

---

# *The JEFLL Corpus* における語彙・意味の分析

## —— *Wmatrix3* の適用と課題 ——

内田 富男

---

### 1. はじめに

第二言語習得研究 (SLA) においてコーパス分析の手法を用いる「コーパスに基づく第二言語習得研究」(Corpus-based second language acquisition research, corpus-based SLA research) では、学習者の産出言語を大規模に収集・電子化して、計量的に調査・分析することによりその言語特徴を記述する。Corpus-based SLA research は、ここ 10 数年間で急速に関心を集めている新しいテーマである。その研究成果は、様々な方法で外国語教育に応用できる。対象は、音韻、語彙、用法、形態素、構文等の言語習得に関わる計量的研究が一般的である。特に、上級学習者の文法、語彙習得に関する先行研究 (例 Granger 1998, Granger et al. 2002, 2009 他) は少なくない。欧州では 2013 年に Learner Corpus Association が発足し、2015 年には学習者コーパスに特化した学術雑誌である *International Journal of Learner Corpus Research* も発行される。

日本でも、2000 年前後にいくつかの日本人英語学習者コーパスが構築され、学習者コーパス研究が盛んになり、主に大学生を対象とした習得研究や母語話者との比較研究 (和泉等 2004, Ishikawa 2013 他) 等が見られるようになった。また、本研究で用いる *The JEFLL Corpus* を使った初・中級の日本人英語学習者に関するコーパス研究 (Tono 2002、投野 2007、小林 2009、能登原 2010、Tono et al. 2012、藤原 2014、飯尾 2013、投野等 2013、Uchida 2012、内田 2014 他) もある。

上記のような学習者コーパス研究では、通常、コーパス分析ツールを用いて学習者データに自動 (又は一部手動) で文法標識等のマークアップ作業を行い、語彙分析や構文解析によってデータを観察する。文法標識の自動付与システムとしてよく使われるものの一つは品詞標識自動付与ツールで、特に 1980 年代初頭にランカスター大学で開発された *CLAWS (the Constituent Likelihood Automatic Word-tagging System)* システムや *Penn Tree Bank* 等は広くコーパス研究で用いられている。*CLAWS* に基づく *CLAWS part-of-speech tagger for English* は代表的な品詞標識自動付与ツールで、*The British National Corpus* (BNC) のオンライン版である *BNCweb* でも用いられている。また、後に詳述する *Wmatrix3* は、単語単位の品詞標識の自動付与機能、

多重単語単位(Multi-word Unit, MWU)の分析、BNC との比較分析もできる。さらに、*Wmatrix3*には品詞標識付与に加え、意味標識の自動付与機能も搭載している。このように *Wmatrix3*は極めて多機能で、学習者英語の分析にも活用が期待されるウェブツールである。

そこで、本論では *Wmatrix3*を使って、*The JEFLL Corpus* のデータに品詞及び意味標識を自動付与し、語彙項目、品詞、意味の分析を通して、初・中級レベルの日本人学習者の英語を検証し、初・中級レベルの学習者英語のコーパス分析における *Wmatrix3*の活用可能性と課題について論じる。

## 2. *The JEFLL Corpus* について

### 2.1 *The JEFLL Corpus* とは

*The JEFLL Corpus*(以下、JLC と略記)が構築された *JEFLL Corpus* プロジェクトは、投野由紀夫氏(東京外国語大学)を中心に、日本の中学、高校の協力のもと、10年間以上の長期間にわたって展開された大規模プロジェクトである。JLC に収集された自由英作文データは日本人中高生延べ1万人分に及ぶ。データの総語数は2014年現在、約60万語であり、公開された初級・中級レベルの英語学習者の産出データとしては世界最大規模である。JLC は日本の英語教育環境に密着したデータであるため日本人学習者の英語習得データとして貴重である。

JLC を用いた研究の流れを概観する。1980年代後半に収集された20万語分の英作文を使った Tono & Kanatani(1996) と Tono(1996) から始まり、本格的に電子化された後、Tono & Aoki (1998)、Tono(2000)、Tono(2002)、Abe & Tono (2005) と続く。その後、JLC の規模は、3倍以上に拡大し、2007年には一般に無償公開され、小学館コーパスネットワークの一部を構成する SCN 版 *JEFLL Corpus* として知られることとなった。現在は、中高生1万人分の自由英作文データ約66万9千語の規模となっている。JLC の一般公開に合わせて、投野編著(2007)『日本人中高生一万人の英語コーパス：中高生が書く英文の実態とその分析』(小学館)が刊行された。同書の目次の一部を転載すると、内容は次の通りである。英語語彙発達のプロセス(Active Vocabulary と Lexical Collocation の発達)、英語品詞使用と発達のプロセス。各章は、品詞発達の概要、名詞、形容詞、副詞、動詞、接続詞、決定詞、助動詞、前置詞の順に品詞毎に研究例を紹介している。そして、英語構文発達のプロセス、品詞連鎖、主要文法事項・構文の発達、英語になりにくい日本語の分析、品詞エラーに着目した研究例も報告されている。

2007年以降には、上記の SCN 版や非公開版<sup>1)</sup>を使った JLC の研究(小林2009, Tono et al. 2012, Uchida 2012, 内田2014a, 2014b 他)が発表されている。さらに、2012年には世界版 *JEFLL Corpus Project* とも言える ICCI プロジェクト<sup>2)</sup>として、*The International Corpus of Crosslinguistic Interlanguage* が構築され、日本語以外の母語話者である初中級レベルの英語学習者のデータを収集し、コーパス化されることとなった。ICCI の基本設計は JLC のそれと同様で、異なる母語話者間の比較が可

能になっている。作文トピック等、異なる部分もあるため、完全に同一基準での比較ができるわけではないが、JLC と同レベルの英語学習者からデータ収集がなされているので有用性は高いだろう。

## 2.2 The JEFLL Corpus (JLC) のタスクとその特徴

JLC のタスクは、授業内に短時間で書いたクイック・ライティングで、課題作文ではあるが、和文英訳や制限作文とは異なり、内容、表現は自由で、自然な発話に近いと言えよう。また、後述のように自由英作文課題は様々な比較検証ができるように注意深くコントロールされているため、通常の教室で個々の授業者が収集する作文とは異なる。

一方で、留意しなければならない点もある。まず、サブコーパス（トピック、学年、学校レベル別）には偏りがあり、作文トピックが6つに限定されているため、研究目的等によってはトピック等の影響に配慮が必要な場合がある。さらに、初中級学習者の英作文で、作文時間が20分に制限されているので、英文エッセイのように十分に時間をかけて書いた作文とは違い、テキスト長は全体的に短めである。また、日本語使用が許容されているため英語で何とか表現しようとする部分は見えにくい。

JLC の英作文は、教室内で実施した辞書無しの自由英作文で、2タイプ・6トピック<sup>4)</sup>からなり、中学1年生から高校3年生まで統一的にデータ収集がなされている。JLC では、どうしても英語で書けない部分は日本語使用を認めている。また、タスクシートには学習者のレベルを考慮して、モデル作文<sup>5)</sup>が与えられている。そのため、モデル作文の影響が懸念され、事実、中1程度の低学年ではモデル作文の模倣が見られるケースもある。しかし、中学2年になると激減し<sup>6)</sup>、高校生の作文では、モデル文をそのまま写すような直接的な影響は少ない。

タスクの自由英作文のタイプは論説文と叙述文で、それぞれのタイプには3つのトピックがある。データ提供する協力校で任意に選択し、授業内に1トピックから6トピックについて授業内に書かせる。全ての参加者が6つ全てのトピックについて書くわけではないため個人の6点の作文を比較できるわけではない。また、JEFLL Corps プロジェクトは長期間に実施されたプロジェクトではあるが、特定の学習者の経年データが蓄積されているわけではない。従って従来のSLA研究のような縦断的なデータ収集方法ではないため、6つの学年と2タイプ・6トピックの作文の疑似横断・縦断的習得データといえる。また、データには日本語<sup>3)</sup>が混在している場合もあるので日英語の交差言語的視点での検証もできる。

## 3. Wmatrix3 : Wmatrix corpus analysis and comparison tool

語の関係性や構造、論理性と意味の組織化を具現化する一つの方法として、本研究では Wmatrix3 を活用する。特に、語彙の意味については、Stubbs(2006) によれば「ある言語における語彙は、多くの語のかたまり (cluster) によって内部的に構造化されていて、それぞれが互いに異なった関係を担っている。それは、同一性・差異性・

含意性といった論理的関係であったり、任意の話題領域ないし意味領域内のより漠然とした関係で…意味領域とは、ある話題に関する語の単なる羅列ではなく、そうした語の間のさまざまな関係によって組織化されたものでもある。本来、個々の語の意味はあいまいなものであるが、語彙としては構造化されている。」(47-48)と言われる。*Wmatrix3*の意味標識システムは談話・意味領域のまとまりを意味標識の組織化によって具体的に計量的に表現していると考えられる。

本節では、*Wmatrix3*の機能について述べる。*Wmatrix3*は、UCREL のウェブ型コーパスアノテーション・検索ツールである。アノテーション機能には、品詞標識付与(精度97%)、意味標識付与(精度92%)、語のレマ化ができる。出力は様々な語彙(レマ化形、非レマ形)、品詞標識(CLAWS7)、そして後述する意味標識(USAS))に基づく頻度リスト、KWIC コンコーダンス、コロケーション抽出と統計的共起性の算出等複数の統計情報、BNC の簡易版である *BNCsampler* との比較が可能で、極めて高機能である。

