

語彙的言い換えに必要な知識の部品化

藤田 篤 乾 健太郎

奈良先端科学技術大学院大学 情報科学研究科

〒 630-0101 奈良県生駒市高山町 8916-5

{atsush-f,inui}@is.aist-nara.ac.jp

我々はこれまでに、語彙・構文的言い換の事例研究を通じて、言い換えに必要な言語知識の種類を明らかにするとともに、それらを実装し、各種語彙・構文的言い換えを実現してきた。本稿では、先行研究ならびに我々のグループの研究で得られた知見を横並びに分析し、さまざまな言い換えに用いることができる部品として再構成することを提案する。具体的には、言い換え後の不適格性を修正する知識に焦点を当てる。まず、先行研究で報告されている言い換え規則と事例分析に基づいて部品を定義した。また、さまざまな語彙・構文的言い換えにおいて、各部品の再利用性を調査した。

キーワード：語彙的言い換え、修正に基づく構造変換、知識の部品化と再利用

Decomposing Linguistic Knowledge for Lexical Paraphrasing

FUJITA Atsushi INUI Kentaro

Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara, 630-0101, JAPAN

{atsush-f,inui}@is.aist-nara.ac.jp

We have focused and explored some instances of lexico-structural paraphrasing. The almost sorts of required knowledge in order to realize paraphrases are built in the transfer rules. In this paper, we discuss decomposing linguistic knowledge into reusable components, focusing on lexical paraphrasing.

Keywords : lexical paraphrasing, revision-based transfer, knowledge decomposition and reusability

1 はじめに

言い換えは、ある言語表現を、できるだけ意味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。言い換えはこれまでに、機械翻訳、要約、推敲支援などの応用分野においてそれぞれ独立に必要性が論じられ、前処理としての導入など個別的な試みがなされてきた。しかし近年、自然言語処理の諸分野においてさまざまな応用が考えられる重要な要素技術として、研究者の関心も高まってきている [3, 18]。

ひとくちに言い換えといっても様々な現象が考えられるが、近年、次の (1) ~ (3) の例のような、内容語の置換に重点を置いた語彙的言い換えに関する試みがいくつか報告されている [15, 7, 8, 6]。我々の研究グループでも、語彙・構文的言い換えのための言い換えエンジン KURA [16] を開発し、これを用いてさまざまな語彙・構文的言い換えの事例研究を進めている [5]。

- (1) 激しい自動車戦争に進む公算が大きい。
⇒ 激しい自動車戦争に進む可能性が大きい。
- (2) 犯人は警察に逮捕された。
⇒ 犯人は警察に捕まえられた。
- (3) 先頭との距離を縮める。
⇒ 先頭との距離が縮まる。

これらの研究では、単語から他の単語、動詞交替など、個別の現象に焦点を当て、事例分析に基づいて言い換え規則などの知識を作成し、計算機上に実装している。言い換えの研究は未だ萌芽的段階にあるので、今後もこのような事例研究を重ねることには意義がある。

しかしながら、あらゆる問題を考慮して頑健な言い換え規則を作ろうとすると、言語的制約や付随的な変形手続きを含む複雑な表現になってしまう。たとえば、近藤らは (2) のような言い換えを実現するために (4) のような規則を示している。

- (4) 「サ変名詞 + する」から動詞相当句への言い換え
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」)の活用形をもとの「れる」に一致させる。
 5. 接尾辞をコピーする

(4) の下線部は、動詞の種類に合わせて接尾辞を定める手続きであり、(5) のような言い換えにも必要である。しかし、変形手続きや語彙に関する制約などが別の言い換えにも利用できるとしても、(4) のようになっては共有は容易でない。

- (5) 犬が彼に噛み付く。
⇒ 彼が犬に噛み付かれる。

事例研究がある程度のバリエーションを持ち、知見が蓄積されてきた現在、さらにさまざまな事例研究を進めるのと並行して、個別の研究で得られた知見を横並びに分析し、統合する作業に取り組むことも必要である。そこで、本研究では、さまざまな語彙的言い換えを計算機上で実現する際に必要となる言語知識を、再利用が容易な部品として再構成することを目的とする。本稿では、部品化のアプローチと知識の同定作業の現状について報告する。

