

動詞述語の抽出による中国語文の構造解析

呉 志剛 * 横山 晶一** 于 素秋 *

*山形大学大学院工学研究科

**山形大学工学部

〒992 山形県米沢市城南 4-3-16 山形大学工学部

E-mail:m96054@eieio.yz.yamagata-u.ac.jp

Tel : 0238-26-3340

中国語の動詞は、その使用法が極めて複雑である。また、動詞は語尾変化や活用をせずに名詞として用いたり副詞的な使い方をすることができる。一方、中国語の文の中で最も重要な役割を果たすのは述語であり、しかも動詞が述語として用いられることが多い。従って、中国語の文を解析する際、動詞の役割の判断は非常に重要であるし、また、困難でもある。本研究では、動詞の接続特徴に基づいて、動詞を5つの型に分類し、それらのうちに優先度(これらを固有優先度と呼ぶ)を導入することによって、文における動詞の役割を機械的に判定し、述語を認定する方法について述べる。また、これらの動詞に修飾語やアスペクトが付加した場合の優先度の上下についても論じる(これらを動詞の総合優先度と呼ぶ)。これらの優先度の導入によって、従来困難であった動詞述語の認定が容易にできるようになった。

キーワード：中国語，構文解析，動詞，述語認定，固有優先度，総合優先度

Chinese-Language Analysis Based on Discrimination of Verb Predicate

Zhigang WU * Shoichi YOKOYAMA ** Suqiu YU *

*Graduate course of Engineering, Yamagata University

**Faculty of Engineering, Yamagata University

4-3-16, Jonan, Yonezawa, Yamagata 992, JAPAN

Email:m96054@eieio.yz.yamagata-u.ac.jp

Tel : 0238-26-3340

The use of Chinese verb is complicated. In Chinese text, verbs can be used as nouns or adverbs without conjugation. In Chinese, verb predicates play very important role, and are mostly used. Therefore, the discrimination of role of verbs is very important and difficult. In this paper, the concept of 'specific priority of verb' and 'comprehensive priority of verb' are introduced for discriminating role of verbs. The introduction of these priorities makes the recognition of verb predicates easier.

Keyword: Chinese, Parsing, Verbs, Discrimination of verbs, Specific priority of verbs, Comprehensive priority of verbs

1 はじめに

文を自動的に解析するためには、構文解析が欠かせない過程の一つである。中国語文の場合は、構文解析する際、構文情報だけでは、うまくできないケースが多い。そのために、単語に含まれた構文解析に役に立つ情報を引き出す必要がある。

本研究では、中国語文の構文の曖昧さによる解析の問題を対象として、動詞を中心とする解析法を提案する。中国語では、述語が文の中で最も重要な役割を果たす。その中で動詞は述語になる場合が多い。中国語の動詞は、その使用法が極めて複雑である。また動詞はそのままの形で、名詞として用いることができる。本研究では、動詞述語文を対象として、文の中に隣接した単語の品詞の変化を考慮することによって、文における動詞の役割を推定する方法について検討する。具体的な方法としては、文における動詞の接続特徴によって、動詞を5つの型に分類し、それらのうちに優先度(これを固有優先度と呼ぶ)を導入することによって、文における動詞の役割を機械的に判定し、述語を認定する方法について論じる。また、これらの動詞に修飾語やアスペクトが付加した場合の優先度の上下についても論じる。これらを動詞の総合優先度と呼ぶ。これらの優先度の導入によって、従来困難であった動詞述語の認定が容易にできるようになった。

2 中国語の解析特徴

中国語では、単語と単語間の文法関係は主に単語の順序と虚詞により表され、単語と単語間の具体的な意味関係はそれぞれの意味素性の組み合わせから定められる。文の構成成分はすべて単語またはフレーズからなる。そのために、解析は文の意味に強く依存する。機械による自動解析の際には、文の意味を明らかにする前に、文の構造を解明しなければならない。その問題を解決するために、単語そのものに含まれた構文に影響のある“意味”要素を抽出する必要がある。