*Wmatrix3*における意味標識(UCREL Semantic Analysis System, USAS)のシステムでは、およそ37,000語(タイプ)に付与できる意味標識は、一般的に同一の心的概念に関連付けられる語義に分類し、談話・意味領域の構造を示す。また、同意語、反意語、さらには上位概念語、下位概念語と16,000種類の多重単語単位(Multi-Word-Unit, MWU)を含む。MWUには、句動詞、名詞句(複合名詞)、固有名詞、イディオム(true idiom)で構成される。

USASのタグセットは、以下のように21の英字大文字(D,R,U,Vの5文字を除く)の談話・意味領域を大分類<sup>4)</sup>とし、232の語彙範疇のラベルが付けられている。意味標識は、大分類である意味領域、小分類は意味領域を表す英字と数字と正負記号の組み合わせで付け、意味標識には正負記号が付くもの(例 E2+, E5-)と付かない標識(例 A13, C1 K1)がある。

## 4. 本研究

### 4.1 目的

本研究は、*Wmatrix3*の自動標識付与機能を用いて、初級・中級レベルの日本人英語学習者コーパスにおける品詞及び語彙の意味カテゴリーの使用傾向について調査・分析する。また、学習者産出文とデータ収集のためのタスクにおけるモデル作文との比較分析により、初級・中級レベルの英語への*Wmatrix3*の適用とその問題点について検討する。

### 4.2 方法

#### (1) コーパスデータ：The JEFLL Corpus(高校生データ)

調査対象はJLCの内、投野由紀夫研究室(東京外国語大学)版の高校3学年分のサブコーパスデータ(延べ296,370語)である。本研究では日本語多く含む中学生の作文は対象としない。

(2) コーパスツールと機能: *Wmatrix3*

*Wmatrix3*を使って、1) 単語のレマ化、アノテーション (① CLAWS7による品詞標識自動付与、② USAS 意味標識自動付与) と、2) 単語単位及び多重単語単位 (MWU) の頻度のリスト化、3) Keyness 分析による BNCsampler との比較統計を行う。

## (3) 分析の観点と手順

5つの観点 (語彙項目、品詞、意味領域全般、感情語彙の分類、形容詞コロケーション) から以下の①から⑦の手順で、対象データを分析する。

手 順

- ① JLC 内の高校生データ (296,370語) のみを *Wmatrix3* の特定フォルダーに格納
- ② 語彙項目 (単語単位・MWU) リストの出力
- ③ 品詞標識自動付与とリスト化
- ④ 意味領域の分析のための意味標識自動付与とリスト化
- ⑤ 感情語彙分析のための感情領域 (E 領域) の意味標識付与済データのリスト化
- ⑥ 感情語彙を含む KWIC コンコーダンスの出力
- ⑦ 形容詞コロケーションの抽出

なお、上記⑤、⑥については E(EMOTIONAL ACTIONS, STATES & PROCESSES) が対象となる。E 領域の下位分類には、例えば、E2 (Liking) があり、この領域に含まれる語彙の心的概念は、fondness, affection, partiality, attachment とその欠如である。E4とE5の意味構造を見ると、表1のような階層構造になっている。英字 (E) に枝番号が付く標識 (E4.1) と付かない標識 (E5)、正負記号を付け、肯定的な感情、否定的な感情を分けている場合がある。

表1 意味標識の構造 (例): E4, E5(emotion)

正負記号付	E4.1+	E4.1-	E4.2+	E4.2-	E5+	E5-
小分類	E4.1		E4.2		E5	
大分類	E4					
意味領域	E (emotion)					

詳細は Archer,D., Wilson, A., Rayson, P. (2002)

## 4.3 結果

## 4.3.1 語彙項目の分析

## (1) 単語単位

単語単位の分布状況 (図1) をみると、単語タイプに関する顕著な結果は、59% が頻度1の語であるという点である。つまり対象コーパス全体で、たった1回しか使われていない単語が半分以上を占めていることになる。そして頻度1から頻度3までを累計すると79%にも及ぶ。一方、頻度11以上の語が10%を占めている。すなわち、多くの高校生が使う語で10%程度構成するということであろう。単語トークンで見ると頻度11以上の語が91%を占めており、頻度1から頻度10までの語は残りの9%である。極少数の書き手により様々な語彙を使用する一方で、多くの書き手は狭い範

囲の語彙を使用していると言える。

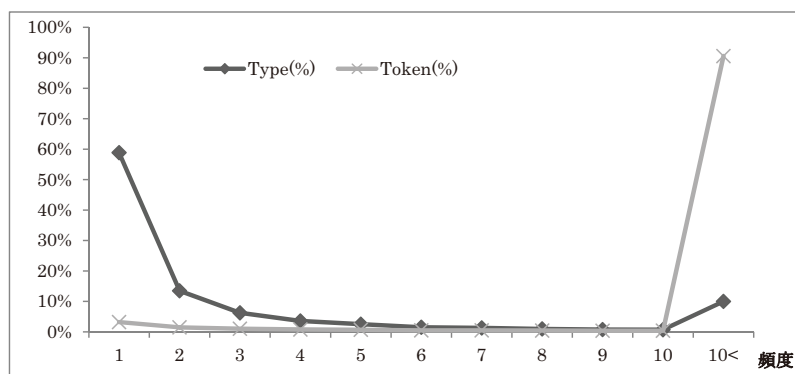


図1 単語単位の頻度分布

そこで、実際にどのような単語がよく使われるのかを知るために、*Wmatrix3*における *BNCsampler* の話し言葉<sup>7)</sup>サブコーパス (*BNC sampler spoken*) の比較機能を使って、特徴語を統計値 (対数尤度比) と共に出力してみた (表2)。これを *JLC* におけるモデル作文の使用語彙 (表3) と比較すると、*my*, *I*, *very*, *but*, *want*, *people*, *money* 等の語が一致していることがわかる。特に、*very* ( $LL=2,745.44$ ) はモデル文で、頻度9となっており、明らかにモデル文の影響である。(但し、すべての学年で同程度の影響があるのかという点については明らかではない。) また、*my*, *I*, *our* 等は対数尤度比が高く、これらの語は書くトピックから必然的に多用される語であろう。

表2 *JLC* の特徴語 : *BNC sampler spoken* との比較

語 彙	<i>JLC</i> 高校生		<i>BNC sampler spoken</i>		対数尤度比 ( <i>LL</i> )
	粗頻度	相対頻度	粗頻度	相対頻度	
my	5,132	1.73	2,354	0.24	6,928.04
i	19,676	6.64	31,907	3.25	5,782.93
our	2,420	0.82	1,271	0.13	2,994.39
very	2,944	0.99	2,201	0.22	2,745.44
money	1,252	0.42	643	0.07	1,572.74
was	4,725	1.59	8,052	0.82	1,227.05
is	5,446	1.84	9,985	1.02	1,154.17
because	1,667	0.56	1,920	0.20	932.65
but	3,624	1.22	6,561	0.67	797.06
so	3,143	1.06	5,570	0.57	734.53
to	7,279	2.46	16,611	1.69	670.32
want	1,346	0.45	1,788	0.18	596.95
will	1,407	0.47	1,993	0.20	553.62
he	3,645	1.23	7,890	0.80	428.03
had	1,476	0.50	2,554	0.26	368.19
than	579	0.20	730	0.07	280.96
when	1,524	0.51	2,939	0.30	275.82
time	727	0.25	1,090	0.11	254.89



take	596	0.20	821	0.08	247.35
make	502	0.17	645	0.07	235.95
went	498	0.17	718	0.07	189.23
have	2,739	0.92	6,624	0.67	184.26
me	1,336	0.45	2,861	0.29	164.18
people	999	0.34	2,001	0.20	158.81
by	716	0.24	1,293	0.13	158.63
first	454	0.15	689	0.07	155.10
his	688	0.23	1,243	0.13	152.23
some	886	0.30	1,934	0.20	100.33
ca	681	0.23	1,417	0.14	93.95
for	2,289	0.77	6,028	0.61	84.97
in	3,779	1.28	10,563	1.07	78.99
much	391	0.13	769	0.08	66.25
things	488	0.16	1,038	0.11	61.67
other	490	0.17	1,105	0.11	47.81
thing	480	0.16	1,090	0.11	45.29
could	807	0.27	2,030	0.21	42.26
must	313	0.11	674	0.07	37.58
more	532	0.18	1,319	0.13	30.68
only	482	0.16	1,191	0.12	28.41
like	1,337	0.45	3,743	0.38	27.46
good	627	0.21	1,678	0.17	20.27

表3 モデル文の語彙集計

頻度5以上 (頻度)		頻度4～3	頻度2	頻度1		
i	(30)	about	always	after	happy	said
my	(11)	book	around	ago	help	save
and	(10)	bought	be	am	home	saw
it	(10)	go	big	anyone	honey	school
t	(10)	his	bread	anything	house	sell
a	(9)	month(s)	bring	are	how	ski
very	(9)	on	class(es)	bad	hundred	slim
was	(9)	one	do	beautiful	hungry	sometimes
to	(8)	skiing	dream	became	interesting	started
don	(7)	there	family	better	last	story
the	(7)	with	fat	boys	lived	summer
but	(6)	have	father	breakfast	lose	tells
he	(6)	in	first	butter	lot	ten
snake	(6)	is	fish	buy	love	them
me	(5)	like	fishing	by	makes	then
		want	girl	caught	married	they
		will	many	cm	milk	thought
			morning	come	money	trip
			rice	couldn	much	up
			say	day	new	use
			shocked	didn	next	usually
			so	difficult	of	wanted
			some	every	or	wasn
			that	everybody	own	way
			were	feel	pay	wear
				finished	people	well

		friends	remembered	went
		from	rent	winter
		give	rich	write
		gloves	road	writing
		happily	sad	year

