

# 中国語コントロール構文の解析

周振<sup>‡</sup> Alastair Butler\*<sup>†</sup> 吉本啓<sup>‡</sup>

\*科学技術振興機構 さきかけ

†東北大学高等教育開発推進センター

‡東北大学大学院国際文化研究科

syusin3@yahoo.co.jp

## 要旨

中国語は機能語や屈折形態素などに乏しい言語であるため、シンタクスの曖昧性を排除することが容易ではない。その一例としては、中国語においてコントロール構文および非コントロール構文についての判断が困難であることが挙げられる。本研究は、中国語のテキストに対して統語解析を行うために、先行研究を踏まえて、中国語のコントロール構文および非コントロール構文の区別方法を考える。また、意味処理の要請から従来の中国語のコントロール構文に対する解析方法について再考察し、その不足を検討しつつ本研究の解決策を提示する。これによって、従来の解析方法では捉えられないコントロール構文における統語・意味情報が得られるようになり、中国語コーパス研究に貢献できると期待している。

## 1 はじめに

中国語の無制約のテキストに対して、論理意味表示（述語論理式）を付加した中国語の意味表示コーパスを構築している。その作業は二つの段階に分けられている。すなわち、(1)分析データとして選ばれた中国語の自然テキストに対する統語解析情報の追加および(2)それをバトラー (Butler 2010) が提唱するスコープ制御理論 (Scope Control Theory; SCT) を実装したシステムで処理することによる自動的な文の論理意味表示の獲得である。

本発表は、中国語におけるコントロール構文と非コントロール構文の判断方法を検討しその基準を決めた上で、統語情報アノテーションを行う際にゼロ代名詞に対するタギングの基本方針を紹介し、従来の解析方法と比べながら本研究における中国語のコントロール構文に関する解析方法および処理結果を示していきたい。

本論文の構成は以下の通りである。第1章では、研究の概要および目的を述べる。第2章で

は、中国語のコントロール構文および非コントロール構文の区別に関する先行研究を紹介しつつ本研究の判断方法と基準を決定する。第3章では、意味処理の要請から中国語のコントロール構文に対する新しい解析方法を具体例で提示する。第4章では、第3章で提唱した解析方法を用いて、コントロール構文が再帰的に埋め込んでいるような複雑な構文を対象にして意味処理を行う。第5章でまとめを行う。

## 2 コントロール構文および非コントロール構文の区分

中国語のツリーバンクを開発する時一つの大きな課題は、ゼロ代名詞が含まれる補文がコントロール構文か非コントロール構文かという判断である。(1a),(1b)に示すように、英語のコントロール補文は必ず不定形節で、しかもその不定形節の主語は、主文の主語または目的語と同一指示を有し顕在的主語と交換することができずPRO（代名詞類の性質と照応形の性質を合わせ持つもの）として分析される。一方、(1c)では、定形節の主語は、主文の要素によってコントロールされず主文以外の人物を指すことも可能なのでpro（純粋な代名詞類の性質を持つもの）として扱うべきだと考えられる。

- (1) a. Marry intended to go to Tokyo.  
b. Marry forced John to go to Tokyo.  
c. Marry said that she will go to Tokyo.

(2)は(1)に対応する中国語の文である。英語の場合、不定形のマーカーのtoおよび補文標識のthatを手掛かりとして、補文がコントロール構文であるか非コントロール構文であるかということが簡単に区別できるが、周知のように中国語は機能語や屈折形態素などに乏しい言語であるため、そういう判断が容易につけられない。また、(2c)に示すように、中国語の非コントロール補文は英語とは異なり、補文の主語である「她」も省略できる

ので、その表層構造がコントロール構文と同じようになってしまう。よって、中国語の場合、文の補文を解析する際に、proとPROのどちらかが適用されるべきかということが決められるように、補文の性質（コントロール構文か非コントロール構文か）を明確にしておく必要がある。

