



## Kobe Shoin Women's University Repository

Title	中間言語と言語変異：KYコーパスを使った「を」格ゼロマーク化の分析 Interlanguage and Variation: Zero-marking of (o) in KY Corpus
Author(s)	松田 謙次郎 (Kenjiro Matsuda)
<i>Citation</i>	Theoretical and applied linguistics at Kobe Shoin, No.4 : 57-76
Issue Date	2001
Resource Type	Bulletin Paper / 紀要論文
Resource Version	
URL	
Right	
Additional Information	

# 中間言語と言語変異: KY コーパスを使った「を」格ゼロマーク化の分析\*

松田 謙次郎

---

## Interlanguage and Variation: Zero-marking of (o) in KY Corpus

Kenjiro Matsuda

### Abstract

Recent surge in variationist analysis of interlanguage variation (Bailey & Preston (1996)) blazed a way for a detailed analysis of variable constraints in L2 acquisition. An interesting possibility here is a correlation between the language varieties (L1 / L2 acquisition, L1 variation) and the distinction between language-universal constraints and language-particular constraints. This paper analyzes a zero-marking of (o) (Matsuda 1995, 2000) in Japanese speech by Korean L2 speakers using the KY Corpus. While zero-making of (o) in Tokyo Japanese involves both a language-universal constraint (morpho-syntactic form of the object NP) and language-particular constraints (adjacency between the object NP and

---

\*本稿の執筆に当たっては、平成 12 年度日本学術振興会科学研究費補助金（基盤研究 (B)）課題番号 11480052/「現代日本語の音声・意味・文法・談話における変異と日本語教育」（研究代表者: 日比谷潤子）を受けている。また、本稿の内容は 2000 年 9 月に慶應大学で開催された上記プロジェクトの研究発表会での口頭発表に基づくものである。当日フィードバックを下された方々に感謝したい。  
*Theoretical and Applied Linguistics at Kobe Shoin* 4, 57-76, 2001.  
© Kobe Shoin Institute for Linguistic Sciences.

the verb, and speech style), a quantitative analysis of the L2 speech found the language-universal constraint as the only significant factor. Moreover, its distribution showed a clear correlation with the speakers' proficiency level. In particular, the difference between lexical NPs and pronouns is already evident in mid-level L2 speakers, suggesting an important role played by the language-universal constraint both in the L2 acquisition and the L1 variation.

近年 Bailey & Preston (1996) に見られるように、変異理論的枠組みによる中間言語変異の分析が盛んになって来つつある。第2言語習得という観点を取り入れることにより、制約条件のありようについて第1言語習得、第1言語変異や言語普遍性・個別言語の特殊性と組み合わせた仮説設定が可能になる。本稿ではKYコーパスを用いて韓国語母語話者の日本語発話に見られる「を」格ゼロマーク化現象の分析を行い、松田(1995, 2000)に報告されている第1言語変異との比較を試みた。「を」格ゼロマーク化には、言語普遍的要因(目的語名詞句形式差)とともに、日本語(東京語)個別的な要因(目的語名詞句と動詞の隣接性、スタイル差)も絡むが、KYコーパスの韓国語母語話者データの分析からは、このうち目的語名詞句形式差のみが統計的に有意な要因として検出された。さらに日本語習熟度別の分布からは、習熟度の上昇に従って、この要因のレベルの分布が母語話者のそれに接近するという傾向が認められた。とりわけ普通名詞 > f 代名詞という差は、すでに中級レベルから習得されている事実は、第1言語変異と第2言語習得における言語普遍的要因の役割という点から大いに注目に値する。

## 1. はじめに

方言接触、ピジン・クレオール化現象などが如実に示すように、言語変異が異なる言語システムとの接触により引き起こされることもあることは周知の事実である。

しかし異言語システムとの接触というのであれば、第二言語学習者がターゲットとなる言語に接触し、その文法体系を習得する過程も異言語システムとの接触として捉えることができるはずである。言語体系を揺れを含んだものとしてとらえ、揺れを支配する文法を日常言語の分析を通して明らかにしていこうとしていたのが変異理論 (Variation Theory) であるならば、当然こうした第二言語習得の途上で現れる変異現象には、うってつけのパラダイムとなる (Bailey & Preston

(1996))。近年中間言語 (interlanguage) として注目を浴びつつあるこのような言語体系の研究は、言語接触現象として一般化することで、ピジンクレオール現象を初めとする各種言語接触現象との差異と類似性などをも絡めた、肥沃な変異研究のリサーチプログラムへと繋がって行くことになるわけである。

こうした問題意識の下に、本稿では韓国語母語話者の発話コーパスを用いて、特定日本語文法変異に見られる諸制約がどのように彼らの中間言語に反映されるのかを、数量的手法を通じて検証してみたい。取り上げる現象は、すでに第一言語変異で詳細な事実が明らかになっている「を」格ゼロマーク化である。すでに第一言語変異でのありようが明らかなることを用いて、中間言語変異の分析が容易になるからであり、また第一言語変異との詳細な比較が可能になるからである。コーパスデータの統計的分析を通して、いずれの変異にも、言語的普遍的 / 個別言語特有的特徴という差が大きく関与していることを示し、第一言語・中間言語という二言語変種に見られる変異現象について、ある種の一般化が可能であることを実証する。

