

「ロシア極東・東シベリア日本語弁論大会」 における審査基準作成

成田高宏・猪狩英美・森本由佳子・坂本裕子

〔キーワード〕 スピーチ、パフォーマンス評価、マルチプルトレイト・スコアリング、評価者
間信頼性、個人差

〔要 旨〕

本稿では、「ロシア極東・東シベリア日本語弁論大会」の新審査基準作成過程およびその試用結果について報告する。ロシア極東・東シベリア地域では、毎年、国際交流基金の助成を受け、ウラジオストク、ユジノサハリンスク、ハバロフスクの3地域持ち回りで、日本語を学ぶ大学生による日本語弁論大会が開催されている。しかし、開催地により審査基準が異なり、その基準も曖昧であったため、共通した審査基準の作成が望まれた。作成に際しては、信頼性の高い評価が行える基準を作成することをねらいとしたが、副次的に学習・指導の一助となるよう配慮した。試用結果から、新基準は適切な手続きをふまえて使用すれば良好に機能することが確認されたが、質的な検証も含めた信頼性のさらなる検証が今後の課題である。

1. 「ロシア極東・東シベリア日本語弁論大会」について

1.1 大会の概要

ロシア極東・東シベリア地域では、毎年10月上旬ごろに「ロシア極東・東シベリア日本語弁論大会」(以下、極東大会)が実施され、2009年度大会で14回を数える。国際交流基金(以下、基金)の海外日本語弁論大会助成を受け、総領事館の所在地であるウラジオストク、ユジノサハリンスク、ハバロフスクの3地域が開催地域となり、毎年3地域の持ち回りで開催している。

出場学生は、サハリン州、沿海地方、ハバロフスク地方、サハ共和国、イルクーツク州、ブリヤート共和国、チタ州、カムチャッカ州、マガダン州から、各地方大会を勝ち抜いた学生もしくは大学内で推薦された学生が、毎年14~15名出場する。極東大会の上位入賞者に対しては、10月末にロシア全土およびCIS地域から学生が集まりモスクワで開催される「CIS学生日本語弁論大会」(以下、CIS大会)への出場権が与えられる。極東大会は、各地方大会と全国大会(CIS大会)の中間に位置づけられる大会である。

審査員は、日本語教育関係者(出場者の中に指導学生がいない教師)、総領事館員等、7~

8名で構成されているが、メンバーは毎年入れ替わることが多い。また、ロシア人が審査員として加わるかどうかは地域によって、または年度によって異なっている。基金派遣の日本語専門家(旧日本語教育ジュニア専門家)は質問員を担当している。極東大会では、5分間のスピーチのあと、質疑応答が行われる。質問は、日本語専門家が事前に原稿を読み、作成したものである。各日本語専門家は、自分の指導学生の質問作成には関わらない。

1.2 問題の所在

極東大会では、2007年度の大会まで共通した審査基準が設けられていなかった。しかし、3地域の日本語専門家の間で、以下のような問題点の認識から、まず大会の意義を再検討し、その意義を反映するような審査基準作成の必要性が確認された。

(1) 3地域の持ち回り開催

1.1で述べたように、極東大会は3地域の持ち回りで開催されている。「極東大会」という名前が掲げられているものの、これまで開催担当地域のみが準備・運営に携わり、3地域間での話し合いは行われていなかった。そのため、地域や実施年度によって審査方法が異なり、極東大会としての方針の統一は成されていなかった。極東大会は広くロシア極東・東シベリア地域を対象にした大きな大会であるため、その審査結果は、出場学生のその後の進路や人生をも左右する影響力をもつ可能性があると言える。同じ「極東大会」という名前でありながら、その審査基準が統一されたものでないということは、審査結果について明確な説明ができなくなると筆者らは考えた。

(2) 極東大会の意義

これまで、極東大会の実施要項等に大会の趣旨が明記されていなかったために、実施意義が不透明であり、人によって、また地域によって、弁論大会のとらえ方はさまざまであった。

第一に、大会が日本語の教育とイベントの、どちらの側面を重視すべきであるかが明確ではなかった。そのため、大会へ向けての取り組み方も指導者によってまちまちであった。

第二に、スピーチに求められるものは、日本語そのものの能力と内容とに大きく分けられるのだが、後者に関し、「内容のいいスピーチ」の概念が人によって異なることが問題であった。そのため、弁論大会で上位になったスピーチに対しては、賛否両論の意見が出てくる。万人に共通する「内容のいいスピーチ」というものは存在しないかもしれないが、教育を目的に掲げて審査をする以上、「内容のいいスピーチ」がどのようなスピーチなのかは、大会運営側が示さなければならない。万人に共通のものがないとするならば、少なくとも極東大会ではどのような内容のスピーチをいいスピーチとするかを規定しておけばよいのではないかと考えた。

(3) 審査方法

極東大会の運営は開催地域にゆだねられていたため、大会ごとに異なる基準が用いられていた。それらの基準は開催地域の大会運営者で議論され、従来の大会の審査方法をふまえながら検討を重ねたものではあったが、評価方法として妥当性・信頼性を備えたものであったのか、検証はされていなかった。

具体的には、目安として審査項目が掲げられてはいるものの、最終的には審査員個人の裁量によって自分なりの順位が決められればよいとする順位点方式を採用していた⁽⁴⁾。各審査項目の点数を合計して順位を決定するのか、あるいは各審査項目の重みづけを調整するのも個々の判断に任されていたため、結果から評価の妥当性・信頼性を検証することが困難であった。

また、各点数の条件が示されていないため、5点満点中、「1や2はつけにくい」という審査員心理が働き、審査員がつける点数は3や4に集中してほとんど同じ点になり、順位がつけにくくなっていた。その結果、各審査員が苦勞して、3+、4-のように+や-をつけたり、3.5、4.5というふうには0.5ずつの段階を設けたりしていた。そうすれば、審査員にとっては、自分の中での点数の差はつけやすくなり、順位が決めやすくなる。しかしながら、その得点は、他の審査員と比較対照できるものではなく、順位をつけるための参考程度のものでしかなくなっていた。

