

# ニューラルテスト理論の第二言語習得研究への応用可能性<sup>1</sup>

山川健一・杉野直樹・清水裕子・大場浩正・中野美知子

## 1 はじめに

テストの実施後、結果を得点化したり、様々な分析を行う際には、古典的テスト理論の枠組みを用いるのが一般的である。受験者集団の特性を知るためには、平均点などの代表値やちらばりを示す分散などの基礎統計量が役立つ。またテストを開発する際には、各項目が受験集団に対して適切であったか否かを検討するための情報を提供する項目分析 (item analysis) が有用であり、テストの精度を上げることにも役立てることができる。ところで、古典的テスト理論では、ある観測変数の得点が真の得点と誤差から構成されると仮定し、正答数に基づく得点や素点を用いて分析が行われる。しかしながら、この方法は受験者集団やテスト項目に依存した統計処理であるため、テスト結果を過剰に一般化できないという点や、集団における受験者の相対的な位置の比較には標準誤差を考慮に入れた慎重な検討が必要である点など、留意すべきことが多々ある (大友・中村, 2002: 104)。そこで、この欠点を補う新しいテスト理論として、項目応答理論 (Item Response Theory: IRT) が注目され、言語テストの分野でも注目されるようになってきた。IRTでは、正答数に基づく得点ではなく、ロジット得点を使用することで、等間隔の目盛りをもつ間隔尺度上に数値情報を示すことができるようになる。その結果、受験者やテストの難易度に影響されることなく、受験者の能力やテスト項目の難易度を連続尺度で推定することを可能にしてくれる。本研究グループの研究 (Yamakawa et al. 2008) においても、様々な文法項目の文法性判断タスクを日本人英語学習者に実施し、得られたデータをIRTを用いて分析することで、広範囲な文法能力の発達過程を明らかにしてきた。また、Yamakawa et al. (2008) は、各文法性判断タスクを構成する各種文法カテゴリーの困難度の相対的な順序を明らかにし、学習者の習得段階を考察する際、動詞に関する3つの習得段階を発見し報告した。しかしながら、学習者の発達段階については、恣意的な区切りにしか過ぎないのではないのかという疑問がどうしても残った。つまり、IRTを用いた研究では、学習者の習得段階を考える際に、ある種の区切りのようなものを見つけ出すことは困難であった。

本論では、上記の問題点を克服すべく、新たな試みとして、潜在的な順序尺度によるニューラルテスト理論 (Neural Test Theory: NTT)<sup>2</sup>を援用することで、連続尺度で位置づけを行うIRTを用いた研究の弱点を補い、第二言語習得の研究における分析手法としてのNTTの可能性を探ることを目的とする。

## 2 教育測定の問題点

何かを測定するには、測定道具が必要である。また、測定した結果を表現したり解釈するための計量単位も不可欠である。例えば、重さの測定では体重計や秤などを使い、キログラムやポンドなどで測定結果を示すことになる。距離や長さも、ものさしや巻尺、ウォーキングメジャーなどを用いて、センチメートル、フィート、ヤードなどの単位で表現することになる。また、用いる道具によって、測定範囲が決まっており、最大計量値も異なってくるのが普通である。ところで、重さや長さのような物理的な特性の測定では、測定道具や計量単位についての共通理解が我々の中にあり、その結果を信頼して解釈をしている。では、言語テストや様々な科目の試験の場合はどうであろうか。確かに、入学試験や検定試験、さらには教師が作成したテストなど、多くの測定道具が存在して、素点や偏差値の形で得点情報が提供できるであろう。しかしながら、これらの教育測定においては、様々な問題が潜在している。例えば、英語のテストを例にとると、「英語力」を測定すると言っても、英語力の構

成概念が明確でないことが多く、普遍的に認められた構成概念も存在しない。また、一つのテストで測定に用いることのできる項目数は限られており、また、それらの項目がすべて代表性をもったテスト項目とは必ずしも言えないため、これらの要因はテスト結果の信頼性に影響を与える可能性もある。信頼性の問題は、物理的な特性に比べて、教育測定における測定誤差が大きいこととも関係している。

このように、テストの精度や信頼性を考えると、連続得点として出力されるテスト結果に基づいて直線上に受験者を並べた場合、わずかな得点差が必ずしも高い精度をもつ能力差とは言えず、誤差によって能力差があるように見えてしまっている危険性も高いことになる。例えば、100点満点のテストの結果を仮に1点刻みで分類したとすると、そのテストで測定しようとする能力は、100の水準に分けられることになるが、果たして、45点と46点や、90点と94点の差が実際に能力差と断定できるのであるか。そこで、荘島(2009a, 2009b)は、連続尺度上に受験者を位置づけるのではなく、潜在的な順序尺度によって5~20程度の段階に位置づけるための手法として、ニューラルテスト理論を展開すると共に、分析のためのソフトウェア「Exametrika (エグザメトリカ)」<sup>3</sup>を開発した。

### 3 NTTとは

#### 3.1 NTTと潜在ランク

NTTは、統計的学習理論の一つであり、観測した対象を数値化し、そこから特定のパターンを認識して記述するものである。ニューラルネットワークモデルの一つである自己組織化マップ(self-organizing mapping: SOM)や生成トポグラフィックマッピング(generative topographic mapping: GTM)のメカニズムを応用した統計モデルであり、学力を段階評価するために連続尺度ではなく順序尺度を仮定したテスト理論である(荘島, 2009a, 2009b)。この段階評価を潜在ランク(latent rank)とし、能力が高い受験者ほど、数値の大きいランクに属するように分析、配置される。ランク数をいくつにするか、つまり、何段階に受験者をグループ分けしたいかは、分析者が決めることができるが、荘島(2009a, 2009b)は、一般的に言って、テストによって識別できるには、5~20程度の段階の差であるとしている。実際のデータを分析してランク数を決定する際には、テストの構成概念や得られたデータの特質を考慮しながら、分析者が経験的、恣意的に行うことになるが、Exametrikaが算出する適合度指標を参考にすることもできる。適合度指標は、データがどの程度NTTモデルに適合しているかを見るもので、潜在ランク数を適宜変えて分析を繰り返しながら、最適なランク数を発見していくことになる。ただし、荘島(2009a, 2009b)は、ランク数が多くなればなるほど適合度指標は高くなるが、適合度指標が最適値を示さなくても、意味があると判断できれば、分析者の判断によるランク数に設定した方がテストの目的に適うとしている。