## 2. 中間言語・第一言語習得・(母語内) 変異

さて第1・2言語習得、第一言語変異の三体系にわたる変異の説明原理を求めに当たり、ここではその一つの可能性として言語普遍性の概念を取り入れてみたい。

当然言語変異現象もそれとの関わりで解明される部分が少なくないはずである。普遍的制約なり規則があるとするならば、先の三体系における変異の発現にも普遍の部分と個別の区別が反映されると考えるのが自然である。ここで仮説的に普遍性と三体系における変異のありよう単純に考えた場合、その関係は以下のようになろう：

表 1: 普遍・個別の区別と三言語変種 (第一言語習得・第二言語変異・第一言語変異) の関わり

	普遍的原理	各言語特有要因
第一言語習得	早くから現れる	遅くに現れる
第二言語変異	要因として存在する	話者の習熟レベルによる。
第一言語変異	要因として重要 (比較的重みが大きい)	周辺的 (比較的重みが小さい)

つまり、普遍的部分に含まれる制約や規則（普遍的原理）は第一言語習得では当然話者の発話では早くから観察されるであろうし、そうした要因は習得後の変異でも何らかの形で少なからぬ影響力を発揮する可能性が高いであろう。また第二言語話者と言えどもそのような制約・規則の存在は言語的知識の一部として持っているはずであり、その発話の変異にも反映されているはずである。逆に特定言語特有な要因は、第一言語習得では普遍部分の制約に比して発現により時間がかかろうし、第一言語変異においてはそのウェイトも比較的低いものとなる。そして第二言語話者にとってもそうした制約・規則を習得するのは困難なはずである。三言語変種と普遍・個別という区別を組み合わせた場合、もっとも簡単な仮説は以上のようなものと思われる。

本稿では、この表のうち第二言語習得に関わる部分に集中して、<sup>2</sup>以下の3仮説を検討してみたい：

第二言語話者の中間言語に見られるバリエーションの制約条件においては、

1. 言語普遍的制約・規則は早くに習得され、要因としての重みも重い
2. 個別言語の特殊なもの（各言語個別要因）は遅くに習得され、要因としての重みが軽い
3. 制約条件の重みは第二言語話者のレベルが上昇するにつれ母語話者のそれに接近する

これらの仮説を検討するに当たり、ここでは韓国語母語話者の日本語発話コーパスデータを用いて、「を」格ゼロマーク化の変異に注目し、第一言語話者の同一変異現象との比較を行う。以下で述べるように、「を」格ゼロマーク化現象は日本語において様々なアプローチにより分析されてきているという歴史を持ち、そのありようについてもかなりなことがわかってきている。その点でこうしたデータを検討するには都合がよい。比較の中心は、変異を司る制約条件 (constraints) に置き、第一言語と中間言語の変異それぞれをコントロールする言語的制約条件の間に、どのような共通性が見られ、どのような差異が見られるかを分析することとする。

<sup>2</sup>宮田(1992)は日本語格助詞の第一言語習得を扱い、公刊されているデータとともに自身で収集したデータを用いて興味深い議論を展開している。

### 3. 日本語「を」格ゼロマーク化現象

日本語の「を」格ゼロマーク化現象は、これまでに生成文法や語用論、自然言語処理、そして社会言語学の立場から様々な分析がなされてきており (Tsutsui 1984, Saito 1985, Masunaga 1988, Hosaka et al. 1992, Matsuda 1995, 松田 2000)、その制約条件のありようについても、自然談話資料を使った分析 (Matsuda 1995) によりかなり詳細な部分についてまで判明している。Matsuda (1995)、松田 (2000) によれば、その主たる言語内的 (性別、年齢といった社会的要因を除いた純粋に文法の枠内に収まる要因) は、以下のように整理される:

**目的語名詞句と動詞の隣接性:** 隣接している方が、していない場合よりも統計的に有意にゼロマーク化が起きやすい。ただし、目的語名詞句と動詞の間に入る要素のうち、オノマトピアと数量詞は隣接性を阻害しない、「透明」な要素である。

**目的語名詞句の形式差:** 目的語名詞句の形式によって、疑問詞 > 普通名詞 > 代名詞 > 節 の順でゼロマーク化を引き起こしやすい (統計的有意差あり)。

**スピーチスタイル:** 砕けたスタイルの場合に、改まったスタイルの時よりもゼロマーク化が統計的に有意に起きやすい。

先に述べた日本語特有の要因 / 間言語的要因という区別に従うのであれば、主要 3 要因 (動詞と目的語の隣接性、目的語名詞句の形式、スピーチスタイル) のうち、目的語名詞句の形式は間言語的要因と考えられる。すでに Matsuda (1995) に指摘があるように、系統的に無関係とされる言語間にあっても、格標示と名詞句形式に高い相関性が認められるからである。<sup>3</sup>あとの 2 つは日本語特有と考えて良い。「動詞と隣接していない場合に名詞句の格標示を明示的にし、隣接の場合は随意とせよ」という規則は、もちろん日本語 (の一方言) 特有の規則であり、言語普遍的なものではない。それどころか、同じ日本語方言であっても関西方言では隣接していない場合でもゼロマーク化が頻繁にされており、むしろ明示的にされているケースの方が珍しい。つまり隣接性は日本語特有な規則であるから、

