

改訂版・構造分類による日本語文法知識テストの開発 —中国人日本語学習者のデータによるテスト評価—¹

早川 杏子²

玉岡 賀津雄³

要約: 宮岡・玉岡・酒井(2014)が開発した調査・研究用の文法知識テストを改訂し, 当該のテストの信頼性を検討した。改訂にあたって, 機能語が語彙と文法どちらの知識に帰属すべきかを確認するために, 確認的因子分析を用いて, 2つのモデルを検討した。第1モデルは機能語を語彙知識に含んだ場合, 第2モデルは, 機能語を文法知識に含んだ場合である。分析の結果, 第1モデルの適合度が高く, 機能語は語彙知識に含まれるべきであることを明らかにした。これを受けて, 改訂版では機能語のカテゴリーは文法知識に含まず, 形態素変化 (morphological inflections), 局所依存 (local independency) および構造的複雑性 (complex structure) の3つの構造分類からなる各12問の合計36問の改訂版テストを作成した。中国人日本語学習者143名を対象に改訂版の文法知識テストを実施したところ, クロンバック信頼度係数 ($\alpha=.820$) は非常に高く, 本テストが, 日本語学習者の文法知識を簡便かつ安定的に測定することができることを確認した。

キーワード: テスト評価 機能語 文法知識 語彙知識 確認的因子モデル

¹ English title: Development of the revised version of the structure-based Japanese grammatical knowledge test: The evaluation by data from native Chinese speakers learning Japanese

² HAYAKAWA, Kyoko; hayakawa@kwansei.ac.jp

³ TAMAOKA, Katsuo; ktamaoka@lang.nagoya-u.ac.jp

1. はじめに

第2言語習得研究における関心は、学習者が目標言語を習得していくプロセスを明らかにすることである。その過程で、学習目標の言語知識の向上にともなって段階的に発達していく項目もあれば、母語の知識を援用させ、目標言語の知識量や学習時間とは関係なく、容易に一定の習熟度に達する言語項目もあると思われる。たとえば、教師や研究者にとって、学習者の母語の知識や第2言語そのものが持つ習得の難しさの諸要因を区別することは、言語の理解のメカニズムを解明する上で、非常に重要である。特に、第2言語習得研究の目的が現場での教授・学習への提言にあると考えれば、このような知見の蓄積は、効果的なカリキュラムを組む上での基本情報となろう。

第2言語習得の調査や実験では、研究者が仮定した要因を検討することが多い。しかし、目的とした課題を実施しただけでは、第2言語習得の研究としては不十分である。学習者の第2言語習得の過程の全体の様相をとらえるためには、どのような学習段階に、どのような傾向が見られるのかを観察する必要がある。仮定した諸要因と目標言語の言語知識との関連性を明確化するために、調査や実験と並んで、被験者の第2言語の知識をきちんと測定して統制しておかねばならない。そこで、日本語能力を統制するために、簡便かつ効率良く実施できるテストが不可欠である。本研究は、こうした研究上の希求性から、これまでに開発された調査・実験用の日本語能力文法テストを改訂し、日本語学習者のデータによって本テストの信頼性の観点から評価する。

すでに開発されている調査・実験用の日本語能力テストは、語彙知識を語種にしたがって、和語、漢語、外来語、機能語の4つに分類して、難易度を統制した語彙知識テスト(宮岡・玉岡・酒井, 2011)と、文法知識を形態素変化、局所依存および構造的複雑性の観点から作成した構造分類に従った文法知識テスト(宮岡・玉岡・酒井, 2014; 2006・2008年度版)がある。宮岡・玉岡・酒井(2014)では、2010年から2012年版では、「～に至るまで」のような機能語を文法知識として扱っている。これは、日本語能力試験の出題基準(国際交流基金・日本国際教育協会, 2002)と同じ分類である。この機能語の扱いをめぐっては、日本語の能力を総体的に測定する上で、語彙と文法のどちらの知識に帰属するかを明らかにしておかなければならない。したがって、2008年に実施された文法知識および語彙知識のテスト(宮岡・玉岡・酒井, 2011)の受験者データを使って、それを検証することにした。方法として、機能語を文法知識あるいは語彙知識に含んだモデルを想定し、確認的因子

分析を行う。

さらに、確認的因子分析の結果を受けて刷新した文法知識テストの内容について詳述する。改訂にあたって、以前の版から削除した項目、新たに追加した項目、または修正した項目を挙げる。さらに、2014年9月に中国の2つの教育機関で行った受験者データに基づいて本改訂テストの信頼性および問題項目の分析結果を記述し、本改訂版の構造分類による日本語文法知識テストの評価を行う。最後に、従来の文法テストの結果と比較し、調査・実験用のテストとしての有用性について述べる。

2. 文法テストの構成

文法知識を測定するテストには、宮岡・玉岡・酒井 (2014) によって開発された2006年版および2008年度版がある。まず、その概要について説明する。この文法テストは、形態素変化、局所依存および構造の複雑性の3つの観点から、学習者の日本語の文法知識を測定している。宮岡・玉岡・酒井 (2014) は、文法能力を図1に示したように「一つのまとまった文を理解するのに必要な単語内外の一連の規則に関する知識」と定義している。形態素変化は、一つの単語内における形態的变化や助動詞との複合的接続と配列に関する知識であり、「形態素」+「形態素」で一つの単語を作るレベルの規則である。局所依存は、隣接する「単語」+「単語」の適切な共起関係の知識である。最後に、構造的複雑性で、英語では *complex structure* と呼ばれるため、厳密には「複雑な構造」であるが、日本語訳では、「構造の複雑性」としている。一つの文の中で、離れて位置する単語間の相互関係に関する知識である。これら3つの関係は、すべて1つの文の中で生じる規則性であり、これら3つを文法知識の下位カテゴリーとした。以下、それぞれの下位カテゴリーについて実際のテスト問題を例に説明していく。

第1の下位カテゴリーである形態素変化に関するテスト問題では、「さっきの説明は()。」という文に対し、「ややこしくありませんでしたか」「ややこしくありませんでしたか」「ややこしいじゃなかったですか」「ややこしいじゃないでしたか」の4つの選択肢から、正しいものを一つ選ばせるものである。この場合、「ややこしい」という単語が否定及び過去を示す場合に、どのような形態素に変化させるのが適切であるのかという知識を問う問題である。日本語では、形容詞否定の形態素は、*-arimasen* と *-nai* の形があり、過去の形態素では *-deshita* と *-ta* の2つがある。前者と後者の形態素は、「* ありませんた」や「* ないでした」の

