

英日機械翻訳における自然な和文生成のための 英語名詞句の書き換え

吉見 毅彦[†]

英日機械翻訳システムによる翻訳に対して感じる不自然さの原因の一つとして、動詞的意味を含む英語の名詞句がそのまま日本語でも名詞句として訳されているということがある。この不自然さを解消するために本稿では、動詞的意味を含む名詞句を文に近い形式に書き換える自動前編集方法を示す。動詞的意味を含む名詞句のうち、属格名詞と of 前置詞句の両方を修飾句として持つ名詞句を主な対象として実験を行なった。提案方法によって書き換えた名詞句を含む文を我々のシステム Power E/J で処理し、書き換えを行わない場合の翻訳と比較したところ、67.3%の文においてより自然な翻訳が得られた。従来、この不自然さの問題に対しては、システム内部の変換過程で対処されることが多かった。従来の方法に比べて、前編集による方法の利点は、特定のシステムへの依存性が低く、実践上の適用範囲が広いことである。実験を通じて、市販されている幾つかのシステムにおいても、書き換えによってより自然な翻訳が得られることを確認した。

キーワード: 原文書き換え, 自動前編集, 名詞化形, 機械翻訳

Automatic Rewriting of English Deverbal Noun Phrases to Generate More Natural Translations by English-to-Japanese MT Systems

TAKEHIKO YOSHIMI[†]

One of the major factors causing unnatural translation in English-to-Japanese MT systems is the literal translation of verb-derived nominal constructions. This paper shows a method of automatically rewriting the nominalization (the packed form) into the one less packed, leading to the generation of a more natural and suitable Japanese. We have carried out an experiment mainly centered upon nominal constructions where the head deverbal noun is pre-modified by a genitive noun and post-modified by an “of” prepositional phrase. Having combined the proposed method with our system *Power E/J*, we found that the method improved the quality of translation for 33 out of the tested 49 sentences (67.3%) which include the rewritten constructions. In previously proposed methods, measures against the unnaturalness have often been taken at the transfer stage embedded in the systems. The advantage of our method over the previous approach is that it is applicable not only to our MT system but also to several other systems. An experiment with commercial MT systems other than ours shows that the incorporation of the pre-editing module has satisfactorily improved the quality of translations in two out of the three systems.

KeyWords: *Rewriting, Automatic Pre-editing, Nominalization, Machine Translation*

[†] シャープ (株) 技術本部 システム開発センター, Corporate Research and Development Group, System Technology Development Center, SHARP Corporation

1 はじめに

英語と日本語は、英語が名詞文法であり日本語が動詞文法であると言われるように、言語的特徴が著しく異なる言語である。このため、英語の名詞句をそのまま日本語の名詞句に直訳すると、違和感を感じる事が少なくない。例えば、文 (E1) を実用に供されているある英日機械翻訳システムで処理すると、文 (J1) のような翻訳が出力される。

(E1) The BOJ's buying of new government bonds is banned under fiscal law.

(J1) 新しい国債の BOJ の購入は、会計法の下で禁止される。

文 (E1) において名詞句 “The BOJ's buying of new government bonds” が伝える命題的な内容は、「BOJ が新しい国債を購入すること」であるが、日本語の名詞句「新しい国債の BOJ の購入」をこの意味に理解することは、訳文を注意深く読まなければ難しい。このような問題を解決するためには、英語と日本語の言語的特徴の違いを考慮に入れ、日本語として自然な表現が得られる処理を実現することが重要となる。

しかしながら、従来の機械翻訳研究では、主に原文解析の正しさに焦点が当てられており (成田一 2000)、訳文の自然さについてはあまり議論されてこなかった (吉村裕美子 平川秀樹 1995; Yamamoto 1999)。訳文の自然さに関する研究としては、文献 (長尾真 辻井潤一 1985; Somers, Hirakawa, Miike, and Amano 1988; Matsuo, Shirai, and Ikehara 1995) などがある。これらの文献に示されている方法では、直訳すると日本語として不自然になる英語の名詞句を適切に翻訳するための処理が原言語から目的言語への変換過程で行なわれる。

ところで、人間による翻訳では、文が表わす命題内容を含む名詞句を日本語に直訳した場合の違和感を解消するために、英語の名詞句を日本語に翻訳する前に、文またはそれに近い形式に言い換えるという処置がとられることがある (Nida, Taber, and Brannen 1973; 安西徹雄 1983)。

本研究では、人間のこの翻訳技法を機械的に模倣し、このような名詞句を前編集の段階で文に近い形式に自動的に書き換えることによって自然な訳文を生成することを試みる。日本語として自然な表現を得るための処理を変換過程で行なう方法に比べて、前編集の段階で行なう方法の利点は、前編集系は特定のシステムの内部に組み込まれていないため、システムへの依存性が低く、実践上の適用範囲が広いことである。実際、対象名詞句が現れる英文を提案手法によって書き換えて既存システムで翻訳し、元の英文の翻訳と比較する実験を行なったところ、我々のシステムだけでなく、市販されている他のシステムにおいても、より自然な翻訳が得られることが確認された。このことは、提案手法が様々なシステムの前編集系として利用可能であることを示している。

