

共起関係を元にした動詞の意味分類の試み

玉野健一 松本裕治

奈良先端科学技術大学院大学 情報科学研究科
6300101 奈良県生駒市高山町 8916-5
0743 72 5248

{kenit-t,matsu}@is.aist-nara.ac.jp

あらまし

事象の時間的な性質を捉える上で、事象の主たる伝達語彙範疇である動詞の意味の分類は重要である。ここでは、抽象的な意味分類とは異なる方向から分類を行う試みとして、動詞を修飾する範疇である副詞との共起関係から動詞を分類することを試みた。その結果として、単なる共起情報だけでも、ある程度の傾向を掴むことはできることがわかった。

キーワード 動作相、動詞、副詞、共起関係。

A study of classification of verbs according to co-occurring adverb

TAMANO Ken'iti and MATSUMOTO Yuji

Graduate school of information science, NAIST
8916-5, Takayama-tyou, Ikoma-si, Nara-ken, 6300101, Japan.
+81-743-72-5248

{kenit-t,matsu}@is.aist-nara.ac.jp

Abstract

In order to understand temporal properties of events, classification of verbs, which mainly describe events, is important. Here, we try to classify verbs according to co-occurrence with adverbs, which modify verbs. As a result, we found that only with co-occurrence information, we can classify verbs, although it is a rough classification.

key words Aktionsart, verb, adverb, co-occurrence.

1 事象の分類

事象 (event) は、我々が言語によって伝達している「出来事」や「状態」などの総称である。その中核的な意味は、主に動詞によって担われている。

これら事象にもいくつかの種類がある。その分類は、主にその時間的な特徴から行われてきた。その劈頭は Vendler によって提唱された states、actions、accomplishments、achievements の四分類である ([Vendler57])。このような事象の時間的な特徴に基づく分類は Aktionsart と呼称される。

例えば、

- She is walking.
- She is dying.

の二文は、walk と die だけの差異である最小対であるが、その意味は「歩いている」と「死にかけている」となって最小対であるとは考えにくい¹。walk と die の指示している事象の差異にこの意味の差異を生起する要因があると考えられる。

このような意味の差異に対して、事象の内部の構造を考え、時制などの形態的な形式の作用をその構造に則して考えることができる。

その構造は例えば、事象に達成性 (telicity) がある場合とない場合を区別したりするような事象の性質に近いものもあり、また複数の下位事象に事象を分解して考え、それらの間の関係を論じるもの ([Pustejovsky95]) などもある。

このような事象の分類は、深い意味的な考察やそれに基いた基準を用いることによって行われるが、一方機械で処理を行う上では困難が多い。すなわち、機械による言語処理は依然表層の形式より深い段階を扱うことには大きな困難を抱えており、文の意味によって事象を分類することは不可能に近い。

そこで、コーパスからこのような情報を獲得しようとの試みがなされた。大石は、[大石97]において日本語の動詞を、「～ようとする」「～続ける」「～始める」などの表現 (アスペクト形式と呼んでいる) との共起関係、ならびに副詞との共起関係によって6つの Aktionsart (同稿中アスペクト) に分類する試みを行っている。

これは、意味的な取り扱いではなく、表層に出現する情報のみによって²動詞の分類を行う試みであ

¹ 意味の基本要素 (それがあるとして) のとりかた如何ではある。

² アスペクト形式は動詞の直後に出現するため、構文解析の必要すらない。

る。これは現存する自然言語処理のための道具によって充分実現できる方法であった。

ここではこのような手法を参考にしつつ、英語を対象として副詞との共起関係から動詞を分類する試みを行った。

2 共起関係

共起関係にある語同士は何らかの意味的な関連を持っているものと考えられる。これまでも、動詞と名詞の共起関係から名詞のクラスタリングを行うような試みはされてきた。

2.1 修飾関係による語の分類

修飾する語と修飾される語の間には何らかの意味的な関連がある。修飾語は統語的に可能な語ならば如何なる語をも修飾できるか、といえそうではない。例えば「a cup of」は「paper」を修飾できない。また「quickly」は「(I know)」を修飾しない。

修飾語は被修飾語の何らかの属性に対して、その値を詳述するための情報を付加するものである、と考えられる。そうすると、修飾語が与える情報に適合した属性を被修飾語が持っていなければ修飾関係は成立しない。

つまり、修飾語は被修飾語の意味的なクラスを示唆するといえることができる。³

さて、副詞の主な修飾先は動詞である。副詞は動詞に比べてその語彙数が少なく、修飾する副詞という観点から修飾される動詞の意味に関する分類が可能ではないかと考えられる。勿論、副詞は動詞の他にも、形容詞、副詞、あるいは文全体など、色々な統語範疇を修飾することがあるが、動詞を主に修飾する副詞というものを選ぶことはできると思われる。

本稿では、主に時間的な意味合いを持ち、動詞を修飾することが多いであろう副詞を選んで、それに基づいて動詞を分類する試みを行った。従って得べき動詞の分類は、時間に関するものである。

なお、ここでは修飾関係の近似として共起関係を捉えた。本来は修飾関係は構文解析或いは掛受解析などを行わなければ抽出できないものである。ただ、修飾語を注意して選べば、この近似はそれほど