(4) 審査員

地方大会および極東大会の審査員は、日本語教師を含め、評価の専門家であることはまれで、採点技術が高いとは必ずしも言えない。しかしそのような審査員であっても、信頼性の高い評価が可能となるようにすることが、審査基準の目指すところであり、存在意義であると考えた。

(5) 地方大会の位置づけ

ロシア極東・東シベリア地域内の地方大会は極東大会の予選会を兼ねていることになるが、極東大会の方針が定まらなければ、地方大会においても審査方法が決められない。各地方で決めたとしても、それが極東大会の審査基準と異なれば、出場学生にとって不利になることもありうる。このような状況をふまえ、地方大会と極東大会は同じ基準で実施されるのが望ましいと考えた。また、各地域の日本語教師会からも、極東大会の統一基準があるのであれば、それを地方大会でも活用したいという声が聞かれ、日本語専門家が派遣されていない地域からも、「統一した基準がほしい」という要望があった。

2. 第二言語としての日本語スピーチの評価基準に関する先行研究

笠原・小山(2003)は、情報伝達のための言語表現能力養成において、学習者自身にも達成目標が認識できるように、また、教師にとっても指導を単に問題点の指摘や訂正で終わらせないために、評価項目や評価基準を学習者と共有することの必要性を述べている。「トピック選

択」「情報選択」「情報整理」「本論構成」「序論・結論構成」「表現」の6つの評価項目のうち、「トピック選択」と「本論構成」はフィードバックと繰り返しにより改善されやすく、「情報選択」「情報整理」については目標達成が困難であることが報告されている。

古賀・中島 (2005) では、教師、教師以外の日本人、学習者自身のそれぞれが、スピーチ発表において何を評価していたのかを質的に分析している。設定された評価項目は、「内容」「聞き取りやすさ」「文法の正確さ、語彙・表現」「発表の仕方」「ハンドアウトと語彙表」「質疑応答」の6つである。分析結果として、教師は「聞き取りやすさ」の点が甘くなり、文法等を重視する傾向にあること、教師以外の日本人は学習者の発音が不明瞭な場合、他の項目の評価に影響し、また社会文化的な面を含めた総合的な評価をすること、学習者は「質疑応答」での自己評価が低いことなどが報告されている。また、学習者の気づきや、教師以外の日本人の評価の視点などを学習者と教師が共有し、次の活動へつなげていくことの必要性を指摘している。

一二三 (2007) はスピーチコンテストで発表されたスピーチを一つ取り上げ、日本語母語話者 (NS) と非母語話者 (NNS) にそのビデオを見せて、28の評価項目について5段階評価をさせている。そこから、スピーチの評価基準がどのような因子から構成されているか、また NS と NNS とでどのように評価基準が異なるのかを調査している。因子分析の結果、「態度」「テーマ・言語」「内容」「発音」の4因子が得られた。評価については、「態度」「テーマ・言語」「内容」および「総合評価」(感動したか、おもしろかったか等5項目)において、NNS よりも NS のほうが高めに評価することが示された。「発音」については両者に有意差が見られなかった。また、NS、NNS とも「態度」→「発音」→「テーマ・言語」→「内容」のように評価は段階的に推移し、最後に「総合評価」にいたることが検証された。以上の結果から、スピーチ指導において「態度」などの外面的評価と「内容」などの「内面的評価」の2側面を学習者に意識させることの必要性が指摘されている。

いずれの研究においても、学習者の能力向上のために評価基準を明示することの重要性が示唆されている。これは、海外における弁論大会においても取り入れるべき視点だろう。

3. ロシアおよび CIS 地域における地方日本語弁論大会の審査基準

ここでは、ロシアおよび CIS 地域の各地方で開催されている日本語弁論大会の審査基準について紹介する。例年10月末にモスクワで全国大会とも言える CIS 大会が開催され、各地域もしくは管轄公館ごとに1～5名の出場枠が与えられる。各地方の日本語弁論大会は、CIS 大会の出場学生選考を兼ねているが、審査方法、審査基準および審査員の選定は開催地によってさまざまである。また、CIS 大会も独自の選考基準となっているが、詳細は公表されていない。

表1に、主に2007年度の各地方日本語弁論大会の審査基準を記す。それぞれ審査項目と配点、具体的に審査するポイントを記した基準が記載されている。

「ロシア極東・東シベリア日本語弁論大会」における審査基準作成

〔表1 CIS 地域における主な日本語弁論大会の審査基準〕

ウクライナ07 (計8項目40点)		
日本語能力	①文法・語彙の正確さと適切さ	5点
	②発音・イントネーションの正確さと自然さ	5点
	③話し方のなめらかさ 問の取り方 意味のまとまり	5点
発表内容	①テーマの深さ 内容の面白さ	5点
	②主張の明確さ 説得力	5点
伝達力	①声の大きさ、スピード、アピール力、視線など	5点
質疑応答	①質疑の日本語力	5点
	②質疑の的確さ	5点
ベラルーシ07 (計5項目60点)		
内容	テーマに独創性がある／一貫性がある／論の組み立てがよい／論旨が明確／主張が深い	20点
言語	語彙や文法が正確	10点
	自然な発音・アクセント・イントネーション	10点
表現	弁論技術 (説得力のある話し方・表情)	10点
	質問をよく理解している、正確な日本語で答える	10点
ノボシビルスク07 (計4項目40点)		
内容	オリジナリティ、面白さ、内容の深さ等があるか。聴衆の興味や関心をひきつけるものであるか。構成力 (主張の明確さ、全体の組み立て、内容の一貫性) があるか。	10点
表現力(プレゼンテーション力)	スピーチにふさわしい発表態度であるか。声の大きさ、速さ、ポーズ、表情、視線、姿勢が適切であるか。説得力があるか (暗記したものを棒読みするのではなく聴衆に訴えるような工夫した話し方ができているか)。	10点
日本語運用力および発音	適切な文法および語彙を使っているか。聞き取りやすい発音・アクセント・イントネーションであるか。	10点
質疑応答	質問を正しく理解し、的確に答えているか。自然な日本語の表現で答えているか。(特に中、上級レベル以上において) 単純な Yes.No ではなく、内容のある回答ができているか。	10点
ハバロフスク07 (計5項目25点)		
音声	聞き取り易い発音、イントネーションか。速すぎず遅すぎず、わかり易い速さで話しているか。問の取り方は適切か。	5点
日本語運用力	適切な文法や語彙を使っているか。自然な日本語の表現を使っているか。	5点
内容	弁論大会の発表内容として適切であるか。聴衆に感動を与えたり、考えさせられるテーマか。主張がはっきりしているか、それをわかりやすく伝えているか。説得力があるか。	5点
発表態度	テキストを暗記し、理解した上で話しているか。弁論大会にふさわしい発表態度であるか (声の大きさ、姿勢、視線、顔の表情など)。	5点
質疑応答	質問を正しく理解しているか。質問に対して的確な答えができているか。自然な日本語の表現で答えているか。	5点
ウラジオストク07 (計8項目80点)		
技術審査	日本語表現、語彙及び文法の正確さ	10点
	イントネーション、アクセント等発音の正確さ	10点
	内容	10点
	質問への応答	10点
総合審査	論旨の明快さ	10点
	発表態度及び弁論技術	10点
	内容	10点
	質問への応答	10点
カザフスタン06 (計3項目30点) 内容20点、発音5点、質疑応答5点		
CIS 大会07～09 論旨の明確さ、主張内容の深さ、独創性を評価点とする。(項目・配点は非公開)		

