

「複雑さ・正確さ・流暢さ」指標の構成概念妥当性の検証：日本語学習者の発話分析の場合

The Construct Validity of the Measures of Complexity, Accuracy, and Fluency: Analyzing the Speaking Performance of Learners of Japanese

桜木ともみ

Tomomi Sakuragi

広島大学大学院教育学研究科

Graduate School of Education, Hiroshima University

It is important for teachers and researchers to be able to assess L2 learners' proficiency through their performance. The measures of complexity, accuracy, and fluency (CAF) have been used for over 30 years in L2 research to analyze language performance (Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Wolfe-Quintero, Inagaki, & Kim, 1998). However, there remain unanswered questions about CAF measures. For example, the length measure (number of words per syntactic unit) needs to be investigated because it has been used inconsistently: Some researchers have used it as a syntactic complexity measure, while others have used it as a fluency measure. Koizumi (2005) pointed out this discrepancy as a serious problem because the interpretation of a single measure varies depending on researchers' orientations. In addition, the results of factor analysis across some studies (e.g., Sheppard, 2004; Tavakoli & Skehan, 2005) showed that the measures of fluency could be divided into

two types: speed and disfluency. These discrepancies in the construct are key issues pertaining to the measures and need to be investigated. Moreover, although CAF measures have often been investigated in Indo-European languages, they have not been sufficiently investigated in other languages; thus, it is important to determine whether CAF measures can be applied to the Japanese language in the same way.

Accordingly, this study examined the construct validity of CAF measures in Japanese as a second language (JSL) from the following three perspectives.

1. As expected, do the measures represent distinct factors of the three CAF dimensions?
2. Does the length measure (the number of words per syntactic unit) have the same construct as the syntactic complexity measure does, rather than having that of the fluency measure?
3. Can the speed measure (the number of words per minute) and the disfluency measure (the number of disfluency markers per minute) be explained as one construct?

To investigate these research questions, 10 general measures were calculated from the narrative production of 113 university-level students learning JSL. Their Japanese language proficiency level ranged from intermediate to advanced on the ACTFL-OPI (American Council on the Teaching of Foreign Languages–Oral Proficiency Interview). Subsequently, a factor analysis was conducted to investigate the construct validity of CAF measures. The initial solution was extracted using the principal factor method, followed by Promax rotation. A three-factor solution was adopted using the Kaiser criterion of eigenvalues greater than one.

The results of this factor analysis indicated the following.

1. The validity of CAF measures was partially demonstrated in terms of syntactic complexity measures (number of clauses per Analysis of Speech unit [AS-unit], and number of subordinate clauses per AS-unit) and accuracy measures (percentage of error-free AS-units, number of errors per AS-unit, and number of errors per clause) but was not demonstrated in terms of lexical complexity measures (number of word types per 100 words, and the Guiraud index) or fluency measures (number of words per minute, and number of disfluency markers per minute).
2. The length measure indicates syntactic complexity because of the high loading (.83) on the same factor with the general syntactic complexity measures.
3. The speed measure and the disfluency measure did not have the same factor as one construct of the fluency, which in turn supports the findings of previous studies (Sheppard, 2004; Tavakoli & Skehan, 2005). The results of this study suggest that further research must be conducted to establish the validity of the fluency measure and the validity of the lexical complexity measure, especially for the Japanese language, which has an agglutinating morphology.

近年、複雑さ (complexity)、正確さ (accuracy)、流暢さ (fluency) の側面から言語運用を分析する指標 (以下、CAF指標) が多くの研究で用いられているが、研究者によって独自の測定方法が用いられており、その妥当性は明らかではない。そこで、本研究では10種類の測定方法について、CAF指標の構成概念妥当性を検証するため、日本語学習者113名の発話データを用いて探索的因子分析を行った。

その結果、構文的複雑さ、正確さについては共通因子による妥当性が示されたが、流暢さは測定方法のタイプ (速さと非流暢さ) によって同じ構成概念を示すとは言えないことが分かった。さらに、構文的な単位に含まれる語数 (長さ) については先行研究で解釈が割れていたが、分析結果では流暢さではなく構文的複雑さを示した。本研究の結果から、流暢さの指標と、日本語の語彙指標に対する検証が今後の課題として得られた。

Keywords: Construct validity, measurement, complexity, accuracy, fluency

学 習者の言語伝達能力を適切に把握することは、教育現場においても研究分野においても重要課題の1つであり、指標はその抽象的な言語伝達能力を実際の言語使用場面から具体的に数値化する手段として用いられている。近年、言語伝達能力を構成する概念については、言語知識だけでなく心理的・生理的メカニズムの側面についての把握が重要であるとの見方から測定方法も多様化し、特に「複雑さ、正確さ、流暢さ (complexity, accuracy, & fluency. 以下、CAF)」に代表される複数の側面を測る研究が増えている (Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Wolfe-Quintero, Inagaki, & Kim, 1998)。今後、言語テストや教育分野への応用のためにもCAF指標を用いた学習者言語研究の増加が予想される。

