

深い構文解析のための 高被覆な中国語 HPSG 文法の開発

岩澤俊弥* Xiangli Wang* Kun Yu*
松崎拓也* 宮尾祐介* 辻井潤一*†‡

本稿では、現実の中国語テキストに対する深い構文解析の基礎となる、基本的な文法現象を高被覆する中国語 HPSG 文法の設計について報告する。始めに、文法枠組み全体について述べたのち、いくつかの具体的な中国語文法現象に関して、我々の文法がどのように解析するかを述べる。また、現在の文法の評価として、Penn Chinese Treebank の文に対する被覆率を報告する。

The development of broad-coverage Chinese HPSG grammar for deep parsing

Shunya Iwasawa* Xiangli Wang* Kun Yu*
Takuya Matsuzaki* Yusuke Miyao* Junichi Tsujii*†‡

In this paper, we describe the development of broad-coverage Chinese HPSG grammar for deep parsing. First, we show the whole design of our grammar, and then explain how our grammar analyzes some basic grammatical phenomena of Chinese. Finally, we evaluate the coverage of our grammar on sentences in Penn Chinese Treebank.

1. はじめに

我々は、実世界の中国語テキストに対する深い構文解析を目標として、HPSG に基づく中国語文法および構文解析器の開発を進めている。従来は HPSG 文法の開発はコ

* 東京大学大学院情報理工学系研究科コンピュータ科学専攻
Department of Computer Science, University of Tokyo
† School of Computer Science, University of Manchester
‡ National Center for Text Mining, UK

ストが高く大規模文法の開発は困難と考えられていたが、近年の研究では既存のコーパスから HPSG 文法を獲得する手法が一定の成功を収め、実際に大規模な英語 HPSG 文法が開発されている。中国語についても既にコーパスが整備されており、我々はこれらのコーパスを利用して大規模な中国語 HPSG 文法を獲得することを目指している。

ところが、中国語に対する HPSG 理論は現在までに確立していない。中国語の文法に関する言語学的研究は数多く存在するが、特定の文法現象のみに着目した研究に限られている。我々の目標は実世界のテキストを解析することであるので、個々の文法現象の分析の厳密性より、多くの文法現象に対する網羅性の方が重要である。したがって、様々な文法現象を網羅する統一的な文法理論を構築することが必要である。

本稿では、実世界のテキストに出現する文法現象を網羅する中国語 HPSG 文法について報告する。まず、中国語の基本的な文の構造は英語と類似しているため、Sag et al. (2003) の英語 HPSG 文法を基盤とした。そして、実際の中国語文を調査した上で中国語特有の文法現象に対して文法の拡張を行い、網羅性の高い中国語 HPSG 文法を開発を行った。辞書は将来研究においてコーパスから獲得することを想定しているので、現時点では文法の基本設計（語彙項目テンプレートや文法規則などの定義）を行っており、本稿においてこの基本設計について詳述する。また、Penn Chinese Treebank を用いて本文法の被覆率を評価した結果について報告する。

2 節では、HPSG の概要と中国語文法に関する関連研究を紹介する。3 節では、我々の文法の基本設計について説明する。4 節では、中国語特有の文法現象について我々の文法による分析を示す。5 節では、現時点での文法の被覆率を評価した実験について報告する。

2. 背景

2.1 Head-driven Phrase Structure Grammar (HPSG)

本研究では、Sag et al. (2003) の HPSG 理論を基盤とし、中国語特有の文法現象に対する拡張を行うことで中国語 HPSG の開発を行う。HPSG は語彙化文法の一つであり、単語の統語的・意味的性質を記述する語彙項目と、複数の句を結合してより大きな句を作るための文法規則から構成される。

HPSG では、語彙項目、文法規則、構文木を sign という素性構造で記述する。図 1 に他動詞 “see” の語彙項目の sign を示す。PHON は単語列の音韻（単語列の表層文字列）を表す。HEAD は句の主辞となる単語の品詞を表す素性である。MOD は句の修飾先に関する制約を表す素性である。SPR, COMPS は単語の下位範疇化フレームを表す素性であり、それぞれ主辞の左側の項、右側の項に対する制約を表す。図 1 では、“see” は他動詞なので、SPR と COMPS にそれぞれ一つずつ要素を持つ。GAP は Wh 移動や

