

文間の弱い対立関係の認識

大木環美^{†1} 村上浩司^{†2} 松吉俊^{†1}
水野淳太^{†1,†3} 乾健太郎^{†3,†1} 松本裕治^{†1}

意見マイニング等の意見を集約するようなタスクでは、賛成意見や反対意見を分類することが行われている。実際には、これらの意見の中には部分的に賛成意見のものや部分的に反対意見のものが存在しており、完全な賛成意見や完全な反対意見とは異なる観点からの有用な情報を述べている。これらを認識するために有用な含意関係認識 (RTE) のタスクでは、文対は含意・矛盾・不明の3種類の関係に分類される。しかし、現状ではまったくの賛成意見や反対意見と区別されず同等として扱われてしまうため、ユーザが知見としてそれらの情報を得ることが難しい。本稿では、このような有用な情報を認識するための意味的關係を**弱対立**の關係と呼ぶ。この弱対立關係を新たに定義し、与えられた2文間の關係が弱対立關係であるかを認識するための弱対立認識の手がかりを作成する。さらにその手がかりを用いたルールベースのシステムを実装し、弱対立認識の評価と現状の問題点を示す。

Defining and Recognizing Confinement in Texts

MEGUMI OHKI,^{†1} KOJI MURAKAMI,^{†2}
SUGURU MATSUYOSHI,^{†1} JUNTA MIZUNO,^{†1,†3}
KENTARO INUI^{†3,†1} and YUJI MATSUMOTO^{†1}

The task of researches for opinion mining and other such tasks is classifying agreement opinions and contradiction opinions. For instance, there are partial agreement or contradiction opinions except for full agreement or contradiction opinions. These partial contradiction sentence pairs contain useful information that it is different fact of full agreement or contradiction opinions.

However it is difficult for Internet users to access this knowledge because current frameworks do not discriminate between full contradictions and partial contradictions. In the Recognizing Textual Entailment (RTE) tasks, sentence pairs are classified into one of three semantic relations: [ENTAILMENT], [CONTRADICTION], [UNKNOWN]. We find some sentence pairs hold full contradictions or entailments, but there are a number of pairs that partially contradict or entail one another depending on a certain situation. In this paper, we define a

new semantic relation as [CONFINEMENT] in order to recognize useful information. This relation is recognized as either [CONTRADICTION] or [ENTAILMENT] under current approaches to semantic relation recognition. We develop *semantic templates* to recognize relations in Internet texts, and then implement a system for recognizing [CONFINEMENT] between sentence pairs, present empirical evaluation, and conduct error analysis.

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。複数文書要約の目的の一つはこうした広範囲に散在する情報を集めて整理することである。Web 上で特定の事項を検索した際に、我々が知りたいことはその検索事項に対する人々の考えや意見である。例えば、ユーザが文 (1) をクエリとして検索した場合、(2a) のような賛成意見や (2b) のような対立意見を見つけることができるかもしれない。

- (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。

二文間の類義關係や対立關係を認識する技術は、自然言語処理、特に、含意認識 (Recognizing Textual Entailment, RTE) の分野において盛んに研究されている^{1),3),8),9)}。RTE の枠組みでは、システムは、与えられた文章 (Text) と仮説 (Hypothesis) に対して、含意・矛盾・不明のいずれかの關係名を出力することを求められる。

Web 上には上記の完全な賛成意見や完全な対立意見のみが存在するのではなく、例えば、(3) に示すように、クエリに対する部分的な賛成や部分的な対立を述べるような意見も存在する。

- (3) a. キシリトールの虫歯予防効果は、**悪くはない程度です**。
b. キシリトールは **1日3回毎食後摂取しないと**虫歯を抑制できない。

(3a) は、(1) を認めてはいるが、その効果の程度が「悪くない程度」であると述べている。(3b) は、(1) と部分的に対立しており、「1日3回毎食後」の摂取が必要であると述べている。実際、観

^{†1} 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

^{†2} 楽天技術研究所
Rakuten Institute of Technology

^{†3} 東北大学
Tohoku University

察してみると、Web 上には (2) のような完全な賛成・対立意見は少なく、(3) のような部分的な賛成・対立意見が多数存在する。(3) において、太字で示した箇所はユーザにクエリに対する新たな観点(クエリの命題の成立には 条件が必要である, 成立程度や範囲について制限がある, など)を提供してくれるため、完全な賛成・対立意見に比べ、部分的な賛成・対立意見はユーザにとって有益であると言える。

部分的な賛成・対立意見は、新たな観点を教えてくれる有益な情報にも関わらず、これまでこのような意見を積極的に抽出しようとした研究はほとんどない。我々は、昨年より、このような部分的な賛成・対立意見を抽出するための研究に取り組んでおり¹⁰⁾、いくつかのトピックを含むクエリに対する部分的な賛成・対立意見を収集し、調査を行うためのコーパスを作成した¹¹⁾。

本稿では、ある文と、その文と部分的に賛成・対立するような文の間を、**弱対立**と呼ぶ。本研究の目的は、二文間の弱対立関係を自動的に認識するシステムを実現することである。以下、本稿は次のように構成される。まず、2 節において、関連研究について紹介する。次に、3 節において、論理学の記法を用いて二文間の弱対立を定義する。続く 4 節において、弱対立の文対を認識するための「文内の手がかりの組」を導入し、小規模なコーパスを用いてその妥当性について検証する。5 節において、自動的に手がかりの組を検出することにより、与えられた 2 文が弱対立の関係にあるかどうかを判定するシステムについて述べ、その評価実験結果について報告する。最後に、6 節において結論を述べる。

