

B4IM2011

修士論文

文間弱対立関係認識に関する研究

大江貴裕

2016年3月25日

東北大学 大学院  
情報科学研究科 システム情報科学専攻

本論文は東北大学 大学院情報科学研究科 システム情報科学専攻に  
修士(工学) 授与の要件として提出した修士論文である。

大江貴裕

審査委員：

乾 健太郎 教授 (主指導教員)

木下 哲男 教授

篠原 歩 教授

岡崎 直観 准教授 (副指導教員)

# 文間弱対立関係認識に関する研究\*

大江貴裕

## 内容梗概

意見分析において，二文間の同意・対立関係を判定することは重要な役割を担っており，含意関係認識をはじめとする多くの関係認識の研究が行われてきた．しかし，これらは部分的に同意・対立する関係や特定の条件が成立する場合のみ同意・対立する関係を認識することができない．このような意見を認識することができれば，今より詳細な意見分析が行えるようになる．

部分的に同意・対立する関係または特定の条件が成立するときのみ同意・対立する関係を弱対立関係と呼ぶ．現状の弱対立関係の定義は，曖昧なもので広範囲のものになっているため，必要としない文対まで認識してしまう．そこで本研究は弱対立関係を構成している要素を調べ，判断するために必要な手がかりとなる調べた．手がかりとなる必要な要素を整理し，弱対立関係とみなせる要素のパターンを列挙することにより，新たな認識の枠組みを作成した．評価実験では，作成した枠組みを用いて同意・対立関係の文対に適応することで，正確に弱対立関係を認識できることを示した．また，既存の手法を組み合わせた弱対立関係認識器を構築し，評価実験を行い，エラー分析を行った．

## キーワード

自然言語処理, 含意関係認識, 条件認識, 意味関係

---

\*東北大学 大学院情報科学研究科 システム情報科学専攻 修士論文, B4IM2011, 2016年3月25日.

# 目次

<b>1</b>	<b>はじめに</b>	<b>1</b>
1.1	本研究の背景	1
1.2	本研究の目的	2
1.3	本論文の構成	2
<b>2</b>	<b>関連研究</b>	<b>4</b>
2.1	含意関係認識	4
2.1.1	Partial Entailment	4
2.1.2	Natural Logic による含意関係認識	6
2.2	二文間の類似・関連箇所の同定	6
2.3	弱対立関係の定義	8
2.4	弱対立関係の認識	10
<b>3</b>	<b>弱対立関係の構成</b>	<b>12</b>
3.1	弱対立関係の詳細	12
3.2	弱対立関係の構成要素	14
3.3	手がかりとなる要素	16
3.3.1	条件の差	16
3.3.2	程度差	17
3.3.3	確信度の差	18
<b>4</b>	<b>弱対立関係認識の枠組み</b>	<b>20</b>
4.1	弱対立関係認識の枠組み構築の手順	20
4.2	弱対立関係認識の枠組み	21
<b>5</b>	<b>弱対立関係認識の枠組み評価実験</b>	<b>23</b>
5.1	実験設定	23
5.2	実験 1	24
5.3	実験 2	25

5.4 エラー分析 . . . . .	25
<b>6 弱対立関係認識器の構築</b>	<b>28</b>
6.1 概要 . . . . .	28
6.2 条件の差 . . . . .	29
6.2.1 条件の探索範囲の拡大 . . . . .	30
6.2.2 使用する素性 . . . . .	30
6.3 程度の差 . . . . .	33
6.4 確信度の差 . . . . .	33
<b>7 弱対立関係認識器の評価実験</b>	<b>35</b>
7.1 条件の差の認識の評価実験 . . . . .	35
7.1.1 実験設定 . . . . .	35
7.1.2 実験結果 . . . . .	36
7.1.3 エラー分析 . . . . .	36
7.2 弱対立関係認識の評価実験 . . . . .	38
7.2.1 実験設定 . . . . .	38
7.2.2 実験結果・考察 . . . . .	39
<b>8 おわりに</b>	<b>41</b>
謝辞	42

## 目 次

1	Natural Logic による文間関係の演算 . . . . .	6
2	文節アライメントの選択 <small>[1] より引用</small> . . . . .	8
3	大木らの弱対立関係定義 . . . . .	14
4	本研究の弱対立関係定義 . . . . .	15
5	弱対立関係認識器の概要 . . . . .	29
6	条件認識実験の評価結果：P-R 曲線 . . . . .	36
7	弱対立関係認識実験の評価結果：P-R 曲線 . . . . .	39

## 表目次

1	Natural Logic で定義される意味関係 . . . . .	5
2	条件文中の意味関係の射影 . . . . .	5
3	弱対立関係定義の比較 . . . . .	16
4	弱対立関係認識の枠組みの適用例 . . . . .	21
5	弱対立関係認識の枠組み . . . . .	22
6	提案する枠組みによる弱対立関係認識の評価 . . . . .	24
7	実験 2 の評価結果 . . . . .	25
8	使用する素性 . . . . .	32

# 1 はじめに

## 1.1 本研究の背景

近年ブログや SNS などに自分の意見を書く人が多くなり、インターネット上には多種多様な意見が存在するようになった。これらの意見を分析することで、任意のトピックについて多くの人の考えを知ることができる。このように現代の社会では言語資源を活用した意見の集約、整理は有用なタスクとなっている。意見分析には二文間の同意・対立関係などの意味関係を理解する必要があり、含意関係認識などが用いられている。含意関係とは、与えられた二つの文の一方が真であるときに、他方の文も真と推論可能であることを指し [2]、与えられた二つの文が同時に真と成り立たない関係という [3]。二文間の含意関係認識は、近年 NTCIR-11 RITE-VAL [4]<sup>1</sup>などをはじめ盛んに研究が行われている。

これらの研究では同意・対立・その他の 3 値分類を行っており、条件付きで同意・対立関係になる関係を認識することはできない。(1) の仮説はテキストに対して「有機栽培」という条件が付くことで同意している。解釈するとテキストは「有機栽培のコーヒー」については健康に良いと述べている一方、「有機栽培ではないコーヒー」については健康に良くないことを示唆している。これは部分的に同意している一方、部分的に対立していることを示唆していることとなる。

(1) H コーヒーは健康に良い

T 有機栽培のコーヒーは健康に良い

このような部分的に同意・対立するまたは、特定の条件が成立するときに同意・対立する関係を弱対立関係と呼ぶ。弱対立関係を認識することによって、より詳細な意見分析が行えるようになる。

大木らはテキストを前件と帰結に分割し、それぞれの付加情報によって弱対立関係を定義した [5, 6]。しかし、付加情報に関する記述が曖昧なため広い範囲で認識してしまい、必要ではない文対も認識してしまう問題点がある。また、現状

---

<sup>1</sup><https://sites.google.com/site/ntcir11riteval/>

の弱対立関係認識はルールや簡単な素性を用いた機械学習を用いた簡単なものとなっている。

## 1.2 本研究の目的

現状の弱対立関係の定義が曖昧で広い範囲を認識する問題点に対して、定義を明確にし必要な文対だけを認識させることを目的とした。多くの事例を分析し、弱対立関係かどうかを判定するために必要な手がかりとなる要素を調べた。必要な要素として条件の差、程度の差、確信度の差の三つの要素を取り上げる。この認識の手がかりである3要素の各要素がどのような組み合わせの場合に弱対立関係となるかを整理することで、新しい認識の枠組みを提案する。3要素を新たに取り上げることで、曖昧だった定義が明確になり、不要な文対を除き、必要な文対だけを認識することができる。提案する枠組みを用いることによって、正確にかつ必要な弱対立関係となる文対を認識することができた。

既存の弱対立関係認識よりも誤った認識を減らした高精度な認識器の構築を試みた。認識には以下の3つの手順で行う。

1. 局所構造アライメントで二文間の対応する箇所の同定を行う。
2. 前件と後件のそれぞれの意味関係と条件の差、程度の差、確信度の差の3要素に関して調べる。
3. 提案する枠組みに適応することで弱対立関係か否かを判定する。

この認識器の評価を行い、エラー分析を行うことでこの手順の中にある問題点を述べる。

## 1.3 本論文の構成

本論文の構成は以下のようなものである。第2章で本研究と関連する含意関係認識の手法と、弱対立関係認識の先行研究を述べ、本研究の立ち位置を明確にする。第3章で弱対立関係の構成要素を考えることで認識の手がかりとなる要素の整理を

行い，第4章で手がかりとなる要素がどのような組み合わせのときに弱対立関係と認識されるかをまとめた認識の枠組みを提案する．第5章で提案した枠組みが正しく弱対立関係を認識できるかの評価実験を行い，残った課題を述べる．第6章で新たな枠組みを用いた弱対立関係認識器を構築し，第7章で認識器の評価実験を行い，分析，考察する．最後に第8章で本研究のまとめと残された課題について述べる．

## 2 関連研究

弱対立関係認識などの二文間の関係認識に関する研究として、含意関係認識があり、その中でも、本研究に近い研究をいくつか紹介し、弱対立関係との違いを明確にする。また、弱対立関係認識のこれまでの研究の推移を説明する。

### 2.1 含意関係認識

含意関係とは、与えられた二つの文の一方の文から他方の文が推論可能なことを指す。含意関係認識とは、文間の意味関係を理解する上で重要なことである。RTE や RITE のワークショップも多く開催され、自然言語処理の分野で様々な取り組みが行われている。

#### 2.1.1 Partial Entailment

Dzikovska らは、条件に限らず文中の表現のわずかな差異によって含意関係にない文対に対して、部分的な含意関係を認識する課題 (Partial Entailment) を提案した [7]。仮説 H をいくつかの要素に分解して考え、T が分解した要素の全てではなく一部を推論できる関係を部分的な含意関係とした。分解には Nielsen の *facets* モデルを活用している [8]。*facet* は二つの単語のペアと単語間の関係から構成される。例えば、(2) の H には (*muscles, move*) と (*move, bones*) の二つの *facet* に分解できる。

(2) H The main job of muscles is to move bones.