しかし、CAF指標には重要な課題が残されている (Housen & Kuiken, 2009; Pallotti, 2009)。その1つとして、測定方法の選択と解釈が研究者によって多様である点があげられる。これまでの研究において、「複雑さ、正確さ、流暢さ」という抽象的な能力を、具体的に何の数値を用いて測定しているかは研究者によって多様である。さらに、同一の測定方法を用いても、その結果の数値が「複雑さ、正確さ、流暢さ」のどれを示すものかについて解釈が異なる研究が見られる (Koizumi, 2005)。こういった、指標の使用方法和解釈の不一致は研究結果を左右する重大な問題である。したがって、CAF指標の適切な応用へ向けて、どの測定方法が何を測定しているのか、つまり構成概念妥当性を検証することがCAF指標の重要な課題であるといえる。

そこで、本研究では学習者データ分析へのCAF指標の応用可能性を検討するため、日本語学習者を対象にCAF指標の構成概念妥当性の検証を行うことを目的とする。

CAF指標に関する先行研究と課題

CAFという略称は、2009年の *Applied Linguistics, Special Issue "Complexity, Accuracy, and Fluency in second language acquisition research"* で用いられた比較的新しい用語であるが、学習者のデータを客観的に測定する方法自体は古くから数多くの研究で使用されてきた (e.g., Harrington, 1986; Larsen-Freeman, 1978)。関連する研究には大きく2つのタイプがある。1つは、どの測定方法が第二言語の発達

やレベル差を特徴づけるかを調べた研究(Iwashita, 2006; Koizumi, 2005; Larsen-Freeman, 1978; Wolfe-Quintero et al., 1998)、もう一つは、特定の要因(例えば、プランニング時間の違い等)の影響を測るための従属変数としての妥当性を検証した研究(Mehnert, 1998; Nakakubo, 2011; Tavakoli & Skehan, 2005)である。これらの研究結果から、検証された測定方法が第二言語能力を評価できるかどうか(発達指標としての妥当性)、あるいは、測ろうとする能力を測っているかどうか(構成概念妥当性)について、傾向を知ることができる。

第二言語習得研究では、長い間、発達を示す基準としての指標が模索される中で第一言語習得研究を応用した研究が行われてきた。例えば、Larsen-Freeman(1978)は、第一言語の作文研究における分析単位T-unitを応用し、212人の英語学習者の作文を分析した結果、「誤用のないT-unitの割合」と「誤用のないT-unitにおける語数」という2つの「正確さ」の指標がレベル差を最もよく示したと述べている。T-unit (Minimal terminable unit)とは、1つの主節からなる単位であり、従属節もこれに含まれる(Hunt, 1965)。他にもT-unitを応用した研究は多く、Wolfe-Quintero et al.(1998)は39の先行研究の結果を統合し、作文を対象とした場合の適切な指標を表1のように示している。

表1. Wolfe-Quintero, Inagaki, and Kim (1998, p. 119)による作文のための指標

発達を示す最適な指標	
流暢さ	1 T-unitあたりの語数 1節あたりの語数 誤用のない1 T-unitあたりの語数
構文的複雑さ	1 T-unitあたりの節数 1節あたりの従属節数
語彙的複雑さ	$\sqrt{2 \times \text{延べ語数}}$ あたりの異なり語数 異なり語数あたりの洗練された語数
「誤用」を示す最適な指標	
正確さ	誤用のないT-unitの割合 1 T-unitあたりの誤用数

注. 日本語訳は本研究による

同じく25の作文研究の結果を統合したOrtega(2003)でも、「1 T-unitあたりの語数」「1文あたりの語数」「1節あたりの語数」「1 T-unitあたりの節数」がレベル差を示す測定方法であると結論づけている。「構文的な単位(T-unitや節)あたりの語数(以下、長さ)」の発達指標としての妥当性が示されたことになるが、Wolfe-Quintero et al.(1998)とOrtega(2003)では、同じ長さ指標に対して「流暢さ」と「構文的複雑さ」とい

う異なったラベルが与えられている。つまり、測定方法は同じであるが、何を測っているか(構成概念)が2つの作文を対象とした研究において異なっている。このような解釈の不一致は発話を扱った研究においても見られる。例えば、「長さは従属節などの構文的な複雑さを示唆する」と解釈した研究(e.g., Bygate, 2001; Mehnert, 1998; 瀧口, 2004)に対して、「長さは発話量を反映するため流暢さを示す」と解釈した研究(e.g., 阿野, 2002; Kimura, 1995; Robinson, 2001)があり、同一の測定方法に対して異なる解釈が存在している。Koizumi (2005)はこの解釈の不一致を指標の重大な問題であると指摘し、日本人英語学習者74名の発話を対象に、長さ指標が「流暢さ」と「構文的複雑さ」のどちらであるかを検証した。そして、相関係数による検証の結果、長さ指標は流暢さよりも構文的複雑さを示す指標であると結論づけている。また、探索的因子分析を用いて複数の測定方法を調べた研究(Mehnert, 1998; Nakakubo, 2011; Sheppard, 2004)においても、結果を見ると、長さは「1 T-unitあたりの節数」といった構文的複雑さの指標と同一因子を示しており、流暢さの因子では因子負荷量が低いことが分かる。しかし、長さに関しては、基とする語数のカウント方法や、検証される他の測定方法の種類やデータサイズに影響を受ける可能性も考えられるため、今後の更なる検証が必要であろう。