<sup>3</sup>この相関関係は正確には、「ある明示的な格表示を持つ言語があれば、少なくともその言語体系における代名詞には必ず格表示があるが、普通名詞には必ずしも格標示があるとは限らない」という一般化として定式化される。

中間言語変異において現れなくてもそう不思議ではない。スタイル差についてはいうまでもないであろう。よって先の仮説をゼロマーク化に乗っ取って設定し直すと、次のようになる:

#### 中間言語話者発話における「を」格ゼロマーク化の仮説

1. 目的語名詞句の形式差は中間言語話者の発話にも重要な要因として早くから存在するのに対して、
2. 隣接性やスタイル差は中間言語話者の発話には要因としてはそれほど重要ではなく、また早くから存在しない。
3. 変異の習得差は、日本語の習熟レベルに比例する

今回の考察対象からは、スタイルは除外することにした。第二言語習得者のスタイルはそれ自体興味深いテーマであるが、(1) KY コーパスをざっと読み通した上であまり大きなスタイル上の変化が見あたらない、(2) レベルの異なる学習者すべてに適用できる、信頼し得るようなコーディング法が見あたらない、という2つの理由から今回は要因としては見送ることにしたわけである。試みにデータを各々のインタビューを単純に文字化行数で二分し、前半と後半で分けた上での分析も施したが、全く何らの相関も見いだせなかった。学習者にとってインタビューはかなり緊張を強いる状況であり、慣れるに従ってスピーチスタイルに変化が生じるものと予想されたためこのようなコーディングを行ったものである。結果は習熟レベルとクロスさせた上でも同様であった。自己の緊張や改まりの度合いに応じて話し言葉を変化させるほどには学習者の日本語能力が習熟していないものと考えることができよう。将来的には変異理論研究で用いられている、Labov and Sankoff (1985) のように、トピックを中心にしてスタイルを分類する方法も試みる価値があるだろうが、第一言語話者の場合とは異なる捉え方で臨まない限り第二言語習得者のスタイルは十分に分析し得ないというのが筆者の印象である。

スタイルを要因と含めることは見送ったが、そのかわりにここでは人間性(humanness)を要因として考察することとした。Matsuda (1995) では、少なくとも東京方言では目的語名詞句の人間性は全くゼロマーク化に関わりがないことが確認されているが、巨視的に見て、日本全国の方言分布を視野に入れた場合、この要因が大きく作用していることも事実である。『方言文法全国地図』第6図「酒を(飲む)」および第7図「おれを(連れて行ってくれ)」(国立国語研究所 1989: 6-7)

に現れた分布では、「おれ」の方が明らかに「を」がより多くの地点で使われており、人間性がゼロマーク化に一役買っている可能性を示唆している。<sup>4</sup>今回は分析対照が中間言語であり、他方言で強い要因となっているファクターが効いてくる可能性もあるので、探索的見地からこの要因を含めておくことにした次第である。

#### 4. 要因・データ・分析手順

##### 4.1 要因

まとめると、本研究で検討する要因は以下ようになり、それぞれについてコーディングに当たって従ったルールを記しておく。

1. 動詞と目的語の隣接性 I: 隣接 / 非隣接
2. 動詞と目的語の隣接性 II: 隣接 / 非隣接
3. 目的語名詞句の形式: 普通名詞 / 代名詞 / 指示代名詞+名詞 / Wh 句 / WH+名詞句 / 節
4. OPI レベル (習熟度) : S / A / I / N
5. 人間性: 人間 / 非人間

「動詞と目的語の隣接性 I」では、ほぼ機械的に動詞と目的語名詞句が直接隣接していれば「隣接」とし、それ以外を「非隣接」とコード化した。これに対して、「動詞と目的語の隣接性 II」では、Matsuda (1995)、松田 (2000) の知見に鑑み、オノマトピアと数量詞が隣接性への例外であると見なし、これらの要素のよってのみ動詞と目的語名詞句が分けられている場合は、「隣接」とコード化している。ただし、いずれの場合も相手のうなずきのみで分けられている場合は、隣接と判断した。話者自身のうなずきや、「えーと」の類はこの限りではない。

また、全般的に以下のような場合にはデータとして採用することを見合わせている:

- ゼロ形が「は」や「が」と紛らわしい場合。述部が状態動詞などで目的語を「が」で受けうる場合は一律に除外した。

<sup>4</sup>ただし実はこの分布が人間性のみによるものとは断言できない。「おれ」という代名詞の使用による効果 (代名詞→ゼロマーク化され難い) と考えても説明可能だからである。