3 中国語文の要素

中国語の書き言葉は、日本語と同じように、文に述語、述語の主体、述語の対象があり、各成分に修飾語が付く可能性がある。各成分に関する単語の集合は、一定の順序に並ぶ。中国語の語順は次の通りである。

(文の修飾語)動作主体+動作+動作の対象語

動詞主体:主体修飾語+主体中心語

動作:述語前置修飾語+述語+述語後置修飾語

動作の対象語:対象語修飾語+対象語中心語

3.1 中国語の文成分

中国語の文において、述語はとても重要な存在である、単語(又は単語グループ)の役割は文の中の位置によって決まるが、この“位置”は、述語に対する相対位置である。中国語文を解析する際に述語が明確になれば、文の他の成分も明らかになる。

本論文で対象とする中国語の文成分は述語(本論文では動詞述語だけについて論じる)を中心に、名詞句(名詞修飾成分と名詞)、文の修飾成分、述語の修飾成分である。

3.1.1 名詞句

名詞句とは最後の単語が名詞である単語の集合である。最後にある名詞以外の単語はこの名詞の修飾語に当たる。述語との相対位置によって、動作の主体や、動作の対象になる。名詞句の構造は以下のようなになる。なおアルファベットの太文字は中国語電子辞書(参考書[5])の分類記号(付録参照)を示す。

構造特徴: 名詞修飾成分+名詞

名詞句の構文:

名詞句

名詞修飾語

中心語

3.1.2 文の修飾語

文の全体を限定したり、修飾したりする部分である。

例 3.1: 昨天, 他们去滑雪了。(昨日, 彼らがスキーに行ってきた)

例文中では、“昨天”と言う成分が、文中の事実の発生した時間を表している。そのような動作の様子ではなくて、ある“事実”の場所や時間などを表す部分は“文の修飾語”と言う。

構造特徴:

介詞+時間や場所などを表す名詞句

そのうちに“介詞”と言う部分は省略できる。

3.1.3 述語の修飾語

述語に加えて、動作の具体的な様態や結果や

アスペクトなどを表す部分を指す。このような修飾語はほとんどが特定の形をしているために、比較的分けやすい。また、述語との位置関係が決まっているので、述語を推定するための有力な手がかりにもなる。

(1)前置述語修飾語の構文

前置述語修飾語

副詞(CADMS, CADMV)

時間名詞

場所名詞

介詞フレーズ

地フレーズ(CAD4+地)

否定語(不, 不能, 没, 没有)

(2)後置修飾語の構文

後置修飾語

不フレーズ

得フレーズ

介詞フレーズ

了フレーズ

過フレーズ

个フレーズ

着, 極, 死, 透, 壞

4 動詞述語の抽出

一つの文に“動詞”である単語が一つしかないとは限らないし、たとえ“動詞”が一つだけであっても、この動詞は必ず述語になるとは言えない。動詞が述語のほかに、別の文成分として現れることもできる。自動解析する際に、他の役割の動詞から述語の動詞を見分けなければならない。広い意味で述語の抽出は、文にあるすべての動詞の役割を判定するということを意味する。

4.1 文における動詞の役割

中国語の単文における動詞の役割は次の通りである：

動詞：

非述語：名詞修飾語，述語修飾語，名詞化

述語：主文の述語，対象語としての文の述語

動詞役割を判定するのに、次の要素が利用できる。

(1)動詞自身の性質

(2)文脈による動詞への影響

4.2 動詞の分類

動詞の使い方はその動詞の性質に決められたものである。ここで、動詞の意味によって、次のように動詞を分類する。これらの分類に基づいて、後に動詞の優先度を設定する。(4.5 参照)

(1)人間の自分自身の意志を表す動詞。(ここでは、第一類の動詞と呼ぶ。他の動詞と連続できる。その場合には、後ろの動詞が表示する動作を開始すると言うことを意味する)。

