

日本語事実性解析課題の経験的分析

成田 和弥[†] 水野 淳太[‡] 乾 健太郎[†]

[†] 東北大学 [‡] 奈良先端科学技術大学院大学

{narita, junta-m, inui}@ecei.tohoku.ac.jp

1 はじめに

事実性とは、文中のある事象が実際に起こったことなのか、あるいは起こる可能性を述べただけなのか、に関する情報である。本稿では、事象は行為、出来事、状態の総称であると考えられる [11]。例えば、(1a) の事象「出る」は実際に起こったことである。(1b) の事象「帰る」は「だろう」という表現により、起こった可能性が高いと判断できる。(1c) の事象「発生する」は「防いだ」という表現によって、実際には起こっていないことだとわかる。

- (1) a. 彼はさきほど部屋を出た。
- b. もう遅いから、彼は先に帰ったんだろう。
- c. 問題が発生するのを防いだ。

事象に対する事実性の解析は、情報抽出や含意関係認識などに有用であるが、いまだ十分な解析性能は実現されていない。その原因の一つに、内在する問題の分析・整理が不十分であることがあげられる。

本研究の最終的な目的は、高精度な日本語事実性解析器の実現である。そのためには、まず問題の分析を行う必要がある。本稿では、パターンベースの日本語事実性解析器を構築し、内在する問題の分析を行った。

具体的には、Sauri(2008)[4] の英語を対象とした事実性解析モデルをもとに、日本語事実性解析器を構築し、その誤り分析を行った。Sauriの解析モデルは、事実性に影響を与える表現を手がかりとし、それらの組み合わせにて事実性解析を行う。このモデルは、辞書情報の組み合わせによって事実性を解析できる構成性をもち、事実性を分割して個別に分析を行えるといった特長がある。そのため、内在する問題の分析という本研究の目的に適している。分析の結果、事実性解析における問題、特にリソースを利用する上で何が問題となるのかを報告する。

2 関連研究

事実性およびその周辺情報をマークアップするための体系および付与基準の関連研究として、Sauriらによる FactBank[5, 7] や松吉ら (2010)[11]、川添ら (2011)[14, 15] などがある。Sauriらは、事象を対象とし、事実らしさに対する modality(certainty) とその確信の方向を表す polarity の組によって事実性を定義した。そして、事象とその時制、肯否、モダリティが付与された TimeML[6] に対して、確信度と肯否極性を態度表明者 (source) ごとに付与する枠組みを提案し、FactBank と呼ばれるコーパスを構築した。本研究では modality(certainty) を確信度、polarity を肯否極性と呼ぶ。松吉らは、態度表明者、相対時、仮想、態度、真偽判断、価値判断の 6 項目からなる拡張モダリティタグ体系を設計し、それを付与したコーパスを構築した。彼らは拡張モダリティの項目の組み合わせによって事実性を表現している。川添らは、様相表現・否定表現・条件表現を「確実性」の側面から分析・分類し、それに従ったアノテーションスキーマの作成を行った。近年このような研究は、上で述べた純粋な自然言語処理の分野ばかりでなく、生物医学分野においても行われている [2, 3, 8]。

解析手法としては、機械学習に基づく手法 [1, 2, 3, 9, 10, 13] や、人手で構築した語彙的・統語的知識に基づくパターンベースの手法 [2, 4] があげられる。例えば江口ら (2010)[10] は、拡張モダリティにおける項目間および事象間の依存関係を考慮できる条件付確率場を利用した解析システムを提案した。松吉ら (2011)[13] は、最大エントロピーモデルを用いた拡張モダリティ解析システムを試作し、その中の 1 つのタグである「態度」にのみ注目して誤り分析を行った。

Sauri(2008)[4] は、事象の成立に影響を与える手がかり表現を利用し、確信度と肯否極性で表される事実性を、依存構造木の根から伝搬させて解析するアルゴ

リズムを提案した。このモデルは、手がかり表現の情報を持つ辞書が与える事実性を組み合わせ、事象の事実性解析を行うという構成性をもつため、どのような解析過程でその結果が出力されたのかが明確である。また、事実性を確信度と肯否極性という2軸に分割するため、個別に分析することができる。

