

動詞語釈文を利用した動詞とその結果状態を表す形容動詞との対の 収集および人手による整理

石原 靖弘 竹内 孔一

岡山大学大学院 自然科学研究科

{ishihara, koichi}@cl.cs.okayama-u.ac.jp

1 はじめに

動作が引き起こされた後に状態の変化を伴うことがあり、その変化した状態を結果状態と呼ぶ。動詞には結果状態を持つものがあり、動作と結果状態を結びつけるデータを構築することができれば言語理解に有効であると考えられる。文献[1]によると動作の前状態、結果状態を明確に構造化することができればスクリプトを通して、プラン推定や意図理解、談話予測、談話理解に効果的であることが述べられている。一方、動詞の分析や動作と結果状態を結びつける研究[4]が進められている。文献[1]では人手で正確な知識として与えていたが言語の知識として動作と状態の関係を書くことができれば、スクリプトの自動獲得やプラン推定、談話予測、行動予測に有効な知識ベースになると考えられる。

ところで動作の結果状態には1語で表現可能なものとそうでないものがあり、前者はその状態を形容詞や形容動詞などで表現できる。例えば動詞「完治する」の結果状態は「病気や怪我が完全に治った状態」であり、その状態は形容動詞「健康だ」に対応付けることができる。一方で動詞「歩く」の結果状態は「歩いた状態」という他なく、「歩いた状態」に対応する語はない。本研究では表層における対応関係を記述することを目標に、動作の結果状態を表現する語を形容動詞だけに絞り、動詞とその結果状態を表現可能な形容動詞とを対応付けたデータの構築を行った。

2 動詞形容動詞対の獲得

2.1 動詞語釈文を利用した候補の獲得

本研究では結果状態の関係にある動詞と形容動詞の組を獲得するにあたり、国語辞典中の動詞語釈文に着目した。例えば、動詞「全快する」と形容動詞「健康だ」には結果状態の関係があると考えられるが、Lexeed[3]における「全快する」の語釈文は「病気や傷が完治し、元の健康な体になること。」とある。このとき「全快する」の語釈文にはその結果状態を表す形容動詞である「健康」が含まれている。このように動詞の語釈文にはその動詞の結果状態に関する形容動詞が含まれていることがある。これを利用し動詞とその動詞語釈文中の形容動詞の組を、結果状態の関係にある動詞と形容動詞の候補として集める。集めた候補には関係のないものも含まれているため候補の絞り込みを行い、結果状態の関係にある動詞と形容動詞の組を獲得する。

語釈文を利用する国語辞典にはLexeedを用いた。対象とするのはLexeedに含まれる動詞6777語(12998語義)、形

容動詞1673語(2175語義)である。ここで、動詞と形容動詞を結びつける単位は動詞形容動詞共に語義を単位とした。動詞も形容動詞も多義性を持つことがあり、語義が異なれば対応関係も変わってくる。そのためどの語義において対応しているかを識別できる必要があり、対応付けは語義単位である必要がある。動詞12998語義の語釈文を形容動詞2175語義の形容動詞語幹で検索し、その形容動詞語幹を語釈文に含む動詞とその形容動詞の組を候補として集めた。これにより8785組の動詞と形容動詞の組を候補として得た。

2.2 人手による候補の絞り込み

集めた候補の手作業による絞り込みを作業員一人で行った。絞り込みは動詞および形容動詞の語義を考慮して関係があると思われる候補を残した。項の対応関係の有無や形容動詞が述語として用いられるかどうかは考慮していない。考慮していないことによる構築したデータが持つ特徴は3.4節で説明する。また、絞り込みで除外する対象には形容動詞語幹が語釈文中の語の一部に含まれるために候補として選ばれたものがある。例えば、形容動詞「不安だ」の語幹である「不安」は語釈文中に「不安定」などの語が含まれていても候補として選択される。このように語釈文に形容動詞そのものが使われていない明らかな誤りも含まれるので除外する。関係の有無の判別にあたっては、判断に迷った場合にはその候補は残すこととした。絞り込みにおいてはデータの信頼性よりも可能性のあるものの取りこぼしを防ぐことを重視した。この絞り込みにより8785組の候補のうち1609組を関係のある動詞形容動詞対として得た。