(2)人間の客観世界に対する客観的な反映(第二类動詞。他の動詞と連続できる。後ろの動詞で記述される現象を表現する)

(3)実在物に加える動作(他動詞)

(4)自然現象と普通の自動詞。

(5)道具動詞:何かをするために、方法、手段、道具などを表す動詞

4.3 文脈の動詞への影響

中国語の表現特徴の一つは、単語自身が変形しないということである。文において、述語動詞と非述語動詞との区別をつけるのは動詞自身ではなく、動詞の前後にあるほかの単語である。すなわち、動詞は文にある他の単語または字から影響を受けて、非述語成分になる。このような状況には二つの場合がある。

(1)非動詞からの影響

例 4.1：我买的书丢了。

(私が買った本はなくなった)

例文の“买”と言う単語はもともと動詞である、しかし、ここでは“的”と連続して、“书”と言う名詞の修飾語に当たる。すなわち“的”のような非動詞が動詞の役割を決めることができる。

(2)動詞からの影響

例 4.2：他用筷子吃饭

(彼が箸でご飯を食べる)

例文には、“用”と“吃”と二つの動詞があるが、“吃”は述語となって、“用”は述語“吃”の様態を表わす修飾語となる。ここでは、“用”が“吃”と言う動詞に影響され、修飾語となるわけである。もし後ろの動詞“吃”によってもたらす構造“吃饭”がなければ、例 4.2 は次のようになる、

例 4.3：他用筷子

(彼が箸を使う)。

例 4.3 では、動詞“用”に影響する成分はないので、“用”が述語になる。

4.4 文型に対する動詞の判断条件

前述のように文脈によって、動詞の役割が変わる。従って、それぞれの文に対応する動詞の判断条件も異なる。

4.4.1 動詞を中心とした文の分類

本研究で提案する解析法は動詞に基づいた方法なので、文の中の動詞数が重要である。すべての動詞を述語候補と見なす。

文は動詞の数によって、単動詞文と多動詞文に分けられる。多動詞文の場合は、動詞の接続形態により、連続型と断続型に分かれる。

連続型：文中の動詞が連続的に現れる。動詞の間に動詞以外の単語は入っていない。

断続型：文にある動詞の間に他の品詞の単語がある。一般的に、動詞の役割がそれらの動詞の間にある単語によって決められる。

単動詞文において、判断の重点は文の中の動詞は述語であるか、それとも名詞の修飾語であるか、ということである。それに対して、多動詞文の場合は、判断の重点は、幾つかの動詞または動詞構造のうちから述語を選ぶことである。

4.4.2 単動詞文の動詞役割判定

単動詞文は述語推定の立場から見ると、他の動詞に影響されないので、最も単純な文である。このような文では、動詞の役割を非述語と述語とに分けることが出来る。

(1) 非述語の条件

単動詞文において、動詞が非述語となる条件は次の通りである。

動詞名詞化の条件(下線の動詞が名詞化)

CNN+的+CVN

CPR1+的+CVN

例：他的建议很好

(彼のアドバイスは素晴らしい)

CVT+的+NULL (NULL 文末)

CVT+的+是

名詞修飾語になる条件(下線を引いた動詞が名詞または代名詞の修飾語になる)。

CVT+的+CNN(名詞)

CVT+的+CPR(代名詞)

(2) 述語の条件

中国語文において、動詞以外の単語が述語になるケースもある。このようなケースは本論文

の範囲を超えるために、ここでは論じない。

動詞述語の前にあるのは、名詞句と前置述語修飾語であるが、後ろにあるのは、名詞句と後置述語修飾語である。そのため、名詞句、前置述語修飾語と後置述語修飾語、その三つの文成分の構造から、動詞述語の接続特性が分かるはずである。文において、その三つの成分が動詞述語と接続するパターンは次の通りである。

名詞句+述語

前置修飾語+述語

述語+名詞句

述語+後置修飾語

実際に、名詞句の先頭と終わりにある単語、前置修飾語の終わりにある単語、後置修飾語の先頭にある単語と述語との接続関係が重要である。すなわち、それらの接続形態を満たせば、動詞は述語であると言うことが判断できる。

