

簡易型中日機械翻訳実験システムについて

寺田 栄男 孫 東依 田町 常夫
福岡工業大学

我々は限定された分野で質の高い翻訳ができ、かつ小型機に適したシステムの実現を目指して英日機械翻訳システムの研究を行ってきた。さらに非英語圏諸国間の翻訳の重要性を感じ、1988年より中国大連理工大学との共同研究により、中国語日本語間の翻訳にも着手した。

本稿では、中日翻訳に必要な中国語の構文データの整理と、パーソナル・コンピュータによる中日機械翻訳実験システムの概要および予備実験の結果について述べる。

On the Experimental Chinese-Japanese
Machine Translation System

Hideo TERADA Donghui SUN Tsuneo TAMACHI

Fukuoka Institute of Technology
3-30-1, Wajiro-Higashi, Higashi-ku, Fukuoka, 811-02 Japan

We have been studying the English-Japanese machine translation system suited for small size machine for the purpose of realizing accurate translation in the limited field. We also started the research on translation between Chinese and Japanese in collaboration with Dalian University of Technology in China from 1988.

This paper describes: (1) the arrangement of Chinese syntactic data required for translation, (2) the outline of experimental Chinese-Japanese machine translation system by using personal computer and some results of translation.

1. はじめに

最近、機械翻訳の実用化研究が盛んであるが、現状では英語と他国語間を対象とした汎用を狙いとするものが多い。我々は過去において自然言語理解システム、英日、日英機械翻訳システムの開発を行ってきたが、非英語圏諸国間の翻訳の重要性を感じ、中国語、韓国語と日本語間の翻訳にも着手してきた。中国語に関しては、1988年より中国大連理工大学との共同研究により、中国語日本語相互間の小型翻訳システムの開発を行なってきた。分野はとりあえず人工知能関係の文献に限定し、質の高い翻訳を行うことを目指している。本研究では中国語から日本語への翻訳を担当し、共通の対訳テキストを使用し、漢字入出力、翻訳の方式、各種辞書の3つの問題についてそれぞれ検討を進めている。最初は分節された文章入力を対象とし、それぞれの翻訳実験システムのプロトタイプを完成し、以後文法辞書の精密化、辞書の拡充改善を図るものである。中国語の特色として、文章の分節法自体が形態素レベルから文脈処理を必要とすること、従って、漢字、単語の意味属性に関する処理を本格化する必要がある。その他、中国漢字と日本語漢字の両方を混在させること、機種種の相違の問題がある。

本稿では、構文情報を中心にデータを収集し整理を行ない、基礎資料の妥当性を確かめるために、小規模な実験システムをパソコン上に作成し、予備実験を行った結果の概要について述べる。

2. 中国語語法

翻訳処理に必要な文法事項に関する整理を行った。まだ十分とは言えないが、翻訳実験と並行して不備を補いつつある。以下では品詞の分類、連語、文型、時態について述べる。

2. 1 中国語の品詞

中国語の品詞を次のように分類した。

1) 名詞

- ①普通 → 学生
- ②抽象 → 成功
- ③固有 → 北京
- ④時間詞 → 昨天
- ⑤方位詞
 - イ. 単純 → 上、前
 - ロ. 複合 → 前面

2) 動詞

①一般

- イ. 繫詞 → 是、叫
- ロ. 存在詞 → 有
- ハ. 自動詞 → 走、坐
- ニ. 他動詞 → 看、打
- ホ. 二重目的語をとる → 教、問
- ヘ. 動詞句を目的語にとる → 喜歡
- ヘ. 主述連語を目的語にとる → 知道、希望
- ト. 兼語構造をとる → 使、請

②方向 → 来、去

3) 代名詞 (代詞)

- ①人称 → 我、他
- ②指示 → 那、这
- ③疑問 → 什么

4) 数詞 → 一、五

5) 量詞

- ①名量詞 → 个、張
- ②動量詞 → 次、回

6) 助動詞 (能願動詞)