2 関連研究

本研究の弱対立と、RTE の含意・矛盾・不明の間を、図 1 に示す。ここで、「含意」は、RTE の枠組みに則り、Text が Hypothesis を含意する関係を表し、「逆向き含意」^{*1}は、この逆の関係、すなわち、Hypothesis(情報検索ではクエリに相当)が Text(情報検索では検索された 1 文)を含意する関係を表す。Text と Hypothesis が含意かつ逆向き含意の関係にある場合が、「同値」である。弱対立の論理的定義は 3 節で行うが、我々の弱対立は、主に、RTE の矛盾(論理的に両立しない関係)の一部である^{*2}。前節で述べたように、ある文と弱対立する文は、新たな観点を提供する有益な情報を持つが、RTE の枠組みにおいては、この弱対立を対象としていない。

RTE を拡張した関係を定義する村上らの研究¹¹⁾において、弱対立は「限定」という関係名

*1 一般に、RTE の枠組みでは、この関係は不明に含まれる。

*2 調査の結果、逆向き含意の一部であるような弱対立も存在することが分かっている。

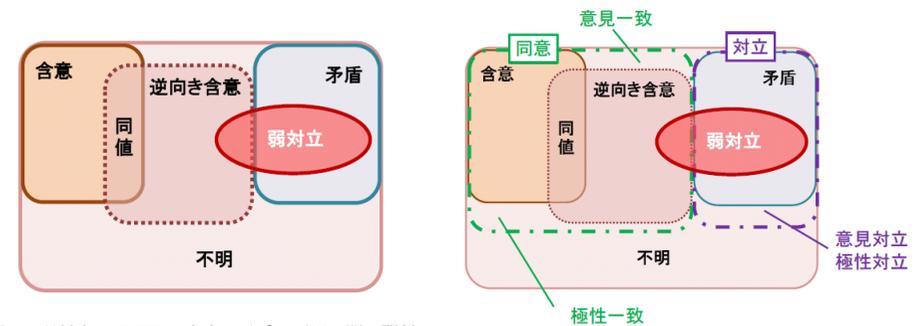


図 1: 弱対立と RTE の含意・矛盾・不明の間関係

図 2: 弱対立と村上ら¹¹⁾の同意・対立の間関係

で扱われている。弱対立(=限定)と、彼らの同意・対立との間を、図 2 に示す。弱対立は、彼らの対立と、一部、同意に含まれる。

さらに多くの文間関係を扱う研究として、Radev らの Cross-Document Structure Theory(CST)⁷⁾が挙げられる。CST では、単文書内の談話構造を複数文書間の構造に拡張しており、彼らは英語において 18 種類の文間関係を定義している。日本語においては、衛藤ら¹³⁾が、Radev らの 18 の関係を 14 種類の関係に再定義している。我々の弱対立は、主に、彼らの対照(Contradiction, Contrast)と対応するが、この対照においては、弱対立と RTE における矛盾が区別されない。

弱対立を認識するための主な手がかりは、文の帰結成立に関する前提条件を捉えるための条件節に関する知識と、文の帰結部分に付加された、程度や範囲の限定を表す副詞句に関する知識である。これらについては言語学において数多く研究されており^{12),17)-20)}、例えば、井上²¹⁾は、条件節とモダリティの関係を明らかにするとともに、条件節を仮定条件と実現前提条件とに分類することで、条件節の主文への制約を一般化している。森ら¹⁵⁾は、日本語のマニュアルにおける 4 つの条件表現「と」、「れば」、「たら」、「なら」がどのような制約を示すのかを論じている。一方、安部¹⁶⁾は、「だけ」という助詞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報告している。

3 弱対立の定義

この節では、論理学の記法を用いて二文間の弱対立を定義する。

ある文対 (S_0 と S_1) が弱対立の関係にあるのは、主に、ある文 S_1 がもう一つの文 S_0 に対して次のような差異を持っている場合である。

- 付加的な前提条件
- 帰結に対する成立範囲や程度の制限

前者の文対の例は、(1)と(3b)であり、後者の文対の例は、(1)と(3a)である。

本稿では、二文間の弱対立を扱う最初のステップとして、条件節を持つ単文程度の文を対象とする^{*1}。実際、情報検索においてクエリとして現れる表現はこのような単文であり、応用として情報検索や情報抽出を想定する場合には、この対象範囲で十分であると思われる。以下、本稿では、便宜上、文対 (S_0, S_1) において、文 S_0 に対して文 S_1 がある種の差異を持っていると仮定する。

我々は、 S_0 を表す論理式として $\forall x(P(x) \rightarrow Q(x))$ を仮定する。 S_0 が条件節を持たない場合、主文の主語が全称量化されていると考える。これにより、例えば、「キシリトールは虫歯予防に効果がある」という文は、次のような論理式で表すこととする。

$$\forall x(\text{キシリトールである}(x) \rightarrow \text{虫歯予防に効果がある}(x))$$

S_0 が条件節を持たず、かつ、主文の主語が全称量化されていると考えにくい、という場合は、前提条件は常に成立する恒真命題であると考え、 $P(x) \equiv T$ (恒真) として形式的に上記の論理式で表すこととする。