ように両者を組み合わせて使うことはできない。さらに、形容詞の否定形の *-nai* は、文末に *-desu* を付加しなければ丁寧体とはならない。したがって、提示文の単語のイ形容詞否定過去の形態素変化の正しい形は、語形変化部分の「い」を「く」に変え、「ややこしくありませんでした」か、「ややこしくなかったです」のいずれかということになる。ここでは、単語の品詞を判別した上で、助動詞における複合的接続形の適切さの知識が問われる。

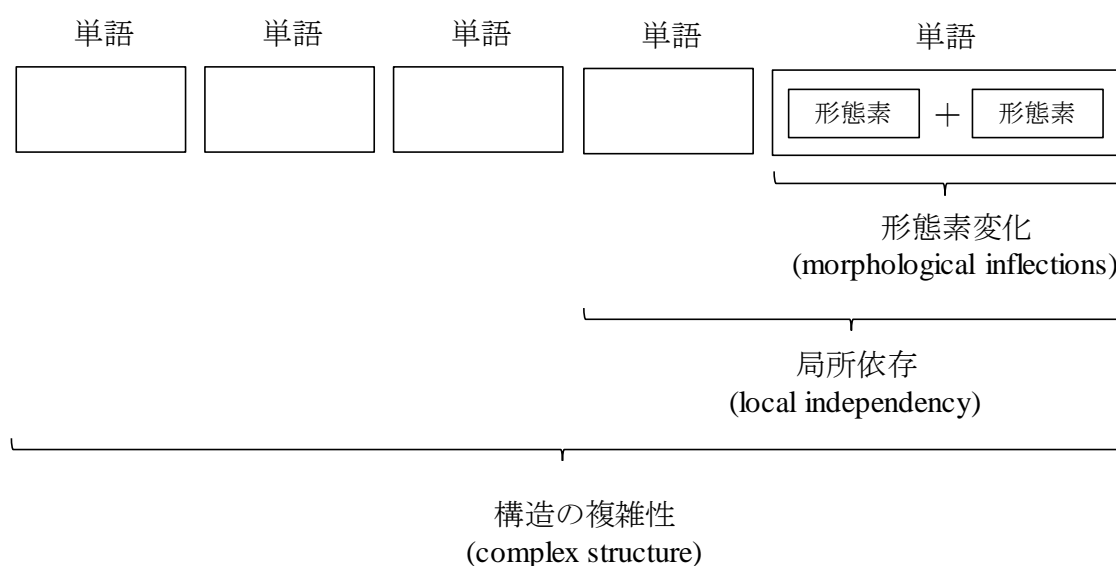


図 1 文法知識を構成する3つの構造分類 (宮岡・玉岡・酒井, 2014, p. 34, 図1を転載)

第2の下位カテゴリーの局所依存のテスト問題の例としては、「私は()人が好きだ。」の「人」という名詞に隣接する形容詞が、「かっこいい」「かっこいいの」「かっこよく」「かっこいいな」のどれがふさわしいかを判断させる問題が挙げられる。これは名詞修飾の問題であり、修飾される名詞の前に置かれる品詞はイ形容詞なので、原形の「かっこいい」が適切な答えとなる。学習者は、まず修飾する単語(かっこいい)の品詞(イ形容詞)を判別し、修飾される単語(人)が名詞であることが判断できた上で、はじめて適切な形の選択が可能となる。「かっこいい」はイ形容詞なので、「の」や「な」の付加はできない。そして、イ形容詞の正しい活用は、「かっこいい」と「かっこよく」であるが、後者は形容詞あるいは動詞が非修飾語となる場合の活用なので、選択肢から排除される。形態素変化と異なる点は、形態素変化の単語の活用が単語内で独立しており、隣接する単語に影響を受けない点である。その点、局所依存は、「かっこいい人」のように、隣接する単語との関係によって活用の形の選択に複数の可能性があり、隣接の単語は相互依存的である。

第3の下位カテゴリーの構造的複雑性のテスト問題としては、「彼女がいつ日本に行くと()、田中さんに聞いてみてください。」のように、共起的制約の知識を問うものなどがある。括弧内に「思う」「思うか」「思うかどうか」「思いますか」のいずれか一つを選ぶ問題などがある。日本語では、疑問詞に係る句末には、疑問助詞「か」を付加し、疑問詞のない句には「かどうか」を付加しなければならない。局所依存と異なる点は、局所依存が隣接による結合関係であるのに対し、構造的複雑性は、単語が呼応を要求し、結合関係にある単語同士が文中で離れた位置関係にある。

2006年および2008年度版における宮岡・玉岡・酒井(2014)の文法テストは、これら3つの下位カテゴリーが、各12問ずつ配置され、合計36問から構成されている。日本語学習者の文法に関するテストであるという特性が考慮され、難しい語彙は使わず、旧・日本語能力試験の配当級で4級から3級の初級段階で学習する語彙項目のみが使用されている。

3. 文法テスト改訂の背景

冒頭で触れた通り、調査や実験において、学習者の日本語の語彙知識を測定するテストについては、すでに宮岡・玉岡・酒井(2011)によってその内容とテスト評価が報告されている。この語彙テストのクロンバックの信頼性係数(α)は、2006年は0.737であったが、2008年は0.896と非常に高い信頼性を示した。また、数々の調査や実験で実際に日本語学習者の能力統制のためのテストとして使用されている実績がある(初・玉岡・早川, 2013; 早川・玉岡, 2012; 大和・玉岡, 2013a, 2013b; 大和・玉岡・初, 2013)。

一方、日本語学習者の文法知識を測定するテストも、上記の語彙テストと共に作成され、実施されている。しかし、文法テストのクロンバックの信頼性係数は、2006年が0.795であり比較的高いが、2008年では0.733であった。また、2010年から2012年には3度目の改訂を行ったが、このテストは、李・宮岡・林(2013)の研究に使用する目的で改訂されたものである。李・宮岡・林(2013)では、学習者作文コーパスデータが日本語能力と相関するかどうかを検討するために、2008年実施の文法テストから、形態素変化・局所依存・構造的複雑性の項目各4問を除外し、文法の構成概念に、以前までは語彙知識にあたるとしていた宮岡・玉岡・酒井(2011)の語彙テストの下位項目である機能語12問の問題を含んだ。この2010年～2012年度版文法テストは、2008年改訂版の文法テストの項目を減らして機能語の項目をそのまま使用したため、問題数は36問と変わらない。結果的に、3回の調査をし

