

日本語文末表現の言い換え*

本 田 聖 晃 **
 田 辺 利 文 ***
 高 橋 雅 仁 ****
 吉 村 賢 治 **
 首 藤 公 昭 **

Paraphrasing of Japanese Sentence-final Expressions

Takaaki HONDA, Toshifumi TANABE, Masahito TAKAHASHI,
 Kenji YOSHIMURA and Kosho SHUDO

Abstract:

Non-propositional contents (NPCs) is concerned with notions such as tense, aspect, voice, modality, and illocutionary-act, which incorporate temporal, contingent, epistemic or attitudinal information into the propositional content of the natural sentence. In Japanese, NPCs are typically indicated in the V-final position by auxiliaries, ending-particles and various multiword expressions (MWEs). We refer to these as non-propositional content indicators (NPCIs). Each NPCI have some non-propositional functions (NPFs).

On the other hand, it is important for future NLP systems to formulate the semantic similarity of natural language expressions. We have described in (Tanabe et al., 2001) some similarity rules between sequences of NPFs.

The main purpose of this paper is to show the result of a preliminary experiment of the new paraphrasing framework and to discuss the problems left for future work.

Key Words: Natural Language Processing, Multi-Word Expression (MWE), Non-propositional Content, Paraphrasing, Similarity Rules.

1. はじめに

自然言語処理の応用技術の一つである言い換え技術は、翻訳への前処理やテキストの曖昧さの解消などが期待できるため、研究者の注目が高まっている(佐藤, 2001, 乾ら, 2004)。例えば KURA に代表される言い換えエンジ

ンは、各種表現間の言い換えを可能にしている(KURA, 2001)。

一方、自然語文の非命題的内容(Non-Propositional Content; NPC)とは、相、態、様相などの時間的、偶発的、認識的、発話行為的情報を自然語文に添え、発話を現実化するものと考えられている。日本語文においてこれらの情報は、文末の助動詞、終助詞、その他の複数単語表現(Multi-Word Expression; MWE)で表される場合が多い。我々はそれらの表現の辞書を構築しており、さらに、文の非命題的意味構造(NPS)およびその類似性

* 平成19年5月31日受付

** 工学研究科情報・制御システム工学専攻

*** 電子情報工学科

**** 久留米工業大学

等についても既に報告してきた(田辺ら, 2001, Shudo et al., 2004). 本稿では, 日本語文の文末表現の言い換えについて提案する⁽¹⁾. 具体的には, 分かりやすい表現への言い換え, および日本語文末表現における意味の類似性規則について説明する. 類似性規則には, 論理的類似性規則と語用論的類似性規則を設けており, 各々の規則を適用した言い換え処理について実験および考察を行う.

2. 非命題的意味構造(NPS)

2.1 NPCIの概要

日本語文における非命題的内容(NPC)は, 多くの場合, 文の末尾の助動詞, 終助詞その他多種の複単語表現(MWE)により表される. 我々はそれらの表現を, 非命題的内容指示表現(NPCI)と呼び, それらの表現の辞書には約1,450個の複数単語表現⁽²⁾と50個の助動詞, 終助詞を収録している. 我々は網羅性の高さを目標の一つと考えており, 例えば, (森田ら, 1997)の収録表現のうち該当表現は本辞書ですべてカバーされている.

2.2 日本語文の構造の概形

複単語のNPCIを適切に設定することで, 日本語文の構造の大枠を次の生成規則で表現することができる.

- (1) $S_0 \rightarrow BP^* \cdot PRED,$
- (2) $S_i \rightarrow S_{i-1} \cdot NPCI_i, (1 \leq i \leq n)$

ここで S_0 , BP, "*"はそれぞれ骨格文, 文節, 閉包演算子を表す. 以後, $PRED \cdot NPCI_1 \cdot NPCI_2 \cdot \dots \cdot NPCI_n$ を文の述部と呼ぶことにする.

2.3 文の非命題的意味構造(NPS)

自然語文のNPSは, 次のような入れ子型表現で表すことができる.

$$(3) M_m [M_{m-1} \dots [M_2 [M_1 [S]]] \dots],$$

但し, Sは命題的な骨格文, $M_i (1 \leq i \leq m)$, はNPFである. 我々は, 150個程度のNPF, 例えば疑問, 必要性, 推量, 依頼, 起動, 否定, 過去時制などを提案している. これらのNPFは当面, 日常会話程度の日英機械翻訳を目的としたものである. 我々は, NPFを, 時制, 相, 態, 様相, 発語内行為など9つのカテゴリに分類している(S hudo et al., 2004).