長距離依存を含む構文において、単語の項の中で痕跡となったものに関する制約を表す素性である。痕跡は痕跡を埋める句（フィラー）と結合しない限り、親ノードの sign に伝播する。STOP_GAP は痕跡を埋める（痕跡の伝播を止める）句に関する制約を表す素性である。INDEX と RESTR は意味構造を表す素性である。INDEX は意味的主辞の意味を表現し、RESTR はその他の修飾関係を表す素性である。

Sag et al. (2003) の英語 HPSG 文法では、Head-Specifier, Head-Complement, Head-Modifier, Head-Filler, Coordination の 5 個の文法規則が定義されている。図 2 に Head-Specifier 規則を示す。これは、右側の sign が主辞となって、左側の sign を指定詞（主語など）として取る文法規則である。主辞の SPR 素性の要素と左側の sign が単一化されることで、主辞が指定詞を選択している。

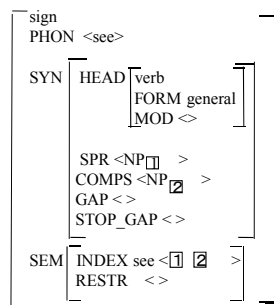


図 1 他動詞“see”の語彙項目

Figure 1 The lexical entry of transitive verb “see”

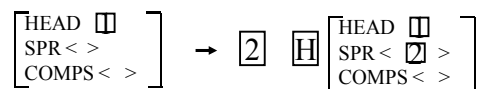


図 2 Head-Specifier 規則

Figure 2 Head-Specifier Rule

2.2 関連研究

中国語の構文解析に対しては、Penn Chinese Treebank などのツリーバンクを学習データとして用いて、句構造解析や依存構造解析の研究が進められている。深い構文解析を目的としたものでは、LFG に基づく構文解析の研究 (Fang et al., 2007) が挙げられるが、まだ十分な構文解析精度は得られていない。一方、HPSG に基づく中国語文法

の研究では Gao (2000) が挙げられるが、これはトピック構造を中心とした言語学的研究であり、現実のテキストの解析を目標として幅広い文法現象を扱うものではない。また、名詞句の構造や「把」「被」構造など、特定の構文構造に着目した言語学的研究はいくつかあるが、文法現象を網羅的に扱うことを目指した中国語 HPSG 文法は現在までに存在していない。

3. 中国語文法の設計

我々の最終的な目標は高被覆な深い構文解析を行うことであるため、本研究で開発する文法は実世界のテキストに現れる文法現象を網羅的に解析できる必要がある。しかし、中国語の文法現象を網羅する HPSG 理論は現在までに存在しない。したがって、我々はまず英語の HPSG 理論を参考に中国語の基本文法を作成した上で、以下のようなアプローチで文法理論の拡張を行った。

1. 中国語の文法書 (Fan, 1998) より、それぞれが異なる文法現象を含む 65 文を選び、これらの文を解析できるように文法を拡張する。
2. Penn Chinese Treebank (CTB) (Xue et al., 2005) より抽出した 90 文に対して同様の作業を行う。

語彙項目や文法規則を設計する際には、Sag et al. (2003) の HPSG 理論をできるかぎり踏襲し、中国語に特有な文法現象の場合は独自の拡張を行った。以下、3.1 節では sign と文法規則について、3.2 節では HEAD 素性の型階層の設計について説明する。

3.1 Sign と文法規則

Sign の定義は、基本的に Sag et al. (2003) を踏襲する。ただし、後述するように中国語には英語にはないトピック文が存在するため、Gao (2000) の分析に倣い、TOPIC 素性を追加する。トピックを SPR 素性で扱うという解決策も考えられるが、文と動詞句の sign が区別できなくなる問題があるため、SPR 素性とは別に TOPIC 素性を導入した。TOPIC は SPR, COMPS と同様に、単語の項を表す素性である。図 3 に、主語の他に名詞句のトピックを項としてとる形容詞性動詞“長/long”の語彙項目を示す。TOPIC を除く他の素性は、2.1 節で述べた英語の HPSG 理論における各素性と基本的に同様である。意味表現では、トピックも“長/long”の述語項構造に加えられている。

