

## 語彙・構文的言い換えにおける変換誤りの分析

藤田 篤<sup>†</sup> 乾 健太郎<sup>†</sup>

本論文では、世界知識や発話の文脈に関係なく言語学的知識で実現可能な語彙・構文的言い換えを対象とし、これらを構文トランスファ方式で実現する際に生じる変換誤りについて述べる。本研究では、次の2段階の調査を行った。(1) まず、先行研究で示されている知識・知見に基づいて構文トランスファによって生じる変換誤りを洗い出し、その種類を整理・分類し、(2) 次に、さまざまな種類の構文トランスファ規則を用いて言い換え事例を生成し、事例分析に基づいて、各種変換誤りが生じる傾向を調査した。これらの調査によって、多くの変換誤りが、構文トランスファ規則の種類に依存することなく生じることが明らかになった。また、問題の性質を分析し、解消に向けての課題を整理した結果、まずは頻繁に生じる傾向がある形態的な不適格性、および構文的な不適格性を優先的に解消すべきであるという結論を得た。

## Exploring Transfer Errors in Lexical and Structural Paraphrasing

ATSUSHI FUJITA<sup>†</sup> and KENTARO INUI<sup>†</sup>

In lexical and structural paraphrasing, meaning-preserving linguistic transformations are performed such as lexical or phrasal replacements and alternations in case, voice. In this paper, we report the results of our investigation into transfer errors, which reveals: (1) what types of errors tend to occur in generating lexical and structural paraphrases of Japanese sentences, and (2) which of them should be given preference as the subject of further research. We found that most types of errors occurred irrespective of the types of transfer. The sorts of lexical errors and syntactic (collocative) errors should be tackled firstly, since they not only occurred more frequently, but also seems to be solved by maintaining revision patterns or utilizing statistical language models.

### 1. はじめに

言い換えは、ある言語表現を、意味や内容を保持したまま同一言語の別の表現に変換する処理である。言い換えの工学的研究は、機械翻訳などの自然言語処理アプリケーションにおける部分タスクとして、個別的に導入する形で始まった。機械翻訳や手話生成では、入力文を機械処理にしやすい(翻訳しやすい、あるいは手話で表現可能な)表現に言い換えることによって訳質の向上が図られた<sup>37),40)</sup>。また、情報検索や質問応答においても、質問や、検索対象の文書中の表現を言い換えることで、よりの確な回答を発見することができるようになる<sup>24),25)</sup>。このように、多くは言語表現の多様性を吸収するために、機械処理の前処理として用いられてきた。一方、高齢者や幼児、外国人、障害者など、利用者の言語能力に合わせて言い換えれば、それ自体が読解支援アプリケーションとなる<sup>5),15)</sup>。さ

らに、読みやすい文章を書く、スタイルを統一する、あるいは規定の語彙・構文(制限言語<sup>28)</sup>)を使って文書を書くといった推敲作業を支援する場合でも、読みにくい文や制限言語に合わない文を自動的に言い換える技術が必要である。このように、言い換えは近年、自然言語処理の諸分野にわたって重要な要素技術ととらえられるようになってきた<sup>1),11),16),30)</sup>。

我々は、さまざまな種類の言い換えるのうち、例(1)、(2)のように、世界知識や発話の文脈に関係なく、言語学的知識で実現できる語彙・構文的言い換えの実現を目指している。

- (1) s. 激しい自動車戦争に進む公算が大きい。  
t. 激しい自動車戦争に進む可能性が大きい。
- (2) s. 先頭との距離を縮める。  
t. 先頭との距離が縮まる。

例(1)や(2)の言い換えは、次の(3)のような言い換えパターンを用意すれば実現できると考えられる。

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科  
Graduate School of Information Science, Nara Institute  
of Science and Technology

s は言い換え前の文, t は言い換え後の文, t' は t を修正した文を表す。  
N, V はそれぞれ名詞句, 動詞句を表す変数とする。

- (3) a.  $N \Rightarrow N'$  ( $N'$  は  $N$  の同義語)  
 b.  $N$  を  $V$  ( $V$  は他動詞)  $\Rightarrow N$  が  $V'$  ( $V'$  は  $V$  に対応する自動詞)

ただし、このような言い換えパターンを用いた語彙・構文レベルの構造変換（構文トランスファ）方式でテキスト中の一部の表現を同義の表現に置換するだけでは、言い換えた箇所と周辺の表現とのつながりが不自然になったり、言い換え前の意味が変わったりする可能性がある。たとえば、例文 (4.s) に (4.r) の言い換えパターンを適用すると、(4.t) のような不適格な文が生成される。

- (4) r.  $N$  しか  $V$  ない  $\Rightarrow N$  だけ  $V$   
 s. ルイ・ヴィトンカップの優勝艇しかアメリカズカップに出場できない。  
 t. \*ルイ・ヴィトンカップの優勝艇だけアメリカズカップに出場でき。

言語的に適格で、かつ言い換え前と同じ意味のテキストを生成するためには、たとえば次のような修正を施す必要がある。

修正 1. 「でき」の活用形を基本形（終止形）に変える ((4.t)  $\Rightarrow$  (5)).

- (5) ?ルイ・ヴィトンカップの優勝艇だけアメリカズカップに出場できる。

修正 2. 副助詞「しか」は格助詞「が」とは共起できないが、副助詞「だけ」の場合は「が」と共起できるため、明示する方が自然である ((5)  $\Rightarrow$  (6)).

- (6) ルイ・ヴィトンカップの優勝艇だけがアメリカズカップに出場できる。

ところが、語彙・構文的言い換えのためにどのような修正処理を実現すべきか、すなわち、どのような変換誤りがどのような傾向で生じるかについては、包括的な調査もなく、不明なままである。

以上より、本論文では、

さまざまな種類の語彙・構文的言い換えを構文トランスファ方式で実現する際に、どのような種類の変換誤りが、構文トランスファ規則の種類に対してどのような傾向で生じるかを経験的に明らかにし、

変換誤りを修正する技術を研究する際に、どの種類の誤りから優先的に取り組むべきかを論じる。語彙・構文的言い換えにともなう変換誤り・修正処理については、特定の種類の言い換えの実現方法に関する先行研究の中に参考にするべき知見を見つけることができる<sup>12),18),21),22),35),41)</sup>。そこで、本研究では、次の 2 段階の調査を行った。

調査 1. まず、先行研究で示されている知識・知見に

基づいて構文トランスファによって生じる変換誤りを洗い出し、種類を整理・分類する。

調査 2. 次に、さまざまな種類の構文トランスファ規則を用いて言い換え事例を生成し、事例分析に基づいて、各種変換誤りが生じる傾向を調査する。以下、まず、本論文が対象とする修正処理の位置付けについて 2 章で述べる。次に、変換誤りの分類（調査 1）について 3 章で、変換誤りの分布の傾向に関する調査（調査 2）について 4 章で述べ、各変換誤りがどのようなレベルで不適格であるととらえられるかについて 5 章で議論する。6 章で関連研究に言及し、最後に 7 章でまとめる。