では、構文的な長さによる測定方法が流暢さの指標として妥当でないとする、流暢さを測るには他にどのような測定方法を用いるのが妥当なのだろうか。流暢さに関しては、測定方法だけでなく、概念の定義づけも課題として残されている。流暢さは、発話や作文における雄弁さや滑らかさなどによって特徴付けられる一般的な言語能力を示すことが多いが、その定義は研究者によって様々である(Housen & Kuiken, 2009)。複雑さや正確さのような言語の形式に関わる側面と対比させる場合、流暢さは「リアルタイムで、意味を重視して目標言語を使用する能力を指す」(Skehan & Foster, 1999, pp. 96-97)とされている。具体的な流暢さの測定方法には、発話量、スピード、ポーズの長さや回数、言い間違いの回数、繰り返し回数、言い直しの回数などがあり、複雑さや正確さと比べて測定方法の種類が多いことが特徴的である。ただ、多様な測定方法全てが「流暢さ」という同一の構成概念を示すといえるかどうか疑問が残る。この点に関して、Tavakoli and Skehan (2005)は12種類の測定方法について80名分の発話データを用いて探索的因子分析を行った結果、流暢さに関する10種類の測定方法は、2つの因子(「速さ・発話量」に関するもの¹と「非流暢さ」に関するもの²)に分かれた。Sheppard (2004)でも、27種類のCAF指標について82名分の発話データを用いて探索的因子分析を行った結果、流暢さに関する16種類の測定方法については3つの因子(「速さ・発話量」に関するもの³、「ポーズ」に関するもの⁴、「非流暢さ」に関するもの⁵)に分かれた。これら2つの研究は、検証した指標の種類やその数に違いがあるものの、両者において「速さ・発話量」「非流暢さ」は独立した因子、つまり異なる構成概念を示す傾向が示唆された。またKoizumi (2005)の結果においても、「1分あたりの語数」と「1分あたりの非流暢さのマーカー」は相関が低いことから、同じ傾向が認められる。これらの結果から、「速さ・発話量」と「非流暢さ」が同じ構成概念を示しているとは断言できず、また、これまで多くの研究で用いられた「流暢さ」の結果は測定方法のタイプによって結果が異なっている可能性がある。測り方の違いによって研究結果に差が生じるということは指標としての重大な問題であるため、流暢さを構成する概念の整理と代表的測定方法の検証が課題とされる。

さらに、「複雑さ」についても分類上の課題が残されている。複雑さは多くの研究で「構文的複雑さ」と「語彙的複雑さ」の2つに下位分類され説明されることが多い。具体的には、構文的複雑さは、節や従属節の割合、特定の文法項目の使用頻度や多様性によって測られ、また語彙的複雑さは、延べ語数における異なり語数の割合や使用頻度が低い語(洗練語)の使用割合など、多数の測定方法が見られる。しかし、構文的複雑さと語彙的複雑さの2つが実際に同一の構成概念を示しているのか、別の構成概念であるのかは明らかではない。Skehan(2009)は、非母語話者の運用場面において、語彙的複雑さは構文的複雑さとは異なる側面であるという見方を示し、語彙を構文的複雑さと区別して調べる必要性を指摘した。Sheppard(2004)の因子分析による結果では、語彙の多様性に関する測定方法(「延べ語数あたりの異なり語数」)は、構文的複雑さと正確さに同程度の因子負荷量を示した。また、日本語を対象としたNakakubo(2011)では、2つの語彙の測定方法、「延べ語数あたりの異なり語数」と「 $\sqrt{2} \times$ 延べ語数あたりの異なり語数」のうち、前者は独立因子を示し、後者は構文的複雑さと共通の因子にまとまっていた。また、2つの測定方法は同じ因子を持たず、相関も見られなかった($r = .11, ns$)。このように、語彙の多様性については、構文的複雑さとの関係や、語彙の多様性を測る指標間の違いなど、明らかにすべき点は多い。語彙の多様性が構文的な能力や他の能力と関係しているのか、または独立しているのかは、学習者の発話能力全体の把握のために重要な課題である。

以上述べたように、CAF指標として使用されてきた測定方法であっても、それが実際に予想する構成概念を示すかどうかは明確ではない。そのため、CAF指標の構成概念妥当性の検証が急務となる。特に日本語においては、これまで学習者のレベルとの関係を調べた研究が多く(e.g., Harrington, 1986; Iwashita, 2006; 田丸・吉岡・木村, 1993)、流暢さを含めた複数の測定方法の構成概念妥当性を検証した研究はNakakubo(2011)の他に見られない。Nakakuboは、タスクのプランニングの違いによる影響を測るための従属変数として代表的な指標を選択することを目的に、143名の日本語学習者の発話データを対象に12種類の測定方法を探索的因子分析と相関係数により分析した。その結果を見ると、流暢さ(速さ・発話量)と、構文的複雑さの指標については構成概念妥当性が認められるが、語彙的複雑さと正確さについては明確な妥当性が認められない。そのため、今後は上記で述べた長さや流暢さの指標の確認も含め、構成概念の妥当性を更に検証していく必要がある。