弱対立の文対を観察すると、弱対立には明示的なものと暗示的なものが存在することが分かった。明示的な弱対立では、 S_0 の帰結が成立する場合と不成立となる場合を S_1 が明示的に示す。その一方で、暗示的な弱対立では、 S_1 は、 S_0 の帰結が成立する場合か、帰結が不成立となる場合のいずれかしか示さない。例えば、(1)と(4a)は明示的な弱対立の関係にあり、(4a)は、「毎日摂る」場合に(1)の帰結「キシリトールが虫歯予防に効果があること」が成立し、そうでない場合にその帰結が成立しないことを述べている。

- (4) a. キシリトールは毎日摂って初めて虫歯予防に効果がある。
b. キシリトールは食後に摂ると虫歯予防に効果がある。

(1)と(4b)は暗示的な弱対立の関係にあり、(4b)は、「食後に摂らない」場合には(1)の帰結に関して何も述べていない。双方の弱対立の細分類を区別することで、より確実な弱対立となるもののみをユーザに示すといったユーザの意図にあった有効な情報の提示が可能となると考えている。

我々は、これらの明示的・暗示的な弱対立の関係の文対には(i)文の成立に必要とされる前提条件に付加された条件、(ii)帰結の成立する程度や範囲の制限が存在していると考えており、

*1 実際には、このような単文が等位接続詞により並列している文も本論文で扱う。

本稿では、前提条件に対する(i)と帰結に対する(ii)の2種類の弱対立の関係に細分類して定義する。

3.1 前提条件に対する弱対立

前提条件に存在する文の成立に必要な条件を論理式で表し、明示的・暗示的双方の弱対立についての定義を示す。

前提条件に対する明示的な弱対立

この弱対立の関係にある文対では、次の式のように S_0 に対する成立・不成立の状況が明示的に示されている。このタイプの文では論理的に唯一の条件を示す「ONLY IF」に相当する表現がしばしば用いられる。

$$P(x) := P_1(x) \vee P_2(x) \quad (1)$$

$$S_0 : \forall x(P(x) \rightarrow Q(x)) = \forall x(P_1(x) \vee P_2(x) \rightarrow Q(x)) \quad (2)$$

$$S_1 : \forall x(P_1(x) \rightarrow Q(x)) \wedge (P_2(x) \rightarrow \neg Q(x)) \quad (3)$$

P_1 を $P(x) \wedge A_P$ 、 P_2 を $P(x) \wedge \neg A_P$ とする。 A_P は前提条件に対する付加情報を表している。例えば、表1の S_1 の例では、キシリトールを毎日摂取した時には虫歯予防の効果があり、そうでない時には効果がないことを明示的に示し、 $P(x) \cdot A_P$ は表1に示す弱対立認識のための手がかりとなる。

前提条件に対する暗示的な弱対立

この弱対立にある文対は帰結の成立を満足させるために必要な条件が前提条件に加えられ、不成立になる場合については明示的に示されていない。このような文対は以下の論理式で表すことができる。

$$P(x) := P_1(x) \vee P_2(x) \quad (4)$$

$$S_0 : \forall x(P(x) \rightarrow Q(x)) = \forall x(P_1(x) \vee P_2(x) \rightarrow Q(x)) \quad (5)$$

$$S_1 : \forall x(P_1(x) \rightarrow \neg Q(x)) \quad (6)$$

表1の S_3 の文では「食後にキシリトールを食べなければ虫歯予防に効果がない」ということを暗黙的に意味すると考えており、論理式上では明示的な弱対立の定義上での論理式の前半部分しか示されないが、後半部分を暗黙的に示唆しているものと捉えられることから暗示的な弱対立としている。

3.1.1 帰結に対する弱対立

帰結の成立の程度や範囲を制限する帰結における付加情報を論理式で表し、明示的・暗示的双方の弱対立の関係についての定義を示す。

帰結に対する明示的な弱対立

このタイプの関係は帰結が成立する部分を特定し、帰結が成立・不成立となる範囲や程度をそれぞれ明示的に示す。

$$Q(x) := Q_1(x) \vee Q_2(x) \quad (7)$$

$$S_0 : \forall x(P(x) \rightarrow Q(x)) = \forall x(P(x) \rightarrow Q_1(x) \vee Q_2(x)) \quad (8)$$

$$S_1 : \forall x(P(x) \rightarrow Q_1(x)) \wedge (P(x) \rightarrow \neg Q_2(x)) \quad (9)$$

Q_1 を $Q(x) \wedge A_Q$, Q_2 を $Q(x) \wedge \neg A_Q$ とする。 A_Q は帰結に対する付加情報を表している。表1の例 S_2 では、「キシリトールは子どものみに効果がある」というキシリトールの虫歯予防に対しての限界についての価値ある情報を提供している。

帰結に対する暗示的な弱対立

このタイプは帰結が成立するための範囲や程度が単に制限されている。

$$Q(x) := Q_1(x) \vee Q_2(x) \quad (10)$$

$$S_0 : \forall x(P(x) \rightarrow Q(x)) = \forall x(P(x) \rightarrow Q_1(x) \vee Q_2(x)) \quad (11)$$

$$S_1 : \forall x(P(x) \rightarrow Q_1(x)) \quad (12)$$

例えば、表1の S_4 は重傷の虫歯に対してキシリトールの予防効果がないことについて言及しているが、他の虫歯に対する効果には言及していない。そのため、 $Q_1(x)$ は重傷の虫歯の範囲のみを示す論理式として定義される。

