

# 談話情報を用いた日本人英語学習者による論説文の計量分析

小林雄一郎

大阪大学大学院言語文化研究科

E-mail: [kobayashi0721@gmail.com](mailto:kobayashi0721@gmail.com)

## 1. Introduction

学習者が実際に使用した言語の特徴や誤用についての知識を持つことは、英語教師にとって重要なことである。第2言語習得の分野では、誤用は、(1) その学習者が対象言語をどれくらい学んだかについての情報を教師に提供する、(2) どのように対象言語が学習されるかについての情報を研究者に提供する、(3) 学習者が正用と誤用に関する情報を基に対象言語の規則を発見できる、という3つの点で有意義であるとされている (Corder 1967)。

近年、英語教育の現場では実践的コミュニケーション能力の育成を図ることが求められており、中学校や高校の学習指導要領にも同様の記述が見られる。しかし、円滑で効果的なコミュニケーションをするためには、「何を」伝えるかよりも、「いかに」伝えるかが必要不可欠となる (e.g. Conner & Mbaye 2002; Young 2002)。具体的に効果的なコミュニケーションを達成する1つの方法は、対比 (e.g. *while, on the other hand*)、理由 (e.g. *because, so*)、結果 (e.g. *therefore, as a result*)、列挙 (e.g. *firstly, secondly*)、例示 (e.g. *for example, in particular*) といった接続語 (connectives) や談話標識 (discourse markers) によって、談話のユニット間の論理関係や意味関係を表すことである (Altenberg & Tapper 1998)。それらの表現には、談話の結束上の「手がかり」を与えて (Leech & Svartvik 1994)、読み手や聞き手がユニット間の一貫性を見つけることを手助けし、テキストの意味理解ができるようにする働きがある。しかしながら、学習者にとって、談話標識の「適切な」使用は非常に難しい。談話標識の使用数が増えても結束性の質が上がるわけではなく (McCarthy 1991)、むしろ「人工的で機械的な文章」(Zamel 1983) となってしまうことで、テキストの理解を妨げることもある (Crewe 1990)。従って、実践的コミュニケーション能力の育成を図る上で、学習者による談話標識の使用傾向を調査し、彼らの談話構造における特徴や誤用を究明することは極めて重要である。しかしながら、これまでの研究では、手作業による談話分析のコストが高いこともあって、限られた数の学習者データしか扱うことができず、そこから得られた結果がどこまで普遍的なものかを検証することが難しかった。さらに、大規模な調査を行なう場合は、多くの分析者が必要となり、どうしても結果が個々の分析者による主観に影響されてしまうという欠点があった (e.g. Baker 2006)。

それに対して、本研究では、日本人中学生、高校生、大学生の英作文を集めた「学習者コーパス」(learner corpus) をテキストマイニングの手法を用いて客観的に解析し、そこから得られた結果を母語話者のコーパスと比較する。具体的には、以下のような点に主な焦点を当てる。

- 日本人中学生、高校生、大学生の学習者コーパスに関する縦断的研究

- 日本人学習者英語と母語話者英語の差異に関する分析
- 日本人学習者の談話標識に関する特徴抽出, 誤用分析
- 日本人学習者の英作文における談話能力の測定と評価

そして, 本研究の独創的な点は, 主に以下の2点である。

- 学習者談話の計量的分析:  
計量的な談話分析は, 節の境界を越えた言語特性を同定する必要があるため, 語彙分析よりも技術的に難しく (e.g. Biber *et al.* 1998; Conrad 2002), 本格的にコーパスを利用した先行研究も少ない。一方, 本研究では, コーパスに談話標識を自動付与するプログラム (小林・山田 2008) を作成し, 英作文における談話標識の頻度を抽出し, 多変量解析を用いて結果を視覚化する。
- 中学生から大学生までの縦断的研究:  
学習者コーパス研究では, 特定の学校や学年を分析したものが多い。一方, 本研究では, 中学生, 高校生, 大学生による大規模コーパスを縦断的に解析することで, 「初心者ほど苦手な表現は何か」や「上級者になってもあまり習得されない表現は何か」といった点を炙り出す。

本稿では, 上記の研究の出発点として, テキストマイニングの技法を用いて, 学習者作文における談話標識の使用傾向を統計的に俯瞰する。また, 量的分析から得られた結果を手がかりに, 質的分析を有機的に組み合わせ, 学習者の談話における特徴や誤用パターンを抽出する。

## 2. Background

### 2.1 Corpus linguistics

「コーパス」(corpus) という語は, 元来「身体」(body) を表すラテン語であるが, *Oxford English Dictionary* によれば, “The body of written or spoken material upon which a linguistic analysis is based” という定義の初例は 1956 年のものであり, 「言語分析のための言語資料の集積」を意味する言語学的用法は, 比較的新しいものである。また, 現代のコーパス言語学 (corpus linguistics) において, 「コーパス」とは, 「機械可読な」(machine-readable) 形式で集積された大規模で「真正な」(authentic) 言語データベースのことを指し, 特定の言語変種やジャンルに対する「代表性」(representativeness) を持つ点で単なるアーカイブとは異なるものである (e.g. Baker *et al.* 2006; McEnery *et al.* 2006)。そして, コーパス言語学は, Leech (1992) によれば, 次のような特徴を持つとされている。

- 言語能力よりも, 言語運用に中心を置く
- 言語の普遍的特性の解明よりも, 個別言語の言語記述に中心を置く
- 質的な言語モデルのみならず, 量的な言語モデルにも中心を置く
- 言語研究における合理主義的な立場よりも, より経験主義的な立場に中心を置く

これまで世界中で数多くのコーパスが構築されてきた。それらのコーパスは、大まかに、「汎用目的コーパス」(general-purpose corpus) か「特殊目的コーパス」(specific-purpose corpus) か、「書き言葉コーパス」(written corpus) か「話し言葉コーパス」(spoken corpus) か、「共時的コーパス」(synchronic corpus) か「通時的コーパス」(diachronic corpus) か、という点で分類される。本研究で扱うコーパスは、学習者の英作文を集積したものであり、「特殊目的コーパス」で「書き言葉コーパス」である。また、同一の学習者を経年的に追跡調査したデータではないため、厳密な意味では「共時的コーパス」である。しかし、中学生・高校生・大学生という言語学習の過程を時系列なモデルで分析することができるという点では、緩やかな意味での「通時的コーパス」とも言える。

## 2.2 Learner corpus

学習者コーパスとは、外国人学習者によって産出された言語データを機械可読な形式で集積したものである (Leech 1998)。1990 年代以降、様々な学習者コーパスが構築されてきたが、第 2 言語習得 (SLA) の研究のために言語データを比較検討するという方法論は必ずしも新しいものではない。1950 年代後半、対照分析 (Contrastive Analysis) の研究者たちは、言語データこそ使っていなかったものの、目標言語と第 1 言語の類似度が言語学習の困難度と関係があると予想していた (e.g. Lado 1957)。1960 年代に入ると、誤用分析 (Error Analysis) の研究者たちが、学習者言語の発達過程を解明するために、学習者がどの段階でどのような誤用をするかに注目し、そのパターンの記述を試みた (e.g. Corder 1967; Richard 1971)。1970 年代には、運用分析 (Performance Analysis) の研究者たちは、誤用だけでなく、正用も含めた「中間言語」(Selinker 1972) の運用全体を分析対象とするようになった (e.g. Dulay & Burt 1973; Hakuta 1976)。Tono (2005) がまとめているように、学習者コーパスとは、このような SLA における歴史の流れの中で生み出されたものに他ならない。

学習者コーパスの先駆的な試みとしては、コペンハーゲン大学で 1970 年代に行なわれた PIF Corpus (Færch *et al.* 1984)、1970~80 年代にドイツの移民を対象に調査した ZISA Project (Clahsen 1980; Meisel *et al.* 1981)、ヨーロッパ 5 カ国における移民を対象に調査した ESF Database (Perdue 1993)、幼児の母語習得データベースである CHILDES (MacWhinney 1995) などがある。そして、コーパス言語学の知見を充分に取り入れた現代の学習者コーパスとして、代表的なものとして、様々な母語を背景とする英語学習者の作文を集めた ICLE (International Corpus of Learner English) (Granger 1998) を始め、LLC (Longman Learner's Corpus) や CLC (Cambridge Learner's Corpus) といった商用コーパスがある (e.g. Granger 2002; Pravec 2002)。また、日本人学習者を対象とする代表的なコーパスとして、書き言葉を集めた JEFLL Corpus (投野 2007) や話し言葉を集めた NICT JLE Corpus (和泉ほか 2004) などがある。

## 2.3 Discourse markers

「談話標識」(discourse markers) とは何か。その定義や構成要素をめぐって、これまで多くの議論が成されてきた。まず、その名称に関しても、discourse markers (e.g. Jucker & Ziv 1998; Schiffrin 1987) 以外に、cue phrases (Knott & Dale 1994)、discourse connectives (e.g.

