Jess: 日本語小論文の自動採点システム

— 入社試験における作文デ– タの評価 —

石岡 恒憲*　鷺坂 由紀子**　二村 英幸**
* 大学入試センター・研究開発部
** HRR 株式会社・測定技術部

1　はじめに

Jess は大学入試センター石岡らのグループが開発したわが国で最初の日本語小論文を自動採点するシステムである (石岡,2002)．小論文試験が Writing Ablity を測定しているするならば，コンピュータによる自動採点は，小論文試験に固有な以下の誤差要因を排除することができる．

- 評定者による違い
- 文字の巧拙 (文字の上手さ，綴りの正確性) による違い
- 評定の系列的効果 (ある小論文の評定が答案の中で何番目に行なわれたか)
- 課題選択 (異なる課題に基づいて書かれた小論文をどう評価するか)
- その他種々の誤差要因 (書き手の性別，人種など)

他にも，採点の手間を軽減させるという目的で，あるいは公平性，また最近では説明責任といった点からその開発が広く望まれていた．

Jess ではアメリカの経営大学院での入学試験 Graduate Management Admission Test, GMAT における小論文試験での採点基準をほぼそのまま踏襲する；文章の形式的な側面である「修辞」と，アイディアの理路整然とした表現の程度を示す「論理構成」と，トピックに関連した語彙が用いられているかを示す「内容」の 3 つの観点から小論文を評価する．Jess は毎日新聞の社説およびコラム (余録) を学習し，これを模範とした場合に適切でないと判断される採点細目に対して減点することで採点を行なう．また書かれた小論文の診断情報を提示する．システムは UNIX 上で動作し，800–1,600 字の小論文を通常能力のパソコン (Plat'Home Standard System 801S, Intel Pentium III 800MHz, RedHat7.2) で 1 秒程度で処理する．現在，Web(http://zaza.rd.dnc.ac.jp/jess/) で公開されており，原則として休日を除き終日，無料で使用することができる．

さて，小論文試験はいまや面接試験と並んできわめて一般的な試験の一つとなりつつあるが，その適用は大学入試に限られるものではない．企業の採用/昇進/昇格などの選考場面においてもしばしば実施される．特に近年，募集や初期選考段階においてウェブ上でエントリーシートと呼ばれるフォーマットに応募者の属性情報と共に小論文を書かせ，その内容によって第一次選抜を行う企業が増加している (二村・村井,1999)．作文評価が今後も採用選考に適用され，更に拡大する可能性を鑑み，著者らは先にある会社の採用選考において，言語能力検査に携わる専門家によって評価した結果と，Jess による自動採点の結果とを比較することを試みた．また，採用選考時に実施した基礎能力検査 GAT(HRR 株式会社) のうち，言語的理解力の得点との関連を示す．この言語的理解力は「文の要素である語の意味を正しく把握し，文章の構成や要旨を的確に理解する能力」と定義されている．

2　実験方法

作文のテーマは 1.「学生時代に力をいれたこと」と，2.「あなたにとって仕事とは」の 2 種類である．

1. についてはある会社の 1999 年度新卒の採用選考において，(a) 会社説明会時に 20 分の時間制約で課した作文 100 編と，(b) インターネットウェブ上で時間制約を設けずに行った作文 590 編とがある．いずれにおいても明示的な字数の制限はないが，(a) においては A4 一枚という物理的な制約が，またウェブ上では予め用意されているテキスト入力領域の大きさが暗黙的に字数を制限していると考えられる．結果として (a) の文字数の平均は約 540 字であり，(b) のそれは約 670 字であった．(a) においては 1 つの作文を 4 名の評定者が，作文から読みとった人物が大学 4 年生としてどの程度の水準にあるかを 5 段階で総合的に評価した．事前に採点者には約 2 時間の訓練を実施した．(b) における採点の観点は達成動機の高さ，文章力，内容の魅力などの観点から総合的に評価した．これら観点についてのスコアの与え方に採点者の主観がなるべく反映しないように基準を定め，採点者には約 1 時間の訓練を実施した．採点者は全部で 8 名であるが，1 つの作文には 4 名ずつ評価を行った．

2. については別の会社の最近の新卒の採用選考において，インターネットウェブ上で時間制約を設けず作文 480 編を収集した．専門家による評価では，評価の観点を文章力におき，それを更に 3 つの観点，すなわち適切な表現をする力 (的確な語彙を使っているか/ 効果的な言い回しを用いて豊かな表現をしているか/ 文法や言葉使いに誤りがないか/ 誤字はないか)，文章を組み立てる力 (構成のバランスがよいか/ 文章の論理展開に一貫性があるか)，説得性 (テーマにふさわしい話題を取り上げているか/ 説得力のある事例を挙げるなど，訴求力を高める工夫をしているか) に細分化した．採点者は全部で 5 人いるが，1 つの作文については 3 人が独立で採点した．字数の平均は 630 字であった．

3 結果

2 種類の作文に対して，Jess および専門家の平均，および GAT の言語的理解力 (GAT(V) と略記) の 3 者間の相関係数を表 1 に示す．また参考として表 1 欄外に専門家の評定値間の級内相関係数を示す．

表 1: Jess/専門家の平均/GAT(V) との相関

	1.(a)		1.(b)		2.	
	Jess	専門家の平均	Jess	専門家の平均	Jess	専門家の平均
専門家の平均	0.46		0.43		0.57	
GAT(V)	0.10	0.28	0.06	0.11	0.08	0.13

(参考) 専門家の評定値間の級内相関係数	0.59	0.45	0.48

これを見るに，小論文の作文能力を評価していると考えられる 2. については，Jess と専門家の平均との相関は 0.57 と小さくないことがわかる．この値 (0.57) は，単純な比較は難しいものの，専門家の評定値間の級内相関係数 (0.48) よりも大きく，Jess は作文の良さの評価において専門家の一人に代替しうることが可能であると言ってよいと考えられる．しかしその一方で，1.「学生時代に力を入れたこと」のような，書かれている内容そのものを人間がある価値観を持って評価するような作文の評価については，(当然のことではあるが) 専門家には及ばず，Jess による利用の限界も確認された．そもそも 1. における採点基準そのものが，Jess の採点基準とは全く異なっていることは弁明してよいであろう．企業における採用選考においては，このような経歴評価，はたまた人物評価を作文に求めることが多く，そのような目的で Jess は使うべきではないことが改めて確認された．また，専門家，あるいは Jess のいずれにおいても，GAT(V) との相関はほとんどなく，作文能力と言語的理解力とは異なった別の能力であることは興味深い事実であろう．

参 考 文 献

石岡 恒憲・亀田 雅之 (2002): コンピュータによる日本語小論文の自動採点システム, 信学技報 Vol.102, No.491, TL2002-40 (2002-12), 43–48. Available online: http://www.rd.dnc.ac.jp/~tunenori/doc/jerater_ieice.{ps,pdf}

二村 英幸・村井 智恵子 (1999): 採用選考における作文評価, 経営行動科学学会第 2 回年次大会発表論文集, 56–62.