³ 恐らく逆方向も可能であると思われるが、一般に修飾語の方が被修飾語よりも数が少ないため、逆方向はあまり役に立ちそうにない。

大きな誤差を発生しないのではないかと考えている。

3 実験

以上のような観点から、実際のデータに照して副詞との共起関係から動詞の Aktionsart 的分類が行えるかについての簡易な実験を行った。

3.1 材料と方法

本実験では、英語を対象とした。本来、多くの言語においてこのような実験を行って比較対照したいところであるが、処理するための道具の制限から、残念ながら英語以外の言語を対象にできなかった⁴。

実験は以下のように行った。

1. コーパスを前処理し、形態素解析を施す。
2. 副詞を抽出し、その最近傍の動詞を選ぶ。
3. 共起頻度に従って動詞を序列化する。

コーパス

今回用いたコーパスは、Sir Arthur Conan Doyle の Sherlock Holmes シリーズ全巻 ([Doyle1887] 他) である。Oxford Text Archive⁵ に収録されていたもので、約 67 万語 (約 3.5MB) からなる、19 世紀末から 20 世紀初頭の小説である。

これを用いたのは、入手の容易さに加え、新聞や雑誌などから構成されたコーパスに比べて、物語の進行や事柄の推移などを表すことが多く、従って事象の分類に有用なデータが得られるのではないかと推定したためである。

新聞記事と比較したところ、副詞の割合がおおよそ三、四割方高いという結果も得られた (表 1)。

	副詞数 / 全語数	割合
Sherlock Holmes	29140/667249	0.0437
朝日新聞社説英語版	22205/676219	0.0328
Wall Street Journal	45213/1985267	0.0228

表 1: 副詞の数の対照

⁴ 日本語でも可能であるかとも思えるが、ここでは基本的に印欧語を対象としている。

⁵ <http://www.ota.ox.ac.jp/>

前処理と形態素解析

形態素解析には、Brill Tagger ([Brill94] 他) を用いた。同形態素解析器は一行一文などの仮定があるため、入力に適合するような前処理を行った後、形態素解析を行った。なお、ここでは形態素解析の誤りについては一切の顧慮を行わなかった。

副詞と動詞の抽出

形態素解析の結果から副詞を抽出した。副詞は 29140 語抽出され、異り語数は 833 語であった。あまりに出現頻度が低いものは雑音の元になるのではないかと考え、出現頻度が 10 に満たないものをまず切捨てた。

出現頻度が 10 以上の副詞について、主に動詞を修飾すると考えられ、かつ今回の実験の目的である時間的な内容を有する副詞を選抜した。選抜した副詞は 43 語である。この部分については筆者の主観によって行った。こうして選抜された副詞を本稿の末に掲げた。

続いてコーパスを形態素解析した結果から、これらの副詞を検索した。得られた副詞から、前後 5 単語以内の最近傍の動詞を検索し、発見した場合、それを共起している動詞として選抜した。この際、be、have、do、およびこれらの活用形は除外して、更に動詞を検索した。こうして得られた対は 4371 あった。

結果と評価

前節において得られた対について、最も多く出現した「heard」を基準に、共起頻度のパタンに近いもの順に動詞を並べた。近さの尺度としては、一つの動詞に対する共起副詞の頻度を並べたものを 43 次元のベクトルと考え、長さを正規化した後、これらの間の距離を計算して用いた。結果を図 1 に示す。図中、adverb は副詞の、verb は動詞の番号である。

また、夫々の副詞について、共起頻度の高いものから順に動詞を挙げ、これも考察の材料とした。

この結果を元にして、若干の考察を行った。

4 考察

分布図を観察すると、ある程度の排他的なまともりが観察できる。楕円で囲んだ部分がそれである。

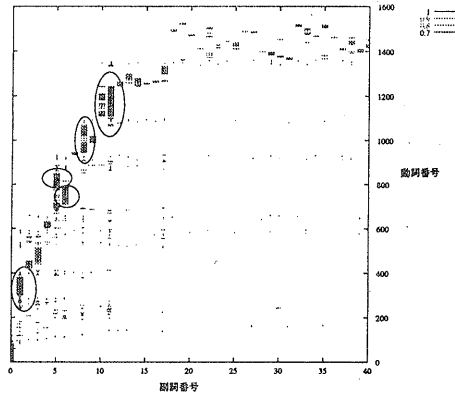


図 1: 共起頻度の分布

すなわち、これらの動詞と副詞の間には比較的強い共起関係が存在していると考えることができ、これは動詞の区分けに有用な情報であると言える。

4.1 共起

次に、夫々の副詞とよく共起した動詞の中で特徴的なものを掲げて考察を加える。

suddenly

turned, sprang, vanished, broke, bent, burst

ここに多く現れた「曲がる」「跳ねる」「消える」などの動詞は、瞬間的あるいは変化に類する事象を表わす動詞である。suddenly は事象の生起が突然であるという情報を付与する副詞である。従ってこのような副詞と共起する動詞は、前触れなどがある可能性のある事象を表わすものであると考えられる。