#### 3.2 項目参照プロファイル

各項目がどのような振る舞いをするかを知りたい場合、IRTでは項目特性曲線(item characteristic curve: ICC)を用いるが、NTTにおいては、図1や図2のような項目参照プロファイル(item reference profile: IRP)を用いる<sup>4</sup>。これらの図ではランクを10に設定しており、X軸は潜在ランクを示し、数値が大きいほど高いランク(上位ランク)となる。また、Y軸は、各ランクに属する受験者集団が、それぞれの項目に正解する確率を示している。図1の場合は、最も下位に属する集団(潜在ランク1)でも、70%以上の確率で正解するような項目であることがわかる。また、ランク5以上が正答する確率に変化があまり見られず、ランク5以上にある受験者の識別力が低い項目であるとも言えよう。

図2の項目はどうであろうか。ランク6の周辺で曲線の勾配が変化しており、ランク6以上の集団は正答確率が高くなっており、ランク5以下はあまり差がない。つまり、この項目は、ランク6の周辺で識別力が高いと言える。

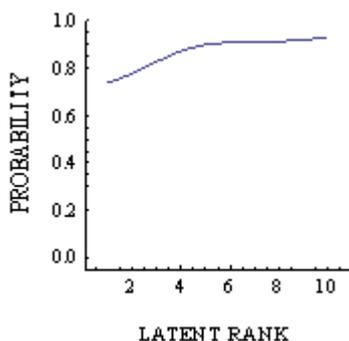


図1 項目参照プロファイル (1)

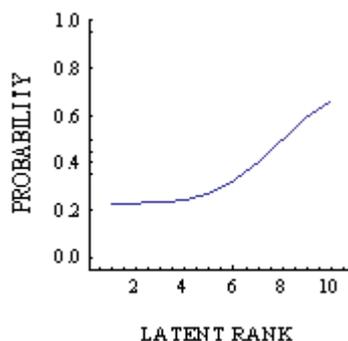


図2 項目参照プロファイル (2)

### 3.3 テスト参照プロファイル

NTTでは、テスト参照プロファイル (test reference profile: TRP) が算出される。これは、前述の項目参照プロファイルの重み付きの和であり、各ランク内に収まる受験者の期待得点である。図3は35項目からなるテストを分析した結果であるが、X軸は項目参照プロファイルと同じく潜在ランクを示し、Y軸は期待得点を示している。なお、この例では全ての項目の重みが1のため、期待正答数を示していると考えればよい。例えば、ランク1に属する受験者は、35問中11問程度に正解するし、ランク10の者は23問程度に正解すると解釈できる。

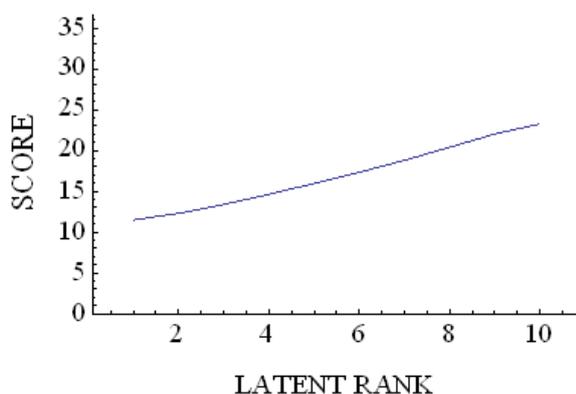


図3 テスト参照プロファイル

### 3.4 ランク・メンバーシップ・プロファイル

NTTにおける各受験者に関する情報として、ランク・メンバーシップ・プロファイル (rank membership profile: RMP) がある。受験者一人ひとりが、各潜在ランクに所属する確率を示すものであるが、図4のような結果を示す受験者は、潜在ランク6に属する確率が最も高いと言えるし、図5の受験者は、潜在ランク3ではあるが、潜在ランク8のような高いレベルに移る可能性もあり、基礎的な学習の強化の必要性を示す結果となっている。このような情報は、学習者に対する診断的情報を提供するものであり、教育的な意義が高い。

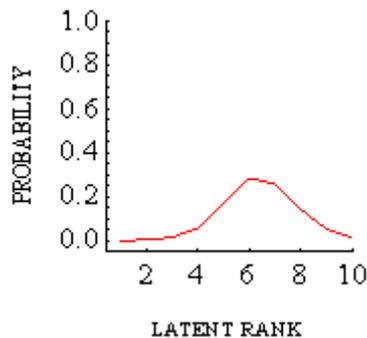


図4 ランク・メンバーシップ・プロファイル (1)

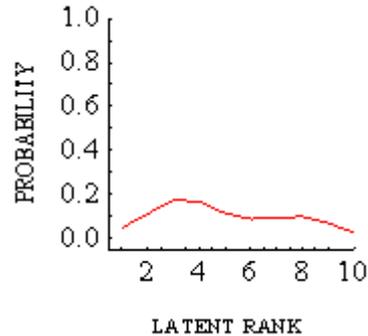


図5 ランク・メンバーシップ・プロファイル (2)

### 3.5 NTT と教育的意義

このように、NTT では、多くの指標や情報が提供されるが、Shojima (2009a, 2009b) が述べているように、NTT の教育的意義は、学力進捗表 (can-do chart) の作成補助となる情報の提供にある。Can-do chart は、連続尺度では作成が困難であるが、NTT により能力段階が明らかになると、各段階に対応する学力水準の記述文 (can-do statement) を作成することができるようになる。これにより、各学力段階において学習者が何ができて何ができないかを明確にすることができ、テストの説明責任を果たすことができるとしている。