Blakemore 1987, 1992; Warner 1985), discourse connectors (e.g. Biber 2006), discourse operators (e.g. Redeker 1990), discourse particles (e.g. Aijmer 2002; Schorup 1985), discourse signaling devices (e.g. Polanyi & Scha 1983), phatic connectives (e.g. Bazanella 1990), pragmatic connectives (e.g. Stubbs 1983), pragmatic expressions (e.g. Erman 1987), pragmatic formatives (e.g. Fraser 1987), pragmatic markers (e.g. Aijmer 2008; Fraser 1990), pragmatic operators (e.g. Ariel 1994), pragmatic particles (e.g. Östman 1995), semantic conjuncts (e.g. Quirk *et al.* 1985), stance connectives (e.g. Halliday & Hasan 1976) など、枚挙に暇がない。

そして、談話標識の定義についても、多くの議論が存在する。例えば、Levinson (1987) は、談話標識とは「ある発話と先行する談話との間の関係を表す語や句」であり、「発話の冒頭で用いられる」としているが、Fraser (1999) などは、談話標識が発話や文の途中や最後に現れることもあるとする。さらに、談話標識の研究の多くは話し言葉を分析対象としているが、書き言葉でも談話標識は多く用いられる。但し、書き言葉と話し言葉では、現れる談話標識の種類や頻度が明らかに異なっている (e.g. Rühleman 2007; Swan 2005)。この点に関して、Biber (2006) は、discourse connectors の中には、話し言葉で使われる discourse markers (e.g. *OK, all right*) と、書き言葉で使われる linking adverbials (e.g. *however, thus*) があるとしている。しかしながら、前者が具体的に何を含まのかは必ずしも明確ではなく、後者が接続副詞以外の表現を排除している点にも、議論の余地が残る。また、上記のいずれの研究においても、限られた数（多くの場合は極めて少数）の談話標識しか分析対象とされておらず、網羅的なリストが提示されていないのが残念である。

## 2.4 Metadiscourse markers

Metadiscourse markers も、いくつかの相違点はあるものの、広義での談話標識の一種である。かつて、metadiscourse は「discourse についての discourse」(Vande Kopple 1985) と考えられていたが、最近の研究では「書き言葉、あるいは話し言葉のテキストにおける言語要素 (linguistic material) で、命題内容に何かを付け加えるものではなく、聞き手や読み手が与えられた情報を系統立て、解釈し、評価することを助けるためのもの」(Crismore *et al.* 1993) と定義されるようになった。

Metadiscourse markers の研究において、最もよく使われる枠組みは、恐らく Hyland list (Hyland 2005) であろう。このリストは、Vande Kopple (1985) や Crismore *et al.* (1993) による研究をベースとして、10 種類のカテゴリー (表 1) に分類される約 400 種類の談話表現を網羅的に収録したものである。また、このリストは、コーパスに基づく統計的研究を想定して作成されたものであり、これまでにアカデミック・ライティングを始め、教科書、学位論文、ビジネスレターなど、様々な言語データの分析で成果を上げている (代表的な先行研究に関しては、Hyland 2005 にまとめられている)。

表 1: metadiscourse markers のカテゴリー (Based on Hyland and Tse 2004)

Category	Function	Examples
<b>Interactive resources</b>	<b>Help to guide reader through the text</b>	
Transitions (TRA)	Express semantic relation between main clauses	in addition / but / thus / and
Frame markers (FRM)	Refer to discourse acts, sequences, or text stages	finally / to conclude / my purpose here is to
Endophoric markers (END)	Refer to information in other parts of the text	notes above / see Fig / in section 2
Evidentials (EVI)	Refer to source of information from other texts	according to X / (Y, 1990) / Z states
Code glosses (COD)	Help readers grasp functions of ideational material	namely / e.g. / such as / in other words
<b>Interactional resources</b>	<b>Involve the reader in the argument</b>	
Hedges (HED)	Without writer's full commitment to proposition	might / perhaps / possible / about
Boosters (BOO)	Emphasize force or writer's certainty in proposition	in fact / definitely / it is clear that
Attitude markers (ATM)	Express writer's attitude to proposition	unfortunately / I agree / surprisingly
Engagement markers (ENG)	Explicitly refer to or build relationship with reader	consider / note that / you can see that
Self-mentions (SEM)	Explicit reference to author(s)	I / we / my / our

前述のように、談話標識には書き言葉に顕著なものと話し言葉に顕著なものがあるが、Hyland list は書き言葉の分析を想定したリストである。また、多くの研究で「談話標識」とみなされている接続詞や接続副詞の多くは、Transitions (TRA) や Frame markers (FRM) に含まれているが、Hedges (HED) や Boosters (BOO) のような stance markers (Biber *et al.* 1999), Self-mentions (SEM) のような人称代名詞のみならず、書き手の態度や評価を表す動詞なども含まれているため、非常に多角的な分析が可能である。以下、本研究においては、Hyland によってリスト化された metadiscourse markers を「談話標識」と呼び、分析の対象とする。

### 3. Method

#### 3.1 Research questions

本研究の目的は、日本人中学生、高校生、大学生の英作文を集めた学習者コーパスをテキストマイニングの技法を用いて解析し、そこから得られた結果を母語話者のコーパスと比較することである。その目的を達成するために、本研究では、以下の 2 つの research questions (RQ) を設定する。

- 中学生、高校生、大学生と学年が上がるにつれて、談話標識の頻度や使用傾向はどのように変化するのか
- 学習者（中学生、高校生、大学生）と母語話者の間には、談話標識の頻度や使用傾向にどのような違いがあるのか

これら 2 つのアプローチに関して、Granger (1998) の表現を用いるならば、前者は「中間言語の異なる段階の比較」(IL-IL comparison) であり、後者は「母語と中間言語の比較」(NL-IL comparison) である。また、本稿では、テキストマイニングを用いて談話標識の全体的な使用傾向を俯瞰したのち、書き言葉における談話標識であると多くの先行研究でみなされてきた接続詞 (e.g. *and*, *but*, *so*, *because*) に光を当てて、詳細な量的・質的分析を行なう。

### 3.2 Data

本研究では、JEFL (the Japanese EFL Learner) Corpus<sup>1</sup>, ICLE-JP (the Japanese component of the International Corpus of Learner English)<sup>2</sup>, LOCNESS (the Louvain Corpus of Native English Essays)<sup>3</sup> の3種類のコーパス (総語数は661043語) をデータとして使用する。

JEFL Corpus は、日本の中学生と高校生による自由英作文を集めた学習者コーパスである (約60万語)。本研究では、ICLE-JP および LOCNESS とのデータの整合性を考慮し、論説文 (argumentative essay) のデータのみを分析対象とする。なお、著作権者の許諾を得てプレリリース版を分析した小林・山田 (2008), 小学館コーパスネットワーク (SCN) で無償公開されている全データ (叙述文を含む) を分析した Kobayashi (2009) および小林 (2009a, 2009b, 2009c) とは、分析対象としているデータのサイズが異なる。

ICLE-JP は、日本の大学生による英作文 (論説文) を集めた学習者コーパスである (約17万語)。まもなく一般公開される予定だが、本研究では、著作権者の許諾を得てプレリリース版を分析対象とする。

LOCNESS は、英米の母語話者による英作文を集めたコーパスであり、ICLE の参照コーパス (reference corpus) として設計された (約30万語)。これは、コーパス作成者である S. Granger か S. De Cock にコンタクトを取ることで入手可能 (有償) である。本研究では、アメリカ人大学生による論説文のみを分析対象とするため、LOCNESS の全データを分析対象とした Kobayashi (2009) および小林 (2009a, 2009b) とは、分析対象としているデータのサイズが異なる。

表2は、本研究で使用するデータとその総語数をまとめたものである。なお、表中の JH, SH, UNI, NS は、それぞれ中学生、高校生、大学生、母語話者を表している。また、作文タスクの詳細については、投野 (2007) および Granger (1998) を参照されたい。

表2: 使用データ

	JH	SH	UNI	NS
Corpus	JEFL		ICLE-JP	LOCNESS
Tokens	162919	179750	168800	149574

### 3.3 Procedure

本研究は、テキストマイニングの技法を駆使し、英作文における談話構造に解析するものである。また、コーパス中の文章に品詞情報・構文情報・談話情報などを自動付与し、それらの情報から得られる頻度パターンに対して様々な量的分析を行なう。しかしながら、コーパス言語学において、量的分析と質的分析は常に相補的な関係になければならない (McEnery & Wilson 2001)。従って、テキストマイニングによって全体的な傾向が把握されたのちは、「テキスト」そのものに戻って、コンコーダンスの精緻な読みがなされなければならない。そして、テキスト読解から得られる知見は、新たな量的分析のための手がかりを与えてくれる。このようにして、量的分析と質的分析は有機的に循環していく (図1)。

<sup>1</sup> [http://scn02.corpora.jp/~jefll03/jefll\\_top.html](http://scn02.corpora.jp/~jefll03/jefll_top.html)

<sup>2</sup> <http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/icle.htm>

<sup>3</sup> <http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/LOCNESS1.htm>

