

文間の限定関係認識：課題設計および分析と予備実験

大木環美[†] 村上浩司[†] 水野淳太[†] 増田祥子^{†§} 乾健太郎^{†‡} 松本裕治[†]
奈良先端科学技術大学院大学[†] 東北大学[‡] 大阪府立大学[§]
{megumi-o, kmurakami, junta-m, shouko, inui, matsu}@is.naist.jp

1 はじめに

Web 上には多種多様な文書が存在し、同じトピックに対しても複数の相対立する意見や異なる情報が存在する。複数文書の目的の一つはこうした散在する情報を集めて整理することであり、我々も言論マップ生成 [3] の研究を通して Web 上の多様な情報の俯瞰を支援する技術の開発を目指している。複数の文書から得られる意見や情報を整理する有効な方法の一つは、類似の情報と対立する情報を区別して集約することである。言語情報間同義や含意、矛盾関係を認識する技術は、近年 Pascal RTE Challenge[1] を契機に研究が進んでおり、それらの成果を言論マップのような情報の集約に利用できる環境が整いつつある。

例として、「キシリトールの虫歯予防効果」について調べたいユーザが、(1) のような文をクエリとして検索する状況を想定する。Web 上には多様な言明が存在し、その中には (2a) のようにクエリ文とほぼ同義なものや (2b) のようにクエリと矛盾するものが見つかるかもしれない。TAC の RTE[2] では、入力文対の関係を「含意」「矛盾」「不明 (unknown)」の 3 種類に分類する課題を想定しており、この技術が実現すれば、(1) と (2a)、(2b) のような関係を自動認識することができる。

- (1) キシリトールは虫歯予防に効果がある
- (2) a. キシリトールは虫歯を防ぐことができる
b. キシリトールは虫歯予防にまったく効果がない
- (3) a. キシリトールが50 %以上配合されていないと虫歯を抑制できない
b. キシリトール だけでは、十分には 虫歯を防げません

(3a) では「50 %以上配合されている」ことが (1) を成立させるための必要条件として付加されており、(3b) では「だけ」という副助詞が虫歯予防の効果をもつものの範囲を、「十分には」という程度副詞が虫歯予防の効果の程度をそれぞれ制限する情報として (1) と対応する命題に付加されている。(1) の文を「キシリトールはどんなときでも虫歯予防に効果がある」という強い意味で捉えればこれらの文も矛盾の関係となるが、我々は普通そのようには捉えない。では、(1) を「キシリトールは虫歯予防に効果がある場合もある」という弱い意味での肯定と捉えた場合、(3a) や (3b) は含意の関係となると考えられるが、そうすると、これらの文が (1) に

対応する命題が成立するための制限を含み、弱い対立の関係になっていることを捉えることができない。

このような弱い対立を示す文間の意味的關係を認識し、命題の成立を制限する着目すべき条件や程度、範囲などの表現を捉えることは、意見や情報をより正確に整理する上で重要であるが、これらの文間の関係は RTE の 3 種類では対象外となっている。そこで、我々は条件や程度、範囲などの制約情報によって弱い対立の関係になっているものを便宜的に「限定」関係と呼び、本稿では、(1) のようなユーザの知りたい命題をクエリ、Web 上でクエリを検索した結果の文書集合中の文を検索対象文とし、クエリと検索対象文間の限定関係を識別する課題について論じる。

2 関連研究

2 文間の関係を認識する RTE の研究では「含意」「矛盾」「不明」の 3 種類の関係のみが定義され、これらについては数多くの研究が行われているが、限定に相当する関係については扱われていない。

より多くの文間関係を扱う研究として Radev らの Cross-Document Structure Theory(CST)[4] が挙げられる。RTE では単一文書内の構造が対象とされてきたが、CST は談話構造を複数の異なる文書間の構造に拡張しており、最終的には 18 種類の文間の関係が定義されている。一方、日本語においては、衛藤ら [6] が CST をベースに 14 種類の関係を再定義している。しかし、いずれの関係でも限定に相当する関係は存在しない。これは、CST では文間の同一性と差異に着目しているためであると考えられる。「限定」関係は一見すると一方の文に対して差異であるが、付加された条件の実現によっては同一となる場合が多く、同一・差異の双方の関係を持つため、CST の範疇でないと考えられる。

