

## 動詞の結合価を用いた中国語構文解析

張 銀峰 横山 晶一 西原 典孝  
山形大学工学部

## 1 はじめに

中国語の単語は語尾変化が無く、単語の性質を決められる形態もない。ゆえに、一つの単語で文法的意味を明確に表すことができない。従って、多くの言語学者は中国語の構文は単語の語順と少数の虚詞によって研究すべきだと主張している。しかし、実際語順は活用により変わり、虚詞も活用が複雑で省略される場合が多い。そのため、単語に含まれている構文解析に役に立つ情報を引き出す必要がある。

中国語の動詞は文の中で極めて重要な役割を果たしているが、その使い方は複雑である。動詞は場合により名詞の文法的特徴や文法的機能を有するが、語尾変化がないため判定が難しい。文のパターンを用いて述語動詞を抽出する研究[1]はすでに行われているが、パターンに合致しない場合の動詞抽出は困難であった。結合価を用いることによって、この判定を行い、構文解析を行うことができる。これに関する研究はすでに行われているもののまだ少ない。本研究では中国語動詞をその動詞の持つ意味によって動作動詞、致使動詞、心理動詞等に分類し、その結合価を調べる。さらに動詞ごとに共起する成分の意味素性を定め、これらの情報をデータベース化して構文解析の機械処理に用いた結果を示す。

## 2 中国語文の構造特性[2]

中国語の文成分は主語、述語、目的語の三大主種類と定語(連体修飾語)、状語(連用修飾語)、補語の概ね6種類に分類される。各成分の間の係り受けを表現すると図1のようになる。大きな点線四角で囲まれた部分が文の

骨格を示し、矢印がその係り受け関係を示す。四角で囲まれた成分は主成分で、楕円形で囲まれたのは修飾成分である。

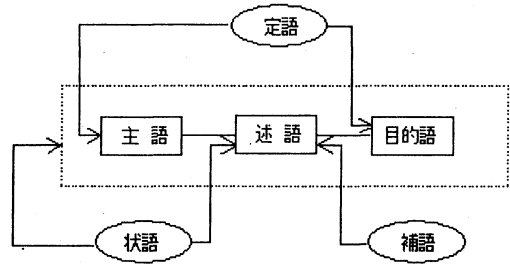


図1 中国語の文成分

中国語文を解析する際、中国語の文は主部と述部に分けられる。主語とその修飾語は文の主部であり、述語と目的語そしてそれらの修飾語は述部である。目的語は述語のみと関係があり、主部は述部と関係があるので、文のなかで述語が中心となる。従って、述語の判定は、文全体の解析にとって、非常に重要である。主に動詞が述語となるので、動詞の構文的属性についての研究は重要である。

## 3 動詞の分類と結合価の確定

動詞の構文情報はその意味と関連があるので、まず動詞をその意味によって動作動詞、致使動詞、心理動詞、性状動詞などに分類する。次にその動詞の結合価を決める。結合価文法[3,4]はフランスのテニエルが最初に使ったが、ドイツ語やロシア語の研究者を中心に、この考え方が受け継がれて発展した。中国語には、1978年朱徳熙が結合価文法の考えを最初に導入した。90年代に、結合価文法は発展し多くの論文が発表された[5,6]。

### 3.1 動詞の分類

まず動詞を動作動詞、致使動詞、心理動詞、性状動詞、関係動詞、交互动詞に分類する。そして数の一番多い動作動詞をさらに図2のように分類する。

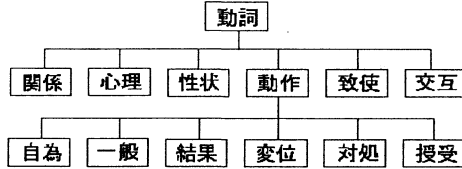


図2 動詞の分類

### 3.2 結合価の確定

分類された各動詞について結合価を定める。その成分は陈昌来[7]に従って、動詞の共演成分を価質、共演成分の数を価類、動詞の一番基本的な文の中での構文構造を価位、そして活用された構文構造を価用とする。動詞の価類は一価、二価、三価までにする。