- (2) a. 玛丽 打算 去 东京。  
メアリーつもり行く東京  
メアリーは東京に行くつもりだ。
- b. 玛丽 逼 约翰 去 东京。  
メアリー強制するジョン行く東京  
メアリーはジョンに東京に行くように強制した。
- c. 玛丽 说 (她) 去 东京。  
メアリー言う(彼女)行く東京  
メアリーは(彼女が)東京にいくと言った。

中国語のコントロール構文と非コントロール構文を区別するために、従来多くの研究(Huang 1989, Li 1990, 湯 2000)がなされてきた。Huang は、中国語では、補文に助動詞的(auxiliary)要素(アスペクトやモダリティーなど)を許すか否かによって両者の区別ができるとした。それによると、コントロール構文は不定形節で助動詞的要素を受け入れないが、非コントロール構文は定形節で助動詞的要素が出現できる。この対立は、次の(3)と(4)から観察できる。従って、(3a)はコントロール構文であるのに対して、(4a),(4b)は非コントロール構文であるということが分かる。

- (3) a. 我 逼 李四 来 。  
私 強制する李四 来る  
私は李四に来るように強制した。
- b.\* 我 逼 李四 会/能 来  
私 強制する李四 だろう/できる来る  
。
- c.\* 我 逼 李四 来 着 。  
私 強制する李四 来るASP
- (4) a. 张三 说 (他) 来 了 。  
张三 言う(彼) 来るASP  
張三は(彼)が来たと言った。
- b. 张三 相信 (他) 会 来 。  
张三 信じる(彼) だろう来る  
張三は(彼が)来ると信じている。

しかし、黄(1992)、徐(1994, 1999)、Hu et al (2001)などが指摘しているように、実際にはそういう対立が存在しないこともある。例えば、(5a)と(5b)はコントロール構文であるにも関わらず、非コントロール構文と同様に、補文にモダリティーやアスペクトなどの要素が来られる。

- (5) a. 我 准备 明天 要 来 。  
私 つもりだ明日 MOD 来る  
私は明日来るつもりだ。
- b. 妈妈 逼 小明 吃 过 药。  
母 強制する名前 食べるASP 薬  
お母さんは小明に強制して薬を飲ませた。

以上に見られたように、補文に何らかの要素を許すか否かによってコントロール構文か非コントロール構文かという判断をすることは不十分である。そのため、中国語の場合、コントロール構文になれるかどうかの決め手はむしろ主文の動詞の意味に掛っていると考えられる。そこで、本研究では、まずPenn Chinese TreebankにおけるPRO/proの区別にもとづいて、中国語の動詞(Penn Chinese Treebankの中に現れたもの)を以下の三種類に分けた。

- 1) コントロール構文しか作れない動詞：逼(強制する)、愿意(願う)、喜欢(好む)、拒绝(断る)、感到(感じる)、坚持(堅持する)、企图(企む)、涉嫌(疑われる)、爱(愛する)、足以(十分に足りる)、恢复(回復する)、尝试(試みる)、获准(許可を得る)、など
- 2) 非コントロール構文しか作れない動詞：相信(信じる)、指出(指摘する)、听说(聞く)、导致(導く)、研判(判断する)、发现(発見する)、下令(命令を下す)、看到(見かける)、提到(言及する)、了解(了解する)、裁定(裁定する)、想象(想像する)、证明(証明する)、など
- 3) 両方とも作れる動詞：表示(示す)、赞成(賛成する)、希望(希望する)、知道(分かる)、确定(確定する)、禁止(禁止する)、答应(承知する)、考虑(考える)、要求(要求する)、觉得(思う)、记得(覚える)、证实(実証する)、避免(避ける)、など

その結果、1)に属する動詞の数は4720個であるのに対して、2)に属する動詞の数は1)の半

分に近い2408個になっている。また、コントロール構文でも非コントロール構文でも作成できる動詞の数は196個であった。Penn Chinese Treebank における区別に対しては言語学的な検証がさらに必要だが、少なくとも大まかな傾向は示していると考えることが出来る。これによると、中国語の場合、表層構造が同じ構文（コントロール構文および非コントロール構文）は、コントロール構文になる可能性が非コントロール構文になる可能性より高いことと、両方とも可能な動詞は確かに存在しているが、その数がそれほど多くはないということ（アノテーションに大きな負担をかけたりはしない）が分かる。

