

中韓機械翻訳におけるスーパー関数の構築について

金 珍玉 任 福継 黒岩 眞吾

徳島大学大学院 博士前期課程 知能情報工学専攻

Email: kin, ren, kuroiwa@is.tokushima-u.ac.jp

世界規模のネットワークの発達によって、世界中の人達の距離が非常に近くなった現在、コミュニケーションツールとしての機械翻訳システムの需要が高まっている。また、中国朝鮮族の人口は220万を越え、中韓の交流も近年増加傾向にあり、中韓機械翻訳システムのニーズが高まっている。しかしながら、英語を中心とした日英機械翻訳や中英機械翻訳などの研究は盛んに行われているが、中韓機械翻訳についてはあまり研究されていない。そこで本研究は、計算機上の自然言語コーパスを利用し、スーパー関数に基づく中韓翻訳システムの構築を行うことを目的とし、本稿では、中韓機械翻訳におけるスーパー関数及び対応動詞ベースについて説明する。

キーワード： 機械翻訳，第一、第二スーパー関数

Building Chinese-Korean Machine Translation System With Super-Function

Zhenyu Jin, Fuji Ren, Shingo Kuroiwa

Faculty of Engineering, University of Tokushima Tokushima

Email: kin, ren, kuroiwa@is.tokushima-u.ac.jp

Abstract : In this paper, we propose a Super Function Based Chinese-to-Korean Machine translation. This system uses Super Function (SF) to translate without through syntactic and semantic analysis as most MT systems usually do. By this research, it explains how the First Super Function, the Second Super Function and Verb-base is extracted about Chinese-to-Korean Machine translation.

Keywords: Machine Translation (MT), First Super-Function (FSF), Second Super-Function (SSF)

1. はじめに

世界規模のネットワークの発達によって、世界中の人達の距離が非常に近くなった現在、コミュニケーションツールとしての機械翻訳システムの需要が高まっている。また、中国朝鮮族の人口は220万を越え、中韓の交流も近年増加傾向にあり、中韓機械翻訳システムのニーズが高まっている。しかしながら、英語を中心とした日英機械翻訳や中英機械翻訳などの研究は盛んに行われているが、中韓機械翻訳についてはあまり研究されていない。そこで本研究は、計算機上の自然言語コーパスを利用し、スーパー関数に基づく中韓翻訳システムの構築を行うことを目的とし、本稿では、中韓機械翻訳におけるスーパー関数及び対応動詞ベースについて説明する。

2. スーパー関数

簡単に言えば、スーパー関数は自然言語表現を関数化して、自然言語におけるいろいろな現象をパターンに類別するものである。従来の構文解析、意味解析、言語変換、文生成等の複雑な処理（無理な一般化といってもよい）を避けると同時に、メモリーベース翻訳における探索空間の爆発をも回避した新しい方法であり、人間の思

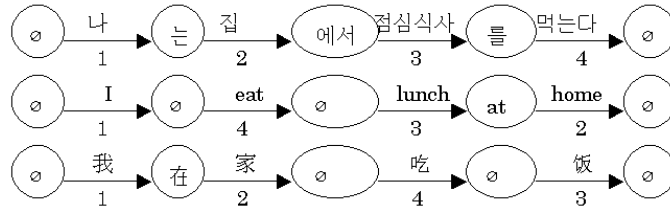
考過程に非常に類似した方法となっている。

3. スーパー関数のアーキテクチャー

スーパー関数を表わして、2つのアーキテクチャーを考慮する：方向グラフ (DG) および変形テーブル (TTB) である。

方向グラフ (DG)

スーパー関数について記述していた場合、次の記法を使用する。STRINGは円ノードによって表している；変数は辺によって表している；また、方向グラフを導入する目的は、スーパー関数と文の間に一致する際に有限状態技術を使用することです。



変形テーブル (TTB)

変形テーブルは節点テーブル (NTB) および辺テーブル (ETB) から成る。節点テーブルはさらにSTRINGテーブル (STB) と呼ばれる、また辺テーブルは変数テーブル (VTB) と呼ばれる。