まず、全地域において共通した審査項目として「内容」と「質疑応答」がある。「内容」については、独自性、面白さ、深さ、聴衆に感動を与えるものであるか等の「ストーリーの魅力」と、主旨の明確さ、説得力、全体の構成等の「論理展開」を問うものの二つに分けられる。前者は審査員の背景知識が影響し、主観的判断によるところが大きいと思われるが、後者については審査員による判断の差異は少ないのではないと思われる。両方を盛り込んでいる場合が多いが、前者のみの地域もみられる。また、「質疑応答」については質問の理解度、答えの的確さ、応答での日本語力の3つが主な審査基準としてあげられており、ほぼどの地域にも共通している。

また、発音やイントネーション、声の大きさや視線などの発表態度・プレゼンテーション力も、ほぼ共通して取り上げられている。ただし、審査対象が同じ「発音（イントネーション、アクセント等）」であっても、「発音」もしくは「音声」として独立した項目を設けている地域や、文法や語彙力と同じ項目に分類している地域がある。

地域や年度により変動がみられるのは、スピーチの日本語運用力を問う項目である。カザフスタンでは審査項目に含まれておらず、CIS大会においても日本語力よりも内容重視の傾向がみられる。

次に、各審査項目の配点であるが、①5段階か10段階か、②各項目への配点比重を同一にするか否か、という2点において違いがみられる。一般的に、大学における成績の採点方式は5段階であるため、5段階方式の地域が多いと思われるが、採点に幅ができ点数がつけやすい、もしくは同点になりにくいという考えから10段階方式を採用している地域もある。配点比重に差を設ける場合、「内容」に比重を置く例が多く、カザフスタン、ベラルーシでは「内容」に対して20点の配点としている。

以上、ロシア・CIS地域における地方日本語弁論大会およびCIS大会の審査基準について概観した。審査項目の内容・数・配点のすべてについて、地域により違いがあることが明らかである。また、実際に審査対象となっている項目は同様であるにもかかわらず、審査基準で使用されている用語がさまざまであることも注目すべき点であろう。さらに、どの地域においても、これら審査基準が流動的であることが、各地域への聞き取りからうかがえた。日本語専門家の交代や現地の日本語教育状況、CIS大会の動向などに合わせて、随時変更されている。

弁論大会開催の意義および目的は地域により異なるかもしれないが、CIS大会に出場する代表者選考を兼ねている点においては一致している。当該地域において、CIS大会に出場し上位入賞することは、留学や進学、就職の際にアピールできるポイントの一つとなり、学生のその後の人生に大きな影響を与えかねない。また、大学や現地教師にとっては名誉と威信をかけた大会であり、皆、並々ならぬ意気込みで臨んでくる。このような現地の認識には賛否両論聞かれるが、学生にとって最大かつ唯一の、自身の日本語能力を披露できる場として重要な位置づ

けにあることには違いがない。そのため、多くの参加者が費用と時間をかけ、練習に励み、大会に臨むのである。したがって、彼らのその努力が正当・公平に評価され、同時に望ましい方向へ努力が向けられるためにも、参考にできるような基準が必要であろう。

しかし、前述したように、現状では地域や大会によって審査基準の差異が大きく、参加学生にとっても審査員にとっても、わかりにくく取り組みにくい状況であることは否めない。一地域の大会運営上の問題としてではなく、CIS大会に参加するすべての地域が採用可能な、透明性および信頼性の高い統一基準が定められ、採用されることが望まれる。

本稿で報告する審査基準改定は、表1のうちハバロフスクとウラジオストクを含むロシア極東・東シベリア地域の弁論大会（極東大会）についてのものである。極東大会の審査基準改定については次章で詳述する。

4. 作成過程

4.1 改善目標の設定

3章で述べたような状況をふまえて、極東大会および極東地域の各地方大会に共通の審査基準を作成した（完成した基準の「机上使用版」を資料2として、また、「机上使用版」や本稿中からもれた諸事項をまとめて資料1として添付した）。

ウラジオストク、ユジノサハリンスク、ハバロフスクに派遣された各日本語専門家が、それぞれの提案をメールにより提示し、それについて他の2名が意見を返し、できるだけそれらを反映させながら最終的には評価研究を専門とする1名が判断を下す、という形で作成が進行した。作成にあたっては、第二言語としての英語および日本語のライティング評価研究を主な理論的背景として用いた。スピーチの評価とライティングの評価とは異なる点もあり配慮が必要ではあるが、より蓄積のあるライティング評価研究の成果をスピーチ評価の分野に応用することは、多くの可能性をもつ試みであると思われた。

