口頭産出測定のための正確さの指標
－妥当性の検証と今後の研究課題－

田　島　ますみ

要　旨

第二言語学習者の口頭産出がどの程度に正確であるのかを表す客観的な指標に関しては、現在、研究者の間でも統一的な見解が確立されておらず、その妥当性、信頼性に関する実証研究もないまま、個々の研究において様々な指標が使われているのが実情である。本研究では、日本語学習者の口頭産出を文字化したKYコーパス中36人分のデータを使用して、正確さを表す指標として用いられている三種の指標の妥当性を、OPIの主観的評価を基準として統計的に検証した。三種の指標は、これまでの研究で区別されている特定指標、全体指標、さらに全体指標の中でも誤用のない数量系と誤用数系の三つのカテゴリーから一つずつを選んだ。結果はどの指標の妥当性にも肯定的な実証証拠を示さなかった。これに基づいて、今後の研究課題を明確にするとともに、さらに正確さの妥当な指標が確立されるための前提となる発話の正確さと中間言語の発達の関係についても論及する。

キーワード：口頭産出、正確さ、指標、妥当性、KYコーパス

1．はじめに

第二言語学習者の発話の正確さはどのような指標によって測れるのであろうか。最近の第二言語習得研究では、学習者の口頭産出を分析する際に正確さ、流暢さ、複雑さの三つの範疇がよく用いられている。正確さは、流暢さや複雑さに比べればそれほど曖昧な用語ではなく、学習者の発話が正確であるかないかの判断は比較的明確につけられる。しかしながら、単語や文レベルではなく談話レベルの、ある程度まとまった長さのある産出がどの程度に正確であるのかを測る指標に関しては、研究者の間でも統一的な見解がなく実証研究もないため、個々の研究において様々な指標が用いられているのが現状である。異なる指標を用いて得られた研究結果は一般化の論拠に乏しい。また個々の研

究の信頼性も、指標が確立していないために低くなる。指標の信頼性、妥当性を実証的に検討し、より適切な指標を見つけ出していく基礎作業は、今後の研究のためにも十分になされていくべきである。

本研究では、日本語学習者の口頭の産出を文字化したKYコーパスをデータとし、正確さを客観的に表す指標として用いられている三つの指標の妥当性を、OPIの主観的評価を基準として統計的に検証する。KYコーパスはOPI（Oral Proficiency Interviews）で得られた日本語学習者の自然発話、90人分を文字化したもので、ACTFL（American Council on the Teaching of Foreign Languages）のガイドラインに従った個々の学習者の能力判定の結果もついている。本研究では、このOPIの判定結果を基に、正確さを表す指標の数値がOPIの各レベル間で有意差を示すかどうかを分析し、指標がOPIの主観的評価を反映した識別度を持つ妥当なものであるかのかを検証する。特に特定指標と全体指標の違い、及びこの二種類の指標に関する議論（Foster & Skehan, 1996; Ortega, 1999）を視野に入れ、さらに全体指標の中でも誤用のない割合と誤用数という区別（Bygate, 2001; Mehnert, 1998）を考慮し、これら種類の異なる正確さの指標がどのような数値を出してくるのかを検討する。また、実際の数値を出す際には、口頭産出分析のための基本単位が重要な問題となる。基本単位に関する議論、問題点にも言及し、その定義を明確にした上で、本研究のデータ分析、統計処理を行い、その結果を基に正確さを表すより適切な指標について考察する。

2．正確さを表す指標に関する問題点

2.1．これまでの研究に使用された指標

正確さを数量化するために使われている指標は実に様々である。以下に、最近の口頭産出の量的分析を含んでいる研究で用いられた正確さの指標の例を、正用率系、誤用のない数量系、誤用数系の三つの種類に分けて示す。正用率系は、ある特定の項目に着目して学習者がその項目を母語話者と同じように使えているかを割合で表すものである（表1参照）。誤用のない数量系は、誤用のない部分の量を数や割合で表すもの（表2）、また、誤用数系は、何らかの条件下の誤用の数を数えるものである（表3）。

表１：正用率系の指標の例

指　　標	対象言語	使用した研究
動詞の過去形の正誤	英語	Ellis（1987）
冠詞の正用率	英語 スペイン語	Crookes（1989） Ortega（1999）
名詞複数形（*s*）の正用率	英語	Crookes（1989） Wigglesworth（1997）
動詞活用形の正用率	英語	Wigglesworth（1997）
名詞修飾部の正用率	スペイン語	Ortega（1999）
助詞の正用率	日本語	Nagasawa（1999）

