

作文に用いられる語彙と評価について —語彙のレベル別語数という観点から—

長谷川守寿

1. 目的

本研究の目的は、プレースメントテストの一環として行われる作文テストと、その評価の関係を作文で使用される語彙レベルから考察し、作文テストの評価システム構築に活かしていくことである。

プレースメントテストに作文テストを含めた場合、クラス運営の都合上などから、大量の枚数の評価を短時間に、しかも限られた人数で行うことが求められる。これは評価者への負担が非常に大きく、評価する際の客観的な指標が望まれる。本稿では、作文評価の際に作文を構成する語彙のレベル情報が一つの基準となりうると考え、語彙のレベルと評価の関係を考察し、自動作文評価システム作成への一歩としたい。

作文の評価には、もちろん実際に作文を読み、内容を吟味することが重要であり、書かれている文の意味の側面からの評価を欠くことはできないのは当然であるが、本研究では、語彙のレベルから作文の評価を見た場合、どのようなことがいえるのか明らかにし、本研究の成果を自動評価に繋げていきたいと考えている。

2. 先行研究

まず、作文評価基準案として、菊池(1987)がある。菊池(1987)は、外国人が日本語の作文を書くための能力は「日本語能力」と「文章能力」に大きく分けられるとし、留学生（特に初級後期から中級）の作文を点数で評価する場合の試案を提案している。その中で、評価する項目として五つを挙げ、語彙自体としては「正確さ」という項目で調べられている。その中で、難しい語彙を使用したという積極性については、加点の対象となり、つまらない表現や稚拙な表現の使用などは減点の対象となっている。しかし、配点の中心となる内容における語彙の扱いについては不明である。この試案の問題点として、採点に時間がかかるのではないかという点が挙げられる。

また、作文評価の観点に関する研究としては、田中・坪根・初鹿野(1998)、田中・初鹿野・坪根(1998)がある。田中・坪根・初鹿野(1998)では、日本語教師と一般の日本人が初級から上級までの6レベルの作文を評価した時の観点(22の評価項目)を因子分析し、作文評価の基本構造が示されている。さらに、田中・初鹿野・坪根(1998)

では、実際の作文評価において作文評価構造のどの要素が重視されるか検討し、「正確さ」などを重視するグループと高度な文型などを重視するグループがあり、ともに「内容」「主旨の明確さ」「構成」が良い作文の決定要因と考えている可能性があることを明らかにした。この中では、教師により語彙は「よい作文」の評価に決定要因となっているが、どのように作用しているかについては、明らかではない。

長谷川(2004)では、プレイスメントテスト内の作文と評価の関係について考察する中で、延べ語数、異なり語数は、評価と高い相関があることを示した。また品詞数と作文評価では、助詞・動詞・名詞がより高い相関を持つことが分かった。長谷川(2006)では、作文に含まれる文の数、一文中の格成分の数・従属節の数という構文情報と評価の関係を考察した。結果として、これらの項目の中に評価と強い相関があるものは見あたらなかった。これらの研究から、作文の評価基準として形態素情報と評価が強い相関があることが分かったが、さらに関連性の高い項目があるかどうか、検討する必要がある。

3. 方法

本研究では、自動作文評価システム・作文評価補助システムを目標とした研究の一環として、作文評価と作文の語彙レベルについて考察する。作文評価の手掛かりとしては、作文の表層情報と深層情報が考えられる。表層情報には文字列情報と形態素情報が考えられ、深層情報には、構文・意味情報が含まれると考える。長谷川(2004)では、作文テストの表層情報に着目し、形態素情報と評価の関係について考察し、長谷川(2006)では文の深層構造の中の構文情報に注目し、作文における構文情報と評価の関係を考察した。

本研究は、これらの一連の考察に続くものとして、作文に使用された語に与えられたレベルを観点に分析を行う。ここでいう語のレベルとは、日本語能力試験での級別分類を指す。プレイスメントテストの受験者は、日本語レベルがさまざまな学習者の集団であるが、多少の差はあっても、初級から上級に進むに従って導入された語彙と、その導入時期はおおむね同じと考えられる。従って、学習者が書いた作文の語彙を分析するには、日本語能力試験での級別分類という、広く知られた明確な基準を手がかりとすることにより、評価との関係を一定の客観性を保って観察できると考える。

本研究では、レベル別語数と評価の関係を明らかにし、作文テーマや課題に依存しない、作文評価と強い相関がある項目を探る。

3. 1. 作文データ

筑波大学留学生センターで実施されたプレイスメントテストの中の、作文テストとその評価を対象とする。プレイスメントテストでは、日本語レベルを判定し、受講できる授業を決定するため、留学生センターで開講している授業の履修を希望する学習者は全て受験しなければならず、受験者のモチベーションは高いと思われる。作文データは、1999年2学期に実施された、筑波大学の学生の起床時間と就寝時間に関する調査結果を見て書かれた67人分（以後「99年作文」と呼ぶ）と、2002年1学期に実施された、4コマ漫画の説明をする課題112人分（以後「02年作文」と呼ぶ）である（参考資料参照）。ともに15分間で書かれたもので、対象者の国籍、日本語歴、レベルは様々である。全数調査を予定したが、プレイスメントテスト全てを受験していない学生のデータや、長谷川(2004,2006)で明らかに外れ値となったデータも除外したため、99年作文は66人分、02年作文は106人分である¹⁾。

