

## 中国語三人称代名詞照応現象解析の一手法

于 素秋<sup>1</sup> 横山 晶一<sup>2</sup> 西原 典孝<sup>2</sup><sup>1</sup>山形大学大学院工学研究科システム情報工学専攻<sup>2</sup>山形大学工学部電子情報工学科

## 1.はじめに

照応解析における重要な問題は照応対象と指示対象をどう同定するかということである。中国語の構文構造は極めて複雑であるため、照応関係の同定が非常に難しい。中国語の照応についてはすでに、語義[1]、文構造[2]、統語機能[3]の立場から照応表現の形式、照応の制約条件などが分析されている。しかしながら、これらの研究は照応現象について検討しているのみで、解決方法は提示していない。一方、日本語においては、用例や表層表現、名詞の指示性[4]、語用論的、意味論的制約[5]、センター理論[6]と言った観点から解析がなされている。これらは中国語における照応現象解析にはある程度参考にはなるが、そのままでは適用しにくい。本研究は中国語の代名詞照応現象を分析し、その結果に基づいて、顕在性ゼロ照応<sup>1</sup>を含む三人称代名詞照応現象を解析する手法を提案する。その手法は人間の照応理解のプロセスに基いて、最初に語義関係、統語関係、文構造関係を用いる。またこれとは別に中国語における名詞の指示性、情報の新鮮度、文構造上の特徴を利用し、照応解析の「優先度」を設定し、それを規則ベースに組み込む。最初の解析で未解決のまま残されたものに対して、これらの規則を適用し、解析する。その結果、本手法が三人称代名詞照応現象解析に有効であることが確認された。

## 2.照応現象の依存性

照応現象は談話の構造に強く依存する。その依存

条件は名詞、述詞、話題の情報、構文関係などが考えられる。

## 名詞への依存性

中国語の三人称代名詞は体言の一つなので、統語機能も名詞と同じである。名詞で表現できるものは具体的なものでも抽象的なものでも全て代名詞で表現できる。一度表現されたものに再び言及する場合には、強調以外では、代名詞が省略か的手段で表現することが多い。そこで名詞は照応対象を解析するときの最も重要な参照点になる。

## 述詞への依存性

中国語では主に動詞、形容詞などが述詞になる。文の述詞と名詞句との間には語義上でも統語上でも種々の繋がりがある。述詞に依存して名詞句を代替または省略することができるが、ある名詞句の有無によって、文の意味がが変換することがある。文意に影響しないときに限って省略可能であり、ゼロ照応となることができる。

例1: a 老張: 有一個女兒, b  $\phi$  出嫁了(張は娘が一人あり、もう花嫁になった)。

例1の文aには、文b  $\phi$ の候補名詞として、「老張」と「女兒」の二つがある。文bにある述詞「出嫁」の意味属性(女性だけできる動作・状態)により、その主体は「女兒」しかできない。従ってこの文の省略は可能である。

例2: a 老張: 有一個女兒, b  $\phi$  身体不太好。(張氏は娘が一人あり、体があまり丈夫でない)。

文bでは主題部が省略され、述題部「身体不太好」が「老張」でも「女兒」でも可能な状態であるので、この省略は曖昧になる。意味関係を明確的に示すためには代名詞の「他」か「她」を用いて、「…, 他(她)身体不太好」のように表現すべきである。

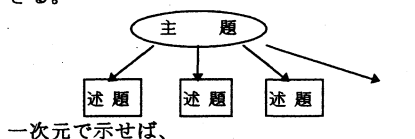
従って、述詞は中国語文における代名詞照応、特に顕在性ゼロ照応の参照点の一つと考えられる。

<sup>1</sup> 中国語の省略は顕在性と隠在性に分けられる。隠在性省略は照応性語句に反映されるものではなく、文結構の制約形式により実現する。顕在性省略は構文結構に反映されるだけでなく、照応性語句の制約形式により実現する。

