

命題－モダリティ構造に対応する階層構造の解析

小橋 洋平 坂野 達郎

東京工業大学 社会理工学研究科社会工学専攻
{kobashi.y.aa,sakano.t.aa}@m.titech.ac.jp

本稿では、日本語学の分野において古くから指摘されてきた日本語文の階層構造について議論する。モダリティ論によると、日本語文は客観的な事柄を表す命題と心的態度を表すモダリティからなり、モダリティが命題を内包する構造を取る。この階層構造を解析することで、命題とそれに対する心的態度の情報を取り出すことが可能になると考えられる。我々は、階層構造の解析を、従来の係り受けを部分的に解析するものとして捉え、朝日・毎日・産経・読売新聞の社説記事を対象に、SVMに基づく既存の統計的な係り受け解析を行った。そして、階層構造と重要な関わりのあるモダリティと従属節の情報を素性に加えてその有効性について検証した。

Statistical Analysis of Japanese hierarchical structure that express relation between proposition and modality

Kobashi, Yohei and Sakano, Tatsuro

In this paper, we discuss the method to analyze Japanese hierarchical structure suggested by some researchers of Japanese Linguistics. According to the theory of modality, Japanese sentence consists of “proposition” that expresses objective thing and “modality” that expresses speaker’s attitude, and modalities put proposition around. If we analyze the hierarchical structure, we can get the proposition and speaker’s attitude toward it. In our study, we selected a general method for analyzing Japanese dependency structure based on Support Vector Machines for the analysis.

1 はじめに

1940年代から、日本語学の分野では、山田孝雄の語排列上の遠心性、三上章の係り結び性など、日本語文の構造が階層的な性質を持っていることが多くの研究者によって指摘されてきた。そして、日本語文の階層構造を体系的にまとめたものとして南の階層モデル¹⁾が存在する。南は一般的な述語文をA~Dの4つの段階からなる重層的な性格のものとして捉えるモデルを提案している(図1)。このモデルは、述語を中心とした階層構造であり、文節と述部要素がそれぞれどの階層に属するかを示している。この文節と述部要素における階層上の対応関係は、英語のように必須の条

件ではないものの、文節の順序関係に傾向があることを表している。

この日本語文の階層性について、日本語文法学では、意味構造と相関関係があるという学説が存在する。日本語文法学では、日本語文の構造は客観的な事柄を表す「命題」と表現者の心的態度を表す「モダリティ」で構成され²⁾、以下の(1)のようにモダリティが命題を内包する階層構造を成している、という階層モダリティ論が多くの研究者に支持されている。

(1) [モダリティ [命題] モダリティ]
例えば、次の文(2)では、「新ミサイル構想と

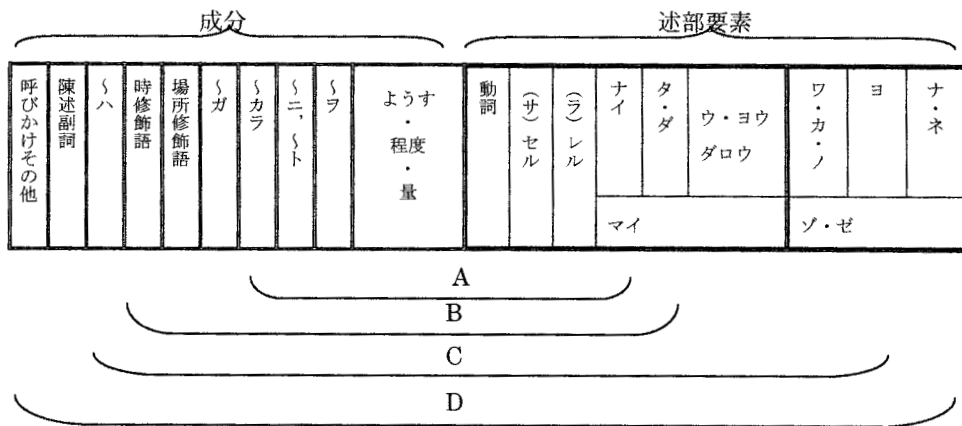


図1 述語動詞文の階層構造 (南 1993)

