



自然言語処理2014

No.14

東京工科大学
コンピュータサイエンス学部
亀田弘之

今日の内容(各種ツール紹介)

- テキスト処理

- OpenNLP(<http://opennlp.apache.org/>)

- Python + NLTK(<http://www.nltk.org>)

- 形態素解析

- Juman

- Chasen

- MeCab

- 統語解析

- OpenNLP

- 機械翻訳

openNLP

- The Apache OpenNLP library is a machine learning based toolkit for the processing of natural language text.
- It supports the most common NLP tasks, such as tokenization, sentence segmentation, part-of-speech tagging, named entity extraction, chunking, parsing, and coreference resolution. These tasks are usually required to build more advanced text processing services. OpenNLP also includes maximum entropy and perceptron based machine learning.

NLTK

- NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to [over 50 corpora and lexical resources](#) such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning.

NLTK(2)

- Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.
- NLTK has been called “a wonderful tool for teaching, and working in, computational linguistics using Python,” and “an amazing library to play with natural language.”

機械翻訳 (Machine Translation; MT)

歴史（概略）

- コンピュータ発明当初から研究された
- その後、ALPACレポートにより実用性が否定（研究が下火）
- 質問応答システムの研究が別に進行
- 積み木の世界システム（SHRDLU）によりAI研究活性化
- 各種翻訳手法の提案・研究
- Web等の電子化テキストが大量に出現
- 用例に基づく翻訳
- 統計的機械翻訳
（まだまだ解決すべき問題は多い）

機械翻訳とは

1. 全自動翻訳
2. 翻訳支援システム
3. 翻訳補助システム
4. 手作業による翻訳

機械翻訳とは

1. 全自動翻訳
2. 翻訳支援システム
 - 前編集 (pre-edit) と後編集 (post-edit)
3. 翻訳補助システム
 - 電子辞書
 - 文法チェック
 - ワープロ (翻訳文書の清書)
4. 手作業による翻訳
 - 大昔のやり方

機械翻訳処理方式の種類

- トランスファー(transfer)方式
- ピボット(pivot)方式(あるいは、中間言語方式)

