

複文構造における言い換えについて

中野 滋 徳[†] 足 立 顕^{††} 牧 野 武 則[†]

日本語文はその語順を変えることで複数の表現方法を持つことができ、文の内容を変えず表現方法を変える研究が注目を集めている。本論文は、依存ネットワークを用いた複文の言い換えについて述べる。複文構造を依存ネットワークで表現し、ネットワーク上のノードの関係を書き換える規則を与えることで複文構造の言い換えが可能であることを示す。

キーワード: 言い換え, 依存関係, ネットワーク, 複文

Automatic Paraphrase Generation for complex sentences

SHIGENORI NAKANO[†] AKIRA ADACHI^{††} and TAKENORI MAKINO[†]

Japanese sentence has various way of expressions which are rewritten by its word sequences. Recently, many papers reported a method of paraphrase sentence generation. This paper describes a method of paraphrase sentence generation for a complex sentence structure based on dependency network. The method derives paraphrase sentence by rewriting the relations between nodes from the dependency network of a complex sentence.

Keywords: *paraphrase, dependency, network*

1. はじめに

自然言語では一つのことを表するのに複数の表現方法がある。ある表現を内容を変えずに別の表現へ変換する処理を言い換えと位置づけている¹⁾。

言語自体に用意された多様な選択肢を適切に組み合わせることにより、文中の任意の語を文末とする表現が自由に得られるならば、強調表現の自由度が広がることになる。さらには文の内容を保存したまま、その語に対する修辞関係を正確に取出せた時、情報検索や情報要約⁶⁾のタスクにおける前処理として、有用な要素技術となる。このような任意の語を文末とする言語操作を、以後、語にフォーカスを当てると呼ぶ。

本稿では依存ネットワークに基づく言い換え文生成モデル⁵⁾を適用し、複文における従属節内の

任意の語にフォーカスを当てた文生成ツールを実装した。さらにそのツールに必要な語ならびに単文間における多様な接続形態の関係情報を定義した。これらをもとに、EDRの文例から言い換えた実験結果を評価する。

本稿の構成は、まず2節で先行研究について述べ、3節では文生成方法のモデルを紹介する。4節で関係情報の取得方法について取り上げ、5節で文生成ツールの実装について述べる。6節で本手法による評価実験を行い7節でまとめを行う。

2. 先行研究

複文の言い換えに関する先行研究として連体修飾節や連用節の言い換えがある。これらはいずれも文を分割して、必要ならば適切な接続詞を添えるという取り組みがなされている。

神田らの連用主節化に関する規則の追試²⁾では、複文の分割の実現にあたり連用節主節化の規則を追試し、文の分割候補点の決定規則、接続詞の選

[†] 東邦大学 大学院理学研究科 情報科学専攻
Department of Information Sciences Toho University

^{††} 富士通株式会社 情報メディアソリューション本部
Division of Media Solution, Fujitsu Co. Limited

択基準の決定方法，連用節を主節化するための接続表現についての考察がなされている．

野上らの連体修飾節の言い換え³⁾では，連体節を主節から切り離して主節化するために対象文を分割して，2文の順序と文間の接続表現の選択基準決定のアプローチが報告されている．

依存構造を利用した言い換えの研究として，単文ならびに連体修飾を含む複文の言い換への報告がなされている⁴⁾．これによると依存構造から依存ネットワークに展開したとき，ノード間の関係情報が設定できればエントリノードを文末にした言い換えが可能としている．しかし関係情報としては述語に対する格助詞に限定しているために，副詞節や形式名詞を含んだ複文に対する言い換え文の生成にまで至っていない．

文の構造には格助詞以外にも多様な関係表現が存在することから，複雑かつ多重に文が入り組むにしたがい，文の分割タスクは非常に重要な課題である．

しかし文を分割することは，文を相互に結び付けるための規則が複雑になることになることから，我々は複文のままの言い換えがどこまで可能かについて取り組んだ．

3. 文生成方法

複文に対応した言い換え文として，柔軟な文生成方式についての報告⁵⁾がなされている．語と語の直接の関係を表層表現により表示したネットワーク型の構造から文を生成する手法で，ネットワーク上の任意のノードを文生成のエントリノードとして指定することで，文生成を可能とする手法である．