3. 2. 「リーディングチュウ太」について

本研究で考察の対象となるデータは、全て「リーディングチュウ太」²⁾内のレベル判定ツールを使用した（2006年12月2日実行）結果である。「リーディングチュウ太」

（川村他(2000)）は学習者による読解の自習を支援するツールである。その中の語彙チェッカーは、日本語能力試験を基準にして単語の難易度を判定するプログラムで、テキストに含まれる語彙のレベル判定(1~4級)を行うことが出来る。本研究では、各単語のレベルをこのツールを使用して調べ、作文に含まれるレベル別語数を集計する。

このツールはインターネット上で使用することが出来、(1)のような作文を入力すると、ブラウザ上に図1のような結果が出力される。

- (1) 筑波大学の学生に「あなたは普段何時に寝ますか。そして何時に起きますか」という調査をしたところ、つぎの答えがでた。筑波大学の学生は、平均的に夜中の1時位に寝て、朝の8時に起きる。この時間帯が一番高パーセンテージ

1 学生の国籍は、99年作文では中国・26人、韓国・17人、アメリカ・10人、台湾・4人、マレーシア・3人、スロベニア・2人、オーストラリア・スリランカ・ブラジル・メキシコ・各1人で、02年作文では、中国・42人、韓国・35人、タイ・4人、ハンガリー・ペルー・各3人、アメリカ・マレーシア・ロシア・各2人、アゼルバイジャン・イギリス・オーストラリア・カナダ・スリランカ・セネガル・チリ・パキスタン・フィリピン・ブラジル・フランス・メキシコ・台湾・各1人である（調査用紙より）。

2 (<http://language.tiu.ac.jp/tools.html#input>)

を占めている。勉強が忙しいあるいはアルバイトの帰りが遅いことから、1時に寝る者は45%以おり、授業の限目は8時40分開始なので、8時に起きて充分間に合うと思って、その時間に起きる学生は37%位いる。

中にも、夜明けの4時に寝る者もあり、全体の2%という位置を占めている。バランスよく、午後の2時に起きる者も見られ、それは3%位である。おそらく、朝の4時あるいは5時まで起きている者は、遊びではなく、夜中の方が静かで、勉強や研究に集中できると言って、夜起きて仕事して、昼間は寝る。あまり健康的ではないが、一時的な習慣で研究の為ならよいだろう。

(99年作文より)

図 1. リーディングチュウ太の処理結果画面

Reading Tutor Homepage - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(O) お気に入り(A) ツール(T) ヘルプ(H)

アドレス(A) http://language.tiu.ac.jp/result/jsol/29712AFE.html

リーディング チュウ太

日本語 / English / Deutsch

入力された文章

筑波大学の学生に「あなたは普段何時に寝ますか。そして何時に起きますか」という調査をしたところ、つぎの答えがでた。筑波大学の学生は、平均的に、夜中の1時に寝て、朝の8時に起きる。この時間帯が一番高パーセンテージを占めている。勉強が忙しいあるいはアルバイトの帰りが遅いことから、1時に寝る者は45%以おり、授業の限目は8時40分開始なので、8時に起きて充分間に合うと思って、その時間に起きる学生は37%位いる。

中にも、夜明けの4時に寝る者もあり、全体の2%という位置を占めている。バランスよく、午後の2時に起きる者も見られ、それは3%位である。おそらく、朝の4時あるいは5時まで起きている者は、遊びではなく、夜中の方が静かで、勉強や研究に集中できると言って、夜起きて仕事して、昼間は寝る。あまり健康的ではないが、一時的な習慣で研究の為ならよいだろう。

級外 6 (5)
筑波大学 2
答え 1
パーセンテージ
以
限

1級 4 (2)
的 3
や 1

2級 29 (22)
者 4
夜 3

単語レベル: ★★ やさしい

総数	語彙総数	級外	1級	2級	3級	4級	その他
256	205	6	4	29	16	150	51
-	100%	2.9%	2.0%	14.1%	7.8%	73.2%	-
(114)	(101)	(5)	(2)	(22)	(14)	(58)	(13)
-	(100%)	(5.0%)	(2.0%)	(21.8%)	(13.9%)	(57.4%)	-

結果は、3つの区画に分かれて表示される。①は作文を形態素解析した結果で、級別に色分けされて表示される。②では、作文に含まれるレベル別語数が辞書形とともに表示される。例えば図1の場合、1級の語は延べ語数で4語、異なり語数で2語、そして実際には「的」が3回、「や」が1回現れていることを表している。③は当該文に対する判断と、②の数値が集計され表示される。本研究では、③の部分より「総数」から「その他」、さらにその2段下で丸括弧“()”で囲まれた部分を抜き出す。

上段が延べ語数、2段下が異なり語数を表している。この語数の部分を「異なり総数」「異なり総語彙数」「異なり級外」「異なり1級」「異なり2級」「異なり3級」「異なり4級」「異なりその他」「延べ総数」「延べ総語彙数」「延べ級外」「延べ1級」「延べ2級」「延べ3級」「延べ4級」「延べその他」として抜き出し、データとする。以後「異なり1級」などの表現は、「数」という言葉を用いないが、レベル分けされた語数を表す。さらに表中の「総数」は「総語彙数」と「その他」の合計で、「総語彙数」は「級外」「1級」「2級」「3級」「4級」の各語数の合計である。例えば、図1の「総数」256語は、「総語彙数」205語と「その他」51語の合計であり、205語は、「級外」6語「1級」4語「2級」29語「3級」16語「4級」150語の合計である。「その他」には、括弧等の記号・アラビア数字・句読点などが含まれる。

