

評価の「ゆらぎ」を問い直す：

評価観・評価プロセスを探る研究

宇佐美 洋（国立国語研究所）

smudr@kokken.go.jp

1. 「評価基準グループ」における評価研究

一般に「評価」というと、「試験・テスト」、あるいは学校における「成績」のようなものを思い浮かべる人が多いかもしれない。「試験・テスト」や「成績」とは、被評価者の行動や能力が、教師など権力を持つ専門家によって値踏みされ、ランク付けされる行為である。この種の評価においては、被評価者が不当な不利益をこうむることがないように、可能な限り公正で、一貫性のある評価結果を出すことが重視される。つまり、「妥当性・信頼性」の保証された評価、というものが、「よりよい評価」、「あるべき評価」ととらえられることになるのである。

しかし、われわれ国立国語研究所日本語教育基盤情報センター 評価基準グループでは、「よりよい評価」、「あるべき評価」というものを、もう少し広い意味でとらえたい。われわれの考える「よりよい評価」、「あるべき評価」とは、

生活の中で、よりよい問題解決に結びつきうる評価

である。一般の日本人と、日本語を母語としない人々とが日本語を使ってともに生活する場面において、「よりよい問題解決に結びつきうる評価」が広くおこなわれるようになるには何が必要なのか。そういうことについて、われわれは考えていきたい。

これが具体的にどういうことなのかということ、本論では述べていくこととする。

2. 責任を伴わない評価、責任を伴う評価

小出(2005)によれば、「評価」とは以下のように定義づけられる。

評価主体が、何らかの目的のもとに、評価対象に関する情報を収集し、何らかの基準に従ってその情報を解釈し、価値判断をすること（小出 2005:777）

「評価」をこのように広くとらえるならば、私たちは日々、絶えず「評価」という行為をおこないつつ生きている、ということになるだろう。例えば、「あの人はすばらしい人格者だ」「この味噌汁はおいしいが、ちょっと味が濃い」云々。

そして時には、こうした価値判断としての評価をもとにして次の行動につなげる、という行為

がおこなわれることがある。例えば、「あの人は人格者だから、お金を貸しても大丈夫」「味噌汁の味が濃いから、だし汁を足そう」のように。このように評価には、「心の中で考えて終わり」の評価と、「考えた結果を次の行動につなげる」ための評価があることになる。

鹿毛(2000)は、前者の評価を「価値判断としての評価」、後者の評価を「問題解決のための評価」と呼んでいる。「価値判断としての評価」は、対象に注目し理解する「把握」の段階と、自らの持つ判断基準に即して対象を価値付ける「判断」の段階で構成される。そして鹿毛(2000)は、「同じ対象に対しても把握する視点や判断する基準には個人差があるため、価値判断の結果は人によってさまざまである」と、評価が本質的にゆらぎを含むものであることを認めている。

一方で「問題解決のための評価」とは、「価値判断としての評価」によって得られた情報を次の行動に生かすという、「活用」の段階をもつ評価である。鹿毛(2000)の挙げている例を用いるなら「問題解決のための評価」とは、風景画を描いていてイメージどおりの空を表現するため、試行錯誤的に絵の具を混ぜ合わせているような過程である。自分が作り出した色を見て（対象を把握し）、自分のイメージどおりであるかと考え（基準に即して価値判断し）、もし色が濃いと判断されれば白を混ぜてみる、というように次の行動に生かしていく（活用する）。そして活用の結果は再び把握、判断、活用、というサイクルに乗せられ、問題が解決するまで（納得の行く空色が得られるまで）このサイクルは循環する。

単なる価値判断ならだれにでもできる。そしてどのような評価をしようとも、それは基本的に個人の自由であり、それぞれ尊重されるべきものである。しかしその評価の結果を次の行動に結び付ける場合は、どんな評価をしても自分の勝手、といって済ませるわけにはいかない。評価に基づく「次の行動」は、好ましい問題解決につながる場合も、問題解決にはつながらない（むしろ状況を悪化させるような）場合もあるからである。つまり、「価値判断としての評価」については優劣を問うことはできないが、「問題解決のための評価」は、得られた問題解決が好ましいものであるかそうでないか、という意味で優劣を問われ得る（つまり評価の結果が評価される）ということができるのである。

