

## 第2言語コーパスからみた言語普遍性\*

松田謙次郎

---

### Language Universal in an L2 Corpus

MATSUDA Kenjiro

#### Abstract

One of the interesting possibilities suggested by the development of interlanguage studies is a correlation between the L1/L2 variation and the distinction between language-universal/language-particular constraints governing the variation within a language system. If there were a language-universal constraint such that all the (known) languages conform to it, then one would expect that interlanguage, as a language system in itself, should conform to such a constraint as well. This paper takes up the zero-marking of an accusative case marker *o* in Japanese to explore the possibility mentioned above. Zero-marking of *o*, which is most sensitive to the adjacency between the object NP and the verb (a language-particular constraint), has at least one constraint that is known to be a language-universal: the distinction between the lexical NP and the pronoun for the object NP, with the pronoun significantly disfavoring zero-marking. This is easily explainable as a reflection of a language universal such that a language with an overt case-marking system should case-mark the pronouns, but not necessarily the lexical NPs. Using a corpus of transcribed OPI interviews with Chinese and Korean speakers, I collected relevant accusative constructions to see if the variable zero-marking by the L2 speakers is accountable by the language-particular/ universal factors. The result shows that: (1) the form of NP showed up as the significant factor in predicting the zero-marking by the L2 speaker and (2) the difference between the pronoun and the lexical NP, also was as expected, with pronoun disfavoring zero-marking. The result clearly shows that at least in the early and mid stage of L2 acquisition, the interlanguage system reveals language-universal constraint, in this case the implicational relationship between the case-marking on NP and pronoun.

---

\*本稿は、2002年3月22日に米国サンフランシスコ州立大学(SFSU)で開催された、国際実用日本語学会(International Conference on Practical Linguistics of Japanese)でポスター発表したものに大幅に加筆・修正を加えたものである。当日コメントを頂いた参加者諸氏に謝意を表したい。なお、本稿で使用したデータの収集・分析に当たっては、1997年度および2005年度の松蔭特別研究助成(「言語コーパスの構築・収集とその教育的応用」)を受けている。

## 1. はじめに

松蔭特別研究助成を受けて発足した松蔭コーパスプロジェクトでは、1997年、2005年の両年にわたりコーパス研究とその教育への応用を視野に入れて、さまざまな国内外のコーパスを収集し、また独自のコーパスを作成してきた。独自作成のコーパスは、例えば樋口(1999)、宮永(2000)、宮永・松田(2001)といった神戸方言を中心に方言コーパスや佐々木(2005)、竹内(2005)のような英語教師の英語朗読コーパス、さらにおねえことば発話集(衣川, 2005)、国政選挙候補者による選挙演説コーパス(佐野, 2006)、松波(2005)、大西(2006)、森辺(2006)のような各種国会会議録データなどである。こうした多彩なコンテンツのコーパスに基づいた研究の開く可能性は無限大であり、上述の諸論文以外にも Nambu(2005)、Nambu and Matsuda (forthcoming)、松田 (forthcoming) といった独自の成果に結実しており、今後の研究においても重要なデータとしての役割を果たすことが期待される。これまでのコーパスには、文学作品や新聞データ、講演、インタビュー、対話などといったコンテンツが多いが、コーパス言語学の展開が進むにつれてますます重要になってきているのは、こうしたコーパス内容の多様化と思われる。またコンテンツと言うことで言えば、現在入手可能な言語コーパスの多くは第1言語話者の産出を対象としたものであるが、中には KY コーパスや上村隆一氏による「日本語会話データベース」(<http://www.env.kitakyu-u.ac.jp/corpus/>) のように、第2言語話者のそれを対象としたものも存在する。条件を揃えた上で十分なデータを集めにくい話者群を対象とした研究の場合、こうしたデータは大きな威力なることは間違えない。本論文では、第2言語としての日本語発話コーパスの嚆矢であり、松蔭コーパスプロジェクトの一環として入手した KY コーパスを取り上げ、第2言語話者の発話中の言語変異に注目し、そこに見られる言語普遍性の探求という視点から行った分析の報告を行い、コーパス言語学的手法が持つ可能性の一つを追究することとしたい。

## 2. 中間言語変異の普遍性と特殊性

第2言語習得は、とかく第一言語(L1)を話す話者の頭の中の、別な場所に第2言語(L2)の文法が出来上がっていく過程と捉えられ勝ちであるが、こうした捉え方だと、発達途中のL2文法は単なる不完全なものとし映らなくなってしまう。

