

英語入試問題の語彙からみた分析

福 島 孝 博

Analysis of English Entrance Examinations in Terms of Vocabulary

Takahiro FUKUSHIMA

1. はじめに

本学の入学試験の英語問題をそこで使われている単語の難易度の観点から分析を行う。まず、問題文中にどのような難易度を持った単語がどの程度使われているのかを調査する。調査は過去数年間の各種の入試問題に対して行なう。この調査・分析を行うことにより、入試問題種別ごとの単語レベルにおける単年度内での差異および経年変化が読み取れることになる。また、同様の分析がセンター入学試験の英語問題において既に行われており、センター入試問題との比較を行う。

また、問題文の読みやすさ (Readability) の点からも同様の調査・分析を行う。以下に分析の対象、使用した英語語彙リスト、分析の手法、その結果、および考察を記述する。考察においては、単語リストを使っての分析結果と読みやすさの分析結果を受けて、双方それぞれの結果の考察だけでなく、それらが入試問題の正答率とどのような関係にあるのかについても調べている。

2. 分析対象

追手門学院大学の英語入試問題、2002年度入試から2006年度入試の中から著作権上問題のない、過去の問題集に掲載されているものを調査・分析の対象とした。また、英語の問題のうち読解問題としてまとまった文章のみを対象とし、その他の文法問題やアクセント問題、並び替え問題で使われている英語の語句や文章は対象としていない。

詳細は、以下の表のとおりである。尚、○印のないところは、対象となる文章がなかったことを示している。

英語入試問題の語彙からみた分析

年度 \ 入試種別	公募制推薦 入試 A 日程	公募制推薦 入試 B 日程	一般入試 A 日程	一般入試 S 日程	一般入試 B 日程
2006 年度		○	○	○	○
2005 年度	○	○			○
2004 年度	○	○	○	○	○
2003 年度	○	○	○	○	○
2002 年度	○	○	○	○	○

表 1 分析対象の入試問題

3. 英語語彙リスト

英語の語彙リストとしては大学英語教育学会基本語リスト（JACET List of 8000 Basic Words、通称 JACET 8000）を使用した。語彙リストは、今までに種々のものが発表されているが¹⁾、本研究では、大学英語教育学会英語語彙研究会の語彙研究の成果である JACET 8000 を使用している²⁾。

JACET 8000 では、8000 語を頻度順に 1000 語毎に 8 つのレベルに分けている。各レベルの内容の程度は大まかに言って次のようである³⁾。

Level 1	中学校の英語教科書に頻出する基本単語
Level 2	高校初級のレベルの単語
Level 3	高校英語教科のレベルの単語
Level 4	大学受験、大学一般教養の初級に相当
Level 5	難関大学受験、大学一般教養に相当
Level 6	英語を専門としない大学生やビジネスマンが目指すレベル
Level 7	英語専攻の大学生やビジネスマンが到達目標とするレベル
Level 8	日本人英語学習者の一般的な単語学習の最終目標

表 2 JACET 8000 の 8 レベル

JACET 8000 には、レベル分けされた 8000 語以外に、固有名詞など 250 語を追加したプラス 250 があるが、今回の分析では使用していない。

4. 分析の手法

第一に、JACET 8000 を使い、英語問題における文章中に使われている単語のレベルを調べ、

どのレベル単語が何度出現しているのか（頻度）を入学試験の問題毎に集計している。この分析には、Web で公開されているプログラム (<http://www.01.tcp-ip.or.jp/~shin/j8web/j8web.cgi>) を使用した。

第二に、英語読解問題の文章を単語数など以下の項目について算出している。

- 文字数
- 単語数
- 文数
- 一単語における平均文字数
- 一単語における平均シラブル数
- 一文における平均単語数

次にこれらの結果を利用して、以下の読みやすさ（Readability）に関する3個の指標について計算する。

- GF Index : Gunning Fog Index
- FK Grade : Flesch Kincaid Grade level³⁾
- Flesch RE : Flesch Reading Ease⁴⁾

Gunning Fog Index は英文を読みやすさをその英文を理解するために必要な教育年数で算出する。また、Flesch Kincaid Grad level は英文を理解する米国の学校における学年レベルにして示すものである。最後の Flesch Reading Ease は、読みやすさに関する代表的な指標の一つであり、この数値が低いほど、読みやすさが低くなる、つまり読むのが難しくなる²⁾。