3. 3. 「リーディングチュウ太」で処理を行う場合の問題点

「リーディングチュウ太」は元来自習用のツールで、日本語の読解教材の自習に活用することを想定したツールである。そこで本研究のように、日本語としての誤りを多く含む文を処理した場合、正確な結果が出てこないという問題が発生する³。例えば、以下の文を対象とした場合を検討する。

(2) 学生だち中で八時におきる人だちがもっとも大いです。(中略) たいていの学生だちがまい日7時間くらいねるとかていすれば、(以下、略) (99年作文より)

(2)には、「学生だち」「大い」のような語彙の間違いや、「まい日」のような交ぜ書きが見られる。この場合、「だち」の「ち」が「ちる」(2級)の活用形としてカウントされ、「まい日」の「まい」が「まう」(2級)の活用形としてカウントされる。また、表記によってレベル判定が異なるという問題が発生する。例えば、時刻の表記には、前記の(2)内の「八時」の他に、(3)(4)のような形が見られる。

(3) いちばんの学生が8時に起きます。 (99年作文より)

(4) 約37%の学生は、毎日8:00おきます。 (99年作文より)

「リーディングチュウ太」では数字「8:00」は「その他」へ分類されるが「8時」は「8」が「その他」、「時」が「4級」に分類される。また「八時」は「八」「時」ともに「4級」に分類される。同様に「5%」は「5」「%」ともに「級外」に分類さ

³ 他の問題点には「1限目」「パーセンテージ」が級外と判定されたり、「為」が現在の表記ではあまり使われないが3級と判定されるなど、日本語能力試験自体の問題も見られるが、今回は研究の目的と異なるため、指摘にとどめる。

れるが、「5パーセント」の場合は、「5」（級外）と「パーセント」（2級）になる。

また、文として正しく最後まで書かれた文とそうでないものでは、形態素解析の結果が異なり、語彙のレベル判定も異なってしまう。たとえば、「もっとも大いです。」では「もっとも／大／い／です。」となるが、「もっとも大いです」と句点「。」がない場合は「もっとも／大／い／で／す」となる。

このように、表記や文法に関する誤りが多く、また通常漢字で表記される文字がひらがなで書かれることが多い作文を解析した場合、得られる結果は我々が期待するものとは大きく異なることが多い。ただ、作文評価の自動化を想定した時、人手で修正する処理を加えることは自動化という考え方に反する。また長谷川(2004)では、品詞数と評価の関係を見た時、形態素解析の結果への修正の有無には有意な差がなかったことが明らかになっている。本研究では、全てをゼロから作り出すことが不可能であることもあるが、システム構築の方針として、使用するツールの解析結果に多少誤りが含まれていても、それらを使用して構築することとしたい。

3. 4. 作文の評価について

作文データの評価について説明する。評価は、日本語教育歴が長くさらに作文教育担当の経験がある教師によって行われ、0から6までの7段階でつけられている。

99年作文の評価者は1名、02年作文の評価者は2名である。02年作文に対して2名が下した評価の相関は、相関係数は0.83で、2名の評価者間には強い相関があるといえる。本研究では、02年作文に関しては2名の評価者を使用して、99年作文と02年作文というテーマによる違いと、2名の評価者（以後評価者A、Bと呼ぶ）による個人差も考察する⁴。

3. 5. 考察の観点

本研究では以下の観点から、考察を行う。まず、作文評価では、色々な語彙を使うことがいいのか、そうではなくたくさん書くことがいいのかを観察するために、99年作文と02年作文を対象として、レベル別語数と評価の関係を、異なり語数と延べ語

⁴ 99年作文（66人分）の評価別データ数は、評価0が6、評価1が5、評価2が16、評価3が13、評価4が6、評価5が9、評価6が11である。02年作文（106人分）は評価者Aの場合、評価0が0、評価1が2、評価2が32、評価3が24、評価4が18、評価5が20、評価6が10である。評価者Bの場合、評価0が2、評価1が3、評価2が19、評価3が12人、評価4が20、評価5が21、評価6が29である。

数という観点から考察する。さらに 99 年作文と 02 年作文の結果を比較することで、作文の課題による違いも考察する。またこれに関連して、評価者の違いが存在するの
かを確認するために、02 年作文の評価者 A・B の評価とレベル別語数の関係を見る。

二番目の観点は、複数の項目と評価の関係である。一番目の観点では、各レベル別
語数と評価の関係であったが、ここでは複数のレベル別語数と評価の関係をみる。日
本語能力試験（国際交流基金他(2002)）では、1 級を「日本語を 900 時間程度学習し
たレベル」、2 級を「日本語を 600 時間程度学習し、中級日本語コースを修了したレベ
ル」、3 級を「日本語を 300 時間程度学習し、初級日本語コースを修了したレベル」、
4 級を「日本語を 150 時間程度学習し、初級日本語コース前半を修了したレベル」と
設定している。このように、レベル別語彙は導入時期に違いが見られるため、プレ
イストメントテストでは、大きく見てどの時期の語彙が評価と相関があるのか、検討する。
考察する項目は、導入時期順に連続したもの、つまり 4 級と 3 級、3 級と 2 級、2 級
と 1 級というように二つのレベルに含まれる語数と評価の関係を調べる。三つの要素
に増やしていくことも考えられるが、その数値は「総数」「総語彙数」に近いものとな
り、考察に値しないと考える。また「その他」は 4 級の下に置くこととする。4 級に
は記号や数字などが含まれ、これらは他の言語にも存在し、日本語を学習する以前か
ら知っているものなので、4 級以下と考える。また「級外」には固有名詞の他に、文
法的に間違っているために語の連続としては通常考えられない解析結果である「り」
「よ」「べ」「る」や、“_”⁵が多く含まれ、誤りに起因しているとして学習レベルは
4 級以下とした。また、数値だけであれば、「延べ 1 級」と「異なり 1 級」の合計と評
価の関係や「延べ 1 級」と「異なり 4 級」の合計と評価の関係を考察することも可能
だが、語彙の導入時期と関係づけた解釈が出来ないため、これらは不要と考える。