これに対して「中間言語」(interlanguage)という考え方では、学習途中に出来るL2文法をL1との接触言語と捉える。これがまさに中間言語である。中間言語の考え方に従うと、中間言語自体もひとつの言語体系であり、それなりの「文法」に従うことになる。当然この「文法」は自然言語としての文法の制約に従うわけである。<sup>1</sup>

生成文法風にこれを言い換えれば、中間言語は普遍文法(UG)の制約からも自由ではあり得ないはずだということになり、これは検証可能な数々の仮説を生み出す可能性を持った考え方である。言語普遍的な制約や規則は、習得過程においても早期に現れることが予測され、習得途上の変異においても要因としての重みも相対的に高いことが考え

<sup>1</sup>中間言語理論の詳細については、迫田(1998)、水野(2000)を参照のこと。

られよう。逆に各言語に特殊なものは、遅くに習得され要因の重みも低くなるはずである。つまり、中間言語変異においては、以下のような予測が成立するはずなのである：

第2 言語話者の中間言語変異の内的制約条件においては、

1. 言語普遍的制約・規則は早期に習得され、要因としての重みも相対的に高い
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相対的に低い
3. 制約条件の重みは第2 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

### 3. 日本語「を」格ゼロマーク化現象: 内的条件を中心に

日本語の「を」格ゼロマーク化現象は、以下のように対格助詞「を」が自然談話中でゼロとして実現される現象であり、その内的制約条件については経験的データから詳細な部分についてまで判明している (Tsutsui, 1984; Saito, 1985; Masunaga, 2001; Hosaka, Takezawa, & Uratani, 1992; Matsuda, 1995; 松田, 2000; Fry, 2003)。

- (1) あいつがパソコンを/φ 壊した
- (2) あいつがパソコンを/?φ きのう壊した

Matsuda (1995), 松田 (2000) による東京語における主たる言語内的制約条件を挙げてみよう：

#### 東京語ゼロマーク化の主要内的制約条件

**目的語名詞句と動詞の隣接性:** 目的語名詞句と動詞が隣接している場合の方が、していない場合よりも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存在しても隣接と効果が変わらない、「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)。

**スタイル:** くだ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。

さて、このうち普遍的要因と考えられるのは、目的語名詞句の形式である。その理由は類型論的な証拠から来るもので、系統的に無関係とされる言語間であっても、格標示と名詞句形式に高い相関性が認められ (Matsuda, 1995)、次のような一般化として定式可能だからである。

#### 格標示と名詞句形式差の関連に関する一般化

ある明示的な格標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明示的な格標示があるが、普通名詞には必ずしもそうした格標示があるとは限らない。

一方、「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、日本語一方言に特有な規則であり、普遍的なものではないものと考えられる。現に、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いる（むしろ日常会話では明示的に実現される場合の方が珍しい）。こうした個別言語特有な規則は、中間言語変異において観察されなくても不思議ではない。スタイル差についても同様であり、言語普遍的なものとは考えにくい。<sup>2</sup>

言語普遍的な要因として考えられるものは、実はもう一つ考えられる。それは人間性 (humanness) である。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は、Matsuda (1995) で確認されているが、『方言文法全国地図』第6・7図 (国立国語研究所, 1989) に見られるように、方言によってはこの要因が大きく作用している場合がある。つまり、方言によってはこの要因が効いている可能性があるわけで、ここから特定言語特有の言語内的制約条件として人間性を考察する必要が出てくることになる。

以上をまとめると、中間言語におけるゼロマーク化の変異については、言語普遍的・特有な要因を中心として次のような仮説が設定できる:

#### 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である
2. 隣接性や人間性は中間言語話者の発話には要因としては相対的に重要性が低い。

## 4. 要因・データ・分析手順

### 4.1 データ

ここで使用する KY コーパスは、英語、韓国語、中国語母語話者との対話を文字化したものである。<sup>3</sup> 話者の日本語レベルを、初級、中級、上級、超上級に分類し、日本語母語話者との対話を収めている。3言語それぞれ30名ずつ（初級 (Novice) 5人、中級 (Intermediate) 10人、上級 (Advanced) 10人、超級 (Superior) 5人ずつ）、計90名について OPI (Oral Proficiency Interview) 方式のインタビューに則り、テスト (T) と被験者 (S) の会話という形態で対話を行っている。<sup>4</sup> 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

