

**NTCIR-13**  
**MedWeb** コーパス・ガイドライン  
Medical Natural Language Processing for **Web** Document



Version 1.0 (2017/04/03)

**Organizers**

Eiji Aramaki (Nara Institute of Science and Technology)  
Shoko Wakamiya (Nara Institute of Science and Technology)  
Mizuki Morita (The University of Okayama)  
Yoshinobu Kano (Shizuoka University)  
Tomoko Ohkuma (Fuji Xerox)

## 1 分類の手順

対象となる症状は以下の通りである。

インフルエンザ	風邪	花粉症	下痢
頭痛	咳・たん	熱	鼻水・鼻づまり

<sup>1</sup>該当症状が、<sup>2</sup>今現在、<sup>3</sup>本人とその周囲にあるかどうかを判別し分類した。全症状に共通の<sup>1</sup>症状に関する基準、<sup>2</sup>時制に関する基準、<sup>3</sup>空間に関する基準を設けた。また、症状ごとに、罹患性や原因性を加味しながら 症状別の判断基準 を設けた。それら判断基準から症状ありと判断されたものを「p (positive)」，それ以外を「n (negative)」とした。なお、判定不明なものには「?」を付与した。

なお、本ガイドラインでは、分かりやすさのために、ある症状の判定結果が p (positive) の場合には (症状+)， n (negative) の場合には (症状-) の形式のラベルとともに発言例を示す。また、「インフルエンザ」「咳・痰」「鼻水・鼻づまり」に関しては、スペースの都合上、それぞれ略した症状名を用いている（「インフル」「咳」「鼻水」）。

## 2 大部分の症状に共通の基準と症状別の例外

### 2.1 大部分の症状に共通の基準

ラベル付けでは、実用性を重視し、古すぎる情報（数週間前の情報など）や重篤でない症状（飲み過ぎによる頭痛など）はネガティブとするなど、医学的重要性を考慮する必要がある。このため、以下の3つの基準を原則とする。

**【症状（の事実性）の適合】** まず、疾患に罹患したという事実が確認される必要がある。単に、疾患名を含んでいるだけの話題としての発言は除く（ネガティブとみなす）。詳細は2.3節を参照いただきたい。

**【時制の適合】** 次に、古すぎる情報は、サーベイランスという観点からは意味がないと考える。どれくらいの時間を古いとみなすかであるが、24時間を目安として、本日、または、前日の情報であれば、これをポジティブとみなす。詳細は2.4節を参照いただきたい。

**【空間の適合】** 最後に、罹患者の位置情報が特定される必要がある。発言者が罹患者本人である場合は、発言者の位置情報が発症場所になる。伝聞などの場合は、発言者の近傍にいると推測される場合はポジティブとし、そうでない場合はネガティブとした。近傍判定は同一県内かどうかを基準とした。詳細は2.5節を参照いただきたい。

以下の表に、これらをまとめたものを記す。

症状	<ul style="list-style-type: none"><li>• 症状を直接的に表現している場合</li><li>• 症状を間接的に表現している場合</li><li>• 否定されているが症状の事実が認められる場合</li><li>• 回復過程の場合</li><li>• 原因不明の場合</li><li>• 症状の疑いのみが認められる場合</li><li>• 症状が完全になくなった場合</li><li>• 予想や過程のモダリティ表現の場合</li><li>• 罹患とは関係ない話題の場合</li><li>• 比喩表現・ジョークの場合</li></ul>
時制	<ul style="list-style-type: none"><li>• 今日を含む24時間以内の場合</li><li>• 昨日の症状である場合</li><li>• ごく近い過去の症状と推測できる場合</li><li>• 継続または頻繁に起きている症状で、今現在も症状としてみられる場合</li><li>• 一昨日以前の症状である場合</li><li>• 過去の症状である場合</li><li>• いつの時点の症状か不明な場合</li></ul>
空間	<ul style="list-style-type: none"><li>• 本人と同じ地域にいると分かる場合</li><li>• 主語がなくても身近にいることが推測できる場合</li><li>• 家族に関する症状の場合</li><li>• 周囲での症状だが人以外の症状の場合</li><li>• 不特定多数への呼びかけ、近傍にはいない人の症状の場合</li><li>• RT（リツイート）等のうち相手方の症状の場合</li></ul>

