

意味解析システム SAGE の精度向上 —モダリティと副詞節について—

梅澤 俊之[†] 加藤 大知[†]
松田 源立^{††} 原田 実^{††}

概要：我々は、文章中の単語の語意の決定および係り受け関係にある2文節間の深層格の決定を行う意味解析システム SAGE の開発を行ってきた。本研究では、文中の語意に現れない話し手の命題に対する主観的認識や発話態度を表すモダリティを、命題の確からしさを捉える真偽判断のモダリティ、命題の実現をどの程度望ましいと捉えているかを表す価値判断のモダリティ、話し手の発話態度のあり方を表す発話のモダリティ、命題内容の実現の難易度を表す程度のモダリティの4つに大きく分類し、それぞれにおいて、詳細なモダリティを設定し、これらを主として表層表現をベースに自動的に判定できるようにした。また、副詞節に対する深層格の精度を向上するために、EDR 日本語コーパスから約 11000 文を用いて決定木ルールを学習した。これらの改善により語意に現れない話者の意図の把握が可能になり、また深層格全体の精度は 1.1% 向上し 89.9% になった。

Improvement of precision of the semantic analysis system SAGE - Focused on modality and adverbial clause case -

Toshiyuki Umezawa[†], Hiroaki Kato[†]
Yoshitatsu Matuda^{††} and Minoru Harada^{††}

Abstract: We have developed semantic analysis system SAGE which decides the meaning of a word and the deep case between two clauses which exist in the dependency relation in the sentences. In this research, we classify the modality which is not represented by the meaning of a word of a sentence but shows the subjective recognition of the proposition and the utterance attitude of the talker into four of modalities: the modality of truth judgment, the modality of value judgment, the modality of utterances and the modality of difficulty level. In addition, detailed modalities are set in of each. Further, we made the rules which can judge each detail modality chiefly based on the surface expression of a word. Moreover, the decision tree rule was studied by using about 11000 sentences from EDR Japanese corpus to improve the accuracy of a deep case to the adverbial clause. By these improvements, the recognition of the intention of

[†] 青山学院大学大学院 理工学研究科理工学専攻 知能情報コース

^{††} 青山学院大学 理工学部 情報テクノロジー学科

the talker which did not show up to in the meaning of a word became possible, and the accuracy of the entire deep case analysis was improved by 1.1% and became 89.9%.

1. 背景・研究目的

近年の IT の急激な発展によって、大量の文書データからの知識の発掘（テキストマイニング）などの分野で、文章の意味解析への期待が高まっている。

原田研究室では、EDR 電子化辞書^[1]に記載された情報を元に、文章中の単語の語意の決定および係り受け関係にある2文節間（主辞同士）の深層格の決定を行う意味解析システム SAGE^{[2][3][4][5]}の研究開発を行ってきた。

特に、クレーム分類、要約などの応用研究において、文中の語意には現れない話し手の認識や態度など文の意図を把握する必要があることから、本研究においては、文の発話者の命題（文の主要部分）に対する認識や発話態度を表すモダリティの細分類と判定の自動化を行った。さらに、副詞節に対する深層格の精度を向上するため EDR 日本語コーパスから約 11000 文を用いて精密な決定木ルールを学習し、このルールを SAGE に組み込んだ。

2. 基本的考え方

SAGE は日本語を意味解析し、結果を文節や形態素ごとにそれらの意味や品詞や深層格（他の文節との役割関係）などを保持したリストの集合として表現する。これは、文節を頂点、係り受け関係にある文節間の深層格を辺と考え、図1のような意味グラフとして表現される。