実際に生の中国語テキストをアノテーションする際に、以上の動詞リストをデフォルト値とした修正プログラムを作って、統語解析機によって得られた自動解析結果に対して自動修正をしてから、人手によるチェックと訂正を行う。明らかに言語データの実情に違反する時や動詞リストにない動詞が現れる場合は、リストに対する修正・補充も行う。このように、アノテーションの量が増えれば増えるほど、この動詞リストが完璧になり、最終的にはこれを手掛かりとして中国語のコントロール構文および非コントロール構文を区別するもっとも肝心のポイントに接近できると期待される。

### 3 コントロール構文の解析

一般的には、依存関係の表示および述語 - 項関係の再構成に必要なため、主語または目的語が動詞の必要格として求められるにもかかわらず文中で表現されていない場合、ゼロ代名詞の追加を行ってそれらを明示する必要がある。

中国語の文の補文に主語が現れていない場合は、まず当補文がコントロール補文か非コントロール補文かということを決めてから、pro或いはPROの解析方式を適用させる。非コントロール補文の場合はproを追加すればよいが、問題はコントロール構文である。

従来コントロール構文に対するゼロ代名詞のタギングは、直接PROを付け加えることが主流だった。目的語コントロール文の(6a)と主語コントロール文の(6b)に示すように、Penn Chinese Treebank 式の解析スキーム (Xue et al 2000) では、コントロール補文であるIPに対して主語PROのアノテーションを行った。

- (6a) (IP (NP-SBJ (NPR 张三/ちょうさん))  
(NP-TMP (N 今晚/こんばん))  
(VB 请/しょうたいする)  
(NP-OB1 (PRO 我们/わたしたち))  
(IP (NP-SBJ \*PRO\*)  
(VB 吃饭/しょくじする))  
(PU 。))

张三今晚请我们吃饭。  
張三は今晚私たちを御馳走する。

- (6b) (IP (NP-SBJ (NPR 李四/りよん))  
(VB 帮/てつだう)  
(NP-OB1 (PRO 我们/わたしたち))  
(IP (NP-SBJ \*PRO\*)  
(VB 叫/よぶ)  
(NP-OB1 (N 医生/いしゃ)))  
(PU 。))

李四帮我们叫医生。  
李四は私たちのためにお医者さんと呼ぶ。

このようなコントロール構文の解析仕方はある意味で論理的だが、精度の高い述語論理式をまとめるために、PROの値を確定しようとする際に問題が出てしまう。即ちPROをコントロールしている対象（主文の主語か目的語か）の同定ができないのである。

このように、意味処理の要請から考えるとPROがいったい何を指しているかということを確認しなければならない。そのために、本研究では、PROの使用の代わりに、補文の句 (IP, CP) に機能タグを付与する。

- (7a) (IP-MAT (NP-SBJ (NPR 张三/ちょうさん))  
(NP-TMP (N 今晚/こんばん))  
(VB 请/しょうたいする)  
(NP-OB1 (PRO 我们/わたしたち))  
(IP-INF (VB 吃饭/しょくじ))  
(PU 。))

- (7b) (IP-MAT (NP-SBJ (NPR 李四/りよん))  
(VB 帮/てつだう)  
(NP-OB1 (PRO 我们/わたしたち))  
(IP-PPL-OB1 (VB 叫/よぶ)  
(NP-OB1 (N 医生/い  
しゃ)))  
(PU 。))

(7a),(7b)では、IPにそれぞれINF、PPLという機能タグを与えた。これで、二つのIPの区別が可能になると同時にPROをコントロールしている対象の確定もできるようになった。SCTを実装した意味処理システムで二つの機能タグに関する解釈規則をインプリメントすることにより、意味処理により(7a)における主文目的語と補文主語、および(7b)の主文主語と補文主語とが同一指示であるとして関係づけられるので、(8a),(8b)が示すとおり、より精度の高い述語論理式が得られる。