T Muscles generate movement in the body.<sup>2</sup>

上述の (2) の T と H に関して、*facet(muscles, move)* は T から推論可能だが、*facet(move, bones)* は推論できない。また、T が H の構成要素に対して、構成要素を否定する要素を含んでいる場合や構成要素とは関係無い要素を含んでいる場合の組み合わせで Partial Entailment を分類した研究も存在する [10]。このよう

---

<sup>2</sup>[9] より引用

表 1: Natural Logic で定義される意味関係

意味関係	記号	例
同値 (equivalence)	$\equiv$	カウチ $\equiv$ ソファ
前向き含意 (forward entailment)	$\sqsubset$	カラス $\sqsubset$ 鳥
後ろ向き含意 (reverse entailment)	$\supset$	鳥 $\supset$ カラス
否定 (negation)	$\wedge$	男性 $\wedge$ 女性
交代 (alternation)	$ $	日本人 $ $ アメリカ人
包含 (cover)	$\smile$	男性 $\smile$ 日本人
独立 (independence)	$\#$	空腹 $\#$ カバ

表 2: 条件文中の意味関係の射影

connective	projectivity						
	$\equiv$	$\sqsubset$	$\supset$	$\wedge$	$ $	$\smile$	$\#$
negation ( <i>not</i> )	$\equiv$	$\supset$	$\sqsubset$	$\wedge$	$\smile$	$ $	$\#$
conjunction ( <i>and</i> ) / intersection	$\equiv$	$\sqsubset$	$\supset$	$ $	$ $	$\#$	$\#$
disjunction ( <i>or</i> )	$\equiv$	$\sqsubset$	$\supset$	$\smile$	$\#$	$\smile$	$\#$
conditional( <i>if</i> ) (antecedent)	$\equiv$	$\supset$	$\sqsubset$	$\#$	$\#$	$\#$	$\#$
conditional( <i>if</i> ) (consequent)	$\equiv$	$\sqsubset$	$\supset$	$ $	$ $	$\#$	$\#$
biconditional ( <i>if and only if</i> )	$\equiv$	$\#$	$\#$	$\wedge$	$\#$	$\#$	$\#$

[11] Table 6.1 より引用

な部分的な含意関係でない箇所を認識することが Partial Entailment であるため、本研究の提案する条件の厳しさや程度の差を考慮した弱対立関係の認識は Partial Entailment の一部として考えられる。しかし、条件の差や程度の差などをはじめとする部分的に含意関係でない箇所は、T 側の結論に大きな影響を与えるため、部分的な含意の認識だけでなく、その定義を明確にすることは重要である。



メントされた文節間の構造に対してアライメントを行う．最後に意味関係認識に用いる文節アライメントの選択する．

文節アライメントは，表層的な類似度，語彙資源の知識による類似度に加えて，文構造に基づく推定をする．例えば，(3)ではQの「使用する」とTの「散布する」は同じ意味を表しているが，表層，語彙知識を用いても類似度が高いとは言えない．二文間でトピックを共有している場合には，複数の同じ項を共有している述語は類似語である可能性が高くなり，QとTはこの状況に一致する．この予測に基づき，一定の文節，項を共有している述語にもアライメントをとる．

(3) Q 畑で 農薬を 使用する

T 畑に 農薬を 散布する<sup>3</sup>

局所構造アライメントは，文節アライメントで対応された文節間の関係に対応づける．(4)では， $Q_a - Q_c$ が「ブラックバス<sub>a</sub> - ガー破壊<sub>c</sub>」という格関係が成り立つ，これは $T1_a - T1_c$ と同じ関係であり，対応づけられる．また，T2は「魚類<sub>a</sub> - ガー破壊<sub>c</sub>」の格関係があるが「魚類」は「ブラックバス」に置き換えられるので， $T2_a - T2_c$ も $Q_a - Q_c$ と同様の関係になる．一方， $T3_a - T3_c$ は格関係が成り立たないため，アライメントされない．このようなアライメントをT側で文節アライメントされた文節に対して，直接もしくはいくつかの文節を介した修飾関係にあり，述語項構造関係でもあるものに行う．

(4) Q ブラックバス<sub>a</sub> は 生態系<sub>b</sub> を 破壊する<sub>c</sub>

T1 ブラックバス<sub>a</sub> は有名な外来魚であり，生態系<sub>b</sub> を 破壊する<sub>c</sub>

T2 ブラックバス<sub>a</sub> のように獰猛な魚類が 生態系<sub>b</sub> を 破壊する<sub>c</sub>

T3 ブラックバス<sub>a</sub> を駆除することが 生態系<sub>b</sub> を 破壊する<sub>c</sub> <sup>4</sup>

局所構造アライメントが付与された文節アライメントの中から意味関係認識に有用かを選択する必要がある．クエリの述語の依存構造に対応する局所構造をよ

---

<sup>3</sup>[1] より引用

<sup>4</sup>[1] より引用

り多くもつテキストの述語を選択する．図 2 を例に説明する．横の文がクエリ，縦の文がテキストであり，クエリの「良い」がテキストの「悪いので」と「良い」に対応している．テキスト「悪いので」と対応している「良い」は「マーガリンは」，「体に」を係り元にもち，2 つの局所構造に対応している．一方，テキスト「良い」は 1 つの局所構造にしか対応していない．多くの局所構造をもつアライメントを選択するので，文節アライメント番号 “1”，“2”，“4” が選択され，“3”，“5” は選択されない．この手法で述語が選択できない場合は，文末に近い述語を選択する

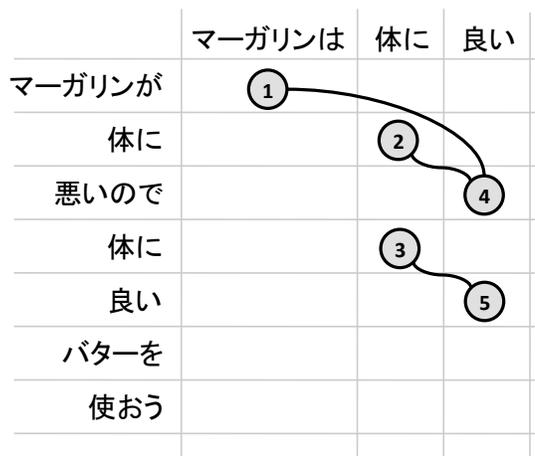


図 2: 文節アライメントの選択 [1] より引用

### 2.3 弱対立関係の定義

大木らは，二文が与えられたとき一方の文が多保の文に対して，条件付きで同意・対立しているまたは，帰結の範囲を制限した二文間の関係を弱対立関係とした [13]．これは一方の文が他方の文を部分的に同意・対立していると言える．付加情報について着目し，この関係を論理式を用いて定義した．

$$\forall x(P(x) \rightarrow Q(x)) \quad (1)$$

仮説を式 (1) と表現した。  $P(x)$  は前提条件，  $Q(x)$  は帰結を表す。文「コーヒーは健康に良い」では前提条件が「コーヒー」，帰結が「健康に良い」となる。この仮説に対して，前提条件，帰結のどちらかに付加情報がついた式 (2)，式 (3) が弱対立関係となる。

$$\forall x((P(x) \wedge AC(x) \rightarrow Q(x))) \quad (2)$$

$$\forall x((P(x) \rightarrow Q_{AR}(x))) \quad (3)$$

$AC(x)$  が付加条件 (*Additional Condition*)，  $Q_{AR}$  程度表現 (*Additional Restriction*) により制限された帰結を表す。式 (2) は式 (1) に比べて帰結の成立に条件が付加されており，式 (3) は帰結が程度表現によって制限されている部分的な同意・対立関係を表している。次に実際のテキストで考えてみる。

(5) H コーヒーは健康に良い

T1 有機栽培のコーヒーは健康に良い

T2 コーヒーは少しは健康に良い

例 (5) で考えると，H の前提条件  $P(x)$  「コーヒー」に対して，T1 の前提条件は「有機栽培のコーヒー」であり， $P(x)$  に付加条件  $AC(x)$  「有機栽培の」がついているため弱対立関係となる。また，H の帰結  $Q(x)$  「健康に良い」に対して，T2 の帰結は「少しは健康に良い」であり，程度表現「少し」が帰結を制限しているため  $Q_{AR}(x)$  と表され弱対立関係となる。

大西らは，付加情報と付加先の関係を限定関係と呼び，付加情報は類似した同様の要素との対比を示唆するとし，限定関係を構成する要素としてとりたて助詞，ノ格，数量・程度表現などを取り上げた [14]。下記の文ではブラックコーヒーがその他のコーヒーの中から取り立てられ対比していることがわかる。とりたて助詞は，「コーヒーの中でもブラックコーヒーは健康に良い」の「は」をはじめとして「も」「こそ」などの助詞によって，要素をとりたてる用法である [15]。「は」は，主語を表す用法である場合と，とりたて助詞となる場合があり，弱対立関係の付加情報となるのは後者である。

しかし、この弱対立関係の定義の付加条件や程度表現に関する記述が乏しく、定義自体が曖昧なものとなっている。そのため、(6)のT1、T2といった部分的に同意していると考えにくい文まで認識してしまう。

(6) H コーヒーは健康に良い

T1 コーヒー飲料は健康に良い

T2 コーヒーは非常に健康に良い

T1は「飲料」を前提条件の付加条件、T2は「非常に」を帰結を制限する程度表現とみなしてしまう。T1は「コーヒー飲料」と表記しているものの、「コーヒー」に関して対立する部分がないため、Hに対して部分的に同意しているとは言いがたく、同意の関係に近い。T2は程度表現「非常に」があるものの、Hの主張をより強く述べており、同意の関係にある。本研究は、これらを条件の差、程度の差などの新たな要素に着目することでより精密な定義を行い、必要のない文を除いている。

## 2.4 弱対立関係の認識

大木らはモダリティ解析[16]と水野らの局所構造アライメント[1]を用いて、二文の前提条件と帰結それぞれの対応する箇所を同定する。次に、付加条件、帰結を制限する程度表現があるかを調べて、弱対立関係か判定した[13]。付加条件の判定には、「場合」「限り」などの条件表現を含み、「および」「または」といった並列を表す単語などの例外的な表現を含まないとしたルールを用いて行った。同様に帰結の文章に程度を制限する語句が含まないかをルールで判定した。これらの組み合わせによって、最終的な弱対立関係判定を行った。