2.3 結果状態以外の関係の獲得

この絞り込みの目的は動詞とその結果状態を表す形容動詞の組を獲得することだったが、絞り込みを行うなかで結果状態以外の関係も含まれていることに気が付いた。そこでそのような関係にある動詞と形容動詞も構築するデータに取り込むことにした。結果状態以外の関係は次の2つに分けることができた。ひとつは「同義」である。例えば動詞「要る」は形容動詞「必要だ」と同義の関係にあり、「パスワードが要る」は「パスワードが必要だ」に言い換えることができる。このような動詞と形容動詞を述語とする文が互いに等価に置き換え可能であるような動詞形容動詞の組を「同義」の関係と定義した。もうひとつの関係は「前提条件」である。例えば動詞「スカウトする」と形容動詞「有能だ」は前提条件の関係にある。「彼は有能だ」という状態は「彼をスカウトする」という動作の前に満たされている状態と考えることができる。このような形容動詞で表される状態が動詞で表される動作の前に成立していると考えられるような関係を前提条件

とした。また、動詞「授かる」と形容動詞「大切に」のように動作とは直接関係ないものも前提条件に含める。例えば、「夫婦は子供が大切に」という状態は「夫婦が子供を授かる」という動作自体には関係しない。しかし、「授かる」という動詞は対象が動作主にとって「大切に」と発話者が認識しているときに使用される語であり、発話者が動作主は対象が「大切に」と考えていないときには使われることはない。このように、動詞の表す動作に直接は関係しないが、発話者がその形容動詞によって表される状態を認識しているときにのみ使われる動詞と形容動詞の関係も前提条件とした。

3 獲得した動詞形容動詞対に対する関係の付与

絞り込みにおいて前章で述べたように結果状態以外の関係にある動詞形容動詞対も残したが、具体的にそれらどの関係に分類されるかまでは決定していなかった。そこで3人の作業員により獲得した1609組の関係を分類した。3人の作業員がそれぞれ1609組の動詞と形容動詞の組を、関係がないと改めて判断された場合に付与する「関係なし」を含め、結果状態、同義、前提条件、関係なしの4つに分類した。

3.1 作業手順

作業員には動詞と形容動詞の見出し語およびLexeedにおける語釈文が与えられ、関係の付与においてはその動詞および形容動詞の語釈文の表す語義において適切と思われる関係を付与する。作業員は動詞と形容動詞の組に対し、結果状態、同義、前提条件の4つのうち1つを選択し付与を行う。また、関係の付与にあたって作業員の付与結果に対する確信の度合いを測るため、作業員は付与した関係がどの程度尤もらしいと思うかを3段階で評価する。

付与にあたっては、文脈によって関係の有無が変化する場合も関係があるものとみなすものとした。文脈によって関係の有無が変化する例として、動詞「温める」と形容動詞「熱々だ」という組について考える。「温める」と「熱々だ」の間には結果状態の関係があると考えることができる。しかし、両者の間に常に結果状態の関係が成り立つとは限らない。例えば、「スープを温めた」という文があったときその「スープ」の状態は「スープは熱々だ」と表現することができる。しかし同時に「スープを温めたが温まらなかった」という文も存在することができる。この場合「スープ」の状態は「スープは熱々だ」と表現することはできない。また文の前後の文脈だけでなく、動詞が項にとる名詞という文中の文脈も関係の成立に影響することがある。同じく「温める」と「熱々だ」の組を考える。このとき「アイスを温める」という文を考えるとその結果状態を「アイスは熱々だ」と表現することはできない。アイスを温める」という文自体は成立するが、「アイスは熱々だ」と表現できる状態で存在することはないためである。このように「温める」という動詞の結果状態は必ずしも「熱々だ」と表現できるわけではなく、文の前後や文中の文脈によってその関係が成立するかどうかが決まる。動詞と形容動詞間関係が必然的に成立するものだけでなく、「温める」と「熱々だ」の組のようにあくまで可能性のひとつとして成立する関係も残すこととした。例として挙げた「温める」と「熱々だ」のように必然的に成立するわけでない関係を結果状態として扱い必然的に成立するものと区別しないことで応用的な処理を行う場合に問題になる可能性はある。しかし