- (8a)  $\exists x_4 t_1 e_2 e_3$  (  
 $x_4 =$  我们  $\wedge$   
今晚( $t_1$ )  $\wedge$   
请( $e_3$ , 张三,  $x_4$ , 吃饭( $e_2$ ,  $x_4$ ))  $\wedge$   
 $tmp(e_3) = t_1$ )

- (8b)  $\exists x_4 x_1 e_2 e_3$  (  
 $x_4 =$  我们  $\wedge$   
医生( $x_1$ )  $\wedge$   
帮( $e_3$ , 李四,  $x_4$ , 叫( $e_2$ , 李四,  $x_1$ )))

## 4 複雑なコントロール構文

本研究で提唱した中国語のコントロール構文の解析方法は、従来のやり方に比べると主に二つのメリットがあると考えられる。それは、PROをコントロールする対象の確定ができることと、補文の機能タグにデフォルトの解釈を与えることによってPROの使用が避けられ、文全体の統語構造が簡潔になることである。本研究の解析方法を活用すれば、(9)のようなコントロール構文が再帰的に行う複雑なコントロール構文に対しても精密な解析ができるようになる。その統語・意味の処理結果は以下の(9),(10)に示す。

- (9) (IP-MAT (NP-SBJ (NPR 张三/ちょうさん))  
(CP-ADV (IP-SUB (VB 动用/しようする)  
(NP-OB1 (N 关系/コネ))))  
(VB 让/させる)  
(NP-OB1 (NPR 李四/りよん))  
(IP-INF (VB 帮/てつだう)  
(NP-OB1 (NPR 王五/おおご))  
(IP-PPL-OB1 (VB 归还/かえす)  
(NP-OB1 (N 贷款/しゃっきん))))  
(PU 。))

张三动用关系让李四帮王五归还贷款。  
張三はコネを使って李四を王五に借金を返すようにさせた。

- (10)  $\exists x_1 x_2 e_3 e_4 e_5 e_6$  (  
关系( $x_1$ )  $\wedge$   
贷款( $x_2$ )  $\wedge$   
动用( $e_3$ , 张三,  $x_1$ )  $\wedge$   
让( $e_6$ , 张三, 李四, 帮( $e_5$ , 李四,  
王五, 归还( $e_4$ , 李四,  $x_2$ ))))

## 5 まとめ

中国語ツリーバンクを構築する際に、コントロール構文の解析について論じてきた。中国語の文はいわば漢字の並びだけで、形態素的な手掛かりが欠けているためコントロール構文と非コントロール構文の区別が困難である。

本研究では、中国語のコントロール構文を解析するために先行研究を踏まえてその判断の方法を提示し、アノテーションの作業の量を減らすために動詞ごとに解析のデフォルト値も与えた。本研究の解析方法は、従来のものと比べて、利点を持っている。これによって、従来の解析方法では捉えられないコントロール構文における統語・意味情報が得られ

るようになり、中国語コーパス研究に貢献できると期待している。

## 参考文献

- Butler, A. (2010) *The Semantics of Grammatical Dependencies*. Emerald.  
Hu, J., et al. (2001) Is There A Finite vs. Nonfinite Distinction in Chinese? *Linguistics* 39-6: 1117-1148.  
Huang, J. (1989) Pro-drop in Chinese: A Generalized Control Theory. In: Osvaldo Jaeggli and Kenneth J. Safir (eds.), *The Null Subject Parameter*, 185-241. Kluwer Academic Publishers.  
Li, A. (1990) *Order and Constituency in Mandarin Chinese*. Khwer Academic Publisher.  
Xue, N., et al. (2000) The Bracketing Guidelines for the Penn Chinese Treebank (3.0). Tech. Rep. 00-08, Institute for Research in Cognitive Science, University of Pennsylvania.  
黄衍 (1992) 汉语的空范畴 『中国语文』 5: 383-393.  
汤廷池 (2000) 汉语的限定子句与非限定子句 *Language and Linguistics* 1.1: 191-214.  
徐烈炯 (1994) 与空语类有关的一些汉语语法现象 『中国语文』 5: 321-329.  
徐烈炯 (1999) 从句中的空位主语 『共性与个性-汉语语言学中的争议』, 159-175.