

事実と推論との関係を反映させた「しかし」による文の自動接続

Automatic Conjunction of Sentences by 'shikashi' Reflecting Relationship Between Facts and Inference

佐藤 陽介*1 井原 健紘*2 長井 隆行*1
Yosuke SATO Takehiro IHARA Takayuki NAGAI

*1電気通信大学電子工学専攻

The Department of Electronic Engineering, The University of Electro-Communications

*2電気通信大学情報工学専攻

The Department of Computer Science, The University of Electro-Communications

This paper discusses a way to generate the paragraphs, which describe human movements in the real world, involving Japanese conjunction 'shikashi' (in English 'however' or 'but') without referring to examples written by humans. Three criteria are used to generate paragraphs automatically. The subjective test shows that these criteria yield logical paragraphs that are comparable to those selected manually.

1. はじめに

本稿では、人間が実際におこなった行動を機械に自動的に文章化させる方法について議論する。このとき、行動を羅列するだけでなく、そこに論理性を付与したいと考えている。

昨今、画像や音声の認識技術の向上が著しい。例えば、動画画像から得られた情報から人間の行動を自然言語に変換することができるようになってきている [小島 1998]。また、実環境下においても 8 割から 9 割程度の精度で音声から文字列への変換ができるようになってきている。つまり、十分な語彙知識と十分なセンサーさえ存在すれば、人間の言動と行動を逐次的に文章化できるところまで来ている。

しかしながら、行動や言動を羅列するだけでは文章とは言い難い。人間の書く文章と機械の出力する文の羅列との差はいくつか挙げられるが、文章に論理性を付与したいという目標の下、本稿では、接続詞を適切に用いる方法について考える。

接続詞をとりあげた先行研究には、「ため」という語に注目した [乾 2002]、「にもかかわらず」という語に注目した [富川 2005] などがある。これらは新聞やウェブ記事のデータを事例として扱っており、人間がその事例を作っているために事例に間違いがないという長所がある。非常に強力な研究戦略であるといえよう。しかしながら、機械に文章を出力させることを考えたときに、事例を集めて分類したりルールを作ったりしていても、生産的な言語運用はできないと予想できる。ゆえに、人間の作った事例が存在しない場合についての方法論を考えなければならない。

人間はいかにして、接続詞を使っているのか。[野矢 2003]によれば、接続詞「しかし」は真理条件的意味論では捉えきれないとしている。そして、接続詞「しかし」は、先行了解が裏切られたときに使われるとしている。また、[森 1993]によれば、接続詞「しかし」の前後の文には二項対比の関係があるとしている。本研究は、これら二つの原理をベースとして、自動的に「しかし」という接続詞を使用する機械の実現を目的とする。

2. 問題の定義

研究室内での人間の実際の行動から文を生成し、その文を「しかし」で論理的に適切に繋ぐことを考える。このときに必要な要素は、行動を描写した文、および、接続詞「しかし」である。この行動を描写した文を「単文」と呼ぶことにする。また、「しかし」の前の文を「前文」と呼び、後ろの文を「後文」と呼ぶことにし、二つの単文を「しかし」で繋いだものを「文章」と呼ぶことにする。

と呼ぶことにし、二つの単文を「しかし」で繋いだものを「文章」と呼ぶことにする。

単文をなるべく正確に生成したかったため、本研究では超音波センサーを用いた。これによって室内における人間の位置を誤差 5cm の精度で求めることができ、この位置情報を用いて移動を記述する単文を作ることができる。単文の要素は常に「誰が (6)・何日の (13)・何時頃に (9)・どこで (16)・何を (10)」の 5 つである。なお、括弧内の数字はそれぞれの語彙数である。それぞれの語彙の意味は既知であるとした。ただし、「東の反対は西である」といった語彙間の関係性は未知である。また、突発的に起きた行動に関して議論するのではなく、習慣性のある行動を扱いたかったために、頻度の高い行動のみを単文として生成し、「よく」という副詞を用いた。

本研究で考えるべき問題は、作成した単文の中からいかにして前文と後文を選出するかである。適切に選出できなかった場合、違和感の生じる文章が作成されることとなる。

3. 3つの基準

前文と後文を適切に選択するために、3つの基準を作り、それらの基準全てを満たす前文と後文の組を出力とした。

説明の例として、以下の3つの単文を用いる。なお、「期待文」については後述する。

- 佐藤は 12 月 3 日の朝によく東側を歩いた (前文)
- 佐藤は 12 月 3 日の昼によく東側を歩いた (期待文)
- 佐藤は 12 月 3 日の昼によく西側を歩いた (後文)

3.1 二項対比

単文には 5 つの要素があるが、前文と後文とで 2 つの要素が異なるようにした。例えば、例に示した前文と後文では、「何時頃に」の要素である「朝」と「昼」が異なっており、「どこで」の要素である「東側」と「西側」が異なっている。異なる要素の数は 1 つの場合でも 3 つの場合でも不自然に感じられることが調査されている [森 1993]。例えば、例に示した前文と期待文を「しかし」で繋ぐと違和感が生じる。これは、次のような理由によるものと考えられる。接続詞「しかし」というのは、前文からある行動 (これを「期待文」と呼ぶことにする) が期待されやすく、そのある行動から後文が期待されにくいにもかかわらず実際に起きてしまった場合に、用いられると考えられる。この期待文の生成のときに前文と期待文で要素が 1 つ変わり、期待文から後文を生成するときに要素が 1 つ変わる。つ

まり、単文から単文を期待するときに要素が1つ変わるの、異なる要素の数が2つになることの原因であろう。ただし、この二項対比の基準のみでは充分ではない。普段の行動を観察しなければ、内容的に不適切な文章が出来上がってしまうことがあるため、「期待と裏切り」の基準を設ける。

