

カルテの文章における副作用表現の分析

大熊智子⁽¹⁾ 小山田由紀⁽²⁾ 外池昌嗣⁽¹⁾ 三浦康秀⁽¹⁾ 増市博⁽¹⁾

荒牧英治⁽²⁾ 篠原恵美子⁽³⁾ 大江和彦⁽³⁾

⁽¹⁾富士ゼロックス(株)研究技術開発本部コミュニケーション技術研究所

⁽²⁾東京大学知の構造化センター

⁽³⁾東京大学医学部附属病院

1. はじめに

近年テキストマイニングの手法が盛んに研究されている。医療分野においても、大量に存在する電子カルテから有用な情報を獲得することを目標としてツールの開発がなされている。

我々は医薬品投与による副作用症状の出現状況の調査を支援するために、カルテを対象にした言語処理技術の研究開発を行っている(今までの研究については参考文献[4][5])。

副作用は医療行為において必ず発生する現象ではない。カルテのテキスト上において副作用の記述は頻出しないため、機械学習のみでは副作用表現の特定が難しい。そこで、本稿では現状の副作用関係抽出の性能を向上させることを目指し、副作用について記述されている文の構造について分析を実施した。

本稿では、**副作用表現**を「医薬品の投与によって本来の目的とは異なる別の有害な作用が生じ、症状が出現することを述べる表現」と定義する。例えば、「医薬品A投与により蕁麻疹が出現した」というような表現を副作用表現と呼ぶ。

一般のテキストでは、イベントについてのパターンを網羅することは非常に困難であると思われるが、医療テキストでは、時間軸に沿って記述されるという性質、かつ、医師が患者について記述するという性質から、見通しのよい整理が可能である。本稿では、副作用表現を

((副作用が患者に出現し) 始めた) ことを医師が認識した)

といった入子の構造として捉え、そのパターンと構成する述語の整理を行った。

2. 分析

2.1. 対象データ

本稿で扱う退院時サマリとは、患者の入院から退院までの経緯が簡潔に記載された文章である。時間軸に沿って、患者の状態、治療内容、検査結果などのうち臨床的に重要だと判断されたものが記載される。そのデータから副作用表現を含む文(合計203文、副作用表現264か所)を手で洗い出し、それらを対象に分析を行った。なお、前章の定義より、分析対象文には必ず副作用症状とその原因となった医薬品が含まれる。

2.2. 結果の概要

副作用表現の構文パターン(本稿では**副作用パターン**と呼ぶ)は、医薬品の投与(**投薬イベント**)と症状の認識(**症状イベント**)という2つのイベントから構成されるものとする。本稿では、副作用パターン(2.2.3項)、投薬イベント・症状イベントのパターン(2.2.2項)、各イベントを構成する述語とその分類(2.2.1項)について述べる。

2.2.1. 述語の分類

まず、もっとも基本の単位である述語について分類すると、以下の3種に分類できる。

- ◆ **Drug 述語 (D 述語)**
: 医薬品を格にとり治療行為を表す。
(例) 内服、投与、開始、中止
- ◆ **Symptom 述語 (S 述語)**
: 症状を格にとり症状の状態を表す。
(例) 出現、悪化、改善、認める、疑う
- ◆ **Relation 述語 (R 述語)**
: 医薬品と症状の両方を格にとり、それ自体で副作用表現となる。

(例) 起因、誘発、関与

本稿では、R 述語については詳しく述べない。実例は表 1 を参照されたい。

2.2.2. イベントの構文パターン

次に、個々のイベントを構成する構文パターンを分析した。

南(1993)や益岡(1989)、野田(1989)らに代表される文の階層構造という考え方がある。野田(1989)では、述語成分を [実質的意味] - [ボイス] - [アスペクト] - [肯否] - [テンス] - [ムード] という階層構造で考えるが、動詞や形容詞などの内容語が実質的意味を、助動詞や補助動詞などのいわゆる機能語が他のカテゴリを担う。