NPFの生起順序には言語に依存しない何らかの規則性が存在するという報告がされており(Cinque, 1999, Cinque, 2006), NPSには言語依存性がないと考えられることから, NPSは言い換えや機械翻訳を行う際の間表現として有効であると考えている.

また, それぞれのNPCIは, 何らかの対応するNPFを持つように決められており, (1), (2)に示される構文構造と, (3)に示される意味構造とには一種の同型性があるといえる.

NPCIが文の述部にいくつも並んだ複雑な文末表現の場合でも, NPFとの対応をつけることにより, NPSを求めることが可能であるといえる. このように(3)は構造のシンプルさと同時に対応可能な表現の多様さの点で工学的に重要な性質を示していると考えている.

3. 非命題的意味構造(NPS)間の類似性規則

NPS間の類似性規則に関しては既に報告している(田辺ら, 2001, Shudo et al., 2004). 類似性規則は, 論理的類似性規則および語用論的類似性規則に大別される. それらの規則と適用例を表1に示す⁽³⁾. 我々はこれまでに, このような類似性規則を41個設けている.

NPS間の類似性規則には, 条件によっては適用できないものもある. 例えば, 「守ら/なければなりません」

表1 類似性規則の例

論理的類似性規則	
類似性規則	規則の適用例
否定 ₁ [否定 ₁ [S]] ⇔ [S]	「行か/ない/ことはない」 ⇔ 「行く」
否定 ₁ [必要性 ₁ [S]] ⇔ 許容 ₂ [否定 ₁ [S]]	「食べ/なければならぬ/ことはない」 ⇔ 「食べ/なく/てもよい」
推量 ₄ [否定 ₁ [S]] ⇔ 否定 ₁ [推量 ₁ [S]]	「読ま/ない/かもしれない」 ⇔ 「読む/に違いない/とはいえない」
過去時制[許容 ₁ [否定 ₁ [S]]] ⇔ 過去時制[否定 ₁ [必要性 ₂ [S]]]	「書か/なく/てよかつ/た」 ⇔ 「書く/必要がなかつ/た」
語用論的類似性規則	
類似性規則	規則の適用例
願望[受動態 ₃ [S]] ⇔ 依頼 ₁ [S]	「見て/いただき/たい」 ⇔ 「見て/ください」
可能性 ₂ [S] ⇔ 許容 ₁ [S]	「入る/ことができる」 ⇔ 「入っ/てもよい」
疑問 ₁ [否定 ₁ [可能性 ₂ [S]]] ⇔ 依頼 ₂ [S]	「行く/ことが出来/ない/か」 ⇔ 「行って/ください」
疑問 ₁ [否定 ₁ [S]] ⇔ 勧誘[S]	「見/ない/か」 ⇔ 「見よ/う」

≡「守り/なさい」に対する規則、

必要性₁[S] ≡ 命令[S]

は、補文の動作主が2人称でなければ適用できない。このような条件についてはまだ明確に定義しておらず、厳密には、各類似性規則における適用条件を定義する必要がある。

4. 実験

4.1 実験手順

日本語文末表現をNPSに変換し、NPS間の類似性規則の適用後、各NPFを代表表現に変換する言い換え実験を行う。

代表表現については、人手により各NPFに対応する日本語代表表現を選定し、辞書を作成した(表2)。表2では、NPFと対応する代表表現だけを抜粋したが、実際に辞書に表記されている情報としてはこの他に、活用接続に関する情報などがある。

表2 NPFの代表表現辞書の一部

NPF	代表表現
疑問 ₁	のか
依頼 ₁	てください
願望	ないかな
否定 ₁	ない
必要性 ₁	なければならない
必要性 ₂	ひつようがある
推量 ₁	にちがいない
推量 ₂	はず
受動態 ₁	れる
受動態 ₂	てくれる
受動態 ₃	てもらう

4.2 実験結果

ブログ、小説等からランダムに抽出した651文に対して、分かち書きシステム、NPS構築システムにより正しく出力されたNPSは572個であった。1述部に対して正解NPSは1つしか存在しないため、言い換えシステムへの入力となる述部は572個である。

この572個の述部に対して言い換え処理を行った。その結果、類似性規則を適用できた述部は63個、適用できなかった述部は509個であった。

類似性規則を適用できた述部63個について正誤判定を行ったところ、正解述部数は48個、不正解述部数は15個であった。また、類似性規則を適用できなかった述部509個について正誤判定を行ったところ、正解述部数は228個、不正解述部数は139個であった。ここで、残った述部142個に関しては入力と出力が同じであったため評価の対象外としている。

