

語彙性判断に基づく英語語彙処理テストの開発

長谷 尚弥*・氏木 道人**・門田 修平***

The Development of Computer-based English Lexical Processing Test Based on the Lexical Decision Task

Naoya HASE, Osato SHIKI, Shuhei KADOTA

要旨：近年の「使える語彙力」の必要性に鑑み、語彙の流暢性測定法の開発が不可欠である。私達の研究グループでは、語彙処理能力測定ツールとして「コンピューター版英語語彙処理テスト Lex 版」(以後 CELP-Lex) を新たに開発した。このテストは、英単語親密度リストから選定された語彙をもとに、提示された単語が実際に存在するか否かを判断するという語彙性判断課題を採用することで語彙処理能力を測定するものである。本稿では、CELP-Lex の開発のコンセプト及び手順について報告する。

Abstract :

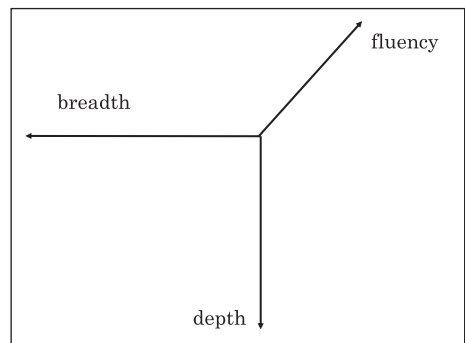
Given the necessity of a workable knowledge of vocabulary in English, it is inevitable that a vocabulary test that can measure ‘lexical processing fluency’ be developed. We have developed a ‘Computer-based English Lexical Processing Test Based on the Lexical Decision Task (hereafter CELP-Lex).’ Using the list of words taken from the Familiarity List of English Vocabulary, this test adopts the lexical decision task and measures lexical processing fluency. This paper reports the theoretical concepts behind and procedures of the development of CELP-Lex.

キーワード：コンピューター版語彙テスト、語彙運用力、語彙性判断課題

1. はじめに

国際共通語としての英語の地位が確立している現状においては、国際コミュニケーションのツールとして実用に耐えうる英語力がますます求められている。語彙力についても同様である。ではこのようなコミュニケーションに耐えうる語彙力とはどういったものであろうか。

Daller et al. (2007) は、語彙を以下の図 1 で示す breadth (広さ)、depth (深さ)、fluency (流暢さ) の 3つの要素で説明している。



(Daller et al., 2007, p.8 に基づく)
図 1 語彙の 3要素

* 関西学院大学国際学部教授

** 関西学院大学理工学部准教授

*** 関西学院大学法学部教授

すなわち、breadth（広さ）とはどのくらい多くの語彙を知っているかという指標であり、これを測るためには例えば Vocabulary Levels Test (Nation, 2001) などがある。そして depth（深さ）とは語彙の深さであり、どれだけその語彙の用法を詳しく知っているかという指標である。これを測るには例えば Word Association Test (Read, 2000) などがある。最後に fluency（流暢さ）であるが、これはどれだけ速く、そして安定的に語彙処理（メンタルレキシコンにアクセスし、その単語の意味を取り出す）ができるかということであり、この能力を測定するためには例えば Computer-based English Lexical Processing Test Based on the Semantic Judgment Test（以後 CELP-Sem）（門田他、2010）などがある。これらのうち、breadth と depth は「知識領域」に属するものであり、fluency は「運用領域」に属するものであると考えられる。以下の図2に示した通りである。

真の意味で使える英語力を身につけるためには宣言的知識（declarative knowledge）を手続き的知識（procedural knowledge）へと移行させなければならない。前者は意識によって統制された（controlled）形で利用され、後者は自動化された（automatic）形で利用可能である。同様のことが語彙知識においても言える。つまり、「使える語彙力」とは、意味アクセスが自動的に行える語彙知識のことである。語彙における宣言的知識（declarative knowledge）とは、図2中の「知識領域」、手続き的知識（procedural knowledge）とは「知識領域」に「運用領域」を併せたものと考えることが出来る。