図1: SAGE の意味解析結果を示す意味グラフ

(1) 真偽判断のモダリティ

真偽判断のモダリティ 7 種の分類体系を図 4 に示す。

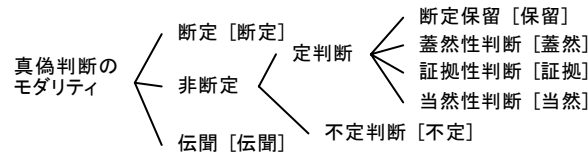


図 4：真偽判断のモダリティ

真偽判断のモダリティは主に、命題を確かなものとして捉える断定と不確かなものとして捉える非断定、情報源が発話者以外である伝聞に大別される。さらに、非断定には、真偽性の度合いに応じて下位分類が存在する。

(2) 価値判断のモダリティ

価値判断のモダリティ 5 種の分類体系を図 5 に示す。

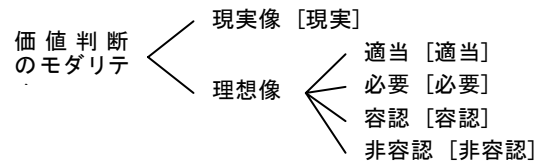


図 5：価値判断のモダリティ

価値判断のモダリティは主に、命題を現実のものとして捉える現実像と、現実となることを理想として捉える理想像とに大別される。さらに、理想像は理想のあり方に応じて適当、必要、容認、非容認に分類される。

4.2.2 発話のモダリティ

仁田^[8]による、発話のモダリティ 18 種の分類体系を図 6 に示す。

話し手の発話態度のあり方を表す発話のモダリティは、大きく四つに分類され、要求、情意、演述、問い掛けからなる。さらにそれぞれに対して、下位分類が存在する。

4.2.3 程度のモダリティ

「この棒は折れる」と「この棒が折れにくい」では、同一の命題でも語意に現れない違いがある。しかし、判断・発話のモダリティに分類されるものではない。

そこで本研究では、命題内容が肯定か否定かあるいはその実現の難易度などを表すものとして、程度のモダリティを定義する。

図 7 に示すように、困難、容易、過度、否定、外部否定の 5 種類に分類した。

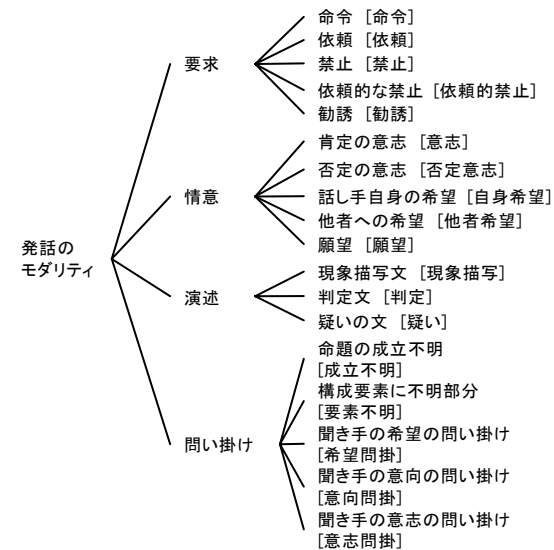


図 6：発話のモダリティ

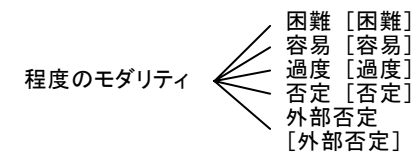


図 7：程度のモダリティ

4.3 モダリティの判定

モダリティ判定処理の手順を図 8 に示す。まず、対象となる文節を選択する。基本的にモダリティが現れるのは文の述語節である。従って、判断のモダリティと程度のモダリティについては主節と接続節(副詞節、連体節、補足節、並列節)に付与する。一方、発話のモダリティは接続節には現れない、これは発話行為の基本単位として文が存在しているからである。よって、発話のモダリティについては主節の述語節のみ付与する。SAGE における述語節には、文節品詞として動詞節、動名詞節、形容詞節、形容動詞節、断定節、形容名詞節、形容動名詞節、断定名詞節がある。