## (2) 多重単語単位

多重単語単位 (MWU) の分布状況をみると、タイプの顕著な結果は、57%が頻度1の語であるという点である。つまり対象コーパス全体で、1回しか使われていない多重単語単位が半分以上を占めていることになる。そして頻度1から頻度3までを累計すると80%にも及ぶ。一方、頻度11以上の語は8%を占めている。すなわち、多くの高校生が使う多重単語が8%程度を構成しているということである。トークンで見ると頻度11以上の語が67%を占めており、頻度1から頻度10までの語は33%である。

この結果を上述の単語単位の結果と比較すると、図1と図2の曲線のプロファイルは酷似しており、単語単位の値の方が顕著である。

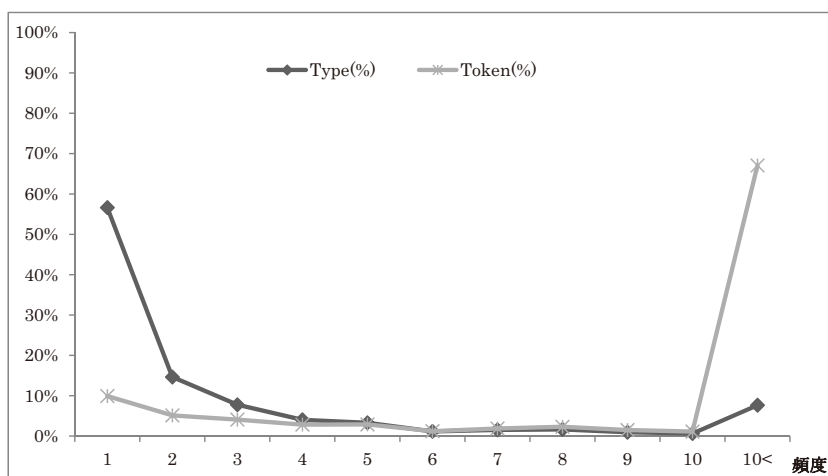


図2 多重単語単位の頻度分布

多重単語単位の具体的な項目のリスト (表4) を見ると、モデル文で使用されている多重単語単位で、コーパスに高頻度に出現する項目 (例 a lot, very much, have to) は、多くはない。しかし、作文トピックに密接に関連すると思われる多重単語は多い。例えば、take out(地震), every morning(朝食), high school(文化祭), bring out(地震), every day(朝食), get up(朝食) 等の多重単語は、括弧内に示したトピックに密接に関連する項目であり、高頻度であることは当然であろう。しかし、一般的な多重単語も高頻度に使われているのでトピックやモデル作文の影響を強く受けていると断定することは難しい。例えば、I thinkのようなフレーズは、自分の意見を述べるタイプの作文課題は、いかなるトピックでも頻出し得る多重単語単位である。こ



れ以外にも同様な定型的な多重単語単位（例 every year, next year, came to, last year, one day, had to, going to, long time, such a）の例は少ない。

表4 多重単語単位の頻度

MWU	粗頻度	相対頻度	MWU	粗頻度	相対頻度
a lot	487	0.16	woke up	96	0.03
take out	473	0.16	wake up	93	0.03
very much	337	0.11	long time	90	0.03
I think	298	0.10	such a	88	0.03
have to	280	0.09	come to	85	0.03
Urashima Taro	204	0.07	by the way	81	0.03
every year	188	0.06	miso soup	73	0.02
every morning	187	0.06	summer vacation	70	0.02
high school	182	0.06	new year	67	0.02
next year	162	0.05	so many	66	0.02
came to	160	0.05	in fact	62	0.02
last year	138	0.05	run away	61	0.02
one day	131	0.04	at that time	60	0.02
get up	131	0.04	make it	55	0.02
had to	119	0.04	mobile phone	51	0.02
bring out	116	0.04	take part	49	0.02
every day	101	0.03	in dream	49	0.02
going to	100	0.03	came back	48	0.02

#### 4.3.2 品詞分析：品詞タグ別構成率

品詞分析の結果を概観すると、品詞標識が付与された語の総数は296,370語で、計137種類の品詞標識が付与されている。内訳は表5の通り、頻度10以上の標識が119種類で、約87%を占める。なお、頻度1の標識が5種類（3.65%）あるが、標識付与のエラーである。

表5 頻度別品詞構成

頻度	タイプ	%	頻度	タイプ	%
1	5	3.65%	6	0	0.00%
2	2	1.46%	7	1	0.73%
3	4	2.92%	8	1	0.73%
4	2	1.46%	9	1	0.73%
5	2	1.46%	10	0	0.00%
			10 <	119	86.86%

そこで、主な品詞標識を詳細に見ることにする。まず、最頻度品詞は相対頻度16(47,763)を占めるNN1である。第2位のPPIS1以降は大きく下がり、1ケタ台の標識が26種類続く。28位以降は1.00未満と、小数点以下の値になる。紙幅の都合により、付録に頻度20以上の結果のみを掲げる<sup>8)</sup>。なお、動詞類(VV等)の品詞標識が27種類で、名詞類(NN等)の13に比べ、種類が多いため数値が分散する。反対に形容詞類(JJ等)は5種類と少なく、JJ(原形)を除けば極めて低頻度、低順位である。

表6 JLCにおける出現語彙の品詞別頻度

品 詞	相対 頻度	粗頻度	品 詞	相対 頻度	粗頻度
NN1(単数名詞)	16.12	47,763	VBZ( <i>is</i> )	2.05	6,077
PPIS1(人称代名詞 <i>I</i> )	6.61	19,604	VM(法助動詞)	2.00	5,925
JJ(形容詞原形)	5.06	14,985	CS(従位接続詞)	1.97	5,845
II(前置詞、除く <i>of</i> )	3.88	11,511	PPH1( <i>it</i> )	1.73	5,130
VV0(一般動詞原形)	3.84	11,393	TO(to 付不定詞)	1.72	5,086
VVI(一般動詞不定形)	3.70	10,960	VBDZ( <i>was</i> )	1.60	4,754
VVD(一般動詞過去)	3.29	9,764	XX(否定辞)	1.59	4,701
NN2(複数名詞)	3.27	9,684	RG(程度副詞、 <i>very</i> 等)	1.46	4,334
AT(定冠詞等、 <i>the, no</i> )	3.00	8,898	PPHS1(人称代名詞 <i>he, she</i> )	1.38	4,092
CC(等位接続詞、 <i>and, or</i> )	2.97	8,812	CCB(等位接続詞 <i>but</i> )	1.22	3,609
APPGE(代名詞所有格)	2.97	8,806	IO(前置詞 <i>of</i> )	1.21	3,580
RR(副詞)	2.74	8,124	PPIS2(人称代名詞 <i>we</i> )	1.06	3,139
AT1(冠詞等、 <i>a(n), every</i> )	2.11	6,240	NNT1(曜、日、年等単数)	1.03	3,051

6種類全てのJLC(表6)とモデル作文(表7)との品詞の頻度と順位を比較すると、僅かな相対頻度の違いはあるが、一部の例外PPI01(*me*), VD0(*do*), MD(序数)を除き、30位あたりまでは品詞構成は類似している。但し、モデル作文が品詞にどのような影響を与えるか明らかではない。

表7 モデル作文における品詞別頻度

品詞標識	相対頻度	粗頻度	品詞標識	相対頻度	粗頻度
NN1	10.32	36	VH0	0.86	3
PPIS1	8.31	29	IW	0.86	3
VV0	6.30	22	MC1	0.86	3
II	4.30	15	PN1	0.86	3
JJ	4.30	15	EX	0.86	3
VVD	4.01	14	NP1	0.86	3
VVI	3.72	13	VVZ	0.57	2
APPGE	3.72	13	VBI	0.57	2
ZZ1	3.44	12	RT	0.57	2
CC	3.15	11	DD	0.57	2
FO	3.15	11	NN	0.57	2
PPH1	2.87	10	VBDR	0.57	2
RG	2.87	10	DA2	0.57	2
AT1	2.87	10	RRR	0.29	1
VBDZ	2.58	9	VBM	0.29	1
NN2	2.01	7	CST	0.29	1
RR	1.72	6	RA	0.29	1
CCB	1.72	6	RRQ	0.29	1
TO	1.72	6	NNU	0.29	1
MC	1.72	6	VHI	0.29	1
AT	1.43	5	PPHO2	0.29	1
PPI01	1.43	5	IO	0.29	1
VD0	1.43	5	PPHS2	0.29	1
VM	1.43	5	VBR	0.29	1
PPHS1	1.43	5	RL	0.29	1
NNT1	1.15	4	DD1	0.29	1
VBZ	1.15	4	VVG	0.29	1
XX	1.15	4	NNT2	0.29	1