- ①意志、願望 → 要、想
- ②可能 → 能、会
- ③当為 → 要、得

7) 形容詞

- ①単音節 → 好、忙
- ②二音節 → 干净
- ③複合 → 鮮紅

8) 副詞

- ①程度 → 最、很
- ②範圍 → 都、只
- ③時間 → 正、就
- ④語気 → 可、倒
- ⑤肯定・否定 → 当然、没有
- ⑥重複・頻度 → 又、再

9) 前置詞 (介詞)

- ①時間 → 在、從
- ②場所・方向 → 在、向
- ③目的 → 为、为了
- ④原因 → 因、由于
- ⑤対象・関連 → 对、关于
- ⑥手段・方式 → 把、用
- ⑦共同 → 和、同
- ⑧比較 → 比
- ⑨排除 → 除、除了
- ⑩受動 → 被、叫

1 0) 接続詞 (連詞)

- ①名詞・代名詞・名詞性連語を接続
→ 和、同
- ②動詞・形容詞・(動詞、形容詞性) 連語を接続
→ 并、而且
- ③各種の賓語や各種の連語を接続
→ 与、或
- ④文を接続
→ 如果、不但

1 1) 助詞

- ①構造 → 的、地
- ②時態 → 了、着
- ③語気 → 吗、呢

1 2) 接辞

- ①接頭 → 第
- ②接辞 → 化

2. 2 連語

中国語では、日本語のような格助詞がなく、また英語のような語尾変化もないので、語順が重要となる。単語が一定の規則に従って結合し、一定の意義を表わすものが連語である。中国語では連語がそのまま文の成分となる、つまり単語と同じ働きをする。単語の基本的な結びつきを示すと次のようになる。

- 1) 修飾構造：修飾語+非修飾語
- 2) 並列構造：2つ以上の語句の並列
- 3) 補充構造：動詞(形容詞)+補語
- 4) 動賓構造：動詞+目的語
- 5) 介賓構造：前置詞+目的語
- 6) 連動構造：動詞(句)+動詞(句)
- 7) 主述構造：主語+述語
- 8) 兼語構造：動詞+(代)名詞+動詞
- 9) 固定構造：熟語的なもの

修飾語は名詞を修飾するもの(限定語)と動詞・形容詞を修飾するもの(状況語)とに分けられる。限定語と被修飾語の接続には、間に構造助詞”的”を用いるものと、用いないで直接接続するものがある。限定語の種類と”的”を付けるか否かをまとめると表1のようになる。状況語の場合も被修飾語の間には構造助詞”地”を用いるものと、用いないで直接接続するものがある。状況語と”地”を付けるか否かをまとめると表2のようになる。

動詞や形容詞の後にあって、補充的に説明している

部分を補語という。動詞の後に置かれる点では賓語(目的語)に似ているが、賓語は一般に名詞性のものであり、補語は数量補語を除いては非名詞性のものである。補語の種類、形態、補語となるものを要約すると表3のようになる。

表1 名詞修飾(限定語)

限定語	”的”を付ける	”的”を付けない
数量詞	——	○
代名詞	領属関係を表わすとき	中心語が親族、所屬單位のとき 指示代名詞+量詞のとき
形容詞	二音節(一般に付ける) 形容詞に修飾語が付くとき 付加成分の付いたもの 重ね型のとき	単音節のとき
名詞	領属関係を表わすとき 時間・場所名詞のとき	中心語の性質を説明しているとき 熟語化しているとき
動詞	○(一般に付ける)	——
連語	○	——

表2 動詞・形容詞修飾(状況語)

状況語	”地”を付ける	”地”を付けない
数量詞	重ね型	○(一般に)
代名詞	——	○
形容詞	二音節(一般に) 重ね型(一般に) 形容詞に修飾語が付くとき	単音節
名詞	○(一般に)	時間・場所を表わす名詞
動詞	○	——
副詞	——	○(一般に)
連語	○	——
介詞連語	——	○