#### 3.2.1 価質

価質は動詞の共演成分の性質を示し、表1のようなものがある。

表1 価質の種類

施事	動作を行う主体
受事	動作の対象
経事	心理活動の感知者、体験者、経験者
感事	心理活動の感知・体験対象
系事	ある性質と状態の係属者・所有者
涉事	性状動詞に関わる事物・事件
致事	行為発生の原因
使事	致事の影響により変化したもの
補事	三価致使動詞の致事の結果
起事	関係動詞の前の成分
止事	関係動詞の後の成分である
成事	動作によって生じたもの、現象
与事	施事の協同者
位事	事物の存在或いは移動時の位置

### 3.2.2 価位と価用

動詞の分類ごとにその構文パターンを決めたが、動作動詞の一般動詞の構文パターンを例として以下のように示す。

A : N p 1 + V + N p 2

B : (N p 1) + 把 / 将 + N p 2 + V p

C : N p 2 + 被 / 由 / 归 + N p 1 + V p

D : N p 2 + (N p 1) + V p

N p 1 : 施事 N p 2 : 受事

同じ分類であっても動詞ごとにその具体的な意味の違いによって構文形式も違うので、その動詞ごとに価位と価用を決める。

## 4 構文解析への応用

中国語の構文解析の難点としては、品詞分類の困難さと文の構造とフレーズの構造の同一性が挙げられる。特に、中国語は動詞が文の中で中心となっているが、役割もたくさんあるのでその判定が難しい。文の中で、動詞や名詞の役割を果たすことのできる動詞を動名詞と言う。動名詞として使える動詞は全部の動詞の中で、39%にも及ぶ[8]。表2は文献[8,9]による日本語と中国語の単語の数の比較である。中国語の動詞は日本語と比べると遥かに多いことが分かる。逆に名詞は日本語の方がずっと多い。

表2 日本語と中国語の辞書の単語数比較

	日本語	中国語
総単語数	58,409	51,696
名詞	50,928	27,828
動詞	4,168	10,787
形容詞	637	2,640
副詞	1,447	1,050

#### 4.1 意味素性

動詞は述語自体の意味の制限の他に、その共演成分に対しても制限を持っている。例えば、動詞「跑(走る)」は施事を持つが、この「跑(走る)」という動作のできる名詞句は「植物」などは不可能で、「動物」しかできない。このような、動詞の共演成分の名詞句に対する制限を「意味素性」と呼ぶ。

意味素性は大きく具体名詞、抽象名詞、空間名詞に分類されるが、詳細は表3に示す。

表3 意味素性

具体名詞	生物 動物 植物 生物の部分 人間 組織団体 自然物 建物 人造物 食品 服装
抽象名詞	事情 事理 外観 性質 情報 精神 言語作品 領域
空間名詞	場所 方角

#### 4.2 データベースの作成

動詞の構文的意味属性を実際の機械処理に反映するにはデータベース化が必要である。そこで、動詞の分類、結合価そして意味素性を反映した表4のような動詞データベースを作成した。主語、目的語1、目的語2、非目的は価質の文の中での対応する文成分である。

表4 動詞の構文的意味属性

見出し語	分類	価類	主語	目的1	目的2	非目的	価位	価用
前进	動作:自為	1	施:人				A	
喜欢	情緒	2	経:人	感			A	
定居	動作	2	施:人	位:場所			A	B
翻译	動作:一般	2	施:人、団	受:言語			A	B
要求	致使	3	致	使:人、団		補[補:動]	A	
告诉	動作:授受	3	施:人	与:人	受:情報		A	B

目的1：間接目的語 目的2：直接目的語 非目的：目的語のほかの成分

#### 4.3 構文解析への応用

近年統計手法による研究が盛んである。しかし、中国語は述語中心となる文であり、統計的手法は文全体の構造を無視するので、その結果が正しいとは言えない。本研究では文の構文解析の時、まず述語判定を行い、動詞中心とする文の構造による構文解析を行うことを提案する。

