

語彙多様性を表す指標の妥当性に関する研究

—日本人大学生の書き言葉コーパスの場合—

田島ますみ* 深田 淳** 佐藤 尚子***

1. はじめに
 - 1.1 語彙の多様性を表す指標
 - 1.2 第一言語の場合
 - 1.3 研究の目的
 2. 方法
 - 2.1 資料
 - 2.2 主観評価
 - 2.3 語彙指標
 - 2.4 語数の算出
 3. 結果
 - 3.1 主観評価
 - 3.2 各群の平均
 - 3.3 主観評価と指標数値の相関
 4. 考察とまとめ
 5. 本研究の制約と今後の課題
- 参考文献
- 資料1：総合的評価基準
- 資料2：全データ

1. はじめに

近年のテクノロジーの進歩に伴い、日本語のコーパス構築、またそれらを使った研究が進んでいる（大曾（2006））。実証研究の資料としては宝庫というべきもので、日本語教

育への応用も含め研究の進展が期待されている。

電子化されたコーパスの利点は、コンピュータを活用して大量のテキストの中からほしい情報を短時間で取り出せることである。本研究では、大量の日本語の文章においてコンピュータ処理により延べ語数と異なり語数を簡便に算出することができるというコーパスの利点を活用し、これらの数値を使って語彙の多様性を表す二つの指標の妥当性を検証する。

1.1 語彙の多様性を表す指標

産出された発話や文章の質を分析する際、語彙の多様性という項目がある。これはどれくらい多様な語が使われているのかをみるもので、数値化する場合、延べ語数における異なり語数の割合という指標を用いるのが一般的である。英語で言えば type/token ratio である。全体量に対して異なる語の数がどれくらいあるのかを示す指標であり、この値が大きいほど多様な語が使われていて語彙が豊富であるということになる。

よく使われている指標ではあるが、弱点としては全体量が多くなる、つまり発話や文章が長くなると分母の延べ語数が多くなり、値

* 本学法学部専任講師

** インディアナ州立パデュー大学外国語外国文学学部准教授

*** 千葉大学国際教育センター准教授

が小さくなってしまふ点が挙げられる。さらに、Wolfe-Quintero, Inagaki, & Kim (1998) は、この指標では短い作文で異なり語数が少ない場合と長い作文で異なり語数が多い場合で同じ値が出てしまい、両者を区別することができないという問題点を指摘した。

これらの問題点への対策として、分母の延べ語数に若干手を加えた指標を Carroll (1967) が提案している。分母を延べ語数の2倍の平方根とするというものである。こちらのほうが産出量を考慮しているために、より適切な指標であることが実証されてきている。前掲の Wolfe-Quintero, et al (1998) は、第二言語のライティングに関する39の量的研究に使用された108の指標を検討した著作である。それらの先行研究の結果を検討する限りにおいて、Carrollの提案した、分母を延べ語数の2倍の平方根とした指標が、中間言語の発達を表すには最適との結論を出している。また田島(2002)は日本語学習者の話し言葉コーパスであるKYコーパスを資料として語彙指標の妥当性を統計的に検証したものである。この研究でも、単純な延べ語数分の異なり語数ではなく、延べ語数を操作した指標がOPIの級の判定と対応する、妥当性のあるものと確認された。

1.2 第一言語の場合

第二言語の研究分野では上記のような状況にあるが、第一言語の場合、語彙の多様性を表す指標の妥当性に関して一石を投ずるような調査結果がある。Sasaki & Hirose (1999) は、第一言語としての日本語の作文評価基準を作成することを目的に、高校の国語教員に対し、作文を評価する際にどんな項目をどの程度重視するかについてアンケートを行い、102人から回答を得た。その結果、全35項目中、11項目が重視されない項目として基準からはずされたが、その中に「語彙は豊富

か」という項目が含まれていた。

この研究結果に従って考えれば、第一言語としての日本語の作文評価に語彙の豊富さはさほど重要ではなく、評点には影響しないという可能性が出てくる。つまり評価の高い作文であっても語彙が豊富とは限らないし、評価の低い作文であっても語彙は豊富かもしれないということである。もしそうならば、語彙の多様性を表す指標は第一言語では有効ではなく、分析や研究に使用すべきではないという結論が導かれることになる。