分析対象とした副作用表現には出現する機能語の種類が少なかったため、本稿ではまず前項で分類した述語について意味・出現の仕方などを分析した。その結果、主たるイベント・極性・態度という3つのレベル¹が存在することが明らかになった。副作用表現に現れるイベントは、これらの述語による階層構造を持つ。図 1 を参照されたい。

D 述語から成るイベント(投薬イベント)では、治療行為という主たるイベントが核となる(レベル[D1])。その外側に極性を表す述語が続き(レベル[D2])、その更に外側にカルテの筆者である医師もしくは患者の態度を表す述語が続く(レベル[D3])という3層構造となる。S 述語から成るイベント(症状イベント)では、治療行為の代わりに、主たるイベントが症状の状態となる(レベル[S1])。その外側に続く極性(レベル[S2])、態度(レベル[S3])はD述語と同様である。

また、この3つのレベルの述語全てが必ず出現する訳ではなく省略可能である。たとえ省略されたとしても、主たるイベント・極性・態度という3層構造は保たれ、意味が失われることはない。

実際にどのような述語があるかは、結合価も併せてリスト化した。抜粋ではあるが、表 1 を参照されたい。

2.2.3. 副作用パターン

前述したとおり、副作用表現は投薬イベントと症状イベントから構成される。副作用パターンの基本形は、前項に述べたイベントの間に接続表現が入る形である。その接続表現が2つのイベントの関係を表す。前後のイベントが投薬イベントと症状イベントのどちらなのかによって大きく2つの構文パターンに整理できる。

- ◆ **方針変更**: 症状イベントが前に、投薬イベントが後に記載される。症状を認識し、それを根拠として治療行為の方針変更がなされる。
- ◆ **観察**: 投薬イベントが前に、症状イベントが後に記載される。治療行為がなされ、それから、症状の状態が観察される。

更に、観察は前後のイベントの関係の強さの違いによって、以下の2つに分類できる。

- ◆ **前後**: 前後のイベントの関係について、単に時間的前後関係しか述べていない。
- ◆ **因果**: 前後のイベントの関係について、積極的に前のイベントが原因だと述べている。

以上3つの構文パターンが代表的なものである。図 2 を参照されたい。

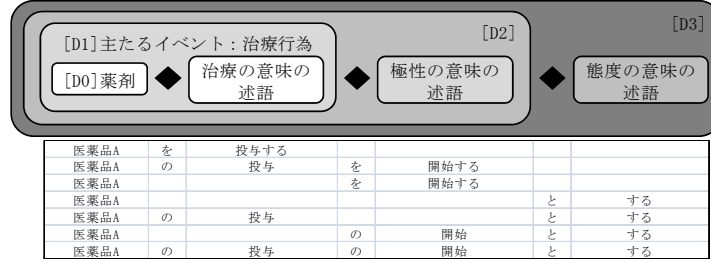
本稿での接続表現は全く品詞に依らず、接続助詞・格助詞・名詞やそれらの連語など様々なものを想定している。前後のイベントを関連付ける表現全般を一括りに扱っている。

また、分析対象データには、もちろん「副作用」などの直接的な表現を使った記述も見られる。それらを[観察-因果]のサブパターンと位置付ける。ただ、本稿では、紙幅の関係でパターンは記載できなかった。

¹野田の用語の以下に相当する。
・主たるイベント: 実質的意味
・極性: 肯否とアスペクト
・態度: ムード

【投薬イベント】

※ ◆は主に表1の結合値で許される格に相当する。



【症状イベント】



図1 副作用表現に現れるイベントの述語による階層構造
表1 述語の分類リスト (抜粋)