		作業員 A				
		結果状態	同義	前提条件	関係なし	合計
作業員 B	結果状態	631	69	60	0	760
	同義	39	286	74	0	399
	前提条件	53	59	297	0	409
	関係なし	5	3	33	0	41
	合計	728	417	464	0	1609

表 1 関係付与結果 (作業員 A-B)

		作業員 A				
		結果状態	同義	前提条件	関係なし	合計
作業員 C	結果状態	701	214	253	0	1168
	同義	12	134	31	0	177
	前提条件	15	66	172	0	253
	関係なし	0	3	8	0	11
	合計	728	417	464	0	1609

表 2 関係付与結果 (作業員 A-C)

		作業員 B				
		結果状態	同義	前提条件	関係なし	合計
作業員 C	結果状態	737	199	207	25	1168
	同義	8	129	37	3	177
	前提条件	15	70	163	5	253
	関係なし	0	1	2	8	11
	合計	760	399	409	41	1609

表 3 関係付与結果 (作業員 B-C)

動詞と形容動詞間にある関係の定義がまだ不完全であり、抽象的な表現による定義になっているため関係の定義を作業員がどのように理解し、付与がどれだけ作業員間で一致して行えるかも分からなかった。付与する関係の種類が増えれば作業のコストが増加し、また作業員間の付与結果の一致度が下がることが考えられた。そのためまずはゆるい分類で関係の付与を行い、そのうえでどのように分類されるかを確認することとした。

データの信頼性を高めるため作業員は付与した関係の尤もらしさを記述するが、それに加えて判断を迷ったものがあればそのことについてコメントを適宜残せるようにした。付与する関係の判断に迷ったものに関する情報は、なぜ作業員にとって難しかった組を考え、今ある関係を増やすべきかなどを考慮する上で役に立つと考えられる。

3.2 付与結果

作業員3人の付与結果に関して、作業員を2人ずつ取り上げて付与結果の比較を行った。結果は表1-3の通りである。

付与結果から作業員Cが結果状態の関係に付与した数は1168であり、他の2人よりもかなり多いことが分かった。そこで作業員間の付与結果の一致度を測るため、作業員2者間の付与結果の κ 値を求めた。結果を表4に示す。