そしてまた、評価によって得られた問題解決の責任は、ほかでもなく評価者自身が負わなければならない。この意味で、「価値判断としての評価」とは「責任を伴わない評価」、「問題解決のための評価」とは「責任を伴う評価」とであると言い換えることができるだろう。

私たちが社会生活の中でおこなう評価とは、単に個人的な価値判断をおこなってそれで終わり、ということは少なく、その判断の結果は多くの場合、何らかの「次の行動」につながっているのではないだろうか。だとすれば、社会生活の中でおこなわれる評価とは、多く「責任を伴う行為」といえる。

3. 生活場面における評価——「非エキスパート」による評価

ここで話を、日本語非母語話者の言語運用の評価、というところに当てはめて考えてみる。

「評価」というものを、小出(2005)の定義のように広くとらえるならば、私たちは日常生活の中で常に、他者の言語運用に対し「評価」という行為をおこなないながら生きていることになる。

そのことは裏を返せば、私たち自身の言語運用もまた、常に周囲の人々からの評価を受けつつ生活している、ということでもある。

ここで「私たち」というのは、いわゆる「日本人」だけではない。近年増加している、日本国内で日本語を使って暮らしている日本語非母語話者（以下、便宜的に「外国人」と呼ぶ）も、常に周囲の人々——例えば、職場の雇用主やともに働く同僚、あるいは学習者と同じ地域で同様に生活者として暮らす地域住民など——からの評価にさらされながら生きている。特に、日本語の学習がまだ十分でない人々については、彼ら・彼女らが話したり書いたりする日本語について、あるいは言語運用に伴うさまざまな言語行動について、周囲の人々から意識的・無意識的な評価を受けることは多いだろう。

こうした「生活場面における言語運用の評価」の結果は、評価者により、場面により、おそらく大きな「ゆらぎ」をもつものであろう。なぜなら個人によって、言語運用を評価する際の基本的な態度は大きく異なっている、ということが考えられるからだ。

自分にとってはまったく違和感がない言語表現に対し、別の人が極めて強い拒否反応を示して驚いた、という経験は、おそらくだれにでもあるのではないだろうか。逆に、周囲の人はなんとも思っていないらしい言語表現が、自分にとっては気になってしょうがない、ということもあるだろう。つまり、どういう言語運用を望ましく感じるか、という基本的な態度——これを「**評価観**」と呼ぼう——には、人によってかなり大きなばらつきがあるのだ。そして人によってはその評価観の中に、極めて個人的、かつ恣意的な観点が含まれていることも考えられる。

また生活場面における評価をおこなう人々は、言語・教育のエキスパートではなく、妥当性・信頼性のある評価をおこなうための訓練は、おそらく受けてはいないものと考えられる。こうした評価においては、そもそもいま、何のために評価をおこなうのか、という「評価の目的」自体、評価者に明確に意識されているとは限らない。また評価対象に対する情報の収集方法も、必ずしも適切でないかもしれない（例えば、対象のほんの一面しか考慮せずに評価結果を出してしまっているかもしれない）。この結果得られる評価とは、決して妥当性・信頼性を備えたものとはいえないだろう。

しかしながら、生活者としての日本在住外国人が、その言語運用について日常的に受ける評価とはまさにそのようなものなのであり、訓練された評価者による妥当性・信頼性の高い評価を受ける機会は非常に少ないといってよい。彼ら・彼女らが、周囲の人々と円滑な人間関係を築いていくためには、**現実に自分の周囲にいる人々によって、自分自身の言語運用がどのように評価されているのか（されてしまうのか）**ということ認識するところからはじめなければならない。

ここで、「（教師などではない）一般の日本人によって、外国人の言語運用がどのように評価されているか」という問題がクローズアップされてくることになる。こうした問題意識に基づく研究として、小林(2004)に掲載された一連の研究などがある。小林は、「一般日本語話者による、外国人の言語運用に対する評価」（以下、便宜的に「日本人評価」と呼ぶ）に関する研究を、以下のように意義付けている。