また、今回の実験で生じた意味関数のうち、代表表現での出現頻度が最も高い意味関数の割合は約60%であったため、代表表現の選定方法が妥当であったのではないかと考えられる。

4.3 考察

正解例、不正解例を元に今回の実験の考察を行う。

<正解例1>

入力：言わ/なかつ/た/ことが/ない

↓

NPS：否定₁[頻度₃[過去時制[否定₁[言う]]]]

↓(NPS類似性規則の適用)

NPS：頻度₁[過去時制[言う]]

↓(代表表現への変換)

出力：言っ/た/ことしかない

適用した規則：

否定₁[頻度₃[過去時制[否定₁[S]]]] ≡ 頻度₁[過去時制[S]]

正解例1は、NPS類似性規則を適用できた言い換える例である。この例では、論理的規則である「否定₁[頻度₃[過去時制[否定₁[S]]]] ≡ 頻度₁[過去時制[S]]」を適用したが、入力文よりも出力文がわかりやすい文であると考えられるため正解であると判定した。

<正解例2>

入力：(そんなこと口)に出す/んじゃありません

↓

NPS：禁止[出す]

↓(代表表現への変換)

出力：(そんなこと口)出し/てはいけません

適用した規則：なし

正解例2は、NPS類似性規則を適用できなかった例である。この例では入力文から生成されるNPF「禁止」を代表表現に変換する処理のみを行っているが、入力文よりも出力文が分かりやすい文であると考えられるため正解であると判定した。

<不正解例>

入力：とらえる/ことができ/なくなつ/てい/た/のか

↓

NPS：疑問₁[過去時制[状況[変化₁[否定₁[可能性₃₁[とらえる]]]]]]

↓(NPS類似性規則の適用)

NPS：疑問₁[過去時制[状況[変化₁[禁止[とらえる]]]]]]

↓(代表表現への変換)

出力：とらえ/てはいけません/になつ/てい/た/のか

適用した規則：否定₁[可能性₃₁[S]] ≡ 禁止[S]

これは、文の体裁が保たれていないため不正解であると判定した例である。この場合「変化₁[禁止[S]]」というNPSを代表表現に変換すると、「S/てはいけません/

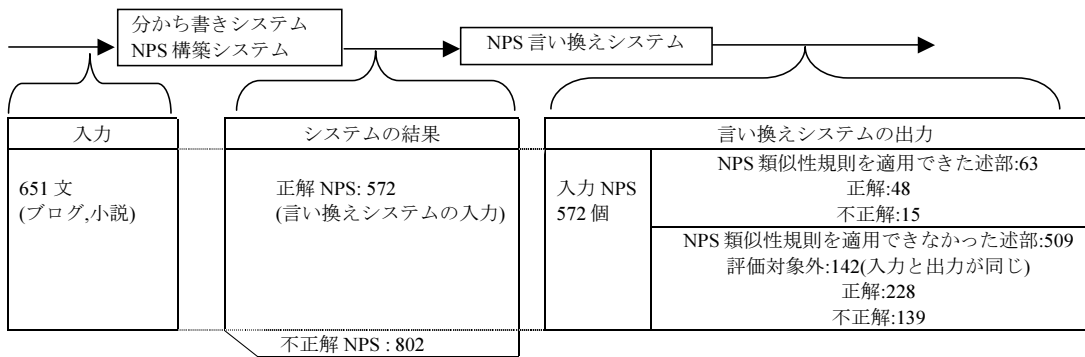


図1 実験の流れおよび評価の概略

になる」となり、体裁を保っていない文が出力された。不正解であると判定した154個(=15+139)は、このように文の体裁を保っていないものや、文自体が自動生成できなかったものである。

対処法として、2つの方法が考えられる。1つ目は、対応する代表表現が接続できないような複数のNPFが連続して存在した場合、それらをまとめて1つの表現に変換する規則を設ける方法である。不正解例におけるNPS「変化:[禁止[S]]」は「S/てはいけな/くなる」と変換する。この方法は、出力された文を人間が判断する必要があり、規則作成のコストが大きく、あまり望ましい方法ではないと思われる。

2つ目は、NPFに対応する代表表現を複数設ける方法である。選定した代表表現は、「禁止:てはいけません」「変化:になる」であった。文脈に応じ、「禁止:てはいけない」「変化:くなる」を追加すればNPS「変化:[禁止[S]]」は「S/てはいけな/くなる」と正しく生成できる。この方法では、文脈に応じ、どの代表表現を使うべきかの情報が必要になり、かつ、情報を用いた場合の処理コストが大きくなるなどが予想される。