<sup>2</sup> スタイル差については、松田 (2001) でも触れたようにコード上の問題もあり、取り上げるのが難しい。

<sup>3</sup> なお、韓国語話者と中国語話者の分析については、松田 (2001, 2002) にも報告がある。

<sup>4</sup> OPI については、牧野成一、鎌田修、山内博之、齋藤真理子、荻原稚佳子、伊藤とく美、中島和子 (2001) を参照のこと。

の上級のみ) という形式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・分ち書きなしであり、同一話者の発話は一行に収めてある。<sup>5</sup>

## 4.2 要因

前節で述べた要因を実際に検討するためには、さらに下位分類を設けコード表を作成する必要がある。今回は、以下のようなコード表を元にコーディングを行った。

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

「動詞と目的語の隣接性」では、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」とした。オノマトピアと数量詞の透明性 (Matsuda, 1995) に鑑み、これら2つの要素ので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。対話相手のうなずきも同様である。ただし話者自身のうなずきや、「えーと」といったフィラーはこうした扱いはせず、非隣接としてコードしてある。<sup>6</sup>

## 4.3 方法論

KY コーパス全巻の学習者発話について、対格「を」が明示的に使用されている部分、また実際に使用されていないが明らかに「を」が使用しうると考えられる部分を手作業で抜き出し、各要因についてコーディングを施した。続いて各要因で単変量分析を行い、カイ自乗検定を行った。仮説通りであれば、英語、韓国語、中国語いずれを母語とする話者にとっても名詞句形式差要因は有意な制約条件なはずである。それに対して、隣接性や人間性要因は、そのような有意性を呈しないはずである。

## 5. 結果

### 5.1 隣接性

まず、隣接性の結果から見てみよう。<sup>7</sup> 中国語と韓国語母語話者にとって、隣接性はまったく要因としては効いていないが、英語母語話者には 0.6% レベルと高度に有意な要因であることが分かる (図 1)。中国語母語話者と韓国語母語話者を比較すると、隣接性とゼロマーク化の方向が逆となっており、ここからもこれらの話者にとって隣接性は何らの要因として機能していないことが窺える。言い換えると、隣接性は異なる母語話者の間で、ゼロマーク化に関する一貫した要因ではないということになるわけである。

<sup>5</sup>KY コーパスの入手・使用にあたっては、作成者である鎌田修氏と山内博之氏のお手を煩わせている。この場をお借りして感謝申し上げます。なお、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。