国は控断すべきではない。



価値判断情報
 命題 国が反省する
 心的態度 負の当為

だが問題は、それだけでは終わらないだろう。



真偽判断情報
 命題 問題がそれだけでは終わらない
 心的態度 推量

図2 命題とモダリティの抽出例

「国は控断すべきではない。」が命題、「べきだろう。」がモダリティとなる。

(2) [[新ミサイル構想と一対のものを見る] べきだろう。]

本稿では、この統語構造と意味構造との階層的な対応関係に着目し、階層構造を解析対象としてその手法を検討する。命題と心的態度の情報を図2のような形で取り出すことで、文章から書き手の意見のみを抽出するシステムや、表現者の主体的な判断に基づいて命題の真偽および蓋然性を評価する手法など、文章要約システムへの応用が期待できる。

以下、2章で南の階層モデルに関する統語上、意味上の特徴についての説明と、本稿と同様に「モダリティ」の概念に着目して心的態度についての分析を行っている先行研究を紹介する。3章では、階層構造を表すタグ(以下、階層構造タグ)の表現方法の提示、4章ではそのタグを分析するモデルのアルゴリズム

と素性を示す。そして、5章で実験と評価、6章で結論と今後の課題を述べる。

2. 階層構造の特徴と先行研究

2.1 南の階層モデルの統語的特徴

南は、自身の提案した階層モデルについて、意味との相関関係を示唆してはいるものの踏み込んだ言及はしておらず、主に統語的な面についてその特徴を示している。南がこのモデルを考える上で主な手がかりとしたのは、各種の従属節(いわゆる接続助詞、あるいは活用語の連用形で終わっている節が主なもの)における、文節間や述部要素間、文節と述部要素間の共起関係である。

例えば、仮定の従属節「～ば」の中に主題の係助詞「は」や推量の述部要素「だろう」は通常現れない。それに対し、逆説の「～が」内には両方現れ得る。このように、従属節はその種類に応じて出現可能な文節と述部要素が異なる。南は、文節と述部要素を、どの従属節内に出現可能かによって、A~D段階に分類している。例えば、仮定の「～ば」内には出現せず、逆説の「～が」には出現する「は」や「だろう」はC段階に位置づけられる。すなわち、文中で出現する位置と、どの従属節の中で出現可能かということの間には相関関係があるということになる。

この統語的特徴は自然言語処理の分野におい

でも応用されており、白井ら²⁾や小野ら³⁾の研究では、係り受け解析の手法に従属節の情報を用いた規則を導入し、係り受けの曖昧性を一部解消することに成功している。

2.2 統語構造と意味構造との相関関係

前述の通り、日本語文法学の分野では、日本語文が持つ階層性を意味構造との相関関係として捉える研究が取り組まれている。前述の日本語文が命題とモダリティによって階層的に構成されているという学説は南のモデルが提示されて以降、南のモデルとモダリティ論とを対比させる試みへと発展を見せた。

益岡⁴⁾は、南のモデルに対応する意味構造として、「事態命名」、「現象」、「判断」、「表現・伝達」の4段階の概念レベルを提唱している。各文節がどの階層に属するかということに対する解釈に相違点があるものの、益岡の「事態命名」、「現象」、「判断」、「表現・伝達」という意味上の階層はそれぞれ、南のA, B, C, Dの統語上の階層と対応した形となっている。例えば、「梅雨前線が雨を降らせたようだ。」という文の統語構造を概念レベルと対応させると、図3のように表せる。

益岡の概念レベルと南の階層モデルとの相関関係についてはまだ十分に議論されていない。ただ、南のモデルを解析することで、概念レベルのような意味構造が明らかになるとすれば、従来とは異なる、心的態度に焦点を当てた意味解析への応用が期待できる。