文法規則は 2.1 節で導入したものに加え、トピック文を扱うための Head-Topic 規則を定義した。Head-Topic 規則は Head-Specifier 規則とほぼ同様の働きをするが、SPR 素性の代わりに TOPIC 素性の値を制約とする点異なる。

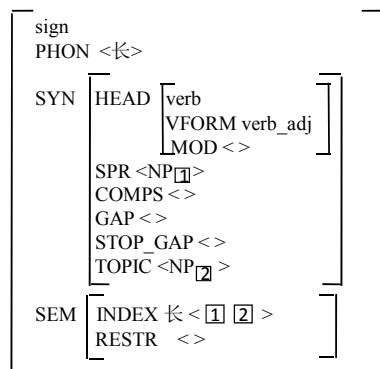


図 3 トピックを持つ形容詞性動詞“长/long”の語彙項目
Figure 3 The lexical entry of “长/long”, which has a topic

3.2 型階層

本文法で HEAD 素性の値となる型の階層を図 4 に示す。小文字は型を表し、大文字はその型で定義される素性を表す。型階層を設計する際には、出来る限り語彙項目の冗長性を排除する点に注意した。

全ての単語はまず noun, verb, prep, particle, adv, det, conj, class, case_marker, relativizer, complement の 11 個の型に分かれる。このうち、noun から conj まではそれぞれ名詞、動詞、介詞（英語の前置詞に相当）、助詞、副詞、限定詞、接続詞を表し、英語の対応する品詞にほぼ相当する。class は量詞（日本語の助数詞に相当）を指し、中国語特有の品詞である。その他の型は、中国語特有の文法現象を解析するための独自の分類である。case_marker は現在「把」「被」の 2 単語である。これらの単語に関しては、動詞とする考え方や介詞とする考え方もあるが、長距離依存と考えられる例文があることから扱いを未だ決定しておらず、暫定的に動詞とも介詞とも異なる品詞とした。relativizer は関係代名詞の「的」のみを含む。「的」は文法書では助詞の一つとされているが、他の助詞とは異なる統語的振る舞いをすることから、別の品詞として定義した。complement は、中国語の結果補語を表す。結果補語は自動詞や形容詞性動詞が主動詞の後ろに置かれて意味を添える文法現象であり、文法書などで品詞は動詞とされるが、結果補語となる単語は単独で述語にはならないため、統語的性質が動詞と異なる。したがって、これらの単語には別の品詞を定義した。

noun, verb, conj, class 型は、それぞれ NFORM, VFORM, CONJ_FORM, CLASS_FORM 素性を用いてさらに細かく分類する。

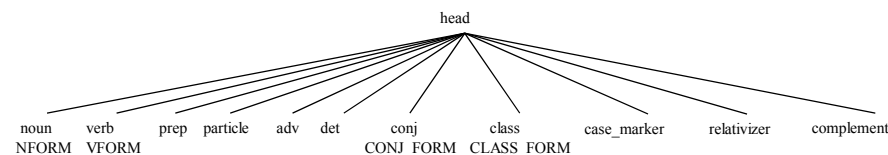


図 4 HEAD の型階層
Figure 4 The type hierarchy for HEAD

名詞 noun は NFORM 素性によって、local, temp, general, proper, pronoun, cd_noun に細分化する。中国語には、一般名詞に加えて場所名詞、時間名詞、方位詞と呼ばれる品詞が存在し、これを全て異なる品詞と考えるか、あるいは名詞の下位型と考えるかについて様々な研究がある（相原, 1988; 朱, 1995）。本文法では、語彙項目を最大限効率的に定義する視点から、一般名詞、場所名詞、時間名詞、方位詞はすべて名詞の下位型であるとした。また、固有名詞、代名詞、数詞も名詞の下位型として定義している。これにより、これらの品詞が名詞として同じように振る舞う場合（例えば、動詞の主語や目的語となる）と、特殊な振る舞いをする場合（4.3 節を参照）の両方を適切に扱うことができる。

動詞 verb は本動詞に加えて助動詞と形容詞性動詞を含む。中国語の形容詞は、単独で述語になれる性質を持ち動詞とほぼ同じ構文構造を持つ。したがって、動詞と統一的に取り扱うために verb 型を持つとした。助動詞と動詞を同じ型で扱うのは英語 HPSG の分析に倣っている。これらの細分化は verb 型に定義される VFORM 素性によりなされる。