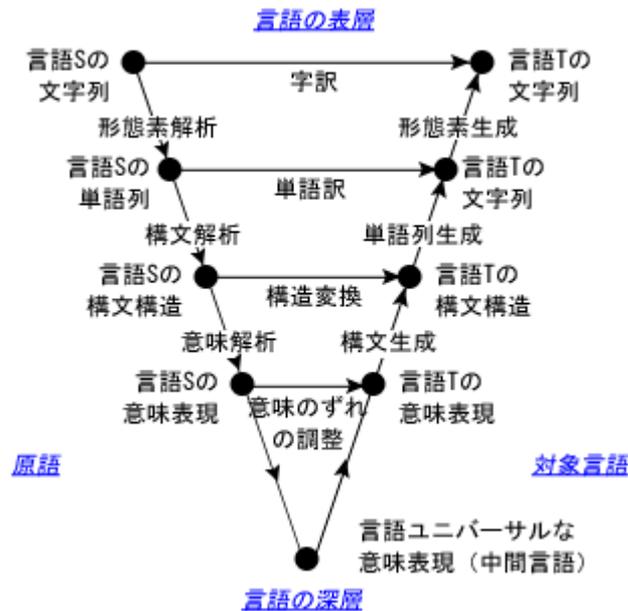


図. 機械翻訳の解析/生成のトライアングル

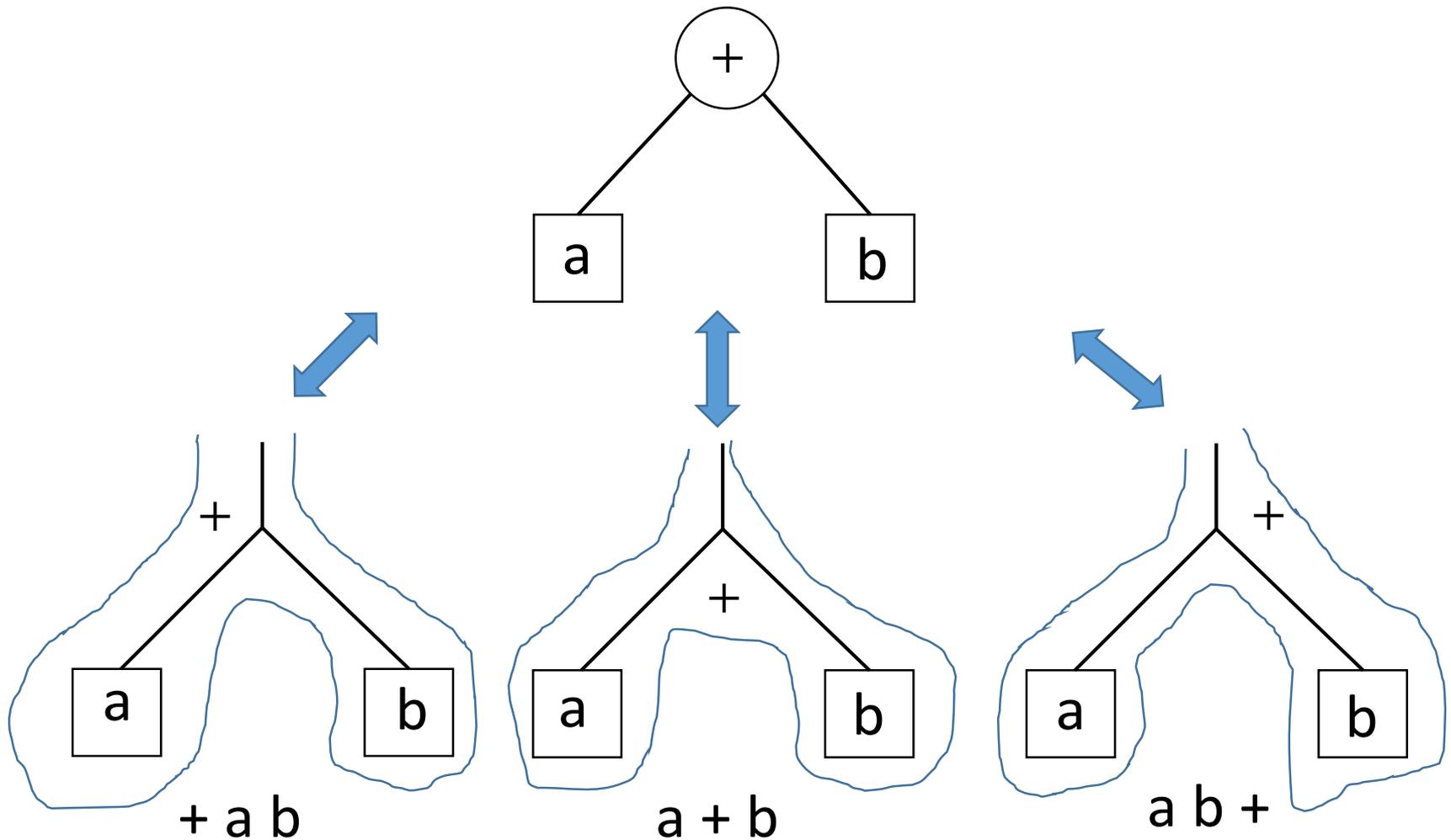
(出典) http://blog.ermitejo.com/wp-content/uploads/2009/01/mt_trianguolo1.png

東京工科大学コンピュータサイエンス学部

機械翻訳システムの例

- Google翻訳
- Yahoo!翻訳
- エキサイト翻訳
- Nifty翻訳
- ATLAS
- LogoVista
- UNL
(<http://www.undl.org/>)
(国連の特許第一号)

翻訳方式の1例紹介



機械翻訳の課題

- (各自で考えてみよう)

そもそも機械翻訳の用途は？

- (価値あるニーズを考えてみよう)

人工無脳について

- AIML(Artificial Intelligence Markup Language)
の紹介

以上で、今日はおしまいです。

- 次回は1月21日(水)です。
- 次回は定期試験対策講座です。