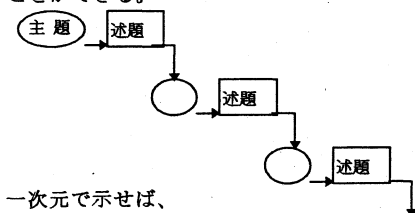
### 話題と情報への依存性

中国語文は主題部と述題部からなる。主題部が談話の出発点であるのに対して、伝達しようとする情報を徐々に展開する部分は述題部である。主題部は文成分の一つであり、既知の、又は分かりやすい情報を表現する。発話者はそこから発話し、新しい情報を聞き手に伝え、その後、元的话题を引き続き陳述するか、それとも新しい情報を話題にして更に新情報を伝えるかによって述題部を展開する[7]。

既知情報を利用し、未知情報を伝達する方式は「主題チェン」式と言い、下図のように示すことができる。



主題部(述題部, 述題部, 述題部……)となる。新情報を主題にし、更に未知情報を引出す伝達方法は「情報チェン」式と言い、図のように示すことができる。



「主題チェン」式でも、「情報チェン」式でも、その話題に最初にはっきり言及し、その後は省略や代替で表現する。それは照応対象存在の依存点になるし、話題又は焦点を判断する重要な参考点になる。

### 構文関係への依存性

幾つかの文をつなぎ合わせる手段として二つの方法が考えられる。それは文の連結と文の圧縮である。圧縮させる手段としては、関係節によって重複する名詞を省略したり、代名詞化したりする方法がある。文を連結させる手段としては、主題、述題、又は主語、目的語、時には定語などの関係節で連結したり、等位関係、主従関係などを示す関連語句で連結したりする方法がある。

関連語句は文と文とを繋ぐために用いられる。それがなければ、文と文或は節と節の間の意味的な繋

がりによって文を理解しなければならない。それがあれば、聞き手は言語知識を用いて、前文と後文との間の意味的關係にある種の予測を立てる[8]。

従って関連語句の位置によって、分句間にある文法関係、意味関係が明確にできる。そこで複文の関連語句、関連語句の位置は省略や代名詞の照応関係を処理する参照点の一つであると考えられる。

### 3. 照応解析の手法

人間は照応関係を理解する時に、まず照応表現の認定、次に先行表現の認定、更に指示対象の認定というプロセスを経る。認定方法は主として代名詞そのものの情報、名詞の指示性、動詞の属性、文構造関係、特に人間の知識などが考えられる。本研究では人間のプロセスに従い、照応解析の流れを設定する。

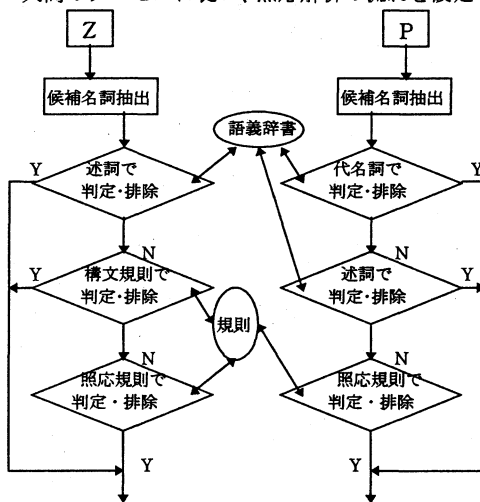


図1 「Z」プロセスと「P」プロセスの流れ

本手法は構文解析が終わった段階からスタートする。その段階では文の統語関係(文成分)、文の構造関係、文の語義関係が明らかになっている。三人称代名詞と省略の有無、あるとすれば、それらの文成分もはっきりしていると仮定する。構文解析によって、文の照応現象は①省略だけのゼロ照応；②三人称だけの代名詞照応；③省略も三人称もある複雑照応に分けられる。本解析手法の流れは先ず第一ステップとして、ゼロ照応対象を解析する。これを「Z」プロセスと称する。次に第2ステップで、三人称代

名詞照応対象を解析する。これを「P」プロセスと呼ぶ。「Z」プロセスと「P」プロセスは図1のとおりである。

### ゼロ照応解析

既に述べたように、文は主題部と述題部からなり、述題部の述詞は主語と目的語を支配する。自動詞述語と非動詞述語は主語と共存し、他動詞述語は主語の他に目的語とも共存する。従って、文中にある顕在性省略成分は主語と目的語（時に定語）である。これに基づいて、ゼロ照応解析では先ず初めに述詞を用い、その支配対象の候補名詞を判定・排除する。

