I, UH, BE, BECAUSE, TO, THE, AND, CHILD, THINK, IN, MY, SO, IT, NOT, LIKE, CAN, VERY, WANT, FOR, OF, ポストテストでは, UH, I, BE, TO, BECAUSE, AND, IT, THINK, SO, CHILD, THE, WE, MY, CAN, FOR, NOT, DOといった語が現れた<sup>4</sup>。言いよどみを表すUHを除いては、自分の考えを言うためであると考えられるI, THINK, BECAUSEが共通に見られた。また、機能語(function word)の割合が大きいことも特徴的である。今回実施した試験が聞こえた質問に対して回答する形式であるということと、回答時間がわずか 20 秒という制約の中での発話であることなどを考慮しても、使用語彙の種類がかなり限定されていると言えるだろう。

次に、発話語彙と生起頻度の関係について見てみよう。表 1 および表 2 の頻度順位および平均頻度数から分かるように、語彙占有率が上がるにつれて、頻度順位は下降し、平均頻度数は上昇する傾向が見られた。この傾向は、プリテストでもポストテストでも見られた。今回の発話データに関しては、概して(BNC の)生起頻度の高い語を発話に使用する傾向があると言える。

表 3. 発話に見られた British National Corpus 生起頻度上位 3000 語以外の語

語彙占有率	プリテスト	ポストテスト
0～50%	—	—
51～60%	—	—
61～70%	—	—
71～80%	INTERNET	—
81～90%	SOCCER, TRUMP, CASUAL, DELICIOUS, HAYAO, JAPAN, KIMONO, TEENAGER	CASUAL, HOUSEWORK, INTERNET, KIMONO, NOISY
91～100%	ANIMATION, ANNIVERSARY, CALORIE, COMPUTER, COVENIENT, DEMO, ENGLISH EVERYDAY, FRANCE, ITALIA, ITALIAN, LONELY, MADE-UP, MECHANIC, MIYAZAKI, MOBILE, NEIGHBORHOOD, OPENMIND, SUSHI, THIRTY, TWO, WEEKDAY	TWELVE, CASE-BY-CASE, ENJOYABLE, ENJOYMENT GEISHA, GRANDFATHER, GRANDMOTHER, GRANDPARENT, HOMEWORK, JAPAN, OBEY, SUNLIGHT, TEN

このことに関連して、語彙占有率ごとに BNC 生起頻度上位 3000 語以外の語をリストアップしたものが表 3 である。プリテストでは、語彙占有率 0～70%の範囲には、3000 語以外の語は現れず、ポストテストでは、語彙占有率 0～80%の範囲には、3000 語以外の語は現れなかった。また、語彙占有率 90%までに現れた 3000 語以外の語は、全発話データの中で 2～3 回、語彙占有率 91～100%までに現れた 3000 語以外の語は、全

<sup>4</sup> これらの語のうち、UH は、回答中に正解が分からなかったり、言葉が出てこないため、「えー」や「あー」、「んー」といったためらい音すべてを含んだものである。また、BE には、am, is, are, was, were, be が含まれる。

発話データの中ですべて 1 回しか現れておらず、その出現率はきわめて低いと言える。具体的に語を見てみると、MIYAZAKI, HAYAO, JAPAN, FRANCE などの固有名詞、TWO, THIRTY などの数字を除けば、ANIMATION, KIMONO など質問のトピックに関連していると思われる語が見られるが、その割合はきわめて低い。

このことは、発話におけるほぼすべての使用語彙が、BNC 生起頻度上位 3000 語の範囲に収まっていることを示している。総発話語数(token)でカウントすると、プリテストでは、1766 語中 47 語が範囲外の語であり、ポストテストでは、1988 語中 24 語が範囲外の語であった。これらの総発話語数に対する割合は、それぞれ 2.7%, 1.2%である。また、実質語数(types)でカウントすると、プリテストでは、364 語中 33 語が範囲外の語であり、ポストテストでは、370 語中 18 語が範囲外の語であった。これらの総実質語数に対する割合は、それぞれ 9.1%, 4.9%である。実質語数で見れば、プリテストでは、およそ 1 割の語は 3000 語以外の語であったが、ポストテストではその割合は半減しており、上位 3000 語の語彙を多く用いるようになったと解釈することもできるだろう。

