

# コーパスに基づく感性と論理のネットワーク表現

村井 源\* 松本斉子\*\* 往住彰文\*

\*東京工業大学 社会理工学研究科

\*\*九州大学 大学院システム情報科学研究院

本研究では特定領域のコーパス中に含まれる感性語と内容語の係り受け関係を計量し、ネットワーク化する手法を提案する。ネットワーク化により、感性語の背景的な内容的意味、内容語の背景的な感性的意味側面の双方を数理的に解析可能となる。提案手法に基づき日本の政治家の Web サイトからテキストを収集し、自民党と民主党での感性的・内容的な語彙の意味差異を計量的に抽出した。

## 1. まえがき

人の感性のメカニズムを解明する上で、感性語に着目した研究はテキスト解析・分類の分野をはじめ多くの分野でおこなわれてきている。

感性的表現の特徴としては、論理ではなく感情や直感的な価値判断に訴えかける、どちらかという主観的な表現である点が挙げられる。これは、複雑で、理解に思考的操作を必要とする、どちらかという客観的で論理的な言説とは対照的な性質である。

感性語が効果的に使われてきた分野としては、読者に感情的に訴えかける詩や種々の文学が挙げられる。また感性的表現は、読者の感情を喚起する効果を狙って用いられることが多く、広告やマスコミにおいては頻繁に用いられ、それらに関する研究も盛んである。

マスコミや広告以外で、近年感性的表現との関係がクローズアップされてきた分野として政治が挙げられよう。小泉前首相(首相在任 2001～2006)の言語表現は感性的・感覚的で短く、小泉語と評されマスコミにも大きく取り上げられた。小泉首相の用いた、短く簡潔で感性に訴える表現には、理解しやすいという好意的な受け止め方も多かった。しかし、感性的な政治の表現を、論理的思考の欠如、思考停止、空疎なスローガンの蔓延であるとして、問題視した見解も少なくない[1]。本研究では、政治分野における感性的な表現を対象として取り上げ、感性と論理の関係性を語彙レベルでの解析を試みる。

## 2. 研究の目的

感性的な表現は、簡潔で平易であると感じさせる一方で、意味の全体を構成する論理的構造や、文脈が見えにくい側面も持つ。しかし、政治領域では、非常に短い選挙期間に、政治的知識や関心の薄い層も含め、多くの人々に対して自己の優位性をアピールする必要があり、スローガ的な感性的表現が多用される傾向にある。

政治家側としては、論理の把握困難な感性的表現によって、市民が政治家の意図を推測せざるを得ない状況を築けるため、明言を避けたい

状況などでは好都合であろう。しかし、逆に市民の側からすれば、政治家の真意が不明瞭な状態で投票を行うことになりかねない。このため、感性的表現が多用される政治状況中では、各候補者・政党の論理構造と将来的目標を、市民側が的確に読み取る必要がある。政治領域の感性的表現に込められた意図を、計量的により正確に解釈し、再表現することが可能になれば、種々のミスコミュニケーションを防ぎ、より効率的で豊かな意図の伝達が実現するであろう。また、政治的領域で有効な手法が確立できれば、将来的には詩・文学・音楽・美術や広告などの他の分野にも適用が可能であると考えられる。

本論文では、テキストに出現する感性的表現、中でも頻出する感性語が、意味を伝える内容語とどのような関係にあるかに着目する。感性語と内容語の係り受けの関係性から、ある内容に関連する感性的構造、およびある感性に関連する意味的構造の抽出と解析を可能にする。

人が物事を理解し判断する上で機能する、論理的・意味的な処理と、直感的・感性的な処理は明示的に分割できるわけではないが、感性語と内容語の関係性を解析することで、論理の裏に隠れた感性、感性の裏に隠れた論理の構造を明示化することが可能になると考えられる。

## 3. 対象となるデータ

本研究では、日本の政治家のホームページ(ホームページを公開している、自由民主党・民主党・日本共産党・社会民主党の国会議員 615 人)におけるテキストファイル(html, xml, txt, その他cgi 等のプログラムで自動生成されたタグ付テキストとプレーンテキスト)を対象とする。これらのデータを、WWW自動巡回ソフトを用いてダウンロードする。文字コードをShift JISに統一し、タグの除去を行い、全 134616 ファイル、521MBの日本語政治WWWテキストコーパスを構築した[2](表 1)。

これらのテキスト中の単語を、Chasen を用いて分かち、政治テキストで頻繁に用いられる形態素のリストアップを行った。形容詞の頻出

10 語を表 2 に示す。

#### 4. 感性語と内容語の係り受け解析

構築された日本語政治 WWW テキストコーパスを用いて、日本の政治家ホームページにおける感性語と内容語の関係を解析する。

表 1 収集したテキストデータ

	自民	民主	共産	社民
議員数	419	196	18	13
HP 所有者	394	190	18	13
HP 所有率	94.0	96.9	100.0	100.0
ファイル数	74740	37671	18831	3374
テキスト量(MB)	209	191	99	22
平均テキスト量(MB)	0.5	1.0	5.5	1.7

