

3T-10 前置詞の意味に着目した名詞分類の試み

石井 康毅

東京外国語大学 外国語学部

1 メタファ表現と名詞の分類

既に語彙分類データが数多く作成されている [1]. それらのデータは語彙規模も大きく, 言語を分析する上で有用な情報を提供してくれる. 一般に, 語の分類は分析者の分類視点に基づくため, 語彙数が多くなると細部の語の分類では, 分析者の言語直観の影響が大きいことが予想される. したがって, (1) 言語の運用体系全体から得られる分類視点を抽出すること, (2) 語の分類作業過程の効率化 (自動化) を行うことが必要である.

コーパス言語学 [2] の分野の成果が示すように, 言語の運用体系の記録であるコーパスの分析により, 語と語の新たな関係が明らかになる. 例えば, 例文 (1)a~c の *sight*, *way*, *field* が示すように, 一般的な語彙の意味分類では十分には記述されない関係がある.

- (1) a. I have him *in* sight.
- b. I can't see him — the tree is *in* the way.
- c. That's *in* the center of my *field* of vision.

上記の例文に表れる前置詞表現の前置詞と名詞の関係は, いわゆるメタファ (metaphor) を通じた理解を必要とする. メタファは人間の根本的な概念構造と結びつき, 人間の思考に欠かすことのできないものであり [3], 語彙の「意味」の大きなサブセットを構成している. 一般の語彙分類は名詞や動詞などの概念語を中心にした分類だが, 上記のような例では, 機能語を中心にした分析が必要になる.

著者はメタファ表現の意味を説明する語彙分類モデルの構築を目標として, 特に機能語である前置詞と名詞から構成される句に着目した名詞分類システムの構築を目指している. 第一段階として, 本稿では, コーパスデータから収集した前置詞とその補語 (名詞) の結びつきを基準にした名詞分類の分析の結果を報告する.

1.1 メタファ

メタファとは, ある事柄をほかの事柄を通して理解し, 経験することであり, 単なる表層的な言語現象ではなく, 人間の概念体系の本質を規定するものであり [3], 言語運用のみならず, 思考にも欠かすことができない. 例文 (1)a~c は, VISUAL FIELDS ARE CONTAINERS というメタファによって理解される. この基本的な文を理解するにもメタファが必要である. 語の「通常の意味」[4] を中心に理解を試みる英語学習者にとって, 成句と言えるほどには固定的になっていない (辞書に載っていない) 表現を理解するためには, メタファの理解が重要である. メタファは類似性をつくりだす [4] ことから文学表現でも多用される.

1.2 容器のメタファ

例文 (1)a~c の理解には, 容器のメタファ (Container Metaphors) (ここで言う「メタファ」はより正確には「イメー

ジ・スキーマ」(image schema)[5] と呼ばれる) が必要である. 容器のメタファは, 先行研究の結果, ある程度確立されている概念である. 人間は, 自然界の物理的境界が存在しないところにも, 境界を想定し, 内側と境界面を持つような領域に区分する (容器を規定する)[3]. これは人間の最も基本的な習性のひとつであるという.

2 分析

英語は容器のメタファをしばしば前置詞によって表現する. 典型的な前置詞 (相当句) は, *in*, *into*, *out*, *out of* 等である. これら前置詞の補語になる名詞は, 容器とみなすことができ, 共通の意味素性を共有する範疇を形成するはずである. 本稿では, 前置詞相当句 *out of* の補語の調査を行った.

2.1 分析の手順

コーパスデータから, 前置詞 (前置詞相当句 *out of*) の補語を収集し, 補語部分の名詞の性質を検討する. 分析手順を挙げる.

1. 前処理 タグを削除 (タグ付きコーパスの場合) し, 1 文/行のデータ形式にする.
2. 分析 文中に前置詞句 (本実験では *out of* の句) がある場合, 構文解析を行い補語を特定する. また, 全ての文から名詞の出現頻度を計算する.
3. 検定 分析結果の検定を行う.

コンコーダンスには数多くのものがある [2]. しかしコンコーダンスは, キーワードの前後数語を対象にした語彙の頻度調査はできるが, 統語構造は考慮しない. 本実験では, 前置詞句の補語の抽出のために統語解析を行った. 構文解析処理には, 出力結果の処理の容易さ, 辞書が付属していること, フリーで利用できるものという観点から Satoshi Sekine 氏の Apple Pie Parser (version 5.9) ¹ を利用した.

2.2 コーパス

アメリカ英語の典型的なコーパスである Brown Corpus (語数約 100 万語) とイギリス英語の典型的なコーパスである LOB Corpus (語数約 100 万語) を用いた. このうち, 名詞は約 50 万語である.

3 実験

3.1 コーパスからのタグ除去

本実験で利用した 2 つのコーパスは, (1) 行頭にヘッダ情報を含み, (2) 一定の文字数を超えると強制的に改行され, (3) 文中にもタグを含む形式である. このため, Perl スクリプトでタグを除去した.