2 書き換えの分類

本節ではまず、書き換え前の表現(名詞的表現)が文中で果たす構文的機能と、書き換え後の表現の構文的機能が同じであるか異なるかという観点から書き換えを二種類に分類する。本稿では、書き換え前後の表現の構文的機能が変化しない書き換えを扱うが、名詞的表現の中でも動詞的性質を持つもの(後述)を書き換え対象とする。このため、次に、動詞的性質の強さの度合いに着目した名詞的表現の分類を示す。

2.1 書き換え前後の表現の構文的機能に着目した分類

書き換え前後の表現が文中で果たす構文的機能が異なる書き換えには、名詞句として機能している表現を副詞的に機能する表現に書き換えるものなどがある。例えば文(E2)のような無生物主語他動詞構文における主語の名詞句を、文(E2')のように、接続詞 *as* で導かれ理由を表わす副詞的従属節に書き換えることができる(別宮貞徳 1975; 安西徹雄 1983)。このような書き換えを行なうためには、構文構造を大きく変更する必要がある。

(E2) His three year's stay in London brought him a great progress in English conversation.

(E2') As he had stayed in London for three years, he made a great progress in English conversation.

書き換え前後の表現の構文的機能が変化しない書き換えでは、書き換え後の表現も名詞的に機能する。例えば文(E3)における名詞句“his absence from the party”は、文(E3')のように動名詞節“his being absent from the party”に書き換えることや、文(E3'')のように名詞的従属節“that he was absent from the party”に書き換えることができる¹。

(E3) We were disappointed at his absence from the party.

(J3) 我々は、パーティからの彼の不在に失望した。

(E3') We were disappointed at his being absent from the party.

(J3') 我々は、彼がパーティを欠席していることに失望した。

(E3'') We were disappointed that he was absent from the party.

(J3'') 我々は、彼がパーティを欠席していることに失望した。

本研究では、書き換え前後で構文的機能が変化しない書き換えだけを扱う。もちろん、このような制限を設けないほうが望ましいが、構文的機能が変化しない書き換えは、構文構造を大きく変える必要がないため、機械的な前編集による実現が容易である。また、名詞的表現への比較的簡単な書き換えを行なった上で処理するだけでも、書き換え前の文の翻訳に比べてより自然な翻訳が得られる例も少なくない。実際、文(J3')と文(J3'')は、書き換え後の文を我々の

¹ 一般には、元の名詞句の解釈と、書き換え後の動名詞節や *that* 節の解釈とが一致しないことがあり(Chomsky 1970; Thomason 1985; Siegel 1997)、常に書き換え可能であるとは限らない。

システムで実際に処理して得られた翻訳であるが、書き換え前の文の訳文 (J3) に比べてより自然な翻訳になっている。具体的にどの程度の文において翻訳品質の改善効果が期待できるかについては 3.3 節で述べる。

2.2 名詞化の階層

名詞的表現の中には、文が表わす命題内容を含むものがある。そのような名詞的表現と文の間には、動詞的性質の強さ (名詞的性質の弱さ) に応じて、名詞化の階層をいくつか設定することができる (Jelinek 1966; Lees 1968)。ここでは、Jelinek の分類に準じて、文から動詞由来名詞句までに次のような五段階の表現形式が可能であるものとする。文が最終的に動詞由来名詞を主辞とする名詞句に縮退していく過程で、動詞的性質は徐々に弱まり、逆に名詞的性質が徐々に強まっていく (時制や相、態、法などの動詞的範疇に対する制限が強くなっていく)。本研究は、以下に挙げる表現間での書き換えを実現することを目的とする。

縮退度 0 縮退度 0 の表現形式である文では、時制や相、態、法などのあらゆる動詞的範疇が明示される。

例 He refused our invitation to a party.

縮退度 1 縮退度 1 の表現は、接続詞 *that* などで導かれる名詞的従属節である。従属節は時制や法などの選択肢が制限された形式であり、時制の区別は時制一致の法則に従う。

例 We were disappointed *that he refused our invitation to a party.*

縮退度 2 縮退度 2 の表現を動詞的動名詞節と呼ぶ (安井稔 1982; Declerck 1994)。動詞的動名詞は、次のような構文的性質を持つ。(a) 冠詞や限定詞を伴うことができず (ただし属格は可能)、(b) 副詞による修飾が可能であるが形容詞による修飾は不可能であり、(c) 意味上の主語や目的語が前置詞 *of* で導かれず、(d) 時制の区別が可能であり、(e) 複数形が存在しない (乾亮一 1954; 太田朗 梶田優 1974; 安井稔 1982; Declerck 1994)。

例 We were disappointed at *his having refused our invitation to a party.*

縮退度 3 この縮退度の表現は名詞的動名詞句と呼ばれる。名詞的動名詞句は、動詞的動名詞節の構文的性質 (a) ないし (e) とすべての点で逆の性質を示す。

例 We were disappointed at *his refusing of our invitation to a party.*

縮退度 4 縮退度 4 は、接尾辞 *-ion* や *-ment*などを伴う動詞由来名詞を主辞とする名詞句である。また、接尾辞 *-ness*などを伴う形容詞由来名詞を主辞とする名詞句もこの縮退度の表現に分類される。