3 日本語事実性解析器

日本語事実性解析における問題点を分析・整理するため、Sauri(2008)[4]の英語を対象とした事実性解析モデルをもとに、日本語に対する解析器を構築した。本節では、3.1で事実性の定義、3.2で使用するリソース、3.3で解析モデルについて述べる。

3.1 事実性の定義

Sauriは確信度を Certain(CT)・Probable(PR)・Possible(PS)・Underspecified(U)の4段階、肯否極性を positive(+)・negative(-)・underspecified(u)の3値として扱い、それらの組によって事実性を定義している。例えば(1a)の事象「出る」は著者から見て実際に起こったことであり、その事実性はCT+と表される。同様に(1b)における事象「帰る」は起こった可能性が高いことであり、その事実性はPR+、(1c)における事象「発生する」は実際には起こっていないことであり、その事実性はCT-と表される。

英語では、PRは *probable*、PSは *possible* といった表現によって判断しているが、日本語では表現が多彩であるため、PRとPSの境界が曖昧で、その区別は容易でないことが予想される。例えば(2)において、事象「着く」の確信度がは明らかにCTでもUでもないが、PRであるのかPSであるのかは判断できない。

(2) 彼が無事に着いたと信じている。

そこで我々は、PRとPSを1つの値PRとしてまとめ、それ以外はSauriの定義を利用した。即ち、確信度を Certain(CT)・Probable(PR)・Underspecified(U)の3段階、肯否極性を positive(+)・negative(-)・underspecified(u)の3値として扱い、これらの組み合わせによって事実性を表す。

Sauriは上記に加え、著者以外の態度表明者 (source) に対する事実性も考慮しているが、タスクを簡略化し分析をしやすくするため、今回は著者の事実性判断を認識する問題に限定した。

表 1: 機能表現辞書の意味クラス例と事実性への影響

機能表現辞書中の意味クラス	該当する機能表現	事実性への影響
否定	ない ぬ わけでない	肯否極性: + → -, - → +
推量	だろう かもしれない らしい	確信度: CT → PR
疑問	か かな ではないか	U _u への更新

表 2: 手がかり表現辞書の見出し語例と対応する事実性

見出し語	下位事象の事象形式	上位事象の肯否環境	上位事象に対する真偽判断	上位事象に対する事実性
防ぐ	スル形	肯定 否定	不成立 成立	CT- CT+
	タ形	肯定 否定	- -	- -
忘れる	スル形	肯定 否定	不成立 成立	CT- CT+
	タ形	肯定 否定	成立 成立	CT+ CT+

3.2 使用するリソース

Sauriのモデルは、確信度と肯否極性の組で表される事実性を、factuality marker と呼ばれる事実性に影響を与える表現を利用して解析する。例えば *not* は肯否極性を反転させる、*may* は確信度を下げる factuality marker である。機能語だけでなく述語についても、例えば *know that* という表現は *that* 節を前提とした発言であるので、*know* は *that* 節内の事象をCT+とする factuality marker と考えることができる。これらのような語彙的な marker だけではなく、目的節や関係詞節等の統語的な marker も考慮している。

日本語においても「～ない」は肯否極性の反転、「～だろう」は確信度の減少、というように同等な表現が存在する。このような表現を集めたリソースとして、機能表現辞書 [12] および手がかり表現辞書 [10] を利用した。

機能表現辞書は、主に文の構成に関わる要素である機能表現を意味・文法的機能・活用などに応じて階層的なクラスに分類し、収録した辞書である。機能表現辞書の意味クラスおよびその表現の例、それらの事実性に対する影響を表1に示す。この辞書は意味クラスによって表現を分類しており、それぞれの表現に対して活用や表記の異なりなどを考慮し構成されている。この中で「否定」と分類される表現は肯否極性を反転、「推量」と分類される表現は確信度を減少、「疑問」

と分類される表現は U_u を割当，といった役割を持つ factuality marker として利用できる．このような事実性に影響を与える手がかり表現として，機能表現辞書の一部である 5,345 表現を利用した．ここでは「否定」「推量」「疑問」「意志」「願望」などの事実性に影響を与える意味クラスに基づいて表現を選出した．