- 決まり切ったいい方、イディオムの類 (例: お気をつけて)
- 名詞句だけで、動詞がない場合
- 「漢語+する」の形で、漢語の前に修飾語がない場合 (例: 日本語の勉強をする → ○; 勉強する → ×)
- 「に」との混同など、文法的に誤りと思われるもの<sup>5</sup>

#### 4.2 データ: KY コーパス概要

KY コーパスは鎌田修氏と山内博之氏によって編集された、英語・韓国語・中国語を母語とする日本語学習者のインタビュー発話を文字化した電子化テキストである。<sup>6</sup> 3言語それぞれ30名づつ、計90名についてOPI (Oral Proficiency Interview) 方式のインタビューを行い、それをテキストフォーマットで文字化している。

内容としては、テストター (T) と被験者 (S) の会話という形態をとっており、日本語学習歴、ロールプレイなども含まれる。30人のOPIの判定結果別の内訳は、それぞれ、初級 (Novice) 5人、中級 (Intermediate) 10人、上級 (Advanced) 10人、超級 (Superior) 5人ずつである。各話者のデータは個別ファイルに収められており、ファイルは「3つのローマ字 (被験者の母語 (English, Chinese, Korean) + 言語能力レベル+サブレベル (L/M/H)) + 2桁の通し番号」 (但し超級にサブレベルはなく、上級もHとただの上級のみ) というフォーマットでファイル名が付されている。コーパスの表記は、漢字かな混じり・分かち書きなしであり、同一話者の発話は一行に収めてある。その他、以下のような文字化表記規則を持つ:

- || : 非言語的な行動 例: {笑い}
- <> : 相づち表記 (同一行内に表記) 例: <あ、そうですか>
- [] : 固有名詞 例: [大学名] [人名]
- 、 : ポーズ表記。「、」の数はポーズ長に比例する。
- () : 記録上不明瞭な箇所

<sup>5</sup>文法的エラーについてのエラーアナリシスは、当然今後の課題の一つである。

<sup>6</sup>以下の記述は、KY コーパス version 1.1 配布時に同梱されている readme.txt ファイルに基づく。

- \*: 全く聞き取れない箇所

ただし、KY コーパスには、テスターに関する説明がない点は指摘しておく必要がある。テスターの言語的・社会的背景、またテスターは結局何人いて誰がどのインタビューをしているのか、といった情報も望まれるのである。

### 4.3 分析手順

まず KY コーパスの韓国語話者のファイルから、該当部分を話者単位で Microsoft Excel 2000 にコピーし、必要最小限のテキスト部分へ編集した。<sup>7</sup>次に各データについてそれぞれの要因について上記の要領でコーディングを施した。コーディングを終えた段階でデータをテスターと被験者に分離し、それぞれで各要因別のゼロマーク化率を計算し、必要に応じてクロス集計を行った。最終的に分析に使われたデータ総数は 691 件である。

次に、ここまでの二変量による分析からさらに踏み出して、複数の独立変数を使って中間言語話者の「を」格ゼロマーク化の予測を試みる。独立変数は順序尺度である OPI レベルを除いて三つ（隣接性、目的語名詞句形式差、人間性）が名義尺度であり、従属変数であるゼロマーク化・非ゼロマーク化も名義尺度である。こうした複数の名義尺度によって名義尺度変数の予測をし、各パラメータの推測をするには、ロジスティック回帰を行うのがよい。<sup>8</sup>ここでは GoldVarb version 2.0 (Rand and Sankoff 1990) を使ってこの処理を行った。以下被験者データののみ絞って報告する。

## 5. 分析

### 5.1 隣接性

まず、隣接性 I から検討してみよう。次表は被験者全体の目的語・動詞の隣接・非隣接別とゼロマーク化の相関を示したものである。72.6% (502/691) と、大方のデータは隣接・「を」に集中しており、そもそもゼロ化自体が多くは見られないが、それでもこの表からは2つの統語的ポジションの差によるゼロマーク化率の差はたかだか6%程度であり、母語話者の時のように劇的な差は見られないことがわかる。事実カイ自乗検定でも有意差は検出できなかった ( $p < 0.10$ )。では母

<sup>7</sup>この作業で、鎌田尚子、中村勝、西村麻里の協力を得た。

<sup>8</sup>ロジスティック回帰の詳細については、エバリット (1977)、Fienberg (1989) などを参照のこと。

語話者で隣接性に関して透明性を発揮したオノマトピアと数量詞を隣接として計算した、隣接性 II ではどうだろうか。

表 2: 隣接性 I によるゼロマーク化の分布 (カッコ内は実数を示す)

隣接・非隣接	を	ゼロ	合計
隣接	84.94% (502)	15.06% (89)	100.00% (591)
非隣接	91.00% (91)	9% (9)	100.00% (100)
合計	85.82% (593)	14.18% (98)	100.00% (691)

隣接性 II でデータを整理し直したのが次の表である。2つのポジションの差は隣接性 II になって5%程度とむしろ縮小しており、言うまでもなくカイ自乗値からも全く有意な差は見られない ( $p < 0.19$ )。隣接性の差が出ていないのは、隣接性のまとめ方によるものではなく、そもそも隣接性自体が無関係なわけである。上記2つの表から、母語話者の場合と大きく異なり、中間言語話者にとって隣接性がゼロマーク化の要因として全く機能していないことになる。

