

論理関係に基づく複文間の言い換え / 含意関係の認識と生成

阿部修也 乾健太郎 松本裕治

奈良先端科学技術大学院大学 情報科学研究科

1 はじめに

文 (1) と文 (2) は言い換えの関係にあるが、(1) は時間の修辞関係を表す複文で (2) は目的の修辞関係を表す複文である。このように、修辞関係が異なる 2 つの複文の間に言い換え関係が成り立つ場合がある。

- (1) 赤ん坊が飲み込む前に、ビー玉を片付ける。
- (2) 赤ん坊が飲み込まないように、ビー玉を片付ける。

本稿ではこうした異なる修辞関係を持つ複文間の言い換えや含意関係を認識したり生成したりする計算モデルについて論じる。

一方、文 (3),(4) は文 (1),(2) は同じ修辞関係と表層的パターンを持つが、文 (4) が不適格な文であるために言い換え関係にならない。ここからわかることは、修辞関係や表層的パターンを捉えただけでは異なる修辞関係間の言い換えを認識することができないということである。

- (3) 金庫を開ける前に、暗証番号を聞く。
- (4) ?! 金庫を開けないように、暗証番号を聞く。

そこで、異なる修辞関係を持つ複文間の言い換えや含意関係を認識するという問題に取り組むために、文が含む 2 つの事態と事態間の論理的な関係から文の生成パターンを推定するモデルを仮定する²。生成モデルの正しさを証明するために、論理関係の下で生成可能な文の集合を表すコーパスを作成し、実験を行う。

2 論理関係に基づく生成モデル

2 つの事態と論理関係から複文を生成するモデルを仮定する (図 1)。このモデルは、論理関係と 2 つの事態が与えられたとき、修辞関係、接続表現、モダリティ³、

¹文頭に“?”を付けた文は不適格な文

²文 (1),(2) は「ビー玉を片付ける」と「赤ん坊が (ビー玉を) 飲み込む」という事態を持ち、事態間の関係は回避の動機である。文 (3),(4) は「(X が) 暗証番号を聞く」と「(X が) 金庫を開ける」という事態を持ち、事態間の関係は前提条件である。この例から事態間の論理的な関係が文の生成可能性を決めているのではないかと考える。

³ここでは、接続表現を除いた複文を構成する表現をテンス、アスペクト、ヴォイス等を含めてモダリティと呼ぶ。そして、このモダリティの定義から、モダリティと接続表現を組み合わせることで全ての複文の可能性を扱うことができる。

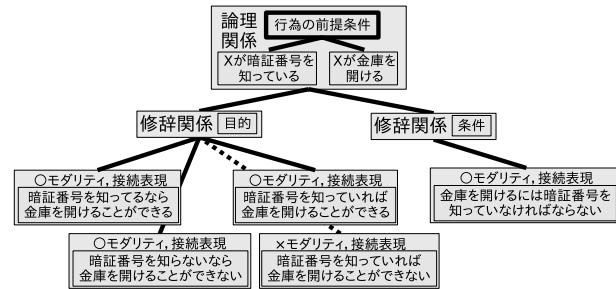


図 1: 論理関係に基づく文の生成モデル

事態と節の対応⁴の組み合わせを変えることで様々な文を生成することができる。ただし、修辞関係とモダリティと事態と節の対応の組み合わせは論理関係に制約されるため必ずしも全ての組み合わせが可能なのではない⁵。

文の表出可能性を判断できるように、論理関係ごとに修辞関係、接続表現、モダリティ、事態と節の対応の組み合わせを変化させたときの文の表出可能性を表すようなパターンを作成する。この文の表出パターンを用いることで、論理関係がわかればどのような文が生成できるのかを判断できるようになる。

3 言い換え関係と含意関係の定義

生成モデルに論理関係と 2 つの事態を与えて文 A と文 B を表出したとき、A の情報が B の情報を包含していれば、A は B を含意しているとする。「文 A の情報が文 B の情報を包含している」とは、事態が表す可能性、真偽、時間、蓋然性等のモダリティが表わす情報が包含関係にあるか、修辞関係も含めた情報に矛盾がない場合を言う。

「ビー玉を片付ける」(前件)と「赤ん坊が (ビー玉を) 飲み込む」(後件)という 2 つの事態と論理関係「動機 (回避)」から表出した文 (5) と文 (6) の間に含意関係が成立する⁶。