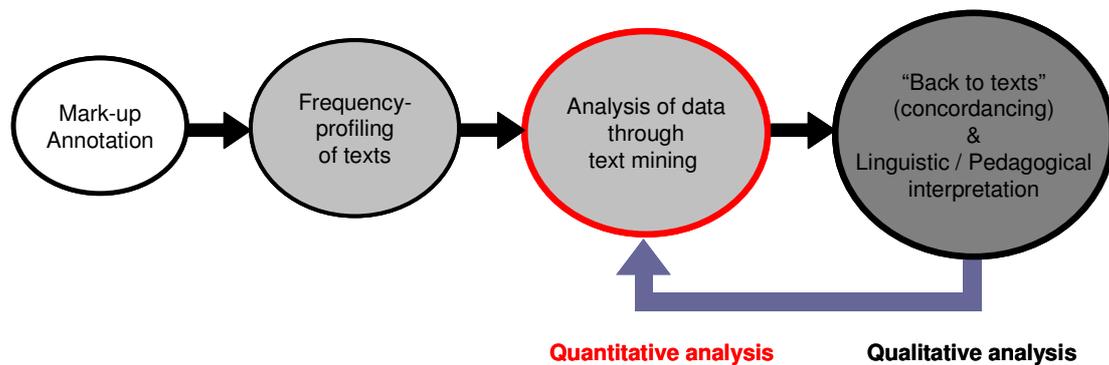


図 1: 分析の流れ (田畑 2009 を一部改変)

## 4. Results and discussion

### 4.1 Statistical analysis

#### 4.1.1 Frequency-profiling

まず、一部の記号（括弧，疑問符，ハイフンなど）を除いて，Hyland list における全ての談話標識の頻度を調査したところ，355 種類の談話標識が延べ 138988 回使用されていた。表 3 は，355 種類の談話標識を 10 種類のカテゴリー（表 1）に分類した粗頻度を集計したものである。また，図 2 はそれを 10 万語あたりの相対頻度に変換したのちに視覚化したものである。

表 3: 談話標識のカテゴリー別の頻度集計表（粗頻度）

	JH	SH	UNI	NS	chi-squared	p-value
COD	568	1133	1475	1175	1480.964	***
END	114	195	334	306	496.623	***
EVI	0	4	44	82	1685.416	***
FRM	2573	2442	2213	853	162.497	***
TRA	11788	11325	8443	6455	741.850	***
ATM	419	985	887	503	419.840	***
BOO	1550	2506	3170	1766	1735.963	***
SEM	23622	22549	8869	2018	12992.040	***
ENG	1985	2261	3242	2470	2422.122	***
HED	1343	2385	2291	2645	3037.304	***

df = 3, \*\*\* =  $p < .001$

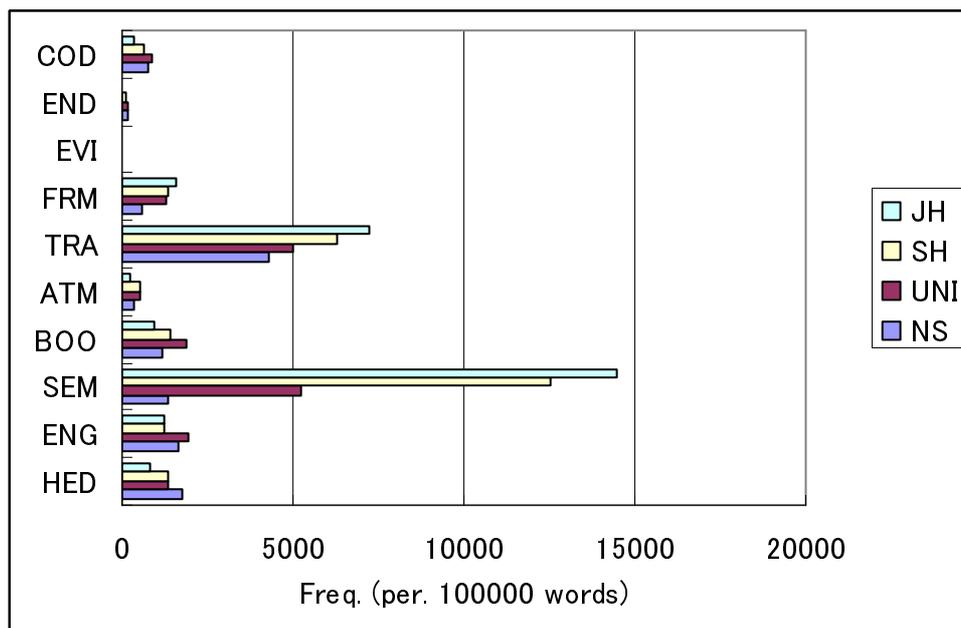


図 2: 談話標識のカテゴリー別の頻度集計グラフ (相対頻度)

図 2 を見ると、中学生や高校生が SEM や TRA を極端に過剰使用していることがよく分かる。また、表 3 における“chi-squared”と“p-value”は、それぞれ  $\chi^2$  統計量と有意確率を表している。全てのカテゴリーに 0.1%水準の有意差が見られるが、コーパスのような大きな標本を扱う場合は有意差が出やすい傾向にあることが知られているため (Kilgarriff 1996)、この結果のみで何かを結論するのは早計である。

#### 4.1.2 Correlation matrix

中学生、高校生、大学生、母語話者の 4 グループは、談話標識の使用パターンに関して、どの程度の類似性を持っているのであろうか。表 4 は、談話標識のカテゴリー別頻度集計表 (表 3) に対して、各個体間における Pearson の積率相関係数を求めた結果である。また、図 3 は、個体の相関係数行列の中で相関係数  $r$  が  $0.7 \leq r$  を満たす個体のネットワーク図を示す。

表 4: Pearson の積率相関係数

	JH	SH	UNI	NS
JH	1.000			
SH	0.998	1.000		
UNI	0.910	0.916	1.000	
NS	0.455	0.466	0.760	1.000

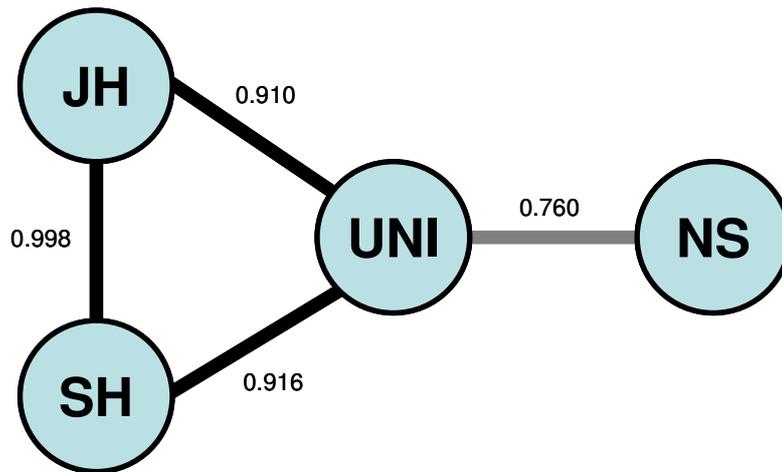


図 3: 相関係数ネットワーク

表 4 によれば、学習者（中学生，高校生，大学生）の間では相関係数は 0.910～0.998 と極めて高く，談話標識の使用傾向という点に限って言えば，これら 3 つのグループはかなり類似した言語特徴を持っている。それに対して，母語話者と学習者の相関係数は 0.455～0.760 と相対的に低い。

#### 4.1.3 Data reduction

次に，分析データ（中学生，高校生，大学生，母語話者）における談話標識の全体的な使用傾向を俯瞰するために，対応分析 (e.g. Abe 2007; Kobayashi 2009; Nakamura 1993, 1994, 1995, 2002; Nakamura & Sinclair 1995; Tabata 2002, 2005, 2008; Tono 2000; 小林 2007, 2008a, 2008b; 小林・山田 2008) を行なう。この手法は，大量のデータを分類，整理，縮約することでデータの全体像を掴み，個体間の関係，変数間の関係，個体と変数の関係を多次元空間上に視覚化するものである。数学的には林知己夫の数量化 III 類や西里静彦の双対尺度法と同じ計算方法で，クロス表における行と列の相関（正準相関）を最大化する解析法である。なお，この解析法は，1 回の固有値計算や特異値分解によって解が得られるために計算が簡便であり，広く利用されている。さらに，計算過程のオプションがないため，因子分析や主成分分析といった解析法よりも再現性が高い。

図 4 は，データ中に生起する全て（355 種類）の談話標識を変数として，対応分析を実行した結果である。この図は，解析の結果として得られる得点のうち，最も寄与率の高い第 1 次元（77.32%）と第 2 次元（18.93%）を 2 次元散布図に布置したものである（第 2 次元までの累積寄与率は 96.25%）。図中で近接する項目は類似した性質があることを示し，図中の項目間を隔てる距離が大きいほど異質性が高いことを示す。但し，解析結果を解釈するにあたっては，各次元が直交している（無相関である）ことに注意しなければならない。なお，視認性を重視し，変数（談話標識）のラベルは非表示としている。この図を見ると，第 1 次元（横軸）の右から左へ向かって，中学生，高校生，大学生，母語話者の順で個体（サブコーパス）が布置されており，談話標識の使用傾向と英語習熟度が関連性を持っていることが分かる。また，変数の分布を見ると，第 1 次元の負の帯域（左側）に数が多く，学年が上がって，習熟度が上がっていくにつれて，産出できる談話標識の種類が

