

機械翻訳システムの評価と改善

3N-3

田中 康仁

兵庫 大学

E-mail: yasuhito@humans-kc.hyogo-dai.ac.jp

(0) はじめに

幾つかの機械翻訳システムのテスト結果を述べ、その改良点を示す。

(1) 機械翻訳システムの評価

3社の機械翻訳システムの訳文による理解度評価を示す。

(1) 訳文の理解度による評価

機械翻訳システムでは「理解容易度」による評価と「忠実度」による評価がある。ここでは理解容易度により評価方法を行ってみた。

評価に用いた英文はEDRの英文コーパスを採用し、単語数が2~7のものを用いた。採点は英語の教員が行った。

評価結果 A社の英日翻訳システム

	5	4	3	2	1	合計	平均値
2 単語文	15	1	1	1	1	19	4.47
3 単語文	336	97	16	11	0	460	4.65
4 単語文	1,369	336	157	25	2	1,889	4.61
5 単語文	3,655	808	510	53	4	5,030	4.60
6 単語文	4,742	1,379	595	81	1	6,798	4.58
7 単語文	4,471	2,331	870	118	1	7,791	4.43
合計	14,588	4,952	2,149	289	9	21,987	4.53
%	66.35	22.52	9.77	1.31	0.04	100%	

評点の高いものが良い結果を示す。

平均値をみると4.5程度で大変良いように思える。しかし、評点5のものだけを○とし、他のものは×とすると100点満点で平均66点となる。これは我々の感覚とほぼ同じであることがわかる。

次にB社、C社の製品についても同様の方法で評価してみた。

評価結果 B社の英日翻訳システム

	5	4	3	2	1	合計	平均値
2 単語文	6	9	4	0	0	19	4.10
3 単語文	259	127	54	14	6	460	4.34
4 単語文	1,099	551	167	67	5	1,889	4.41
5 単語文	2,555	1,680	645	150	0	5,030	4.32
6 単語文	3,585	2,087	953	174	0	6,798	4.33
7 単語文	2,222	3,541	1,638	368	22	7,791	3.97
合計	9,725	7,995	3,461	773	33	21,987	4.21
%	44.23	36.36	15.74	3.52	0.15	100%	

評点の高いものが良い結果を示す。

評価結果 C社の英日翻訳システム

	5	4	3	2	1	合計	平均値
2 単語文	8	5	4	2	0	19	4.00
3 単語文	232	151	58	19	0	460	4.30
4 単語文	966	634	228	60	1	1,889	4.33
5 単語文	1,645	2,104	1,047	228	6	5,030	4.02
6 単語文	1,907	2,965	1,598	328	0	6,798	3.95
7 単語文	2,149	3,575	1,667	400	0	7,791	3.96
合計	6,907	9,434	4,602	1,037	7	21,987	4.01
%	31.41	42.91	20.93	4.72	0.03	100%	

評点の高いものが良い結果を示す。

このA社はEDRプロジェクトで中心的役割をはたした会社である。単語数が少ない文であるためか、非常に良い翻訳結果が得られた。しかし、単語数が多くなると少しずつ平均値が下がっている。

B社の製品はA社に比べ劣っていることがわかる。C社の製品はまだ改良しなければならないことがわかる。

この方法では、翻訳結果の良い文と悪い文を分けることがなされるので、悪い結果を集め分析することにより、改良方法がわかる。個々の文固有の問題点、知識データの不足、文法体系、文解析等の問題であるかを判別することができる。

言語学者や辞書学者、語用論学者へ機械翻訳で処理できないデータや問題点を提供し、研究が発展することを期待する。

また単語数別の分析は、問題点の単純化をはかり、原因をつかみやすくする。翻訳ソフトウェアのような大きなシステムは、一度に全てを調べあげることにはできない。個々の単純な部分を分析し、誤りを修正し、より複雑な誤りをみつけだすようにすべきである。

A社以外の数社について試したところ、A社より平均点で約0.5点程度低いことがわかった。この0.5点の差は大きな差である。今後の改良を期待する。

A社、B社、C社とも、この論文を作成した時点から版(バージョン)が更新され少しずつ良くなっている。このような実験を行うと、このデータについては改良がなされるため、さらに新しいテストデータを準備しなければならない。新しい多量の言語データを標準化しなければならない。どのような条件があれば標準化されたデータかは大きな課題である。最近では大量のWWWが作られているのでこの英文を集め、単語数別に整理し、同一の文はまとめ、頻度をつけてテストデータとして利用することもできる。

(2) この評価方法について

単語数別ファイルを作り機械翻訳システムを評価する方法は、問題点の抽出が容易で興味ある方法である。

しかし、翻訳結果の評点付けは人間の作業であり、大変労力のかかる作業であった。この作業を自動的に行うように機械翻訳システムに組込んでおくことが重要である。最終的な判断は人間の作業であるが、ある程度のところまでは機械的に可能である。これにより翻訳結果の