表２：誤用のない数量系の指標の例

指　　標	対象言語	使用した研究
誤用のない1 T-unitあたりの語数	英語	Crookes（1989）
誤用のない節の割合	英語	Crookes（1989） Foster（2001） Foster & Skehan（1996） Iwashita, *et al.*（2001） Robinson（2001） Skehan & Foster（1999）
誤用のない節の数	ドイツ語	Mehnert（1998）

表３：誤用数系の指標の例

指　　標	対象言語	使用した研究
1節あたりの誤用数	英語	Williams（1992）
1 T-unit あたりの誤用数	英語	Bygate（2001）
100語あたりの誤用数	ドイツ語	Mehnert（1998）
語彙に関する誤用数	ドイツ語	Mehnert（1998）
語順に関する誤用数	ドイツ語	Mehnert（1998）

2.2. 特定指標と全体指標

上記三種の分類は、これまでの研究の中でしばしば言及されている区分であるが、まず、正用率系と後二者、誤用のない数量系と誤用数系は、特定指標と全体指標という二つのカテゴリーに分類される。特定指標はある項目だけに絞って正用かどうかを見るもので、全体指標は発話全体を対象として正確さを数量化しようとする指標である。この二種のうちどちらを使うかに関しては実験条件の違いや中間言語の発達などの観点から

の議論がある。

初期の研究では特定指標が使われることが多かったが、この問題点を指摘したのがFoster & Skehan（1996）である。特定指標は、学習者の発話にその項目が相当量含まれていれば適切であるが、その項目があまり使用されない場合は適切な指標とは言えない、つまり、特定指標は実験条件の違いに左右されやすい傾向があるので、もっと一般化した指標が望ましいであろうという主張であった。Foster & Skehanがこの時に採用したのが、誤用のない節の割合（percentage of error-free clauses）である。特定の項目の正用率を見るのではなく、いわば発話の全体量における正用率を見る指標の採用であった。全体量を節の数で表し、誤用のない節がどれくらいの割合を占めるかを示した。この指標は後に続く研究でも頻繁に使用され、現在最もよく使われている正確さの指標と言っていいだろう。

これに対して、Ortega（1999）は、全体指標の長所を、学習者の発話が対象言語に従う用法になっているのかについてのより現実的な全体図が得られることとして認めながらも、一方で、その短所を、中間言語の発達にとって重要となる文法領域での小さな変化が捉えられないことと指摘した。実際、Ortegaの研究では全体指標は採用されず、特定指標二種を用いての分析となっている。

いずれの立場もそれぞれの指標の特徴を明確にしていて否定的な要素は見当たらない。現段階では、Ortegaの言うように、全体指標及び中間言語発達にとって鍵となる項目を見る特定指標の併用が、量的分析には最良の選択と言える。しかしながら、これらの議論を踏まえた上で、指標の妥当性を実証的に検証してみることも、また違った角度からの指標選択の論拠となるであろう。本研究の意義は、これらの議論の上に実証証拠を加えることにある。

2.3. 誤用のない割合、あるいは誤用数

特定指標、全体指標の区分とは別に、誤用のない割合と誤用数という区分もしばしば指摘される問題点である。この区別は全体指標の下位区分としてなされる。Mehnert（1998）は、誤用のない節の割合が誤用を一つしか含まない節と複数の誤用を含む節を区別しないことを指摘し、自身の研究においては全体指標として、1節あたりの誤用数、さらに節は長さに違いがあるので、その違いの影響を消すために100語あたりの誤用数を採用した。Bygate（2001）も、誤用のない節の割合は記録される誤用の数を減

らして識別度を下げてしまうとし、同じ問題点を指摘している。Bygate は、誤用のない割合と誤用数という二種類の指標の妥当性と有効性に関する研究がないことにも言及し、その調査研究を促している。

2.4. 基本単位としての AS-unit

故にこれまでの問題点を整理すれば、特定指標、全体指標、その中でも誤用のない割合と誤用数を示す指標、という三種類の指標の検証が必要であることがわかる。本研究では、当初、日本語を対象言語とするので特定指標として助詞の正用率を、全体指標では最も使用頻度の高い誤用のない節の割合、及び 1 節あたりの誤用数を選んで検証することを計画していた。しかしながら、実際の数値を出す段階になって、節を基本単位としては全体指標にならない可能性が出てくることが確認された。今回データとした KY コーパスは、学習者と試験官の対話型の口頭産出であり、対話型である場合、学習者の 1 回の発話が節にならない場合は少なくない。質問に対して「はい」「いいえ」のような短い応答ですんだり、節にならない名詞句の形で答えても正確な発話である場合は多い。特に初級者においては、節の産出自体が非常に少なくなる傾向が顕著である。よって、全体の発話の中で非常に数の少ない節だけを見ていくことは、全体指標ではなくて、むしろ特定指標に近くなるケースが出てきたわけである。本研究では全体指標と特定指標の区別に焦点を当てているので、限りなく特定指標に近い全体指標を検証しては意味がない。そこで、当初の計画を若干変更し、全体指標の基本単位を節ではなく、最近提案された AS-unit とすることにした。すなわち、誤用のない AS-unit の割合、1AS-unit あたりの誤用数を二種類の全体指標として検証する。

