

日英機械翻訳における原文前編集の内容と効果について

4K-9

池原 悟* 白井 諭* 小見佳恵**

* NTTコミュニケーション科学研究所

** NTTアドバンステクノロジー株式会社

1. まえがき

機械翻訳システムを現実の文書翻訳に適用する際、翻訳の品質を向上させる方法として、①原文をそのシステムで翻訳しやすいように前編集する[文書をシステムに合わせる]方法⁽¹⁾、②翻訳システムを翻訳対象とする文書の領域(分野、文体)にチューンアップする[システムを文書に合わせる]方法^(2, 3)、の2つの方法が考えられる。

本報告では、日英機械翻訳実験システムALT-J/E⁽⁴⁾を用いて、両者の訳文品質向上効果を比較評価する。

2. 実験比較の方法

技術者が書き下ろした(テクニカルライティングされていない)ハードウェアのシステム設計書の一部(1,800文)を対象に、図1に示すような3種類の翻訳実験を行う。

システム設計書 (翻訳対象原文) 日本語(1,800文)

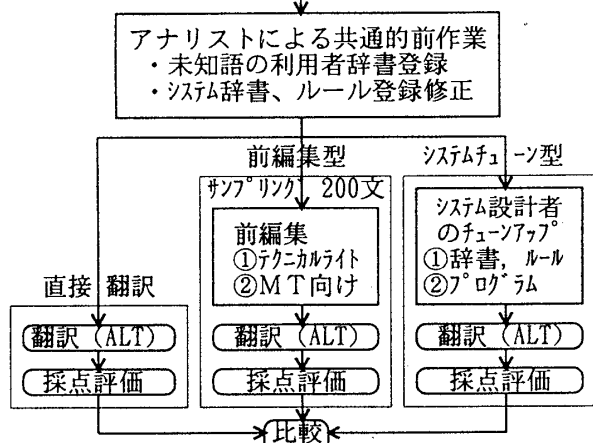


図1. 実験評価の方法

表1. 4段階評価基準の概要

区分	グレード	説明
合格	◎ 秀訳	グレード4 ほぼ修正不要の英訳
	○ 可訳	グレード3 意味が正しく通じる
不合格	△ 借訳	グレード2 借しい
	× 駄訳	グレード1 翻訳失敗

ALTシステムは辞書、ルールのデバッグが未完の実験システムであるため、実験に先立ち図にあるような共通的前

作業として、アナリストによる辞書ルールの追加修正作業を実施する。これは、試験文の範囲でシステムのデバッグ完了状況を実現するためのもので、システム運用グループ内で可能な作業レベルを想定したものである。その後、3つの方法による翻訳実験を行い、それぞれの訳文品質を、表1に示す4段階評価基準によって比較評価する。

なお、前編集型の実験では、省力化のため、その中からランダムに抽出した200文を対象とする。

3. 前編集の方法と内容

3.1 前編集の方法

以下の2ステップに分けて前編集を行う。まず、翻訳結果が不合格(△, ×)となった文に対して、前編集を行い、翻訳を行う。その翻訳結果が合格となった文は、前編集を終了する。次に、翻訳結果が不合格のままである文を対象に、前回の前編集を見直して、再度、前編集を行い、その結果を翻訳する。

3.2 前編集の項目

前項で実施された前編集の項目を、テクニカルライティングの項目とシステム依存の項目の2つに分けて分類し、標本文200文中の適用回数を調べた。その結果を表2に示す。

この表より以下のことが分かる。テクニカルライティングの項目はALT向けの前編集項目よりも多く(全体の約6割)、機械翻訳にとっては、まず日本語として分かりやすい原文であることが大切である。次に、T1, T3, S1の適用回数が特に多く、この3項目だけで全体の半分以上を占める。これらの3項目はいずれも文の分割に関する項目であることから、文を短くすることは、テクニカルライティング、システム依存のいずれの観点からも効果的である。

4. システムチューンアップの方法と内容

4.1 チューンアップの方法

システム設計者が、訳文が不合格であった文に対して途中の翻訳過程を調べ、辞書、ルールの追加修正、プログラムデバッグ等を行う。運用責任者のレベルでは実行不能な作業内容が対象となる。但し、チェックを含む修正等の作業時間が、1項目あたり20~30分以内の項目を対象とし、それ以上となるもの(システム改造のレベル)は対象外とする。

4.2 チューンアップの項目

チューンアップは、実験標本に選定した1,800文を対象としたが、実験標本文にチューンしたことによる他の分野で

Experimental Studies on Pre-editing of Source Texts in Japanese to English MT

Satoru Ikehara*, Satoshi Shirai* and Yoshie Omi**

*NTT Communication Science Laboratories,

**NTT Advanced Technology Corporation

表2. 編集項目と適用回数・品質向上(標本文数:200文)