大西らは係り受け関係にある文節ペアに対して、限定関係であるかを調べ、アライメントされている文節が限定関係の付加先である場合に弱対立関係であると判定した[14]。限定関係認識には、ルールベースと機械学習の二つの手法で実践した。ルールベースでは大木らと同じ考えで、「場合」「時」などの条件を表す語か「限り」「だけ」などの制限された条件を表す語を含んでいる場合限定関係と

した．一方機械学習では，とりたて助詞，ノ格，数量・程度表現などを素性としている．

上記の手法では，ルールベースでは適合率が高い代わりに精度が低くなってしまい，機械学習の手法でも精度は決して高いとは言えない．使用箇所として，インターネットなどからの大量の言語資源を認識させると考えられるため，網羅して認識するのではなく，確実に弱対立関係と言える文対がとれてほしい．そのため，現状よりも高い精度の認識器の構築を目的とし，精度を上げるために，定義をより明確なものとした枠組みを適用させた．

### 3 弱対立関係の構成

弱対立関係とは、条件や程度表現がつくことによって、部分的に同意・対立する関係である。この文章からでは、弱対立関係と認識される文対に含まれる条件や程度表現はどのような特徴を持つのか、部分的に同意とは一部と同意関係であるだけなのか、それとも対立関係にある部分も存在しているのかなどの疑問点が生じる。この章では弱対立関係をもう一度詳しく考え直してみるとともに、定義を明確にし、疑問点の改善を試みる。さらに、その過程において考えた構成要素から、認識するうえで手がかりとなる要素を見つける。

#### 3.1 弱対立関係の詳細

条件が付加された同意の関係とは、同意関係とどのように異なるのかを次の仮説 H と様々なテキスト T 間の関係を見ながら考えてみる。便宜上、本研究は前提条件を前件、帰結を後件と呼ぶ。(これ以降は前件、後件と表記する。)

(7) H コーヒーは健康に良い。

T1 コーヒーは体に良い効果がある。

T2 コーヒーは胃に良い。

H は前件が「コーヒー」、後件が「健康に良い」となる一般的な意見文である。最初に、同意の関係にある文との関係を考える。T1 は前件が「コーヒー」、後件が「体に良い効果がある」であり、H を同意する文になっている。さらに、H は T1 を同意しているため H と T1 は同値の関係にある。一方、T2 は後件の「胃に良い」が「健康に良い」と推論可能なので、H を同意している。しかし、「健康に良い」から「胃に良い」は推論できないので同値の関係とは言えない。

(8) T3 インスタントコーヒーは健康に良くない。

T4 濃いコーヒーは胃に良い。

次に条件がつくことで同意・対立関係である弱対立関係となる文との関係を考える。T3は前件が「インスタントコーヒー」であり、Hと比べて「インスタント」という条件が付加されている。この文は「インスタントコーヒー」に限定して述べており、Hの前件「コーヒー」のうち「インスタントコーヒー」の部分に限定して考えると対立している。「インスタント以外のコーヒー」に関しては何も言及していないが、ここでわざわざ「インスタント」という条件をつけたことを考えてみよう。条件をつけることによって、「コーヒー」を背景に「インスタントコーヒー」を「インスタント以外のコーヒー」と対比させることで、特別な意味を与えている。この現象は条件が成立する場合は後件が成立し、条件が不成立の場合は後件が成立しないことを示唆する。すなわち「(インスタント以外の) コーヒーは健康に良い」が「インスタントコーヒー (に関しては) 健康に良くない」と解釈することができる。T3はHに対して「インスタントコーヒー」の部分は対立し、「インスタント以外のコーヒー」の部分では同意していると解釈することができる。

T4をT3と同様に考えると「濃いコーヒー (に関しては) は胃に良い」から「濃くないコーヒーは胃に良いわけではない」と解釈することができる。「濃くないコーヒー」に関しては「胃に良いわけではない」と解釈できるだけで、「健康に良い」、または「健康に良くない」と解釈することはできない。すなわちT4はHに対して「濃いコーヒー」の部分は同意しており、他の部分に関しては同意も対立もしていない。

(9) T5 コーヒーはわずかに健康に良い。

T6 コーヒーはわずかに健康に悪い。

最後に後件に程度表現がついている文との関係を考える。T5の前件はHと同じで、後件は「わずかに健康に良い」と程度表現「わずかに」を含む形となっている。程度表現「わずかに」を付与したということは、健康に良い効果はあるが、程度表現を何も付与していない「健康に良い」と表記するよりも健康に良い効果は少ないことを表している。T5はHに対して、完全に同意しているとはいえないが、全く同意していないともいえない。つまり効果の有無という部分において

は同意しているが、効果の程度においては完全には同意していない(ただし、対立しているとも言い難い)。T6もT5と同様に考えると、対立しているが「わずかに」により、完全に対立しているとはいえない。T6はHに対して、効果の有無に関しては対立しているが、効果の程度に関しては完全に対立しているとは言えない。

以上のことをまとめると弱対立関係とは部分的に同意(対立)し、部分的に対立(同意)する関係と、部分的に同意(対立)しているが、他の部分に関しては言及していない関係が存在する。

### 3.2 弱対立関係の構成要素

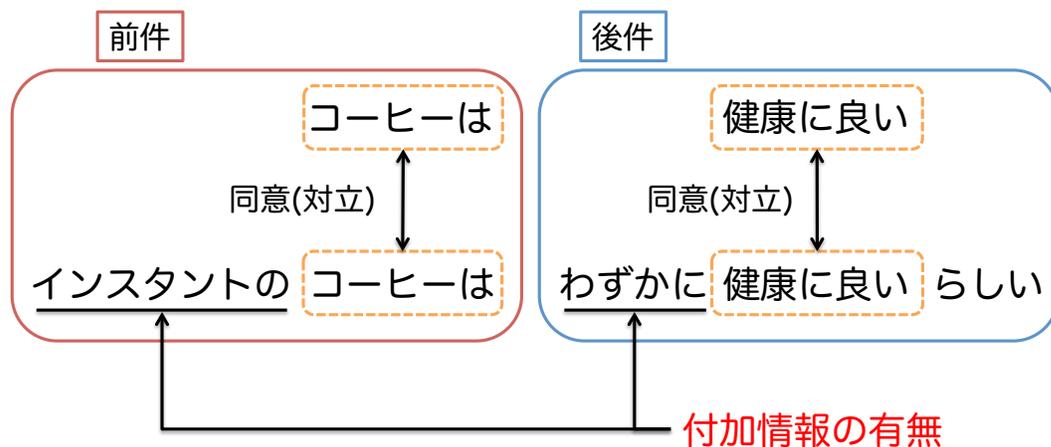


図 3: 大木らの弱対立関係定義

これまでの弱対立関係の定義と比べながら、本研究で考える弱対立関係の構成要素について述べる。弱対立関係認識する上で文H、Tをそれぞれ前件と後件に分けて考える。大木らの弱対立関係定義を図3で示す[13]。Hの前件、後件にそれぞれ対応するTの要素を同定し、その要素に対して付加情報の有無を調べ、付加情報があると判断されれば弱対立関係と認識される。ここで注意しておきたいことは、対応する要素の意味関係は同意もしくは対立の関係に限定される点である。この定義によるとT側の弱対立関係の構成要素は、Hの前件と後件に加えて、

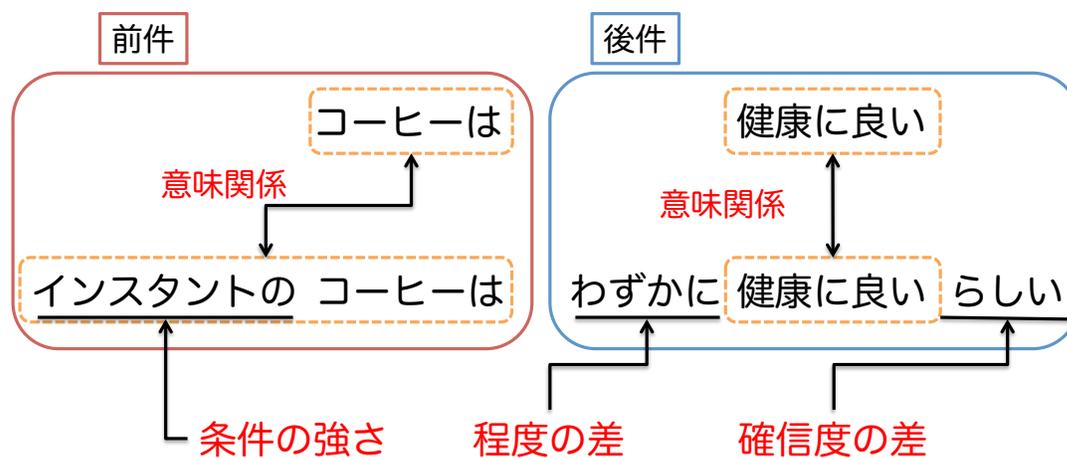


図 4: 本研究の弱対立関係定義

付加される情報の3つである。しかし、前述したように付加情報に関する制約はないため、どのような付加情報でも弱対立関係と認識してしまう。そこで、ここで取り上げた構成要素をより明確にした上で、さらに新たな構成要素を考える。

本研究は図4のように定義を行う。前件と後件に分けて考えるが、扱う要素を増やすとともに、意味関係を以前よりも詳細に考える。今まではHの前件と後件にそれぞれ対応する要素の意味関係は同意、対立のみとしていたが、例8のように後件の意味関係が同値と同意では、解釈が異なることがわかる。そのため、同意、対立のみから拡張し、より広い意味関係を扱う必要がある。そこで本研究では Natural Logic を用いた含意関係の手法 [11] で用いられる意味関係を用いる。

付加情報として考えられてきた前件に付加される条件について、条件の差という要素を提案する。条件付きの同意関係の場合、条件が不成立の場合に対立を示唆するが、条件の差が小さいとその限りではない。また、どのような程度表現も扱っていたが、「わずかに良い」と「かなり良い」では、同意・対立の具合が異なってくる。程度の差を考え、程度表現の差異による認識の違いを捉える。最後に新たな要素として確信度を提案する。図4では「らしい」という表現がついており、後件が成立する確信度が小さい。確信度が小さいと完全に同意するとは言いがたく、部分的な同意になる。また、「場合がある」などの確信度の小さい表現のときに、「成立しない場合もある」と解釈されることもあるため、弱対立関係の認識す

