

# タグ付きコーパスの句構造の解析について

古 橋 聰  
高 橋 薫

## 1 はじめに

コンピュータを用い英語の文章を様々な面から解析した研究報告は数多くある。そのためには、なによりも文章をコンピュータで解析可能なデータベース化する作業を経て、解析に関しては主に語彙レベルのものが中心であった。すなわち分野別の語彙の使用状況や特定の単語の出現頻度等であった。しかしながら、さらに高度な解析を目指す時、語彙レベルの解析では自ずと限界がある。文章を文体や構文といった単位で扱うことが必要となる。

本報告は筆者が入手した代表的な英語文章データベースを用い、そのデータベースの特徴を生かし、句レベルの解析を中心に行う。また、このデータベースは文章が論説文や小説等さまざまな分野別に区分されているので、分野別の解析結果の比較も興味深い報告になるとと思われる。

## 2 タグ付きコーパスについて

中等教育で普通に教える文法は学校文法あるいは規範文法と呼ばれるものであり、品詞についての理解が大変重要な位置を占める。品詞は一般的に8品詞、すなわち、名詞、代名詞、形容詞、動詞、副詞、前置詞、接続詞、間投詞であり、さらに冠詞を加え、9品詞とする考えもある。ただし、これ以上の分類はそれぞれの品詞内を細分化することであり、大まかにはこの程度が妥当であるとされている。中学、高校ではこれらの品詞についての役割を習得させることにより、講読や作文などの総括的な英語力を高

めることを意図している。

ところで、それぞれの品詞についてさらに理解を深めるために、さまざまな分類がされている。名詞であれば意味上の分類から普通、集合、物質、抽象、固有名詞、また動詞であれば、その活用により規則、不規則動詞に分けられる。このような分類はどの文法書でも基本的には同じであるが、その文法書の目的とする習得の達成度に応じて異なるものである。英語教師もその学生の既習得レベルを考慮しつつ、教壇に立つ訳である。

そのようなこともあり、8品詞あるいは9品詞を越えて単語にいったい何種類の分類をするのが適切であるのかという議論は無意味であったとも言える。100種類であるとか、もっと細分化して200種類であると言った議論はある意味では歯止めの利かぬ議論であろう。

ところが近年コンピュータが発達し、文科系の教師が容易にそれを扱える時代となると、事態が急変する。コンピュータが大量の文字データを解析する、すなわち単語間の配列のなかに普遍的な規則性を持つ事を経験則を越えて統計的に立証できることが期待できる。1980年、エジンバラ大学のモートン教授は当時イギリスの大英博物館に眠っていた作者不祥の文学作品について、特定の語の位置や頻度をコンピュータを用い調査することによって、それがシェークスピアの作品であることを立証し<sup>①</sup>、そのことが文法学者にも新しい時代が訪れたことを暗示したのである。

その後コンピュータで解析が可能な形式の文章データベースの構築が盛んになりコンピュータの得意とする用語検索、ワードリスト等を用い、語彙に関する研究が盛んになり、それについての研究報告も数多く出ている。しかしこれらは、文章データベースが単語以外に情報を持っていないことから文章解析にも限界があった。仮に各単語に8品詞に関する情報が付加されるだけでもかなり飛躍的な解析が行えることが期待できる。

Computer Corpus in English Language Research (1982) ではすでに構築した文章データベースに文法的役割を示すタグを多数設定した点に注目できる。それ以後英語文章データベース<sup>②</sup> (以後コーパスという) にタグを付加することが今後の英語文章の解析の必須条件となり、研究の中心はどのような種類のタグを設定するかに移った。

Brown Corpus はアメリカ英語のタグ付きコーパスで百万語からなる

データベースである。ここではタグを 87 個に設定し各単語に付加している。また、LOB Corpus はイギリス英語の同規模のコーパスであり Brown Corpus のタグ設定を参照し、その種類を 134 個に増やした。このタグは基本的には Brown Corpus のタグに基づいているが新たな固有名詞をさらに 4 種類に分類する等、特定の品詞をさらに細分化し、後の解析にとってより理想的な形式としている。また、これらのタグのおかげで、様々な先進的な文章解析の研究報告が出されている。

ところでタグが文章解析に大変有効な手段である理由として次のことが挙げられる。すなわち、文型についての検索が容易に行える点である。たとえば受動態の文章を探したい場合、be 動詞のタグと過去分詞のタグを中心に検索すればよい。このような検索をタグのない文章で行うことに比べてみると解析がはるかに容易に行える。なお、LOB Corpus の他の特徴として、コーパスが社説、評論に始まり、小説にいたる 15 種類のカテゴリからなる点にある。そのため、様々な点でのカテゴリ間の比較が行える。問題となるのは語彙レベルの分析を越えた句レベルでの解析のためにコンピュータのハード及びソフトに関わる技術面である。それらに投資すべき費用も従来の英語学の研究に比べると比較にならないものがある。次にコンピュータを中心とした解析方法に触れていきたい。

### 3 解析方法

英文のコーパスをコンピュータで扱う場合、ハード的な問題点としてデータベースの容量が挙げられる。LOB Corpus は 11 メガバイト (フロッピーディスク 11 枚相当) 以上の容量のデータベースなので、もはやフロッピーディスクを抜き差ししてデータの交換を行うわけにはいかない。解析結果のデータの蓄積等も考慮して 100 メガバイト程度のハードディスクが必要となる。これがコーパスの解析のためのハード的な必要条件だと言える。

次に具体的な解析のためのソフトとして OCP (Oxford Concordance Package) がある。これは英国の研究機関で主流となっている文章解析ソフトで、現在ではパソコンで使用可能となっている最も強力な解析用ソフ

トである。これは、ワードリスト、インデックスの用語検索を始め、接頭語、接尾語、句、など、文章解析に必要な作業を簡単なコマンドで処理することができる。特にタグの活用によって、目的とする句構造の検索が容易に行える。

以上のような設備を整え、文章解析を行うものである。

## 4 LOB Corpus について

### 4-1 カテゴリーについて

表1に分野別のカテゴリーと大まかな文章数、使用されている単語数、MS-DOS管理上での容量が示されている。なお、データはテキストファイルで扱われている。<sup>(3)</sup>