なお、CAF指標の構成概念妥当性の検証では、相関や因子分析などを用いて検証されることが多いが、分析の際にはデータサイズへの留意が必要である(Sheppard, 2004)。例えば、Mehner(1998)では学習者31人のデータに対し13種類の指標で探索的因子分析を行っているが、データサイズが十分ではなく、それが結果に影響した可能性が考えられる。Sheppard(2004)やTavakoli and Skehan(2005)のように、前提としてデータが因子分析の使用に適しているか検定し報告することが望ましい。

研究課題

先行研究から得られたCAF指標の課題に基づき、本研究では以下の3つの研究課題を設定する。

RQ1: CAF指標は、日本語学習者の発話データを用いて探索的因子分析による妥当性検証を行った場合に先行研究から予測される複雑さ、正確さ、流暢さの

分類に沿った独立因子を示すか。つまり、CAF指標の構成概念妥当性は言語によらず同じように示されるか。

RQ2: 構文的な長さは、構文的複雑さと流暢さのどちらの因子グループと共通因子を示すか。つまり、構文的複雑さと流暢さのどちらを測る指標だといえるか。

RQ3: 流暢さの指標のうち、速さ(1分あたりの語数)と非流暢さ(言い直しなど)は共通の構成概念を示す指標だといえるか。

分析

データ

日本語学習者113名による、ストーリーテリングの独話データ⁶を用いた。被調査者は、海外在住の日本語学習者で、ACTFL-OPI判定基準による中上級(中級84名、上級29名)レベルを対象とした。調査は調査者と被調査者との対面式で行われた。レベルチェックのインタビューの後、合計3種類の発話タスクが行われたが、本研究では3番目に行われた図1の絵を用いた発話タスクのデータを分析対象とした。タスクを行う際、準備時間や発話時間に制限はなかった。音声データは第二言語習得研究を専門とする日本語母語話者によって文字化され、更に別の母語話者1名が再度音声データを聞きながら確認し、一致率99%を得た。

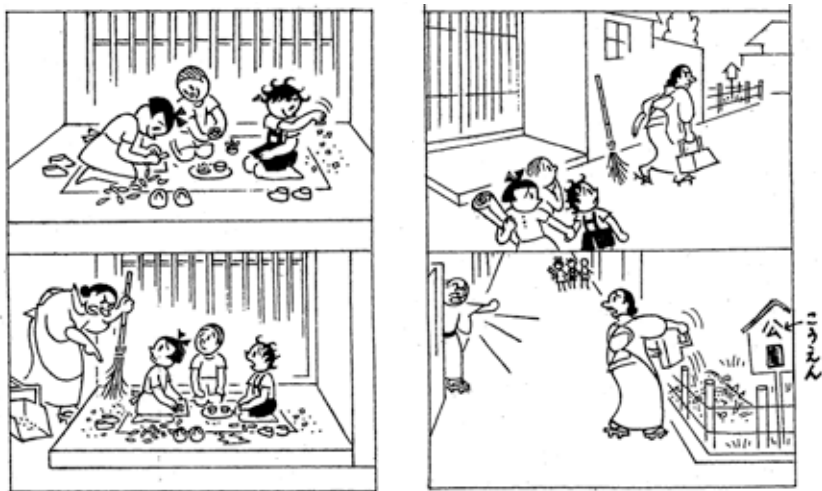


図1. 調査材料の4コママンガ

根本進.(1970).『クリちゃん みどりの本』東京:さ・え・ら書房.

注. 図中の「公園」の記述は本研究者による

コーディング

113名分の文字化資料を用いて、まず基本的な単位となる8項目(AS-unit数、従属節数、節数、延べ語数、異なり語数、非流暢さ数、誤用数、誤用のないAS-unit数)をカウントした。AS-unit(Analysis of Speech unit)、従属節、節の定義は、Foster, Tonkyn, and Wigglesworth(2000)に従った。Foster et al.では、AS-unitの定義は「独立節または節に準ずるものから成る発話であり、従属節もこれに含まれる」(p. 367)とされている。本研究では日本語の特徴を考慮し、「テ形接続」で文が続く場合の判断基準について、Foster et al.の接続(coordination)の扱いを参考に、主語が同じ場合は1つのAS-unitとし、主語が異なる場合は別のAS-unitとしてカウントした。例を以下に示す。

(例1)韓国語母語話者K3の発話例(一部)

K3: | そして子供たちは部屋を出て、他の場所に行って |、| お母さんはどこかでそのごみを持って、捨てるつもりで::どこかへ行きます |

(| AS-unit | =2, 節 = 5, 従属節 (::) = 1)

語数については、本研究では田丸・吉岡(1994)を参考に自立語のみをカウントの対象とした。自立語の認定については南(1998)を参考に、文節を作ることができるかどうかで判断した。形態素レベルの語数としなかった理由については、(a)膠着語である日本語の場合、英語と異なり、語と語の間にスペースなどの切れ目がなく、自立語(名詞など)と付属語(活用語尾など)が連結しているために、形態素レベルのカウントが非常に困難である(田丸・吉岡)ため、コーディングの信頼性が低くなることを考慮した、(b)李(2009)は、現時点の日本語を対象とした形態素解析ツールの使用について、現状では学習者の会話データに対しては十分に妥当な結果を出すまでには至っておらず、その結果を鵜呑みはできないと注意を喚起している、という2点からである。なお、本研究での語数は、Koizumi(2005)等で用いられた「非流暢さを示す発話やフィラーなどの発話部分を除いた語数」とし、その総数を延べ語数とした。これに対し、異なり語数は語のタイプ数とし、同じ語の繰り返しがある場合は1とカウントした。例えば上記のK3の発話例の「行って／行きます」は1とカウントされる。