### 4.3.3 意味分析

意味領域の分析結果について述べる。まず、意味標識の出力結果を概観してみよう。図3を見ると明らかなように、50%弱を占めるのが NAMES & GRAMMATICAL WORDS(Z 領域)である。この領域は以下のような12種の標識で構成されている。Unmatched proper noun(Z0), Personal names(Z1), Geographical names(Z2), Other proper names(Z3), Discourse Bin(Z4), Grammatical bin(Z5), Negative(Z6), If(Z7), Unconditional(Z7), Pronouns(Z8), Trash can(Z9), Unmatched(Z99)。

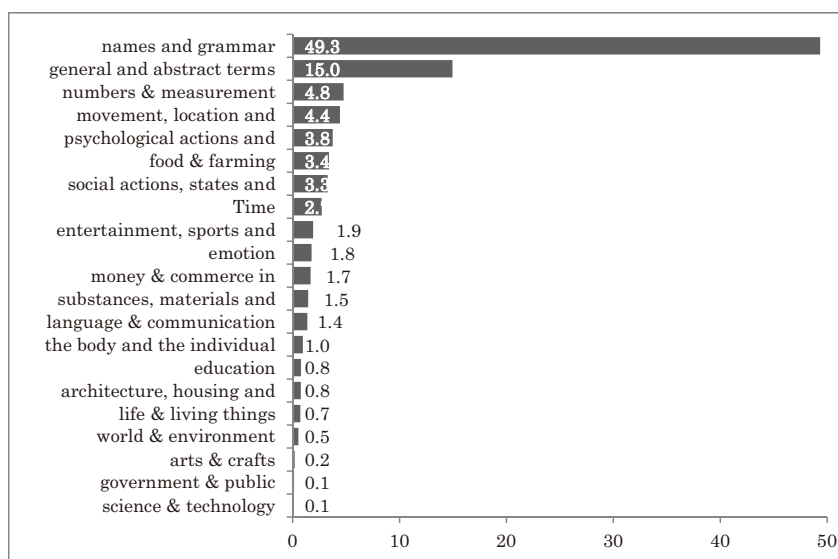


図3 意味領域別構成率 (%)

Z 領域の多くは意味領域とは言い難く、closed class と呼ばれる品詞類と名詞で、具体的には前置詞や限定詞、人名等の固有名詞である。特に Z0 は Unmatched proper noun で、JLC では、その性格上、日本人の名前が高頻度で、Z2 の地名も同様となる。Z4 は談話標識等である。Z5 は文法語（前置詞、副詞、接続詞等）、Z6 は主に、否定辞類 (neither, no, non, not, not at all, not really,) である。Z7 は条件節を形成する語句 (例 if, providing, as long as, even if, as, whether or not) や Z8 は代名詞 (例 he, hers, it, its, itself, my, one, ones, anything at all, this that and the other) である。誤綴りを含む判定不能語は Z9 と Z99 が充てられる。全体の 15.0% に当たる A 領域は、広範囲の意味領域をカバーする。詳細な集計結果について主な標識の相対頻度 (%) を見ると、Z5 (22.4) が突出する。以下に主な標識の相対頻度を括弧内に示す。

Z5 (22.4) 文法語 (冠詞、前置詞、接続詞、一部の動詞が含まれる)

Z8 (17.2) 代名詞 (関係代名詞が含まれる)  
 A3+ (4.4) 存在 (存在を表す一般語・抽象語)  
 F1 (3.0) 食品・食事 (仕度を含む)  
 A1.1.1, (1.2) 一般的行為・行動 (行動・行為を表す一般語・抽象語)  
 A1.5.1 (0.2) 使用・不使用を表す一般語・抽象語  
 A5.1+ (0.4) 評価: 良  
 A6.1- (0.29) 比較: 類似・相違  
 A9 + (1.9) 所有 (授受を含む)  
 Z6 (1.8) 否定  
 P1 (1.8) 教育一般  
 A13.3 (1.3) 程度: 強調詞  
 T1.3 (1.2) 時間・期間

表8 意味標識別頻度 (頻度順)

意味 標識	意味領域	相対 頻度	粗頻度	語彙項目
Z5	Grammatical bin	22.41	66,425	<i>the, and, to, a, in</i>
Z8	Pronouns	17.18	50,914	<i>I, it, my, he, we</i>
Z99	Unmatched	6.31	18,707	<i>Otoshidama, err</i>
A3+	Existing	4.45	13,177	<i>is, was, are, be, were</i>
F1	Food	3.05	9,032	<i>eat, rice, breakfast, bread, food</i>
A9+	Getting&possession	1.90	5,630	<i>have, had, take, has, get</i>
Z6	Negative	1.84	5,468	<i>n't, not, no</i>
P1	Education	1.78	5,272	<i>school, class, students</i>
A13.3	Degree: Boosters	1.49	4,411	<i>very, so, very_ much, more</i>
T1.3	Time: Period	1.23	3,651	<i>year, morning, day</i>
A1.1.1	General actions/ making	1.22	3,620	<i>made, do, make</i>
A7+	Likely	1.21	3,599	<i>can, could, ca(n't), would</i>
M1	Moving, coming and going	1.21	3,573	<i>go, went</i>
K1	Entertainment generally	1.11	3,286	<i>festival</i>
M6	Location and direction	0.82	2,435	<i>this</i>
N5+	Quantities: many/ much	0.76	2,251	<i>many, a_lot, much</i>
X2.1	Thought, belief	0.74	2,181	<i>think, thought</i>
N4	Linear order	0.71	2,105	<i>then, first</i>
E2+	Like	0.69	2,053	<i>popular, love, enjoy, like</i>
X7+	Wanted	0.65	1,932	<i>want</i>
N1	Numbers	0.64	1,901	<i>one, two</i>
S4	Kin	0.63	1,879	<i>family, mother</i>
M2	Putting, pulling, pushing, transporting	0.63	1,875	<i>bring, take_out</i>
O2	Objects generally	0.62	1,837	<i>thing, things</i>
S2	People	0.60	1,779	<i>people, children</i>
I1	Money generally	0.58	1,705	<i>money</i>
I2.2	Business: Selling	0.55	1,636	<i>buy</i>

B1	Anatomy and physiology	0.50	1,469	<i>face, sleeping, wake_up, body, tired</i>
Z4	Discourse Bin	0.49	1,467	<i>i_think</i>
Z7	If	0.49	1,449	<i>If</i>
A2.1+	Change	0.47	1,393	<i>became</i>
S6+	Strong obligation or necessity	0.47	1,383	<i>must, need</i>
Q2.1	Speech:	0.43	1,271	<i>said</i>
	Communicative			
N5	Quantities	0.42	1,251	<i>some</i>

●ケーススタディ 1: JLC における感情語彙: EMOTIONAL ACTIONS, STATES & PROCESSES

JLC における感情語彙について分析した。*Wmatrix3*では、感情語彙は E 標識が付与される。コンコーダンス機能により以下のようなコンコーダンスラインを出力し、E 領域標識の集計結果をリスト化した(表9)。

E1

because\_Z5/A2.2 having\_A9+ foods\_F1 in\_Z5 morning\_T1.3 make\_A1.1.1 me\_Z8mf **feel\_E1** bad\_A5.1- .\_PUNC

E2+

I\_Z8mf **like\_E2+** to\_Z5 eat\_F1/B1 but\_Z5 I\_Z8mf ca\_A7+ n't\_Z6 eat\_F1/B1 anything\_Z8 every\_N6+[i131.2.1 morning\_N6+[i131.2.2 .\_PUNC

E2+, E3+

I\_Z8mf **love\_E2+** nature\_W5 ,\_PUNC peace\_E3+ ,\_PUNC and\_Z5 earth\_W3 .\_PUNC

E4.1+

Mail\_Q1.2 is\_A3+ very\_A13.3 **fun\_E4.1+** .\_PUNC

If\_Z7 <jp>\_PUNC okane\_Z99 ga\_Z3c </jp>\_PUNC <jp>\_PUNC areba\_Z99 </jp>\_PUNC very\_A13.3 **happy\_E4.1+** !\_PUNC

E5

Urashima\_Z1mf[i38.2.1 Taro\_Z1mf[i38.2.2 was\_A3+ very\_A4.2+ **shock\_E5-** .\_PUNC

表9 感情語彙の小分類別頻度

感情領域	感情語彙標識	相対頻度
E1 General: emotional actions, states and processes	E1	0.016
E2 Liking: fondness/affection/partiality/attachment, or the lack of	E2	0.001
	E2-	0.026
	E2+	0.693
	E2++	0.046
	E2+++	0.035
E3 Calm/Violent/Angry(): (level of) serenity/composure/anger/violence	E3-	0.073
	E3-	0.001
	E3+	0.050

E4.1 Happy/sad: Happy : (level of) happiness	E4.1-	0.167
	E4.1+	0.379
	E4.1++	0.001
	E4.1+++	0.003
E4.2 Happy/sad: Contentment : (level of) contentment	E4.2-	0.013
	E4.2+	0.066
E5 Fear/bravery/shock : (level of) trepidation/courage/surprise, etc	E5-	0.140
	E5+	0.004
E6 Worry, concern, confident : (level of) apprehension/confidence...	E6-	0.057
	E6+	0.005