表1 カテゴリーの種類と文章数

カテゴリー	文章の数	単語数	容量 (バイト)
A: 報道 (ルポルタージュ)	44	107,852	1,051,310
B: 報道 (社説)	27	64,203	630,806
C: 報道 (評論)	17	24,336	237,458
D: 宗教	17	24,298	229,600
E: 技能, 手仕事, 趣味	38	90,184	868,842
F: 著名な伝説	44	105,666	1,023,271
G: 純文学, 伝記, 随筆	77	182,200	1,780,694
H: その他 (政府書類, 学生便覧, 工業報告書, 大学紹介, 社内報)	30	69,964	700,608
J: 学術書, 科学書	80	190,196	1,874,822
K: 小説全般	29	73,672	675,354
L: 怪奇小説, 推理小説	24	61,747	562,190
M: 科学小説	6	15,528	143,540
N: 冒険小説, 西部小説	29	76,181	686,065
O: 空想小説, 恋愛小説	29	0	0
P: ユーモア小説	9	75,703	683,389
計	471	1,161,730	11,147,949

LOB Corpus の各分野の文章を一部タグを省いた形式で示す。

(A) 報道 (ルポルタージュ)

'Stop Electing Life Peers.'

By Trevor Williams.

A move to stop Mr. Gaitskell from nominating any more labour life peers is to be made at a meeting of labour MPs tomorrow.

Mr. Michael Foot has put down a resolution on the subject and he is to be backed by Mr. Will Griffiths, MP for Manchester Exchange.

Though they may gather some left-wing support, a large majority of labour MPs are likely to turn down the Foot-Griffiths resolution.

(B) 報道 (社説)

The President's speech.

Sir, - we are much indebted to the Times for publishing yesterday, in full, the broadcast of the President of the United States to his people - and to the world - on account of his recent visit to Europe.

This address, and the President's inaugural speech, has brought a voice and an authority to the councils of the free world 'and outside it' that speaks in frank, clear, and unambiguous terms, enabling those who hear and read to appreciate the dangers and the immense issues involved. His language, more than that of any other, reminds me of the great utterances of Sir Winston Churchill during 'and immediately after' the war.

(C) 報道 (評論)

Television.

Life of Miss Nightingale.

Skilful picture.

The BBC's dramatised documentary on Florence Nightingale last night cleverly managed to suggest the person behind the legend.

While never minimising the immensity of her work, it lifted the saintly halo which usually surrounds her name to reveal a warm, dedicated person who accomplished most by perseverance and hard work.

Most stories of Miss Nightingale begin and end with her work in the Crimea. This one started from that point and devoted itself to her lifelong campaign to improve nursing in this country. The documentary managed to show the obstacles and her devotion.

#### (D) 宗教

With so many problems to solve, it would be a great help to select some one problem which might be the key to all the others, and begin there. If there is any such key-problem, then it is undoubtedly the problem of the unity of the gospel. There are three views of the fourth gospel which have been held.

Some critics, not many, argue that the gospel is the product of one mind and one hand. For them the problems of the fourth gospel exist only in the mind of its detractors.

#### (E) 技能, 手仕事, 趣味

Introduction

What a world of graceful accomplishment lies in a piece of finely worked hand-made lace! The very word lace has a charming derivation, stemming through the old French *las*, coming from the Latin *laqueus*, a snare, allied to *lacere*, to entice.

Certainly all who make a study of lace-craft become enmeshed in a highly coloured history — a pattern of proud queens and pious ecclesiastics, of statesmen, burghers and poor, under-

privileged people. All the human passions are drawn in its threads, from saintliness to downright sinfulness, for it shows the smuggler running 'laces for a lady' and the highwayman meeting death upon the gallows 'in a fine show of hand-made lace.' As to the origins of lace-making in this c.

(F) 伝記

The 'ladders of the mind' are the clues which we use to track down items of knowledge which can not be immediately remembered.

The 'organisation' and the 'shelves' will form important topics in our later discussion. The work that goes on at the bench must also be considered. For here the items which go into the store may be taken to pieces and reassembled, and a sketch may be made of their internal construction. Both the item and the sketch can then go into the store. In the mind it is the stored items which constitute our memory and it is the stored sketches which constitute our understanding. Thus reading for understanding means taking items of knowledge to pieces as we read them and seeing how the pieces are connected.

(G) 純文学, 伝記, 随筆

A Scottish knight 'Sir John Mercer' was imprisoned in England. His son, in revenge, was hurrying English shipping as far away as Cherbourg, and doing it to some purpose.

John Philpot, one of that new class of merchant financiers which the city of London was now producing, fitted, equipped and manned a fleet from his own resources, and captured the young Mercer in a brilliant Channel fight. It was naturally a highly popular victory with the Londoners, but it brought heavy censure from nobles who still believed that they had a monopoly of leader-

ship. But, at last, Gaunt sailed. Opposing him was the French Admiral, 'Jean de Vienne' a great sailor and an able strategist. Obedient to the policy of his King, de Vienne avoided trouble at sea as cleverly as Du Guesclin avoided it on land. Gaunt was compelled to give up his search for an elusive foe, and, afraid to return home without something to show, he foolishly attempted to besiege the well-protected fortress of St. Malo.

(H) その他 (政府書類, 学生便覧, 工業報告書, 大学紹介, 社内報)

Questions about marriage and children were again included, as they had been at the 1911 and 1951 censuses. The former had asked for marriage details for all married women, the latter for all married women under the age of 50. The 1961 census questions related to all women who were or had been married, and so repeated the enquiry made fifteen years earlier by the 1946 family census conducted on behalf of the royal commission on population. The questions about children were the same as in 1951, except that they extended to all women who were or had been married. It was not regarded as practicable within the limits of the census to include particulars about the date of birth of each child, as had been done in the specialised enquiry for the royal commission. The innovation at the 1961 census, though the family census had previously included the question, was the date of termination of the first or only marriage.

