

CIEC 第101回研究会報告

日 時 2014年3月8日(日) 13:00~17:00

会 場 大学生協杉並会館 B103・B106

テーマ 「機械翻訳との上手なつきあい方」

参加者 17名

■概要

会誌「コンピュータ&エデュケーション」Vol.35 巻頭インタビューで、ロゴヴィスタ株式会社を取り上げたことが紹介された。その後、会誌英文アブストラクトの現状説明があり、学生のみならず教員や研究者が上手に機械翻訳を利用してほしいという趣旨で、この研究会が開催されたと参加者に伝えられた。

■講演

機械翻訳の仕組みとその利用法

講師 長田 孝治 氏 (ロゴヴィスタ株式会社)

1. 機械翻訳の歴史について、スライドを使ってかなり詳しく説明された。

a. 黎明期 (1980年以前)

1954年 ジョージタウン大学とIBMによるMT公開実験

1959年 機械翻訳専用機「やまと」(電気試験所)完成

1959年 翻訳実験用計算機「KT-1」(九大)

1966年 ALPAC (Automatic Language Processing Advisory Committee) 報告書、機械翻訳に否定的な報告内容であった

1978年 初の日本語ワープロ(JW-10)発売(東芝)、1978 JIS漢字コード制定

b. メインフレーム期 (1980年~1985年)

1982~85年 Muプロジェクト、科学技術庁が実施した機械翻訳プロジェクト、科学技術文献抄録の日英・英日の機械翻訳システムの開発を目指す

1984年 初の商用英日・日英自動翻訳システム(メインフレーム版)が発売、これに続き、機械翻訳システムの発表が相次ぐ

1986~2002年 電子化辞書の構築(日本電子化辞書研究所)

c. 第一次実用期 (1985~1994年)

1985年 初のミニコン版翻訳ソフト発売

1985年 初のPC版翻訳ソフト発売

1987年 「第1回MT SUMMIT(電子協主催)」箱根で開催

1989年 海外ニュース自動翻訳放送試行開始(NHK)

1990年 パソコン通信での機械翻訳サービス開始

1994年 安価な翻訳ソフト登場、カテナからPC版英日翻訳ソフト「コリャ英和!」発売

d. 第二次実用期 (1995年~)

1995年 初のブラウザ連動型翻訳ソフト発売開始

2000年 機械翻訳による日本特許英文検索サービス開始(特許庁)、日本のポータルサイト初のインターネット無料翻訳サービス開始

2006年 Google が、統計ベース翻訳を使用した翻訳サービス開始

2. 機械翻訳の方法について、概略を説明された。

a. ルールベース

ルールベースとは、文法ルール+辞書（一般辞書+専門用語辞書）を利用した機械翻訳のことであり、以下のような手法が説明された。

- ・最も単純なのは「単語変換方式」
- ・構文構造を解析し、ターゲット言語に言語生成する「トランスファー方式」
- ・中間言語経由で翻訳する「中間言語方式」

b. 統計ベース

英語文と日本語文のような大量の言語ペア（コーパス）データを統計処理あるいは、類推して翻訳する方式についてその手法が紹介された。

- ・「統計的機械翻訳」：大規模なコーパスを統計的に分析し確率に基づいて翻訳する
- ・「用例翻訳」：コーパスから入力文に類似する文を選択し対応言語を修正する

さらに、それぞれの方法について、図を使いながらより詳しい説明があった。

○ルールベースについて

- ・日本語の場合、形態素解析を使った分かち書きから始まり、辞書引き・構文解析をとおして訳が生成される。
- ・英語の場合、構文解析の結果として構文木が得られ、それを変形した日本語の構文木を通して日本語に訳される。
- ・中間言語（メタ言語）を媒介して、訳を作成する方法もある。

○統計ベースについて

- ・大規模なコーパスから比較的単純な規則を大量に自動生成させ、それをを用いて最も適切と思われる解を選び出す。
- ・大量のサンプル文（2言語並行コーパス）を活用する場合
文同士の対応はあるが、単語列同士の対応は不明
文内部の対応がうまく付く文同士が高い対応付け確率を持つ
文の長さが類似しているほど、よく対応する文のペアとなる
- ・大量のサンプル文（2言語非並行コーパス）を活用する場合、計算量が莫大となる。

○統計的機械翻訳について

対訳コーパス・翻訳モデルと目標言語コーパス・言語モデルから、原文 S に対する事後確率 $P(T | S)$ が最大となる翻訳文 T を見つける。

○用例翻訳について

対訳コーパス・シソーラス・対訳辞書をとおして翻訳文が作られる。

c. 機械翻訳の精度について

- ・機械翻訳の精度は、言語間の距離による
英語とフランス語は、言語間の距離が近いので精度 98%と非常に高い。英語と日本語（東アジアの言語の多く）は、かなり距離が遠い。日本語と韓国語は、文法が近く精度 93%程度。日本語と中国語は遠いようだが、漢字の借用や文法などで比較的近い部分がある。
- ・言語によって構造維持的か否かの差がある：文の主成分が削除されているなど
- ・分かち書きの問題：曖昧性の問題がある
- ・意味の文脈依存性：機械翻訳では、文脈の評価は困難である

精度の評価ポイント

- ・文の構造

- ・個別言語に固有な文法要素（時制、活用、冠詞など）
- ・語彙、慣用句などが正しく翻訳されているか
機械的評価をすると、統計ベースの評価が高いが、人間が評価すると、ルールベースのほうが高い
機械翻訳精度の向上には ー利用者の立場からー
- ・前編集の重要性：主語の追加、短文・単文化など
- ・機械翻訳結果の後編集：現在 ISO で後編集の要求事項などが議論