また、非流暢さについては、Koizumi(2005)を参考に、非流暢さを示すマーカー(言いよどみ、言い直し、意味のない繰り返し)をカウントした。

誤用数については、語彙、文法、語順、表現等、意味伝達に関わる誤用を対象とし、アクセントやイントネーションの間違いは含めなかった。

判定の信頼性については、2名の判定者によってデータの約半数(53名分)のコーディングを行い、98%以上の一致率が得られた。

その後、本研究で検証する測定方法10種類の数値を算出した。表2に指標の種類を、表3に記述統計量を示す。

表2. 本調査で検証した測定方法

測定方法の種類	コード	先行研究でのカテゴリー
1 1 AS-unitあたりの節数	[SC1]	構文的複雑さ
2 1 AS-unitあたりの従属節数	[SC2]	
3 100延べ語数あたりの異なり語数	[LC1]	語彙的複雑さ
4 ギロー指数 ($\sqrt{\text{延べ語数あたりの異なり語数}}$)	[LC2]	
5 誤りのないAS-unitの割合	[A1]	正確なunitの数
6 1 AS-unitあたりの誤用数	[A2]	誤用数
7 1節あたりの誤用数	[A3]	
8 1分あたりの延べ語数	[F1]	速さ
9 1分あたりの非流暢さ(言い間違い、繰り返し、自己訂正)の数	[F2]	非流暢さ
10 1 AS-unitあたりの語数	[SC3 or F3]	構文的複雑さか、流暢さか、先行研究で意見が割れている指標: unitの長さ

注. 各指標のコード([SC]など)のアルファベットの意味は、SCが構文的複雑さ、LCが語彙的複雑さ、Aが正確さ、Fが流暢さを示す

表3. 記述統計量

	平均値	標準偏差
基本単位		
AS-unit数	5.91	2.41
節数	11.29	5.04
従属節数	3.42	2.70
延べ語数	36.26	14.51
異なり語数	25.35	8.77
誤用数	3.00	2.45
誤りのないAS-unit数	3.61	2.12
非流暢さの回数	5.14	3.80
時間(秒)	76.54	40.16
指標		
[SC1]	2.10	1.17
[SC2]	0.69	0.76
[SC3 or F3]	6.65	2.96
[LC1]	71.83	8.85
[LC2]	4.19	0.64
[A1]	0.60	0.27
[A2]	0.53	0.43
[A3]	0.27	0.21
[F1]	31.99	12.09
[F2]	4.02	2.24

N = 113

測定方法の選択理由は以下の通りである。

まず複雑さについて、構文的複雑さは一般的な測定方法である「1 unitあたりの節数[SC1]」「1 unitあたりの従属節数[SC2]」とした。語彙的複雑さについては、先行研究(Koizumi, 2005; Nakakubo, 2011; Sheppard, 2004)を参考に、語彙の多様性を示す「100延べ語あたりの異なり語数[LC1]」と、「ギロー指数[LC2]」(Koizumi, 2005; 小泉, 2009; Vermeer, 2000)を選択した。「ギロー指数[LC2]」は延べ語数の平方根を用いるもので、「異なり語数[LC1]」よりもテキストの長さに比較的影響を受けないといわれる。

また、長さが構文的複雑さと流暢さのどちらを示すものかを検証するために(RQ2)「1 AS-unitあたりの語数[SC3 or F3]」を含めた。

正確さについては、Wolfe-Quintero et al. (1998)を参考に、最も一般的とされる「誤用のないunitの割合[A1]」と「1 unitあたりの誤用数[A2]」、「1節あたりの誤用数[A3]」を選択した。特定の文法項目(例えば、助詞や活用形など)の正用率はタスクやレベルに影響を受ける可能性があるため今回は扱わなかった。

最後に、流暢さについては、「速さ」と「非流暢さ」の2つのタイプが「流暢さ」として1つの因子を示すかを確認するため(RQ3)、先行研究(Koizumi, 2005; Sheppard, 2004; Tavakoli & Skehan, 2005)を参考に、速さに関する「1分あたりの延べ語数[F1]」と、非流暢さに関する「1分あたりの非流暢さ数(言いよどみ、繰り返し、自己訂正など)[F2]」を選択した。

分析手順

先行研究(Nakakubo, 2011; Sheppard, 2004; Tavakoli & Skehan, 2005)を参考に、以下の手順で分析を行った。(a) データが因子分析の使用に適しているか検定し、確認する、(b) 探索的因子分析を行い、得られた因子グループとこれまで各測定方法が示すと考えられてきたCAFの構成概念が一致するかどうかを確認する(RQ1, RQ2, RQ3)、(c) 因子分析の結果から特定の指標間に明確な関係が見られなかった場合は、特定の指標間の相関を確認する(RQ2, RQ3)。