言語学の分野では 1 文内での制約を示す言語表現に対する研究として、森ら [7] が日本語のマニュアルにおける「と」「れば」「たら」「なら」の 4 つの条件節を示す表現がどのような制約を示すかを論じている。また、安部 [9] は「だけ」による個体や事態の限定や数量詞による量の規定について研究しており、これら以外にも数多くの言語学における「限定」の表現についての研究がある。このような 1 文内での「限定」を示す表現は同義や矛盾を制約する表現を認識する上で有用であるが、言語学の分野では 2 文間における限定の関係についての研究は我々の知る限り見当たらない。

3 限定関係モデルの一般パターン

言語学における「限定」表現の整理は、文内構成素間の関係の分析に留まっただけのもの、文間の関係を整理するための良い出発点にもなると考えられる。例えば、次の文(4)を現在のクエリと仮定しよう。

(4) キシリトールは虫歯予防に効果がある
 このとき、(4)に条件表現を加えた(5)のような文が元の(4)に対して限定の関係を持つことは、条件表現の意味から明らかである。

(5) キシリトールは**1日3回毎食後食べて初めて**虫歯予防に効果がある

また、同様に、(4)と同義の文に条件表現を加えた(6)のような文も、やはり(4)に対して限定関係を持つ。

(6) キシリトールは**1日3回毎食後食べて初めて**虫歯予防を抑制できる

さらに、(4)を否定する文に条件表現を加えても、限定関係が成り立つ場合がある。

(7) キシリトールが**50%以上配合されていないと**虫歯抑制効果はない

以上より、限定関係を示す文対の一般的な構造として図1のようなパターンを考えることができる。これは、一方の文をクエリ、他方の文を検索対象文とすると、

- クエリと同義、含意、あるいは矛盾の関係を持つ言明(以下、クエリ対応言明)が検索対象文に含まれており、さらに、
- クエリ対応言明の成立を制限する条件、程度、範囲等の付加情報(以下、限定表現)が検索対象文に含まれている

という2点から成り立っているパターンである。

図1では「キシリトールは虫歯予防に効果がある」と「キシリトールは虫歯を防げる」が同義の関係であることが1のパターンを満たし、「毎食後にとれば」や「そこそこ」といった条件や程度を示す表現が付加されていることが2のパターンを満たすことで、文対が限定の関係となることを示している。

もし限定関係を示す文対の多くに図1のパターンを適用できるとすれば、限定関係認識の問題を次の2つの部分問題に帰着させることができる。

- 言明間の同義や矛盾の関係を判定する RTE の問題
- 条件や範囲や程度の制限などの限定表現を識別する問題

RTE については各所で活発な研究が進んでおり、限定表現についても2節で述べたように言語学的分析の成果を自然に利用することができる。そこでまず、我々が冒頭の複数文書要約の目的で認識したい「弱い対立」の文対がどの程度このパターンで網羅できるのかを調査した。

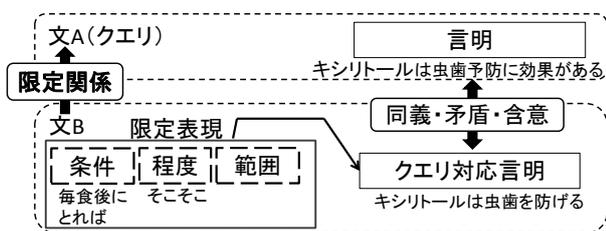


図1: 限定関係認識の一般パターン

4 限定関係一般パターンに基づく事例分析

村上らの言明間意味の関係コーパス [5] などから上述の意味で弱い対立関係にあると認められた99対を集め、図1に示した限定関係認識の一般パターンに当てはまるかどうかを調べた。以下に適用事例を示し、表1にクエリ対応言明を特に制限している限定表現が99対中でのどのような内訳になっているのかを示す。下線部はクエリ対応言明を、太字は限定表現箇所をそれぞれ示している。

表1: 限定関係を示す言語表現の分類

限定関係表現	特徴語彙	文数	
条件	順接条件	+ 理由・原因節	1
		+ とりたて助詞	1
		+ 程度表現	1
	逆接表現	+ 程度表現	1
		理由・原因節	3
		目的節	2
		その他条件	5
範囲	とりたて助詞	+ 程度表現	2
		+ 理由・原因節	4
	連体修飾	2	
	その他	1	
	程度表現		3
部分否定	程度表現 + 理由・原因節	2	
部分否定	わけではない	4	