ているが、クロンバックの信頼度係数はちょうど 0.800 であった。ただし、機能語を語彙テストと併用した場合には、両方に同じ分類の項目が含まれることになり、構造方程式モデリングなどで、語彙と文法の知識を含む因果関係モデルを検討する場合には、文法テストの中の機能語が使えないことになる。

2010 年から 2012 年度の改訂の際に、機能語を文法テストに組み込んだのは、『日本語能力試験 出題基準』(2002) では、機能語が文法類のリストとして列挙されており、近年の「日本語能力試験」でも、言語知識の「文字・語彙」にではなく「文法・読解」のほうで出題されている傾向に準拠したためである。ただし、『日本語能力試験 出題基準』(2002) でも、「機能語が語彙、文法のどちらの類に属するかの境界を決めることは簡単ではない」(p. 158) と説明しており、その上で機能語を文法知識に含んでいる。機能語が文法知識を反映しうるのか、あるいは語彙知識に帰属させるべきなのかについて、実証的に検討したわけではない。そこで、機能語が文法知識と語彙知識のどちらに帰属する知識であるのかを確認するために、文法知識、語彙知識のそれぞれの潜在変数に対し、機能語を観測変数とした場合におけるモデルの適切性を確認的因子分析によって検討する。

4. 確認的因子分析

機能語が語彙知識あるいは文法知識に含まれるべきなのかを検討するために、確認的因子分析を行う。確認的因子分析とは、潜在変数を構成する観測変数が、どのくらい概念構成に貢献しているかを検討するための統計的手法である。本研究では、機能語が語彙知識に含まれた場合と文法知識に含まれた場合の両方のモデルを比較検討する。

4.1 データ概要

対象とするのは、西安外国語大学で同時に実施した 2008 年度版の語彙知識ならびに文法知識のテストに参加した日本語学習者 224 名のデータである。被験者は、3年生と4年生で、各 112 名ずつで、平均年齢は 21 歳5ヶ月(標準偏差1年2ヶ月)であった。文法テスト(36 点満点)の平均は 27.87 点(標準偏差 4.17 点)で、語彙テスト(48 点満点)の平均は、32.10 点(標準偏差 7.28 点)であった。

4.2 分析方法

本分析では、各テストの下位項目を観測変数として、潜在変数である文法能力あるいは

語彙能力がそれぞれに寄与する因果構造モデルを想定する。第1のモデルは、機能語の理解を語彙知識に含んだモデルである。第2のモデルは、機能語を文法知識に含んだモデルである。

4.3 分析と結果

語彙と文法の知識が機能語に影響する2つのモデルは図2と図3に示したとおりである。2つのモデルを比較した結果、AIC (Akaike Information Criterion) の飽和モデルは 56.000 で、モデル1は 36.773, モデル2は 64.082 であり、モデル2は飽和モデルよりも大きい値であった。AIC にサンプル数の影響を考慮した CAIC (Consistent Akaike Information Criterion) の飽和モデルは 179.526 で、モデル1は 102.948, モデル2は 130.257 とやはりモデル1のほうが値が小さかった。ECVI (Expected Cross Validation Index) の飽和モデルは .251 で、モデル1は .165, モデル2は .287 であった。これもまた、モデル2は飽和モデルよりも値が大きくなっていった。さらに、BIC (Bayes Information Criterion) の飽和モデルは 151.526 で、モデル1は 87.948, モデル2は 115.257 であった。モデル1の方がモデル2よりも値が小さく、より適したモデルであった。以上のように、AIC, CAIC, ECVI, BIC の4つのモデル比較指標すべてにおいて、図2のモデル1の方が図3のモデル2よりも本研究で得たデータに適合したモデルであることが示された。

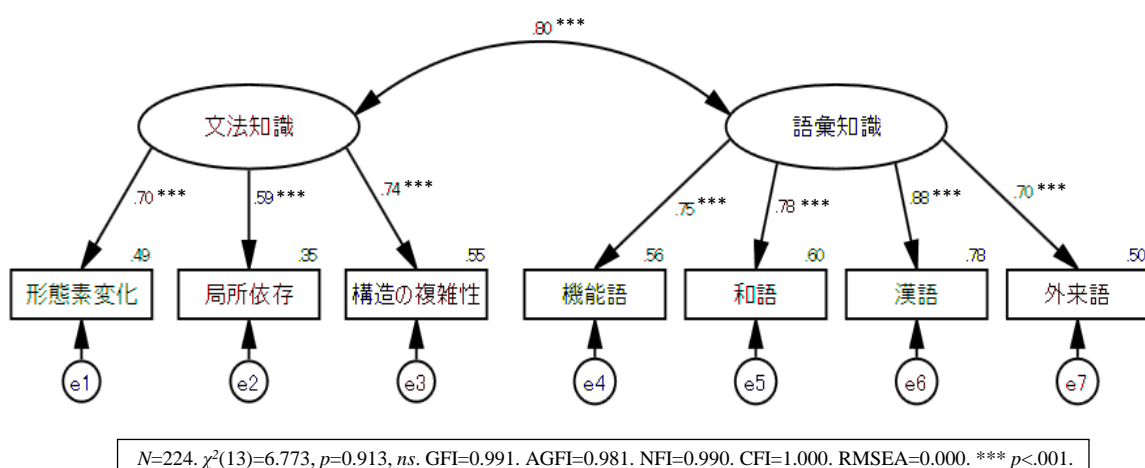


図2 モデル1: 語彙知識から機能語への理解を予測するモデル

さらに2つのモデルそれぞれの適合度を比較検討した。モデル1では、 $\chi^2(13)=6.773,$

$p=0.913$, $GFI=0.991$, $AGFI=0.981$, $NFI=0.990$, $CFI=1.000$, $RMSEA=0.000$ とすべての指標において基準とされる値よりも高く、いずれもあてはまりの良さの基準を満たしていた。

