

共起関係を用いた多義動詞格フレームの自動獲得\*

4M-01

片岡 浩一, 唐澤 博†

山梨大学 工学部‡

E-mail: {koichi, karasawa}@jewel.yamanashi.ac.jp

1 はじめに

日本語を解析する上で、未知語が出現した場合の対処方法が問題になることが多い。その中でも動詞の未知語が出現した場合、それに付随する情報が多いため、処理が難しくなる。そこで、未知語動詞が出現した場合でもその深層格を推定し、解析の処理が止まらないようにする。未知語が入力され、品詞推定によって動詞と判定された場合に、その意味推定を行い、その結果で文解析を続行する一方で、動詞エントリを追加しよう、というものである。

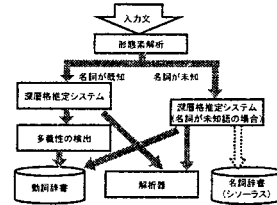


図 1: システムの概要の概要

2 自動獲得処理の概要

文が入力されると、まず形態素に分解する。その中に動詞以外の未知語が含まれない場合、名詞が既知語の場合の深層格推定システムによって深層格を推定し、その結果を用いて構文解析以降の処理を続行する。その一方で出現した動詞を辞書に問うるが、その時に動詞を語義ごとに分類して登録する必要があるため、多義性を検出し、語義ごとに分類する必要がある。このようにして辞書に登録した動詞を半既知語動詞と呼ぶ。また一方で、名詞が未知語の場合には、それまでに獲得した半既知語動詞の辞書を基に深層格を推定する。以上の処理の流れを図 1 に示す。図 1 における名詞が既知語の場合の「深層格推定システム」及び「多義性の検出」はそれぞれ [4], [5] の方法を用いることとする。

か、という仮定に基づいている。図 2 の例文で考えると、「近く」と共起しているものと、「に走る」と共起しているものを深層格ごとに分類し、どれが類似しているかを調べる、という手法 (図 3) を用いて推定する。

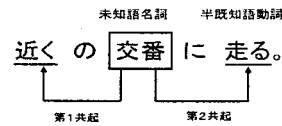


図 2: 推定文例

3 名詞が未知語の場合の深層格推定

多義性の検出及び分類によって辞書に登録していくと、その動詞についての事例も辞書中に蓄積されていく。名詞が未知語の場合には、その事例と、EDR 共起辞書を基に、その第 2 共起の深層格を推定する。これは、第 1 共起の係り側要素と、第 2 共起の動詞を固定すると、その深層格もある程度固定されるのではない

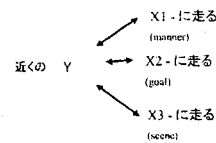


図 3: 推定方法

\*Automatic acquisition of polysemous verb case frames by co-occurrence relation information

†Koichi Kataoka, Hiroshi Karasawa

‡Yamanashi University, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

3.1 具体的な推定方法

説明の便宜上、図 2 の「近く」と「交番」の共起を第 1 共起、「交番」と「走る」の共起を第 2 共起と表記す

る。

1. 第2共起の係り側と共起しているものを探す。  
EDR共起辞書から第2共起の係り側と共起している語を全て抜き出す。
2. 第2共起の動詞と共起する語を探す。  
名詞が既知語の場合の深層格推定で半既知語となった動詞と共起し、かつ第2共起と同じ表層格を持つ語をEDR共起辞書から探す。

### 3. 概念距離の計算

受け側概念と係り側概念の間の距離を全ての組合せで計算する。この時、概念間の距離の計算には松永の方法[2]を用いる。

### 4. 概念距離に重みを付加

3で計算した概念距離に重みを付加する。重みの付け方として以下の2つの方法が考えられる。

- 自然対数を用いた計算方法  
式(1)を用いて計算する。概念距離が大きいものの影響を小さくする。

$$Dis_w = a \times \ln(Dis + 1) \quad (1)$$

$Dis_w$ : 重み付き概念距離

$Dis$ : 概念距離

a: 適当な実数を代入(この値によって順位は変動しない)

- シグモイド関数を用いた計算方法  
式(2)を用いて計算する。概念距離が大きいものはほぼ同じ値を取るようになり、自然対数関数よりさらに概念距離が大きいものの影響を小さくする。

$$Dis_w = \frac{b}{1 + e^{-Dis}} - \frac{b}{2} \quad (2)$$

b: 適当な実数を代入(この値によって順位は変動しない)

これらの関数を視覚的に比較すると図4のようになる。

### 5. 深層格ごとに平均

4で求めた重み付き概念距離を深層格ごとに平均を計算する。

## 4 評価

語の共起関係データ[3]を仮登録辞書とし、50件の入力文を作成し、含まれる動詞を推定し、表1の結果を得た。

## 5 問題点と考察

概念距離をそのまま用いるよりは重み付き概念距離を用いる方が良い結果が得られることがわかった。自然対数関数を適用した結果とシグモイド関数を適用した結果を比べると、シグモイド関数を適用した結果の方が高正解率であった。

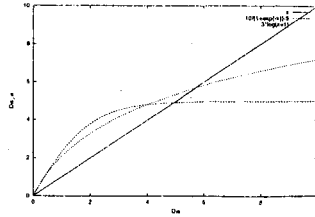


図4: 各関数の比較

表1: 名詞が未知語の場合の推定結果

		第1候補	第2候補	第3候補	第4位以下
概念距離	数	18	11	7	10
	確率	39.1%	23.9%	15.2%	21.7%
自然対数	数	24	9	5	8
	確率	52.2%	19.6%	10.9%	17.4%
シグモイド	数	30	7	4	5
	確率	65.2%	15.2%	8.7%	10.9%

## 6 課題

- 未知語名詞のシソーラスへの登録

現段階では、未知語名詞が出現した場合、その未知語名詞を登録していないが、これをシソーラスに登録することができれば、次回からは半既知語名詞として扱うことができるので、その方法を考案する。

- 評価実験

名詞が未知語の場合の深層格推定について、複数共起のものについても評価実験を行う。

## 参考文献

- [1] (株)日本電子化辞書研究所: EDR 電子化辞書, Version1.6
- [2] 松永, 田中: 文章に対する重要度付与処理における視点情報の有効性の分析, 情報処理学会第51回全国大会講演論文集(3), pp119-120, Sep1995
- [3] 田中: 語の共起関係データ
- [4] 片岡, 唐澤: 共起関係のみに基づく未知語動詞の格フレームの推定, 情報処理学会第60回全国大会講演論文集(2), 2ZA-03, pp.269-270 (2000.3)
- [5] 片岡, 唐澤: 未知語動詞の語義推定における多義性の検出, 情報処理学会第62回全国大会講演論文集(2), 8N-06, pp.321-322 (2001.3)