

クラウドソーシングによる単調推論データセットの構築

Monotonicity Dataset Creation on Crowdsourcing

谷中 瞳^{*1*2}
Hitomi Yanaka戸次 大介^{*2}
Daisuke Bekki峯島 宏次^{*2}
Koji Mineshima関根 聡^{*1}
Satoshi Sekine乾 健太郎^{*1*3}
Kentaro Inui^{*1}理化学研究所 AIP センター
RIKEN Center for Advanced Intelligence Project^{*2}お茶の水女子大学
Ochanomizu University^{*3}東北大学
Tohoku University

Large crowdsourced datasets are widely used for training and evaluating neural models on recognizing textual entailment (RTE). However, it is still unclear whether neural models can capture logical inferences, including monotonicity reasoning, for which no large naturalistic dataset has yet been developed. To investigate this issue, we introduce a method of creating a dataset for monotonicity reasoning by crowdsourcing and report the result of the first run. The error analysis indicates that workers tend to provide different answers from what logical entailment defines, for some downward monotonicity reasonings involving pragmatic reasoning.

1. はじめに

含意関係認識 (Recognizing Textual Entailment, RTE) [Dagan 13] とは、前提文 T が仮説文 H の意味を含むか否かを自動判定する基盤技術であり、コンピュータによる自然言語理解の実現に向けて、きわめて重要な技術である。近年、自然言語処理においては、クラウドソーシングを活用して大規模な言語資源を低コストで構築できるようになった。含意関係認識の研究においても例外ではなく、クラウドソーシングによって人手で構築された大規模な含意関係認識データセット [Bowman 15, Williams 18] から、ニューラルネットワークを用いて文の意味を学習し、高精度で含意関係認識を解く手法 [Devlin 18, Wang 18] が活発に研究されている。しかし、クラウドソーシングによって構築されたデータセットには量化や否定、比較表現、条件文といった、論理的な意味の計算を要する推論現象のデータが含まれていないことが指摘されている [Geiger 18]。そのため、ニューラル含意関係認識モデルがこれらのデータセットから論理語の意味を学習できているのかは自明ではない。

そこで [谷中 19] は、形式意味論において分析が進められている単調推論に着目して、外部辞書を用いて量子子を含む文の語彙を置き換えることで、単調推論のデータセットを自動構築する手法を提案した。この手法ではニューラル含意関係認識モデルの学習に足るデータセットが構築可能である。しかし、この手法では、意味論的推論としては妥当でも、人間の日常的な推論としては不自然な例が作られることがある。例えば、次の例のような downward monotone の推論は直観的には、人でも困難な推論とされている [Geurts 05]。

(1) T : I saw a dog
 $\Rightarrow H$: I saw a dog or a small red cat with green eyes

この推論は、一階述語論理の意味論においては妥当な推論であるが、人は仮説文 H から緑の目をした小さな赤い猫が存在するという前提があると読み込み、語用論的に含意しないと推論する。このように、一階述語論理の意味論において妥当な推論と、語用論的に日常的に人間が正しいと判断する推論にはギャップがあり、含意関係認識モデルの評価ではこのギャップを考慮する必要がある。

既存の単調推論を扱う評価用データセットとしては、FraCaS [Cooper 94] や GLUE diagnostic dataset [Wang 18] がある。しかし、いずれのデータセットも単調推論の問題は 80 件、56 件と非常に少なく、語彙の置き換えも辞書的な語の置き換えと修飾語の追加に限られている。語彙の置き換えはこれらのパターンに限らず、等位接続表現、to 不定詞、関係節の追加など、様々なパターンが考えられる。このような自然な置

き換えパターンのバリエーションと数を担保するためには、クラウドソーシングなどの人手による構築手法が求められる。しかし、専門家ではないクラウドソーシングの作業者が推論の問題を設計することは、前述のギャップも考慮すると容易ではなく、作業内容が明快かつ単純なタスクを設定する必要がある。

本研究では、形式意味論の知見とクラウドソーシングを組み合わせることで、実テキストから単調推論の評価用データセットを構築する手法を提案する。具体的には、Combinatory Categorical Grammar (CCG) [Steedman 00] に基づく統語解析によって単調推論が関わる表現を特定し、その表現をより詳細な表現にするクラウドソーシングタスクを実施することで、単調推論が成り立つ多様な前提文と仮説文のペアを効率的に構築する。なお、構築したデータセットは研究利用が可能な形式で公開する予定である。