<sup>6</sup>なお、データとして採用しない基準については松田 (2000) と同様であり、ここでは詳説を避ける。

<sup>7</sup>このセクションで報告する結果の一部は、松田 (2001, 2002) でも報告されている。

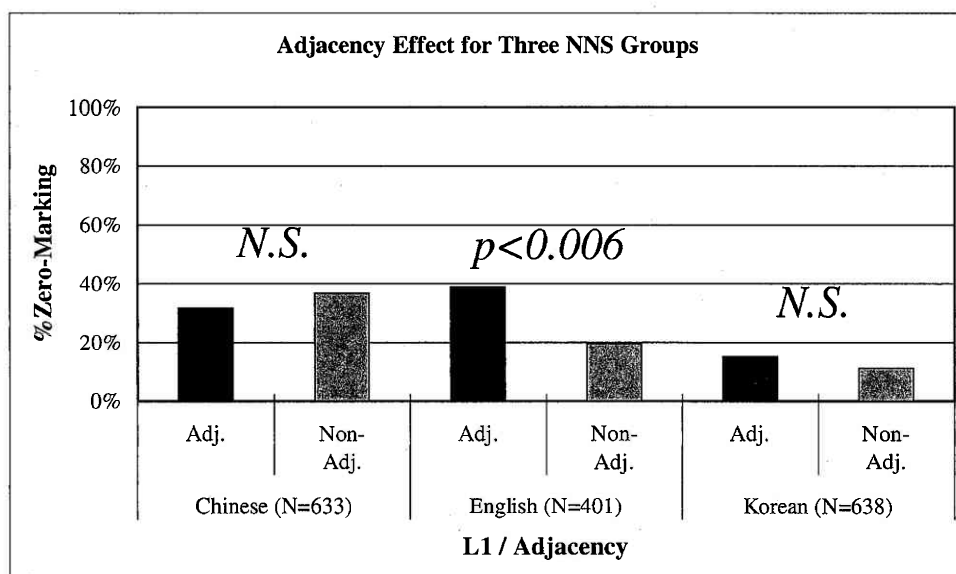


図 1: 隣接性による分布

## 5.2 名詞句形式差

名詞句形式差は、3 母語話者集団でいずれも有意である (図 2)。グラフからも分かる通り、WH > 節 (S) > 名詞句 (NP) > 代名詞 (Pro) という分布状況が、3 集団で一致している点が特に注目し得る点である。これは差異の方向性も一致していなかった前節の隣接性とは大きな対照をなしている。ここから名詞句形式差は、母語のいかに関わらず、第 2 言語話者のゼロマーク化を司る大きな要因であることが結論づけられることになる。

## 5.3 OPI レベル

OPI レベルによる分布を示したのが、図 3 である。これによると OPI のレベルによる差が有意だったのは中国語母語話者グループのみであり、英語と韓国語母語グループではレベル間にゼロマーク化についての有意な差は検出されていない。その有意差が見出された中国語話者にしても、中間 (intermediate) のグループがもっともゼロマーク化率が高いという結果であり、レベルの上下差と一貫した相関があるとは解釈できない結果となっている。

## 5.4 人間性

人間性要因は、中国語母語話者グループで非有意、英語で有意、韓国語話者で差異の傾向が見られるという具合に、結果が大きく分かれている (図 4)。3 グループで差の方向は一致しているものの、要因としては存在するとしてもきわめて微弱なものと言わざるを得ない。