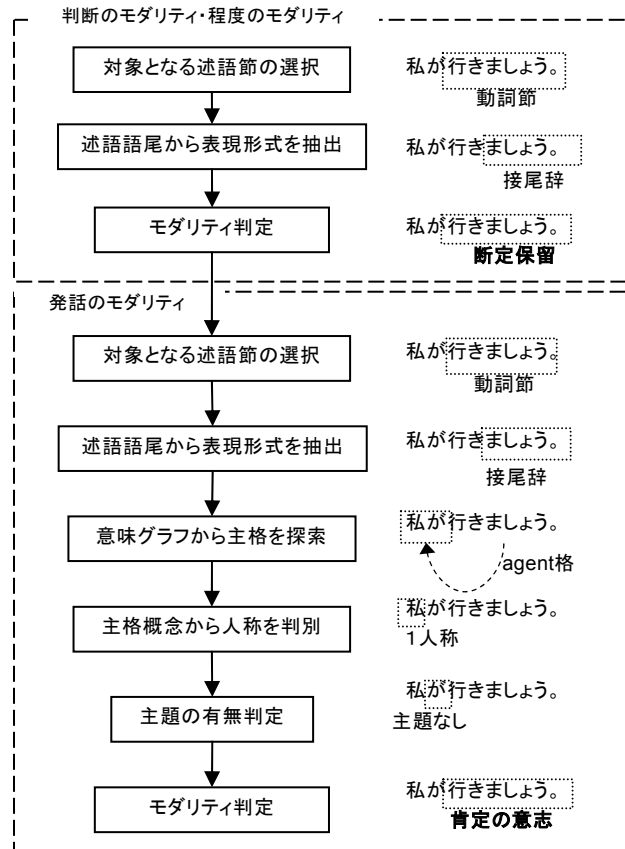


図8：モダリティ判定処理手順

次に、選択した文節に対して、表現形式と主語格および主題との関係によってモダリティを判定する。

判断・程度のモダリティは述語語尾に現れる為、述語語尾を構成する形態素およびその品詞・活用形が表1、表2、表4に示すような表現形式を持てば、表の左側に示すモダリティを持つとする。

発話のモダリティでは述語の語尾に加え、聞き手の存在が重要になるため、表3に示す表現形式の抽出に加え、表5に示す主語格の人称、主題の有無により判定する。このように JUMAN・KNP の形態素解析と係り受け解析、及び SAGE の意味解析の結果を用いることにより、if_then 型ルール形式でモダリティが判定できる。動詞の活用形、助動詞、接尾辞は JUMAN 出力の品詞、活用形から、主語格の人称は SAGE 出力の述語節から agent 格、o-agent 格または a-object 格でつながる文節の主辞の概念から、主題の有無は主語格の助詞の種類によりそれぞれ判断する。

表1：真偽判断のモダリティの表現形式

モダリティ	表現形式
断定	無標
断定保留	だろう/まい/たろう/(よ)う/でしょう
蓋然性判断	かもしれない
	にちがいない
証拠性判断	ようだ/みたいだ/らしい/(し)そうだ/という
当然性判断	はずだ/はず
不定判断	か/かな/かしら
伝聞	そうだ

表2：価値判断のモダリティの表現形式

モダリティ	表現形式
現実像	無標
適当	べきだ/ほうがよい/ればよい/のだ/ことだ /ものだ/~たい
必要	なければいけない/(せ)ざるをえない/ しかない/なければならない
容認	てもよい/て構わない
非容認	てはいけない/だめだ

5. 副詞節の解析精度向上

5.1 副詞節について

副詞節は、述語の修飾、又は主節全体に対する修飾として働くものであり、「テレビを見ながらご飯を食べる」といった文における「テレビを見ながら」がそれにあたる。副詞節が表す意味には、時、原因・理由、条件・譲歩、付帯状況・様態、逆接、目的がある。副詞節は、文節の係り受けの観点から見れば、述語節が述語節に係るということになる。