	作業員 A, B	作業員 A, C	作業員 B, C
κ 値	0.62	0.42	0.40

表 4 付与結果の作業員間の κ 値

付与結果が最も一致していたのは作業員 A と B であった。

3.3 最終的な動詞形容動詞対の関係の決定

作業者 3 人による付与結果をもとに動詞形容動詞間関係の最終決定を行った。関係の最終決定は 1 人の最終決定者が行った。最終的な付与結果として κ 値の最も高かった 2 人、作業者 A, B の付与結果のうち一致するものはそのまま最終結果として採用することとした。2 人の付与結果のうち一致していないものについては最終決定者が関係を決定した。一致していないものの最終決定においては作業者 2 人の付与結果を考慮に入れて、2 人のつけた関係のうちどちらかを選択するというようなことはせず、新たに関係を付与した。そのため 2 人の付与結果のどちらとも異なる関係を付与することがある。3 人の作業者の付与結果のうち作業者 A, B の作業者の結果を採用した理由は、作業者 A, C および作業者 B, C 間の κ 値が作業者 A, B 間の κ 値に比べ低く、作業者 A, C または作業者 B, C の付与結果の一致が偶然の一致である可能性が作業者 A, B の付与結果の一致に比べて高いことから、作業者 3 人の付与結果の多数決で最終的な関係を決定するというようなことはせず、より信頼性が高い作業者 A, B の付与結果の一致するものを使うこととした。最終的な付与結果は、結果状態が 748 組、同義が 457 組、前提条件 382 組、関係なしが 21 組で、結果状態とも同義ともれるものが 1 組あった。結果状態とも同義ともれるものは動詞を他動詞として使う場合は結果状態になり、自動詞として使う場合は同義になるものであった。対応付けにおいては動詞の自動詞と他動詞は区別していなかったため、このようなものが含まれていると考えられる。このような動詞と形容動詞との対応の存在から、対応付けにあたっては自動詞と他動詞両方の使い方ができる動詞についてはそれぞれ分けて対応付けを行う必要があるといえる。

3.4 構築したデータの特徴

動詞と形容動詞それぞれの組およびその組の関係（結果状態、同義、前提条件）を記述している。動詞と形容動詞は Lexeed における語義単位で結びつけており、Lexeed における見出し語、語釈文をそれぞれ載せている。構築したデータは関係のある動詞と形容動詞を結びつけた、言い換えなどの応用的な処理のためのデータを構築するための基礎的な知識を記述したデータであるといえる。このデータの構築においては、言い換えなどの処理は目的としておらずそのような処理の際に必要な情報の付与は行っていない。例えば、言い換えにとって必要な動詞と形容動詞の項同士の対応関係は記述していない。そのためこのデータを利用して言い換えなどを行うには項同士の対応関係の知識が必要になる。また、形容動詞は必ずしも述語となるものとは限らない。例えば形容動詞「新ただ」は「新たな生活」のように修飾語として使い、述語としては使わない。「新ただ」に対応する動詞としては「新調する」があり、「服を新調する」を「新ただ」を使い言い換えると、「服は新ただ」という日本語として違和感のある文になる。一方で、「新調した服」を「新たな服」と言い換えても問題ない。構築したデータを使い言い換えを行うには「この形容動詞は述語として使わない」などの語の用法に関する知識が必要になる。また、前述したようにデータの構築においては必然的に成立しない関係も残している。そのため構築したデータには文脈によっては関係が成立しないことがある動詞と形容動詞の組も含んでいる。現在はそのような必然的に成立するとは限らない関係も必然的に成立する関係と区別はしていない。そのようなものを区別するべきか

どうかを含めどのような関係を定義するべきかは将来的な課題であるが、それは別にしても関係が必然的に成立しないものについては関係が成立しているかどうかを判断するためにどのような文脈で関係が成立するかに関する知識が必要である。構築したデータを利用し含意関係認識や言い換えなどを行うためには、どのような文脈で関係が成立するかに関する知識を事例などから得る必要がある。

4 統計的学習モデルを利用した自動獲得の可能性

前章までで動詞と関係のある状態を結果状態、同義、前提条件という 3 つに分類し、その状態を表現する形容動詞とその動詞との組を動詞の語釈文を用いて候補を抽出し、その後人手で整理を行い動詞と形容動詞の組からなるデータを構築した。構築したデータをもとに統計的学習モデルを利用した新たな動詞と形容動詞の対、さらに本研究では扱っていない形容動詞との対の獲得を目指す。本章では、統計的学習モデルによる新たな動詞形容動詞対獲得の可能性を探るための予備実験として、構築したデータを学習およびテストに用いて実験を行った結果を報告する。