Foster, Tonkyn, & Wigglesworth（2000）は、AS-unit を「独立節、あるいは節に準ずる単位、及びそれに結び付く従属節も含む」（p. 365、原文英語、訳は著者）と定義している。この定義は、今までの研究で最も頻繁に使われている T-unit、その問題点を指摘して提案された c-unit をより明確に統語的に定義し直すという形で提案された。T-unit はもともと第一言語習得の分野で使われている単位で、「主節、及びそれに従属する節を含む」と定義されている。この単位に対して、話し言葉にはそもそも省略の傾向があり、節にならない発話も多いのだから、これを分析の基本単位とすることは不適切だという議論がなされ、代わりに c-unit（communication unit）という単位が提案される。しかしながら c-unit の定義は様々で、特に確立された定義はなく個々

の研究者が自分なりの基準で判断して使っているというのが実状である。例えば、Pica, Halliday, Lewis, & Morgenthaler（1989: cited in Foster *et al.*, 2000）は「文法的、非文法的に関わらず、指示的な、あるいは語用論的な意味を提供する発話、例えば、語、句、文など」（原文英語）と定義しているが、「指示的な、あるいは語用論的な意味（referential or pragmatic meaning）を持つ語、句、文」では、ほぼすべての発話を指してしまってあまり意味のある定義とは言えない。Loban（1966: cited in Foster *et al.*, 2000）の定義、「文法的な独立の述部……あるいは質問に対する答えで、その質問の要素の繰り返しとなるもののみを欠いているが、独立述部である基準は満たしているもの」（原文英語）も、慣用表現的な語や名詞句を含まずすべての発話を包含できそうにない。

このような経緯から、話し言葉の分析の際の基本単位としてより明確な統語的な定義を試みたのが上記の AS-unit の定義である。「節に準ずる単位」（a sub-clausal unit）とはどんなものなのか、「それに結び付く」とは何をもって結び付くとするのか、不明な部分もあるが、ともかく節を基にした T-unit の問題を補う定義が出されたわけである。節に準ずる単位が入ることで、話し言葉では頻繁に使われるものの、T-unit ではカウントされない表現、例えば「はい」「ありがとう」「すみません」などがこの単位を使えばカウントされることになる。本研究では、AS-unit を基本単位として採用する[1)]。

3．研究方法

3.1．資　料

KY コーパスの中から初級、中級、上級、超級各９人、計36人分を選んでデータとする。各９人のデータは、それぞれ母語が中国語、英語、韓国語の話者から３人ずつ選ぶ[2)]。

3.2．指　標

前述のように、本研究では正用率系の特定指標、誤用のない数量系と誤用数系の二種の全体指標という三種のカテゴリーからそれぞれ一つずつ、以下の三つの指標を選択してその妥当性の検証を行う。詳細は次項に記す。

特定指標：＜正用率系＞

主要助詞の正用率

全体指標：＜誤用のない数量系＞

AS-unit の総数における誤用のない AS-unit の割合

＜誤用数系＞

1 AS-unit あたりの誤用数

3.3. 各指標の詳細と数値算出法

3.3.1. 主要助詞の正用率

本研究では、特定指標の代表として主要助詞の正用率を取り上げる。他にも特定指標の項目となるものは動詞の活用形や名詞の修飾部など様々なものが考えられる。しかしながら指標選定の際には、中間言語の発達を表す意味のある項目を選ぶことが重要となる（Ortega, 1999）。さらにその上で、使用頻度がある程度見込めるものであることが必要である。ヨーロッパ言語を対象とした研究では冠詞の正用率が代表的である。日本語を対照とする際に、この指標に相当するような指標で、ある程度の使用頻度が見込めるものとして、助詞の正用率を選択する。

さらに今回、助詞の正用率は主要助詞の「は」「が」「を」「に」に絞ることにする。「が」「を」「に」は日本語の構文が構成される際の基本の助詞であり、使用頻度はかなり高い。信頼性のある正用率の指標とするには、初級レベルの学習者でもある程度使用して適当な数のトークンを産出するような助詞でなければならない。初級者が全く使わないような助詞では正誤が見られず比較が困難となるし、使用頻度が低ければ数値の信頼性は下がる。使用頻度の高さ、また日本語構文における基本助詞であることを考慮し、さらにもう一つの基本助詞であり格助詞「が」との誤用が起きやすい「は」を加えて、上記の四助詞を選択する。

