

# 発話行為タグ作成のための格フレーム辞書の利用

野澤元<sup>1</sup>・河原大輔<sup>1</sup>・横森大輔<sup>2</sup>・黒田航<sup>1</sup>

1. 情報通信研究機構・2. 京都大学大学院人間・環境学研究科

## 1. はじめに

対話コーパスに発話行為や談話機能のタグを付与するためのスキームは、海外では DRAMA(Carletta et al. 1997) や DAMSL(Allen and Mark 1997) が、また、国内においても「発話単位タグ」(荒木・他 1999) が提案されている。これに対して、野澤・他 (2007) は、文の意味を複数の状況フレームの統合体として記述する意味タグ付けの手法である、「複層意味フレーム分析 (MSFA)」(黒田・他 2004) を利用して、対話コーパスに発話行為タグを付与することを試みた。しかし、タグ付与の基準に関するいくつかの問題から、作業は容易には進まなかった。そこで、本稿では、発話行為タグの定義や付与基準の指標として、「格フレーム辞書」(河原・黒橋 2005) を利用する手法を提案する。

## 2. 発話行為タグをめぐる問題点

MSFA では、文の意味は状況水準のフレームを単位として記述されるが、その際、(i) 粒度の高い状況フレームの分類と、(ii) 状況フレーム間の関係性の記述を目指している。例えば、(1) と (2) は、共に「<捕らえる>」という状況を表しているが、より具体的には、前者では、漁師は魚を「<漁>」として捕獲しているのに対し、後者では、警官は容疑者を「<逮捕>」しており、これらは異なるフレームとして記述される («> は状況フレームを表す)。また、(1) では、「師」という語によって、「<漁>」がたまたまではなく、「<専門とする>」行為として行われたことが表されており、二つのフレームの関係性が記述される。

- (1) 漁師は魚を捕らえた。
- (2) 警官は容疑者を捕らえた。

しかし、MSFA によって対話コーパスに発話行為タグを付与する場合、この要件を満たすことはしばしば困難である。その最も大きな理由は、遂行される発話行為が明示的に言語化されることが少なく、多くの場合、暗示されるだけだからである。

- (3) A: やって頂けますか。

B: いいでしょう。

例えば、(3) の対話では、A の発話において英語の "request" に相当する働き掛け行為が遂行されているが、同様の行為を表す日本語のサ変名詞には、「依頼」、「要望」、「要求」、「要請」、「請求」などがある。しかし、この発話によって、どの行為が遂行されたのかは、明示的に表されておらず、「頂けますか」によって暗示されているだけである。日本語母語話者であれば、日常の文脈の中では、これらの語を躊躇なく使い分けることができるが、対話コーパスに発話行為タグを付与する場合には、特に行為の区分が明確でないために、前後文脈を参照しても判断が困難であることが多い。

また、B の発話では、「承諾」、「承知」、「受け入れ」、「同意」などと呼べる、ある種の応対行為が遂行されているが、通常この種の行為は、義務として何らかの「<履行>」を伴う。さらに、B に課せられた「<履行>」の内容は、A の「やって」が表す行為と一致しているはずであり、この対話を深く理解するためには、両者の関係性を記述する必要がある。しかし、あるフレームに対する関連フレームの種類と、それらの間の関係性が、通常どのようなものであるのかは自明ではなく、作業者は判断に困ることが多い。

MSFA では、タグの集合を事前に限定せず、作業者が必要に応じて新たなタグを追加するという手順を採用しているが、タグ付与作業を安定的かつ効率的に行うには、やはり、作業者の直感に頼るだけではなく、これらの行為を区別し、また、一般的な関連フレームとその関係性を示す、何らかの指標が必要であると考えられる。発話行為の分類については、いくつかの利用可能な先行研究がある。しかし、久保・他 (2002) による言語学的な分析は、小説や新聞の実例に基づいているものの、網羅性に乏しく、また、分類が内省のみに基づいているため、信頼性の点で満足できない。荒木・他 (1999) が提案する発話単位タグは、実験において一定の一致率を示しており、信頼性は保証されているが、例えば、「依頼」、「要望」、「要求」、「要請」、「請求」に相