確かに、第二言語では習得が進むにつれ産出できる語彙が増え、発話や文章の使用語彙を見るだけでも習得段階についてある程度判断をすることは可能である。この程度の語彙を使っているならば初級、これほど多くの語が使えるのは上級などというような判断である。しかし、成人の第一言語の場合、語彙の多様性に関してはそれほど顕著な個人差は出てこないことも考えられる。個人差が全くないということは考えにくいだが、数値化したときに指標として機能するほどの違いがないことは十分にありえる。さらに数値に出るほどの違いがあっても、それが発話や文章の総合的な質とは無関係の可能性もある。第一言語と第二言語では個別に検証する必要があるだろう。

1.3 研究の目的

このような状況を踏まえ、本研究では、第二言語の研究分野で使われている語彙の多様性を表す指標が、第一言語の場合でも有効な指標となるのかを検証する。実際の手続きとしては第一言語で書かれた文章の主観評価と語彙指標数値の相関関係を分析する。

予想される結果は二つある。一つは、第一言語の場合も第二言語と同様、指標として有効であるというものである。もう一つは、第一言語の場合、語彙の多様性は発話や文章の

質とは無関係であり、指標としては使えないというものである。

2. 方法

2.1 資料

日本人大学生が授業の課題に対して提出した日本語の文章100人分を資料とした。以下、単位を「作文」とし、「100作文」というように示す。「作文」という呼び方は、国語・日本語教育において、比較的簡単なレベルでの出来事の記述や私的感想文などを指す狭義の意味で使われることが多く、高等教育の文脈ではあまり使用されないが、今回は便宜上「作文」を単位として使用する。

文章産出者の学生の所属は首都圏にある私立A大学商学部・法学部で、学年は1年生から4年生すべてにわたる。資料収集を行った授業は、日本語力の向上を目的として設置されている「日本語操作法」という自由選択科目である。2007年度の前期に、担当教員に何かを紹介する文章を400字以上で書くよう指示した課題に対して提出された文章を資料として使用した。

すべて電子ファイル化して書式を統一し、主観評価にはプリントアウトしたもの、語数のカウントにはテキストファイルを用いた。主観評価に印刷したものを使用したので、手書きの場合、印象や評価に影響を与える表記上の観点、たとえば字が丁寧に書いているか、漢字が正確に書いているかなどといった点は、今回の研究では評価の対象外となる。

2.2 主観評価

大学で日本人学生を対象に第一言語としての日本語教育に携わる教員3人が、あらかじめ作成した採点基準(資料1)をもとに総合的絶対評価で5段階の点数をつけた。評価者の第一言語は全員日本語で、第一言語とし

ての日本語教育歴は評価を始めた前月である2008年3月末時点で、評価者Aが4年、Bが21年、Cが2年であった。第二言語としての日本語教育も含めた日本語教育歴では、Aは11年、Bは21年、Cは7年である。3人の第一言語としての日本語教育歴の平均は9年、第二言語としての日本語教育も含めた教育歴の平均は13年である。

今回、分析的評価よりも信頼性が低いと言われる総合的評価をあえて用いたのは、比較的多くの文章の採点をしなければならないため利便性を重視したことも一因であるが、文章全体に対する評価と語彙の指標数値がどのように関連するのかを調べたいという研究目的による。項目ごとの点数の合算である得点よりも文章全体につけられる単一の点数が本研究には適切であると判断した。信頼性の確保のため、わかりやすい評価基準を作ることを目指した。また評価例を示し、評価のトレーニング・セッションを設けるという手順も設けた。

日本語の文章の評価基準は、一般に広く普及しているというものは存在せず、たとえば英語におけるTOEFLのライティング・セッションのためのスコアリング・ガイドに相当するようなものはない。日本留学試験の記述問題の採点基準も現在改訂作業中である。ライティング関係の先行研究を見ても、その都度基準が作成されているという状況である。

本研究では、TOEFLのライティング・スコアリング・ガイドを参考に今回の課題に対応する基準としてあらたに作成したものを使用した。TOEFLでは0から6の7段階評価であるが、第一言語であれば第二言語ほど質の低いものは出ないという点と利便性、信頼性を考慮して1から5の5段階評価とした。