(J) 学術書, 科学書

Unfortunately the accuracy with which an impurity dependent physical or chemical property of sodium can be measured decreases with decreasing impurity concentration. To get over this difficulty Alcock has suggested that instead of measuring directly the concentration of oxygen in the flowing sodium its thermo-



dynamic potential should be measured by a suitable galvanic cell incorporated in the circuit. The principal advantages of this should be continuous monitoring of the sodium and an accuracy of monitoring which, if the sodium-oxygen system obeys Henry's law, should increase with decreasing concentration of the impurity.

#### (K) 小説全般

'Are you sure you're quite fit? It's terrible weather.' He turned round to face his colleague.

For some esoteric reason Fairbanks always completed the buttoning of his flies in the main area of the lavatory. 'Good morning, Harold,' he said. 'I'm pretty chipper, thanks, considering.' He was a tiny man, of fanatical neatness, his remaining hair snowy, and cropped like a Prussian's. His white shirt cuffs were actually starched: he protruded from them his surprisingly thick and hairy wrists and began to wash. 'As a matter of fact a good hard frost seems to clear the old tubes. Much better for me than the rain.

'Good,' said Colmore.

#### (L) 怪奇小説, 推理小説

He did well. He got in touch with the woman Pete was passing off as his mother. Starmouth managed to win her confidence. It seems that she was an honest enough woman, only her mind wasn't as clear as it could have been. She showed him photographs. He found out that the name of her house 'Grand Greve' was taken from a bay in one of the Channel Islands. One of the small ones. Sark, that's it. The Caxtons used to have their holidays there. Starmouth went there. He dug out some people who remembered the family. In the end he pieced it all together.

**(M) 科学小説**

A shudder, more mental than physical, ran through him, and his mind seemed to melt away into emptiness. His bulging eyes caught the reclining form of Heather, who was still repeating in sing-song: ' . . . I will not give in . . . . ' He stared at her blankly, mouthing an incoherent gabble of half words . Then he broke into a crazy laugh that made rolling echoes through the house, and trailed-off into a long-drawn-out unearthly wail. The wail should have been despairing; but its eerie note, even in its senseless gagging babble, was jubilant, triumphant.

**(N) 冒険小説, 西部小説**

Alastair was a bachelor. All his life he had been inclined to regard women as something which must necessarily be subordinated to his career. Now he realised that he was entrusting not only his own life, but Geoffrey's as well, to a strange girl whom he had never met and on whom after no more than a couple of telephone calls he was prepared to place complete reliance. It was a novel experience.

Alastair was quite unable to explain why he should feel so much confidence in her.

**(P) 空想小説, 恋愛小説****Chapter 29.**

The black candle Saturday night cellar party was on. The time was approaching midnight. Margot and Jasmine had left together shortly after ten. Before doing so at Mr. Copthorne's invitation they had spent a pleasant half hour with Con O'Shea, Mr. Butler and Mr. Narain Khuma at a table for four. At another table nearby Mr. Smith and two of his officers in plain clothes were quietly chatting together.

#### 4-2 タグの種類について

LOB Corpus において単語を語類に分類する方法は変形文法の分類法に類似している。大まかに8つの品詞, すなわち名詞, 動詞, 形容詞, 前置詞, 接続詞, 代名詞, 限定詞, 副詞を中心とした分類である。第一段階でこれらをベースタグとして, さらに細分化している。タグの先頭の1文字, あるいは2文字目までが大まかな分類を示している。さらに, 第二段階では接尾辞が加わる。表1にタグの分類を文法用語とともに示す。第三段階の分類は後述のような役割を持つ接尾辞がさらに加わり, タグの総数は134種類となる。なお, イディオムの場合はタグに double quotation ” を付けて表現する。

表1 接尾辞一欄

#### I 限定詞, 代名詞, 名詞, 数詞につく接尾辞

A: 主格 O: 目的格 I: 単数または複数 S: 複数 \$: 所有格

R: 関係詞

#### II 動詞につく接尾辞

D: 過去形 G: 現在分詞, 動名詞 N: 過去分詞 Z: 3人称単数

#### III 形容詞, 副詞につく接尾辞

R: 比較級 T: 最上級

表2に各タグに対応する単語を示す。<sup>(4)</sup>

表2 タグ一欄

A...

ABL	前位修飾語	quite, rather, such
ABN	前位数量詞	all, half <sup>(5)</sup>
ABX	相関接続詞	both, etc
AP	後位限定詞	few, fewer, former, last, latter, least, less, little, many, more, most, much, next, only, other, own, same, several, very <sup>(6)</sup>

AP"	"	a few, a little
AP\$	" (所有格)	other's
APS	" (複数)	others
APS\$	" (複数所有格)	others'
AT	単数冠詞	a, an, every
ATI	単複数冠詞	the, no
BE	動詞、助動詞の be	be
BED	" (過去)	were
BEDZ	" (過去)	was
BEG	" (ing )	being
BEM	"	am, 'm
BER	"	are, 're
BEZ	"	is, 's
CC	等位接続詞	and, and/or, but, nor, only, or, yet
CD	基数	2, 3, two, three, etc ; hundred, thousand, etc ; dozen, zero
CD1\$	" (所有格)	one's
CD-CD	" (ハイフオン)	2-1 (football score)
CD1S	" (複数)	ones
CDS	" (-s )	tens, millions, dozens, etc
CS	従位接続詞	after, although, etc
CS"	"	in that, so that, etc
DO	本動詞、助動詞	do
DOD	" (過去)	did
DOZ		does
DT	単数限定詞	another, each, that, this
DT\$	" (所有格)	another's
DTI	" (単複)	any, enough, some
DTS	複数限定詞	these, those
DTX	限定詞、二重接続詞	either, neither
EX	存在の there	there
HV	have	have