三番目の観点は、評価別に作文を分けた場合の評価と語数の関係である。一般的に、
上位の学生の方が文法的に正しい文が書けると考えられるため、正しく形態素解析が
行われ、作文に含まれるレベル別語数も正しくカウントできると考えられる。そのた
め、レベル別語数と評価の相関は、下位の学生の場合より強いと推測される。また、
この観点が有効ならば、大まかに作文を上位・下位に分け、上位と予想されるもの
のみに適用するなど、様々な方法が考えられるため、作文データを評価により上位・下
位のグループに分けて考察を行う。二つに分ける理由は、データ数による制約もある

⁵これは、作文に書かれたひらがなや漢字に、日本語に相当するものがない場合に入力
したものである。

が、実際面での作業の効率なども考慮に入れた結果である。なおデータを二つに分ける際、7段階評価（0 から 6）であるため、評価 3 を上位・下位のどちらに入れるか問題である。99 年作文では、評価 3 を上位に入れた場合、上位が 39 人分（評価 3～6）、下位が 27 人分（評価 0～2）、評価 3 を下位に入れた場合、上位が 26 人分（評価 4～6）、下位が 40 人分（評価 0～3）となる。どちらを採用しても、半数に近い数値にはならないため、両方の場合について考察する。

4. 結果

作文データを構成する個々の作文の書き手が異なり、書き手の構成も異なるため、単純に比較できないが、データ全体を比較した時、表 1 のような結果となった⁶。

表 1. 99 年作文と 02 作文の語彙情報(単位:語)

		99 年作文(n=66)			02 年作文(n=106)		
		平均	比率	SD	平均	比率	SD
異なり総数		70.2	100.0%	21.4	73.7	100.0%	20.9
異なり総語彙数		60.1	85.6%	20.7	69.6	94.4%	20.1
内	異なり級外	5.2	7.4%	2.7	6.4	8.7%	3.3
	異なり1級	0.8	1.1%	1.2	1.3	1.8%	1.2
	異なり2級	7.6	10.8%	4.4	8.5	11.5%	4.6
	異なり3級	5.7	8.1%	3.9	10.1	13.7%	5.8
	異なり4級	40.7	58.0%	12.5	43.4	58.9%	10.7
異なりその他		10.1	14.4%	4.0	4.2	5.7%	2.0
延べ総数		152.7	100.0%	58.2	147.9	100.0%	47.0
延べ総語彙数		119.0	77.9%	47.5	130.0	87.9%	41.3
内	延べ級外	6.8	4.5%	3.8	9.0	6.1%	5.2
	延べ1級	1.0	0.7%	1.4	1.5	1.0%	1.5
	延べ2級	9.8	6.4%	5.9	10.9	7.4%	5.9
	延べ3級	7.2	4.7%	5.2	12.7	8.6%	7.7
	延べ4級	94.3	61.8%	37.8	96.0	64.9%	29.5
延べその他		33.6	22.0%	18.6	17.9	12.1%	8.9

⁶分析には柳井久江『4Steps エクセル統計』(1998 年、オーエムエス出版)内の Statcel を使用した。

99年作文(66人分)では、1作文あたりの「異なり総数」が平均70.2語、「異なり総語彙数」が60.1語、「異なりその他」が10.1語である。「異なり総語彙数」の内訳は、「異なり級外」が5.2語、「異なり1級」が0.8語、「異なり2級」が7.6語、「異なり3級」が5.7語、「異なり4級」が40.7語である。延べ語数、02年作文も同様である。

99年作文と02年作文では、「異なり総数」・「延べ総数」ともそれ程大きな違いがない。単語数に限っていえば、両者とも作文の量的側面には大きな違いはないといえる。それに対して、級別語彙では、99年作文の「異なりその他」「延べその他」が多く、逆に02年作文の「異なり3級」と「延べ3級」は99年作文に比べ多い。そこで、レベル別語彙において延べ語の出現頻度を調べる。方法としては、「リーディングチュウ太」の領域②の情報、例えば前出の図1では「1級4(2)的3や1」の「的」が3回現れ「や」が1回現れたという情報を集計した。上位10位までを見ると、表2・表3のようになる。1級の語彙「的」が、99年作文全体で12回現れたことを示している。

表 2. 99年作文の詳細・出現頻度の多い語(同頻度の場合は文字コード順)