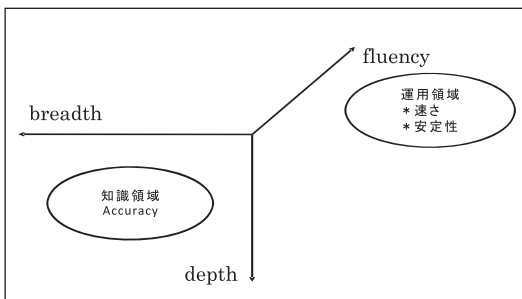


図2 語彙の3要素と語彙力の2領域

2. CELP-Sem について

2.1 CELP-Sem の概要

門田他（2010）では、前述の CELP-Sem の開発経緯やその妥当性、信頼性について報告した。これは、コンピュータ上で語彙処理速度（メンタルレキシコンへのアクセス速度）とその正確さを測るものであり、その際のタスクとしては類義性判断課題（semantic judgment task）を用いた。使用した単語については、英単語親密度リスト（横川他、2006）から日本人英語学習者にとってのなじみ深さという指標をもとに3000語レベル以内の語を選定した。手法としてはプライミング法を用いたが、この方法は、先行刺激（プライム語）が後続刺激（ターゲット語）の処理にどのような影響を与えるかを検討するためにしばしば利用されるパラダイムである。CELP-Sem の判定手順をプライム語とターゲット語としてそれぞれ house と home を例にとって説明する。はじめにコンピュータ画面に house が提示され、次に home が提示される。連続して提示された2つの単語が意味的に類似しているかどうかの判断を学習者に求め、判断の正誤とともに判断するまでに要した時間（反応時間 response time）が記録される。両者の値をデータとすることで、学習者がターゲット語の意味を正しく理解しているかどうかだけでなく、どのくらい速くその語の意味を取り出すことが出来るかを測定できる。以上の流れを図示すると次の通りとなる。



図3 CELP-Sem の判定手順

2.2 CELP-Sem の妥当性・信頼性

CELP-Sem の妥当性、信頼性についてはすでいくつかの実証研究がある。まず、門田（2010）は、(1) CELP-Sem の A 版と B 版は極めて均質的で等価なテストである、(2) CELP-Sem は、語彙の顕在的な知識領域と語彙の潜在的な運用領域の両方をデータとして算出するものであるが、両

データ間にはまったく有意な相関はなく、両領域は実は全く異なる能力である、(3) CELP-Sem と Vocabulary Size Test total (Nation & Beglar, 2007) の相関は、 $r = .55$ 、 $r = .51$ の相関を示すことを明らかにしている。また、野呂 (2010) によると、CELP-Sem と、多読により養成されると考えられている英語読解力を測る EPER (Edinburgh Project on Extensive Reading) の Reading Test との相関が $r = -.42$ となっており、中程度の相関を示している。最後に Shimamoto & Ikemura (2011) によると、CELP-Sem と TOEFL が $r = .45$ という中程度の相関関係を示している。以上のように、語彙力標準テストとの相関性が中程度であるということは、CELP-Sem が、英語語彙の潜在的な運用能力を測定するだけでなく、その知識領域テストとしての妥当性もある程度兼ね備えていることを示唆している。

3. CELP-Lex の開発

3.1 開発の必要性

これまで説明してきた CELP-Sem はタスクとして類義性判断課題を採用している。つまり、プライム語とターゲット語の間の意味的関連性を判断させるものであるが、これにはどうしても判断における曖昧さが残る。意味的関連性判断には個人差があり、また、プライム語とターゲット語の組み合わせによっては、個人差とは別に、明瞭な判定が困難な場合もあるからである (例えば student-pupil など)。

一方、これまでメンタルレキシコン (心内辞書) に関する多くの研究は、語彙性判断課題 (lexical decision task) を採用している。語彙性判断課題とは、ある単語を呈示し、それが実在する実単語であるか、または存在しない非単語 (疑似単語) であるかを瞬時のうちに判断させるものである。この課題について Harrington and Carey (2009) は、第二言語としての英語語彙運用テストにおいては語彙性判断課題を採用した方がより妥当性が高いとし、語彙性判断課題を利用した語彙テストのスコア (正答率と反応時間の両方) と Proficiency テストとの相関が、語彙テストの正答率と全体のスコアでは $r = .54$ 、Listening のスコア

では $r = .52$ 、Grammar のスコアでは $r = .53$ 、Writing のスコアでは $r = .53$ 、Speaking では $r = .41$ と、いずれも有意な相関を示したと報告している。また語彙テストの反応時間との相関においても、各々 $r = -.42$ 、 $r = -.27$ 、 $r = -.29$ 、 $r = -.30$ 、 $r = -.27$ と有意な相関があると報告している。