こうした結果は、中学校および高等学校における英語教科書で学習する語彙が最大で約 2900 語程度であることを考え合わせると非常に興味深い。検定教科書の語彙選定にあたっては、BNC のような言語コーパスデータも参考にされることが多いため、BNC の高頻度 3000 語と検定教科書に登場する語の一致率は比較的高いものと思われる。しかし、教科書にはそれ以外の語も多く登場しており、大学入学以前の語彙学習は教科書に限られていないにもかかわらず、いわゆる発表語彙(productive vocabulary)は、きわめて少数の高頻度の語彙にとどまった。このような意味においては、教科書に登場する語の多くを、まず発表語彙のレベルに高める学習活動を充実させることが求められるだろう。

最後に、英単語親密度(familiarity)との関係を見てみよう。表 1 および表 2 の「親密度(文字)」、「親密度(音声)」は、横川編(2006, 準備中)に収録されている、日本人英語学習者を対象に実施された英単語親密度評定値を使用したものである。数値は、「当該単語を「よく見聞きする度合い」を 7 段階評定した平均値であり、1 に近いほど親密度が低く、7 に近づくほど親密度が高くなる。プリテストおよびポストテストのいずれにおいても、語彙占有率が高まるにつれて、徐々にではあるが、「親密度(文字)」は段階的に下がる傾向が見られた。また、語彙占有率 90%までは、ほぼすべての語が親密度評定値 6 ポイント台の語であった。一方、語彙占有率と「親密度(音声)」の間には、このような相関関係が見られず、必ずしも音声としての聴覚的親密度の高い語を優先的に用いるということにはなかった。つまり、横川編(2006, 準備中)で調査された語は、BNC における高頻度語 3000 語であるが、その中でもきわめて親密度の高いと言える評定平均値 6 ポイント以上の語の使用率が高く、それ以下の親密度の語はほとんど用いられなかった。上述したように、教科書に登場する語の多くを、まず発表語彙のレベルに高める工夫が必要であるが、指導する語の優先度を決定するには、横川編(2006, 準備中)に掲載されている親密度評定値を一つの指標として活用することができるかもしれない。

Versant for English は口頭による英語能力試験であるので、音声での英単語親密度が高い語を優先的に用いる傾向が見られるのではないかと予想したが、語彙占有率と文字による英単語親密度との間に相関が見られたことは非常に興味深い。単なる予測の域を出ないが、日本人英語学習者の場合には、音声によるコミュニケーションの場合にも、リスニングの際には、音声を文字にいったん変換し、スピーキングの場合には、発話計画を音声化する場合に語を文字として表象するために、このような結果が得られたのかもしれない。

#### 4. まとめと今後の課題

本研究では, Versant for English の発話データを, 主として語の生起頻度および親密度の観点から分析した. その結果, 以下のことが明らかになった.

- 上位 17-20 語で総発話語数の 50%を占める.
- 使用語彙のほぼすべての語が BNC 頻度上位 3000 語の範囲に入る.
- 文字としての親密度の上位語を発話に使用する傾向が見られる.
- 総発話語数の 80%までは, ほぼ視覚親密度 6 ポイント台の語彙を使用.

今後は, 以下の点について考慮し, さらに調査・分析を行いたい.

- 被験者の英語の熟達度 (proficiency) による発話語彙の違いの検討.
- 発話開始までの時間などを操作することによる発話語彙の検討.

#### 文 献

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1993). Accessing words in speech production: Stages, processes and representation. *Cognition*, 42, 1-22.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.

Ordinate Corporation, *SET-10 Test Description and Validation Summary*, Menlo Park, CA, 20043.