⁴前件が従属節に後件が主節に対応しているのか、それとも後件が従属節に前件が主節に対応しているという事態と節の対応関係のことである。

⁵図 1 のリーフで が適格な文、× が不適格な文

⁶文 (6) は論理関係 (回避の動機) を述べているだけである。一方

- (5) 赤ん坊が(ビー玉を)飲み込む前に、ビー玉を片付けなさい!
- (6) ビー玉を片付けないと、赤ん坊が(ビー玉を)飲み込んでしまうかもしれない。

含意関係の定義を用いて言い換え関係を定義すると、2つの事態と論理関係から文Aと文Bを表出したとき、文Aが文Bを含意しかつ文Bが文Aを含意しているならば、文Aと文Bは言い換え関係である。

文(7)と文(8)は互いに含意関係にあるので、定義に従い文(7)と文(8)は言い換え関係にある。

- (7) 赤ん坊が飲み込むといけないから、ビー玉を片付ける。
- (8) 赤ん坊が飲み込む前に、ビー玉を片付ける。

4 論理関係に基づく生成と認識

論理関係に基づく生成モデルを用いて、文が与えられたときに言い換え/含意関係にある文を生成する手順を示す。

1. 入力文が与えられる。入力文は適格な文であると仮定する。
2. 入力文を表出可能な論理関係を推定する⁷。このとき、文が含む事態も推定する。
3. 入力文が含む事態と論理関係の下で表出可能な文の集合を求める。
4. 入力文と表出可能な文の集合の間で言い換え/含意関係が成り立つかどうかを判定し、言い換え/含意関係が成立する文を出力とする。

次に、2つ文が与えられたときに言い換え/含意関係を認識する手順を示す。

1. 入力として2つの文が与えられる。
2. それぞれの入力文に生成モデルを適用して、2つの事態と論理関係を推定する。このとき論理関係が一意に定まらないこともある。
3. 入力文から推定した事態と論理関係の組の集合をA、Bとする。このとき、集合Aと集合Bの積集合を残す。これは2つの入力文に共通する事態と論理関係の組の集合である。この処理で論理関係を推定したときの曖昧性を一部解消することができる。
4. 事態と論理関係の組が空集合の場合、つまり2つの入力文の間で事態または論理関係が異なっていた場合は、言い換え/含意関係にならない。

文(5)は論理関係を述べた上で加えて「ビー玉を片付ける」という行動を命じている。よって、文(5)は文(6)を含意している。

⁷複数の論理関係が選ばれることもある

5. 事態と論理関係の組が空集合でない集合、2つの入力文が言い換え/含意関係が成り立つかどうかを確認し、その結果を出力とする。

5 コーパス

論理関係と表出パターンを作成するために、事態の組とそこから表出可能な文の集合から成るコーパスを作成する。

本研究はあらゆる接続表現とモダリティの組み合わせを考慮しているが、コーパスを作成するにあたり接続表現とモダリティを制限した。主節と従属節のモダリティをそれぞれ(モダリティ)なし、否定(～しない)、義務(～しなければならない)、可能(～できる)、可能の否定(～できない)に限定し、接続表現を「ために」「ので」「れば」「なら」「けれども」に限定する。

250組の事態に対して、制限を加えた接続表現とモダリティ、さらに事態と節の対応を様々に変化させて、人手で複文を表出した。コーパスには表出できた複文のみを記述し、表出できない複文は空欄とした(表1⁸)。

本コーパスにおける事態は基本的にモダリティを除いた単文である。しかし、現在は接続表現を制限しているが、本当は様々なモダリティを網羅的に観察したいという理由がある。そのため、本コーパスでは、受け身形を含む形も事態とする。

6 実験

生成モデルに基づいて、適切に論理関係と表出規則を作成することで、論理関係から文を表出できることを示す。

コーパスから18種類の事態の組を抽出し、事態間の論理関係⁹と事態に対して意思性と状態性を持つかどうかというタグを付与した。ここから、人手で論理関係と論理関係ごとの表出規則を生成した。

5種類の論理関係を作成¹⁰した。論理関係の説明と例を示す。

結果 前件の結果として後件が満される「夜になれば月が見える」(夜になる 月が見える)、「運動したので汗をかく」(運動した 汗をかく)

動機 後件を満すために前件を行なう「海外旅行に行くためにパスポートを取る」(パスポートを取る 海外旅行に行く)、「汗をかくために運動する」(運動した 汗をかく)

⁸行を接続表現、列を従属節のモダリティとして表出可能な全ての文を作成した。同一枠内に、接続表現と従属節のモダリティは等しいが、主節と事態と節の対応が異なる文が複数記載されることもある。

⁹ある事態の組が2つ以上の論理関係を持つこともある。

¹⁰論理関係を作成するにあたり [6] を参考にした

表 1: コーパスの例: 「目が悪い」(前件), 「兵役を免除される」(後件), 結果(論理関係)

接続表現\従属節のモダリティ	モダリティがつかない	否定	義務	可能	可能の否定
ために	目が悪いために兵役を免除される.	目が悪くないために兵役を免除されない.			
ので	目が悪いので兵役を免除される.	目が悪くないので兵役を免除されない.			
れば	目が悪ければ兵役を免除される.	目が悪くなければ兵役を免除されない.			
なら					
けれども	目が悪いけれども兵役を免除されない.	目が悪くないけれども兵役を免除される.			
ても	目が悪くても兵役を免除されない.	目が悪くなくても兵役を免除される.			

