

自然言語文を構成する各単語に関する 意味素と表層格とに基づく注視の度合の定量化

Quantification of the Degree of Gazing at Each Word

that Composes a Natural Language Sentence Based on Semantic Primitives and Surface Cases

轟木 美穂 鈴木 寿

Miho Todoroki Hisashi Suzuki

中央大学

Chuo University

1 はじめに

本研究の目的は、自然言語文から任意の情報を抽出することである。自然言語文を構成する各単語に対する注視の度合を定量化し、文中から任意の情報を名詞として抽出することを目指す。注視の度合を注視度とし、得たい情報を指す名詞を注視語とする。文中に含まれる名詞の中で付与される注視度が最大の名詞を注視語と判断することで、文中から任意の情報の抽出が可能となる。本稿では注視度の定量化について述べ、実験結果から情報の抽出において有効であることを示す。

2 注視と注視度

本研究において注視とは自然言語文に含まれる名詞の1つを得たい情報として注目することとし、名詞の深層格[1]が得たい情報であると判断することとする。このため、得たい情報が変化すれば同一の文であっても注視語は変化する。これは文の各名詞の注視度についても同様である。なお、本研究で扱う深層格は Fillmore[1] が提案した動作主格、経験者格、道具格、対象格、源泉格、目標格、場所格、時間格の8種類とする。

文に含まれる各名詞の注視度は、意味素と表層格[1]からそれぞれ独立に求めた値の合計とする。意味素を基にする注視度を意味注視度とし、表層格を基にする注視度を表層注視度とする。自然言語文“東京で友達に会う。”における意味素、表層格、意味注視度、表層注視度の例を表1、表2、表3に示す。意味注視度と表層注視度がそれ

表 1: 意味注視度の例

意味素	意味注視度
地域	0.8
一般	0.4

表 2: 表層注視度の例

表層格	表層注視度
デ格	0.8
ニ格	0.5

表 3: 文に含まれる名詞の意味素と表層格の例

名詞	意味素	表層格	意味注視度	表層注視度	合計
東京	地名	デ格	0.8	0.8	1.6
友達	一般	ニ格	0.4	0.5	0.9

ぞれ表1、表2に示した値のとき、各名詞の意味素と表

層格は表3のとおりなので、注視語は付与される注視度が最大の“東京”となる。

3 注視度の定量化

自然言語文に含まれる各名詞の深層格を定め、それらの表層格と意味素を求める。これを元に、注視度の定量化をおこなう。各名詞の深層格は判断する人によって異なる場合があるので、複数人でこれを定めるものとする。

図1に、ある小説テキストにおいて、深層格が場所格となる各名詞とその表層格の関係を示す。図1においてA、Bは各名詞の深層格を定めた人物である。図1から

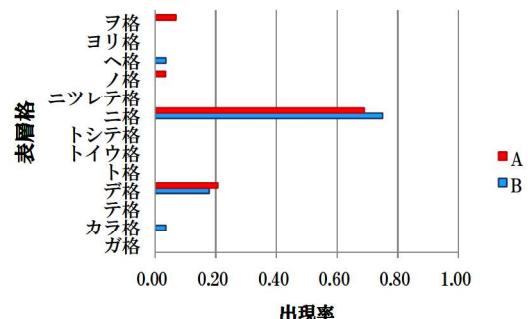


図 1: 場所格と各表層格の関係

場所格となる名詞の多くの表層格は二格であることがわかる。このように深層格と表層格の組合せによって出現率に特徴があり、これを表層注視度としたとき注視語の特定が可能である。なお、意味素と深層格についても同様の結果が得られ、各意味素の出現率を意味注視度として定めることができる。

4 おわりに

自然言語文から情報を抽出するための手法として注視度について述べ、注視度の定量化実験をおこなった。この結果、深層格に対応する各表層格や意味素の出現率から注視度を定量化し、注視語を抽出した。

参考文献

- [1] C. J. Fillmore, “The case for case,” *Universals in linguistic theory*, Holt, Rinehart and Winston, New York, 1968.