以下、2 節では提案する言い換えのモデルと部品の定義を示し、3 節で具体的な部品化の方法について述べる。4 節では部品の再利用性についての調査結果を示す。5 節で知識の部品化と語彙的言い換えにおける課題を整理し、6 節でまとめる。

2 言い換えと知識の部品化

2.1 部品化の方針

共有できる知識、部品化された知識は先行研究の中ですでにいくつか示されている。

鍛冶らは、テキスト中の単語を国語辞典の語釈文に言い換える際に、次の 4 つの整形処理を用いている [6]。

- (6) 語釈文を整形する処理
 - a. 入力文の態によって、格助詞を変換する。
 - b. 格助詞を提題化する。
 - c. 「など」を削除する。
 - d. 用言が埋め込まれる場合は、接尾辞を整形する。

これらがどのような処理を担うのかを想像することはできるが、どのような条件で用いるのか、出力はどのような適格性を保証されているのかなどは明示的に示されていない。

また、織らが文献 [12] で提案している『テキストの合成演算』も部品化された知識であるといえる。それは、この演算では、2 つの文、文間の関係などの入力データの具体例、出力テキストが満たすべき制約が明示的に定義されているからである。再利用

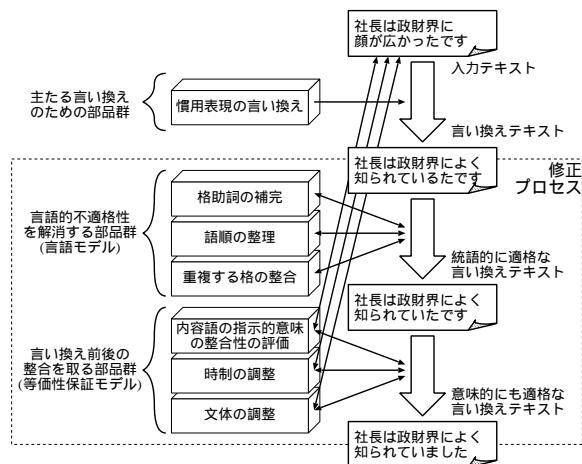


図 1: 想定する言い換え生成モデル

を容易にするためには仕様を定める必要があるの
で、このような定義は参考になる。

部品としては、さまざまな入出力、処理単位を持つものを定義することが可能だが、我々は、さまざまな言い換えで共有できるように可能な限り単純な(プリミティブな)処理を部品として定義する。

2.2 言い換え生成モデル

知識の部品化のアプローチは、次に示すような仮説に基づいている。

- 複雑に見える言い換えるの多くは、比較的単純な(プリミティブな)言い換えるの組み合わせで実現できる。
- 原文の一部を言い換えただけでは不適格な文が生成される可能性がある。これを修正したり、棄却したりするための知識は、言い換えるの種類をまたいで共有できる可能性が高い。
- 辞書的知識も、言い換えるを生成する制約、言語的適格性を保証するための知識として、応用横断的に利用できる可能性がある。

とくに、(4)、(6)で例示したような、言語的・意味的に不適格な文を修正する処理に焦点を当て、図1の言い換え生成モデルを考える。

図1では、修正のために言い換え前のテキストを参照する必要があるか否かで修正プロセスを大きく2つのステージに分類している。一つは、言い換え後のテキストの言語的な不適格性を解消するための部品の集合(言語モデル)であり、もう一方は、言い換え前後の整合を取るための部品の集合(等価性保証モデル)である。それぞれのステージは、部

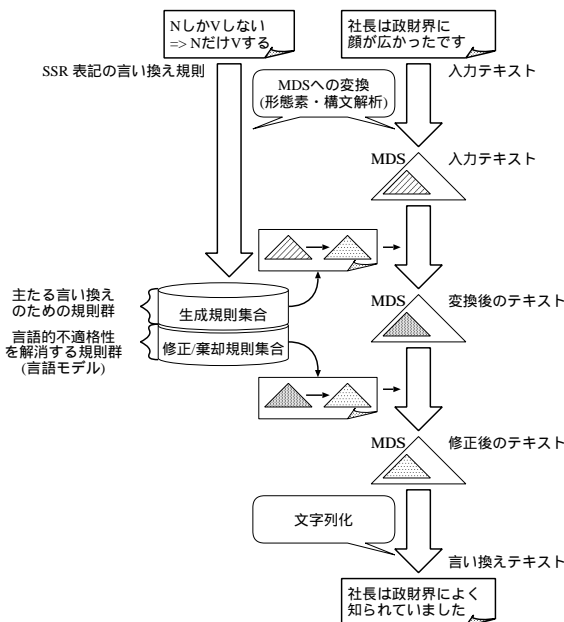


図 2: KURA の言い換え生成モデル

品が自律的に修正・棄却処理を行う黑板モデルを考える。

本研究では、図1のモデルに準じた図2のようなモデルを持つ言い換えエンジン KURA[16]を用いて、語彙的言い換えるの実現に取り組む。KURAは具体的には次のような特徴を持っている。