表 2 全コーパスでの頻出形容詞

形容詞	出現数	比率	累積比率
多い	39799	5.6%	5.6%
強い	33511	4.7%	10.4%
大きい	33016	4.7%	15.0%
新しい	31801	4.5%	19.5%
高い	30283	4.3%	23.8%
厳しい	24608	3.5%	27.3%
よい	23979	3.4%	30.7%
早い	19992	2.8%	33.5%
良い	18815	2.7%	36.2%
難しい	16914	2.4%	38.6%

コーパス中における語彙間の関係は、多様で複雑になるが、複雑な関係性を表現するのに適した手法の一つとしてネットワーク表現が挙げられる。テキスト中の語彙の関係性を計量する方法として一般的なものは、LSA[3]などの、周辺での共起単語を計量し、その頻度によって単語間の関係性の強さとするという方法である。ただし、LSA のように二つの語彙が周辺で共起したことのみでは、実際に単語間に意味的關係性があることの指標としては問題が残る。これに対し、語彙間の係り受け関係を計量し数値化した場合、語彙間には確実に意味的關係があり、先の共起情報に比べてよりノイズの少ない関係性のデータを得られると期待される。

係り受けから語彙の関係性を得る場合、名詞に対しては形容詞、動詞に対しては副詞という感性的な語彙が得られる。本論文では、特定の内容的な語彙（本論文では名詞に限定）が用いられた場合に、その語彙が指し示す背景的な感性構造を、コーパス中の当該語彙と感性語との係り受け関係から計量的に推測し、ネットワー

クとして表現する。また逆に、ある感性的語彙（本論文では形容詞に限定）が用いられた場合に、その語彙が指し示す背景的な内容的構造を、同様にコーパスから計量的に抽出する。

まず、特定の感性語、あるいは内容語と係り受け関係にある語をコーパス中から抽出し、頻度の多いものから順に（本論文では 20 個）選択する。次にこれらの語彙間の関係性の強さをコーパス中の共起関係から数値化し、共起順に上位の関係性を（本論文では上位 100 ペア）抽出する。これより、頻出語彙をノード、語彙間の強い関係性のペアをエッジとする語彙ネットワークが構築できる。

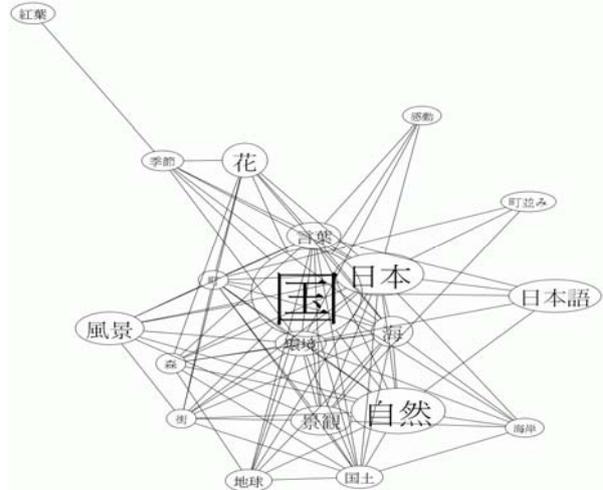


図 1 自民党の「美しい」の内容的意味

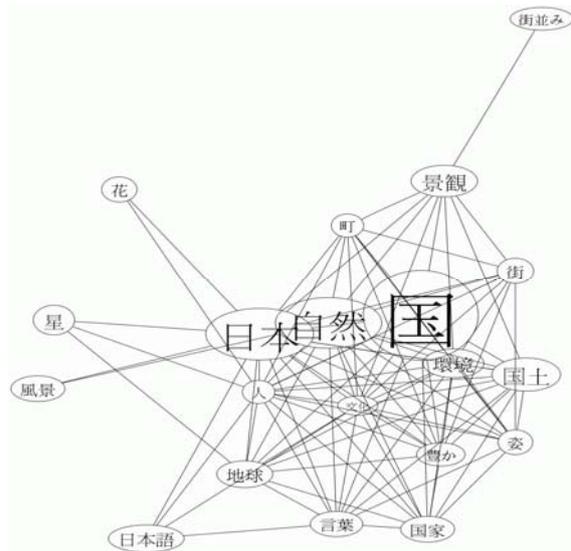


図 2 民主党の「美しい」の内容的意味

本手法の意義を示すために、安部前首相のキャッチフレーズとして有名になった「美しい国、日本」を例に挙げて、感性と論理のコーパスに基づく関係性を解析する。「美しい国、日本」というフレーズは、その感性的表現がさす内容