述語の接続特性を前置条件と後置条件に分けると次のようになる。

(a) 前置条件

副詞+動詞

CADMS+動詞

CADMV+動詞

地+動詞

「的」構造+動詞

的+動詞(名詞化不能の)

不能+動詞

不+動詞

得+動詞

(b) 後置条件

動詞の後にくるのは目的語と補語である。補語は必ず動詞の直後に現れるので、動詞を同定する手がかりとなる。

CVC0の動詞の後ろ

CVC0+CP1(+了)

CVC0+得+CP2

.....

動詞+于+名詞(フレーズ)

動詞+向+名詞(フレーズ)

動詞+自+名詞(フレーズ)

4.4.3 多動詞文の動詞役割の判定

多動詞文は動詞の数が一つ以上ある文のことである。複数の動詞が存在するために、それらの動詞間の相互影響を考えなければならない。

(1) 非述語の条件

単動詞文の動詞名詞化の条件、名詞修飾語の条件はそのまま多動詞文に適用できる。

(2) 述語の条件

述語の判断は、文型によって、やり方が異なる点があるので、以下では、文型別の判断条件を論じる。

4.5 連続型多動詞文における動詞の優先度

連続型の接続は次のようになる。

...+動詞+動詞+

例 4.4:

机器 翻译 研究 要求 建立 统一 标准。

(機械翻訳のための標準を作る必要がある。)

文の中に下線を引いた単語はすべて動詞になる可能性がある。このような文では、品詞が動詞である単語が連続して現れるので、述語の特徴を表わす接続がまったくないと言うことになる。ここで、このような文に対して、述語を判断するために、“動詞優先度”と言う概念を導入する。

優先度とは、動詞が述語になる可能性を評価する基準である。具体的に言えば、一つの文に幾つかの動詞があるとする、優先度の一番高い動詞が述語となる。たとえば、

例 4.5: 我决定考大学

(私は大学の入試を受けることにした)
と言う文には“決定”と“考”と二つの動詞がある、どちらも、述語になる構造を持っていないので、構造による判断が不可能である。日本語の場合は、文において動詞述語はほとんど“を、は、が、へ”の後にある、言い換えれば、“を、が、は、へ”は動詞述語の印だとも言える。それに対して、中国語の場合は、“を、は、が、へ”などと同じ役割を果たすものは全部単語に含まれて、文を構成する際に現れないのである。“優先度”はこれらの文字に隠れた性質の一つである。

4.5.1 動詞優先度の判定基準

本研究において、動詞優先度を判定するための基準として、動詞の対象成分の複雑さを使う。動詞対象の複雑さはその動詞によって表現できる客観世界の幅に比例する。たとえば、“説”(言う)と“製造”(作る)との対象語を比較する。“説”の対象(内容)はとても広く、“説”の対象語にすべての文成分が現れる可能性がある。それに対して、“製造”(作る)の対象語は“物”だけで、対応する文成分は、名詞句だけである。そのために、“説”の優先度が“製造”より高い。

4.5.2 各々の動詞の固有優先度

前述のような5種類の動詞のうち、(1)類と(2)類は人間の意識に関わる動詞である、これらは、表現内容の範囲が広く、動詞の“優先度”が一番高い。たとえば、“決定”(決める)は(1)類であり、“説”(言う)は(2)類動詞である。(3)類動詞は実在物を対象して行う動作である。前の二種の動詞と比べて、対象語の範囲は少し狭いので、優先度も低い。(4)類の動詞は、動詞自身である現象を記述できて、対象語は要らないので、優先度は三番目である。(5)類の道具動詞は、ある行為を行うための手段や、使った道具などを表す動詞である。この動詞だけで構成された文は文法上では問題ないが、現われた意味は不完全である。文脈を参照すれば、このような動詞が単動詞文の場合、述語として使える。