なお、前文と期待文との間では「誰が・何日の・何時頃に」の要素が変わるものとし、期待文と後文との間では「どこで・何をした」の要素が変わるものとした。

3.2 期待と裏切り

文章の内容的な側面である「前文から期待文への期待されやすさ」と「期待文から後文への期待されにくさ」を計算する必要がある。この期待されやすさは次のように間接的に計算することとした。

前文と期待文で異なる要素・「どこで」の要素・「何をした」の要素の3つの要素を固定し、残りの2つの要素をパラメータとする。例の場合では「誰が」と「何日の」の要素がパラメータとなる。このパラメータを変化させることにより、前文から期待文への期待されやすさ、期待文から後文への期待されやすさを計算した。例の場合でいえば、「A氏はB日の朝によく東側を歩いた」と「A氏はB日の昼によく東側を歩いた」の頻度が、パラメータAとBを変化させたときに共に高いかどうかを判定することにより、期待されやすさを推測した。

3.3 かけ離れない行動

「しかし」の前後にまるで関連のない「何をした」の要素が来ると違和感のある文章が出来上がる。今回の枠組みでは、特に習慣性についての文章を生成しようとしているため、その「何をした」の要素そのものの頻度が低いと違和感が生じてしまう。そこで、頻度の低い「何をした」の要素は期待文と後文の間で変化させないものとした。こうして、頻度の高い「何をした」の要素同士を比べることにより、行動がかけ離れないようにした。

4. 出力形態

単文をそのままの形で出力すると不自然になるため、あらかじめルールを作り、助詞の変更・繰り返し省略・単語の言い換えをおこなった。また、テンプレートにしたがって、前文と後文のみならず期待文も出力するものとした。テンプレートは3種類である。以下に実際の出力例を示す。

- 「肯定型」最後が肯定文で終わる
長井は11月19日の夜に東側をよく歩いた。青森も東側をよく歩いた。しかし、杉山はその時東側にある井原席の前でよく立ち止まった。
- 「否定型」最後が否定文で終わる
青森は12月8日の夕方頃に東側をよく走った。しかし、11月19日の夕方頃には東側を全然走らなかった。
- 「説明型」習慣の傾向を途中で説明する
板本は11月22日の夕方頃によく東側を歩いた。杉山も長井も青森もよく東側を歩いた。ほとんどみんな11月22日の夕方頃に東側を歩いたので、木村もよく東側を歩くと考えたのに、木村はよく西側を歩いたのだった。

5. 主観評価実験

7人の被験者を用いて主観評価実験をおこなった。次の4種類の文章を読んでもらい、0から10の11段階で論理性を評価してもらった。4種類の文章については20文ずつ用意した。

表 1: 主観評価実験の結果

文章の種類	(1)	(2)	(3)	(4)
平均点	3.9	6.5	7.0	7.3

- (1) 頻度の高い全ての単文の中からランダムに前文と後文を選んで作った文章
- (2) 本研究により出力された「肯定型」の文章の中で期待文を含まないもの
- (3) 人間によって選ばれた単文によって作られた「肯定型」の文章
- (4) 本研究により出力された様々な型の文章

それぞれの平均点を表1に示す。ここから、ランダムに生成された文章と人間によって生成された文章には明らかな差があることが分かる。また、本研究の枠組み内では、機械の出力と人間によって生成された文章との間にほとんど差がないことが読みとれる。

6. まとめと考察

人間の作った事例を参照することなく、接続詞「しかし」を用いて、実際の行動を描写した文章を生成することを考えた。適切に文を接続するために、3つの基準を作り、その基準を満たすように文章を生成した。その結果、ランダムに文と文を繋いだときに比べて論理的な適切さの高い文章を生成することができた。

何がおこなわれたのかという情報のみでは接続詞「しかし」を適切に用いることができないため、行動の推測に基づく「期待と裏切り」の基準を用いた。今回の枠組みではここからさらに「二項対比」の基準が導かれたが、枠組みを変えた場合に「二項対比」の基準は用いることができなくなるため、本質は「期待と裏切り」にあるものと考えられる。また、第三番目の基準として「かけ離れない行動」の基準が必要となった。この必要性については直感的には理解できるものの説明をつけることができていないため、今後、考察を深める必要がある。

本稿の手法は、推測される行動と現実起きた行動との比較であると言い換えることができる。もし、「推測手法」「現実起きた行動の取得方法」「比較手法」の範囲が広くなれば、より表現の幅が広がると予想される。

参考文献

- [小島 1998] 小島篤博, 田原典枝, 田村武志, 福永邦雄, “動画像における人物行動の自然言語による説明の生成,” 電子情報通信学会論文誌 (D-II), vol.J81-D-II, No.8, pp.1867-1875, Aug. 1998.
- [乾 2002] 乾孝司, 乾健太郎, 松本裕治, “接続詞「ため」を含む複文から因果関係知識を獲得する,” 自然言語処理研究報告, 2004-NL-150-25, Jul. 2002.
- [富川 2005] 富川昌則, 石川勉, “接続詞「にもかかわらず」を含む文からの因果知識の獲得,” 情報処理学会第 67 回全国大会, pp.2-323-2-234. 2005.
- [野矢 2003] 野矢茂樹, “「しかし」の論理,” 月刊「言語」, Vol.32, No.3, pp.18-25, Mar. 2003.
- [森 1993] 森辰則, 中川裕志, “接続詞「しかし」の意味論,” 情報処理学会第 46 回全国大会, pp.3-175-3-176. 1993.