国立国語研究所学術情報リポジトリ

統語・意味情報付きコーパスの開発に関する研究：中国語名詞句の解析について

メタデータ	言語: Japanese 出版者: 国立国語研究所 公開日: 2019-07-25 キーワード: コーパス, 中国語, 統語解析, 主題, 主語 作成者: 周, 振, 吉本, 啓, ZHOU, Zhen, YOSHIMOTO, Kei メールアドレス: 所属:
URL	https://doi.org/10.15084/00002223

国立国語研究所論集 (NINJAL Research Papers) 17: 35–65 (2019)
ISSN: 2186-134X print/2186-1358 online

統語・意味情報付きコーパスの開発に関する研究
——中国語名詞句の解析について——

周　振[a]　　吉本　啓[b]

[a] 東北大学／国立国語研究所 共同研究員
[b] 東北大学／国立国語研究所 研究系 理論・対照研究領域 客員教授

要旨

この論文は，統語・意味情報付きコーパスを開発するに当たって，中国語名詞句の解析を考察するものである。名詞句の解析をめぐっては，二つの課題がある。それは，名詞句の内部構造を明らかにし形式的に解析すること，および名詞句の担う類似した統語的役割を区別することである。名詞句は，もっとも一般的に使われている句の一つだが，様々な修飾部を持つゆえに，それに対して一貫した解析を与えることは容易ではない。本研究は，名詞句の修飾部を中国語の実情に合わせた形でより細かく分類し，本研究の採用しているアノテーション方式の枠組みの中で，各修飾部の含まれる名詞句のパターンを網羅的に考察していく。一方，名詞句は文の中で多彩な統語的役割を果たしているが，類似したものの見分けが困難な場合がある。形式的な手がかりに欠ける中国語にとっては，この問題が特に顕著だと思われ，主題と主語の区別はその典型的な例である。本研究は，主題と主語の定義をコーパス構築という視点からそれぞれ明確に行う。その上で，中国語の様々な実例を考察し主題と主語を決定する基準を明確にする。名詞句の解析は，コーパス構築作業および構築できたコーパスを基にする言語研究の基本的かつ重要な一環を成しており，それを明らかにすることによって，研究の基盤を固めることができると考えられる *。

キーワード：コーパス，中国語，統語解析，主題，主語

1. はじめに

電子化された文章に対して言語分析情報を付加（タグ付け）したコーパスは様々な形で言語の研究や教育に利用されている。かつてはテクストを単語ごとに分けて品詞情報だけをタグ付けしたものが多かったが，現在は品詞情報だけでなく構文情報も付加したコーパス（ツリーバンク）が増えつつある。これにより，コーパス利用の効率は飛躍的に高まった。しかし，中国語に関しては，文系の言語研究者が簡単に利用できるツリーバンクはこれまでどこにも存在しなかった。さらに，ツリーバンクそのものにも限界がある。通常のツリーバンクでは表層的な統語情報が示されるだけである。複文が現れた場合，明示的に示されていない動詞の取る主語や目的語を捉えることは困難である。このような背景の中，筆者たちは，言語分析情報の付加（アノテーション）を意味のレベルまで進め，統語情報（ツリー）と意味情報（述語論理式）が両方提供できる中国

* 本稿は，NINJAL International Symposium 2017: Exploiting Parsed Corpora Applications in Research, Pedagogy, and Processing での発表を基にしたものである。本研究は，JSPS 科研費 18K12440「中国語学習を支援するためのデータベースの構築」（研究代表者：周振）の助成を受けている。また，本稿の一部は国立国語研究所機関拠点型共同研究プロジェクト「統語・意味解析コーパスの開発と言語研究」（プロジェクトリーダー：プラシャント・パルデシ）の研究成果である。本稿の作成において，査読者の先生から有益なご意見を沢山承りました。英文要旨に関しては Alastair Butler 先生にご助言を頂きました。この場を借りて厚く御礼申し上げます。

語コーパスを開発している。その作業は二つの段階に分けられる。即ち，1）分析データとして選ばれた中国語の自然テクストに対し統語情報を付与すること（統語解析），および 2）スコープ制御理論（Scope Control Theory; SCT, Butler 2010）を実装したシステムで 1）の結果を処理することによる自動的な文の論理意味表示の獲得（意味処理）である。

本論文は，統語・意味情報付きコーパスを開発するに当たり，中国語名詞句の解析を論じようとするものである。名詞句の解析は主に二つの課題があると思われる。それは，①名詞句の内部構造を明らかにした上でそれを形式的に解析すること，および②名詞句の持つ類似した統語的役割を区別することである。本論文では，この二つの課題を本研究の枠組みの中でそれぞれ明確にしていきたい。なお，頁数の制限もあり，名詞句の持つ統語的役割の区別については，もっとも代表的だと考えられる主題と主語をめぐる問題を取り上げることにする。

本論文の構成は以下の通りである。第 1 節では，研究の背景と目的を述べた。第 2 節では，統語解析を中心に本研究の使用するアノテーション方式を説明する。第 3 節では，名詞句の主要部を明らかにし，先行研究を踏まえて本研究における名詞句修飾部の種類を決めた上で，第 2 節で紹介したアノテーション方式を用いて各修飾部を持つ名詞句の内部構造を解析する。第 4 節では，主題と主語の定義を本研究の研究目的に合わせた形でそれぞれ明確に行い，中国語の様々な用例を実際に考察・解析していく。第 5 節では，まとめを行う。

2. 本研究のアノテーション方式

本節では，本研究で使用するアノテーションの方式について説明する。なお，実際の作業では，意味処理は統語解析の結果を基に全自動的に行われている（手作業を必要とするのは統語解析の部分だけである）ため，本節の説明は統語解析のみとする。意味処理については，Butler（2015）を参照してほしい。

本研究は，統語解析を行う際に，基本的にはペン通時コーパス（Penn Historical Corpora; PHC）式の解析規約（Santorini 2010）に従う。当解析スキームは，ペンツリーバンク式の解析スキーム（中国語版の場合は，Xue and Xia 2000）を修正したものであり，文の統語構造をラベル付きの括弧で表示する。ラベルには単語レベルの品詞タグ（例えば，普通名詞は N，形容詞は ADJ）と句レベルのカテゴリー（例えば，名詞句は NP，形容詞句は ADJP）の二種類がある。文の全ての末端要素（即ち，語と句読点）は，単語レベルのラベルによってタグ付けされているが，必ずしも句レベルのカテゴリーを持つ（即ち，句へ投射する）必要はない。

また，実用性重視という原則を貫き，同解析スキームは特定の言語理論に依拠していない。統率・束縛理論（GB; Chomsky 1981, Chomsky 1982）に拘ったペンツリーバンク式のものと異なり，PHC 式の解析規約では，中間レベルの構造（N´ や ADJ´ など）は，いかなる場合も明示的にタグ付けされることはない。その最大の理由は，これによりツリーの構造がシンプルなままに保てることにある。さらに，PHC 式の解析規約においては，言語学者たちがこれまで使い続けてきた幾つかの句の構造が廃止されている。その典型的な例は動詞句（VP）である。PHC 式の解析規約は，VP の境界を決めることはほぼ不可能に近いと考え，アノテーションの一貫性を保つた

めに，VP の使用を認めていないのである。PHC 式の解析スキームは，高い適用性を持ち，言語ごとに必要なある程度の修正を施した上で，これまで多くの言語（英語，フランス語，アイスランド語，ポルトガル語，ギリシア語，日本語など）に適用されてきた。

2.1 節の内部構造

（2）は，PHC 式の解析スキームを応用し実際に中国語の無制約の文（1）を対象として統語解析を行った一例である。

（1）　嗯 ，　所以　张三 不　想　买　那　一 款 新　车。
　　うん　だから 張三 否定 たい 買う 指示詞 -あの 一 種 新しい 車
　　うん, だから張三はあの新車を買いたくない。

（2）　(IP-MAT (INTJ 嗯)
　　(PU ，)
　　(CONJ 所以)
　　(NP-TPC-SBJ (NPR 张三))
　　(NEG 不)
　　(AX 想)
　　(VB 买)
　　(NP-OB1 (D 那)
　　(NUMCLP (CARDP (CARD 一))
　　(NUMCL 款))
　　(ADJ 新)
　　(N 车))
　　(PU 。))

中間レベルの I′ や VP は存在しないので，節（IP と CP）は常に動詞（VB）と節レベルの構成要素を直接支配する。動詞の他，中国語の場合，形容詞（ADJ）もそのまま述語になれるため，節に直接支配されることが許される。また，他にも少数の単語レベルの要素が節に直接支配される形で現れることができる。それは，(2)に見られる間投詞(INTJ)，接続詞(CONJ)，否定辞(NEG)および助動詞（AX）の他に，アスペクトマーカー（AS）とモーダル助動詞（MD）がある。以上のいずれの品詞タグも，句を投射しても常に二分木（二股）にならないという点において，共通している。