2. 3 中国語の文型

中国語の単文は、述語の性質によって名詞述語文、形容詞述語文、動詞述語文、主述述語文に分けられる。

1) 名詞述語文

S + N

S + ”是” + N

2) 形容詞述語文

S + Adj

3) 動詞述語文

①SV型

・目的語をとらない

②SVO型

・目的語を1つとる

③SVOO型

・目的語を2つとる

④S1+V1+S2+V2型

・主述連語(S2+V2)がV1の目的語となる

4) 主述述語文

・S1+S2+V(Adj)型

動詞述語文には次に示すような特殊な構造をとるものがある。

1) 是～的文

～には、動詞述語、主述述語がはいる。

2) 兼語式

N1+V1+N2+V2

〔主動者+使役性の動詞+受動者+動詞句〕

・N2はV1の目的語であり、V2の主語

・V1は使役性の動詞(使、叫、讓・・・)、

または有、是(無主語の文もある)

3) 連動式

V1+(N1)+V2+(N2)

・V1とV2は主語を共有する

意味

①V2がV1の目的を表わす

②動作の前後関係を表わす

③V1がV2の方法手段を表わす

④2つの同時的な動作を表わす(V1に"着"を伴う)

⑤V2がN1を対象とする

4) 受動式

N1+Pre+N2+V

〔受動者+受動の前置詞+主動者+動詞〕

・受動の前置詞(被、叫、讓)

・"被"の場合はN2がなくてもよい

5) 処置式

N1+"把"+N2+V

目的語を動詞の前に出す

・N2は特定されたもの、既知のもの

・Vには使用できない動詞がある

①自動詞

②非動作

③心理活動、感覚、知覚

④方向

・Vには目的語、状況語、補語あるいは動態助詞の"了"、"着"などがつく

・否定副詞、時間詞、能願動詞は"把"の前

6) 存現文

場所詞(時間詞)+V+N

〔場所詞+動詞+人(事物)]・・・存在を表わす

〔場所詞(時間詞)+動詞+名詞]・・・出現・消失を表わす

・Nは特定のもの

・Vは結果補語・方向補語、助詞の"了"、"着"を伴う

2. 4 時態

中国語には時制(tense)を表わす文法範疇はなく、時間を表わす名詞、副詞などで表わす。中国語では動作の開始・進行・持続・完了といった時態(aspect)が重視される。中国語の時態についてまとめたのが表4である。

3. 中日翻訳実験システム

2で述べた構文情報に関する基礎資料の妥当性の検証するため、および翻訳システム構築に必要な情報を得るために小規模な実験システムをパソコン上に作成し予備実験を行なった。

3. 1 実験システムの概要

実験システムは、過去に開発した英日翻訳用のシステムFILTS¹⁾をもとにして、入力部と解析部を中国語用に手直したものを使用した。使用言語はPASCALである。翻訳処理の概要は次のようになる。

1) 中文入力

入力は中国漢字を使用し、分かち書き(原則としてピンインの分かち書きに合わせた)とした。中国漢字の入力には、市販の中国語ワープロを使用した。中国語ワープロにより予め文書ファイルを作り、エディタにより修正したものを入力ファイルとするか、フロントエンドプロセッサとして使用し、キーボードより直接入力する。

2) 辞書引きおよび形態素解析

単語辞書は中国漢字の第1文字の内部コードをキー

表3 補語（動詞、形容詞の後に置かれる補充成分）

種類	形態	補語となるもの	備考
1. 結果	(動詞/形容詞) + 補語	動詞 (見、成、走、・・・) 形容詞 (大、小、多、少・・・)	・動詞や補語の間には助詞や目的語がはいらない ・複合動詞とみなされるものが多い ・否定は動詞の前に"没(有)"を加える
2. 状態	(動詞/形容詞) + 得 + 補語	動詞 形容詞 主述連語	注B
3. 程度	① (形容詞/動詞) + 補語 ~ 了 ② (形容詞/動詞) + 得 + 補語	①の場合 形容詞 (多、速、死・・・) ②の場合 形容詞、副詞、句や節	・注B ・否定は"不"を使い、補語を否定する
4. 方向	①単純 動詞 + 補語 ②複合 動詞 + 補語	①の場合 来、去、上、下・・・ ②の場合 {上、下、出、・・・} + {来、去}	・賓語の位置 ・ {上、下、出、・・・} の時はA1 {来、去} の時は状況によりA2 またはA1 ・場所が賓語の時必ずA2
5. 可能	①動詞 + {得/不} + {結果補語/可能補語} ②動詞 + {得/不} + 了		・否定は"得"の代わりに"不"を用いる ・賓語の位置、短い時はA1、長い時はA3
6. 数量	①動量 (回数) 動詞 + 補語 ②時間量 動詞 + 補語 ③比較数量 形容詞 + 補語	①の場合 数詞 + 動量詞 ②の場合 時間を表す語 ③の場合 数詞 + 名量詞	・賓語の位置、名詞の時A1、人称代名詞または人を表す名詞の時A2 ・②の場合、名詞が賓語の時B