評価を始める前に、研究で使用する資料とは別の文章で作成した採点例を評価者に示し

た。評点5に相当するのはこのような作文、評点4はこの程度のもの、というように各評点の一つずつの例を提示した。さらに、その例においてどのような点をどのように判断するのかを具体的に述べた「評価のポイント」も書面にして合わせて示した。

評価者には採点基準と評価例、「評価のポイント」を読んで十分検討する時間を与えた後、7作文を使って評価トレーニングを行った。まず、3作文をそれぞれの評価者が採点し、その後お互いの点数を確認して話し合いながら一致した点や違う点などを明らかにした。さらに別の3作文を用いて、同じ手順を繰り返し、最後に1作文を採点して確認した。トレーニング・セッションは評価者全員が集まって行うのが最善であるが、今回は評価者の都合のため、評価者Aと評価者B、評価者Aと評価者Cで、2組2回に分けて行った。

2.3 語彙指標

本研究で取り上げる語彙指標は多様性を表す以下の二つである。

指標1：異なり語数／延べ語数

指標2：異なり語数／ $\sqrt{2 \times \text{延べ語数}}$

2.4 語数の算出

延べ語数、異なり語数は、日本語形態素解析システム「茶釜」(奈良先端科学技術大学院大学松本研究室開発。http://chasen.naist.jp/hiki/ChaSen/)を使用した形態素分析をもとに計算機処理によって算出した。

3. 結果

3.1 主観評価

3名の評価者がそれぞれつけた評点の平均を、その文章の主観評価の得点とした。評価者間信頼性は、Spearman-Brownの公式を使う指数で0.65であった。信頼性は高いとは言えないが、この点に関しては本研究の制約

として後述する。

100作文すべての平均は3.11点で、標準偏差が0.96点であった。この結果から、上位群を区切る得点を平均に0.96点を足した4.07点以上、下位群を区切る得点を平均から0.96点を引いた2.15点以下とし、その間の点数のものを中位群とした。この分類の結果、各群の作文数は、上位群12作文、中位群71作文、下位群17作文となった。

3.2 各群の平均

まず各群の異なり語数、延べ語数の平均を示す。異なり語数、延べ語数ともに上位群が多く、次に中位、下位の順番であった。

表1 各群の語数平均値

群	n	異なり語数	延べ語数
上位	12	190.33	430.67
中位	71	143.78	304.96
下位	17	123.53	251.12

主観評価の際、評価者全員で、文章の長さは評価の対象としない、すべての文章が課題であった400字以上という条件を満たしているものとして、あくまでも文章の質を評価するということを確認したが、結果は長く書けているもの、語数が多いものの評価が高いということを示した。

次に語彙指標数値の平均を示す。指標の1と2は全く逆の方向を示した。

表2 各群の指標数値平均値

群	指標1	指標2
上位	0.48	6.46
中位	0.49	5.84
下位	0.50	5.52

指標 1 では上位から下位に行くに従い、わずかではあるが数値が大きくなる。差はほとんどないと言えるが、数値の示す方向としては上位よりも下位に行くに従って語彙多様性が増すことを示すほうへ向いている。一方、指標 2 では上位から下位に行くにしたがい数値が小さくなるという結果が出た。指標 1 とは逆の方向を示す指標 2 の結果のほうが、下位から上位に行くにつれて語彙の多様性が増すという傾向を示すものとなった。

3.3 主観評価と指標数値の相関

以上のように求められた主観評価の得点と各数値の関係を、Spearman の順位相関係数を使って表す。

表 3 主観評価と各数値の相関行列

	得点	type	token	指標 1	指標 2
得点	1.00				
type	0.51**	1.00			
token	0.47**	0.91**	1.00		
指標 1	-0.16	-0.33**	-0.66**	1.00	
指標 2	0.46**	0.87**	0.62**	0.11	1.00

**：1%水準で有意な相関であることを示す。

主観評価の得点と、異なり語数、延べ語数、指標 2 の間に、それぞれ中程度の相関がみられる。得点と指標 1 の相関係数は -0.16 で、ほとんど相関がなく、また方向としては負の相関となっている。得点と指標 2 との係数は 0.46 であり、上記のように中程度の相関関係にあると言える。