- 対象テキストを、内部では形態素の係り受けに基づく依存構造 MDS (Morpho-syntactic Dependency Structure) の形式で扱う。
- 主たる言い換えるのための規則(変換規則)、不適格な表現を修正/棄却する規則(修正規則)の2種の言い換え規則を扱う。変換規則の適用後に、適用可能なすべての修正規則を適用して、言語的な不適格性を取り除く。
- 言い換え規則は、言い換え規則記述のために用意された SSR (Simplified MDS Rewriting) 表記を用いて記述する。(7)のように、“=>”の両側に、文字列と単純な記号でパターンを記述したものを宣言的に記述する。規則の適用条件にあたるのは表現対の左辺であるが、(7b)のようにアノテーションを付加することで細かく指定できる。言い換え規則も内部では MDS の形式で扱われる。

(7) 否定表現を言い換える規則の SSR 表記 [17]

- a. NしかVしない => NだけVする。

- b. N(表記 ≠ 何とか)Vする(原形 ≠ する)し
かない => VするのはNだけ.

3 部品化する知識の同定

まず, 先行研究で提示されている言い換え規則から次のような手順で部品化する知識を探る.

1. 複雑な言い換え規則に埋め込まれている修正処理に相当する知識をすべて取り出し, 部品の候補とする.
2. 言い換えエンジン KURA を用いて修正処理を組み合わせて適用し, 言い換えテキストを網羅的に生成する.
3. 複雑な規則でも, 細かい修正処理の組み合わせでも, 同じ言い換えを生成できるのであれば, 修正処理は, 変換規則から切り離して部品化することができるかとみなす. 変換規則を適用する際の情報を引き継ぐ必要がある場合は, 変換規則に依存しているとみなし, 変換規則に再度埋め込む.
4. 2と3を, 可能な限り繰り返す.

分析に用いた例文の規模は次に示す通りである.

- 格交替(態・使役の交替, 動詞交替): 文献 [8] で例示されている 26 文
- 否定表現の言い換え: 文献 [17] で規則作成に用いられた京大コーパス中の 500 文
- 機能語相当表現の言い換え: 文献 [4] で規則作成に用いられた例文 185 文
- 「サ変名詞+する」から動詞相当句への言い換え: 文献 [7] で例示されている 23 文

たとえば, (4) の言い換え規則は, 次の 4 つの処理に分割した. (8a) が主たる言い換えのための規則であり, (8b) ~ (8d) が修正処理にあたる.

- (8) a. 「サ変名詞+する」を対応する動詞相当句に置き換える.
- b. 対象文の主動詞の活用形を修飾先の語に合わせて修正する.
- c. ヴォイス表現がある場合は, 直前の動詞にあわせて「れる」「される」を選択する.
- d. ヴォイス表現「れる」(または「られる」)の活用形を修飾先の語に合わせて修正する.

KURA でこれらを組み合わせて言い換え事例を生成した結果 (4) と同じ言い換えを生成できたので, (8b) ~ (8d) すべてを (8a) に依存しない処理であると同定した.

次に, (8b) と (8d) のように, 修正処理が同じものは, 対象を抽象化し, 「活用語の活用形を係り先の語に合わせて修正する」のような部品としてまとめあげ, 修正処理の仕様を定めた. 図 1 において部品は自律的に作用するものを仮定しているので, 次の 2 点を明示的に定めた.

- 局所的な適格性(あるいは不適格性)に基づく部品の発動条件 C
- 部品が担う処理内容 F

仕様を定めた修正部品の一部を付録 A に示す. ここでは, 図 1 と同様に, 言い換え前のテキストを参照する必要があるか否かで部品を分類してある.