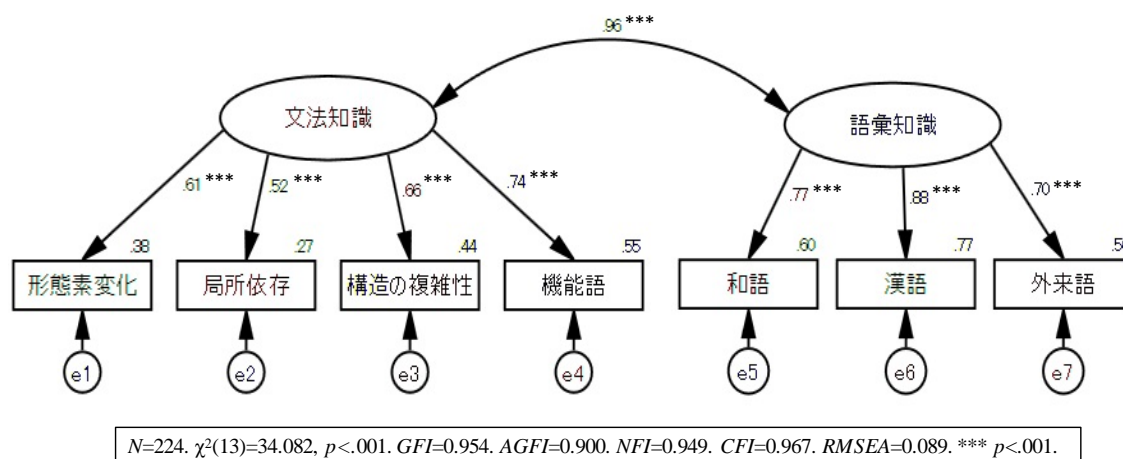


図3 モデル2: 文法知識から機能語への理解を予測するモデル

一方、モデル2は、 $\chi^2(13)=34.082$, $p<.001$, $GFI=0.954$, $AGFI=0.900$, $NFI=0.949$, $CFI=0.967$, $RMSEA=0.089$ と、 χ^2 検定では帰無仮説が棄却できず、 $RMSEA$ も1に近い値になっていた。 GFI と $AGFI$ をみても、適合の指標とされる 0.95 および 0.90 の値を超えていた。 NFI と CFI では許容できる程度の適合 (NFI が 0.90, CFI が 0.95) とみなされる値を示した。

したがって、2つのモデルを比較すると、モデル1のほうがすべての適合度指標においてより適した値を示しており、機能語は文法ではなく、語彙に位置付けるべきであることが実証された。

5. 文法知識テストの改訂

確認的因子分析の結果から、文法テストにおいては、2006年および2008年版の文法知識の構造分類による定義に立ち戻り、機能語を含めない、形態素変化・局所依存・構造的複雑性の構成で、改めて文法テストの改訂を行うことにした。改訂にあたっては、2006年および2008年頃に使用した文法テストには、一部の項目において弁別力の低いものがいくつもあったので、それらの項目や選択肢を精査することにし、より信頼性の高いテストの

開発を目指した。

5.1 2006/2008 年度文法テストの問題点

このテストが初めて実施された 2006 年度文法テストのクロンバック信頼性係数は、0.795 と悪くはなかったが、2008 年の改訂で項目困難度の高かった(すなわち、易しかった)問題 12 問(形態素変化1問, 局所依存 10 問, 構造の複雑性1問)が差し替えられた。しかし、結果として信頼性係数は 2008 年度の改訂では、全体として、難易度や識別度が上がらず、0.733 と 2006 年度よりも下がってしまった。

さらに、調査・実験にあたって簡便に測定ができるよう、同時に開発された日本語学習者の語彙テスト(宮岡・玉岡・酒井, 2011)では、日本語能力試験の配当級が 1 級および 2 級の語彙が使用されているため、両テストを実施した場合に、文法と語彙の難易度が一致せず、対象群によっては文法テストに天井効果が出てしまう場合があったことも、問題点の一つであった。

そこで、2006 年および 2008 年度改訂版の基本的な構成概念や項目数は変えずに、再度文法テストの改訂を行うことにした。その際、旧文法テストの項目分析において、困難度や識別度、実質選択指数の観点から、低い値のものを取り除いて新たに項目を追加することにした。なお、4章で明らかになったように、2010 年度版で加えられた機能語の項目は文法能力とするには不適合であることから、これらの項目は文法テストから排除した。

5.2 改訂において追加した新項目

宮岡・玉岡・酒井(2014)の 2008 年度改訂版テストでの形態素変化、局所依存、構造の複雑性のカテゴリで作成された項目において、項目困難度、項目弁別力の数値が低かったものを見直し、差し替えた。差し替えた問題は、問題そのものを替えたもの、錯乱肢のみを変更したものを含めて、全部で 18 項目である。改訂した文法テストの項目を表 1 に示す。網掛け部分は、今回の改訂で差し替えた問題である。天井効果が出てしまわないよう、旧・日本語能力試験で 2 級に配当級されているものを各カテゴリに 1 つないし 2 つ加えた(問題番号 6, 22, 25, 31, 表 1 を参照)。

形態素変化の項目では、5 項目の問題を差し替えた。特に、文法の難易度としては 3 級程度であるものの、複数の助動詞(ヴォイスやモダリティ)が組み合わせられた形で文末に用いられる複合助動詞の問題を 2 問追加した。なぜなら、助動詞が複数組み合わせられる場合

に、どのような配列が文法的に適格であり、意味的に通ずる形なのかを知っていることも重要な文法知識だからである。個々の助動詞の知識については、初期段階で学ぶことが多い。ところが、日本語では、単独で使用されるばかりでなく、それらが複数で組み合わせられた形が頻繁に用いられる。文における形態素配列の正確さを問う問題は、個別の形式のみを問う問題よりも深い文構造のレベルの知識を問うものであり、日本語学習者の文法能力をより高次に測ることができるようになっている。