once, ever

saw, see, seen, heard, said

ここに現れた動詞には、感覚に関するものが多い。「見る」や「聞く」などである。once や ever といった副詞は、過去の経験の有無に関する情報を付加するものである。感覚の動詞は経験を表わすことがあると考えられ、こうした動詞は once や ever といった副詞と共起することが多い。

already

convinced, reached

already は、ある時点⁶において事象が既に起っているという情報を付加する副詞である。一方、already と共起した動詞群で特徴的なのは、ここに掲げたある状態に達したことを表わす動詞である。already と共起するものの特徴には、達成性 (telicity) を有しているということが考えられる。

still

lay, left, knew, lying

ここに現れた動詞には、状態を表わすものが多い。また、lying といった現在分詞、すなわち進行時制が出現したことも特徴的である。still は過去から現在に到るまでの事象の継続という意味を付与する副詞であると考えられ、そういったことが可能な事象、すなわち状態を表わす動詞とよく共起するものと考えられる。

slowly

walked, moved, said

ここに現れた動詞は、主に何らかの動作を表わすものである。slowly は、動作が進む速度に関する意味を付与する副詞であると考えられる。従って、動作を表わす動詞とよく共起すると考えられる。

⁶ Reichenbach[Reichenbach47]に言う Reference point.

その他

suddenly などの場合、過去形と過去分詞形が違う動詞の場合は、過去形が現われている。これは、完了時制に共起することが稀であることも示している。

このことは、時制との共起情報を加味することによって、もう少し有効な分類ができるであろうことを示唆している。

4.2 結論

以上のように、副詞との共起関係を用いるだけでも、ある程度分類ができることがわかった。これは、動詞に対して事象の構造などの深い意味的考察を加えることを行わなくても、より数の少ない副詞などの範疇について考察することで、分類をすることが可能であることを示唆している。また、動詞そのものではなく、副詞という修飾範疇を視点とした分類方法であるとも言える。

5 今後の課題

今回の実験では副詞句や動詞句といった句の階層の情報が全く利用できなかった。これによって、例えば「for a while」などの副詞句が取得できなかった。また、動詞句についても、完了形などにおいて若干の考慮はしたものの、不正確さが増した。更に近傍の動詞を修飾先であると仮定している。これらの要因によって共起関係を修飾関係の近似と考えた場合の誤差が増してしまったと考えられる。やはり形態素解析だけでなく、ある程度の構文解析を行うことが必要である。Penn tree bank などを利用することも考えている。

また、副詞の分類については、より精密な考察を経たものを用いる必要がある。

今回用いたコーパスでは、最終的に得られた結果における共起頻度が最大で18と、定量的な評価を行うには少な過ぎた。より大規模なコーパスで行う必要がある。

副詞や副詞句だけでは捉えきれない動詞の意味範疇がある。例えば「walk」と「walk to the station」では、その共起制限が異なる。このように、名詞句との関係も今後考察すべきである。

6 まとめ

本稿では、副詞との共起関係という、表層的な情報に基いて動詞の分類を行う試みについての報告を行った。その意義について述べ、未だ定性的なものに留まるものの、若干の実験を行ってそれに基づく知見を述べた。

これは、Aktionsart等の動詞の意味の分類に関する機械処理からのアプローチであると言え、深い意味的な考察などによって生成された理論的な枠組の検証等に用いることができると考えられる。

本稿における実験程度では依然精度は不十分であるが、このような動詞のAktionsart情報は、事象の連続における背景事象と前景事象や、事象の繋がり方、あるいは事象間の前後関係などの知識の抽出などに利用することができるものと考えられる。

謝辞

材料収集について助言くださった Dee A. Worman 博士、ならびに usenet news の読者の方々に感謝を申し上げます。

参考文献

- [Brill94] Brill, E.: Some advances in rule-based part of speech tagging, *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)* (1994).
- [Doyle1887] Doyle, A. C., Sir: *A Study in Scarlet* (1887).
- [Pustejovsky95] Pustejovsky, J.: *The Generative Lexicon*, The MIT Press, Massachusetts, USA (1995).
- [Reichenbach47] Reichenbach, H.: *Elements of symbolic logic*, Collier-Macmillan, London, England (1947).
- [Vendler57] Vendler, Z.: Verbs and Times, *The Philosophical Review*, Vol. 66, pp. 143-160 (1957).
- [大石 97] 大石亨, 松本裕治: コーパスからのアスペクト情報の獲得, 情報処理学会研究報告(自然言語処理研究会), NL-118, pp. 15-20 (1997).

A 採用した副詞一覧

730	now	40	rapidly
466	once	38	occasionally
454	yet	38	late
379	ever	37	quickly
368	still	37	forever
305	always	35	immediately
268	already	28	first
257	again	25	frequently
173	suddenly	24	presently
161	soon	21	halfway
134	later	21	continually
110	ago	19	gradually
92	often	19	early
89	slowly	18	eventually
78	finally	16	thoroughly
67	sometimes	15	recently
61	twice	14	steadily
53	instantly	14	fast
46	usually	13	speedily
41	to-morrow	12	upward
40	shortly	11	invariably
		10	newly