4 再利用性の検証

4.1 実験の概要

部品化する知識を同定する過程で, 結果としてさまざまな種類の語彙的言い換えの変換規則が KURA 上に実装された. 変換規則作成および部品化する知識の同定作業に用いていない未知のデータを用いて, 定義した修正部品の再利用性を調査した.

具体的には, まず京大コーパス 1 月 9 日分の 1,220 文を対象とし, 次の 7 種の変換規則を適用して 792 件の言い換え事例を得た.

<分> 分裂文から無標文への言い換え

- (9) 収録されているのは約一〇〇〇人の人物だ。
⇒ 約一〇〇〇人の人物が収録されている。

<格> 格交替(態・使役の交替, 動詞交替)[8]

- (10) 米国の多くの地域が悪天候に見舞われた。
⇒ 悪天候が米国の多くの地域を見舞った。

<否> 否定表現の言い換え [17]

- (11) 参加すべきでない。
⇒ 参加してはいけない。

<機> 機能語相当表現の言い換え [4]

- (12) 先輩とはいえ一軍枠を争うライバル。
⇒ 先輩だけ一軍枠を争うライバル。

<動> 動詞句間の言い換え

表 1: 修正部品の要求頻度

対象は未知の文 1,220 文, “*” は分析対象中では当該事例が確認できたもの

変換規則の種類	<分>	<格>	<否>	<機>	<動>	<語>	<慣>	合計
変換規則数 (SSR)	18	34	284	261	4,221	4 ¹	3,942	8,764
のべ事例数	19	110	67	165	45	297	89	792
評価事例数	19	108	67	162	43	262	85	746
修正を必要としない事例数	2	3	16	40	5	59	18	143
語の活用形修正	6	51	28	26	21	58	15	205
機能語の連接修正	3	9	6	2		8		28
格助詞の補完	6		*	1				7
重複する格の整合		2				10	1	13
兄弟格要素の整合	4	4				*		8
修飾語同士の整合			1			*		1
格と動詞の整合		64	1		8	14		87
修飾語と動詞の整合		1		*		*	3	4
言語モデル: その他		2		109		5	2	118
内容語の指示的意味の整合					14	92	10	116
テンス・アスペクトの整合	3			2				5
モダリティの整合						2		2
文体の整合		*		2		*		2
語順の整理		1		*	2	*		3
主題・陳述構造の整合	5	*	1					6
節間, 文間の修飾的関係の整合	3		1					4
等価性保証モデル: その他	5	15	16			21	7	64
慣用表現・固有表現の誤認	1	4	1			32	5	43
規則・辞書の不備		1	7		4	16	31	59
KURA のバグ (構造変換)		6	11	1	1	16		35
KURA のバグ (パーズング)		6		28		33	4	71

- (13) 得票総数が投票者数を上回ることが判明。
⇒ 得票総数が投票者数を超えることが判明。

< 語 > 単語 (「サ変名詞 + する」も含む) から同概念語, 語釈文への言い換え [7, 2, 6]

- (14) 大統領は十七日付の書簡で了承した。
⇒ 大統領は十七日付の書状で了承した。

< 慣 > 慣用表現から字義通りの表現の言い換え

- (15) 事態に早急に終止符を打つ。
⇒ 事態を早急に終わらせる。

次に, 各事例について, 不適格性を修正/棄却するために定義した部品を用いることできるか否かを判定した。部品ごとに必要とされた回数を数上げた結果が表 1 である。なお, 部品として仕様を定めていない修正処理を必要とする場合もあったので, これは“その他”とした。

“< 格 >” ~ “< 機 >” と “< 語 >” の一部は, 分析結果に対する追試になるが, その他の言い換えを取り上げた理由について触れておく。“< 分 >” は, 語彙的言い換える事例を対象として定義した部品が, より構文的な言い換えにも再利用が見込めるか否かを調査するために取り上げた。“< 動 >” は翻訳

¹ 単語を対象とする言い換える規則数が少ないのは, 対応する表現対を辞書として実装しているためである。辞書のエントリーは 19,990 語。

用の辞書 [11] から同じ英語格フレームを持つ動詞句対を自動獲得したものである。格フィラーに選択制限が付与されているので単純な動詞の言い換えとは異なる種類とみなして用いた。“< 慣 >” も国語辞典の語釈文 [14] から自動獲得したものであるが, 慣用表現か字義通りの意味かの判断基準を設けていないため, 完全に他の言い換えとは排反な種類の言い換えとはみなせない。したがって, この規則を用いることによって得られた事例と部品の要求回数は参考値とみなす。