依存ネットワークは原文におけるノードの位置情報をもたず，ノードとノードとの直接の関係を表層表現を用いて表示した構造である．図1に依存ネットワークのモデルを示す．

依存ネットワークは w_i はノードを表し w_i と w_j を結ぶアークはノード間の関係を示す．ノード間の関係情報 $R(w_i, w_j)$ を文の生成時にネットワークをトラバースするとき，その関係情報を規則に基づきスタックしながら線形変換を行う．その結果言い換え文を得ることができる．

したがって複文構造において，ノード間の関係

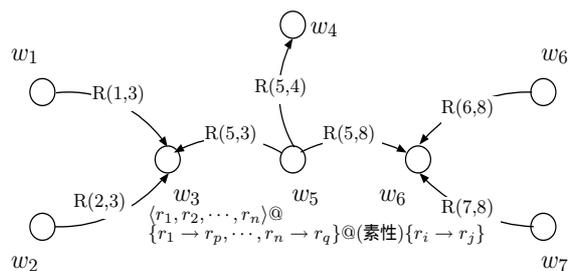


図1 aaa

情報をいかに定義できるかが課題となる．

4. 関係情報

複文を構成する単文の接続形式は多様である．そこで文をつなぐ構文の形式的分類として，寺村⁷⁾は以下のように分類している．

- (1) 一つの文の文末の述語の活用形をつなぐ
- (2) 接続助詞をつなぐ
- (3) 一つの文をほかの文の修飾節にする (文の連体節化)
- (4) 一つの文をほかの文の名詞的構成要素にする (文の名詞節化)
- (5) 引用 (文の引用節化)

今回の実験では，語と語の直接の表層表現が得られるものを優先し，(1)の文末の述語の活用形をつなぐ接続形式の複文は対象から除外する．(3)連体修飾節については先の報告⁴⁾と重複するので(2)，(4)，(5)を中心に関係情報取得方法について述べる．

従属節の語をフォーカスするためには従属節を主節にする必要がある．この操作を可能にする関係情報を取得するために，以下に示す取り組みを行った．

- 接続助詞に対応した言い換えの調査
- 「のは強調構文⁷⁾」を利用した調査 (分裂文⁸⁾と呼ばれることもある)

接続助詞に対応した言い換えについては，主節と従属節の接続表記を変えたり文末表記を追加することで，文の内容が保存できるケースを調査した．

「のは強調構文」について，日本語の文法(下)⁷⁾によれば，文中の補語のどれかを取り出して特にほかのものと際立たせて叙述する強調構文であると述べている．

そこで関係情報を定義するために EDR⁹⁾ の文例を調査した。この「のは強調構文」は EDR⁹⁾ の中に 207,802 件中 6,120 件 (全体の 3%) 存在している。この文例で注目したのは、接続助詞と同じ表記の「とき」「から」「あと」「ため」などの形式名詞が強調構文の文末に多数存在していることである。そこで「のは強調構文」を含んだ例文に対して、想定される原文の従属節が主節化したものと想定したとき、強調構文を適用しない文に逆変換が可能と仮定して調査を行った。この逆変換操作を通して従属節を主節に変換するための言い換え表現に相当する関係情報を報告する。

- (1) それを思い知らされたのは、この秋、東シベリア・クラスノヤルスク地方を視察したときだった。
この秋、東シベリア・クラスノヤルスク地方を視察したときそれを思い知らされた。
- (2) 竹下さんに指名を決めたのは最終の段階で、白紙一任をいただいた後だ。
最終の段階で、白紙一任をいただいた後竹下さんに指名を決めた。
- (3) 急に中国が態度を変えたのは、われわれが中国に対して自立した政策を持ったからだ。
われわれが中国に対して自立した政策を持ったから急に中国が態度を変えた。
- (4) 私が国会図書館を利用するようになったのは、卒業論文作成の文献探しのためだった。
卒業論文作成の文献探しのため私が国会図書館を利用するようになった。
- (5) 彼らに必要なのは、まず自分たちを組織化し、団結した力で社会を変えていくことだ。
まず自分たちを組織化し、団結した力で社会を変えていくことが彼らに必要なだ。
- (6) 忘れてはいけないのは、売上税の建前は最終的にすべて消費者が負担するということだ。
売上税の建前は最終的にすべて消費者が負担するというのを忘れてはいけない。

実例で示すように、接続助詞の表記と一致する「とき」「後」「から」「ため」や名詞節を構成する形式名詞「こと」が強調構文に出現している。これらの結果をもとに、接続助詞や形式名詞とを対応づけた関係表現を定義した。