2.2 句の内部構造

句のヘッド（主要部）は各句によって直接支配されるため，中間レベルの構造が明示的に示されていない。(2) では，数助詞句（NUMCLP）のヘッドである数助詞（NUMCL）の“款”は，

NUMCLP によって直接支配されている。このように，原則として，ヘッド（NUMCL や ADJ など）は句レベルのカテゴリー（NUMCLP や ADJP など）と一致しなければならないが，二つの例外がある。それは，ヘッドが一般的なカテゴリーの下位カテゴリーのラベルを持っているか，またはヘッドが省略されたか，のどちらかの場合である。例えば，(2) では，固有名詞（NPR）の“张三”は名詞の下位カテゴリーであり，主題・主語名詞句（NP-TPC-SBJ）を投射している。同じように，普通名詞（PHC 式の解析規約では，普通名詞（normal noun）のタグは NN などではなく，単なる N になっている）の“车”が直接目的語名詞句（NP-OB1）を投射できるのも，そもそも普通名詞が名詞の一つの下位カテゴリーであるためである。

以上のように，ヘッドは通常句を投射するが，これにも二つの例外がある。まずは，前述したように，INTJ，CONJ，NEG，AX，AS，MD，VB および述語になった ADJ は常に句を投射しない。さらに，(2) における NP-OB1 の内部構造の示す通り，単一の単語からなる前置修飾部（当例の場合は限定詞（D）と ADJ）も句を投射しない。これは，単一の単語からなる前置修飾部の投射する句の可能性は予測できる（D は DP，ADJ だけでは ADJP を投射することしかできない）ため，枝分かれのない投射を減少することによって句の内部構造を単純化するためである。これに対して，複数の単語から構築される前置修飾部は，相当する句を投射する。例えば，(2) では，NUMCL の“款”は数詞句（CARDP）を取り NUMCLP を投射してから N を修飾している。一方，補部および修飾部が後置する場合は，いずれも相当する句を投射する必要がある。以上を踏まえて，句（IP と CP は節として扱われるため両者を除く）の内部構造を図 1 に一般化することができる。中間レベルの構造は存在しないので，従来のいわゆる指定部および付加部は修飾部に統合され，しかも修飾部も補部も常にヘッドと同じレベルにある。Y と YP の順番は問わない。さらに，中国語の場合，ヘッドが補部の後ろに来ることも稀にある。(2) の数詞（CARD）“一”がヘッドの NUMCL に前置し，しかも単一語からなるものであるにもかかわらず，CARDP を投射しなければならないのは，それが NUMCLP の補部になっているからである（中国語の場合，数助詞は基本的には数詞と共起しなければならない）。

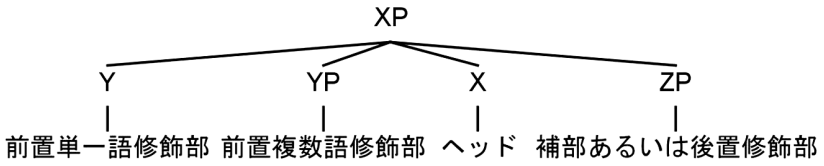

図 1　句の内部構造

2.3 機能情報

ラベルには，形式と機能を示すものがある。単語レベルの品詞タグは決して機能を示すものを持つことはないが，句レベルのカテゴリーは形式と機能を示すものを両方持つことができる。例えば，(2) における NP-OB1 の場合，NP は句のタイプが名詞句であることを表すと同時に，OB1 はこの名詞句の機能を直接目的語に限定している。一般的には，句レベルのカテゴリーは機能タグを一つしか持たないが，二つ以上持つ場合（例えば，NP-TPC-SBJ：主題であり，主語

でもある名詞句を表すタグ）もある。

節レベルの構成要素（即ち，IP・CP によって直接支配されるもの）の中では，NP は常に機能タグを持つことになる。(2) では，IP-MAT（主節）によって直接支配される二つの NP は，それぞれ TPC，SBJ および OB1 によってマークされている。また，全ての節は機能タグによってタイプ分けがなされている（IP-MAT の他，CP-ADV（副詞節）や CP-REL（関係節）などがある）。

3. 名詞句の構造

一般的に言えば，自然言語の中でもっとも重要でしかも頻繁に使われている文法カテゴリーは名詞句（NP）および動詞句（VP）である。第 2 節で紹介したように，VP はその境界を決めるのが困難であるという理由もあり，PHC 式の解析規約では，その使用を控えている。一方，NP はその内部構造が煩雑なわりには，句の境界が常にはっきりしている。本節では，名詞句の主要部と修飾部を考察した上で，第 2 節で紹介したアノテーション方式を用いて各修飾部を持つ名詞句の内部構造を解析する。

3.1 名詞句の主要部

言うまでもないが，名詞句の主要部は名詞である。普通名詞（N），固有名詞（NPR）および代名詞（PRO）が，名詞の下位カテゴリーであるため，名詞句の主要部になることができる。また，中国語の場合，名詞句の主要部の省略も可能である。

(3)　你　是　　北京大学　的　　学生　吗　　?
君　である　北京大学　助詞　学生　文末助詞 -か
君は北京大学の学生であるか。

是　，　我　是　　北京大学　的　　(学生) 。
はい　　私　である　北京大学　助詞　(学生)
はい，私は北京大学の（学生）である。

(3) から分かるように，名詞句主要部の省略は，1）名詞句の修飾部に助詞“的”を含む助詞句が存在すること，および 2）聞き手は文脈から省略された主要部の復元が可能であることを条件とする。名詞句の主要部の省略は，関係節構文からもよく見られる。

(4)　这里　有　　我　要　　的　　。
ここ　ある　私　ほしい　補文標識
ここに私のほしいものがある。

(3) における“的”と (4) の“的”とは別のものであるということに注意が必要である。省略された主要部の指示対象の同定を文脈に依存している (3) に対して，(4) の場合，文脈がなくても，このような同定は主要部を修飾する関係節の統語構造からある程度行える（その詳細は

3.5.1 節を参照のこと）。以上の他，後の 3.3 節で示したように，指示詞修飾部を持つ名詞句が文の最初（間投詞や接続詞が文の最初に現れる場合はその後ろ）に現れる時に限って，その主要部名詞の省略が可能である。

3.2 本研究における名詞句修飾部の種類

名詞の修飾語・句については，従来様々な観点から分類され分析されてきた。以下，Li and Thompson（1981）をはじめとする先行研究を紹介し，それを踏まえて，本研究での分類を述べる。Li and Thompson（1981）は，名詞を修飾する要素を数助詞句（classifier phrase）・度量詞句（measure phrase），連合句（associative phrase），限定句（modifying phrase）の三つに分けている。数助詞句は，数助詞（classifier）を中心とする句である。Li and Thompson（1981）は，（5）のような例を挙げ，中国語の場合，数助詞は数詞（numeral），指示詞（demonstrative）あるいは量化詞（quantifier）と共起しなければならないため，数助詞句は，主要部の数助詞のほか必ず数詞，指示詞と量化詞のどちらかを含むことになると主張している。

（5a）　一/这/每　　个　苹果
　　一/指示詞 -この/量化詞 -それぞれ　個　リンゴ
　　一個の/この/それぞれの/リンゴ

*（5b）　个　苹果
　　個　リンゴ

数助詞句の主要部の数助詞を度量詞に変えると，数助詞句は度量詞句になるが，（6）に示すように，度量詞句は基本的には数助詞句と同じような特徴を持つと考えられる。

（6a）　一/这/每　　斤　苹果
　　一/指示詞 -この/量化詞 -それぞれ　斤　リンゴ
　　一斤の/この斤の/それぞれの斤の/リンゴ

*（6b）　斤　苹果
　　斤　リンゴ

連合句は通常，名詞プラス助詞“的”という形を取る。（3）では，“北京大学的”は主要部名詞“学生”を修飾する連合句である。associative とは，二つの名詞が何らかの関係でつながっているということを意味する。この関係は二つの名詞の意味によって決められるものであるが，もっとも重要な関係の一つとして所有（possession）が挙げられると Li and Thompson（1981）は述べている。“北京大学的学生”の場合，“学生”は“北京大学”に所属していると考えられるため，両者の関係は所有になる。また，所有関係を構築する二つの名詞の間の“的”は省略することもある。この場合，両方の名詞はみな人間を指すものであり，しかも一番目の名詞が代名詞である必要がある。この条件を満たした次の（7）では，“的”の省略が可能である。