NTT のこの考え方は、上述した Yamakawa et al. (2008) の問題点を解決する糸口を与える可能性があるといえる。すなわち、IRT によって複数の文法項目のそれぞれの内部カテゴリーの困難度の相対的順序は明らかになったが、習得段階という視点でとらえた場合、各段階にいる学習者が何ができて何ができないのかについて示唆を得ることができなかった。そこで、NTT を用いることによって、連続尺度上にあった文法カテゴリーが、順序尺度上に置き換えられ、学習者の習得段階が分けられるようになる。そして、それぞれの段階に存在する文法カテゴリーを分析することによって、各段階の学習者が何ができて何ができないかを記述することが期待できるのである。本論文は、NTT を第二言語習得研究に応用した初めての試みである。本論文の調査においては、日本人英語学習者の非対格動詞と非能格動詞の習得に関して、NTT を用いて分析を行った。次のセクションでは、第二言語習得研究における英語の非対格動詞と非能格動詞の先行研究についてまず概観することにする。

## 4 非対格動詞と非能格動詞

### 4.1 2種類の自動詞—非対格動詞と非能格動詞

Perlmutter (1978) は、自動詞を非対格動詞 (unaccusative verbs) と非能格動詞 (unergative verbs) に分類した最初の研究である。関係文法 (Relational Grammar) の枠組みの中で行われたこの研究によると、非対格動詞は意味的には、①形容詞・状態動詞、②対象物 (theme) や被動者 (patient) を主語に取る動詞、③存在・出現を表わす動詞、④五感に作用する非意図的な現象を表わす動詞、⑤アスペクト動詞が属するとし、一方、非能格動詞は、①意図的・意志的行為を行う動詞や生理的現象を表わす動詞が属するとしている。そして、非能格動詞の主語は最初から主語であるのに対し、非対格動詞の主語はもともとは目的語であり、後の段階で主語になると分析されている。

生成文法 (Government-Binding (GB) 理論) の枠組みでは、非能格動詞の主語 (動作主 (agent) の

意味役割を持つ) は、(1) のように D 構造と S 構造で主語 (外項 : external argument) の位置にあるのに対し、非対格動詞の主語 (対象物 (theme) の意味役割を持つ) は、(2) のように D 構造では目的語 (内項 : internal argument) の位置にあり、S 構造で主語の位置に移動すると考えられている。よって、表層的には非能格動詞と非対格動詞はどちらも「NP + V」の構造を持ち、その統語的相違を指摘するのは難しくなっている。

- (1) Mary danced.  
非能格動詞 [NP [VP V]] : [Mary [VP danced]]
- (2) The door opened.  
非対格動詞 [e [VP V NP]] : [e [VP opened the door]]
- (3) The boy opened the door.  
他動詞 [NP [VP V NP]] : [The boy [VP opened the door]]

このように、非対格動詞の主語がもともとの統語構造では内項として規定されているという仮定は「非対格性の仮説 (Unaccusative Hypothesis)」と呼ばれる。

#### 4.2 英語学習者に見られる非対格動詞の誤りとその原因

第二言語習得研究においては、これまで概観してきたような非対格動詞と非能格動詞の区別に関して、学習者がどのように理解しているかについて一連の研究が進められてきた (Balcom, 1997; Hirakawa, 1995, 1997, 2003; Ju, 2000; Oshita, 2000, 2001; Shomura, 1996; Sorace, 1993, 1997; Yip, 1994, 1995; Yuan, 1999; Yusa, 2003; Zobl, 1989)。その中でも、Zobl (1989) は先駆的な研究であり、この研究では、異なる L1 の英語学習者の英作文における非対格動詞の誤りについて、GB 理論の枠組みから詳細な分析が行なわれた。そして、学習者は非能格動詞よりも非対格動詞において、次の (4) や (5) のような誤りを多く犯していることが発見された。

- (4) \*The most memorable experience of my life was happened 15 years ago.  
(5) \*Most of people are fallen in love and marry with somebody. (Zobl, 1989: 204)

この誤りの原因について Zobl (1989) は、D 構造において受動文と非対格動詞の文は、次のように類似の構造を持つためであると説明している。

- (6) English is spoken in many countries.  
受動文 : [e [VP be spoken English]]
- (7) The accident happened 15 years ago.  
非対格動詞 : [e [VP happen the accident]]
- (8) \*The accident was happened 15 years ago.  
(9) \*The driver happened the accident 15 years ago.

(6) と (7) のどちらにおいても、theme の意味役割を担う NP が D 構造では内項の位置にあり、S 構造が派生される際に外項位置 (主語位置) に移動する。Zobl (1989) は、学習者は D 構造において受動文と非対格動詞の文が持つこの類似性に気づき、受動文の動詞の形態的特徴 (be+p.p.) を過剰般化して、非対格動詞の内項が主語位置へ移動するという現象と結びつけてしまい、(8) のような非文を産出すると説明した。このように、非対格動詞の内項位置にある NP が外項位置に移動する際に、自動詞を受動化した誤りが生じてしまうという仮説は、「NP 移動説 (NP movement account)」と呼

ばれている。

一方、(8) のような誤りの原因を説明する他の仮説として「語彙使役化説 (causativization account)」がある (e.g., Ju, 2000; Yip, 1995)。これによると、(8) のような非文は、(9) のように、学習者が使役主 (例えば *the driver* のようなもの) を勝手に創造し、自動詞を他動詞化したものが前提にあると仮定する。そして、この誤った他動詞文に通常の受動態生成規則を適用して、結果的に (8) のような非対格動詞を受動化した非文を産出することになると説明している。