4. 考察とまとめ

今回の結果は、指標 2 が第一言語の場合であっても文章の質を表す指標としてある程度妥当なものであることを示している。語彙の多様性を表す指標 2 が文章の主観評価と相関

があり、指標として有効であって、研究や評価に使えるのではないかというのが本研究の結論である。

第一言語の場合の妥当性への懸念は、1.2 の項で述べたことであるが、語彙の多様性の個人差は第二言語の場合ほど顕著ではなく数値として現れないのではないかと、さらに数値に出るような差があったとしても、それが文章の総合的な質と相関関係にない恐れもあるのではないかと 2 点であった。今回の結果を見る限り、指標 2 の表す数値は主観評価による上中下の差に対応していて、語彙の多様性は第一言語の場合であっても文章の質の高さと関連することが統計的に示されている。

その一方で、指標 1 の数値は文章の評価と逆行する値を出してしまう。異なり語数の平均値だけ見れば下位から上位にかけて増加していくのであるから、全体量を考慮しないという条件のもとで上位のほうがより多くの語を使っていることは示されている。それでも指標 1 の数値は逆行して上位のほうへ行くにつれ多様性が減少するかのような値を示す。これは、下位から上位へかけての延べ語数の増加が大幅であり、その大きな延べ語数をそのまま分母として異なり語数を割ってしまうと指標 1 の値が小さくなってしまふからである。

文章の次元で考えてみるとよりわかりやすい。一般的な状況において、延べ語数が倍になっても異なり語数は倍にはならない。これは助詞や助動詞などの機能語が繰り返し出てくるためと考えられる。文章が長くなれば、異なり語数で 1 としかカウントされない助詞、たとえば「は」は何回も繰り返し延べ語数を増やす。故に異なり語数を単純に延べ語数で割る指標 1 の数値は文章が長くなると小さくなっていく。他方、指標 2 の分母は平方根をとっているため増加率が抑えられ、数

値が大きな分母で割られることなく主観評価と対応する関係を示すことになる。

このように、単純な type/token ratio ではなく分母を操作して全体量を考慮した指標 2 のほうが発話や文章の質と対応するという傾向は、今までの第二言語の場合の研究結果と同じである。広く普及している type/token ratio だが、何も考えずに指標として使ってしまうと誤った結果を導くことにもなりかねない。教育効果を測るような計量的研究において、type/token ratio の値が事前事後で小さくなったことをとらえて語彙の多様性には逆効果であった、などのような解釈がなされてはならないのである。

本研究の結果によって語彙の多様性を表す指標 2 の妥当性に一定の証左が与えられた。その妥当性は第一言語、第二言語の違いに左右されない。産出物としての日本語の測定を行う際には、指標 2 を使うべきであることを強く主張したい。

5. 本研究の制約と今後の課題

最後に本研究の制約を確認し、今後の課題を明らかにしておきたい。

まず挙げられる点は、主観評価の評価者間信頼性が低いことである。一般の相関からいえば、3 人の評価者の一致度を表す係数が 0.65 というのは決して高い数値ではない。しかし、主観という観点からいえば、それほど高い数値が出る必要があるかどうか、検討課題の 1 つとしたい。

総合的評価が一般的でない中で、1 作文に単一スコアをつけてみるとこれぐらいの一致度であるということは本研究で得られた一つの結果であり成果である。評価者 3 人は全員第一言語としての日本語の教育経験があり、評価者間信頼性を高めるべく方法の項で説明

したような手順も設けた。それでも単一スコアでは、重視する観点、その度合いによって点が割れてしまう弱みは否めない。それぞれ自分の中で採点基準に従い判断していった結果ではあるので、平均点をとった主観評価の得点はある程度信頼性のあるものと考ええる。

現在、同じ 3 人の評価者によって主観評価の再検討の作業を続けている。この作業でもっと評価者間信頼性の高い主観評価のデータを出し、より信頼できる研究結果を報告する予定である。

次に、文章の種類についての制約を述べておく。今回資料とした文章は、何かを紹介するという比較的書きやすい簡単な課題である。授業中に書かせたものではなく、1 週間の猶予がある提出課題であった。違う種類の文章では違う結果が出ることも考えられる。レポートや論述試験の答案といった、もっと論述力を問われる文章、あるいはタイム・プレッシャーがある中で産出される文章では、また別の結果が出て、指標の妥当性に疑義の出ることも十分考えられる。妥当性の検証は、文章の種類にも留意しながら繰り返し行われていくことが必要である。