HVD	" (過去)	had, 'd
HVG	" (-ing)	having
HVN	" (過去分詞)	had
HVZ	"	has, 's
IN	前置詞	about, above, etc
IN"	"	as to, in spite of, etc
JJ	形容詞	略
JJB	" (限定的用法のみ)	chief, entire, main, etc
JJR	(比較級)	略
JJT	(最上級)	略
JNP	" (文頭大文字)	English, German, etc
MD	法助動詞	'll, can, could, etc
N	名詞	
NC	引用語	' * * * '
NN	普通名詞	略
NNP	文頭大文字 (単数普通名詞)	Englishman, German, etc
NNU	測量単位の省略形	hr, lb, etc
NNU"	"	per cent
NNUS	" (複数形)	gns, yds, etc
NP	単数固有名詞	略
NP\$	" (所有格)	略
NPL	単数名詞 (文頭大文字で場所に関する)	Abbey, Bridge
NPL\$	" (所有格)	略
NPLS	" (複数形)	略
NPLS\$	" (所有格)	略
NPS	複数固有名詞	略
NPS\$	" (所有格)	略
NPT	称号等の名詞 (文頭大文字で称号等を示す)	Archbishop, Captain
NPT\$	" (所有格)	略
NPTS	" (複数形)	略
NPTS\$	" (所有格)	略

NR	副詞的名詞 (单数)	January, February, etc ; Sunday, Monday, etc ; east, west, etc ; today, tomorrow, tonight ; downtown, home
NR \$	" (所有格)	略
NRS	" (複数形)	略
NRS \$	" (所有格)	略
OD	序数	1st, 2nd, etc ; first, second, etc
OD \$	" (複数形)	略
PN	不定代名詞	anybody, anyone, anything, everybody, none, nothing ; somebody, someone, something ; so
PN"		no one, some one
PN \$		略
PN \$ "		略
PP	人称代名詞	略
PP \$	所有限定詞	my, your, etc
PP \$ \$	所有代名詞	mine, yours
PP1A	人称代名詞	I
PP1AS	"	we
PP1O	"	me
PP1OS	"	us, 's
PP2	"	you, thou, thee
PP3	"	it
PP3A	"	he, she
PP3AS	"	they
PP3O	"	him, her
PP3OS	"	them, 'em
PPLS"	複数再帰代名詞他	each other, one another
QL	修飾語 (副詞)	as, awfully, less, more, so, too, very, etc
QLP	後位修飾語	enough, indeed

RB	副詞	略
RB"		at last, in general, etc
RB\$	" (所有格)	else's
RBR	" (比較級)	略
RBT	" (最上級)	略
RI	前置詞的副詞	below, near, etc
RN	名詞的副詞	here, now, there, then, etc
RP	副詞辞	back, down, off, etc
TO	不定詞の to	to
TO"		in order to, so as to
UH	感嘆詞	略
VB	本動詞	略
VB"		略
VBD	過去形	略
VBG	現在分詞、動名詞	略
VBN	過去分詞	略
VBZ	3人称、単数形	略
WDT	WH 限定詞	what, whatever, whatsoever, which, whichever, whichsoever
WDT"		略
WDTR	WH 限定詞、関係詞	which
WP	WH 代名詞、疑問詞	who, whoever
	" (主格、目的格)	略
WP\$	" (所有格)	whose
WP\$R	関係詞 (所有格)	whose
WPA	WH 代名詞 (主格)	whosoever
WPO	疑問詞 (目的格のみ)	whom, whomsoever
WPOR	関係詞 (目的格)	whom
WPR	" (主格、目的格)	that, who
WPB	WH 副詞	how, when, etc
XNOT		not, n't
ZZ		e, pi, x, etc

これらのタグをコーパスの単語一語一語にコンピュータで自動的に割り付けて表3の形式のタグ付きコーパスが完成するのである。

表3 タグ付き LOB Corpus の一部

^ the\_ATI film\_NN is\_BEZ a\_AT well-made\_JJ variation\_NN  
 on\_IN that\_DT sinister\_JJ yarn\_NN in\_IN  
 which\_WDTR half\_ABN the\_ATI cast\_NN try\_VB to\_TO  
 persuade\_VB  
 heroine\_NN that\_CS she\_PP } A is\_BEZ out\_RP of\_IN her\_PP \$  
 (中略)

^ this\_DT perhaps\_RB is\_BEZ what\_WDT gave\_VBD  
 fascination\_NN with\_IN bodies\_NNS that\_WPR  
 were\_BED agile\_JJ ,\_, bodies\_NNS that\_WPR could\_MD  
 do\_DO what\_WDT

## 5 具体的な解析について

語彙レベルの分析についての研究報告以降、文体や構文についての解析についての研究が望まれるところではあったが、進展はあまりないようである。しかし、タグ付きコーパスをOCPを用いることによってそれが可能となった。ここでは具体的にどのような形式の文章についてそのような解析が可能であるかを分野別のコーパスに現れる特徴の比較について仮説をたてて、述べていきたい。特にコーパスの特徴については主に文語体と口語体の2系統での相違点について検討したい。

### 5-1 受動態

受動態にできる他動詞文の意味的な特徴の1つは、動作をするもの(動作主)と動作を受けるもの(対象)があることである。自然言語は人間中心にできており、能動態のように、動作主(典型的には人間)が主題になって、それについて何かを述べる形が自然である<sup>(7)</sup>。受動態はこの一般的な傾向とは逆になるわけで、その複雑な操作はかなり意図的である<sup>(8)</sup>。

したがって、論説文と小説では明らかに受身文の出現頻度に違いが出てくるはずである。ここで第一の仮説として「小説に現れる受動態は論説文におけるそれよりも出現頻度が少ない傾向にある」をたてて検証していくことにする。



### 5-2 受動態に現れる副詞について

ある文章中のワードリストを作成する場合それが語彙レベルの検索であるため如何なる文体でどのような副詞が現れるかといったことは全く無関係となっていた。しかし、5-1にあるように能動態についての検索を行いつつ、副詞に注目することがここでは可能である。したがって、受動態の分野別の副詞の特徴を分析していく。