4.2 考察

表 1 から, 定義した修正部品のいくつかは, 変換規則の種類をまたいで利用できることが確認できた。変換規則が, 単語の単位に依存するもの, 活用語を対象とするもの, 新しく内容語を生成するかなどに依存する傾向も見られたが, 少なくとも, 種類の言い換えを対象に規則を作り込むのに比べると再利用性は高いといえる。

対象テキストのドメインが替わったり, 新たな種類の変換規則を持ち込んだり場合でも, 定義済みの部品がある程度の修正処理を担うことができるとは期待できる。しかし, 定義済みの部品では扱えない処理も同様に増えるのであれば, 知識作成のコストが抑えられず, 部品化の効果が高いとはいえない。

これについて、現状では実際にどれだけ対応できるかの予測は立てられていないが、少なくとも、修正すべき問題をときほぐして共通の処理を部品化することで、残されている問題も単純化できているとは考えられる。

5 議論

共通の修正処理は、仕様を定めて部品化して共有しようというのが、本稿の主張である。この主張に対し、部品の仕様を定める、それらを組み合わせる際にいくつかの議論すべき点がある。

5.1 発動条件をどのように設定するか

まず、各部品の発動条件となる適格性、あるいは不適格性の記述方法を定める必要がある。

Mitamura らが提案する KANT は、文法的に適切な文を文法・語彙が制限された言語へ言い換えるシステムである [9]。KANT では、目標となる制限言語を満たさない部分を次の 2 種類の手法を組み合わせで検出している。

- a) 制限言語とみなす文のみを受理するような文法を用意しておく。これを用いてパーズングできなければ、理由を推論しユーザに書き換えを要求する。
- b) 制限言語とみなさないパターンを列挙しておく。このすべてを用いてエラーチェックを行い、いづれかにマッチしたら自動的に書き換える。

我々は Mitamura らのように具体的な言い換えの目的を設定していないので、a) のようなアプローチは漠然としすぎていて採用できないように見える。しかし、部品化によって問題を切り分け、部品ごとに修正の目的となる局所的な適格性を定義することは可能であると考えられる。たとえば、『語の活用形修正』では、対象を抽象化して活用型のレベルで扱えば、正しい出現パターンを網羅することができる。一方、『格と主動詞の整合』のように、内容語 1 語 1 語の共起に基づくため抽象化が難しい問題は、すべての正例を書き尽くせない。かつ、b) の手法で負例を書き尽くすことも不可能である。

現状では、各部品が発動条件を満たすか否かを判定する機構を実装できていないが、したがって、a)、b) 両方を柔軟に組み合わせることができるように設計する必要がある。

5.2 何を指して修正するのか

『語の活用形修正』が担う「係り先に対して適切な活用形に修正する」など、統語的な適格性を保障する処理は、言語生成や、推敲支援の先行研究で、実現方法がある程度示されている。この他の適格性として、前後文脈との結束性や、意味的な適格性、等価性を計算する仕組みを実現しなくてはならない。

今回語彙的言い換えとして扱った中には、態・使役の交替のように、むしろ構文的な構造変換に重点がおかれた言い換えもあった。すなわち、連体節を主節から切り離したり、分裂文を無標文にする言い換えと同様に、主題展開を評価したり、前後文との修辭的關係が適切、かつ言い換え前から保存されていることの評価が必要であった。前後文脈との結束性、あるいは文脈との修辭的關係の適格性の評価については、文分割を題材に、再利用可能性を考慮した評価モデルの構築が始められており [10]、これを部品として実装することも考えられる。

また、修正知識の一部を、言い換え前後の等価性を保証するためのモデルとして示したが、こちらもモダリティ要素、文体など、比較的抽象度が高いレベルの等価性は容易に保証できると考えられる。しかし、言い換え前後の文の命題や内容語の持つ指示的意味が変化していない、ということを保証するということのハードルは高い。アプローチとしては、1) 「任意の単位のテキストの意味を計算する」と普遍的な意味を扱う、2) 「言い換え前後の意味の差分が所与の文脈で無視できるか否かを計算する」と近似的に扱う、の 2 つが考えられる。1) のアプローチを取るための枠組として、Generative Lexicon [13] がよく知られている。しかし、前提となる大規模な語彙知識は白紙の状態であるし、既存の辞書は言い換えを目的に編纂されたものではないため、そのまま持ち込むことはほとんどできない。我々は、2) のアプローチで、普通名詞に焦点を当て、任意の 2 語の意味の共通部分と差分を国語辞典の語釈文の共通部分と差分で近似する手法を提案した [2] が、必ずしも文脈に関わらない静的な言い換え対とするのは適切でないことが明らかになった。意味の計算に既存の言語資源を持ち込む際には、静的知識 (辞書) と動的知識 (処理部品) をうまく分類しなくてはならない。