その他の主要助詞に絞る理由としては、終助詞などは正誤の判断が揺れるものが多いことが挙げられる。「ね」や「よ」の使用ははっきりと正誤を断定しかねるものが多い。また終助詞の他に、助詞の「の」を見るには名詞の代用としての「の」や動詞を名詞化する際の形式名詞「の」と区別せねばならず、誤認の可能性が高まり数値の信頼性が低くなる、というような事情もある。また軽視されがちではあるが、時間的な効率の問題も研究に使用する指標としては重要な要素である。いくら信頼性、妥当性の高い指標で

あってもその数値を出すのに不適当と思われる程の長大な時間を要するようでは、実際の指標として研究に使用できない。研究に適切な指標を見い出すには効率面での要素も配慮して然るべきである。以上のような理由で本研究では主要助詞のみとする。

また、実際のカウントの際には、基準を明確にするために、助詞の省略・欠落は今回正誤を見る数に入れない。実際に産出された上記四種の助詞の使用が正用か誤用かのみを判断して、全体数における正用の割合をパーセンテージで出す。また学習者自身が誤用を言い直している場合（self-correction）は、言い直した方の助詞で正誤の判断をし、最初に言った方の助詞は採らない[3]。なお、「を」「に」に関しては「～ができる」「～になる」に現れるものは慣用句化していると見なし、カウントしない[4]。

3.3.2. 全体指標二種

全体指標の一つとしては、前述したように誤用のない節の割合の問題点を考慮して、本研究では誤用のない AS-unit の割合を検討する。全体指標としての特徴をさらに強調するために、学習者の発話はすべていずれかの AS-unit に属するようにカウントし、AS-unit にカウントされない発話はないようにする。そのようにして出した AS-unit の総数を分母とし、誤用のない AS-unit の数を分子として割合をパーセンテージで出す。

もう一つの全体指標、1 AS-unit あたりの誤用数は、学習者の発話の全誤用数を AS-unit の総数で割って計算する。誤用には、発音[5]、文法、語彙のレベルで誤用と判断されるものはすべて含める。ただし、デスマス体とデアル体の不統一、性差のある表現の不適切な使用などは、正誤判断が揺れるものも多いので今回誤用としない。また方言の影響も筆者が方言と判断できる限りにおいて誤用とはしない。

なお、全体指標の二つでは、主要助詞の正用率で見ない助詞の省略・欠落も誤用の対象とする。助詞が落ちていることに不自然さを感じるものは、誤用のない、すなわち error-free の AS-unit とはしない。また誤用数を数える際にも、不自然な助詞の省略は誤用としてカウントする。

3.4. 分　析

上記三種の正確さの指標が言語運用能力を反映する適切な指標であるかを検証するため、各指標の平均値が OPI の初級、中級、上級、超級の 4 レベルの間で有意差を示すかどうかを分析する。各セルのサンプルサイズが同数であるためテューキー法による多

重比較を行って各レベル間の有意差を検討する。

4．結　果

4.1．平均値

分析を行う前に、三種の指標の各レベルにおける平均値を示す。個々の数値に関しては、稿末資料を参照されたい。

まず主要助詞の正用率は、初級から上級に上がるにつれて高くなっていくものの、それほど差があるわけではない。誤用のない AS-unit の割合においては、初級、中級がほぼ同じような値、上級でやや高くなり、超級ではかなりよい数値が出ている。1 AS-unit あたりの誤用数では、中級の発話に最も多くの誤用がカウントされ、それに上級が続き、初級の誤用数はさらにその次となった。

表4：3指標の各レベルにおける平均値

	助詞の正用率	誤用のない割合	単位あたりの誤用数
初 級	89.2138	59.9491	0.6868
中 級	89.7787	59.4449	0.7925
上 級	92.5124	67.4675	0.7292
超 級	97.9155	84.0488	0.2505

4.2．多重比較

次にテューキー法による多重比較の結果を以下に示す。

まず、主要助詞の正用率は四つのレベルのどこにおいても有意差を示さない。F 比＝1.67、P 値＝0.1932 で、帰無仮説は棄却されず、四つのレベルの平均値は有意には異ならないという結果となった。

誤用のない AS-unit の割合は、超級者とその他のレベルで有意差が出た（表5参照）。超級者の発話では、他のレベルとは有意に異なる高い割合で正確な産出がなされたということになる。ただし、超級者以外のレベルの差は検出されなかった。上級者はこの指標の数値を見る限り、初、中級者と変わりが無い。また、初級と中級の間にも有意差は無いことから、平均値は初級の優位を示しているが、初級が中級より有意に上ということではない。