分類	項番	編集内容の説明	適用回数	品質向上
テクニカルライターから見た編集項目	T1	長文は分割する。 例: 一主題(一動作)一文, 接続詞等の挿入	54	19.7
	T2	簡潔な表現とする。 例: 同種の意味を持つ語を併用しない	19	6.7
	T3	箇条書きを使う。 例: 物体の構成要件, 条件, 順序など	54	75.0
	T4	長い埋め込み文は書き換える。	7	2.7
	T5	中止法には語句を補完する。 例: ~し~して, ~し~した時	4	0.7
	T6	受動態は能動態に書き換える。	2	1.0
	T7	「の」と「で」の意味を明確にする。 例: 格助詞「で」、所有以外「の」の変更	11	5.0
	T8	読点の追加と削除。 例: 修飾先の明確かがポイント。	4	0.5
	T9	適切な動詞への書き換え, 動詞の追加 例: 視点統一。具体的な動詞を使用	14	7.0
	小計	-----	169	118
機械翻訳システムから見た前編集項目	S1	できる限り短い文にする。	26	10.2
	S2	主語目的語を補完する。	4	1.5
	S3	目的語を明確にする。 例: 語順はなるべくSOVの順。	18	6.0
	S4	数詞はなるべく名詞の前におく。 例: 数詞の副詞的用法は避ける。	6	1.0
	S5	修飾先を省略しない。	2	0.0
	S6	「~に(が)できる」→「~できる」 「~が可能」→「~可能」	2	1.0
	S7	並列表現には、助詞を補完する。 例: 「録音再生」→「録音と再生」	4	1.7
	S8	助詞、助動詞を補完する。 例: 「最大4chを」→「最大で4chを」	10	2.5
	S8	記号を使用しない。 例: 中点“.”は“/”に替える。	8	6.0
	S1	その他: 複合述語、接辞の連鎖を避ける	4	1.2
	小計	-----	84	31.1
合計		-----	235	149

[備考]・編集は初回の機械翻訳結果が不合格となった文のみを対象。
・前編集は1800文中の200文が対象。
・品質向上の項は各前編集項目に該当する文の訳文品質の向上した段階数(グレード)の累積値を示す。

表4. 前編集による訳文品質向上効果

前	後	○(秀訳)	○(可訳)	△(惜訳)	×(駄訳)	合計
△(惜訳)	1.6	14.7	19.7	0.0	36文	
×(駄訳)	0.5	15.3	10.7	0.5	27文	
合計	2.1	30.0	30.1	0.5	63文	

[備考] 数値は文数を示す。但し、前編集によって1文が2文以上に分割された場合は、分割された文毎に評価し、分割数で割った値を文数として集計した。

表5. 訳文品質向上効果の比較

評価結果	初回の翻訳		前編集後		チューンアップ後	
秀訳(◎)	25%	合格率	27%	合格率	45%	合格率
可訳(○)	43%	68%	58%	85%	39%	84%
惜訳(△)	18%	不合格	15%	不合格	10%	不合格
駄訳(×)	14%	32%	0.2%	15%	6%	16%
合計	100%	100%	100%	100%	100%	100%

の「グレード」が生じないことを確認するため、機能試験文(3,700文)や新聞記事試験文等を併用した。

実施したチューンアップ項目は、構文解析ルールの修正、意味解析規則の修正、プログラムとの相互関係で調整し直された辞書情報等合わせて、約100項目であった。

5. 訳文品質向上効果

前編集各項目による訳文品質向上効果を表2、前編集対象文の訳文品質の変化を表4に示す。また、前編集、システム設計者によるチューンアップの効果を表5で比較する。これらの結果から、以下のことが分かる。

- ①前編集でも、システムチューンアップ(84%)とほぼ同等の訳文合格率(85%)を得ることができた。
- ②しかし、不合格文を秀訳(◎)まで向上させるにはシステムチューンアップの方が効果が大きい。
- ③前編集、チューンアップいずれの場合も、惜訳(△)に比べて駄訳(×)の改良は効果が少ない。

このうち②から、前編集は、原文解析失敗を回避する点で効果的できるが、英文生成段階の問題(表現の適切性の向上)には弱いことが分かる。

なお、前編集とシステムチューンアップの項目数を見ると、前編集項目は文当たり平均約4件(200文中、不合格だった63文に対して延べ235項目)であったのに対して、チューンアップ項目は、文当たり平均0.2件(1,800文全体の不合格文約600文に対して約100項目)であった。

6. あとがき

日英機械翻訳システムを用いた技術仕様書の翻訳において、訳文品質を向上させるための方法として、①文書をシステムに合わせる方法と、②システムを文書に合わせる方法の効果を比較した。その結果、①(前編集)の効果は予想以上に大きく、訳文合格率では、②と同等の効果を得られることなど分かった。秀訳率を上げる場合

や、対象とする文書が大量少種の場合は、やはりシステムのチューンアップが望まれるが、文書が少量多種の場合は、原文前編集が適していると考えられる。

今後、前編集では自動化、システムチューンアップの場合はシステム設計者から離脱できる簡易な仕組みが望まれる。

謝辞

終わりに、原文前編集とデータの集計を担当していただいたNATC社の渡辺瞳さん、およびシステムチューンアップを担当された方々に感謝する。

参考文献

- (1)「機械翻訳セミナーテキスト」, AAMT, 1994. 6. 22
- (2)長尾、田中、辻井: 制限言語に基づく文章作成援助システム, 自然言語処理研究会44-5(1984)
- (3)辻井、長尾: 日英翻訳過程での処理とその翻訳結果への反映: 自然言語処理研究会47-10(1985)
- (4)池原、宮崎、白井、林: 言語における話者の認識と多段階翻訳方式, 情報論Vol. 28, No. 2, 1269-1279(1987)