例 We were disappointed at *his refusal of our invitation to a party.*

3 名詞句の書き換え

機械翻訳システムの前編集でどの縮退度の表現をどの縮退度の表現に書き換えるべきかは、書き換え結果を入力として受け取るシステムに依存する²。以下、我々が実験に利用したシステムにおいて必要な書き換えについて述べる。

3.1 書き換え対象

我々のシステムでは、動詞的動名詞節 (縮退度 2) の翻訳には問題はないが、名詞的動名詞句 (縮退度 3) と動詞由来名詞句 (縮退度 4) は直訳されて不自然な翻訳となることが多い。このため、本研究では名詞的動名詞句と動詞由来名詞句を書き換え対象候補とする。さらに、訳文の不自然さの度合い (後述) を考慮して形式 (1) に適合する名詞句に候補を限定する。この候補の中から実際に書き換える名詞句を選択するための条件については、3.4 節で述べる。

$$[NP_1 's ADJ^? NOM \text{ of } NP_2] \quad (1)$$

形式 (1) の名詞句は、NOM を主辞とする名詞句である。NOM は、動名詞または動詞由来名詞を表す。NP₁'s は、属格名詞句または限定的所有代名詞である。形容詞 ADJ に付されている上付き記号[?]は一回以下の出現を意味する。形式 (1) の名詞句を含む文の例と、我々のシステムによる翻訳を以下に挙げる。各文において、斜字体の部分が書き換えるべき名詞句であり、ボールド体の部分がそれに対応する翻訳である。文 (E4) には名詞的動名詞句が含まれ、文 (E5) には動詞由来名詞句が含まれている。文 (E6) の動詞由来名詞句では動詞由来名詞を修飾する形容詞が存在する。

- (E4) Relocation of the Futenma air-base is totally unrelated to *Okinawa's hosting of the 2000 Group of Eight Summit*.
- (J4) Futenma 空軍基地の再配置は、沖縄の 2000 年先進 8 ヶ国サミットのホスティングに完全に無関係である。
- (E5) *Your fulfillment of this obligation* will be important to the renewal of our dealership agreement with you.
- (J5) この義務のあなたの達成は、あなたとの我々の販売権合意の更新にとって重要であろう。
- (E6) *Your negative consideration of our position* will not facilitate mutually rewarding marketing efforts.
- (J6) 我々のポジションのあなたの負の考慮は、相互に報いるマーケティング努力を促進しないであろう。

² しかし、4.2 節で示すように、本稿の書き換えは我々のシステムだけでなく他のいくつかの市販システムの前編集系としても有効である。

書き換えは、名詞的動名詞あるいは動詞由来名詞 NOM を主辞とする名詞句のすべてに対して行なう必要はなく、日本語に直訳した場合の問題が大きい名詞句に対してだけ行なえばよい。このため、次のような理由から主な書き換え対象を形式 (1) の名詞句とした。形式 (1) は、属格名詞句と of 前置詞句の二つの句が主辞 NOM に結合した名詞句を表わしている。NOM を主辞とする名詞句のうち形式 (1) 以外の名詞句としては、(a) “fulfillment of this obligation” や単に “fulfillment” のように、NOM に結合する句が一つ以下である名詞句や、(b) “the city’s destruction by the enemy” や “destruction of the city by the enemy” のように、NOM に結合する句が二つであってもそれらが属格名詞句と of 前置詞句の組み合わせでない名詞句などがある。形式 (1) の名詞句の翻訳は、次のようなことが原因で (a) や (b) の名詞句の翻訳よりさらに読みにくい。(a) や (b) の名詞句では、助詞「の」によって NOM に結び付けられる句は高々一つしか存在しない。例えば、(a) の “fulfillment of this obligation” の翻訳は「この義務の達成」となり、(b) の “the city’s destruction by the enemy” と “destruction of the city by the enemy” の翻訳は共に「敵による都市の破壊」となる。これに対して、文 (J4) ないし文 (J6) に示したように、従来システムで形式 (1) の名詞句を処理した場合には、「NP₁ の NP₂ の NOM」と翻訳され、NP₁ と NOM の関係及び NP₂ と NOM の関係は共に助詞「の」で示される。形式 (1) の名詞句では、NP₁ が NOM の意味上の主語、NP₂ が目的語であることが多い³が、このように翻訳されると、NP₁ と NP₂ のどちらが主語でどちらが目的語であるのかが認識しにくくなる。従って、形式 (1) のように、動詞由来名詞 NOM に属格名詞句と of 前置詞句の両方が結合している場合は、NOM を動詞に書き換えて、日本語として自然な翻訳にする必要性が高い。

なお、本稿の主な書き換え対象候補は形式 (1) の名詞句であるが、この他に形式 (2) の名詞句も扱う。

$$[\text{NP}_1\text{'s ADJ}^? \text{ NOM to ADV}^? \text{ VERB}] \quad (2)$$

形式 (2) における NOM は、to 不定詞を支配する動詞由来名詞、例えば attempt や failure, request などである。形式 (2) の名詞句を含む例を文 (E7) に示す。