これらの3個の指標は、文中の単語数や単語のシラブル数を計算の基礎としている。例えば、Gunning Fog Index は以下の式で計算する。

$$\text{Gunning Fog Index} = 0.4 \times \text{文長 (1 文中の単語数) の平均} + 100 \times \text{全単語中のシラブルが 3 以上の単語の割合}$$

読みやすさの指標の計算は、Web 上で公開されているソフトを利用した。(http://www.online-utility.org/english/readability_test_and_improve.jsp)

5. 分析結果

まず、JACET 8000 を使った語彙レベル別の使用頻度であるが、各レベルに毎に出現の割合（パーセント）を異なり語数（Indexes）と延べ語数（Tokens）で集計した。詳細な結果は Appendix に載せている。

先行研究では、レベル4（頻度の順位が1から4000まで）の単語で、分析の対象となる単

英語入試問題の語彙からみた分析

語のどれくらいをカバーするかを算出している^{6,7)}。同様に、今回のデータをレベル 4 までの単語が語彙全体のどれだけカバーしているかをまとめ次の表のとおりとなる。

結果を示した表の第一列目の記号は入試問題種別を示している。KB は公募制推薦、IP 一般入試を、A、B、S は日程の別を、その直後に付く 1 または 2 は、何日目かを示している。最後の丸括弧内の数字は、読解問題が 2 題ある場合、何題目の問題かを表している。例えば、KB A-1 (I) は、公募推薦入試 A 日程の 1 日目の読解問題の第一問目を意味する。尚、2006 年度入試の一般 S 日程は、第二日目のものである。

	2002	2003	2004	2005	2006	Average
KB A-1	85.714	92.064	86.033	85.992		87.451
KB A-2	93.814	94.521	82.265			90.200
KB B	94.012	84.259	82.353	89.163	85.714	87.100
IP A-1 (I)	87.865	84.771	82.301			84.979
IP A-1 (II)	84.545	86.701	76.271		87.672	83.797
IP A-2 (I)	92.432	85.646	86.931		84.236	87.311
IP A-2 (II)	87.685	92.147	89.352		83.853	88.259
IP S (I)	93.479	87.755	80.477		84.906	86.654
IP S (II)	90.452	86.238	85.909			87.533
IP B (I)	82.467	92.771	88.203	90.909	88.726	88.615
IP B (II)	89.081	88.462	78.612		86.606	85.690

表 3 レベル 4 までの単語の割合 (異なり語数)

入試種別によりばらつきは見られるが、レベル 4 までの異なり語では、84% 弱から 88% を超える程度の出現割合となっている。

Tokens	2002	2003	2004	2005	2006	Average
KB A-1	90.909	95.122	92.548	90.261		92.210
KB A-2	95.239	96.315	87.528			93.027
KB B	95.705	89.325	89.884	91.422	90.909	91.449
IP A-1 (I)	91.436	88.55	89.877			89.954
IP A-1 (II)	90.661	91.215	85.439		92.721	90.009
IP A-2 (I)	94.582	90.808	91.512		87.229	91.033
IP A-2 (II)	91.709	95.909	93.763		90.697	93.020
IP S (I)	96	92.84	86.512		88.838	91.048
IP S (II)	94.6	90.971	90.719			92.097
IP B (I)	83.696	94.445	93.466	95.086	93.334	92.005
IP B (II)	92.857	93.154	87.03		91.177	91.055

表 4 レベル 4 までの単語の割合 (延べ語数)

延べ語数では、レベル 4 までの単語の割合は、90% 弱から 93% となっている。

次に、読みやすさの 3 種類の指標（Readability）に関する分析結果を示す。

年度	2002			2003			2004		
	GF Index	FK Grade	Flesch RE	GF Index	FK Grade	Flesch RE	GF Index	FK Grade	Flesch RE
KB A-1	11.13	9.48	60.01	6.82	5.55	77.44	9.21	7.09	67.36
KB A-2	5.90	4.49	80.94	8.01	5.82	72.14	11.05	8.72	65.13
KB B	7.93	6.55	67.53	8.07	6.65	69.12	14.73	12.24	44.31
IP A-1 (I)	11.81	10.26	54.26	9.52	7.92	67.62	13.18	12.08	43.29
IP A-1 (II)	15.35	13.79	51.39	8.85	7.35	62.11	11.25	9.13	49.93
IP A-2 (I)	9.37	7.80	64.84	9.91	7.41	68.11	13.96	11.53	42.91
IP A-2 (II)	10.89	9.32	61.92	14.30	11.71	48.62	12.36	9.94	51.90
IP S (I)	12.56	10.82	52.33	12.57	10.19	57.62	8.27	7.54	67.56
IP S (II)	10.29	9.80	49.31	9.82	7.65	72.53	12.13	10.19	59.47
IP B (I)	8.23	6.84	67.2	11.11	9.05	60.38	11.66	10.23	56.33
IP B (II)	14.57	12.1	46.78	14.02	11.9	52.62	9.4	8.82	56.68