述詞で判定可能ならば、直接文生成段階に入り、文を生成する。判定できないなら、次の段階に入り、構文規則で判定・排除する。できるなら、文を出力する。構文規則でも解析できない場合、照応規則を用いて処理する。

### 三人称代名詞照応解析

三人称代名詞照応では、代名詞そのものから先ず情報（例えば、「她们」は「人間、女性、複数」である）を取り出す。その情報と適合するものを先行詞として判定し、文を出力する。候補名詞が複数あり、代名詞で処理できない時は、述詞、規則で処理する。

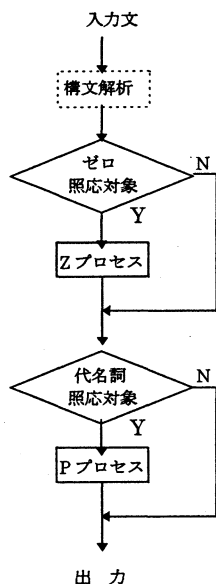


図2 照応解析の流れ

### 複雑な照応解析

実際の文には代名詞照応とゼロ照応現象が両方ともある場合が多い。その場合には、まずゼロ照応対象、その次代名詞照応対象という順に処理する。

従って、複雑な照応現象をまず「Z」プロセスに送り、ゼロ照応処理をする。「Z」プロセスが終わってから「P」プロセスに送り、代名詞照応処理をする。

本手法の全システムの流れを図2に示す。

流れ図2のように、構文解析によって出てきた照応現象をまずステップ1「ゼロ照応対象」に入れ、ゼロ照応現象があれば、Zプロセスで解析を行う。ゼロ照応現象がなければ、そこから出てステップ2「代名詞照応対象」に入れて、解析を行う。ゼロ照応現象を解析してからも、文をステップ2に入れ、代名詞照応現象があれば、引き続き解析を行い、なければ、それから出て文を出力する。

### 4.適用例と結果

前述の通り、本手法は先ず代名詞、述詞の語義属性、及び代名詞、述詞が名詞との間にある語義関係、更に文の構文構造関係を用いて顕在性ゼロ照応を含める三人称代名詞照応関係を解析する。それだけでは半分程度しか処理できない。よって本手法は名詞の指示性、情報の新鮮度、「主題チェン」などのような文構造に現われる特徴を捕まえ、人間の知識、論理関係をも利用して、照応解析の「優先度」を設定した照応規則ベースを作成した。候補が複数か曖昧性のある場合、照応規則を適用し、処理する。照応規則ベースは開放式である。ユーザがそれを補充したり、修正したりすることができる。

本手法の有効性を確認するために、雑誌「神州学者」(97年第87,89期)から100文(単文15文、複文85文、代名詞照応文(「P」文)50文、顕在性ゼロ照応文(「Z」文)25文、複雑な照応文(「ZP」文)25文を選び、この手法を適用して検証した。

例3: a 小張 || 不仅 会 英語, b φ || 也 会 日語, c 他 || 还 会 (几句) 俄語<sup>2</sup>。(張君は英語だけではなく日本語もでき、彼はロシア語も少し話せる。)

<sup>2</sup> abc…は複文の各分句番号を示す。||は主題部と述題部の標識であり、前の部分は主題部、後の部分は述題部である。下線の単線は主語、二重線は述語、波線は目的語であり、その他( )は定語、[ ]は状語、< >は補語を示す。

構文解析により、次のことが分る。

①「累加複文」である。

②関連語句は「不仅…，也…，还…」であり、第一関連語句「不仅」は第一文の主語「小張」の後に位置される。

③文 b にはゼロ照応対象があり、省略されたのは文 b の主語である。

④文 c には三人称代名詞「他」があり、それは文 c の主語である。

処理プロセスとしてまず文をゼロ照応ステップに送り、「Z」プロセスでゼロ照応解析をする。次に代名詞照応ステップに送り、「P」プロセスで代名詞照応解析をする。

文 b の候補として「小張」と「英語」の 2 個、文 c 「他」の候補として「小張」、「英語」と「日本語」の 3 個がある。

「小張」：固有名詞、姓(男・女)。

「英語」：固有名詞、言語名。

「日本語」：固有名詞、言語名。

ステップ 1：述詞「会」で判定・排除する。

「会」：可能動詞、N1 会 N2 (N1：主体、機械、商品等；N2：人間活動、抽象物(精神)変動等)