分析に使用したソフトは、SPSS 16.0J for Windowsであった。

結果

まず、本調査で用いたデータが因子分析に適しているかをKaiser-Meyer-Olkin (KMO)の標本妥当性の測度検定、及びBartlettの球面性検定を用いて確認した。その結果、KMOは.66であり、基準とされる.60(“mediocre”: Kaiser, 1974)に達していた。また、Bartlettの球面性検定では有意性が認められた($\chi^2 = 850.14$, $df = 45$, $p < .01$)。このことから、本調査データはSheppard (2004)やTavakoli and Skehan (2005)と同じく、より多くのサンプルサイズでの追検証が望ましいレベルではあるが因子分析を行うには問題がないと判断し、主因子法・プロマックス回転による探索的因子分析を行った。その結果、因子数は、カイザー-ガットマン基準に基づき固有値1.00以上の3因子解が抽出された(表4参照)。因子相関行列を見ると、3つの因子間に相関は見られず、因子が独立していることが確認された。次に、各因子の内的一貫性を調べるためCronbachの α 係数を調べたところ、第1因子では.77、第2因子では.85であり、あ

程度の信頼性が得られた。しかし、第3因子では2つの項目しか含まれないこともあり.08と非常に低く、信頼性が得られなかった。そのため、本研究では第2因子までを得られた因子として認め、結果を分析した。

表4. 因子分析結果

	I	II	III	共通性
[SC1]	1.02	-0.04	-0.05	1.00
[SC2]	0.96	-0.02	0.02	0.93
[SC3 or F3]	0.83	0.03	0.07	0.72
[A3]	-0.35	0.93	-0.04	0.89
[A1]	-0.13	-0.84	-0.08	0.75
[A2]	0.33	0.83	-0.07	0.93
[F1]	0.05	-0.05	0.70	0.52
[LC2]	-0.01	0.12	0.60	0.34
[F2]	-0.08	0.27	0.24	0.09
[LC1]	0.16	-0.17	-0.11	0.04
因子寄与率	32.27	21.89	7.97	
因子相関行列				
I	-	0.19	0.19	
II		-	-0.23	
III			-	

注. 因子抽出法は主因子法、回転法はプロマックスによる
枠線で囲った数値は、因子負荷が0.50以上であることを示す

第1因子は、3種類の測定方法「1 AS-unitあたりの節数[SC1]」「1 AS-unitあたりの従属節数[SC2]」「1 AS-unitあたりの延べ語数[SC3 or F3]」において高い因子負荷量(1.02～0.83)が得られたことから「構造的複雑さ」因子と解釈された。このことから、RQ2に関して先行研究で構成概念の解釈が割れていた「1 AS-unitあたりの延べ語数[SC3 or F3]」は、本データにおいては一般的な構文的複雑さの指標([SC1][SC2])と共通の構成概念を測る測定方法だといえる。

第2因子は、正確さに関する3種類の測定方法で構成された。「誤りのないAS-unitの割合[A1]」で-0.84、「1 AS-unitあたりの誤用数[A2]」で0.83、「1節あたりの誤用数[A3]」0.93の因子負荷量が得られた。このことから、第2因子は「正確さ」因子と解釈された。

以上の因子分析で得られた2つの因子から、RQ1の結果として、6つの測定方法(SC1, SC2, SC3 or F3, A1, A2, A3)については予想された構成概念による独立因子が確認され、妥当性が示された。しかし、残りの4つの測定方法(LC1, LC2, F1, F2)については、今回の因子分析では先行研究から予想された「流暢さ」や「語彙的複雑さ」といった構成概念のまともには確認できなかった。

さらに、項目間の相関係数から見ても、流暢さを示すと予測された「1分あたりの延べ語数(F1)」と「1分あたりの非流暢さの数(F2)」の相関は $r = .06$ と非常に低く、同じ構成概念を測っているとは考えにくい結果となった(RQ3)。また、語彙的複雑さを示

すと予測された「100延べ語数あたりの異なり語数(LC1)」と「ギロー指数(LC2)」の相関は $r = .02$ と非常に低かったが、その一方でLC2とF1に中程度の相関($r = .42, p < .01$)が見られた。このように、語数を用いた測定方法(SC3 or F3, LC1, LC2, F1)のうち、LC1とLC2よりもLC2とF1の相関の方が高いという、予想に反する結果となった。

表5. 相関行列

	[SC1]	[SC2]	[SC3 or F3]	[LC1]	[LC2]	[A1]	[A2]	[A3]	[F1]	[F2]
[SC1]	-	.95**	.85**	.16	.12	-.26**	.48**	-.19*	.14	.01
[SC2]		-	.83**	.12	.14	-.26**	.44**	-.19*	.18	.01
[SC3 or F3]			-	-.04	.11	-.32**	.38**	-.17	.19*	.03
[LC1]				-	.02	.08	.00	-.17	-.03	-.13
[LC2]					-	-.01	.02	-.07	.42**	.14
[A1]						-	-.81**	-.68**	.08	-.23*
[A2]							-	.71**	-.16	.15
[A3]								-	-.26**	.12
[F1]									-	.06
[F2]										-