当する行為は、全て「依頼」にまとめられているように、発話行為の分類の粒度が低い。

### 3. 格フレーム辞書の利用

#### 3.1. 格フレーム辞書

格フレームとは、用言とその格要素の共起パターンであり、今回使用する格フレーム辞書とは、そのような共起パターンを大量のコーパスの構文解析結果から抽出し、さらに類似性に基づいてクラスタリングするという手法によって、自動構築したデータである。この辞書にはいくつかの利用法があるが、例えば、名詞をキーワードとして検索すると、その名詞を項とする格フレームとその頻度の一覧を見ることができる。

#### 3.2. 発話行為の分類

格フレーム辞書で発話行為を表す名詞、例えば、「依頼」をキーワードとして検索すると、「受ける(ヲ格), 来る(ガ格), ある(ガ格), 出す(ヲ格), ...」のように、その名詞を項とする格フレームの一覧が出力される。また、異なる名詞をキーワードとすれば、異なる格フレームの一覧が出力されるが、概念的に類似した名詞の間では、結果的に、類似した格フレームの一覧が得られることになる。さらに、一覧のそれぞれの格フレームには、キーワードの名詞に対する頻度が付属して出力される。そのため、発話行為を表す複数の名詞をキーワードにして検索し、出力された格フレームを素性とし、その頻度を重み付けとしてクラスタリングを行うと、発話行為を分類することができる。

「依頼、希望、願望、嘆願、要望、要求、要請、請求、請願、申請」の10種類の名詞を分析したところ、図1に示すように、直感的に自然な分類を得ることができた。三つのクラスターに分類した場合、[A: 希望、願望、嘆願、請願]、[B: 要求、要請、要望]、[C: 申請、請求、依頼]となった。また、項となる格フレームの比較によって、発話行為の特性を知ることができる。例えば、Cは「受け付ける(ヲ格)」の項となるが、Bはそうでないことから、前者は、聞き手がその行為に対して開かれている必要があることがわかる。また、Bのうちの「要望」と「要求」では、前者のみが「頂く(ヲ格)」の項になることから、要望は聞き手にとって尊重すべき行為であることがわかる。

#### 3.3. 関連フレームの推定

検索によって得られた、特定の発話行為を表す名詞を項とする格フレームを、手作業で時系列的に配置すると、一種のスクリプトを作成することができる。例えば、聞き手から見た<依頼>についていえば、<依頼が来る>、<依頼を受け付ける>、<依頼を引き受ける>、<依頼に応える(依頼内容を遂行する)>といった流れがあることがわかる。このように、格フレーム辞書の情報を参照することで、<依頼>の関連フレームとして、最低限記述すべきだと思われる一連のフレームを推定することができる。

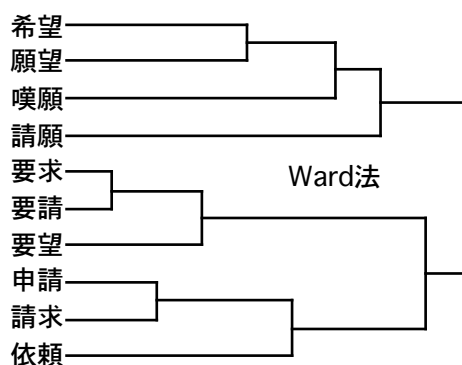


図1

#### 4. おわりに

この手法は発展段階であり、その精度は格フレーム辞書に直接的に依存している。また、発話行為の特性の分析には、主成分分析などの統計的手法を用いることも可能である。発話行為の分析は、これまでの内省的な手法に加え、心理実験的な手法も可能であるが、発話行為に対する判断は、自明ではない文脈に依存することが多く、実験の統制は容易ではないだろう。コーパスを用いた統計的手法の導入により、粒度の高い行為カテゴリーについての、一定の見通しを得ることが出来るのではないかと考えられる。

#### 参考文献

- 河原大輔・黒橋禎夫. 2005. 「格フレーム辞書の漸次的自動構築」, 『自然言語処理』, Vol.12(2), pp.109 - 131.
- 黒田航・他. 2004. 「日本語のための意味タグ体系を定義する試み: FrameNet の視点から」, 『自然言語処理第10回大会発表論文集』, pp.148-51.
- 野澤元・他. 2007. 「自然な対話文に対する複層意味フレーム分析の試み」, 『自然言語処理第13回大会発表論文集』, pp.250-253.