順位	1級		2級		3級		4級		級外		その他	
	語	頻度	語	頻度	語	頻度	語	頻度	語	頻度	語	頻度
1	的	12	グラフ	81	思う	51	の	434	筑波大学	73	。	490
2	べし	4	調査	21	こと	45	は	381	ごろ	31	、	355
3	割	4	間	18	れる	18	時	368	—	19	0	252
4	まえる	3	数	18	さ	14	に	331	によって	17	1	242
5	えい	2	約	16	について	14	学生	277	起る	9	%	100
6	かする	2	ば	15	あ	12	ます	269	つくば	8	5	92
7	はじる	2	はねる	12	ので	12	が	220	に対して	8	2	85
8	上回る	2	帯	12	ほとんど	12	時間	217	べ	7	8	83
9	正午	2	より	11	生活	12	と	210	る	7	:	75
10	正常	2	下	11	以上	10	て	188	早	7	3	68

99年作文では、筑波大学の学生150人に「おきる時間」と「ねる時間」について訊いた結果であるグラフを見て、説明する課題である。1級の「割」、2級の「グラフ」「調査」、4級の「時」「学生」「時間」、級外の「筑波大学」「ごろ」「つくば」、「その

他」の「0」「1」「2」「3」というローマ数字や「%」は、グラフを読み取らせると
いう課題の影響と思われる⁷。

表 3. 02 年作文の詳細・出現頻度の多い語(同頻度の場合は文字コード順)

順	1級		2級		3級		4級		級外		その他	
	語	頻度	語	頻度	語	頻度	語	頻度	語	頻度	語	頻度
1	体力	14	より	135	思う	123	た	655	田中	68	。	889
2	先	8	老人	56	こと	86	は	603	—	52	,	564
3	あせる	7	ば	31	しまう	62	て	574	A	34	「	70
4	主人公	7	ら	21	さ	58	ます	539	って	22	」	69
5	がん	6	—	21	運動	41	の	490	り	21	”	58
6	若者	6	日	20	れる	37	が	411	ジョギング	16	.	41
7	的	5	かれる	19	彼	36	に	372	けど	15	!	36
8	べし	4	後	19	ので	33	を	359	速	15	O	24
9	固	4	汗	17	よう	31	人	288	年上	14	.	22
10	駆る	4	今日	16	考える	27	と	273	よ	13	2	15

02 年作文は、ジョギングをしていた主人公が若い人に追い抜かれ、さらに老人に追
い抜かれたので、老人を追い抜くが道端にへたり込んでしまうという 4 コマ漫画であ
る。1 級の「体力」「主人公」「若者」、2 級の「老人」「汗」、3 級の「しまう」「運動」
「彼」などが作文の課題からの影響と思われる⁸。作文の課題によってレベル別語数が
異なることについては表 1 から明らかで、比率を見た時、99 年作文では「延べその
他」、02 年作文では「延べ 3 級」の割合が多い。作文全体の異なり語数⁹を検討するた
めに表 5 を挙げる。これは領域②をもとに筆者が集計したものである。02 年作文では
「延べその他」に比べ「異なりその他」の比率が小さいことが分かるが、これはロー

⁷ リーディングチュウ太では、図 1 に示した①と②が連動しており、作文中のどの語
が、辞書形にどのような語を持つものと解析されたかが画面上で確認できるようにな
っている。表 2 において、1 級の「まえる・えい・かする・はじる」は、作文に含ま
れる誤りから起こったもので、2 級の「はねる」は「学生はねる」(「寝る」の意味)
のような表記によって起こったものである。

⁸ 1 級の「がん」や 2 級の「かれる」は「がんかえて」(「考えて)」「がんしん」(「感
心)の一部、「かれは」(「彼は」のひらがな表記)等、誤りから起こったものである。

⁹ ここでの異なり語数は全作文データの異なりであり、表 1 の異なり語数は学生 1 人
ずつの作文に含まれる異なり語数を平均したものである。数値の性質が異なる。

マ数字の使用が多いことの表れである。

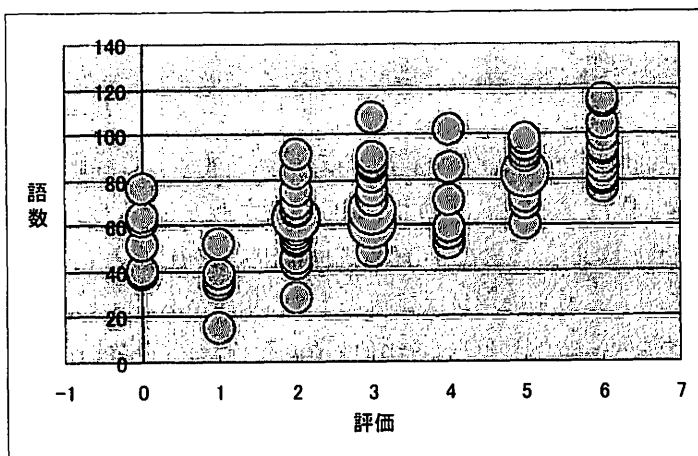
表 4. 99 年作文・02 年作文の全データの異なり語数と比率

	1級	2級	3級	4級	その他	級外	合計
99 年作文	38 4.2%	223 24.8%	135 15.0%	303 33.7%	32 3.6%	167 18.6%	898 100%
02 年作文	81 5.6%	351 24.3%	213 14.7%	414 28.7%	37 2.6%	349 24.2%	1445 100%

4. 1. レベル別語数と評価

99 年作文について「異なり総数」と評価の関係をみる。図 2 に示したように、評価が上がるに従って「異なり総数」も上昇しているように見え、評価と「異なり総数」は一見相関がありそうである。そこで作文の評価とレベル別語数の関係をスピアマンの順位相関係数により調べる¹⁰と、表 5 のような結果となった。これは「異なり総数」と評価の相関係数が 0.68 で、“*” は 5%水準で有意であることを示している。

