

論 題 英文リーダビリティ測定システムの開発

福井正康^{*1} 小篠敏明^{*2}

<概要>英文リーダビリティの測定にはこれまで単語数と音節数によるものが多く考えられて来たが、我々はこれらに単語と熟語の難易度を加えた新たな指標を開発した。その開発には単語と熟語の難易度を与える辞書とこの辞書を利用して文を解析するプログラムが不可欠である。この報告ではこのプログラムの主な機能とこれを使った新しい指標の開発方法を説明する。

<キーワード>英文, リーダビリティ, プログラム

1. はじめに

英文リーダビリティの測定については、これまで様々な指標が与えられて来たが、ほとんどは単語数と音節数によるものであり、文の構造や単語の難しさなどを考慮したものはほとんど見られなかった。最近、熟語の難易度を取り入れた指標の開発も進められているが¹⁾、まだ具体的なシステムとしては完成していない。

この状況をふまえ、我々は英語学習者のための英文リーダビリティ指標の開発を始めた。手始めとして単語と熟語の難易度を加えた新たな指標作りを目指し、単語辞書と熟語辞書の作成、及び英文の難易度を測定するプログラムの製作を開始した。単語辞書には出発点としてJACET 8000を使い、単語の難易度はその中にある単語のランクを利用することにした。また単語の変化形については、JACET 8000の各単語にその変化形を加え、原形を見出し語とした。単語の音節数については変化形ごとに異なるため、プログラムにより計算することにした。熟語辞書については、雛形とするものが見つけにくいいため、旺文社の英語基本熟語集大学入試1000熟語を利用することにした。

この研究には日本学術振興会科学研究費補助金 基盤研究C (19520535)の援助を受けており、謝意を表したい。

2. プログラムの概観

プログラムの実行画面を図1に示す。画面の右上がメインメニュー、右下が辞書ユーティリティ、左上がサンプルデータ、左下及び

中央下が出力データである。

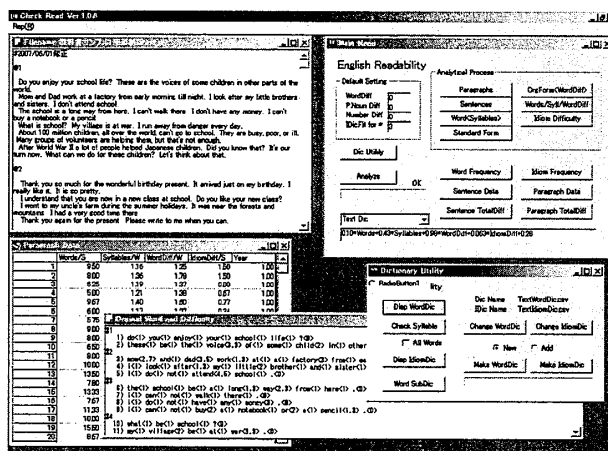


図1 プログラム実行画面

出力は表形式とテキスト形式のどちらかであり、サンプルデータには@1(中学1年)や@2(中学2年)のように学年のデータが入っている。

メインメニューの「Analyze」ボタンをクリックすると解析が始まる。最初は文章を段落に分け、段落を文に分ける。その後1文の単語数と各単語の音節数を計算し、文単位で単語を原型に変換する準備を始める。これには「It's」や「I'd」などの省略形を元に戻し、「,」や「.」などを単語と分離する処理などが含まれる。これらの後、単語辞書を検索して単語を原型に変換し、難易度の割付けを行う。次に原形への変換結果を利用して熟語辞書の検索を行う。処理時間は実用上重要であるが、中学1年から高校2年までの一連の教科書で10秒程度である。