見出し	分類				意味素性	結合値							極性(時間)	態度	備考	
	表現の種類	意味	方向	増減多少		が	を	から(→)	まで	に	へ(へと)	と				
内服投与	D	治療	-	増	[D1]	-	[D0]	-	-	-	-	-	-	-	-	
	D	治療	-	増	[D1]	-	[D0]	-	-	-	-	-	-	-	-	
減量	D	治療	-	減	[D1]	-	[D0]	[num]	[num]	-	-	-	-	-	-	
	D	治療	-	減	[D1]	-	[D0]	[num]	-	[num]	-	-	-	-	-	[num]: 数値。
悪化	S	症状	悪	増	[S0]	[item]	-	-	-	-	-	-	-	-	-	[item]: 「血圧」「血糖」などの検査項目。
	S	状態	悪	増	[S1]	[S0]	-	-	-	-	-	-	-	-	-	
改善出現	S	状態	良	増	[S1]	[S0]	-	-	-	-	-	-	-	-	-	
	S	状態	悪	-	[S1]	[S0]	-	-	-	-	-	-	-	-	-	
開始	D	極性	-	増	[D2]	-	[D0]	-	-	-	-	-	-	開始	-	
	D	極性	-	-	[D2]	-	[D1]	-	-	-	-	-	-	開始	-	
中止	D	極性	-	減	[D2]	-	[D0]	-	-	-	-	-	-	終了	-	
	D	極性	-	-	[D2]	-	[D1]	-	-	-	-	-	-	終了	-	
認める	S	態度	悪	-	[S3]	-	[S0]	-	-	-	-	-	-	-	確実	
	S	態度	-	-	[S3]	-	[S1]	-	-	-	-	-	-	-	確実	
	S	態度	-	-	[S3]	-	[S2]	-	-	-	-	-	-	-	確実	
考える	D	態度	-	-	[D3]	-	[D1]	-	-	-	-	-	-	-	不確実	[D0]は格要素にならない。ヲ格よりト格の方が確信度が高い。
	D	態度	-	-	[D3]	-	[D2]	-	-	-	-	-	-	-	不確実	
	D	態度	-	-	[D3]	-	-	-	-	-	-	-	[D1]	-	確実	
	D	態度	-	-	[D3]	-	-	-	-	-	-	-	[D2]	-	確実	
	S	態度	悪	-	[S3]	-	[S0]	-	-	-	-	-	-	-	不確実	
	S	態度	-	-	[S3]	-	[S1]	-	-	-	-	-	-	-	不確実	
疑う	S	態度	悪	-	[S3]	-	[S0]	-	-	-	-	-	-	-	不確実	連用中止形から転成した名詞あり。・疑い
	S	態度	悪	-	[S3]	-	[S1]	-	-	-	-	-	-	-	不確実	
	S	態度	悪	-	[S3]	-	[S2]	-	-	-	-	-	-	-	不確実	
起因	R	関係	悪	-	[R]	[S0]	-	-	-	[D0]	-	-	-	-	-	
	R	関係	悪	-	[R]	[S0]	-	-	-	[D1]	-	-	-	-	-	
	R	関係	悪	-	[R]	[S1]	-	-	-	[D0]	-	-	-	-	-	
	R	関係	悪	-	[R]	[S1]	-	-	-	[D1]	-	-	-	-	-	
誘発	R	関係	悪	-	[R]	[D0]	[S0]	-	-	-	-	-	-	-	-	ある一つの作用・動きが他の(よくない)作用・動きを誘い起こすこと。
	R	関係	悪	-	[R]	[D0]	[S1]	-	-	-	-	-	-	-	-	
	R	関係	悪	-	[R]	[D1]	[S0]	-	-	-	-	-	-	-	-	
	R	関係	悪	-	[R]	[D1]	[S1]	-	-	-	-	-	-	-	-	