(E7) *The manager's failure to properly supervise Palmer's dealings caused the losses.*

(J7) Palmer の取引を適切に監督することに関するマネージャの不履行は、損失を引き起こした。

3.2 動詞的動名詞節への書き換え

書き換えるの目的は、形式 (1) の名詞句において NP₁ が NOM の意味上の主語であり NP₂ が目的語であることが機械翻訳システムに認識できるような構造に書き換えることである。このためには、我々のシステムでは、動詞的動名詞節に書き換えればよい。

形式 (1) の名詞的動名詞句を動詞的動名詞節に書き換えるには、名詞的動名詞 NOM の直後

³ 3.4 節で述べるように、NOM が活動の結果などを表わす場合はこの限りではない。

の of を削除し、形容詞が存在する場合にはそれを副詞化する．例えば、文 (E4) に現れる名詞句 “Okinawa’s hosting of the 2000 Group of Eight Summit” は、“Okinawa’s hosting the 2000 Group of Eight Summit” に書き換える．このように書き換えた文 (E4’) からは、文 (J4’) のように、文 (J4) よりも自然な翻訳が得られる．

(E4’) Relocation of the Futenma air-base is totally unrelated to *Okinawa’s hosting the 2000 Group of Eight Summit*.

(J4’) Futenma 空軍基地の再配置は、沖縄が 2000 年先進 8 ヶ国サミットを主催することに完全に無関係である。

形式 (1) の動詞由来名詞句を書き換えるには、of の削除と形容詞の副詞化を行なう他に、動詞由来名詞をそれに対応する動名詞に書き換える．例えば、文 (E5) に現れる名詞句 “your fulfillment of this obligation” は、“your fulfilling this obligation” に書き換える．このように書き換えた文 (E5’) から得られる文 (J5’) は、文 (J5) より自然な翻訳である．

(E5’) *Your fulfilling this obligation* will be important to the renewal of our dealership agreement with you.

(J5’) あなたがこの義務を果たすことは、あなたとの我々の販売権合意の更新にとって重要であろう。

形式 (2) の名詞句についても、動詞由来名詞の動名詞への書き換えによって文 (J7’) のようなより自然な翻訳が得られるようになる．

(E7’) *The manager’s failing to properly supervise Palmer’s dealings* caused the losses.

(J7’) マネージャが Palmer の取引を適切に監督することができないことは、損失を引き起こした。

3.3 期待される改善度

名詞的動名詞句や動詞由来名詞句を動詞的動名詞節に書き換えることによって、どの程度の翻訳品質 (訳文の自然さ) の向上が期待できるのかをあらかじめ確認しておくために、訓練文集を作成し、形式 (1) または (2) の名詞句を手で動詞的動名詞節に書き換えた文を我々の実験システムで処理して、書き換え前の文の翻訳と比較した．評価値は品質の向上、低下、同等の三値とした．評価の実施は第三者一名に依頼した．

訓練文集の作成は、単語列 [$X Y^? Z$ (*of/to*)] を含む文を文字列照合によって新聞記事から抽出することによって行なった． X は、アポストロフィ s で終わる単語 (属格名詞と想定) または my などの単語 (限定的所有代名詞) と適合する． Y は形容詞であることを想定している． Z は -al, -ance, -ing, -ion, -ment, -sis で終わる単語、または、予備調査において比較的高い頻度で出現した語 attempt, failure, pledge, request のいずれかである．

この単語列に適合し、かつ品詞の条件を満たす (Z が動名詞あるいは動詞由来名詞であり、

もし Y が存在するならば形容詞である) 表現を含む文として, 218 文が抽出された. このうち 30 文は, 書き換え前の文の翻訳品質が翻訳の自然さを比較するのに適切ではなかったため, 除外した. 残りの 188 文のうち構文的, 意味的に書き換え可能であるものは 95 文であった. 95 文についての評価結果を表 1 に示す. 95 文の 57.9% に当たる 55 文において, より自然な翻訳が得られている. このことから, 動詞的動名詞節への書き換えを正しく行なうことができれば, その効果は高いと考えられる. なお, 31 文の翻訳品質が低下しているが, この原因は, 動詞由来名詞句や名詞的動名詞句を動詞的動名詞節に書き換えると, 構文構造がシステムにとって複雑になりすぎ, 正しく認識できなかったことにある.

表 1 書き換えによって期待される翻訳品質の改善度

向上	同等	低下
57.9%(55/95)	9.5%(9/95)	32.6%(31/95)

3.4 書き換え条件

3.2 節で述べた書き換え操作の適用条件は, 以下で述べるように, 書き換え対象名詞句の文中での構文的機能に関する制約と, 動詞由来名詞の意味に関する制約から主に構成される.

3.4.1 名詞句の構文的機能に関する制約

名詞的動名詞句と動詞由来名詞句は, 主語, 補語, 動詞・前置詞の目的語として機能する. これに対して, 動詞的動名詞節は, 主語, 補語としては動詞の種類によらず機能するが, 動詞・前置詞の目的語になれるかどうかは動詞や前置詞の種類に依存する⁴. このため, 書き換えが構文的に可能であるかどうかを判定しなければならない. このような判定を厳密に行なうためには構文解析が必要であるが, ここでは次のように近似的に判定する.