表 5 Readability Scores (2002-2004 年度)

年度	2005			2006		
	GF Index	FK Grade	Flesch RE	GF Index	FK Grade	Flesch RE
KB A-1	9.77	8.44	63.08			
KB B	11.67	10.56	50.56	7.99	6.64	71.47
IP A-1 (II)				10.35	8.16	58.15
IP A-2 (I)				12.28	10.36	57.16
IP A-2 (II)				12.17	9.82	48.33
IP S-2 (I)				10.77	8.62	57.88
IP B (I)	7.88	6.21	72.51	10.47	8.99	63.32
IP B (II)				9.88	8.11	57.12

表 6 Readability Scores (2005-2006 年度)

(注：表 6 において、データの存在しない入試種別の行は削除している。)

6. 考 察

JACET 8000 からみた単語出現度であるが、レベル 4 までの単語の出現の割合を、公募制推薦入試（A、B 日程）、一般入試（A、S 日程）および一般入試 B 日程の 3 種類に大別して、それぞれの平均値を出すと以下のようなものである。

英語入試問題の語彙からみた分析

差は少しであるが、3種類の入試のなかで一番易しいとされる公募制推薦入試は、異なり語数、延べ語数の双方において、レベル4までの単語が多く使用されていることが分かる。同時に、一番難しいとされる一般入試のA、S日程の問題では、レベル4までの単語の使用の割合が一番低くなっている。

北尾ら⁷⁾提供するデータは、大学入試センターが実施するセンター入試の読解問題（1990年度から2005年度まで）を、本研究と同じソフトを使ってJACET 8000のレベル4までの単語の出現度合を分析しているのでセンター入試との比較を行う。対象は2001年度から2005年度までの5年間分とした。

年度	2002	2003	2004	2005	2006	Average
異なり語数						
KB-A, B	91.180	90.281	83.550	87.578	85.714	88.250
IP-A, S	89.410	87.210	83.540		85.167	86.422
IP-B	85.774	90.617	83.408	90.909	87.666	87.153
延べ語数						
KB-A, B	93.951	93.587	89.987	90.842	90.909	92.229
IP-A, S	93.165	91.716	89.637		89.871	91.193
IP-B	88.277	93.800	90.248	95.086	92.256	91.530

表7 レベル4までの単語の出現割合（3区分）

年度	2001	2002	2003	2004	2005	Average
異なり語数	86.018	83.581	87.058	87.714	85.989	86.072
延べ語数	90.779	89.792	90.949	91.716	91.068	90.861

表8 センター入試におけるレベル4までの単語の出現割合（%）

表7と表8を比較すると、レベル4までの単語の出現割合においてセンター入試の読解問題は、一般入試（A、S日程）に良く似た数字を残している。単語のレベルだけから言うと公募制の入試や一般入試（B日程）の問題は、センター入試の問題よりも易しいと推測できる。

次に、読みやすさ（Readability Score）の点からの分析であるが、データの中で一番最近の年度である2006年度入試を更に分析する。2006年度入試では、公募制が1個、一般入試のA、S日程として4個、一般B日程2個の英文入試問題が分析の対象となっており、GF IndexとFK Gradeをグラフにすると図1のとおりとなる。

これら2つの指標とも値が小さく、それだけ読みやすさが増すことになるが、傾向として、公募制入試が一番易しく、次に、一般B日程そして一般入試A、S日程が続くことが見て取れる。これは、JACET 8000を使ってのレベル4までの単語の出現度の傾向と同じである。

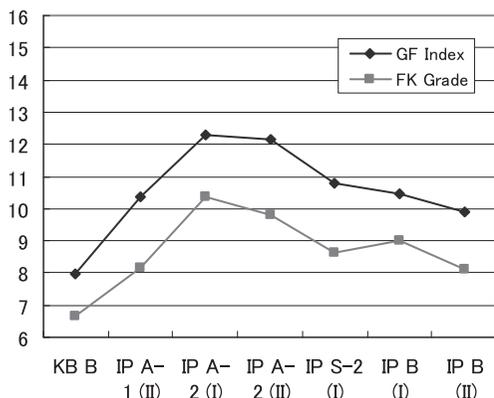


図 1 Readability Scores (2006)