2.3 先行研究

自然言語の分野において、本稿と同様に、

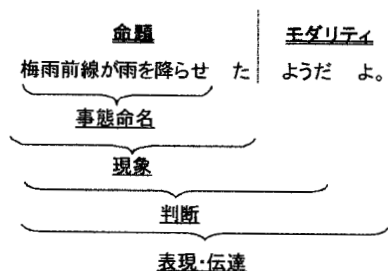


図3 概念レベル(益岡 1997)の一例

モダリティという概念に着目して心的態度の分析を試みた研究に永野ら⁵⁾の研究が挙げられる。永野らは、談話文の主張単位ごとに、客観的事態を述べる部分と、それに対して話者が主観を述べる部分とがあるとし、前者を命題(propositional content)、後者を様相(modality=モダリティ)としている。その上で、命題と様相との対応関係を係り受け構造と見なし解析手法を提案している。

永野らの研究では表1のような様相表現を独自に定め分析を行っている。それに対し、本稿で解析する南の階層モデルは日本語に普遍的なモデルであるとされており、研究対象とするコーパスに依存しない一般性の高いモデルとして適用できると考えられる。

3 階層構造の表現方法

3.1 解析対象となる構造の特徴

前述の南のモデルを解析する上で問題となる文は複文・重文である。単文の場合は、南のモデルに従うと形態素解析の結果のみで一意に階層構造が定まる。それに対し、2つ以上述語が存在する文においては、各文節がどちらの述語に係るかを解析することが求められる。この問題については、これまで、係り受け解析の研究としていくつかの手法が提案されてきており、一定の精度が得られている。そこで本稿では工藤ら⁶⁾など、多くの係り受け解析で採用されているSVMを用い、階層構造の解析を試みた。

階層構造の解析を従来の係り受け解析と比較した場合、階層構造の解析は係り受け関係

賛成	承知しています/よくわかります
反対	疑問であります/反対をいたします/賛成しがたい
確信	確信をいたします/重要であります/当然であります/言をまちません
推測	考えます/思います

表1 様相を示す語句の例(永野ら 2001)

のうち、述語と一部の文節との係り受けのみを解析することに相当する。従来の係り受け解析の場合、文単位では50%程度だが、文節単位では90%近い精度が得られている。階層構造では、一文ごとに解析する係り受けの数が減少するため、従来の手法をそのまま適用するだけでも文単位で高い正解率が得られる可能性がある。

3.2 構造タグの人手による付与

前節で、複文、重文の係り受けの一部を解析すると述べたが、具体的には、従属節がどの文節で構成されているかを明らかにすることが求められる。そこで、本稿では図4で示すように、従属節に対して番号タグをつけることによって、その構造を表す。主節の文節に係る従属節の文節には①、①の文節に係る従属節の文節には②と、従属節の深さに応じて番号を振る。このタグ付与の方法は、係り受けタグを全て付与する作業と比べ、同じ分量に対し大幅に短い時間で行うことができる。

そして解析時には、この従属節を表す番号に基づき、任意の文節は同じ番号かつ間に自身の番号より小さい（主節は0とみなす）文節が存在しない文節のうち、最も文末側にあるものに属するという形で係り受け関係と表記を合わせる（図5）。以後、ある文節が別の文節に属している関係を従属関係と呼ぶ。

なお、この従属関係を用いた場合、連体詞、連体節は他の文節と異なり、従属関係と係り受け関係が一致しない可能性が極めて高い。今回は連体詞・連体節を分析対象から外し、直後の名詞と結合して扱い、その扱いは今後の課題とする。

4 SVMによる階層構造解析

4.1 学習・解析アルゴリズム

本稿では、SVMを用いたチャンキングの段階適用による係り受け解析のアルゴリズムを採用した。SVMによる学習には、TinySVMを用いた。カーネル及び外生パラメータは全て工藤らに準じ、多項式カーネルを採用して次元数は3とした。