表1 改訂文法テストの項目と問題文

問題番号	級	品詞・文法項目		問題文	
形態素変化 (Morphological Inflections)					
1	4	動詞	活用	過去・否定	誤って花びんを壊した私を、父は()。
2	4	動詞	活用	て形	外は寒そうだから、コートを()行こう。
3	4	形容詞	活用	過去・否定	さっきの説明は()。
4	4	形容詞	活用	て形	食事を作るのが()、ほとんど外食ですませています。
5	4	形容動詞	活用	現在・否定	私はあなたのわがままにつきあっているほど()。
6	2	助動詞	活用	ヴォイス+否定	今日、教科書を忘れちゃったから、ちよつと()?
7	4	形容動詞	活用	現在・肯定	彼の答えは()ので、彼の本心はわからない。
8	3	助動詞	接続	受身	道でだれかに肩を()、振り向いたら友達だった。
9	3	助動詞	接続	う・よう	このお金を銀行に()と思う。
10	3	複合助動詞	接続	ておく	事前に会議室に資料を()ください。
11	3	複合助動詞	接続	て+しまう+否定+ば	早くレポートを()、間に合わない。
12	3	複合助動詞	接続	使役受身	もう少しで留学を()。
局所依存 (Local Dependency)					
13	4	格助詞		に(対象)	私は毎朝7時発の()乗って大学に行きます。
14	4	格助詞		を(対象・経路)	私の弟は来年()卒業する。
15	4	格助詞		に(方向・目的)	昨日の放課後、()本を読みに行きました。
16	4	格助詞		で(範囲)	申し訳ありませんが、アルバイトの募集は先着()締め切りました。
17	4	形容詞+名詞			私は()人が好きだ。
18	4	形容詞+動詞			その布を()広げてください。
19	3	形容動詞+形容動詞			私の故郷は景色が()、静かです。
20	4	形式名詞+時		ばかり	仕事を始めた()とき、よくミスをしました。
21	3	自他動詞		わく・わかす	やかんでお湯を()、ポットに注いでください。
22	2	自他動詞		入る・入れる	はさみは引き出しに()あるから、自由に使っていいですよ。
23	4	疑問詞			青木氏が初めて書いた小説は()本ですか。
24	3	動詞の連体修飾		た(状態)	あそこのぼうしを()方が鈴木さんです。
構造の複雑性 (Complex Structure)					
25	2	条件		たとえ～ても	たとえそのような事実が()、許すことはできない。
26	4,3	副詞・アスペクト		まだ～ていない	今朝はまだ朝ごはんを()。
27	3	主述の一致		私は～と思う	私は彼が腹を立てたのは()。
28	3	疑問詞+～か・～かどうか			彼女がいつ日本に行くと()、田中さんに聞いてみてください。
29	3	主語の一致		～がる	私の妹は有名人を見ると必ずいっしょに写真を()。
30	3	名詞句の中の「が」			()提案した話を岡田さんは部長に伝えることにした。
31	2	敬語+可能		謙譲+否定	今回、妻が松本先生に()残念です。
32	4	疑問詞+も+否定		どこかも～ない	山で道に迷って、ここが()わからなくなった。
33	3	副助詞		までに	明日は朝8時半()学校に来てください。
34	3	授受		～に～もらう	私は昨日、()宿題を手伝ってもらった。
35	4	時制		仮定	あ、傘がない。雨が()時は友だちに借りましょうか。
36	3	アスペクト		～ところだ	私はこれから大学に()。

注1: グレーの部分は宮岡・玉岡・酒井(2014)から変更した項目を表す。

局所依存の項目では、7項目の問題を差し替えた。2006年度版では、局所依存に項目難易度が高い項目が多く、2008年における改訂では10問差し替えている。ただ、変更後の項目においても、全体的に項目難易度は高く、特に7つの項目の項目難易度が、0.89から1.00と非常に易しいとみられる問題であった。そこで、この7つの項目に関して、文法項目に大きな変更こそないものの、難易度が上がるように問題文や選択肢を見直した。

構造の複雑性の項目では、5項目の問題を差し替えた。特に、これまで2006年・2008年度版では、敬語の項目が含まれていなかったため、1問加えることにした。敬語は、動作主との人間関係がウチであるかソトであるかという基準から、動作主の行為を表す動詞に対して、尊敬語もしくは謙譲語という2つの異なる形態から適切にどちらかを選択しなければならない。敬語は、動作主が動詞の呼応を要求する文法であり、今回の例(問題番号31)のように、「今回、妻が松本先生に()、残念です」という文は、「妻」が動作主であるので、関係は「ウチ」となり、動詞は「会う」の尊敬語である「お会いになる」という形は間違いであり、謙譲語である「お会いする」というのが正しい呼応の形となる。それに加えて、この文は「会うことができなかった」という不可能の事実を表すので、「お会いできなくて」が正しい形である。したがって、錯乱肢の「お会いになれなくて」は尊敬語＋可能で呼応の違反となる(錯乱肢は表3を参照)。

以上のように、今回の文法テストの改訂では、2008年度改訂版の半数にわたる項目を差し替えたことになる。この改訂文法テストを、2014年に中国・上海において中国人日本語学習者に対して実施した。次章で調査の概要ならびにテストの信頼性と項目の分析の結果を報告する。

6. 文法テストの実施

6.1 調査協力者

中国の上海にある東華大学および上海商学院で、日本語を専攻する1年終了生30名、2年終了生63名、3年終了生42名、4年終了生8名、合計143名(男性23名、女性120名)に対して、2014年9月に文法テストを行った。調査協力者の平均年齢は21歳0ヶ月で、年齢が最も低かった学生が18歳8ヶ月、最も高かった学生が25歳5ヶ月であった。日本滞在歴の平均は5ヶ月で、最も短い滞在期間は2週間、最も長い期間が1年6ヶ月であり、95名は日本への滞在歴はなかった。

6.2 文法テストの得点と信頼性係数

今回の文法テストの平均点は、20.83 点(36 点満点)で、標準偏差は 0.7 点であった。カテゴリー(下位範疇)別に得点(各 12 点満点)を見ると、形態素変化の平均点は 7.27 点で、標準偏差は 0.26 点、局所依存の平均は 6.89 点で、標準偏差は 0.21 点、構造の複雑性の平均が 6.67 点で、標準偏差は 0.25 点であった。平均得点に違いがあるかどうかを調べるために、学年(被験者間要因)とカテゴリー(被験者内)の 2 元配置による分散分析を行ったところ、学年およびカテゴリーの主効果がみられ [学年の主効果; $F(3, 139)=15.343, p<.001$, カテゴリーの主効果; $F(2, 278)=3.255, p<.05$], 交互作用も有意であった [$F(3, 273)=2.420, p<.05$]。学年の主効果とカテゴリーの主効果が有意であったので、多重比較を行った。学年別にみると、1 年終了生と 2 年終了生の得点には差がなく、3 年生と 4 年生の得点差にも違いはなかった。

表 2 改訂文法テストの平均、標準偏差および分散分析の結果

文法テストの 下位カテゴリ	合計 ($N=143$)		1年終了生 ($n=30$)		2年終了生 ($n=63$)		3年終了生 ($n=42$)		4年終了生 ($n=8$)	
	M	SD	M	SD	M	SD	M	SD	M	SD
形態素変化	7.27	0.26	5.50	0.43	6.49	0.29	8.07	0.36	9.00	0.83
局所依存	6.89	0.21	6.07	0.35	6.03	0.24	7.60	0.30	7.88	0.68
構造の複雑性	6.67	0.25	4.97	0.41	5.25	0.28	7.60	0.35	8.88	0.80
合計	20.83	0.73	16.53	1.19	17.78	0.82	23.26	1.01	25.75	2.31
ボンフェローニ の多重比較	4つの学年 下位カテゴリ		<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>Mo</u> <u>L</u> <u>C</u>		<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>Mo</u> <u>L</u> <u>C</u>		<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>Mo</u> <u>L</u> <u>C</u>		<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>Mo</u> <u>L</u> <u>C</u>	