増えていくことも分かる。

図 5 は、10 種類のカテゴリ別の頻度行列（表 3）に対して、対応分析を実行した結果である。この図は、解析の結果として得られる得点のうち、最も寄与率の高い第 1 次元 (93.84%) と第 2 次元 (4.60%) を 2 次元散布図に布置したものである（第 2 次元までの累積寄与率は 98.44%）。

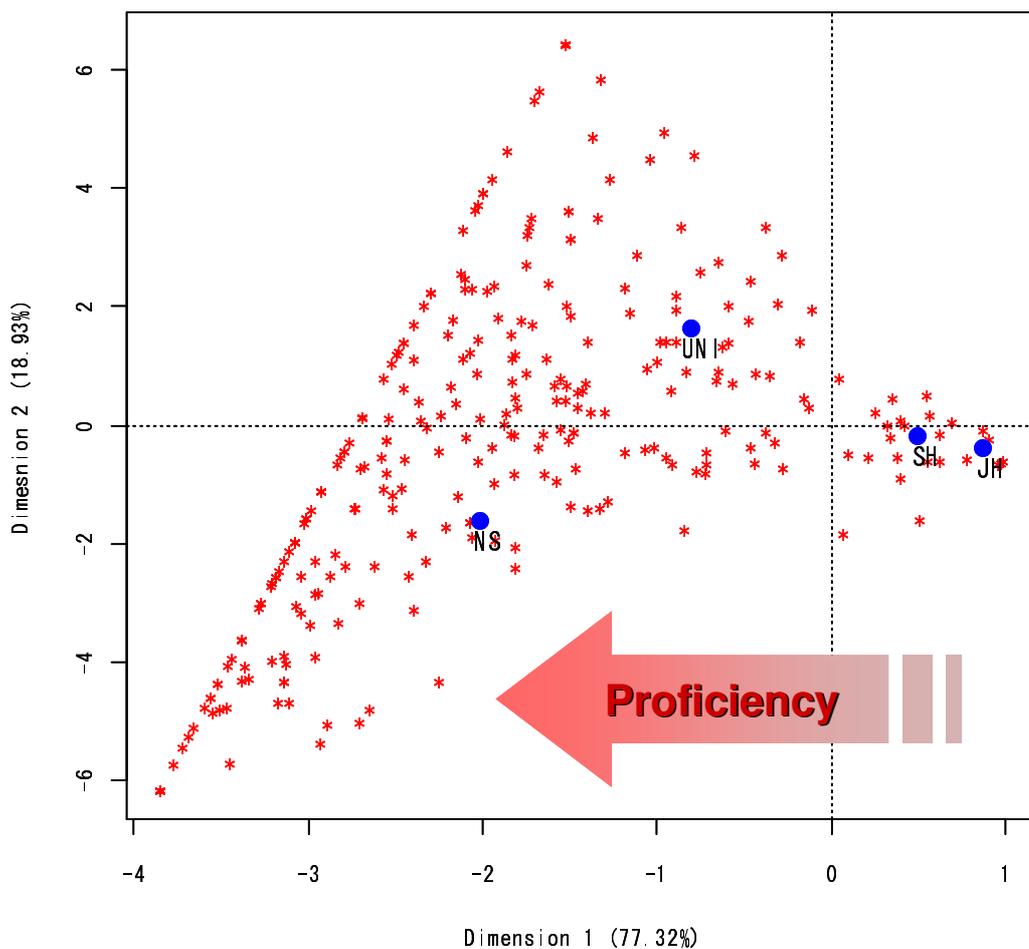


図 4: 355 種類の談話標識を変数とする対応分析

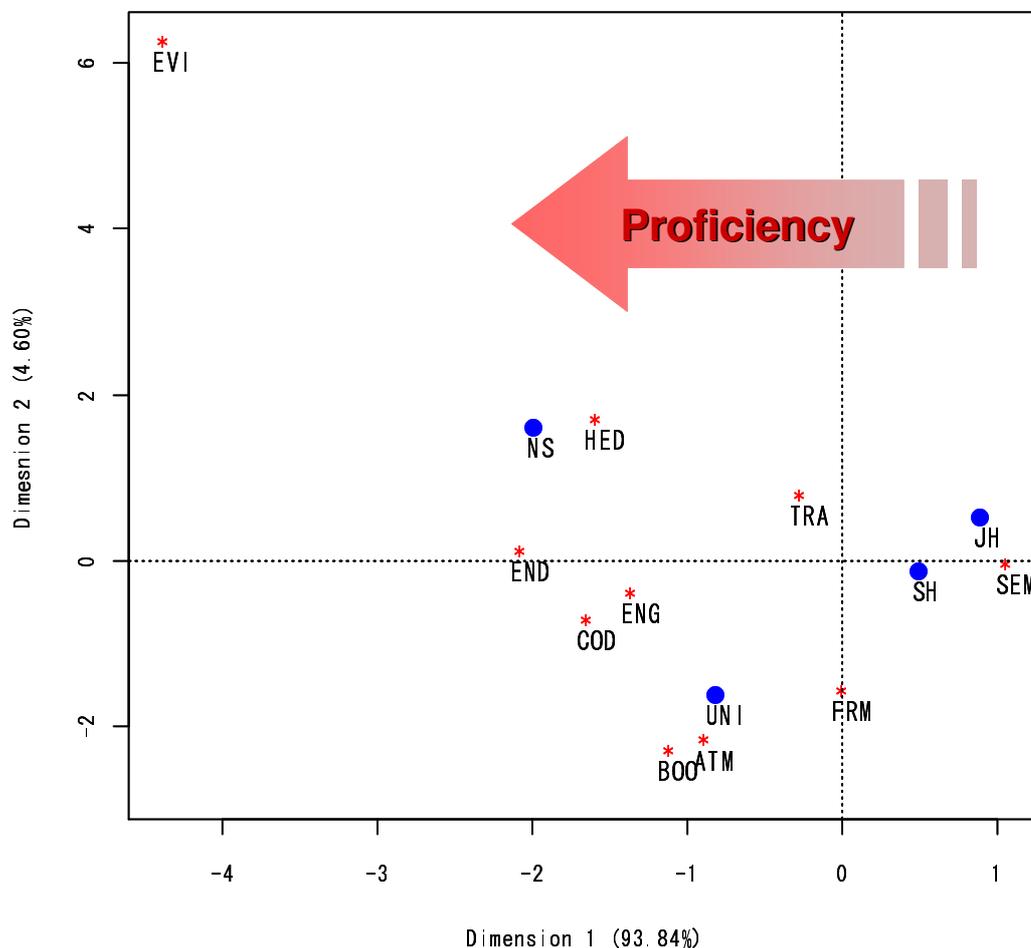


図 5: 10 種類の談話標識カテゴリーを変数とする対応分析

図 5 においても、第 1 次元に習熟度が反映されており、談話標識の使用傾向と習熟度が強い関連性を持っていることが分かる。そして、サブコーパスと変数の分布を見ると、中学生や高校生の談話に SEM, FRM, TRA などと顕著である一方、母語話者の談話には HED や END が顕著であることが分かる。

#### 4.2 Case study: Transitions

学習者による産出言語の中核を担うのは内容語であるが、機能語も文構成上で重要な働きを持っている。とりわけ、接続詞は、初期段階から英語学習教材などで多く提示され、学習者に馴染みのある語彙である。同時に、接続詞は、学習者が英作文などで多用する表現でもあり、いかなる学習者コーパスにおいても高い頻度で現れている (e.g. 小林 2009a, 2009b, 2009c)。しかしながら、接続詞や接続副詞のような接続語を適切に使用することは、

語彙に関する知識のみならず、構文や意味、そして談話に関する多様な知識が要求されるため (Tankó 2004), 必ずしも容易ではない。

学習者による接続詞や接続副詞の使用を分析した先行研究は数多い (e.g. Altenberg & Tapper 1998; Anping 2002; Bolton *et al.* 2002; Chiu 2004; Damascelli 2004; Field & Yip 1992; Granger & Tyson 1996; Haywood & Wilcoxon 1994; Milton 1999; 2001; Milton & Tsang 1993; Müller 2005; Narita & Sugiura 2006; Tankó 2004; Ventola 1992; Ying 2007)。これらの先行研究の多くは、母語話者と学習者のデータを統計的に比較し、学習者が有意に過剰使用/過少使用 (overuse / underuse) している表現を報告したものである。そこで用いられている方法論は、異なる母語を持つ学習者のデータを比較しているために、第2言語習得における母語の影響などを探るには有効である。その一方、同一の母語を持つ異なる学習段階の学習者データを分析しているものは極めて少なく、あったとしても、扱っているデータ規模は数十人程度と極めて小さい。その限りで、中学生、高校生、大学生による大規模コーパスを縦断的に解析する本研究の意義は大きい。

#### 4.2.1 Quantitative analysis

分析データにおける TRA の頻度は 38011 回であったが、その中の 34222 回 (90.03%) は頻度上位 4 語 (*and, but, so, because*) によるものである。従って、本稿では、その 4 語を中心に分析を進めていく。