図 2. 99 年作文における「異なり総数」と評価の散布図



¹⁰相関関係をもとにした相関関係の強さの目安には色々異なる種類のものがあるが、本研究では柳井(1998,p164)に従うものとし、相関係数 (0.0 ~ ±0.2) を「ほとんど相関がない」、(±0.2 ~ ±0.4) を「やや相関がある」、(±0.4 ~ ±0.7) を「相関がある」、(±0.7 ~ ±0.9) を「強い相関がある」、(±0.9 ~ ±1.0) を「極めて強い相関がある」とする。

表 5. レベル別語数と評価の相関係数

	99 年作文	02 年作文	
		評価者 A	評価者 B
異なり総数	0.68*	0.64*	0.76*
異なり総語彙数	0.65*	0.65*	0.77*
異なり級外	0.39*	0.16*	0.31*
異なり1級	0.25*	0.22*	0.26*
異なり2級	0.56*	0.54*	0.69*
異なり3級	0.64*	0.72*	0.71*
異なり4級	0.55*	0.54*	0.63*
異なりその他	0.33*	0.13	0.22*
延べ総数	0.73*	0.56*	0.68*
延べ総語彙数	0.71*	0.57*	0.69*
延べ級外	0.37*	0.10	0.26*
延べ1級	0.26*	0.17*	0.23*
延べ2級	0.52*	0.47*	0.61*
延べ3級	0.60*	0.72*	0.72*
延べ4級	0.71*	0.52*	0.60*
延べその他	0.49*	0.29*	0.40*

(*:5%水準で有意)

99 年作文では、「延べ総数」「延べ総語彙数」「延べ4級」と評価の相関係数がそれぞれ 0.73、0.71、0.71 となり、評価と「強い相関がある」といえる。02 年作文については、評価者 A・B 共に、「異なり3級」の相関係数が 0.72 と 0.71 で、「延べ3級」が共に 0.72 となり、これらの項目は評価と「強い相関がある」といえる¹¹。

表 1 で比率が高かったのは「延べその他」(99 年作文)、「延べ3級」(02 年作文)で、評価と強い相関がある項目は異なる結果となった。さらに 99 年作文については「延べ4級」、02 年作文については「延べ3級」「異なり3級」というように「強い相関」を示す項目が異なっている。したがって、作文評価に関係が深い項目を抽出する目的からは、「強い相関がある」項目を特定できる結果とはならなかった。

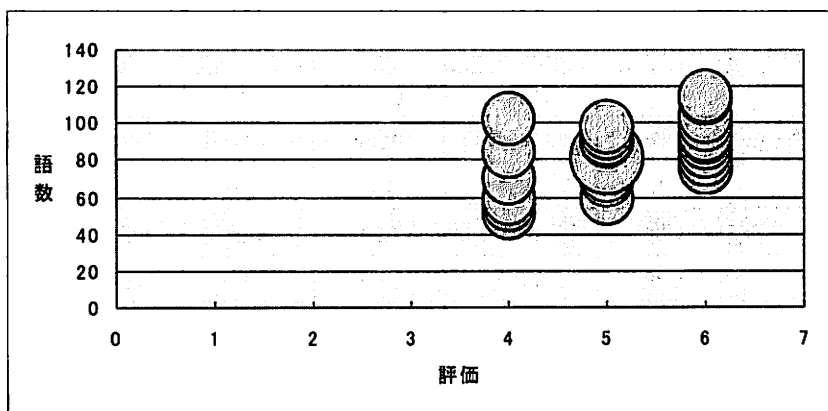
¹¹ なお評価者 B のみ、「異なり総数」(相関係数 $r_s=0.76$)「異なり総語彙数」(相関係数 $r_s=0.77$) が、評価と「強い相関がある」となった。

さらに評価者による評価の差異を検討する。評価者 A と B では、数値の面では全般的に評価者 B の方が数値が高く、レベル別語数と評価という観点からは、評価者間の個人差も伺われるようであるが、「強い相関」を示す項目は共通しており、99 年作文の項目と一致することはなかった。

次に、レベル別語数と評価の関係を調べる際に、延べ語数でみるか異なり語数でみるかについて考察する。表 5 で示したように、99 年作文では、延べ語数の総数・総語彙数・4 級と評価に「強い相関」があり、02 年作文の場合、異なり語数でも延べ語数でも 3 級に「強い相関」があった¹²。前節と同様、二つの作文データに共通して「強い相関がある」項目は見あたらない結果となった。

4. 2. レベル別語数と上位・下位評価

図3. 99 年作文の「異なり総数」と評価の散布図(評価上位グループ)



99 年作文において、「異なり総数」と評価の関係を、評価の上位グループ(26 名分)に限って見たのが図 3 である。図 2 に比べ評価 0~3 の部分がないため、作文評価が上がるに従って、「異なり総数」も上昇傾向に見える。そこで、評価上位グループと評価下位グループに分け、レベル別語数と評価の関係を 4.1 と同様に調べると、表 6 のようになる。結果として、99 年作文・02 年作文の評価との相関係数では、99 年作文の評価下位グループと「異なり 3 級」の相関係数 0.64 が最高で、評価と「強い相関がある」と考えられる、相関係数が 0.70 以上の項目はなかった。なお、99 年作文の評価上位グループに評価 3 の作文 (13 名分) を加えて、上位 39 名・下位 27 名とした場合も結果は同様であった。このことから、評価の上位・下位と関わるかたちで作