なお寺村⁷⁾ は、この「のは強調構文」の述部の位置にくる述部は名詞化しなければならないと述べている。したがってこれを関係情報として利用するときは、文末を名詞化することになり、関

係情報には文末表記との両方をもつことになる。

以上の調査結果から、接続助詞と名詞節・引用節を構文化する接続形態ごとに定義した関係情報を述べる。なお名詞節・引用節を構文化する接続形態を、以後の説明のため一括して結合標識と呼ぶ。

4.1 接続助詞

接続助詞で時、原因・理由、逆接を表す従属節は適切な関係情報を定義することで言い換え可能である。その関係情報を表 1 に示す。文末表記の述語と名詞は、文末の素性(素性)でそれぞれ言い換えること示している。

なお、副詞節を形成する接続助詞の分類は基礎日本語文法¹⁰⁾ の分類に従った。

表 1 接続助詞の関係情報

種類	接続助詞	関係情報	文末表記	
			述語	名詞
時	とき	のは	ときだ	のときだ
	から	のは	からだ	だからだ
	まえ(に)	のは	まえだ	のまえだ
	あと(に)	のは	あとだ	のあとだ
原因理由	から	のは	からだ	だからだ
	ため	のは	ためだ	のためだ
	おかげで	のは	おかげだ	のおかげだ
	せいで	のは	せいだ	のせいだ
逆接	けれど(も)	けれども		
	のに	けれども		
その他	反面 一方	反面 一方		

時を表す接続助詞では、時間の幅を問題にするような「まで」「あいだ」などは関係表現をつけるのは困難であり、時の 1 点をきめて関係付ける場合に言い換えが可能である。また、原因・条件の接続助詞や事態を対比的に見るような逆接を表す接続助詞や「反面」「一方」の接続助詞も安定した関係情報が得られた。

条件・譲歩、様態、目的、程度、並列を表す接続助詞は、このような安定した関係情報が見つからなかった。とくに因果関係をもったり文の流れが時の流れを維持する場合に、時の流れをさかのぼるような言い換えができない。

4.2 結合標識

名詞節を構文化する接続形態として、「こと」「の」「ところ」「か」をはじめとする形式名詞が存在する。「こと」にともなう格助詞は「ことが」「ことを」「ことは」に限定し、「ことに」「ことで」「ことから」などは安定した関係情報を得ることができないため除外した。「の」は「のを」に限定する。また引用節を構文化する接続形態は「と」「という」「との」に限定した。これらの関係情報を表2に示す。

表2 結合標識の関係情報

種類	結合標識	関係情報	文末表記	
			述語	名詞
名詞節化	こと(が,を,は)	のは	ことだ	だ
	のを	のは	ことだ	だ
引用節化	と	ところによれば		
	との	とは		
	という	とは		

接続助詞や結合表記が依存ネットワークでノードとなったときの関係情報を「こと」を例に図2に示す。図の中の関係情報の2段目,3段目はエントリがそれぞれ述語,名詞のときの文末表記である。

そのがれきの中から新たな歴史が始まることを、人々はしっかりと感じとっている。

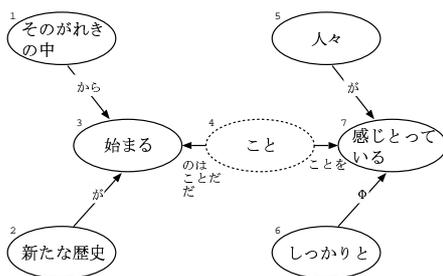


図2 依存ネットワーク上の結合標識

5. 文生成ツールの実装

柔軟な文生成方式⁵⁾で提案されたモデルに基づき、文生成ツールを実装した。全体のプロセスフ

ローを図3に示す。

なおこの文生成ツールではエントリの表層格が「に」格、「から」格が連体修飾されるとき、受身変換で「が」格または「を」格に変換可能なもののみを取り扱った。述部の語尾修正として、テンス並びに活用形については行ったが、アスペクトについては行っていない。