条件+矛盾

- (8) a. イソフラボン**は**健康に効果がある
 b. イソフラボン**も、過剰摂取すると、健康障害を引き起こす**

(8b)では、(8a)と(8b)の下線部がそれぞれ矛盾の関係にあり、太字箇所の条件表現が下線部のクエリ対応言明に付加され、言明を制限していることから、図1のパターンと一致する。

範囲+程度+矛盾

- (9) a. ステロイド剤は症状の改善効果が高い
 b. **吸入**ステロイド剤の症状の改善効果は**あまり望めない**

また、(9b)でも「吸入」が症状の改善効果が望めないステロイドの種類を、「あまり」が改善の望めない程度を制限し、それらが(9a)と矛盾の関係となるクエリ対応言明に付加されていることから、図1のパターンと一致する。この他に以下のような例が挙げられる。

条件+同義/含意

- (10) a. キシリトールは虫歯予防効果がある
 b. キシリトールに虫歯予防効果があるとした場合は、**ガムをかむ回数と量**がかなり求められる

(10b)のクエリ対応言明と同義となる箇所が(10b)では主節部分(「キシリトールに虫歯予防効果があるとした場合は」)に相当し、下線部の帰結節に(10a)の条件となっている。

また、条件には必要条件と十分条件が存在し、(1b)のように、必要条件(「50%以上配合されている」)がク

エリ対応言明に付加されている文が限定関係となるのは明らかである。一方、(4a)のように十分条件(「1日3回毎食後食べれば」)がクエリ対応言明に付加されている場合もある。このとき、(4a)は「1日3回毎食後食べないと虫歯予防効果がない」ことを暗示しており、言語表現では十分条件も程度の差はあるものの必要条件の役割を果たしていることから、限定表現となると考えられる。

範囲+同義/含意

- (11) a. イソフラボン¹は体によい
b. イソフラボンの効果には 個人差 がある

(11a)の下線部のクエリ対応言明が成立するには「個人差」があり、どんなときにも成立するわけではないため、事象の成立範囲を制限している文となっている。

程度+矛盾

- (12) a. 抗がん剤は副作用がある
b. 抗がん剤には必ずしも強い副作用はない

(12b)ではクエリ対応言明が(12a)と矛盾の関係にあり、下線部の程度表現が付加されることで「副作用がある」頻度を制限する限定関係の文となっている。

このように、図1のパターンは99文対のうち少なくとも80文対はカバーできたが、このパターンでは説明できない事例も存在した。その事例を以下に示す。

- (13) a. ステロイド剤は副作用が問題とされている
b. 鼻に直接入れるステロイド剤は、副作用も 少ない

- (14) a. 毎食後にキシリトールをとれば虫歯予防効果がある
b. 毎食後にキシリトールを食べても、虫歯予防に絶大な効果があるわけではない

(13b)ではクエリ対応言明内に「少ない」という程度の表現が含まれており、3節の一般パターンに当てはまらない。これは同義や矛盾を制約する情報が、付加されるだけでなく言明内に存在する場合もあることを示している。(14b)では「わけではない」という部分否定によって、弱い対立を示す文となっており、この例から条件、範囲、程度以外の限定関係を示す表現として、部分否定の表現が存在することがわかった。

上記のような例外のパターンも存在したが、多くの場合3節で定めたパターンの限定関係表現に部分否定を加えたもので限定の関係を把握することが可能であることがわかった。この一般パターンにより、限定関係が言明間の同義や矛盾、含意の関係を求める問題と言明に付加されている語彙が限定表現であるかを求める問題に帰着できることとなる。

5 予備実験:限定関係認識例と今後の課題

3節で定めた一般パターンと4節で分析した限定表現で限定関係を認識できるかの実験を行った。本実験では、まだ調査段階のためルールベースでの認識実験を行い、事例からの手統計的学習による最適化は今後の課題とする。使用したデータは3トピックのクエリ対実文で全701文対あり、以下の手順で処理を行う。

1. 図1の言明間の同義や矛盾、含意の同定にあたる処理を行う。本研究では後藤らの文内構造アライメント手法[8]を用いて、言明クエリと実文中の文節ごとのアライメントを行い、