注 賓語の位置
A1. V + 補語 + 賓語
A2. V + 賓語 + 補語
A3. 賓語を文頭に置く

注B 賓語がある時は動詞を繰り返す
S + V + O + V + 補語

表4 中国語の時態

種類	表現	主な訳語	備考
1. 将動態	要 ~ 了 (得、就、快) 要 ~ 了 快 ~ 了	~しようとしている まもなく~する ~しそうだ	
2. 開始態	動詞 + 起來 動詞 + 上	~しだす	
3. 進行態	① ~ 呢 ② {在、正、正在} + 動詞 ③ ①と②の併用	~している ~しているところだ ~しつつある	・過去、未来でも使う ・否定 "没(有)"
4. 持続態	動詞 + 着	~している ~である	・"正在~呢"と併用されることもある ・"着"の後に付くのは賓語だけである ・否定 "没(有) ~ 着"
5. 経験態	動詞 + 過 動詞 + 過 ~ 了	~したことがある ~してしまった	・語順 動詞 + {結果・方向補語} + "過" + 賓語 動詞 + "過" + 動量補語 ・否定 "没(有) ~ 過"
6. 完了態	動詞 + 了	~した	・現在、過去、未来すべてに用いる ・語順 動詞 + {結果補語} + "了" + 賓語 ・否定 "没(有)"、"了"は付けない

としたランダムファイルとして構成してある。単語辞書を引き、品詞、訳語、属性を求める。熟語、慣用句については最長一致法により検索を行なう。分かち書きされない動態助詞などは形態素として扱い処理をする。

3) 多義解消

辞書引きで複数の品詞・訳が得られたときは、前後の語句の品詞、属性情報により接続関係辞書を参照して品詞・訳語を決める。決まらないときは、決定を保留して先へ進む。

4) 構文解析

構文辞書は文脈依存構造を基本としたプロダクションルールの形で与えられている。処理はボトムアップに行い、構文辞書を参照し句になるものを順次まとめているという操作を繰り返す。

5) 訳文生成

訳文を生成する。4) の処理で解析に失敗して最終的な文の形が得られない場合にも、部分的な訳は生成されているので、修正の便のために部分訳をスラッシュ"/"で区切って表示するようにしている。

実験システムでは、ステップ実行が可能で、規則が適用される過程を逐次見ることができる。また規則の使用頻度のデータを収集するのも可能である。

翻訳処理を行なう本体の他には次のようなプログラムを用意している。

- ・単語辞書作成プログラム
- ・訳文出力プログラム
- ・統計データ出力プログラム

3. 2 翻訳実験

連語を作る規則を精密化したものと動詞のパターンから文法規則を作成した。構文規則はプロダクションルールの形で与え、条件部で意味的あるいは構文的な適用条件を指定した。

構文規則の例

(ad/n)(ad/c)(aj/)

→ (aj/);E;(+あまり)(N)(HITEI)

部分否定

(v/g)(pro/h)(n/) → (v/);E;()(+に)()(#1-#2)

授与動詞

訳語の選択は、共起する語との関係で決め、3. 1で述べた多義解消段階で処理している。

”看”： 見る・・・映画など
読む・・・本など

翻訳実験は初歩の中国語³⁾より、約150文について行なった。文型は2. 3で述べた名詞述語文、形容詞述語文、動詞述語文、主述述語文の単文のみを対象とした。単語辞書の登録語数は約200語で、その内訳は次の通りである。