横川博一(編) (2006). 『第二言語教育・研究のためのデータベース: 日本人英語学習者の英単語親密度<文字編>』. くろしお出版.

横川博一(編) 『第二言語教育・研究のためのデータベース: 日本人英語学習者の英単語親密度<音声編>』. (準備中).

横川博一 (2006). 「語彙と文法はいかに関連しているか」 門田修平・池村大一郎(編)『英語語彙指導ハンドブック』大修館書店. 259-271.

横川博一・藪内智・鈴木正紀・森下美和 (2006). 「日本人英語学習者の発話における使用語彙分析と発話潜時—海外語学研修参加者の発話データに基づく分析—」 電子情報通信学会『信学技報』 2006-11, 13-18.

横川博一. (2007). 「アウトプット活動で語彙力を伸ばす: 言語運用能力につながる語彙力を」『英語教育』 2月号, 20-23. 大修館書店.

Appendix: 頻度順にみたプリテストにおける使用語彙<sup>5</sup>

使用語彙	頻度	累積%	BNC 頻度順位	BNC 頻度数	親密度 (文字)	親密度 (音声)
I	127	7.19	11	928957	6.82	6.65
UH	124	14.21				
BE	90	19.31	2	4239632	6.13	4.11
BECAUSE	58	22.59	99	103003	6.40	6.43
TO	57	25.82	5	2674162	6.76	5.23
THE	46	28.43	1	6187267	6.79	3.37
AND	37	30.52	4	2687863	6.80	6.07
CHILD	37	32.62	123	71008	6.32	5.11
THINK	35	34.60	66	153881	6.43	5.35
IN	29	36.24	7	1975967	6.51	4.76
MY	27	37.77	68	152619	6.68	6.34
SO	27	39.30	44	246352	6.29	5.64
IT	26	40.77	10	1090186	6.64	5.69
NOT	26	42.24	21	465486	6.65	5.18
LIKE	24	43.60	64	162673	6.42	5.38
CAN	22	44.85	39	268699	6.34	4.16
VERY	21	46.04	88	123080	6.75	6.16
WANT	20	47.17	104	95142	6.49	6.50
FOR	19	48.24	12	898683	6.71	5.08
OF	18	49.26	3	3093444	6.41	3.57
WE	17	50.23	31	358039	6.60	4.08
FAMILY	16	51.13	221	42773	6.55	5.43
GOOD	16	52.04	93	111103	6.86	6.81
SHOULD	16	52.94	81	128393	6.16	4.00
THEY	16	53.85	26	433441	6.60	4.14
A	15	54.70	6	2186369	6.35	5.42
DO	14	55.49	18	559596	6.41	5.76
HAVE	14	56.29	8	1375636	6.43	4.41
MANY	14	57.08	110	89659	6.35	4.14
MORE	14	57.87	48	213227	6.45	4.18
WITH	14	58.66	16	675027	6.54	4.07

<sup>5</sup> 紙幅の都合で、プリテストにおける発話語彙のみ掲載する。また、表中で数値が空白になっている語は、原則として BNC 生起頻度上位 3000 語に入っておらず、横川編(2006, 準備中)で調査対象になっていない語であることを示す。

