

震災・原発に関する言説空間の物語論的分析及び物語生成システムへの 応用—Web 等のテキストからの情報自動獲得機構の開発を中心として—

小方 孝 (ソフトウェア情報学部、教授)

<要旨>

本研究は、震災・原発に関する言説空間の物語論的分析を目標とするが、その一つの方法として、筆者らが開発中の物語生成システムによる言説シミュレーションを用いた検討を目標としている。今年度は、Web 等のテキストからの各種情報自動獲得機構の開発を中心に研究を進める。この成果は、将来的には、震災・原発に関連するテキストからの情報獲得に利用される。

1 研究の概要

本研究は、震災・原発に関する言説空間の物語論的分析を目標とするが、その一つの方法として、筆者らが開発中の物語生成システムによる言説シミュレーションを用いた検討を目標としている。今年度は、大量文書からの事象連鎖の半自動獲得及び自動獲得、獲得された知識の半自動構造化及び大量文書に基づく語彙の頻度情報及び語彙間の共起情報の自動獲得と利用を中心に研究を進める。この成果は、将来的には、震災・原発に関連するテキストからの情報獲得に利用される。

2 研究の内容

今年度の主要な研究内容は以下の通りである—

- 1) 大量文書からの事象連鎖の半自動獲得及び自動獲得、獲得された知識の半自動構造化：物語における文章を詳細に解析すると様々な要素からできていることが分かる。その中で、定型的な事象連鎖もしくはスクリプトは、認知心理学におけるスキーマ理論由来の概念として、従来から研究されて来た。本研究では、青空文庫などの小説や随筆から、事象連鎖知識を自動的に獲得し、これを統合物語生成システムにおいて利用可能な形式に半自動的に構造化する。それと共に、人間によるスクリプト入力やそれをベースとした拡張など、半自動的に事象連鎖獲得も行う。
- 2) 大量文書に基づく語彙の頻度情報及び語彙間の共起情報の自動獲得と利用：現在の統合物語生成システムでは大規模な概念辞書を利用しているが、物語生成に当たって使用する語彙の読みやすさ／読みにくさなどを制御するために、語彙の頻度及び共起情報を青空文庫などの小説を中心としたテキストにより計量し、その数値データにより物語生成における使用概念・語彙を制御・調整する研究を、前年度から引き続き行った。

3 これまで得られた研究の成果

本研究においては、主に以下の成果が得られた—

- 1) 大量文書からの事象連鎖の半自動獲得及び自動獲得、獲得された知識の半自動構造化：自動的もしくは半

自動的に取得された概念をもとに、統合物語生成システムにおいて利用可能な格構造形式に半自動的に変換するツールを作成し、各種実験を行った。また、青空文庫をテキストとして n-gram の方法で動詞の語彙を取得し、これを上記ツールを利用して格構造化することを試みている (途中)。さらに、一旦自動的もしくは半自動的に獲得された事象連鎖知識をさらに組み合わせてそのバリエーションを作り出す方法を考案している。

- 2) 大量文書に基づく語彙の頻度情報及び語彙間の共起情報の自動獲得と利用：統合物語生成システムにおいて利用する概念及び語彙の選択のために、語彙の頻度情報と語彙どうしの共起情報を用いる試みを前年度から継続した。

以上は、主に以下の論文として公表されている。

Ogata Takashi & Ono, Jumpei (2016). A Way for using the Verb Conceptual Dictionary in an Integrated Narrative Generation System: The Use of Co-occurrence Information on Verb Concepts, *Journal of Robotics, Networking and Artificial Life*. Vol. 3, Issue. 3, 209-212.

荒井達也・小野淳平・小方孝 (2016). 物語生成のための事象連鎖知識の半自動生成—統合物語生成システムにおける利用—. 『人工知能学会全国大会 (第 30 回) 予稿集』. 3P1-1in2.

照井和舎・小野淳平・小方孝 (2016). 語の共起情報による概念・単語選択の改善—統合物語生成システムにおける利用—. 『人工知能学会全国大会 (第 30 回) 予稿集』. 3P1-7in2.

吉田和樹・小野淳平・小方孝 (2016). 語の頻度情報による概念・単語選択の改善—統合物語生成システムにおける利用—. 『人工知能学会全国大会 (第 30 回) 予稿集』. 2P1-6in1.

Asakawa, S. & Ogata, T. (2017). Comparison Between Variational Autoencoder and Encoder-Decoder Models for Short Conversation. *Proc. of the 2017 international conference on artificial life and robotics*. 639-642.

荒井達也・小野淳平・小方孝 (2017). 統合物語生成システ

ムにおけるスクリプト型事象連鎖の獲得と利用. 『言語処理学会第23回年次大会 予稿集』. 703-706.

4 今後の具体的な展開

「大量文書からの事象連鎖の半自動獲得及び自動獲得、獲得された知識の半自動構造化」については、現在まだ獲得された事象連鎖知識のすべてを実際に使用できる状態になっていないため、格構造構成の効率化、意味的整合性の検討などによって、これを進める。「大量文書に基づく語彙の頻度情報及び語彙間の共起情報の自動獲得と利用」については、実際の使用実験による検証を続ける。予想としては、この方法に加えて概念辞書におけるカテゴリ分類の改善などを組み合わせることが必要になると考えられる。なお、震災・原発情報そのものについては、定点観測のための新聞と雑誌保存を続けており、次年度以降その情報を整理し、上記方法等により利用することが課題となっている。