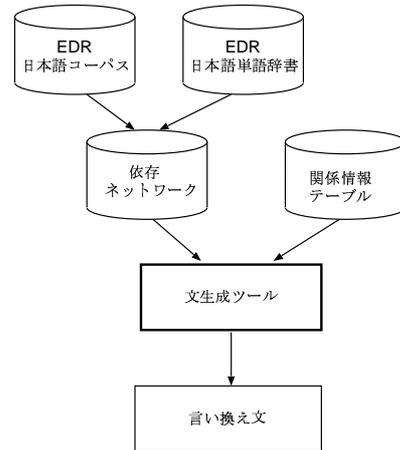


図3 文生成のプロセスフロー

6. 実験と評価

本手法が実際の言い換えにどこまで有効かを調査した。まず実装した文生成ツールで得た言い換え文の中で、従属節に存在する語を文末にした文に対して、成功例と失敗例を紹介する。なお、言い換え文の中での太字は関係表現を示し、< >内は文末に指定した語句(エントリ)である。

6.1 成功例

- (7) 広島に原爆が落とされたとき、彼は近郊に駐屯していた。
- 彼が近郊に駐屯していたのは、原爆が落とされた<広島>のときだった
 - 彼が近郊に駐屯していたのは、広島に落とされた<原爆>のときだった
 - 彼が近郊に駐屯していたのは、広島に原爆が<落とされた>ときだった

- (8) 職員が乗務をすっぱかしたり、飲酒運転したりするから、内部の規律をうんぬんされるんだ
- 内部の規律をうんぬんされるのは乗務をすっぱかしたり飲酒運転したりする<乗務>だからだ
 - 内部の規律をうんぬんされるのは職員がすっぱかしたり飲酒運転したりする<乗務>だからだ
 - 内部の規律をうんぬんされるのは職員が乗務を<すっぱかしたり飲酒運転したりする>からだ
- (9) 株では損の方が大きいけれども、株はもうコリゴリとは私は思わないのです
- 株はもうコリゴリとは私は思わないけれども損の方が大きい<株>
 - 株はもうコリゴリとは私は思わないけれども株では大きい<損の方>
 - 株はもうコリゴリとは私は思わないけれども株では損の方が<大きい>
- (10) そのがれきの中から新たな歴史が始まることを人々はしっかりと感じとっている。
- 人々がしっかりと感じとっているのは新たな歴史が始まる<そのがれきの中>だ
 - 人々がしっかりと感じとっているのはそのがれきの中から始まる<新たな歴史>だ
 - 人々がしっかりと感じとっているのはそのがれきの中から 新たな歴史が<始まる>ことだ
- (11) 実質的な核軍縮は簡単なことではない、と軍縮交渉の研究を 長年 続けてきた前田寿氏という。
- 軍縮交渉の研究を長年続けてきた前田寿氏がいうところによれば簡単なことではない<実質的な核軍縮>
 - 軍縮交渉の研究を長年続けてきた前田寿氏がいうところによれば実質的な核軍縮は<簡単なことではない>
- (12) 警官が一方的に襲撃されたことは、警察が暴力団にナメられた、との批判も 生んでいる。
- 警官が一方的に襲撃されたことが生んでいる批判とは暴力団にナメられた<警察>
 - 警官が一方的に襲撃されたことが生んでいる批判とは警察がナメられた<暴力団>
 - 警官が一方的に襲撃されたことが生んでいる批判とは警察が暴力団に<ナメられた>

6.2 失敗例

- (13) ゲリラは、駆けつけた警官隊と銃撃戦を展開したあと、バスや車などに分乗して逃走した。
*バスや車などに分乗して逃走したのは駆けつけた警官隊と銃撃戦を展開した<ゲリラ>のあとだった
- (13) は自然な文とは言えない。文末が述語の場合はこのような問題は発生していないが、文末が

名詞の時に不自然な文が発生するケースがある。

- (14) ある国民に、その国のことばを話してはいけないということは、民族文化の抹殺に通じる。
民族文化の抹殺に通じるのはその国のことばを話してはいけない<ある国民>だ
- (14) では「あの」「その」という指示詞が文が転置しているため、不自然な文となっている。指示詞についての書き換え規則が必要である。
- (15) イルカが汚染されるということは、人間も同じくらい化学物質漬けになるということでもある。
人間も同じくらい化学物質漬けになるというのは汚染される<イルカ>だ
人間も同じくらい化学物質漬けになるというのはイルカが<汚染される>ことだ
- (15) は「同じくらい」はもともと主節にあったものである。主節に比較を表す表現を含んでいるときは、言い換え後に比較表現が従属節に移動するので不自然な文になる。