### 4.3 非対格動詞の習得の道すじ

非対格動詞の習得の道すじについては、Oshita の一連の研究のみといっても過言ではないくらい数は少なく、不明な部分の多い分野である。Oshita (2001) は「非対格性わな仮説 (Unaccusative Trap Hypothesis: UTH)」を提唱し、非対格動詞の縦断的習得段階について次の 3 つの段階に分けて説明を加えている。

- (10) \* $[ \text{the accident} [ \text{VP happened} ] ] \rightarrow \text{The accident happened 15 years ago.}$
- (11)  $[ e [ \text{VP happened the accident} ] ] \rightarrow * \text{The accident was happened 15 years ago.}$
- (12)  $[ e [ \text{VP happened the accident} ] ] \rightarrow \text{The accident happened 15 years ago.}$

これによると、学習者は習得の第一段階 (10) では、非対格動詞と非能格動詞の区別をしておらず、非対格動詞の場合にも非能格動詞と同じ D 構造を想定し、内項にあるはずの名詞句 (*the accident*) が外項の位置にあり、誤った D 構造となっている。そして、これが S 構造になる際は表面的には文法的な非対格動詞の文が産出されることになる。次の第二段階 (11) になると、学習者は非対格動詞の正しい D 構造をもつようになる。そしてこの段階になって D 構造と受動文との類似性に気づき始め、D 構造から S 構造になる時に誤った受動化規則を適用して非文を産出する。これはすなわち NP 移動説を支持した考え方になっている。最後の第三段階 (12) になると、D 構造においても S 構造においても非対格動詞の正しい表象構造を持つことになる。以上見てきたように、非対格動詞を受動化した文を産出したり容認したりする現象や、また非対格動詞の正しい自動詞用法を拒否する現象は中級レベルで表出し、非対格動詞の習得状況は全体として U 字曲線 (U-shaped curve) を描くとされている。

## 5 調査

### 5.1 目的

本調査の目的は、上述したように第二言語習得研究に NTT を用いることによって、英語学習者の文法能力の習得段階を明らかにするものである。具体的には、非対格動詞と非能格動詞の習得に関して、文法性判断タスクを用いて、ある学習者グループの潜在ランクを明らかにし、それぞれの潜在ランクの特性、すなわち *can-do list* を記述するものである。

### 5.2 方法

本調査の実験参加者は、我々の過去の研究 (山川他, 2005) で収集した 369 名の大学生の日本人英語学習者のデータである。これらの学習者は、我々が開発した文法性判断タスクに解答した。このタスク (以下、Unit Y) は、英語の非対格動詞と非能格動詞の習得について調べるためのものであり、48 問から構成され、5 段階 (1: 完全に不可能、2: 多分不可能、3: わからない、4: 多分可能、5: 完全に可能) で各テスト項目の文法性を判断するものである。集計時には、実験参加者の判断と正解

との「距離」によって得点化した。例えば、適格文を「5」と判断すれば4点が与えられ、「1」と判断すれば0点が与えられる。非文の場合は逆に、「5」には0点、「1」は4点が与えられた。Unit Yには、6つの非対格動詞 (appear, arrive, die, exist, fall, happen) と6つの非能格動詞 (cry, dance, laugh, play, sing, work) の計12の動詞が、「適格文 (NP + V)」「非文の受動文 (NP + be + Ven)」「非文の他動詞文 (NP + V + NP)」の3つの構造の中で用いられ、全部で36のテスト項目 (カテゴリ A ~ F) となり (表1)、これにフィラーが12項目 (break, burn, close, dry, grow, melt の6つの能格動詞を用いて適格文の他動詞文5問と適格文の受動文5問) 加わり、全部で48項目となっている。

表1 Unit Y におけるテスト項目の例

Category	Verb	Construction	Example
A	unaccusative	NP+V	Your letter arrived yesterday.
B	unergative		Her father cried at her wedding ceremony.
C	unaccusative	*NP+be+Ven	*Because of the rain, the train was arrived late.
D	unergative		*He was cried when he heard of his mother's death.
E	unaccusative	*NP+V+NP	*Finally the waitress arrived the salad to us.
F	unergative		*The boy hit his little sister and cried her.

上記の369名のデータを、前述のNTTによる分析のために開発された統計ソフト Exametrika を用いて分析した。本論文のセクション3では、テスト項目の解答が正解か誤答のどちらか(2値)になるものを処理するためのNTTの理論(2値モデル)について概観した。一方、我々の今回の調査では、テスト項目は2値ではなく、5段階の文法性判断であり、4点から0点までが配点される。したがって、このようなテスト項目を分析するために、NTTの下位モデルの1つである段階ニューラルテスト(graded neural test: GNT)を用いた。

NTTは前述したように、テスト結果に基づいて学習者をいくつかの潜在ランクに分けるが、この際のランク数は分析者が決定することができる。今回の分析では、いくつかのランク数を試した結果、ランク数を4つに設定した。これは、これ以上ランク数を増やしても各ランク間での違いに発達段階上の意義が見いだせなかったためである。

### 5.3 結果

表2は、非対格動詞と非能格動詞36問の項目参照プロファイル(IRP)である。分析の結果、4つの潜在ランクの人数は、それぞれ88人、99人、95人、87人となった。次に、項目の平均点(4点満点)に従って、項目を4つのグループに分けた。表の最も濃い網掛けの部分は、平均点が3.0以上の項目、2番目に濃い網掛けの部分が2.5から3.0の項目、次の網掛けが2.0から2.5の項目、網掛けのない部分が、2.0以下の項目である。各ランクごとに見ると、上から下に向かって平均点の高い項目から低い項目となっている。ただし\*印の付いた項目は順序が例外となっている。