4 弱対立認識の手がかり

前節では、我々の考える弱対立についての定義を行った。

この節では、定義した弱対立の関係の細分類を認識するために関数を作成し、作成した弱対立認識の手がかりの妥当性を評価する予備調査について述べる。

4.1 弱対立認識の手がかりの作成

この弱対立認識の手がかりは、弱対立認識の手がかりの有無を入力とし、文対の関係を返り値とする。実際の文には一文として成立する複数の文が組み合わさっている場合があるため、一文として内容の成立する文内の箇所を言明と呼び、原則として弱対立認識の手がかりは言明ごとの関係を返すものとする。4節では、弱対立認識の手がかりが必要かつ十分であることを文対の小さなセットを用いて人手で確認し、5節では、文対の表現から手がかりの有無を自動的に判定して入力として与え、文対を弱対立と弱対立以外の関係に分類する実験を行う。

3節の定義から我々は文対が前提条件と帰結から構成されると想定しており、弱対立関係に

ある文対は少なくとも付加条件を含んでいるか帰結に対する程度や範囲の制限を含んでいるものであると考えている。

これらの制限や条件となる表現は様々である。例えば、(3b)の内容を示すには、「キシリトールは1日3回毎食後使用して初めて虫歯を予防できます」や「キシリトールは1日1回の使用では虫歯を予防できません」といった様々な表現を用いることができる。

このような弱対立関係の文対に存在する表現について調査している文献は数多く存在するが、対象となるすべての表現を書きつくすことは不可能なため、我々は前提条件と帰結それぞれに存在する2種類の手がかり、(1) 帰結に対する種類、(2) 付加情報の有無、の計4つの素性に着目して弱対立の関係を認識することを目指す。上記で述べた変数の詳細な情報については以下に示す。

前提条件に対する手がかり

明示的な状況を示す表現の有無

この特徴値は S_0 の前提条件に対応する箇所のがを示す。「P」は S_1 の帰結が S_1 の前提条件が成立したときに成立することを示し、「ONLY-P」は S_1 の前提条件が成立したときのみ帰結が成立し、他の場合には成立しないことを示す。「 ϕ 」は帰結が成立する場合がない、もしくは不明の場合を示している。この特徴値により明示的・暗示的の判定が行われる。

付加情報

S_0 の帰結に相当する S_1 の帰結箇所が成立するための前提条件以外に帰結の成立に必要な付加情報があるかどうかを示す。「 A_P 」は前提条件に付加情報が存在することを示し、「 ϕ 」は付加情報が存在しないことを示す。この付加情報により、 S_1 の帰結が成立する状況を特定し制限されるため、 S_1 が弱対立関係になるかどうかを判定する特徴値となる。

帰結の特徴値

帰結成立に対する帰結の種類

この特徴値は S_0 に対応する S_1 の帰結の種類を示す。「Q」は S_0 に対応する帰結が成立することを示し、「ONLY-Q」は S_0 に対応する帰結のみが存在し、前提条件に対して他の帰結は存在しないことを示している。この特徴値も弱対立関係の細分類に対する判定に影響を与える。

付加情報

S_0 の帰結に相当する S_1 の帰結箇所に帰結の成立する範囲や程度を制限する付加情報があるかどうかを示す。「 A_Q 」は帰結に付加情報が存在することを示し、「 ϕ 」は付加情報が存在

表 1: 弱対立認識関数

弱対立認識関数				文対の関係	言明間の検証結果		文対の例	特徴値			
前提条件		帰結			関係の	不一致数		前提条件		帰結	
種類	付加情報	種類	付加情報					種類	付加情報	種類	付加情報
ONLY-P	A_p	Q/ONLY-Q	*	明示的な弱対立	16	0	S_0 :キシリトールは虫歯予防に効果がある. S_1 :キシリトールは毎日摂って初めて虫歯予防に効果がある.	キシリトールは毎日摂って初めて	毎日	虫歯予防に効果がある	—
P/ONLY-P	*	ONLY-Q	A_q	明示的な弱対立	17	0	S_2 :キシリトールは子どもの虫歯予防にのみ効果がある.	キシリトールは	—	虫歯予防にのみ効果がある	子どもの
P	A_p	Q	*	暗示的な弱対立	65	0	S_3 :キシリトールは食後に摂ると虫歯予防に効果がある.	キシリトールは摂ると	食後に	虫歯予防に効果がある	—
P	*	Q	A_q	暗示的な弱対立	6	0	S_4 :キシリトールは重傷の虫歯には予防効果がない.	キシリトールは	—	虫歯には予防効果がない	重傷の
P/ONLY-P	*	Q/ONLY-Q	Q	同意/対立	284	0	S_5 :キシリトールを食べると虫歯予防に効果がある.	キシリトールを食べると	—	虫歯予防に効果がある	—
							S_6 :キシリトールを食べただけ虫歯予防に効果がある.	キシリトールを食べただけ	—	虫歯予防に効果がある	—
							S_7 :キシリトールを食べないと虫歯予防に効果がない.	キシリトールを食べないと	—	虫歯予防に効果がある	—