表 3: 弱対立関係定義の比較

	前件，後件の意味関係	前件の付加情報	後件の付加情報	確信度
大木らの弱対立関係定義	同意，対立	任意の条件	任意の程度表現	扱いなし
本研究の弱対立関係定義	NL の 7 つの意味関係	条件の差	程度の差	確信度の差

る上で重要な要素である．これらの要素は弱対立関係を認識する上での手がかりとなる要素になる．

大木らの定義と本研究の定義を比較すると，表 3 のようになる．意味関係や付加情報に関して，差を扱うなどして詳しく調べ，新たな要素として確信度を追加した．次の節では，前述した条件の差，程度の差，確信度の差について詳しく述べていく．

### 3.3 手がかりとなる要素

#### 3.3.1 条件の差

条件は前件および後件のどちらにも付加されうるが、「子どもにはコーヒーが健康に良い」と「コーヒーは子どもの健康に良い」が言い換え可能であるように，前件に付加された条件と後件に付加された情報の両方をまとめて取り扱う．条件は，それが付加される前後で，後件の成立・不成立が大きく変化する場合に「差が大きい」，ほぼ変化しない場合に「差が小さい」という。「差が大きい」場合は条件が不成立のとき，後件が不成立になり，同意・対立関係が逆転する．一方「差が小さい」場合は条件が不成立のときも貢献が成立し，同意・対立関係はそのままである．

(10) H コーヒーは健康に良い

T1 1日3杯のコーヒーは健康に良い

T2 コーヒーを飲むと健康に良い

T3 ブラックコーヒーならば健康に良い

T4 ブラックコーヒーでさえも健康に良い

### T5 ホットコーヒーは健康に良い

(10) において、T1には「1日3杯」という条件が付加されており、この条件は一般的に飲まれる量を超えていると考えられるため、有無によって「健康に良い」の成立・不成立は変化する。一方で、T2はHと比較して「飲むと」という条件が付加されているが、コーヒーは一般的に飲むものであるため、条件の有無によって「健康に良い」の成立・不成立は変化しない。

条件の強さは、T2のように内容語によるものだけでなく、T3, T4のように文脈によっても表される。T3は「ならば」という表現で「ブラック」という条件を強めているが、T4の「でさえも」という表現は「ブラック」以外でも「健康に良い」ことを示唆しており、T3と比較すると弱い条件となっている。

条件の差は文脈によって表される場合は明確に現れる。T3は「ブラック以外のコーヒーは健康に良くはない」、T4は「ブラック以外のコーヒーも健康に良い」と推論することは容易である。しかし、内用語による条件の差は明確には現れない。T5には「ホット」という条件が付加されているが、条件の差を大きいと見るか小さいと見るか人によって異なってくる。「ホットではないコーヒーは健康に良くない」と解釈可能であり、また「ホットではないコーヒーも健康に良い」とも解釈可能である。このように内用語の条件の差をどのように扱うかは難しい問題である。

### 3.3.2 程度差

後件に付加されうる「かなり」「少し」「間違いなく」といった程度表現は、語彙によって程度の強さが異なる。

#### (11) H コーヒーは健康に良い

T1 コーヒーは少し健康に良い

T2 コーヒーはかなり健康に良い

T3 コーヒーはそこそこ健康に良い

(11)において、T1には「少し」という程度表現が「健康に良い」に付加され、「健康に良い」が成立しない場合があることを示唆しており、Hと比較して程度量は小さい。ここで程度量とは、文の後件の効果の範囲・具合について指す尺度と考える。一方で、T2は「かなり」という程度表現が付加されているが、Hに対してより強調しているだけであり、この表現の有無によって「健康に良い」の成立・不成立は変化しなく、Hと比較して程度量は大きい。T3では「そこそこ」という程度表現が付加されているが、これは後件の程度が普通のことを表しているため、この表現の有無によって後件の成立・不成立は変化しなく、程度量はHとほぼ同等である。このように、TがHよりも程度量が小さい場合は後件の成立・不成立が変化し、程度量が同等または大きい場合には後件の成立・不成立が変化しない。これは程度表現が付加されている文同士でも比較可能である。例えば、T1はT2に比べて程度量が小さい、T2はT3に比べて程度量が大きいと判定することができる。

### 3.3.3 確信度の差

後件の述語に対する著者の確信度に差がある場合に、弱対立関係になりうる。程度の差と重なる部分があるが、確信度の差は成立／不成立に対する推量、伝聞といった差であり、程度の差は確信度の差の高低に関わらずその程度の強さである。従って、「かなり良いと聞いたことがある」は、程度と確信度の両方とも差があると判断する。

(12) H コーヒーは健康に良い

T1 コーヒーは健康に良い可能性がある

T2 コーヒーは健康に良いのは証明されている

T3 コーヒーは健康に良い場合がある

(12)では、T1は「可能性がある」によって「健康に良い」の確信度が低く、「健康に良い」が成立しない場合があることを示唆していることから、Hと比較して

確信度の差がある。一方で、T2は「証明されている」によって確信度の差が高いことを示唆しており、Hと比べて確信度の差は少ない。

T3は「場合がある」を含んでいるため確信度が低く、後件が成立しない場合があることを示唆している。大木らはある暗黙的な条件が成り立つ場合に後件が成立として、条件が付与されていると考えたが、本研究では条件が不明瞭なため確信度として扱う。

## 4 弱対立関係認識の枠組み

### 4.1 弱対立関係認識の枠組み構築の手順

3.2 で述べた構成要素の組み合わせで弱対立関係と認識できるパターンの整理を行う。各組み合わせに対応する文対を作成し、二文間の意味関係により弱対立関係か否かを判定した。

H の前件と後件、それぞれに対応している T の要素間の意味関係を考える。意味関係は、MacCartney の Natural Logic に基づく手法 [11] と同等の 7 つの意味関係記号を用いる (表 1)。なお、意味関係は MacCartney の手法では T から H への変換を行うことに合わせ、T[意味関係]H の並びでの意味関係を考える。条件の差は、内容語によるものと文脈によるもので分けて考え、[差が大きい]、[差が小さい]、[差がない] の 3 値で扱い、それぞれ [強]、[弱]、[無] と表記する。程度の差は、T の程度量が H の程度量に比べて [大きい]、[小さい]、[差がない] の 3 値で扱い、それぞれ [大]、[小]、[無] と表記する。確信度の差は、T の確信度が H の確信度に比べて [低い]、[差がないもしくは高い] の 2 値で扱い、それぞれ [P ( Probable )]、[C ( Certain )] と表記する。

弱対立関係の判定は、大木らと同様に [明示的, 暗示的] の 2 値とした。下記の (13) を例にそれぞれについて述べる。

(13) H コーヒーは健康に良い

T1 ブラックコーヒーならば健康に良い

T2 ブラックコーヒーは健康に良い

明示的な弱対立関係とは明らかに付加条件の前後で、後件の成立・不成立が変化する関係である。(13) の T1 は「ならば」が含まれており、条件を強く限定している。このとき、T2 は条件「ブラック」が不成立の「ブラック以外のコーヒー」は「健康に悪い」と明確に解釈できる。このように T1 は部分的に同意し、部分的に対立していることが明確に分かる。このような関係を明示的な弱対立関係という。

表 4: 弱対立関係認識の枠組みの適用例

	T	H	意味関係 (語彙)	条件の差 (語彙)	条件の差 (文脈)	程度の差	確信度の差	弱対立
前件	コーヒー, 1日3杯	コーヒー	□	強	無	-	-	暗示的な
後件	胃腸に良い	健康に良い	□	-	-	小	C	弱対立

一方, 暗示的な弱対立関係とは上記の現象が明確にわからない関係である。(13)のT2は条件「ブラック」が与えられている。T2と同じように「ブラック以外のコーヒー」は「健康に悪い」と解釈できるが, この解釈は絶対ではない。このように明確には部分的な同意, 部分的な対立が混同しているとは判断できない関係を暗示的な弱対立関係という。

前述した要素の各組み合わせの文対を作成し, 人手で弱対立関係か判定を行うのを繰り返し行った。ここで文対を作成し, 弱対立関係か判断を行う一例をあげる。前件の意味関係が[□], 後件の意味関係が[∧], 条件の差の語彙, 文脈とも[強], 程度の差, 確信度の差がともに[無]の文対では, (??)のような文対が作成される。この文対は条件「ブラック」が付加されており, 付加前後で文の成立・不成立が変わるため, 弱対立関係である。この作業を全ての組み合わせに対して行い, 弱対立関係と判断された組み合わせを整理することで認識の枠組みを作成した。

## 4.2 弱対立関係認識の枠組み

表5に提案する弱対立関係認識の枠組みを示す。

(14) H コーヒーは健康に良い

T コーヒーを1日3杯飲むと, わずかだが胃腸に良い効果がある

例えば, 上述の(14)に対して表5の弱対立関係の認識を適用すると, 表4の通り, TとH間は暗示的な弱対立であると得られる。まず, Tの前件に「1日3杯」という条件が付加されているため, 前件の意味関係は[□]となり, Tの後件からHの後件が推論可能なことから, 後件の意味関係は[□]と判定される。また, 「1日に3杯」が「コーヒー」に強い限定条件を与えていることから, 語彙の条件の差は[強], Tの文脈に前件の条件を強く限定する表現が存在しないことから, 文

脈の条件の差も [無] となる。程度の差は「わずかに」と程度表現が与えられており、程度量が H より小さいと考えられることから、[小] となる。一方、確信度に関しては、それらと関連する表現が T に存在したいため、[C] が割り当てられる。

このようにして得られた各要素が表 5 に存在する場合は、T と H が弱対立関係であることを意味し、明示的もしくは暗示的に分類される。(14) の T と H は最終的に暗示的な弱対立であると判定される。