例 4.6A: 他用钢笔

(彼がペンを使って)

例 4.6B: 他用钢笔写信

(彼がペンを使って手紙を書く)

例文 4.6A の動詞“用”は典型的な道具動詞である、この例文は、読み手に“ペンで何かするか”と思わせる傾向がある。すなわち、不完全感をもたらす。日本語の“ペンを使って”や“ペンで...”などのような構造に相当する。多動詞文の場合は、不完全感をもたらす道具動詞が述語になるのは不可能である。後ろにある動詞で何のために“道具”を使うかを説明するからである。そのような文の場合は、多分道具動詞の後ろにある動詞が述語になる。例 4.6B の場合は、“写”と言う(3)類動詞があり、この動詞が文の述語になる。

動詞自身の性質により決められた“優先度”の順位は次の通りである:

優先度 1: (1)類動詞と(2)類動詞

優先度 2: (3)類動詞

優先度 3: (4)類動詞

優先度 4: (5)類動詞

このような優先度をここでは“固有優先度”と呼ぶ。

連続型の文においては、動詞の固有優先度を利用して、述語を判断する。

4.6 断続型多動詞文における述語の判断

断続型文の接続特徴は次の通りである

...+動詞+...+動詞+...

例 4.7: 下棋, 在他的生活中占了很长时间。

(碁を打つのは、彼の生活において、多くの時間を占めている)

このような文では、動詞がいくつかあるが、接続しないという点が特徴である。動詞と接続する単語の中に、動詞の役割を示す成分も含まれている。動詞が述語となるかどうかを判断するとき、これらの成分は、重要な要素となる。一方、これらの成分は、動詞の“固有優先度”にも影響する。文の述語を推定する際に、動詞自身の優先順位のほかに、これらの影響を考えなければならない。

4.6.1 動詞の総合優先度

接続型によって動詞の固有優先度があがったり、下がったりすることがある。詳細はつぎの通りである

(a)優先度を下げる接続

(1)名詞化の接続

動詞+着

例 4.8 A:他抽着烟看电视

(彼はたばこを吸いながら、テレビを見る)

例 4.8 B:他抽烟看电视

(彼はたばこを吸ったり、テレビを見たりする)

例 A の場合は“動詞+着”と言う形で、テレビを見るときの様子を表わして、“抽着烟”と言う構造は修飾語になる、それに対して、例 B の場合は、“抽烟”は“看电视”と同じように、述語である。構造“動詞+着”の優先度は動詞の固有優先度より一段下がる。

(2)名詞修飾語の接続

動詞+的

動詞述語文+的

(b)優先度をあげる接続

動詞+了

動詞+得

動詞+過

(c)動詞の総合優先度

動詞の接続の影響を考慮すると、動詞の総合優先度は次の通りである(数字の若い方が優先度が高い)：

(1)総合優先度 0：

是：“是”は特殊な動詞である、

(2)総合優先度 1：

固有優先度 1 の動詞+得

固有優先度 1 の動詞+了

固有優先度 1 の動詞+過

(3)総合優先度 2

固有優先度 1 の動詞

固有優先度 2 動詞+得

固有優先度 2 動詞+了

固有優先度 2 動詞+過

(4)総合優先度 3

優先度 1 の動詞+着

固有優先度 2 の動詞

固有優先度 3 の動詞+得

固有優先度 3 の動詞+了

固有優先度 3 の動詞+過

(5)総合優先度 4

固有優先度 2 の動詞+着

固有優先度 3 の動詞

固有優先度 4 の動詞+得、了、過

(6)総合優先度 5

固有優先度 4 の動詞+着

優先度 5 の動詞(道具動詞)+得、了、過

(7)総合優先度 6

固有優先度 5 動詞(道具動詞)