$N = 113, ** p < .01, * p < .05$

考察

以下、CAFのカテゴリーごとに妥当性検証の結果をまとめ、考察を行う。

構文的複雑さ

構文的複雑さを示すとされる2つの測定方法(SC1, SC2)、及びRQ2に関連する長さ(SC3 or F3)は、第1因子において3種類とも高い因子負荷量を示し、更にこれらの3つの指標は他の因子では因子負荷量が低い(全て.07以下)ため、10種類の指標においては構文的複雑さを示す指標として妥当性が期待できる測定方法だといえよう。

この結果は、中級ドイツ語学習者を対象としたMehnert(1998)や、初中級英語学習者を対象としたSheppard(2004)、Koizumi(2005)と共通しており、中級レベルの学習者の発話データについては、1 unitあたりの節数や語数が構文的複雑さの指標として、言語の違いを越えて汎用性の高いことが示唆された。

正確さ

正確さに関する3つの測定方法(A1, A2, A3)については、3つとも同一因子において高い負荷量を示し、妥当性が示された。この結果は、Housen and Kuiken(2009)の「正確さは明確で一貫性のある指標である」という指摘を、構成概念妥当性の側面において支持するものだと見える。

流暢さ

流暢さについて、本研究では「速さ(1分あたりの延べ語数)」と「非流暢さ(1分あたりの言いよどみ、繰り返し、自己訂正などの非流暢さ数)」について検証した結果、2つの測定方法は共通因子を持たず、同じ構成概念を持つとは言えないことが分かった(RQ3)。この結果は、同じく絵の描写タスクによる発話データを用いたTavakoli and Skehan(2005)やSheppard(2004)と共通しており、絵の描写タスクのように、独話でストーリーを説明するような場面においては、英語学習者でも日本語学習者でも速さと非流暢さは異なる現れ方をすることが示唆される。このことから、流暢さとして使用されてきた指標のうち、速さと非流暢さのどちらのタイプを用いるかによって、流暢さとして得られる結果が異なる可能性が高い。これは研究結果の解釈や研究間の比較の際に大きな問題となる。したがって流暢さについては、概念自体が広いうえ、言語能力に関わる様々な要素を含みうるという点を考慮し、流暢さに関わる要素の「何」を測るかによって測定方法を使い分ける必要がある。指標としての妥当性という点からは、今後は本研究で扱わなかったポーズなども含め、流暢さに関する多様な測定方法がタイプごとに独立した因子を示すかどうか、更なる検証が必要である。

語彙的複雑さ

最後に、語彙的複雑さについて、今回の検証では「100延べ語数あたりの異なり語数[LC1]」と「ギロー指数[LC2]」を用いたが、他の構成概念との混同した結果が得られた。

まず、複雑さの枠組みで構文的複雑さとの関係について見てみると、今回の分析結果ではLC1とLC2は、構文的複雑さ(SC1, SC2, SC3 or F3)とは共通因子を持たなかったことから、同じ構成概念を示すとは考えにくい。一方Sheppard(2004)やNakakubo(2011)では語彙的複雑さの指標は部分的に構文的複雑さとの共通性が示されており、構文的複雑さとの関係については全体的には統一した結果が得られていないといえる。

また、同じ語彙の多様性を示すはずのLC1とLC2についても予想に反して相関が見られなかった($r=.02, ns$)。本研究と同じく日本語学習者の発話を分析したNakakubo(2011)でも語彙の多様性を測る2つの測定方法(「延べ語数あたりの異なり語数」と「 $\sqrt{2 \times \text{延べ語数あたりの異なり語数}}$ 」)の間に相関が見られなかった($r=.11, ns$)。これらの結果は、「延べ語数における異なり語数の割合」という一般的に使用されている測定方法が、その計算方法の違いによって結果が大きく異なることを示すとともに、他の構成概念との関係が明確ではないことが分かる。その理由として、発話データの長さやタスクの違いが影響していることが考えられる。一般的な異なり語数の指

標は、テキストの長さに影響を受けやすい(Vermeer, 2000)という点をふまえると、本研究においても選択した指標(LC1とLC2)の関係や、発話データの特徴が影響した可能性は否定できない。今後、指標の種類や、発話内容、データサイズ等を変えて再検証を行うことでより明確な結果が得られると考えられる。その際、Koizumi(2005)やNakakubo(2011)のように、タスクの違いに関する結果が蓄積されることは指標の妥当性を多角的に検証するためにも重要であるとする。また、語彙研究全体の課題として、膠着語である日本語において形態素による語数の測定をいかに正確にできるかが実践上の課題として残されている。

まとめと課題

CAF指標の構成概念妥当性を調べるため、10種類の測定方法について日本語学習者113名の発話データを用いて探索的因子分析による検証を行った。その結果、6種類の測定方法については、構文的複雑さ、正確さの指標としてそれぞれ妥当性が示されたが、4種類の測定方法については予測と異なる結果となり、語彙的複雑さと流暢さの指標としての妥当性が示されなかった。CAF指標を用いた先行研究と合わせて考えると、妥当性が得られなかった語彙的複雑さと流暢さは、今後構成概念の定義自体の見直しや、その測定方法の検証が課題として残されている。また、これまでの研究結果を適切に解釈し比較するために、CAF指標の留意点として、流暢さが測定方法のタイプ(「速さ」または「非流暢さ」)によって異なる可能性があること、「構文的な単位(T-unitや節)における語数(長さ)」が流暢さではなく構文的複雑さを示すことなどが指摘できる。