表 3: 隣接性 II によるゼロマーク化の分布

隣接・非隣接	を	ゼロ	合計
隣接	85.17% (517)	14.83% (90)	100.00% (607)
非隣接	90.48% (76)	9.52% (8)	100.00% (84)
合計	85.82% (593)	14.18% (98)	100.00% (691)

しかしこれらの表はあくまでデータ全体をまとめたものであり、日本語習熟度の差によって隣接性の効き方に差が出てくる可能性も考えられる。日本語の習得が進むに従って、より細かな「変異の文法」も修得していると考えerことは自然であろう。つまり OPI レベルが N → I → A → S と高くなるに従い、中間言語話者が統語的なポジションとゼロマーク化の関連に気付き、徐々に2つの隣接性のレベル差が拡大する、というシナリオである。この仮説は、下掲の表のように隣接性 I/II と OPI レベルをクロスさせることで確認できる。

表 4: 隣接性 I と OPI レベルのクロス表

OPI レベル	隣接・非隣接	を	ゼロ	合計
S	隣接	83.51% (157)	16.49% (31)	100.00% (188)
	非隣接	89.74% (35)	10.26% (4)	100.00% (39)
A	隣接	82.76% (168)	17.24% (35)	100.00% (203)
	非隣接	90.32% (28)	9.68% (3)	100.00% (31)
I	隣接	89.19% (165)	10.81% (20)	100.00% (185)
	非隣接	93.33% (28)	6.67% (2)	100.00% (30)
N	隣接	80.00% (12)	20.00% (3)	100.00% (15)
合計		85.82% (593)	14.18% (98)	100.00% (691)

初級(N)レベル話者のデータが少なく、特にここで非隣接のデータがないことが惜しまれるが、他の3レベルを比較してもその差に何ら一貫した方向性がないことは明らかである。OPI レベルとクロスしても、ゼロマーク化率と隣接性の相関は浮かび上がってこないということは、中間言語話者は初級レベルから超級レベルに至るまで徹頭徹尾隣接性を制約条件として隣接性を受け入れてはいないことになるわけである。

### 5.2 目的語名詞句の形式差

目的語名詞句も母語話者のゼロマーク化では大きな要因であった。それが中間言語話者でも同様に制約条件として機能していることを示したのが以下の表である。ここでは疑問詞+普通名詞(例: どんな考え)と疑問詞(例: 何を紹介すればいいでしょう)、そして名詞(例: TV見たり)と指示代名詞+普通名詞(例: そういう親をおっぼり出すのは)をそれぞれ一つのグループにまとめてある。よってカテゴリーは計4つ(普通名詞・代名詞・疑問詞・節)である。表を見ると、もっともゼロ化率の高い疑問詞とゼロ化率最低の代名詞では18%もの差があり、疑問詞 > 普通名詞 > 節 > 代名詞の順でゼロ化率が低下している様子がよくわかる。つまりこれが中間言語話者全体での目的語名詞句の制約条件のランキングになるわけである。この差についてはカイ自乗検定でも有意差が検出されており

表 5: 隣接性 II と OPI レベルのクロス表

OPI レベル	隣接・非隣接	を	ゼロ	合計
S	隣接	84.10% (164)	15.90% (31)	100.00% (195)
	非隣接	87.50% (28)	12.50% (4)	100.00% (32)
A	隣接	82.78% (173)	17.22% (36)	100.00% (209)
	非隣接	92.00% (23)	8.00% (2)	100.00% (25)
I	隣接	89.36% (168)	10.64% (20)	100.00% (188)
	非隣接	92.59% (25)	7.41% (2)	100.00% (27)
N	隣接	80.00% (12)	20.00% (3)	100.00% (15)
合計		85.82% (593)	14.18% (98)	100.00% (691)

( $p < 0.039$ )、全くと言っていいほど差のなかった隣接性と好対照をなしている。

表 6: 名詞句形式差によるゼロマーク化の分布

名詞句形式	を	ゼロ	合計
普通名詞	84.48% (490)	15.52% (90)	100.00% (580)
代名詞	97.62% (41)	2.38% (1)	100.00% (42)
疑問詞	79.17% (19)	20.83% (5)	100.00% (24)
節	95.56% (43)	4.44% (2)	100.00% (45)
合計	85.82% (593)	14.18% (98)	100.00% (691)

さて、すでに見たとおり、母語話者のランキングは、ゼロ化率の高い順に疑問詞 > 普通名詞 > 代名詞 > 節となっており (松田 2000)、最後の 2 つが中間言語話者と異なっている。そこで、隣接性の場合と同様、OPI レベルとの関連性を調べてみよう。習得順とランキングの相関を考えるのであれば、N から始まり、徐々に制約条件が習得され、I → A と経て S に至って母語話者と同様な制約条件のランキングが習得されている可能性が考えられる。すでに全体的にはほぼ期待された分布をしているが、N レベルでいきなり母語話者と同様な分布はしていないはず