Node	Korean	English	Chinese
1	∅	∅	∅
2	는	∅	在
3	에서	∅	∅
4	를	at	∅
5	∅	∅	∅

Table1. Example of NTB

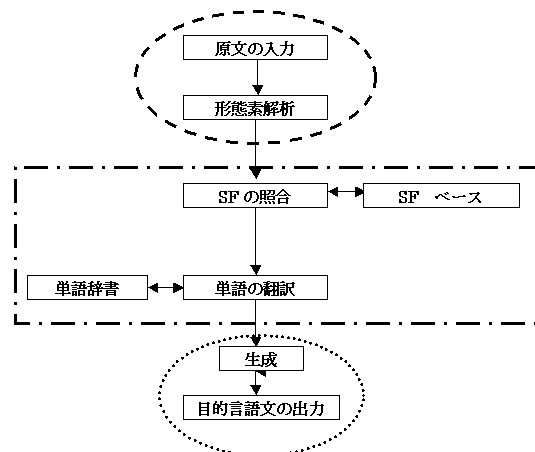
Edge	location(K)	location(E)	location(C)
1	1	1	1
2	2	4	2
3	3	3	4
4	4	2	3

Table2. Example of ETB

スーパー関数を表わすために変形テーブルを使用する。しかしながら、変形テーブルと方向グラフの間に転送することは容易である。

4. 翻訳システムのプロセス

スーパー関数に基づく翻訳システムはソース言語の翻訳を行うために辞書およびスーパー関数を使用する。入力文は最初に、形態素解析されて、次に、原文およびSFとマッチする。このシステムのプロセスは、3つの主要部分から成る：形態素解析、スーパー関数のマッチする、と形態素処理である。

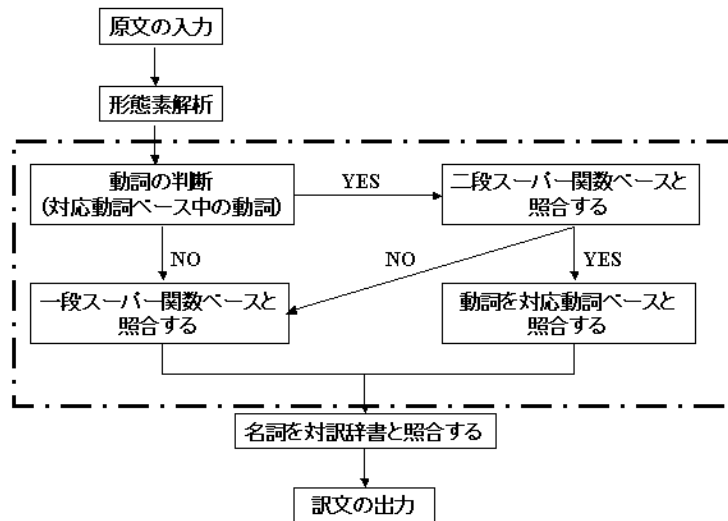


4.1 スーパー関数とのマッチ

スーパー関数は NTB と ETB によって表される。スーパー関数とマッチするでは、NTB の各節点とマッチし、

ETB に各辺の種類を確認することである．入力文は最初に、形態素解析され、名詞部分と残ったスーパー関数部分に分け、名詞部分が対訳辞書と照合する．

スーパー関数部分の動詞を先判断して、もし動詞は対応動詞ベース中の動詞の場合、原言語の名詞、動詞を全部分けます．残りの関数を二段 SF ベースとマッチする．動詞は対応動詞ではない場合は名詞と分けた一段スーパー関数が一段スーパー関数ベースとマッチする．



スーパー関数の抽出

例えば、

나는 집에서 점심식사를 먹는다. (i)

I eat lunch at home.

我在家吃午饭。

어머니는 클럽에서 테니스를 한다. (j)

Mother plays tennis at club.