2. 単調推論

単調推論は語の置き換えに基づく推論であり、語彙知識と統語構造の統合的な理解が必要となる、重要な言語現象の一つである。この推論は、否定や一般化量子子などの表現が持つ単調性に基いて文中の語を意味的に上位もしくは下位の表現に置き換えた文と、元の文との含意関係が成り立つという推論である [Benthem 83, Icard 14]。上位の表現への置き換えは、ある表現を辞書的に上位の表現に置き換える、修飾表現を除く、or で結ぶ等位接続表現を加える、といった操作によって可能である。同様に、下位の表現への置き換えは、辞書的に下位の表現に置き換える、修飾表現を加える、and で結ぶ等位接続表現を加える、といった操作によって可能である。

polarity は、単調性に依って決定される、言語表現の構文木中における性質であり、↑が付与されている場合、その表現を意味的に上位の表現に置き換えても、元の文との間に含意関係が成り立つ。また、↓が付与されている場合は、その表現を意味的に下位の表現に置き換えても、元の文との間に含意関係が成り立つ。

ある表現が upward monotone である (項関係にある表現の横に “+” を付与して表す) とは、その表現と項関係にある表現の polarity を維持するという意味である。逆に、ある表現が downward monotone である (項関係にある表現の横に “-” を付与して表す) とは、その表現と項関係にある表現の polarity を反転させるという意味である。

一般化量子子は単調推論と関連が深く、一般化量子子の単調性に依って、その一般化量子子の項となる名詞句 (NP) と動詞句 (VP) の polarity が決定する。例として、Every spectator buys a movie ticket という文を考えてみよう。every は、項にとる名詞句に対しては downward monotone であり、動詞句に対しては upward monotone である。このため、(2) に示すように、spectator の polarity は ↓ とな

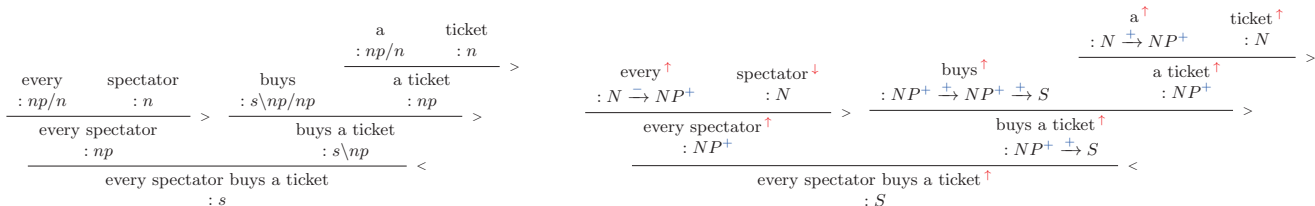


図 1: *Every spectator buys a ticket* に対する CCG の導出木 (左) とそれを変換し各ノードに polarity (↑, ↓) が付与された木 (右)

り、これを下位の表現に置き換えても元の文と含意関係が成り立つ。また、*buys a movie ticket* の polarity は ↑ となり、上位の表現に置き換えても、元の文と含意関係が成り立つ。

- (2) *Every* [_{NP} *spectator* ↓]⁻ [_{VP} *buys a movie ticket* ↑]⁺
 ⇒ *Every* [_{NP} *female spectator*] [_{VP} *buys a movie ticket*]
 ⇒ *Every* [_{NP} *spectator*] [_{VP} *buys a ticket*]

downward monotone の表現には一般化量子子に限らず、様々な表現がある。表 1 に downward monotone の性質を持つ表現の例を示す。

表 1: downward monotone 表現の例

量量子	<i>every, all, any, few, no</i>
否定表現	<i>not, n't, never</i>
動詞	<i>deny, prohibit, avoid</i>
名詞	<i>absence of, lack of, prohibition</i>
副詞	<i>scarcely, hardly, rarely, seldom</i>
前置詞	<i>without, except, but</i>
条件節	<i>if, when, in case that, provided that, unless</i>

こうした downward monotone の性質をもつ表現が複数埋め込まれた文では、polarity の反転が起こる。例えば、条件節のスコープ内に *every* が現れる以下の例では、polarity が以下に示すように反転する。

- (3) *If* [_{every} [_{NP} *spectator* ↑]⁻ [_{VP} *buys a ticket* ↓]⁺]⁻,
 [_{that film will be sold out} ↑]⁺
 ⇒ *If* [_{every} [_{NP} *person*] [_{VP} *buys a ticket*]],
 [_{that film will be sold out} ↑]⁺
 ⇒ *If* [_{every} [_{NP} *person*] [_{VP} *buys a movie ticket*]],
 [_{that film will be sold out} ↑]⁺