接続詞は CONJ_FORM 素性によって、等位接続詞と従属接続詞を分けている。量詞は CLASS_FORM 素性によって、class_adj と class_adv に細分化される。class_adj は名詞句を修飾する量詞句を構成する量詞であり、class_adv は動詞句を修飾する補語となる数詞句を構成する量詞である。本文法では、数詞と量詞の結合における主辞が、結合後に生成する句の文法的役割によって異なるため、不要な曖昧性を生成しないために量詞を細分化している。

4. 中国語特有な文法現象の解析

本章では、中国語文法における基本的かつ高頻出ないくつかの文法現象を、本文法がどのように解析するかについて述べる。

4.1 トピック構造

本文法では、Wang et al. (2009) に従い、主語より前に置かれる句をトピックと考える。トピック構造は、トピックの後ろに来る文に痕跡が存在する場合と存在しない場合の2種類に分けられる。まず、痕跡がある文の例を以下に示す。

- (1) 飯/meal 我/I 吃/eat (I eat a meal)
- (2) 汽车/train 小李/Li 喜欢/like 日本/Japanese 的/one (Li likes Japanese trains)
- (3) 苹果/apple 小李/Li 吃/eat 了 三/three 个/numeral classifier (Li eats three apples)

例 (1) は“吃/eat”の目的語が文頭に移動した文で、英語にも存在する構文である。したがって、英語と同様に“吃/eat”が GAP 素性に目的語に相当する要素を持つとし、“飯/meal”と Head-Filler 規則で結合されるとする。

例 (2)、(3) は名詞句から主辞の名詞だけが文頭に取り出された構造である。この構文は中国語特有のものであるが、「的/one」や「个/numeral classifier」の語彙項目に痕跡を持つものを定義することで、例 (1) と同様に解析することができる。

次に、トピックに対応する痕跡がない場合の例を示す。この場合は、トピックが文内の他の単語と照応関係にある場合とない場合の2通りに分けられる。

- (4) 小李/Li 他/he 吃/eat 饭/meal (Li eats a meal)
- (5) 今天/today 他/he 吃/eat 饭/meal (He eats a meal today)
- (6) 大象/elephant 鼻子/nose 长/long (Elephant's nose is long)
- (7) 唱/sing 歌/song 小李/Li 很/very 拿手/good (Li is very good at singing a song)

例 (4) は照応関係がある例で、“他/he”が“小李/Li”を指している。この場合は、代名詞が痕跡を導入すると考え、例 (1)~(3)と同様に Head-Filler 規則で解析することで“他/he”と“小李/Li”の関係が捉えられる(図5参照)。

一方、例 (5)~(7)は照応関係がない例で、それぞれ、トピックが時間詞句、名詞句、動詞句の場合である。例 (5) のようにトピックが時間詞句などの修飾句の場合は、それらが修飾句の本来の位置(主語と述部の間)に現れる場合と同様に Head-Modifier 規則で解析する。一方、例 (6)、(7) のようにトピックが非修飾句の場合は、Gao (2000) の分析に倣い、新たに導入した TOPIC 素性と Head-Topic 規則を用いて解析する。例 (6)、(7) の構文を取る述語は限られているので、そのような述語の語彙項目として TOPIC 素性を持つものを与え、その要素とトピックが Head-Topic 規則で単一化されるとする(図6参照)。

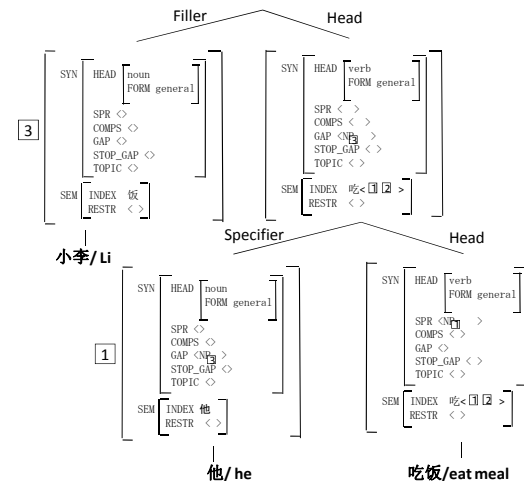


図5 トピック構造の構文木(1)