であるから、この仮説は上の隣接性の時よりは現実性が高いものと考えられよう。

表 7: 名詞句形式差と OPI レベルのクロス表

OPI レベル	名詞句形式	を	ゼロ	合計
S	普通名詞	84.49% (158)	15.51% (29)	100.00% (187)
	代名詞	92.31% (12)	7.69% (1)	100.00% (13)
	疑問詞	54.55% (6)	45.45% (5)	100.00% (11)
	節	100.00% (16)	0.00% (0)	100.00% (16)
A	普通名詞	80.43% (148)	19.57% (36)	100.00% (184)
	代名詞	100.00% (20)	0.00% (0)	100.00% (20)
	疑問詞	100.00% (11)	0.00% (0)	100.00% (11)
	節	89.47% (17)	10.53% (2)	100.00% (19)
I	普通名詞	88.72% (173)	11.28% (22)	100.00% (195)
	代名詞	100.00% (9)	0.00% (0)	100.00% (9)
	疑問詞	100.00% (1)	0.00% (0)	100.00% (1)
	節	100.00% (10)	0.00% (0)	100.00% (10)
N	普通名詞	78.57% (11)	21.43% (3)	100.00% (14)
	疑問詞	100.00% (1)	0.00% (0)	100.00% (1)
合計		85.82% (593)	14.18% (98)	100.00% (691)

表の分布を OPI レベル別にランキングとしてまとめてみよう (“=” はランクが等しいことを示す)。

- N: NP > WH (???)
- I: 普通名詞 > 代名詞 = 節 = 疑問詞
- A: 普通名詞 > 節 > 疑問詞 = 代名詞
- S: 疑問詞 > 普通名詞 > 代名詞 > 節

ここでも乏しい初級者レベルデータがネックになっており、特に初級者レベルは一応除外してみたい。また他のレベルでも僅少なデータが見られるので、将来的なデータの積み上げ・検証が待たれることは言うまでもない。その上で言うの

であれば、中級 (I) ではまず目的語名詞句というカテゴリー全体が普通名詞とそれ以外に分化している。次にそれが上級者 (A) に達すると、普通名詞、節とそれ以外 (疑問詞と代名詞) と三分化し、最終的には超級話者 (S) において疑問詞と代名詞が分離して母語話者と全く同じランキングに到達している。つまりこれは中間言語話者がはじめは一つの固まりとして受け入れてきた名詞句のカテゴリー全体から、異なるカテゴリーを析出していく過程と考えられる。よってやはり仮説通りに、レベルの上昇に伴って、制約条件は徐々に習熟度が上がるにつれて S レベル話者、ないしは母語話者のそれに接近しているわけである。注目すべきは、これらの全段階 (データ数の少ない初級者を除く) を通して、普通名詞 > 代名詞というハイアラーキーが保持されている点である。Matsuda (1995)、松田 (2000) での主張ではこれが間言語的に見られる類型的普遍性ということであったが、それがここでも現れていることになるわけである。

### 5.3 人間性 (humanness)

人間性によってデータを分類したのが以下の表である。一見して分かる通り、人間性は全く要因として機能していない。実際にカイ自乗値を計算してみても  $p < 0.08$  と有意なレベルには達していない。名詞の意味的カテゴリーとしては基本的なものと考えられ、上述した通り日本語方言でも全国分布をコントロールする要因であるだけに大きな作用を持つものと期待されたが、中間言語話者は要因として選択しなかったわけである。

表 8: 人間性によるゼロマーク化の分布

人間・非人間別	を	ゼロ	合計
人間	93.22% (55)	6.78% (4)	100.00% (59)
非人間	85.13% (538)	14.87% (94)	100.00% (632)
合計	85.82% (593)	14.18% (98)	100.00% (691)

人間性がゼロマーク化の重要な要因でないことは、このことは OPI レベルとクロスした結果でも変わらないことから窺える。

表 9: 人間性と OPI レベルのクロス表

OPI レベル	人間性	を	ゼロ	合計
S	人間	100.00% (27)	0.00% (0)	100.00% (27)
	非人間	82.50% (165)	17.50% (35)	100.00% (200)
A	人間	80.95% (17)	19.05% (4)	100.00% (21)
	非人間	84.04% (179)	15.96% (34)	100.00% (213)
I	人間	100.00% (11)	0.00% (0)	100.00% (11)
	非人間	89.22% (182)	10.78% (22)	100.00% (204)
N	非人間	80.00% (12)	20.00% (3)	100.00% (15)
合計		85.82% (593)	14.18% (98)	100.00% (691)

#### 5.4 OPI レベル

すでに作成しているクロス表からも窺えるが、OPI レベル自体はゼロマーク化率ときれいな相関を示していない（カイ自乗検定で  $p < 0.23$ ）。特に初級レベルでは格助詞「を」自体を十分に習得していないことが考えられるわけで、たとえ格助詞「を」がゼロとして現れていても、それが学習者のエラーなのか母語話者の変異文法習得の結果なのかは、にわかに判断しがたい部分がある。よってゼロマーク化率が OPI レベルと単純な相関を示す必要性はないが、それにしても OPI レベルによる分布はランダムであり、何らの規則性も浮かび上がってこない。OPI レベルとゼロマーク化率には単純な相関はないが、名詞句形式差とクロスさせた時にのみ上で見たような規則性が見えてくるわけである。OPI レベルと一貫した関連性を示すことで、名詞句形式差の制約条件としての重要性が再確認されるのである。