コミュニケーションに関わる要素は多岐にわたるので、優先的にシラバスに盛り込むべき項目とそうでない項目を選別しなければならない。その際、日本人母語話者が学習者の日本語に接したときに、どのような点に注目しどのような評価を下すかを知ること（＝日本人評価研究）は、シラバスへの取り込みや教える際の優先準備を決める上で意義がある（小林 2000:149）。

前述の通り、従来「評価」という行為は、基本的には「学習者をランク付けし、値踏みするための手段」としてとらえられてきた。このため、「妥当性・信頼性」の保証された評価をおこなうためにはどうすればいいか、ということが、評価研究の重要な課題であったのである。

一方で小林らの研究においては、評価を「より効果的なシラバスを作成する」という有用な目的のために応用されるべきもの、としてとらえられている。こうした問題意識においては、「妥当性・信頼性」という概念は必ずしも重視されないであろう。「あるべき評価」とは何か、ということを追及するのではなく、実際のところ一般の日本人によってどのような評価がおこなわれているのか、という「評価の実態」を明らかにし、それを有用な目的のために活用していく、ということを目指しているといえる。こうした見方は、「評価」をめぐる新しい動きのひとつとして注目すべきものであろう。

4. 「日本人評価」について論じる際の問題点

4.1 「平均値」だけでなく、「ばらつき」の全体像に着目を

しかしながら、こうした「日本人評価」を学習や教育に活用していくためには、いくつかの点で議論が必要であると考えられる。

第1の論点は、「一般日本人の評価」というものが、どの程度一般化できるものなのか、というものである。そもそも、「一般日本人」というものをどう定義し、どうサンプリングするのか、ということがまず問題となる。仮にサンプリングができたとしても、実際のところ、外国人の日本語運用に対する一般日本人の評価の観点はあまりに多様であり、また個々の要因が複雑に絡み合っていることが予想される。同一の言語運用に対し、互いに矛盾しあう評価観点が使用されている可能性もある。そのような中で、「一般の日本人の評価には～という傾向がある」ということを平均的に導き出せたとしても、その平均的な傾向から外れる評価者は必ず存在するだろう。「一般日本人の評価」を平均的にとらえていくことは、汎用的なシラバスを作成する際には一定の意味を持ちうるが、その扱いには十分注意しなければ、ステレオタイプの誤解を招くことにもなってしまいうだろう。

そしてある評価者の評価結果が、平均的な傾向から外れていたからといって、それを好ましくない評価として排除してしまうことは当然ながらできない。なぜならば、評価結果のばらつきは多くの場合、評価者の価値観のばらつきに基づいており、また、異なる価値観の間で優劣を論じることは基本的には不可能と考えられるからである。それは、関東風の色の濃いそばつゆを好む価値観と好まない価値観との間で優劣を論じることができないのと同じである。であれば、「平

均的な評価結果」だけでなく、「平均から外れる評価結果」もまた、同様の価値を持つものとして扱われなければならないといえるだろう。

日本人評価を研究の対象とする際には、平均的に「一般日本人の評価傾向」をとらえるだけでは不十分であり、また危険でもある。われわれはまず、**ある言語運用に対する一般日本人の評価というものが、現実問題としてどのようにばらつきうるものなのか、また評価に用いられる観点がどのように多様であるか、その全体像を虚心坦懐にとらえる**、というところからはじめる必要があるのではないだろうか。

4.2 評価の結果だけでなく、そのプロセスにも着目を

前節で、評価のばらつきの全体像をとらえることの重要性について言及した。しかし、単に評価結果にこれだけのばらつきが見られた、ということを示すだけでは意味が薄い。次の段階としては、なぜそのようなばらつきが生じるのか、という問いに具体的に答えることができないならぬ。