Figure 5 A parse tree for the topic construction (1)

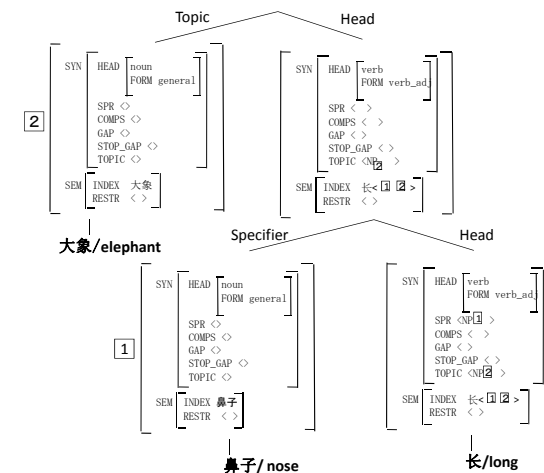


図6 トピック構造の構文木(2)

Figure 6 A parse tree for the topic construction (2)

4.2 関係節構造

中国語には、「的」を用いた関係節構造が存在する。関係節の構造は、主辞となる名詞句の関係節中での文法的役割、主辞となる名詞句の有無、関係代名詞「的」の有無によって場合分けができる。まず、主辞となる名詞句の文法的役割としては、主語、目的語、トピック、副詞句の4種類がある。

- (8) 买/buy 书/book 的/relativizer 人/man (a man who buys a book)
- (9) 我/I 买/buy 的/relativizer 书/book (a book which I buy)
- (10) 鼻子/nose 长/long 的/relativizer 大象/elephant (an elephant whose nose is long)
- (11) 我/I 买/buy 书/book 的/relativizer 钱/money (money with which I buy a book)

例(8)~(10)の文では、関係節の文に痕跡が存在し、主辞の名詞句はその痕跡を埋めていると考えることができる。一方、例(11)では主辞の名詞が関係節に対して副詞的役割を持つ。この場合、関係節との関係は介詞を伴わなくては分からないため、現在の文法では例(11)は関係節構造とは考えず、単に「文+的+名詞句」という構造として解析する。この際の問題点として、同じ構造が同格を表す場合もあるが現在はその区別はできない。

続いて、主辞となる名詞句の有無と関係代名詞「的」の有無について考える。ただし、主辞となる名詞句と関係代名詞がともに省略される場合は存在しないため、以下の3つの場合が考えられる。

1. 主辞となる名詞句、関係代名詞ともに存在する場合
2. 主辞となる名詞句は存在しないが、関係代名詞は存在する場合
3. 主辞となる名詞句は存在するが、関係代名詞は存在しない場合

主辞となる名詞が持つ文法的役割は、1と3の場合は全てのケースがありうるが、2の場合は副詞的関係になる場合は存在しない。

ケース1は最も基本的な構造である。関係節の文に存在する痕跡は、「的」との結合の際に「的」のSTOP_GAP素性によって埋められる。また、「的」のSTOP_GAP素性とMOD素性における制約を構造共有させておくことで、主辞となる名詞句は、関係節により修飾される関係と、関係節内の痕跡を埋めているという二つの関係がどちらも表現できている(図7参照)。

次に、ケース2の場合では、「的」が関係節の主辞となる点がケース1と異なる。「的」と関係節が組み合わせられるところまではケース1と同様に解析するが、「的」は他の名詞を修飾せず、自ら主辞となるため、MOD素性を空リストとする。この場合に完成する句は名詞句であるので、ケース2の場合の「的」の語彙項目のHEAD素性はnoun

とした。

最後に、ケース2は、現在の文法ではまだ実装されていないが、「的」が省略されていると考え、省略された「的」と関係節が結合するような単項規則を定義することで解決する予定である。