#### 5.5 ロジスティック分析によるゼロマーク化予測・変数選択・パラメータ推定

ロジスティック回帰分析でステップアップ・ダウン式に変数選択を行うと、名詞句形式差のみが有意な変数として選択された。<sup>9</sup>この変数選択を行った段階では、上の分析のように疑問詞や普通名詞をまとめていなかったため、これら2つのゲ

<sup>9</sup>ステップアップ・ダウンでの閾値は 0.05 に設定した。



表 10: OPI レベルによるゼロマーク化の分布

OPI レベル	を	ゼロ	合計
S	83.76% (192)	16.24% (35)	100.00% (227)
A	84.58% (196)	15.42% (38)	100.00% (234)
I	89.77% (193)	10.23% (22)	100.00% (215)
N	80.00% (12)	20.00% (3)	100.00% (15)
合計	85.82% (593)	14.18% (98)	100.00% (691)

ループにまとめた上で、対数尤度差を用いた検定を行い、レベルをまとめることの統計的妥当性を検討した。この結果二つのグループをまとめた場合でも予測に有意差はないと判定され、実際に予測を行うと極めて高い精度で予測が可能であることがわかった。最終的なパラメータの推定値は、定数 = 0.131; 代名詞 = 0.140; 節 = 0.236; 普通名詞 = 0.550; 疑問詞 = 0.636 であり、<sup>10</sup>上の分析を裏付けるものとなっている。

## 6. 考察

上の分析では、結局名詞句形式差だけが有意な要因であるということであり、これは取りも直さず、第1の仮説、「目的語名詞句の形式差は中間言語話者の発話にも重要な要因として早くから存在する」を支持するものと解釈できる。レベル的には中級話者ですでに普通名詞 > 代名詞という差を習得しており、これがそのまま超級に至るまで保持されることは、この差の発現が単なる偶然ではないことを物語っている。類型的分布でも確認された普通名詞と代名詞の格標示差が、韓国語話者の日本語習得でも現れたことは非常に興味深いことと思われる。データ数の僅少さから詳細な議論は難しいが、I → A → S とレベルの上昇に従って制約ランキングが母語話者のそれに接近して行く様は、変異の習得差が日本語の習熟レベルに比例するという3つ目の仮説を裏付けるものと言えよう。

全く要因としての有意性が見られなかった隣接性 I/II と人間性についてはどう

<sup>10</sup>GoldVarb のパラメータは、0.5 を境にして、それより大きければ従属変数（この場合ゼロマーク化の割合）を増進する方向に、小さければその反対の方向に関連を持つよう設定されている。

だろうか。まず隣接性については、第2の仮説が実証されたものと考えられる。もちろん、被験者のレベルがさらに上昇した場合に要因として発現することは考えられるが、それは名詞句形式差の発現の早さとは比べ物にならないのであり、やはり普遍的制約・規則/個別言語特有的制約・規則の差は明らかである。ただし東京語方言については隣接性が大きな要因であったことから、第一言語変異ではそうした個別言語特有的の要因が重要な要因に成長する可能性は考えなければならない。中間言語での発現の早さと、第一言語変異での最終的な要因の重みは必ずしも相関しないわけであり、この点で上記の仮説は訂正されるべきである。

探索的に含めた人間性要因が全く要因として作用していなかったことは、中間言語話者がゼロマーク化の習得に当たって、語彙的意味を手がかりにしていないことを示唆するものと考えられる。東京語のゼロマーク化と同様に (Matsuda 1995、松田 2000) 中間言語でもゼロマーク化は形式を手がかりに行われていると推察される。

ここでここまでの分析に対して考えられる反論について考えてみよう。まず考えられるのは、韓国語母語からの干渉、と言う可能性である。韓国語においても「を」に相当する格助詞の実現には変異があることは知られているが、それは「代名詞には格助詞が付加され、普通名詞には付加されない」といったカテゴリー的なものではない。よって韓国語母語話者は、もしも母語を頼りとするにしても、確率論的な制約であり、それにしても当然日本語と同様なランキングであるはずである。すでに見たとおり、その中心部分は、日本語独自の制約というよりは、広く言語に見られるものである。韓国語の目的格ゼロマーク化の詳細は不明だが、いずれにしても最大限母語干渉があったとしても、それは言語一般に普遍的に見られる特徴と同様なものとなるわけであり、古典的な母語干渉とは言い難いものである。

次に被験者が関西方言などの、隣接性を要因として持たない方言を学習のターゲットとしていたのではないかという可能性を考察してみよう。この場合、隣接性にその影響が見られるのであれば、語彙・音韻・文法など他のレベルでもそうした方言差が観察されるはずであるが、KY コーパス内の発話を検討した印象からは、そうした方言的特色は見られなかった。よって方言差に帰するのには無理があろう。