## 2. 言い換えの生成モデル

言い換えは、その目的や入出力を一般化すると、図 1 のようにモデル化することができる。(i) のインタフェースレイヤは、表層テキストと計算機の内部表現の間の変換や、入出力メディアの違いを吸収する役割を担う。(ii) の目的適合性の評価レイヤは、アプリケーションに依存した処理を担う。たとえば読解支援を目的としてテキストを単純化するために言い換えるのであれば、入力側では、テキストのどの部分がユーザにとって読みにくいのかを検出し、どう言い換えるかをプランニングする。一方、出力側では、言い換え結果がユーザにとって読解可能になっているかどうかを評価する。これらのレイヤに対し、(iii) の言い換えレイヤは、目的に依存しない言い換え機構を指す。このような設計は、テキストの評価基準を取り換えることによってさまざまな用途の違いを吸収できる、より

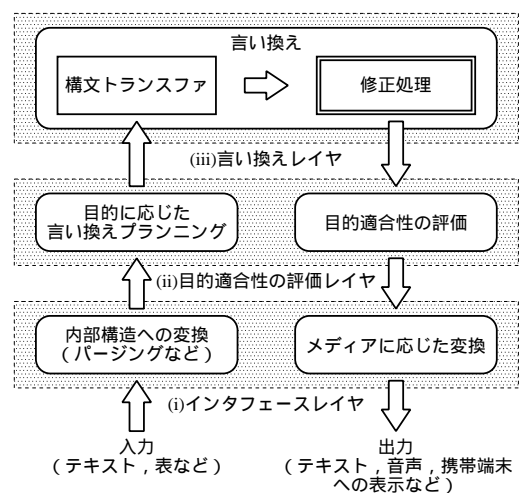


図 1 メディア、目的を考慮した言い換えるモデル

Fig. 1 A generic model of paraphrasing.

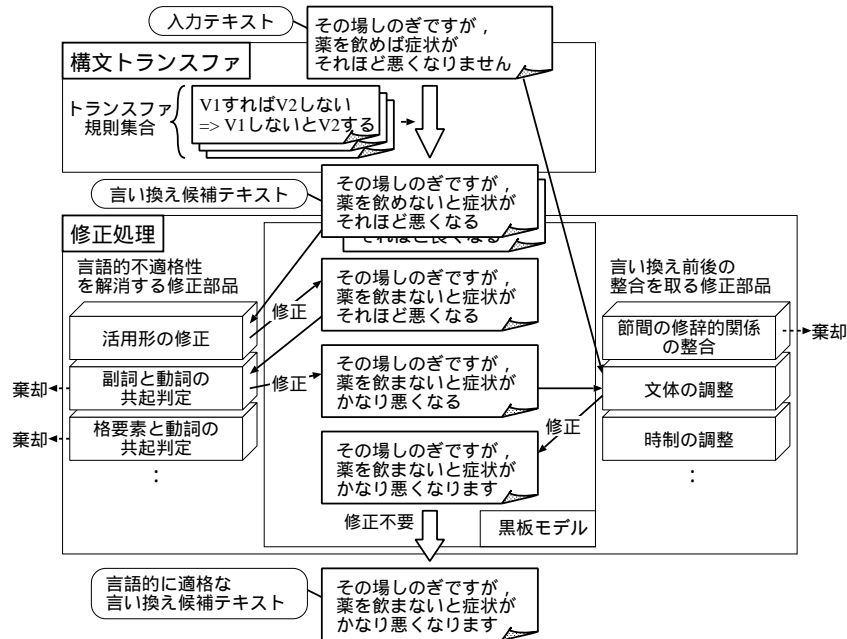


図 2 構文トランスファと修正処理に基づく言い換えの生成モデル

Fig. 2 A paraphrasing model bases on syntactic transfer and revision.

汎用的な枠組みを提供することができる。本論文では、さまざまなアプリケーションに応用可能な (iii) の言い換えレイヤの実現を目的とし、とくに修正処理に着目する。

機械翻訳の研究においては、構文レベルから中間言語に近い意味レベルまで、どのレベルで構造変換を行うべきかについてさまざまな議論があった。たとえば文献 6) では、意味レベルで構造変換する手法(意味トランスファ)を用いた機械翻訳は、言語に依存した語彙・構文情報を捨象でき、構造変換の負荷が軽くなるという利点について述べている。これに対し、単言語内翻訳である言い換えには、次の 2 つの特徴がある。

- 原文の語彙・構文構造を特定の構造に変換すること自体が目的になる場合がある。したがって、意味トランスファのように、入力の解析過程で単語の表層や構文の情報を過度に捨象することは望ましくない。
- 原文の一部のみを変換するため、一部の意味情報しか必要としない場合も多い。したがって、文全体にわたる意味解析は不要である。

これらをふまえ、本論文では、語彙・構文レベルの構造変換(構文トランスファ)方式を採用する。

我々が考える言い換えの生成モデルを図 2 に示す。このモデルでは、構文トランスファと修正処理は次のような役割を果たす。

**構文トランスファ:** ある適格な表現に対し、構文トランスファ規則を適用し、言い換え候補を生成する処理

**修正処理:** 各言い換え候補を評価し、不適格な表現を含む候補を修正、または棄却する処理

### 3. 調査 1: 変換誤りの分類

まず、先行研究で示されている言い換え知識に基づいて変換誤りを洗い出し、種類を整理・分類した。調査の手順は次のとおりである。なお、言い換え事例の生成には、構文トランスファ方式に基づく日本語言い換えエンジン KURA<sup>39)</sup>を用いた。

**手順 1. 変換誤りとそれを修正・棄却するためのパターンの収集:** 2 章で述べた構文トランスファと修正処理の役割分担に基づいて、先行研究で示されている言い換え知識を、言い換え候補を生成するためのパターン(トランスファ規則)と変換誤りを修正・棄却するためのパターン(修正・棄却パターン)に分解する。

**手順 2. 変換誤りとその修正・棄却処理の分類:** 収集した個々の修正・棄却パターンは、なんらかの変換誤りを含む表現とその誤りをなくした表現の対である。ここで、修正規則を構文パターンに基づいて分類し、種類を整理・分類する。

### 3.1 先行研究に基づく調査

日本語を対象とした語彙・構文的言い換えに関しては、特定の種類の言い換えを対象とした事例研究がすでにいくつか報告されている<sup>1</sup>。たとえば、文献 21) では、例 (7) のような言い換えを実現するための知識として、(8) のアルゴリズムが示されている。

<サ> 「サ変名詞+する」から動詞相当句への言い換え<sup>21)</sup>

- (7) s. 犯人は警察に逮捕された。  
t. 犯人は警察に捕まえられた。
- (8) サ変名詞 + さ + れる + 0 個以上の接尾辞 ⇒
1. 「サ変名詞 + さ」を動詞相当句に置き換える。
  2. 動詞相当句の活用形を「さ」に一致させる。
  3. 動詞相当句の主辞がサ変動詞か子音動詞なら「れる」、それ以外なら「られる」を接続させる。
  4. 「れる」(もしくは「られる」)の活用形もとの「れる」に一致させる。
  5. 接尾辞をコピーする。

また、文献 22) では、例 (9) のような言い換えを実現するために、(10) の言い換えパターンが示されている。

<格> 格交替(態・使役の交替, 動詞交替)<sup>22)</sup>

- (9) s. 悪天候が米国の多くの地域を見舞った。  
t. 米国の多くの地域が悪天候に見舞われた。
- (10) N1 が N2 を V する ⇒ N2 が N1 に V される。  
制約 a 名詞句の主辞 N2 は有情,  
制約 b 動詞 V は受身可,  
制約 c 動詞 V はヲ格を必須格とする,  
制約 d 動詞 V は二格を非必須格とする。