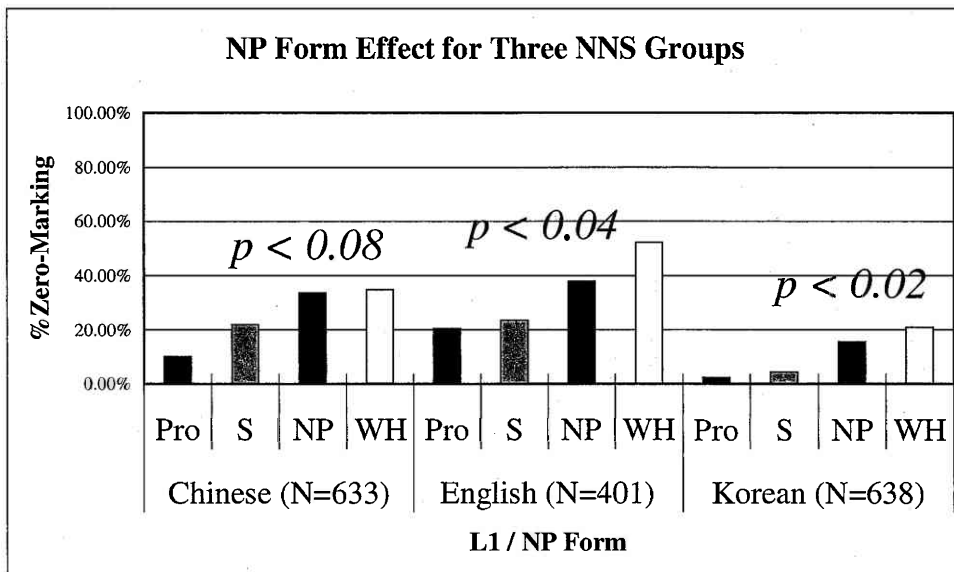


図 2: 名詞句形式差による分布

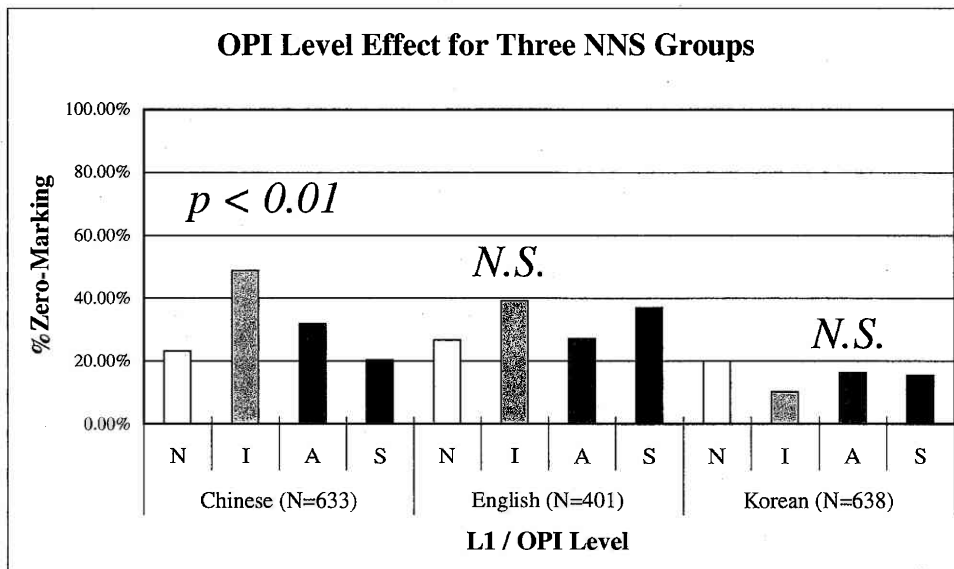


図 3: OPI レベルによる分布

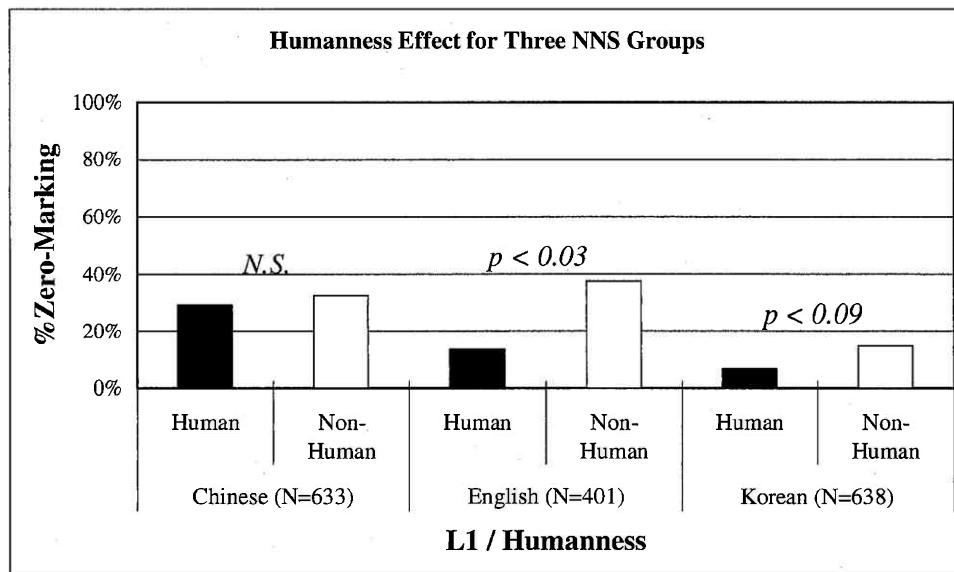


図 4: 人間性による分布

## 6. 議論

前節の4要因に関する結果をまとめてみよう(表1)。この表から言えることは、すべての母語話者で一致して有意差、ないしは傾向を示している要因は、検討した4要因のうちでは名詞句形式差のみだということである。これは冒頭で示した仮説に照らし合わせると、「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である」、および「隣接性や人間性は中間言語話者の発話には要因としては相対的に重要性が低い」という予測とぴったりと一致する結果である。ここから、もともとの仮説であった、「言語普遍的制約・規則(すなわち名詞句形式差)は早期に習得され、要因としての重みも相対的に高い。個別言語の特殊なもの(各言語個別要因、すなわち隣接性と人間性)は遅くに習得され、要因としての重みが相対的に低い」という仮説の、後半部分を実証したことになる。表2<sup>8</sup>に見るような類型論的分布は、第2言語習得にも反映されていたわけである。

表 1: 分析結果のまとめ

要因 / L1	中国語	英語	韓国語
隣接性	N.S.	$p < 0.006$	N.S.
名詞句形式差	$p < 0.08$	$p < 0.04$	$p < 0.02$
OPI レベル	$p < 0.01$	N.S.	N.S.
人間性	N.S.	$p < 0.03$	$p < 0.09$