従来審査基準を使用した2006、2007年度の大会を経験し、その反省をふまえたうえで改善すべきだと意識された点は、以下の3点である。

- ① 「大会の意義を明らかにし、それが学習者の日本語能力向上にあることを明示すること
＝意義の明確化」
- ② 「審査基準に沿って採点が行われれば、たとえ基準を使いこなさきれていない審査員が少数含まれていたとしても、あるべき結果が導かれるようなものにする
＝適切な手続きの選択」
- ③ 「各評価項目の意味を明確にし、審査員の好みなど個人差に影響されるケースを可能な限り排除できるようにすること
＝公正さの維持」

以下、これら3点について詳述する。

4.2 意義の明確化

まず①についてだが、これまでは手続き上、意義について言及されてはいても、それが出場者や指導する教師にまで行き渡っているとは言いがたい状況にあった。しかし、各地方大会に始まり、各領事館の管轄区域を経て CIS 大会にいたるまで良く整備され、浸透している弁論大会は、日本語教育の機会としてさらなる有効利用の可能性に満ちていた。その実現のために、筆者らはまず、出場者および指導者に学習の指針となるべき意義を周知させる必要があると考えた。

重視したのは「学習者の努力によって向上可能な目標を設定すること」である。たとえば、これまでは会場を沸かせたスピーチが上位に入賞しがちだった。もちろん、公正な審査の結果そういったものが上位に来るのであれば問題はないが、初めから会場を沸かせることを出場者に目指させるような方針を、意義として示すべきではないと考えた。なぜなら、会場を沸かせることが努力によって可能なことであるかという点に、疑問符がつくからである。さらに留意した点は、海外の大学（もしくはそれと同等の教育機関）で日本語を学ぶ者に求められる「総合的な力」とは何か、それを分類し言葉で表現するとすれば、どのようなものになるか、ということだった。

以上の点について議論を重ねた結果、「日頃の日本語学習の成果をスピーチという形により表現し、競い合うことによって学習の励みとすること」を開催の意義とし、具体的には、Ⅰ「日本語そのものの力（言語力）」、Ⅱ「大学生に求められる視点や思考の深さ（内容）」、Ⅲ「それらを伝えるための表現力やコミュニケーション力（運用力）」という3要素の総合的な力を競うことを明記した。

4.3 適切な手続きの選択

次に②についてだが、まずこれまでの極東大会の審査基準で採用されていた「多項目をスケールで評価しつつも、それを参考程度にしか扱わず順位点方式で採点する」という方法をやめ、「多項目のスケールでの評価を基本とし、同点だった場合のみ順位点で決める」という形にあらためた。出場者全員に順位点をつけ、最後に全審査員の総和を算出する順位点方式の欠点は、他の審査員からずれた評価を行っている審査員がいた場合、その結果が全体に及ぼす影響が大きいこと、そして出場者数が増えるほどその影響力が増すことである。従来も各出場者につけられた最高順位と最低順位を除く手続きが取り入れられてはいたが、それだけで信頼性を保つことは心もとない。たとえば、出場者15名の大会で、ずれた評価を行う審査員が2名おり、本来上位に来るはずの出場者に、この2名が13位、15位という順位をつけたと仮定すると、13位という点数は残ることとなり、それが致命的な結果を招く可能性がある。

またスケールの段階については、5段階方式の採用を決定した。地方大会で10段階を採用しているところもあり、「100点満点で採点することに慣れている人は10段階のほうが審査しやすい」「5段階では3～5点に集中し出場者間に差がなくなるかもしれない」という意見もあった。しかし、これらについては、特に相対評価のしにくい項目（「内容」や「質疑応答」）の1点から5点までの各段階に、「その点数であるための条件」を付記することで解決できると判断した。今回採用した評価方式は、評価研究の分野においてマルチプルトレイト・スコアリング⁽²⁾と呼ばれるものであり、その信頼性を保つためには、段階の多すぎるスケールは設定すべきでないとされている。また段階数を多くすると、他とずれた評価をする審査員の全体への影響力が増すことも予測され、そのようなデメリットを避けるためにも5段階程度が適切だろうと判断した。なお、各項目の点数の比重を変えることは、信頼性を低めるとして、マルチプルトレイト・スコアリングにおいては否定されている。

4.4 公正さの維持

次に③についてだが、まずこれまで使用されてきた審査基準をもとに、4.2の意義であげた3要素を包含する以下の6項目をたてた。その際、極東大会上位入賞者がCIS大会に進んでも不利にならないよう、CIS大会の審査基準も参考とした。

- I 言語力に関するもの：「発音」「文法・語彙」
- II 内容に関するもの：「聞き手への意識」「内容の深さ・説得力」
- III 運用力に関するもの：「プレゼンテーション力」「質疑応答」

「文法・語彙」は、指導者の介入が真の力の測定を妨げていることを理由に、一部の地方では採りあげられていない項目であるが、審査対象となるのは、あくまでその時のパフォーマンスであり、またどのようなケースにも対応できるよう準備しておくことも審査基準の存在意義の一つであると考え、独立した項目として採用した。

これらのうち相対評価が比較的容易な「発音」「文法・語彙」および「プレゼンテーション力」に関しては、基準を可能な限り簡素化することも考慮し、従来どおり、特に説明のない5段階のスケールによって採点する方式をとった。一方、相対的に評価することが困難な「聞き手への意識」「内容の深さ・説得力」および「質疑応答」に関しては、1点から5点までの各段階に「その点数であるための条件」を付記することで対応しようと考えた。これはマルチプルトレイト・スコアリングにおいて一般的に用いられる手法である。

これらの項目の相対的評価を困難にしている一大要因は、評価結果に評価者の背景が影響し、個人差が反映されてしまうことであろう。たとえば、これまでの審査基準には「主張」（今回作成した新審査基準の「内容の深さ・説得力」にあたる）や「内容」（同じく「聞き手への意識」にあたる）という項目があり、前者には「説得力がある」という説明が、後者には「考え