感情語彙について BNC sampler spoken と比較し、対数尤度比を出力し、顕著な例を肯定的感情語彙、否定的感情語彙、それぞれについて示す(表10)。肯定的感情群では E2+( $LL=1230.2$ ) と E4.1+( $LL = 1117.0$ ) が最も顕著で、否定的感情語彙群では、E4.1-( $LL = 429.6$ )、E5-( $LL=295.2$ ) が目立つ。対立する2つの感情語彙の頻度と順位の関係は特筆すべき点である。肯定的感情語彙は項目数が多いが、否定的感情は相対的に少ない。しかも、後者は第1位でも頻度は低い (*sad*: 0.07、*shocked*: 0.04), 肯定的感情語彙の第1位は5倍以上の頻度 (*like* 0.36, *happy*: 0.17) である。

表 10 感情語彙の過剰使用: BNC sampler spoken との比較 \*相対頻度 0.01 以上のみ掲載

<u>E2+ (Like)</u>	相対頻度	粗頻度			
<i>Like</i>	0.36	1,057			
<i>Enjoyed</i>	0.09	260			
<i>Enjoy</i>	0.08	247			
<i>Love</i>	0.06	182			
<i>Popular</i>	0.02	62			
<i>Liked</i>	0.01	36			
<i>Precious</i>	0.01	35			
<i>Likes</i>	0.01	30			
<i>Enjoying</i>	0.01	25			
<i>Loved</i>	0.01	19			
<i>live_with</i>	0.01	15			
<u>E4.1+ (happy)</u>			<u>E4.1- (sad)</u>	相対頻度	粗頻度
<i>Happy</i>	0.17	498	<i>Sad</i>	0.07	197
<i>Fun</i>	0.08	230	<i>Cried</i>	0.04	111
<i>Funny</i>	0.02	55	<i>Crying</i>	0.01	40
<i>Happily</i>	0.02	48	<i>Cry</i>	0.01	33
<i>Comics</i>	0.01	31	<i>Unhappy</i>	0.01	18
<i>Enjoyable</i>	0.01	28	<i>Regret</i>	0.01	16
<i>Smile</i>	0.01	23			
<i>had_a_good_time</i>	0.01	18			
<i>Happiness</i>	0.01	17			
<i>Laughed</i>	0.01	16			
<u>E5- (fear/shock)</u>					
<i>Shocked</i>	0.04	114			
<i>Afraid</i>	0.02	70			
<i>Scared</i>	0.01	27			
<i>Fear</i>	0.01	27			
<i>Scary</i>	0.01	23			
<i>Panic</i>	0.01	23			
<i>Horror</i>	0.01	21			



<i>Frightened</i>	0.01	20
<i>Fearful</i>	0.01	15

最後に、品詞標識と意味標識の対応関係を明らかにするために語彙を軸にして、それぞれの頻度と標識を比較した(表11)。これにより1対1の対応関係にある標識とそうでない標識があること分かる。

表11 モデル作文使用されている語の品詞と意味標識の対応(一部抜粋)

語彙	品詞 標識	頻 度	意味 標識	頻 度	語彙	品詞 標識	頻 度	意味 標識	頻 度
<i>a</i>	AT1	8	Z5	8	<i>they</i>	PPHS2	1	Z8	1
<i>many</i>	DA2	2	N5+	2	<i>very</i>	RG	8	A13.3	8
(日本語)	FO	1	Z99	1	<i>home</i>	RL	1	H4	1
<i>about</i>	II	2	A13.4	1	<i>well</i>	RR	1	A5.1+	1
<i>by</i>	II	1	Z5	1	<i>usually</i>	RR	1	A6.2+	1
<i>to</i>	II	1	Z5	7	<i>always</i>	RR	2	N6+++	2
<i>with</i>	IW	3	Z5	3	<i>how</i>	RRQ	1	Z5	1
<i>difficult</i>	JJ	1	A12-	1	<i>better</i>	RRR	1	A5.1++	1
<i>bad</i>	JJ	1	A5.1-	1	<i>sometimes</i>	RT	1	N6	1
<i>sad</i>	JJ	1	E4.1-	1	<i>to</i>	TO	6	Z5	7
<i>shocked</i>	JJ	2	E5-	2	<i>was</i>	VBDZ	9	A3+	9
<i>hungry</i>	JJ	1	F1-	1	<i>be</i>	VBI	2	A3+	2
<i>rich</i>	JJ	1	I1.1+	1	<i>is</i>	VBZ	4	A3+	4
<i>fat</i>	JJ	1	N3.2+	1	<i>do</i>	VD0	5	Z5	3
<i>slim</i>	JJ	1	N3.5-	1	<i>have</i>	VH0	2	A9+	3
<i>beautiful</i>	JJ	1	O4.2+	1	<i>have</i>	VHI	1	A9+	3
<i>interesting</i>	JJ	1	X5.2+	1	<i>will</i>	VM	4	M45.1.3	4
<i>fish</i>	NN	1	K5.1	1	<i>could</i>	VM	1	A7+	1
<i>people</i>	NN	1	S2	1	<i>wear</i>	VV0	1	B5	1
<i>bread</i>	NN1	2	F1	2	<i>love</i>	VV0	1	E2+	1
<i>house</i>	NN1	1	H1	1	<i>save</i>	VV0	1	I1.1	1
<i>money</i>	NN1	1	I1	1	<i>sell</i>	VV0	1	I2.2	1
<i>skiing</i>	NN1	3	K5.1	3	<i>go</i>	VV0	2	M1	3
<i>snake</i>	NN1	2	L2	2	<i>say</i>	VV0	1	Q2.1	2
<i>trip</i>	NN1	1	M1	1	<i>want</i>	VV0	3	X7+	4
<i>fat</i>	NN1	1	O1	1	<i>wasn</i>	VV0	1	Z99	1
<i>class</i>	NN1	1	P1	1	<i>became</i>	VVD	1	A2.1+	1
<i>school</i>	NN1	1	P1	1	<i>bought</i>	VVD	3	I2.2	3
<i>story</i>	NN1	1	Q2.1	1	<i>said</i>	VVD	1	Q2.1	1
<i>book</i>	NN1	3	Q4.1	3	<i>married</i>	VVD	1	S4	1
<i>girl</i>	NN1	1	S2.1	1	<i>finished</i>	VVD	1	T2-	1
<i>father</i>	NN1	2	S4	2	<i>thought</i>	VVD	1	X2.1	1
<i>dream</i>	NN1	2	X4.1	2	<i>remembered</i>	VVD	1	X2.2+	1
<i>otoshidama</i>	NN1	1	Z99	1	<i>saw</i>	VVD	1	X3.4	1
<i>gloves</i>	NN2	1	B5	1	<i>wanted</i>	VVD	1	X7+	1
<i>boys</i>	NN2	1	S2.2	1	<i>writing</i>	VVG	1	Q1.2	1
<i>friends</i>	NN2	1	S3.1	1	<i>use</i>	VVI	1	A15.1	1
<i>winter</i>	NNT1	1	T1.3	1	<i>pay</i>	VVI	1	I1.2	1
<i>year</i>	NNT1	1	T1.3	1	<i>bring</i>	VVI	2	M2	2
<i>everybody</i>	PN1	1	Z8	1	<i>write</i>	VVI	1	Q1.2	1
<i>them</i>	PPHO2	1	Z8	1	<i>say</i>	VVI	1	Q2.1	2

<i>want</i>	VVI	1	X7+	4
<i>lose</i>	VVI	1	X9.2-	1
<i>tells</i>	VVZ	1	Q2.2	1
<i>n't</i>	XX	4	Z6	4

#### 4.3.4 形容詞分析

##### ●ケーススタディ 2：形容詞を含む2語連鎖

本節では、代表的な形容詞に隣接する共起語を抽出し、コロケーション分析を行った結果について報告する。まず、*Wmatrix3*のコロケーション機能を用いて、対数尤度比及び *t* 値に基づく2語のコロケーションを出力すると、複合名詞 (school festival)、語彙コロケーション (eat breakfast)、隣接または近接する文法構造 (I n't,

I want, n't much, fire earthquake) が抽出できる。これによりある程度の広いスパンで共起関係にある語が統計的にわかる。しかし、2語のみで意味関係のまとまりを見出すことが難しいケースが多い。意味関係が容易に判断できる2語の組み合わせは多くのケースで隣接していることを目視で確認した。

そこで形容詞を含む2語で構成される名詞句、形容詞コロケーション、を明らかにするために AntConc3.2.3 に同じコーパスデータを読み込ませ、cluster 機能を用いて *t* 値に基づき有意な共起語 (*t*>2.0) をリスト化した。表12では、一例としてその中から6語を取り上げ、共起名詞を肯定・否定的形容詞に分け、整理した。表中の CO は共起頻度 (co-occurrence) を、下線を施した語は単数複数両方の語がリストされている語を表す。