¹² 評価者 B のみ評価と「異なり総数」「異なり総語彙数」にも「強い相関」があった。

文自動評価の指標となるような「強い相関」を示す項目はないといえる。

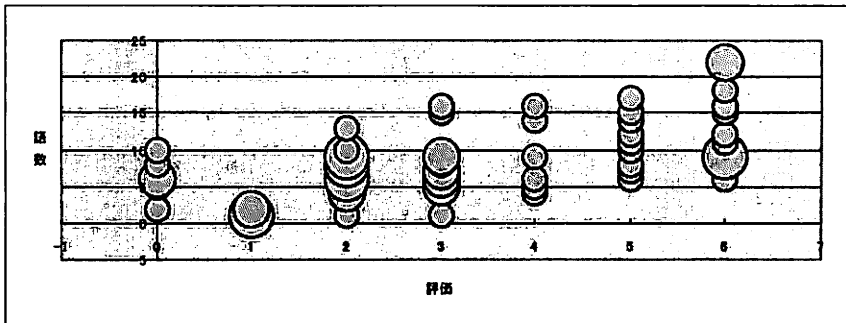
表 6. レベル別語数と評価の相関係数(上位グループと下位グループに分けた場合)

	99年作文		02年作文			
			評価者 A		評価者 B	
	上位:26名	下位:40名	上位:48名	下位:58名	上位:70名	下位:36名
異なり総数	0.51*	0.51*	0.34*	0.28*	0.55*	0.44*
異なり総語彙数	0.50*	0.46*	0.36*	0.29*	0.55*	0.44*
異なり級外	0.30	0.41*	0.05	-0.21	0.27*	0.15
異なり1級	0.30	0.21	0.29*	-0.19	0.19	-0.18
異なり2級	0.35*	0.22	0.28	0.24	0.49*	0.39*
異なり3級	0.50*	0.64*	0.31*	0.57*	0.47*	0.38*
異なり4級	0.42*	0.35*	0.37*	0.25	0.41*	0.39*
異なりその他	0.03	0.19	-0.14	-0.07	0.12	0.03
延べ総数	0.43*	0.56*	0.34*	0.38*	0.44*	0.46*
延べ総語彙数	0.47*	0.43*	0.38*	0.38*	0.42*	0.43*
延べ級外	0.12	0.39*	0.02	-0.08	0.14	0.20
延べ1級	0.34	0.18	0.29	-0.21	0.15	-0.19
延べ2級	0.29	0.19*	0.31*	0.18	0.40*	0.35*
延べ3級	0.52*	0.57*	0.39*	0.62	0.43*	0.42*
延べ4級	0.48*	0.38*	0.36*	0.38*	0.31*	0.40*
延べその他	0.21	0.42*	0.02	0.30*	0.25	0.34*

(*:5%で有意)

4. 3. 連続する二つのレベル別語数と評価

図4. 99年作文「異なり1級」「異なり2級」の合計語数と評価の散布図



ここでは、作文の評価と、複数のレベル別語数の関係を見る。例として、99年作文の評価と「異なり1級」「異なり2級」の合計の散布図を図4に示す。評価が上がるに従って「異なり1級」「異なり2級」の合計も上昇しているように見え、相関がありそうである。そこで作文の評価と連続する二つのレベル別語数の関係を4.1と同様に調べると、表7のような結果となった。結果的には、やはり共通して「強い相関がある」項目はなく、作文評価に有効な項目は抽出できなかった。0.70以上の強い相関があった項目は、99年作文では「延べ3級」と「延べ4級」の合計と評価(0.70)、「延べ4級」と「延べその他」の合計と評価(0.71)、「延べ4級」と「延べ級外」の合計と評価(0.70)であり、以上から「延べ4級」が含まれた場合、評価との「相関関係が高い」と言える。一方、02年作文では、評価者AB共通して、「異なり2級」と「異なり3級」の合計と評価(評価者A:0.72、B:0.78)、「延べ2級」と「延べ3級」の合計と評価(評価者A:0.70、B:0.76)で「強い相関がある」結果となり、99年作文との共通性はみられない。

99年作文では、初級前半の語をたくさん使うことが評価と「強い相関」があり、02年度作文では、初級後半から中級の色々な種類の語を使うことが評価と「強い相関」がある結果となった。こうした違いがみられることについては、テーマや課題そのものを考慮すること、たとえば、テーマや課題により深く関連する語彙とそのレベルを調べていく必要があるだろう。ただ、この点を検討するには、さまざまなテーマや課題で書かれた作文データと比較しなければならず、今後の考察課題としたい。

表 7.連続する二つのレベル別語数と評価の相関係数

	99年作文	02年作文	
		評価者A	評価者B
異なり1級と異なり2級	0.57*	0.53*	0.68*
異なり2級と異なり3級	0.67*	0.72*	0.78*
異なり3級と異なり4級	0.62*	0.65*	0.72*
異なり4級と異なりその他	0.63*	0.53*	0.63*
異なり4級と異なり級外	0.55*	0.51*	0.64*
異なりその他と異なり級外	0.50*	0.18	0.34*
延べ1級と延べ2級	0.55*	0.47*	0.61*