IMPORTANT	11	59.29	242	39265	6.67	5.56
PLAY	11	59.91	204	46303	6.89	6.27
COMPUTER	10	60.48	614	16976	6.80	6.40
OR	10	61.04	29	373808	6.49	5.38
PARENT	10	61.61	531	20060	5.45	4.31
FRIEND	9	62.12	301	31927	6.75	5.31
GET	9	62.63	45	220940	6.60	5.67
STUDY	9	63.14	226	41886	6.48	6.38
EAT	8	63.59	662	15446	6.38	6.30
LIFE	8	64.04	136	64423	6.55	5.75
MONEY	8	64.50	255	37892	6.61	6.72
BUT	7	64.89	23	459622	6.50	6.05
FROM	7	65.29	24	434532	6.49	5.53
INFORMATION	7	65.69	246	38656	6.13	6.68
KNOW	7	66.08	27	417737	6.49	6.53
MAKE	7	66.48	46	218145	6.62	3.19
ON	7	66.87	13	749101	6.67	5.88
READ	7	67.27	358	28216	6.58	5.81
BY	6	67.61	20	520069	6.76	5.71
DOG	6	67.95	820	12406	6.31	4.06
ENJOY	6	68.29	710	14527	6.41	6.69
GO	6	68.63	42	251301	6.87	6.30
HE	6	68.97	15	681255	6.63	5.42
IF	6	69.31	40	261089	6.40	6.28
ME	6	69.65	76	138151	6.71	4.02
NEWSPAPER	6	69.99	1162	8544	6.37	6.73
ORGANIZATION	6	70.33	473	22509	5.65	5.83
TALK	6	70.67	233	40244	6.62	5.56
THEIR	6	71.01	41	260919	6.21	5.29
TIME	6	71.35	55	185193	6.85	5.54
WILL/WOULD	6	71.69	36	305887	6.67	4.68
WORK	6	72.03	80	130090	6.49	6.72
AGREE	5	72.31	443	23497	5.86	6.31
BETTER	5	72.59	658	15626	6.47	6.27
FOOD	5	72.88	511	21044	6.76	5.33
GIVE	5	73.16	78	131417	6.69	5.42

LOT	5	73.44	292	32733	6.11	4.05
PRESENT	5	73.73	262	36806	5.85	6.40
THERE	5	74.01	34	329350	6.39	4.67
THIS	5	74.29	22	463227	6.23	5.00
WHAT	5	74.58	43	249466	6.78	6.74
AS	4	74.80	17	667715	5.96	5.78
COOK	4	75.03	1701	5183	6.49	5.74
EASY	4	75.25	503	21480	6.80	6.57
FEEL	4	75.48	140	63186	6.13	4.52
HAPPY	4	75.71	787	12854	6.75	6.42
HELP	4	75.93	183	51871	6.81	5.21
HOW	4	76.16	98	103508	6.75	5.45
LOVE	4	76.39	338	29319	6.65	5.71
NEW	4	76.61	91	115523	6.81	4.44
OTHER	4	76.84	54	185308	6.23	6.36
OWN	4	77.07	1888	4492	3.52	4.70
SCHOOL	4	77.29	181	52227	6.80	6.62
THAN	4	77.52	71	147618	6.00	5.02
THAT	4	77.75	9	1149171	6.56	5.20
THEM	4	77.97	58	173414	6.16	3.74
TOO	4	78.20	5	2674162	6.76	5.23
BEST	3	78.37	1022	9790	6.55	5.91
CHANGE	3	78.54	135	66094	6.61	6.08
CHOOSE	3	78.71	621	16731	5.56	5.50
DIFFICULT	3	78.88	482	22033	6.63	6.21
EACH	3	79.05	150	61492	5.80	5.41
EXPERIENCE	3	79.22	340	29191	6.42	6.68
HER	3	79.39	35	326968	6.05	4.24
INTERESTED	3	79.56	1265	7605	5.89	5.25
INTERNET	3	79.73				
LIVE	3	79.90	293	32675	6.38	6.23
MAYBE	3	80.07	964	10472	5.67	5.70
NOISE	3	80.24	1649	5455	5.59	5.19
OUR	3	80.41	85	125219	6.44	6.47
OUT	3	80.58	51	204252	6.44	4.56
PROBLEM	3	80.75	168	56483	6.58	5.96

