

イベントの相互関係を表すシステム*

入江 和磨 横山 晶一

山形大学大学院理工学研究科

1.はじめに

イベントの相互関係とは、「Aする前にBする」型の文中に含まれる2つの動詞の関係のことである。人間は「歯を磨く前に寝てしまった」を「寝たせいで歯を磨けなかった」と解釈する。このような言語理解を機械的に実現することで、人間の動作にかかわる知識を蓄積し、データベース化できる。動詞を分類した言語資源には、大規模格フレーム[1]や動詞項構造シソーラス[2]などが存在する。これらには、格や意味分類が記述されているが、どのような状態ならば動作が可能なのかという点には着目されていない。

本研究では、動作が有効な時間を考慮して、動詞を「生⇔死」「起⇔寝」など、動詞の意味を正負対に分類した結果を用いて文を解析し、イベントの相互関係を表示する。結果、システムの性能は目標に達しなかった。しかし、辞書に動詞が登録されている場合には判定に成功し、今回作成した辞書が、格フレームでは捉えられない動詞の時間的変化を表せる可能性があることがわかった。

2.研究内容

2.1 動詞分類の必要性

動詞を正負グループに分類する必要性を述べる。例として、「歯を磨く前に寝てしまった。」という文の動作の時間的な流れを以下に示す。

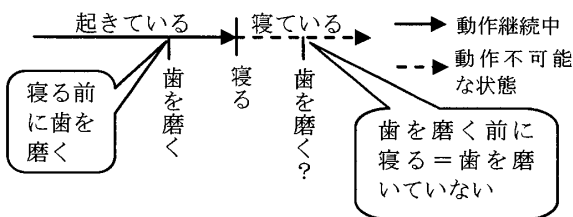


図1 時間の流れと2つの動詞の関係

図1から「歯を磨く」という動作は起きている時のみ可能である。そのため、その前に寝てしまうと歯は磨けないことがわかる。同様に「桜を見る前に亡くなった」では「見る」という動作が死んでしまうとできないため、桜を見ていないことがわかる。

2.2 動詞の分類

2.1 で述べた特徴から動詞を正負対に分類する。

(1) 正負のグループを決定

グループの決定と動詞の分類は辞書を参考に人手で行った。正負のグループは以下の通りである。

- 生きる⇔死ぬ
- 覚える⇔忘れる
- 起きる⇔寝る
- 就く⇔辞める
- 始まる⇔終わる

(2) 正負それぞれに動詞を分類

例えば「生きる」グループに対しては、生きている間にできる動作(食べる、読む、寝る、など)を、「死ぬ」グループには「死ぬ」の類義語(亡くなる、死亡する、など)を分類する。同様に「生きる」グループ内の動詞に対して、「起きる⇔寝る」のように階層的な動詞分類をおこなう。動詞の語義は1つに限定し分類した。人間の行動を対象に動詞を収集するので、存在を表す「ある」「いる」、単独では意味を持たない「する」などは除外する。図2で動詞とグループ分けの概念を示す。

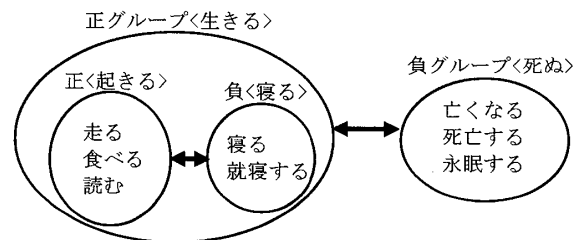


図2 動詞分類のイメージ

2.3 動詞分類辞書の具体例

表 1 は動詞分類辞書の登録例である。動詞のグループ辞書の登録数は、約 530 語である。

表1 動詞辞書エントリ

グループ	分類	包含関係	登録動詞	語数
生⇔死	生	-	産む、成長する	28
	死	-	死ぬ、亡くなる	25
起⇔寝	起	起⇔始	読む、食べる	413
	寝	寝⇔始	寝る、就寝する	10
覚⇔忘	覚	覚⇔起	覚える、伝える	10
	忘	忘⇔起	忘れる、失念する	4
就⇔辞	就	就⇔生	習う、働く	34
	辞	辞⇔生	辞める、引退する	4
始⇔終	始	始⇔生	始める、続く	4
	終	終⇔生	終わる、止める	7

このような構造にすることで、例えば、

• 動詞「読む」⇔グループ[起]⇔グループ[生]

• 動詞「覚える」

⇔グループ[覚]⇔グループ[起]⇔グループ[生]