必要条件 前件が後件の必要条件になっている¹¹。「海外旅行に行くならばパスポートを取らなければならない」(パスポートを取る 海外旅行に行く)、「汗をかくためには運動しなければならない」(運動した 汗をかく)

十分条件 前件が後件の十分条件になっている。「パスポートを取れば海外旅行に行くことができる」(パスポートを取る 海外旅行に行く)、「運動すれば汗をかくことができる」(運動した 汗をかく)

許可 前件が許可されるために後件を満す。「株価が上るならば株を買うことができる」(株を買う 株価が上る)、「毎日世話をするのでペットを飼うことができる」(ペットを飼う 毎日世話をする)

論理関係ごとに、接続表現と主節のモダリティと従属節のモダリティと事態と節の対応が与えられたとき表出可能かどうかという規則を作成した(表 2)。

さらに、事態(前件, 後件)がそれぞれ意思性を持つかと状態性を持つかに基づいて、次に示す統語的制約を与える規則を作成した。

- 主節と従属節のモダリティがなく、接続表現「れば」「なら」、主節と前件が対応している文を生成できない。
- 従属節の意思性がない場合、後件と従属節が対応し、主節のモダリティが「義務」「可能」「可能の否定」になる文を生成できない。
- 従属節の意思性があり、かつ状態性がない場合、主節と従属節のモダリティがなく、接続表現「ために」、主節と後件が対応している文を生成できない。
- 従属節の意思性があり、かつ状態性がない場合、主節のモダリティが「否定」「可能」「可能の否定」、従属節のモダリティが「可能」「可能の否定」となる文を生成できない。
- 従属節の意思性と状態性がなく、かつ主節の意思性がない場合、モダリティに関わらず、接続表現「れば」「ても」、主節と後件が対応している文を生成できない。

¹¹ 厳密な意味での必要条件ではない。海外旅行に行くためにパスポート以外にも必要な条件があるが、パスポート以外の条件は満たしていると仮定して、海外旅行に行くならばパスポートを取らなければならないと言える。

表 2: 表出規則: 接続表現の集合「ために」「ので」「れば」「なら」を A, 接続表現の集合「けれども」「ても」を B, 接続表現全てを含む集合を * とする。主節が前件に従属が後件に対応していれば「1」、主節が後件に従属が前件に対応していれば「0」とする。

論理関係	接続表現	主節	従属節	事態と節の対応
結果	A	なし	*	0
結果	B	否定	*	0
動機	A	なし	*	1
動機	B	否定	*	1
必要条件	A	義務	*	1
十分条件	A	可能	*	0
十分条件	B	可能の否定	*	0
許可	A	可能	*	1
許可	B	可能の否定	*	1

論理関係に基づいて文が表出され、統語的制約によって表出された文の一部が削除され、この結果が事態とその間の論理関係ごとの文の表出パターンとなる。

結果 表出規則の作成に用いた 18 種類の事態の組(540 種類の文の表出可能性)に対して表出規則と統語的制約を捉える規則を適用し、人手で作成した表出パターンを十分に捉えているかを観察した。再現率は 0.80, 精度は 0.75, F 値は 0.77 であった。

7 考察

5 種類の論理関係に基づく表出規則と統語的制約を用いて、人手で作成した表出パターンを十分に捉えられるかという実験を行なった。無作為に表出パターンを判定しても F 値が 0.50 であることを考えると、特に良い結果ではなかったが、簡単な規則だけで表出パターンを捉えられることがわかった。

結果が悪い理由について、コーパスの問題と統語的制約の問題に分けて考察する。

コーパスの揺れ コーパスの揺れ¹²と論理関係の解釈が一貫していないという現象がある。

コーパスの作成方法を改善することで問題に対処する。コーパスの漏れに対しては、予め文の候補を提示しておき、作業者は文の適格性のみを判断するという

¹² コーパスが揺れているというのは、例えば「(後件)するので(前件)する」は言えているのに「(後件)しなければならないので(前件)する」が言えないといった現象のことを言う。特にモダリティ「義務」に関係する部分が大きく揺れている。モダリティ「義務」を除いた場合は、再現率は 0.84, 精度は 0.88, F 値は 0.86 であった。なお、統語的制約を用いない場合は、再現率は 0.80, 精度は 0.51, F 値は 0.63 であった。