例 (8), (10) のように、先行研究で示されている言い換え知識には、変換誤りが生じるのを未然に防ぐため、あるいは変換誤りが生じた場合にこれを修正するための処理や制約が含まれている。ただし、それらは多くの場合、特定の種類の言い換えに固有の処理や制約ではない。たとえば、(8) のステップ 2 は例 (11) のような修正のための手続きであるが、この処理は、1 章で示した例文 (4.t) から例文 (5) への修正と同じ、活用形の修正処理である。

- (11) t. \*犯人は警察に捕まえるられた。  
t'. 犯人は警察に捕まえられた。

同様に、動詞に合わせて機能語を選択する (8) のステップ 3 のような処理も、しばしば必要となることが

予想できる。

このように、言い換えにともなう変換誤りとその修正処理については、先行研究においていくつかの考察が見られる<sup>12),18),21),22),35),41)</sup> もの、修正に関する知識の共有は容易ではない。しかし、特定の種類の言い換えに固有でない修正処理は、言い換えの生成とは切り離して共有すべきである。そうすれば、新しい種類の言い換えの事例研究や、機械的に獲得したトランスファ規則を用いる場合<sup>2)</sup>でも、人手で規則を洗練するコストを減らすことができると期待できる。

我々は、個別の言い換えの事例研究で得られた知識・知見をもとに、変換誤りの種類を調査・整理した。変換誤りには、語彙・構文レベルから意味・談話レベルまで性質の異なるさまざまなものが含まれると考えられるので、変換誤りを細かく分解して整理した。この細分化によって、以下にあげる効果が期待できる。

- 各変換誤りを修正する手続きを、異なる種類のトランスファ規則の種類、あるいは個々のトランスファ規則の間で共有できる可能性が高くなる。
- 修正に必要な知識が単純になり、限られた量のコーパス、あるいは非パラレルコーパスからでも、必要十分な語彙知識を獲得できる可能性がある。

### 3.2 リソース

次に示す 4 つの先行研究で示されている知識を参照した。すなわち、

<サ> 「サ変名詞+する」から動詞相当句への言い換え<sup>21)</sup>

<格> 格交替(態・使役の交替, 動詞交替)<sup>22)</sup>

<否> 否定表現の言い換え<sup>41)</sup>

<機> 機能語相当表現の言い換え<sup>12)</sup>

ここで、<サ> は例 (7) で、<格> は例 (9) で示した種類の言い換えである。作成された言い換え知識が詳しく示されており、実験環境への導入が容易であったため採用した<sup>3)</sup>。一方、<否> と <機> は KURA を用いて行われた言い換えの事例研究であり、KURA とともに言い換え知識が公開されている<sup>4)</sup>。以下、各研究および知識について簡単に述べる。また、今回用いた言い換え知識、および例文の規模を表 1 に示す。詳細については各文献を参照されたい。

<サ> 「サ変名詞+する」から動詞相当句への言い

<sup>1</sup> 単語から他の語句への言い換え<sup>9),18),21)</sup>, 動詞交替<sup>17),22)</sup>, 格交替<sup>22)</sup>, 複文の分割<sup>23),31)</sup> など。

<sup>2</sup> 近年、人手による複数の翻訳結果や異なる新聞社の記事を一種の言い換えコーパスと見なし、機械翻訳で用いられたテキストアライメントの技術<sup>27)</sup> を用いて、大規模な言い換えパターンを機械的に獲得する試みも報告されるようになってきた<sup>4),36)</sup>。

<sup>3</sup> 例文については文献 21), 22) に掲載されているもののみを用いた。

<sup>4</sup> <http://cl.aist-nara.ac.jp/lab/kura/doc/>

表 1 変換誤りの調査に用いた言い換え知識と例文  
Table 1 Knowledge resources for error exploration.

言い換の種類	言い換え知識の形式, 規模	調査に用いた例文の種類, 規模	文献
<サ>	言い換え規則 (手続き) 5 個, 規則選択規則 4 個	「通信に関する現状報告」中の 22 文	21)
<格>	言い換えパターン 40 個	IPA 単文コーパス中の 25 文	22)
<否>	トランスファ規則 291 個, 修正・棄却規則 30 個	京大コーパス中の否定辞「ない」を含む 500 文	41)
<機>	トランスファ規則 261 個, 修正・棄却規則 42 個	3 種の日本語学習教材中の例文 185 文	12)

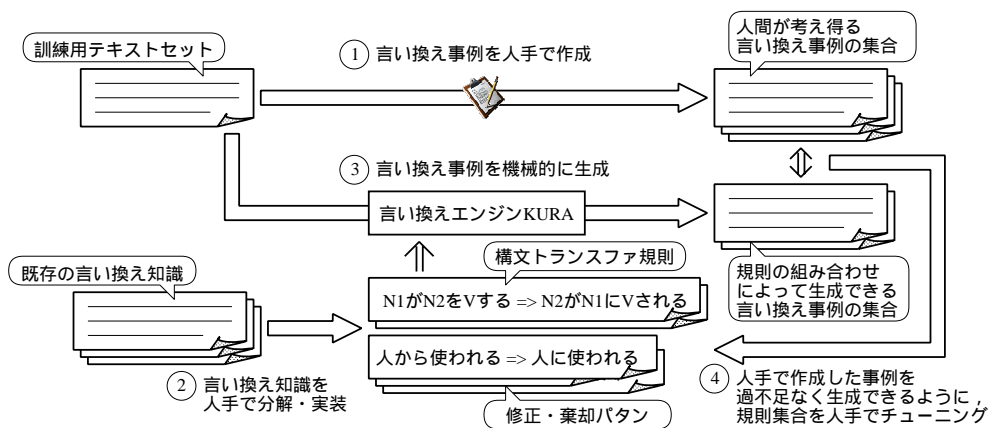


図 3 修正・棄却パターンの収集

Fig. 3 Knowledge decomposition scheme for exploring revision patterns.

#### 換え

文献 21) では, (8) のような言い換え規則と, 適用する言い換え規則を選択するための規則適用アルゴリズムが示されている. 規則適用アルゴリズムは, 「サ変名詞+する」に対応する動詞相当句の主辞が自動詞であるか他動詞であるか, 「サ変名詞+する」に後続する接尾辞の種類などによって, 適用する言い換え規則をただ 1 つ選択する.

<格> 格交替 (態・使役の交替, 動詞交替)

文献 22) では, 単文中の指定した格要素をガ格化するタスクに取り組んでいる. 指定した格要素をガ格化するために, (10) のような格交替の言い換え規則を繰り返し適用している.

<否> 否定表現の言い換え

- (12) s. 太郎しか合格点を超えなかった.  
t. 太郎だけが合格点を超えた.

文献 41) は, 例 (12) のような, 否定辞「ない」と取り立ての意味を持つ副助詞 (「しか」「だけ」など) との呼応に着目している. 新聞記事中の文から人手で作成した事例の分析をもとに, トランスファ規則と修正規則を KURA に実装している.

<機> 機能語相当表現の言い換え

- (13) s. 洪水のごとき宣伝によるところも大きい.  
t. 洪水のような宣伝によるところも大きい.  
機能語相当表現とは, いわゆる「表現文型」とい

う術語で一括りにされる助詞・助動詞相当表現である. 文献 12) では, 日本語能力試験出題基準に準拠した日本語学習教材に掲載されている解説や例文に基づいて, 例 (13) のような言い換えを実現するためのトランスファ規則と修正規則を KURA に実装している.