妈妈在俱乐部打网球。

例 (i)、名詞と分けたら、スーパー関数は

\varnothing_n / 는 / \varnothing_n / 에서 / \varnothing_n / 를 / 먹는다

\varnothing_n / 在 / \varnothing_n / 吃 / \varnothing_n

\varnothing_n / eat / \varnothing_n / at / \varnothing_n

例 (j)、名詞と分けたら、スーパー関数は

\varnothing_n / 는 / \varnothing_n / 에서 / \varnothing_n / 를 / 한다

\varnothing_n / 在 / \varnothing_n / 打 / \varnothing_n

\varnothing_n / **play** / \varnothing_n / at / \varnothing_n

そこで、二つのスーパー関数の中に、動詞と分けたら同じの関数が得られる．

\varnothing_n / 는 / \varnothing_n / 에서 / \varnothing_n / 를 / \varnothing_v

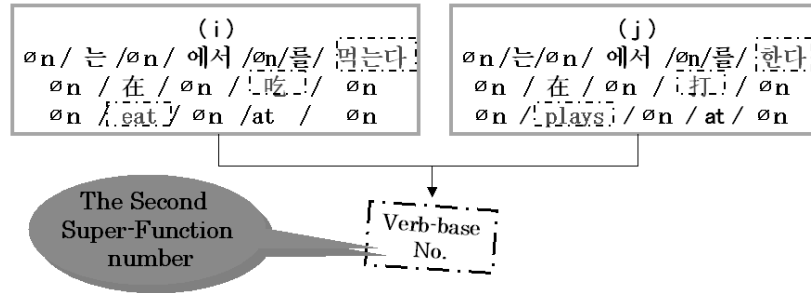
\varnothing_n / 在 / \varnothing_n / \varnothing_v / \varnothing_n

しかし、動詞は名詞のように簡単に変数で表すことができません．

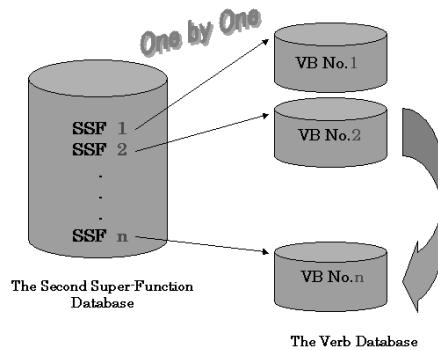
動詞ベース

原文を最初に、形態素解析し、名詞部分とスーパー関数部分に分ける．さらに、そのスーパー関数部分を動詞部分と分ける．複数の原文から同じ二段スーパー関数が得られたら、それを二段スーパー関数ベースに入れ、動

詞部分をその二段スーパー関数の対応動詞ベースに入れる。



The Second Super-Function base



4.2 形態素処理

形態素処理は原言語および目標言語にカテゴリーと特徴セットおよび値の一マッチについて記述するテーブルによって行われる。原言語の各カテゴリーや単語の個々の形態素特徴では、目標言語のカテゴリーおよび形態素特徴に、写像される。簡単に言えば、スーパー関数と名詞と動詞の並び替えで、訳文を出力することである。

5. まとめと今後の予定

今回、計算機上の自然言語コーパスを利用し、スーパー関数に基づく中韓翻訳システムの構築を行うことを目的とし、第一、第二スーパー関数、対応動詞ベース及びシステムについて説明しました。現在までに、1986文から1428個の一段スーパー関数、413個の動詞ベースと二段スーパー関数を得た。今後の予定としてはスーパー関数を集めることである。

謝辞

本研究は科学研究費補助金基盤研究B (No. 14380166) によって実施された。

参考文献

- [1] Ren Fuji 1999 Super Function based Machine Translation, Communication of COLIPS, vol.9 No.1, pp.83-100.
- [2] 電子情報通学会 田中穂積監修 「自然言語処理 規則と応用」 pp. 248-278
- [3] 日本認知科学会編 佐藤 理史著 「アナロジーによる機械翻訳」 pp. 1-37
- [4] 第一回大学と科学公開シンポジウム組織委員会編 日本語の特性と機械翻訳 pp. 92-144
- [5] 吉村賢治 「自然言語処理の基礎」 pp. 52-80
- [6] 工藤 拓、松本 裕治 2002 チャンキングの段階適用による係り受け解析、情報処理学会論文誌、Vol 43、No. 6 pp. 1834-1842
- [7] Fuji Ren 1996 A Japanese-Chinese Machine Translation System Based on Translation Rules Acquired from Corpora, LASSO-96: The 25th annual meeting of the Linguistic Association of the South-west, (Baton Rouge, USA, 1996), pp.1-8