ランク1の学習者を見てみると、平均点が3.0以上(文法性判断の正解の度合いが高いもの)の項目は皆無であり、2.0から2.5の項目が最も多い。加えて平均点が2.0以下のもの(つまり文法性判断が間違っているもの)が6項目ある。ランク2では、2.0以下の項目数は変化していないが、3.0以上の項目が出現し始め、2.5から3.0の項目の数も増加している。これは文法性判断が全体的に正しい方向に変化していることを意味する。次にランク3では、3.0以上の項目と2.5から3.0の項目の数が増加しており、さらに正しい判断の方向へ移行していることを示している。最後にランク4の学習者を見てみると、2.0以下の項目は2項目のみとなり、3.0以上の項目が相対的に増加している。

表 2 非対格動詞と非能格動詞の項目参照プロファイル

Item No.	Category	Verb	Item Reference Profile (IRP)			
			Rank 1	Rank 2	Rank 3	Rank 4
			n=88	n=99	n=95	n=87
Y46	B	sing	2.991	3.365	3.672	3.822
Y42	B	cry	2.890	3.232	3.475	3.592
Y32	B	dance	2.830	3.203	3.380	3.517
Y06	A	die	2.949	3.086	3.299	3.460
Y23	B	laugh	2.817	3.123	3.367	3.383
Y04	E	arrive	2.813	3.013	3.217	3.282
Y36	A	happen	2.674	2.946	3.137	3.278
Y26	F	work	2.549	2.865	3.143	3.236
Y09	B	work	2.621	2.696	2.808	3.010
Y24	E	die	2.571	2.706	2.915	*2.931
Y03	B	play	2.405	2.554	2.888	3.169
Y16	A	appear	*2.539	2.607	2.775	3.048
Y18	D	sing	2.414	2.661	2.819	3.005
Y28	A	exist	2.324	*2.477	2.846	3.142
Y08	F	sing	*2.566	2.637	2.601	2.745
Y31	F	cry	2.416	2.525	2.688	2.907
Y20	A	arrive	2.421	2.613	2.805	2.751
Y13	A	fall	2.409	2.583	2.675	2.799
Y10	E	happen	2.345	2.524	2.755	2.818
Y17	F	dance	2.348	2.423	2.678	2.936
Y07	C	exist	2.358	*2.582	2.694	2.804
Y48	F	play	2.268	2.395	2.643	2.858
Y34	C	arrive	2.243	2.328	2.627	2.866
Y45	E	appear	2.310	*2.521	2.612	2.661
Y35	E	exist	2.126	2.380	2.594	2.841
Y29	D	play	2.125	2.293	2.627	2.874
Y14	D	cry	2.209	2.315	2.561	2.588
Y47	D	work	2.206	2.287	2.442	2.665
Y38	D	dance	2.152	2.214	2.367	2.581
Y39	F	laugh	2.270	2.347	2.316	2.300
Y12	D	laugh	1.937	1.982	2.073	2.171
Y02	C	appear	1.822	1.899	2.041	2.343
Y41	E	fall	1.933	1.938	1.873	2.085
Y43	C	die	1.790	1.780	1.916	2.116
Y27	C	fall	1.928	1.831	1.841	1.847
Y22	C	happen	1.736	1.623	1.685	1.848



次に、表 3 は 36 問の項目を 6 つ (A~F) のカテゴリごとに並び変えたものである。この表では 6 つのカテゴリはカテゴリ全体の平均点によって、平均点の高いものから低いものの順に配置されている。したがって、難易度順は、簡単なものから  $B < A < F < E < D < C$  となっている。

また、各カテゴリにおける 6 つの動詞の難易度順序も、平均点の高いものから低いものの順に配列されている。カテゴリ B のように、どのランクでも動詞の難易度順序がほとんど一致しているものもあれば、他の多くのカテゴリに見られるように、ランク間で変動の見られるものもある。

表 3 から 4 つの潜在ランクの学習者の発達段階を見ると、学習者がランク 1 からランク 2 に移行するにつれ、カテゴリ A と B の平均点が上昇している。すなわち、これらのカテゴリのすべての動詞について、2.5 以上の正しい判断ができるようになってきている。次に、ランク 2 からランク 3 に移行するにつれ、特にカテゴリ E と F の得点が上昇している。最後に、ランク 3 からランク 4 に移行するにつれ、カテゴリ A と B の判断がほぼ正しく (3.0 以上) なってきている。

加えて、表 3 から明らかなのは、同一の動詞にもかかわらず、用いられる構造 (カテゴリ) によって、難易度が変化しているものがあるという特徴である。例えば、非対格動詞の *die* や *happen* はカテゴリ A や E においては、比較的 average 点の高い動詞となっている一方で、カテゴリ C においては、かなり困難な項目となっている。

カテゴリ C は最も困難なカテゴリであり、ランク 3 やランク 4 の学習者でさえ、多くの項目については半数近い項目が 2.0 以下、つまり非文であるのに適格文であると判断されている。

表 3 各ランクにおける各カテゴリの動詞の困難度順序

Category	Difficulty order	Rank 1	Rank 2	Rank 3	Rank 4
B (unerg.) (NP+V)	1	sing	sing	sing	sing
	2	cry	cry	cry	cry
	3	dance	dance	dance	dance
	4	laugh	laugh	laugh	laugh
	5	work	work	play	play
	6	play	play	work	work
A (unacc.) (NP+V)	1	die	die	die	die
	2	happen	happen	happen	happen
	3	appear	arrive	exist	exist
	4	exist	appear	arrive	appear
	5	arrive	fall	appear	fall
	6	fall	exist	fall	arrive
F (unerg.) (*NP+V+NP)	1	sing	work	work	work
	2	work	sing	cry	dance
	3	cry	cry	dance	cry
	4	dance	dance	play	play
	5	play	play	sing	sing
	6	laugh	laugh	laugh	laugh