#### 3.3 手順 1: 修正・棄却パターンの収集

まず, 先行研究で示されている言い換え知識を分解し, 修正・棄却パターンを収集する. 収集の手順を図 3 に示し, 以下, 説明する.

(Step 1) 表 1 の例文集 (訓練セット) に対して可能な言い換え事例の集合を人手で作成する.

(Step 2) 2 章で述べた構文トランスファと修正処理の役割分担に基づき, 先行研究で示されている言い換え知識を, 言い換えを生成するためのパターン (トランスファ規則) と変換誤りを修正・棄却するためのパターン (修正・棄却パターン) に分解する. そして, 修正・棄却パターンは, 訓練セットの各文に対して正しく適用されるように, 例文中の単語を用いて図 3 中央下のような形式で KURA に実装する.

(Step 3) 訓練セットに対し, KURA によってパター

両者を可能な限り細かく分けるが, これは変換誤りを修正する処理を洗い出すためであり, 規則に対する注釈付けを否定するものではない.

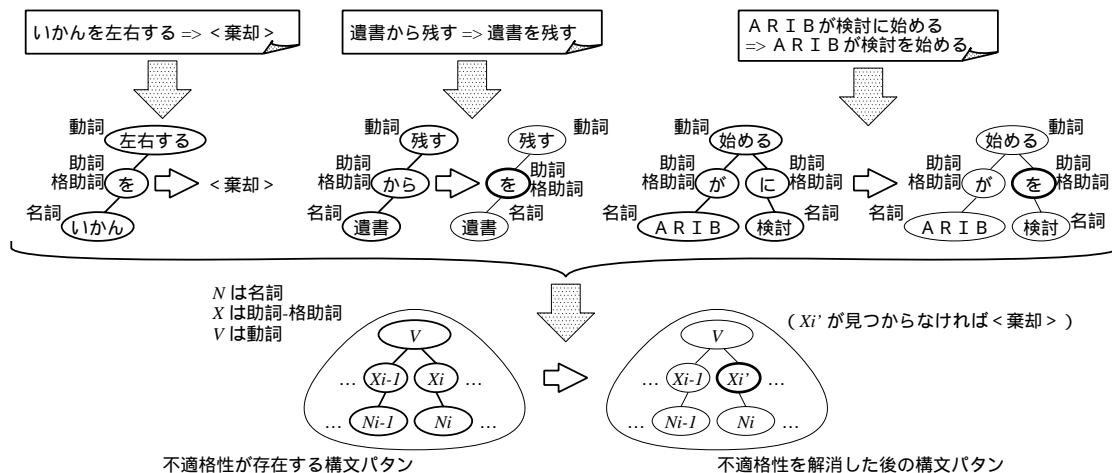


図 4 修正・棄却パターンの抽象化

Fig. 4 Abstraction of revision patterns.

ンを網羅的に組み合わせて適用し、言い換え事例を機械的に生成する。

(Step 4) Step 1 で作成した言い換え事例を過不足なく生成できるように構文トランスファ規則に対して修正・棄却パターンが示す制約を注釈として付与する。

(Step 5) Step 3, 4 を繰り返して、構文トランスファに依存しない修正・棄却パターンを得る。

Step 2 では、たとえば、(10) の規則のうち、構文パターンの対のみを構文トランスファ規則とし、名詞や動詞の各々に関する制約 a ~ d は、変換誤りを含む文が生成されるのを防ぐためのパターンと見なす。

このとき、Step 4 の作業はたとえば次のようになる。言い換え規則 (10) の制約 c と制約 d は、構文トランスファ規則を適用する際の、表層的な必須格、任意格に関する制約である。(10) の受動態への言い換えのように表層格が能動文における標準的な格から別の格に変化する場合、言い換え後のテキストにおいてこれらの条件を評価することはできない。すなわち、構文トランスファの制約としなければならない。一方、制約 a (ガ格の格要素が動作主としての役割を担うことができるか否か) や制約 b (動詞が受身を表す接尾辞「られる」(あるいは「れる」) を後続できるか否か) は、構文トランスファとは関係なく、テキストが適格か不適格かを評価するための制約と見なせる。

ただし、制約 a は、態交替、動詞の自他交替など可能な言い換えの中から、最適な言い換えを選択する目的のために導入されたものである。我々は、2 章で述べたように、可能な言い換えの中から最適な 1 つを選択するタスクは、言い換元の生成とは独立して取り組むべきと考えているため、この知識は扱わない。

### 3.4 手順 2: 変換誤りと修正・棄却処理の分類

テキスト中のある表現が適格であるか否かは、多くの場合個別の単語に依存する。このため、手順 1 では、修正・棄却パターンが訓練セットに対して正しく適用されるように単語を埋め込んであった。手順 2 では、手順 1 で得られた変換誤り、およびその修正例を構文パターンに基づいて分類し、種類を分類・整理した。構文パターンとしては、KURA の内部表現である形態素をノードとする依存構造を用いた(図 4 参照)。各形態素の情報は、形態素解析器『茶釜』<sup>26)</sup>、および日本語辞書 IPADIC<sup>3)</sup> に従っている。

以上の作業の結果、変換誤りとそれを直したパターンは 14 種類に分類できた。これらはおおよそ、形態的な不適格性、構文的な不適格性、意味的な不適格性、談話的な不適格性による誤りからなっている。各々の詳細については、5 章で例をあげて議論する。

## 4. 調査 2: 変換誤りが生じる傾向の調査

3 章において変換誤りの種類を洗い出したが、各誤りがどれだけ深刻であるか、どの誤りから優先的に解決すべきかについては見通しを立てることはできない。そこで、次に、変換誤りの分類に用いたトランスファ規則の集合に別の種類の規則集合を加え、新しいテキスト集合(評価セット)を用いて各変換誤りが生じる傾向を調査した。

### 4.1 言い換え事例の生成

さまざまな種類の構文トランスファ規則(約 28,000 規則)を言い換えエンジン KURA 上に実装し、京大コーパス 1 月 9 日分の 1,220 文に対して適用した結果、言い換え事例 680 件を得た。規則、および言い換

表 2 構文トランスファの種類と規則および事例の規模

Table 2 Numbers of rules and paraphrase candidates for each type of transfer.

構文トランスファの種類	<格>	<否>	<機>	<サ>	<分>	<動>	<語>	<慣>	合計
構文トランスファ規則数	33	291	248	6,642	18	3,630	13,348	3,942	28,152
生成された言い換え事例数	148	77	19	46	20	60	252	58	680
本文中の例文	(9)	(12)	(13)	(7)	(14)	(15)	(16)	(17)	-

え事例の数を表 2 に示す。表中、<格> ~ <サ> の 4 種類は、3 章で用いたトランスファ規則である。ただし、<サ> については、国語辞典<sup>33),34)</sup> の語釈文から「サ変名詞+する」に対応する動詞相当句を機械的に抽出し、トランスファ規則とした。一方、<分> ~ <慣> の 4 種は、新規に導入したトランスファ規則である。以下、導入のねらいおよび規則の実装方法について述べる。

<分> 分裂文から無標文への言い換え

- (14) s. 収録されているのは約一〇〇〇人の人物だ。  
 t. 約一〇〇〇人の人物が収録されている。  
 分裂文から無標文の言い換えについては、言い換える可否が前後文脈の主題や焦点に依存することが文献 38) で示されている。3 章における調査対象としては語彙的な変換が主となるトランスファのみを用いていたが、より構文的な言い換えにも同様の変換誤りが生じるか否かを調査するために取り上げた。トランスファ規則は、文献 38) 中の例文をもとに人手で作成した。