手がかり表現辞書は，拡張モダリティに影響を与える 3,692 述語を収録しているが，これは事実性とも大きな関わりをもっている．手がかり表現辞書の項目の例を表 2 に示す（実際の辞書には態度・真偽判断・価値判断の 3 つ組が記述されているが，表では事実性と関わりの深い真偽判断に関する項目のみ示した）．この辞書は，各述語が直前の事象に与える影響を，直前の事象の時制（未来：スル形「～する」・非未来：タ形「～した/していた/している」）および肯否環境ごとに収録している．例えば (1c) の「防いだ」という述語は，肯定環境下では不成立 (CT-)，否定環境下では成立 (CT+) というように，事象「発生する」の肯否極性を反転させている．この辞書は述語の直前の事象の形式も考慮しており，(3) の「忘れる」はその一例である．「発言する」に対して，(3a) では「発言する」の事実性は CT+ となるが，(3b) では「発言する」ことが実際には起こっておらず，事実性は CT- となる．

- (3) a. 彼は発言したのを忘れている。
 b. 彼は発言するのを忘れている。

3.3 解析モデル

前節であげたような事実性に影響を与える手がかりとなる表現を利用し，確信度と肯否極性を依存構造木の根から伝搬させて事実性を解析する．Sauri の英語システムでは，単語をノードとした依存構造木を走査するが，日本語では文節を意味の区切りと考えるのが一般的であるため，文節をノードとした依存構造木を走査する．

このアルゴリズムを Algorithm 1 に示す．ここでは，文全体で伝搬させる事実性を contextual factuality，各事象に割り当てる事実性を event factuality と呼ぶ．入力是一文の係り受け解析結果，出力は各 event factuality であり，contextual factuality は event factuality を伝搬させるための変数である．

まず，contextual factuality に初期値 CT+ を割り当て (line 2)，最後の文節から依存構造木の走査を開始する．各文節で行うことは，機能表現辞書を用いた

Algorithm 1 日本語事実性解析システム

```

1: chunk = root node in DEPENDENCY TREE
2: set initial value (CT+) of contextual-factuality
3: function traverseDependencyTree(chunk)
4:   # Step1: 機能表現による更新
5:   if chunk has functional expressions then
6:     update contextual-factuality with
       functional expressions dictionary
7:   end if
8:   # Step2: 事象をもつかの確認
9:   if chunk has an event then
10:    get event-factuality from contextual-factuality
11:  end if
12:  # Step3: 手がかり表現による更新
13:  if chunk has key expressions then
14:    update contextual-factuality with
       key expressions dictionary
15:  end if
16:  for all chunks c adjacent to chunk not visited do
17:    traverseDependencyTree(c)
18:  end for
19: end

```

更新 (Step1)・事象がその文節内に存在するかの確認 (Step2)・手がかり表現辞書による更新 (Step3) の 3 ステップである．以下では，それぞれのステップについて述べる．

Step1 (line 4-7) では，機能表現辞書と着目している文節とを照らし合わせ，機能表現が見つかった場合はそれに応じて contextual factuality を更新する．例えば「～ない」のような否定表現があれば肯否極性を反転させ「～ではないか」のような疑問表現があれば U_u を割り当てる．Step2 (line 8-11) では，その文節内に事実性を割り当てるべき事象が存在するかどうかを確認し，存在するならば現在の contextual factuality をその事象の event factuality として割り当てる．本来であれば事実性を割り当てるべき事象を特定する必要があるが，今回は分析対象を増やすため，すべてに事実性を割り当てた．Step3 (line 12-15) では，手がかり表現辞書と着目している文節内の述語とを照らし合わせ，マッチした場合にはその文節より前の文節に影響を与えるために contextual factuality を更新する．この順番で処理を行うのは，文節内の事象は機能表現に影響を受け，その事象自体がそれ自身に影響を与えることはなく，現在見ている文節より前の文節に影響を及ぼすことになるからである．