## 2.2 症状別の例外

上記の原則には例外が存在する。例えば、「頭痛」の症状では、疾患を原因としない（「飲み過ぎ」など）発言が多い。疾患サーベイランスとしては、これらの発言はノイズとみなされるべきである。また、ある種の疾患に関しては、24時間より前であっても、臨床的に重要視される場合がある。例えば、既存のインフルエンザのサーベイランスは一週間単位で配信されているため、これよりも細かい時間粒度であれば意味を持つ。また、疾患ごとに多様な例外がある。前述の例では「飲み過ぎ」の他に、「服薬」「刺激物」「精神的」「乗り物酔」「つわり」「生理痛」などによる症状も除かれるべきであろう。このような例外を次の表にまとめる。

	症状				時制 前日の症状を含めるか	空間 RT相手の症状を含めるか	
	症状疑い 表現を 含めるか ※ <sup>1</sup>	治癒表現を 含めるか ※ <sup>2</sup>	単語のみ 表現を 含めるか	例外			
				症状とみなす			症状とみなさない
インフル エンザ		○			○	○	
風邪	○	○	○		○		
花粉症	○	○	○				
下痢	○		○	食べ過ぎ 消化不良 飲酒 服薬 刺激物	○	○	
頭痛	○	○	○			視覚嗅覚	
咳・たん	○	○	○	飲酒 刺激物			
熱	○	○	○ 微熱 のみ	花粉症 注射の副作用			
鼻水・ 鼻づまり	○	○		花粉症		気温変化	

※<sup>1</sup>「っぽい」「気がする」「かもしれない」「可能性がある」「前兆」「しそう」「予感」等の疑い表現

※<sup>2</sup>「治った」「回復」「克服」「良くなった」「止んだ」「消えた」等の治癒表現

## 2.3 症状の事実性の基準

### ● 症状が直接的に表現されている場合

症状を直接的に表現している場合が該当する。症状の強さの有無は問わず、軽い症状も含める。

ううむ、腹の痛みと下痢はあいかわらずだ… (下痢+)

朝喉が痛くなるのは花粉症で口呼吸になってるとなることがあるようだ (花粉症+)

### ● 症状が間接的に表現されている場合

通院、投薬、他症状併発などから症状の事実が間接的に推測される場合が該当する。

風邪薬を飲んで今から寝ます。 (風邪+)

食べたくもないラーメンを胃に入れ頭痛薬を飲む (頭痛+)

### ● 否定されているが症状の事実が認められる場合

文面では症状が否定されていても、それと類する症状が伺える場合はポジティブとみなす。多くの症状について、一般の患者が十分な知識を持っておらず、見過ごしてしまう可能性を考慮した。

にしても、頭が痛いwww風邪・・・じゃないよな、あんなに食べたし (頭痛+, 風邪+)

時期的にはヒノキ?目の痒みもあるので、単純に風邪と**言い切れない**のよね (花粉症+, 風邪+)

### ● 回復過程の場合

症状から回復する過程においても、症状は弱いとはいえ罹患しているといえるため、完全に症状がなくなっていない場合は、ポジティブとみなす。ただし、症状が完全になくなっている場合はネガティブとみなす。

頭痛吐き気状態から回復したっぽい (頭痛+)

週末をまるまる熱出して寝てた。週明けの今日、鼻水が少々出るけど**ほぼ復活**。 (熱-, 鼻水+)

### ● 症状の疑い原因不明の場合

最近すぐ頭が痛くなる。なんでだ (頭痛+)

### ● 症状の疑いのみが認められる場合

患者が十分な知識を持っておらず、症状・疾患を看過してしまう可能性を考慮し、症状・疾患の疑いのみが認められる場合においてもポジティブとみなす(表現例:「っぽい」「らしい」「似た」「のような」「気がする」「かもしれない」「ぞみ」「っぽい」「可能性がある」「なのかな」「前兆」「しそう」「予感」)。

風邪引く前兆出てるナウ (風邪+)

すごい肩こり。風邪かも? (風邪+)

### ● 症状が完全になくなった場合

あ、下痢症状は治まりました。 (下痢-)

頭痛や気だるさも水飲んで一晩寝たらなおるよ! って事でなおったよ! (頭痛-)

### ● 予想や過程のモダリティ表現の場合

「つもり」「しかける」「したら」「予定」「だったら」など事実性でないモダリティの場合はネガティブとみなす。

なんかだるい・・・もしこれで咳と熱が出たら●●を疑うかも。校内で感染者が出て休校中なわけだしw (熱-)

風邪じゃありませんように \*mb\* (風邪-)