以上は、本研究が検証したデータや測定方法による限られた結果であり、今後、CAF指標の検証が蓄積されることにより、語彙的複雑さや流暢さに関してより妥当な測定方法が得られると考えられる。学習者の能力の特徴や変化を適切に把握するために、CAF指標の応用に向けて今後も引き続き研究が必要である。

注

1. speech rate, total silence, time spent pause, no. of pauses, mean length of pauseの5種類を含む。
2. reformulations, false starts, replacements, repetitionsの4種類を含む。
3. 1分あたりの語数、1分あたりの非流暢さを除いた語数、1分あたりの音節数、1分あたりの非流暢さを除いた音節数の4種類を含む。
4. Mean length of run (MLR)。発話の切れ目となるポーズ時間を0.25秒、0.60秒、1秒とする3つのMLRを用いた。
5. 繰り返し、言い直し、フィラーに関する測定方法で、それぞれ1分またはT-unitあたりの数で算出される計6種類を含む。
6. 本研究で分析する調査データは、広島大学の迫田久美子先生のサコダコーパス及び広島国際学院大学の川崎千枝見先生の調査データの一部であり、本研究において分析する許可をいただいた。調査は、第二言語習得研究を目的とし、2008年1月から3月に、韓国・台湾・タイの大学機関で日本語を外国語として

学習している大学生118名を対象に、第二言語習得研究・日本語教育を専門とする4名の調査者によって実施された。本研究では初級の5名を除く113名を分析対象とした。

謝辞

本論文の作成にあたり、資料提供をはじめ終始ご助言をいただいた広島大学の迫田久美子先生、広島国際学院大学の川崎千枝見先生に感謝の意を表す。

桜木ともみ(さくらぎともみ)は、現在広島大学大学院教育学研究科博士課程に在籍し、日本語教育を専攻し、主に第二言語習得や評価に関する研究に取り組む。非常勤講師として広島修道大学及び福山大学で留学生のための日本語教育に携わる。

引用文献

- 阿野幸一. (2002). 「高校生英語学習者の発話における流暢さと正確さの関係」*STEP Bulletin*, 14, 39-49.
- 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). Essex, UK: Pearson Education Limited.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354-375.
- Harrington, M. (1986). The T-unit as a measure of JSL oral proficiency. *Descriptive and Applied Linguistics*, 19, 49-56.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30, 461-473.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. NCTE Research Report No. 3. Champaign, Ill: National Council of Teachers of English.
- Iwashita N. (2006) Syntactic complexity measures and their relation to oral proficiency in Japanese as a foreign language. *Language Assessment Quarterly*, 3, 151-169.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kimura, Y. (1995). An analysis of English communicative competence of high school returnees and non-returnees. *Annual Review of English Language Education in Japan*, 6, 123-133.

- Koizumi, R. (2005). Speaking performance measures of fluency, accuracy, syntactic complexity, and lexical complexity. *JABAET Journal*, 9, 5-34.
- 小泉利恵. (2009). 「常磐大学国際学部の学生の英語スピーキングの特徴: 日本人英語上級者と英語母語話者との比較から」『常磐国際紀要』, 13, 53-70.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12, 439-448.
- 李在鎬. (2009). 「タグ付き日本語学習者コーパスの開発」『計量国語学』, 27, 60-72.
- Mehnert, U. (1998). The effect of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 52-83.
- 南雅彦. (1998). 「第5章分かち書き」. In Y. Oshima-Takane, B. MacWhinney, H. Sirai, S. Miyata, & N. Naka (Eds.), *CHILDES Manual for Japanese* (pp. 39-43). Montreal: McGill University. Retrieved from <http://www.cyber.sist.chukyo-u.ac.jp/JCHAT/JChatman.pdf>
- Nakakubo, T. (2011). *The effects of planning on second language oral performance in Japanese: Processes and production*. Available from ProQuest Dissertations and Theses database. (UMI No. 3461207)
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24, 492-518.
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30, 590-601.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22, 27-57.
- Sheppard, C. (2004). The measurement of second language production: The validity of fluency, accuracy, and complexity. 『語学研究』, 19, 139-156.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510-532.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93-120.
- 瀧口均. (2004). 「日本人EFL中学生のスピーキング能力の発達研究－「流暢さ」「複雑さ」「正確さ」の指標を用いて－」『関東甲信越英語教育学会研究紀要』, 18, 1-13.
- 田丸淑子, 吉岡薫, 木村静子. (1993). 「学習者の発話に見られる文構造の長期的観察」『日本語教育』, 81, 43-54.
- 田丸淑子, 吉岡薫. (1994). 「日本語発話資料分析の単位をめぐる問題－第二言語習得仮定観察の立場から－」*The Language Programs of the International University of Japan- Working Papers*, 5, 84-100.

- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-277). Amsterdam: John Benjamins.
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*, 65-83.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawai'i.