させられる」という説明が付記されていた。しかし、「説得力がある」、あるいは「考えさせられる」スピーチとは、具体的にはどのようなスピーチを指すのか、という判断は、個々の審査員にゆだねられていた。どのようなものにどの程度「説得力がある」と感じ、どのようなものにどの程度「考えさせられる」かは、個人によって異なり、そこに普遍性や絶対性は存在しない。したがって、その中身を明らかにしない状態で複数の審査員が評価したならば、結果には大きなずれが生じやすい。このような状態では、学習者の努力が評価結果に直結することも望みがたい。

しかし、普遍的・絶対的な条件が存在しないからずれを生じさせる要素を外す、という選択はできなかった。なぜなら、上にあげた要素は各地の大会の審査基準に含まれており、これらがスピーチに対して一般に求められる最大公約数的なものであることは間違いないからである。とはいえ、妥当性・信頼性のある評価のためには曖昧なままにしておくこともできない。そこで、普遍的・絶対的な条件を探るのではなく、極東大会の意義に照らし、その目的の達成のために必要とされる条件を設定することにした。すなわち、普遍性・絶対性よりも教育的配慮を優先し、「少なくとも極東大会では、このような条件を満たすスピーチを説得力があり、考えさせられるスピーチとみなすので、それを目指して能力の向上に努めてほしい」と明言する、ということである。

条件を示す文中に用いる言葉は、議論を重ねたうえで決定された。たとえば、「聞き手への意識」という項目において「魅力」という言葉が採用されたが、これに代わる候補として「面白さ」「エンターテインメント性」といったものもあげられていた。使用する言葉は、誰にでもわかりやすいものでありながら、イメージされるもののずれが複数の審査員の間で小さくなるようなものが理想的である。「面白さ」ではそのずれが大きすぎ、「エンターテインメント性」は誰にとってもわかりやすい言葉ではないという理由から、結果として「魅力」が採用されることとなったのだが、「魅力」も決してずれが小さいとは言えない。しかしその言葉が意味するところを基準内で詳細に説明するわけにもいかず、また他に適当な言葉も見つからないため、苦渋を伴う決定となった。

また、「内容の深さ・説得力」の項目では「大学生らしい視野の広さや思考の深さ」が、「主張」に属するものなのか、もしくは「主張をサポートする根拠」に属するものなのかという点が問題となった。たとえば「趣味を持つ」と主張するスピーチがあったとする。それ自体は、自身の周囲にしか目が向けられていない主張とも感じられる。しかし、その根拠として、趣味を持つことの医学的効用や、趣味から偉大な発明が生まれたなどといった歴史的事実にまで言及していれば、十分に「大学生らしい視野の広さや思考の深さ」を備えたものになり得るだろう。逆に、環境保全の大切さを取りあげたスピーチのように、世界的な時事問題を扱っていても、根拠にあげた事柄が身の回りの出来事に終始していれば、説得力のあるスピーチにはなら

ないかもしれない。一見稚く思える主張ほど、説得力を持たせるための技巧には高いものが求められる。それに挑戦する機会を奪うべきではなく、また、奪ってしまえば斬新なアイデアも生まれにくくなる。そのため、最終的に、視野の広さや思考の深さは、主張ではなく根拠に求められるべきだと判断した。

さらに、意義に明示されている「コミュニケーション力」の中身も、議論の対象となった。この弁論大会の出場者は当然、海外在住の学習者であり、その多くが将来、日本の文脈において日本語を使用したいと望んでいる。しかし、そこで必要とされると承知しながらも自国においては意識しにくい事柄もある。そこに目を向けてもらいたいという思いから、コミュニケーション力にはいわゆる社会文化的能力も含まれると解釈し、それを「質疑応答」の審査基準に盛り込むこととした。その際、専門用語をそのまま使用するのを避けるため、「日本的な常識から外れる（相手に対する配慮がないなど）ことなく、友好的かつ自然な対話を成立させることに貢献しているか」という形で記述した。「[日本的な常識]が何を意味するのか漠然としている」、[ロシア人の審査員にはわからないのではないか]といった意見もあり、確かに各審査員のイメージにゆだねる部分が大きくなるためにずれも生じると思われたが、かといって外すべき条件ではないと判断した。

以上のような議論を経て、これまでより明確なものとなった審査基準は、評価の信頼性を高めるだけでなく、どのような評価が行われた結果の序列であるのか、出場者および指導者に明示することを可能とする。また、明示されれば、出場者に足りなかった要素も明らかとなり、それを今後の学習の指針とすることもできるであろう。

4.5 複雑化した審査基準への対応

ここで、やむを得ず複雑化させた審査基準を、審査員に理解してもらうために講じた対策についても付け加えたい。まず日本語教育関係者だけではなく、領事館員や日本企業関係者が審査員となることもあるため、専門用語を基準内で使用することは避けた。また通常、小論文の大規模試験では、評価者間のずれをなくすためのミーティングおよび評価トライアルが評価者に課されるのが一般的であるが、多忙な中、協力してもらっている審査員にそこまで依頼するのは事実上困難なため、これを設定しなかった。

ただし、今回作成した審査基準は、初めて審査を担当する者が会場で見、いきなり使いこなせるほどの簡易性を有してはいない。そこで、審査基準と出場予定者の原稿を事前に配付し読んでおいてもらうことを基準内に示すこととした。また、基準内に規定してあるわけではないが、過去のスピーチ原稿の審査例（特に「聞き手への意識」と「内容の深さ・説得力」について、なぜその点数になったのか、理由まで説明されているもの）も5件ほど用意した。これらについても事前に配付し目を通してもらうことが望まれる。これは、評価者間信頼性（inter

-rater reliability) を少しでも確保するため、評価トライアルに代わって必要となる手続きである。

審査員全員に完全に対応してもらえるかどうかはわからないが、たとえ半数にしかこれらの事前配付物を読んでてもらえないとしても、あるいは完全に理解されていなかったとしても、項目のみで説明の乏しい審査基準を適用するよりは、はるかに信頼性を確保しやすくなることをここで強調しておきたい。審査員に理解してもらるのが難しいから簡素な基準のほうがいい、読んでてもらえるのが難しいから審査例を送らないほうがいい、ということにはならないだろう。