このように、文の構成素の polarity を決定する上では、文の統語構造が重要な役割を果たしている。

CCG に基づく統語解析は、文の構成素を関数と項に分類する点で、polarity の計算に適した枠組みとなっている。CCG に基づいて polarity を付与するシステムとして、ccg2mono [Hu 18] がある。図 1 に、その解析例を示す。ccg2mono では、まず入力文に対して CCG に基づく統語解析を行い、CCG の導出木を得る。その上で、CCG の統語範疇 (*s, n, np*) を *e* と *t* を基底型とする意味論の型 *1 に変換し、*every* や *a* などの機能語がもつ monotonicity (+, -) と構成素間の関数・項の関係に着目して、最終的に各構成素の polarity (↑, ↓) が決定される。

3. 提案手法

2 節で述べたように、単調推論が成り立つ前提文と仮説文のペアをクラウドソーシングで構築するためには、語彙が持つ性質と統語構造に基づいて前提文中の表現の polarity を計算し、適切に語彙を意味的に上位・下位の表現に置き換えて仮説文を構築する必要がある。

また、既存のクラウドソーシングによる含意関係認識データセットの構築手法 [Bowman 15, Williams 18] では、否定表現を入れて矛盾となる仮説文を作成するといった、ワーカーの無意識な作業バイアスが生じてしまうという問題が指摘されている [Gururangan 18, Poliak 18, Tsuchiya 18]。既存の手

法ではワーカーに前提文を提示し、前提文と含意・矛盾・非含意の関係になる仮説文を作成するというタスクを課している。しかし、このタスク設定では、ワーカーにそれぞれの意味の関係が成立する条件を指定できていないことが、作業バイアスが生じる原因の一つとして考えられる。

そこで提案手法では、まず CCG に基づく統語解析結果に基づいて各表現の polarity を計算し、upward monotone, downward monotone の性質を持つ表現を含む前提文の選定と置き換え対象の表現を特定する。そして、(i) 前提文中の置き換え対象の表現をより詳細な表現に置き換えて仮説文を作成してもらうタスクと、(ii) (i) でできた前提文と仮説文のペアの含意関係を判定してもらうタスクという 2 つのクラウドソーシングタスクを用意することで、ワーカーの作業バイアスを防ぎ、単調推論のデータセットを構築する。クラウドソーシングのプラットフォームには、figure eight を用いた。次節で各手順の詳細を述べる。

3.1 前提文の選定と置き換え対象の表現の特定

本研究ではまず、統語・意味解析情報つき多言語コーパスである Parallel Meaning Bank (PMB) [Abzianidze 17] のうち、5 トークン以上からなる平叙文約 6 万 4 千件から前提文を選定した。PMB は様々なジャンルの実テキストから構成され、語彙的・構文的に多様な文を扱っているため、多様な単調推論のデータの構築が期待できる。

ワーカーに語彙の置き換えを依頼する際に、抽象的な概念を表す表現や、*no one, every day* といったイディオム化された表現は、語彙の置き換えが困難であると考えられる。そこで 500 件でトライアルタスクを実施し、ワーカーが仮説文を作成できていなかった例から、置き換えが困難な表現を特定した。これらの表現を含む文は前提文として採用しないこととした。特定した語彙の置き換えが困難な表現を表 2 に示す。

表 2: 語彙の置き換えが困難な表現

*one, body, person, people, human,
 idea, money, thing, part, time,
 day, week, month, year, ages, chance, wonder*

次に、各文に対して depccg [Yoshikawa 17] を用いて CCG に基づく統語解析を行い、ccg2mono [Hu 18] を用いて CCG の導出木から文の各構成素の polarity を計算し、↑, ↓ の polarity が付与された文を選定した。この中で、upward monotone の問題と downward monotone の問題の割合を調整するため、表 1 にある downward monotone 表現を含む文を積極的に選定した。最終的に、upward monotone の表現を含む文例 1485 件、downward monotone の表現を含む文例 1982 件の合計 3467 件を前提文として選定した。

3.2 仮説文作成タスク

3.1 節で選定した文を提示し、↑ か ↓ の polarity が付与された表現をより詳細な表現に置き換えて仮説文を作成してもらうことで、単調推論に関わる前提文と仮説文のペアを収集するタスクを設計した。ワーカーはインストラクションにある回答例を参考にして回答するため、より高品質の仮説文を収集するには、わかりやすい回答例を用意することが重要である。そこで