表 5: 弱対立関係認識の枠組み

意味関係 前件	意味関係 後件	条件の差		程度の差	確信度の差	例文 (T)	例文 (H)	弱対立
		(語彙)	(文脈)					
≡	≡, C, ↑, ↓	-	*	大, 無	P	コーヒーはかなり健康に良いかもしれない	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	-	*	小	*	コーヒーは少し健康に良い	コーヒーは健康に良い	暗
C	≡, ↑	強, 弱	強	無	*	コーヒーを1日3杯飲むならば, 健康に良い	コーヒーは健康に良い	明
	C, ↓	強, 弱	強	無	*	コーヒーを1日3杯飲むならば, 血糖値が下がる	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	強, 弱	強	大, 小	*	コーヒーを1日3杯飲むならば, 少し健康に良い	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	強	無	*	*	コーヒーを1日3杯飲むと健康に良い	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	弱	無	大, 無	P	美味しいコーヒーは健康に良いかもしれない	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	弱	無	小	*	美味しいコーヒーは少し健康に良い	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	強, 弱	弱	大, 無	P	コーヒーを1日3杯飲んだ場合でさえ, 健康に良いかもしれない	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	強, 弱	弱	小	*	コーヒーを1日3杯飲んだ場合でさえ, 少し健康に良い	コーヒーは健康に良い	暗
C	≡, C, ↑, ↓	-	*	大, 無	P	飲料は健康に良いかもしれない	コーヒーは健康に良い	暗
	≡, C, ↑, ↓	-	*	小	*	飲料は少し健康に良い	コーヒーは健康に良い	暗
∪	≡, ↑	-	強	無	*	温かい飲料ならば, 健康に良い	コーヒーは健康に良い	明
	≡, ↑	-	無	*	*	温かい飲料は健康に良い	コーヒーは健康に良い	暗
	≡, ↑	-	強, 弱	大, 小	*	温かい飲料ならば, かなり健康に良い	コーヒーは健康に良い	暗
	C, ↓	-	強, 無	*	*	温かい飲料は血糖値を下げる	コーヒーは健康に良い	暗
	C, ↓	-	弱	大, 無	P	温かい飲料でさえ, 血糖値を下げるかもしれない	コーヒーは健康に良い	暗
	C, ↓	-	弱	小	*	温かい飲料でさえ, 少し血糖値を下げる	コーヒーは健康に良い	暗

## 5 弱対立関係認識の枠組み評価実験

提案した枠組みを用いることで正しく弱対立関係が認識できるのかを評価する。まず、各要素の組み合わせによる導出結果が、二文間の関係が弱対立関係であるかを直接判断した結果と合致するかを調べる(実験1)。実験1において、弱対立関係を直接判断するためには、二文間の情報の差異を総合的に判断する必要があることから、付与者は、評価実験のデータ以外で十分な訓練を行ってから付与作業を実施する。次に、弱対立関係について詳しくない付与者であっても、提案する枠組みを用いれば、正しく弱対立関係を判定できるかを調べる(実験2)。実験1と比較して、文全体ではなく手がかりとなる要素のみに注視した場合でも、正しく弱対立関係の認識が行えるかを検証する。

### 5.1 実験設定

実験に利用するデータの元は、言論マップ [1] の基準で新しく構築された含意、矛盾、同意、対立、その他の5種類のラベルが付与された文対のうち、含意、矛盾、同意、対立のいずれかのラベルが付与された文対である。実験1では1000文対(データ1)、実験2では弱対立関係認識の難しさで2種類に分類したデータ(データ2、データ3)それぞれ75文対の合計150文対を対象とする。データ2の75文対は、含意、矛盾、同意、対立のいずれかのラベルが付与された文対のうち、明示的に条件または程度表現を含むものをランダムサンプルしたものである。データ3の75文対は、含意、矛盾、同意、対立のいずれかのラベルが付与された文対のうち、明示的な条件や程度表現を含まないものをランダムサンプルしたものである。二つのデータの大きな違いは、条件・程度表現が含まれるかであり、これが提案する枠組みの適用性能にどう影響するかを調査する。データ1についてはどちらの種類の日データも含まれている。

実験1は付与者A、実験2は二人の付与者A、Bによって行った。Aは、弱対立関係について十分詳しく、与えられた二文について弱対立かを判断することが可能である。Bは、弱対立の定義のみを知っており、本実験までに弱対立関係の付与作業をしていない。両付与者とも、提案する枠組みの各要素については十分

表 6: 提案する枠組みによる弱対立関係認識の評価

導出法	Precision	Recall	F1
以前の定義	0.66(217/327)	0.97(217/223)	0.79
提案する枠組み	0.82(187/230)	0.84(187/223)	0.83

に説明をするが、どのような組み合わせが弱対立になるかは伝えない。これは、提案する枠組みによる演算結果によって弱対立になるように意図的に操作することを防ぐためである。弱対立かの判断は、実験 1 は「1. 弱対立, 2. 弱対立関係ではない」実験 2 では「1. 弱対立, 2. どちらかと言えば弱対立, 3. どちらかと言えば弱対立ではない, 4. 弱対立ではない」の 4 段階で行う。提案する枠組みによって導出される明示的/暗示的な弱対立との対応は、1 と 2 が明示的/暗示的な弱対立に対応し、3 と 4 がそれ以外、すなわち弱対立関係にないことに対応する。弱対立関係に詳しくない付与者 B にとって、2 値で分類することが困難であることが想定されたため、実験 2 では 4 値で判断させた。

## 5.2 実験 1

付与者 A に、データ 1 に対して、二文間の関係が弱対立関係にあるかというラベル、大木らの定義の要となる前件、後件に付加情報の有無、枠組みの要素である表 5 の各要素を付与してもらった。直接二文間の関係が弱対立関係にあるか判断したラベルを正解として、大木らの弱対立関係定義、提案する枠組みから判断したラベルのそれぞれの結果の精度、再現率を表 6 に示す

提案する枠組みが以前の定義よりも F 値が高かったことから、枠組みを使用する方法は、弱対立関係を正確に認識できることを確認した。また、大木らの定義よりも高精度であることから、弱対立関係とならない事例の認識を減らすことに成功した。しかし、再現率は改善の余地があることから、本枠組みでとらえられていない要素が存在することが分かった。このような要素を明らかにするのが今後の課題の一つである。

表 7: 実験 2 の評価結果

対象データ	判定手法	Precision	Recall	F1
データ 2	直接判断	0.85(28/33)	0.51(28/55)	0.64
	枠組み	0.88(36/41)	0.65(36/55)	0.75
データ 3	直接判断	0.82(14/17)	0.45(14/31)	0.41
	枠組み	0.8(8/10)	0.26(8/31)	0.39

### 5.3 実験 2

弱対立関係について十分な付与訓練を行った付与者 A と行っていない付与者 B に、データ 2 とデータ 3 に二文間の関係が弱対立関係にあるかというラベルと、表 5 の各要素を付与してもらった。A が弱対立関係かを直接判断した結果を正解として、前者と、後者から導出された弱対立関係の精度、再現率を表 7 に示す。

データ 2 では、枠組みを利用することで精度および再現率が向上している。弱対立関係に詳しくない場合は、直接判断するよりも、枠組みを利用する方が正しく弱対立関係を認識できることが確認できた。一方で、データ 3 は、精度、再現率ともに悪化してしまった。データ 3 は、明示的に条件や程度といった表現が含まれていないことから、各要素を付与することが難しく、文全体を見て弱対立関係かを直接判断する方が容易であったと考えられる。条件や程度の差を判断するために、文中のこういった情報を見るべきかということを明確にすることが今後の課題の一つである。

### 5.4 エラー分析

以前の定義では誤って弱対立関係と認識していたが、本枠組みを使うことで弱対立関係ではないと認識できた事例を紹介する。

(15) H ウコンは花粉症に効く

T ちなみに、花粉症がヒドイ母親のために 2 年前から写真の「秋ウコン茶」を送っています

(15) は、A は弱対立関係ではないと判定した事例である。この H と T の前件はそれぞれ「ウコン」、「秋ウコン茶」となる。T の前件は H の前件に条件が付加されたものであるため、以前の定義では弱対立関係と認識される。しかし、T の文からは「ウコン」が花粉症に効くために「秋ウコン茶」を送っていると読み取れるため、弱対立関係とはならない。枠組みを使用した方法では、T の前件に条件が付加される前「ウコン」に対しても「花粉症に効く」ということが推論されるため、条件の差が小さくなり、正しく弱対立関係ではないと認識できる。

次に、本枠組みで弱対立関係の判定を誤ってしまう事例を紹介し、エラー分析を行う。

(16) H 整骨院の電気治療は効果がない

T 電気治療の場合は人によっては効果があまり実感できないこともあるよ  
うなので

(16) は A は弱対立関係であると判定したが、枠組みを利用した場合には弱対立と判定できなかった事例の一つである。T の「人によっては」が条件の差を生じる部分であるが、要素ごとに詳細に付与していく際には条件の差が小さいと判断されてしまった。条件に差があるとはどういうことなのかを、より明確にする必要がある。

(17) H 花粉症は食生活の改善で治る

T 花粉症の症状を緩和するためにも、アレルギー体質になりやすい食生活を避けるようにしましょう

(17) は (16) と同様に A は弱対立関係と判定したが、枠組みからは弱対立関係と判定できなかった事例である。この H と T の後件がそれぞれ「花粉症が治る」、「花粉症の症状が緩和する」となる。後件間の意味関係は同意の関係に近いが、厳密に考えると NL の 7 つの意味関係のどれに該当するかの判定が難しく、人の捉え方によって、 $[=]$ 、 $[<]$ 、 $[||]$  の 3 種類に分かれるものと思われる。A は  $[<]$  と付与したため、弱対立関係ではないと誤って判定された。このように意味関係を厳密に考えると、判定が難しく弱対立関係とは判定されない事例が存在する。

(18) H 偽薬は病気の治療に効果がある

T 値段の高い偽薬は低価格の偽薬より効果がある

次に、(18)は、弱対立関係ではないが、枠組みからは弱対立関係と導出された事例の一例である。Tは、「値段の高い偽薬」と「低価格の偽薬」とを比較しており、文全体を見ると両方とも「効果がある」ことが読み取れることから弱対立関係ではない。しかし、要素ごとに詳細に付与していくと、「値段の高い」や「低価格」によって条件に差があると判断されるため、弱対立関係となってしまう。比較文中の条件は、その文中での比較であり、文間では差がないことがあるため、枠組みの改善が必要であることが分かった。