4.1 自動獲得に向けた予備実験の設定

学習モデルとして SVM を使用し学習を行った。実験は何らかの関係のある組かどうかを識別するものとどの関係にあるかを識別するものの 2 種類を行った。何らかの関係のある組かどうかを識別する実験では学習データとして動詞の語釈文を利用して収集した動詞形容動詞対 8785 組を用い、構築したデータにおいて結果状態、同義、前提条件のいずれかの関係を付与されたものを正例として、それ以外のものを負例とした。どの関係にあるかを識別する実験では学習データとして構築したデータの動詞形容動詞対 1609 組を用い、それぞれの組に対して付与されている関係を正解クラスとした。特徴量には動詞、形容動詞の Lexeed における語釈文の bag-of-words、すなわち動詞、形容動詞それぞれの語釈文中の各形態素の数をを用いた。さらに、関係のある動詞と形容動詞の語釈文同士は意味的に近い関係にあり、その語義を表現するのに同じような語句が使用されていると仮定した。そしてその意味の近さを表現する特徴量として両方の語釈文で共通する名詞の数および動詞の数をを用いた。これらの特徴量は両実験で共通して使用し、学習およびテストは 2 分割交差検定、10 分割交差検定をそれぞれ行った。

4.2 実験結果

結果は表 5-8 に示す。

結果から語釈文の bag-of-words や動詞と形容動詞の語釈文で一致する動詞の数および名詞の数といった比較的単純な特徴量では未知の動詞形容動詞対の獲得は難しいことが分かった。今回語釈文同士の意味の近さの表現するため、語釈文間の動詞と名詞の一致数を特徴量として入れたが、効果は得られなかった。しかし、動詞と形容動詞の語釈文の意味的な近さを特徴量に取り入れること自体は有効だと思われる。日本語 WordNet[5] のような語の意味の近さを求めることができる辞書を使用し、より詳細に意味の近さを表す特徴量を加えることも視野に入れている。

	適合率	再現率	F 値
関係あり	0.603 (913/1515)	0.575 (913/1587)	58.9
関係なし	0.904 (6338/7012)	0.913 (6338/6940)	90.9

表 5 関係の有無の識別結果 (2 分割交差検定)

	適合率	再現率	F 値
関係あり	0.662 (971/1466)	0.612 (971/1587)	63.6
関係なし	0.913 (6445/7061)	0.929 (6445/6940)	92.1

表 6 関係の有無の識別結果 (10 分割交差検定)

	適合率	再現率	F 値
結果状態	0.796 (631/793)	0.844 (631/748)	81.9
同義	0.679 (319/470)	0.698 (319/457)	68.8
前提条件	0.698 (233/334)	0.610 (233/382)	65.1
関係なし	0.636 (7/11)	0.333 (7/21)	43.7

表 7 関係の種類の識別結果 (2 分割交差検定)

	適合率	再現率	F 値
結果状態	0.813 (652/802)	0.872 (652/748)	84.1
同義	0.710 (331/466)	0.724 (331/457)	71.7
前提条件	0.712 (230/323)	0.602 (230/382)	65.2
関係なし	0.588 (10/17)	0.476 (10/21)	52.6

表 8 関係の種類の識別結果 (10 分割交差検定)

5 関連研究

先行研究として青山 [2] が動詞語釈文の構造化による事態間関係の獲得を行なっている。青山の研究では国語辞典の動詞語釈文中の述語と見出し語との項の対応関係および結果状態などの意味関係を付与している。見出し語の動詞に対応する述語として動詞および形容詞があることは確認したが、形容動詞を結びつけているかどうかは確認できなかった。青山の研究では動詞と形容動詞の対を獲得しているかどうかは不明であり、また辞書に岩波国語辞典を用いているのに対し、我々は辞書に Lexeed を用いたという違いはあるが、仮に青山の研究が動詞と形容動詞の対を獲得しているとするれば、我々が獲得した動詞と形容動詞の対と重複はあるものと思われる。例えば、我々の構築したデータに動詞「要る」と形容動詞「必要だ」の組があり、関係は「同義」である。Lexeed における「要る」の語釈文は「無いと不都合である。必要である。」とあり、青山の手法においても「要る」と「必要だ」の組は獲得できるものと思われる。一方で、青山の手法では獲得できないものと思われる動詞と形容動詞の組も存在する。例えば、我々の構築したデータには動詞「裏付ける」と形容動詞「本当だ」の組があり、関係は「結果状態」である。ここで「裏付ける」の Lexeed における語釈文は語釈文は「あることが確かに本当の事を、他の面から探って、証拠を作る。」とあるが、語釈文中の「本当」は述語として使われていない。このため青山の手法では「裏付ける」と「本当だ」の組を獲得することはできないものと思われる。青山の手法が構造化を行うことを主眼において動詞語釈文を利用しているのに対し、我々は動詞の語釈文は動詞と関係のある形容動詞を集めるために利用しており、それによって集めた動詞と形容動詞の組が関係のある組かどうかを判別する際に語釈文を判断材料として用いるものの、語釈文の内容を構造化してい