### 5-3 代名詞+of+名詞

句について解析を行う場合、名詞句についての検索をどのように行うかが問題となる。初歩的には名詞+前置詞+名詞の形で前置詞句が形容詞的な働きを持って前の名詞を修飾し、全体として名詞句を構成する場合が挙げられる。しかしながらこの形の検索をすることは唯単に膨大な単語リストを作成することにほかならない。分野別の特徴を表し、百万語のデータベースのなかで検索によって適量の結果が現れ、分野別に比較対照できるものとして代名詞+of+名詞に着目できる。また、many, much は口語体、または形式ばらない文章で用いられることが少ないとされているので<sup>9)</sup>、その点について観察し検証していきたい。また、LOB Corpus の課題の一つとして考えられることとして、LOB Corpus はタグの種類豊富さという点においては、BROWN Corpus の発展形であるが必ずしもそれでタグが十分だとは言えない。たとえば、限定詞が代名詞の用法を共有する単語においても代名詞のタグが付くことはない。そのような単語として little, least, more, most, many 等が挙げられる。

これらの単語の代名詞としての用法を検索する手だてとして代名詞後に of が来る名詞句の形式を検索することによってある程度使用頻度の傾向がつかめることも期待できる。

### 5-4 形容詞+of+名詞

形容詞は分野別で使われる種類に違いが出てくるのは当然である。小説であれば、他に比べ人間の感情に関する形容詞、性質、性格を表す形容詞が多く使われることは十分予想される。ここでは単独の形容詞のそれぞれの分野における頻度を調査するとともに、句構造の形容詞にも着目し、それが形容詞全体に占める割合についても分野別に比較していく。

句構造の形容詞とは形容詞の直後に前置詞句が続き全体として形容詞句

を作っているものでこのタイプでは以下のような前置詞句が続くことが予想される。

- a. Pity is akin to love.
- b. He is eager for success.
- c. She is ignorant of the danger.
- d. My opinion is different from yours.

ここでは特に of の前置詞句を持つ文を検索し頻度を調べる。また、その場合、次の構文が同時に検索されるが、この構文にも特に注目することにした。

it+be 動詞+形容詞+of+名詞の構文ではその形容詞が、ある人物の行為を通じてその人物に対する話者の主観的評価を述べるものであり、これは分野別に使われ方に明らかに違いが出てくることが予想される。これについては具体例を列挙する。

## 6 解析結果

### 6-1 受動態

Micro-OCP を用い受身文の検索を分野毎に行った<sup>(10)</sup>。ただし、不定詞の形容詞用法で受身の形、また、be+being+過去分詞は今回は除外した。さらに次の様にグループ分けをし、出現率についての比較を行った。

- ア) be 動詞現在形のもの: am are is
- イ) be 動詞が過去形のもの: was were
- ウ) 現在, 過去完了形のもの: have+been+過去分詞, had+been+過去分詞
- エ) 助動詞+be 動詞+過去分詞のもの, to 不定詞+be 動詞+過去分詞

表4に分野毎のア)～エ)についての出現度数を示す。これは分野別の出現頻度をそこで使用されている単語数で割ったものである。たとえば、Aのア)では、ルポルタージュの分野でのア)のタイプの受身文の頻度が191個であり、これをこの分野で使用されている全単語数107,852で割ったものを出現率とした。

最終段の小数値がそれぞれの分野別の出現率を表していて、AからJま

表4 受動態の出現度数

上段 出現度数×0.00001  
 : 数値の177とは10万語中に177回の出現率を表す  
 下段 (出現個数  
 : 実際の出現頻度を表す  
 計 : .008とは1000語中8回の出現率を表す。

	A	B	C	D	E	F	G	H	J	K	L	M	N	P
ア)	177	227	214	482	439	248	159	354	426	49	24	39	39	38
	(191	(146	(52	(117	(396	(262	(289	(248	(810	(36	(15	(6	(30	(29
イ)	325	134	160	235	224	314	344	273	311	185	232	110	152	130
	(351	(86	(39	(57	(202	(332	(506	(191	(592	(142	(143	(17	(116	(99
ウ)	117	120	82	82	86	81	115	170	131	61	100	103	49	62
	(127	(77	(20	(20	(78	(86	(209	(119	(250	(45	(62	(16	(37	(47
エ)	202	312	156	313	367	182	197	386	351	56	78	90	46	79
	(218	(200	(38	(76	(331	(192	(359	(270	(668	(1	(48	(14	(35	(60
計	.008	.008	.006	.011	.011	.008	.008	.012	.012	.004	.004	.003	.003	.003

A B C D E F G H J K L M N P

で(AJ群)は千語中6回から最高11回の頻度で、またKからPまで(KP群)は4回以下の頻度で現れることがわかる。また、KP群についてア)からエ)についてはいずれの分野もイ)の過去形での出現率が他よりもはるかに多く、これはAP群には見られない傾向である。その他、AJ群のなかのA(ルポルタージュ)とG(純文学、伝記、随筆)はKP群と同じ傾向が過去形に現れている。

6-2 受動態に現れる副詞について

be動詞と過去分詞の間に挟まる副詞については、あ)無いもの、い)1語のもの、う)2語のものを現在形、過去形について調べた。

現在形と過去形の受身文に現れる副詞はKP群がもともと受身文の頻度が少ないために表5での数値が低い他は1単語のもの、2単語のものともほぼ同じ傾向を示している。特に1単語の副詞は副詞無しを受動態に比

表5 副詞の出現度数

数値は表4同様10万語につきの出現率を表す

		A	B	C	D	E	F	G	H	J	K	L	M	N	P
現在	あ)	138	173	160	383	361	185	120	280	351	39	19	32	34	38
	い)	33	47	33	91	68	50	35	61	59	10	3	0	5	9
	う)	6	7	16	8	10	1	3	7	5	0	2	0	3	0
過去	あ)	290	113	119	177	185	279	162	240	266	158	207	97	130	155
	い)	34	19	41	45	38	30	46	33	41	31	23	13	2	30
	う)	6	1	0	1	1	5	4	0	4	3	2	0	3	3

べ1単語の副詞の受動態は4, 5倍程度の出現率で一定している。

また、実際に使用されている副詞について、同様の群分けをして考察するとAJ群についてはそれぞれの分野及び現在形、過去形で全般的に、ある共通の副詞が使用されていて、下記のnotにおいてはAJ群の現在形、過去形で17箇所にわたり使用されている。同様に比較的多く使用されている副詞を頻度順に示す。( )内の数値は上述の箇所数である。

