辞書語義文を利用した対訳辞書の拡充*

6M−06

釜谷 聡史　小川 泰弘　稲垣 康善†

名古屋大学大学院工学研究科§
kamatani@inagaki.nuie.nagoya-u.ac.jp

1 はじめに

日本語とウイグル語は，共に膠着語に分類され，また，語順がほぼ同じであるなど，構文的類似性が高いという特徴がある．そこで，日本語－ウイグル語間の機械翻訳においては，構文的類似性を利用し，形態素解析結果を逐語訳する手法が考えられ，それに基づく日本語－ウイグル語機械翻訳システム [4] が，我々の研究室で開発されている．しかし，一般に，辞書の構築には多大な時間と人手を要し，構築コストが高く，翻訳に用いられる辞書が充実していないために，翻訳できない事例がしばしば見られる．

本研究では，辞書語義文を用いて，既に対訳辞書に存在する単語に言い換えることで，日本語－ウイグル語対訳辞書を拡充する．本稿では，特に日本語の「サ変名詞＋する」の言い換え獲得と辞書の拡充について述べる．

以下では，言い換える前の見出し語を言い換え元と呼び，言い換え結果を言い換え先と呼ぶ．

2 言い換え獲得

辞書の語義文を利用して言い換え先の候補を獲得することを考える．実際の手法としては，語義文が類似しているもの同士を言い換え対としてみる手法 [1] と，語義文内の語，あるいは句に，見出し語を直接言い換える手法の2つが考えられる．本研究では，より多くの言い換えを獲得できる，後者の手法を基本とする．

しかし，語義文に直接言い換えると，表現が冗長になり，翻訳にはそぐわない場合がある．そこで，以下で述べる規則を用い，言い換え先として，語義文の中で必要な部分を過不足なく取り出す．この時，言い換え先を獲得するためのデータとして，EDR 日本語単語辞書 [3] から抜き出した，見出し語と日本語概念説明 (語義文) の組を用いた．

2.1 獲得規則

言い換えを獲得する際に，語義文中の次のような語，あるいは節を除外する．

[規則 **1**] 「末尾の動詞に係るガ格」の削除
例 **1**：収納する
語義文　:“役所が金銭を受納する”
⇒ 言い換え:“金銭を受納する”

[規則 **2**] 「動詞＋ために」の削除
例 **2**：足継ぎする
語義文　:“高くするために足を継ぎ足す”
⇒ 言い換え:“足を継ぎ足す”

[規則 **3**] 「動詞 or 名詞＋など＋ (格助詞)」の削除
例 **3**：解毒する
語義文　:“体内で毒物の働きを薬などでなくする”
⇒ 言い換え:“体内で毒物の働きをなくする”

[規則 **4**] 「名詞＋において」の削除
例 **4**：返球する
語義文　:“野球などの球技においてボールを投げ返すこと”
⇒ 言い換え:“ボールを投げ返す”

[規則 **5**] 読点より前の文を削除
例 **5**：有人飛行する
語義文　:“宇宙船などが，人間を乗せて飛ぶこと”
⇒ 言い換え:“人間を乗せて飛ぶ”

[規則 **6**] 文末の「こと」「さま」の削除

2.2 共起情報を利用した規則

見出し語「相乗りする」を例に考える．「相乗りする」は，「自転車に相乗りする」のように，ニ格を伴って現れる事例が多く見られることが期待される．しかし，語義文は“一つの乗り物にいっしょに乗る”であり，これをそのまま言い換え先とすると，「自転車に一つの乗り物にいっしょに乗る」となってしまう．よって，言い換え先にはニ格をとるものを含まないように，即ち，“一つの乗物に”を含めないようにすべきである．本研究では，次の手法を用いて，これを解決する．

1. 言い換え先の末尾の動詞に係る名詞を探査．
2. 言い換え元の語 W について，次式により値 θ を計算．

$$\theta = \frac{\text{コーパス中で語 W が C 格を伴って現れた頻度}}{\text{コーパス中の語 W の出現頻度}}$$

3. θ が，ある閾値よりも低ければ，語義文中において C 格で共起する語を削除し，言い換え先に含めない．

この時，共起情報は EDR 日本語共起辞書 [3] によった．

2.3 係り受けを利用した削除

可能な限り過不足無く，語義文から言い換え先を取り出したい．そこで，語義文を係り受け解析しておき，前節で述べたような削除作業の際に，削除される語に係る語も併せて削除する．本稿では，語義文の係り受け解析に KNP[2] を用いた．係り受け解析結果を用いた削除は，前節までに挙げた規則適用後に行なう．

以上の言い換え規則を適用した結果の例を，図 1,2に示す．図中の破線は削除された係り受け，実線は言い換え先中の係り受け，太枠は言い換え先をそれぞれ表す．

---

* **Dictionary expansion using lexical descriptions**
† Satoshi Kamatani , Yasuhiro Ogawa and Yasuyoshi Inagaki
§ Graduate School of Engineering,Nagoya University