E (unacc.) (*NP+V+NP)	1	arrive	arrive	arrive	arrive
	2	die	die	die	die
	3	happen	happen	happen	exist
	4	appear	appear	appear	happen
	5	exist	exist	exist	appear
	6	fall	fall	fall	fall
D (unerg.) (*NP+be+p.p.)	1	sing	sing	sing	sing
	2	cry	cry	play	play
	3	work	play	cry	work
	4	dance	work	work	cry
	5	play	dance	dance	dance
	6	laugh	laugh	laugh	laugh
C (unacc.) (*NP+be+p.p.)	1	exist	exist	exist	arrive
	2	arrive	arrive	arrive	exist
	3	fall	appear	appear	appear
	4	appear	fall	die	die
	5	die	die	fall	happen
	6	happen	happen	happen	fall

#### 5.4 考察

前節では、NTTによる分析結果を報告した。上記の分析結果から、本研究における非対格動詞・非能格動詞の習得に関して、以下の4つの習得上の特徴を主張することができる。

##### ①学習者の潜在ランク

本調査の実験参加者に関しては、潜在ランク数は試行の末4つに設定された。その結果、上記で述べたような各ランク間の発達上の特徴がみられた。ここから各ランクの can-do list の記述を試みることにする。まず、ランク1の学習者はカテゴリBの判断が正しくできており、次にランク2ではカテゴリAの判断が正しくできるようになる。ランク3ではカテゴリEとFの判断が正しくできるようになり、ランク4ではカテゴリDの判断が正しくできるようになる。しかしながら、ランク4の学習者でさえ、カテゴリCの判断に関してはまだ十分な正確さになっていない。これは本研究の実験参加者の英語能力の上限が十分に高くなかったことによるものかもしれないが、指導の際には、カテゴリCのような誤りへの対応については、指導者側は十分に配慮する必要がある。

##### ②カテゴリの習得順序

上述したように、Unit Yの各カテゴリの難易度順序は、 $B < A < F < E < D < C$ であった。このことから、同じ文法構造であれば、学習者は常に非能格動詞よりも非対格動詞の方を困難に感じている ( $B < A$ ,  $D < C$ ,  $F < E$ ) ということの意味する。また学習者は、自動詞構文の適格文 (カテゴリAとB) から正しい判断ができるようになり、次に他動詞構文の非文 (カテゴリEとF) の判断が正しくなり、最後に受動態の非文 (カテゴリCとD) の判断ができるようになる。しかしながら、特にカテゴリC (非対格動詞が受動態の構造で用いられたもの) の判断はランク4の学習者にとっても困難であることがわかる。これらのことは、我々の過去の研究 (例えば、Yamakawa et al. 2003; 山川他, 2005; Yamakawa et al. 2008 など) でも同様に見られた結果である。このことから、今回NTTを用いた分析結果は過

去の分散分析 (ANOVA) や IRT を用いた分析結果と類似していることから、NTT で明らかになるテスト項目の困難度順序はかなりの妥当性があると主張できる。

### ③非対格動詞の誤りの原因について

4.2 で述べたように、先行研究では非対格動詞の誤りの原因に関して、NP 移動説と語彙使役化説が存在する。本研究は、誤りの原因について直接的に調査する性質のものではないため、これら 2 つの説に関して断定的にどちらかの優位性を主張することはできない。しかしながら、誤りの原因について、興味深い事実がいくつか観察された。

まず、*fall* という非対格動詞に注目してみる。表 3 からわかるように、この動詞はカテゴリ E とカテゴリ C において、ほとんどすべてのランクで平均点が 2.0 以下 (つまり非文であるのに適格文の判断) となっている。カテゴリ E (他動詞用法の非文) においては、4 つのどのランクでも最も困難な項目となっている。カテゴリ E の平均点が低いということは、学習者は *fall* を他動詞化しているということであり、カテゴリ C の場合はこの *fall* の他動詞用法を受動化していると考えられる。すなわち、*fall* に関しては、語彙使役化が誤りの原因と考えることができる。このことは、*fall* に対応する日本語の「落ちる・落とす」の自他関係の存在が影響している可能性がある。

一方、*die* という動詞に関しては、カテゴリ E においては、どのランクでも比較的平均点が高いにもかかわらず (2.5~3.0 の範囲内)、カテゴリ C においては、ランク 1~3 で 2.0 以下である。すなわち、*die* の場合は、他動詞化していないにもかかわらず、受動化が行われていることになる。したがって、*die* に関しては、NP 移動説の可能性の方が高いと考えられる。同様の現象は、*happen* でも見られる。加えて、*die* と *happen* のどちらも、日本語では、「死ぬ・死なす」「起きる・起こす」のような自他関係の語彙が存在するにもかかわらず、他動詞化が生じることなく、受動化が生じている。このことは上で見た *fall* の場合と異なっている。

以上のことから、非対格動詞の誤りにについて少なくとも本研究からは、動詞によって誤りの原因が異なる可能性を指摘できる。言い換えれば、「非対格動詞」というカテゴリに属する成員 (各動詞) が、同じような習得の道すじに従うのではなく、ある程度の多様性を持つことが示唆されている。ただ、非対格動詞が同じ自動詞の非能格動詞と比較してより困難であることは、本研究のみならず先行研究からも支持されていることから、非対格動詞という「カテゴリ」の存在をすべて否定することはできない。しかしながら、カテゴリ内での何らかの *asymmetry* (不均整) は想定することができ、これに対する理論的説明が待たれるところである。

### ④発達の U 字曲線について

4.3 では、非対格動詞の習得の道すじとして、Oshita (2001) の非対格わな仮説 (UTH) を概観した。本調査においても、この仮説を支持するようなデータが観察された。以下の図 6 と図 7 はそれぞれ、Unit Y の項目である次の (13) と (14) の項目カテゴリ参照プロファイルである。

(13) My grandmother *died* when I was five. (項目 Y06: カテゴリ A)

(14) \*I was not there when the accident *was happened*. (項目 Y22: カテゴリ C)