表 12 形容詞を含む名詞句

+				-			
中心語 (意味標識/ 頻度)	共起名詞	t値	CO	中心語 (意味標識/ 頻度)	共起名詞	t値	CO
GOOD (A5.1+/643)	<i>time</i>	5.80	(38)	BAD	<i>dream</i>	14.21	(204)
	<i>memory</i>	4.75	(23)	(A5.1-/390)	<i>dreams</i>	4.77	(23)
	<i>taste</i>	3.96	(16)		<i>man</i>	2.46	(7)
	<i>dream</i>	3.58	(16)	WORST (A5.1-/23)	<i>dream</i>	3.59	(13)
	<i>idea</i>	3.43	(12)				
	<i>thing</i>	3.33	(13)				
	<i>memories</i>	3.26	(11)				
	<i>movie</i>	2.61	(8)				
	<i>things</i>	2.46	(8)				
	<i>point</i>	2.21	(5)				
	<i>job</i>	2.19	(5)				
	<i>experience</i>	2.18	(5)				
	<i>dreams</i>	2.15	(5)				
	<i>friends</i>	2.11	(6)				
	<i>thing</i>	3.37	(12)				
	<i>memory</i>	2.81	(8)				
BEST (A5.1/174)	<i>friend</i>	2.16	(5)				
	<i>earthquake</i>	16.94	(289)				
	<i>brother</i>	2.40	(6)				
	<i>event</i>	2.38	(6)				
	<i>problem</i>	2.20	(5)				
	<i>sister</i>	2.18	(5)				
OLD (T3/501)	<i>man</i>	13.22	(176)	YOUNG	<i>man</i>	7.32	(54)
	<i>woman</i>	4.22	(18)	(T3-/193)	<i>boy</i>	2.22	(5)
	<i>books</i>	3.07	(10)	NEW	<i>year</i>	10.34	(109)
	<i>people</i>	2.68	(10)	(T3-/134)	<i>years</i>	5.23	(28)
	<i>days</i>	2.43	(7)		<i>house</i>	2.43	(7)
					<i>cd</i>	2.42	(6)
					<i>life</i>	2.31	(6)
					<i>shoes</i>	2.21	(5)
					<i>clothes</i>	2.17	(5)

形容詞コロケーションに使われる形容詞の中で最も顕著な語は肯定的形容詞で意味標識 A5.1+ が付与される good(頻度 643)/best(同 174) で、t 値が 2.0 を上回る該当のコロケーションでは、共起名詞は抽象名詞が多く、バリエーションは最も多い。但し、

原級の good(643) の共起名詞の内、memory, dream, thing は単数形、複数形が含まれ、共通しており、しかも最上級の best に対する thing, memory, friend は good と同じであるので、レマ形にすると good のコロケーションは12種類になる。また、その半数以上は共起頻度は粗頻度で10以下に過ぎない。一方、否定的形容詞で、A5.1 – の意味標識が付与される bad(390)/worse(23) は共起名詞のバリエーションは極めて少ない (dream(s), man)。しかも、作文トピックの影響が表れている。

good 以外の形容詞コロケーションも、1、2位以下は共起頻度は粗頻度10以下と低頻度である。形容詞コロケーションのレパトリが極めて限定的であることが分かる。

## 5. 結論

### 5.1 本研究の総括と課題

本研究では、*Wmatrix3*を用いて、初級・中級レベルの日本人英語学習者コーパスにおける品詞及び語彙の意味カテゴリーの使用傾向の検証を試みた。主な結果を整理し、考察を述べる。まず、語彙分析の結果からは、単語単位、多重単語単位のいずれも同様の使用傾向があることが分かった。タイプ、トークンそれぞれを見ると語彙は個人によりばらつく傾向があり、使用語彙の偏りも観察された。JLC の語彙と BNC の話し言葉との違いを分析すると、JLC におけるタスク、作文のトピックから起因するもの、そして限定的ではあるが作文課題におけるモデル文の影響が示唆される。但し、およそ40パーセントを占める機能語や文法語についてはそうした影響は少ないと考えるのが妥当であろう。

学習者の英文とタスクにおけるモデル作文の比較分析により、初級・中級レベルの英語への *Wmatrix3* の適用とその問題点について検討することも本研究が目指すところであった。モデル作文との比較も踏まえた品詞分析の結果から、JLC では名詞は頻度が、動詞は種類が多い。*Wmatrix3* に搭載されている *CLAWS7* の品詞標識は130以上と種類が多い。一部修正をすることでより適した方法で JLC の品詞使用の特徴を記述することできるだろう。

次に、意味分析の結果からは、文法領域の語彙が半分を占め、意味分析の主対象は残りの語彙であることが分かった。意味領域は A 領域が最も構成率が高いが、A 領域は対象範囲が広いことに起因する。小分類による分析がよいだろう。特定の領域、例えば、F 領域、の語彙が中頻度に出現することから、作文トピック (例 朝食における F 領域の頻度) の影響が考えられる。トピック別にサブコーパス分析を行うことでより明らかになるだろう。

今後の課題と展望について最後に述べる。*Wmatrix3* の自動表記付与システムの適用性について、学習者コーパスの場合、精度については問題が少なくなく、完全自動化には至っておらず、手動での修正が不可欠であるといった課題が見えた。特に、初級、中級レベルの学習者の場合、文法、語彙、形態素等エラーが多く、品詞自動付与を実施した後に精度を上げる手作業に要する研究コストは膨大である。特に、JLC は、既述のように初・中級レベルの学習者産出データで、多くの文法、語彙のエラーを含

むことが想定していた。

データ処理過程でのエラーも散見されるため、品詞と意味標識付与の正確性の解釈については留意が必要である。いくつかの頻度の語彙項目についてはエラーが目視により観察され、また、品詞標識付与の出力結果では、unclassified(FU) と foreign word(FW) を合算しても 0.08%(200 件) のみとなっている。また、意味標識による分析でも unmatched(Z99 6%) は JLC のタスクやレベル等の特徴を考慮すると少なめである。こうした曖昧な領域の品詞付与は課題である。但し、*Wmatrix3* に日本語の単語を独自に搭載させることで一部は解決されるであろう。

最後に、今後の展望と教育的示唆について述べる。意味分析を精緻化するために、トピックによる意味標識の頻度の違いについて検証を進める。例えば、「地震」のトピックの作文ではマイナス記号が付与される語彙がどの程度占めるのか。「朝食」では、F 領域の標識が多いのか。といった、トピックの意味領域への影響の度合いを詳細に調査・分析したい。また、語彙発達の視点から学年による違いも今後の重要な課題である。本研究では、初級・中級学習者による英語コーパスへの品詞及び意味標識付与を試みた。上述のように精度については目視、手作業による確認の必要がある。

## 5.2 教育的示唆

本研究の結果から得られる教育的示唆については慎重に検討しなければならない。コミュニケーションな教育の流れのなかで意味交渉を中心とした学習を成功させるためには、意味に関する研究から示唆を得ることができる。筆者の知る限り本研究のようにコーパスに基づく初級中級レベルの学習者データに基づく語彙意味の検証は稀で、本研究は、その研究の初期段階である。

語彙習得・学習上の課題は、学習者にとっても、指導者にとっても少なくない。Nation(2007) は、語彙学習の実践的な課題 (learning burden) について、意味 (meaning)、形式 (form)、用法 (use) の 3 つを挙げ、学習者の母語を念頭に置いて以下の点に触れている。意味領域の課題については、目標語が借用語か、母語における意味との類似性、対応語における語のふるまいの類似、用法については、母語におけるコロケーションとの一致と語彙の選択制限の問題を指摘している (49 頁)。

本論のまとめとして、意味中心の語彙学習・指導の重要性について述べる。本研究の結果から示唆されるように日本人高校生の語彙、コロケーションのレパトリーは極めて貧弱で、語彙の意味理解も狭く、1 語 1 義と思われる用例が多くを占めている。これは語彙学習や指導方法に起因することが考えられる。Folse(土屋訳) (2009) は、語義の理解については訳語の効用を強調している。語彙の中核的な意味は訳語で与える方が学習効果は高かったという実証研究を諸例挙げている。少なくとも JLC における初級・中級レベルの学習者を対象とする語彙の学習・指導では、まず、新出語彙について、訳語で語義を理解させる。シラバス上の工夫で、再び同じ語彙が異なる語義で接触するようにし、新出語に準じる指導語彙として扱うような教材構成が望ましい。また、多義語についてはコロケーションや定型表現の一部として指導する方法が考えられる。とりわけ Lexical Approach (Lewis 1993, 1997 他) で提唱されているよ

うに、意味のかたまりでの語彙指導の重要性を指摘したい。こうして語彙の意味がマッピングできるような語彙指導が必要なのではないだろうか。そのためにも意味中心の語彙習得の検証が重要となるだろう。

本研究は科研補助研究(学振 研究課題番号25370704)における研究の一部である。

## 注

- 1) 同プロジェクトには筆者自身は2005年から参画しており、本研究では投野由紀夫研究室の所蔵する非公開版を使用させていただいた。
- 2) ICCI プロジェクトの詳細は Tono, Y., Kawaguchi, Y., and Minegishi, M., (Eds.). (2012) を参照されたい。
- 3) プロジェクト初期の収集、転写したデータについては、日本語部分はローマ字表記になっており、ここ数年は漢字かな混じりで表記しているため検索する場合は特に注意が必要である。
- 4) JLC のデータ収集で使用された作文タスク(タイプ・トピック) は以下の通りである。

### タイプ1: 論説文 自分の意見を論理的に書く

- ・朝食 (breakfast): 「朝食は、パンがいいか、ご飯がいいか。その理由は?」
- ・地震 (earthquake): 「大地震が来たら何を持って逃げるか。その理由は?」
- ・お年玉 (Otoshidama): 「お年玉をもらったら、何を買うか?」