## 6 弱対立関係認識器の構築

第4章で提案した枠組みを用いた弱対立関係認識器について述べる。

### 6.1 概要

(仮説, テキスト)の二文からなる文対を入力とし, テキストが仮説に対して弱対立関係にあるかないかを判定する。入力 of 仮説は単文であると想定しており, 複雑な文には対応できない。弱対立関係認識器の概要は図5のようになり, 手順を下記に示す。

1. 仮説の前件, 後件に対応するテキストの箇所を同定する
2. テキスト側の前件に付加される条件の有無を調べ, 含む場合は条件の差を3値で分類する。
3. テキスト側の後件に付加される程度表現の有無を調べ, 含む場合は程度の差を3値で分類する。
4. テキスト側の後件の述語の確信度により, 確信度の差を2値で分類する。
5. 2, 3, 4で付与された構成要素を第4章で提案した枠組みに適用し, 弱対立関係にあるか判定する。

仮説を前件, 後件に分割し, 水野らの局所構造アライメント解析の手法 [1] を用いて, 前件, 後件それぞれに対応するテキストの箇所を同定する。なお, このとき文間の意味関係を [関連あり] と [その他] に分類される。[その他] に分類された文対は, 二文間で関連性がないため, 弱対立関係にはならないとみなし, この時点で弱対立関係ではないと判定する。アライメントされた仮説とテキストの文節間の意味関係は, 類似度が高いと予測されるので, 後件の意味関係は [≡] または [^], 前件の意味関係は付加条件を含む場合は [□], 含まない場合は [≡] とする。本来ならば, 上記以外の意味関係においても弱対立関係となるパターンが存在す

るが、そのような文対を集めることが困難なため、今回はこのような仕様となっている。

次に、仮説側の前件、後件に対応付けられたテキスト側の箇所に対して、構成要素である [条件の差]、[程度の差]、[確信度の差] を調べ、それぞれ枠組みに合う形式に分類する。求めた構成要素の組み合わせが枠組みに存在する場合は弱対立関係と判定し、存在しない場合は弱対立関係ではないと判定する。なお、弱対立関係は [明示的]、[暗示的] の 2 値で分類する。次の節からは構成要素を分類する手法に関して述べる。

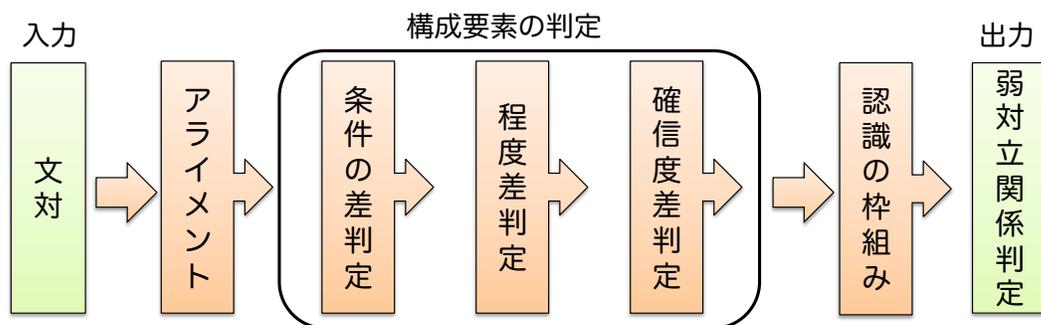


図 5: 弱対立関係認識器の概要

## 6.2 条件の差

テキスト側の前件と後件に対する条件の有無を、大西らの機械学習による限定関係認識手法 [14] にいくつかの変更点を加えた手法を用いて調べる。枠組みでは条件の差は [差が大きい]、[差が小さい]、[差が無い] の 3 値で分類するが、今回の手法では前件、後件に [差が大きい条件がある] と [差が大きい条件が無い] の 2 値分類を行う。すなわち、[差が大きい] と [差が小さいあるいは差が無い] の 2 値に分類する。次から、本手法の変更点である条件の探索範囲の拡大と機械学習に用いる素性について説明を行う。

### 6.2.1 条件の探索範囲の拡大

大西らの手法では、条件の探索範囲は前件/後件の文節に対して、係り元の文節が係り先である前件/後件の文節に対して、条件であるか否かを判定している。T1では前件「コーヒー」の係り元である「インスタントの」が条件となっているため、認識することが可能である。しかし、(19)のT2では前件「コーヒー」に対しての条件である「1日1杯飲む」が係り元ではなく前件の直後に位置しているため、条件として認識することができない。

(19) H コーヒーは健康に良い

T1 インスタントのコーヒーは健康に悪い

T2 健康に良いのは、コーヒーを1日1杯飲むことだ

この問題を解決するために条件の探索範囲の拡大を行う。(19)でもわかるように、条件となる表現は制限する対象の前後に位置することが多いと予測できる。この予測に従って、前件/後件の係り元の文節に加えて、直前と直後の文節も条件の探索範囲に加える。探索範囲の拡大によって、(19)のT2の条件「1日1杯飲む」が認識できるように既存の手法では認識できない範囲に存在する条件も認識され、網羅率が上がることが期待できる。

### 6.2.2 使用する素性

機械学習に用いる素性として、大西らの手法 [14] での素性を基に新たな素性を加えるなどした表8を用いる。ここで前件/後件を対象、前件/後件の条件か否かを調べる範囲を候補と表記する。新素性の列には、大西らの手法 [14] の素性は [既]、新たに追加した素性は [新] と表記する。次にいくつかの新たに追加した素性の内容や追加した狙いについて説明を行う。

対象と候補の関連性を測る素性としてPPMI(正の相互情報量)を追加した。PPMIは次の式で表される。

$$PPMI(x, y) = \log \frac{P(x, y)}{P(x)P(y)} \quad (4)$$

PPMI は関連性が高い語句同士だと高い値を示し，関連性が低い語句同士だと 0 に近い値を示す．

(20) H コーヒーは健康に良い

T1 コーヒーを 飲むと 健康に良い

T2 インスタントの コーヒーは健康に良い

(20) は T1, T2 とも付加情報が含まれているが，H と T2 が弱対立関係となり，H と T1 は弱対立関係とはならない事例である．T1 は付加情報「飲むと」が含まれているが， $PPMI(\text{コーヒー}, \text{飲むと}) = 0.55$  と低い値を示す．一方，T2 では付加情報「インスタントの」を含んでおり， $PPMI(\text{コーヒー}, \text{インスタント}) = 2.37$  と高い値を示す．このように，条件の差が大きい場合には関連性が高くなり PPMI も高い値を示し，逆に条件の差が小さい場合には関連性が低くなり PPMI も低い値を示す．この素性の追加により，対象と候補の関連性を捉えることにより，条件の差が小さい事例を正しく認識することを期待する．

候補が条件になりやすい依存構造や条件とはなりにくい依存構造があると予測した．そこで，候補の前後，または係り受け前後に前件/後件があるかを調べる素性を追加した．

(21) H コーヒーは健康に良い

T1 コーヒーを 常に飲むと 健康に悪い

T2 コーヒーを 常に飲むと，カフェインの取りすぎで健康に悪い

(21) は T1, T2 共に付加情報「常に飲むと」が含まれている H と弱対立関係になる文である．T1 の付加情報「常に飲むと」の前後には，前件「コーヒー」と「健康に悪い」が位置しており，T2 の付加情報「常に飲むと」の係り先には後件「健康に悪い」が位置している．このことから条件の周辺には前件/後件が位置しやすいことが予測される．

表 8: 使用する素性

素性	素性の入れかた	新素性
対象が前件であるか	2 値	新
対象が後件であるか	2 値	新
候補が対象より前にあるか	2 値	新
候補が対象より後ろにあるか	2 値	新
候補の係り先が対象であるか	2 値	新
対象の係り先が候補であるか	2 値	新
次の単語のいずれかが候補に含まれているか (として, とした, では, には, れば, にも, でも での, ときは, による, により, によって, から, しか, 的に)	単語ごとに 2 値	既
次の単語が候補の最後か (に, と, で, も, の, は, ば)	単語ごとに 2 値	新
次の単語が対象の最後か (に, と, で, も, の, は, ば)	単語ごとに 2 値	既
対象に条件名詞 (とき, 時, 限り, かぎり, 場合, ばあい) が含まれているか	2 値	既
候補の最初の品詞	品詞ごとに 2 値	新
対象の最初の品詞	品詞ごとに 2 値	新
候補の最後の品詞	品詞ごとに 2 値	新
対象の最後の品詞	品詞ごとに 2 値	新
対象に限定名詞が含まれているか	2 値	既
対象に限定副詞が含まれているか	2 値	既
対象に限定助詞が含まれているか	2 値	既
対象に限定名詞, 限定副詞, 限定助詞のいずれかが含まれているか	2 値	既
候補の前後に前件/後件があるか	前件/後件と前後の 組み合わせごとに 2 値	新
候補の前後に前件もしくは後件があるか	前, 後ごとに 2 値	新
対象の前後に前件/後件があるか	前件/後件と前後の 組み合わせそれぞれで 2 値	新
対象の前後に前件もしくは後件があるか	前, 後ごとに 2 値	新
候補の係り受けパス前後に前件/後件があるか	前件/後件と前後の 組み合わせごとに 2 値	新
候補の係り受けパス前後に前件もしくは後件があるか	前, 後ごとに 2 値	新
対象の係り受けパス前後に前件/後件があるか	前件/後件と前後の 組み合わせごとに 2 値	新
対象の係り受けパス前後に前件もしくは後件があるか	前, 後ごとに 2 値	新
対象と候補に含まれる単語間の PPMI の最大値	実数	新
対象と候補に含まれる単語間の PPMI の最小値	実数	新

### 6.3 程度の差

テキストの程度表現による程度量と仮説の程度量を比較して，[大きい]，[小さい]，[差が無い]の3値に分類する．程度表現の有無は，川端らの程度評価用語・尺度感覚を表す語彙群 [17] を元に，程度表現が表す量・尺度に基づき3値に分類した程度表現辞書を用いる．「たいへん」や「特に」など程度が大きいことやその表現を強調する程度表現には「3」，「わずかに」や「やや」など程度が小さいことやその表現を控えめにする程度表現に「1」，「まあまあ」や「まずまず」など程度が一般的な様を表す程度表現には「2」とした．仮説とテキストの後件と，前件を除いた後件の係り元の文節に対して，程度表現辞書にマッチした程度量の値の平均をとり，仮説の値よりテキストが大きいときには「大きい」，反対に小さい時は「小さい」，また，同じ値のときは「差が無い」とする．また，程度表現が存在しないときは程度量を「2」としている．