外国語による弁論大会は、言語能力を審査する場である。そして言語は、適切な手続きをふみさえすれば信頼性ある評価を行うことが可能であるとされている。だからこそ芸術の評価とは異なり、その道の大家ではない人物が評価者となり得るのであり、また審査の場は審査員の考えを反映する場ではなく、大会の方針すなわち審査基準を反映する場となる。大会を運営する日本語専門家には、言語の専門家として、学習者の努力の成果を評価し序列をつける責任の重さについて、共に運営に関わる皆に理解を求めていくことが望まれる。

5. 試用結果と分析

改定された審査基準は、極東大会に先立って、2008年ハバロフスク地方大会と2008年ユジノサハリンスク地方大会において試用された。後者では審査基準、出場予定者の原稿、審査例の3点の事前配付を行ったが、前者の実施段階ではこれらの事前配付の方針が決定していなかったため、審査基準のみ事前配付された。表2、表3にそれぞれの審査結果を示す。表中の値は、6つの評価項目の合計点とその平均値、および標準偏差である。

ハバロフスク地方大会においては、出場者に満点の30点をつける審査員が多く見られた(表2)。7名の審査員のうち、出場者5人に満点をつけた審査員が1名、3人に満点をつけた審査員が1名、2人に満点をつけた審査員が2名いた。7名の各審査員が全出場者に対してつけた点数は、審査員Aが16~29、Bが15~29、Cが21~30、Dが17~30、Eが10~25、Fが22~30、Gが24~30と適度にばらついており、審査基準が良好に機能しているように見える。しかし、逆に9人の各出場者に対する全審査員の点数は、出場者aに対して25~30、bに23~30、cに24~30、dに23~30、eに23~27、fに21~28、gに14~30、hに14~25、iに10~24と、審査員によってつけた点に大きな差が見られ、改善すべきであることが示唆された。

〔表2 2008年ハバロフスク地方大会審査結果〕

	出場者 a	b	c	d	e	f	g	h	i
審査員 A	29	29	26	23	23	21	20	20	16
B	28	28	29	23	25	21	19	20	15
C	30	30	30	30	26	24	30	24	21
D	30	30	29	26	27	23	26	21	17
E	25	23	24	23	23	23	14	14	10
F	30	30	30	29	26	28	25	23	22
G	30	30	29	29	26	28	26	25	24
平均値	28.86	28.57	28.14	26.14	25.14	24.00	22.86	21.00	17.86
標準偏差	1.73	2.38	2.10	2.95	1.46	2.73	5.03	3.38	4.45

大会運営者による反省会においては、審査基準に関して「1から5までの各点の条件が詳述してある部分があった」とする意見があった一方、「審査の素人である審査員にとって難しい」という意見があった。また、「個々の審査員が抱く各点の条件についてのイメージが異なっているように感じられ、やはり審査例が必要である」という意見も提示された。

その後行われたユジノサハリンスク地方大会においては、ハバロフスク地方大会での反省をふまえ、上述の審査例の事前配付が行われた。5名の審査員のうち出場者に満点をつけた審査員はなく、5名の各審査員の全出場者に対する点数も、審査員 A が15～27、B が14～26、C が15～24、D が17～27、E が18～24と適度にばらついた点をつけていた（表3）。また11人の各出場者に対する全審査員の点数は、出場者 a に対して24～27、b に21～26、c に21～24、d に19～25、e に19～24、f に18～24、g に18～21、h に16～21、i に16～19、j に16～18、k に14～19であり、6点以内のばらつきに抑えられた良好な結果となった。

〔表3 2008年ユジノサハリンスク地方大会審査結果〕

	出場者 a	b	c	d	e	f	g	h	i	J	k
審査員 A	27	21	21	23	24	24	20	20	16	16	15
B	26	21	22	20	20	21	21	20	19	18	14
C	24	22	21	19	19	18	18	16	18	17	15
D	27	26	24	25	23	24	19	21	19	17	19
E	24	21	23	22	22	21	20	21	18	18	19
平均値	25.60	22.20	22.20	21.80	21.60	21.60	19.60	19.60	18.00	17.20	16.40
標準偏差	1.36	1.94	1.17	2.14	1.85	2.24	1.02	1.85	1.10	0.75	2.15

これらの結果から、改定された審査基準は手続きをふまえて使用すれば、良好に機能するも

のであると言い得るだろう。しかしながら、出場予定者の原稿や審査例の事前配付なしで評価者間信頼性を保つことが難しいことも同時に示唆された。原稿や審査例の事前配付や大会前のブリーフィング等により、審査員に事前に審査になじんでもらうという定められた手続きを省略すれば、評価者間信頼性が下がることは容易に推測される。しかし、これは運営に携わった者の単なる主観に過ぎないが、審査基準のみ事前配付されたハバロフスク地方大会においてでさえ、決定された順位に関しては、以前の極東大会と比べ、より順当だという印象を抱かせるものであった。

6. 今後の課題

ハバロフスク地方大会後の運営者による反省会において、「審査の素人である審査員にとっては使いこなすことが難しい、実際に審査するのは素人なのだから審査方法を素人向きにしたほうがいい」という意見が示された。この言葉には、今後の課題が集約されていると言えるだろう。素人であるからこそ、公正な評価がなされるためには詳細な手続きが必要となることを、運営や審査に携わる者皆が理解しなければならないと思われるのだが、評価研究、特にパフォーマンス評価に関して知識や経験の豊富な教師は決して多くはなく、なかなか周知されない現状にある。

地方大会、極東大会を問わず、弁論大会の前後には、大会運営者であるその地域の教師会による意見交換会が行われる。大会前には審査基準等の手続きの見直しや確認が、大会後には反省点が話し合われるのが習慣となっている。それ自体は奨励されるべきことであるのだが、筆者らがある場を経験して危惧を抱くのは、そこで提出された意見を根拠とする改変が、安易に行われすぎていないかという点である。

審査に関する改変が「手続き上の利便性」だけでなく、「より公正な評価」にまで配慮したうえで行われているのであれば問題はない。しかし実際には、前者を強調する意見を耳にすることのほうが多かった。だが、どちらが本来優先されるべきであるかは言うまでもない。改変するのであれば、改変しても後者を守ることができるとする根拠が示されなければならないはずである。