- (18) あなたが手伝ってくれたおかげで、早く仕事が片づいて時間の節約になったわ。
早く仕事が片づいて時間の節約になったのは手伝ってくれた<あなた>のおかげだ
早く仕事が片づいて時間の節約になったのはあなたが<手伝ってくれた>おかげだ

終助詞について本研究での文生成ツールの処理対象ではないが、語尾につけることにより、より正確な言い換えが可能となる。

6.3 評価結果

EDR から選択した文例は 125 文である。4 節で掲げた接続詞と結合標識を含んだ文で 3 つ以下の従属節を構成する文を対象とした。

実験にあたっては EDR の構文解析済みの文例に対して、格の復元とノードの単位とする文節のまとめる作業を手作業で行い、依存ネットワーク構造に変換した。

さらに図 3 に示した関係情報テーブルをともに文生成ツールによる言い換え文を生成し、従属節内のエントリノードに対する言い換え文 318 文を得た。この言い換え結果を人手により、成功・失敗を評価した結果を表 3 に示す。

対象とした文例の範囲では言い換え文 318 文中、303 文 (95%) を正常に言い換えることができ

表3 実験結果

接続表記	文例数	言い換え数	成功数	失敗数
とき	5	15	11	4
から	4	9	9	0
あと(に)	4	12	11	1
から	5	11	11	0
ため	11	27	26	1
おかげで	2	5	5	0
ので	6	16	15	1
けれど(も)	6	12	12	0
のに	6	13	13	0
反面	5	10	10	0
一方	4	10	10	0
こと	30	74	69	5
のを	8	19	19	0
と	9	29	26	3
という	9	22	22	0
との	11	34	34	0
合計	125	318	303	15

た．ネットワーク構造を正確に反映していることから，関係情報と文生成順序情報が適切である条件下では，本研究での生成手法は有効に機能していることが確認できた．

7. まとめ

本稿では複文構造における言い換えに必要な関係情報の取得を EDR の例文から調査した．その際「のは強調構文」の文例が多数存在しており，それらが従属節が主節化したものと想定したとき，強調構文を適用しない文に逆変換が可能と仮定して調査を行った．その結果，接続助詞と名詞節を構文化する接続形態，引用節を構文化する接続形態に関する 16 の関係情報を得ることができた．

依存ネットワークに基づいた文生成モデルにもとづいて実装した文生成ツールを使って，EDR の文例の言い換えに適用した．その結果，従属節のエントリノードを文末にした文生成結果で 95% の言い換えに成功した．関係情報を定義できた接続詞または結合標識を含んだ単純な複文に対して有効であることを確認した．

しかし失敗したケースもあり，関係表現の定義以外に語順が入れ替わることによる影響も配慮しなければならないことなどが，この実験結果からわかった．

不自然な文となった失敗例に対して分析と対策を講じるとともに，対象件数を増やし並列を表す接続表現等を含めた関係情報の取得に今後取り組みていきたい．

参考文献

- 1) 乾健太郎. “言語表現を言い換える技術”. 言語処理学会, 第 8 回年次大会チュートリアル, (2002)
- 2) 神田慎哉, 藤田篤, 乾健太郎: “連用主節化に関する規則の追加と追試の洗練”. 人工知能学会, 第 15 回人工知能学会全国大会, 1A1-06 (2001)
- 3) 野上優, 藤田篤, 乾健太郎: “文分割による連体修飾の言い換え”. 言語処理学会, 第 6 回年次大会発表論文集, pp.215-218, (2000)
- 4) 中野滋徳, 足立顕, 牧野武則. “依存構造を利用した言い換え文の自動生成”. 情報処理学会, 研究会報告書, 2004-NL-159, pp.185-191. (2004).
- 5) 足立顕, 牧野武則. “柔軟な文生成方式”. 情報処理学会, 研究会報告書, 2004-NL-160, (2004).
- 6) 山脇拓, 中野滋徳, 足立顕, 牧野武則. “依存ネットワークをもとにしたパラグラフ要約”. 情報処理学会, 研究会報告書, 2004-NL-160, (2004).
- 7) 寺村秀夫: “日本語の文法(下)”. 国立国語研究所, (1981)
- 8) 井上和子: “変形文法と日本語(下)”. 大修館, (1976)
- 9) EDR 電子化辞書使用説明書: “EDR 電子化辞書使用説明書”. (株) 日本電子化辞書研究所, (1993)
- 10) 益岡隆志, 田窪行則. “基礎日本語文法”. くろしお出版, (1992)