1 AS-unit あたりの誤用数では、超級と上級、超級と中級の間でのみ有意差が認めら

れた（表6参照）。しかしながら、これは超級と初級は有意には変わらないということになる。

表5：誤用のないAS-unitの割合

レベル	レベル	平均値の差（左項－右項）
超級	上級	16.581**
超級	中級	24.604**
超級	初級	24.100**
上級	中級	8.023
上級	初級	7.518
中級	初級	－0.504

**：5％水準で有意

表6：1 AS-unitあたりの誤用数の結果

レベル	レベル	平均値の差（左項－右項）
超級	上級	-0.479**
超級	中級	-0.542**
超級	初級	-0.436
上級	中級	-0.063
上級	初級	0.042
中級	初級	0.106

**：5％水準で有意

5．考　察

結果は、初級から級が上がるにつれて数値がよくなっていくという単純なものではなく、また有意差もそれほど捗々しくは見られない。この結果を見る限り、どの指標も適切な指標とは言い難く、指標の数値はOPIの級の判断とそれほど関連しているものではない。以下、特定指標と全体指標に分けて考察する。

5.1．特定指標

特定指標である主要助詞の正用率は、かろうじて級が上がるにつれて数値が高くなるが、有意差を示すまでには至っていない。一つの要因としては、今回助詞の省略、欠落は見なかったことが考えられる。KYコーパスを見ていくと、級の低いレベルでは、間違った助詞を使うよりもむしろ助詞を使わない、省略する傾向がある。助詞の省略、欠

落は、話し言葉ではかなり厄介な問題で、省略可能な場合もあるし、また省略しないとおかしい場合もある。つまり、省略しても使ってもいいケースがある一方で、使うと誤用となるケースがあり、さらに使わなければ誤用である場合がもちろんある。今回は数値を出す際の基準を明確にするために欠落を誤用とはしなかったが、あるいは、欠落も含めて正用率を見ればまた違った結果が出たかもしれない。また今回は、主に使用頻度という理由で主要助詞の「は」「が」「を」「に」に絞ったが、主要助詞であるとレベルに関係なく正用率が高いという可能性も考えられる[6]。今回扱わなかった「で」や「へ」など、他の助詞の正誤を見ることで正用率が変わってくることもあるだろう[7]。

またこれらの要素よりももっと直接的に結果に影響しているのが、量の問題である。各レベルの助詞の平均使用回数は、初級、19.11、中級87.67、上級147.78、超級190.89（資料参照）で、かなりの開きがある。率になると有意差の出ない同じような平均値となるが、助詞の度数だけ見れば各レベルに明らかな違いがある。極端な例を挙げれば、初級学習者の一人は主要助詞を５回しか使っていないが、その５回すべて正用であるために正用率は100％となる。一方、超級でこれらの助詞を307回使っているデータでは、１回の誤用があるために正用率は99.67％となって、100％の初級者より低いということになる。かといって、この超級学習者が初級学習者より正確さにおいて劣るということは、彼ら二人のコーパスを読む限り全くないのである。この指標の妥当性は疑わしいということになる。また別の例としては、初級で主要助詞の使用はわずかに 2回、そのうちの一つが間違っているために正用率50％の値が出ている。これではあまり信頼性の高い指標とは言えない。以上の数値は、Foster & Skehan（1996）の指摘する、特定指標は、検討する特定の項目が十分に使われて相当数のトークンがなければ適切な指標としては機能しないということの例となるだろう。

5.2. 全体指標

それでは、全体指標の問題点は何なのだろうか。本研究の結果では、誤用のない AS-unit の割合、1 AS-unit あたりの誤用数とも、OPI の級とはうまく連動していない。なぜであろうか。まず、誤用のない AS-unit の割合では、僅差ではあるが、初級と中級が逆転する。この理由に、初級学習者のそれぞれの AS-unit がかなり短く、あまり誤用の余地のないものが多いことが挙げられる。「はい、そうです」「わかりません」「すみません」「ありがとうございます」のような慣用表現、名詞に copula「～です」