### ● 罹患とは関係ない話題の場合

罹患とは関係なく、単なる話題として、症状・疾患表現が出現した場合は、ネガティブとみなす。ニュース、一般論、広告などが該当する。

鼻水が固体化し始めたら風邪寛解の合図 (鼻水-, 風邪-)

News: 【新型インフル】大阪の感染者83%が38度以上の高熱 - MSN□□ニュース <http://□□> (インフル-)

### ● 比喩表現・ジョークの場合

誤表記を含む。

絶賛発熱中 (熱-)

台風並みに風邪が強いな (風邪-)

## 2.4 時制の適合

### ● 今日を含む 24 時間以内の場合

朝微熱→研究室ミーティング欠席→熱さがる→受け取れないはずの宅急便受け取り	(熱+)
高熱が出てこれはもうダメだと思ったら4時間後に平熱に戻ったので仕事しに来たでござるの巻	(熱+)

### ● 昨日の場合

症状別に判断が異なるため、症状別に判断する。

昨日からの風邪が治らない。	(風邪+)
食欲がない以外は昨日熱があったのが不思議なくらい元気なだけだな。平熱に戻ったし。	(熱-)

### ● ごく近い過去の症状と推測できる場合

最近すぐ頭が痛くなる。なんでだ	(頭痛+)
謎の高熱で3日ほど死んでました。	(熱+)

### ● 継続または頻繁に起きている症状で、今現在も症状としてみられる場合

冬の間はずっと風邪をひいてる気がしてきた。痰が黄色い。	(風邪+, 咳+)
うう……。豚肉食ったら、なぜかいつも下痢をする・・・orz	(下痢+)

### ● 一昨日以前の症状である場合

こないだサウナに入って一切汗が出なくてそのあと少し激しいスポーツをしたらその夜かなりの高熱を出したから今日はサウナに入る前にポカリスエットをたくさん飲んだ	(熱-)
---	------

### ● 過去の症状である場合

更に発表会当日に39度近い熱が出たのも思い出した。一番目にしてもらって、速攻帰ろうと思ったんだけど、誰も順番変わってくれなかった。	(熱-)
---	------

### ● いつの時点の症状か不明な場合

状況の具体性から推測して判断する。

焼き肉の翌日は下痢	(下痢+)
今年のインフルエンザは咳系統らしいですよ！って熱が出て咳が出まくって倒れた俺が言ってみる	(インフル-, 熱-, 咳-)

## 2.5. 空間の適合

### ● 本人と同じ地域にいると分かる場合

電車の中に激しく咳している人がいる。風邪うつさないでね。	(咳+)
やばいやばいやばいよ・・・!!近くで新型インフルエンザ発症者出たらしい、ああ、恐いよ・・・	(インフル+)
でも学校閉鎖は・・・正直おいしいです^p^	

### ● 主語がなくても身近にいることが推測できる場合

風邪なのに仕事とかなんなの？移すの？	(風邪+)
咳込んでわざと風邪をうつそうと頑張るのやめてほしいわー。うつらないけど*Tw*	(咳+, 風邪+)

### ● 家族に関する症状の場合

同居別居問わず周囲の症状とする。

今、家族で風邪が流行！！親父を除いて！！うがい手洗い大事な！！	(風邪+)
娘が発熱しだした。昨日じいさんばあさんと豊川稲荷に行った時にもらったっ風邪か？	(熱+)

### ● 周囲での症状だが、人以外の症状の場合

今日は子犬さんが風邪気味なので早めに帰ってあったかいものを作ってあげたいと思います。...	(風邪-)
朝からバンが熱暴走...例の電源が高熱になって落ちるヤツ。	(熱-)

- 不特定多数への呼びかけ、周囲ではない人の症状の場合

---

風邪なみなさんは、耳鼻科へ行くといいです。全部解決します。 (風邪-)

なんか風邪ひいてる人が多い気がする[mb] (風邪-)

---

- RT（リツイート）等のうち相手方の症状の場合

RT や@□□等、特定の人との会話において、相手方の症状に対する励ましや警告が多々見受けられる。この場合の相手方を周囲の人と解釈するか否か、症状別に判断する。明らかに見知らぬ人に対しての RT はネガティブとみなす。

---

毎日のように熱がどう薬がどう体調が悪いだの mixi で書いてる子が、病院で休みを潰したくないから行かないとか書いてて、流石にちょっとツッコミを入れることにきめた (熱-)

---