(7)　张三　爱　　他　(的)　　儿子。
　　張三　愛する　彼　(助詞)　息子
　　張三は彼の息子を愛している。

実は，助詞“的”と組み合わせることができるのは，中国語の場合，名詞の他にも幾つかあると考えられる。Li and Thompson（1981）は，形容詞プラス“的”というようなパターンを名詞化（nominalization）の一種となる関係節として扱っているが，それ以外のパターンについては言及していない。

限定句は，関係節（relative clause）と限定形容詞（attributive adjective）のことを指す。(8) と (9) に示すように，中国語の形容詞は，名詞を修飾する際に“的”を付けなくてもいいものと“的”を付けなければいけないものがある。

(8a)　红　　花
　　赤い　花
　　赤い花　　（Li and Thompson 1981）

(8b)　红　　的　　花
　　赤い　形式名詞　花
　　赤い花　　（Li and Thompson 1981）

*(9a)　舒服　　椅子
　　心地よい　椅子　　（Li and Thompson 1981）

(9b)　舒服　　的　　椅子
　　心地よい　形式名詞　椅子
　　心地よい椅子　　（Li and Thompson 1981）

両者の違いについて，Li and Thompson（1981）は，“的”を付けない形容詞修飾部と修飾される名詞とはより高い一体性を持ち，その結果全体が一種類の存在物（entity）の名前になる傾向があると述べている。つまり，(8b) は「赤色の花」という通常の意味であるのに対して，(8a) は一語としての傾向が著しく，ある種の花（例えば，表彰式に使う花など）全体を指している。その上で，“的”を付けない形容詞修飾部と“的”を付ける形容詞修飾部をそれぞれ形容詞句と関係節として分析している。

先行研究に比べて，本研究は中国語のデータを網羅的かつ形式的に扱う必要があるので，より詳しい対応が求められる。本研究は，本研究のアノテーション方式の枠組みの中で名詞の修飾部を八つの種類に分類する。それは，量化詞，指示詞，数助詞句，助詞句，形容詞，関係節と同格節（両方合わせて連体修飾節とする）および同格句である。次節からは，以上の八種類の名詞句修飾部の含まれる名詞句の構造をそれぞれ解析していく。

3.3 量化詞，指示詞と数助詞句

Li and Thompson（1981）は，数助詞は数詞，指示詞，量化詞のいずれかと共起しなければならないという理由で，指示詞も量化詞も数助詞句の一部であると見なしている。しかし，（10）に示す通り，（5a）と（6a）のような例の場合，実は数詞“一”が省略された（明示的に現れていない）に過ぎないのである。

（10） 这/每 (一) 个 苹果
指示詞 -この/量化詞 -それぞれ (一) 個 リンゴ
この/それぞれの/リンゴ

中国語では，数助詞は基本的には数詞と共起しなければならないが，指示詞や量化詞の修飾部が同時に現れており，しかも数詞が“一”になっている場合のみ，数詞が省略できるのである。さらに，（11）と（12）から分かるように，指示詞が文頭に現れる時は数助詞句および主要部名詞の省略が可能であり，また量化詞が数助詞句を経由せずにそのまま主要部の名詞を修飾することもできる。

（11） 这 (一 位 小姐) 是 田中。
指示詞 -この (一 名 お嬢さん) である 田中
こちらは田中さんだ。

（12） 所有 学生 都 非常 优秀 。
量化詞 -あらゆる 学生 みな 非常に 優秀だ
全ての学生はみな非常に優秀である。

従って，指示詞と量化詞は数助詞句の一部ではなく，より高い独立性を持つ名詞修飾部であると考えられる。本研究は，指示詞と量化詞にそれぞれ D と Q という品詞タグを与え，（10）（“一”が省略された場合）の統語解析を以下の（13）と（14）のように付与する。

（13） (FRAG (NP (D 这)
(NUMCLP (NUMCL 个))
(N 苹果)))

（14） (FRAG (NP (Q 每)
(NUMCLP (NUMCL 个))
(N 苹果)))

（13）と（14）は類似した構造を取っている。（13）では，名詞句の主要部名詞“苹果”は二つの修飾部を持ち，D は独立した修飾部として扱われている。しかも，前置修飾部が一語からなるものの場合は，句への投射を要求しないので，D は DP を投射していない。一方，NUMCLP は，主要部 NUMCL が，（13）と（14）では明示的に現れていないが，潜在的な CARDP と結びつい

て，修飾部の NUMCLP を投射するものとする。なお，FRAG は断片（fragment）を意味する。

3.4 助詞句と形容詞

本研究における名詞の修飾部としての助詞句は，助詞“的”と組み合わされる要素を名詞のみに限定しないため，Li and Thompson（1981）の分類に見られる連合句に比べて，適用の範囲がより広くなっている。中国語の場合，名詞の他，形容詞，量化詞および前置詞句も助詞“的”と一緒に修飾部助詞句を構築することができる。その例を（15）〜（18）に示す。

（15）　北京大学　的　　学生
　　北京大学　助詞　学生
　　北京大学の学生

（16）　优秀　　的　　学生
　　優秀だ　助詞　学生
　　優秀な学生

（17）　所有　　的　　学生
　　量化詞 -あらゆる　助詞　学生
　　全ての学生

（18）　在　　校　　的　　学生
　　前置詞 -で　学校　助詞　学生
　　学校にいる学生

また，一般的には助詞と組み合わされる要素は助詞の補部と見なされるため，単一語が助詞“的”に前置する場合でも，相応する句への投射が必要である。（15）〜（18）の統語解析は次の（19）〜（22）のようになる。

（19）　(FRAG (NP (PP (NP (NPR 北京大学))
　　(P 的))
　　(N 学生)))

（20）　(FRAG (NP (PP (ADJP (ADJ 优秀))
　　(P 的))
　　(N 学生)))

（21）　(FRAG (NP (PP (QP (Q 所有))
　　(P 的))
　　(N 学生)))

（22）(FRAG (NP (PP (PP (P 在)
(NP (N 校)))
(P 的))
(N 学生)))

（19），（20），（21）では，NPR，ADJ と Q はそれぞれ NP，ADJP と QP を投射してから P の補部となって PP を構築している。（22）では，助詞“的”の補部はもう一つの PP“在校”である。なお，PHC 式の解析規約では，助詞に異なる品詞タグを付与しないため，助詞の“的”といわゆる格助詞の“的”（所有格）との区別はしないこととなる。というよりも，そもそも両者の区別が困難な場合がほとんどである。しかも，意味処理の観点から考えても両者を区別しなければならない理由がない。また，（7）で見られた“的”の省略については，PHC 式の解析スキームなら，省略された主要部の統語特徴は親の句のタイプから推測できるため，統語解析の段階では特に別の工夫をしなくても済む。従って，（7）の“他儿子”の部分に関する統語解析は次の（23）のようになる。PP のヘッドである P が省略されていると考えられるため，NP が PP によって直接支配されている。

（23）(FRAG (NP (PP (NP (PRO 他)))
(N 儿子)))

一方，主要部名詞をそのまま修飾する形容詞に関しては，関係節を構築する必要がないと本研究では考える。自然言語は余剰性を極力排除し，最短の派生ともっとも簡潔な表示を優先させるシステムだと考えられる。本研究のアノテーション方式では，形容詞が主要部名詞を直接修飾することも関係節を構築してから修飾することも可能だが，本研究のアノテーション方式を自然言語の文法処理システムの一種のシミュレーションと見なすのであれば，それは経済性を重視しなければならない。従って，名詞句の内部構造に対しても，より簡潔な表示を取るべきである。であれば，形容詞がわざわざ関係節を構築してから主要部名詞を修飾する必要はないと考えられる。しかも，本研究および本研究の採用している PHC 式の解析スキームはそもそも実用性を重視するものであり，アノテーション作業を煩雑にしてしまうことやアノテーターのミスを増やしてしまうようなことはむしろ極力回避すべきだと思われる。以上を踏まえて，（8）の統語解析を（24）のように付与する。

（24a）(FRAG (NP (ADJ 红)
(N 花)))

（24b）(FRAG (NP (PP (ADJP (ADJ 红))
(P 的))
(N 花)))

形容詞が直接主要部名詞を修飾する時は，ADJ は N の前置単一語修飾部になるため，ADJP への投射を必要としない。これに対して，形容詞と名詞との間に“的”が現れると，ADJ は

ADJP を投射し，さらに P の補部として P と一緒に PP を構築してから N の前置複数語修飾部になる。また，（25）が示すように，形容詞が程度副詞を伴うこともあり得る。なお，この場合は，一種類の存在物の名前になる傾向が極めて低いと思われるため，“的”が不可欠になる。