*1 [Hu 18] に従って、*t* を *S*、*e* → *t* を *N*、(*e* → *t*) → *t* を *NP* と表す

表 3: 構築した単調推論データの例

現象	前提文	仮説文	正解
upward monotone	<i>All men have equal rights</i>	<i>All men have equal rights to express oneself in public</i>	非含意
	<i>There is a cat on the chair</i>	<i>There is a cat sleeping on the chair</i>	非含意
	<i>If you heard her speak English, you would take her for an American</i>	<i>If you heard her speak English, you would take her for a native American</i>	非含意
	<i>Dogs have all the good qualities of people without at the same time possessing their weaknesses</i>	<i>Dogs and cats have all the good qualities of people without at the same time possessing their weaknesses</i>	非含意
	<i>It is not tea but coffee that I want</i>	<i>It is not tea but coffee with milk that I want</i>	非含意
downward monotone	<i>Tom hardly ever listens to music</i>	<i>Tom hardly ever listens to rock 'n' roll</i>	含意
	<i>You don't like love stories</i>	<i>You don't like love stories and sad endings</i>	含意
	<i>Tom doesn't like to eat fish</i>	<i>Tom doesn't like to eat fish cooked in restaurants</i>	含意
	<i>I never had a girlfriend before</i>	<i>I never had a girlfriend taller than me before</i>	含意
	<i>Rock climbing without proper equipment is dangerous</i>	<i>Rock climbing without proper equipment of security is dangerous</i>	含意

インストラクションには修飾表現を追加した例、and や or で結ぶ等位接続表現を追加した例、関係節を追加した例、辞書的に上位・下位の語に置き換えた例をそれぞれ提示した。本研究では多様な仮説文を収集するため、1 問あたり\$5 の報酬を与えるとして、1 問あたり 3 名に、1 つの文につき 3 通りの書き換え作業を依頼した。また、低品質な回答を除くため、1 問当たり最低 200 秒以上時間をかけた回答のみを採用した。仮説文作成タスクのインタフェースを図 2 に示す。

Please make the underlined part with a more specific expression in three different ways in English while keeping the naturalness of the whole sentence. Please fill in only a new expression (a more specific expression) and do not fill in the whole sentence.

Sentence: tom seldom listens to music at home.

tom seldom listens to
(required)

at home.

図 2: 仮説文作成タスクのインタフェース

合計 128 名のワーカーに作業を依頼し、upward monotone の推論で非含意となる例 7179 件、downward monotone の推論で含意となる例 8160 件と、合計 15339 件の仮説文を構築した。

3.3 含意関係ラベル判定タスク

ワーカーは専門家ではないため、仮説文作成のような自由記述のタスクでは、意味的に不自然な文が作成されている場合がある。意味的に不自然な文である場合、そもそも含意関係ラベルの判定が困難である。そこで作成した前提文と仮説文のペアについて、含意関係が成り立つか否かに加えて、文が自然な文か否かをチェックするタスクを用意した。

また、ワーカーに含意関係が成り立つか否かを正しく判定させるためには、含意関係の定義をワーカーに正しく理解させる必要がある。そこで 3 種類のラベル (含意、非含意、不自然) について、それぞれ次のような説明をワーカーに指示した。

- 1. 含意: 前提が真であるどのような状況においても、結論がかならず真であるケース
- 2. 非含意: 前提が真であったとしても、結論が常に真であるとは限らない (前提が真である特定の状況では、結論が真とならない) ケース
- 3. 不自然: 文法的に、あるいは意味的に不自然であるケース

また、インストラクションには、含意、非含意、不自然の例をそれぞれ 5 件ずつ提示した。

高品質のアノテーション結果を収集するため、1 回の作業につきチェックテストを 10 問用意し、チェックテストに対して 3 割以上誤答したワーカーの回答、1 問あたりにかかった時間が 200 秒未満の回答は棄却した。1 問あたり\$4 の報酬を与えるとして、3 名にアノテーション作業を依頼した。合計 1237 名のワーカーが参加し、15339 件のアノテーションを行った。

4. 構築したデータセット

4.1 データセットの例

表 3 に構築した単調推論データの例を示す。downward monotone の 1 つ目の例のような *music* と *rock'n'roll* といった辞書的な語彙の置き換えだけでなく、2 つ目の例のような等位接続表現を挿入した例、3 つ目の例のような関係節を挿入した例、4 つ目の例のような比較表現を挿入した例と、クラウドソーシングを用いることによって多様な仮説文が構築できている。