をつけた「机です」「3年生です」「刺し身です」など、誤用の起きる可能性が低い表現が1 AS-unitとして数えられ、さらに誤用がないユニットとして数えられる。中級ではもう少し長くて複雑なAS-unitが出てくる。例えば「あのー日本に来た日は4月6日から、あー韓国のソウルから来ました」「あのシンガポールはー、小さいくてー、あのー、とても暖かいんですけど、あの、シドニ、シドニーは、の方が、大きい町ですけど、あのーわたしはシドニーの方が好きです」「うん、そうですねー、暇な時間は、やーだいたい、自分好きなところで、遊びます」などは、初級学習者のレベルよりは長く複雑なAS-unitとなっているが、それぞれただ一つの誤用があるためにerror-freeとはならない（下線部、誤用）。初級と中級のコーパスを比べると、初級は正確さを云々する前に、まず＜長く言えない＞＜複雑なことは言えない＞＜ごくごく簡単なことしか言えない＞というレベルにあるようである。またそれ故にOPIで初級と認定されるのであろう。これに対して中級は＜より長い複雑な発話にはなるが誤用が含まれる＞ということになる。

同様のことが1 AS-unitあたりの誤用数の結果にも影響していると考えられる。この指標では、さらに複雑な逆転現象となって、超級と中・上級が有意に違うが、超級と初級の差は有意ではないという結果が出た。初級学習者は上級学習者よりも正確な発話をしているということになる。しかし実態は、中・上級では誤用の起きる確率の高い長くて複雑な発話ができるが、初級ではまだそこまでに至らないということであろう。

このように見てくると、本研究で取り上げなかった誤用のない節の割合（percentage of error-free clauses）があらためて有効な指標の候補として浮上してくる。この指標は現在最もよく使われている指標ではあるが、本研究が特定指標と全体指標の違いに主に着目したために、検証の対象とはしなかった。しかし、誤用のない節の割合は、全体指標とも特定指標ともつかない指標となりながら、極めて興味深い数値を出してくるものと考えられる。初級学習者の全体指標の数値がよくなるのは、誤用の余地のない、あるいは誤用の起きる可能性の低い短い発話が多いことが一因と考えられる。一方、節は用言の要素を含む発話しかカウントせず、短い慣用表現などが数に入らないことになる。その上で、誤用のない節の割合を見れば、AS-unitよりは複雑な発話部分に関しての正用率となり、今回取り上げた誤用のないAS-unitの割合よりも妥当な数値を出してくる可能性は十分にある。今後の研究で検証する必要はおおいにあると考えられる。

さらに発展的に考えれば、指標の妥当性という問題を超えて、言語運用能力と正確さ

がそもそも関連しているのかという疑問が出てくる。客観的な指標で測定した場合、初級者がより正確さの高い数値を出すことがあり得るのである。この結果で、すぐ想起されるのは、第二言語習得で以前から言われている中間言語の発達は線状ではなくU字型の曲線を描くという仮説である（Lightbown, 1983: cited in Gass & Selinker, 1994）。この仮説を支えているのは、言語習得は中間言語の再構築を繰り返しながら現在の段階から次の段階へと移行して進む、という考え方である。もし正確さの発達がU字型を描くのであれば、本研究で検証した二つの全体指標は、初級よりも中級でやや後退し超級において有意に異なる正確さに達するというU字の推移を示して、発達曲線に忠実な妥当性のある指標ということにもなりうる。本研究はこの問題に関しての実証研究ではないので、これについて議論することは本論の主旨を超えてしまうが、中間言語の発達過程と発話の正確さの関係はこれからの興味深い研究課題となるのではないだろうか。

6．まとめと今後の課題

本研究では、正確さの指標の妥当性は実証されなかった。検証した三種の指標の優劣は、今回の結果ではとうてい論じ得ない。現時点での結論としては、第二言語学習者の口頭産出の測定・評価は、正確さの三種の指標を組み合わせること、さらに、正確さの指標と中間言語の発達過程との関係が解明されない限り、正確さの指標だけでなく複雑さ、流暢さなどその他の指標も組み合わせて総合的になされるべきであるということになるであろう。今回の結果を基に研究方法、あるいは指標を変えての研究が、今後必要である。考察の項で述べた、特定指標、全体指標の今後の研究課題、正確さと言語習得のプロセスの関係を視野に入れた指標の研究の必要はここで繰り返さないが、最後に本研究の制約を記して今後の研究への示唆としたい。

まず、データとしたKYコーパスの問題である。もちろんKYコーパスが問題のあるデータだと言うのではなく、指標検証にこのデータを使うことの問題点である。KYコーパスでは、レベルによる産出量の差が激しい。超級と上級では、そこまでの差は感じられないが、初級と超級の発話量の差はかなりのものがある。KYコーパスを使って語彙的複雑性の指標を検証した研究（田島，2002）でも、一つの指標においては量の影響によって指標の数値が予想とは逆の順番を示す結果が出た。言語運用能力を基準にして指標検証を試みる場合は、もっと量的に近似するデータの使用が望ましい。KYコー