しないことを示す。この付加情報により、 S_1 の帰結の成立する程度や範囲が制限されるため、 S_1 が弱対立関係になるかどうかを判定する特徴値となる。

すべての弱対立認識の手がかりを表 1 に示す。表 1 の*は対応する素性のすべての場合を示し、*が存在する関数については上のものが優先される。表 1 の最後の列にそれぞれの文対の例から得られる特徴値を示す。ここで「—」は対象となる値が存在しないことを示す。この特徴値の組み合わせから表 1 を照合することで弱対立関係のタイプを返す。

さらに、(5) や (6) のような実際の文には複数の言明が存在するものがある。

(5) S_1 . ステロイドは副作用がある

S_2 . ステロイドの副作用はステロイド剤を長期に使用した場合におこることが多いですが、短期間の使用では副作用の心配はありません。

(6) S_1 . キシリトールは虫歯予防に効果的だ

S_2 . 虫歯予防はキシリトールだけで済むわけではなく、基本的には規則正しい食生活とキシリトールを毎食後とることで虫歯の予防ができます。

(5) の S_2 では、逆接表現「～が」の前の言明「副作用はステロイド剤を長期に使用した場合におこりやすい」と後の言明「副作用はステロイド剤を短期に使用した場合には心配ない」がそれぞれ前提条件に対する暗示的な弱対立を示している。しかし、文全体では長期とその他の状況の短期の場合の帰結の成立について述べられるため、明治的な弱対立の関係にある文対とみなすことができる。さらに、(6) の S_2 では「～ではなく」という等位節を表す表現の前の言明「虫歯予防はキシリトールだけで済むわけではない」がクエリ全体を否定しているため矛盾の関係となっており、後の言明部分はキシリトールに加えて「規則正しい食生活とキシリトールを毎食後とる」という付加情報を持つ暗示的な弱対立を示している。双方を組み合わせることで、全体の否定から付加条件によって一部が肯定されるため、(6) は明示的な弱対立の

関係となっている文対である。

このように実際の文対には複数の弱対立関係が組合わさることで各々の言明と S_0 との関係と文対の関係が異なることがある。このことから複数の言明を持つ文との関係を判定する場合には、新たに言明ごとの関係の組み合わせを考慮する必要がある。そのために弱対立認識の手がかりの返り値の組み合わせを考慮する特徴値として、言明をつなぐ語となる逆接表現と等位節を表す表現を用いることとする。言明ごとの関係は弱対立認識の手がかりで求め、実際の文対の関係を求める場合には逆接表現や等位節の表現の有無を認識することで複数の言明を考慮すべき文にも対処する。これら以外の複雑な論理式で表される文対については今後の課題とし、本稿では対象としないこととする。

4.2 意味テンプレートに対する予備調査

この節では、弱対立認識の手がかりの効果を実際のデータを用いて検証する。検証のために我々は Web からランダムに取得したユーザーのクエリと検索文書文対のデータセットを作成した。クエリは $\forall x(P(x) \rightarrow Q(x))$ の論理式で示すことのできるものを対象としている。村上⁶⁾ の手法を用いて、各文対に同意・対立・弱対立の関係を付与した。図 2 に示すように、我々が認識すべき弱対立は村上⁶⁾ の定義するその他の関係には存在しないと想定しているため、同意と対立、弱対立の関係となる文対のみのデータセットを作成する。これらの関係は 2 人の日本語母国話者によって確認し、賛成が得られない文対に関しては使用しないこととした。さらに検証で用いるデータセット A については、4.1 節の定義に基づいた弱対立の細分類を含めた 4 種類の関係を各言明ごと付与し、4.1 節で定めた特徴値の正解を手でつけた。これらのデータの内訳を表 2 に示す。データセット A の () 内はそれぞれの言明数を示し、() 外とデータセット B では文対数を示す。

データセット A を用いて、弱対立認識の手がかりの特徴値となる文内の節や単語を人手

表 2: データセット

	同意	対立	弱対立	合計
A	258(272)	8(12)	79(104)	345(388)
B	230	170	200	600

で調査した。

表 1 の 6 列目にそれらの意味テンプレートが示す関係と実際の事例の関係が一致している事例数を示し、7 列目に一致していない事例数を提示する。弱対立認識の手がかりには調査したデータでは対象となる文対が存在しないものも存在したが、実際の文対の例がない関数については人手で対象となる文対を作成することで弱対立認識の手がかりの妥当性を確認した。

我々が得たデータではすべての文対の例を網羅できないため、例外が存在する可能性は否定できないが、特徴値が存在するものがすべて実際の関係と一致することから、弱対立認識の手がかりは弱対立関係を認識するには有効であると考えられる。

5 弱対立の認識実験

前節では、4.1 節で定義した弱対立認識関数に正しい特徴値を入力された場合に弱対立関係を正しく分類できるかについて検証した。この節では、我々は 4.1 節で定めたそれぞれの特徴値を生成するためのルールベースシステムを作成し、より実際の状況のもとで認識関数を用いた弱対立関係分類の実験的な評価を行い、弱対立関係認識の実現性を検証する。

本実験では含意を包括する同意と矛盾を包括する対立の関係を弱対立関係の分類を行う。その他の関係については、図 2 で示したように弱対立の関係が存在しないと考えていることから対象外とし、同意・対立・その他の関係についてはすでに関係分類が行われているものとする。同意や対立の情報を捉えることは引き続き研究が必要であるが、これらの関係からユーザーの必要とする情報のある弱対立関係を認識することを目的としているため、本実験では弱対立関係の認識のみに焦点を当て、同意と対立として認識されてしまう弱対立関係についての分類を行う。さらに同意と対立では弱対立の関係を認識する上で難しさに違いが生じるかを調べるために、3 種類の関係の分類の他に同意と弱対立関係の分類と対立と弱対立関係の分類についての評価も行う。

