3W－1

機械翻訳システムの評価
－品質向上にむけて－

田中康仁
兵庫大学

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

〔0〕はじめに

機械翻訳システムは、多くのコンピュータ企業やソフトウェア会社で作成されている。しかし、これらの品質はさまざまである。また、これについての評価方法、評価方針は決まっていない。これらについて色々と検討する必要がある。このことについて考えてみる。

〔1〕評価の方法について

評価については色々な方法がある。幾つかのものを次に示す。

(1)機械翻訳システムを対象に、A，B，Cのランクを付けるための評価

(2)機械翻訳システムの利用と使わない方法での費用の分析による評価

(3)品質向上のために各機械翻訳システムにはどのような問題点があるか、大量のデータによるテストにより問題点を明確にする。どのような改善をすれば品質が向上するかを評価する。

(4)その他

等の方法がある。

ここでは(3)の方法を基にして考えてみる。しかし、この方法は問題点の指摘であるため、開発者にとってはあまり良い気持ちを持てない。

しかし、問題点を明確にすることが、品質向上に結びつくことを理解していただきたい。

〔2〕機械翻訳システム用評価データ

機械翻訳システムを評価するにあたって重要な点はどのような評価データを集めるかということである。ここでは日本電子化辞書株式会社（ＥＤＲ)の英文コーパスを用いることにした。この英文コーパスはタグ付けされているが、このタグを取り除いた文を用いた。これを単語数別に分類する。単語数別文の数は次の通りである。

単語数	1	2	3	4	5	6	7	8	9
文の数	0	19	460	1,889	5,030	6,798	7,791	8,416	8,698

10	11	12	13	14	15	16	17	18	19
9,114	9,018	9,176	9,050	8,815	8,446	8,245	7,466	6,559	5,283

20	21	22	23	24	25	26	27	28	29
3,645	562	464	321	235	142	77	46	23	12

30	31	32	33	合計
2	0	0	1	125,803

このコーパスの中で？（疑問符）と！（感嘆符）の付いている文の数を単語数別に調べると次のようになる。

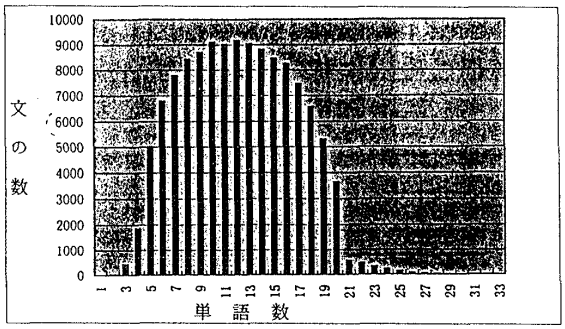
単語数	2	3	4	5	6	7	8	9	10	11
?の文数	0	50	314	872	1,105	1,092	965	811	566	475
!の文数	1	20	29	45	34	34	17	23	19	16

Evaluation Method for Machine Translation System.
－quality up－
Yasuhito Tanaka
Hyogo University

12	13	14	15	16	17	18	19	20	21	22
322	251	206	156	101	72	71	42	38	5	3
10	21	7	5	6	2	1	4	1	0	0

23	24	25	26	27	28	29	30	33	合計
0	0	2	1	0	0	0	0	0	7,652
1	0	0	0	0	0	0	0	0	296

疑問文は6単語のものが最も多く、全体的に他の文と比べ短い。感嘆文も5単語のものが最多である。



ＥＤＲの英文コーパスを分析した結果、次のような特徴がある。

文の総数	125,670文	
最長の文	33単語の文	1文
最短の文	2単語の文	19文
平均単語数	12.22単語	
単語数の中央値	12単語	
疑問文の数（？の数）	7,652文	6.09%
感嘆文の数（！の数）	296文	0.23%

また単語数が２０以上になると急に少なくなっている。さらにEDRコーパスには次の特徴がある。

(1) 新聞の文に比べ人称代名詞が主語になる文が多い。

(2) 疑問文、感嘆文の数が多い。

(3) 単語数が２０単語以内のもので９９％を占めている。

(4) 単語数が２０以上がきわめて少なくなっているのは、ＥＤＲコーパス作成時の何らか操作を行ったと思われる。

〔3〕評価の具体的方法について

(1) ソフトウェアとしての安定性

機械翻訳システムはコンピュータ上で動くソフトウェアである。このためどのようなデータに対しても、ソフトウェアが何の応答もなく終了するということは避けなければならない。

Ａ社のシステムでは、ある特定のデータを処理させると、ソフトウェアが何の応答もなく終了していた。しかし、Ａ社は直ちに新しい版（ヴァージョン）を出してしまい、今では問題なく処理す

るシステムになっている。

(2) 問題点だけの収集

B社のシステムは構文解析、その他の処理で処理不能となると、一定の記号列を表示し、訳語の単語列を表示する処理を行い、次の文の処理に移ってしまった。このため大量のデータを処理してみると、多量の誤りが出現した。そこでＥＤＲの英文コーパスの７単語以内の約２万文の処理中に1,226文見つかった。これらを開発者に送ったところ、次のような返信を得た。

1）ＥＤＲの英文コーパスの中に英文字の誤り、編集上の誤りがあった。これは筆者の問題点であった。

2）省略形　例えば、you're のようなものについての対応が不十分であることが判明した。また、文の一部が省略されているものもある。これについては対応されていない。

Yes, I certainly will (do) . do の省略

3）一部の文法上の欠点があることが、開発者によって認められた。倒置文に対して弱いということである。

例　Behind every caltural image exists those taboos.

以上のことは開発者が誤った処理によるデータを検討すれば、ただちに判明するものであった。これは誤った処理を約１,２００文集めたからである。約５％の文が処理できていない。誤って処理したデータを分析する中で、一文毎の個別の問題か、それらの中にある共通の問題であるかが判明した。

(3) 訳文の理解度による評価

機械翻訳システムでは「理解度」による評価と「忠実度」による評価がある。ここでは理解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　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	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。

このC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。

この方法では、翻訳結果の良い文と悪い結果の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。C社以外の数社について試したところ、C社より平均点で約１点程度低いことがわかった。この１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（ヴァージョン）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。

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。

筆者の友人もWWWから242MBのデータを集めてくれた。これは参考になるものである。

〔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ついた。これにより機械翻訳システム（英⇒日）の品質の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。

〔５〕参考文献

(1) 安田賀計　らくらく使えるビジネス文書1,230文例ＣＤ－ＲＯＭ　日本経済新聞社

(2) 田久保浩平、橋本光憲　英文ビジネスレター文例大辞典　ＣＤ－ＲＯＭ　日本経済新聞社　15,000文

(3) 塩澤　正、スコット シェフェルバイン　インターネット英語表現辞典　三修社　1998年1月

(4) 新編英和活用大辞典（自然な英文を書くための３８万例）ＣＤ－ＲＯＭ版　研究社　1996年

(5) 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

(6) 牧野武則　評価技術「機械翻訳」Ｂｉｔ別冊　共立出版　1988年９月

(7) 長尾　真　「機械翻訳はどこまで可能か」岩波出版　1986年６月

〔６〕データについて

英文データは日本電子化辞書㈱の英文コーパスを利用した。

日本電子化辞書㈱のプロジェクトに参加した企業がＥＤＲ英文コーパスをあまり利用していない。翻訳システムにもっと活用してほしいものである。

ＥＤＲコーパスには日本語のコーパスもある。これを日英機械翻訳システムに利用すべきである。