他方、語彙運用力を測定するタスクとしての語彙性判断にも問題がないわけではない。その一つが、語彙性を判断する際に単語の意味をメンタルレキシコンから検索しないで、つまり、意味処理なしにその単語が実際に存在するものかどうかを判断してしまうのではないかという問題である。多義語への意味アクセスを調査した Miki (2010) によると、複数の意味を持つ多義語のうち、日本人英語学習者により馴染みのある意味を提示した方が、そうではない意味を提示したときよりも意味関連性判断が迅速に行われた。しかし、同条件で語彙性判断をさせた場合、その反応時間に差が出なかったとしている。実験では、money と bank の間に意味関連性があると判断するほうが、river と bank の間での判断よりも速いが、語彙性判断ではいずれの組み合わせにおいても反応時間には有意差がなかったと報告している。これは、語彙性の判断をする際に、その語の意味アクセスをしない可能性を示唆したものである。この結果からすると、語彙の意味処理能力を測定する上で語彙性判断課題を用いることには慎重でなければならない。

3.2 開発の手順

CELP-Lex は、CELP-Sem が類義性判断課題を利用しているのとは異なり、視覚提示された単語が実際に存在するか否かを判断する語彙性判断課題 (lexical decision task) を採用した語彙処理能力測定テストである。本語彙テストの開発に着手した理由は、前節で既に指摘したように、連続提示された単語ペア間の類義性の判断において問題となった曖昧性がこのテストでは解消できると考えたからである。

テスト形式は、プライミング手法は用いていないものの、CELP-Sem と同様、提示されるターゲット語が実在するかどうかを判断するまでにかか

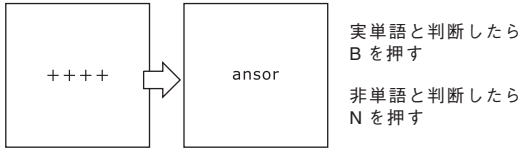


図4 CELP-Lex の流れ

る時間（反応時間）及び判断の正答率を測定し、それにより語彙処理能力を測定しようとするものである。単語の提示方法については、はじめに注視点（++++）が3秒提示された後、ターゲットとなる単語が提示され、「実単語」ならキーボード上の“B”を押し、「非単語」なら“N”を押し（図4参照）。練習問題が8問提示された後に1語ずつ計100語（実単語50、非単語50）が提示され、テスト終了後に各単語に関して正答数と判断に要した反応時間がエクセルで読み込める。なお、スペースキーを押すことで次の単語へと進むため、受験者が自分のペースで適宜休憩も取りながら解答できる点も CELP-Sem と同様である。

以下、テスト開発の手順とテストの信頼性向上にいかに取り組んだかについて述べる。

3.2.1 基本語彙リスト（実単語）の作成

本テストに使われた単語リストは、CELP-Sem と同様に英単語親密度リスト（横川他, 2006）にもとづく。本テストは語彙知識へのアクセス効率をもとにした潜在的な運用能力を測定することが主な目的である。このことから、顕在的な知識の有無を問うことを目的とした語彙サイズの測定とは本質的に異なっている。親密度リストでは日本人学習者が実際に馴染みのある単語と感じるかどうかを基準に単語が列挙されていることから、この親密度リストを使用することは、日本人学習者のメンタルレキシコンにおける語彙のネットワークの実態を反映したものであると考えられる。従って、日本人学習者の平均的なメンタルレキシコンの様相を反映した、横川他（2006）による親密度リストを基にターゲット語を選択することが適当であると言えよう。一般的に出現頻度が高い（学習者が遭遇する可能性が高い）とされている単語でも、必ずしも日本人学習者が英語学習経験の中で繰り返し遭遇し、馴染みがあるとは限らない。そうなると、学習済の単語の「処理速度」を

測定するには、その単語の馴染みの程度が重要な要因となることから、親密度にもとづいて単語の選定の指標とすることが適当であると考えた。

以上の考察に基づき、本語彙テスト作成の第一段階として CELP-Sem と同様に、親密度順位2999以下の語を300順位毎のセクションに分け、リスト内の親密度順位がすべて均等に反映されるように留意した。また、スペリングの長さについても長短を均等に含めた。さらに、品詞については名詞と動詞が中心となるように調整し、意味の重複やスペリングの類似性を避けるように選択した。その結果、最終的に200語が選択され、実単語の語彙リストが作成された。