- (1) 書き換え対象名詞句の直前に be 動詞以外の動詞が存在すれば, それが主語付き動詞的動名詞節を目的語として支配しうる場合に限り, 動詞的動名詞節に書き換える.
- (2) 書き換え対象名詞句の直前に be 動詞か前置詞が存在すれば, 対象名詞句をそれぞれ be 動詞の補語, 前置詞の目的語とみなして動詞的動名詞節に書き換える.
- (3) 書き換え対象名詞句の前方 3 語以内に動詞が存在しなければ, 対象名詞句を主語とみなして書き換える.

⁴ この動詞的動名詞節は, 形式 (1) の名詞句を書き換えたものであるため, 意味上の主語 NP₁ を伴う. 主語付き動詞的動名詞節は after や before など時間を表わす前置詞の目的語にはならない (Quirk, Greenbaum, Leech, and Svartvik 1985).

判定手続き (1) において、動詞が主語付き動詞的動名詞節を支配できるかどうかは、Hornby の分類 (Hornby 1977) を拡張した動詞型に基づいて判断する。

書き換え対象名詞句の直前が前置詞である場合、前置詞の種類による制約を考慮せずに書き換えを行なう。前置詞の種類を考慮しないと、英語の適格性に問題が生じることがあるが、本研究では適格な英語表現に書き換えることは二次的な目的であり、主な目的は書き換え結果を入力として受け取る機械翻訳システムによって自然な日本語表現が生成されるように英語表現を書き換えることである。実験に用いたシステムでは、動詞的動名詞節が前置詞の目的語になりうる位置に出現した場合、前置詞の種類によらず、動詞的動名詞節は前置詞の目的語と解釈されるため、構文解析に失敗しない。このため、システムのこの性質を利用して、前置詞の種類によらず動詞的動名詞節への書き換えを行なうことにする。

3.4.2 動詞由来名詞の意味に関する制約

文は 2.2 節で述べた過程を経て動詞由来名詞句 (縮退度 4) に縮退しうる。名詞的動名詞句 (縮退度 3) までの段階では、動詞の特徴を残して出来事や活動、行為を表わすが、動詞由来名詞句になると、この他に、行為の結果や場所、原因など様々な具体的あるいは抽象的なモノを表わすようになる (Grimshaw 1990; 影山太郎 1999)。例えば、organization には「組織すること」の他に「協会」の意味があり、decoration には「飾ること」か「勲章」かという曖昧性があり、destruction には「破壊の原因」という意味も含まれる。ここで、文献 (影山太郎 1999) に倣い、動詞由来名詞が出来事や行為を表わす場合をデキゴト名詞と呼び、具体的あるいは抽象的なモノを表わす場合をモノ名詞と呼ぶことにする。

動詞由来名詞が動詞的意味を表すデキゴト名詞である場合は、動詞由来名詞句を動詞的動名詞節に書き換えることができるが、モノ名詞である場合にはできない。このため、動詞由来名詞がデキゴト名詞であるかモノ名詞であるかを判定する必要がある。この判定に利用できる手がかりとしては、動詞由来名詞句の内部から得られるものと、外部から得られるものがある。名詞句内部からの手がかりを利用する方法 (Hull and Gomez 1996) では、動詞由来名詞に対応する動詞の格パターンの各スロットに記述されている共起制約を主語 NP₁ と目的語 NP₂ が満たせばデキゴト名詞であると判定され、満たさなければモノ名詞であると判定される。他方、外部からの手がかりとしては、動詞由来名詞と、それを支配する動詞やその動詞に支配されている他の格要素との意味的な関係がある (Vendler 1967)。

しかし、本研究のように前編集の段階では、このような手がかりを正確に得ることは容易ではない。また、動詞と動詞由来名詞との意味的な対応関係は体系的なものではなく、単語ごとに個別である (太田朗・梶田優 1974)。このため、次のような素朴な方法でモノ名詞かデキゴト名詞かを区別する。すなわち、動詞由来名詞が形式 (1) の名詞句の主辞になっている場合に、動詞由来名詞の解釈として、(a) 文脈によらずデキゴト名詞が優勢か、(b) 文脈によらずモノ名

詞が優勢か、(c) どちらの解釈かが文脈により決定されるかを主観的に判断する。(a) に該当する動詞由来名詞は書き換え対象候補とし、(b) の動詞由来名詞は対象外とする。(c) については、書き換え漏れを減らすことよりも不適切な書き換えを抑えることを重視するという方針の下で実験を繰り返して経験的に書き換え対象候補に追加していく。このような方針に基づいて、接尾辞 -al, -ance, -ion, -ment, -sis を伴う動詞由来名詞に、訓練文集において比較的高い頻度で出現した語 attempt, failure, pledge, request を加えて、786 語を書き換え対象候補とした。