注: M は平均, SD は標準偏差を示す。Mo は形態素変化, L は局所依存, C は構造の複雑性を示し、各 12 点で合計 36 点満点である。

このことから、本調査では、2 年生と 3 年生の間に文法知識に大きな開きがあることがわかった。カテゴリー別にみると、形態素変化と局所依存の間に差がなかったが、2つのカテゴリーは構造の複雑性よりも有意に得点が高かった。旧文法テストと比較すると、2006 年度版では局所依存の得点が高く、易しすぎたために天井効果が出てしまっていたが、今回は形

態素変化と同程度の難易度となった。また、交互作用が有意であったので、ボンフェローニの検定で学年別に多重比較を行ったところ、表 2 のようになった。学年が上がるにしたがって、形態素変化の得点は段階的に伸びていって局所依存との差が縮まり、徐々に構造の複雑性の得点が上がっていくことで、局所依存との差も小さくなる。構造の複雑性は、結合関係にある単語同士が位置的に離れていたり、文全体を通して文法的な解析が必要であったりすることから、もっとも難易度が高いカテゴリーであるといえる。多重比較の結果から、今回の改訂文法テストは、日本語学習者が日本語の文構造の知識を段階的に習得していく様相を適切に反映していると考えて差し支えないであろう。

また、本テストの信頼性分析を行ったところ、クロンバックの α 係数は 0.820 で、これまでに改訂された文法テストの中で、最も高い数値が示された。項目数が 36 問と少なく、総被験者数も 143 名と小規模だったことをふまえると、本テストは高い信頼性を持つものであるといえよう。

6.3 改訂文法テストの項目分析

本改訂文法テストの項目分析には、TDAP (Test Data Analysis Program) Ver.2.0(大友・中村・秋山, 2002)を用いた。項目分析の内容としては、一般的に項目困難度 (item difficulty: DIFF), 項目弁別力 (item discrimination power index: DISC), 実質選択肢数 (actual equivalent number of options: AENO) が考慮される (大友, 1996)。TDAP には、これらの3つの指標が示されるようあらかじめプログラムが実装されているため、本文法テストの項目分析にあたって、この各指標を中心に項目の診断を行っていく。なお、項目ごとの各指標は、表 3 に示した。

6.3.1 項目困難度

項目困難度 (DIFF) は、0.000 から 1.000 までの値をとり、1.000 に近いほど易しい項目となる (大友, 1996; 中村, 2002)。項目困難度を求める式は、正答者数/受験者総数で、いわゆる正答率である。項目の診断にあたっては、選択肢の数によって、どのくらいの困難度が適切かを算出することができる。本テストのように選択肢が4つの場合の最適困難度は、 0.5 (理想困難度) + 0.5 (偶然に正解した確率) $\times 1/4$ (選択肢数) = 0.625 となる。このように算出された最適困難度の値に近いほど、その項目は適当な難易度を持っているということになる。項目難易度は正答率と同じなので、被験者の 70%以上が正解した問題は、易しい問

表3 改訂文法テストの項目分析結果

カテゴリ	級	正答	錯乱肢	項目困難度 (DIFF)	項目弁別力 指数(DISC)	実質選択指数 (AENO)
形態変化 (Morphological Inflections)	4	責めなかった	責めなかったでした	0.804*	0.565	2.026
	4	着て	着いて	0.308	0.495	2.711
	4	ややこしくありませんでしたか	ややこしいじゃなかったですか	0.322	0.241*	3.854
	4	面倒くさくて	面倒くさかって	0.462	0.489	3.267
	4	ひまじやない	ひまじくない	0.664	0.476	2.452
	2	見せてもらえない	見せてもられない	0.699	0.473	2.498
	3	あいまいな	あいまいに	0.755*	0.348	2.125
	3	たたくれて	たたかられて	0.573	0.419	2.878
	3	預けよう	預けろ	0.538	0.522	2.743
	3	持って行っておいで	行って持っておいて	0.580	0.097*	2.280
局所依存 (Local Dependency)	3	書き上げてしまわなければ	書き上げなければいけません	0.538	0.245*	3.171
	3	あきらめさせられそうになった	あきらめられそうになった	0.643	0.360	2.735
	4	バスに	バスで	0.573	0.478	2.641
	4	大学を	大学に	0.538	0.386	2.900
	4	図書館に	図書館を	0.434	0.284*	2.216
	4	10人で	10人を	0.273	0.314	2.808
	4	かっこいい	かっこよく	0.748*	0.356	1.950*
	4	大きな	大きい	0.713*	0.510	2.027
	3	きれいで	きれいの	0.902*	0.278*	1.484*
	4	ばかりの	ばかりな	0.706*	0.219*	2.358
構造的複雑性 (Complex Structure)	3	わかつて	わいて	0.210	0.394	3.910
	2	入れて	入れられて	0.231	0.274*	3.983
	4	どの	どう	0.846*	0.112*	1.699*
	3	かぶった	かぶり	0.427	0.268*	3.010
	2	あっても	あるのに	0.552	0.525	3.235
	4,3	食べていない	食べなくて	0.392	0.577	2.188
	3	当然だと思っ	当然です	0.650	0.336	2.538
	3	思うか	思いますか	0.399	0.419	3.218
	3	とりたがる	とりたがるです	0.573	0.602	3.082
	3	森山さんが	森山さんこそ	0.671	0.600	2.593
3	お会いできなくて	お会いにならなくて	0.287	0.304	3.736	
4	どこかも	どこにも	0.455	0.325	3.426	
3	までに	にまで	0.741*	0.088*	2.242	
3	兄に	兄を	0.476	0.238*	2.438	
4	降った	降って	0.259	0.082*	2.463	
3	行くところです	行っているところですよ	0.629	0.535	2.743	

注1：グレーの部分は富岡・玉岡・酒井(2014)から変更した項目を表す。

注2：*は項目困難度、項目弁別力指数、実質選択指数の各指標において、問題の適切さを示す望ましい値を満たしていないことを表す。

題であるといえよう。表 3 の項目難易度の欄の*印は、0.7 以上の数値を示した項目であり、36 問中 8 問であった。そのうち、0.9 を超えた項目は1つのみ(問題番号 19, 局所依存, ナ形容詞+ナ形容詞「私の故郷は景色が(きれいで)、静かです。」)であり、2006 年度版のテストで 0.9 以上の問題が 7 問あったことと比較すると、各項目の難易度は大幅に上がったといえる。