表の最左列に示した深層格をこの係り受けに与えることにする。

この中で、赤字で示した共起関係子や活用形は複数回表6に出現している。この場合は一意に深層格を決定できない。このため、たとえば、共起関係子「ながら」の場合は表7に示す様な他の素性に基づく判定ルールを個別に作成し判定していた。しかしこの判定ルールは、少数の事例を基に構築したルールで精度上問題があった。

表7：素性に基づく判定ルール

共起関係子	SAGE格	係り側の上位概念			
ながら (defaultは cooccurrence)	reverse	「状態」 3aa963			
	cooccurrence	「変化」 3f9856	「行為」 30f83e	「移動」 30f801	「現象」 30f7e5

5.3 決定木学習法による副詞節の深層格決定

EDRには、副詞節に関する深層格の共起事例は共起辞書中には含まれていないが、コーパス中には例文を手で解析した結果が存在している。調査の結果、全コーパス例文40207文中、16348個の副詞節をなす係り受けの解析結果を抽出できた。そこで、これらを、表6の曖昧性のある共起関係子や活用形計10個のそれぞれのケースごとにまとめ、係り側と受け側双方から素性を抽出し、深層格の決定ルールをWeka^[11]を用いて決定木として学習することにした。ここで、抽出する素性としては、曖昧性の元になる共起関係子や活用形ごとに若干異なるが、共通的には、まず係り側からは、係り側の主辞の上位概念が「変化 3f9856」であるなどの特定の上位概念の有無に関する素性を42個、係り側の形態素の活用形が「タ系連用テ形」であるなどの活用形に関する素性を12個、係り側の文節品詞が「形容詞節」であるなどの文節品詞に関する素性を7個、計61個の素性を、受け側からは、受け側の主辞の上位概念が「思考する 444dda」であるなどの特定の上位概念の有無に関する素性を16個、受け側の形態素の活用形が「形容詞」であるなどの活用形に関する素性を2個、受け側の文節品詞が「動詞節」であるなどの文節品詞に関する素性を5個、受け側と係り側が「共に動詞節で連続」などの素性を3個、計26個の特徴を取り出すことにした。事例として、共起関係子「から」の場合の学習データの一部を表8に示した。

学習と評価のために、副詞節を含む係り受け解析事例16348件を、7割を学習用、3割をテスト用とした。

表8：学習データ例

kind	henka	idou	riyousuru	kakari_zyou	zyoutai	sonzaizyou	taisyoukou
cause	0	0	1	0	0	0	1
cause	0	0	0	0	1	1	0
cause	0	0	1	0	0	0	1
cause	0	0	1	0	0	0	1

学習には、WekaのJ4.8と呼ばれるQuinlanのC4.5に基づいた決定木の作成機能を

用いた。学習データから決定木を作成し、SAGE本体中の副詞節の解析プログラムにおいて、生成された決定木ルールをC#に書き換えてそれぞれの曖昧性の判定個所の従来のプログラムを取り換えた。その後評価実験を行った結果、表9～表18のような結果を得た。ここでこのように曖昧性のある判定に関する結果をそれぞれの深層格の事例数をかけた加重平均をした結果、従来の手作りルールから学習した決定木ルールに変更した結果、精度は70.0%から75.7%に向上し、再現率は73.7%から77.4%に向上、F値は70.4%から75.7%に向上することができた。

表9：共起関係子「から」における実験結果

共起関係子	判定法	深層格	精度	再現率	F値	総合件数
から	手作りルール	cause	0.94	1.00	0.97	472
		time-from	0.99	0.86	0.92	194
	学習した決定木	cause	0.99	0.97	0.98	472
		time-from	0.94	0.98	0.96	194