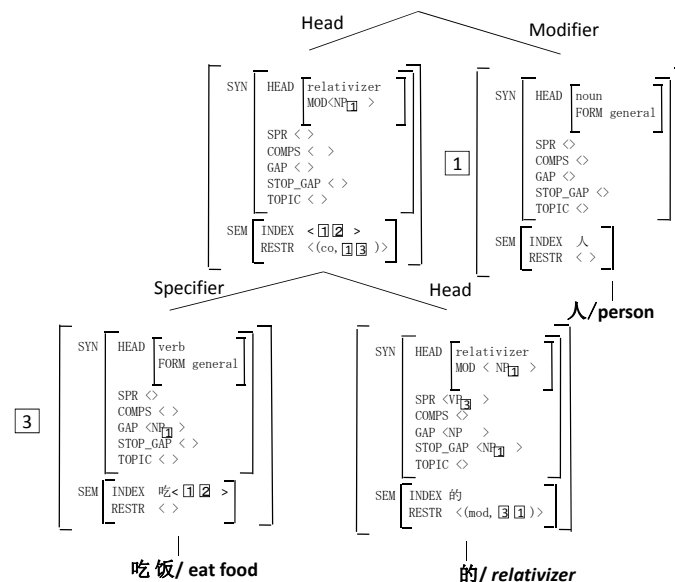


図7 関係節の構文木

Figure 7 A parse tree for the relative clause construction

4.3 方位詞句構造

中国語では、場所や時間の意味を持つ句の一つとして、方位詞句がある。方位詞は、名詞句、動詞句、文と結合して、主語や目的語、あるいは副詞句として働く。本文法枠組みでは、方位詞はnoun型とし、NFORM素性で方位詞と他の名詞を区別するようにしている。

方位詞は(12)のように、単独あるいは名詞句、動詞句、節と結合して、他の名詞と同様に動詞の主語や目的語となることができるため、名詞句と同様の文法的性質を持っている。一般名詞とは異なる点として、方位詞は句として働くために名詞句、動詞句、節と結合して飽和する必要がある場合があるが、この問題は方位詞のsignの下位

範疇化フレームに対する制約で解決することができる。また、方位詞句は(13)のように副詞句として働くことができるという一般名詞には存在しない、時間名詞などに特有の構文も存在するが、この現象は MOD 素性に修飾する動詞句の制約を記述することで、Head-Modifier 規則によって実現できる。

- (12) 桌子/table 上/*locative* 有/*exist* 苹果/apple (there is an apple on the table)
 (13) 我/I 走/go 后/*temporal*, 他/he 吃/eat 饭/ meal (he eats a meal after I go)

本文法で方位詞を名詞の下位型として定義している利点として、場所や時間の意味を持つ名詞句をまとめて表現できるという点が挙げられる。例えば、介詞「在」の目的語には、時間、場所の意味を持つ名詞句しか取れないという性質が、「在」の COMPS 素性で NFORM を指定すると、簡潔に表現できる。その結果、本文法では、(14)は非文ではないが(15)は非文であると正しく解析できる。

- (14) 他/he 把/ba 书/book 放/put 在/on 桌子/table 上/*locative* (he puts the book on the table)
 (15) *他/he 把/ba 书/book 放/put 在/on 桌子/table

4.4 接続詞を用いた複文構造

中国語には等位接続詞と従位接続詞がある。等位接続詞は、「X + 接続詞 + Y」の形で用いられるもので、X と Y は同じ統語範疇であることが多い。(16)は名詞句の並列の例であり、(17)は痕跡を含む関係節の並列の例である。これらの解析は、X と接続詞が接続詞を主辞として結合した後に、Y を修飾すると解析する(図 8 参照)。

- (16) 小李/Li 和/*and* 小王/Wang (Li and Wang)
 (17) 我/I 买/buy 的/*relativizer* 鼻子/nose 长/long 的/*relativizer* 大象/elephant (the elephant that I buy and has a long nose)

一方、従属接続詞は、(18)のように「接続詞 1+X+接続詞 2+Y」のように用いられ、複文構造を構成する。X, Y は動詞句か文のいずれかであり、また接続詞 1, 2 はどちらも省略されうる。更に、接続詞 1 と接続詞 2 は呼応関係にある場合が多い。本文法では、現在呼応関係を扱うことはせず、従属接続詞は単なる副詞として解析する方針を取る。

- (18) 因为/*because* 他/he 说得/talks 太/*very* 快/*fast*, 所以/*then* 我/I 听不懂/*don't listen clearly* (Because he talks very fast, I can't listen clearly.)