6.3.2 項目弁別力指数

項目弁別力指数 (DISC) は、テストにおけるある項目が、能力の高い学習者と低い学習者を識別できるかどうかを見出すための指標である。上位グループの正解者数が下位グループの正解者数よりも多いほど、項目弁別力が高いということになり、上位グループと下位グループの正解者数が同じであれば、項目弁別力がない(つまり、被験者を識別しない)問題だといえる。能力とは反対に、下位グループの正解者数が上位グループを上回れば、項目弁別力がないばかりか、負の弁別を行っていることを示す。したがって、項目弁別力指数は、-1.000 から+1.000 までの値をとる。項目弁別力の目安としては、0.3 以上が望ましいとされている(大友, 1996)。本改訂文法テストで 0.3 以下であった項目は 12 問で、とりわけ局所依存のカテゴリーが多かった。項目困難度の値が 0.7 以上で、項目弁別力が 0.3 以下の項目は、正答率が高く、上位下位グループの識別をしていないことを意味し、非常に簡単な問題であったと診断できる。この部類に入る項目は、問題番号 19, 20, 23, 33 の 4 問であった。それ以外の問題番号 3, 10, 11, 15, 22, 24, 34, 35 は、項目困難度が低く、上位下位グループ関係なく、難しい問題であったと診断できよう。

6.3.3 実質選択肢数

実質選択肢数 (AENO) とは、項目の選択肢が、受験者にどれだけ選ばれたかを示す指標であり、言い換えれば、実質的にどれだけ魅力的な錯乱肢があったかを判断する材料ともいえる。実質選択肢数は、選択肢数が 4 つならば、1.00 から 4.00 までの値をとる。例えば、問題番号 14 の実質選択肢数は 2.900 なので、3 つの選択肢に回答が分布しており、1 つの選択肢は錯乱肢としての機能をあまり果たさなかったことを意味する。単純に、回答が均等に分布していると考えると、理想的な実質選択肢数は 2.50 となるが、実際はそのようにきれいな分散とはならないので、回答が 2 つの選択肢のみに集中したものを、選択肢に改善の余地がある項目とみなすことにする。本テスト

で、実質選択肢数が 2.00 以下であったものは問題番号 17, 19, 23 の 3 つのみであった。問題番号 17 は、イ形容詞+名詞の問題で、正答「かっこいい」と「かっこいいな」に回答が集中しており、問題番号 19 は、ナ形容詞+ナ形容詞の問題であり、正答「きれいで」と「きれい」に回答が集中していた。問題番号 23 は、疑問詞の問題であり、回答は、正答「どの」と「なん」に集中していた。「なん」のほかには、「なに」も若干選択されていた。これは、疑問詞の後項が「本」であり、問題文全体を読まないで、「どの本」なのか、助数詞としての「何本」なのかを区別できない問題となっている。また、「本」は中国語では書籍を数える「～冊」という意味で使われ、そのように解釈すると「何」(なに・なん)を選択してしまう可能性があり、一種のひっかけ問題である。項目困難度をみると、これら 3 問はすべて 0.7 以上で高い数値を示しており、そのうち問題番号 19 と 23 は項目弁別力指数でも識別力が低かった。つまり、この問題は下位グループの受験者にとっても非常に簡単であったということがわかる。このことから、3 つの項目はそれほど魅力的ではなかった錯乱肢であったと考えられる。したがって、問題 23 は錯乱肢を変更し、問題 19, 23 は除外して別の問題と差し替えれば、さらに本テストの信頼度係数が高くなると考えられる。

7. おわりに

本テストは、新たに宮岡・玉岡・酒井 (2014) の文法テストの改訂版として開発し、日本語学習者の文法能力を適切に測定するテストであるかどうかを、信頼性の観点から評価した。その結果、これまでで信頼性係数が最も高い数値を示した。本文法テストは、動詞や形容詞の変化などの知識を問う形態素レベル、隣り合う語の関係についての知識を問う局所依存、文中である程度の距離にある語の関係を問う構造の複雑性という 3 つの構造分類により構成されている。今後、従来のように助詞や文法形式・機能などの知識を問う文法の範疇分類によるテストを開発して併用することで、構造分類の妥当性を検討する必要もあろう。また、信頼性係数は、受験者母集団によって変化するものであり、絶対的な指標ではない。しかし、どのような受験者集団であっても安定した結果であるということは、そのテストが内的整合性と一貫性を持つものであることを示している。その点で、さまざまな学習者を対象に実施して、同様に安定した指標が得られるかどうかを確認することが今後の課題であろう。今後、これらを考慮して改訂を加えていけば、より安定した文法テストとなると考えられる。

[引用文献]

- 大友賢二 (1996) 『項目応答理論入門』 大修館書店
- 国際交流基金・日本国際教育協会 (2002) 『日本語能力試験出題基準 改訂版』 凡人社
- 中村洋一 (2002) 『テストで言語能力は測れるか～言語テストデータ分析入門』 大友賢二
監修, 桐原書店
- 初相娟・玉岡賀津雄・早川杏子 (2013) 「中国人日本語学習者の場所を表す格助詞「で」と
「に」の習得に影響する諸要因」『日中言語研究と日本語教育』 6, 59-70.
- 早川杏子・玉岡賀津雄 (2012) 「中国人・韓国人日本語学習者による聴覚・視覚提示の言
語間同形義・言語間異形同義の二字漢字語の処理」『小出記念日本語教育論集』 20,
17-32.
- 宮岡弥生・玉岡賀津雄・酒井弘 (2014) 「日本語の文法能力テストの開発と信頼性—日本
語学習者のデータによるテスト評価」『広島経済大学研究論集』 36(4), 33-46.
- 李在鎬・宮岡弥生・林炫情 (2013) 「学習者コーパスと言語テスト—言語テストの得点と作
文のテキスト情報量の関連性—」『言語教育評価研究(AELE)』 3, 22-31. 言語教育評
価共同研究所
- 大和祐子・玉岡賀津雄 (2013a) 「中国人日本語学習者による外来語処理への英語レキシ
コンの影響」『レキシコンフォーラム』 6, 229-267.
- 大和祐子・玉岡賀津雄 (2013b) 「中国語母語話者と韓国語母語話者の日本語テキストの
読み処理における言語的類似性の影響」『小出記念日本語教育研究会論文集』 21,
61-73.
- 大和祐子・玉岡賀津雄・初相娟 (2013c) 「中国人日本語学習者のテキストのオンライン読
みにおける語彙と文法能力の影響」『日本教科教育学会誌』 36(1), 33-43.