（25）很　　红　　的　　花
　　とても　赤い　助詞　花
　　とても赤い花

（25）についても，副詞と形容詞とが関係節を構築しているとは見なさない。（25）の統語解析は（26）のようになる。

（26）(FRAG (NP (PP (ADJP (ADV 很)
　　(ADJ 红))
　　(P 的))
　　(N 花)))

ADV が ADJ の前置単一語修飾部であり，両者が ADJP を構築している。

3.5 連体修飾節

主要部名詞の修飾部は，語と句のほか，節となることもある。このような節のことは，日本語学では，連体修飾節と呼ばれることが多い。Li and Thompson（1981）は，主要部名詞を修飾する節のことを名詞化として扱っている。さらに，中国語の名詞化は，助詞“的”が動詞句などに後置するという形を取ると述べている。しかし，実は中国語では，（27）に示すように動詞句（（27）では，“学中文”）がそのまま主語名詞句や直接目的語名詞句として使用されることが多い。

（27）学　　中文　　不　　难　　。
　　学ぶ　中国語　否定　難しい
　　中国語を学ぶことは難しくない。

このような言語の実情から，本研究は日本語学の用語を借用し，主要部名詞句を修飾する節のことを連体修飾節とする。中国語の連体修飾節の判定については，従来補文標識（complementizer）“的”の有無および被修飾部の省略可否を基準として行われるものが多いが，本研究ではより網羅的にデータを考察するために，基本的には必ずしもこの両方を満たさなくてもよいようにしている。ただし，（27）のようなものと区別するには，“的”あるいは被修飾部のどちらかが具現化しなければならない（即ち，両方が同時にゼロ形式を取ってはいけない）という注意をアノテーターに対して与える。なお，（27）は，本研究では，動詞連続構文（serial verb construction）の一種として扱われており，その詳細については，周他（2015）を参照してほしい。連体修飾節は，その被修飾部の名詞が修飾部の中で文法的な役割を果たすか否かによって，内の関係と外の関係に分けられることが一般的である（寺村 1975, 1977a, 1977b）。本研究の場合もその考え方を踏襲

するが，内の関係を関係節（relative clause），外の関係を同格節（appositive clause）とそれぞれ称する。

3.5.1 関係節

関係節とは，被修飾名詞がこれを修飾する節の中で文法的役割を担っているような連体修飾節のことである。中国語の関係節は一般的には被修飾名詞の前に現れており，（28）はその一例である。

（28） 之前 失去 的 机会 又 来 了 。
この前 失う 補文標識 チャンス また 来る 完了
この前失ったチャンスがまたやってきた。

“机会”は，関係節“之前失去的”によって修飾され，しかもその中で直接目的語という文法的な役割を担っている。（28）に対しては以下の（29）のような統語解析を与える。

（29） (IP-MAT (NP-TPC-SBJ (CP-REL (IP-SUB (NP-SBJ *pro*)
(NP-OB1 *T*)
(ADVP (ADV 之前))
(VB 失去))
(C 的))
(N 机会))
(ADVP (ADV 又))
(VB 来)
(AS 了)
(PU 。))

（29）では，関係節は CP-REL というタグが与えられ，その内部に二つの空要素が付け加えられている。「NP-SBJ *pro*」は明示的に現れていない“失去”の主語を示しているのに対して，トレース *T* は直接目的語 NP-OB1 として扱われている。なお，NP-TPC-SBJ に関しては，主題と主語の解析を扱う第 4 節を参照してほしい。（29）を意味処理システムに入力して自動意味評価することにより，（28）の述語論理式は，（30）のように得られる。

（30） ∃ x4 x1 e2 e3 (x4 = pro ∧ 机会(x1) ∧ 之前(e2) ∧
失去(e2, x4, x1) ∧ 又(e3) ∧ 来_了(e3, x1) ∧ topic(e3) = x1)

（30）では，二つの述語“机会”と“失去”は連言結合子∧によって結びつけられており，x1“机会”は“失去”の直接目的語項の位置にも現れている。このように，関係節の場合，修飾部関係節と主要部名詞との意味的関係は並列関係になる。

また，中国語では，関係節が独立して使われることも可能である。しかし，これはあくまでも

関係節の話で，3.5.2 節で論じる同格節には通用しない。つまり，中国語では，主要部ぬきで現れる連体修飾節は必ず同格節ではなく関係節である。さらに，具現化されていない主要部名詞が関係節の内部でどのような文法的な役割を果たしているかについて，Li and Thompson（1981: 577–579）は次のようにまとめている（前述したように，Li and Thompson（1981）は，この種の関係節も含め，主要部名詞を修飾する節のことを全部名詞化として扱っている）。

① To be used alone as a noun phrase, a nominalization must contain a verb with at least one of its participants unspecified.
② If there is only one participant unspecified, then the referent of the nominalization is the same as that of the missing participant.
③ If both the subject and direct object participants are unspecified in a nominalization, then that nominalization will generally be understood to have the same referent as the unspecified direct object participant of that verb.
④ A nominalization used alone as a noun phrase never refers to the indirect object participant.

以上を用いて，（31）～（33）の具体例を分析してみる。

（31）从事　服务业　的　很　辛苦。
従事する　サービス業　補文標識　とても　辛い
サービス業に従事する人がとても辛い。

（32）难民　缺少　穿　的　。
難民　欠ける　着る　補文標識
難民は着るものに欠ける。

（33）卖　的　比　租　的　贵　。
売る　補文標識　前置詞 -より　レンタルする　補文標識　高い
売るものはレンタルのものより高い。

まず，（31）の関係節において，その動詞“从事”は他動詞（二項動詞）であり，しかもその主語が指定されていないため，②により，具現化されていない主要部名詞が（31）の関係節において，主語の文法的な役割を果たしているということが分かる。次に，（32）の関係節における二項動詞“穿”はその項が両方とも未指定のままであるが，③に基づいて主要部名詞が担っている文法的役割は主語ではなく直接目的語であるという結論に至る。最後に，二重目的語他動詞（三項動詞）が含まれる（33）についても，②と③に④を加えることにより，同様に正しい推論が得られる。

このような（①から④までの）一連の規則を意味処理システムにインプリメントすれば，この種の関係節を解析する時に，統語解析の段階でトレースを付けなくても要素間の意味関係が正確に捉えられるはずだが，アノテーションの一貫性や意味処理システムの汎用性（処理できるのは

中国語だけでなく，英語や日本語など全ての言語に関する処理ができるようにしている）への配慮もあり，本研究では，この種の関係節に対しても，トレースのアノテーションを求める。(33) の統語解析は以下の（34）のように行う。

(34) (IP-MAT (NP-TPC-SBJ (CP-REL (IP-SUB (NP-SBJ *pro*)
(NP-OB1 *T*)
(VB 卖))
(C 的)))
(PP (P 比)
(NP (CP-REL (IP-SUB (NP-SBJ *pro*)
(NP-OB1 *T*)
(VB 租))
(C 的))))
(ADJ 贵)
(PU 。))

（34）では，CP-REL によって修飾される二つの主要部名詞はともに明示的に現れていないが，両方の CP-REL に対して，共に「NP-OB1 *T*」のアノテーションを行っている。

(35) ∃ x7 x6 x1 x2 e3 e4 e5 (x6 = pro ∧ 卖(e3, x6, x2) ∧
x7 = pro{x2} ∧ 租(e4, x7, x1) ∧ 贵(e5, x2) ∧ 比(e5) = x1 ∧ topic(e5) = x2)

意味処理の結果 (35) では，二つの個体変項 x1 と x2 (即ち，関係節で省略された被修飾部名詞の項) が追加されており，それぞれ述語“租”と述語“卖”の直接目的語項の位置に来ている。また，x2 は形容詞述語“贵”((33) の主節の述語）の主語項であり，x1 は，前置詞 P の“比”と結びついて PP となって，“贵”と関係づけられている。

3.5.2 同格節

関係節と異なり，同格節によって修飾される主要部名詞は，その中で何の文法的な役割も担当せず，抽象的な概念を示すものが多い。また，前節で触れたように，同格節は主要部名詞の具現化が常に求められており，単独では使用することができない。(36) は前節で見られた関係節構文（28）に対応する同格節構文の例である。

(36) 去 美国 的 机会 又 来 了 。
行く 米国 補文標識 チャンス また 来る 完了
米国に行くチャンスがまたやってきた。

“机会”は抽象名詞で，その内容が修飾部“去美国的”によって補充され，“机会”自体は同格節“去美国的”の中で文法的な役割を果たしていない。(37) は（36）の統語解析である。