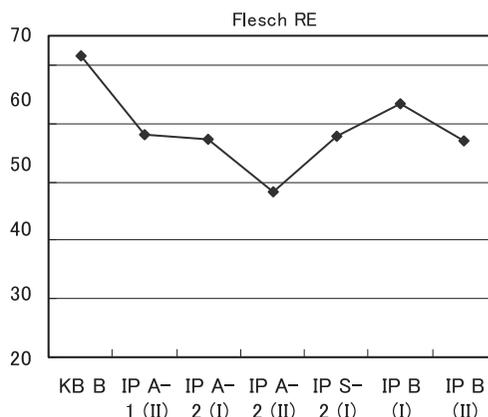


図 2 Readability Score (FRE, 2006)

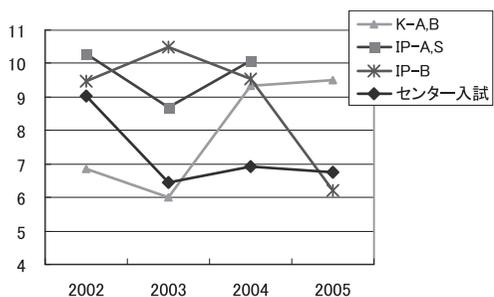


図 3 FK Grade 値の比較

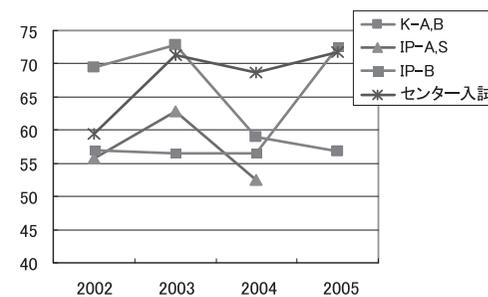


図 4 Flesch RE 値の比較

Flesch Reading Ease の値においても、公募制入試の問題が一番読みやすく、一般入試 A、S 日程の問題が一番読んで理解するのが難しいものとなっている傾向を示している（図 2）。前出の北尾ら⁷⁾提供するデータには、センター入試問題の readability scores（FK Grade と Flesch RE）も含まれている。本学の入試との比較を表にまとめると図 3、4 のようになる。ここでも、単語出現割合で行ったように、公募制、一般入試 A、S および一般入試 B 日程の 3 区分し平均した値との比較を行っている。

図 3、4 のグラフから分かるのは、2 つの Readability 指標において 2002 年度、2003 年度の公募制入試（A、B 日程）を除いては、センター入試問題のほうが理解しやすい英文であったということである。また、一般 A、S に関しては、JACET 8000 を用いたレベル 4 までの単語の割合の分析結果とは異なり、センター入試の英文のほうが読みやすい傾向を示している。

考察の最後に入試の正答率との関係を分析する。単語リストを利用しての分析結果である単語の出現割合（L4 までの Index のものと L4 までの Token ベースのもの 2 種類）、Readability の指標 3 種類（GF Index, FK Grade, Flesch RE）と入試の問題の正答率との相関関係を算出し

英語入試問題の語彙からみた分析

	L 1-L 4 Index	L 1-L 4 Token	GF Index	FK Grade	Flesch RE
Ave (1-8)	-0.1440	-0.0540	0.4534	0.3756	-0.5678
Ave (9-12)	-0.8339	-0.8834	0.7441	0.7346	-0.6003
Ave (1-12)	-0.4429	-0.3924	0.6483	0.5839	-0.6804

表9 問題正答率との相関関係

た。入試問題は、2006年度の一般B日程、A日程、S日程とB日程（一問目のみ）とした。問題は、各入試とも読解対象の英文1個につき12問出されるが、ここでは、問1から問8まで、問9から問12まで、それに全体である問1から問12までの3区分にて正答率の平均を出し、その平均値との相関係数を算出している。問1から8までは、特定の語句や表現に関する問題であり、問9から12は文章全体を読んだの内容理解を問うものであり、別扱いとした。