(22) H コーヒーは健康に良い

T コーヒーはわずかに健康に悪い

(22) を例に程度の差を測定する流れを説明する．Hの後件は「健康に良い」であり，係り元の文節は前件を除いて存在しない．Hの後件には程度表現が含まれていないので，Hの程度量は「2」となる．次にTの後件は「健康に悪い」であり，係り元の文節に「わずかに」が存在する．「わずか」が程度表現辞書にマッチし，程度量「1」が得られる．この他に程度表現が含まれていないので，Tの程度量は「1」となる．Tの程度量が「1」とHの程度量「2」より小さいため，程度の差は「小さい」と分類される．

### 6.4 確信度の差

仮説とテキストの述語に対する著者の確信度の差を [C]，[P] の2値で分類する．仮説とテキストの述語に対する著者の確信度をモダリティ解析により得られる事実性から判断する．事実性は [成立]，[高確率]，[0]，[低確率]，[不成立] の5値に分類される．[成立]，[不成立] は後件の言及を断定しているため確信度が高く，[高

確率]，[低確率]は推量や伝聞を含むため確信度が低いと考えることができる．一方，[0]は詳細不明のため扱うことができないが，一般的に二文間の確信度の差はないと推測されるため，仮説とテキストどちらかが[0]の場合は確信度の差は[C]とする．仮説の事実性が[成立]，テキストの事実性が[成立]の場合などの確信度の差が無いもしくは，仮説の事実性が[高確率]，テキストの事実性が[成立]などテキストの確信度の方が高い場合は[C]と分類する．一方，仮説の事実性が[成立]，テキストの事実性が[高確率]など仮説の事実性の方が高い場合は[P]と分類する．なお，確信度の差は事象の成立/不成立に対する推量，伝聞の差であり，成立/不成立かどうかは考慮しない．例えば，[高確率]と[低確率]は前者は成立，後者は不成立に対するものだが，それぞれの言及に対する推量，伝聞の差はないため，確信度の差は現れない．

## 7 弱対立関係認識器の評価実験

構築した弱対立関係認識器の性能を評価する．弱対立関係の認識にはいくつかの手順が存在するが，2つの実験で構成要素の判定の部分の評価を行う．1つ目の実験では，条件の差の認識の評価を行う．これは，程度表現や確信度の差がなく条件が付加されただけのものが弱対立関係となる文対の大半を占めており，弱対立関係の認識で一番重要な手順であるためである．また，程度の差，確信度の差は既存手法のものを流用しただけのものであるため本研究では評価を行わない．2つ目の実験では，弱対立関係認識全体の評価を行う．アライメントは，難しいタスクであり，認識器に用いている手法でも決して高い精度ではなく，エラーが集中する恐れがあることや本研究で新たに行った手法でもないため，人手で作成した正解データを与える．認識の枠組みの評価を第5章で評価したので，結果的にこの章では構成要素の判定の評価を行う．

### 7.1 条件の差の認識の評価実験

#### 7.1.1 実験設定

(仮説，テキスト)の文対と人手で作成した前件/後件の対応させる文節アライメントのデータを入力とする．テキスト側の前件/後件に対して，各手法の条件の探索範囲の場所にある文節が [条件の差が大きい条件] か [条件の差が小さい条件，もしくは条件では無い] の2値で分類する．条件の認識の手法は，大西らの機械学習による限定関係認識手法 [14] をベースラインに第6.2章で記述した手法の比較を行う．

機械学習に用いる学習データは，医療・健康ドメインのアライメントされた文節とその文節の条件かを調べる文節間において，条件となる文節ペアを正例，条件ではない文節ペアを負例と付与した全454文対を用いた．学習器は，多項式カーネルのSVMを用いて，パラメータは次元数を2，コストパラメータを1000とし，他のパラメータはデフォルト値と設定した．評価データは学習データと同様のドメインの弱対立関係の文対から，各クエリからランダムで5~10文対サンプリングした合計119文対を使用した．評価データには合計83個の条件が含まれている．

### 7.1.2 実験結果

学習器の出力の閾値を変更し，P-R 曲線を描いたところ図 6 になった．提案手法は大西らの手法に比べて，再現率が 0.15 付近の精度では劣っているが，おおよそ精度，再現率ともに上昇した．条件の探索範囲の拡大により，既存手法より再現率が上昇し，また，係り受け関係にある文節が前後の文節よりも条件になる場合が多いために，既存手法より精度が低い箇所が現れたと思われる．しかし，依然として精度，再現率ともに有効な数字とは言いがたいものである．

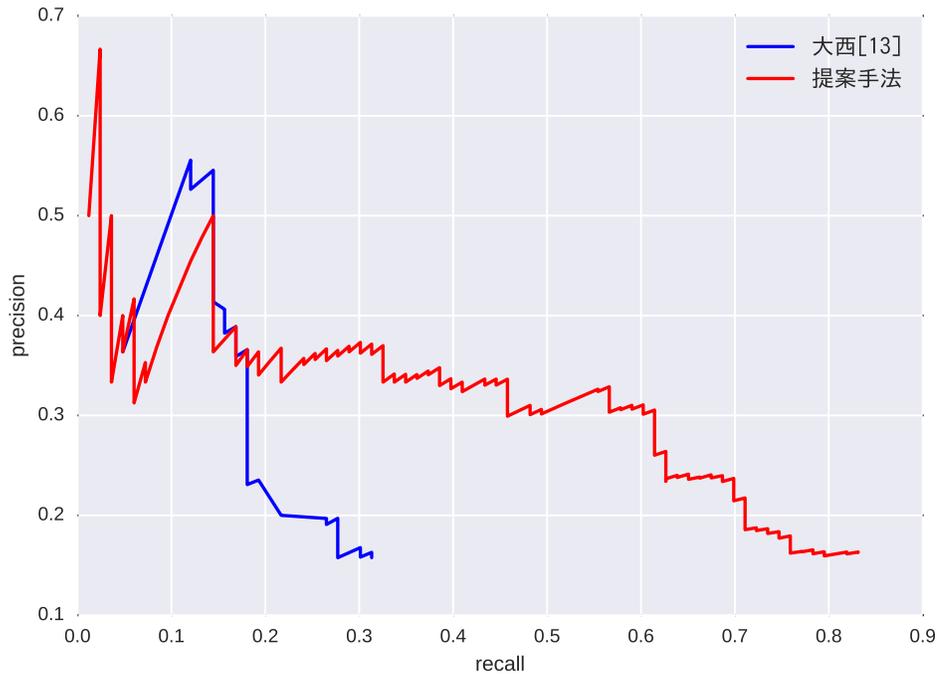


図 6: 条件認識実験の評価結果：P-R 曲線

### 7.1.3 エラー分析

ここではいくつかの代表的な条件認識のエラーを紹介し，どのような課題が残されているのかを明確にする．

(23) H コーヒーは健康に悪い

T1 コーヒーは 飲むと 健康に悪い

T2 コーヒーは 飲みすぎると 健康に悪い

(23) は H と T1 は弱対立関係ではなく、H と T2 が弱対立関係となる例である。T1 は H と比べて「飲むと」が付加されているが、「コーヒー」に対する制限が小さいため、条件の差が小さい。一方、T2 は「飲みすぎると」が付加され、「コーヒー」に対する制限が大きく、条件の差が大きくなる。「飲むと」と「飲みすぎると」を比較すると、同じ動詞「飲む」が含まれており、表層の違いは少ない。しかし、表層の違いは少なくとも条件の差に違いが生まれる。このように表層の小さな違いにも、条件の差が生じてしまい、認識するのが困難となる。

(24) H コーヒーは健康に悪い

T1 カフェインを含む コーヒーは健康に悪い

T2 乳成分を含む コーヒーは健康に悪い

(24) は H と T1 は弱対立関係ではなく、H と T2 が弱対立関係となる例である。T1 は「コーヒー」に対して「カフェインを含む」が付加されているが、あらゆるコーヒーはカフェインを含んでおり、「コーヒー」に対する説明をしている表現であり、条件とはならない。一方、T2 は「コーヒー」に対して「乳成分を含む」が付加されており、T1 とは異なり、あらゆるコーヒーに乳成分が含まれているわけではなく、「コーヒー」に対して限定する表現であり、条件となる。このように、ある事柄に対して説明している場合とある事柄に対して制限する条件である場合が存在する。付加情報が、ある事柄に対する説明か、条件であるかを区別するためには、その事柄と付加情報間の知識が必要となる。

(25) H コーヒーは健康に悪い

T1 コーヒーを 多く飲むのは 健康のためだ

T2 コーヒーを 多く飲むのは 健康に悪い

(25) は、T1 と T2 はどちらも「多く飲むのは」が付加されているが、H と T1 は弱対立関係ではなく、H と T2 が弱対立関係となる。T1 は、後件が「健康のためだ」であることから、書き手が健康を理由にコーヒーを多く飲んでいることがわかる。これは「コーヒーが健康に良い」との前提があり、「多く飲む」は条件にならない。T2 は、後件が「健康に悪い」であることから、「多く飲む」は条件となる。このように、付加情報が同じ表現でも後件のわずかな違いにより、条件になる場合とならない場合が存在する。このことから、条件認識には文脈を考慮する必要があると言える。

これまでで述べた例のように、条件認識には付加される対象と付加情報の表層だけでは解けないことがわかる。また、付加された前件/後件だけでなく、その他の前件/後件や文脈も考慮する必要がある。現状の手法では、表層しか捉えられてなく、文脈などが捉えられてはいない。今後は表層だけではなく、分散表現などの深い言語情報が必要とし、文脈などの情報も捉えなければならない。

## 7.2 弱対立関係認識の評価実験

### 7.2.1 実験設定

実験 7.1 と同様に (仮説, テキスト) の文対と人手で作成した文節アライメントのデータを入力とする。なお、どの文節アライメントを意味解析に使うのかは明記されていないため、述語の選定を水野らの手法 [1] で行うが、文節アライメントは正解データを与えているため高確率で正しく判定できていると推測される。次に第 4 章で述べた手法で構成要素の判定を行い、弱対立関係認識の枠組みにあてはめることで最終的に弱対立関係か否かの判定を行う。