<動> 動詞句から動詞句への言い換え

- (15) s. 得票総数が投票者数を上回ることが判明。  
 t. 得票総数が投票者数を超えることが判明。  
 機械翻訳用に作成された動詞格フレーム辞書<sup>13)</sup> から、同じ英語格フレームに訳される日本語格フレーム対をトランスファ規則として自動的に抽出した。必須格とその格要素の選択制限を含めた規則であるため、<サ> および後述の<語> よりも適格な言い換えを生成できると期待できる。

<語> 単語から同概念語への言い換え<sup>9)</sup>

- (16) s. 大統領は十七日付の書簡で了承した。  
 t. 大統領は十七日付の書状で了承した。  
 我々<sup>9)</sup> は、シソーラス中の同じ概念に属するあらゆる 2 語の組合せの中から、語釈文を照合することで言い換え可能な単語対を抽出する方法について報告している。この方法を採用し、EDR 日本語単語辞書<sup>8)</sup> の同概念語の中から、言い換え可能

な単語対(名詞、動詞、形容詞どうし)を獲得した。なお、語釈文の照合には、岩波国語辞典<sup>34)</sup>、角川類語新辞典<sup>33)</sup> をそれぞれ用いた。

<慣> 慣用表現から字義どおりの表現への言い換え  
 (17) s. 早急に終止符を打つ。

t. 早急に終わらせる。

国語辞典の語釈文には、慣用表現と字義どおりの表現がある特定のパターンで記述されている場合がある。我々は、角川類語新辞典<sup>34)</sup> からこのようなパターンをトランスファ規則として抽出した。このようなヒューリスティクスで獲得できる知識にはしばしばノイズが含まれるが、言い換える規模を拡大するには、このように機械的に獲得した知識を導入する必要がある。今回は、機械的に獲得したトランスファ規則をそのまま用いた場合の変換誤りを調査するために導入した。

#### 4.2 言い換え事例の評価と修正処理の必要性

次に、文の複数箇所と同じ規則が適用されている事例を除く 630 件を対象とし、言い換える良否を人手で評価した。変換誤りが生じている場合は、3 章で整理した変換誤りの分類に基づきラベルを付与した。不適格な言い換え事例については、可能ならば修正して正しい言い換えを付与し、修正不可能な場合は「棄却」というラベルを付与した。トランスファ規則の種類ごとに各変換誤りを含む事例の数を表 3 に示す。

表中「その他」は、新出の変換誤りを含む事例の数を示す。また、次に示す 4 つは、トランスファに用いた要素技術において生じている誤りである。これらは、3 章の手順では誤りであるかどうかを評価するためのパターン、および解決方法を示すことができないため、別途数え上げた。

- (A) 慣用表現・固有表現の誤認  
 (B) 機械的に獲得したトランスファ規則における矢印や括弧などの辞書特有のメタ表現によるノイズ  
 (C) 形態素・構文解析の誤り  
 (D) 言い換えエンジンの書き換え操作の誤り

言い換え事例の内訳は、修正・棄却処理が不要な事例が 114 件、なんらかの修正処理が必要な事例が 222 件、棄却すべき事例が 294 件であった。トランスファ直後のテキストをそのまま言い換える出力と

文獻 21) で用いられた各種辞書は公開されていない。そこで、サ変名詞の語釈文中の「~すること」で終わる文を手がかりにし、末尾の「こと」を取り除いた表現を「サ変名詞+する」に対応する動詞相当句として取り出した。

表 3 各変換誤りを含む言い換え事例数

Table 3 Error distribution for each type of transfer.

トランスファの種類	<格>	<否>	<機>	<サ>	<分>	<動>	<語>	<慣>	合計
評価事例数	138	75	19	39	20	60	221	58	630
不適格性を含む(修正を必要とする)事例数	137	57	9	35	17	53	172	36	516
(a) 《活用形の誤り》	125	41	3	31	7	43	47	6	303
(b) 《不適格な機能語接続》	42	14	2	3	5		8	4	78
(c) 《格助詞の欠損》		*			6	2			8
(d) 《同じ格要素の重複》				7				4	11
(e) 《節内の格要素と動詞の不整合》	66	*	*	8		28	57	3	162
(f) 《修飾語の重複, 競合》				*					0
(g) 《(e) 以外の共起の不整合》				*		3	28	5	36
(h) 《内容語の意味の変化で文の意味が変わる》							30	1	31
(i) 《モダリティの持つ意味の変化で文の意味が変わる》	1	5		3			13		22
(j) 《時間情報が等しくない》	2	1			3				6
(k) 《文体が等しくない》	*		1	*					1
(l) 《すわりが悪い語順》	23	*	*	*	2	7		2	34
(m) 《主題・陳述構造の不整合》	10	1			10	1			22
(n) 《節間, 文間の修辭的關係の不整合》	2	4	2						8
その他	38	16	2	7	8	3	19	22	115
(A) 慣用表現・固有表現の誤認	9			1			26	4	40
(B) 辞書特有のメタ表現によるノイズ							18	20	38
(C) 形態素・構文解析の誤り	7	5	5	1			22	1	41
(D) 言い換えエンジンの書き換え操作の誤り	8	1	1		1	1	1	2	15

\*は、訓練セットでは生じたが評価セットでは生じなかった変換誤りを指す。

した場合、その再現率は 33.9% (114/336)、精度は 18.1% (114/630) となる。このことは、言い換えにおける修正処理の重要性を示唆している。

## 5. 議 論

表 3 から、多くの変換誤りが、構文トランスファの種類に依存することなく生じていることが分かる。すなわち、特定の変換誤りを検出・修正する修正モジュールを実現すれば、構文トランスファ規則の種類を越えて利用できる。以下、変換誤りの種類ごとに事例を示し、解決に向けての考察を述べる。

### 5.1 形態的な不適格性

例文 (4.t), (11.t) のような (a) 《活用の誤り》は、さまざまな種類の構文トランスファにおいて共通して、頻繁に生じていた。ただし、解決はそれほど難しくないと考えられる。ある活用語の活用形は、活用型とそれが係る語の品詞に対してほぼ一意に決定できるし、活用語の 1 語 1 語に対して活用型を定義したリソースも存在するためである。また、自然言語生成タスクにおける選択点の 1 つとして、活用形を決定するさまざまな手法も提案されている。

表中 (b) の《不適格な機能語の接続》の誤りも頻出した。この誤りの修正処理として、例 (18), (19) のような、同じ役割の機能語への置き換えが考えられる。

(18) s. 予想外の負け方にあっけにとられた。

t. \*意外の負け方にあっけにとられた。

t'. 意外な負け方にあっけにとられた。

(19) s. 犯人の自宅を同署員が調べた。

t. \*同署員に犯人の自宅が調べられた。

t'. 同署員に犯人の自宅が調べられた。

「比例される」のように受動態になれない動詞が統語的受身「れる」をとともなう場合に棄却する場合もある。この種の不適格性も、(a) と同様に品詞や活用形に基づくレベルで扱えるため、比較的容易に解決できると考えられる。ただし、接尾辞「れる」が受身、可能、丁寧の意味を持つように (i) の《モダリティの持つ意味の変化で文の意味が変わる》問題が残る場合もある。