項目カテゴリ参照プロファイル (item category reference profile: ICRP) とは、本調査で用いたような段階ニューラルテストにおいて、それぞれの潜在ランクにおいて当該カテゴリを選択する確率を表現している。一般的に、上位の潜在ランクに所属する受験者ほど上位のカテゴリを選択するようになる。例えば、項目 Y06 は、非対格動詞の中で最も平均点の高かった項目である。この場合、「完全に可能」(4 点を示す線) は、ランクが上がるにつれて高くなり、ランク 4 では、約 80% の学習者が「完全に可能」の判断を示す可能性を表している。逆にランクが上がるにつれて、他の選択肢は下降の一途をたどり、「多分可能」(3 点) を選択した学習者でさえもランク 4 では約 10% となっている。図 6

はランクが上がるにつれて、文法性判断が正しくなる典型的なパターンである。

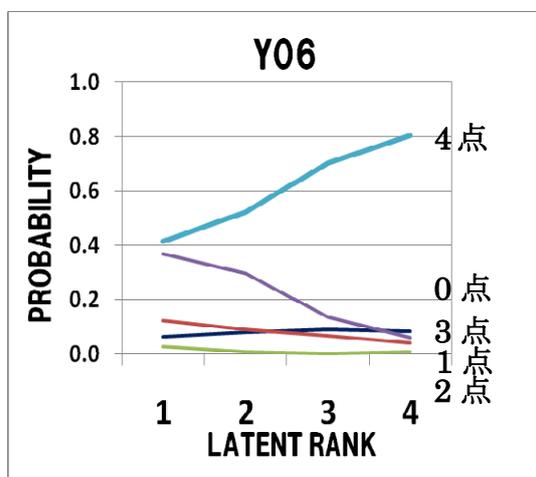


図6 項目 Y06 の項目カテゴリ参照プロファイル

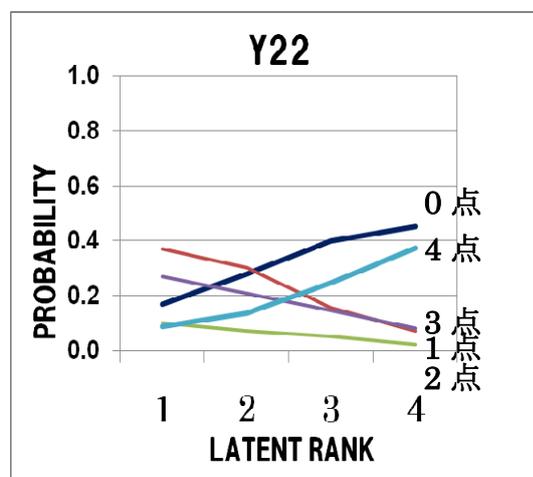


図7 項目 Y22 の項目カテゴリ参照プロファイル

一方、図7の項目 Y22 は最も困難度の高かった項目であるが、ここでは、「完全に可能」(4点)を選択する確率はランク4で約40%となっているので、項目 Y06 と比較するとかなり困難であることが分かる。加えて、非文である項目 22 を「完全に可能」(0点)と判断する確率も、ランクが上がるにつれて増加しており、ランク4では40%以上となっている。このことは、当該文法項目に関する学習者の能力が上がるにつれ、4点を獲得するものと、0点を獲得するものに二分されていくことを意味する。別の言い方をすれば、ランクが上がるにつれて、学習者の多くは非対格動詞について何らかの「再解釈・再構築」を行い、結果として、誤った判断をするようになることを意味する。このような反応を示す項目は、特にカテゴリ C と D に多かった。すなわち、困難度が相対的に高いカテゴリにこの現象は観察されると言える。

このことは、4.3で概観した Oshita (2001) の UTH を反映していると解釈することが可能である。すなわち、UTH では第一段階から第二段階へと学習者の発達段階が進むにつれて、非対格動詞の D 構造と受動態の D 構造の類似性に気付き始め、非対格動詞の受動化を容認しているのであり、本調査のランク4の学習者の約40%は、まさしくこの第二段階にいと仮定することができる。ただし、「完全に不可能」(4点)という判断をしている学習者もほぼ同数存在するので、ランク4にいる学習者全員が Oshita (2001) の UTH の第二段階に属しているわけではない。そしてさらに言うのであれば、本調査で判明した4つのランクのうち、Oshita (2001) の発達段階の第三段階のものが大部分属するランクというものはなかったといえる。つまり本調査の実験参加者は、非対格動詞の習得段階が相対的に高くなかったということが考えられる。これは本セクションの上記①の議論にも反映されていると考えられる。

## 6 結論

本研究では、日本人英語学習者の非対格動詞と非能格動詞の習得研究において、NTTの援用を試みた。分析の結果、学習者の潜在ランクに基づいて学習者を4つのランクに分けることとなった。そして、それぞれのランクにおける習得上の特性の記述をすることができた。加えて、文法カテゴリやテスト項目に関しては、我々の過去の研究と整合性を示す結果が出た一方で、項目カテゴリ参照プロファイルの分析を通して、多くの学習者の習得段階において、習得のU字曲線の存在を支持するような

現象を新たに観察することができた。

ANOVA や IRT を用いた従来の我々の研究では、各項目の難易度順序は把握できたが、今回のように、学習者を意味のある発達段階（潜在ランク）に分けながらテスト項目の難易度順序を明らかにすることはできなかった。今回 NTT を用いることによって、恣意的な判断ではなく、統計的処理において学習者を差異のある段階に分けることが可能になった。今後は、実験参加者数を増やした際に、今回と異なるランク数になるのか興味を引くところである。また、今回は単一の文法項目のみに焦点を当てたが、今後は複数の文法項目のデータを用いて分析することにより、Yamakawa et al. (2008) では判明しなかった、より広範囲の文法の発達段階について明らかになることが期待される。