手法は以下の 3 つを用いて比較する。1 は条件認識と弱対立関係判定を既存のものを利用し、2 では弱対立関係判定のみ本研究の枠組みを使用して行う、3 では条件認識も第 6.2 章の提案手法を用いる。

1. 大西らの手法による条件認識 (既存手法) + 付加条件の有無による弱対立関係判定 (既存手法)

2. 大西らの手法による条件認識 (既存手法) + 枠組みを使用した弱対立関係判定 (提案手法)

3. 提案手法による条件認識 + 枠組みを使用した弱対立関係判定 (提案手法)

条件の差の判定の機械学習は実験 7.1 と同様の設定で行う。評価データは学習データと同様のドメインの 656 文対を使用した。評価データは、文節アライメントが人手で付与され、かつ人手で二分間の意味関係が同意・対立と判断されるものである。評価データのうち 155 文対が弱対立関係にある。

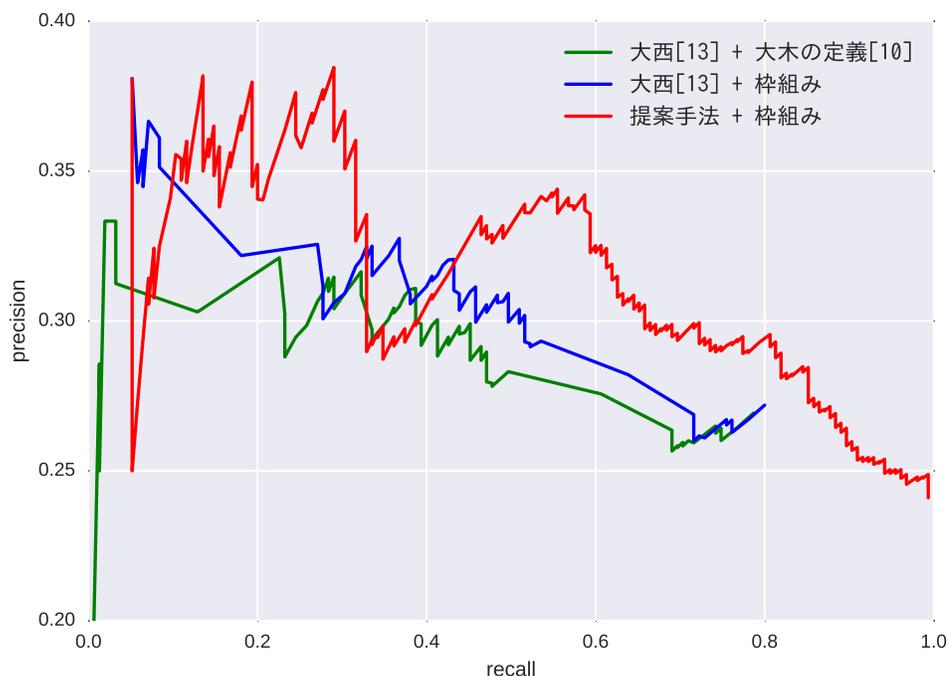


図 7: 弱対立関係認識実験の評価結果：P-R 曲線

### 7.2.2 実験結果・考察

学習器の出力の閾値を変更し、P-R 曲線を描いたところ図 6 になった。条件認識を提案手法、弱対立関係を枠組みにより認識する方法が一番高い性能を示した。

条件認識が既存手法より改善されたことに加え，枠組みを使用することで [程度の差]，[確信度の差] がある事例を認識でき，誤った認識も減ったことにより性能が上がったものと思われる．また，条件認識を既存手法を用いて，枠組みを使用した方法が，大木らの定義で認識した方法より高い性能を示したことから，枠組みの優位性が確認できた．どの手法も，精度が低い値を示しており，条件認識の精度の低さが影響していると思われ，条件認識の性能を上げることが今後の課題である．実用的には，再現率はそこそこでも精度が高い方が良いため，提案手法の精度 0.4，再現率 0.3 のポイントが使える．

## 8 おわりに

本論文では，弱対立関係の構成を考え，どのような現象が起こっているのかを分析することで，認識したい事例を明確にした．弱対立関係とは条件つきで同意・対立関係にある，もしくは部分的に同意・対立している関係である．詳しく述べると，条件が付くことで，条件が成立の場合には同意(対立)し，条件が不成立の場合には対立(同意)することを示唆する．または後件に程度表現を含むことで，効果や範囲の程度においては完全に同意しないことを意味する．

上記で認識したい事例に対して，認識の手がかりとなる要素に「条件の差」，「程度の差」，「確信度の差」を提案した．「条件の差」は条件が付加される前後で後件の成立・不成立の変化の大きさ，「程度の差」は二文間の後件の効果・範囲に関する程度量の差，「確信度の差」は二文間の後件に対する著者の確信度の差のことをそれぞれ表す．この3要素と前件，後件それぞれの意味関係の組み合わせが，弱対立関係と認識されるパターンをまとめた認識の枠組みを提案した．

既存の手法を組み合わせることで弱対立関係認識器を構築した．認識の手順は，二文間のアライメントを行い，前件，後件を対応させる．次に，手がかりとなる要素である「条件の差」，「程度の差」，「確信度の差」の判定し，最後に認識の枠組みに適応することで弱対立関係の判定する．しかし，この認識器は精度，再現率共に低い結果となった．これは，条件を表層だけの情報だけで認識するのが難しいためである．今後はより深い言語情報を捉えることで，条件認識の性能を上げ，認識器全体の性能を上げることが必要となる．

## 謝辞

本研究を進めるにあたり，多くの方々からご協力，助言を頂きましたこと心より感謝申し上げます．主指導教員である乾健太郎教授には，ご多忙の中，研究活動全般にわたりご指導，助言を頂きましたこと，心より感謝申し上げます．水野淳太氏には，研究テーマ決めで迷っているところに，温かく相談に乗って頂き，研究活動でつまづいたときにも，適切な助言や温かく見守って頂き大変お世話になりましたこと心から感謝申し上げます．稲田和明には，研究活動全般で適切な助言を多くして頂き，研究を進める上で大きな助けになりましたこと心から感謝申し上げます．お忙しい中，審査委員をお引き受け下さいました木下哲男教授，篠原歩教授に深く感謝いたします．最後になりましたが，日々の議論を通して，研究のアドバイスをしていただき，また研究生活を支えていただきました研究室の皆様感謝申し上げます．

## 参考文献

- [1] 水野淳太, 渡邊陽太郎, エリックニコルズ, 村上, 乾健太郎, 松本裕治. 文間関係認識に基づく賛成・反対意見の俯瞰. *情報処理学会論文誌*, Vol. 52, No. 12, pp. 3408–3422, 2011.
- [2] Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proceedings of the First PASCAL Machine Learning Challenges Workshop*, pp. 177–190, 2005.
- [3] Danilo Giampiccolo, Bernardo Magnini, Ido Dagan, and Bill Dolan. The Third PASCAL Recognizing Textual Entailment Challenge. In *Proc. of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pp. 1–9, 2007.
- [4] Suguru Matsuyoshi, Yusuke Miyao, Tomohide Shibata, Chuan-Jie Lin, Cheng-Wei Shih, Yotaro Watanabe, and Teruko Mitamura. Overview of the ntcir-11 recognizing inference in text and validation (rite-val) task. In *Proceedings of the 11th NTCIR Conference*, 2014.
- [5] 大木環美, 村上浩司, 松吉俊, 水野淳太, 乾健太郎, 松本裕治. 文間の弱い対立関係の認識. *情報処理学会研究報告. 自然言語処理研究会報告*, pp. 1–9, 2010.
- [6] Megumi Ohki, Eric Nichols, Suguru Matsuyoshi, Koji Murakami, Junta Mizuno, Masuda Shouko, Kentaro Inui, and Yuji Matsumoto. Recognizing confinement in web texts. In *Proceedings of The 9th International Conference on Computational Semantics (IWCS2011)*, pp. 215–224, 2011.
- [7] Myroslava O Dzikovska, Rodney D Nielsen, Chris Brew, Claudia Leacock, Danilo Giampiccolo, Luisa Bentivogli, Peter Clark, Ido Dagan, and Hoa T Dang. Semeval-2013 task 7: The joint student response analysis and 8th recognizing textual entailment challenge. Technical report, DTIC Document, 2013.

- [8] Rodney D Nielsen, Wayne Ward, and James H Martin. Recognizing entailment in intelligent tutoring systems. *Natural Language Engineering*, Vol. 15, No. 04, pp. 479–501, 2009.
- [9] Omer Levy, Torsten Zesch, Ido Dagan, and Iryna Gurevych. Recognizing partial textual entailment. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 451–455, 2013.
- [10] Yi Zhou and Yan Zhang. A logical study of partial entailment. *Journal of Artificial Intelligence Research*, Vol. 40, p. 25, 2011.
- [11] Bill MacCartney. *Natural language inference*. PhD thesis, Stanford University, 2009.
- [12] Bill MacCartney and Christopher D Manning. Natural logic for textual inference. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pp. 193–200. Association for Computational Linguistics, 2007.
- [13] 大木環美. 文間の弱対立関係の定義と認識. 2011.
- [14] 大西真輝, 水野淳太, 福原裕一, 渡邊陽太郎, 乾健太郎. 文節限定関係認識に基づく文間弱対立関係認識. 言語処理学会第 19 回年次大会発表論文集, pp. 948–951, 2013.
- [15] 日本語記述文法研究会. 現代日本語文法: 第 9 部. とりたて; 第 10 部. 主題, 第 5 巻. くろしお出版, 2009.
- [16] 江口萌, 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治. モダリティ, 真偽情報, 価値情報を統合した拡張モダリティ解析. 言語処理学会第 16 回年次大会発表論文集, pp. 852–855, 2010.
- [17] 川端元子. 程度副詞を分類する視点の考察. 2012.

## 発表文献一覧

### 国内会議・研究会論文

大江貴裕, 水野淳太, 稲田和明, 乾健太郎. 文間弱対立関係認識のための Natural Logic の拡張. 人工知能学会 全国大会, June 2015.