※必須格に色づけ。実際の表現が係助詞・無格・連体修飾の場合は、必須格とみなしてよい。

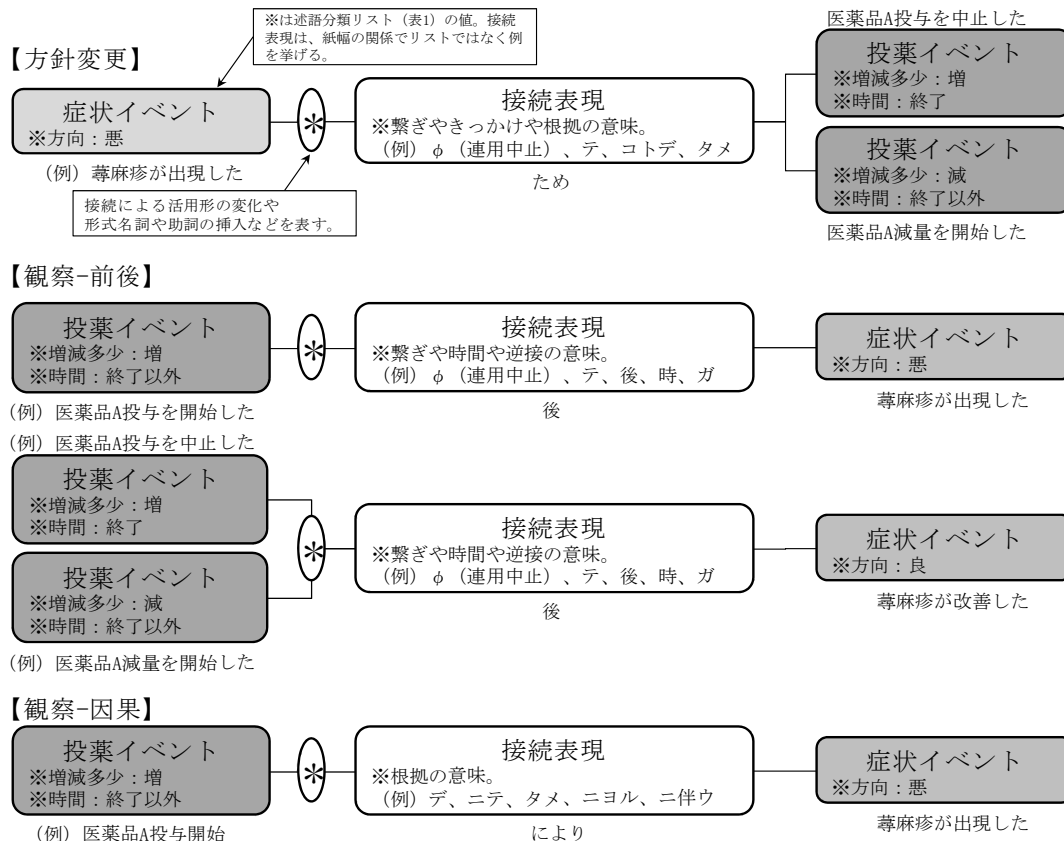


図2 副作用パターン

3. おわりに

本論文では、カルテに出現する副作用表現の分析を行った。分析の結果得られた構文パターンは、分析対象データの約9割に適応できた。

今後はカルテ全体に分析対象を広げ、述語の分類リストや構文パターンを拡張したい。文という単位を超えてカルテ内のイベントの連鎖が抽出できるようになると、より多くの副作用表現が発見できるだけでなく、例えばヒューマンエラーや適切な治療方法の確認につながるなど、安全な医療業務の遂行に貢献できると考える。

また、文の階層構造を考えるうえで、助動詞や補助動詞などの機能語の分析は欠かせない。今後は併せて検討したい。

【参考文献】

[1] 野田尚史(1989)「文構成」宮地裕編『講座日本語と日本語教育 1 日本語学要説』明治書院

[2] 益岡隆志(1989)「モダリティの構造と疑問・否定のスコープ」仁田義雄、益岡隆志編『日本語のモダリティ』くろしお出版

[3] 南不二男(1993)『現代日本語文法の輪郭』大修館書店

[4] 大熊智子、外池昌嗣、三浦康秀、荒牧英治、増市博、篠原(山田)恵美子、大江和彦(2012)「医薬品の副作用調査を目的とした統合的言語処理システム」言語処理学会第17回年次大会

[5] Yasuhide Miura, Eiji Aramaki, Tomoko Ohkuma, Masatsugu Tonoike, Hiroshi Masuichi and Kazuhiko Ohe(2010) “Adverse-effect relations extraction from massive clinical records” In COLING2010 Workshop (In cooperation with Infoplosion) The Second International Workshop on NLP Challenges in the Information Explosion Era (NLPPIX 2010)