るわけではない。そのため「裏付ける」と「本当だ」の組のように動詞の語釈文に「本当になる」のように結果状態が明示されていない場合でも、人がその結果状態を「本当だ」と表現できると判断すれば関係のある組として獲得する。このように我々の手法は候補として集めた動詞と形容動詞の関係を個別に判断することで、語釈文で表現されている範囲を超えた関係を獲得できるものになっている。

6 おわりに

本稿では動詞とその結果状態を表現する形容動詞の組を獲得を目的として動詞の語釈文を利用して可能性のある候補を収集し、その後人手による候補の絞り込みおよび結果状態に同義と前提条件を加えた 3 つの関係を付与したデータを構築したことを報告した。3 章で述べたように構築したデータは応用的な処理のための基礎データという位置づけられ、例えば言い換えを行うときに必要になる項同士の対応関係の記述がないなど言語処理を行う上で必要な情報が不足している。我々は現在構築したデータを基に、項同士の対応関係を記述した言い換えのための辞書の構築を行なっている。また、統計的学習モデルによる新たな動詞形容動詞対を獲得を目指し、獲得したデータを学習データおよびテストデータとして利用し実験を行い、語釈文の bag-of-words といった比較的に単純な特徴量では難しいことを明らかにした。動詞と形容動詞の組を関係のあるものとそうでないもの、また具体的にどの関係を識別するためには、どのような特徴量を用いればいいかを明らかにすることは今後の課題である。また、実験では学習データとテストデータは同じものを利用し、2 分割交差検定および 10 分割交差検定を行い識別の性能を確認したが、実際に新たな対を獲得するためには対象となる新たな候補が必要である。本研究では候補の収集においては動詞語釈文中に形容動詞を含むものを候補とするという手法をとったが、類語辞書を利用して構築したデータを基に新たに候補を獲得することも視野に入れている。特に、動詞項構造シソーラスのように動詞をその結果状態の観点から分類している辞書を用いることは、すでに獲得した対の形容動詞側を固定してそれに対応する動詞の類語をその形容動詞と対となる候補として獲得することに有効であると考えられる。

参考文献

- [1] James Allen. Natural Language Understanding. Benjamin/Cummings Pub. Co., pp. 378, 1987.
- [2] 青山桜子, 阿部修也, 大西良明, 乾健太郎, 松本裕治. 事態間関係の獲得のための動詞語釈文の構造化. 言語処理学会第 13 回年次大会発表論文集, pp.286—289, 2007.
- [3] 笹原要, 佐藤浩史, Francis Bond, 田中貴秋, 藤田早苗, 金杉友子, 天野成昭. 「基本語意味データベース:lexeed」の構築. NTT コミュニケーション科学基礎研究所, pp.1-3, 2004.
- [4] 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. 言語処理学会 14 回年次大会論文集, pp.344-347, 2008.
- [5] Kyoko Kanzaki, Francis Bond, Noriko Tomuro, Hitoshi Isahara. Extraction of Attribute Concepts from Japanese Adjectives. In LREC-2008, Marrakech, 2008.