### 5.2 構文的な(共起の)不適格性

表中 (c) ~ (e) は、動詞格フレームに関する不適格性である。(f), (g) は、句または節内の構文的な不適格性で、(c) ~ (e) に該当しないものである。この中では、(e) の《節内の格要素と動詞の不整合》が最も頻繁に生じた。例文 (20.t), (21)t のような、選択制限の違反、あるいは動詞の格パターンの誤りである。

(20) s. ネットワークが国境を超えた。

t. \*ネットワークが国境を上回った。

(21) s. 彼に嫁ぐ。

t. \*彼に結婚する。

t'. 彼と結婚する。

例 (20) は、英語動詞「exceed」に対応する「超える」と「上回る」の言い換えである。例 (15) の文脈ではこの言い換えは適格であったが、例 (20) においては「超える」は「exceed」の意味ではないため不適格である。言い換える前に多義性を解消しておけばこのような変換誤りを防ぐことができるが、「国境」が「上回る」のヲ格の選択制限を満たしていない、ととらえることもできる。一方、例 (21) は、英語動詞「marry」



に対応する「嫁ぐ」と「結婚する」の言い換えである。多義性の問題ではないが「結婚する」の格フレームが二格を持たないため不適格である。この例では、格助詞を別の格助詞に書き換えることで、適格な文に修正できる。しかし、このように格助詞の交替によって修正可能な事例は162件のうち22件しかなかった。したがって、修正処理を、

- ある構文パターンが適格であるか不適格であるかを判定する処理（誤り検出）
- 不適格な場合の処理（誤り修正）

の2つの部分問題に分けるとすれば、《節内の格要素と動詞の不整合》に関しては、まず誤り検出の問題について取り組むことが重要であるといえる。

動詞格フレームを対象とした修正処理には、言語解析における統計的アプローチが参考になる。たとえば、河原<sup>19)</sup>は、動詞とその直前の格要素をキーとし、統計的に良質の格フレームを自動収集する方法を提案している。また、宇津呂<sup>43)</sup>、宮田<sup>29)</sup>は、ある名詞がある動詞の下位範疇を満たす確率を最大エントロピー法、ベイジアン・ネットワークといった確率モデルで推定している。これらの手法を応用すれば、適格/不適格、およびどのように修正するか、棄却すべきかの判定を訓練可能なモデルで実現できる可能性がある。ただし、言い換え後の表現が適格であるか否かは、個々の単語に依存していることが多い。このため、従来の意味クラスに基づく手法が有効に働くかどうかは明らかでない。

このほか、名詞とその修飾語の共起（例(22)）、複合動詞の共起（例(23)）、副詞と動詞の共起（例(24)）など、係り受けの関係にある単語の共起によってとらえられる不適格性は(g)に分類した。これらについても、動詞格フレームと同じく統計的なアプローチが有効であると考えられる。

- (22) s. 難解な言葉が法律文の随所にある。  
t. \*難解な言葉が法律文の各地にある。
- (23) s. 貴乃花は引きつけ直さず、勝ちを急いだ。  
t. \*貴乃花は魅せ直さず、勝ちを急いだ。
- (24) s. 毎年同湖にやって来るオオハクチョウ。  
t. \*年年同湖にやって来るオオハクチョウ。

### 5.3 意味的な不適格性

(h)《内容語の意味の変化で文の意味が変わってしまう》という不適格性は、言い換え前後の語の多義性(ambiguity)や曖昧性(vagueness)によって生じる問題のうち、(c)~(g)の変換誤りとして検出できないものである。

例(25)は、EDR 日本語単語辞書<sup>8)</sup>中で「3cf6fc:

仕事や物事を終える」という共通概念を持つ「仕上げる」と「括る」の間の言い換えである。ところが、例(25)の文脈において、「括る」は「3be172:ばらばらの物を1つにまとめてしぼる」の解釈が自然であるため、言い換えとしては不適格になる。

- (25) s. 奥さん方は針と糸で仕上げる浴衣に大喜び。  
t. \*奥さん方は針と糸で括る浴衣に大喜び。

31件中12件は、例(25)のように多義性解消によって誤りをとらえることができると考えられた。

残りの19件には、言い換え前の語に比べて、言い換え後の語は漠然とした意味になっているために不適格になる事例などが含まれる。たとえば、例文(26)tにおいては、「遺伝子」に比べて「因子」は対象としている分野の情報が欠けている。

- (26) s. 遺伝子を抑えることで成長を抑制できる。  
t. \*因子を抑えることで成長を抑制できる。

これらは、2つの語の間の細かい意味の違いがシソーラスや語釈文の照合によって十分にとらえられていないために生じる変換誤りだと考えられる。Edmonds<sup>7)</sup>は、類義語(near-synonym)間の細かい意味の違いを記述するために、よりプリミティブな意味素からなるオントロジを提案し、機械翻訳における訳語選択に用いている。必要となる大規模な知識を安価に獲得・記述する手法については、Inkpen<sup>14)</sup>、岡本<sup>32)</sup>のような、類義語の使い分けを記述した辞書から類義語間の意味の差分を取り出すという試みはあるものの、多くの問題が未解決のまま残っている。

### 5.4 談話的な不適格性

表中(j)~(n)は、言い換え前と同様の談話構造が得られない、あるいは結束性が壊れてしまうといった不適格性である。

(j)《時間情報が等しくない》と(k)《文体が等しくない》は、例(27)のように、各節の述語表現が言い換え前と同じ意味を持たなくなる不適格性もある。

- (27) s. ニューオリンズの本屋の店員はトウキョーを知らなかった。アメリカがどれだけ広いかを知るためには格好の挿話だろう。その“広い”アメリカを代表するのがミシシッピ川流域である。  
t. \*ミシシッピ川流域はその“広い”アメリカを代表する。  
t'. ミシシッピ川流域はその“広い”アメリカを代表している。

解決のためには、時間情報、文体などを表す体系、および意味解析を導入し、言い換え前後で意味解析結果を照合する必要がある。ただし、同じテイル形で

あっても動詞、文脈によって指示するアスペクトが異なることなどに留意しなければならない。

(l) 《すわりが悪い語順》は、副詞と格要素、格要素どうし(例(28))など、同じ要素に係る語句の順序の問題である。単文において格間の順序を決定する際に有効な素性については文献(42)で考察されている。ただし、「腹をくくる」のような慣用表現においては特定の格要素が動詞の直前に来ることが必須であるが、このような特殊な例を扱うための素性は用いられていない。また、実際には、個々の単語やその組合せ、主題展開などの文脈にも依存している複雑な問題である。

(28) s. マラナス人民統一戦線が革命を起こし、夢の島は収容所群島に変わる。

t. <sup>?</sup>革命がマラナス人民統一戦線から起こり、夢の島は収容所群島に変わる。

t'. マラナス人民統一戦線から革命が起こり、夢の島は収容所群島に変わる。

(m) 《主題・陳述構造の整合》や(n)《節間、文間の修辭的關係の整合》といった結束性に関する評価には、センタリング理論や修辭構造理論の導入が考えられる。結束性に関する制約、選好基準については、野上らが文献(31)でまとめているので参考になる。ただし、野上らの結束性評価基準は、連体節を主節から切り離すタスクに限定しているため、連用節、副詞節なども含めると必要な選択点の体系が変化する可能性もある。また、連体修飾節の限定/非限定のように、評価に必要な素性を高い精度で解析するアルゴリズムの開発も必要である。