例として，(4) に対するアルゴリズムの動きを図 1 に示す．

- (4) 彼が出場を断念したことを相手は知らない。

まず contextual factuality の初期値 CT+ を割り当て，

Algorithm lines	依存構造木	手がかり表現	contextual factuality	event factuality
2			CT+	
(4-7) 8-11 12-15	知らない	polarity: -(ない)	CT-	知る: CT-
(4-7) (8-11) (12-15)	相手は ことを	modality: CT (知る) polarity: + (知る)	CT+	
(4-7) 8-11 12-15	断念した	polarity: -(断念する)	CT+	断念する: CT+
(4-7) 8-11 (12-15)	彼が 出場を		CT-	出場: CT-

図 1: (4) に対する事実性解析アルゴリズムの動き

依存構造木を文末から走査していく。最初の文節「知らない」に対して、Step1 では極性を反転させる否定の機能表現「ない」が存在するので、contextual factuality を CT- に更新する。Step2 において「知る」は事実性を割り当てるべき事象であるため、そのときの contextual factuality (CT-) を「知る」の event factuality として出力する。Step3 では「知る」という述語が手がかり表現辞書内に存在するので、それに基づいて contextual factuality を CT+ に更新する。各文節についてこれを繰り返し、「知る」の事実性が CT-、「断念する」の事実性が CT+、「出場」の事実性が CT- という出力が最終的に得られる。

4 実験

日本語事実性解析における問題を整理・分析するために、構築した解析器をウェブ上の数千文に対して適用し、その評価を行った。

実験では、拡張モダリティタグ付与コーパス [11] の一部である、Yahoo!知恵袋 (OC) に含まれる 6,404 文 (形態素数: 110,649) に対してアルゴリズムを適用した。このコーパスは現代日本語書き言葉均衡コーパス (BCCWJ)¹ に対して、6 項目からなる拡張モダリティを付与したものである。

¹<http://www.tokuteicorpus.jp/>

表 3: 真偽判断ラベルと事実性との対応

mod \ pol	+	-	u
CT	成立 不成立から成立 (CT+)	不成立 成立から不成立 (CT-)	0 (Uu)
PR	高確率 低確率から高確率 (PR+)	低確率 高確率から低確率 (PR-)	
U			

この中の真偽判断は、肯定か否定か、断定か推量かの組み合わせを表しているものであり、我々の事実性に相当する。態度や相対時といったラベルも事実性に影響を与える場合があるが、今回の実験では単純にこの真偽判断のラベルを正解とした。表 3 に真偽判断ラベルと我々の事実性との対応を示す。事実性と真偽判断ラベルとの対応を考え、CTu, PRu, U+, U-, Uu は全て 0 に対応づけた。また、我々の事実性には、「成立から不成立」や「高確率から低確率」などの変化を伴うラベルに相当するものは存在しない。これら変化を含意するラベルは、変化後が主要要素であると考え、変化後のラベル（「成立から不成立」であれば「不成立」）を割り当てた。

本実験で対象とする事象は、態度表明者が著者となっているものの中で、限定修飾・機能表現のように「対象外」とされていない 14,917 事象である。評価尺度としては、それぞれのラベル (CT+・PR+・PR-・CT-・Uu) に対する二値分類と考える。例えば CT+ に対し

表 4: それぞれのラベルに対する評価

	CT+(成立)	PR+(高確率)	PR-(低確率)	CT-(不成立)	Uu(0)	Micro-Average
TP	7,140	141	7	631	1,893	(9,812)
FP	1,467	286	71	583	2,698	(5,105)
FN	2,791	794	168	408	944	(5,105)
Precision	0.83	0.33	0.09	0.52	0.41	0.66
Recall	0.72	0.15	0.04	0.61	0.67	0.66
F₁	0.77	0.21	0.06	0.56	0.51	0.66

表 5: Confusion Matrix

正解 \ 出力	CT+(成立)	PR+(高確率)	PR-(低確率)	CT-(不成立)	Uu(0)	Total
CT+(成立)	7,140	237	25	408	2,121	9,931
PR+(高確率)	528	141	8	23	235	935
PR-(低確率)	47	6	7	67	48	175
CT-(不成立)	97	1	16	631	294	1,039
Uu(0)	795	42	22	85	1,893	2,837
Total	8,607	427	78	1,214	4,591	14,917