### タイプ2: 叙述文 物語や経験を書く

- ・文化祭 (festival): 「あなたの学校の文化祭について書きなさい。」
- ・浦島太郎 (Urashima) 「浦島太郎のその後について、想像して書きなさい。」
- ・夢 (dream): 「今まで見た怖い夢について書きなさい。」

- 5) 付録1にモデル作文の一部について意味標識を付与済の例を示した。

- 6) 意味領域・大分類

A. general and abstract terms, B. the body and the individual, C. arts and crafts, E. emotion, F. food and farming, G. government and public, H. architecture, housing and the home, I. money and commerce in industry, K. entertainment, sports and games, L. life and living things, M. movement, location, travel and transport, N. numbers and measurement, O. substances, materials, objects and equipment, P. education, Q. language and communication, S. social actions, states and processes, T. time, W. world and environment, X. psychological actions, states and processes, Y. science and technology, Z. names and grammar

- 7) 従来の研究結果から、JLC のデータは BNC 全体に比べ、話し言葉サブコーパスとの比較の方が妥当性が高いことが分かっているため本研究では BNCsampler spoken と比較することとした。

## 参考文献

- Abe, M. and Tono, Y. (2005). Variations in L2 Spoken and Written English: Investigating Patterns of Grammatical Errors across Proficiency Levels. *Proceedings of the Corpus Linguistics Conference Series*, U.K., University of Birmingham pp.1-11.
- Archer, D., Wilson, A., Rayson, P. (2002). *Introduction to the USAS Category System*. <http://ucrel.lancs.ac.uk>.
- 藤原康弘.(2014).『国際英語としての「日本英語」のコーパス研究:日本の英語教育の目標』(シリーズ言語学と言語教育 31) 東京:ひつじ書房.
- Granger, S. (Ed.) (1998). *Learner English on Computer*. Harlow, England: Addison Wesley Longman.
- Granger, S., J. Hung, and Petch-Tyson, S. (Eds.) (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.



- Granger, S., Dangneaux, E., Meunier, F., and Paquot, M. (2009). *International Corpus of Learner English. Version 2 Handbook and CD-ROM*. Louvain-la-Neuve: Press iniversitaires de Louvain.
- 飯尾豊. (2013). 「コーパスを用いた日本人学習者の句動詞の使用に関する研究」熊本大学. 熊本大学社会文化研究 11: 35-53.
- Ishikawa, S. (Ed.) (2013). *Learner Corpus Studies in Asia and the World. School of Languages and Communication*, Kobe University.
- 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
- 小林雄一郎. (2009). 「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紀要』第23号: 11-21.
- Lewis, M. (1993). *The Lexical Approach: The state of ELT and the way forward*. Hove, England: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting theory into practice*. Hove, England: Language Teaching Publications.
- Nation, P. (2005). Teaching Vocabulary. *The Asian EFL Journal. Quarterly September* 7(3): 47-54. British Virgin Islands: The Asian EFL Journal Press.
- 能登原祥之. (2010). 「日本人英語学習者のイベントスキーマと文型への親密度: JEFLL Corpus の分析を通して」. 『英語コーパス研究』17: 33-48
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*. 13:4 519-549. DOI: 10.1075/ijcl.13.4.06ray
- Rayson, P. (2003). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. Ph.D. thesis, Lancaster University.
- 小学館コーパスネットワーク. (2007). *JEFLL Corpus*. [http://scn02.corpora.jp/~jefll03/jefll\\_top.html](http://scn02.corpora.jp/~jefll03/jefll_top.html). 2014年10月20日参照
- Stubbs, M. (2006) (南出康世・石川慎一郎監訳) 『コーパス語彙意味論: 語から句へ』. 東京: 研究社.
- Tono, Y. (1996). Using Learner Corpora for L2 Lexicography. *LEXIKOS 6*, Stellenbosch: Universiteit van Stellenbosch. pp. 116-132
- Tono, Y. (2000). *A corpus-based analysis of interlanguage development: Analysing part-of-speech tag sequences of EFL learner corpora*. Lewandowska-Tomaszczyk, B. and Melia, J.P. (eds.) *Proceedings of PALC'99: Practical Applications in Language Corpora*. Frankfurt am Main: Peter Lang, pp.323-340.
- Tono, Y. (2002). *The Role of Learner Corpora in Second Language Acquisition Research and Foreign Language Learning: The Multiple Comparison Approach*. PhD Dissertation. Lancaster University.
- Tono, Y & Aoki, M. (1998). *Developing the optimal learning list of irregular verbs based on the native and learner corpora*. *First International Symposium on Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching: Symposium Proceedings* (14-16 December, 1998. The Chinese University of Hong Kong), pp. 113-118.
- Tono, Y., & Kanatani, K. (1996). EFL Learners' Proficiency and Roles of Feedback: Towards the Most Appropriate Feedback for EFL Writing. *Annual Review of English Language Education in Japan* 6, pp.1-11.
- Tono, Y., Kawaguchi, Y., and Minegishi, M., (Eds.). (2012). *Developmental and Crosslinguistic Perspectives in Learner Research*. Amsterdam: John Benjamins.
- 投野由紀夫編著. (2007). 『日本人中高生一万人の英語コーパス: 中高生が書く英文の実態とその分析』東京: 小学館
- 投野由紀夫・金子朝子・杉浦正利・和泉絵美編著 (2013). 『英語学習者コーパス活用ハンドブック』. 東京: 大修館

Uchida, T.(2012). Use of Multiword Verbs by Nonadvanced EFL Learners: Focusing on Common Verb + Particle Combinations. *Working Papers in Corpus-based Linguistics and Language Education No. 8*. Tokyo university of foreign studies 303-323.

内田富男 .(2014a).「高校生はCEFR Level-Aの形容詞をどのように使うのか」. 関東甲信越英語教育学会千葉研究大会 ポスター発表

内田富男 .(2014b).「コーパスと英語教育語彙表における基本色彩語の考察：BNC, JEFLC Corpus, CEFR(-J)を用いて」 明星大学研究紀要人文学部 (50), 19-32.

Wmatrix 3.0 <http://ucrel.lancs.ac.uk/claws/trial.html> 2014年10月20日参照

#### 付録1 モデル文の意味標識自動付与(各文前半のみ抜粋)

朝食 I\_Z8mf usually\_A6.2+ have\_A9+ rice\_F1 and\_Z5 J0sTLVA1\_Z99 in\_Z5 the\_Z5 morning\_T1.3 .PUNC But\_Z5 I\_Z8mf don\_P1/S2mf t\_Z5 like\_Z5 it\_Z8 .PUNC I\_Z8mf like\_E2+ bread\_F1 and\_Z5 milk\_F2 better\_A5.1++ .PUNC Bread\_F1 with\_Z5 butter\_F1 and\_Z5 honey\_F1 is\_A3+ very\_A13.3 J0D0W0D0\_Z99 .PUNC Rice\_F1 makes\_A1.1.1 me\_Z8mf fat\_O1 .PUNC I\_Z8mf do\_Z5 n't\_Z6 want\_X7+ to\_Z5 be\_A3+ fat\_N3.2+ .PUNC

地震 I\_Z8mf will\_T1.1.3 bring\_M2 a\_Z5 new\_T3- 0V0J0\_Z99 first\_N4 .PUNC My\_Z8 father\_S4m bought\_I2.2 me\_Z8mf that\_Z5 0V0J0\_Z99 one\_N1 year\_T1.3 ago\_T1.1.1 .PUNC I\_Z8mf like\_E2+ fishing\_K5.1/L2 very\_A13.3[i2.2.1 much\_A13.3[i2.2.2 .PUNC I\_Z8mf go\_M1 fishing\_K5.1/L2 every\_N6+[i3.2.1 month\_N6+[i3.2.2 .PUNC My\_Z8 father\_S4m always\_N6+++ tells\_Q2.2 me\_Z8mf how\_Z5 to\_Z5 fish\_K5.1/L2 well\_A5.1+ .PUNC

お年玉 I\_Z8mf love\_E2+ skiing\_K5.1 .PUNC But\_Z5 I\_Z8mf do\_Z5 n't\_Z6 have\_A9+ 000\_N1 n0\_Z99 g\_Z5 .PUNC 0000\_N1 .PUNC gloves\_B5 or\_Z5 ski\_K5.1 wear\_B5 .PUNC I\_Z8mf always\_N6+++ go\_M1 skiing\_K5.1 in\_Z5 winter\_T1.3 .PUNC but\_Z5 I\_Z8mf have\_S6+[i5.2.1 to\_S6+[i5.2.2 rent\_I2.2 them\_Z8mf and\_Z5 pay\_I1.2 a\_N5+[i6.2.1 lot\_N5+[i6.2.2 of\_Z5 money\_I1 .PUNC

夢 One\_T1.1.3[i9.2.1 day\_T1.1.3[i9.2.2 I\_Z8mf was\_A3+ on\_M6[i10.3.1 my\_M6[i10.3.2 way\_M6[i10.3.3 home\_H4 from\_Z5 school\_P1/H1c .PUNC I\_Z8mf saw\_X3.4 a\_Z5 snake\_L2mf on\_M1[i11.3.1 the\_M1[i11.3.2 road\_M1[i11.3.3 .PUNC I\_Z8mf do\_Z5 n't\_Z6 like\_E2+ snakes\_L2mf .PUNC The\_Z5 snake\_L2mf said\_Q2.1 .PUNC " .PUNC Come\_M1[i12.2.1 on\_M1[i12.2.2 .PUNC everybody\_Z8/N5.1+c ! .PUNC " .PUNC Then\_N4 there\_Z5 were\_A3+ many\_N5+ snakes\_L2mf around\_Z5 me\_Z8mf .PUNC There\_Z5 were\_A3+ about\_Z5 one\_N1[i13.2.1 hundred\_N1[i13.2.2 snakes\_L2mf ! .PUNC