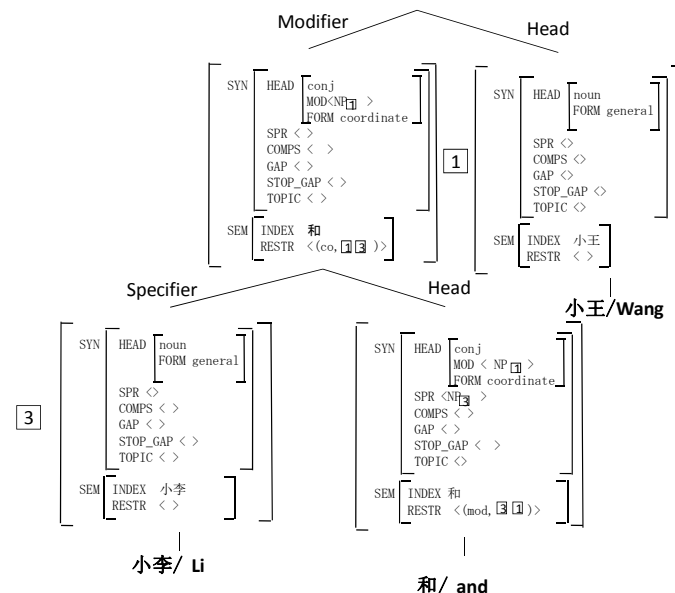


図 8 等位接続構造の構文木

Figure 8 A parse tree for the coordinate conjunction construction

4.5 「把」/「被」構造

中国語に特有な構文構造の一つに、「把」構造、「被」構造と呼ばれる構文構造がある。「把」構造は、動詞の目的語が「把」と結合して動詞の前に置かれる構造である。一方、「被」構造は受動態を表す構造で、「被」によって文の主語が取り出され、目的語が文頭に置かれる構文構造である。

- (19) 他/he 吃/eat 苹果/apple (He eats an apple)
 (20) 他/he 把/ba 苹果/apple 吃/eat (He eats an apple)
 (21) 苹果/apple 被/bei 他/he 吃/eat (An apple is eaten by him)

(19), (20), (21)は全て同じ意味を表す例文である。(20)は「把」を用いて目的語である「苹果/apple」を取り出した文であり、(21)は「被」を用いて受動化した文である。中国語の「把」「被」は特殊な文法現象であり、「把」「被」については、動詞とする考

え方や、前置詞とする考え方など多くの研究がなされてきた(Ding, 2000)。本研究では、「把」「被」を case marker とする考え(Yin, 2004)に倣い、「把」句や「被」句を項として取る語彙項目を定義することでこれらの構文を説明する(図 9, 10 参照)。

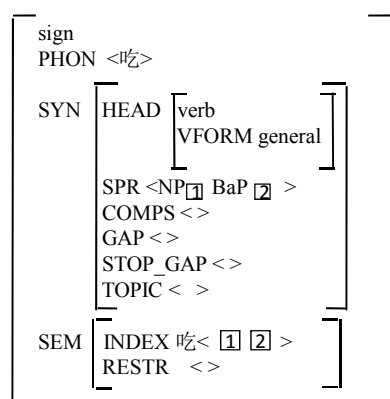


図 9 「把」構造のための動詞の語彙項目
 Figure 9 A lexical entry for the *ba* construction

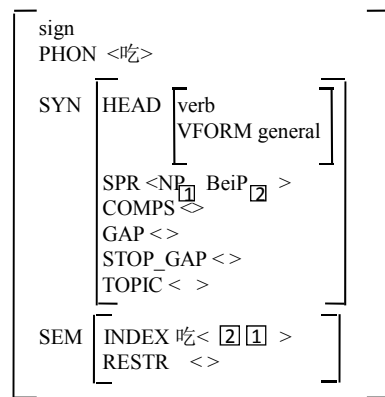


図 10 「被」構造のための動詞の語彙項目
 Figure 10 A lexical entry for the *bei* construction

5. 評価

本節では、現在までに開発した文法が現実のテキストに対してどの程度解析できるかを評価した結果を報告する。まず、文法開発に利用した文法書の例文の被覆率を語彙項目と文ごとに調べ、さらに、文法開発に利用していない CTB の文に対して被覆率を測定した。ただし、現在の文法では大規模な語彙項目辞書がないため、直接被覆率を測定することはできない。そこで、各単語に対して人手により語彙項目テンプレートを与える作業を行い、与えるべき語彙項目テンプレートが現在の文法に存在しない場合に、その単語を被覆していないと判定した。解析不能な文法現象に対して、どの単語の語彙項目不足が原因となっているかを特定することは難しいため、基本的に 1 つの解析不能な構造に対しては、その句の主辞となる単語の語彙項目が不足しているものとして計算した。