そのための方策として考えられることは、**評価の結果だけでなく、評価に到るまでのプロセスにも着目する**、ということである。これが第2の論点となる。

ある言語運用に対し、異なる評価者が、結果としては同じような評価をくださったとする。しかし、その評価にいたるまでのプロセス（評価に使用する観点の種類や、観点の使用手順）は、人によってかなり大きく異なっているかもしれない。例えばある言語運用に対し、一人目の評価者は、Aという観点に着目して高い評価を下す一方で、二人目の評価者は、Aという観点は考慮せず、Bというまったく別の観点を使用して、同様に高い評価を与える、ということがありうるだろう。こうした場合、評価の結果がたまたま似ていたといっても、この2名の評価者の評価態度はかなり異なるものだといわざるを得ない。評価のばらつきを論じる、ということは、評価結果のばらつきを論じるだけでは不十分であり、評価に到るプロセスのばらつきをも論じる、ということではなければならない。

4.3 「日本人評価」に対するクリティカルな問い直しを

第3の論点は、ある意味で極めて根源的なものであるが、「**一般日本人**」が「**いい**」と**評価する言語運用は、ほんとうに「いい」と言えるのか**、ということである。

教師などの言語エキスパートによる評価の観点・プロセスと、一般日本人による評価の観点・プロセスとは異なっている可能性がある。生活者としての外国人が言語使用の場において実際に接するのは、言語エキスパートではない一般の日本人なのであるから、「教師ではなく、一般日本語母語話者に現実が高く評価される言語運用」をこそ学習の目標にすべきだ、という主張は非常によく理解できるものである。

反面、一般日本人の評価には、多くの場合「教育的な配慮」は含まれていないだろう。また前述のように、「データ収集の方法が不適切」「評価基準が個人的・恣意的」「そもそも評価の目的が認識されていない」というような問題点があることも予想される。そのような問題含みの評価を

そのまま学習や教育の指針としてしまうことは、果たしてほんとうに妥当なのだろうか。そのような評価を、日本人と外国人がともに生きていくという場において、「よりよい問題解決」につながることは、果たして可能なのだろうか。

ここに来て議論は循環することになる。教師など言語エキスパートによる評価は、一般日本人の評価とは乖離している可能性がある。そういう理由のために、一般日本人の評価に着目しようとしたところが、今度は、一般日本人の評価には教育上必ずしも適切でない点もある、ということが指摘される。そうすると、学習や教育の指針を得るためにはやはり言語エキスパートの評価に拠るべきではないか、というところに戻ってしまう。こうした不毛な循環から抜け出すには、どのようにすればよいのだろうか。

このことについて、われわれの考えは以下の通りである：

われわれは、生活場面における評価、ということを考えてとき、まず立脚点とすべきなのは「教育や言語のエキスパート『ではない』、一般日本人による評価」である、という立場に立つ。「言語エキスパートによる評価」を基本にすべき、という考え方はしない。

しかしながらわれわれは同時に、**一般日本人の評価に対し、クリティカルな問い直しをおこなうべきだ**、とも考える。

社会の中で日常生活を営んでいる者はだれでも、好むと好まざるに関わらず、また意識的にせよ無意識的にせよ、評価という行為をおこないながら生きている。そしてその評価の結果に責任を負わねばならない場合もある。であれば、自分自身が日常生活の中でおこなっているさまざまな評価について、「**自分の下したこの評価は果たして適切なのか**」「**この評価は、果たしてよりよい問題解決につながりうるものなのか**」という再考をおこなうことは、社会人として極めて重要なことではないか。

そしてこの「問い直し」は、研究者だけがおこなっても仕方がない。その問い直しとは、生活の中で外国人と直接接する一般日本人ひとりひとりが、自らの力でおこなうのでなければならない。われわれ研究者がすべきことは、だから問い直しそのものではない。また研究者の立場から、「あるべき評価の方法論」を作成し、それをすでにできあがったものとして社会に提示することでもない。われわれ研究者としては、**ひとりひとりの日本人が自らの評価観を問い直すことを支援できるような枠組みを提案したい**、と考える。そしてその枠組みを使って、言語運用の評価が**より高い意識に基づいておこなわれるよう、社会的な促しをおこないたい**、と考える。そのことは、言語・文化を異にする人々が日本社会の中で共生し、お互いにとってよりよい問題解決をおこなっていくための重要な基盤となるはずである。