### 5.5 考 察

本論文で述べた調査は、限られた数のトランスファ規則を使った限定的なものであるため、さらに大規模な調査が必要のように見えるかもしれない。しかし、言い換えは個々の単語に依存して生成され、各事例が適格であるか否かも個々の単語に依存していることが多い。このため、言い換えにともなう変換誤りを網羅的に調査することは実際には不可能である。

この点に関して我々は、変換誤りの種類を構文パターンなどで抽象化して整理することにより、今後取り組むべき修正処理の問題に良い見通しを与えることができたと考える。また、新規の変換誤りが生じた場合でも、3章で述べたように構文パターンを付与することで、どのレベルの不適格性であるのかを分類することができる。そして、経験的な分類に基づいて、発生しやすい、あるいはより広範な種類の言い換えにおいて生じる変換誤りから順に修正モジュールを実現していけば、適格な言い換への生成に対して貢献できる。

実際、表3中(a)~(g)の誤りを適切に修正・棄却できたならば、今回の事例の範囲では、132件を修正、109件を棄却でき、言い換への再現率は73.2%(246/336)、精度は47.2%(246/521)となる。(a)~(g)の、形態的な不適格性、および構文的な不適格性は、規則ベース、あるいは大規模なコーパスに基づく統計ベースの言語モデルを用いて解決できる可能性もあるので、まずはこのレベルの不適格性について、修正処理の実現に取り組むべきである。

## 6. 関連研究

この章では、異言語間機械翻訳における翻訳の後処理に関する考察を、言い換えにおける修正処理との比較の観点から整理する。

Allenら<sup>2)</sup>は、仏英機械翻訳の後処理として、活用語の語尾変化、ドメインに依存して決定すべき逐語対訳といった頻出する問題点をあげている。そして、小規模ながら、翻訳前テキスト、翻訳後のテキスト、それを人手で修正したテキストのtri-textから、修正事例を自動的に獲得したと報告している。これに対し、言い換えについては、トランスファによって生じる変換誤りの種類や、必要な修正処理が明らかにはなっていない。本論文では、実際にある程度の規模の事例分析を通じて変換誤りの種類を整理し、各変換誤りが生じる傾向、およびそれらの性質を整理した。これにより、8種類ではあるが、性質が異なるトランスファ規則においても、生じる変換誤りの多くが共通のものであることが明らかになった。すなわち、修正処理をモジュール化することでトランスファ規則の開発コストの削減に貢献できる可能性を示した。

Knightら<sup>20)</sup>は、機械翻訳後の修正処理の例題として、冠詞を補完するタスクに取り組んでいる。名詞句の主辞の中で頻出する1,600語の名詞の各々に対して、前後2単語の文脈情報からa, an, theのどれを補うかを決定するタスクである。Knightらは次の3つの設定によって、問題を単純化している。すなわち、(i)実際の機械翻訳の出力ではなく、人手で冠詞を除いた疑似テキストを用い、(ii)冠詞を補う箇所は特定されていると仮定し、(iii)いずれかの冠詞を必ず補う戦略によって多義性解消と同様の択一問題に帰着している。これに対し、言い換えでは、構文トランスファ後の言い換え候補の各々が適格か不適格かを評価し、不適格な言い換えしか生成できないのであれば言い換えないという戦略を選択する必要がある。このことから、個々の修正処理は、機械翻訳のそれに比べ、より複雑なタスクになると考えられる。

## 7. おわりに

本論文では、語彙・構文的言い換えを構文トランスファ方式で実現する際に生じる変換誤りに焦点を当て、その種類を分類・整理した。まず、個別の言い換えに関する先行研究で示されている言い換え知識を用いて変換誤りを洗い出した。そして、変換誤りの種類を構文パターンなどで抽象化し、不適格性の種類を整理した。次に、さまざまな種類の構文トランスファ規則を用いて変換誤りが生じる傾向を調査した。これらの調査を通じて、多くの変換誤りが、構文トランスファ規則の種類に依存することなく生じることが分かった。また、今後修正処理の実現に取り組むにあたって良い見通しを与えることができ、以下の知見が得られた。

- 形態的な不適格性は、さまざまな種類の構文トランスファ規則を適用する際に共通して発生していた。ただし、言語生成においてさまざまな解決方法が論じられており、リソースも整っているため、比較的容易に修正できると考えられる。
- 構文的な不適格性も、トランスファ規則の種類にかかわらず、比較的頻繁に生じていた。多くは、単語の共起の問題としてとらえることができる。また、言い換える語の多義性や曖昧性によって生じる問題であっても、構文パターンとしては、共起の問題としてとらえられる場合もある。
- 意味的な不適格性はそれほど頻繁には生じないが、単語、句レベルの置き換えである語彙的言い換えの際に生じやすい。この不適格性を修正するためには、単語の多義性や類義語間の細かい意味の違いをとらえる必要がある。
- 談話的な不適格性もそれほど頻繁には生じないが、文の構造を大きく変える構文的言い換えにおいて生じやすい傾向が見られた。この不適格性の修正には、結束性評価に必要な素性を解析する処理の実現が必須である。

以上、まずは頻繁に生じる傾向が認められた形態的な不適格性、および構文的な（共起の問題としてとらえられる）不適格性を優先的に解消すべきであるという結論を得た。この結論をふまえ、我々は、(e)《節内の格要素と動詞の不整合》を検出するタスクに取り組んだ<sup>10)</sup>。5章の議論をふまえ、個々の単語を区別する統計モデルを構築し、人手で洗練した格フレームよりも優れた誤り検出精度を得ている。

謝辞 奈良先端科学技術大学院大学の高橋哲朗氏、飯田龍氏（株）富士通研究所の岩倉友哉氏には、言い換えエンジン KURA における修正処理の実現に関し

て、多くのコメントをいただきました。奈良先端科学技術大学院大学の松本裕治氏には、示唆に富む多くの助言をいただきました。本論文の査読者の方々には、論文を精読していただき、有益なご指摘をいただきました。皆様に深く感謝いたします。

## 参考文献

- 1) ACL: *The 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP)* (2003).
- 2) Allen, J. and Hogan, C.: Toward the development of a postediting module for raw machine translation output: a controlled language perspective, *Proc. 3rd International Workshop on Controlled Language Applications (CLAW)*, pp.62-71 (2000).
- 3) 浅原正幸, 松本裕治: IPADIC ユーザーズマニュアル, 技術報告, 奈良先端科学技術大学院大学 (2002).
- 4) Barzilay, R. and McKeown, K.R.: Extracting paraphrases from a parallel corpus, *Proc. 39th Annual Meeting of the Association for Computational Linguistics and 10th Conference of the European Chapter of the Association for Computational Linguistics (ACL-EACL)*, pp.50-57 (2001).
- 5) Carroll, J., Minnen, G., Pearce, D., Canning, Y., Devlin, S. and Tait, J.: Simplifying text for language-impaired readers, *Proc. 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp.269-270 (1999).
- 6) Dorna, M., Frank, A., van Genabith, J. and Emele, M.C.: Syntactic and semantic transfer with F-structures, *Proc. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL)*, pp.341-347 (1998).
- 7) Edmonds, P.: Semantic representations of near-synonyms for automatic lexical choice, Ph.D. Thesis, CSRI-399, Department of Computer Science, University of Toronto (1999).
- 8) 日本電子化辞書研究所: EDR 電子化辞書仕様説明書, 日本電子化辞書研究所 (1995).
- 9) 藤田 篤, 乾健太郎: 語釈文を利用した普通名詞の同概念語への言い換え, 言語処理学会第7回年次大会発表論文集, pp.331-334 (2001).
- 10) 藤田 篤, 乾健太郎, 松本裕治: 自動生成した言い換え文における動詞結合誤りの自動検出手法, 情報処理学会自然言語処理研究会予稿集, NL-156-8, pp.53-60 (2003).
- 11) 言語処理学会: 言語処理学会第7回年次大会ワー