4 節では言明対での検証を行ったが、実際の文には省略や照応などの複雑な解析を要する問題が残るため、本実験では弱対立認識関数の返り値の組み合わせを用いることで文間の関係についての認識を行う。この組み合わせによる弱対立の判定には 4.1 節で示した逆接表現や等位節の表現を用いる。

以上の条件のもとで我々は弱対立認識関数の特徴値を文対内から判定するルールベースのシ

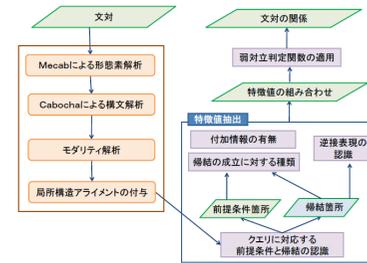


図 3: システム図

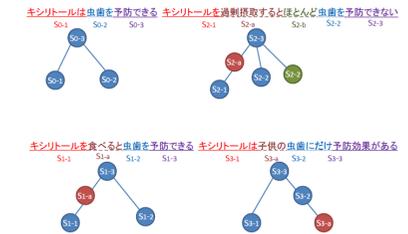


図 4: 文対の構文木例

ステムを作成し、弱対立関係の認識実験を行った。システム図を図 3 に示す。このシステムの素性抽出方法について以下で述べ、実験結果とエラー分析によるルールベースシステムの問題点を示す。

5.1 ルールベースシステム

本節では図 3 で示したルールベースシステムの各処理について述べる。システムはある 2 文対を入力とし、前処理と特徴値抽出を行った後、得られた特徴値を弱対立認識関数に適用して文対の関係を返す。

5.1.1 前処理

2 文対の形態素解析 Mecab¹⁴⁾ を用いて各文をわかち書きにし、各文を構成する単語と品詞を得る。

2 文対の構文解析 CaboCha²⁾ を用いて各文節の係り先を判定し、各文の構文木の構造を得る。

2 文対のモダリティ解析 松吉ら⁴⁾ のモダリティ解析器を用いて、文が持つより充実した情報として各文節の種類¹⁵⁾のモダリティを得る。

2 文対のアライメント 水野ら⁵⁾ の局所構造アライメントを用いて、2 文間の各文節ごとのアライメントを得る。局所構造アライメントでは、各単語のマッチングだけではなく、構文の構造の違いにも対処できるようなアライメント手法を採用している。

5.1.2 特徴値抽出

弱対立認識関数に用いる特徴値を抽出するために、表 2 で示したデータセット A から特徴値となる語彙表現のリストと構造的なパターンを開発用のデータセット A 中の 200 文対から人手で作成した。この語彙リストと図 4 のような構文木による構造パターンを利用すること

で入力された文対の前提条件と帰結から特徴値を自動的に抽出する。各特徴値の抽出方法について以下で述べる。

前提条件の種類

手順1: S_1 の前提条件と S_0 の前提条件の対応に関する判定

S_1 の各文節内の情報が1つめの項目を満たさない場合には S_0 に対応する前提条件でないと判定し「 ϕ 」を返し、それ以外の場合、2つめもしくは3つめの項目が満たされた場合に S_0 と対応する前提条件であると判定し、手順2に進む。

- (1) S_0 の主題とアライメントされる文節が S_0 の帰結とアライメントされる S_1 の文節を構文木の親として持つ
- (2) S_0 の主題とアライメントされる文節もしくはその文節を構文木の子に持つ帰結ではない文節が条件を表す語を持つ
- (3) 「ためには」によって S_1 の前提条件と認識された場合、条件箇所の各文節の内容語すべてがアライメントされている

手順2: 帰結成立に対する前提条件の種類に関する判定

手順1で求めた前提条件が帰結が不成立になる場合について明示的に示しているかどうかについての判定を行う。前提条件の文節の係り先もしくは前提条件の文節内に最も制限された状況を示す語彙を含む場合には「ONLY-P」を返し、含まない場合には「P」を返す。それぞれの手順で使用した語彙は以下の通りである。

条件を表す語: 場合、たら、時、とき、ば、なら、と、で等

制限された状況を表す語 限り、かぎり、だけ、しか、初めて、こそ、(ため)には等

前提条件に対する付加情報の有無

S_0 に対応する前提条件が存在する場合に判定を行う。以下の項目の1つめと2つめ、もしくは1つめと3つめを満たす場合には付加情報は存在しないとして「 ϕ 」を返し、満たさない場合には「 A_p 」を返す。

- (1) 前提条件の文節を構文木の親に持つ文節が存在しないか、存在する場合に S_0 の文節とアライメントされる
- (2) 前提条件の文節構文木の子に持ち、 S_0 の帰結とアライメントされる文節を親に持つ文節が存在しないか、存在する場合に S_0 とアライメントされる
- (3) 前提条件の文節を構文木の子に持つ文節が S_0 の主題の「使用」を表す語彙である

帰結の種類

表 3: 限定関係認識の実験結果

	限定と含意のデータ	限定の矛盾のデータ	全てのデータ
Recall	0.65(129/200)	0.65(129/200)	0.65(129/200)
Precision	0.70(129/185)	0.76(129/169)	0.57(129/225)
F-Score	0.67	0.7	0.61