5. 「一般日本人評価の問い直し」のためにすべきこと

しかしながら、ひとりひとりの日本人が自らの評価観を問い直すことができるようにする、というのはいかにも遠大な目標であり、すぐにこのような目標が達成できるとは思われない。目標は目標として遠くに掲げた上で、われわれはまずどこから作業を始めるべきか、ということもまた考えなければならない。

そこでわれわれは、4.1 節、4.2 節で述べたことについて調査をおこなうところから始めたい、と考えている。すなわち：

- 1) 日本人評価を平均値的にとらえるのではなく、日本人評価というものがどれほどのばらつきをもちうるものか、その全体像をとらえる
- 2) 日本人評価の結果だけでなく、評価のプロセスをとらえる

というふたつのことがらである。

1) の作業によってわれわれは、「自分以外の人々がどのような評価をおこなっているのか」ということを知るための手がかりが得られるだろう。自分自身の評価観について問い直すためには、まず自分以外の人々がどのような評価をおこなっているのか、ということできるだけ広く知り、それと自分自身の評価がどのように異なっているかをとらえることが効果的であると思われる。

そして前述のように、評価の結果について考察するだけでは不足である。「評価観について問い直す」ためには、ひとりひとりの評価者が評価対象を把握し、価値判断を下すまでの間、いったいどのような思考過程を経ているのか、ということをはっきりと示していく必要がある。そこで2) の「評価のプロセスをとらえる」という作業が必要となってくる。この作業においては、例えば、

- ・評価をおこなうときどのような観点が用いられているのか
- ・その観点はどのようにして選択されるのか
- ・場合によって評価の観点が変わりうるとしたら、観点の変化はどのような要因によってもたらされるのか
- ・評価という行為をおこなっている途中で、評価の観点に変化する、ということはないか

などのことを明らかにしていきたいと考えている。

それともう一つ重要なこととして、「いきなり話しことばの評価を対象とするのは困難である」ということを、われわれは認めておかざるを得ない。話しことばの評価には、言語として表現された内容の評価だけでなく、口調・声質・イントネーションなどの「パラ言語的な要因」や、その言語表現が発せられたときの話し手の態度、表情、場合によっては服装などの「非言語的な要因」も関わってくる。さらに話しことばとは一瞬で消えてしまうものであり、発話をあとでもう一度聞きなおすということは、少なくとも日常生活における評価ではありえない。このため、聞き手（評価者）の記憶容量の違いも、評価には影響を及ぼすことであろう。

つまり話しことばは書きことばに比べ考慮すべき変数は格段に多く、いきなりその評価研究に踏み込むことは危険である。まずは書きことばを対象に、評価研究の枠組みを打ち立ててから、その後話しことば研究に進む、という手順を踏むことが適当であるとわれわれは考える。

6. 「評価基準グループ」の調査研究

上記のような考え方にに基づき、われわれ「評価基準グループ」では、具体的に以下のような調査研究をおこなっていく予定である。

6.1 外国人の日本語運用データ（作文データ）の収集

まずは、評価の対象となる「外国人の日本語運用データ」を収集するところから始めなければならない。

すでにわれわれは、外国人の日本語作文データベースとして、「日本語学習者による日本語作文と、その母語訳との対訳データベース¹（作文対訳 DB）」を持っている。もちろんこのデータベースに収録されたデータも評価対象データとして活用していく予定ではあるが、しかし現在までのところ、このデータベースに収録されている作文データは主として教室活動の中で書かれたものであり、外国人が日本での日常生活の中で書くような課題に基づいているとはいえない²。

そこでわれわれは、「生活者としての外国人が実際に書きそうな課題³」を設定し、その課題に基づく作文データの収集を始めている。

6.2 外国人の日本語作文データに対する、日本人の評価データの収集

次にこれらの作文データを、可能な限り多様な日本人に読んでもらい、作文データに対する評価データを収集していく。この「評価データ」には、「個々の作文に対する評点」「複数の作文についての順位付けデータ」「添削情報」「個々の作文に対するコメント、感想などのインタビューデータ」などが含まれる。また、「評価プロセス」について調べる、という目的のため、「評価付けの理由に関する事後インタビュー」「発話思考法によるプロトコル⁴」などのデータを採ることも計画している。