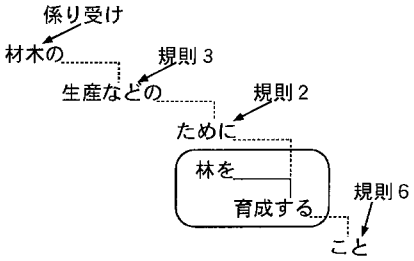

図 1: 「育林する」の言い換え

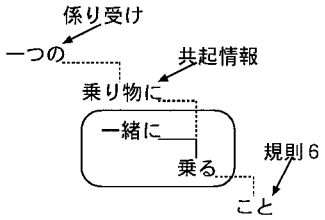

図 2: 「相乗りする」の言い換え

表 1: 言い換え獲得結果

言い換え元語数	言い換え獲得数	訳語獲得数
24,208	22,612	6,922

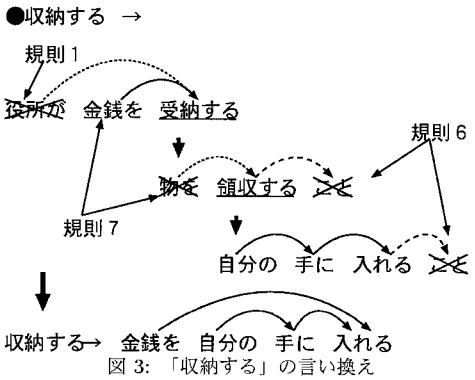

図 3: 「収納する」の言い換え

3 実験

本実験の目的は, 訳語を持たない単語を翻訳可能な語句に言い換えることである. そこで, EDR 日本語単語辞書中のサ変名詞 27,108 語から, 日本語−ウイグル語対訳辞書 (総登録単語数: 約 24,000 語) に登録済みのサ変名詞 2,900 単語を実験対象から外した. 残りの 24,208 語について, θ の閾値を 0.5 とし, 言い換えを獲得した.

3.1 結果と考察

結果を表 1に示す. 言い換え獲得数は, 言い換え元と違う表現に言い換えられた単語を示す. EDR 日本語単語辞書の概念説明では, 見出し語と概念説明が同じ場合があり, この様な場合は言い換えられなかったと判断した. また, 訳語獲得数とは, 言い換え結果が既存の対訳辞書のみで全て翻訳可能であったものを示す.

言い換え元の総数は 24,208 語であり, 内, 新たに翻訳が可能になった, すなわち, 辞書に記載がある語に言い換えられた語は, 6,922 単語である. これらの内, 言い換え先が動詞で終っていない, 意味が十分保存されていないなど, 望ましい言い換えではなかったものが 229 例あった. 229 例中,167 例は必要なガ格で出現した名詞を削除した例であった. また,1,610 例で, 「駐車する」 ⇒ “車を止めて置く” のように, 言い換えとしては正しいが, 実際に翻訳に用いると, 「車を車を止めて置く」のようになるなる可能性があるものが見られた. これは, 共起情報を利用した削除が十分行なわれなかったためである. 改善作として, 削除の根拠とする共起情報を増やす必要がある.

既に日本語−ウイグル語対訳辞書に登録されていた単語, 2,900 語に加えて, 望ましい言い換えではなかった 229 例を除いた 6,693 語が, 新たに翻訳可能な単語となり, 総計 9,593 単語が翻訳可能な単語となった.

4 複数回の言い換え

言い換えが獲得できたもののうち,15,690 語が日本語−ウイグル語対訳辞書中の語で完全に言い換えられなかったものである. 対訳がない語を再度言い換えることによって, さらに訳語を得ることができる. 言い換え結果を走査し, 対訳辞書中に載っていない語を再度言い換える. この時, 次の規則が更に必要である.

[規則 7] 「同じ格で係る語」の削除

実際の言い換えの例を図 3に示す. 1 回目の言い換えで得られた “金銭を受納する” で, 「受納する」が対訳辞書にない場合, 更に言い換える必要がある. 今, 「受納する」の語義文が “物を領収すること” であるとする. この時, 「金銭」=「物」であることから「物を」を削除することが望ましい. そこで, 規則 7 を適用し, 言い換え先から「物を」を削除する. さらに, 規則 6 を適用し, 最終的な言い換え “領収する” を得る. 以下同様に辞書に記載のない単語が見つかるたびに言い換えを繰返し, 翻訳可能な言い換え先を獲得する.

例のように言い換える対象は動詞だけではなく, 他の品詞も考える必要があるが, これらも動詞と同様に考えることができる. 言い換え操作を複数回重ねれば, 更に多くの単語について翻訳が可能になることが期待できる.

5 おわりに

本稿では, 辞書語義文を用いて, 言い換えを行うことにより, 対訳辞書を拡充した. 今後は, 更に質の良い言い換えを獲得すると共に, 他の品詞の単語についても同様の処理を加えることで, 対訳辞書の更なる拡充を図る. また, 複数回の言い換えによる拡充の効果についても検証する.

参考文献

[1] 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 (2000).

[2] 黒橋 禎夫, “日本語構文解析システム KNP version2.0 b6 使用説明書”, 京都大学大学院情報学研究科, (1998).

[3] 日本電子化辞書研究所, “EDR 電子化辞書仕様説明書”, (1996).

[4] 小川泰弘, ムフタル・マフスット, 杉野花津江, 外山勝彦, 稲垣康善, “派生文法に基づく日本語動詞句のウイグル語への翻訳”, 自然言語処理, Vol.7,No.3, pp.57–78, (2000).