手順1: S_1 の帰結と S_0 の帰結の対応に関する判定

S_1 の各文節内の情報が S_0 の帰結の各文節とアライメントされない文節が存在する場合、 S_0 に対応する帰結でないと判定し「 ϕ 」を返し、それ以外の場合には S_0 と対応する帰結であると判定して手順2に進む。

手順2: 帰結成立に対する帰結の種類に関する判定

手順1で求めた帰結が不成立となる場合について明示的に示しているかどうかについての判定を行う。帰結の文節の係り元もしくは帰結の文節内に最も制限された状況を示す語彙を含む場合には「ONLY-Q」を返し、含まない場合には「Q」を返す。対象とする語彙には前提条件で使用したものに加えて、「ほとんど」等の確率的な程度を表す表現を用いる。

帰結に対する付加情報の有無 S_0 に対応する帰結が存在する場合に判定を行う。以下の項目のすべてを満たす場合には付加情報は存在しないとして「 ϕ 」を返し、満たさない場合には「 A_q 」を返す。

- (1) 帰結の文節を構文木の親に持つ文節が存在しないか、存在する場合に S_0 の文節とアライメントされる
- (2) 帰結の文節内に特定の程度や範囲を表す語彙が存在しない
- (3) 帰結の文節を兄弟に持つ特定の程度や範囲を表す語彙が存在しない

さらに複数の言明を持つ S_1 に関しては、帰結箇所の係り先もしくは係り元の文節を対象にリスト内の組み合わせを示す表現が存在するかの判定を行い、存在した場合には言明ごとの弱対立認識関数の戻り値の組み合わせから関係を判定する。

上記で示した特徴値を割り当てるための個々のパターンは自動的に学習されるべきであるが、今後の課題として本稿ではルールベースでの実験結果を示すにとどめることとする。すべてのパターンを書き尽くせないような現状のルールベースでも同意や対立内のデータから弱対立関係を認識するのに効果的なシステムであるという結果が得られている。

5.2 実験結果

この節では、前節で述べた自動的な素性抽出と弱対立認識関数を用いて、文対が弱対立かどうかを認識する実験の結果とエラー分析を示す。データは表2で示したデータセットBを用

表 4: Negative Result の事例

		S_0	S_1
False Negative	a	イソフラボンで健康を回復できる.	イソフラボンの健康への効果に期待しすぎでの過剰摂取は禁物です.
	b	キシリトールは虫歯予防に効果がある.	歯を磨く・規則正しい食生活を送る等がきちんと行われている上でキシリトールを用いることが虫歯予防に効果的となります.
False Positive	c	キシリトールは虫歯を予防することができます.	キシリトールを口にしていれば、虫歯を予防できると考えるのは大きな間違いです.
	d	ステロイドで病気は改善できる.	アトピー性皮膚炎は、ステロイドの使用を止めれば完治する.
	e	ステロイドは副作用が懸念される.	ステロイドの副作用は、どのくらいの量でどのくらいの期間使い続けられれば現れるかは人それぞれです.

いる。この実験では同意と対立、弱対立関係を分類する実験に加えて、同意と弱対立、対立と弱対立の分類精度についても評価を行った。結果を表3に示す。

我々のシステムはF値でおよそ6割の精度を得ており、認識できない弱対立関係の多くは特徴値抽出に用いる語彙リストの不足や構文の複雑さによる現状の特徴値を判定するパターンの未対応によるものである。さらに、同意と対立双方では弱対立関係の分類精度に大きな差異は見られなかった。このことから、弱対立関係の認識の難しさは同意や対立の関係には依存しないと考えられる。実際には同意や対立の関係と弱対立の関係の分類にはそれぞれ留意する点があると考えられ、特徴値にも工夫が必要であることが想定されるが、ルールベースのシステムでは特徴値の判定に極性や否定表現などを用いていないために現状では結果に差異が現れなかったと考えられる。

認識に失敗した中の5文対をクエリを S_0 、検索文書文を S_1 として表4に示す。

文対(a)は「イソフラボンの過剰摂取は健康を回復できない」という意味であり、「過剰摂取は」が(a)の S_1 に対しての付加条件となっている。この場合「過剰」が普通に摂取する場合とを区別する特別な状況を示す語彙であり、「は」でとりたてられている。しかし、現状の枠組みでは対比を示す「は」と主題を示す「は」は区別がつかないため、特徴値に利用していないことからこの文対を認識できていないと考えられる。こうした複数の意味を表す特徴値となる語彙を扱うための拡張した手法を考案する必要がある。

文対(b)の付加情報の条件箇所は「キシリトールを用いる」を修飾しているため、条件部分からみた係り先が「キシリトールを用いる」であると判定してしまっている。この事例を捉えるためには、前提条件に対しての修飾部分が付加情報となり得るかどうかを判定する、付加情報を含む前提条件の範囲の判定を行う必要がある。

(c)の S_1 は「口にしていれば」という「使用」を表す語彙が辞書にないために、前提条件が付加情報を含まないと判定できなかった事例である。「使用」を示す語彙には様々なものが存在し、主題にも依存するものであるため、事例を経験的に学習して幅広く取得する必要がある。一方で、(d)の S_1 では「使用をやめれば」という前提条件が S_0 に対応する前提条件の否定となっていることを捉えられずに、前提条件が付加情報を含むと判定してしまっている。この文対から「使用」を示す反意語に関してもリストを作成し、 S_0 の前提条件と対応する語彙として捉える必要がある。