方法で対処する。コーパスの揺れと論理関係の解釈に対しては、論理関係の明確な定義と論理関係毎の言語試験を導入することで対処する。同時に、作業仕様書の整備を行なう。

統語的制約が不十分 この問題の原因は、統語的制約を行なうための規則が不十分であることと統語的制約に用いる情報が少ないということがある。

規則が不十分であるという点については、人手で規則を整備しているため不十分になってしまった。そこで、機械学習を用いることなどの工夫で対処する。

接続表現と意思性の有無のみで統語的制約を行なっているため、この情報だけでは統語的制約を行なうのに不十分である。そこで、語彙概念構造 [7] を導入し統語的制約に用いる情報を増やすつもりである。

修辞関係間の言い換え / 含意関係の認識 / 生成のための論理関係 本実験で用いた論理関係に基づく表出規則は1つの論理関係のみに依存しているため、表出パターンから一意に論理関係が推定でき、かつ論理関係から一意に表出パターンを推定できるという利点がある。

一方で本実験で用いた表出規則は主節のモダリティに依存しているため、わずかな種類の修辞関係が異なる言い換え / 含意関係しか捉えることができない。そこで、論理関係「必要条件」が満たされていれば論理関係「結果」も同時に満たされているといったような、論理関係間の依存関係を作成する必要がある。

8 関連研究

Knott[4] は複文の接続表現以外の文の構成要素を固定し、接続表現のみを他の接続表現変更したときの可換性から接続表現の分類を行なった。接続表現という観点から見れば、Knottの研究は接続表現の言い換えを収集したことになる。Hutchinson[1][2][3] は、Knottが接続表現を可換性に基づいて人手で分類した作業を、半自動的に行なった。Power[5] は、同一の修辞関係にある言い換え関係の生成を試みた。

本研究は、Knott や Hutchinson のように接続表現だけを可換とするのではなく接続表現とモダリティの両方を可換とし、Power のように同一の修辞関係に限定せずに、異なる修辞関係の言い換え関係を生成する。修辞関係を固定した場合と比較して多様な言い換えを生成することができる。また、生成を用いることで言い換への認識も行なう。さらに、言い換え関係の認識と生成と同様の手法を用いて、含意関係の認識と生成も行なう。

9 まとめ

論理関係に基づいた言い換え / 含意関係の認識と生成という問題に取り組むために、論理関係を用いた生成モデルを仮定した。適切な論理関係を用いることで、表出パターンを推定できること示した。

今回の実験は小規模なコーパスを用いたので、今後は大規模なコーパス¹³で実験を行い、作成した論理関係の規則が適切であることを示したい。また、コーパス作成の際に揺れを押えるために、作業用手引書を用意しようと考えている。

他に、大規模コーパスを用いて、機械的な手法等を用いて統語的な表出規則を作成、論理関係ごとの依存関係を作成、論理関係ごとの言い換え / 含意関係を作成する。

参考文献

- [1] Ben Hutchinson. Acquiring the meaning of discourse markers. In *the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, pp. 685–692, 2004.
- [2] Ben Hutchinson. Modelling the similarity of discourse connectives. In *the 27th Annual Meeting of the Cognitive Science Society (CogSci 2005)*, 2005.
- [3] Ben Hutchinson. Modelling the substitutability of discourse connectives. In *the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, 2005.
- [4] A Knott. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. PhD thesis, Department of Artificial Intelligence, University of Edinburgh, 1996.
- [5] R. Power and D Scott. Automatic generation of large-scale paraphrases. In *Proceedings of the 3rd International Workshop on Paraphrasing (IWP2005)*, pp. 73–79, 2005.
- [6] 乾孝司, 乾健太郎, 松本裕治. 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得. 情報処理学会論文誌, Vol. 45, No. 3, pp. 919–933, 2004.
- [7] 竹内孔一, 乾健太郎, 藤田篤, 竹内奈央, 阿部修也. 分類の根拠を明示した動詞語彙概念構造の構築. 自然言語処理研究会 2005-NL-169, 2005.

¹³ある論理関係について事例を集めるとき重要なことは、似たような関係を持つ事例の組を沢山集めるのではなく、一見異なる関係だが論理関係は等しい事例の組を沢山集めることである。今回は少ない事例で実験を行なったが、ある程度と一見異なる関係の事例を集めた。