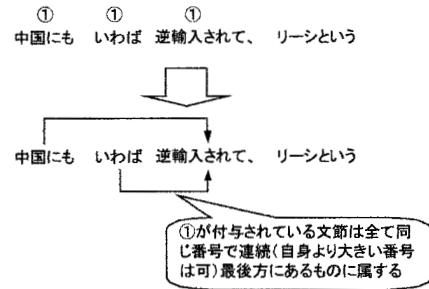


図5 構造タグから従属関係の導出

SVMを用いた係り受け解析には、さまざまな応用的手法が提案されているが、本稿では階層構造の解析がこれまで行われていないことを踏まえ、SVMの標準的な推定方法を用い、従属関係を推定した。

そして、解析された従属関係を元に階層を特定した。その手順を以下に示す。

1. 従属関係で繋がっている2つの文節は同じ階層を構成していると見なす
2. 文末以外でどの文節とも従属関係にないものがあつた場合、その文節をまたぐタグのうち、最も内側にあるタグが付与されている文節と同じ階層と見なす
3. どの文節とも従属関係になく、かつその文節をまたぐタグがひとつもない場合、一番外側の階層に属するものと見なす

4.2 学習・解析に用いる素性

学習・解析で使用する素性を表2に示す。素性の情報は分析対象のコーパスをChaSenによって形態素解析した結果から得る。本稿が工藤らの研究と大きく異なる点は、係り先に接続表現とモダリティを素性として採用している点である。以下、それら2つの素性について説明する。

前述の通り、南のモデルによると従属節はその種類に応じて内部に含みうる文節・述部要素が異なっている。南はその違いに応じて従属節をA~C類に分類している。白井らでは、このA~C類に分類されている従属節の情報をを用いた解析規則を提案している。それ

表2 素性一覧

静的素性	係り元・係り先共通	主辞見出し, 主辞品詞, 主辞品詞細分類, 主辞活用, 主辞活用形語形見出し, 語形品詞, 語形品詞細分類, 語形活用, 語形活用形括弧有無, 句点の有無, 読点の有無, 文節の位置(文頭, 文末) 接続表現, モダリティ
	係り先の文節	接続表現, モダリティ(導入時には語形を除外)
	文節間	距離(1, 2-5, 6), 助詞, 括弧, 句点, 読点
動的素性		係り先に係る文節の静的素性
		係り元に係る文節の静的素性
		係り元が係る文節の静的素性

表3 接続表現一覧

A類	つつ
B類	ば, のに, ので, ないで, と, たら, ず(に)
C類	し, けれど, から, が
D類	と[引用]
A~B類	ながら, 連用形
A~C類	て, で

に対し, 本稿では従属節の接続表現を文節の素性として用いる. 文節単位で素性を与えるため, 付随語のみで構成されている接続表現のみを用いる. また, 南, 白井らともに引用の「と」を取り上げていないが, これは全ての文を内部に持つことができる特殊な従属節としてD類という位置づけにする. 素性として用いる接続表現を表3に示す.

そしてモダリティとは, 述部要素として出現し, 表現者の心的態度を表す表現である. 本稿では, 益岡⁷⁾, 宮崎ら⁸⁾を参考に表4のように分類した. モダリティに関しても, 文節単位に素性を与えるため, 付随語のみで構成されているもののみを用いた.

なお, 接続表現とモダリティについては, 対象となる社説内に出現したものを人手でピックアップして辞書を作成した.

5 実験と考察

5.1 データセット

本実験では, 朝日, 毎日, 産経, 読売新聞4紙が2001年5月に出版した新聞に記載されている社説記事各15記事, 計60記事, 1539文を実験用のコーパスとして用いた. まず, 社説の各文をChaSenで形態素解析した結果を人手にて修正した. 次に, それを用いて機械的に文節に分割した. その上で, 2.2の手

表4 モダリティの種類

モダリティの分類	例
真偽判断	だろう
疑問	(の)か
価値判断	べきだ
表現類型	してください
テンス・アスペクト	(し)た, ている

順に従い人手で階層構造タグを付与した.