3.4.3 その他の制約

3.2 節の書き換えは、形式 (1) の名詞句において NP_1 が NOM の意味上の主語であると解釈できると仮定して行なうものであるが、 NP_1 が時間を表わす場合、この仮定が成り立たないことがある。例えば “last week’s revival of Super 301” という名詞句では、“last week” は “revive” の主語とは解釈できない。一般には、時間名詞でも NOM の主語と解釈し、書き換えるべき場合もありうるので、 NP_1 が NOM の主語になりうるかどうかは、 NP_1 と NOM に対応する動詞との共起的意味解析などによって決定する必要がある。しかし、ここでは、ごく簡単な処理によって不適切な書き換えが抑えられればよいという立場から、時間名詞は書き換えないことにする。 NP_1 が時間名詞であるかどうかの判定は、辞書に記述されている意味標識に基づいて行なう。

また、書き換え対象候補の名詞句が連語や慣用句のように固定的な表現を構成する場合にも書き換えない。

4 実験と考察

提案手法の有効性を検証するために、訓練文集の場合と同じ方法 (3.3 節) で選別した 227 文を対象として実験を行なった。このうち書き換え可能な文は 93 文である。

書き換えの評価は、次の二点について行なう。

単体精度 提案手法自体の精度であり、英語の適格性の観点から、書き換えるべき表現がどれだけ書き換えられ、書き換えるべきでない表現がどれだけそのまま残されたかを表わす。

実用精度 提案手法と機械翻訳システムを組み合わせた場合の精度であり、英日翻訳の観点から、書き換えられた文のうちどれだけの文において翻訳品質が実際に向上したかを表わす。

英語の適格性の観点からの評価は英国人一名に依頼した。英日翻訳の観点からの評価では、3.3 節での事前評価者を主評価者とし、副評価者を二名加えた (いずれも日本人)。

4.1 書き換えの妥当性の評価

提案手法自体の評価結果を表2に示す．表2によれば，71.8%の精度が得られており，ほぼ妥当な書き換えが行なえていると考えられる．

表2 提案手法自体の精度

正解との比較	提案手法		精度
	書き換えた	書き換えず	
一致	17.2%(39/227)	54.6%(124/227)	71.8%
不一致	4.4%(10/227)	23.8%(54/227)	28.2%

書き換えるべきでない表現が誤って書き換えられた10文について，その原因を分析した．10文のうち8文は，書き換え対象名詞句の直前が前置詞である場合に前置詞の種類による制約(脚注4参照)を考慮していないために，不適格な英文となっていた．しかし，この書き換えは，3.4.1節で述べたように意図的に行なっている処理であるので，適格な英語表現に書き換えることに主眼を置くならば，抑えることができる．前置詞の種類による制約を加えれば，精度は75.3%にまで上がる．

残りの2文は，動詞的動名詞節に書き換えると構文構造が複雑になりすぎるため，書き換えない方がよいと考えられるものであった．このような場合に書き換えを行なわないようにするためには，まず構文的複雑さを測る尺度を明確にする必要がある．しかし，たとえそのような尺度が定められたとしても，それを前編集の段階で利用することは容易ではないため，この問題は提案手法の限界を越えている．

4.2 翻訳品質改善度の評価

提案手法によって書き換えられた文を機械翻訳システムで処理した場合，書き換えを行わない場合に比べて翻訳品質がどの程度向上するかを検証する．提案手法によって書き換えられた文のうち4.1節の評価で英語の適格性に問題があるとされた文は前置詞の問題と構文的複雑さの問題であったが，これらの問題は機械翻訳システムでは問題にならないこともあるので，これらの文を評価対象から除外はしないことにする．

4.2.1 我々のシステムでの改善度

我々の実験システムでの翻訳品質の評価結果を表3に示す．表3によれば，33文(67.3%)で品質改善が見られ，比較的簡単な書き換えでほぼ有効な結果が得られている．なお，評価値は評価者三名の多数決によるものである．ただし，判定が三つに分かれ多数決で決まらない場合

は、主評価者の評価値を採用した⁵。

表 3 機械翻訳としての精度

向上	同等	低下
67.3%(33/49)	8.2%(4/49)	24.5%(12/49)

4.1節の表2より、書き換えるべき表現が書き換えられなかった書き換え漏れは、54文存在した。このうちもし書き換えられていれば翻訳品質が向上したのは27文であった。この27文について、書き換え漏れの原因を分析した結果を表4に示す。

表 4 書き換え漏れの原因

原因	文数
構文的制約	13
意味的制約	9
その他の制約	5

構文的制約による書き換え漏れは、3.4.1節で述べた処理の不備によるものである。最も多かった原因は、書き換え対象名詞句の直前に存在する語(着目語)が following や including のように動詞か前置詞かの曖昧性を持つが対象文中では前置詞として機能していることを正しく認識できなかったことにある。3.4.1節の処理では、着目語が動詞か前置詞かの曖昧性をもつ場合を適切に考慮していないため、following などが実際には前置詞として機能している場合でも、主語付き動詞的動名詞節を目的語として支配できないものとして手続き(1)で書き換えが行われなくなる。

