

中日数量詞の機械翻訳規則を開発するための調査分析

邵敏 尹大鵬 任福継 黒岩眞吾

Min Shao Dapeng Yin Fuji Ren Shingo Kuroiwa

(徳島大学)

中日両言語の助数詞を比較した時、中国語の助数詞は日本語より全体として量が多いと言える。そのため翻訳時においては以下の三つが難しい問題として挙げられる。1. 一対一の助数詞。2. 一つの日本語の助数詞はいくつかの中国語の助数詞と対応する。3. 一つの中国語の助数詞はいくつかの日本語の助数詞と対応する。

本稿では主に中国語の数詞、数量詞フレーズの特徴、概数の表示方法、及び特殊な数詞の使い方について論じる。中国語と日本語との助数詞を比較して分析する。最後に中日間における数量表現の機械翻訳ルールについて論じる。これは、数量詞が含まれる中国語を日本語へ変換するとき、数詞と量詞の挿入位置候補が日本語では三通り存在するという問題があるが、中国語文を形態素解析し、品詞の構造を利用することによって、その位置を一意に定めることができるというものである。

キーワード：数詞、助数詞、数量表現、数量詞。

An analysis for machine translation rule development of Chinese-Japanese quantifier

Min Shao Dapeng Yin Fuji Ren Shingo Kuroiwa

Comparing to Chinese-Japanese has more quantifier. Three problems are very difficult to solve during Chinese-Japanese machine translation. 1. One-to-one correspondence quantifier .2. A Japanese quantifier corresponds to many Chinese quantifier .3. A Chinese quantifier corresponds to many Japanese quantifier.

This paper presents the properties of Chinese Numeral and Quantity phrase, and mainly discusses the expression method of approximate number and the usage of special numeral. A comparative analysis for the properties of quantifier between Chinese and Japanese is implemented. It is Carrying Morphological analysis on the Chinese sentence, depending on the structure of part of speech in Chinese sentence, and three kinds of inserted position of Japanese numeral and quantifier in a sentence, expounding rules that a sentence which contains quantifier is translated from Chinese to Japanese.

Keywords: Numeral、quantifier、Expression of amount、quantifier.

1. はじめに

機械翻訳の研究は長い歴史を持ち、特に、最近、中日機械翻訳が盛んに研究されている。しかし中日両言語の数量表現についてはあまり研究されていない。そのため多くの問題が残されている。

中日両言語の助数詞を比較すると、中国の助数詞は日本語より量が多いということである。日本語では、数量詞はほとんどが単に数量を表すにすぎないが、中国語では、数量詞は数量を表すほかに、表現上及び文法上特殊な働きをする場合がある。そのため、特に数量表現のない日本語文を中国語に翻訳するとき、数量表現を付加しなければならない場合が多い。

中国語と日本語両言語の数量表現について以下の三種類のパターンがある。

1. 中国語には数量表現がある、日本語にも対応する数量表現がある。
2. 中国語の文には数量表現がある、対応する日本語の文には数量表現がない。
3. 中国語の文には数量表現がない、対応する日本語の文には数量表現がある。

そこで、本稿では中国語の数詞の特徴、概数の表示方法、及び特殊な数詞の使い方について論じた。次に中国語と日本語との助数詞を比較して分析した。最後に中日間における数量詞の機械翻訳ルールについて論じた。

2. 中国語の数詞

数詞とは数量や順序の概念を表す語である。個数、数量、度数を表す「基数詞」と順序を表す「序数詞」とがある。不定数を表す語も数詞に含める。

2.1. 中国語数詞の概数の表し方

話し手が正確な数を知らないか、或いは言いたくないか、または言う必要のない場合には大体の数を言うことがある。概数には次のような数種類の表し方がある。

- 二つの隣合った数詞を続けて用いる

続けるのはふつう、係数詞で、通常、数の小さい方を前に置き、数の大きい方は後に置く。

例： 八九千（八千から九千ぐらい）

- 数詞の後に概数を表す語句をつける

主なものには「来」(lai)、「多」(duo)、「把」(ba)「左右」(zuoyou)、「上下」(shangxia) などがある。このような語は位置がかなり重要である。以下に「来」を例として位置の決める方を述べる。

