

副詞および形容詞による感情表現性の判定 Judgment of affective expressions by adverb and adjective

佐伯 美香[†] 徳久 雅人[†] 村上 仁一[†] 池原 悟[†]
Mika SAEKI Masato TOKUHISA Jin'ichi MURAKAMI Satoru IKEHARA

1. はじめに

従来の機械翻訳では、入力された原文の文型に対して、決定論的に唯一の訳文の文型が決定されていた。これに対して、最近、原言語と目的言語の文型を意味類型化することにより、意味的に対応する複数の訳文文型の中から、最適な文型を選択し、訳文を生成する新しい翻訳方式が提案されている [1]。新しい翻訳方式では、訳文文型を選択するため、文脈情報など様々な情報を使用することが考えられているが、話者の感情も重要な情報であると考えられている。

そこで、本研究では、日本語を対象に、話者の感情が表現された文がどれくらい存在するか、また、それが計算機の処理によってどの程度抽出可能かについて検討する。また、研究を進めるにあたり、[2][3] で定義されているイメージ値を用いる。

2. 感情の推定方法

2.1 感情表現性

日本語から話者の感情を推定する方法としては、一般的に文脈情報を要し、その前後にある状態から判断する。しかし、使用する対訳データは一文単位であるため、一文で感情を表現した文を扱う。従って、一文から感情を判定するために、まずは、単語の持つ感情的な属性を手がかりとする方法を試みる。そこで、本研究では一般に感情を表すことができる副詞・形容詞に注目し、感情を推定する。また、文の話者や動作主が、快/不快の気持ちを表している文を感情表現性のある文と定義する。

2.2 副詞・形容詞のイメージ値

副詞・形容詞には、感情を表す評価が一定している語が多い。しかし、使用する人や状況によって評価が揺れる語もある。これまで、評価が揺れる語に対し、客観的な評価を下すことが行われなかった。しかし、[2][3] では、文脈に依存しない感情的な評価を定義している。この評価を、イメージ値と呼び、-3 から +3 の 7 段階に区分され、個々の条件によって左右される語のイメージ値は 0 と定義している。

2.3 語義

[2][3] には、単語の意味である語義にイメージ値が付与されており、一単語に語義が複数ある場合は、語義毎に値が付与され、一単語において正負の値が混在することがある。[2][3] より、以下に例を紹介する。なお、[2][3] は学校文法での副詞・形容詞と異なる定義をしている。

・形容詞、語義数 2

– 彼女は愛情 こまやか な人だ
(イメージ値:+3)

– 松の緑 こまやか な季節となりました
(イメージ値:なし)

3. イメージ値を有する単語の出現率調査

3.1 調査目的

複数の訳文文型から最適な訳文文型を選択するために話者の感情も重要な情報と考えられている [1]。しかし、感情を表現した日本語が、対訳データ中にどの程度存在するのか未調査である。そこで、イメージ値のある単語を有する文がどの程度存在するのか調査する。

3.2 調査対象

調査対象には日英対訳辞書の例文を使用する [4]。調査対象のデータは感情の主体を明らかにするため、主に単文を含んだ文を対象とする。そのデータ数は約 17 万文である。

3.3 調査方法

調査データ文から [2][3] で集録されている副詞・形容詞を含む文を抽出する。

3.4 調査結果

調査データ文から語義数が 3 以下の副詞・形容詞を含む単文を抽出し、語義数とイメージ値で分類を行った。結果を表 1 に示す。多義がある場合、[2][3] の見出しで最初に定義されているイメージ値で分類を行った。

表 1: イメージ値のある語を含む単文の数

イメージ値	副詞			形容詞			合計
	語義数 1	語義数 2	語義数 3	語義数 1	語義数 2	語義数 3	
-3	262	1063	1121	1451	697	1386	5980
-2	743	314	460	1028	1618	1007	5233
-1	928	444	1374	247	322	936	4251
+1	358	354	2	56	643	63	1476
+2	482	603	6	241	538	1035	2905
+3	496	115	62	1323	1136	1832	4964
合計	3269	2893	3025	4346	5017	6259	24809
出現頻度	1.92%	1.70%	1.77%	2.55%	2.94%	3.67%	14.54%

表中の出現頻度とは、それぞれの語義数の合計を調査データ文 (170,654 文) で割ったものである。

表 1 の結果より、イメージ値を持つ副詞・形容詞を含んだ文の数の出現率は 14.5% であった。

4. 計算機による抽出可能性の評価

4.1 目的

第 4 章では日英対訳辞書の単文において、[2][3] で定義されているイメージ値と一般の人が感じる感情表現性の関係について調査する。

4.2 評価方法

(1) 第 3 章の調査で得られた文からランダムに文を選択する。選択数は語義数単位で 50 文ずつを選択するため、調査対象となる文は計 300 文となる。

(2) (1) で選択した 300 文に対し、評価者による感情表現性の有無の判定を行う。

[†]鳥取大学工学部 知能情報工学科

(3) [2][3] で定義されたイメージ値と (2) で判定した感情表現性との一致性を調査する。

感情表現性の度合いを判定することは個人的な揺れがあるため、評価の判定は感情表現性が「+」「-」「なし」の3種類とし、評価者3名(A, B, C)でイメージ値の符号とどの程度の割合で一致するかの調査を行う。