RESTAURANT	3	80.92	1723	5100	5.90	6.82
SHE	3	81.09	28	380257	6.78	4.44
SOCCER	3	81.26				
SOMETHING	3	81.43	180	52452	6.11	6.27
SPORTS	3	81.60	1138	8698	6.42	6.48
TELEVISION	3	81.77	1039	9603	6.45	6.81
TELL	3	81.94	118	77245	6.24	4.83
THING	3	82.11	117	77612	6.05	3.14
TRUMP	3	82.28				
US	3	82.45	145	62350	6.15	5.43
WARM	3	82.62	1186	8381	5.41	5.50
ADULTS	2	82.73	1179	8402	6.09	5.19
AIR	2	82.84	559	19046	6.14	6.21
ALL	2	82.96	37	286441	6.50	4.08
ALONG	2	83.07	415	24314	5.78	5.29
ARTICLE	2	83.18	1029	9710	4.35	2.90
AT	2	83.30	20	534162	6.23	5.18
BECOME	2	83.41	130	67219	6.13	5.25
BOOKS	2	83.52	240	39569	6.61	6.68
BROTHER	2	83.64	870	11757	6.15	5.87
BUILDING	2	83.75	516	20770	6.48	4.07
CASUAL	2	83.86				
CELEBRATE	2	83.98	2265	3481	5.62	5.90
CLOTHES	2	84.09	468	22747	5.55	4.89
CONCENTRATE	2	84.20	1325	7157	4.46	5.35
COURSE	2	84.31	161	58614	5.35	4.56
DAY	2	84.43	106	92699	6.63	4.11
DECIDE	2	84.54	413	24380	5.31	5.10
DELICIOUS	2	84.65				
DETAIL	2	84.77	596	17584	5.13	3.60
DIRTY	2	84.88	2607	2769	5.18	4.24
EVERY	2	84.99	235	40114	6.42	5.40
EVERYBODY	2	85.11	1512	6084	6.27	5.94
FACTORY	2	85.22	1475	6259	5.65	3.98
FAVORITE	2	85.33	1607	5623	5.84	4.80
FINISH	2	85.45	740	13902	6.53	6.77

HAYAO	2	85.56				
HEALTHY	2	85.67	2073	3972	6.29	5.31
HIS	2	85.79	25	433574	6.08	5.96
HOME	2	85.90	163	58390	6.73	5.16
IMPORTANCE	2	86.01	1027	9718	5.84	4.60
JAPAN	2	86.13				
KIMONO	2	86.24				
KIND	2	86.35	350	28637	6.09	6.35
LEARN	2	86.47	448	23394	6.02	5.27
LOCATION	2	86.58	1636	5499	5.52	5.48
MIGHT	2	86.69	151	61446	5.45	5.09
MUCH	2	86.81	107	92622	6.59	5.70
NEIGHBORS	2	86.92	1763	4919	4.49	5.95
NICE	2	87.03	769	13183	6.57	6.60
NO	2	87.15	27	417737	6.92	6.53
ONE	2	87.26	114	86364	6.91	5.73
OVER	2	87.37	77	135170	6.55	5.73
PATIENT	2	87.49	398	25553	4.49	4.78
PHONE	2	87.60	988	10151	6.20	3.71
PREFER	2	87.71	1369	6854	5.37	4.60
QUIET	2	87.83	1494	6194	5.96	6.26
RAISE	2	87.94	563	18913	4.89	4.22
RESPECTED	2	88.05	804	12627	5.85	6.38
SEARCH	2	88.17	932	10822	6.04	5.97
SEASON	2	88.28	844	12161	6.47	4.34
SEE	2	88.39	52	203091	6.20	5.62
SINGLE	2	88.51	552	19192	6.45	4.05
SPEND	2	88.62	429	23799	5.43	5.27
TAKE	2	88.73	56	179220	6.75	4.60
TEENAGER	2	88.84				
TENNIS	2	88.96	2684	2635	6.42	6.25
TOGETHER	2	89.07	312	30960	6.23	6.34
USE	2	89.18	94	108820	6.29	6.23
WEAR	2	89.30	89	123048	5.66	6.21
WHEN	2	89.41	47	216276	6.48	6.53
YEARS	2	89.52	61	163930	6.68	6.73