(8)総合優先度 7

動詞(優先度と関わらず)+的

動詞述語文+的

多動詞文においては、総合優先度の高い動詞が述語になる。

4.7 文型と判断条件の対応関係

前述の文の類型と判断条件の対応関係は次のようになる。

文型	単動詞文	多動詞文	
		連続型	断続型
判断条件	単動詞文 条件	動詞の固 有優先度	動詞の総 合優先度

5 動詞述語を推定するプロセス

このプロセスを行うにあたっては、単語の分割は済んでいると仮定する。辞書は、本研究で提案した動詞の新しい分類情報を含んでいる。まず、辞書引して、各単語の品詞情報や、接続特性情報を取得する。次に、品詞情報に基づいて、文が単動詞文か、多動詞文なのかを判断する。単動詞文に対しては、単動詞文の判定条件を利用し、動詞の役割を判断する。多動詞文の場合は、動詞の数や、接続状態などによって、さらに、具体的な分類を推定し、対応する条件で、判定する。その際に、動詞の固有優先度と総合優先度を用いる。全体の解析手順は図 1 示すようになる。

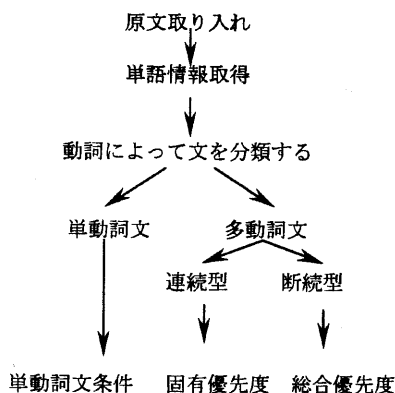


図1 解析手順

6 結び

本研究では、文における動詞の役割を検討し、動詞の新しい分類情報によって、動詞述語を判断する方法を提案した。これによって文構造を解析する時、動詞の品詞変化による曖昧さの一部が解決され、多動詞文の述語抽出が効率的にできるようになった。

本研究では、動詞述語の文だけを対象として、他の品詞の単語が述語となる文の特徴は、論じなかった。実際に、動詞と他の品詞の単語との間に“優先度”と言う問題も存在する。この点について、一層の研究が必要である。

ここで提案した方法では、動詞述語、名詞化、名詞修飾語などの接続特性を明確にしたが、動詞と介詞の同形と言う問題については触れていない。

例 6.1: 他打天津来 (彼は天津から来た)

例 6.2: 他打我来了 (彼は私を殴りに来た)

例 6.1 の“打”は、介詞で、“天津から”という意味を表わす、例 6.2 の“打”は、動詞で、“殴る”に相当する。二つの文の構造は、まったく同じであるので、解析しづらい。このような文については、今後研究の必要がある。

分析を完全に行うためには、全ての文の成分を推定しなければならない。ここでは、動詞の役割を推定するためのプロセスだけを提出したが、他の文成分の判断は、問題として残されている。

参考文献

[1] 劉月華, 潘文娛: 《現代中国語文法総覧》<

ろしお出版(1996)

[2] Dorr: 《Machine Translation》The MIT Press (1993)

[3] 張伯江, 方梅: 《漢語功能語法研究》江西教育出版社(1996)

[4] 長尾 真: 《自然言語処理》岩波書店(1996)

[5] 《中国語電子辞書》財団法人 国際情報化協力センター(1995)

符号の説明

ここで、引用した単語の属性符号は CICC(Center of the International Cooperation for Computerization)によって編集された《中国語電子辞書》からのものである。

CVJCVN 名詞として使える動詞

CVJ 程度副詞に修飾される動詞

CP1 動詞または形容詞について、補語として使える

CP2 「得」のうしろについて補語として使う動詞または形容詞

CVNV1 兼ね語動詞

CVEX1 「着」と連用できない動詞

CVCO 後ろに補語をつけられる動詞

CVT 目的語をつけられる動詞

CVT1 目的語は動作の対象である動詞

CVT2 目的語は動作の対象ではない

CVI 目的語なし

CNN 名詞

CADMS 文を修飾する副詞

CADMV 動詞を修飾する副詞

CPR 代名詞