##### 4.3.1 述語の判定

例1のような一つの文の中に多数の動詞があり、述語となる動詞の判定に本研究のデータベースを用いる。

例1 机器翻译 研究 要求 建立 统一 标准.  
(機械翻訳研究は統一標準を作ることを必要とする)

科学院構文解析[10]の結果：

( IP ( NP ( n(机器) v(翻译) v(要求) ) VP  
( v(建立) NP ( vn(统一) n(标准) ) ) w.  
(. ) ) )

この結果を見ると述語になるはずの「要求」が、「机器」「翻译」と一緒に名詞句となると判定されている。つまり、この構文解析が誤った原因は文の中の中心になる述語動詞の判定の誤りにある。

「翻译、研究、要求、统一」は全部動詞としても名詞としても使える単語である。表4

を調べると「翻译」は2価の一般動作動詞で、共演成分の施事が「人、団」であることを必要とするが、「机器」は違うので述語動詞ではないと判定される。次に動詞「研究、建立、统一」も同じやりかたで述語でないと分かる。一方で、「要求」は三価の致使動詞で、前の共演成分となる「致事」には特別制限がない。しかし、後には必ず動詞成分の「補事」がくるのを制限するため、この形に合うのでその判定が可能である。

#### 4.3.2 「的」フレーズの構文曖昧性解消

例2 清扫 教室 的 工人 在找你。

v n u n

(教室を掃除する人が君を捜していた)

「清扫教室的工人」の抽象構文パターンは[v np u[的] n]でそれを解析する場合次のような2パターンがある。

a. [ [v np u[的] ] n ]

b. [v np u[的] n]

この場合「清扫」は二価自移動作動詞で、「工人」は施事になるべきで、位事になるのは「場所」の「寝室」になり、正しい構文パターンはaだと判定できる。

#### 4.3.3 単語分割への応用

北京大学の統計と規則による単語分割システム[11]による誤りの例を例3に示す。本研究のデータベースを利用して、述語中心とする単語分割を行うと正しい分割結果が出る。

例3 她本人生了三个孩子。(彼女自身は三人の子供を生んだ)

(誤) 她 本 人 生 了 三 个 孩子。

r r n u m q n

(正) 她 本人 生 了 三 个 孩子。

r r v u m q n

## 5 おわりに

中国語の動詞は1万個以上ある。今回データベース化したのは1,200個ぐらいに限られている。これらのデータを増やすことが必要である。そして動詞は、「来、去」等「趋向动词」が付いた時、その活用により元々とは違う役割を果たすので、それについても意味と結合価を利用してデータベース化できる。

一方では、名詞や形容詞などの他の品詞の情報も必要なので、動詞と同様な方法でデータベース化し、計算機で実際の処理ができるようにシステム化し、本研究の構文解析の有効性を検証する予定である。

## 参考文献

- [1] 吳志剛:動詞述語の抽出による中国語文の構文解析、山形大学修士論文(1998)
- [2] 刘月华、潘文娛、故桦:《現代中国語文法総覧》、くろしお出版(1996)
- [3] 水谷静夫:文法と意味I、朝倉書店(1983)
- [4] 横山晶一:言語処理技術——意味解析、機械翻訳の基礎と技術、bit(1988.9)pp.75-88
- [5] 沈阳、郑定欧:《現代汉语配价语法研究》、北京大学出版社(1995)
- [6] 袁毓林、郭锐:《現代汉语配价语法研究——第二辑》、北京大学出版社(1998)
- [7] 陈昌来:《現代汉语动词的句法语义属性研究》、学林出版社(2002)
- [8] 俞士汶:《現代汉语语法信息词典详解》清华大学出版社(1998)
- [9] 横山晶一:電子化辞書に基づく日本語の計算言語学的研究、電子技術総合研究所研究報告第881号(1988)
- [10] 確率構文解析器、中科院計算所:  
<http://mtgroup.ict.ac.cn/parserform.html>
- [11] 汉语文本切分与词性标柱系统、北京大学計算言語学研究所:  
<http://www.icl.pku.edu.cn/nlp-tools/segtagtest.html>