ては、CT+を正しく出力できているものを TP (True Positive), CT+を出力しているが正解は異なるものを FP (False Positive), 正解には CT+がついているが正しく出力できていないものを FN (False Negative) とする。他のラベルも同様にし、TP, FP, FN の数と、各ラベルに対する以下の 3 値を表 4 に示す。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

また、各正解ラベルに対してどのような出力が与えられたかを表す Confusion Matrix を表 5 に示す。表 4 には、多値分類と考えたときの精度 (表 5 の太字部分の和を Total で割った値=Micro-Average) も併記する。これを見ると、特に PR-(低確率) は確信度と肯否極性の両方に対して更新を行わなければならない、難易度が高いことがわかる。

5 誤り分析

3 節で述べた解析器は、確信度と肯否極性を独立して認識し、その組み合わせによって事実性を表している。2 つのうち肯否極性の解析誤りは、成立と不成立のような真逆の誤りにつながるため、確信度の誤りよりも深刻である。そこで肯否極性に関する誤りの分析を優先して行った。ランダムに選出したおよそ 150 の誤り事例について分析し、辞書項目 (語彙知識) の不足に起因する問題 (約 31 %)・辞書にはあるがその解

釈が難しい問題 (約 10 %)・単純に辞書だけでは解くのが難しい問題 (約 33 %) に分類した (残りのおよそ 25 % は、構文解析やアノテーション誤りなどであった)。本節では、それら誤りについて例文とともに述べる。

肯否極性の誤りの多くは語彙知識の不足に起因している。(5) では、「接続」の肯否極性はシステムでは+と認識されたが、正解は-である。これは、事象の肯否極性を反転させる述語である「解除する」が手がかり表現辞書に存在しないことが原因である。

- (5) サーバーは接続を解除しました。(正解:CT-, 出力:CT+)

解釈が難しい問題をもつ表現としては、「すみません」や「申し訳ない」などの、否定表現が含まれているが意味的には否定でない表現があげられる。(6) はその例であり、「質問」の肯否極性が誤っている。

- (6) 初歩的な質問で申し訳ありません。(正解:CT+, 出力:CT-)

肯否極性と同様に、確信度においても語彙知識の不足が原因となる誤りがあった。(7) では「見栄っ張り」に対して、「おそらく」がその確信度を下げているが、今回利用したリソースではカバーできていない。このような推量に関する副詞は川添ら (2011)[14, 15] の枠組みではカバーされており、このような枠組みを利用することが考えられる。

- (7) おそらく、親戚は自慢したいだけの見栄っ張りです。(正解:PR+, 出力:CT+)

以上のようなリソースに起因する問題は、対応するリソースを充実させ、その表現を正しく認識できれば解決できる誤りである。しかし、リソースの問題を解決すれば全ての問題が解けるわけではない。単純に辞書だけでは難しい問題として、表現の語彙的な曖昧性の問題と前提や並列といった節間・事象間の関係を考慮すべき問題が存在する。これらの問題は、 U_u も含めた確信度・肯否極性全体にわたる問題である。

語彙的な曖昧性に起因する誤りの例を(8)に示す。この例では、機能表現「よう」を誤って推量と認識してしまったため、「知る」の確信度をPRと判定してしまった。

(8) 何分かおきに音で知らせてくれるような機械です。(正解:CT+, 出力:PR+)

このような機能表現の曖昧性は、特に確信度の誤りに対して多く見られた。今後の課題として、曖昧性を解消するような枠組みを取り入れる必要がある。

前提や並列のような節間・事象間の関係に起因する誤りを(9)に示す。(9a)は「放棄してしまった」ことが前提となっている。この例では「放棄する」の事実性がCT+とならなければならないが、現在のモデルでは「でしょうか」の U_u がそのまま伝搬されてしまう。(9b)では「証明できます」と「改ざんできません」は独立した事象であり、事実性を伝搬させてはいけない。このような構造を認識し、どのような場合に係り受け構造木の親に依存せず、事実性を伝搬させるべきでないのかを考慮する必要がある。

(9) a. 放棄してしまったのが敗因でしょうか。(正解:CT+, 出力: U_u)

b. アリバイは証明できますし自分の上司は改ざんできません。(正解:CT+, 出力:CT-)

