

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

佐藤 直美[†] 韓 東力[‡] 原田 実[†]

[†]青山学院大学大学院 理工学研究科 経営工学専攻

[‡]青山学院大学 理工学部 情報テクノロジー学科

1. はじめに

原田研究室ではこれまで、EDR電子化辞書に記載された情報を元に、日本語文を意味解析し、格フレーム群に自動変換するシステムSAGE^[1](Semantic frame Automatic Generator)を開発してきた。既存のSAGEは語意(概念ID)と深層格のみを意味グラフとして出力しており、ヴォイス・テンス・アスペクト・ムードなどの属性子の解析は行っていない。本研究では、属性子の判定方法と、ヴォイス(使役と受身)を含む文の語意や格の解析精度向上手法を提案し、新版SAGEに実装する。

2. ヴォイスの特徴

従来のSAGEでは、ヴォイス(使役と受身)を含む文の解析精度が低かった。EDRコーパスでの例文を調べた結果、能動表現と比較して、次の事が分かった。

EDRでは使役表現での動詞は能動表現での動詞と語意が異なる。

受身表現での動詞と能動表現での動詞では、同じ語意が割り振られている。また、能動表現と受身表現では、表層格が変わっても深層格は変わらないという特徴がある。

2.1. 使役表現判定

基本形が「せる」、品詞が「接尾辞」、詳細品詞が「動詞性接尾辞」である場合、使役と判定する。

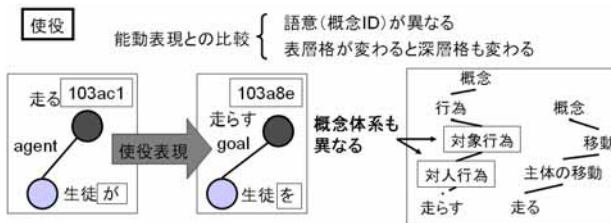


図1 EDRでの使役

2.2. 使役を含む文の辞書引き

従来のSAGEでの辞書引きは、まず形態素分析と係り受け解析を行う南瓜とKNPによって得られた動詞の基本形をEDR単語辞書での入力とし、動詞の不変化部を得る。この不変化部を更に共起辞書で引き、語意を得ていた。しかし、この方法では、使役動詞の語意が得られないので、次の方法で使役動詞の語意を得るようにする。

使役表現と判定された文節があった場合、南瓜とKNP解析器から辞書見出し語を得て、この辞書見出し語を用いて、共起辞書を引けば、使役の意味を持つ語意を得ること

Determination of voice, tense, aspect and mood in the Japanese semantic analysis system SAGE.

Satoh Naomi[†], Dongli Han[‡] and Minoru Harada[†]

[†]Graduate School of Industrial and Systems Engineering, Aoyama Gakuin University.

[‡]Department of Integrated Information Technology, Faculty of Science and Engineering, Aoyama Gakuin University.

ができる。

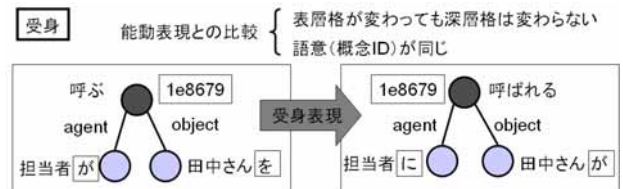


図2 EDRでの受身

2.3. 受身表現判定

受身とは、述語に接辞(r)areruが付くと、表層格は変わるが深層格は変わらない表現である。ただし、この接辞が付くと、受身以外に尊敬と可能表現の場合もある。この曖昧性を解消し受身と判定するには、尊敬や可能との判別をする必要がある。このために、以下に示すように、「主語が敬意を表す対象かどうか」の点数と「受身を表す活用型かどうか」の点数と「語意からして他に影響を与える動詞かどうか」の点数を求め、合計点が2以上であれば受身と判定することにした。

表1 動詞活用型得点

	I類(ラ行以外)	I類(ラ行)	II類	来る	する
変態活用型	五段・～行	五段・ラ行	一段	カ変・来ル	サ変・スル
human活用型	子音動詞	子音動詞	母音動詞	カ変動詞来	サ変動詞
可能	-eru (「れる」以外)	-れる	-れる (ら抜き)	来れる	できる
受身	-れる	-られる	-られる	来られる	される
尊敬	-れる	-れる	-られる	いらっしゃる	なさる

0点 1点 2点

(1) 「主語が敬意を表す対象かどうか」の点数

接辞(r)areruが付いた動詞vに係る主語の上位概念を検索し、「3c4795」「先達」、「444df2」「職業、肩書き、役割で限定した人間」、「444ae9」「敬称に該当する語そのもの」の何れにも該当しない場合、得点に1を加える。

(2) 「受身を表す活用型かどうか」の点数

「受身を表す活用型かどうか」の点数とは、活用型と接辞の組み合わせによって決定される点数である。例えば、一類ラ行で接辞「られる」なら2を加え、「れる」なら加点しない。

(3) 「語意からして他に影響を与える動詞かどうか」の点数

「語意からして他に影響を与える動詞かどうか」の点数では、日本語文法に記載された「持ち主の受身」になりやすい動詞、「まともな受身」になりやすい動詞、もしくはそれらの動詞と語意において概念距離が1の動詞が含まれていた場合は、得点に1を加える。