図 6 は、中学生、高校生、大学生、母語話者の 4 グループにおける TRA 上位 4 語の相対頻度 (10 万語あたり) をまとめたものである。なお、本来ならばヒストグラムで描かれるべきデータ形式であるが、各グループ間の推移を捉え易いように折れ線グラフで描いている。

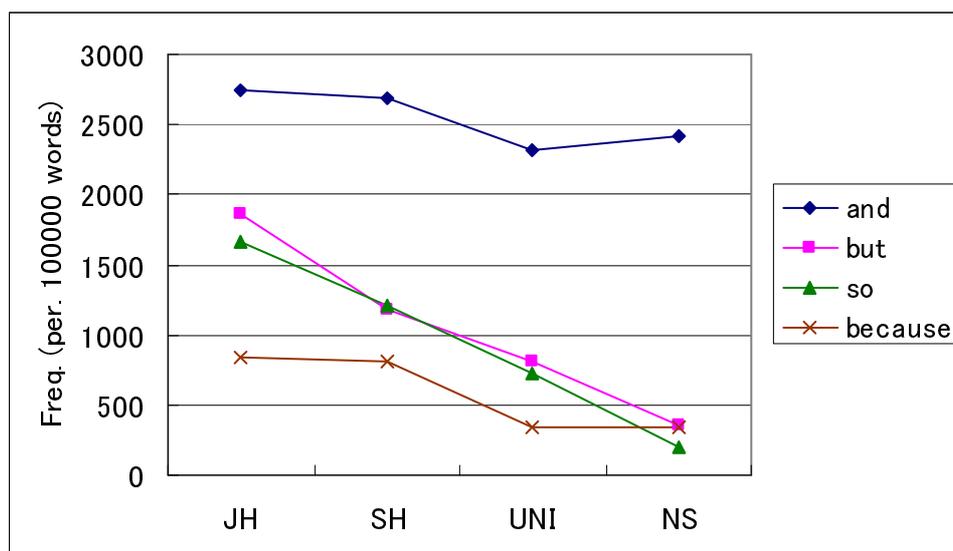


図 6: *and, but, so, because* の相対頻度

図 6 を見ると、4 語とも習熟度が上がっていくにつれて、有意に頻度が減少していく (*and*:  $\chi^2 = 88.407$ ,  $df = 3$ ,  $p < .001$ ; *but*:  $\chi^2 = 1818.366$ ,  $df = 3$ ,  $p < .001$ ; *so*:  $\chi^2 = 1958.101$ ,  $df = 3$ ,  $p < .001$ ).

< .001; *because*:  $\chi^2 = 660.092$ ,  $df = 3$ ,  $p < .001$ ). ICLE のフランス人学習者データと LOCNESS を統計的に比較した Granger & Rayson (1998) によっても報告されているように、接続詞の過剰使用は、学習者言語の顕著な特徴の 1 つである。そして、母語話者は、学習者が *and* や *but* ばかりを産出するのに対して、*also* や *however* などの接続副詞を巧みに使い分けることで、より効果的な文章を構築している。

次に、図 7 は、4 グループにおける TRA の頻度上位 4 語が文頭で生起する割合をまとめたものである。

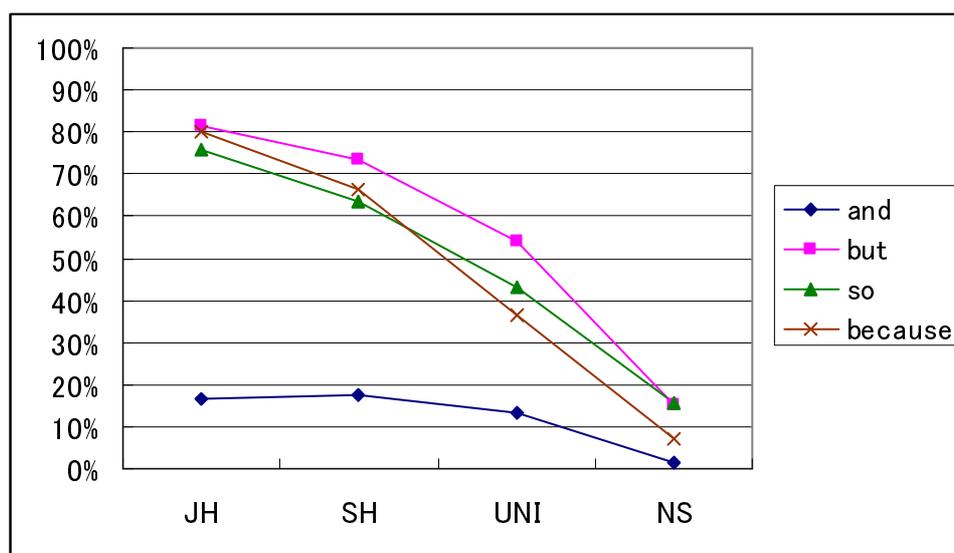


図 7: *and*, *but*, *so*, *because* が文頭で生起する割合

図 7 を見ると、4 語とも、習熟度が上がるにつれて文頭で生起する割合が小さくなっていく。それぞれの語に関して、文頭での頻度と文中での頻度の比率には有意差が見られる (*and*:  $\chi^2 = 588.916$ ,  $df = 3$ ,  $p < .001$ ; *but*:  $\chi^2 = 1104.453$ ,  $df = 3$ ,  $p < .001$ ; *so*:  $\chi^2 = 669.402$ ,  $df = 3$ ,  $p < .001$ ; *because*:  $\chi^2 = 979.620$ ,  $df = 3$ ,  $p < .001$ )。日本語では「また」、「しかし」、「なぜなら」のように、接続表現で文章を始めることが多いため、学習者が文頭の接続詞を好むのは、母語の干渉である。

また、Biber *et al.* (1999) が示しているように、文頭に生起する *and* や *but* は、話し言葉の特徴づけるものであり、一般的にフォーマルな書き言葉で使ってはならないとされている。それにもかかわらず、母語話者が *but* の約 15.14% を文頭で使っている点は注目に値する。これは、恐らく、語用論的用法、つまり語と語をつなぐ統語論的機能ではなく状態と状態をつなぐ語用論的機能を担う用法 (van Dijk 1977) であるが、この点については今後の詳細な分析が必要である。

そして、*so* と *because* に関しては、単に文頭の頻度を見るだけではなく、構文レベルや談話レベルの分析が求められる。まず、*so* という語には、一般的に、副詞としての用法 (e.g. We behaved so stupidly.) と従属接続詞としての用法 (e.g. It was quite windy, so we had to

button our coats up.) があるとされている。<sup>4</sup> さらに, Fraser (1993) は, 第3の用法として, 引用 (1) のように文頭に現れる *so* を談話標識の *so* と定義し, 引用 (2) のような従属接続詞の *so* などと区別している。

(1) John was sick. **So** don't expect him.

(2) John was sick, **so** he went to bed.

しかしながら, Fraser (1993) が談話標識の *so* と定義した用法は, 『ロングマン英和辞典』などにも明記されているように, 通常は話し言葉で用いられる用法である。それにもかかわらず, 学習者が文頭で *so* を多用することは, 彼らの言語使用が話し言葉の影響を受けていることの証左となる。

次に, *because* は, 多くの文法書や辞書にも書かれているように, 「主節の後ろにくるのが普通」(江川 1991) であり, 文頭で使うのは「先に理由や原因を述べる場合に限られる」(リーチほか 2006)。しかし, もっと問題にされるべきは, 学習者が文頭で *because* を用いた場合の多くが, 引用 (3)~(5) のような主節を持たない断片文であることである。

(3) \* **Because** it is my dream. (JH)

(4) \* **Because** I am Japanese! (SH)

(5) \* **Because** English is used all over the world. (UNI)

本研究の分析データでは, 中学生, 高校生, 大学生が文頭で *because* を使った場合, それぞれ 97.55%, 94.75%, 84.58% の割合で断片文を産出している。これも「なぜなら, それは私の夢だからです」のような文章が許容される日本語の影響であると考えられるが, 英語においても, 主節を持たない *because* の文も, 話し言葉における *why* で始まる疑問文の答えとしては許容される。従って, 学習者が断片文を多く産出する原因として, 彼らが書き言葉と話し言葉というモードの違いを理解していないことが挙げられる。さらに, 小林 (2009c) で明らかにされているように, 現在の中学校検定教科書で提示されている *because* の例文の数が極端に少なく, その半数以上がダイアログ部分で提示される主節を持たない文に現れていることも見逃せない。

しばしば, 「日本人学習者は, あたかも話しているように, 英語を書く」(Asao 2006) と指摘されてきたが, 学習者による接続詞の使用頻度や用法などにも, その特徴は顕著に見られる。

#### 4.2.2 Qualitative analysis

前節までは, あらゆる学習者コーパスは「様々な言語形式に関する独特の頻度行列」(Krzyszowski 1990) を持っているという仮説に基づいて, 学習者言語の計量分析を行なってきた。本節では, 個々の使用例を手作業で精査することによって, 質的な特徴抽出およ