はっきりとした相関がみられるのは、問9から12までの正答率の平均値とL4までの単語の出現率（Index, Token 双方とも）およびReadabilityの各指標である。L4までの単語の出現割合との関係は負の関係であり、L4までの単語の出現率が高い程、正答率が低くなるという直感とは反する結果を示している。また、Readabilityの指標との関係においても、読みやすさにおいて易しい英文において正答率が低くなっている傾向がみられる。

なぜ、このような結果となったかを知るには更なる分析が必要である。また、分析対象のデータ数が少ないため、一般化するのには困難であるが、今回のデータの傾向としてL5以上の難易度の比較的高い単語を含む英文においては、内容理解の問題が易しくなっている。

7. ま と め

本学の英語入試問題を語彙のレベルと読みやすさの観点から分析を行った。分析結果から分かる事項をまとめると以下ようになる。

- JACET 8000を使った語彙レベルの分析では、レベル4までの単語の出現割合からみると、公募制、一般B、一般A、Sの順に値が低くなる。
- 語彙レベルの分析結果をセンター入試問題の分析結果と比較すると、センター入試の問題は、レベル4までの語の出現割合が一般A、Sに相当することが判明した。
- 読みやすさ（Readability）の指標においては、2006年度の入試種別での比較をおこない、公募制、一般B、一般A、Sの順に読みやすさが低くなる、つまり読んで理解することが難しくなることが分かった。
- 読みやすさの指標において、センター入試問題と比較すると、2002、2003年度の公募制入試問題を除いて、本学の入試問題のほうが読むのが難しい傾向があることが分かった。

- 問題の正答率との相関関係においては、L4 までの単語の出現割合、Readability 指標において、易しいであろう英文の問題に関して正答率が低いという、直感に反する結果が示された。

今後の検討事項としては、まず、単語分析の際の品詞の問題がある。JACET 8000 を使用しての語彙レベルの分析において、Web 上のプログラムを使用したがる、このプログラムでは、単語の品詞まで考慮されていない。一方、JACET 8000 の単語は品詞まで特定されており、今後は単語の品詞まで考慮した分析が必要である。

また、入試の英語問題には注のある単語（語句）があるが、今回は、それらの単語を全く考慮せずに分析を行った。注の付与されている単語については、今後、調査分析において別の扱いを行う必要がある。

これらの検討事項は2件とも、英文の読み手が実際に感じるであろう英文の難しさ・易しさに関係するものであり、これらを考慮することにより、実際の英文の難しさをより良く反映する研究になると期待される。

参考文献

- 1) 高梨庸雄、卯城祐司編、『英語リーディング事典』研究者出版、2000年。
- 2) 大学英語教育学会基本語改訂委員会 著作編集、『大学英語教育学会基本語リスト JACET List of 8000 Basic Words』、2003年。
- 3) 相澤一美、石川慎一郎、村田年編集代表、『JACET 8000 英単語』桐原書店、2007年。
- 4) Flesch, R., "A new readability yardstick", *Journal of Applied Psychology*, 1948. 32 : p. 221-233.
- 5) Kincaid, J.P., et al., Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. *CNTECHTRA Research Branch Report*, 1975 : p. 8-75.
- 6) 望月正道、JACET 8000 の有効性と問題点：大学入試問題分析から、JACET 8000 活用事例集 大学英語教育学会 基本語改訂委員会（編）2004年。
- 7) Kitao, Kenji, Kitao, S. Kathleen, "Corpus-Based Analysis of Japanese University English Entrance Exams", *Asia TEFL International Conference*, August 18-20, 2006. (データは <http://www.cis.doshisha.ac.jp/~kkitao/Japanese/bio/present.htm#24> に掲載されている)

使用プログラムの URL

- ・ JACET 8000 単語リストに基づく語彙分析
<http://www01.tcp-ip.or.jp/~shin/j8web/j8web.cgi>
- ・ 読みやすさの指標の算出
http://www.online-utility.org/english/readability_test_and_improve.jsp

英語入試問題の語彙からみた分析

Appendix

JACET 8000 を使った語彙レベル別の使用の割合（パーセント）を示す。以下の表では、レベル 1（1000 語レベル）、レベル 1 から 4 まで（4000 語レベルまで）、およびレベル 1 からレベル 8（8000 語レベルまで）までの 3 種類の割合を示す。尚、L 1-L 8 の列の値が 100 とならないが、これは、8000 語レベル以上の単語や、固有名詞、記号があるからである。