名詞 (91)	代名詞 (27)	動詞 (33)
助動詞 (4)	形容詞 (12)	副詞 (14)
前置詞 (6)	接続詞 (1)	数詞 (4)
量詞 (7)	助詞 (5)	

また使用した文法規則の数は93個である。

翻訳結果の一部を図1に示す。

- 1: 中文 → 这是你的课本吗?
訳文 → これはあなたの教科書ですか。
- 2: 中文 → 我念课文, 练习发音。
訳文 → 私は課文を読み、発音を練習する。
- 3: 中文 → 老师教我中文。
訳文 → 先生は私に中国語を教える。
- 4: 中文 → 书房里有书桌、椅子、书架。
訳文 → 書斎の中に文机、椅子、本棚がある。
- 5: 中文 → 我去图书馆借书。
訳文 → 私は本を借りに図書館に行く。
- 6: 中文 → 你看过中国电影吗?
訳文 → あなたは中国映画を見たことがあるか。
- 7: 中文 → 我在听无线电呢。
訳文 → 私はラジオを聞いている。
- 8: 中文 → 你能参加明天的活动吗?
訳文 → あなたは明日の行事に出席することができるか。
- 9: 中文 → 我今天不很忙。
訳文 → 私は今日あまり忙しくない。
- 10: 中文 → 我很想去中国。
訳文 → 私はとても中国に行きたい。
- 11: 中文 → 他工作太忙。
訳文 → 彼は仕事(働く)がたいへん忙しい。
- 12: 中文 → 从前我见过他一次。
訳文 → 以前私は1度彼に会ったことがある。
- 13: 中文 → 这本小说你看完了吗?
訳文 → この小説、あなたは読み終えたか。
- 14: 中文 → 黄河不比长江长。
訳文 → 黄河は長江より長くない。
- 15: 中文 → 我从今年四月起开始学的中文。
訳文 → 私は今年の4月から中国語を学び始めた。

図1 翻訳例

- 例文1 → 名詞述語文
- 例文2 → SVO型
- 例文3 → SVOO型
- 例文4 → 存在を表わす文
- 例文5 → 連動文
- 例文6 → 経験態
- 例文7 → 進行態
- 例文8 → 可能の助動詞
- 例文9 → 形容詞述語文、部分否定
- 例文10 → 願望の助動詞
- 例文11 → 主述述語文
- 例文12 → 経験態、数量補語
- 例文13 → 結果補語
- 例文14 → 比較の表現
- 例文15 → 前置詞の呼応

参考文献

- 1) 寺田他：英日機械翻訳システムFILTSにおける訳文述語部の処理について、福工大エレクトロニクス研究所所報、vol.5(1988)
- 2) 寺田他：簡易型中日機械翻訳システム構築のための予備実験について、平成元年度電気関係学会九支連大
- 3) 金丸：初歩の中国語、昇龍堂
- 4) 三野：中国語文法の基礎、三修社
- 5) 中山他：新しい中国語語法、東方書店

実験で用いたような簡単な例文の範囲ではある程度の翻訳はできるが、実際の技術文献を対象とした場合には解析の段階での曖昧さが多く発生する。曖昧さが生じやすいのは

①多品詞の語が多い

前置詞（介詞）のほとんどは動詞でもある
名詞であつて動詞となるものが多い

②形態的な特徴が少ない

日本語のような格助詞がない
英語のような語尾変化をしない

③虚詞（前置詞、接続詞、助詞など）の省略が多い

などの理由による。

解決には意味処理を導入して中国語に適した解析方法を検討していく必要がある。

4. むすび

中日機械翻訳実験システムの概要について述べた。中国語では、解析の手がかりとなる構文上の特徴が少なく、曖昧性が発生しやすい。今後中国語に適した解析方法を検討しなければならない。予備実験では文法書より抽出した単文について翻訳実験を行なったが、実際の技術文書についての翻訳を行なうためにはさらに、中国語テキストの分析、データ収集を行なう必要がある。また入力に関しては、中国語は分かち書きされないで、ベタ書きの入力を分節する方法も検討しなければならない。以上は今後の課題である。

本研究の一部は文部省科学研究費「一般研究（C）63580031」によつた。