

2007年度ヒューマンインタフェース研究室 卒業研究発表会

研究テーマ「機械翻訳の精度分析」

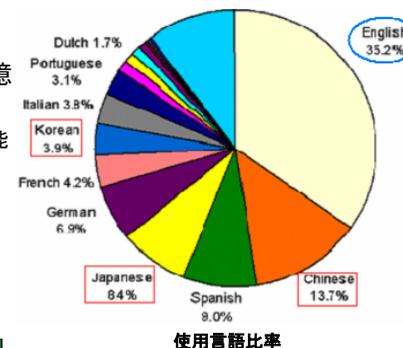
大阪工業大学 情報科学部情報システム学科
学籍番号: B04-081
発表者: 杉田 匠

2008/2/14

1

はじめに

- 全世界のインターネット利用者数は8億人を突破(2004年9月時点)*1
 - 多くの人とコミュニケーションを取る事が可能
- 利用者が増加する一方、利用者の使用言語が多様化



インターネット上の標準言語が英語とは限らなくなっている

*1 Global reach - <http://global-reach.biz/globstats/index.php3>

2

研究目的

- インターネット上には機械翻訳、翻訳サービスが多く存在するが、全世界の言語に対応していない事や、翻訳文に誤訳が多い問題がある*2

- 本研究では主要な言語5種の翻訳サービス品質を比較、評価する

- 日本語
- 韓国語
- 英語
- 中国語
- フランス語
- ドイツ語

*2 言語グリッド(Langrid Input)の構想-異文化コラボレーション基盤の構築を目指して-
アーレム ベン ハッシン、藤原義功、林良彦、菱山玲子、稲葉利江子、石田亨、喜多千草、松原繁夫、村上陽平、
瀬本明代、中口孝雄、中西英之、重信智宏、吉野孝、人工知能学会全国大会(第20回)、2006.

4

研究の流れ

1. 5翻訳言語対の翻訳文のわかりやすさを評価

- ① 日本語→韓国語
- ② 日本語→英語
- ③ 日本語→中国語
- ④ 日本語→英語→フランス語
- ⑤ 日本語→英語→ドイツ語

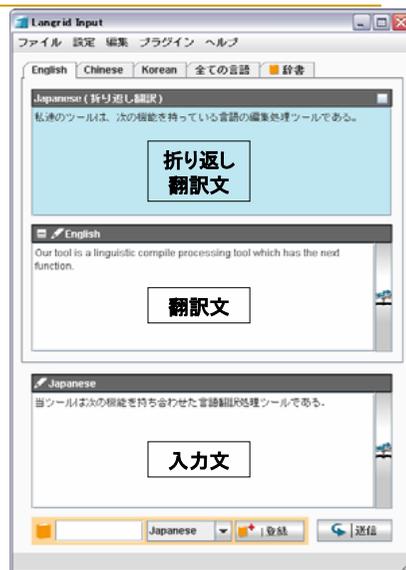
2. 評価点数が特に低い文章の特徴を考察

Langrid Input

■NICT言語グリッドが提供しているツール

■機能

- 標準化されたユーザ・インタフェース
- 多言語翻訳機能
- 折り返し翻訳機能: 翻訳文を入力言語に再翻訳する事で母国語による確認が可能
- 辞書機能: 使用するユーザが必要な言葉を自由に追加できる機能



Langrid Input

5

評価実験(1)

■使用文書

- オープンソースの一般文書『Cathedral and the Bazaar』⁴から抜粋した192個の日本語文

■使用する翻訳言語対の例文

原文(日本語)

リナックスは、私が知っているつもりでいたことの大部分を覆した。

① 日本語→韓国語

【リナックス】は、私が知っているつもりなのであったことの大部分を覆した。

② 日本語→英語

私はLinuxを知り、ここにある最も多くは転覆した。

③ 日本語→中国語

LINUX, 私にご破算になってになる事の大部分を知っているつもりです。

④ 日本語→英語→フランス語

私はLinuxを知り、このほとんどの存在はリバースされている。

⑤ 日本語→英語→ドイツ語

私はLinuxを知り、落ちられたそれは通常ここにあることから適していた。

⁴The Cathedral and the Bazaar - <http://www.tlwg.jp/docs/cathedral-bazaar/cathedral-paper-jp.html>

6

評価実験(2)

■評価基準

- 5: 同じ意味
- 4: 多少問題があるが、大抵原文と同じ意味
- 3: 原文の意味はなんとなく掴める
- 2: 雰囲気は残っているが、原文の意味は伝わらない
- 1: 全く違う意味