2. 3種類の関係にある言明間とみなされた文節とその周辺の文節に、図1での条件、範囲、程度に相当する表現が存在するかを判定し、
 3. 存在すれば限定関係とみなす
- 限定関係を示す表現は以下の言語表現と分析結果によるいくつかのパターンを用いた。

- 順接条件節
限り、場合、たら、なら、(と)、ては、では、のでは、とすると、すれば、したら、なると、なれば
- 程度副詞
あまり、さほど、大して、大抵、全く、全然、さっぱり、ちっとも、少しも、割合、甚だ、割と、多少、いくらか
- とりたて助詞
しか、だけ、ばかり、くらい、ほど、でも、まで、こそ

限定関係認識の成功例と誤認識例を表2に示す。Aでは限定であるものが正しく認識できた事例、Bでは限定であるのにそれを認識できなかった事例、Cでは限定でないのに限定と認識してしまった事例をそれぞれ挙げている。太字部分が限定表現の特徴語彙と認識した個所であり、下線部が認識した限定箇所全体である。

aからdではクエリの成立を制限する特に注目すべき条件などが限定部分として表示できており、これまでの文間関係認識分野では取り扱われていなかった事例を得ることができた。誤認識の例からは以下のような考察が得られた。

- 限定表現の誤認識によって失敗する場合がある(h)。特に連体修飾の表現はその意味を考慮する必要があることから、パターン化することは困難であることため、学習などの手法を適用する必要があると考えられる。
- モデルの構造で認識した限定表現が必ずしも限定関係にはならない(c,e)。このような例外は4節では分析しきれなかった事例のため、今後モデルに反映させ対応していく。
- 限定表現が言明に係っているのかを判別することが重要となってくるが、現在では「高い」や「軽減する」と「ある」が同じ意味を持つことを判定できていない(f,i)。この問題についてはアライメントの精度向上により解決できると考えられる。
- 筆者が述べたいことが帰結部分の「積極的に摂るべきだ」の個所にあたり、このような筆者の意図を汲み取る必要がある(l)。

この他の課題として、とりたて助詞の「と」や「は」には引用など様々な用法が存在し、それらが混在して使用されているため、限定表現で用いられる助詞のみを判定する必要のあることが挙げられる。この問題については、統計的に用法の事例を解析し、パターンに準拠するような判定方法を考案することで対処したいと考えている。

また(13b)が限定の関係になるのに対し、(15)は同義の関係となることから「多い」という存在の範囲を拡大する単語には限定表現となる要素が少なく、反対に「少ない」という存在の範囲を狭めている単語は限定表現となる場合があることがわかる。

- (15) 鼻に直接入れるステロイド剤は、副作用も 多い
このことから2文間の言明間には存在の範囲の違いを示す関係もあり、今後はこの関係も把握していく必要があると考えられる。

表 2: 限定認識の成功例と失敗例

分類	番号	クエリ	検索対象文
true pos.	a	ステロイドは副作用がある	ステロイド剤という副作用を心配する人が多いのですが点鼻療法ではごく少量を局所的に用いるだけなので副作用の心配は殆どありません
	b	キシリトールは虫歯予防に効果がある	キシリトール使用は100%キシリトールのガムであれば1日3回毎食後1枚かむだけで効果があるとされています
	c	キシリトールは虫歯予防に効果がある	キシリトールの虫歯予防効果を発揮するには1日3回3ヶ月以上ガムやタブレットを食べ続ける必要があり虫歯になりやすい人には特に効果的と考えられている
	d	イソフラボン健康に効果がある	大豆イソフラボンをサプリメントとしてこれ以上摂取する場合健康維持には負の影響を与える結果となります
false pos.	e	ステロイドは副作用がある	ステロイド剤と漢方薬を併用することで副作用を軽減し、ある程度症状が治まった時点でステロイド剤を止め漢方薬を前面に出した治療法に切り替えていく方法を取ることでステロイド剤の欠点を補うことが可能です
	f	キシリトールは虫歯予防に効果がある	虫歯予防に最適な甘味料のキシリトールですが、使い方を間違えると全く効果が期待できないものになってしまうのです
	g	キシリトールは虫歯予防に効果がある	歯磨き剤には、フッ素が入っていますので、歯に良い効果を期待出来ますが、甘味料は本来極わずかししか入っていませんのでキシリトールの効果に過大な期待は禁物です
	h	イソフラボン健康に効果がある	しかし近年大豆イソフラボンを原料とした健康食品なども多く出回っている反面でエストロゲン様の働きをもつイソフラボンの過剰摂取はかえって身体に良くないと指摘する声も出ています
false neg.	i	ステロイドは副作用がある	副腎皮質ホルモン剤ステロイド剤は痛みと炎症を抑える効果がありますが副作用も強いので痛みが非常に強い場合にだけ使われます
	j	キシリトールは虫歯予防に効果がある	100%キシリトールガム352の場合通常1日3回歯磨き前に1粒ずつ噛むだけで十分効果があります
	k	キシリトールは虫歯予防に効果がある	キシリトールの含まれている量が多いほどむし歯予防の効果は高いようです
	l	イソフラボン健康に効果がある	簡単にまとめるとサプリとしてイソフラボンを摂りすぎると特に妊婦さんの場合あまりよくない場合があるが普通に大豆製品として食べたり飲んだりする分にはまったく問題なくむしろ大豆の健康成分が有効なので積極的に摂るべきだということになります