2.4 受身文での辞書引きの問題点

既存の SAGE では、能動表現から受身表現に変換するにあたって、表層格が変わっても深層格が変わらない受身表現に対応していなかった。これは、共起辞書の事例は能動系が中心のためである。そこで、受身表現を能動表現と仮定して辞書引きを行えば、深層格の精度が向上できると考えた。相手に動作をさせる側を「受身の主体」とし、動作を行う側を「動作の主体」とし、それぞれの共起関係子の置き換えを行って、辞書引きを行う。

2.4.1 「受身の主体格」の共起関係子の置き換え

受身文「B に(から・によって)C られた」が与えられた時、

- (1) 受身文がどうか判定する
- (2) 受身の主体を見つける

まず、表層格が二格、カラ格、ニヨッテ格を含む文節を見つけ、B を受身の主体の候補とする。次に B の上位概念を調べ、上位概念に「人間」があれば、受身の主体とみなす。「人間」がない場合でも、「位置」「時」「物事」がなければ、また受身の主体とする。

B の共起関係子(に・から・によって)をガ格に変更し、能動態に変換して受身の主体 B の深層格を決定する。

例: B に(から・によって)C された B が C した

2.4.2 「動作の主体格」の共起関係子の置き換え

受身表現「A は C られる」は能動表現にすると「A に C する」または「A から C する」または「A を C する」の 3 通りある。このように、動作の主体 A の表層格は受身表現では「八格」又は「ガ格」であるが、能動表現では「二格」、「カラ格」、「ヲ格」の何れかである。この能動表現の 3 通りの可能性に対する共起情報に基づく語意格組を語意-格総合評価値配列に加え、一意に絞り込む。

例: A は(が)C られた A を(に・から)C した

3 テンス・アスペクト

テンスとは時間的概念を表す部分である述語の中の要素の形態素変化であり、「過去」「現在」「未来」がある。アスペクトは発話者の事象の時間的性質に対する見方を伝える形式であり、基準時以前に動作や出来事が完了したことを表す「完結相」と、基準時に動作や出来事が継続していることを表す「非完結相」がある。

3.1 テンス・アスペクト判定

(1) 過去形完結相「タ」と過去形非完結相「タ」の判別動詞が状態動詞(上位概念が「3aa963」「状態」の場合)か否かを調べることによって判定する。例えば、「太郎は図書館に居た」の場合、述語「居た」の上位概念を検索し、「状態」があるので、状態動詞となる。タ形かつ状態動詞は、テンスが過去、アスペクトが非完結相と判定する。

(2) 現在形の非完結相と未来系の非完結相の判別

未来を表す文節が動詞に係っている場合は、未来と判定する。または、動詞の上位概念が「30f83e」「行為」、「30f7e6」「自然現象」、「3f9856」「変化」の何れかであり、「30f7a2」「事象にかかわる頻度、時間、速度、程度の様態」が係っていなければ未来と判定する。

テンス	過去		現在		未来	
アスペクト	状態	非状態	状態	非状態	状態	非状態
述語の上位概念	状態	非状態	状態	非状態	状態	非状態
完結相	-	タ	-	-	-	ル
非完結相	タ	テイタ	ル	テイル	(ル)	(テイル)

出典: 『日本語の時制とアスペクト』 判定方法1 判定方法2

「居る」の上位概念

- 概念
- 事象
- 状態
- 存在状態
- 居る

図3 テンスとアスペクトの分類

4 ムード

ムードとは、事態や相手に対する話し手の判断・態度を表す文法形式である。日本語文法書でのムードの分類と、EDR 属性子の対応表が表2 である。本研究では 13 種類の属性子を用意した。この対応の一例が表2 である。本研究では、このようにムードを判定し、EDR 属性子を文節につける。

表2 ムードの小属性と EDR 属性子対応表

大属性	小属性	EDR属性子	例	小属性	EDR属性子	例
ムード	断定	conclude	だ、です、である	意志	will	～(よ)う、つもりだ
	非断定	seem	だろう、まい	願望	want	たい
	外観・兆候	appearanceL	そうだ	命令	imperative	なさい、な
	確信	sure	はずだ	可能性	maybe	かもしれない
	状況からの判断	appearanceS	ようだ	勧め・忠告	recommend	～ほうがいい
	義務・必要	duty	～なければならぬ	否定	not	な、ぬ、まい
	伝聞	rumor	そうだ			

4.1 ムードの判定方法

事例調査した結果、ムードは形態素の基本形、品詞、活用形、活用型から判断できることが分かった。例えば、「外観・兆候」の「そうだ」は、基本形が「そうだ」、JUMAN 品詞が「接尾辞」、JUMAN 詳細品詞が「形容詞性述語接尾辞」である時、「外観・兆候」と判断し、EDR 属性子「appearanceL」を文節に加える。また、「伝聞」の「そうだ」は、基本形が「そうだ」、JUMAN 品詞が「助動詞」、JUMAN 活用型が「助動詞そうだ型」である時、「伝聞」と判断し、EDR 属性子「rumor」を文節に加える。

5 おわりに

本研究の手法により、従来の SAGE に比べ、受身表現を含む文の深層格の解析精度と、使役表現を含む文の語意の解析精度が向上した。評価事例は、EDR コーパスと文法書の例文(使役文:83 文、受身文:55 文)から無作為に抽出した文章に対して行った。

表3 本研究の評価実験の結果

	深層格の精度		語意の精度
	受身の主体格	動作の主体格	使役動詞
ロジック組み込み前	67.0%	42.0%	45.8%
ロジック組み込み後	91.0%	76.0%	81.9%

表3 の受身動詞の深層格(「受身の主体格」と「動作の主体格」)の精度と、使役動詞の語意の精度の正解率の結果に示したように、本手法の有効性を示すことができたと言える。

参考文献

[1] 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).