延べ2級と延べ3級	0.64*	0.70*	0.76*
延べ3級と延べ4級	0.70*	0.58*	0.66*
延べ4級と延べその他	0.71*	0.50*	0.60*
延べ4級と延べ級外	0.70*	0.49*	0.60*
延べその他と延べ級外	0.53*	0.22*	0.37*

(*:5%で有意)

5. まとめ

長谷川(2004)では、助詞・動詞・名詞が評価と「強い相関がある」ことが分かり、一方、長谷川(2006)では、文数・従属節数・格成分数といった構文情報と評価の間には、「強い相関がある」ことを示す結果は得られなかった。

そこで本研究では、作文のテーマや課題にかかわらず、単語のレベル別語数が作文評価と強い相関を持つのではないかの予想から、単語の前後のつながりを排除した、単語そのもののレベル別語数と評価について、色々な観点から調べた。その結果、予想に反し、二つの作文データに共通して強い相関がある項目を得ることはできなかった。レベル別語数を、延べ語数で集計するか異なり語数で集計するかにも、大きな違いは見られなかった。また、作文評価を上位・下位に分けた場合の評価と語数には、「強い相関がある」といえる項目はなかった。さらに、語彙の導入時期から見た語数と評価の関係でも、作文データによって違いが見られ、自動作文評価に有効な、強い相関関係を示す共通の項目はなかった。このことは、別の見方をすれば、長谷川(2006)で指摘した、作文のテーマ・課題による違いが大きいことを支持する結果でもある。

6. 今後の課題

本研究は、長谷川(2004、2006)に引き続き、自動作文評価システムの構築にむけて、作文に含まれるさまざまな情報と、作文評価の関係を考察したものである。これら一連の研究で、作文評価と「強い相関がある」項目として抽出されたのは、現在のところ形態素情報のみである。今後は、構文・語彙において評価と相関が高い項目がないか、引き続き別の観点から分析を試みると共に、形態素情報に基づいた自動評価システムの構築とその精緻化を行っていきたい。

また、本研究ではプレイメントテストにおける作文について、語彙のレベルと評価の関係を見てきたが、これがプレイメントテスト特有の問題なのかには検討の余

地がある。そのためにもアチーブメントテスト等における作文と評価の関係などにも考察を行いたい。

謝辞：本研究で使用したデータは、筑波大学留学生センター西村よしみ教授からお借りしたものである。感謝申し上げます。

参考文献

- 川村よし子・北村達也・保原麗(2000)「EDR 電子化辞書を活用した辞書ツールの開発」『日本教育工学雑誌』第 24 号、7-12
- 菊池康人(1987)「作文の評価方法についての一私案」『日本語教育』第 63 号、87-104
- 国際交流基金・財団法人日本国際教育協会(2002)『日本語能力試験出題基準【改訂版】』、凡人社
- 田中真理・初鹿野阿れ・坪根由香里 (1998)「第二言語としての日本語における作文評価—『いい』作文の決定要因」『日本語教育』第 99 号、60-71
- 田中真理・坪根由香里・初鹿野阿れ(1998)「第二言語としての日本語における作文評価基準—日本語教師と一般日本人の比較—」『日本語教育』第 96 号、1-12
- 長谷川守寿(2004)「プレイスメントテストにおける作文の評価と形態的特徴の関係について」、筑波大学留学生センター『日本語教育論集』第 19 号、1-20
- 長谷川守寿(2006)「作文評価と構文的特徴について—留学生のプレイスメントテストを例に—」、『日本語研究』第 26 号、59-73
- 柳井久江(1998)『4Steps エクセル統計』、オーエムエス出版
(はせがわ もりひさ・首都大学東京准教授)

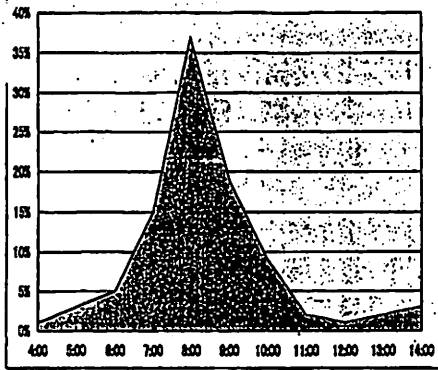
参考資料

99年作文の課題

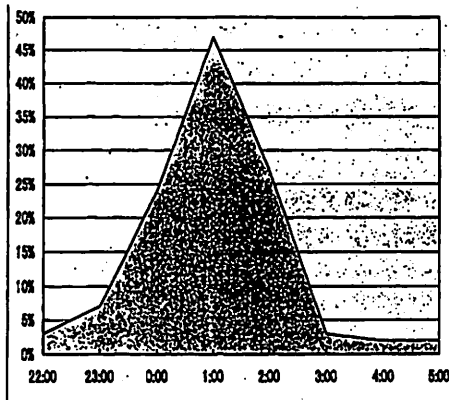
02年作文の課題

下のグラフは筑波大学の学生150人に「おきる時間」と「ねる時間」について聞いたものである。よく見てグラフを説明しなさい。

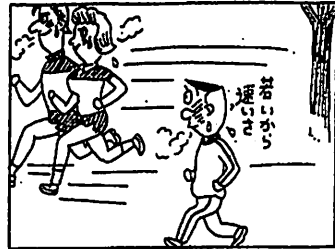
「おきる時間」



「ねる時間」



題 (topic) 絵を見て説明しなさい。
Explain the comic strip



若いわかい
速いはやい