*A System Representing Interrelation among Events

IRIE Kazuma & YOKOYAMA Shoichi, Yamagata University

のように、動詞が階層的に検索できる。

2.4 システムの動作例

「桜を見る前に彼は亡くなった。」という文を入力例としてシステムの動作を説明する。図3はシステムの処理過程を示す。

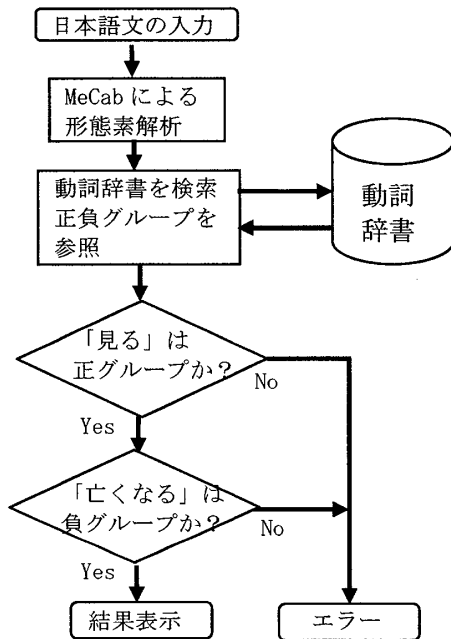


図3 システムの処理過程

- (1) 文を入力: 桜を見る前に彼は亡くなった。
- (2) MeCab[3]を使って文を解析
- (3) 動詞「見る」が動詞辞書の正グループかを検索
- (4) 動詞「亡くなる」が動詞辞書の負グループかを検索
- (5) 「見る⇔亡くなる」が「生⇔死」の対立であるので「見る」動作は行われていないと判断

3. 実行結果・評価

システムを評価するため、予備実験を行った。正誤の判断は人手で行い、入力文とシステムの結果を比較した。

結果、判定に成功した文、辞書に登録がなく判定できない文、今回提案したアルゴリズムが適用できない文が存在した。

(1) 判定に成功した文

・メールする前に寝た。

「寝たせいでメールをしていない」と解釈できる。システムでは、以下のように動詞を検索する。

動詞「メールする」∈グループ【起】

動詞「寝る」∈グループ【寝】

検索した結果、「メールする⇔寝る」の関係が「起⇔寝」の関係と一致するため、システムでも人間と同じように判断する。

(2) 辞書に登録がない文

・打ち明ける前に亡くなられた。

「亡くなったために打ち明けられなかった」という解釈できる文だが、現時点で動詞「打ち明ける」が辞書に登録されていないため、エラーとなり判定に失敗した。

(3) アルゴリズムが適用できない文

・建築基準をクリアする前に許可されていた。

提案したアルゴリズムは、「BするとAできない」という動詞の関係を利用した。「(基準を)クリアする」と「許可する」という2つの動詞には、「許可すると基準をクリアできない」という関係は存在しないのでアルゴリズムが適用できない。また、表1で示した「生⇔死」「起⇔寝」「就⇔辞」「始⇔終」「覚⇔忘」の5グループの対立関係に該当しない。そのため、イベントの相互関係の判断に失敗する。

4. 問題点と今後の課題

本研究では、人間の動作とその動作が可能な状態に着目した辞書を作成し、動作が行われたかどうかを表すシステムを作成した。その結果、システムの性能は高くなかった。原因として、辞書に単語がない場合に判別失敗することや、今回提案したアルゴリズムが適用できない例があることが挙げられる。問題点として、辞書作成を人手で行っているため非効率であることが挙げられる。

しかしながら、今回作成した辞書は、適切な動詞が登録された場合には、格フレームでは捉えられない動詞の意味の時間的変化を表せる可能性があることがわかった。

今後は、動詞の登録数を増やし、システムの性能を向上させることを目指す。動詞分類については、機械学習や、クラスタリング手法を用いることで効率よく分類する必要がある。3の(3)のような文には「AするとBできる」という新たな動詞の関係をシステムに与える必要がある。

参考文献

- [1] 河原大輔, 黒橋禎夫.
“高性能計算環境を用いた Web からの大規模格フレーム構築”. 第 171 回自然言語処理研究会 pp.67-73(2006)
- [2] 竹内孔一, 乾健太郎, 竹内奈央, 藤田 篤.
“意味の包含関係に基づく動詞項構造の細分類”. 言語処理大会第 14 回年次大会(2008)
- [3] 奈良先端科学技術大学院大学 松本研究室,
高速形態素解析器システム “MeCab”.