さらに、本研究の結果の中で研究目的からは外れるものの、特に興味深いと思われた点を指摘しておく。結果の項で主観評価の得点と数値の相関を表 3 に示した。今まで得点と指標 1、指標 2 の相関にばかり着目していたが、表 3 を見れば指標よりも異なり語数と延べ語数のほうが得点との相関が強いことがわかる。指標 1 の相関係数が -0.16 、指標 2 が 0.46 に対し、異なり語数は 0.51 、延べ語数は 0.47 である。指標 2 の数値より、差は若干ではあるが、単純な語数のほうが係数は高い。わざわざ割り算や平方根をとるようなことをしなくても、語数、特に異なり語数を見れば指標となるという結果なのである。

しかしながら、常識的な感覚からは、単純な延べ語数、異なり語数のほうが文章の質とより相関のあるよい指標になるということにはかなり抵抗を感じる。延べ語数が多ければよい文章になるというのであれば、単に言葉が多く書かれているもの、長く書けているものがいい文章ということになってしまう。異なり語数が多ければいい文章というのであれば、文章全体の長さに関係なく多様な言葉が使用されているものがいい文章となる。たとえば原稿用紙1枚で50語の文章と原稿用紙10枚で50語の文章が同じで、原稿用紙1枚で40語のものはそれより劣るという判断につながってしまう。

常識からは疑問を感じるのだが、相関係数の結果は数値として出ている。もし異なり語数、あるいは延べ語数だけである程度妥当性のある判断ができるのであれば、無論利便性、経済性には優れるこれらの数値を指標として認めるべきである。今回の結果では、指標2よりもこれらの語数のほうが主観評価との相関は強かったのであるから、指標2よりも有効な指標とする結論を出すことも可能なのである。この点についても検証を繰り返していく必要のあることを述べておく。

また、表3に示された指標1と指標2の相関係数も注意を払っておきたい項目である。二つの指標の相関係数は0.11であって、相関はほとんど認められない。同じ語彙多様性を表すはずの指標である二つの相関係数がなぜこうも低くなるのか、今回の結果だけの問題であるのか、それとも計算上の問題で常にこれぐらいの値になるのか、これらの問題も非常に興味深い今後の課題である。

以上、本研究では語彙の多様性を表す二つの指標を取り上げて妥当性を検証したが、研究の動機としては、言葉という産出物を測る際に、数値で表す指標は目安なのだからと

いった、あまりにも安易で無自覚な指標使用に注意を喚起したかったということがある。数字にこだわりすぎることは不必要であるが、測定的手段として信頼できる妥当な指標を選ぶことは研究の基本である。本研究で整理された今後の課題を含め、検証を繰り返しつつより有効な指標を探っていきたいと考えている。

参考文献

- 1) 大曾美恵子 (2006) 「日本語コーパスと日本語教育」『日本語教育』130, 3-10
- 2) 国立国語研究所編 (2006) 『世界の言語テスト』くろしお出版
- 3) 田島ますみ (2002) 「KY コーパスを用いた語彙的複雑性の測定に関する研究——語彙的多様性及び密度と言語運用能力との関連——」*The tenth Princeton Japanese Pedagogy Workshop proceedings*, 94-104
- 4) 田島ますみ・佐藤尚子・深田淳 (2008) 「よい文章の語彙は豊富なのか? ——日本人学生の文章における主観評価と語彙指標の関係——」日本リメディアル教育学会第4回全国大会予稿集, 77-78
- 5) 田中真理 (2005) 「日本語教育におけるライティング評価」国立国語研究所編『日本語教育年鑑 2005 年版』42-52
- 6) 田中真理・長阪朱美 (2004) 「日本語と英語を目標言語とするライティング評価基準の展望: 第二言語としての日本語のライティング評価基準作成に向けて」『第二言語としての日本語の習得研究』7, 214-253
- 7) 日本語教育学会編 (1991) 『日本語テストハンドブック』大修館書店
- 8) Carroll, J. B. (1967). On sampling from a lognormal model of word-frequency distribution. In H. Kucera & W. N. Francis (Eds.), *Computational analysis of present-day American English* (pp.406-424). Providence, RI: Brown University.
- 9) Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*, New York: Heinle & Heinle

