

## 英語の読みやすさ指標となる Readability 新公式開発環境の構築

工藤 良一\* 宮崎 佳典\*

\*静岡産業大学国際情報学部

Readability とは元来欧米で開発された、英語の文章の読みやすさを出力する関数として定義される。本研究では、より信頼性のある Readability 公式開発の第一歩として、既存の Readability 公式を実装したプログラムを作成し、さらに新しい公式を導出するための開発環境を作成する。従来の公式には種々の問題点があると著者らは考えており、より信憑性がある、現実的な Readability 公式を考案することを最終目標に掲げている。

### 1. はじめに

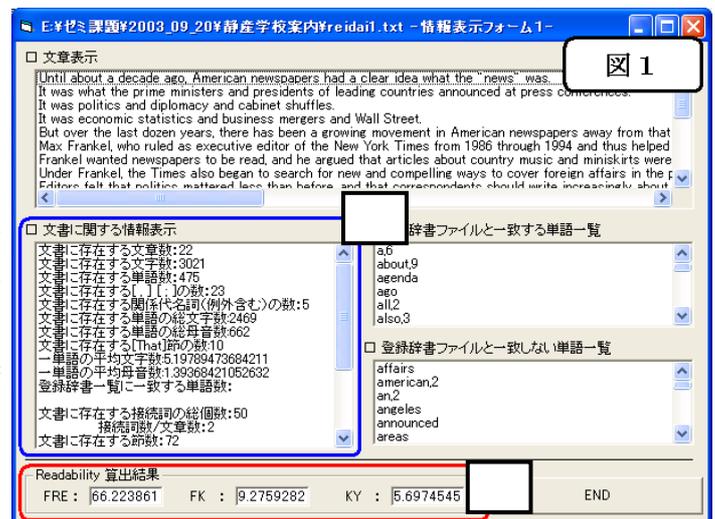
著者らは、日本人英語学習者を支援するプログラムの開発に従事している。英文テキストの難易度を測定する概念として、Readability と呼ばれる、入力を英文テキスト、出力を実数とする関数が考案されている。しかし、この Readability は英語圏で母語使用者の為に作られた歴史があり、日本人には十分に対応しているとは言えない。研究例は数例あるものの、まだ成熟していない感がある。そこで、日本人向けの Readability 開発を最終目標とし、英語教員など、専門知識を有する研究者が Readability を開発する際の支援環境構築を目指す。現在は、新公式完成の際に比較検討する為に使用する、既存 Readability 公式のうち代表的な 3 種を計算する作業を終え、また開発支援環境の大枠が仕上がりにつつある。この開発支援環境には、新公式のパラメータとなる可能性のある各種“(属性)情報”の表示、パラメータや Readability 値間の相関関係、重回帰方程式の作成などが含まれ、Readability 公式開発に必要な最低限の機能が包含されている。なお、本研究の開発プログラムは全て Microsoft 社の Visual Basic 6.0 Professional の環境下で作成された。

### 2. 既存 Readability

新公式開発の際に比較検討で使用するため、3 種の Readability を算出、表示することにした。使用した Readability 公式は、Flesch Reading Ease(FRE)、Flesch-Kincaid Grade Level(FK)、清川(KY)[1]の 3 種である：

- ・FRE :  $R=206.8 - 1.05 \times X - 84.6 \times Y$
- ・FK :  $R=0.39 \times X + 11.8 \times Y - 15.59$
- ・KY :  $R=0.288 \times N + 0.17 \times X + 1.739$

R は Readability 測定値であり、X は 1 文ごとの平均単語数、Y は 1 単語の平均音節数を意味し、N は難語率とし、公式 KY の発案者である清川氏が常用単語として選定した英単語リスト 1726 語に該当しない単語の出現率と定義した。ただ、Y (1 単語の平均音節数) は単語の音節情報がない限り自動化は不可能である。次節でこの問題の対処法について述べている。実行画面は図 1 部となる。実際に公式を見てみると利用しているパラメータは適切なのか？パラメータの種類が少ないのではないかなど様々な点で疑問が生じてくる。図 1 は上記 X,Y,N だけでなく、英文テキストに関する各種情報を羅列したものである。



### 3. 単語属性情報付加プログラム作成

過去そして現在の研究グループ(ゼミ)内で議論した結果、Readability のパラメータ候補として次が提案された：文章数 文字数 単語数 単語の総文字数 単語の総母音数 単語の音節数 ”;”, “,” の個数 節の数&接続詞の数 代名詞