■評価点数

- 原文と折り返し翻訳文を比較、評価した各翻訳言語対192問の平均点

7

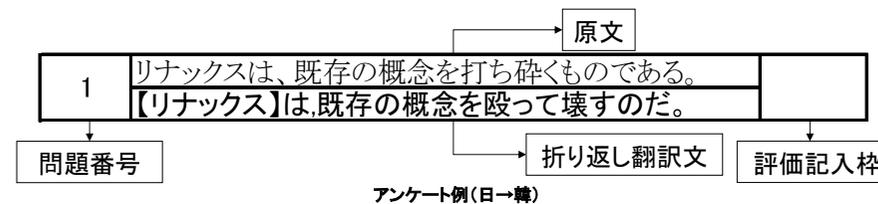
評価実験(3)

■被験者

31名(平均年齢22.1歳, 範囲20-41歳: 男性28名: 女性3名)

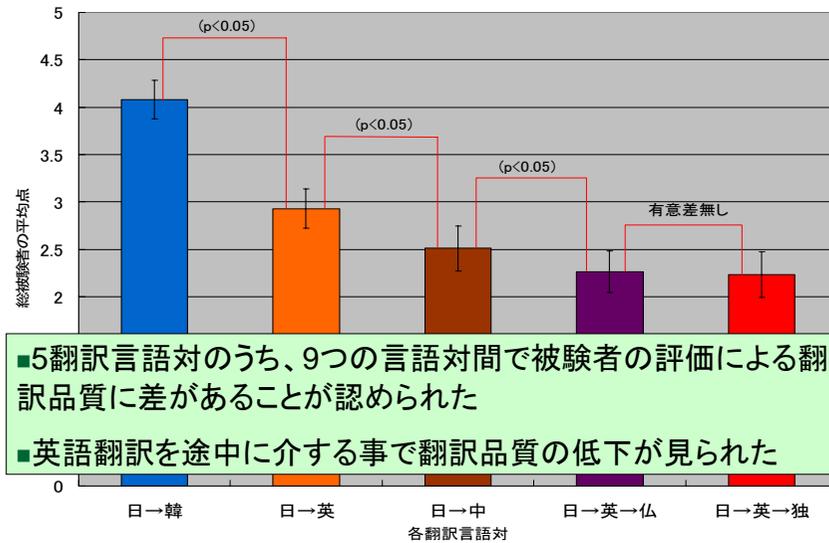
■実験方法

- 5翻訳言語対毎の192問を示した紙面で評価
- 翻訳言語は被験者に明示せず、番号にて区別
- 評価を行う順序を被験者毎に指定



8

評価結果のグラフ



■5翻訳言語対のうち、9つの言語対間で被験者の評価による翻訳品質に差があることが認められた

■英語翻訳を途中に介する事で翻訳品質の低下が見られた

9

考察

■機械翻訳の誤訳

□日本語→中国語 (評価点数36点)

原文: おそらくこれは **オープン・ソース**型ソフトウェアの未来だけではない。リナックスのコミュニティが1つの問題の対応に動員できる人材プールに匹敵する人材を揃えている **クローズド・ソース**型の開発者はいない。

中国語文: 恐怕这个不仅仅是**开放・调味汁**型软件的未来。Linux的地方自治团体统一比得上能动员为1个问题的对应的人材游泳池的人材的**closed・调味汁**型的开发者不在。

折り返し翻訳文: 恐らくこの**開放的**なだけではありません・**汁**型のソフトウェアの未来に**味をととのえます**。Linuxの地方自治の団体は統一的に動員して1つの問題になる対応する人材のプールの人材の**closed**に追いつけて**汁**型の開発者に**味をととのえて**いないことができます。

□日本語→英語→フランス語 (評価点数32点)

基本文: これはいくらかは当たっているかも知れない。

英語文: How much this is **may** be worth.

フランス語文: Combien c'est à peut valoir.

フランス語折り返し翻訳文: How much it is to **can** be worth.

折り返し翻訳文: **缶**が価値であることはいくらにである。

10

考察のまとめ

翻訳言語対	評価総点数平均	評価総点数60点以下の文章数
日本語→韓国語	126.25	1個
日本語→英語	89.56	22個
日本語→中国語	76.24	57個
日本語→英語→フランス語	68.59	87個
日本語→英語→ドイツ語	68.42	95個

■ わかりにくい翻訳文の主な原因の一つ

- 単語の誤訳: 前後の文脈により複数の意味となる単語を正しく翻訳できない

- 評価実験で使った文章は情報分野の用語が多く含まれていたが、今後の展望として日常会話などの専門用語が少ない文章を使用した評価を行う事が考えられる

11

おわりに

■ 本研究の結果

- 一般に使用されている主要言語5種の翻訳サービス品質評価
- 評価点数の低い文章の誤訳を翻訳経過と共に抽出

12