---

<sup>4</sup> だが, 2つの用法の境界は必ずしも明確ではなく, 辞書や文法書によっても定義が分かれる。また, Leech *et al.* (2001) にいっては, linking words としての *so* は, 副詞と接続詞の両方で用いられるとしている。

び誤用分析を試みる。以下、いくつか典型的な特徴や誤用を示す。

前述のように、日本人学習者は、母語話者と比べて、*and*, *but*, *so*, *because* のような接続詞を有意に過剰使用し、とりわけ文頭でそれらの語を使用する傾向がある。これは、単文を連続させ、それらを文頭の接続詞で連結しようという意識の表れであると思われる。さらに、学習者が文頭で接続詞を用いている例の中には、引用 (6)~(17) のような、不必要なコンマを挿入する誤用が多く見られる。なお、本節では、接続表現以外の誤用については言及しない。

- (6) \* And<sub>2</sub> I like books and games. (JH)
- (7) \* And<sub>2</sub> I want to buy a sofa for a person. (SH)
- (8) \* And<sub>2</sub> people in the past couldn't go abroad easily. (UNI)
- (9) \* But<sub>2</sub> it's very interesting. (JH)
- (10) \* But<sub>2</sub> I want to be fat. (SH)
- (11) \* But<sub>2</sub> Japanese business people are very busy, so usually they don't have enough time to do it. (UNI)
- (12) \* So<sub>2</sub> I want it very much. (JH)
- (13) \* So<sub>2</sub> I have only four thousand yen now. (SH)
- (14) \* So<sub>2</sub> English is already second language in Japan. (UNI)
- (15) \* Because<sub>2</sub> we need money. (JH)
- (16) \* Because<sub>2</sub> my feeling is too bad in the morning. (SH)
- (17) \* Because<sub>2</sub> communication is very important thing. (UNI)

この種の誤用は、とりわけ *and* と *but* に多い。接続詞ではなく、*also* や *however* のような接続副詞の場合はコンマが必要となる (e.g. *However<sub>2</sub> it is very interesting.*)。不必要なコンマの挿入は、接続詞と接続副詞の違いに関する知識が学習者に不足していることに起因するものである。

また、個々の語の使用例を精査していくと、様々な誤用が散見される。先ず、引用 (18)~(20) を参照されたい。

- (18) \* I don't have tennis shoes and a hat! (JH)
- (19) \* I don't have rice and miso-soup in the morning. (SH)
- (20) \* Though, we can't still speak and write English fluently. (UNI)

このような場合、否定語のあとでは *and* ではなく、*or* (あるいは *nor*) を用いるとされている (e.g. *I don't have rice or miso-soup in the morning.*)。この用例については多くの辞書や文法書にも記述があるが、学習者の理解度はそれほど高くない。

以下の引用 (21)~(23) は、名詞を前方修飾する 2 つの形容詞をつなぐ *and* は不要であるとする Turton & Heaton (1996) の定義によれば、誤用とみなされる文である。だが、実際には、“It was the first time our national and international network had gathered together in one place and made us all realise just how much the work has grown” (BNC A00 158) のように *and*

が不可欠な文も存在し, “nice / good and [adjective]”という強意表現 (e.g. 岡田 2002) も存在するため, 一概に誤用とは言い切れない。しかしながら, 学習者が, 母語話者と比べて, “[adjective] and [adjective]”という連鎖を好むことは事実である。

(21) ? I want to buy a big **and** cool Sony [JP:CD コンボ]. (JH)<sup>5</sup>

(22) ? First, I'll wear a pair of strong **and** thick shoes in order to protect my feet from something strange. (SH)

(23) ? It is sad **and** miserable reality. (UNI)

そして, 学習者作文には, 引用 (24)~(25) のように, 順接の意味が希薄な *and* や逆接の意味が希薄な *but* の使用例が見られる。

(24) ? I was really fun **and** tired. (SH)

(25) I usually have bread.

? **But** I like rice.

I eat rice sometimes.

? **But** it's [JP:めったにない].

And I don't like milk.

? **But** I like [JP:味噌汁].

? **But** milk and [JP:味噌汁] drink [JP:めったにない]. (JH)

引用 (24) における *fun* と *tired* は対比されるべき表現であることから, この文章は, 逆接の接続詞である *but* などを使って, “I [really had] fun **but** [I was] tired.”や “[It] was [real] fun **but** [I was] tired.”などと書かれるべきものである。そして, 引用 (25) では, 7文のうち, 4つの文頭で *but* を使用しており, 2文続けて用いている箇所もある。しかし, (“I like rice, too.”のようなニュアンスにも読める) 2行目の “**But** I like rice.”のように, 必ずしも逆接とは言えない表現がある。Scollon & Scollon (1995) は, 日本人, 韓国人, 中国人などの学習者が用いる *and* や *but* がしばしば結束性を持っていないと指摘している。日本語の「が」は, 逆接だけでなく順接の場合にも用いられるため (綿貫ほか 2000), 逆接の意味が希薄な *but* も母語の干渉と見ることができる。

前節で言及した断片文にも, いろいろな種類がある。勿論, 引用 (3)~(5) のような単純に主節が脱落したものが数としては多いのだが, 引用 (26)~(31) のような断片文も散見される。

(26) \* **Because** I have a camera, **but** it is not so good. (JH)

(27) \* **Because** it is very useful **and** it has my memories. (SH)

(28) \* **Because** they suddenly barked **and** chased me. (UNI)

---

<sup>5</sup> JEFLF では, 学習者の流暢さを確保するために, 英語が思い浮かばない箇所を日本語で書くことを許可している。

- (29) \* **Because, if** we don't have a breakfast. (JH)
- (30) \* **Because** I have no time every morning, **so** I can eat bread for breakfast very quickly. (SH)
- (31) \* **Because when** I go to foreign country, I will can speak to them (foreigners) and attempt to communication. (UNI)

引用 (26)~(28) は, *because* で始まる従属節が等位接続詞 (*and, but*)によって拡張されているタイプの断片文である。これらは, 文と節の違い, あるいは主節と従属節の違いなどを理解していないために起こる誤りである。そして, 引用 (29)~(31) は, *because* で始まる従属節の中に, 他の従属接続詞 (e.g. *if, when*) によって導かれる従属節が入れ子のように入れ込まれているタイプの断片文である。また, 引用 (30) のような *because* と *so* の二重使用も多い。本来ならば, これは, *because* か *so* のどちらか一方だけで因果関係を表せる文である。この種の誤用は, 学習者が *because* や *so* を等位接続詞と同じように用いていることを反映している。因みに, 若干タイプは異なるが, 接続詞の二重試用という点では, 引用 (32) のような *but* と *although* の二重使用も見られる。

- (32) \* **Although** he died soon, **but** his story, and his name as a story teller, lasted forever. (SH)

さらに, 従属接続詞には主節と従属節の両方が必要であるという認識がない学習者は, 引用 (33) のように, しばしば非常に複雑な埋め込みがなされている文を産出する。

- (33) \* **If** I take all **then** I will die **because** there are very heavy, **so** I can't escape. (JH)

Turton & Heaton (1996) によれば, 最初の節が *if, since, as*, あるいは *because* で始まる場合は, その節に続く節を *so* で始めてはならない。しかし, 引用 (33) には, *if, because, so* という3つの従属接続詞が用いられており, それに加えて, 接続副詞の *then* も用いられている。これまで学習者による接続表現の過剰使用や, 主節と従属節の違いに関する知識の不足について見てきたが, 引用 (33) はその極端な例である。

## 5. Pedagogical suggestion

コーパス言語学の方法論は, これまで様々な言語学の領域でパラダイムシフトを引き起こしてきたが, 言語教育にも大きな影響を与えつつある (e.g. Bernardini 2004; Biber *et al.* 1994; Flowerdew 2001; Ghadessy *et al.* 2001; Gut 2006; Johns 1994; Kettemann & Marko 2002; Leech 1997; Meunier 2002; Mindt 1997; Murkherjee 2006; Partington 1998; Ragan 2001; Römer 2004; Todd 2001; Tribble 2001)。ここまで本稿において, 学習者言語に見られるいくつかの特徴や誤用を指摘してきた。以下, それらを踏まえて, 実際の作文指導に向けて, いくつかの提案を行なう。

まず, 接続詞の生起位置や不要な句読点の挿入のような誤用に対しては, DDL

(Data-driven learning) 的な発想で、母語話者の使用例を KWIC (Key Word In Context) concordance で示すことが有効である。学習者自身に使用パターンを発見させる学習法は、学習者の自律性 (autonomy) をサポートするという利点を持つ (e.g. Osborne 2004)。それと合わせて、引用 (18)~(20) のような誤用に対しては、従来型の選択問題や穴埋め問題を用いて、学習者の理解度を確認する必要があるであろう。

そして、断片文のような構文の誤用に対しては、前の文章と接続させる指導 (書き換え 1) や、文頭の *because* を *That is because* のような表現に置き換えるような指導 (書き換え 2) がもっと必要である。

断片文の例: \* I don't think English should be taught at primary school in Japan. **Because** children need to learn Japanese first.