(37)　(IP-MAT (NP-TPC-SBJ (CP-THT (IP-SUB (NP-SBJ *pro*)
(VB 去)
(NP-OB1 (NPR 美国)))
(C 的))
(N 机会))
(ADVP (ADV 又))
(VB 来)
(AS 了)
(PU 。))

関係節には CP-REL というタグが与えられたのに対して，同格節に付与するタグは CP-THT になる。また，言うまでもないが，CP-THT の内部にトレースの追加はしない。タグの違いにより，意味処理システムでは両者の識別が可能になる。(37) を入力すると，以下の処理結果が得られる。

(38)　∃ x4 x3 e1 e2 (x4 = pro ∧ 机会(x3, 去(e1, x4, 美国)) ∧
又(e2) ∧ 来_了(e2, x3) ∧ topic(e2) = x3)

関係節の意味解析で見られた並列関係と違って，(38) では，述語“去”に関する意味表示は述語“机会”の意味表示の一部になっている。このように，同格節の場合，修飾部の意味が被修飾部名詞の意味の中に埋め込まれるのである。

3.6 同格句

主要部名詞と同格関係を持つ修飾部は，節の他に，句もある。これは，主に（39）のような再帰代名詞（reflexive pronoun）が代名詞に後置するパターンや（40）のような固有名詞（名前）と普通名詞（肩書き）からなるパターンに見られる。

(39)　我　自己
私　自身
私自身

(40)　冠军　　刘翔
チャンピオン　劉翔
チャンピオン劉翔

同格関係を持つ両者のどちらが名詞句の主要部なのかに関する判断はある意味では容易ではない。とはいうものの，先行研究のほとんど（Li and Thompson 1981，など）は，中国語の名詞句は主要部名詞が常に名詞句の最後の位置に現れると考えている。同格関係も修飾関係の一種と見なせば，(39) と（40）の主要部もそれぞれ“自己”と“刘翔”のように思われる。また，従来は，(39) における“自己”のような再帰代名詞を特別なものとして取り扱うものが多い。そのため，

（39）と（40）では，主要部名詞と修飾部名詞とがみな同格関係を持つとしても，両者の区別がツリーバンクの内部ではっきり表現されることが望ましい。実際のところ，PHC 式の解析規約においても，両者の区別がなされている。第 2 節で紹介したように，このような区別は，常に明示的な機能タグの付与によって実現されている。とはいえ，機能タグを持てるのは句レベルのカテゴリーのみなので，修飾部名詞の名詞句への投射が求められるようになる。本研究のアノテーション方式では，単一語修飾部の句への投射が認められるのは，それがヘッドの後ろに現れる場合に限られる。従って，(39)と(40)に対しては，修飾部を担当する名詞をそれぞれ二番目の“自己”と“刘翔”に設定したほうが適当だと思われる。以上の議論を踏まえて，本研究は，(39）と（40）の統語構造をそれぞれ次の（41）と（42）のように与える。

（41） (FRAG (NP (PRO 我)
(NP-RFL (PRO 自己))))

（42） (FRAG (NP (N 冠军)
(NP-PRN (NPR 刘翔))))

（41）と（42）に示す通り，（39）と（40）の主要部はそれぞれ一番目の名詞の“我”と“冠军”となる。一方，二番目の名詞“自己”と“刘翔”は，どちらも後置修飾部として扱われ NP を投射しているが，NP の機能タグが異なる。NP-RFL は再帰名詞句を，NP-PRN は同格名詞句を意味する。

4. 主題と主語

中国語の主題と主語については，従来研究者たちによって頻繁に取り上げられてきた。主語優勢言語（subject-prominent language）である印欧諸語と対立し，中国語は主題優勢言語（topic-prominent language）だとされている（Li and Thompson 1976）。つまり，中国語は文の主題が統語論的に決まった方法で明示される一方，主語はあまり重視されず，「主題（topic）＋説明（comment）」という構造を愛用する言語だと考えられている。類似した考え方は Chao（1968）も持っている。Chao（1968）は，英語などを分析する時によく使われる「主部（subject）」と「述部（predicate）」は，中国語では，それぞれ「主題」と「説明」として扱ったほうが適切だと主張している。中国語の構文を解析する際に主題が重要な役割を果たしているということは確かだと思われる。このような理由で，本研究では主題という概念を導入し実際に使っていく必要があると考える。

しかし，中国語は孤立語（isolating language）であり，主題や主語を明示的に示す形態上の手掛かりがない。また，本研究では，文の表層の統語構造だけではなく，深層の意味構造も同時に扱わなければならないため，語順を基にして主語などの定義を行うような従来のやり方にも限界がある。このように，主題と主語の定義や両者の区別は決して容易な課題ではないが，中国語の構文をシステマティックに論じる場合は，主題と主語の定義を前もってはっきりしておく必要がある。というのも，主題と主語は，中国語の文の根本的な構成要素であり，場合によっては，構

文の解析を決定する際の決め手になることもあり得るからである。本節では，主題と主語の定義を本研究の研究目的に合わせた形でそれぞれ明確に行った上で，実際の用例を考察していく。

4.1 主題

主題とは，文によって陳述される中心的対象のことを言う。即ち，主題は文がそれをめぐって展開していく，予め決めておいた事柄である。このような特徴から，主題は明示される時には文の最初に来るかあるいは左方移動（left dislocation）によって文頭に移動されることが多い。主題のもう一つの特徴は，常に聞き手がすでに分かっていることまたは総称的な存在物を指すことである。従って，主題は必ず定（definite）か総称的（generic）かのどちらかとなる。

中国語に関しても，世界中の多くの言語と同じように，主題は常に文の最初に出現し，しかも定か総称的かであるとされている（Li and Thompson 1981）。しかしながら，文の最初に現れる名詞句が必ず主題なのかあるいは主題は文の最初にしか来られないのかについては，これまで十分な議論がなされていない。この問題は，主題に関する一貫したアノテーションができるかどうかに緊密に関連しているので，主題の定義に踏み込む前に，まず以上の問題を具体的に検討し明確な結論を出しておきたい。

ある要素がある文法カテゴリーに所属するかどうかを知りたい場合，当の文法カテゴリーの持つ幾つかの属性が問題の要素にも見られるかどうかということを検証する方法が有効である。主題に関しては，これまで通言語的に様々な研究がなされており，それが所有している普遍的な属性（Li and Thompson 1976，など）がある程度見つけ出されているし，また中国語に限定するなら Tsao（1977）の研究もある。そのため，上述の手法はもちろん主題にも適用できると考えられる。つまり，主題が所有すべきだとされる全ての属性が，常に文頭の名詞句に見られる（あるいは文頭の名詞句にしか見られない）ということが証明できれば，文の最初に出現する名詞句は必ず主題であるということが言えるのである。とはいうものの，実際のところ，孤立語である中国語は，主題を判断するための明示的な手がかり（即ち，主題の持つ形式上の属性，例えば，日本語では，主題は形式上助詞「は」によってマークされることが一般的である）がほぼ存在しない（「文頭に来る」ということを形式上の属性として考えることができるが，それはまさに今の考察対象であるため，使うことができない）。また，主題は，通常の文法カテゴリーと違って，ディスコース上の概念でもあるため，その属性の考察は明確に行いにくいことがある。例えば，主題の持っているもっとも基本的な属性の一つは，「陳述の中心的対象」である。しかし，「陳述の中心的対象」であるかどうかに関する判断は個人の主観に頼る部分が多く，客観性に欠ける恐れがある。従って，本研究は，中国語の主題の持つ属性のうち，重要かつ明確な考察が可能だと思われる以下の四つのものを抽出し，この四つの属性が文頭の名詞句に見られるかどうかについて考察していくことにする。

（43a）主題は定か総称的かである。（Li and Thompson 1981）
（43b）主題の直後に，“啊 a (呀 ya)”，“嘛 ma”，“呢 ne”，“吧 ba”という四つの休止助詞（pause

particle）の中のいずれかが挿入可能である。（Tsao 1977）

（43c）主題はディスコース上の概念でもあり，その作用域は常に一つ以上の文まで広がっている。（Tsao 1977）

（43d）主題は主題連鎖（topic chain）範囲内の全ての代名詞化（pronominalization）された名詞句と同一指示関係を持っている。（Tsao 1977）