3.2.2 非単語リストの作成

CELP-Sem 同様、CELP-Lex にも A、B の2つの版を設け、各版は実単語50語、非単語50語から成るようにした。実単語の選択と同様、適切な非単語を選択することもテストの信頼性を高めるうえで極めて重要な要因となる。単語にはスペリングのパターン、語長、隣接語など、単語認知に影響を与える諸々の要因があるが、実単語リストの形態的特徴と可能な限り等質となるような非単語リストを作成しなければ、実単語と非単語の間で語彙性判断における難易度が異なり、テストの信頼性が損ねられる可能性がある。この問題を回避する為に、CELP-Lex 作成においては、Washington University, St. Louis により作成され、WEB上で無料公開されている語彙データベース“The English lexicon project（以後 ELP）”「英語語彙プロジェクト」（<http://lexicon.wustl.edu/>）を利用することにした。

ELP のデータベースは、様々な語彙の特性に関するデータを提供するものである。例えば、長さ、隣接語数、出現頻度、語彙範疇などのデータが参照できる。また、実際に800名以上の英語母語話者に対して行われた行動実験で得られた語彙性判断課題および音読課題における反応時間データが参照できる。ELP データベースはメンタルレキシコンに関わる心理実験に利用する刺激語の作成のために多くの研究者に参照され、現在は非単語40,481語、実単語40,481語が収められている。利用者が必要とする語彙の特性を WEB 上で

選択し、実単語の特性に関するデータ（数値）を得ることで、同様の数値を有する非単語が検出できる。つまり、一定の語彙特性において、実験に利用する実単語に匹敵する非単語リストが得られることになる。本テストにおいては、当初の候補となった実単語 200 語に関して、語彙処理に関わる以下の語彙特性データを ELP より取得した。

- 1) 語長（文字数）
- 2) 隣接語数（1 文字のみが異なった近傍語の数）
- 3) 2 文字連鎖の平均頻度
- 4) 音読課題の平均反応時間 標準偏差／正答率
- 5) 語彙性判断課題の平均反応時間 標準偏差／正答率

その結果、上記の特性において実単語に匹敵する 985 語の非単語リストが取得された。

3.2.3 実単語、非単語各 100 語の選定

実単語 200 語を基に ELP より 985 語から成る非単語リストが得られた後、最終的に本テストに使用する単語を実単語、非単語ともに各々 100 語（A 版：実単語 50 語、非単語 50 語；B 版：実単語 50 語、非単語 50 語）に絞り込む作業が行われた。

この最終選定作業を行うためには複数の方法が考えられる。最も理想的な方法は、実単語リスト、非単語リストともに上記の 5 つの特性すべてにおいて等質となる、最適な 100 語を各々に選定する為のある種のアルゴリズムを構築することであろう。しかし、異なった複数の特性を統合する作業は極めて煩雑かつ複雑である。解決策として、最も語彙性判断課題に影響すると思われる特性に絞り、実単語、非単語間でそれらの特性に関して、等質になるように調整することにした。

語彙性判断課題の平均反応時間（RT）／標準偏差／正答率（AC）は、当然ながら重要なデータとして含まれる。加えて、隣接語数を基準とすることにした。つまり、ある単語に対して綴りが似ている近傍語数が多いほどその語は形態的に似通った語彙を多く持つことになり、そのことはその

単語の認知を困難にすると予想できる。この点で、隣接語数は認識速度を左右する重要な要因であると考え、実単語・非単語ともにこの要因で等質になるように留意した。最終的に実単語 50 と非単語 50 の間で隣接語数の平均が等質になればよいことから、実際の選択においては、実単語と非単語ともに隣接語数の値を基に昇順に並べ替え、上から 50 語毎のセクションに区切り、100 語になるまで各セクションから順番に一語ずつピックアップするという方法を採用した。語彙性判断については、ELP から得られたデータを z 値変換し、「反応時間（RT）+ 正答率 AC/2」の数値を算出した。その後、先の隣接語数と同様の方法で 100 語を選定した。