An Experiment in Classifying Nouns Based on Metaphorical Meanings – A Case Study of Prepositional Phrases of *out of* – ISHII, Yasutake
Faculty of Foreign Studies, Tokyo University of Foreign Studies,
4-51-21 Nishigahara, Kita-ku, Tokyo 114-8580 JAPAN

¹ <http://cs.nyu.edu/cs/projects/proteus/app/>. 構文解析処理は 100 % の解析率でなく, 文が長いほど精度が低下することが Apple Pie Parser のマニュアルにも記載されている. しかし, 句単位の解析は精度が保たれており, 本実験の利用目的には問題ないと判断した.

3.2 分析処理

前処理したデータを Apple Pie Parser で分析し、結果を Perl スクリプトでフィルタリングし、*out of* の補語の名詞 (句) を抽出し集計した。尚、現段階のプログラムは単数形名詞と複数形名詞を区別するため、単複の語形の同一視作業は手作業で行なった。

3.3 結果の統計的有意性

out of の補語の名詞の頻度と、それらの名詞のコーパス全体における頻度から、式 (1) を用いて差異係数 (difference coefficient) を求める。なお、ここで言う頻度とは、特定の名詞の全名詞中 (あるいは *out of* の全補語中) に占める割合を指す。

$$\text{差異係数} = \frac{\text{Freq.}_A - \text{Freq.}_B}{\text{Freq.}_A + \text{Freq.}_B} \quad (1)$$

Freq: frequency

式 (1) の差異係数は-1.00 から+1.00 までの間の値をとり、正の値の場合には、その語が A の集合により多く出現し、負の値の場合にはその語が B の集合により多く出現することを示す。

次に、集計結果の度数に対して、式 (2) によりカイ 2 乗検定 (chi-square test) を行う。有意水準 0.01, 自由度 1, 集合の母数 ∞ の場合のカイ 2 乗値 6.63[6] よりも大きく、1 % レベルの統計的有意性が認められるものには、差異係数に α を添える (表 1 参照)。

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

O: observed frequency, E: expected frequency

表 1 には、*out of* の補語として出現した名詞² のうち、5 度以上の頻度で出現したものを挙げる。なお、プログラムによって認定された *out of* の補語は計 680 語であった。

4 評価

イディオム表現に用いられる名詞が多いのは当然の結果として、直観的には内側と境界面を持つような領域を表現する語が挙がっている。定量的には、どの名詞も差異係数が高く、全ての語で 1% レベルの有意性が認められることから、特定の語彙が *out of* の補語として分布していると認められる。

4.1 意味群別に分類した語彙分類との比較

Princeton University で開発された WordNet³ を利用して、表 1 に挙げた名詞の上位 (概念) 語を調べた。上位関係は語義によって異なるが、代表的な語義を選択し、最上位からトップダウンで共通の分類を探した。その結果を以下に示す。

window, pocket, bed, room, car, door, office: /entity, something/object, physical object/artifact, artefact
sight, mind: /psychological feature/cognition, knowledge
way, shape: /abstraction/attribute
town, country, bounds: /entity, something/object, physical object/location

メタファの観点からはグループ化できる語について、語彙分類データからは共通点項目が見出せない。

² 複数形の名詞も、複数形でイディオムとして用いられる bounds を除き、単数形としてカウントした。

³ <http://www.cogsci.princeton.edu/wn/>

語	<i>out of</i> 補語中	コーパス中	差異係数
sight	14	211	0.96 α
window	14	401	0.92 α
way	12	2110	0.61 α
pocket	11	158	0.96 α
hand	10	1490	0.66 α
house	10	1446	0.67 α
mind	10	831	0.80 α
bed	9	373	0.89 α
room	9	955	0.75 α
action	8	582	0.82 α
town	8	628	0.81 α
country	7	1102	0.65 α
car	6	715	0.72 α
date	6	307	0.87 α
income	6	240	0.90 α
bounds	5	18	0.99 α
business	5	628	0.71 α
door	5	738	0.66 α
office	5	597	0.72 α
question	5	811	0.64 α
shape	5	236	0.88 α

表 1: 語彙の出現度数と差異係数

4.2 今後の課題

他の前置詞について同様の調査を実施する予定である。また、MI-score や t-score など、他の尺度でも計測を実施し語彙のグループ化が可能か確認する。最終的には、クラスタリングを利用した、語の自動分類を行なうことを目標としている。

参考文献

- [1] 徳永健伸: 『言語と計算 5 情報検索と言語処理』, 東京大学出版会, 1999.
- [2] 齊藤俊雄・中村純作・赤野一郎編: 『英語コーパス言語学 - 基礎と実践 -』, 研究社, 1998.
- [3] Lakoff, G. and Johnson, M. (1980) *Metaphors We Live By*. Chicago: The University of Chicago Press.
- [4] 赤羽研三: 『言葉と意味を考える I 隠喩とイメージ』, 夏目書房, 1998.
- [5] Lakoff, G. (1987) *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.
- [6] 有馬哲・石村貞夫共著: 『多変量解析のはなし』, 東京図書, 1987.