「手続き上の利便性」と「より公正な評価」とは、反比例的な関係にあり、容易に両立するものではない。今回改定された審査基準は、前者を犠牲にして、後者を優先させたものである。したがって、手続きに携わる運営者のみが集まる会において、より複雑な手続きが必要となった審査基準に対し、否定的な意見が示されるのは当然のことと言えるだろう。しかし、それによって学生の正しく評価される権利が脅かされるかもしれないことについて、顧みることも重要である。

これまでは、「手続き上の利便性」を優先しがちな「評価する側」からのフィードバックの

みが収集され、「より公正な評価」を優先するはずである「評価される側」、すなわち学生たちからのフィードバックが収集されないまま、意見交換がなされていた。今後は、運営に携わった者から聞き取りを行うだけでなく、観客として参加した学生や直接指導に携わらなかった教師、および一般の観客などからもフィードバックを得ることが課題になると思われる。インタビューやアンケートなどの方法が考えられるが、その際、大会の意義や審査基準の一部を示し、それに照らしてどう思うか、という形で尋ねることを忘れてはならない。また、客観的な意見を収集するため、出場者や出場者と近い関係の者を除くといった配慮も必要となるだろう。さらに、本報告で行ったような評価者間信頼性に関する数値上の確認も、継続して行われてゆくことが望ましいだろう。

毎年行われる反省会では、それらのデータを詳細に検討し、改変を加える必要がないのであれば、それまでのあり方を尊重するのが肝要だと思われる。逆に改変を加えるのであれば、理論的背景をふまえて臨むのはもとより、それによって、学生の公正な評価を受ける権利が脅かされることのないことが、明確に保証されなければならないだろう。評価者が一種の権力をゆだねられる存在であるのに対し、評価される者には結果に反論する力が与えられない。したがって、規則に縛られることは評価者としての責務となる。そのことが評価に携わる皆に認識されてゆくためにも、評価する側の視点のみに終始しない、上述のような反省会が行われていくことが望まれる。

〔参考文献〕

- 笠原(竹田) ゆう子・小山慎治(2003)「スピーチの評価基準明示の効果」『日本語教育方法研究会誌』Vol. 10、No. 2、20-21
- 古賀美千留・中島晶子(2005)「スピーチ活動における評価とフィードバック効果」『日本語教育方法研究会誌』Vol. 12、No. 2、6-7
- 一二三朋子(2007)「留学生スピーチの評価基準と評価の過程—日本語母語話者と非母語話者の比較—」『文藝言語研究 言語篇』52号、13-22、筑波大学大学院人文社会科学研究科 文芸・言語専攻
- Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In Hamp-Lyons, L. (ed.), *Assessing second language writing in academic context.*, pp.241-276. Ablex publishing.

〔注〕

- ⁽¹⁾2005年の極東大会の審査項目は以下のようなものであった。(I) 技能審査項目：①発音・イントネーション②文法・語彙③内容④質疑応答、(II) 弁論審査項目：①主張②発表態度③内容④質疑応答。ここでは技能審査を日本語教師が、弁論審査を一般日本人が担当する形を採っているが、年度によっては分けないこともあった。また各項目の点数についても、5点満点を指示する年と10点満点を指示する年とがあった。しかし項目自体の変化はほとんどなかった。また、各出場者につけられた全審査員の順位を合計し、数値が最も少なかった者を優勝者とする「順位点方式」を採ることがほぼ恒例化していた。

各審査員が最終的に出場者全員に順位をつけることができれば、その過程はどのような形でも（極端な場合、採点結果を利用しなくとも）構わないとされた。

⁽²⁾ マルチプルトレイト・スコアリングに関しては、Hamp-Lyons (1991) 等に詳しい。

〔資料1 審査手順および質疑応答に関する規定〕

「審査手順」

- ① 各審査員席の机上にはメモ用紙と筆記具が準備されている。
- ② 審査は「発音」「文法・語彙」「聞き手への意識」「内容の深さ・説得力」「プレゼンテーション力」「質疑応答」の6項目で行われる。
- ③ 各項目の点数は5点から1点（原則的に、5：非常に良い、4：良い、3：普通、2：やや物足りない、1：物足りない）という尺度で採点するが、「聞き手への意識」「内容の深さ・説得力」「質疑応答」については、別に定めた尺度を用いて採点する。6項目の合計は6点（最低点）～30点（最高点）である。
- ④ 審査員は、各項目の点数、および6項目の合計点を審査用紙に記入する。また、6項目の合計点を元に、1位から5位までの順位を決定し、審査用紙の順位欄に記入する。6項目の合計点が同点の場合も、各審査員の判断で順位をつける。その後、係へ提出する。
- ⑤ 審査用紙回収後、集計係はエクセルと手計算による集計を行い、審査員全員の合計点から順位を確定する。同点があった場合のみ、該当する出場者につけられた順位点の合計によって最終順位を決定する（1位＝1点、2位＝2点、3位＝3点、4位＝4点、5位＝5点、順位なし＝6点として計算し、合計点数の低いほうを上位とする）。
- ⑥ 全発表終了後、審査協議を行う（約30分）。集計係は、集計によって算出した順位を審査員に報告する。合計点と順位点で確定しなかった場合は、審査員全員の協議により1位～5位を決定する。
- ⑦ 会場を盛り上げた出場者等のために、順位に関係のない「特別賞」を設けることができる。該当者があるか、審査員の協議によって決定する。
- ⑧ 入賞者決定後、審査結果を司会者に通知する。