最後に、機械翻訳についての考え方と現在の主な利用について述べられた

- ・機械翻訳は、「翻訳お助けソフト」や「辞書引き時間削減ソフト」と考えて利用すると便利さが見えてくると思える。
- ・今、特許や海外事情等のサーベイに利用され、ななめ読みによる情報の判断に利用されている。また、利用者自身の専門分野以外の専門用語理解のために使うと便利である。



■演習

翻訳ソフトの利用法の紹介と翻訳演習

講師 黒石 邦彦 氏 (ロゴヴィスタ株式会社)

1. 製品の紹介

専門家向けの製品 (LogoVista Pro・LogoVista メディカルと医学向けのコリャ英和!) と一般向けの製品 (コリャ英和!) や日英以外の製品 (コリャ英和! マルチリングル) について、その違いなど製品概略が紹介された。

2. 翻訳の場面について

- ・インターネット Web 翻訳
最新情報はインターネットから英語で入手でき、多くの人が Web を利用している。ただし、全体を機械翻訳しても、スムーズに日本語として理解できない場合もある。
- ・大量文書を一括ファイル翻訳
企業等で、機械翻訳の結果を粗訳として利用し、含まれる情報の重要性を判断する際に利用されている。しかし、正確な翻訳とは異なるものである。
- ・正確な翻訳が必要な場合
翻訳エディタ (対訳翻訳) で、翻訳プロセスを丁寧に行うことができる。

3. 翻訳演習

- ・LogoVista Pro2014 の機能紹介をかねて、翻訳エディタや対訳翻訳などについて基本的な操作について説明があった。
- ・翻訳スタイルについて、「です・ます朝」「である朝」などの選択について、具体的な説明があった。
- ・翻訳用の辞書について、システム辞書 (基本辞書) ・専門分野辞書・ユーザー辞書の選択などが話された。
- ・翻訳メモリ (事例文ファイル) ・事例テンプレート・各種条件ファイル (文末判定、原文置換、訳文置換) では、ユーザーがカスタマイズできることで、翻訳の精度が上がっていくことについての説明があった。

基本的な機能や使い方の例を紹介された後、機械翻訳における前翻訳と後翻訳の重要性と、ソフトを使用した実例が説明された。翻訳する日本語の修正・英語に翻訳・翻訳した結果を日本語に翻訳などのプロセスを経てよりよい翻訳を得ることの大切さが、具体的な事例とソフトの機能を使って詳しく説明された。

- ・前編集で利用された機能：未知語検索・原文置換・日本語チェッカー

- ・後編集では、別訳語・別解釈・訳語置換のやり方が紹介された。

休憩前の質疑応答では、英英チェックはできないのか・辞書との連携機能についての質問など、ソフトの機能に関する具体的な質問がいくつかなされ、長田・黒石両氏が答えられた。

■報告

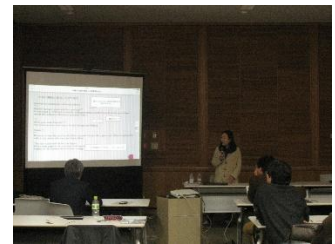
高等学校英語授業での事例紹介

講師 澤田 あゆみ 氏 (神戸国際大学附属高等学校)

- ・高校教師になる前数年間の、海外・日本での日本語教師時代の経験から、google 等の機械翻訳を利用した生徒の文章について、日本語としておかしい点が多々あるなどの問題点が指摘された。
- ・次に、今、授業で取り組んでいる活動から、生徒が実際に翻訳機を使って提出したワークシートを例示し、機械翻訳に関連したポイントについて話があった。1 点目は日本語の入力の質について、2 点目は入力する文章の長さについて問題がある場合、理解できない翻訳となる。(注) 現在取り組んでいるプロジェクト

Asia-Europe Classroom Network

それぞれ準備されているプロジェクトテーマ (世界遺産など) ごとに、参加している学校の生徒がグループワークをとおしてテーマに沿ったコンテンツを作成する。その間、同じテーマを選んだ他国の生徒と SNS を利用してコミュニケーションを図る。<http://aec.asef.org/>



グループワークにおいて生徒の記入したワークシートを使用した例では、

- ・文の頭文字が大文字で書かれていない。
- ・機械翻訳した英語では、日本語で何を言いたいのか分からず、Google 翻訳を使用して英一日で調べたところ、逆に日本語らしい翻訳がでてきた (統計的機械翻訳の特徴との指摘が参加者からあった)。この日本語の内容からして、生徒が最初に質問したかった内容は、「何の目的で、この神社が建てられましたか？」だと想像がついた。
- ・機械翻訳した英語を日本語に翻訳した例「それは松の腕枕のように見えるため松の手の枕を言った理由について外に傾斜している松の手の枕を言っているようだ」しかし、これでは、本当はどのような日本語からこのように訳された質問文ができたのかは謎である。
- ・コンテンツ説明のために、ウィキペディアなどから長文をコピーして翻訳サイトにペーストする。その場合、いくつかの関係代名詞で固められた恐ろしい英文が飛び出してくる。

■質疑応答

翻訳ソフトの機能の詳細について多くの質問が出てきた。それ以外としては、英語の前編集の可能性に関する質問があった。それに対して、英語の前編集は一部組み込まれているが、日本人ユーザーの英語前編集は敷居が高いことが指摘された。学生も機械翻訳をよく使っており、その使い方を教育する必要性も指摘され、機械翻訳で利用されている構文木を学習者が内在化することなどが学習者の課題であるとのしてきもあった。

最後に、小中高での利用の可能性について、小中高の教員に対して講師黒石氏から質問があった。それに対して、デジタル教科書との関連や公立学校での「コリャ英和！」を利用していった例が紹介され、今後の電子辞書での利用可能性が指摘された。

文責 大木 誠一 (神戸国際大学附属高等学校)