これらの問題については今後の課題とする。

5. おわりに

自然語文の非命題的内容は、対話理解、文脈モデル、話者の態度の推定などの自然言語処理で重要な役割を果たす。本稿では、日本語文末表現において非命題の意味構造間の類似性規則を適用する言い換え実験について報告した。

今後の課題としては、NPCIやNPF、NPS間の類似性規則の網羅性の向上や、NPS間の類似性規則を用いた多言語間の翻訳などの応用処理などが考えられる。また、代表表現の選定方法についての考慮も必要ではないかと考えられる。本稿では人手により選定し、ある程度

妥当であると考えられたが、“短い表現”や“硬い表現”などの明確な基準での選定を行い、利用者の用途によって選択できるように改善できるのではないかと考えられる。他に、第3章で述べているが、NPS間の類似性規則には条件によっては適用できないものもある。適用条件についてはまだ明確に定義しておらず本稿では考慮していないため、対処すべき課題である。また、適用できた類似性規則には局所性がみられた。これは、ブログや小説を入力としたことが原因となっていると考えられる。そのため、規則としては存在するが適用されなかった類似性規則もあり実際に適用された類似性規則は、約27%(=11/41)であった。類似性規則そのものの妥当性を検証するためには、様々な入力を考える必要もある。

謝 辞

本論文に対する、松吉俊氏(京都大学)、藤田篤氏(名古屋大学)、竹内孔一氏(岡山大学)の有益なコメントに感謝する。

参 考 文 献

- 佐藤理史.2001.なぜ言い換え/パラフレーズを研究するのか,言語処理学会第7回年次大会.ワークショップ論文集.
- 乾健太郎,藤田篤.2004.言い換え技術に関する研究動向,自然言語処理
- KURA(蔵).2001.言い換えエンジン.
<http://cl.naist.jp/kura/doc/>
- 田辺利文,吉村賢治,首藤公昭.2001.日本語モダリティ表現とその言い換え,言語処理学会第7回年次大会.ワークショップ論文集.
- Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi and Kenji Yoshimura. 2004. *MWEs as Non-propositional Content Indicators*. The Proc. of

- the ACL2004 Workshop on Multiword Expressions: Integrating Processing: 32-39.
- Guglielmo Cinque. 1999. Adverbs and Functional Heads. OXFORD UNIVERSITY PRESS.
- Guglielmo Cinque. 2006. Restructuring and Functional Heads. OXFORD UNIVERSITY PRESS.
- Satoshi Shirai, Satoru Ikehara and Tsukasa Kawaoka. 1993. *Effects of Automatic Rewriting of Source Language within a Japanese to English MT System*. Fifth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-93: 226-239
- 松吉俊, 佐藤理史. 2007. 体系的機能表現辞書に基づく日本語機能表現の言い換え, 言語処理学会第13回年次大会
- 田辺利文, 本田聖晃, 高橋雅仁, 小山泰男, 吉村賢治, 首藤公昭. 2006. 日本語文末表現の取り扱いについて, FIT2006第5回情報科学技術フォーラム.
- Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. 2002. *Multiword Expressions: A Pain in the Neck for NLP*. The Proc. of the 3rd CICLING: 1-15.
- Fellbaum, Christine, ed.: 1998. *WordNet. An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- 森田良行, 松木正恵. 1989. 日本語表現文型用例中心・複合辞の意味と用法, アルク.
- chasen(茶筌). 1996. 形態素解析システム茶筌. <http://chasen.naist.jp/hiki/ChaSen/>

注

- (1) 関連研究として(Shirai et al., 1993)は, 日英翻訳システムの前処理としての言い換え処理のためにNPCIを取り扱っているが, その網羅性および意味は論文中には明記されていない. また, (松吉ら, 2007)は, 助詞, 助動詞, 格助詞, 接続助詞相当の複数単語表現を機能表現として機能表現同士の言い換えを行っているが, 複数の機能表現列の言い換えは行っていない.
- (2) 日本語の述部において調査した結果, 複数単語表現としてのNPCIは, NPCI全体の42%生起していることを明らかにした(田辺ら, 2006). (Sag et al., 2002)に述べられている, WordNet1.7 (Fellbaum, 1999)における英語見出しの複数単語表現の生起の割合が41%であったこととの類似性は興味深い.
- (3) NPFにおける, 程度などの細かい意味の違いは下付き文字にて区別している.