書き換え 1: I don't think English should be taught at primary school in Japan **because** children need to learn Japanese first.

書き換え 2: I don't think English should be taught at primary school in Japan. **That is because** children need to learn Japanese first. (静 2003)

また、英語がある程度できる学生に対しては、従属接続詞が用いられている文の節構造や、等位接続詞が何と何をつないでいるのか (節と節なのか、句と句なのか、それとも語と語なのか)、などを明示的に提示することが有効である。

最後に、接続詞の過剰使用は、学習者が自らの文章の結束性に自信がないことの表れである。それに対して、代名詞のような他の言語表現も結束性を持つことを提示し、文や節をつなぐ際に接続詞以外の選択肢を学習者に与えることも必要である。また、英作文の教科書などには、「論理的な文章を書くための便利な表現」などの見出しで、接続表現のリストがある。このようなリストを見た学習者は、それらの表現を「多く」使うことが論理的な文章を構成することであると勘違いする可能性がある。従って、それらの表現を「多く」使うことが必ずしも「効果的に」使うことを意味しないということも示さなければならない。

## 6. Conclusion

本稿の RQ は、以下の 2 つであった。

- 中学生、高校生、大学生と学年が上がるにつれて、談話標識の頻度や使用傾向はどのように変化するのか
- 学習者 (中学生、高校生、大学生) と母語話者の間には、談話標識の頻度や使用傾向にどのような違いがあるのか

まず、第 1 の RQ に関しては、対応分析の結果にも明らかなように、談話標識の使用傾向と習熟度が強い関連性を持っていることが分かった。さらに、接続詞 (TRA) に焦点を絞って詳細な分析をした結果、*and*, *but*, *so*, *because* といった高頻度の接続詞の頻度、生

起位置、用法に顕著な違いが明らかにされた。そして、第2のRQに関しては、相関分析の結果にも明らかなように、談話標識の使用という点で学習者と母語話者に一定の違いが見られた。また、学習者の習熟度が上がっていくにつれて、徐々に母語話者の用法に近づいていくことも確認された。

このような学習者コーパスの分析は、SLA研究に有益な量的・質的データを提供し、Gass & Selinker (2001) が提起する、学習者言語におけるルールは「母語話者言語のルールのようなものなのか」、「学んでいる言語のルールのようなものなのか」、そして、「母語や学んでいる言語にかかわらず、全ての学習者に共通するパターンはあるのか」、「第2言語学習者によって生み出されたルールは、状況によって変化するものなのか」といった問いに答えようとするものである。それと同時に、言語教育という視点では、コーパス研究は、単に言語使用を記述するだけでなく、新たな言語教材、シラバス・デザイン、教授法を生み出し、ときにはコーパスそのものが教材となる (e.g. Granger 2004; Hunston 2002)。その限りで、学習者コーパス研究は、単なる統計結果などの数値データを見るだけでは不十分であり、常に言語学的考察をともない、教育的効果を想定してなされなければならない。

## References

- Abe, M. (2007). A corpus-based investigation of errors across proficiency levels of L2 spoken production. *JACET Journal*, 44, 1-14.
- Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.
- Aijmer, K. (2008). At the interface between grammar and discourse: A corpus-based study of some pragmatic markers. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente*. (pp. 11-36). Berlin: Mouton de Gruyter.
- Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80-93). London: Longman.
- Anping, H. (2002). On the discourse marker *so*. In P. Peter, P. C. Collins, & A. Smith (Eds.), *New frontiers of corpus research: Papers from the twenty first international conference on English language research on computerized corpora Sydney 2000* (pp. 41-52). Amsterdam: Rodopi.
- Ariel, M. (1994). Pragmatic operators. In R. E. Asher & J. M. Y. Simpson (Eds.), *The encyclopedia of language and linguistics* (pp. 3250-3253). Oxford: Pergamon Press.
- Asao, K. (2006). Spoken and written discourse of EFL learners: Findings from the sound corpus of Japanese learners of English. A paper given at the ICLE / LINDSEI Japanese Sub-Corpus Symposium. Tokyo: Showa Women's University.
- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- Bazanella, C. (1990). Phatic connectives as intonational cues in contemporary spoken Italian. *Journal of Pragmatics*, 14, 629-647.
- Bernardini, R. (2004). Corpora in the classroom. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp.

- 15-36). Amsterdam: John Benjamins.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches in applied linguistics. *Applied Linguistics*, 15, 169-185.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore, D. (1992). *Understanding utterances*. Oxford: Blackwell.
- Bolton, D., Nelson, G., & Hung, J. (2002). A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, 7, 165-182.
- Chiu, Y. (2004). The L2 acquisition of the coordinating conjunction “and” in Taiwanese learners’ interlanguage at the college level. *Asian EFL Journal*, 6, 1-20.
- Clahsen, H. (1980). Psycholinguistic aspects of L2 acquisition. In S. Felix (Ed.), *Second language development: Trends and issues* (pp. 57-79). Tübingen: Gunter Narr.
- Conner, U., & Mbaye, A. (2002). Discourse approaches to writing assessment. *Annual Review of Applied Linguistics*, 22, 263-278.
- Conrad, S. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22, 75-95.
- Corder, S. P. (1967). The significance of learners’ errors. *International Review of Applied Linguistics*, 5, 161-169.
- Crewe, W. J. (1990). The illogic of logical connectors. *ELT Journal*, 44, 316-325.
- Crismore, A., Markkanen, R., & Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish students. *Written Communication*, 10, 37-71.
- Damascelli, A. T. (2004). The use of connectors in argumentative essays by Italian EFL learners. In M. T. Prat Zagrebelsky (Ed.), *Computer learner corpora* (pp. 138-160). Alessandria: Edizioni dell’Orso.
- van Dijk, T. A. (1977). Connectives in text grammar and text logic. In T. A. van Dijk & J. S. Petöfi (Eds.), *Grammars and descriptions* (pp. 11-63). Berlin: W. de Gruyter.
- Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning*, 23, 37-53.
- Erman, B. (1987). *Pragmatic expressions in English: A study of “you know,” “you see,” and “I mean” in face-to-face conversation*. Stockholm: Almqvist and Wiksell.
- Færch, C., Haastrup, K., & Phillipson, K. (1984). *Learner language and language learning*. Clevedon: Multilingual Matters.
- Field, Y., & Yip, L. M. O. (1992). A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal*, 23, 15-28.
- Flowerdew, L. (2001). The exploitation of small learner corpora in EPA materials. In M. A. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 363-379). Amsterdam: John Benjamins.
- Fraser, B. (1987). Pragmatic formatives. In J. Verschueren & M. Bertucelli-Papi (Eds.), *The pragmatic perspective*

- (pp. 179-194). Amsterdam: John Benjamins.
- Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, 14, 383-395.
- Fraser, B. (1993). Discourse markers across language. *Pragmatics and Language Learning*, 4, 1-16.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31, 931-952.
- Gass, S., & Selinker, L. (2001). *Second language acquisition: An introductory course*. Mahwah: Lawrence Erlbaum.
- Ghadessy, M., Henry, A., & Roseberry, R. L. (Ed.) (2001). *Small corpus studies and ELT*. Amsterdam: John Benjamins.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). London: Longman.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.
- Granger, S. (2004). Computer learner corpus research: Current state and future prospects. In U. Conner & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 123-145). Amsterdam: Rodopi.
- Granger, S. & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on computer* (pp. 119-131). London: Longman.
- Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15, 17-27.
- Gut, U. (2006). Learner speech corpora in language teaching. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 69-86). New York: Peter Lang.
- Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning*, 26, 231-251.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Haywood, K., & Wilcoxon, H. C. (1994). Connectives in context. *English Teaching Forum*, 32, 20-23.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. New York: Continuum.
- Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25, 156-177.
- Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge: Cambridge University Press.
- Jucker, A. H., & Ziv, Y. (eds.) (1997). *Discourse markers: Descriptions and theory*. Amsterdam: John Benjamins.
- Kettemann, B., & Marko, G. (2002). *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi.
- Kilgarriff, A. (1996). Which words are particularly characteristic of a text? A survey of statistical approaches. *Proceedings of AISB Workshop on Language Engineering for Document Analysis and Recognition*, 33-40.
- Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18, 35-62.
- Kobayashi, Y. (2009). Profiling metadiscourse markers in native and non-native English. *Lexicon*, 39, 1-17.
- Krzeszowski, T. (1990). *Contrasting language: The scope of contrastive linguistics*. Berlin: Mouton de Gruyter.

- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.
- Leech, G. (1992). Corpora and theories of linguistic performance. In J. Startvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991* (pp. 105-122). Berlin: Mouton de Gruyter.
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, A. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1-23). London: Longman.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xx). London: Longman.
- Leech, G., Cruikshank, B., & Ivanik, R. (2001). *An A-Z of English grammar and usage*. New ed. Harlow: Pearson Education.
- Leech, G., & Startvik, J. (1994). *A communicative grammar of English*. 2nd ed. London: Longman.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk*. 2nd ed. Hillsdale: Lawrence Erlbaum Associates.
- McCarthy, M. (1993). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
- McEnery, T., & Wilson, A. (2001) *Corpus linguistics*. 2nd ed. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Meunier, F. (2002). The pedagogical value of native and learner corpora. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 119-141). Amsterdam: John Benjamins.
- Miesel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3, 109-135.
- Milton, J. (1999). Lexical thickets and electronic gateways: Making text accessible by novice writers. In C. N. Candlin & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 221-243). London: Longman.
- Milton, J. (2001). *Elements of a written interlanguage: A computational and corpus-based study of institutional influences on the acquisition of English by Hong Kong Chinese students*. Hong Kong: Language Center, the Hong Kong University of Science and Technology.
- Milton, J., & Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In R. Pemberton & E. S. C. Tsang (Eds.), *Studies in lexis* (pp. 215-246). Hong Kong: The Hong Kong University of Science and Technology.
- Mindt, D. (1997). Corpora and the teaching of English in Germany. In A. Wichmann, S. Fligelstone, A. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 40-50). London: Longman.
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.
- Murkherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5-24). New York: Peter Lang.
- Nakamura, J. (1993). Statistical methods and large corpora: A new tool for describing text types. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 293-312). Amsterdam: John Benjamins.
- Nakamura, J. (1994). Extended HAYASHI's quantification method type III and its application in corpus linguistics.

- Journal of Language and Literature, Faculty of Integrated Arts and Sciences, 1*, 141-192.
- Nakamura, J. (1995). Text typology and corpus: A critical review of Biber's methodology. *English Corpus Studies*, 2, 75-90.
- Nakamura, J. (2002). A galaxy of words: structures based upon distributions of verbs, nouns and adjectives in the LOB corpus. In T. Saito, J. Nakamura, & S. Yamazaki (Eds.), *English corpus linguistics in Japan* (pp. 19-42). Amsterdam: John Benjamins.
- Nakamura, J., & Sinclair, J. M. (1995). The world of *woman* in the Bank of English: Internal criteria for the classification of corpora. *Literary and Linguistic Computing*, 10, 99-110.
- Narita, M., Sugiura, M. (2006). The use of adverbial connectors in argumentative essays by Japanese EFL college students. *English Corpus Studies*, 13, 23-42.
- Osborne, J. (2004). Top-down and bottom-up approaches to corpora in language teaching. In U. Conner & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 251-265). Amsterdam: Rodopi.
- Östman, J.-O. (1985). Pragmatic particles twenty years after. In B. Wärvik (Ed.), *Organization in discourse: Proceedings from the Turku Conference* (pp. 95-108). Turku: University of Turku.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Partington, A. (1998). *Patterns and meaning: Using corpora for English language research and teaching*. Amsterdam: John Benjamins.
- Perdue, C. (1993). *Adult language acquisition: Cross-linguistic perspectives*. 2 vols. Cambridge: Cambridge University Press.
- Polanyi, L., & Scha, R. (1983). The syntax of discourse. *Text*, 3, 261-270.
- Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, 26, 81-114.
- Ragan, P. (2001). Classroom use of a systemic functional small learner corpus. In M. A. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 207-236). Amsterdam: John Benjamins.
- Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14, 367-381.
- Richards, J. (1971). Error analysis and second language strategies. *Language Sciences*, 17, 12-22.
- Römer, U. (2004). A corpus-driven approach to modal auxiliaries and their didactics. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 185-199). Amsterdam: John Benjamins.
- Rühleman, C. (2007). *Conversation in context: A corpus-driven approach*. London: Continuum.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-231.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schorup, L. C. (1985). *Common discourse particle in English conversation: "Like," "well," "y'know."* New York: Garland.
- Scollon, R., & Scollon, S. (1995). *Intercultural communication*. Oxford: Blackwell.
- Stubbs, M. (1983). *Discourse analysis*. Chicago: The University of Chicago Press.
- Swan, M. (2005). *Practical English usage*. 3rd ed. Oxford: Oxford University Press.
- Tabata, T. (2002). Investigating stylistic variation in Dickens through correspondence analysis of word-class distribution. In T. Saito, J. Nakamura, & S. Yamazaki (Eds.), *English corpus linguistics in Japan* (pp. 165-182). Amsterdam: John Benjamins.
- Tabata, T. (2005). Profiling stylistic variations in Dickens and Smollett through correspondence analysis of low

- frequency words. *ACH/ALLC 2005: The International Conference on Humanities Computing and Digital Scholarship Conference Abstracts*, 229-232.
- Tabata, T. (2008). *Gentleman* in Dickens: A multivariate stylometric approach to its collocation. *Digital Humanities 2008, Book of Abstracts, the 20th Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computer and the Humanities, University of Oulu, Finland, June 24-June 29, 2008*, 199-202.
- Tankó, G. (2004). The use of adverbial connectors in Hungarian university students' argumentative essays. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 157-181). Amsterdam: John Benjamins.
- Todd, R. (2001). Induction from self-selected concordances and self-correction. *System*, 29, 91-102.
- Tono, Y. (2000). A corpus-based analysis of interlanguage development: Analyzing part-of-speech tag sequences of EFL learner corpora. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *PALC '99: Practical applications in language corpora* (pp. 323-340). Frankfurt: Peter Lang.
- Tono, Y. (2005). Corpus-based SLA research: State of art of learner corpus studies. In M. Minami, H. Kobayashi, M. Nakayama & H. Shirai (Eds.), *Studies in language science (4): Papers from the fourth annual conference of the Japanese Society for Language Sciences* (pp. 45-77). Tokyo: Kuroshio.
- Tribble, C. (2001). Small corpora and teaching writing. In M. A. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 381-408). Amsterdam: John Benjamins.
- Turton, N. D., & Heaton, J. B. (1996). *Longman dictionary of common errors*. New ed. London: Longman.
- Vande Kopple, W. (1985). Some explanatory discourse on metadiscourse. *College Composition and Communication*, 36, 82-93.
- Ventola, E. (1992). Writing scientific English: Overcoming intercultural problems. *International Journal of Applied Linguistics*, 2, 191-220.
- Warner, R. G. (1985). *Discourse connectives in English*. New York: Garland.
- Ying, S. (2007). An analysis of discourse markers used by non-native learners: Its implications for teaching English as a foreign language. *Intercultural Communication Studies*, 19, 51-83.
- Young, R. F. (2002). Discourse approaches to oral assessment. *Annual Review of Applied Linguistics*, 22, 243-262.
- Zamel, V. (1983). Teaching those missing links in writing. *ELT Journal* 37, 22-29.
- 和泉絵美・内元清貴・井佐原均 (編) (2004). 『日本人 1200 人の英語スピーキングコーパス』 東京: アルク.
- 江川泰一郎 (1991). 『英文法解説』 改訂三版. 東京: 金子書房.
- 岡田禎之 (2002). 『現代英語の等位構造—その形式と意味機能』 大阪: 大阪大学出版局.
- 小林雄一郎 (2007). 「The NICT JLE Corpus と語彙研究—SST レベルの再検証」 『英文學誌』 49, 17-29.
- 小林雄一郎 (2008a). 「高頻出語を用いた学習者コーパスの分類」 石川慎一郎 (編) 『学習者コーパスの解析に基づく客観的作文評価指標の検討』 (統計数理研究所共同研究レポート 215), 69-82.
- 小林雄一郎 (2008b). 「多変量解析による中間言語の特徴抽出—名詞的発話から動詞的発話へ」 『専修人文論集』 83, 165-180.
- 小林雄一郎 (2009a). 「日本人学習者の英作文における等位接続詞の使用について—“and”と“but”を例に」 『専修大学外国語教育論集』 37, 21-36.
- 小林雄一郎 (2009b). 「日本人学習者の英作文における“so”の統計的分析」 石川慎一郎 (編) 『コーパス言語研究における量的データ処理のための統計手法の概観』 (統計数理研究所共同研究レポート 232),

107-118.

小林雄一郎 (2009c). 「日本人英語学習者の英作文における because の誤用分析」 『関東甲信越英語教育学会研究紀要』 23, 11-21.

小林雄一郎・山田洋文 (2008). 「日本人英語学習者コーパスに見る Metadiscourse Markers の使用傾向」 『英語コーパス研究』 15, 161-173.

静哲人 (2003). 『構造から学べるパラグラフライティング入門』 東京: 松柏社.

田畑智司 (2009). 「多変量アプローチで俯瞰する歴代大統領就任演説の言語変異—Washington から Obama まで」 英語コーパス学会第 33 回大会 (2009 年 4 月 25 日, 神戸大学) 発表資料.

投野由紀夫 (編) (2007). 『日本人中高生一万人の英語コーパス“JEFLL Corpus”: 中高生が書く英文の実態とその調査』 東京: 小学館.

リーチ, G・池上嘉彦・上田明子・長尾真・山田進 (監修) (2006). 『ロングマン英和辞典』 東京: 桐原書店.

綿貫陽・宮川幸久・須貝猛敏・高松尚弘 (2000). 『ロイヤル英文法』 東京: 旺文社.

(註) 本研究の一部は、第 22 回「英検」研究助成 (研究部門) 「テキストマイニングによる学習者作文における談話能力の測定と評価」 (研究代表者: 小林雄一郎) によって行なわれたものである。