4.3 イメージ値と感情表現性の関係の評価

4.3.1 感情表現性の判定調査結果

結果を表2, 表3に示す。

表2: 語義数ごとの感情表現性の有無の判定結果 (副詞)

集合	語義数1				語義数2				語義数3			
	A	B	C	平均	A	B	C	平均	A	B	C	平均
感情あり	49	47	43	46.3	43	41	44	42.7	25	29	32	28.7
感情なし	1	3	7	3.7	7	9	6	7.3	25	21	18	21.3

表3: 語義数ごとの感情表現性の有無の判定結果 (形容詞)

集合	語義数1				語義数2				語義数3			
	A	B	C	平均	A	B	C	平均	A	B	C	平均
感情あり	46	42	38	42.0	39	40	29	36.0	36	40	26	34.0
感情なし	4	8	12	8.0	11	10	21	14.0	14	10	24	16.0

副詞・形容詞共に、語義数が1つである語では、「感情あり」と判断した文が最も多いのに対し、語義数が複数になると「感情あり」と判断した文が少なくなることが分かる。また、判定者による判定の差は、語義数ごとに3から11であり、ばらつきが見られた。

4.3.2 イメージ値と感情表現性の関係結果

第4.3.1節において判定したデータを使用し、イメージ値と感情表現性の関係を調べた。その結果を表4, 表5に示す。

表4: イメージ値と感情表現性の一致 (副詞)

集合	語義数1				語義数2				語義数3			
	A	B	C	平均	A	B	C	平均	A	B	C	平均
一致	47	34	32	37.7	47	42	44	44.3	47	44	43	44.7
不一致	3	16	18	12.3	3	8	6	5.7	3	6	7	5.3

表5: イメージ値と感情表現性の一致 (形容詞)

集合	語義数1				語義数2				語義数3			
	A	B	C	平均	A	B	C	平均	A	B	C	平均
一致	42	39	34	38.3	42	33	32	35.7	43	44	37	41.3
不一致	8	11	16	11.7	8	17	18	14.3	7	6	13	8.7

最も精度の低かった形容詞・語義数2の値については50文中平均35.7文が一致し、割合は71.4%であった。最も精度の高かった副詞・語義数3については、50文中平均44.7文が一致し、割合は89.4%であった。全体で見ると一致した文の合計は726文で900文中の割合は80.7%であった。

5. 考察

第4.3.2節での結果より、イメージ値と感情表現性が不一致となった原因は2通り考えられる。

(1) 形容詞的用法の場合: 43文

計算機の処理でイメージ値のある単語を有する文を抽出することはできたが、イメージ値のある単語が他の語にかかるため、文全体としては感情表現性がないと判断される文が存在する。以下に例を示す。

(例) 彼女は ごつい 鞆を提げている。

(イメージ値: -2, 評価者: なし)

本手法では、この問題に対し、改善することができないと思われる。

(2) モダリティ: 131文

計算機の処理でイメージ値のある単語を有する文を抽出することはできたが、強制、願望、否定、勧誘を表す単語や、一文中に別のイメージ値を持つ語が存在すると、その単語に影響され、イメージ値と感情表現性が一致しないと判断される文が存在する。以下に例を示す。

(例) 改善の見込みは わずか しかない

(イメージ値: なし, 評価者: 「-」)

本手法では、この問題に対し、一文中の他の語に対してそれぞれにイメージ値を付与することによって改善することができるとと思われる。

(2)を改善することにより、95.2%の精度で感情表現性の判定が可能であるといえる。

また、本研究での感情表現性の判断方法は、直感的判断であり、広辞苑などにより厳密な意味を確認しながら行った判断ではない。そのため、方法が不十分であり、快/不快を判断する言語資源が乏しかった。しかし、実際に[2][3]を用いてみると、概ね直感的な感情表現性の判断と一致している。この一致から[2][3]は言語資源の乏しさを解決していると考えられるため、格要素や用言との共起を考慮した抽出方法の開発に取り組み価値があると思われる。

6. 結論

本研究では、日本文を対象に、話者の感情が表現された文がどれくらい存在するか調べ、また、それが計算機によってどの程度抽出可能か調べた。[2][3]を使用して、英和辞書などから集めた例文、約17万文を対象に、調査を行った。その結果、イメージ値をもつ副詞・形容詞を含む単文は14.5%であった。

また、このイメージ値と感情表現性の一致率は、80.7%であった。感情表現性がないと判断された文についても、一文中の他の語に対してそれぞれにイメージ値を付与することによって改善後は95.2%にまで改善されると思われる。今後は、イメージ値と感情表現性のある日本文の対訳に注目し、調査を行う必要がある。

参考文献

- [1] 池原, 佐良木, 宮崎, 池田, 新田, 白井, 柴田: “等価的類推思考の原理による機械翻訳方式”, 信学技報, TL2002-34, pp.7-12, (2002).
- [2] 飛田, 浅田: “現代副詞用法辞典”, 東京堂出版,(1994).
- [3] 飛田, 浅田: “現代形容詞用法辞典”, 東京堂出版,(1998).
- [4] 村上, 池原, 徳久: “日本語英語の文対応の対訳データベースの作成”, 「言語, 認識, 表現」第7回年次研究会, (2002).