- クショップ「言い換え/パラフレーズの自動化」(2001).
- 12) 飯田 龍, 徳永泰浩, 乾健太郎, 衛藤純司: 言い換えエンジン KURA を用いた節内構造および機能語相当表現レベルの言い換え, 第 63 回情報処理学会全国大会予稿集第二分冊, pp.5-6 (2001).
  - 13) 池原 悟, 宮崎正弘, 白井 諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林 良彦 (編): 日本語語彙大系: CD-ROM 版, 岩波書店 (1997).
  - 14) Inkpen, D.Z. and Hirst, G.: Building a lexical knowledge-base of near-synonym differences, *Proc. 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL) Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations* (2001).
  - 15) 乾健太郎: 読解支援を目的とするテキスト単純化の実現に向けて 課題と方法論, 電子情報通信学会思考と言語研究会予稿集, TL2001-8, pp.51-58 (2001).
  - 16) 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資料, pp.1-21 (2002).
  - 17) 影山太郎 (編): 日英対照 動詞の意味と構文, 大修館書店 (2001).
  - 18) 鍛冶伸裕, 黒橋禎夫, 佐藤理史: 国語辞典に基づく平易文へのパラフレーズ, 情報処理学会自然言語処理研究会予稿集, NL-144-23, pp.167-174 (2001).
  - 19) 河原大輔, 黒橋禎夫: 用言と直前の格要素の組を単位とする格フレームの自動構築, 自然言語処理, Vol.9, No.1, pp.3-19 (2002).
  - 20) Knight, K. and Chander, I.: Automated postediting of documents, *Proc. 12th National Conference on Artificial Intelligence (AAAI)*, pp.779-784 (1994).
  - 21) 近藤恵子, 佐藤理史, 奥村 学: 「サ変名詞+する」から動詞相当句への言い換え, 情報処理学会論文誌, Vol.40, No.11, pp.4064-4074 (1999).
  - 22) 近藤恵子, 佐藤理史, 奥村 学: 格変換による単文の言い換え, 情報処理学会論文誌, Vol.42, No.3, pp.465-477 (2001).
  - 23) 神田慎哉, 藤田 篤, 乾健太郎: 連用節主節化に関する規則の追試と洗練, 第 15 回人工知能学会全国大会, 1A1-06 (2001).
  - 24) 黒橋禎夫: 言葉の意味を計算機で扱う, 言語処理学会第 6 回年次大会チュートリアル資料, pp.21-28 (2000).
  - 25) Lin, D. and Pantel, P.: Discovery of inference rules for question answering, *Natural Language Engineering*, Vol.7, No.4, pp.343-360 (2001).
  - 26) 松本裕治, 北内 啓, 山下達雄, 平野善隆, 松田 寛, 高岡一馬, 浅原正幸: 形態素解析システム『茶釜』 version 2.2.9, 技術報告, 奈良先端科学技術大学院大学 (2002).
  - 27) Melamed, I.D.: *Empirical methods for exploiting parallel texts*, MIT Press (2001).
  - 28) Mitamura, T. and Nyberg, E.: Automatic rewriting for controlled language translation, *Proc. 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp.1-12 (2001).
  - 29) 宮田高志, 宇津呂武仁, 松本裕治: Bayesian Network による下位範疇化の確率モデルおよびその学習, 情報処理学会自然言語処理研究会予稿集, NL-119-12, pp.77-84 (1997).
  - 30) NLPRS: *Workshop on Automatic Paraphrasing: Theories and Applications* (2001).
  - 31) 野上 優, 乾健太郎: 言い換えを用いた結束性評価基準の構築, 言語処理学会第 8 回年次大会発表論文集, pp.335-338 (2002).
  - 32) 岡本紘幸, 斎藤博昭: 文脈を考慮した日本語類義表現の言い換え, 言語処理学会第 9 回年次大会発表論文集, pp.97-100 (2003).
  - 33) 大野 晋, 浜西正人: 角川類語新辞典, 角川書店 (1981).
  - 34) RWC: RWC テキストデータベース第 2 版, 岩波国語辞典タグ付き/形態素解析データ第 5 版 (1998).
  - 35) 佐藤理史: 論文表題を言い換える, 情報処理学会論文誌, Vol.40, No.7, pp. 2937-2945 (1999).
  - 36) Shinyama, Y., Sekine, S., Sudo, K. and Grishman, R.: Automatic paraphrase acquisition from news articles, *Proc. Human Language Technology Conference (HLT)* (2002).
  - 37) 白井 諭, 池原 悟, 河岡 司, 中村行宏: 日英機械翻訳における原文自動書き換え型翻訳方式とその効果, 情報処理学会論文誌, Vol.36, No.1, pp.12-21 (1995).
  - 38) 砂川有里子: 日本語における分裂文の機能と語順の原理, 複文の研究 (下), 仁田義雄 (編), pp.353-388, くろしお出版 (1995).
  - 39) Takahashi, T., Iwakura, T., Iida, R., Fujita, A. and Inui, K.: KURA: a transfer-based lexico-structural paraphrasing engine, *Proc. 6th Natural Language Processing Pacific Rim Symposium (NLPRS) Workshop on Automatic Paraphrasing: Theories and Applications*, pp.37-46 (2001).
  - 40) 徳田昌晃, 奥村 学: 日本語から手話への機械翻訳における手話単語辞書の補完方法について, 情報処理学会論文誌, Vol.39, No.3, pp.542-550 (1998).
  - 41) 徳永泰浩: 取り立て詞に着目した否定表現の言い換えと意味解析, 九州工業大学情報工学部知能情報工学科卒業論文 (2002).
  - 42) 内元清貴, 村田真樹, 馬 青, 内山将夫, 関根

聡, 井佐原均: コーパスからの語順の学習, 情報処理学会自然言語処理研究会予稿集, NL-135-8, pp.55-62 (1999).

- 43) 宇津呂武仁, 宮田高志, 松本裕治: 最大エントロピー法による下位範疇化の確率モデル学習および統語的曖昧性解消による評価, 情報処理学会自然言語処理研究会予稿集, NL-119-11, pp.69-76 (1997).

(平成 14 年 9 月 26 日受付)

(平成 15 年 9 月 5 日採録)



藤田 篤

1977 年生. 2000 年九州工業大学情報工学部卒業. 2002 年同大学大学院情報工学研究科博士前期課程修了. 同年, 奈良先端科学技術大学院大学情報科学研究科博士課程入学.

現在にいたる. 自然言語処理の研究に従事.



乾 健太郎 (正会員)

1967 年生. 1995 年東京工業大学大学院情報理工学研究科博士課程修了. 同年より同研究科助手. 1998 年より九州工業大学情報工学部助教授. 1998 年 ~ 2001 年科学技術振興事業

団さきがけ研究 21 研究員を兼任. 2001 年より奈良先端科学技術大学院大学情報科学研究科助教授. 現在にいたる. 博士 (工学). 自然言語処理の研究に従事. 言語処理学会, 人工知能学会, 電子情報通信学会, ソフトウェア科学会各会員.

---