パスだけでなく、様々なレベルの学習者の口頭産出を文字化したコーパス、できれば量的に似通ったコーパスの作成、さらにそれらコーパスへの研究者のアクセスを容易にすることが求められる。

また、36人分のデータでは、サンプルが小さい。もっと大量のデータを統計処理することが望ましいが、指標の数値を出してくる際の時間と労力を考えると、現時点では限界がある。共同研究、あるいはテクノロジーによる数値の簡便な算出方法などが期待される。

さらに、KY コーパスは対話型の口頭産出であることにも注意を払いたい。対話型では、助詞に限らず省略の傾向が強く、モノローグ型の産出とはまた違った特徴がある。モノローグ型の口頭産出で今回の指標を検証してみればまた違う結果が出るかもしれず、対話型では適切な指標ではないがモノローグ型では有効という可能性も否定できない。

以上、本研究は正確さの指標の妥当性の実証とはならず、代わりに今後の研究課題の提示と、さらに第二言語習得における正確さに関しての問題提起となった。正確さの指標に関して何らかの結論が出るのはこれからの研究の積み重ねを待った上でのことになる。

注)

1）ただし AS-unit はまだ定着した用語とはなっていない。Foster（2001）の研究でも AS-unit の定義を c-unit の定義として使っており、混乱が見られる。

2）OPI では、級の中でさらに細かい等級をつけているが、その各級における下位区分も考慮してできるだけ多様なデータが入るようにした。詳細をここに記す。初級は、初級上・中・下から３人ずつ、中級も同様に中級上・中・下から３人ずつ。上級は、単なる上級４人、上級上５人。超級は超級のみ。

3）今回の資料では多くの場合、言い直した方が正用となっているが、言い直して誤用となるケースも皆無ではなかった。

4）その他、「役に立つ」の「に」も慣用句中としてカウントしなかった。ただし「気をつける」「気が合う」「気にする」「気がする」などの「気」を使った慣用句は誤用が目立ったため、一まとまりのユニットとして記憶される慣用句化が学習者の中でなされていないと考え、正誤判断の数に入れた。

5）今回、文字コーパスであるKYコーパスの上で発音の間違いとみなしたものは、文

字で表記されたものが正しい形とは違っているものである。

例：どきゅせい（同級生）、コンビニヤンスストア（コンビニエンスストア）

6）カウントの際、筆者の感覚では特に「を」の正用率は高いように感じられた。

7）KYコーパスを見ていると、「に」と言うべきところに「で」「へ」を使う誤用は少なくない。

参考文献

牧野成一（監修）（1995）『ACTFL-OPI試験官養成用マニュアル』アルク

益岡隆志・田窪行則（1992）『基礎日本語文法－改訂版－』くろしお出版

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* (pp. 94-104).

ACTFL (1999). *The ACTFL Oral Proficiency Interview: Tester training manual.* New York: The American Council on the Teaching of Foreign Languages.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23-48). Harlow: Longman.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11,* 367-383.

Crookes G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics, 11,* 183-199.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9,* 1-20.

Ellis, R. (1994). *The study of Second Language Acquisition.* Oxford: Oxford University Press.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language*

learning, teaching and testing (pp. 75-97). Harlow: Longman.

Foster, P. & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18,* 299-323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354-375.

Gass, S. & Selinker, L. (1994). *Second language acquisition: An introductive course.* Hillsdale: Lawrence Erlbaum.

Halleck, G. (1995). Assessing oral proficiency: A comparison of holistic and objective measures. *Modern Language Journal, 79,* 223-234.

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning, 51,* 401-436.

Lightbown, P. (1983). Exploring relationships between developmental and instructional sequences in L 2 acquisition. In H. Seliger & M. Long (Eds.) *Classroom oriented research in second language acquisition* (pp. 217-243). Rowley: Newbury House.

Loban, W. (1966). *The language of elementary school children.* (Research Report No. 1). Champaign: National Council of Teachers of English.

Makino, S. & Tsutsui, M. (1986). *A dictionary of basic Japanese grammar.* Tokyo: The Japan Times.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11,* 113-128.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20,* 83-108.

Nagasawa, S. (1999). Learning and losing Japanese as a second language: A multiple case study of American University Students. In L. Hansen (Ed.), *Second language attrition in Japanese contexts* (pp. 169-212). Oxford: Oxford University Press.

Ortega, L. (1999). Planning and focus on form in L 2 oral performance. Studies

in *Second Language Acquisition, 21,* 109-148.

Pica, T., Halliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition, 11,* 63-90.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22,* 27-57.

Skehan, P. & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49,* 93-120.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14,* 85-106.