「会」により「小張」を判定し、「英語」を排除する。

述詞だけで判定できた。

ステップ 2：代名詞「他」で判定・排除する。

「他」：人間、単数、男性；

「他」により「小張」を判定し、「英語」と「日本語」を排除する。

解析処理が終わり、文を出力する。

例 3 と似た例 4 への処理は次のようになる。

例 4：a((小李的)朋友)小張 不仅会 英語, b 也会 日語, c 他还 会 (几句)俄語。(李君の友達張君は英語だけではなく日本語もでき、彼はロシア語も少し話せる)。

文 b の候補として「小李」「朋友」「小張」「英語」があり、文 c 「他」の候補として、その外にまた「日本語」もある。

述詞「会」によって、「小李」「朋友」「小張」が候補として残される。そこで、構文規則で処理する。

構文規則 11：[関連語句「不仅…，…」が第一文主語の後に置かれる場合、各分句の主語を同一とする。主語である候補名詞を優先する]

文 b と文 c の候補名詞として「小李」、「朋友」、「小

張」があるが、「不仅」が主語「小張」の後に置かれるので、規則により、各分句の主語は第一文の主語の「小張」と同定される。

ステップ 2 では「他」でも、述詞「会」でもステップ 1 と同様に排除できない。そこで規則で処理する。規則 11 により、「他」も同じ主語である「小張」と同定される。

例 5：a 老張的兒子小張是李教授的学生, b 他日語說得非常好(張氏の息子張君は李教授の教え子であり、彼は日本語が上手に話せる)。

この文にはゼロ照応対象がないので、直接「P」プロセスに送る。代名詞「他」の候補名詞として「老張」、「兒子」、「小張」、「李教授」、「学生」がある。代名詞「他」、述詞「說」からは判定・排除できない。照応規則 8 ([複数の候補名詞が判断説明文「N1 是(N 的・A)+N2, …」にある場合、主語である N1 を優先する])により、主語である候補名詞「小張」を優先とする。

本手法を適用した結果を表に示す。

	P で	V で	規則で	失敗(%)	成功(%)
P 文	21	13	14	2(4%)	48(96%)
Z 文		9	15	1(4%)	24(96%)
ZP 文	12	8	24	3(12%)	22(88%)

失敗した原因は色々あるが、主に文構造に関係があると考えられる。特に複文の構文関係は一層だけではなく二層、三層の時もある。また ZP 文の解析はより一層難しい。これは今後の課題として検討すべき点である。

参考文献：

- [1] 廖秋忠：「現代漢語篇章中指同的表達」、中国語文(1986.2)
- [2] 陳平：「漢語零形回指的話語分析」、中国語文(1987.5)
- [3] 沈陽：「名詞空位的控制性同指、照応性同指与詞匯性同指」《語言工程》清華大学出版社(1997)
- [4] 村田真樹等：①用例や表層表現を用いた日本語文章中の指示詞・代名詞・ゼロ代名詞の指示対象の推定、自然言語処理 Vol.4 No1(1997) ②名詞の指示性を利用した日本語文章における名詞の指示対象の推定、自然言語処理 Vol.3 No1 (1996)
- [5] 中岩浩已等：語用論的・意味論的制約を用いた日本語ゼロ代名詞の文内照応解析、自然言語処理 Vol.3 No4. (1996)
- [6] 田村浩二等：センター理論による日本語談話の省略解析、情報処理学会研究報告 NL107—12 (1995)
- [7] 張伯江等：《漢語功能語法研究》江西教育出版社 (1996)
- [8] 胡壯麟：《語篇的接与連貫》上海外語教育出版社 (1994)