6.3 「評価プロセスモデル」の提案

ここまででわれわれは、日本人評価のさまざまなあり方をデータによって明らかにしていくことを目指し、そのための調査手法について考えてきた。しかしながらただ単に評価の「ゆらぎ」の実態を明らかにするだけでは、そのあまりのばらつきに当惑するばかりで、そうした情報を有効に活用していくことは困難、ということも考えられる。

そこでわれわれは、日本人評価がいかにならつくものか、という「評価観の違い」に着目すると同時に、評価をおこなう際のプロセスを「ひとつのモデルとして表現する」、という試みをおこなってみたい。このことは、「個人の評価プロセスは一見すると極めて多種多様に見えるが、評価

¹ <http://www2.kokken.go.jp/eag/>にて公開。

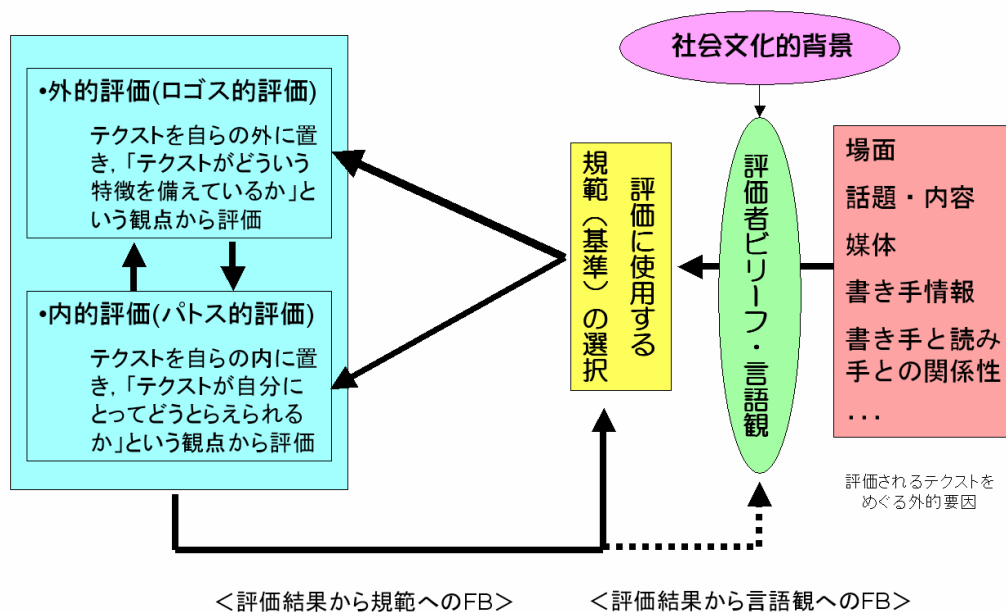
² 例えば、「あなたの国の行事について」「たばこを法律で禁止することについてのあなたの意見」のような課題で書かれている。

³ 例えば、「友人から来たお誘いのメールに断りの返事を出す」「ゴミ出しのことで怒っている隣人にお詫びのメモを書く」のような課題である。

⁴ 評価をおこなうために作文を読む際、考えたことをすべて口に出してもらい、それを録音・文字化したデータ。

の本質的な道筋は共通であり、ただその道筋をたどる順序や、評価の各過程で用いられる変数が異なるだけなのではないか」という考え方に基づいている。

すでにわれわれは、中期計画1年目である06年度、外国人の作文が一般の日本人によって評価される際、どのような観点が、どのような手順で使用されるか、ということ、インタビューを通じて探る、という調査をおこなった(宇佐美2007)。この調査で得られた知見をもとにわれわれは、以下のような「評価プロセスモデル」を試作した。