次はこのプログラムの主な3つの機能であるリーダビリティの測定、新しい指標作成

^{*1} FUKUI, Masayasu : 福山平成大学

e-mail= fukui@heisei-u.ac.jp

^{*2} OZASA, Toshiaki : 福山平成大学

e-mail= tozasa@heisei-u.ac.jp

のためのデータ出力、新しい辞書の作成方法について説明する。

3. リーダビリティの測定

メインメニューの下部には難易度を評価する計算式(指標)が表示されている。指標は1文当りの単語数(Words)、1語当りの音節数(Syllables)、1語当りの単語難易度(WordDiff)、1文当りの熟語難易度(IdiomDiff)の1次式で与えられている。メインメニューの「Sentence TotalDiff」ボタンや「Paragraph TotalDiff」ボタンをクリックすると、「Analyze」で解析した結果を用い、この式を利用して、文または段落ごとのリーダビリティの値が中学1年を1、高校2年を5となるように表示される。

4. 新しい指標の作成

2節で述べた解析から新しい指標を作り出すことは難しくない。メインメニューの「Sentence Data」ボタンや「Paragraph Data」ボタンをクリックすると、英文に付けられた学年データ、解析中に調べた Words と Syllables、WordDiff と IdiomDiff を統計処理用のデータとして出力できる。

この機能を利用して、中学から高校まで通して出版されている教科書をデータとし、各文の学年を目的変数に、他の4つの変数を説明変数にして重回帰分析を行い、結果の重回帰式を新しい指標とすればよい。

5. 新しい辞書の作成

英文データ中に学年データを埋め込んでおけば、検索過程を逆に利用して単語の初出現学年が分かる。この考え方は熟語にも利用できる。この初出現学年を難易度にすれば教科書に基づいた新しい難易度の辞書を作成することができる。そのためにはまずメインメニューの「Dic Utility」ボタンをクリックして辞書ユーティリティを表示させ、「Make WordDic」ボタンや「Make IdiomDic」ボタンをクリックする。出力された結果はそのまま単語辞書や熟語辞書として利用できる。

6. 指標の評価と補正

我々は中学から高校までの教科書を出版

している3社の教科書を用いて、単語辞書と熟語辞書を作成した。その後、これらの辞書と教科書を利用して新しい指標を作成した。その結果は以下の通りである。

$$\text{Diff} = 0.10 \times \text{Words} + 0.43 \times \text{Syllables} \\ + 0.98 \times \text{WordDiff} + 0.063 \times \text{IdiomDiff} + 0.28$$

さて、この指標は教科書の学年を目的変数として作られているが、1つ教科書には様々なリーダビリティの文が含まれている。それゆえ読者が感じる難易度とこの指標による難易度には差が生じるものと思われる。我々はこれを補正するために標準的な英文を130程度集め、専門家4人に各文のリーダビリティを評価してもらい、その平均を我々の指標の結果と比較して補正式を導出した。補正式には非線形の要素を取り入れ、検討の結果、成長曲線のモデルとして知られている以下のゴンペルツ曲線による変換を採用した。

$$\text{NewDiff} = 4.387 \times \exp(-27.762 \times 0.326^{\text{Diff}}) + 1.031$$

この変換により専門家の評価とパソコンの評価との相関係数は0.908となった。

7. おわりに

我々の提案する指標の決定法は、教科書を入力することだけで求められた粗い評価を130程度の例文による詳しい評価で補正する方法である。

この方法は最良の方法ではないが、教科書を入力することは誰にでもでき、130程度の例文を評価することはあまり専門家を煩わせることなくできる方法である。またこれは日本のみならず、英語を学ぶ世界の国で、その国の教育の実情にあった指標を容易に作ることもできる方法でもある。

現在、統計処理は別のソフトウェアを用いているが、今後この部分をシステムに加えて、指標開発ツールとして単独で機能できるようにしてもよい。

参考文献

- 1) Nikolaos K. Anagnostou and George R. S. Weir, Average collocation frequency as an indicator of semantic complexity, ICTATLL Workshop 2007 Preprints, 1-3 August, 2007, 43-48.