以上の手順を経て、CELP-Lex A 版（実単語 50 語、非単語 50 語）、CELP-Lex B 版（実単語 50 語、非単語 50 語）のリストを完成させることができた。

4. CELP-Lex データの検定結果

4.1 等質性の検定

最終的に選定された実単語と非単語各 100 語の間で、ELP から得られた隣接語数、語彙性判断の平均反応時間／正答率の数値について等質性が確保されているかどうかについて統計的に検定した。

t 検定の結果、すべての項目において実単語と非単語の間で有意差があった（ $p < .001$ ）。すなわち、非単語リストの方が反応時間が遅いという結果となった。実際に存在しない非単語は、実単語

表 1 実単語と非単語の隣接語数、語彙性判断反応時間に関する比較

		N	Mean	SD
隣接語数	実単語	100	2.63	4.52
	非単語	100	7.51	2.63
反応時間 (msec)	実単語	100	658.51	61.70
	非単語	100	817.94	71.14
反応時間 (SD)	実単語	100	231.02	70.78
	非単語	100	280.88	74.66
正答率	実単語	100	.97	.02
	非単語	100	.82	.12

表2 CELP-Lex A版とB版の隣接語数、語彙性判断反応時間に関する比較

		N	M	SD
隣接語	Test A	100	4.93	4.28
	Test B	100	5.21	4.58
反応時間 (msec)	Test A	100	737.75	96.77
	Test B	100	738.70	111.08
反応時間 (SD)	Test A	100	259.32	71.65
	Test B	100	252.59	81.74
正答率	Test A	100	.89	.11
	Test B	100	.89	.12

と比較してNOと判定するのに要する時間が長くなり、正答率も低くなると予想できる。また隣接語数にも有意差があった。作成される段階で単語の綴りを基にしていることから、非単語の方がその隣接語数が多くなる傾向があると思われる。また、隣接語数の多さが認識速度の干渉要因となって反応時間を遅延させる可能性もある。これらの要因は相互に語彙処理に影響を与え合っていると考えられる。

4.2 A版とB版の比較

次に、テストのA版・B版間で、上記の項目において有意な差がないかどうか検定した。t検定の結果、いずれの項目においても両者間に統計的有意差は全く検出されなかった。従って、A版とB版はこれらの項目において等質な2バージョンであることが示された。

5. おわりに

潜在的な運用能力（流暢性）という観点を語彙運用力の対象とするという発想においては、CELP-SemとCELP-Lexは非常に似通っている。しかし、CELP-Semとは異なり、CELP-Lexは語彙性判断課題を採用している。CELP-Semに加えてCELP-Lexを開発したことで、語彙処理の自動性や流暢性を語彙性判断を用いて測定することが可能となり、「使える英語力」の評価の実現に向けてさらなる進展をしたと考える。また、前述のようにCELP-Lexは、CELP-Semが抱える問題点である類義性判断における曖昧さを解消したとい

う点でも評価できる。

一方、語彙性判断が、意味へのアクセスを含まない処理プロセスであるという可能性は繰り返し指摘しておくべきである。しかし、この点については、今後CELP-LexとCELP-Semとの相関、あるいはCELP-LexとVocabulary Levels Test (Nation, 2001)との相関を検証することで、語彙性判断と意味へのアクセスとの関連性を精査できると考える。

CELP-SemとCELP-Lexは、お互いの欠点を補完し合うことで更に語彙処理能力の実態に迫ることが可能なテストツールとなり得ると期待できる。そのためにも今後、従来から実施されてきたTOEICなど様々な標準的な習熟度テストのスコアとの関係を検証することで、CELP-LexとCELP-Semの妥当性と信頼性についてさらに探求していく必要があろう。

本論文は、2012年度関西学院大学共同研究一般研究A（研究課題：語彙性判断課題によるコンピュータ版英単語運用能力テスト（CELP-Lex Test）の開発）〈研究代表：門田修平〉の研究助成を受けた研究成果の一部である。

