

5U-04 類似表現検索において周辺文脈を考慮することの効果

熊野 正[†]

後藤 功雄[‡]

浦谷 則好[†]

[†](株) ATR 音声言語通信研究所

[‡]NHK 放送技術研究所

1 はじめに

我々は以前より、ニュース原稿の日→英翻訳現場における翻訳者支援技術の研究を続けている。この研究の一環として開発した翻訳用例提示システム [1, 2] をこれまで試験的に運用してきたが、2000年6月よりNHK国際放送局英語センターにて正式に運用を開始した。これは、過去に翻訳した英語原稿を元となる日本語原稿とともに蓄積し、ユーザの日英表現の入力に応じて類似する表現を含む日英記事対を提示するものである。システムは、ユーザの入力表現から内容語をキーワードとして抽出し、キーワードの出現順序を考慮したAND検索を行う。検索結果は、入力表現におけるキーワード間距離との変位を尺度として順位づける [4]。

本システムは、日英記事対を提示するとき、検索結果の表現とその訳のみをユーザに与えるのではなく、それらを含む記事全体を提示する。これは、ユーザがそれらの表現の記事全体との関わりの中で参照できることを意図したものである。これによって本システムは、単にある表現の訳を調べるだけでなく、前後の文脈による訳のバリエーションを調べたり、ある話題に関する記事を調べるなど、幅広い用途に利用されている。

しかし、そのためにユーザのシステムへの要求意図は自明ではなく、また多くの場合、ユーザの入力する表現はユーザの(暗黙・無意識の)要求を満たす検索を行うには短すぎて情報不足であることがわかってきた。例えば、「森首相」という入力はそれそのものの訳を知りたいという意図でないことはほぼ明白であるが、ユーザがどのような「森首相」入りの記事を求めているのかは、この入力からだけでは知りようがない。また仮に入力そのものの訳を知りたいという意図であっても、前後の文脈に応じて訳し分けがされる場合にどの訳を優先して提示すべきかは、入力表現の現れる文脈情報なしにはわからない。このような直接的な要求を超える情報付加を

ユーザに適切に行わせるのは難しい。

我々はこのような問題を解決するために、以下の手法を提案した [3]。

1. 翻訳元原稿から入力したい表現を「マウスで指定」させるユーザインタフェースを用いて、前後の文脈情報を獲得する手法
2. ユーザが指定した表現中のキーワードと文脈中のキーワードに重みを適切に与えて、
 - i) 指定表現中のキーワードを多く含み、
 - ii) かつ、文脈中のキーワードを多く含むものを優先するように検索結果に順位づけを行うAND検索の拡張手法

本稿では、この手法による検索結果の順位づけを従来のものと比較して、その有用性を検証する。

2 実験

実験は、被験者(3名)に翻訳中に翻訳元の日本語記事中のある表現をシステムに入力した状況を想定させ、検索結果が翻訳作業に有用であるかを主観評価させることで行った。翻訳元記事には、実際に運用中のシステムのログから選んだ短い入力表現を含む14記事を用意し、各々この表現を想定する入力表現とした。用意した翻訳元記事と検索結果の例を図1に示す。本来は検索結果の日本語記事に対して各々対訳記事を提示しなければ翻訳タスクへの有用性を評価することはできないが、その場合有用性は対訳の質に依存し、検索性能を正しく評価することが難しい。そこで、被験者には検索結果の日本語記事のみを与え、各々十分適切な対訳記事が得られることを想定させた。

検索結果の評価は、従来の手法(入力表現のみを用いて検索)と新しく提案している手法の検索結果上位5件を、入力表現に相当する部分を強調表示した形で記事全文を提示し、以下の2つの観点で有用度を5段階評価(5:とても役に立つ ~ 1:まったく役に立たない)させることで行った。

- A: 入力表現そのものの翻訳元記事における適切な訳を得るのに有用な記事か?
- B: 入力表現に相当する部分以外の記事全体が翻訳元記事を翻訳するタスクにおいて有用か?

評価値を被験者ごとに以下の2つの方法で集計した結果を表1に示す。

“The effectiveness of considering the context around the query when searching similar expressions”

KUMANO Tadaki(tkumano@s1t.atr.co.jp),

GOTO Isao, URATANI Noriyoshi

ATR Spoken Language Translation Research Laboratories

2-2, HIKARIDAI, SEIKA-CHO, SORAKU-GUN, KYOTO

619-0288, JAPAN

[翻訳元記事]

ドイツで十四日、全人口の二十パーセントを占める最大の州で州議会選挙が行われ、与党の社会民主党が第一党の座を維持し...