ここで、考えられる反論を2つ取り上げてみよう。

<sup>8</sup>なお、それぞれの言語に関して判断の基礎とした文法書については Matsuda (1995) を参照のこと。



表 2: 語彙名詞・代名詞に対する明示的格標示の類型論的分類

	代名詞に明示的格標示あり	代名詞に明示的格標示なし
語彙名詞に明示的格標示あり	Basque, Burushaski, Garo, German, Khasi, Korean, Kwaikiutl, Sierra Miwok, Serbo-Croatian, Tarascan, Tasmanian, Yukaghir, Yupik	? (実在する言語が確認できない、または存在しない)
語彙名詞に明示的格標示なし	Navajo, Ojibwa, Pawnee, English, French	Chinese, Goajiro, Haitian Creole, Karankawa, Vietnamese, !Xu

### 6.1 反論 1: これは単に目標言語からの学習ではないのか?

最初の反論は、名詞句形式差の分布は、学習者が目標言語である日本語から学習を通じて単に学んだものであり、もともと学習者が持っていた言語知識ではないのではないのか、というものである。名詞句形式差は、L1 話者のゼロマーク化においても重要な要因であるから、それを学習者がピックアップする可能性は否定できないわけである。

しかしながら、この反論では説明できない事実がいくつか存在することから、この反論が受け入れがたいものであることが判明する。まず、L1 話者の自然発話では隣接性をもっとも強い要因なのであり (Matsuda, 1995)、L1 話者の発話からの学習であれば、まず隣接要因こそが学習されるはずである。さらに、隣接性が基本的に隣接か否か、というきわめて簡単な要因であるのに対して、名詞句形式差が、普通名詞、代名詞、WH 句、節と言った多くの形式の区別が必要であることを考えあわせると、この反論の言うような「学習」では説明しきれないことは明白である。もちろん、日本語教育の場でゼロマーク化が（その要因を含めて）学習者に明示的に教授されることはない。

さらに、学習者による授業の場以外での学習を考えると、KY コーパスに登場する L2 話者は異なる日本語方言地域から参加しているのであり、学習者達が日常耳にしている日本語は、ゼロマーク化の規則や頻度が大きく異なることも十分考えられる。しかし、それでも彼らの発話に見られるゼロマーク化の分布には、名詞句形式差による一貫した方向性が存在していたわけである。これを説明するためには、目標言語からの学習（だけ）ではどうしても不十分と言わざるを得ず、何らかの形で学習者の頭の中にあつた（普遍的な）言語知識が活用された、という他はないのである。

### 6.2 反論 2: これは学習者が目的語名詞句と動詞を、ひとつの塊（チャンク）として捉えることから来る現象ではないか?

この説に従えば、頻度の高い要素同士の結びつきほどひとつのチャンクとして認識されるようになるはずであるから、頻度とゼロマーク化の間に相関関係が見いだされるはずである。しかし、データを見てもこれと言った明確なパターンは見られない。よって、

チャンクによって第2言語話者のゼロマーク化パターンを説明しようとするのは、無理だと言うことがわかる。さらに、チャンクということであれば、母語話者の自然談話では目的語名詞句と動詞が直接隣接している場合の頻度が、そうでない場合よりも遙かに高い (Matsuda, 1995)。つまり、学習者も隣接環境のチャンクの方が非隣接環境のそれよりも遙かに多くインプットとして遭遇しているはずであり、ここから隣接性が大きな要因として効くはずだという推論が成立する。これが間違えであることは、前節の結果から明らかであるところである。

よって、いずれの反論も前節の結論を覆すものではなく、3グループに共通して観察された名詞句形式差のパターンは、名詞句に関する言語普遍的原則に帰すべきものであるという結論は、なんら影響を受けず、そのまま指示されるのである。

## 7. おわりに

最後に、今後の課題を書いておきたい。1つは、多変量解析に十分なだけの追加データの収集・コーディングである。今回もロジスティック解析を試みたが、どうしても重要なセルにデータが乏しくなってしまう、分析を断念せざるを得なかった。別な第2言語コーパスを加えるなどして今後は対応していきたいものと考えている。

2つ目は、さらに多彩な母語を持つ話者のL2としての日本語データの分析である。中国語、英語、韓国語はそれぞれ、代名詞・普通名詞共に格標示なし、代名詞のみ格標示あり、そしてどちらも格標示ありという自然言語で確認されている3タイプに当てはまっている。しかし、それぞれのタイプに1言語であるために、これが偶然ではないとはいえない部分が残る。そこで、各タイプにもっと別言語のデータが欲しいところである。本稿の冒頭でも述べたように、まさにこうしたコーパスコンテンツの多様化こそが現在望まれるのであり、松蔭コーパスプロジェクトの成果報告でもある本稿の、これが一つの結論でもある。