- Publishers.
- 10) Sasaki, M. & Hirose, K. (1999). Development of analytic rating scale for Japanese L1 writing. *Language Testing* 16, 457-478.
- 11) Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawaii, Second Language Teaching & Curriculum Center.
- 12) 日本留学試験 <http://www.jasso.go.jp/eju/index.html>
- 13) TOEFL <http://www.ets.org/toefl>

資料 1：総合的評価基準

評定 5

全般に良く書けていて、以下の評価ができるもの

- 課題を把握し、適切に対処して書かれている。
- 構成がしっかりしていて、段落がきちんとつけられている。
- 意味があいまいであったり、わかりにくかったりという個所がない。
- 構文や助詞など文法的な間違いがないか、あってもそれほど問題はない。
- 全体がこなれた表現で、ことばの使い方に統一感がある。

評定 4

良く書けていて 5 に近い評価ができるが、以下のような問題が若干あって最高のレベルとはできないもの

- 構成や段落に不適切な点がある。
- 意味があいまいであったり、不明だったりする個所がある。
- ことばの使い方に乱れや不自然なところがある。
- 構文や助詞など文法的な間違いがある。

評定 3

以下のような点の中で問題として感じられるものが一つ以上あるが、課題に対応する文章となっているもの

- 構成や段落に不適切な点がある。あるいはそれらが考えられていない。
- 意味があいまいであったり、不明だったりする個所がある。
- ことばの使い方に乱れや不自然なところがある。
- 構文や助詞など文法的な間違いがある。

評定 2

以下のような点の中で一つ以上が問題として強く感じられ、課題に対する文章として質が劣るもの

- 課題の把握が弱く、紹介する文章であることがわかりにくかったり、読み手を意識していなかったりする。

- 構成や段落に不適切な点がある。あるいはそれらが考えられていない。
- 意味があいまいであったり、不明だったりする個所がある。
- ことばの使い方に乱れや不自然なところがある。
- 構文や助詞など文法的な間違いがある。

評定 1

以下のような問題点が複数あり、課題に対する文章として質の低さが感じられるもの

- 課題の把握が弱く、紹介する文章であることがわかりにくかったり、読み手を意識していなかったりする。
- 構成や段落に不適切な点がある。あるいはそれらが考えられていない。
- 意味があいまいであったり、わかりにくかったりという個所がある。
- ことばの使い方に乱れや不自然なところがある。
- 構文や助詞など文法的な間違いがある。

資料 2：全データ（主観評価の得点降順）

n	ID	評価者 A	評価者 B	評価者 C	主観評 価得点	異なり 語数	延べ 語数	指標 1	指標 2
1	TH4001	5	5	5	5.0000	126	307	0.4104	5.0849
2	TS1002	5	5	5	5.0000	152	302	0.5033	6.1848
3	MH1001	5	5	4	4.6667	197	380	0.5184	7.1459
4	MH3003	5	5	4	4.6667	268	641	0.4181	7.4850
5	MH4001	5	5	4	4.6667	137	302	0.4536	5.5745
6	MH4003	5	5	4	4.6667	195	481	0.4054	6.2871
7	MS4003	4	5	5	4.6667	158	354	0.4463	5.9380
8	MH1003	5	4	4	4.3333	154	358	0.4302	5.7553
9	MH1009	4	5	4	4.3333	200	420	0.4762	6.9007
10	MS3018	4	5	4	4.3333	210	470	0.4468	6.8494
11	TS2012	4	5	4	4.3333	211	466	0.4528	6.9115
12	TS4001	3	5	5	4.3333	276	687	0.4017	7.4459
13	MH1002	4	3	5	4.0000	111	261	0.4253	4.8583
14	MH1010	4	4	4	4.0000	187	391	0.4783	6.6871
15	MH2002	4	4	4	4.0000	238	510	0.4667	7.4521
16	MH2003	5	3	4	4.0000	134	269	0.4981	5.7772
17	MH3004	4	4	4	4.0000	218	478	0.4561	7.0506
18	MH4002	4	3	5	4.0000	129	227	0.5683	6.0543
19	MH4006	3	4	5	4.0000	136	238	0.5714	6.2335
20	MS1010	4	3	5	4.0000	202	340	0.5941	7.7463
21	MS2005	3	5	4	4.0000	122	239	0.5105	5.5802
22	MS2015	4	4	4	4.0000	187	418	0.4474	6.4675
23	MS3004	5	3	4	4.0000	157	340	0.4618	6.0207
24	MS3019	4	3	5	4.0000	167	380	0.4395	6.0577