5.2 評価尺度

実験を評価するための尺度としては文節間については精度, 再現率とF値, 文単位については正解率を用いた. それぞれの式を以下に示す.

$$\begin{aligned} \text{精度} &= \frac{\text{検出に成功したタグ数}}{\text{解析されたタグ数}} \\ \text{再現率} &= \frac{\text{検出に成功したタグ数}}{\text{評価データのタグ数}} \\ \text{F値} &= \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}} \\ \text{正解率} &= \frac{\text{正解した文の数}}{\text{全文の数}} \end{aligned}$$

実験では4分割交差検定を採用した.

5.3 実験結果

本稿で行った実験の結果を表5に示す. ベースラインとして, 接続表現とモダリティを素性から外したモデルの精度も併記する. この結果からは, 非交差条件の制約を入れない通常のSVMでは階層構造を解析で十分な精

	精度	再現率	F値	正解率
接続表現とモダリティを使用	67.0	71.3	69.0	19.0
語形を使用	77.9	63.1	69.7	17.7

表5 4分割交差検定の結果

度は得られないと考えられる。なお、語形の代わりに接続表現、モダリティを用いた分析では、F値がほぼ同値で、文の正解率は上回っている。今後、新たな分析手法の中でもこの素性を検討に入れる必要があると思われる。

6 おわりに

本稿では、日本語文が持つ階層性を体系化した南のモデルを解析する手法を検討した。そして、従来の SVM による係り受け解析と同様の解析器を採用し、接続表現、モダリティという階層構造と相関を持つと思われる素性をモデルに組み込んで実験を行った。

正解率をより改善していく上で、検討すべきだと思われる課題を以下に示す。まず、解析の前に述語を特定する手続きを加えることの有効性が挙げられる。階層構造タグは文節と述語との関係として表されるので、あらかじめ述語が特定されていれば、階層構造タグ解析の精度が改善すると考えられる。

また、接続表現とモダリティについては、任意のコーパスに適用できるように、網羅性の高い接続表現辞書およびモダリティ辞書の作成が求められる。さらに、今回は付属語のみで構成されている表現を対象としたが、どちらにおいても自立語を含む表現が数多く存在する。特に、モダリティに関しては、一般にモダリティとは見なされないが、表現者の心的態度を表すと思われる表現は多種多様であることが指摘されている⁹⁾。このようなモダリティ以外の心的態度表現の扱いも検討した上で、精度の向上のみならず、今後の応用研究においても重要な課題になると考えられる。

参考文献

- 1) 南不二男：現代日本語文法の輪郭，p.269，大修館書店，東京(1993)
- 2) 白井 諭，池原 悟，横尾 昭男，木村 淳子：“階層的認識構造に着目した日本語従属節間の係り受け解析の方法とその精度”，情報処理学会論文誌，Vol.36，No.10，pp.2353-2361(1995)
- 3) 小野貴博，菅沼明，谷口倫一郎：“日本語文章推敲支援における係り受けを誤解される文の抽出”，言語処理学会研究報告，2006-NL-175,p.206，Vol.2006，No.94，pp.99-104(1997)
- 4) 益岡隆志：新日本語文法選書 2 複文，p.206，くろしお出版，東京(1997)

- 5) 永野敬一郎，辻井潤一，鳥澤健太郎：“談話文からの命題-様相の抽出システム”，言語処理学会第7回年次大会発表論文集，pp.38-41(2001)
- 6) 工藤拓，松本祐治：“チャンキングの段階適用による日本語係り受け解析”，情報処理学会論文誌，Vol.43，No.6，pp.1834-1842(2002)
- 7) 益岡隆志：モダリティの文法，p.232，くろしお出版，東京(1997)
- 8) 宮崎和人，安達太郎，野田晴美，高梨信乃：新日本語文法選書 4 モダリティ，p.325，くろしお出版，東京(2002)
- 9) Narrog, Heiko: On defining modality again, Language sciences, Vol.27, No.2, pp.165-192(2005)