4.2 アノテーション結果の分析

ワーカーの回答の一致率を表 4 に示す。アノテーションの一致率はおよそ 40% であるが、タスクの難易度を考えると妥当な数字である。

表 4: ワーカーの回答の一致率

	一致率/%
3 人一致	40.1
2 人一致	55.7
不一致	4.2

単調推論の性質に基づき自動付与した含意関係ラベルと、ワーカーの回答した含意関係ラベルとが一致した件数と全体における割合を表 5 に示す。自動付与した含意関係ラベルとワーカー全員の回答が一致したデータの割合は全体の約 22% に留まり、構築したデータには人が判断に迷う推論のケースが含まれていることが示唆される。

表 5: 自動付与した含意関係ラベルと一致した件数と割合

	upward /件 (%)	downward /件 (%)	合計 /件 (%)
3 人一致	1069 (7.0)	2285 (14.9)	3354 (21.9)
2 人一致	1814 (11.8)	2301 (15.0)	4115 (26.8)
1 人一致	2295 (15.0)	1915 (12.5)	4210 (27.5)
不一致	1998 (27.8)	1652 (10.8)	3650 (37.8)

4.3 エラー分析

以降はワーカー全員の含意関係ラベルの回答が一致したケース 5941 件を用いて分析を行う。単調推論の性質に基づき自動付与した含意関係ラベルと、ワーカーの回答した含意関係ラベルとが一致した件数と割合を表 6 に示す。

表 6: 自動付与した含意関係ラベルとワーカーの回答との関係

	upward /件 (%)	downward /件 (%)
一致	1069 (7.0)	2285 (14.9)
不一致	1650 (10.8)	937 (6.1)

表 6 のうち、ワーカーの回答が自動付与したラベルと一致していた例は、一見すると正しく単調推論の問題が作成でき

た例であるが、実際に単調推論の問題が作成できているのか、ランダムに 300 件抽出してエラー分析を行った。エラーとしては、以下のように仮説文作成タスクで意味的に下位の表現への置き換えができていなかったケースが約 3 割含まれていた。

類義語による置き換えのケース：次の例は *kid* と *child* が類義語であるために含意関係が成り立つケースであり、意味的に下位の表現の置き換えにはなっていないため、ここで注目している単調推論の問題ではない。

- (4) T: *Tom is no longer a kid*
⇒ H: *Tom is no longer a child*

構成素が異なるケース：次の例の仮説文は、もともと前提文中の *problem* をより詳細な表現に書き換えるよう指示されて作成された仮説文であるが、*at all* は副詞であり、*problem* の構成素ではない。そのため、この例も意味的に下位の表現の置き換えにはなっておらず、単調推論の問題とはみなせない。

- (5) T: *That should be no problem*
⇒ H: *That should be no problem at all*

or を含む等位接続：次の例の仮説文は、前提文中の *strangers* をより詳細な表現に書き換えるよう指示されて作成された仮説文であるが、or を含む等位接続表現の導入は 2 節で述べたように意味的に上位の表現への置き換えであり、下位の表現の置き換えではない。この場合は単調推論の問題にはなっているが、含意ではなく非含意が正しい。

- (6) T: *Tom doesn't trust strangers*
⇒ H: *Tom doesn't trust acquaintances or strangers*

4.4 語用論的推論の影響

表 6 のうち、ワーカーの回答が自動付与したラベルと一致しなかった例について、なぜ不一致が起こるのかランダムに 300 件を抽出し分析を行った結果、類義語による置き換えのケースに加えて、以下のような事例があることが分かった。これらは、様々な語用論的推論によって実際に人間が認識する含意関係が単調推論によって論理的に妥当と予測される含意関係とは異なるケースとみなせる。

対比：以下の例は、単調推論としては非含意と予測される例であるが、ワーカーの回答が含意であった例である。この例では、*cucumber salad without lettuce* という表現によって、*cucumber salad with lettuce* との対比が生じている。

- (7) T: *A horse doesn't eat cucumber salad*
⇒ H: *A horse doesn't eat cucumber salad without lettuce*

このため、H は「レタスが入っていれば食べる」という語用論的な解釈が引き起こされやすく、それがワーカーの非含意の判定に影響したと考えられる。

存在の前提：以下のケースでは、仮説文中の *everyone wearing a blue hat* が存在の前提 (presupposition) をもつと解釈するのが自然である。

- (8) T: *I can't remember everyone*
⇒ H: *I can't remember everyone wearing a blue hat*