意味的制約による書き換え漏れは、動詞由来名詞が3.4.2節で選別した書き換え対象候補に加えられていなかったことによる。書き換え漏れの件数は、慎重な方針で書き換え対象候補を選んだわりには予想外に少なかった。この一因として、訓練文集と試験文集が、同じ種類のコーパス(共に新聞記事)から作成したものであることが挙げられる。今後、他の種類のテキストを対象として実験を行い、書き換え対象候補の検討を行なう必要がある。また、書き換え対象候補の動詞由来名詞をあらかじめ固定しておくのではなく動的に選択する方法などについても検討が必要である。

⁵ 49文のうち、評価者全員の判定が一致したものが32文、多数決により決定されたものが15文、三名の判定が分かれたものが2文であった。

4.2.2 他の市販システムでの改善度

前編集の段階で原文を書き換える方法は、特定のシステムにあまり依存しないため、我々のシステムに限らず他のシステムにおいても品質改善効果があると予想される。この点を確認するために、提案手法を、市販されている三つのシステムの前編集系として利用した場合としない場合で品質がどのように変化するかを調べた。

我々のシステムを評価した際の評価基準と同じ基準で主評価者により評価された結果を表5に示す。表5によると、システムCでは書き換えによって品質が極端に悪化している⁶が、システムAとBでは書き換えによる品質改善の効果が見られる。特にシステムAでは我々のシステムの場合に近い効果が出ている。このことから、提案手法は特定のシステムへの依存性が低く、複数のシステムの前編集系として利用可能であるといえる。

表5 他の市販システムでの翻訳改善度

	向上	同等	低下
システムA	63.3%(31/49)	12.2%(6/49)	24.5%(12/49)
システムB	51.0%(25/49)	16.3%(8/49)	32.7%(16/49)
システムC	12.2%(6/49)	8.2%(4/49)	79.6%(39/49)

5 関連研究

動詞由来名詞句に関する工学的研究としては、文献 (Hobbs and Grishman 1976; Dahl, Palmer, and Passonneau 1987; Somers et al. 1988; Hull and Gomez 1996; Macleod, Grishman, Meyers, Barrett, and Reeves 1998) などがある。このうち、本研究と同様に英日機械翻訳の立場からの研究は Somers らの研究である。

Somers らは、動詞由来名詞句を自然な日本語表現に翻訳するための方法として、英日両言語から中立な内部表現を利用する方法と、変換過程での明示的な構造変換による方法を検討している。これらはいずれも必要な処理をシステム内部で行なうものである。このような方法に比べて、本稿のような前編集による方法は、後編集による方法 (Knight and Chander 1994; Yamamoto 1999; 尾崎正行, 荒木健治, 柄内香次 2001) と同様に、システムからの独立性が高いため、様々なシステムで利用可能であり、実際にシステムと組み合わせる際にシステムの既存部分を修正する必要はほとんどない。

⁶ システムCで悪化した39文で見られた現象の主な内訳は、(a) 例えば “Ieyasu's establishment of the Tokugawa shogunate” という名詞句を書き換えた動名詞節 “Ieyasu's establishing the Tokugawa shogunate” が「徳川の将軍職を確立する Ieyasu」と訳されるなど、形式(1)または(2)の名詞句を書き換えると、NP₁が書き換え後の表現の主辞と解釈されているものが21文(53.8%)、(b) 書き換え対象名詞句自体の翻訳品質には問題ないが、書き換えの影響で他の部分の品質が悪化しているために、文全体としては「低下」となったものが7文(17.9%)などである。

Hullらは、動詞由来名詞が動詞的意味を表すデキゴト名詞であるか、非動詞的意味を表すモノ名詞であるかの判定を行い、さらに、デキゴト名詞であると判定された場合、動詞由来名詞の語義(対応する動詞の格パターン)を決定する方法を示している。これに対して、本稿では、デキゴト名詞かモノ名詞かの判定は行なう(デキゴト名詞かモノ名詞かをあらかじめ区別しておく)が、動詞由来名詞に対応する動詞の語義を決定することは機械翻訳システム本体に委ねるという方針を採っている。

Macleodらは、動詞由来名詞と動詞の対応情報を記述した辞書を構築している。この辞書には、動詞の補足語が動詞由来名詞句では属格となるのかあるいはどのような前置詞を伴うのかなどが記述されている。本稿では書き換え対象を形式(1)に限定したが、この形式以外の表現では、(a) 動詞由来名詞がデキゴト名詞かモノ名詞かの曖昧性や、(b) 動詞由来名詞とその補足語との結合関係の曖昧性、(c) 動詞由来名詞の後方に存在する前置詞句の付加の曖昧性などが生じやすい。Macleodらの辞書を利用すればこのような問題にある程度対処できるので、書き換え対象を拡げること可能であろう。

6 おわりに

本稿では、直訳すると不自然な日本語表現になる英語の名詞的表現を文に近い表現形式に書き換えることによって翻訳品質を改善する方法を示した。小規模ではあるが実験を行なったところ、簡単な書き換えによって比較的良好な結果が得られ、提案した方法の有効性が確認された。また、提案手法は特定のシステムに依存せず、複数のシステムの前編集系としても有効に機能することが実験的に確かめられた。