5.3 どのように組み合わせるか

本稿では、部品を組み合わせで適切な言い換えを含む言い換え候補集合を生成することを目標としていた。しかし、実時間処理を考えると、不適格な言い換えを回避するために、何らかの制約に基づく戦

略的な生成を実現しなくてはならない。

Beale らは、文法的なパタンと単語の共起に基づく制約を与え、この制約を解消するような生成プランニングについて報告している [1]。これもテキスト生成に必要な処理をモジュール化し組み合わせる試みであるので、参考になる。実装を踏まえた上で、部品化によって従来の複雑な制約を排除しきれるのか。語彙的言い換えにおいては、どのような制約に基づいて生成プランをたてればよいかについて考える必要がある。

6 おわりに

本稿では、語彙的言い換えに必要な修正/棄却の知識に焦点を当て、先行研究、事例分析に基づいて部品を定義した。これらの部品は、今回、新聞記事を対象とし、いくつかの言い換えの種類を考えた限りでは、表 1 に示す通りある程度の再利用性を示すことができた。しかし、対象が替わったとき、新たな種類の言い換えを考えたときにどれだけ対応できるかの予測はついていない。しかし、修正処理の単位を細かくし、必要な知識、制約、処理を単純にしておけば、再利用もチューニングも容易になる可能性はある。

現状では、各部品の内部のモデリングについては検証が不十分であるし、本稿で示した部品化の手法、および部品の定義はあくまで「たたき台」に過ぎない。したがって、今後も方法論、部品の評価手法について検討を重ねる必要がある。

参考文献

- [1] Beale, S., Nirenburg, S., Viegas, E. and Warner, L. De-constraining text generation. In *Proceedings of the 9th International Workshop on Natural Language Generation (INLG 1998)*, pp. 48–57, 1998.
- [2] 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.
- [3] 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.
- [4] 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, 3H-03, 2001.
- [5] 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.
- [6] 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, NL-144-23, pp. 167–174, 2001.
- [7] 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.
- [8] 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.
- [9] Mitamura, T. and Nyberg, E. Automatic rewriting for controlled language translation. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS 2001) Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 1–12, 2001.
- [10] 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.
- [11] NTT コミュニケーション科学研究所. 日本語語彙大系. 岩波書店, 1997.
- [12] 織学, 佐藤理史. 日本語テキストの合成演算. 言語処理学会第 8 回年次大会発表論文集, pp. 367–370, 2002.
- [13] Pustejovsky, J. *The generative lexicon*. MIT Press, Cambridge, 1995.
- [14] 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. RWC, 1998.
- [15] 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.
- [16] Takahashi, T., Iwakura, T., Iida, R., Fujita, A. and Inui, K. KURA: a transfer-based lexico-structural paraphrasing engine. In *Proceedings of the Workshop on Automatic Paraphrasing: Theories and Applications, NLPRS2001 Post-Conference Workshop*, pp. 37–46, 2001.
- [17] 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.
- [18] *Workshop on Automatic Paraphrasing: Theories and Applications, NLPRS2001 Post-Conference Workshop*, 2001.