さらに難しい問題として、前後の文脈を考慮しなければ判断できない事例も存在する。(10)において「いる」の事実性は「なら」を順接仮定の機能表現とみるため U_u と認識されたが、正解では成立(CT+)である。これは、その文を見ただけでは「なら」が仮定か前提かを判断できない難しい事例であり、ある程度の文脈を考慮しなければならない。

(10) 教えてくれる人がいるなら、心配する事もないです。(正解:CT+, 出力: U_u)

6 おわりに

本発表では、日本語事実性解析に置ける問題点を分析・整理するため、Saurí(2008)のモデルをもとに、日本語事実性解析システムを構築し、実験を行った。その結果、辞書項目(語彙知識)の不足に起因する問題・辞書にはあるがその解釈が難しい問題・単純に辞書だけで解くのは難しい問題が存在することがわかった。

今回の結果に基づき、今後リソースの整備や節間・事象間の関係の考慮などをする必要がある。リソースの整備は、現在の辞書に不足している表現を、現在の辞書をシードとして獲得してることが考えられる。どのような文脈では事実性を伝搬させ、どのような文脈では伝搬させるべきでないのか、といった節間・事象間の関係に関する知識も明確にしなければならない。今後は更なる分析とともにこれら課題を進め、最終的には事実性も含めた高精度なモダリティ解析に取り組む予定である。

謝辞

本研究を遂行するにあたり、多大な助力をいただきました山梨大学の松吉俊氏、ならびに東北大学 乾・岡崎研究室のメンバーに心より感謝いたします。

参考文献

- [1] Kentaro Inui, Shuya Abe, Kazuo Hara, Hiraku Morita, Chitose Sao, Megumi Eguchi, Asuka Sumida, Koji Murakami, and Suguru Matsuyoshi. Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents. In *the 2008 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 314–321, 2008.
- [2] Marc Light, Xin Ying Qiu, and Padmini Srinivasan. The language of bioscience: Facts, speculations, and statements in between. In *Proceedings of BioLink 2004 workshop on linking biological literature, ontologies and databases*, pp. 17–24, 2004.
- [3] Ben Medlock and Ted Briscoe. Weakly supervised learning for hedge classification in scientific literature. In *Annual Meeting of Association of Computational Linguistics*, Vol. 45, pp. 992–999, 2007.
- [4] Roser Saurí. *A factuality profiler for eventualities in text*. PhD thesis, Brandeis University, 2008.
- [5] Roser Saurí. FactBank 1.0 Annotation Guidelines, 2008.
- [6] Roser Saurí, Jessica Littman, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. TimeML Annotation Guidelines, Version 1.2.1, 2006.

- [7] Roser Saurí and James Pustejovsky. FactBank: a corpus annotated with event factuality. *Language resources and evaluation*, Vol. 43, No. 3, pp. 227–268, 2009.
- [8] György Szarvas, Veronika Vincze, Richárd Farkas, and J. Csirik. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pp. 38–45, 2008.
- [9] 原一夫, 乾健太郎. 事態抽出のための事実性解析. 情報処理学会研究報告, 2008-FI-89, 2008-NL-183, pp. 75–80, 2008.
- [10] 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.
- [11] 松吉俊, 江口萌, 佐尾ちとせ, 村上浩司, 乾健太郎, 松本裕治. テキスト情報分析のための判断情報アノテーション. 電子情報通信学会論文誌 D, Vol. J93-D, No. 6, pp. 705–713, 2010.
- [12] 松吉俊, 佐藤理史, 宇津呂武仁. 日本語機能表現辞書の編纂. 自然言語処理, Vol. 14, pp. 123–146, 2007.
- [13] 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治. 拡張モダリティタグ付与コーパスの設計と構築. 言語処理学会第 17 回年次大会発表論文集, pp. 147–150, 2011.
- [14] 川添愛, 齊藤学, 片岡喜代子, 崔栄殊, 戸次大介. 確実性判断のためのアノテーション済みコーパスの構築. 言語処理学会第 17 回年次大会発表論文集, pp. 143–146, 2011.
- [15] 川添愛, 齊藤学, 片岡喜代子, 崔栄殊, 戸次大介. 言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン Ver.2.4. Technical report, Department of Information Science, Ochanomizu University, OCHA-IS 10-4, 2011.