まずは，以下の（44）を通して，（43a）を検証する。

（44a）书　我　已经　买　　了　。
　　本　私　もう　買う　完了
　　（あの）本は私がもう買った。

*（44b）一　本　书　我　已经　买　　了　。
　　一　冊　本　私　もう　買う　完了

（44a）における“书”は，裸の名詞句であり，冠詞を持たない中国語では理論上定と不定を両方表すことが可能である。それにもかかわらず，文頭に来ている（44a）の“书”は必ず聞き手が特定できるもので定になっている。これに対して，（44b）においては，数助詞句は指示詞ぬきで名詞を修飾しているため，文頭の名詞句“一本书”は不定になっている。（44a）は文法的に正しいのに対して，（44b）は非文になっているという結果から，主題の持っている（43a）の属性を，文頭の名詞句は持っているということが言える。

次に，（43b）について考える。中国語には主題マーカーが存在していないとされてきたが，話し言葉の場合，主題の直後に（43b）に見られる四つの休止助詞を入れることが可能だと Tsao（1977）は述べている。書き言葉の場合，休止助詞を入れることはまずないが，形式上のヒントに乏しい現状では，少なくともそれは主題を判断する際の一つの参考になると考える。以下，休止助詞を入れることのできる位置を確認する。

（45a）书　啊(呀)/嘛/呢/吧　我　已经　买　　了　。
　　本　休止助詞　　私　もう　買う　完了
　　本は私がもう買った。

*（45b）书　我　啊(呀)/嘛/呢/吧　已经　买　　了　。
　　本　私　休止助詞　　もう　買う　完了

文頭の名詞句“书”の後ろに休止助詞を付けても問題がないのに対し，文中の名詞句“我”に休止助詞を付加してはいけないということが，（45）から分かった。さらに，紙幅の都合上（45）の一例しか挙げられなかったが，他の例に関する考察によって，中国語の場合，文頭に来る名詞句が述語に対してどのような意味的役割を果たしていようが，文の述語は何項述語であろうが，休止助詞が出現できるのは，常に文頭の名詞句の直後の位置であるということも明らかにすることができる。従って，主題の持つ（43b）の属性は，文頭の名詞句も持っているということとなる。

続いて，(43c) を検討する。(43c) はディスコース上の属性であり，曖昧な部分もあるが，同じ文脈の下で類似した一連の発話ができるかどうかということを考察することによって，ある程度はっきりさせることも可能だと思われる。

(46a) 我 书 已经 买 了 。 报纸 也 已经 买 了 。
私 本 もう 買う 完了 新聞 副詞 -も もう 買う 完了
私は本をもう買った。新聞ももう買った。

?(46b) 我 书 已经 买 了 。 李四 也 已经 买 了 。
私 本 もう 買う 完了 李四 副詞 -も もう 買う 完了

(46a) と (46b) の前半は全く同一であり（従って，後半は同一の文脈にあると考えられる），両者の唯一の違いは後半の文の最初の位置に現れる名詞句である。(46a) は問題がないが，(46b) は多少不自然な発話になっている。このような違いが生じる理由を考えると，分かってくることがある。つまり，後半の文までその作用域が届いているのは，文中の名詞句“书”ではなく，文頭の名詞句“我”である。(46a) では，文頭の“我”の影響は後半の文にも及んでいる。説明の便宜上その結果を，後半の文の最初に“我”を付け加えた形（即ち，“我报纸也已经买了。”）に単純化する。そうすると，(46a) に関する解釈は，「私は本をもう買った。私は新聞ももう買った。」というように行うことができる。このような解釈は意味上問題が生じないので，自然な発話になる。一方，(46b) でも，後半の文までその影響が行き渡っているのは，文頭の名詞句“我”である。従って，(46b) の後半を単純化した結果は“* 我李四也已经买了。”になるが，(46b) 全体の解釈は「* 私は本をもう買った。私は李四ももう買った。」になってしまう。これは明らかに意味的にはおかしいので，(46b) に対しては不自然さが感じられるのである。(46b) を自然な発話に直すためには，(46c) に示すように，前半の文における“书”と“我”の順番を逆にすればよい。

(46c) 书 我 已经 买 了 。 李四 也 已经 买 了 。
本 私 もう 買う 完了 李四 副詞 -も もう 買う 完了
本は私がもう買った。李四ももう買った。

このように，中国語の場合，主題の持つ (43c) の属性は，文頭の名詞句からも確認できた。

最後に，(43d) について考える。(43d) のチェックは，(43c) と類似したやり方で行うことができる。

(47a) 这 本 书 李四 看 了 三 遍 还 不 懂 。
指示詞 -近称 冊 本 李四 読む 完了 三 回 まだ 否定 分かる

它 实在 太 难 了 。
代名詞 -三人称無生命 実に あまりにも 難しい 文末助詞

この本は李四が三回も読んだが，まだ分からない。それは実に難しい。

?(47b) 这 本 书 李四 看 了 三 遍 还 不 懂 。
指示詞 -近称 冊 本 李四 読む 完了 三 回 まだ 否定 分かる

他 实在 太 笨 了 。
代名詞 -三人称男 実に あまりにも 頭が悪い 文末助詞

(47a) と (47b) は，前半の文には相違が見られず，両者の後半は同一の文脈の下にあると言える。後半の文は構造的には非常に似ているにもかかわらず，(47a) と (47b) との間に容認度の違いが見られる。(47a) と (47b) は，統語上にも意味上にも問題がないので，このような現象が起きる理由は発話の論理上のものだと考えられる。今，(47) における代名詞をコントロールしているのは文の最初に現れる名詞句であると仮定した上で，(47a) と (47b) についてもう一度考える。“它 tā” は無生命のものしか指すことができないため，“它” の指示対象の可能性は，“这本书” に限定されることとなる。“这本书” は前半の文の最初に出現しているので，これはわれわれの仮定と一致している。一方，“他 tā” は，“它” と同じ発音であるにもかかわらず，三人称男性の代名詞である。そのため，“他” と同じ指示対象を持つのは，“李四” に限られる。しかし，“李四” は前半の文の文頭ではなくその文中に現れているので，われわれの仮定と矛盾することになる。(47a) は自然な発話であるのに対して，(47b) には多少の不自然さが感じられる (特に表記なしで, 音声だけ聞いた時に) という事実から, われわれの仮定が妥当だと考えられる。さらに，(47b) から不自然さが感じられる理由としては，以下のように説明することができる。母語話者は，“这本书李四看了三遍还不懂。” という文脈を受け取った状態で “tā” と聞いたら，母語中国語に関する知識によって，これは “这本书” のことを言っているのではないか (即ち，“tā” の正体は “它” である) と頭の中で推測しているにもかかわらず, 次の発話の内容から “tā” の指示対象は実は “李四” である (即ち，“tā” の正体は “他” である) ということが判明する。そのため，論理上に矛盾ができ，不自然さがもたらされたのである。(47c) は，(47b) に対して，前半の文における “这本书” と “李四” との順番を入れ替えたものである。

(47c) 李四 这 本 书 看 了 三 遍 还 不 懂 。
李四 指示詞 -近称 冊 本 読む 完了 三 回 まだ 否定 分かる

他 实在 太 笨 了 。
代名詞 -三人称男 実に あまりにも 頭が悪い 文末助詞

李四はこの本を三回も読んだが，まだ分からない。彼は実に頭が悪い。

このようにすると，不自然さが生じなくなる。このことから以上の説明の正確性を裏付けることができる。これまでの議論を踏まえると，中国語では，主題の持つ (43d) の属性も文頭の名詞句に当てはまるということが言える。

以上主題の持つべき四つの属性は，文頭の名詞句でも全て確認できた。勿論，これによって主題に関する属性の全てを検証したというわけではないが，検討の過程を通して，この四つの属性

を併せ持つことができるのは，文頭の名詞句の他にはないということが分かった。

以上のような主題の持つ属性のチェックの他に，本研究では，先頭名詞句が疑問詞で現れうるか否かに関する考察も行った。疑問詞は話題の既知性（少なくとも存在前提を要求する）に反しており，通常話題化されないとされている。中国語の場合，文頭の位置は主題のためのものだとすれば，そこに疑問詞が現れることがほとんどあり得ないというように推論できると考えられる。この推論が中国語の使用実情に合うかどうかということを検証するために，本研究では，Penn Chinese Treebank (7.0) のデータを対象に調査を行った。その結果，51,447 文（コーパスに含まれている文の総数）中，文の最初の位置に疑問詞が現れる例は 12 文しかなく，その割合が僅か 0.0002 になっているということが分かった。さらに，この 12 文の大部分は，特定の文脈が前提とされる（答えが選択される対象の集合が前提化されている）ものであり，それ以外は次の（48）のようなものである。

（48）　谁　都　　会　　犯　　错　　。
　　誰　みな　助動詞 -可能性　犯す　間違い
　　誰も間違いを犯すものだ。/ 人はみな間違いを犯すものだ。

（48）は疑問文ではない。その中の“谁”は全ての人を指しており，総称的だと言える。これは主題の持つべきだとされる性質に合致しているため，むしろ（43a）を実証できる例だと考えられる。