6 おわりに

Web上の意見や情報には RTE の 3 種類の関係では認識できない弱い対立を示す文が数多く存在し、これらの文には命題が成立するための条件などの着目すべき情報が含まれている。我々は、これを「限定関係」という新たな関係として位置づけ、クエリと検索対象文間の限定関係認識の課題について議論した。本稿では、限定関係の一般パターンに基づいた分析から予備実験を行い、限定関係の認識例を示した。予備実験では命題となるクエリが成立するための条件などを認識することができたが、同時に連体修飾の認識やクエリと検索対象文の新たな関係の識別などの課題も浮き彫りになったため、今後は限定関係の認識方法について機械学習などによる統計的な情報の利用やパターンの増加を行い、限定関係認識を定量的に評価していく。

謝辞 本研究は、(独)情報通信研究機構の委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」の一環として実施した。

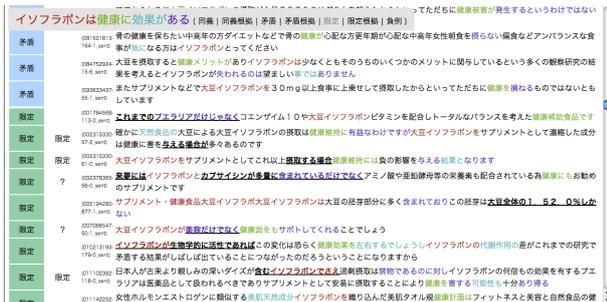


図 2: システムの出力例

参考文献

- [1] Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proc. of the PASCAL Challenges Workshop on Recognising Textual Entailment*, 2005.
- [2] Danilo Giampiccolo, Hoa Trang Dang, Bernardo Magnini, Ido Dagan, Elena Cabrio, and Bill Dolan.
- [3] Koji Murakami, Eric Nichols, Suguru Matsuyoshi, Asuka Sumida, Shouko Masuda, Kentaro Inui, and Yuji Matsumoto. Statement map: Assisting information credibility analysis by visualizing arguments. In *Proc. of the 3rd ACM Workshop on Information Credibility on the Web (WICOW 2009)*, pp. 43–50, 2009.
- [4] Dragomir R. Radev. Common theory of information fusion from multiple text sources step one: Cross-document structure. In *Proc. of the 1st SIGdial workshop on Discourse and dialogue*, pp. 74–83, 2000.
- [5] 村上浩司, 増田祥子, 松吉俊, Eric Nichols, 乾健太郎, 松本裕治. 言明間の意味的關係の体系化とコーパス構築. 言語処理学会 第 15 回年次大会, 2009.
- [6] 衛藤純司, 奥村学. 文書横断文間関係タグ付コーパスの構築. 言語処理学会第 14 回年次大会, 2005.
- [7] 森辰則, 龍野弘幸, 中川裕志. 日本語マニュアル文における条件表現「と」「れば」「たら」「なら」から導かれる制約. *自然言語処理*, Vol. 2, No. 4, pp. 3–18, 1995.
- [8] 後藤隼人, 水野淳太, 村上浩司, 乾健太郎, 松本裕治. 文間関係認識のための構造的アライメント. 言語処理学会第 16 回年次大会発表論文集 E3-7, 2010.
- [9] 安部朋世. ダケによる〈限定〉と数量詞による〈修飾〉: その共通点と相違点. *筑波日本語研究*, Vol. 1, pp. 4–20, 1996.