n	ID	評価者 A	評価者 B	評価者 C	主観評 価得点	異なり 語数	延べ 語数	指標 1	指標 2
25	MS4009	4	4	4	4.0000	125	245	0.5102	5.6469
26	TH3001	4	5	3	4.0000	147	277	0.5307	6.2454
27	TS2001	3	4	5	4.0000	128	302	0.4238	5.2082
28	TS3001	3	5	4	4.0000	268	719	0.3727	7.0673
29	TS4005	3	5	4	4.0000	201	489	0.4110	6.4273
30	MH1007	4	3	4	3.6667	94	163	0.5767	5.2062
31	MH3001	4	4	3	3.6667	171	359	0.4763	6.3817
32	MH3002	4	2	5	3.6667	234	383	0.6110	8.4548
33	MH3005	4	2	5	3.6667	174	358	0.4860	6.5027
34	MS2004	4	4	3	3.6667	123	267	0.4607	5.3227
35	MS3006	4	3	4	3.6667	157	308	0.5097	6.3257
36	MS4001	3	3	5	3.6667	177	433	0.4088	6.0147
37	MS4006	2	5	4	3.6667	293	801	0.3658	7.3204
38	MH4004	4	2	4	3.3333	143	272	0.5257	6.1311
39	MS1007	3	3	4	3.3333	120	201	0.5970	5.9851
40	MS1013	3	4	3	3.3333	111	207	0.5362	5.4554
41	MS2016	3	2	5	3.3333	141	310	0.4548	5.6627
42	MS2020	3	3	4	3.3333	211	556	0.3795	6.3275
43	MS3003	2	3	5	3.3333	114	206	0.5534	5.6164
44	MS3012	2	3	5	3.3333	180	316	0.5696	7.1600
45	MS3016	3	3	4	3.3333	104	197	0.5279	5.2394
46	MS4002	3	3	4	3.3333	115	242	0.4752	5.2273
47	TH3004	4	3	3	3.3333	127	205	0.6195	6.2721
48	TS2008	3	2	5	3.3333	131	286	0.4580	5.4774
49	TS2014	3	2	5	3.3333	102	203	0.5025	5.0622
50	MH1004	4	2	3	3.0000	95	196	0.4847	4.7982
51	MH1005	3	3	3	3.0000	134	218	0.6147	6.4174
52	MS1011	3	2	4	3.0000	127	246	0.5163	5.7256
53	MS2008	4	3	2	3.0000	118	201	0.5871	5.8853
54	MS2018	3	2	4	3.0000	113	232	0.4871	5.2459
55	MS3010	3	3	3	3.0000	136	272	0.5000	5.8310
56	MS3014	2	3	4	3.0000	131	260	0.5038	5.7447
57	MS4007	2	3	4	3.0000	116	240	0.4833	5.2947
58	TS2004	3	2	4	3.0000	115	222	0.5180	5.4577
59	TS2005	3	2	4	3.0000	111	244	0.4549	5.0247
60	TS2010	3	3	3	3.0000	171	380	0.4500	6.2028
61	TS4003	2	2	5	3.0000	186	530	0.3509	5.7129
62	MS1002	2	5	1	2.6667	136	318	0.4277	5.3928
63	MS1003	3	2	3	2.6667	100	185	0.5405	5.1988
64	MS2019	2	3	3	2.6667	166	428	0.3879	5.6738
65	MS2025	3	2	3	2.6667	105	178	0.5899	5.5650
66	MS3013	2	3	3	2.6667	116	225	0.5156	5.4683
67	MS3015	3	3	2	2.6667	93	181	0.5138	4.8880
68	MS3017	2	2	4	2.6667	118	333	0.3544	4.5724