ア) 1単語のもの: not (17) also(13) well(10) clearly(9) now(9) never(8) then(8) generally(8) often(8) first(7) sometimes(7) only(7) usually(7) always(7) thus(6) still(6) commonly(5) all(5) already(5) therefore(5) so(5)

イ) 2単語のもの: no longer(5)

次にその他の副詞について分野別に列挙する。ア)は1単語イ)は2単語を意味する。

現在形

A) 報道 (ルポルタージュ)

ア) currently, dead, equally, ever, evidently, heavily, itself, liberally, more, partly, privately, twice, variously, violently, well

イ) better than, in turn, not yet, now universally, the biggest, very heavily

B) 報道 (社説)

ア) far, invariably, properly, quite

イ) more thickly, no longer, not always, obviously not, of necessity

C) 報道 (評論)

ア) strongly, capitally, fastidiously, more, otherwise, well

イ) at first, most prettily, not entirely, not fully

D) 宗教

ア) frequently, hardly, later

イ) fairly closely, sometimes actually

E) 技能、手仕事、趣味

ア) any, attractively, best, easily, conveniently, especially, firmly, fully, happily, merely, more, normally, properly, readily, repetively, seemingly, seriously, similarly, specially, widely

イ) also closely, but rarely, either not, merely happily, no

過去形

ア) again, breathlessly, closely, each, finally, frequently, highly, later, officially, originally, probably, really, well

イ) no condition

ア) contemptuously, openly, recently, passionately, shockingly, slightly, widely

イ) also widely

ア) again, eagerly, entirely, greatly, excellently, incessantly, quickly

ア) consequently, easily, quite, really, therefore

イ) themselves equally, never actually, not wholly

ア) again, carefully, closely, eventually, finally, normally, originally, reasonably, recently, regularly, somehow, well

イ) not well

longer, not actually, by now, not  
always, not yet

F) 伝記

ア) badly, certainly, chiefly,  
continually, delightfully, equally,  
fervently, locally, newly,  
obviously, picturesquely, simple,  
soon, specially, quickly, strongly,  
tightly, today, variously

イ) now well, no longer, very  
lightly, at present, far easier,  
hardly ever, less often, much  
better

ア) assuredly, completely, each,  
finally, gladly, heavily,  
immediately individually,  
inexorably, much, nearly, newly,  
swiftly, temporarily, undoutedly,  
usually

イ) all but, not already, not  
openly, so widely, so beautifully

G) 純文学、伝記、随筆

ア) aesthetically, basically, best,  
better, fearfully, frequently,  
henceforth, here, highly, legally,  
more, most, normally, nowadays,  
piously, politically, probably,  
rarely, seldom, solely, soon,  
virtually, widely

イ) not normally, no longer, not  
directly, so successfully

ア) accordingly, almost,  
consistently, deeply, diametrically,  
easily, efficiently, either, even,  
eventually, ever, faithfully,  
favorably, finally, fully,  
hurriedly, immediately, largely,  
later, emphatically, warmly,  
naturally, nearly, obviously,  
officially, once, partially,  
persistently, quickly, quietly,  
really, recently, severely, soon,  
sorely, strongly, subsequently,  
too, ultimately, well

イ) even less, as yet, clearly not,  
in 1959, not only, not very, very  
well

H) その他 (政府書類、学生便覧、工業報告書、大学紹介、社内報)

ア) almost, extremely, gratefully,  
highly, much, normally,  
properly rarely, rigorously,  
significantly, strongly,

ア) better, primarily, recently,  
widely

traditionally, virtually

イ) not substantially, of course,  
so closely, not now, not wholly

J) 学術書、科学書

ア) abruptly, actually, amicably,  
apparently, as, automatically,  
best, chiefly, constantly,  
currently, directly, distinctly,  
either, evidently, fairly, falsely,  
formally, freely, fully, here,  
largely, manually, occasionally,  
rarely, seldom, simply, small,  
sound, thoroughly, virtually,  
widely

イ) not all, so lamentably, so  
well, most easily, much more,  
only partly, still not, very well

ア) accidentally, again, amply,  
best, carefully, directly,  
eventually, evidently, expressly,  
finally, greatly, naturally,  
obviously, originally, partially,  
probably, properly, quickly,  
quite, randomly, sharply,  
subsequently, tentatively, widely  
イ) in fact, not all, not closely,  
at length, no longer, not  
significantly, of course, the  
lowest

また、KP 群は少数なので一括して列挙する。

ア) all, almost, already, always, clearly, easily, entirely, ever,  
generally, hardly, hastily, highly, immaculately, jealously, largely,  
literally, never, not, now, often, openly, quickly, quite, rather,  
really, rudely, scientifically, soon, so, still, finally, violently, well,  
wholly

イ) also widely, not only, too deeply, out only, not even

6-3 代名詞+of+名詞

代表的な代名詞+of の句を検索し、表6に示す。

計の数値は総代名詞の各分野における頻度を表している。前述のAJ群とKP群の群分けを参照して比較するとそれぞれの平均値は108.9と52.4であり、このことよりAJ群のほうがKP群より使用頻度が高い傾向にある。

また、特に many に同様の傾向が強く、他に each, most, much, some にもその傾向が現れている。

表6 代名詞+of+名詞

上段 出現度数×0.00001

: 数値の2は10万語中に2回の出現率を表す

: /は無しを示す

下段 (出現個数

: 実際の出現頻度を表す

	A	B	C	D	E	F	G	H	J	K	L	M	N	P
all	2 (2	/	/	8 (2	7 (6	7 (6	6 (19	6 (4	5 (10	8 (6	6 (4	6 (1	7 (5	4 (3
another	1 (1	2 (1	4	/	4 (4	1 (1	1 (2	/	0 (1	/	/	/	1 (1	1 (1
any	2 (2	6 (4	4	8 (2	3 (3	8 (7	4 (8	11 (8	5 (9	5 (/	3 (2	6 (1	4 (3	4 (3
each	2 (2	/	/	21 (5	3 (3	8 (7	1 (2	13 (9	15 (29	/	/	/	/	/
enough	1 (1	/	/	4 (1	/	8 (7	1 (2	3 (2	0 (1	/	/	/	/	/
few	/	2 (1	/	/	4 (4	3 (3	3 (5	3 (2	0 (1	3 (2	/	/	/	/
half	/	3 (2	/	/	1 (1	1 (1	1 (3	4 (/	1 (1	1 (/	6 (1	/	1 (1	
last	/	2 (1	/	/	1 (1	7 (6	1 (1	/	1 (2	1 (1	3 (2	/	1 (1	/
least	1 (1	/	/	/	/	/	/	/	/	/	/	/	/	/
less	/	/	/	/	/	/	/	/	0 (1	/	/	6 (1	/	/
little	2 (2	/	/	4 (1	/	1 (1	2 (4	/	/	3 (2	/	/	3 (2	4 (3