[分析使用ツール]

- 大友賢二・中村洋一・秋山實 (2002) Test Data Analysis Program (TDAP) Ver.2.0 [Windows
版] (中村洋一 (2002) 『テストで言語能力は測れるか～言語テストデータ分析入門』
大友賢二監修, 桐原書店, に添付された解析ソフト)

早川 杏子 - 関西学院大学 日本語教育センター・常勤講師
玉岡 賀津雄 - 名古屋大学大学院 国際言語文化研究科・教授

補記: 改訂文法テスト(2015 年度版) ※問題はランダムにしてある。正解は表3を参照。

次の文の()に入れることばとして正しいものはどれですか。

4 つの中から1つだけえらんで、□に✓を書いて下さい。

- 1 外は寒そうだから、コートを()行こう。
 着て 着いて 着んで 着って
- 2 私は毎朝7時半発の()乗って大学に行きます。
 バスの バスに バスを バスで
- 3 私はあなたのわがままにつきあっているほど()。
 ひまにない ひまじゃない ひまじゃない ひまくない
- 4 今朝はまだ朝ごはんを()。
 食べなかった 食べた 食べなくて 食べていない
- 5 道でだれかに肩を()、振り向いたら友達だった。
 たたかかれて たたかられて たたきられて たたくれて
- 6 私は()人が好きだ。
 かっこよく かっこいい かっこいいの かっこいいな
- 7 私はこれから大学に()。
 行っている 行っているところです 行ったところです 行くところです
- 8 このお金を銀行に()と思う。
 預けるよう 預けろう 預けよう 預けう
- 9 申し訳ありませんが、アルバイトの募集は先着()締め切りました。
 10人が 10人を 10人まで 10人で
- 10 青木氏が初めて書いた小説は()本ですか。
 なに どう なん どの
- 11 誤って花びんを壊した私を、父は()。
 責めないでした 責めないだった 責めなかった 責めなかったでした
- 12 昨日の放課後、()本を読みに行きました。
 図書館を 図書館で 図書館に 図書館と
- 13 明日は朝8時半()学校に来てください。
 で まで までに にまで
- 14 私の妹は有名人を見ると必ずいっしょに写真を()。
 とりたい とりたがる とりたいです とりたがるです
- 15 山で道に迷って、ここが()わからなくなった。
 どこにも どこでも どこへも どこかも
- 16 彼女がいつ日本に行くと()、田中さんに聞いてみてください。
 思う 思うか 思いますか 思うかどうか
- 17 その布を()広げてください。
 大きく 大きな 大きい 大きくて
- 18 私は昨日、()宿題を手伝ってもらった。
 兄が 兄に 兄を 兄から

- 19 彼の答えは()ので、彼の本心はわからない。
 あいまいに あいまい あいまいだ あいまいな
- 20 仕事を始めた()とき、よくミスをしました。
 ばかり ばかりの ばかりな ばかりに
- 21 さっきの説明は()。
 ややこしいじゃないでしたか ややこしいじゃなかったですか
 ややこしくありませんでしたか ややこしくありませんでしたか
- 22 今日、教科書を忘れちゃったから、ちょっと()？
 見せてもらわない 見せてもらいない 見せてもらえない 見せてもられない
- 23 はさみは引き出しに()あるから、自由に使っていていいですよ。
 入れて 入って 入れられて 入られて
- 24 あそこのぼうしを()方が鈴木さんです。
 かぶった かぶり かぶって かぶるの
- 25 私の故郷は景色が()、静かです。
 きれいの きれいで きれいな きれい
- 26 私の弟は来年()卒業する。
 大学に 大学で 大学を 大学から
- 27 やかんでお湯を()、ポットに注いでください。
 わかせば わいて わいたら わかして
- 28 あ、傘がない。雨が()時は友だちに借りますか。
 降る 降り 降って 降った
- 29 食事を作るのが()、ほとんど外食ですませています。
 面倒くさくて 面倒くさいで 面倒くさかって 面倒くさって
- 30 ()提案した話を岡田さんは部長に伝えることにした。
 森山さんなら 森山さんこそ 森山さんが 森山さんは
- 31 もう少しで留学を()。
 あきらめられさせそうになった あきらめそうられさせになった
 あきらめさせられそうになった あきらめそうさせられになった
- 32 たとえそのような事実が()、許すことはできない。
 あるが あっても あるのに あるものの
- 33 私は彼が腹を立てたのは()。
 当然だ 当然だと思う 当然です 当然ではないでしょうか
- 34 事前に会議室に資料を()ください。
 行って持っておいで 持って行っておいで
 持っておいで行って おいで持って行って
- 35 今回、妻が松本先生に()残念です。
 お会いできなくて お会えなくて お会いになれなくて お会えにならなくて
- 36 早くレポートを()、間に合わない。
 書き上げないでしまえば 書き上げてしまわなければ
 書き上げればしまわないで 書き上げなければしまって

Development of the revised version of the structure-based Japanese grammatical knowledge test: The evaluation by data from native Chinese speakers learning Japanese

HAYAKAWA, Kyoko

Instructor of Japanese as a Foreign Language, Center for Japanese Language Education, Kwansai Gakuin University, Japan

hayakawa@kwansai.ac.jp

TAMAOKA, Katsuo

Professor, Graduate School of Language and Cultures, Nagoya University, Japan

ktamaoka@lang.nagoya-u.ac.jp

Abstract: The revised version of the structure-based Japanese grammatical knowledge test is based on the previous version from Miyaoka, Tamaoka and Sakai (2014). Prior to the revision of the test, the present study first investigated which knowledge category functional words belong to, lexical knowledge or grammatical knowledge. Using confirmatory factor analysis, two models were investigated: Model #1 (Figure 2 in the paper) included function words as a part of lexical knowledge, while Model #2 (Figure 3 in the paper) included the category as a part of grammatical knowledge. Four criteria for selecting a model from a set of models (i.e., AIC, CAIC, ECVI and BIC) supported Model #1 against Model #2, suggesting that function words are comprised as a part of lexical knowledge. Based on this result, function words were excluded from the present revised grammatical knowledge test; however, function words were included as question items for the former grammatical knowledge test by Miyaoka et al. (2014). The revised test consisted of 12 multiple choice questions for each of the three sub-categories (a total of 36 questions); namely, morphological inflections, local independency, and complex structure. The test was administered to 143 native Chinese speakers learning Japanese in Shanghai, China and obtained a high Cronbach's alpha reliability coefficient of 0.820. Therefore, this revised test is able to efficiently measure the grammatical knowledge of learners of Japanese as a foreign language.

Keywords: test development, grammatical and lexical knowledge, function words, confirmatory factor analysis