n	ID	評価者 A	評価者 B	評価者 C	主観評 価得点	異なり 語数	延べ 語数	指標 1	指標 2
69	TH3003	1	3	4	2.6667	98	215	0.4558	4.7260
70	TS2003	3	2	3	2.6667	146	296	0.4932	6.0006
71	TS2017	2	2	4	2.6667	148	347	0.4265	5.6180
72	TS3003	3	2	3	2.6667	151	290	0.5207	6.2699
73	MS2002	1	3	3	2.3333	90	211	0.4265	4.3811
74	MS2007	2	3	2	2.3333	120	212	0.5660	5.8277
75	MS2010	2	2	3	2.3333	101	259	0.3900	4.4377
76	MS2024	2	2	3	2.3333	156	447	0.3490	5.2174
77	MS3007	3	2	2	2.3333	162	337	0.4807	6.2400
78	MS3008	3	1	3	2.3333	171	347	0.4928	6.4911
79	MS3009	2	2	3	2.3333	137	307	0.4463	5.5289
80	MS4010	2	2	3	2.3333	109	222	0.4910	5.1729
81	TS2002	2	2	3	2.3333	108	193	0.5596	5.4971
82	TS2006	3	1	3	2.3333	111	214	0.5187	5.3654
83	TS2009	1	2	4	2.3333	134	270	0.4963	5.7664
84	MH1008	2	2	2	2.0000	95	202	0.4703	4.7264
85	MS2009	2	1	3	2.0000	128	235	0.5447	5.9042
86	MS2013	2	1	3	2.0000	177	398	0.4447	6.2736
87	MS2026	3	2	1	2.0000	133	221	0.6018	6.3262
88	TS2007	2	2	2	2.0000	135	237	0.5696	6.2008
89	TS4004	2	3	1	2.0000	144	337	0.4273	5.5467
90	MH1006	2	1	2	1.6667	125	250	0.5000	5.5902
91	MH2001	1	1	3	1.6667	102	244	0.4180	4.6173
92	MH4005	3	1	1	1.6667	151	333	0.4535	5.8511
93	MS1005	2	2	1	1.6667	124	239	0.5188	5.6716
94	MS3005	2	1	2	1.6667	115	216	0.5324	5.5329
95	MS1004	2	1	1	1.3333	115	279	0.4122	4.8683
96	MS1009	2	1	1	1.3333	117	232	0.5043	5.4316
97	MS2017	1	1	2	1.3333	94	177	0.5311	4.9960
98	TS2016	1	1	2	1.3333	125	213	0.5869	6.0563
99	TS4002	1	2	1	1.3333	125	250	0.5000	5.5902
100	TS3004	1	1	1	1.0000	95	206	0.4612	4.6803
	平均	2.99	2.89	3.46	3.1133	145.92	310.89	0.4849	5.8579
	相関	A×B	A×C	B×C					
		0.5527	0.4365	0.4015					

A Study on the Validity of the Indices Expressing Lexical Variation: Using a Corpus of L1 Japanese Writing by College-level Students

TAJIMA Masumi*, FUKADA Atsushi**, SATO Naoko***

* Faculty of Law, Chuogakuin University

** Department of Foreign Languages and Literatures, Purdue University

*** Center for International Research and Education, Chiba University

Abstract

The aim of this study is to investigate the validity of two indices expressing the degree of lexical variation. One is the type/token ratio (TTR), which has been commonly used in the field of applied linguistics including Second Language Acquisition. Recent studies in SLA, however, have questioned its validity, and another index was developed to compensate for its weakness. The new index is a modified version of the type/token ratio, having the denominator changed to the square root of the number of tokens doubled.

This study statistically examined whether these two indices are valid measures of the quality of L1 Japanese compositions. One hundred L1 compositions written by Japanese students were collected and digitized to construct a corpus for quantitative analysis. Each composition was subjectively scored by three independent raters of Japanese language experts, on the one hand, and a computer program computed its type count, token count, and two measures of lexical variation, on the other. Spearman's rank-order correlation coefficients were computed to see how much these five variables correlate with each other.

The results gave statistical support to the validity of the new index. There was almost no correlation between the subjective scores and the traditional type/token ratio. This suggests that the new index should be used for the measurement of language products instead of the traditional measure.