## 参考文献

- Fry, John (2003). *Ellipsis and WA-marking in Japanese conversation*. Routledge.
- 樋口絢子 (1999). 『言語コーパスの現状と問題点』. 卒業論文, 神戸松蔭女子学院大学.
- Hosaka, Junko, Takezawa, Toshiyuki, & Uratani, Noriyoshi (1992). Analyzing postposition drops in spoken Japanese. In *The proceedings of the International Conference on Spoken Language Processing*, pp. 1251–1254.
- 衣川礼子 (2005). 『「おねえことば」とことばの性差—ことばの文末表現・役割語の観点から—』. 卒業論文, 神戸松蔭女子学院大学.
- 国立国語研究所 (1989). 『方言文法全国地図』, 1巻. 東京: 財務省印刷局.
- 牧野成一, 鎌田修, 山内博之, 齋藤真理子, 荻原稚佳子, 伊藤とく美, 中島和子 (2001). 『ACTFL-OPI 入門』. 東京: アルク.

- Masunaga, Kiyoko (2001). Case deletion and discourse context. In Poser, W. J. (Ed.), *Papers from the Second International Workshop on Japanese Syntax*, pp. 145–154 Stanford. CSLI.
- 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』, 51 (1), 61–76.
- 松田謙次郎 (2001). 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. *Theoretical and Applied Linguistics at Kobe Shoin*, No. 4, 57–76.
- 松田謙次郎 (2002). 「を」格ゼロマーク化と中間言語: 中国語母語話者の場合. *Theoretical and Applied Linguistics at Kobe Shoin*, No. 5, 43–52.
- 松田謙次郎 (編) (forthcoming). 『国会会議録を使った日本語研究』. 東京: ひつじ書房.
- Matsuda, Kenjiro (1995). *Variable Zero-marking of (o) in Tokyo Japanese*. Ph.d. dissertation, University of Pennsylvania.
- 松波美也子 (2005). 『関西弁文末表現「や」の成り立ちと普及率についての国会会議録を用いた調査』. 卒業論文, 神戸松蔭女子学院大学.
- 宮永眞紀 (2000). 『ガ行鼻濁音—高砂市における [ŋ] の発音の実態—』. 卒業論文, 神戸松蔭女子学院大学.
- 宮永眞紀・松田謙次郎 (2001). 兵庫県高砂市におけるガ行鼻濁音の分布. 『第72回日本方言研究会発表会発表原稿集』, pp. 25–32.
- 水野光晴 (2000). 『中間言語分析—英語冠詞習得の軌跡』. 東京: 開拓社.
- 森辺明日来 (2006). 『国会会議録に見る「すみません」と「すいません」のバリエーション』. 卒業論文, 神戸松蔭女子学院大学.
- Nambu, Satoshi (2005). Corpus-based study of the change in GA/NO conversion. Master's thesis, Kobe Shoin Women's University.
- Nambu, Satoshi & Matsuda, Kenjiro (forthcoming). Change and variation of *ga/no* conversion in Tokyo Japanese. In *The proceedings of the XVIIth International Conference on Historical Linguistics*. John Benjamins.
- 大西視友子 (2006). 『第163回国会会議録について～参議院～』. 卒業論文, 神戸松蔭女子学院大学.
- Saito, Mamoru (1985). *Some asymmetries in Japanese and their theoretical implications*. Ph.d. dissertation, MIT.

迫田久美子 (1998). 『中間言語研究—日本語学習者による指示詞コ・ソ・アの習得—』. 広島：溪水社.

佐野由美 (2006). 『2005 年第 44 回総選挙演説文字化資料』. 卒業論文, 神戸松蔭女子学院大学.

佐々木優貴子 (2005). 『Australian English — 神戸松蔭女子学院大学英語教師対象の方言調査—』. 卒業論文, 神戸松蔭女子学院大学.

竹内希世子 (2005). 『松蔭教員を対象とした北米英語話者の調査』. 卒業論文, 神戸松蔭女子学院大学.

Tsutsui, Michio (1984). *Particle ellipses in Japanese*. Ph.d. dissertation, University of Illinois at Urbana-Champaign.

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

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