以上の考察結果を踏まえて，中国語の場合，文頭に現れる名詞句は通常主題であるという結論がある程度適切だと考えられる。しかし，先行研究のほとんどはごく一部の中国語データしか考察していない。これは従来の言語学研究の抱える普遍的な問題でもあり，本研究で構築しているコーパスはそれを解決するのに重要な役割を果たすことができると期待される。まだ構築中とはいえ，実際にアノテーション作業を進めていくと，今までのような主題を文頭に現れる名詞句に限定する考え方はやはり不十分であるということが分かってきた。主に，以下の四点がある。

まず，中国語では，3.5 節で見られた（27）のような例も観察されている。（27）では，文の最初に現れるのは名詞句ではなく，一つの名詞節“学中文”である。“学中文”の直後に“啊”などの休止助詞を挿入でき，この節自体は総称的だ（中国語を学ぶこと全般を指す）と考えられるため，主題に見られるほとんどの属性を持つことになる。従って，“学中文”は名詞句ではないにもかかわらず，（27）の主題であるということが言える。

また，2.1 節に現れた（1）が示すように，中国語の場合，文の最初の位置は間投詞や接続詞により占められることもある（特に，書き言葉の場合は接続詞の例が多い）。（1）における“张三”は文頭には現れていないが，それが文の主題であるということは確実だと思われる。

さらに，中国語の生データをアノテーションする際に，次の（49）のような例に出会うことも少なくない。

（49）　今年　年夜饭　　就　　吃　　火锅。
　　今年　年越し料理　副詞 -強調　食べる　火鍋

今年，年越し料理は火鍋を食べる。

（49）では，文の最初に現れるのは時間名詞句の“今年”である。“今年”が（49）によって陳述される中心的対象であるかどうかはともかくとして，もし“今年”を文の主題として扱うと，“年夜饭”に分配できる統語的役割（即ち，機能タグ）の選択肢がなくなってしまう恐れがある。なぜなら，“年夜饭”は明らかに“吃”の目的語ではない（“吃”の目的語は“火锅”である）し，その主語にもならないと考えられる（本研究における主語の定義については，次の 4.2 節を参照のこと）からである。このように，（49）のような時間名詞句が文頭に現れる例の場合，時間名詞句を主題として扱わないほうが望ましい。そもそも，（49）によって述べられる中心的対象は，“今年”というよりも，“年夜饭”のほうが適切であろう。

最後に，文（主節）だけではなく，直接引用の節にも主題が出現できると考えられる。（50）は（49）を直接引用したものである。

（50）　妈妈　说　：“ 今年　年夜饭　就　吃　火锅。”。
母　言う　今年　年越し料理　副詞 -強調　食べる　火鍋
母は「今年，年越し料理は火鍋を食べる。」と言った。

直接引用の節は，PHC 式の解析規約では，文の下のカテゴリーとして扱われているが，文と同じほど高い独立性を持っていると考えられる。従って，直接引用の節に対しても，基本的には文と同じようなアノテーションができることが望ましい。本研究は，直接引用の節と間接引用の節を区別し，それぞれ CP-FINAL と CP-THT というタグを与える。（50）における“年夜饭”は文の最初には現れていないにもかかわらず，CP-FINAL の主題として扱われるべきである。

先行研究を踏まえ，以上の四点も考慮した上で，本研究では，主題の定義を次のように行う。

主題：
主題とは，IP-MAT または CP-FINAL における，述語の前に現れる，時間名詞句以外の一番目の名詞句あるいは名詞節のことである。

なお，主題はそもそもディスコース上の概念でもあり，そのスコープは文のレベルを超えることもある（（46a）と（46c）を参照）が，本研究では，アノテーションが文単位で行われているため，文の範囲を超えた関係などはアノテーションの対象としない。つまり，（46a）や（46c）を実際にアノテーションする際には，それを二つの文に分けてそれぞれ単独に扱い，個々の文はそれなりの主題を持つことになる。本研究は，主題名詞句のタグを NP-TPC とし，第 2 節で紹介したアノテーション方式を用いて，（44a）の統語構造を（51）のように付与する。

（51）(IP-MAT (NP-TPC (N 书))
(NP-OB1 *)
(NP-SBJ (PRO 我))
(ADVP (ADV 已经))
(VB 买)
(AS 了)
(PU 。))

文頭に来る“书”は主題として NP-TPC を投射し，直接 IP-MAT の支配下にある。さらに，NP-TPC のすぐ次に「NP-OB1 *」も付けてある。これは意味処理に配慮するためのものである。主題は統語レベルの概念ではあるが，意味のレベルまで進むと，動作主，対象，経験者，時間，場所など様々な意味役割を担うことができる。即ち，主題の場合，その統語役割と意味役割は一対一の関係にはなっていない。従って，統語解析の段階で主題のことを NP-TPC としかアノテーションしないと（より詳しい情報の提供が本研究の枠組みの中で困難な場合（例えば，（49）における“年夜饭”）を除いて），意味処理によって情報の喪失が発生する恐れがある。情報を減らすためのアノテーションでは絶対ないので，このようなことはどうしても避けたい。大事な言語情報のロスがないように，本研究は，統語解析の段階で主題に対して，（51）のような付加的なアノテーションも行う。（51）を意味処理システムに入れると，出力として（44a）に関する論理意味表示を以下の（52）のように得ることができる。

（52）∃ x3 x1 e2 (已经(e2) ∧ 书(x1) ∧ x3 = 我 ∧
买_了(e2, x3) ∧ arg1(e2) = x1 ∧ topic(e2) = x1)

（52）では，式の最後の部分には新たな述語 topic が追加され，結果として，x1“书”が e2“买 _ 了”の主題であることが示されている。それと同時に，x1“书”は述語“买”に関する述語論理式の直接目的語項の位置には来ていないが，述語 arg1 を新たに作ることによって，“书”と“买”との間の意味関係（即ち，“书”は“买”の直接目的語項であること）が明示されている。これも，（51）において，主題に対して，「NP-OB1 *」というような付加的なアノテーションを実施したことにより可能となる。

4.2 主語

中国語の場合，文の主題は統語論的に決まった方法（即ち，一般的には文の最初に来ること）で明示される一方，主語（subject）にはそれを見付け出すための形式上の明示的な手掛かりは存在しない。この意味では，確かに先行研究の言う通り，中国語では主語はあまり重視されていないのかもしれない。しかしながら，主語という概念は中国語にとってもやはり重要だと思われる。その理由はいろいろあるが，本研究の観点からは，通常明示的に現れていない従属節の主語をコントロールできるのは主題ではなく（Li and Thompson 1976）主節の主語であるということが挙

げられる。中国語に見られる様々なコントロール構文（動詞連続構文）を解析する（その詳細は，周他（2015）を参照）ためには，主語は不可欠である。

形式上の明示的な手掛かりがない場合，言語の表面に止まらずより深いレベルのものに注目するやり方は時に有効だと言われている。例えば，統率・束縛理論では，主語や目的語の文法関係はそれらの要素が表層構造（s- 構造）において占める統語上の位置（即ち，構造的な位置関係）で決まっていると考える。つまり，主語は表層構造において IP の指定部に位置するものだという主張である。このような統語論的な定義方法は，言語学的価値があるものとされており，主語の概念をある程度一貫して規定することができるという点にメリットがある。Greenberg（1963）が提案した自然言語の語順に関する三つのタイプ（VSO，SVO，SOV）のどれにも属していない中国語（Li and Thompson 1981）にとっても有効なように思われる。

しかし，本研究の採用するアノテーション方式はそもそも統語論的な定義方法には向いていない。PHC 式の解析規約のもっとも大きな特徴は，ツリーがフラットなことである。階層が少ないということは，構文の検索などに便利で実用性が高い一方，階層的位置関係で統語要素を決めるような統語論的な方法は使いにくくなる。構造的な位置関係で主語や目的語を決めようとするやり方を取るためには，一つの条件をクリアする必要がある。それは，自分の理論の枠組みの中で，主語や目的語の占める統語上の位置はこうならなければならないということを証明することである。しかし，実際のところ，不都合な用例（その言語の基本語順に合わないような文など）を扱うために，言語学者は意図的に様々な操作（例えば，かき混ぜ（scrambling））を仮定する傾向がある。これはある意味ではやむを得ないことだろうが，このような恣意的な操作を行ってしまうと，場合によっては理論の一般性を失ってしまい，まさに自分の望ましい結果に合わせて論証の過程を巧みに操っているかのように見える。つまり，主語が IP の指定部にしか現れることができないというよりも，自分が主観的に主語だと認定しておいた要素を IP の指定部の位置に出現させることになる。第 2 節で紹介したように，本研究のアノテーション方式では，階層関係の代りに，句に明示的な機能タグを付与している。従って，IP の指定部に現れる要素の統語情報は，本研究においては，それに機能タグ SBJ を付けることによって表出されることになる。このように考えると，「主語は表層構造において IP の指定部に位置するものである」という考え方は，本研究では「主語は SBJ によってマークされるものである」となる。言うまでもないが，これはコーパス構築の段階では意味のあるものだとは思えない。このように，コーパスの構築を目標とする本研究の場合は，研究者（またはアノテーター）の解釈が入った二次的なものを極力回避し，言語データ自体に頼って一次的な定義を行うことが望ましいと考えられる。