引用文献

- 門田修平 他 (2010) 「第二言語における語彙処理と文処理のインターフェイス：日本人学習者への実証研究」平成19年度～平成21年度科学研究費補助金〈基盤研究(C)〉研究成果報告
- 門田修平(2010)「Computer-Based English Lexical Processing Test (CELP Test) の妥当性の再検討」外国語外国文化研究 XV (関西学院大学法学部外国語研究室), 91-106.
- 野呂忠司 (2010) 「第二言語読解における単語認知の働き：効率的な単語処理能力は流暢なリーディングを可能にするか」門田修平, 他 (2010) 「第二言語における語彙処理と文処理のインターフェイス：日本人学習者への実証研究」平成19年度～平成21年度科学研究費補助金〈基盤研究(C)〉研究成果報告, 104-112.
- 横川博一 他 (2006) 「日本人英語学習者の英単語親密度 文字編－教育・研究のための第二言語データベース」東京：くろしお出版
- Daller, H., Milton, J. & Daller, T. J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

- Harrington, M. (2006). The Lexical Decision Task as a Measure of L2 Lexical Proficiency. *EUROSLA Yearbook* 6, 147–268.
- Harrington, M. & Carey, M. (2009). The On-line Yes/No Test as a Placement Tool. *System*, 37, 614–626.
- Miki, K. (2010). *An Access to English Homographic Words of Japanese EFL Learners*. An MA Thesis Presented to the Graduate School of Language, Communication, and Culture, Kwansai Gakuin University.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Shimamoto, T. & Ikemura, D. (2011). *Accuracy and Fluency in Lexical Processing: How Do They Relate to English Proficiency?* Presented at JACET 50th Symposium, Seinan Gakuin University.

付表：CELP-Lex A/B で使用された実単語、非単語リスト

A

W	abandon	NW	alab
W	accurate	NW	beon
W	anticipate	NW	berk
W	assault	NW	bicked
W	assess	NW	bolled
W	attend	NW	bovers
W	bind	NW	burler
W	childhood	NW	burses
W	colleague	NW	caunt
W	consequence	NW	chasts
W	contact	NW	corch
W	crucial	NW	cour
W	detect	NW	crail
W	different	NW	cran
W	dose	NW	dat
W	end	NW	feans
W	expect	NW	freg
W	expose	NW	geals
W	factor	NW	glawed
W	famous	NW	gover
W	generation	NW	grag
W	hero	NW	gree
W	hypothesis	NW	happed
W	improvement	NW	hotted
W	initial	NW	hutter
W	isolated	NW	jile
W	known	NW	keeds
W	lad	NW	lann
W	lake	NW	loast
W	natural	NW	lounded
W	option	NW	marns
W	overall	NW	matty
W	paper	NW	nables
W	pretty	NW	nars
W	pursue	NW	ralt
W	radical	NW	routh
W	request	NW	seach
W	rural	NW	slan
W	scheme	NW	spen
W	seem	NW	spunt
W	sing	NW	stort
W	snow	NW	sumped
W	suffer	NW	talf
W	trail	NW	tarns
W	treaty	NW	tigs
W	unusual	NW	vaked
W	voice	NW	werd
W	vote	NW	wonk
W	wear	NW	zears
W	widespread	NW	ziles

B

W	absolute	NW	bandy
W	audience	NW	bice
W	believe	NW	blawing
W	birth	NW	bosts
W	bottom	NW	brying
W	chapter	NW	bule
W	cigarette	NW	caker
W	client	NW	cipped
W	compare	NW	coan
W	conclude	NW	coom
W	consent	NW	crasp
W	constraint	NW	dall
W	conversion	NW	daped
W	customer	NW	dases
W	democratic	NW	datched
W	employ	NW	dounded
W	encourage	NW	dure
W	enormous	NW	farned
W	established	NW	gats
W	external	NW	gides
W	faith	NW	glaying
W	game	NW	gues
W	great	NW	hable
W	ground	NW	hunner
W	hear	NW	jopped
W	human	NW	marned
W	intellectual	NW	mure
W	just	NW	nallow
W	landlord	NW	nells
W	lesson	NW	noiled
W	mill	NW	nort
W	observation	NW	palked
W	opposite	NW	pandy
W	original	NW	pite
W	payment	NW	plazed
W	pupil	NW	plick
W	reckon	NW	rost
W	retail	NW	rubes
W	score	NW	rulled
W	shout	NW	seans
W	sole	NW	shile
W	special	NW	sice
W	stable	NW	spink
W	substance	NW	tark
W	tell	NW	tropping
W	tension	NW	vaving
W	ticket	NW	volder
W	tough	NW	voles
W	university	NW	vone
W	writer	NW	yans

W：実単語

NW：非単語