テストの発話による影響はどうだろうか。テストが日本語の教師であり、

被験者がかなり緊張している様子は文字化テキストからも明白に読みとることができる。そうした状況で、被験者がテストターの発話を模倣していたという可能性も捨てきれないものであろう。しかしながら試みにテストターの発話を分析した結果、隣接性はテストターについても全く要因として作用していないことが確認されている（隣接性 I:  $p < 0.47$ 、隣接性 II:  $p < 0.10$ ）。また全体的なゼロマーク化率から言えば、被験者の 14.18%に対して、テストターのゼロマーク化率は 8.09%であった。ここからテストターの低いゼロマーク化率に引きずられたという説明も一見可能に見える。しかし同じことは名詞句形式差にも言えるのであり、テストターでは名詞句形式差も有意な要因ではなかった ( $p < 0.54$ )。よってテストターの影響という説明では、隣接性要因のみが被験者による模倣されたことになってしまい、説得力を欠くことになってしまう。

スタイル差を抽出し逃したことが分析結果に影響した、という反論も考えられよう。すでに述べたように、今回はスタイル差分析を見送らざるを得なかったのであり、スタイル差は今後さらに検討を要する課題である。ただし、KY コーパス全体を見渡した限りスタイル差は母語話者よりは遙かに小さいものと判断された。また当然スタイル差は OPI レベルと共に広がるものと予想されるが、名詞句形式差は、中級レベルでも観察されたものである。こういった事実からも、スタイル差のいっそうの分析によっても上記分析の大勢に変化はないものとして良いであろう。

## 7. おわりに

最後に今後の課題を述べておきたい。まず KY コーパスに含まれている、中国語および英語母語話者の発話データの分析があげられる。本稿で扱った議論は、これら 2 種類のデータ分析を通して、さらに密度の高い検証が可能になるであろう。この場合、当然名詞句形式差が同じような形で要因として浮かび上がってくるのが予想されることは言うまでもない。もちろん、KY コーパス以外の発話データの分析も必要である。細かなクロス分析のためには、どうしても大量の発話データが必要になる。この点で KY コーパスのような第二言語コーパスは、今後ますますその必要性が高まって来るであろう。

本稿では全く取り上げていないが、助詞使用のエラーアナリシスはどうしても必要な分析である。「を」は「に」や「が」等助詞との混同が顕著だが、こうした

誤用の分布はゼロマーク化と分かち難く絡み合っている。本稿では前提としてゼロマーク化が「習得」の結果だとしていたが、それが「未習得」の結果でなかったという保証はどこにもないわけである。この難問を解く鍵の一つは、こうしたエラーアナリシスにあるのではないかという見通しを筆者持っているが、これについてはすべて今後の課題とせざるを得ない。

### 参考文献

- Bailey, Robert & Dennis R. Preston (eds). 1996. *Second language acquisition and linguistic variation*. Amsterdam: John Benjamins.
- Fienberg, Stephen. 1989. *The Analysis of cross-classified categorical data*. Cambridge: MIT Press.
- Hosaka, Junko, Takezawa Toshiyuki, and Uratani Noriyoshi. 1992. "Analyzing post-position drops in spoken Japanese." *The proceedings of the International Conference on Spoken Language Processing 1992*. 1251-1254.
- Labov, William, & Gillian Sankoff. 1985. "Variation Theory". Ms.
- Masunaga, Kiyoko. 1988. "Case deletion and discourse context." In *Papers from the Second International Workshop on Japanese Syntax*, edited by W. J. Poser. Stanford: CSLI. Pp. 145-154.
- Matsuda, Kenjiro. 1995. *Variable zero-marking of (o) in Tokyo Japanese*. Doctoral Dissertation, University of Pennsylvania.
- Rand, David, and David Sankoff. 1990. GoldVarb (version 2.0). Montréal: Centre de recherches mathématiques.
- Saito, Mamoru. 1985. *Some asymmetries in Japanese and their theoretical implications*. Doctoral Dissertation, Department of Linguistics and Philosophy, MIT.
- Tsutsui, Michio. 1984. *Particle ellipses in Japanese*. Doctoral Dissertation, University of Illinois at Urbana-Champaign.

エヴァリット、B.S. (山内光哉監訳、弓野憲一、菱谷晋介訳)、1977.『質的データの解析: カイ二乗検定とその展開』, 東京: 新曜社.

国立国語研究所 (編)、1989.『方言文法全国地図』, 東京: 大蔵省印刷局.

松田謙次郎. 2000. 「東京方言格助詞「を」の使用に関わる言語的諸要因の数量的検証」. 『国語学』第 51 卷 1 号, Pp. 61-76.

宮田裕子. 1992. *A study of the acquisition of the case particles in Japanese: With special reference to 'ga' and 'wo'*. 大阪大学大学院言語文化研究科修士論文.

**Author's E-mail Address:** kenjiro@icis.shoin.ac.jp

**Author's web site:** <http://www.shoin.ac.jp/~kenjiro/>