「質疑応答」

- ① 質問は、開催地の教師会等が候補となる質問を用意しておく。最終的な選択・決定は、質問者3名が行う。ただし、質問者の所属機関からの出場者に対する質問の作成に、当該質問者は関与することができない。
- ② 質問は1人の出場者に対し2人の質問者で行う。1人の質問者からの質問事項は1件とし、計2件の質問とする。ただし、質問者は、回答後、確認のため同一の質問事項についてさらに詳しく尋ねることができる。2件の質問のうち、1つは発表内容への理解を確認するための比較的易しいもの、1つは発表内容を元にした発展的なものとする。
- ③ 3名の質問者が交替で行う。当該出場者の所属機関の講師が質問者にあたらぬようなローテーションで行う。
- ④ 質問は日本語能力試験3級程度のレベルの日本語により行われる。ただし、スピーチ内で使用された語彙・文型については、質問において使用してもかまわない。
- ⑤ 1つの質問は、1文または2文で行われる。
- ⑥ 出場者は2回まで質問の繰り返しを要求することができる（質問回数の合計は3回）。
- ⑦ 質問を繰り返す場合、2回目の質問は1回目の質問と同じ文言で速度を遅くして質問する。3回目の質問は、質問者の判断でよりわかりやすい文言に変えてもかまわないこととする。繰り返したこと自体は減点の対象としないが、繰り返しを求めた際の日本語や態度は、審査の対象となる。

[資料2 極東・東シベリア地域共通 弁論大会評価基準]

項目	基準説明
①発音	<p>「日本語の母音・子音の発音が正しくできているか」「アクセント、イントネーションは適切か」「話す速さ、間の取り方は適切か」</p> <p>5：非常に良い 4：良い 3：普通 2：やや物足りない 1：物足りない</p>
I 言語力	<p>②文法・語彙</p> <p>「正しい文法や語彙が使えているか」「適切な表現や語彙（スピーチにふさわしいスタイルや言葉使いなど）が使えているか」「同じ言葉のくり返しではなく様々な表現が使えているか（意図的に同じ言葉を使用している場合を除く）」「文と文のつながりは自然か」</p> <p>5：非常に良い 4：良い 3：普通 2：やや物足りない 1：物足りない</p>
③聞き手への意識	<p>「聞き手の興味を引く内容か」「論旨や話す順序に問題がなく構成・展開的にかかりやすいか」「聞き手の背景知識を考慮し、説明が必要なら（自国文化や出身地域の特徴、一般的な用語等）がある場合に、スピーチの中で説明しているか」</p> <p>5 感情表現や描写などに独創性や魅力が感じられ、また、構成・展開的にかかりやすく、聞き手の背景知識への配慮にも意図がない</p> <p>4 感情表現や描写などに独創性や魅力が感じられ、また構成・展開的にかかりやすいが、聞き手の背景知識への配慮にやや欠ける部分がある</p> <p>3 構成・展開的にかかりやすいが、独創性や魅力があまり感じられず、平凡である</p> <p>2 独創性や魅力の有無にかかわらず、論旨や話す順序に問題があるなど、構成・展開的にかかりやすいが、聞き手の背景知識への配慮にやや欠ける部分が生じている</p> <p>1 独創性や魅力の有無にかかわらず、聞き手への配慮が全く感じられず、全体的にかかりにくい</p>
II 内容	<p>④内容の深さ・説得力</p> <p>「主張（一番いたいこと）＝～し～よう、～すべきである、～しないほうがいいなど、一言で表すことができるもの。ただしその一言がスピーチ内になければならないという（ことではない）がはっきりしているか」「主張をサポートするための根拠（ここでは広く解釈し、主張を補強することにつながる部分はずべて根拠に当たるものとする）は適切かつ十分か」「根拠に大学生らしい視野の広さや思考の深さが感じられるか」</p> <p>5 主張が明確で一貫している。また主張をサポートするための根拠に量的な根拠に不足が感じられず、かつ質的にも客観性あるいは具体性を備えている。さらに根拠に視野の広さや深い思考が感じられる</p> <p>4 主張が明確で一貫している。また主張をサポートするための根拠に量的な根拠に不足が感じられず、かつ質的にも客観性あるいは具体性を備えている。しかし、根拠に視野の広さや深い思考が感じられず、平凡である</p> <p>3 主張は一貫しているが、主張をサポートする根拠に正確さを欠く部分があったり、客観性または具体性を備えていないかあったりする。あるいは量的な不足が感じられる</p> <p>2 主張はあるが、一貫性が欠く</p> <p>1 主張がない</p>
⑥プレゼンテーション力	<p>「スピーチの態度は適切か（姿勢、視線、顔の表情、声の大きさなど）」「視覚および聴覚的にアピール力があるか」（なお、小道具等の使用や歌、ダンス等のパフォーマンスを取り入れることは認められない。また、質疑応答でのプレゼンテーション力はここに含まれない）</p> <p>5：非常に良い 4：良い 3：普通 2：やや物足りない 1：物足りない</p>
III 運用力	<p>⑦質疑応答</p> <p>「正しく適切な日本語を用いているか」「質問に対する回答として十分か」「ふさわしい内容を発言しているか」「自身のスピーチをより詳しく理解してもらうために積極的に話しているか」「日本的な常識から外れる（相手に対する配慮がないなど）ことなく、友好的かつ自然な対話を成立させているか」（なお、回答により会場を沸かせたかどうかはこの項目では評価しない）</p> <p>5 正しく適切な日本語を用いており、また、質問に対して必要かつ十分な情報を発言している。積極的でありながら日本的な常識から外れることもなく、結果として友好で自然な対話が成立している</p> <p>4 部分的に正しさと適切さに欠ける部分があるが、理解できないほどではなく、質問に対して必要かつ十分な情報を発言している。積極的でありながら日本的な常識から外れることもなく、結果として友好的で自然な対話が成立している</p> <p>3 正しく適切な日本語を用いており、また、質問に対して必要かつ十分な情報を発言している。あるいは部分的に正しさと適切さに欠ける部分があるが、理解できないほどではなく、質問に対して必要かつ十分な情報を発言している。ただし、積極性に欠ける。あるいは日本的な常識から外れた印象を与えている</p> <p>2 2問中1問のみ、聞き返しを含め、正しく適切に回答が行われた。もう1問は、正しさと適切さに欠けている、もしくは必要かつ十分な情報がないなどの理由により質疑応答が成立していない（回答拒否も含む）</p> <p>1 正しさと適切さに欠けている、もしくは必要かつ十分な情報がないなどの理由により、2問とも質疑応答が成立していない（回答拒否も含む）</p>