今回の実験では形式(1)または(2)の名詞句を動詞的動名詞節へ書き換える処理を扱ったが、今後は書き換え対象名詞句の拡張や動詞的動名詞節以外の表現形式への書き換えも行なっていく。さらに、名詞的表現を副詞的表現に書き換えるなどのより大きな構造変換を伴う書き換えを前編集の段階でどの程度実現できるかを検討していく必要もある。

謝辞

議論に参加頂いた英日機械翻訳グループの諸氏に感謝します。また、本稿の改善に非常に有益なコメントを頂いた査読者の方に感謝いたします。

参考文献

- 安西徹雄(1983). 英語の発想. 講談社.
別宮貞徳(1975). 翻訳を学ぶ. 八潮出版社.
Chomsky, N. (1970). "Remarks on Nominalization." In Jacobs, R. and Rosenbaum, P. (Eds.), *Readings in English Transformational Grammar*, pp. 184-221. Ginn and Company.

- Dahl, D., Palmer, M., and Passonneau, R. (1987). “Nominalizations in PUNDIT.” In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pp. 131–139.
- Declerck, R. (1994). 現代英文法総論. 開拓社. 安井稔 訳.
- Grimshaw, J. (1990). *Argument Structure*. Linguistic Inquiry Monographs. The MIT Press.
- Hobbs, J. and Grishman, R. (1976). “The Automatic Transformational Analysis of English Sentences: An Implementation.” *International Journal of Computer Mathematics, Section A*, **5**, 267–283.
- Hornby, A. S. (1977). 英語の型と語法. オックスフォード大学出版局. 伊藤健三 訳注.
- Hull, R. and Gomez, F. (1996). “Semantic Interpretation of Nominalizations.” In *Proceedings of the 13th National Conference on Artificial Intelligence and the 8th Innovative Applications of Artificial Intelligence Conference*, pp. 1062–1068.
- 乾亮一 (1954). 分詞・動名詞. 研究社.
- Jelinek, J. (1966). “Construct Classes.” *Prague Studies in Mathematical Linguistics*, **2**, 167–182.
- 影山太郎 (1999). 形態論と意味. 日英語対照による英語学演習シリーズ 2. くろしお出版.
- Knight, K. and Chander, I. (1994). “Automated Postediting of Documents.” In *Proceedings of the 12th National Conference on Artificial Intelligence*, pp. 779–784. <http://xxx.lanl.gov/abs/cmp-lg/9407028>.
- Lees, R. (1968). *The Grammar of English Nominalizations* (Fifth Printing edition). Mouton.
- Macleod, C., Grishman, R., Meyers, A., Barrett, L., and Reeves, R. (1998). “NOMLEX: A Lexicon of Nominalizations.” In *Proceedings of EURALEX'98*. <http://cs.nyu.edu/cs/projects/proteus/nomlex/index.html>.
- Matsuo, Y., Shirai, S., and Ikehara, S. (1995). “Changing Syntactic Classes in Transfer-based Machine Translation.” In *Proceedings of the 3rd Natural Language Processing Pacific Rim Symposium*, pp. 432–437.
- 長尾眞 辻井潤一 (1985). “機械翻訳における訳語選択と構造変換過程.” *情報処理*, **26** (11), 1261–1270.
- 成田一 (2000). “英日・日英機械翻訳の実力.” 言語処理学会第6回年次大会発表論文集, pp. 51–54.
- Nida, E., Taber, C., and Brannen, N. (1973). 翻訳—理論と実際. 研究社出版. 沢登春仁, 升川 潔 訳.
- 太田朗 梶田優 (1974). 文法論 II. 英語学大系 4. 大修館書店.
- 尾崎正行, 荒木健治, 栃内香次 (2001). “帰納的学習を用いた自然な日本語文生成手法の評価.”

研究報告 NL141-10, 情報処理学会.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.

Siegel, L. (1997). "Gerundive Nominals and the Role of Aspect." In *Proceedings of ESCOL'97*. <http://citeseer.nj.nec.com/228617.html>.

Somers, H., Hirakawa, H., Miike, S., and Amano, S. (1988). "The Treatment of Complex English Nominalizations in Machine Translation." *Computers and Translation*, **3**, 3-21.

Thomason, R. (1985). "Some Issues concerning the Interpretation of Derived and Gerundive Nominals." *Linguistics and Philosophy*, **8**, 73-80.

Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press.

Yamamoto, K. (1999). "Proofreading Generated Outputs: Automated Rule Acquisition and Application to Japanese-Chinese Machine Translation." In *Proceedings of the 18th International Conference on Computer Processing of Oriental Languages*, pp. 87-92.

安井稔 (1982). 英文法総覧. 開拓社.

吉村裕美子 平川秀樹 (1995). "理解容易性を指向した訳語/統語構造選択規範に基づく文生成." 情報処理学会論文誌, **36** (2), 247-257.

略歴

吉見 毅彦: 1987年電気通信大学大学院計算機科学専攻修士課程修了. 同年よりシャープ(株)にて機械翻訳システムの研究開発に従事. 1999年神戸大学大学院自然科学研究科博士課程修了.

(2001年4月20日 受付)

(2001年6月7日 再受付)

(2001年6月29日 採録)