表10：共起関係子「ながら」における実験結果

共起関係子	判定法	深層格	精度	再現率	F値	総合件数
ながら	手作りルール	cooccurrence	0.95	0.96	0.96	487
		reverse	0.32	0.32	0.32	28
		condition	0	0	0	3
		manner	0	0	0	1
	学習した決定木	cooccurrence	0.95	1.00	0.97	487
		reverse	0.67	0.14	0.24	28
		condition	0	0	0	3
		manner	0	0	0	1

表11：活用形「**タ系連用テ形」における実験結果

活用形	判定法	深層格	精度	再現率	F値	総合件数
**タ系連用テ形	手作りルール	cause	0.71	0.96	0.82	26
		and	0	0	0	8
		manner	0.33	0.25	0.29	4
		purpose	0	0	0	2
	学習した決定木	cause	0.71	0.92	0.80	26
		and	0.50	0.38	0.43	8
		manner	0	0	0	4
		purpose	0	0	0	2

表 12：共起関係子「のに」における実験結果

共起関係子	判定法	深層格	精度	再現率	F値	総合件数
のに	手作り ルール	reverse	0.77	0.79	0.78	38
		purpose	0.52	0.80	0.63	15
		object	0	0	0	9
	学習した 決定木	reverse	0.69	0.76	0.73	38
		purpose	0.63	0.67	0.65	15
		object	0	0	0	9

表 13：共起関係子「と」における実験結果

共起関係子	判定法	深層格	精度	再現率	F値	総合件数
と	手作り ルール	object	0.91	0.92	0.91	4812
		condition	0.58	0.87	0.70	1615
		goal	0	0	0	221
		purpose	0	0	0	121
		sequence	0	0	0	433
		manner	0	0	0	43
		and	0	0	0	25
		cooccurrence	0	0	0	11
		cause	0	0	0	22
		a-object	0	0	0	2
	time-from	0	0	0	1	
	学習した 決定木	object	0.91	0.95	0.93	4812
		condition	0.70	0.82	0.75	1615
		goal	0.73	0.30	0.43	221
		purpose	0.37	0.44	0.40	121
		sequence	0.52	0.17	0.25	433
		manner	0	0	0	43
		and	0.60	0.48	0.53	25
		cooccurrence	0	0	0	11
		cause	0	0	0	22
a-object		0	0	0	2	
time-from	0	0	0	1		

表 14：共起関係子「ため」における実験結果

共起関係子	判定法	深層格	精度	再現率	F値	総合件数
ため	手作り ルール	cause	0	0	0	69
		purpose	0.22	1.00	0.36	19
	学習した 決定木	cause	0.86	0.88	0.87	69
		purpose	0.53	0.47	0.50	19

表 15：共起関係子「ために」における実験結果

共起関係子	判定法	深層格	精度	再現率	F値	総合件数
ために	手作り ルール	cause	0	0	0	2
		purpose	0.82	0.31	0.45	29
	学習した 決定木	cause	0	0	0	2
		purpose	0.94	1.00	0.97	29

表 16：活用形「タ系連用テ形」における実験結果

活用形	判定法	深層格	精度	再現率	F値	総合件数
タ系連用 テ形	手作り ルール	manner	0.42	0.85	0.56	664
		cause	0.42	0.45	0.43	423
		condition	0.50	0.21	0.30	71
		sequence	0.74	0.38	0.51	1142
		and	0	0	0	56
		implement	0.78	0.75	0.76	132
		purpose	0	0	0	9
		cooccurrence	0	0	0	41
	学習した 決定木	manner	0.47	0.63	0.54	664
		cause	0.52	0.30	0.38	423
		condition	0.40	0.08	0.14	71
		sequence	0.59	0.64	0.61	1142
		and	0.23	0.05	0.09	56
		implement	0.77	0.75	0.76	132
		purpose	0	0	0	9
		cooccurrence	0	0	0	41