5.1 文法開発に使用した例文に対する被覆率

文法書(Fan, 1998)の例文 65 文に対する被覆率を表 1 に示す。本評価はクローズドテストであるため、本来は被覆率が 100%になるべきだが、正反疑問文や一語文など、まだ解析方法を検討中である文法現象が存在しているため、被覆率は 91.59%に留まっている。

表 1 文法書の例文に対する被覆率

Table 1 The coverage on the sentences in the grammatical textbook

語彙項目	316/345 単語	91.59%
文	47/65 文	72.31%

5.2 CTB の未知の文に対する被覆率

次に、CTB の未知の文に対する被覆率の評価を行った。テストデータは開発に用いた文と重複がないようにランダムに 90 文を抽出した。表 2 に被覆率を示す。語彙項目の被覆率は 95%を超えており、本文法が現実のテキストに対して高い被覆率を達成することが確認された。ただし、文単位での被覆率は 36.67%に留まっており、さらに文法の拡張が必要である。CTB の例文 90 文の中で語彙項目が存在しない単語は 127 単語であった。この中には重複した単語や、同じ文法構造も含まれているため、実際に本文法が扱えていない文法現象の種類はより小さい数になる。

表 2 CTB の未知の文の被覆率

Table 2 The coverage on unseen sentences in CTB

語彙項目	2996/3123 単語	95.93%
文	33/90 文	36.67%

6. まとめ

本論文では、深い構文解析を目標とした中国語 HPSG 文法の開発状況について述べた。未知の CTB の例文に関して、語彙項目の被覆率が 95%を超えたことは、現在の文法が基本的な文法現象について被覆出来ていることを示していると考えられる。今後の課題としては、まず文法の更なる拡張が挙げられる。現実のテキストを解析する際には、1 単語でも語彙項目が割り当てられない単語を含むと解析できなくなってしまうため、現在解析ができない文法現象については、特にその出現頻度の高いものから解決していく必要がある。また、今回評価に用いたデータは CTB のごく一部であり、今回の評価は十分なものとは言えないため、より多くの文を調査することで、現実の

テキストに現れるが本文法が解析できない現象を洗い出すことが必要である。さらに、本文法を用いて、文法開発の次の段階である文法獲得の実験を行う予定である。

参考文献

- 1) Fan, X. 1998. 汉语的句子类型(Sentence Types of Chinese). 书海出版社(China: Shuhai Press)
- 2) Fang, J. and King, T. H. 2007. An LFG Chinese grammar for machine use. In GEAF 2007.
- 3) Gao, Q. 2000. *Argument Structure, HPSG and Chinese Grammar*. Ph.D. thesis, Ohio State University
- 4) Yin, H. 2004. Grammaticalization of mandarin transfer verbs *gei* and *bei* as passive markers. In *Annual Conference of the Canadian Linguistic Association*.
- 5) Xue, N., Xia, F., MacIntyre, F. M. P.R. 2005. The Penn Chinese treebank: Phrase structure annotation of a large corpus. In *Natural Language Engineering 11*.
- 6) Ding, P. S. 2000. A Computational Study of the Ba Resultative Construction: Parsing Mandarin Ba Sentences in HPSG. In PACLIC 14.
- 7) Pollard, C. and Sag., I. A. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.
- 8) Sag, I. A., Wasow, T., and Bender, E. 2003. *Syntactic Theory*. CSLI Publications.
- 9) Wang X., Iwasawa, S., Yu, K., Matsuzaki T., Miyao Y. and Tsujii J. 2009. Design of Chinese HPSG Framework for Data-Oriented Parsing. In preparation.
- 10) 相原 茂. 1988. 現代中国語文法総覧 (上). くろしお出版.
- 11) 朱 德熙. 1995. 文法講義—朱德熙教授の中国語文法要説. 白帝社.