YOUNG	2	89.64	258	37278	6.71	5.16
ABLE	1	89.69	317	30454	6.00	5.44
ABOUT	1	89.75	53	197115	6.47	6.41
ACTIVITY	1	89.81	455	23105	5.75	6.48
AFTER	1	89.86	90	116794	6.87	5.50
AGO	1	89.92	539	19808	6.20	4.86
ALSO	1	89.98	86	124884	6.16	5.47
ALWAYS	1	90.03	206	46228	6.20	6.45
ANIMATION	1	90.09				
ANNIVERSARY	1	90.15				
ANOTHER	1	90.20	157	60182	6.40	5.38
ARTIST	1	90.26	1212	8105	6.06	4.90
AVAILABLE	1	90.32	376	27184	3.98	3.64
BAD	1	90.37	397	25608	6.45	5.84
BEFORE	1	90.43	113	88275	6.56	6.17
BENEFIT	1	90.49	547	19513	4.71	5.73
BETWEEN	1	90.54	108	91141	5.89	5.49
BIG	1	90.60	289	33300	6.58	4.98
BORROW	1	90.66	2312	3353	5.13	3.71
BREAKFAST	1	90.71	1947	4314	6.21	3.45
BRINGS	1	90.77	215	43894	5.30	3.34
BUILT	1	90.83	427	23931	5.64	5.68
BUSY	1	90.88	1694	5221	6.31	5.84
BUY	1	90.94	19	545651	6.09	5.71
CALORIE	1	91.00				
CARE	1	91.05	406	25015	6.17	5.35
CAUSE	1	91.11	327	30097	5.39	4.70
COLD	1	91.17	808	12598	6.33	5.34
COMFORTABLE	1	91.22	2080	3957	5.64	4.71
COMFORTABLE	1	91.28	2080	3957	5.64	4.71
COMMUNICATED	1	91.34	2686	2634	6.13	6.65
COMMUNICATION	1	91.39	1042	9590	6.47	6.20
COMPARED	1	91.45	809	12591	4.73	4.73
COMPLICATED	1	91.51				
COMPUTER	1	91.56				
CONDITION	1	91.62	432	23742	6.11	5.38

CONTACT	1	91.68	741	13867	5.84	5.46
CONVENIENT	1	91.73				
COUNTRY	1	91.79	194	48177	6.35	6.37
DEMO	1	91.85				
DEPEND	1	91.90	992	10125	5.34	5.65
DEPENDING	1	91.96	2922	2278	4.04	4.06
DINNER	1	92.02	1404	6667	6.53	4.69
DIRECTED	1	92.07	641	15985	5.31	3.90
EDUCATION	1	92.13	391	25987	5.55	6.13
EDUCATIONAL	1	92.19	1587	5710	4.71	5.99
ELDER	1	92.24	1750	4973	3.60	4.30
ENGLISH	1	92.30				
ENOUGH	1	92.36	294	32593	5.55	5.73
ENTER	1	92.41	730	14141	6.06	4.91
EVERYDAY	1	92.47				
EXAMINATION	1	92.53	1480	6232	5.47	5.75
EXAMPLE	1	92.58	219	43402	6.11	6.87
EXERCISE	1	92.64	798	12721	5.78	6.52
EXPLORE	1	92.70	1860	4586	4.91	5.02
FACTOR	1	92.75	684	14960	4.65	4.07
FAIRLY	1	92.81	1398	6702	4.23	4.48
FAMOUS	1	92.87	1437	6485	6.10	6.51
FASHION	1	92.92	1791	4839	6.44	5.68
FATHER	1	92.98	453	23216	6.29	6.33
FIRSTLY	1	93.04	1812	4772	6.80	6.72
FOREIGN	1	93.09	635	16234	6.06	4.27
FRANCE	1	93.15				
FRIENDSHIP	1	93.20	2854	2353	5.31	4.89
FUN	1	93.26	1720	5107	6.00	5.81
GIFT	1	93.32	1960	4283	6.36	6.03
GROUP	1	93.37	149	61728	6.48	5.63
HARDLY	1	93.43	1126	8801	5.24	5.12
HEALTH	1	93.49	412	24527	6.38	6.47
HEAR	1	93.54	96	107522	6.16	6.52
HOURS	1	93.60	85	125219	6.44	6.47
HOUSE	1	93.66	192	49022	6.75	5.38