Williams, J. (1992). Planning discourse marking and the comprehensibility of international teaching assistants. *TESOL Quarterly, 26,* 693-711.

Wolf-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity* (Technical Report No. 17). Honolulu: University of Hawaii, Second Language Teaching & Curriculum Center.

資料：総　デ　ー　タ

ID	# of Correct P	# of P	TLU of P	Error-Free AS-units	# of AS-units	% of EFAS	# of Errors	# of Errors /AS-unit
CNL01	14	14	100.0000	29	49	59.1837	37	0.7551
CNM01	5	5	100.0000	16	34	47.0588	32	0.9412
CNH01	26	33	78.7879	46	97	47.4227	78	0.8041
ENL01	1	2	50.0000	15	24	62.5000	13	0.5417
ENM01	8	8	100.0000	53	72	73.6111	29	0.4028
ENH01	38	43	88.3721	99	188	52.6596	147	0.7819
KNL01	8	8	100.0000	52	67	77.6119	23	0.3433
KNM01	24	27	88.8889	47	79	59.4937	58	0.7342
KNH01	31	32	96.8750	39	65	60.0000	57	0.869
N-Average	17.2222	19.1111	89.2138	44.0000	75.0000	59.9491	52.6667	0.6868
CIL01	23	24	95.8333	66	110	60.0000	78	0.7091
CIM01	48	63	76.1905	8	173	50.2890	216	1.2486
CIH01	99	114	86.8421	134	215	62.3256	161	0.7488
EIL01	53	59	89.8305	104	138	75.3623	41	0.2971
EIM04	81	8	93.1034	58	82	70.7317	38	0.4634
EIH03	61	68	89.7059	102	190	53.6842	183	0.9632
KIL01	50	58	86.2069	25	81	30.8642	118	1.4568
KIM01	122	129	94.5736	90	134	67.1642	72	0.5373
KIH01	179	18	95.7219	93	144	64.5833	102	0.7083
I-Average	79.5556	87.6667	89.7787	84.3333	140.7778	59.4449	112.1111	0.7925
CA01	114	124	91.9355	113	171	66.0819	129	0.7544
CAH01	160	186	86.0215	54	102	52.9412	103	1.0098
CAH02	67	77	8.0130	132	200	66.0000	129	0.6450
EA01	153	177	86.4407	84	125	67.2000	71	0.5680
EAH01	121	133	90.9774	96	141	68.0851	88	0.6241
EAH02	173	180	96.1111	98	129	75.9690	47	0.3643
KA01	172	172	100.0000	96	114	84.2105	26	0.2281
KA02	126	132	95.4545	32	63	50.7937	123	1.9524
KAH01	147	149	98.6577	82	108	75.9259	45	0.4167
A-Average	137.0000	147.7778	92.5124	87.4444	128.1111	67.4675	84.5556	0.7292
CS01	226	227	99.5595	123	131	93.8931	13	0.0992
CS02	147	151	97.3510	82	113	72.5664	56	0.4956
CS03	116	121	95.8678	80	93	86.0215	16	0.1720
ES01	242	253	95.6522	93	144	64.5833	93	0.6458
ES02	143	146	97.9452	128	141	90.7801	16	0.1135
ES05	103	106	97.1698	69	84	82.1429	21	0.2500
KS01	219	221	99.0950	98	108	90.7407	13	0.1204
KS03	184	186	98.9247	77	91	84.6154	22	0.2418
KS06	306	307	99.6743	133	146	91.0959	17	0.1164
S-Average	187.3333	190.8889	97.9155	98.1111	116.7778	84.0488	29.6667	0.2505

Accuracy Indices for Measuring L2 Oral Production: Examination of Validity and Basis for Future Studies

Masumi Tajima

The indices expressing accuracy of L2 oral production have not yet been established, and no empirical study has been conducted to investigate the validity and reliability of currently available indices. Researchers have not reached an agreement on which indices should be used when analyzing oral output quantitatively, and thus, have employed various and different indices in each study to date. This study investigated the validity of three kinds of accuracy indices, by statistically examining the data of 36 learners of Japanese in KY Corpus, who are at four different proficiency levels. Three indices were selected from three categories of accuracy measures: specific measures and two kinds of general measures, one of which uses the ratio of error-free portion to a total amount of production, and the other of which uses the raw number of errors per a certain unit. The results did not suggest any empirical evidence for the validity of these indices. Based on the results, directions for future research are discussed. Moreover, the need of more research is emphasized to clarify the relationship between the accuracy of L2 utterances and the development of interlanguage, as a premise of the establishment of valid accuracy indices.

Keywords: oral production, accuracy, indices, validity, KY Corpus