表 17：活用形「基本連用形」における実験結果

活用形	判定法	深層格	精度	再現率	F値	総合件数
基本連用形	手作り ルール	sequence	0.90	0.87	0.89	851
		and	0	0	0	175
		cause	0.60	0.69	0.64	508
		condition	0.84	0.78	0.81	172
		cooccurrence	0	0	0	19
		manner	0.41	0.86	0.55	137
		reverse	0.77	0.71	0.74	42
		goal	0	0	0	4
		purpose	0.39	0.64	0.49	14
		object	0	0	0	41
	学習した 決定木	sequence	0.90	0.86	0.88	851
		and	0.45	0.26	0.33	175
		cause	0.60	0.78	0.68	508
		condition	0.84	0.78	0.81	172
		cooccurrence	0	0	0	19
		manner	0.58	0.65	0.61	137
		reverse	0.74	0.74	0.74	42
		goal	0	0	0	4
		purpose	0.77	0.71	0.74	14
		object	0.32	0.17	0.22	41

表 18：共起関係子「ように」における実験結果

共起関係子	判定法	深層格	精度	再現率	F値	総合件数
ように	手作り ルール	manner	0.83	0.86	0.85	29
		purpose	0.20	0.17	0.18	6
	学習した 決定木	manner	0.83	1.00	0.91	29
		purpose	0	0	0	6

6. 実験及び評価

本研究の最終の評価実験では、無作為に抽出した日本語コーパス 101 文を使用した。語意、深層格について、SAGE2007 と比較した精度の結果を表 19 に示す。

表 19：SAGE2008 の評価実験の結果

	語意の精度	深層格の精度
SAGE2007	93.3%	88.8%
SAGE2008	95.0%	89.9%

副詞節の深層格決定ルールを見直したことにより、深層格の精度が 1.1%向上した。また、語意の精度については辞書データの更新により精度が向上した。

モダリティの付与により、文中の語意に現れない話者の意図の把握の精密化が可能となり、応用研究での有用性が高まった。

今後の課題として、今回の副詞節学習での素性の追加による更なる副詞節の精度向上、及び引き続き EDR 辞書データを整理することに焦点をあてて、利便性の向上を目指す。

参考文献

- 1) (株)日本語電子辞書研究所: EDR 電子化辞書仕様説明書(第2版), (株)日本語電子辞書研究所(2002).
- 2) 原田実, 尾見孝一郎, 岩田隆志, 水野高宏: 日本語文章からの意味フレーム自動生成システム SAGE(Semantic frame Automatic GEnerator)の開発研究, 人工知能学会第 13 回全国大会論文集, pp. 213-216 (1999).
- 3) 原田実, 水野高宏: EDR を用いた日本語意味解析システム SAGE, 人工知能学会論文誌, Vol.16, No.1, pp.85-93 (2001.1).
- 4) 原田実, 田淵和幸, 大野博之: 日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価, 情報処理学会論文誌, Vol.43, No.9, pp.2894-2902(2002.9).
- 5) 川口純一, 青木洋, 松田源立, 原田実: 意味解析システム SAGE の精度向上, 情報処理学会第 69 回全国大会論文集, 1C-04, 第 2 分冊, pp. 77-78 (2007.3).
- 6) 京都大学情報学研究科知能情報学専攻能メディア講座言語メディア研究室(黒橋研究室), <http://nlp.kuee.kyoto-u.ac.jp/>
- 7) 益岡隆志: 日本語モダリティ探求, くろしお出版(2007).
- 8) 仁田義雄: 日本語のモダリティと人称, ひつじ書房(1991).
- 9) 佐藤直美, 韓東力, 原田実: 日本語意味解析に伴うヴォイス・テンス・アスペクト・ムードの決定, 情報処理学会第 67 回全国大会論文集, 1J-03, 第 2 分冊, pp.69-70 (2005.3).
- 10) Richard O.Duda, Peter E.Hart, David G.Stock, 尾上守夫(訳): パターン識別, 新技術コミュニケーションズ(2003).
- 11) Weka, <http://www.cs.waikato.ac.nz/ml/weka/index.html/>