IMAGINE	1	93.71	1196	8300	5.65	6.64
ITALIA	1	93.77				
ITALIAN	1	93.83				
JANUARY	1	93.88	985	10200	6.18	6.23
KEEP	1	93.94	191	50092	6.59	5.00
LANGUAGE	1	94.00	478	22117	6.35	6.29
LARGE	1	94.05	199	47185	6.46	6.77
LEND	1	94.11	2496	2980	4.61	5.80
LETTER	1	94.17	502	21488	6.20	4.35
LIMITED	1	94.22	755	13495	5.90	3.89
LISTENING	1	94.28	847	12080	6.17	6.34
LONELY	1	94.34				
LONG	1	94.39	121	73075	6.59	5.60
LOW	1	94.45	342	29131	5.87	4.76
MADE-UP	1	94.51				
MANAGE	1	94.56	688	14894	6.02	6.13
MARRIAGE	1	94.62	1144	8668	5.78	4.90
MEAL	1	94.68	1407	6654	5.08	4.93
MECHANIC	1	94.73				
MEET	1	94.79	245	38742	6.45	5.79
MEMBERS	1	94.85	201	47141	6.35	4.57
MIND	1	94.90	302	31872	6.12	6.33
MISS	1	94.96	949	10627	6.56	5.10
MIYAZAKII	1	95.02				
MOBILE	1	95.07				
MOTHER	1	95.13	364	27784	6.60	6.60
MOVIE	1	95.19	2569	2822	6.69	5.02
MUST	1	95.24	122	72059	6.20	5.83
MYSELF	1	95.30	818	12444	5.91	6.08
NEAR	1	95.36	517	20754	6.53	6.14
NEED	1	95.41	100	99217	6.36	6.28
NEIGHBORHOOD	1	95.47				
NEWLY	1	95.53	2638	2724	3.41	3.73
NEWS	1	95.58	768	13223	6.75	5.00
NOTHING	1	95.64	284	34064	6.25	6.54
NOW	1	95.70	74	143801	6.81	6.54

ONLY	1	95.75	67	152903	6.62	6.10
OPENMIND	1	95.81				
OPPONENT	1	95.87	2486	2993	3.30	4.46
OPTIONS	1	95.92	1079	9189	5.51	6.37
OUTSIDE	1	95.98	486	21947	5.46	5.17
PEOPLE	1	96.04	84	125430	6.73	6.09
PERSON	1	96.09	344	28981	6.02	5.68
PERSONALLY	1	96.15	2195	3621	5.42	3.83
PERSUADE	1	96.21	1751	4970	3.96	3.46
PLACE	1	96.26	129	67929	6.55	4.57
PLAN	1	96.32	259	37265	6.77	6.16
POLLUTION	1	96.38	2133	3789	4.58	5.35
POSSIBLE	1	96.43	283	34178	6.27	5.59
PROGRAMMING	1	96.49	300	32068	6.14	6.48
QUESTION	1	96.55	220	43178	6.45	6.85
QUICK	1	96.60	1374	6838	5.60	5.69
QUICLY	1	96.66	822	12381	6.00	5.62
REAL	1	96.72	461	22982	6.15	5.27
RELATIONSHIP	1	96.77	564	18866	5.58	4.74
RELAXING	1	96.83	2145	3768	6.49	4.80
RETURN	1	96.89	288	33324	5.53	4.51
RIGHT	1	96.94	70	148022	6.48	5.78
ROOM	1	97.00	266	36360	6.69	4.84
SAME	1	97.06	144	62402	6.16	3.47
SECOND	1	97.11	1055	9445	6.31	6.45
SELECT	1	97.17	1328	7134	5.83	4.23
SELF	1	97.23	2130	3795	5.96	5.61
SET	1	97.28	176	54072	6.00	5.54
SIMPLY	1	97.34	592	17756	5.70	4.85
SISTER	1	97.40	1158	8592	6.36	6.69
SITUATION	1	97.45	537	19856	5.66	6.47
SMALL	1	97.51	184	51626	6.70	5.90
SMALL	1	97.57	184	51626	6.70	5.90
SOCIAL	1	97.62	228	41617	6.08	4.75
SOME	1	97.68	57	178175	6.35	4.80
SOON	1	97.73	471	22703	5.95	4.94