この存在の前提は前提文から意味論的には導かれなため、ワーカーの判定との間にずれが生じたものと考えられる。

5. 結論

本稿では、形式意味論の知見とクラウドソーシングを組み合わせることで単調推論のデータセットを効率的に構築する手法を提案した。構築したデータセットを分析した結果、クラウドソーシングを用いることで幅広いパターンの単調推論の問題をある程度構築できることが示唆された。また、論理的に妥当な推論と語用論的に人間が正しいと判断する推論とが食い違うケースを収集することができた。今後の展望として、構築したデータセットから 4.3 節で述べたエラーを除き、評価用データセットとして精緻にした上で、含意関係認識モデルが論理語の意味を学習できるのかについて、評価と分析を進めていく。

謝辞：本研究は JST、AIP-PRISM、JPMJCR18ZM の支援を受けたものである。

参考文献

- [Abzianidze 17] Abzianidze, L., Bjerva, J., Evang, K., Haagsma, H., Noord, van R., Ludmann, P., Nguyen, D.-D., and Bos, J.: The Parallel Meaning Bank: Towards a Multilingual Corpus of Translations Annotated with Compositional Meaning Representations, in *Proc. of EACL-2017*, pp. 242–247 (2017)
- [Benthem 83] Benthem, J. V.: Determiners and Logic, *Linguistics and Philosophy*, Vol. 6, No. 4, pp. 447–478 (1983)
- [Bowman 15] Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D.: A large annotated corpus for learning natural language inference, in *Proc. of EMNLP-2015*, pp. 632–642 (2015)
- [Cooper 94] Cooper, R., Crouch, R., Eijck, van J., Fox, C., Genabith, van J., Jaspers, J., Kamp, H., Pinkal, M., Poesio, M., Pulman, S., et al.: FraCaS—A Framework for Computational Semantics, in *Deliverable*, Vol. D6 (1994)
- [Dagan 13] Dagan, I., Roth, D., Sammons, M., and Zanzotto, F. M.: *Recognizing Textual Entailment: Models and Applications*, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers (2013)
- [Devlin 18] Devlin, J., Chang, M., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018)
- [Geiger 18] Geiger, A., Cases, I., Karttunen, L., and Potts, C.: Stress-Testing Neural Models of Natural Language Inference with Multiply-Quantified Sentences (2018)
- [Geurts 05] Geurts, B. and Slik, van der F.: Monotonicity and Processing Load, *Journal of Semantics*, Vol. 22, No. 1, pp. 97–117 (2005)
- [Gururangan 18] Gururangan, S., Swamyamdipta, S., Levy, O., Schwartz, R., Bowman, S., and Smith, N. A.: Annotation Artifacts in Natural Language Inference Data, in *Proc. of NAACL HLT-2018*, pp. 107–112 (2018)
- [Hu 18] Hu, H. and Moss, L. S.: Polarity Computations in Flexible Categorical Grammar, in *Proc. of *SEM 2018* (2018)
- [Icard 14] Icard, T. and Moss, L.: Recent Progress in Monotonicity, *LILT*, Vol. 9, No. 7, pp. 167–194 (2014)
- [Poliak 18] Poliak, A., Naradowsky, J., Haldar, A., Rudinger, R., and Van Durme, B.: Hypothesis Only Baselines in Natural Language Inference, in *Proc. of *SEM 2018*, pp. 180–191 (2018)
- [Steedman 00] Steedman, M.: *The Syntactic Process*, MIT Press (2000)
- [Tsuchiya 18] Tsuchiya, M.: Performance Impact Caused by Hidden Bias of Training Data for Recognizing Textual Entailment, in *Proc. of LREC-2018* (2018)
- [Wang 18] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. R.: GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding (2018)
- [Williams 18] Williams, A., Nangia, N., and Bowman, S.: A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference, in *Proc. of NAACL HLT-2018*, pp. 1112–1122 (2018)
- [Yoshikawa 17] Yoshikawa, M., Noji, H., and Matsumoto, Y.: A* CCG Parsing with a Supertag and Dependency Factored Model, in *Proc. of ACL-2017*, pp. 277–287 (2017)
- [谷中 19] 谷中 瞳, 峯島 宏次, 戸次 大介, 関根 聡, 乾 健太郎, Abzianidze, L., Bos, J.F ニューラル自然言語推論に向けた Monotonicity に基づく含意関係認識データセット自動構築, 言語処理学会第 25 回年次大会発表論文集 (2019)