(e)の S_1 の条件は「副作用の現れ方」について述べられており、「副作用が懸念される」ことに対する条件とはなっていない。これはアライメントだけでは捉えきれない意味を含んだ判定が必要であるため、弱対立関係と他の関係とを区別することが困難な事例である。

6 おわりに

Web上の情報や意見の多くには、それらが成立するための特定の条件や制限された情報が存在することがある。この情報はインターネット上で検索を行うユーザがクエリに対して妥当性や有効性を知りたい場合に有効な情報であり、これらの情報を捉えることでユーザはクエリに対してより豊かな知見を得ることができる。しかしながら、情報や意見の関係を認識するための手法である RTE の枠組みでは、現状でそのような情報は含意や矛盾としてしか捉えられていない。

本稿では、我々は同意や対立と同様に扱われてしまう、重要な情報を含む文と命題の関係を弱対立関係とし、その関係の論理的な定義を示し、この弱対立関係を認識するための関数を提示した。この弱対立認識関数は文対から抽出した特徴値を組み合わせることで用いられる。予備調査として正しい特徴値を与えた場合に弱対立認識関数が正しく機能するかどうかを示し、さらにこの特徴値を文対からルールベースで自動的に抽出した場合の弱対立関係の認識の効果について評価し、エラー分析を行った。現状では我々のシステムは人手で取得した語彙のパターンを用いて弱対立認識関数の特徴値を抽出しているため、すべての語彙に網羅していないため認識できない事例が存在する。今後の課題としては、制限や条件を示す様々な語彙に幅広く対応するために、弱対立認識関数のそれぞれの素性を事例から学習することを考えている。

謝辞 本研究で使用したデータのタグ付けにおいて大阪府立大学所属の増田祥子さんに多大な助力を頂いた。さらに本研究は、(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。

参 考 文 献

- 1) Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. Recognizing textual entailment with lcc's groundhog system. In *Proceedings of the Second PASCAL Challenges Workshop*, 2005.
- 2) Taku Kudo and Yuji Matsumoto. Japanese dependency analysis using cascaded chunking. In *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*, pp. 63–69, 2002.
- 3) Bill MacCartney, Trond Grenager, Marie-Catherine deMarneffe, Daniel Cer, and ChristopherD. Manning. Learning to recognize features of valid textual entailments. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006)*, 2006.
- 4) Suguru Matsuyoshi, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. Annotating event mentions in text with modality, focus, and source information. In *The seventh international conference on Language Resources and Evaluation (LREC)*, 2010.
- 5) Junta Mizuno, Hayato Goto, Yotaro Watanabe, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. Local Structural Alignment for Recognizing Semantic Relations between Sentences. In *Proceedings of IPSJ-NL196*, 2010.
- 6) Koji Murakami, Suguru Matsuyoshi, Kentaro Inui, and Yuji Matsumoto. A corpus of statement pairs with semantic relations in Japanese. In *Proceedings of the 15th Annual Meeting of the Association for Natural Language Processing*, 2009.
- 7) DragomirR. Radev. Common theory of information fusion from multiple text sources step one: Cross-document structure. In *Proceedings of the 1st SIGdial workshop on Discourse and dialogue*, pp. 74–83, 2000.
- 8) Mark Sammons, V.G.Vinod Vydiswaran, Tim Vieira, Nikhil Johri, Ming-Wei Chang, Dan Goldwasser, Vivek Srikumar, Gourab Kundu, Yuancheng Tu, Kevin Small, Joshua Rule, Quang Do, and Dan Roth. Relation alignment for textual entailment recognition. In *Proceedings of Recognizing Textual Entailment 2009*, 2009.
- 9) Idan Szpektor, Eyal Shnarch, and Ido Dagan. Instance-based evaluation of entailment rule acquisition. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 456–463, 2007.
- 10) 大木環美, 村上浩司, 水野淳太, 増田祥子, 乾健太郎, 松本裕治. 文間の限定関係認識: 課題設計および分析と予備実験. 言語処理学会第 16 回年次大会発表論文集 D3-1, 2010.
- 11) 村上浩司, 水野淳太, 後藤隼人, 大木環美, 松吉俊, 乾健太郎, 松本裕治. 文間意味的關係認識による言論マップ生成. 言語処理学会第 16 回年次大会発表論文集 PA2-22, 2010.
- 12) 今仁生美. 否定量化文を前件にもつ条件文について. 日本語の条件表現. くろしお出版, 1993.
- 13) 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.
- 14) 工藤拓. 形態素解析器 mecab. <http://chasen.org/~taku/software/mecab/>, 2005.
- 15) 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.
- 16) 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.
- 17) 原田康也野田直彦. とりたて助詞の機能と解釈-量的解釈を中心にして-. 日文研叢書 10 制約に基づく日本語の構造の研究. 国際日本文化研究センター, 1996.
- 18) 益岡隆志. 条件表現と文の概念レベル. 日本語の条件表現. くろしお出版, 1993.
- 19) 益岡隆志. 日本語の条件表現について. 日本語の条件表現. くろしお出版, 1993.
- 20) 益岡隆志. 日本語における条件形式の分化. 条件表現の対照. くろしお出版, 2006.
- 21) 井上和子. 日本語の主文のモーダリティと条件節. *Scientific approaches to language*, Vol.6, pp. 39–73, 2007-03-31.