この評価プロセスは、評価をおこなう際の基本的な手順を示したものであるが、途中で枝分かれをおこなうこと、評価手順が一部でループを起こす(つまり、評価の結果に基づいて評価に用いられる基準が選び直され、その基準によって改めて再評価がおこなわれる)可能性を想定していることに特徴がある。こうした特徴により、個人がたどるさまざまな評価プロセスを説明できるような工夫がなされている(ただこのモデルはあくまでも現段階における試案であり、今後6.2の調査が進むにつれて改良が加えられていくものであることをお断りしておきたい)。

紙幅の関係で、本論においてはこのモデルについて詳細を述べることはできないが、公開研究発表会の場においてはやや詳細な解説を加えることとする。そして研究の最終段階には、こうしたモデルも使用しながら、「日常生活における外国人の言語運用の評価」をテーマとしたワークショップの開催につなげたい、と考える。このワークショップは、日常生活において外国人の日本語学習支援に関わっている人々を当面の対象とするもので、「自分自身がどのような評価観をもち、どのようなプロセスをたどって評価をおこなっているのか」ということを自覚し、他の人の評価観との比較検討をおこなうことを通じて、「日常生活における評価」というものを根源的に問い直すことを目的とするものである。このワークショップは、4.3節で述べた、「言語運用の評価がより高い意識に基づいておこなわれることを目指す、社会的な促し」とは何か、ということに対する、われわれなりの回答となるはずである。

7. おわりに

ここまででわれわれは、主として、外国人の日本語運用に対する日本人評価の問い直しについて論じてきた。つまり、評価するのは日本人、評価されるのは外国人、という枠組みで話を進めてきたことになる。

しかしながら現実の社会生活においては、評価者と非評価者の組み合わせとしては上記のものが唯一ではない。日本人が日本人の言語運用を評価する、さらには、外国人が日本人の言語運用を評価する、ということも考えられる。

果たして日本人は、自分と同じ日本人の言語運用を適切に評価できているといえるだろうか。四字熟語をたくさん知っていること、漢字を正確に書けること、敬語を適切に使えること、これらはそれぞれに重要なことであるにせよ、日本人の言語運用を評価する観点として、もっと違うものもあっていいのではないか。そのような問い直しも、もっと広くなされていように思う。

また、日本人の言語運用が外国人によって評価される、という可能性についても、もっと積極的に考えていいのではないか。円滑なコミュニケーションの成立には、基本的に情報の発信者、受信者双方に責任がある。受信者が、送信者の意図を十分に把握することができなかった場合、それは受信者の能力にだけ問題があるのではなく、送信者が受信者側の事情を考えずに情報を発信した、ということこそが最大の問題である可能性がある。そうすると、日本人の日本語運用が外国人によって、「情報の受け手の事情に十分配慮しているか」という観点で評価されることも十分考えられることであろう。「あなたの日本語はそれでいいのか」という問い直しだが、外国人の側からなされる可能性もある、ということである。

外国人の言語運用の評価、というものについて深く考えれば考えるほど、結局は日本人側の言語観や言語能力が問われているのだ、ということに気づく。「外国人の言語運用の評価」について論じることは、日本語教育だけの問題ではない。それは日本人自身の言語問題なのだ。

参考文献

- 宇佐美洋(2007)「学習者作文の評価における観点の多様性—一般日本語母語話者の場合—」『日本語学習者の書き言葉に対する対照言語学的・文章論的研究』, 平成 17-18 年度科学研究費補助金 基盤研究(C)(2) 研究成果報告書 (課題番号 17520360) , 29-67
- 小出慶一(2005)「評価とは」『新版 日本語教育事典』(日本語教育学会編) , 777-779, 大修館書店
- 小林ミナ(2000)「「何を」教えるかの再吟味—日本人評価研究の意義と限界—」『北海道大学留学生センター紀要』, 4, 149-159 (小林 2004 にも再掲)
- 小林ミナ(2004)『日本人は何に注目して外国人の日本語運用を評価するか』平成 12 年度～平成 15 年度科学研究費補助金 基盤研究(B)(2) 研究成果報告書 (課題番号 12480058)
- 鹿毛雅治(2000)「評価と測定」『教育評価重要用語 300 の基礎知識』(森俊昭・秋田貴代美編) , 15, 明治図書