「来」はすべての名助数詞の後ろに用いることができるが、「来」の置かれる位置は名助数詞の種類に関係している。なお、名助数詞を二種類に分けて考えると、一つは、連続した量を表すもので、「斤」、「両」などの度量衡単位や、時間を表す「年」「天」などである。この種の助数詞が表す単位は「さらに小さい単位から成り立っている」という特徴を持つ。もう一種の名助数詞は不連続量を表すもので、主に「只」、「把」など「個体助数詞」である。

「来」が連続量を表す助数詞と一緒に用いられる場合には、「来」の置かれる位置は二つある。それを下記の<1><2>式で示す。

数詞(条件1) + 「来」 + 助数詞 + 「名詞」 <1>

数詞(条件2) + 助数詞 + 「来」 + 「名詞」 <2>

ただし、条件1：10以外の0で終わるもの

条件2：1, 2, 3..., 9で終わるもの及び10

例： 五十来里路（五十華里の道のり）

五里来路（五華里の道のり）

- 数詞の前に概数を表す語句をつける

主なものには「成」(cheng) , [上](shang), [近](jin) [约](yue) などがある。

「成」(cheng) は「百」以上の位数詞の前にもみ用いられ、一定の単位に達していることを表し、数が大きいという含意がある。[上](shang) も「百」以上の位数詞の前にもみ置かれ、数が「とどく」「達する」という意味を表す。[近](jin) は数量フレーズの前に用いられ、「その数に達してはいないが近い」ということを表す。[约](yue) は数量フレーズの前に用いて、後ろの数量フレーズの示す数からあまり隔たっていないことを表す。

例： 约三天的时间。（約三日間。）

近三年的时间。（三年に近い時間。）

- 「几」(ji), 「两」(liang) の活用による概数表現

「几」は疑問代詞であるが、時には疑問を表さず、概数を表すこともある。

例： 你去拿几本书来。（君本を数冊もって来てくれたまえ。）

「两」を活用して概数を表す用法は基本的に「几」と同様である。

例： 过两天我再去看你。（二、三日してから君に会いに行くよ。）

「两」(liang) は主に肯定の場合に用いられることが多いのに対し、「几」は肯定と否定の両方に使用可能である。

2.2. 特殊な数詞

中国語では、「二」、「两」、「半」、「一」など特殊な数詞があり、機械翻訳の観点から、これらの数詞の用法

を検討することが重要であると考える。「二」、「兩」とはどちらも「2」を表すものであるか、用法に違いがある。「半」は二分の一という意味である。「一」は目的語を限定する場合は省略が可能である。「一」には多くの派生的用法があるが、それらの場合必ずしも数を表さず、また品詞も変化が生じる。

3. 助数詞

日本語の数体系は中国語から輸入されたもので、10 が「十」、11 が「十一」というように完全に十進法で合理的だが、物を数える際には適切な助数詞を用いる必要がある。

3.1. 助数詞の定義

(1) 数詞の後につけられる (2) 数えられている事物を示す名詞と共起得る (3) 個別化された事物の数を示す際に用いられる。

3.2. 助数詞の機能

助数詞が異なれば指示する形状も異なる。

例： この手袋は1 {本/玉/巻き/束/山} の毛糸編んだ。
珍しい花を1 {本/輪/束/鉢/株} もらった。

名詞(句)の意味が分からなくても助数詞である程度推測が可能。

例： 「たほいや」

- (1) これは1本の「たほいや」だ。
一次元的細長い物体？
- (2) これは1個の「たほいや」だ。
比較的小さい三次元的な物体？
- (3) これは1枚の「たほいや」だ。
二次元的に平坦な物体？
- (4) これは1軒の「たほいや」だ。
比較的な小さい建物、家屋、小屋？
- (5) これは1匹の「たほいや」だ。
虫、比較的小さい動物？

3.3. 助数詞の選定

日本語の助数詞は中国語由来のものが多いが、中国語とずれていることがよくある。例えば日本語の「本」は細長い物の助数詞だが、中国語の「本(běn)」は書物の助数詞だ。日本語では書物自体を「本」と呼ぶ。中国語で細長い物を数えるのは「条(tiáo)」である。日本語の「隻」は船の助数詞だが、中国語の「隻(只)(zhī)」は「(双)(shuāng)」に対する言葉で、二つ組のうちの片方を意味する。漢語の「隻腕」、「隻眼」はそれぞれ片腕、片目のことだ。