many	12	20	4	16	12	13	15	17	12	/	5	/	/	/
	(13	(13	(1	(4	(11	(12	(28	(12	(22		(3			
more	2	3	/	4	4	3	1	11	3	3	/	6	3	3
	(2	(2		(1	(4	(3	(2	(8	(6	(2		(1	(2	(2
most	20	11	25	4	19	20	19	14	16	5	8	6	13	10
	(23	(7	(6	(1	(17	(18	(35	(10	(30	(4	(5	(1	(10	(8
much	10	3	21	12	7	3	12	9	5	4	6	/	4	4
	(11	(2	(5	(3	(6	(3	(22	(6	(10	(3	(4		(3	(3
other	/	/	/	/	1	3	1	/	/	1	/	/	/	/
					(1	(3	(2			(1				
several	/	3	/	/	/	/	2	2	2	3	/	/	/	/
		(2					(4	(1	(3	(2				
some	21	21	16	16	36	27	31	31	25	13	6	6	9	16
	(23	(13	(4	(4	(33	(24	(56	(22	(49	(10	(4	(1	(7	(12
that	5	/	37	45	1	13	13	13	24	5	3	13	4	1
	(5		(9	(11	(1	(12	(34	(9	(46	(4	(2	(2	(3	(1
this	/	/	/	/	/	/	/	/	/	/	/	/	1	/
													(1	
計	83	75	110	140	110	92	120	134	116	61	44	58	50	49
	A	B	C	D	E	F	G	H	J	K	L	M	N	P

#### 6-4 形容詞+of

次の内容について表7に示す。

- ア) 形容詞が各分野の総単語数に対して使用される割合
- イ) 形容詞+of+名詞の構造を持つ形容詞句が各分野の総単語数に占める割合
- ウ) 形容詞+of+名詞の構造を持つ形容詞句が総形容詞に占める割合
- エ) it+be 動詞+形容詞+名詞の構文の総形容詞に占める割合

表7 形容詞の出現率

- ア) 上段 出現度数は $\times 0.0001$   
: 数値の446は1万語中に446回の出現率を表す
- イ) 上段 出現度数は $\times 0.00001$   
: 数値の25は10万語中に25回の出現率を表す  
下段 (出現個数)  
: 実際の出現頻度を表す
- ウ) 上段 出現度数は $\times 0.0001$   
: 数値の56は1万語中の形容詞中に56回の出現率を表す
- エ) (数値は実際の頻度を表す)

	A	B	C	D	E	F	G	H	J	K	L	M	N	P
ア)	446	530	675	473	583	536	572	590	321	426	403	493	387	380
イ)	25	47	33	38	48	49	44	33	32	58	40	77	43	59
	(27	(30	(8	(9	(43	(52	(81	(23	(61	(43	(25	(12	(33	(45
ウ)	56	88	49	78	82	92	78	56	100	137	100	157	112	156
エ)	/	/	/	/	/	/	(1	/	/	(2	(1	(1	(2	(6

形容詞に関して言えることは、すべての分野ではほぼ同じ頻度で形容詞が使われていることがア)の数値より分かる。また、形容詞句についても同様の傾向がイ)より分かる。しかし、ウ)より形容詞中での形容詞+ofの割合はKP群の方が高い傾向にある。これは、前置詞句をつくる前置詞to, for, of, fromの内、ofがKP群によく出てくることを意味している。

エ)については実際にはAJ群では1個、KP群では12個であり次に具体例をすべて示す。

G) 純文学, 伝記, 随筆

'It was nice of him to say so,' I replied unhappily,

K) 小説全般

He picked her up somewhere. 'That sounds very kind of him,'

'Nice of you to have dropped in,' said Dad with terrible sarcasm.

L) 怪奇小説, 推理小説

'That is very nice of you.' 'Nice, nothing .....

M) 科学小説

'Stupid of me!'

N) 冒険小説, 西部小説

it was the more awful of the pair.

'Sorry, that was silly of me,' she said.

P) 空想小説, 恋愛小説

'Most kind of you to ask me.'

'That is not very charitable of you, Kay.'

it is awfully decent of you to care so much about me, Kay.

But I must say I think it's rather mean of you,

it was very rude of me.'

'That is sweet of you to say that,'

また, of が後続するがウ), エ) のタイプのいずれにもあてはまらない文体としては the most+形容詞の原級+of の形が若干あったがこれらは検索後, 除外した。それら 2 例を示す。

(1)

new yoghurt department, equipped with new plant, the most important of which is a large new automatically controlled yoghurt incubator.

(2)

Most important of all they believe in themselves, their company and their industry.