## 謝辞

本研究の実施にあたっては、科学研究費助成（基盤研究 B）（平成 13 年～15 年：課題番号 1348064、平成 16 年～18 年：課題番号 16320078 及び平成 19 年～21 年：課題番号 19320091）を得た。また、NTT の理論的背景の解説については、大学入試センターの荘島宏二郎先生に多大なご支援をいただいた。ただし、本論文での誤りはすべて筆者らの責任の負うところである。

## 注

- 1 本論は、AAAL (American Association for Applied Linguistics) のアトランタでの大会において 2010 年 3 月 6 日に行った発表 (The Possibility of Application of Neural Test Theory to SLA Research) をもとに草稿化したものである。
- 2 NTT についての詳細は、Shojima (2009a, 2009b) などを参照のこと。
- 3 Exametrica は <http://www.rd.dnc.ac.jp/~shojima/exmk/jindex.htm> からダウンロードが可能である。
- 4 図 1～6 は、荘島氏のサイト (<http://www.rd.dnc.ac.jp/~shojima/ntt/index.htm>) から引用したものである。

## 参考文献

- Balcom, P. (1997). Why is this happened? Passive morphology and unaccusativity. *Second Language Research*, 13, 1-9.
- Deguchi, A. & Oshita, H. (2004). Meaning, proficiency and error types: Variations in nonnative acquisition of unaccusative verbs. In S. Foster-Cohen, M. Sharwood Smith, A. Sorace, & M. Ota (Eds.), *EuroSLA Yearbook 2004* (pp.41-65). Amsterdam: John Benjamins.
- Hirakawa, M. (1995). L2 acquisition of English unaccusative constructions. *Proceedings of the 19th Annual Boston University Conference on Language Development* (pp.291-302). Mass: Cascadilla Press.
- Hirakawa, M. (1997). On the unaccusative/unergative distinction in SLA. *JACET BULLETIN*, 28, 17-27.
- Hirakawa, M. (2003). *Unaccusativity in second language Japanese and English*. Tokyo: Hituzi Syobo.
- Ju, M. K. (2000). Overpassivization errors by second language learners. *Studies in Second Language Acquisition*, 22, 85-111.
- Kim, J-T. (2006). Unaccusativity and L2 passive construction. *English Language & Literature Teaching*, 12, 69-89.

- Kondo, T. (2005). Overpassivization in second language acquisition. *IRAL*, 43, 129-161.
- Kwak, H-Y. (2003). A study of Korean high school students' overpassivization errors. *Journal of Pan-Pacific Association of Applied Linguistics*, 7, 169-187.
- 中村洋一 (著)・大友賢二 (監修). (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』 東京: 桐原書店.
- Oshita, H. (1997). *The unaccusative trap: L2 acquisition of English intransitive verbs*. Ph.D. dissertation. University of Southern California, LA, California.
- Oshita, H. (2000). *What is happened* may not be what appears to be happening: A corpus study of “passive” unaccusatives in L2 English. *Second Language Research*, 16, 293-324.
- Oshita, H. (2001). The unaccusative trap in second language acquisition. *Studies in Second Language Acquisition*, 23, 279-304.
- Oshita, H. (2002). Uneasiness with the easiest: On the subject-verb order in L2 English. *Second Language*, 1, 45-61.
- Oshita, H., & Deguchi, A. (2005). Looking for “U”: Evidence for the unaccusative trap hypothesis. Paper presented at SLRF 2005, Teacher's College, Columbia University
- Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. *Proceedings of the Berkeley Linguistics Society*, 4, 157-189.
- Shojima, K. (2009a). Neural test theory. In K. Shigemasu et al. (Eds.), *New Trends in Psychometrics* (pp.407-416). Tokyo: Universal Academy Press.
- 荘島宏二郎. (2009b). 「ニューラルテスト理論 — 資格試験のためのテスト標準化理論—」『電子情報通信学会誌』 92号, 1013-1016.
- Shomura, Y. (1996). Perspectives on pedagogic grammar in English language teaching: A study of the acquisition of ergatives by Japanese learners. *Annual Review of English Learning and Teaching*, 1, 17-35.
- Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in on-native grammars of Italian. *Second Language Research*, 9, 22-47.
- Sorace, A. (1997). Acquiring linking rules and argument structures in a second language. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage* (pp. 153-175). Amsterdam: John Benjamins.
- Yamakawa, K., Sugino, N., Kimura, S., Nakano, M., Ohba, H., & Shimizu, Y. (2003). The development of grammatical competence of Japanese EFL learners: Focusing on unaccusative/unergative verbs. *Annual Review of English Language Education in Japan*, 14, 1-10.
- 山川健一・杉野直樹・木村真治・中野美知子・大場浩正・清水裕子. (2005). 「日本人英語 学習者による非対格動詞と非能格動詞の習得」『大学英語教育学会中国・四国支部研究 紀要』第 2 号, 91-110.
- Yamakawa, K., Sugino, N., Ohba, H., Nakano, M., & Shimizu, Y. (2008). Acquisition of English grammatical features by adult Japanese EFL learners: The application of Item Response Theory in SLA research. *Electronic Journal of Foreign Language Teaching (e-FLT)*, 5, 13-40.
- Yip, V. (1994). Grammatical consciousness-raising and learnability. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 123-139). Cambridge: Cambridge University Press.
- Yip, V. (1995). *Interlanguage and learnability*. Amsterdam: John Benjamins.
- Yuan, B. (1999). Acquiring the unaccusative/unergative distinction in a second language: Evidence from English-speaking learners of L2 Chinese. *Linguistics*, 37, 275-296.
- Yusa, N. (2003). ‘Passive’ unaccusatives in L2 acquisition. In P. Clancy (Ed.), *Japanese/Korean*

*Linguistics* 11 (pp. 246-259). Stanford: CSLI Publications.

Zobl, H. (1989). Canonical typological structures and ergativity in English L2 acquisition. In S. M. Gass, & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 203-221). Cambridge: Cambridge University Press.