

# シソーラス利用による童話文からの感情情報読み取りシステム

堂腰 裕明<sup>†</sup> 小山 聡<sup>‡</sup> 栗原 正仁<sup>‡</sup>

北海道大学 工学部<sup>†</sup> 北海道大学 大学院情報科学研究科<sup>‡</sup>

## 1 はじめに

コンピュータに人間の感情を読み取らせることを目的とした試みは以前から為されている [1][2] が、その一環として、文章から発信者及び受信者が感じるであろう感情に関して読み取らせるというものがある。

本研究では、文章の受信者が感じる感情に焦点を当てることを考える。またその一つとして、小説文章から読者がどのような情報を読み取って感情へと反映させるかといったことをコンピュータに解釈させるシステムの構築を目的とする。

ここではその前段階として、文章表現が単純で易しめな、子供向けの童話文を対象とし、登場人物の感情を読み取るシステムについて構想する。

## 2 本研究における童話文の定義

本研究では、青空文庫 [3] に収められている、カテゴリー「児童書」内の分野「文学」に分類されている作品を童話文と位置づけ、その中から幾つかの作品を選び、解析に利用する

扱う対象として童話文を選んだのは、感情情報の読み取りが容易である文章から始め、段々と複雑な文章を読み取らせていこうという考えと、童話文は一般の小説文よりも簡単な表現であり、感情情報の読み取りが容易であろうという推測によるものである。

すなわち、現段階ではできる限り単純な文章を扱うことを目的としている。それに沿い、解析の難度を下げるために、文章に対して幾つかの前処理を施すことを考えている。

以下にその一例を記す。

- 文章の主語が曖昧な場合に補う。
- 文章の主語が代名詞になっている場合に、対象がなにものであるかを補う。
- 物語内の時系列がばらばらになっている場合、正しく並べ替える。
- 登場人物のリストをあらかじめ与えておく。

システムの初期段階では、このような前処理を施すことにより、どの程度単純化された文章であれば感情情報を正しく読み取ることができるのか、という点についても調査することを考えている。

## 3 日本語 WordNet

本システムは、日本語 WordNet [5][6] を利用することを考えている。以下に日本語 WordNet の概要を述べる。

日本語 WordNet とは、大規模な英語の意味辞書である WordNet を、日本語に対応するよう書き換えたものである。

WordNet は登録している語を synset という類義関係のセットでグループ化しており、それら synset ごとの関係も記述している。以下に一例を示す。

- 上位語：すべての A が B に含まれているなら、B は A の上位語である。
- 下位語：すべての A が B に含まれているなら、A は B の上位語である。

本システムでは、主に上位語をたどることを重要視する。

## 4 感情読解システムの構成

システムの具体的な構成を以下に述べる。

まず、システムは登場人物ごとにそれぞれどのような感情でいるのかを示すモデルをあらかじめ作成しておく。

与えられた童話文章に対しては、一文ごとに分割し、形態素解析エンジン MeCab [4] を用いて

Use of Thesaurus for Reading Emotions in Fairy Tales

<sup>†</sup>Hiroaki DOKOSHI; Faculty of Engineering Hokkaido University

<sup>‡</sup>Satoshi OYAMA, Masahito KURIHARA; Graduate School of Information Science and Technology, Hokkaido University

形態素に分ける。このとき、主語を記憶しておくことで、その文章がどの登場人物の行動であるか、あるいはどの登場人物にも関係しない文章であるかを判別する。

形態素解析された文章が、登場人物のいずれかが主語となっている文章であった場合、感情の変化を示す語が含まれているかどうかを判断する。この判別は、文章に含まれる品詞の中でも感情情報が含まれる可能性が高いと思われるもの、すなわち名詞、動詞、形容詞に対して行われる。

また、この判別は、日本語 WordNet における分類を利用する。

一例として、「嬉しさ」という名詞を考える。日本語 WordNet を用いてこの単語の上位語をたどっていくと、図 1 にあるような結果が得られる。

嬉しさ joy  
 情感 emotion  
 感 feeling  
 動静 state  
 アトリビュート attribute  
 抽象 abstract\_entity

図 1：日本語 WordNet によって「嬉しさ」という単語の上位語をたどった結果

上位語になればなるほど、下位の語の意味を包括した抽象概念になっていることがわかる。

このとき検出される「情感 emotion」や「感 feeling」に注目したい。これらは「嬉しさ」をはじめとした、感情の動きを表す名詞の多くに対する上位語となっている。

これより、ここでは、ある語が与えられたとき、その語の上位語をたどってこれらの有無を判別することで、その語が感情の動きを表すものかどうかを判別できると考える。

判別の対象が動詞である場合も、名詞の場合と同様に上位語をたどることで判別することを考えている。

また、判別対象が形容詞の場合に対しては、形容詞の synset に上位語や下位語の関係性が適用されていないため、まったく同じ方法では判別できない。

そのため形容詞に関しては、形容詞の synset には意味的に関係のある名詞の synset との繋がりが構築されているということに注目している。図 2 に示したように、形容詞そのものではなく、意味的に関係のある名詞の synset の上位語をた

どっていくことで、擬似的に名詞として扱い、名詞と同様の処理によって判別することができないかと構想している。

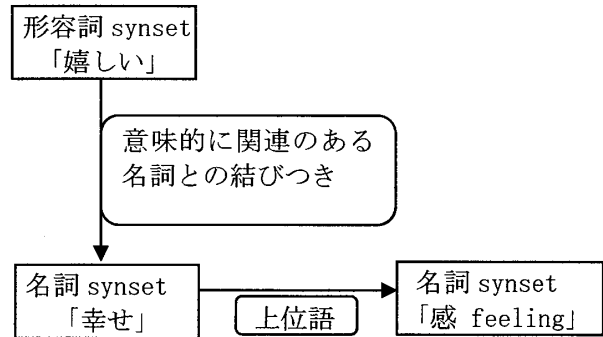


図 2：形容詞に対する感情の動きを表す語の判別

## 5 まとめ

童話文の登場人物の感情の動きを読み取るシステムを考案し、その内容を示した。

今後は実装と、今回扱うことを想定したものよりも更に複雑な、一般的な小説文章を解析するための手法について考察を重ねていきたい。

## 参考文献

- [1] 松本和幸：“会話文からの話者感情推定に関する研究”，人工知能学会誌，24(1)，pp. 120, 2009
- [2] 熊本忠彦，田中克己：“Web ニュース記事からの喜怒哀楽抽出”，情報処理学会研究報告. 自然言語処理学会研究報告，2005(1)，pp. 15-20, 2005
- [3] 青空文庫，〈<http://www.aozora.gr.jp/>〉，2010年1月15日
- [4] MeCab，〈<http://mecab.sourceforge.net/>〉，2010年1月15日
- [5] 日本語 WordNet，〈<http://nlpwww.nict.go.jp/wn-ja/>〉，2010年1月15日
- [6] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi and Kyoko Kanzaki: “Enhancing the Japanese WordNet”, The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP 2009