付録 A 修正部品の外部仕様

ここでは、定義した部品の外部仕様とその部品の適用例を示す。 C は局所的な適格性 (あるいは不適格性) に基づく部品の発動条件、 F は部品が担う処理内容である。部品が不適格であると判断しても修正できない場合は言い換えを棄却する (<棄却>)。各部品は、発動条件 C を満たすか否かを判定するために、対象パタンをすべて監視するものとする。

共起に基づく不適格性を解消するためには、単語を別の単語に言い換えるというアプローチも考えうる。これを実現するためには、修正部品からの言い

換え処理全体のカスケード呼び出し、または単語から単語への言い換えに関する知識を特別に扱う必要がある。今後、部品の内部仕様を具体的に定めるにあたって、このように図 1 のモデルで説明できない部分の拡張も検討が必要である。

A.1 言語モデル

語の活用形修正

C: 活用語 W_1 が語 W_2 に係るまたは接続するとき、 W_1 の活用形が不適切である。

F: W_1 と W_2 に基づいて、 W_1 の活用形を修正する。

例: 終わりますた ⇒ 終わりました

機能語の接続修正

C: 語 W と機能語 S が接続しえない。

F: S に相当し W と接続可能な機能語 S' があれば S を S' に置き換え、そうでないならば棄却する²。

例: 泳いている ⇒ 泳いでいる
意外の負け方 ⇒ 意外な負け方
比例される ⇒ <棄却>

格助詞の補完

C: 名詞 N が助詞を介さずに動詞 V に係っている。

F: V の格フレームに基づき、 N と V の間に適切な助詞を P を補完する。

例: 山田さんりんごを食べる ⇒ 山田さんがりんごを食べる

重複する格の整合

C: 動詞 V の格フレームにおいて、同じ格スロット P に対して 2 つのフィラー N_1, N_2 がある³。

F: N_1, N_2 の関係に基づいて、いずれかを選択するか名詞句 N_i の $N_j (i \neq j)$ を生成する。

例: タクシーに乗り物に乗る ⇒ タクシーに乗る
話をじゃまをする ⇒ 話のじゃまをする

兄弟格要素の共起

C: 動詞 V の格フレームにおいて、格スロット P_1 のフィラー N_1 と P_2 のフィラー N_2 が共起しえない³。

F: N_1 と N_2 のいずれかをそれと同義の語 N に言い換えることができなければ棄却する。

例: 涙が彼の目から流れた。 ⇒ <棄却>

² 「変えるべきだ流行 ⇒ 変えるべき流行」のように語を削除したり、逆に挿入したりする処理も考えられるが、現状ではこれらは仕様を含めていない。

³ 3 つ以上の要素が対象となる場合は同部品が複数回作用することになる。

修飾語同士の整合

C: ある語 W を修飾する 2 つの語 W_1, W_2 が共起しえない³。

F: W_1, W_2 のいずれかをそれと同義の語 W に言い換えることができなければ棄却する。

例: 来年以降も厳しく厳重に監視していくべきだ ⇒ 来年以降も { 厳しく; 厳重に } 監視していくべきだ

格と動詞の整合

C: 動詞 V の格フレームにおいて、 V が伴う接尾辞、あるいは V 自身が持つモダリティを考慮した上で、ある格スロット P のフィラー N が選択制限を満たしていない。

F: N が別の空いている格スロット P' の選択制限を満たすならば P を P' に置き換え、そうでなければ棄却する。

例: 現代社会に原因となっている ⇒ 現代社会が原因となっている
「三ちゃん農業」が兼業に農業を変えられませ ⇒ <棄却>

修飾語と動詞の整合

C: ある語 W が動詞 V を修飾しえない。

F: W をそれと同義の語 W' に言い換えることができなければ棄却する。

例: 就業条件が大きく悪くなる ⇒ 就業条件 { 極めて; 非常に; 実に; 大変 } 悪くなる

A.2 等価性保証モデル

内容語の指示的意味の整合

C: 内容語 C が新しく生成された。

F: 言い換え前の対応する語 C_s を同定し、指示的意味の差分が文脈に照らして無視できない場合は棄却する。

例: s. 昔は一つも通らなかった法案が連続して通った、t. ... 法案が継続して通った ⇒ <棄却>

テンス・アスペクトの整合

C: ある動詞 V について、テンス・アスペクト表現 T が一意に決まらない。

F: 原文中の V を参照し、テンス・アスペクト表現を修正する

例: s. せっかく宴会に行ったのにジュースしか飲まなかった、t. せっかく宴会に行ったのに飲まのはジュースだけだ ⇒ せっかく宴会に行ったのに飲んだのはジュースだけだった

文末表現の整合

C: 文末の述語 $Pred$ が新しく生成された。

F: 言い換え前の述語 $Pred'$ を参照し、文末表現を修正する。

例: s. 金利の変動を分析しろ、t. 金利の変動を細かく細かく調べる ⇒ 金利の変動を細かく調べる