（注：表 12、13 において、該当データの存在しない入試種別の行は空白としている。）

年度 レベル	2002			2003			2004		
	L 1	L 1-L 4	L 1-L 8	L 1	L 1-L 4	L 1-L 8	L 1	L 1-L 4	L 1-L 8
KB A-1	71.921	85.714	89.163	78.307	92.064	93.122	73.184	86.033	88.827
KB A-2	82.990	93.814	95.36	84.932	94.521	95.891	68.473	82.265	84.728
KB B	73.054	94.012	97.605	69.907	84.259	91.667	63.235	82.353	87.255
IP A-1 (I)	66.990	87.865	93.689	69.543	84.771	90.355	53.54	82.301	88.495
IP A-1 (II)	57.727	84.545	90.908	72.34	86.701	89.361	51.977	76.271	81.921
IP A-2 (I)	76.216	92.432	95.135	69.378	85.646	89.473	66.477	86.931	92.612
IP A-2 (II)	61.084	87.685	93.596	76.44	92.147	95.289	68.519	89.352	92.13
IP S (I)	76.087	93.479	96.196	73.469	87.755	92.857	62.381	80.477	90.953
IP S (II)	66.834	90.452	97.488	68.807	86.238	93.578	67.727	85.909	89.091
IP B (I)	67.532	82.467	88.311	75.301	92.771	96.987	70.225	88.203	92.698
IP B (II)	70.690	89.081	95.403	65.385	88.462	95.055	57.803	78.612	83.814

表 10 レベル別の単語出現割合（2002-2004 年、異なり語数）

年度 レベル	2002			2003			2004		
	L 1	L 1-L 4	L 1-L 8	L 1	L 1-L 4	L 1-L 8	L 1	L 1-L 4	L 1-L 8
KB A-1	76.818	90.909	93.41	87.073	95.122	95.61	79.087	92.548	94.231
KB A-2	89.116	95.239	95.92	91.155	96.315	96.807	78.231	87.528	88.663
KB B	84.487	95.705	98.808	79.956	89.325	93.247	78.161	89.884	92.413
IP A-1 (I)	78.086	91.436	94.711	79.644	88.55	91.603	67.654	89.877	93.334
IP A-1 (II)	72.210	90.661	94.534	82.432	91.215	93.242	70.055	85.439	89.285
IP A-2 (I)	86.946	94.582	96.06	78.027	90.808	92.826	76.127	91.512	95.491
IP A-2 (II)	71.357	91.709	95.729	85.909	95.909	97.273	79.834	93.763	95.219
IP S (I)	85.067	96	97.6	83.771	92.84	96.182	75.349	86.512	94.186
IP S (II)	78.169	94.6	98.591	80.813	90.971	96.164	78.654	90.719	92.575
IP B (I)	69.837	83.696	88.587	84.314	94.445	96.733	82.102	93.466	96.022
IP B (II)	80.745	92.857	97.204	78.869	93.154	97.321	68.942	87.03	90.102

表 11 レベル別の単語出現割合（2002-2004 年、延べ語数）

年度	2005			2006		
	L 1	L 1-L 4	L 1-L 8	L 1	L 1-L 4	L 1-L 8
レベル						
KB A-1	66.184	85.992	86.958			
KB A-2						
KB B	71.429	89.163	93.597	71.921	85.714	89.163
IP A-1 (I)						
IP A-1 (II)				63.014	87.672	92.694
IP A-2 (I)				65.517	84.236	90.639
IP A-2 (II)				64.583	83.853	88.021
IP S-2 (I)				64.623	84.906	90.095
IP S-2 (II)						
IP B (I)	73.797	90.909	92.514	70.098	88.726	94.607
IP B (II)				67.857	86.606	90.624

表 12 レベル別の単語出現割合 (2005-2006 年、異なり語数)

年度	2005			2006		
	L 1	L 1-L 4	L 1-L 8	L 1	L 1-L 4	L 1-L 8
レベル						
KB A-1	75.059	90.261	90.974			
KB A-2						
KB B	81.49	91.422	95.711	76.818	90.909	93.41
IP A-1 (I)						
IP A-1 (II)				78.373	92.721	95.719
IP A-2 (I)				73.976	87.229	92.29
IP A-2 (II)				73.643	90.697	93.022
IP S-2 (I)				75.626	88.838	94.077
IP S-2 (II)						
IP B (I)	85.749	95.086	96.315	80.714	93.334	96.429
IP B (II)				77.54	91.177	94.386

表 13 レベル別の単語出現割合 (2005-2006 年、延べ語数)