

LDOCE 上の分野分けを用いた Pictorial Dictionary に基づく

3H-6

場面情報の解析\*

羽柴 正輝, 角田 達彦, 長尾 眞†  
京都大学 工学部

1 はじめに

構文的・意味的曖昧性を減らす文脈情報の一つとして場面が有効であるが、その性質はいまのところ未知である。空間的な連想をフレームにしたものと考えられるが、物体が各場面に登場する理由づけを行なった例はない。そこで本稿では、そのような場面の知識源として OXFORD-DUDEN の Pictorial English Dictionary (OPED) を用い、各場面に現れる単語の意味の特徴を Longman Dictionary of Contemporary English (LDOCE) の機械可読版を用いて明らかにする。機械可読版には各語義に分野を示すコードが付与されているため、OPED の各場面に現れる単語の語義のうち人間が判断した語義に対応する分野コードを収集し、場面ごとの偏り/広がりについて分析を行なった。

2 LDOCE の分野コードによる OPED の解析

場面情報の知識源として OXFORD-DUDEN の Pictorial English Dictionary (OPED) を用いる。これは、Kitchen などの日常生活に出てくる場面を 384 カテゴリに分け、絵に描き、登場する物の名前を対応づけて列挙してある辞書であり、言語では明示的には通常扱われない常識の一部を含んでいる。

OPED の形態を、Kitchen の場面を取り上げて、図 1 に示す。

OPED の各場面は以下のように分類できる。

1. 上位下位関係 - 種類
2. 部分全体関係 - 単体説明 情景 行為描写 模式図
3. 組成構成物関係 - 材料 組成
4. 目的手段関係 - 道具 職場

今回は部分全体関係の中の情景に絞って解析を行なった。

OPED の解析には、LONGMAN 社の Longman Dictionary of Contemporary English (LDOCE) の機械可読版を用いる。LDOCE のこの版では、各単語の語義に分野 (Subject) が与えられている。この分野には主分野 (Main Subject) と、主分野をさらに細かく分けた副分野 (Sub Division) の二つがある。主分野は 100 個程度、副分野は 300 個程度ある。本論文では、とくに断りがなければ、分野は主分野のことを指す。分野の一例を表 1 に示す。

次に、LDOCE で単語を索いた場合の一例を示す。

dish n(名詞)  
1 a large flat (often round or OVAL) vessel from which food is put onto people's plates [CE]  
2 (an amount of) cooked food of one kind [FO]

\* Analysis of Scene Information based on Pictorial Dictionary by LDOCE's Subject Code

† Masateru HASHIBA, Tatsuhiko TSUNODA, Makoto NAGAO

Faculty of Engineering, Kyoto University  
{hashiba,tsunoda,nagao}@kuee.kyoto-u.ac.jp

表 1: LDOCE の各単語に与えられる分野の一例

分野	意味	分野	意味
AF	芸術	BE	化粧
BO	植物	BZ	商業
CE	陶磁器	CL	衣服
FO	食糧	FU	家具
HH	家の物	HR	時間
EG	工学	RE	記録
TE	織物	TO	煙草

[ ] で囲まれたものが分野を表わしており、上の例では、dish の 1 の語義の器には CE (CERAMICS) が、2 の語義の食べ物には FO (FOOD) が割り当てられている。しかし、すべての語義に分野が与えられているわけではない。

本実験では、OPED の各場面にでてくる単語に対して、人手で最も適切な語義を選択し、この語義に与えられている分野を調べる。OPED の各場面にでてくる単語を、次の二つに分けて考える。

(1) 場面に登場するもの自体を表わす名詞

(2) (1) を修飾する名詞

これ以外に、(1) を修飾する形容詞もあるが、数が少ないため、本論文では扱わない。その分野が何回現れたかを (1)(2) について各々数え、それを表にし、解析結果とする。この解析結果から OPED の各場面の特徴を考察する。OPED の各場面にでてくる単語に対して、その場面に直接関係がないと思われる単語 (例えば前置詞など) は数えなかった。単語は、同一単語の同一語義のものであっても、その単語が出た回数だけ、その単語の語義に与えられている分野を数える。これは、同一単語の同一語義のものが多く現れることが、その場面の特徴を顕著に表わすからである。また、OPED の各場面にでてくる単語の原型化は人手で行なった。

3 解析結果と考察

OPED の中から選んだ 7 つの場面 Kitchen (台所), Hall (玄関の間), Living room (居間), Bed room (寝室), Dining room (食堂), Bank (銀行), Restaurant (レストラン) について、LDOCE を用いて解析を行なった。

この結果を表 2 に示す。一番上の行の枠は、各場面を表わし、一番左の列の枠は、各場面に現れた分野の総計の順位を表わす。分野の総計が 2 つ以上の分野について扱ったが、Restaurant の場面だけは表の関係上、3 つ以上の分野についてのみ扱っている。その他の枠は四つに分かれており、左上が分野、右上が意味を表わす。また、左下が場面に登場するもの自体を表わす名詞に与えられた分野の数、右下がその名詞を修飾する名詞に与えられた分野の数を表わす。

