

3G-8

類似文検索機能を備えた翻訳支援システム

伊藤悦雄 長谷部浩一 武田公人 天野真家
(株)東芝 総合研究所

1. はじめに

翻訳する際、以前の翻訳結果を参照するとユーザ特有の表現や文体の統一を容易に行うことができる。さらに、参照した文をそのまま利用することができれば翻訳作業が効率的に行える。しかし現在、以前の機械翻訳の結果を参照する実用的なシステムはない。

この機能を実現するためには、ユーザは、あらかじめデータベースを作成する必要がある。しかし、こういったデータベースの検索は、単一の単語をキーワードとして検索する方法しかなく、適切な例文を簡単に得ることができなかつた。

本論文では、翻訳支援システムとして、翻訳対象文に形態的に類似する文を以前の翻訳結果から簡単に高速検索する方法を提案する。

2. 類似文の検索対象

現在、翻訳支援として利用可能なデータベースは、電子化辞書、用例集等がある。しかし、これらは、単一の要素をキーとした検索に重点が置かれている。また汎用データベースであるため、修正せずに利用可能な文は少ない上、翻訳システムへ直接取込むことができない。

そこで、我々は、検索された類似文をそのまま利用するため、既に翻訳された文を対象とする。

この方法には以下の利点がある。

- (1) ユーザは意識してデータベースを作る必要がない
- (2) ユーザ特有の表現・訳語等が利用できる

また、翻訳で得られた情報を文と共に記憶し、検索条件に利用することで柔軟な検索が可能となる。

3. 文の類似性

類似文には様々な定義がある[1][2][3]。しかし、本方式では類似文として検索された文を極力そのまま

翻訳に利用することを前提とし、従来とは異なる定義を以下の通り提案する。

[類似文]

以下の条件を満たすものを類似文とする。ただし、(a)から順に類似性が高い文とする。

(a)訳文の言語において出現しない、あるいは必須でない情報のみ相違する文

例: 冠詞、名詞の単数・複数のみ異なる文

異形語への置換のある文

(b)翻訳に於ける意味変換に与える影響の少ない部分が相違する文

例: 時制が異なる文

複合語が代表する一語に置換された文

名詞が代名詞に置換された文

(c)原文の意味の差の少ない文

例: 同一品詞の他の語に置換されている語がある文

余分な修飾語が付加されている文

この定義に従って、検索を行うために以下に示す類似度判定条件を設定する。この条件を品詞毎に設定し、これを組合せることにより、上記の類似文の検索が可能である。

[類似度判定条件]

(A)完全一致・・・検索文に出現するすべての単語が出現し、それ以外の単語が出現しない。

(B)包含一致・・・検索文に出現するすべての単語が出現する。

(C)部分一致・・・検索文に出現する単語のうち一部の単語が出現する。出現の割り合

いを%で指定する。

(D)品詞一致・・・検索文に出現する単語の一部が同一品詞の他の語に置換されている。

(E)一致不要・・・この品詞は一致の必要はない。

また、単語の検索においては、動詞などの活用や名詞の単複変化などに対応して柔軟な検索が可能であるように「同一形」「原形」「その他の活用形」の順で検索を行う。また、必要に応じて語順一致条件を付加できる。

4. 実験および評価

上記の類似文検索を行うため、図2の実験システムを作成した。

約1,600文の翻訳結果を検索対象とした。このデータに対する類似文の検索所用時間は1~2秒程度である。また、有効な検索を行うために類似度判定条件を以下の設定にする必要があることが判明した。

(1) 動詞は出現割合70%の部分一致より厳しい条件とする。

類似度判定条件の組み合わせ例						
名詞	動詞	副詞	形容詞	冠詞	前置詞	その他
1: A	A	A	A	A	A	A
(検索文と同一文のみ)						
2: A	A	B	B	E	A	A
(余分な修飾語がついてもよい)						
3: D	B	E	E	E	B	B
(修飾語が無くてもよい。名詞は他の名詞への置換可)						
4: C(50)	B	E	E	E	E	E
(名詞の半数と、動詞が含まれていればよい)						
5: E	B	E	E	E	E	E
(この動詞が使用されていればよい)						

上記の条件で「The man has a big apple.」の検索時に発見される文の例

条件1: The man has a big apple.
 2: The young man has a very big apple.
 3: The woman has fish.
 4: He has a red apple.
 5: His mother has a lot of money.

図1 類似度判定条件の例

(2) 前置詞は出現割合50%の部分一致より厳しい条件とする。

(3) 英日翻訳に利用する際は語順を考慮し、日英翻訳に利用する際は語順は考慮しない。

5. おわりに

本論文では、以前の翻訳結果を利用可能な機械翻訳システムのための類似文検索方式について述べた。今後は、同義語の置換を取扱えるような拡張、構文的情報を利用する検索方式の検討を行う予定である。

参考文献

[1] 隅田、堤「構文の照合による柔軟なテキスト検索機能を備えた翻訳支援システム」情報処理学会第37回全国大会 (1988)

[2] 佐藤、長尾「実例に基づいた翻訳」情報処理学会第38回全国大会 (1989)

[3] 中村「用例検索翻訳支援システム」情報処理学会第38回全国大会 (1989)

[4] 伊藤、長谷部、武田、天野：「過去の翻訳結果を利用した翻訳支援システム」、情報処理学会第38回全国大会 (1989)

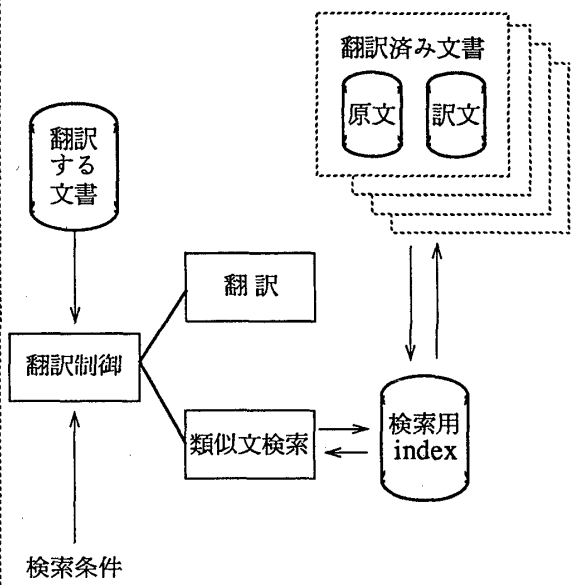


図2 類似文検索システムの概要