[従来の手法による検索結果 (上位 5 件)]

1. 証券取引等監視委員会は、ドイツ銀行グループのドイチ証券東京支店が日本の金融機関に対し...
2. ... 現場にはオランダのほか国境を接するドイツからも消防隊が応援にかけつけ消火活動にあたり、...
3. ... それによりますと武装グループは前回の交渉で、健康状態が悪化しているドイツ人女性の解放について二十四時間以内に回答するとしていたにもかかわらず、...
4. 先月 (四月) 一か月間に国内で販売された輸入車は、人気の高いドイツ車の売れ行きが落ち込んだことから、去年の同じ時期と比べて四点五%減少し...
5. ... フィリピン政府が武装グループとの間で人質解放の交渉を続けていますが、発生からおよそ二週間たった今も、ドイツ人三人、フランス人二人、フィンランド人二人のEU加盟国からの観光客七人を含めて、あわせて二十人余りが、...

[文脈を考慮した手法による検索結果 (上位 5 件)]

1. ことし九月に総選挙を控えたドイツで、その前哨戦ともいえる東部の州議会選挙が二十六日行われ、連邦議会の野党社会民主党が第一党の座についたのに対して、...
2. ... 今回、外国人に選挙参加が認められたことに、ドイツ人の間に反対の声もある他、...
3. ... この後、旧西ドイツのシュミット元首相が基調講演を行い、先進国には、公害物質の排出の削減計画を、また発展途上国には人口抑制のための家族計画の実行を義務づけることや...
4. 十四日投票が行われたのは、ドイツ西部のデュッセルドルフやケルンなどの大都市とルールなどの工業地帯を抱えるドイツ最大の州...
5. コール首相の長期政権が続くドイツで、政権復帰を狙う最大野党・社会民主党は、一日、今年九月の次の総選挙での首相候補に...

図 1: 評価に用いた翻訳元記事と検索結果の例 (入力表現に相当する部分の周辺を抜粋)

被験者	観点	評価値平均		評価値順との変位平均	
		従来手法	新手法	従来手法	新手法
1	A	2.957	3.514	0.671	0.443
	B	1.343	1.857	0.529	0.529
2	A	3.929	4.171	0.514	0.186
	B	1.329	1.914	0.243	0.271
3	A	4.414	4.543	0.214	0.186
	B	1.243	1.714	0.543	0.457

表 1: 上位 5 件に対する有用度評価の結果

評価値平均: 上位 5 件の評価値の平均
(全体的な有用性の評価尺度)

評価値順との変位平均: 上位 5 件を評価値順に並べ替えたものを理想の順位づけとしたとき、各検索結果について求めた理想の順位と実際の順位との変位 (絶対値) の平均
(相対的な順位づけの評価尺度)

評価結果から、以下の事柄が考察できる。

全体的な有用性について: 評価値平均は、どの被験者も観点 A, B のどちらも値が増加していることから、入力表現そのものの訳を調べる、検索結果全体を翻訳タスクに利用する、いずれの場合にも、上位 5 件の有用性は向上していると考えられる。

相対的な順位づけ性能について: 評価値順との変位平均は、観点 A については減少しており、上位 5 件の中での相対的な順位づけ性能は向上していると考えられる。しかし、観点 B については有意な差は見られない。これは、入力表現に近い文脈だけ

では、翻訳元記事の内容を詳細には反映できないためではないかと思われる。

3 まとめ

翻訳支援のための類似表現検索において、ユーザーの入力表現の前後の文脈を考慮して検索結果の順位づけを行う手法は、入力表現そのものの適切な訳を得る、入力表現のまわりの全体的な翻訳の参考になる記事を得る、いずれの観点においてもある程度有効であることが確かめられた。

今後は、より大規模な実験を行ってその有効性が有意であることを確認するとともに、文脈中のキーワードへの重みづけ手法の妥当性を検討していく予定である。

参考文献

- [1] 熊野, 田中, 松田, 浦谷, 江原. 日英放送原稿翻訳者のための類似用例提示型翻訳支援システム. 第 55 回情報処理学会全国大会, pp. 2:582-583, 1997.
- [2] Kumano, Tanaka, Uratani, and Ehara. Translation examples browser: Japanese to English translation aid for news articles. In *Natural Language Processing and Industrial Applications*, pp. I:96-102, 1998.
- [3] 熊野, 後藤, 江原. *Translators' Workbench*: 対訳エディタを中心とした統合翻訳支援環境. 第 6 回言語処理学会年次大会, pp. 143-146, 2000.
- [4] 田中, 熊野, 浦谷, 江原. 放送ニュース文を対象とした効果的類似用例検索法. *自然言語処理*, Vol. 6, No. 5, pp. 93-116, 1999.