4. 数量表現の翻訳ルール

数や量に言及した表現として、日本語では、例えば、豚を数えるとき「1匹」、「2匹」、「3匹」のように、数字「1」、「2」、「3」の後に「匹」という助数詞をつける。「1匹」、「2匹」、「3匹」のように「数詞+助数詞」の形を取るものを「数量詞」と呼び、数量詞とそれに対応名詞が含まれた句や文を「数量表現」と呼ぶこととする。

日本語の数量表現について、次の三つの型をあげる。

1. 昔ある所に3匹の仔豚が住んでいました。 (QのNC型)
2. 昔ある所に仔豚が3匹住んでいました。 (NCQ型)
3. 昔ある所に仔豚3匹が住んでいました。 (NQC型)

数量詞をQ、名詞をN、格助詞Cとすると、1.は「QのNC型」、2.は「NCQ型」、3.は「NQC型」

と表示することができる。

4.1. 数量詞が含まれる文の翻訳ルール

1. 中国語の文の構造： 「時間の数量詞+動詞+数量詞」.
対応する日本語文の構造：「時間の数量詞+に+数量詞+動詞」.
中国語の文の構造： 「数量詞1+数量詞2+名詞」.
対応する日本語の文の構造：「数量詞1+名詞+数量詞2+ずつ」.
例： 一天去一次。
一/m 天/q 去/v 一/m 次/q .
一日に一回いく。
2. 中国語の文の構造：「再」+…+数量詞+….
対応する日本語文の構造：もう+数詞+…+動詞.
翻訳するとき、中国語文の数量詞に対して特定の訳語があれば、それに従って翻訳する。
例： 能再给我一个枕头吗？
能/v 再/d 给/p 我/r 一/m 个/q 枕头/n 吗/y ？
枕をもう一つもらえますか？

4.2. 数量詞「一」が含まれる文の翻訳ルール

1. 中国語の文の構造：動詞（助詞、名詞、形容詞、代名詞）+「一」+助数詞+名詞（代名詞、動詞、副詞前置詞、形容詞）。
翻訳する時日本語の文の数量詞を翻訳しない。しかし数量詞の後に時間名詞（週、年、時間、学期、度量など）が来る場合、必ず翻訳する。
例：1. 请看一下这单子上担当者姓名。
请/v 看/v 一/m 下/q 这/r 单子上/n 担当者/n 姓 名/n .
このメモの受付者名を見てください。
2. 中国語の文の構造：名詞+動詞+「一」+助数詞。
翻訳する時日本語の文の数量詞を翻訳しないです。
3. 中国語の文の構造：「一」+助数詞+「一」+助数詞+「的」+名詞。
あるいは 「一」+量詞+量詞+（「的」）+名詞。
日本語の文の構造：「一」+量詞+「一」+量詞+・・・+動詞。

5. まとめ

本稿では中日機械翻訳における数量表現処理について述べてきた。特に中国の数量表現の文法上の特殊な働きについて考察した。数量詞が含まれる中国語を日本語へ変換するとき、中国語文を形態素解析し、品詞の構造を利用することによって、その位置を一意に定めることができるというものである。

今後は中国語の文を構文解析した後、格助詞の確認、語順の変換、時制を確認し、翻訳精度を高めるルールを収集していく予定である。

参考文献

- [1] 任福継 (1992) : JCMT システムにおける訳文の特殊な数量表現に関する一考査, 自然言語処理 9 1 - 1 .
- [2] 飯田朝子 (1996) : 日本語分類辞の選定における形状の関与—「1本の毛糸」と「1玉の毛糸」をめぐって— 日本言語学会 第112回大会.
- [3] 林佩芬 (2003) : 数量表現「NQC型」における集合と分配について『ことばの科学』16号 p p 55~65 .
- [4] 木村英樹 (1996) 、中国語はじめの一步、筑摩書房、ISBN 4-480-05666-1 .