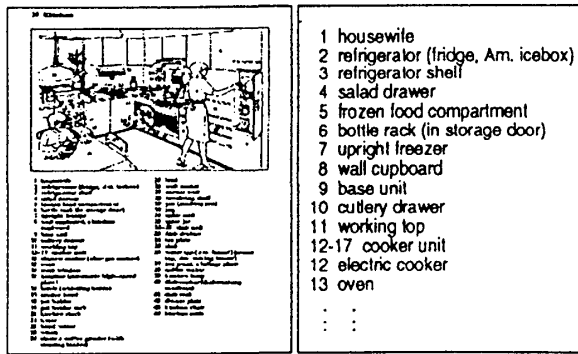


図 1: (左) OPED での台所の図版, (右) テキスト分析部分

実験に用いた 7 つの場面の基礎データを表 3 に示す。a は、場面に登場するもの自体を表わす名詞と、その名詞を修飾する名詞の個数を足した数である。b は、a の中で分野が与えられている単語の個数で、前にも述べたように、分野が与えられていない単語もある。c は、b の単語に分野を与えた総数で、b で示された数以上なのは、1 つの単語の語義に 2 つの分野が与えられることもあるからである（例単語 cupboard のある語義には、HH, FO の二つの分野が与えられている）。

表 2 から、各場面は LDOCE の単一の分野だけでは表わすことができず複数の分野にまたがって表わされること、すなわち、分野の広がりを持っていることが分かる。分野の広がりとは、一つの場面が様々な概念によって表現されていることを意味する。人間がある場面を思い浮かべたときに出てくる事実、すなわち一般的な常識もこの中に含まれる場合がある。

また、各場面は表わされる分野によって分類できるものである。このことを、数量化三類と呼ばれる分類法を用いて、7 つの場面の解析を行なうことで示す。この分類法によって、似ている場面は互いに近くの座標に現れる。この結果を図 2 に示す。この図から、x 軸方向の値を見ると、1, 2, 3, 4, 5, 6 と固まっており、7 だけこの集団から孤立している。この結果は表 2 の結果と矛盾しないものである。

#### 4 おわりに

単語の語義の曖昧性を減らす文脈情報としての場面情報を与える辞書として OPED というものがあり、その解析を LDOCE の分野を用いて行なった。今後の課題として

- (1) 本実験の解析結果を多義性解消にどのように役立てられるのかの検討
- (2) 分野の自動付加の問題
- (3) 分野を用いた場面の自動推定の問題が挙げられる。

謝辞: OPED の場面に対する LDOCE の分野の割り当てに対してお手伝い下さった京都大学工学研究科電子通信工学専攻の加藤哲哉君にお礼を申し上げます。

#### 参考文献

- [1] 角田達彦, 田中英彦 場面情報に基づく英語名詞の語義の優先づけ方法と評価. 情報処理学会第 50 回全国大会, vol.3, pp83-84, 1995.
- [2] Oxford University Press. The OXFORD-DUDEN Pictorial English Dictionary. 日本貿易株式会社, 1981.
- [3] Longman Group Ltd. Longman Dictionary of Contemporary English, 1978.

表 2: 七つの場面の解析結果: ア 道具 イ 職業 ウ 加工対象 エ 建物 オ 目的 カ 場所

	Restaurant		Kitchen		Bed-room		Living-room	
1	FO	オ	HH	ア、イ	HH	ア	HH	ア
	17 個	11 個	18 個	10 個	9 個	2 個	9 個	2 個
2	HH	ア	FO	ア、ウ	FU	ア	FU	ア
	18 個	7 個	6 個	7 個	5 個	1 個	6 個	0 個
3	BV	オ	FU	ア	TE	ア	BV	オ
	1 個	6 個	6 個	0 個	3 個	0 個	0 個	3 個
4	BW	オ	BV	ウ	BE	ア	RE	ア
	0 個	6 個	0 個	3 個	0 個	3 個	1 個	1 個
5	CL	ア	HR	ア	CL	ア	TO	ア
	1 個	3 個	2 個	0 個	0 個	3 個	1 個	1 個
6	BD	エ	CE	ア	HR	ア		
	2 個	1 個	0 個	2 個	1 個	1 個		
7	FU	ア	EG	ア	AF	ア		
	3 個	0 個	1 個	1 個	1 個	1 個		
	Hall		Dining-room		Bank			
1	CL	ア	HH	ア	BZ	ア、カ		
	6 個	5 個	14 個	4 個	6 個	3 個		
2	HH	ア	FU	ア	ON	イ		
	6 個	3 個	7 個	2 個	2 個	1 個		
3	BD	オ	FO	オ				
	0 個	2 個	1 個	5 個				
4	EG	ア	BV	オ				
	0 個	2 個	0 個	4 個				

表 3: 七つの場面の基礎データ: a 場面に関係する全単語数 b 分野が与えられた単語数 c 与えられた分野の延べ数

	Restaurant	Kitchen	Dining-room	Living-room
a	128 個	68 個	52 個	45 個
b	87 個	53 個	39 個	34 個
c	90 個	58 個	39 個	34 個
	Hall	Bed-room	Bank	
a	43 個	38 個	34 個	
b	31 個	30 個	17 個	
c	31 個	31 個	17 個	

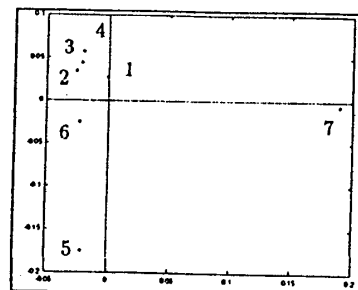


図 2: 数量化三類による場面の解析結果: 1 Restaurant 2 Kitchen 3 Dining-room 4 Living-room 5 Hall 6 Bed-room 7 Bank