SPEAK	1	97.79	393	25788	6.38	6.43
SPECIAL	1	97.85	481	22040	6.39	6.71
STAY	1	97.90	521	20495	5.82	6.22
STILL	1	97.96	120	74740	5.73	5.01
STORY	1	98.02	575	18418	6.84	6.65
SUBJECT	1	98.07	297	32392	6.25	5.03
SUSHI	1	98.13				
SYSTEM	1	98.19	148	61912	6.56	6.57
TALE	1	98.24	1504	6132	4.88	4.89
TASKS	1	98.30	784	12883	5.38	6.00
TASTE	1	98.36	1562	5813	6.16	6.43
TEACH	1	98.41	983	10248	5.91	6.55
TEACHER	1	98.47	541	19744	6.49	6.82
TEMPERATURE	1	98.53	1561	5823	5.43	5.70
THEMSELVES	1	98.58	438	23673	4.86	3.96
THEN	1	98.64	65	160652	5.61	3.98
THESE	1	98.70	83	125442	6.19	3.83
THINKING	1	98.75	2959	2221	5.94	6.29
THIRTY	1	98.81				
TIRED	1	98.87	2149	3762	5.69	4.23
TOPIC	1	98.92	1922	4376	6.20	4.64
TOWN	1	98.98	525	20170	6.16	4.18
TRADITIONAL	1	99.04	1019	9819	4.98	5.40
TRAIN	1	99.09	528	20127	6.48	5.21
TRY	1	99.15	172	55799	6.55	6.21
TWO	1	99.21				
UNIVERSITY	1	99.26	893	11367	6.25	6.30
UP	1	99.32	49	211935	6.59	4.14
VARIETY	1	99.38	990	10138	5.48	3.91
VOICE	1	99.43	367	27665	5.98	5.95
WATER	1	99.49	268	35767	6.77	6.70
WAY	1	99.55	92	114866	6.39	5.62
WEEKDAY	1	99.60				
WEEKEND	1	99.66	1301	7308	6.16	4.52
WHO	1	99.72	50	205432	6.82	5.55
WORKING	1	99.77	994	10123	6.16	4.34

YES	1	99.83	155	60592	6.80	6.83
YOU	1	99.89	14	695498	6.92	6.47
YOUR	1	99.94	75	138334	6.50	4.23
YUKATA	1	100.00				

# Frequency and Familiarity of Vocabulary in the Spontaneous Speech of Japanese EFL Learners

Hirokazu YOKOKAWA

**Abstract** This study reports the results of a lexical analysis and measurement of real-time response latencies in the spontaneous speech of Japanese EFL learners who participated in a short-term (three week) overseas study tour. Thirty-four subjects took Versant for English, Spoken English Test, developed by Harcourt/Ordinate Corporation before and after the study tour. This test evaluated the ability of test-takers to understand spoken English and express themselves clearly and appropriately in English. We analyzed the speech data (Versant for English, Part E: Open Questions) in terms of (1) word frequency, and (2) word familiarity. The results are discussed in terms of cognitive processing in speech production by Japanese EFL learners and teaching implications.

**Keyword** Speaking proficiency, Short-term overseas English study tour, Word frequency, Familiarity