## 7 結果の考察

今回の一連の解析において, 分野別に結果を比較する試みはある意味では的をのべていたと言える。

解析結果 6-1, 受動態の結果において, AJ 群, KP 群と大きく二つに

群分けして傾向を述べた。これは、AJ群が論説文などの比較的文語体に近い文章の構成であるのに対して、KP群は小説等が中心であり、口語文に近い文章構成であることによるという考えからである。

受動態に関しては、受動態の頻度がKP群においてAJ群よりも低く、仮説は立証された。また、KP群では過去形の現れ方が他の現在形、現在分詞、過去分詞、助動詞形よりも多い。ただし、AJ群のなかでもA（ルポルタージュ）とG（純文学、伝記、随筆）は過去形の出現率ではKJ群に近い傾向を示している。すなわち、より口語的な文章であることが分かる。

次にこの群分けで比較できるものとして、代名詞+ofの出現率がある。ここでは代名詞+ofはAJ群の方がKP群より多く使われる傾向にあり、特に文語体によく現れる代名詞も判明した。また、形容詞+ofの句では口語体によく現れる傾向を示した。ただし、単独の形容詞の頻度に注目すると群別の差異が全く現れないことが言える。

いずれにしても、口語体、文語体では文体に明らかに比較対照できる差異があることが膨大なデータベースを解析することにより判明したわけである。

また、その他の傾向として注目すべき点は、今回の解析結果でそれぞれの項目に現れる頻度がほぼ一貫した値を示している点にある。

たとえば、受動態の出現率に関しては文語体であれば、千語中に10回前後、口語体であれば5回以下ということが言える。

また、受動態での副詞の現れる割合も群別に明らかに一定している。代名詞+of+名詞の構造の名詞句においても、少ない頻度ではあるが群別に一定している。形容詞+of+名詞の構造の形容詞句についても同様である。このような事実が判明したことにより今後研究を進めるなかでさまざまな解析のヒントを得ることができる。

## 8 まとめ

今回の解析の主旨は句レベルでの解析をいかに行うかであった。解析におけるハード的、ソフト的な解説は本報告では最小限にとどめているが、解析を句レベルにおいて、大きく3項目の句構造についてと1項目の句構

1993. 2 タグ付きコーパスの句構造の解析について (古橋・高橋) 145 (657)  
造に現れる副詞について、貴重な数値を得ることができた。特に、各分野が口語的、文語的と大きく2つにデータ上で区別することができたのは大変有意義であった。そして、それぞれの文体で副詞の出現頻度、代名詞+of+名詞の構造の名詞句の、出現頻度、及び形容詞+of+名詞の構造の形容詞句の出現頻度に一貫性があることも分かり、今後この方面は利用価値が特に高いと思われる。

また、解析のソフト的な面も大多数割愛したが句レベルの解析がこのように容易にできたのはコーパスがタグ付きであったことによるもので、このコーパスが大変先進的な英語文章のデータベースであることが実証されたわけである。さらに文章解析プログラムである OCP が解析を最小限の労力にしてくれたこともあった。いずれにせよ、この組み合わせにより今後益々解析を発展させていくことが可能であると思える。

#### 注

- (1) 長瀬真理, 西村弘之 1986 コンピュータによる文章解析入門— OCP への招待—, オーム社, p.2.
- (2) 実際には入手したソースデータベースは圧縮ファイルの形式であったために後に解凍し、テキストファイルとした。そのため、厳密には加工データベースと言うのが正しい。
- (3) カテゴリーの I は存在しない。また O はマニュアルには解説があるが実際にはまだデータベース化されていない。
- (4) Stig Johansson 1986 THE TAGGED LOB CORPUS —Users' Manual—, Norwegian Computing Centre for the Humanities, Bergen, pp. 141–149.
- (5) 学校文法では前位限定詞と呼ぶ。
- (6) next, own, very は学校文法では後位限定詞には入らない。
- (7) 安井稔 1978 新しい聞き手の文法, 大修館, p.19.
- (8) 福地肇 1985 談話の構造, 大修館, p.66.
- (9) 池内正幸 1985 名詞句の限定表現, 大修館, p.161.
- (10) Micro-COP によりここで目的とするプログラムは4, 5行の非常に簡単なもので済む。また、検索にワイルドカードも使用できるのも特徴である。

#### 参考文献

- (1) Oxford University Press 長瀬真理訳, 1988 Micro-OCP 文章解析プロ

- グラム, 沖田電子技研.
- (2) 竹蓋幸夫 1981 コンピュータの見た現代英語—ボキャブラリーの科学—, エデュカ株式会社.
  - (3) 高橋薫 1991 英語教育のパソコン利用における諸問題, 計測自動制御学会, 教育工学論文集.
  - (4) 斉藤武生, 鈴木英一 1984 冠詞・形容詞・副詞, 研究社.
  - (5) 高橋薫, 古橋聰 1992 タグ付きコーパスの文章解析について, 中部地区英語教育学会口頭発表.
  - (6) 三浦敏明 1991 英語副詞の研究—副詞の多様性—, 文化書房博文社.
  - (7) 渡辺登士 1989 英語の語法研究・十章, 大修館.
  - (8) STIG JOHANSSON 1982 "Computer Corpora in English Language Research", Norwegian Computing Centre for the Humanities, Bergen.
  - (9) STIG JOHANSSON AND KNUT HOF LAND. 1990 Frequency Analysis of English Vocabulary and Grammar, Based on the LOB Corpus, VOLUME 1: Tag Frequencies and Word Frequencies, CLARENDON PRESS, OXFORD.
  - (10) STIG JOHANSSON AND KNUT HOF LAND. 1989 Frequency Analysis of English Vocabulary and Grammar, Based on the LOB Corpus. Volume 2: Tag Combinations and Word Combinations, CLARENDON PRESS, OXFORD.
  - (11) STIG JOHANSSON. 1978-1979 ICAME NEWS Nos. 1, 2, 3.
  - (12) ICAME Journal 1980-1992 Computers in English Linguistics, Nos. 4-16, Norwegian Computing Centre for the Humanities, Bergen.
  - (13) KAORU TAKAHASHI. 1982 Problem Areas in the Tagged LOB Corpus, Journal of Toyota College of Technology.

#### 謝辞

本論文においてコンピュータのハード面に関わる技術的な援助については豊橋技術科学大学の草間晴幸氏に, また文法に関しての助言を南山短期大学の堀部憲夫氏に頂き, 記して謝意を表する次第です。

なお, 各種の資料提供については Norwegian Computing Centre for the Humanities の Torill Revheim 氏, 各種の情報については The University of Bergen の Arne Svindland 氏, 及び Kjell Morland 氏に御尽力を頂いた。改めて深甚なる謝意を表する次第です。