Evaluation and Improvement of
Machine Translation System

Yasuhito Tanaka

Hyogo University

大まかなグループ分けが可能である。このようにして作業の迅速化が図られる。

〔3〕機械翻訳システムの改良についての提案

前述までのように機械翻訳システムをコーパスを用いて、単語数別に分類することにより短い文から順次評価する方法を提案した。この経験より、次のような改良方法を述べる。

1) 大量のコーパスを用いて、機械翻訳により結果の良くないデータを見つけ出す。

結果の良くないデータを20,000文程度集めるとすれば、30%の翻訳ミスが有る機械翻訳システムでは約6万7千文のコーパスを準備すればよい。

どのような分野の内容を重点にしたコーパスを準備すればよいか考えなければならない。

2) 利用者からのクレームとデータの収集

利用者から翻訳できない文を集めテスト・データとする。又は、機械翻訳システムの特別ユーザ(モニター)からのデータ収集を行う。

3) 単語についてのデータ収集

コーパスや辞書の見出し語や、WWWから英単語や日本語の単語を抽出し、機械翻訳辞書の見出し語と照合し不足のデータを補う。

4) 複合語の補強

我々、日常生活の中で使われている複合語を集め、機械翻訳用辞書と照合し、不足しているものを補う。複合語は日本語と英語のように対訳になったものでなければならない。また対訳語が複数ある場合は、使用条件も付けなければならない。集めた複合語がどの程度機械翻訳システムの辞書と一致したか、一致しなかったものはどの程度かを測定しなければならない。また追加される複合語がどの程度実際に使われているかも調べねばならない。

筆者は日本語の複合語として四文字漢字列、五文字漢字列を数十万語整理して持っている。これらの中で頻度が高いものに訳語を付けるのも一つの解決方法である。

5) 専門用語について

専門用語については専門用語辞書を作成している出版社、ソフトウェア会社、学会や協会と協力し、購入することが得策と考えられる。

機械翻訳システムに組み込まれているもの以外を購入すべきである。このためには購入時に内容を評価してくれる第三者が必要である。

6) 慣用表現の補強

慣用表現の補強も重要なテーマである。我々は約3万種の慣用表現を集めた。日本語と英語が対になったものである。

7) 例文の収集

このごろの機械翻訳システムは用例ベースの機械翻訳機能が付け加えられている。このための例文を集めることが重要である。

我々は学生のタッチタイプの練習として日本語、英語の対を入力させている。これを定期的集め、重複を整理し、誤りを修正すれば良い例文データとなることが分った。一年間に約6万文程度の日本語と英語の

対を集めた。これを単語数別に整理し、1~12単語程度の文が用例ベースの機械翻訳機能に有効であることが分った。

8) 結合価文法の文型データの収集

結合価文法の文型パターンは慣用表現の収集中に、動詞句の中からかなり抽出できることが分った。また、1)で述べた、大量の翻訳誤り文の中から特定の動詞に注目し、検索し、KWICを作成することにより資料が得られる。

9) その他

その他、色々な文法上の改良については1)の大量のコーパスを翻訳した結果の中から、分析するのが良い方法である。

1)~9)までの改良方法についての指針を示すことができた。これらを実行し、新しい版を作り、何%の翻訳結果の改善が得られるか、評価をしなければならない。

〔4〕おわりに

機械翻訳システム(英⇒日)のテストを行い評価を行った。大量の英文データを単語数別に分析し、整理し、単語数の少ないものから順次テストするという方法を考えた。これにより機械翻訳システム(英⇒日)の品質向上をはかる一つの方法ができた。

さらに色々な方法を考えてみたい。これら試行から、大量の日英対訳付コーパスを作成しなければならないこともわかった。また、市販されているCD-ROMを機械翻訳のテストに利用するのも一つの方法である。

実際に稼働している機械翻訳システム上で2万件強のテストデータで翻訳を行ってみた。数社の協力が得られた。さらに、評価ばかりでなく改良のための指針が具体的に得られた。

〔5〕参考文献

- (1) 社 日本電子工業振興協会
「自然言語処理システムの動向に関する調査報告書」
平成9年4月
- (2) 牧野武則 評価技術
「機械翻訳」Bit別冊 共立出版 1988年9月
- (3) 長尾 真 「機械翻訳はどこまで可能か」
岩波出版 1986年6月
- (4) Language and Machines: Computers in Translation and Linguistics, National Academy of Sciences National Research Council (1966)
- (5) 高橋文子 3語で通じる英会話
明日香出版社 1998年5月

〔6〕データについて

英文データは日本電子化辞書(株)の英文コーパスを利用した。

この論文の翻訳評価基準について

理解容易性の詳細な内容については参考文献(3)P55を参照されたい。

この原文は Language and Machines: Computer in Translation and Linguistics, National Academy of Sciences-National Research Council (1966) . (ALPAC レポート) の中にある。