の数 関係代名詞の数 登場人物の数 など。この中には、単語属性情報がなければ計算できないものが多い。ただ、清川氏のリストや北海道リストなど、他によく知られている単語リスト（例えば [3] を参照）には単語以外の情報がほぼ皆無であることが大半である。そこで今回新たに単語リストを作成するにあたり、Web 上辞書サイト Goo 辞書を利用し、同サイトから必要な情報を引き出すプログラムを作成した。手始めに北海道リスト 7454 語に対し、単語の品詞情報と音節数の 2 種類の属性情報を自動的に付加させた。実行画面は図 2 を参照のこと。プログラムには WebBrowser コンポーネントを利用し、サイト内情報を操作出来るようにした。アルゴリズムを簡略的に記述すると

1. 単語リスト内の単語を一つずつ Goo 辞書サイト内の単語入力用テキストボックスに挿入
2. Goo 辞書サイト内の検索ボタンを押下
3. 検索結果画面の音節数ならび品詞情報の抜き取り & 同単語と併せて外部ファイル出力

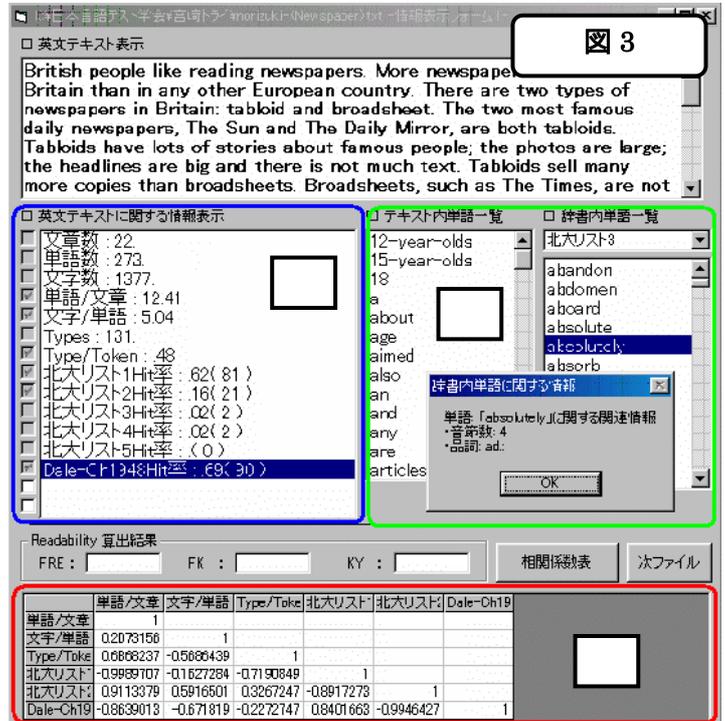
のようになる。このプログラムの完成によって、今回は 2 種類の情報を吸い上げることに成功したが、辞書サイト内のあらゆる情報を付加できる可能性がでてきた。これは研究全体に幅を持たせる。



#### 4. 新公式開発支援プログラム

新公式開発はパラメータの決定、必要に応じて検定、重回帰方程式の作成（各パラメータ係数決定）といったプロセスを経る。本プログラムはこの流れを重視し、仮公式作成を支援する役割をもち、また、そこに至るまでに専門家が必要とする（と思われる）英文テキストに関する様々な情報を表示するものである。図 3 部には既存 Readability でも使用されているパラメータを始め、各種情報を候補に加えている。部は難語率計算の元となる単語リストに関する情報である。

ここに 3 節で取得した各単語の属性情報が表示され、また統計情報計算にも役立っている。部では部の各種パラメータのうち、ユーザーが任意に選択したもののみを抽出し、相関係数を計算したものである。これをもとにユーザーは公式開発をよりスムーズに進めることが可能となる。



#### 5. 結語

日本人英語学習者を支援するプログラムの作成を目指して Readability 公式開発研究が始まりつつある。まだ問題点は数多く指摘され、単語リスト用の単語選定、重回帰方程式の式妥当性など課題は多い。ただ、新しく始まりつつある研究領域の一助となれば望外の喜びである。

**謝辞** 各方面でご支援・ご指導を頂いた静岡産業大学国際情報学部 法月健 助教授に感謝申し上げます。なお、本研究は平成 15 年度静岡産業大学研究活動助成金の補助を受けている(法月助教授と共同受給)。

**参考文献** [1] 清川英男, 高校・大学生用リーダビリティ公式の開発, 和洋女子大学英文学会誌 (21), 43-63 (1988). [2] 宮崎佳典, 工藤良一, Readability 新公式提案に向けてのプログラム実装, 日本言語テスト学会第 18 回研究例会(共催第 16 回英語教育をおもしろくする会)(2003). [3] <http://www.hiroshima-pu.ac.jp/~umamoto/voca.htm>