が不明であるとして、マスコミ等で大きな非難の対象となった表現である。

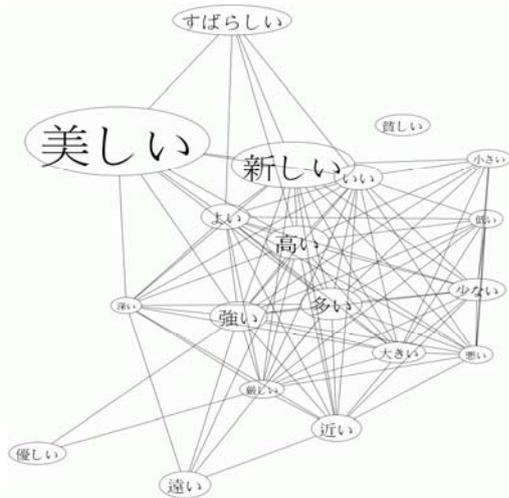


図3 自民党の「国」の感性的意味

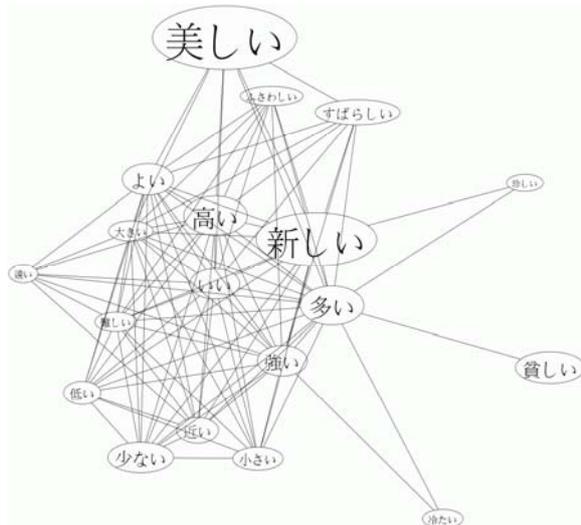


図4 民主党の「国」の感性的意味

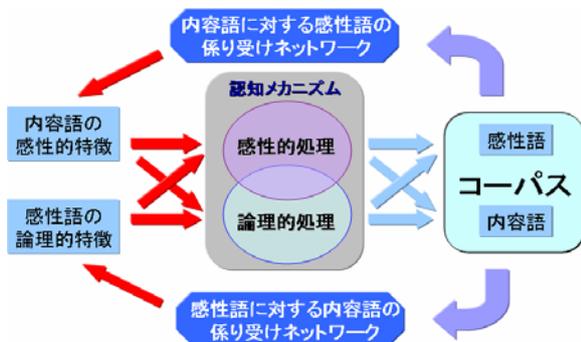


図5 コーパスとネットワークの関係

図1および図2に、「美しい」という感性語の背景にある内容的語彙との関係性の、自民と民主での差異を示す。図3および図4に、

「国」という内容語と感性的語彙の関係性の両党での差異を示す。なお、ネットワーク図中のノードの面積は、コーパス中の対象語彙（「美しい」、「国」）との係り受け頻度に比例する。最大の係り受け語彙数は自民で358回、民主で175回であった（どちらも「国」-「美しい」の組み合わせが最大）。

ネットワーク図から理解できる語彙の背景的ニュアンスの相違としては、例えば、自民では「美しい」といった場合に、「国」以外に「自然」や「日本」が大きなウェイトを占め、かつこれらは非常に近い関係と考えられている。しかし、民主では、「国」以外に「自然」や「日本」がやはり大きなウェイトを占めるものの、「自然」は「国」や「日本」とは、自民党ほどには密接に関連する位置にはない。自民党が「美しい国」といった場合には、国土や自然のニュアンスが強いが、民主党の場合は相対的に言葉などのニュアンスが強いことが推測できる。

また、図3と図4のネットワーク図を考察すると、国が「美しい」ということは、「新しい」、「すばらしい」、「よい」国であるというニュアンスを、自民党・民主党がほぼ同様な形で持っているということを解釈可能である。関連するテキストのコーパスを用いることで、内容的表現の背後の感性、感性的表現の背後の論理を計量的に推測することが可能となる。

人間の思考において、論理的意味処理と、感性的意味処理は単純に分割することは困難である。そのため、自然言語のテキストには、人間が行っている感性的情報処理と、論理的情報処理の結果が複雑に織り込まれていると推測される。ただし、語彙的には、感性的情報と論理的情報は、感性語とそれ以外のよう形で分離して生成される。感性語のように表層的には感性のみを表現する語彙の背後に、論理的な意味もまたニュアンスとして影響を及ぼしていると考えられる。感性語の背景にある論理は、使用者にとって特有の感性語に付随する意味の構造である。これは、語彙レベルとしては、感性語と係り受けする非感性的語彙との関係として抽出できると考えられる。また、同様に論理的表現に用いられる語彙と係り受けする感性的語彙を解析することで、背景にある論理的な意味処理を抽出可能であると考えられる。

テキスト生成の元となる、背景的な感性的処理と論理的処理の関係性を、コーパスに基づいて推測する手法の関係性を概念的に図5に示す。

## 5. 名詞の感性的関係性

係り受け関係のネットワークにおいて、ノードの選択には係り受け関係を用いたが、内容語間および感性語間との関係性は共起関係を用いた。