浦島 Urashima\_Z1mf[i14.2.1 Taro\_Z1mf[i14.2.2 was\_A3+ very\_A13.3 shocked\_E5- .PUNC But\_Z5 he\_Z8m remembered\_X2.2+ his\_Z8m Qz\_Z99 .PUNC He\_Z8m didn\_Z99 t\_Z5 give\_T2-[i15.2.1 up\_T2-[i15.2.2 .PUNC He\_Z8m thought\_X2.1 .PUNC " .PUNC I\_Z8mf will\_T1.1.3 write\_Q1.2 a\_Z5 book\_Q4.1 about\_Z5 my\_Z8 trip\_M1 and\_Z5 sell\_I2.2 it\_Z8 .PUNC " .PUNC

#### 付録2 モデル文の語彙分析：品詞標識(除く、MWU)と意味標識の対応関係(標識付与エラーを含む)

語彙	品詞標識	頻度	意味標識	頻度	語彙	品詞標識	頻度	意味標識	頻度
a	AT1	8	Z5	8	months	NNT2	1	T1.3	1
a lot		1	N5+	1	morning	NNT1	1	T1.3	1
about	II	2	A13.4	1	morning classes		1	T1.3	1
about	RG	1	Z5	2	my	APPGE	9	Z8	9

<i>after</i>	II	1	Z5	1	<i>my_own</i>		1	Z8	1
<i>ago</i>	RA	1	M45.1.1	1	<i>nagatoro</i>	NN1	1	Z99	1
<i>always</i>	RR	2	N6+++	2	<i>new</i>	JJ	1	T3-	1
<i>am</i>	VBM	1	A3+	1	<i>next_month</i>		1	M45.1.3	1
<i>and</i>	CC	10	Z5	10	<i>n't</i>	XX	4	Z6	4
<i>anyone</i>	PN1	1	Z8	1	<i>of</i>	IO	1	Z5	1
<i>anything</i>	PN1	1	Z8	1	<i>on_my_way</i>		1	M6	1
<i>are</i>	VBR	1	A3+	1	<i>on_the_road</i>		1	M1	1
<i>around</i>	II	2	Z5	2	<i>one</i>	MC1	1	N1	1
<i>bad</i>	JJ	1	A5.1-	1	<i>one_day</i>		1	M45.1.3	1
<i>be</i>	VBI	2	A3+	2	<i>one_hundred</i>		1	N1	1
<i>beautiful</i>	JJ	1	O4.2+	1	<i>or</i>	CC	1	Z5	1
<i>became</i>	VVD	1	A2.1+	1	<i>otoshidama</i>	NN1	1	Z99	1
<i>better</i>	RRR	1	A5.1++	1	<i>pay</i>	VVI	1	I1.2	1
<i>big</i>	JJ	2	N3.2+	2	<i>people</i>	NN	1	S2	1
<i>book</i>	NN1	3	Q4.1	3	<i>remembered</i>	VVD	1	X2.2+	1
<i>bought</i>	VVD	3	I2.2	3	<i>rent</i>	VVI	1	I2.2	1
<i>boys</i>	NN2	1	S2.2	1	<i>rice</i>	NN1	2	F1	2
<i>bread</i>	NN1	2	F1	2	<i>rich</i>	JJ	1	I1.1+	1
<i>breakfast</i>	NN1	1	F1	1	<i>sad</i>	JJ	1	E4.1-	1
<i>bring</i>	VVI	2	M2	2	<i>said</i>	VVD	1	Q2.1	1
<i>but</i>	CCB	6	Z5	6	<i>save</i>	VV0	1	I1.1	1
<i>butter</i>	NN1	1	F1	1	<i>saw</i>	VVD	1	X3.4	1
<i>buy</i>	VVI	1	I2.2	1	<i>say</i>	VV0	1	Q2.1	2
<i>by</i>	II	1	Z5	1	<i>say</i>	VVI	1	Q2.1	2
<i>caught</i>	VVD	1	A9+	1	<i>school</i>	NN1	1	P1	1
<i>class</i>	NN1	1	P1	1	<i>sell</i>	VV0	1	I2.2	1
<i>cm</i>	NNU	1	N3.3	1	<i>shocked</i>	JJ	2	E5-	2
<i>come_on</i>		1	M1	1	<i>ski</i>	VV0	1	K5.1	1
<i>could</i>	VM	1	A7+	1	<i>skiing</i>	NN1	3	K5.1	3
<i>didn</i>	VV0	1	Z99	1	<i>slim</i>	JJ	1	N3.5-	1
<i>difficult</i>	JJ	1	A12-	1	<i>snake</i>	NN1	2	L2	2
<i>do</i>	VD0	5	A1.1.1	2	<i>snakes</i>	NN2	4	L2	4
<i>do</i>	VD0	5	Z5	3	<i>so</i>	RR	2	Z5	2
<i>don</i>	VV0	4	P1	4	<i>some</i>	DD	2	N5	2
<i>dream</i>	NN1	2	X4.1	2	<i>sometimes</i>	RT	1	N6	1
<i>every_month</i>		1	N6+	1	<i>started</i>	VVD	1	T2+	1
<i>everybody</i>	PN1	1	Z8	1	<i>story</i>	NN1	1	Q2.1	1
<i>family</i>	NN1	2	S4	2	<i>t</i>	ZZ1	6	Z5	6
<i>fat</i>	JJ	1	N3.2+	1	<i>tells</i>	VVZ	1	Q2.2	1
<i>fat</i>	NN1	1	O1	1	<i>ten</i>	MC	1	N1	1
<i>father</i>	NN1	2	S4	2	<i>that</i>	CST	1	Z5	1
<i>feel</i>	VV0	1	X2.1	1	<i>that</i>	DD1	1	Z8	1
<i>finished</i>	VVD	1	T2-	1	<i>the</i>	AT	5	Z5	5
<i>first</i>	MD	2	N4	2	<i>them</i>	PPHO2	1	Z8	1
<i>fish</i>	NN	1	K5.1	1	<i>then</i>	RT	1	N4	1
<i>fish</i>	VVI	1	L2	1	<i>there</i>	EX	3	Z5	3
<i>fishing</i>	NN1	2	K5.1	2	<i>they</i>	PPHS2	1	Z8	1
<i>friends</i>	NN2	1	S3.1	1	<i>thought</i>	VVD	1	X2.1	1
<i>from</i>	II	1	Z5	1	<i>to</i>	II	1	Z5	7
<i>girl</i>	NN1	1	S2.1	1	<i>to</i>	TO	6	Z5	7
<i>give_up</i>		1	T2-	1	<i>trip</i>	NN1	1	M1	1
<i>gloves</i>	NN2	1	B5	1	<i>urashima</i>	NP1	1	Z99	1

go	VV0	2	M1	3	urashima_ taro		1	Z1	1
go	VVI	1	M1	3	use	VVI	1	A1.5.1	1
happy	JJ	1	E4.1+	1	usually	RR	1	A6.2+	1
have	VH0	2	A9+	3	very	RG	8	A13.3	8
have	VHI	1	A9+	3	very_much		1	A13.3	1
have_to		1	S6+	1	want	VV0	3	X7+	4
he	PPHS1	5	Z8	5	want	VVI	1	X7+	4
help	NN1	1	S8+	1	wanted	VVD	1	X7+	1
his	APPGE	3	Z8	3	was	VBDZ	9	A3+	9
home	RL	1	H4	1	wasn	VV0	1	Z99	1
honey	NN1	1	F1	1	wear	VV0	1	B5	1
house	NN1	1	H1	1	well	RR	1	A5.1+	1
how	RRQ	1	Z5	1	went	VVD	1	M1	1
hungry	JJ	1	F1-	1	were	VBDR	2	A3+	2
i	PPIS1	29	Z8	30	will	VM	4	M45.1.3	4
i	ZZ1	1	Z8	30	winter	NNT1	1	T1.3	1
in	II	4	Z5	4	with	IW	3	Z5	3
interesting	JJ	1	X5.2+	1	write	VVI	1	Q1.2	1
is	VBZ	4	A3+	4	writing	VVG	1	Q1.2	1
it	PPH1	10	Z8	10	year	NNT1	1	T1.3	1
last_summer		1	M45.1.1	1	0	MC	4	N1	4
like	II	1	E2+	3	30	MC	1	T3	1
like	VV0	2	Z5	1	(日本語)	FO	1	Z99	1
like	VVI	1	Z5	1	(日本語)	FO	3	Z99	3
lose	VVI	1	X9.2-	1	(日本語)	NP1	1	Z99	1
love	VV0	1	E2+	1	(日本語)	FO	1	Z99	1
makes	VVZ	1	A1.1.1	1	(日本語)	FO	1	Z99	1
many	DA2	2	N5+	2	(日本語)	FO	1	Z99	1
married	VVD	1	S4	1	(日本語)	FO	1	Z99	1
me	PPIO1	5	Z8	5	(日本語)	FO	2	Z99	2
milk	NN1	1	F2	1					
money	NN1	1	II	1					