主語の定義を明白に行うために，ここで本研究の全体像をもう一度確認する。本研究は表層的な統語情報しか文に付与しない。一方，深層情報を捉えるためには，表層的な統語情報を入力とする意味処理を行う。述語の格フレームに関する情報は，本研究では，述語－項構造を通して表出されているが，述語－項構造における主語項（arg0; 最初の項）の位置に来る要素は，一般的には述語の動作主（agent）であると考えられている。能動文の場合，述語の動作主は文法関係上の文の主語となるため，SCT では，統語情報を意味情報に変換する際に，述語－項構造の主

語項位置に移されるのは，統語解析の段階において SBJ という機能タグが付与された要素である。これに対して，受身文の場合，文法関係上の文の主語は述語の対象（theme）や経験者（experiencer）に変わるため，統語解析の段階において SBJ という機能タグが与えられたものは，述語－項構造の直接目的語項（arg1; 二番目の項）の位置かあるいは間接目的語項（arg2; 三番目の項）の位置に移動される。このような事実を踏まえて，本研究は主語に対して次のような定義を与える。

主語：
能動文の主語とは，意味表示において，述語に対する最初の項（arg0）に対応する名詞句あるいは名詞節のことである。
受身文の主語とは，意味表示において，述語に対する二番目の項（arg1）かまたは三番目の項（arg2）に対応する名詞句あるいは名詞節のことである。

以上の定義は，一見二次的なもののように見えるが，アノテーターが実際にアノテーションをする時に，これをさらに単純化して「SBJ という機能タグを付与すべきものは，能動文においては，述語との間に「〜がする」あるいは「〜が〜である」という関係が成立する要素であり，受身文においては，述語との間に「〜がされる」という関係が成立する要素である」というように考えることができる。これだと個人的な解釈はほぼ必要としないため，アノテーターは一貫して主語を見付け出すことが可能である。前節で見られた（44a）の場合，“我”は述語“买”との間に「〜がする」という関係が成り立っているので，その統語解析である（51）においては，文の主語として NP-SBJ を投射している。さらに，その意味処理の結果である（52）では，個体変項 x3“我”は，“买”に関する論理式の最初の項の位置に現れている。

4.3 主題・主語の含まれる文の解析

以上の 4.1 節と 4.2 節では，本研究における主題と主語の定義をそれぞれ明確にした。本節では，以上の定義に基づいて，主題・主語の含まれる文の解析を考察する。

4.3.1 主題と主語が両方含まれている文

既に見てきた（44a）のような主題と主語が同時に現れる文に対して，本研究で行った定義を用いることにより，両者の区別を簡単に行うことができる。また，中国語には，（53）と（54）のような二重主語構文（double-subject constructions）と呼ばれる例もある。

（53）　象　鼻子　长　。
　　象　鼻　　長い
　　象は鼻が長い。

（54）　朋友　老　　的　　好　。
　　友人　古い　助詞　よい
　　友人は古いのがよい。

以上のような例に対して，従来の研究では，主語を二つ設定した上でその構造を論じるものが多い。例えば，Teng（1974）は，いわゆる二重主語構文の述部 VP を文 S で書きかえることができると提案している。従って，（53）の主部と述部はそれぞれ“象”と“鼻子长”になる。（53）の述部に当たる“鼻子长”自体は S であり，その主部は“鼻子”によって担われることになる。このような分析の妥当性はともかくとして，本研究で提示した主題と主語の定義を用いると，（53）と（54）は主題と主語が両方含まれている文として分析することができる。つまり，“象”と“朋友”は主題で，“鼻子”と“老的”は主語であるとそれぞれ考えられる。この種の例の特徴は，主題と主語との間の特別な意味関係にある。即ち，“鼻子”は“象”の一部であり，“老的”は“朋友”の部分集合である。Li and Thompson（1981）は，このような関係を部分と全体（part-whole）にまとめている。このように考えると，いわゆる二重主語構文は主題と主語が両方含まれている文に属する特殊なタイプに過ぎないのである。

無論，（53）と（54）のような例を二つの主語が同時に現れる構造として扱うことも可能である。しかし，本研究の主題と主語に関する定義を一貫させるためには，二重主語を主題プラス主語として扱ったほうが望ましいと考えられる。即ち，二番目の主語と述語との間に「〜がする」あるいは「〜が〜である」という関係が成り立っているというのは確実だが，このような関係が一番目の主語と述語との間にも成立しているかについては，微妙なところがある。特に，二番目の主語が一番目の主語の部分集合であるような例の場合はこのことは一層問題となる。例えば，（54）では“朋友”と“好”との間に「〜が〜である」という関係が必ずしも成立しているとは限らないと考えられる。以上を踏まえて，本研究はいわゆる二重主語構文を全部主題と主語が両方含まれる構文として扱い，（53）に（55）のような統語解析を与える。意味処理の結果は（56）に示す。

（55） (IP-MAT (NP-TPC (N 象))
(NP-SBJ (N 鼻子))
(ADJ 长)
(PU 。))

（56） ∃ x1 x2 e3 (象(x2) ∧ 鼻子(x1) ∧ 长(e3, x1) ∧ topic(e3) = x2)

4.3.2 主題と主語とが同一である文

既に幾つか見てきたように，これまで提示した主題と主語の定義を両方満たすような要素が含まれる文も数多く存在している。

（57） 田中 看 过 越剧。
田中 見る 経験 越劇
田中は越劇を見たことがある。

（57）では，“田中”は，主節レベルにおける一番目の名詞句であり，しかも述語“看”の前に現れているため，主題となる。また，“田中”と動詞“看”との間に「〜がする」という関係が成

立するため，“田中”は文の主語でもある。このように，本研究の枠組みの中では，主節レベルの一番目の名詞句にも当たる主語は文の主題としても扱われることになる。（57）の統語構造は（58）のように付与される。

（58）(IP-MAT (NP-TPC-SBJ (NPR 田中))
(VB 看)
(AS 过)
(NP-OB1 (N 越剧))
(PU 。))

（58）では，文の最初の名詞句には，TPC と SBJ という二つの機能タグが付与されている。これにより，“田中”は，“看”の主語であるとともに文の主題でもあるという情報が正確に捉えられている。（58）を入力とする意味処理の結果は（59）に示す。

（59）　∃ x1 e2 (越剧(x1) ∧ 看_过(e2, 田中, x1) ∧ topic (e2) = 田中)

（59）では，“田中”は e2 の主語項の位置に現れている。それに加えて，topic という関数によって，“田中”と e2 との意味関係がより詳しく捉えられている。

4.3.3 主題しか含まれていない文

中国語には，主題は含まれているが，主語は明示的に示されていないような文もある。（60）はその一例である。

（60）　论文　发表　了　。
論文　発表　完了
論文は発表された。

（60）における“论文”は必ず特定的で既知のものである。そのため，それが主節レベルの一番目の名詞句（即ち，主題）の位置に来ることができる。また，（60）は受身文ではないため（中国語の受身文は“被”などの受身マーカーを含まなければならないとされている），“论文”と述語“发表”との間に「〜がされる」の関係が成立するにもかかわらず，“论文”は文の主語ではない。さらに，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，主語などの動詞の必須格が文中で明示的に表現されていない場合，本研究はゼロ代名詞（pro（small pro））の追加を行っている。なお，PRO（big PRO）は本研究のアノテーション方式では使われていない。（60）の統語解析は次の（61）の通りである。

(61) (IP-MAT (NP-TPC (N 论文))
(NP-OB1 *)
(NP-SBJ *arb*)
(VB 发表)
(AS 了)
(PU 。))

(61) では，主題名詞句 NP-TPC に関する付加的なアノテーションの次に主語のゼロ代名詞「NP-SBJ *arb*（一般的非人称指示に用いるゼロ代名詞）」が追加されている。(61) を入力とする意味処理の結果は (62) に示す。

(62) ∃ x1 e2 (论文(x1) ∧ 发表_了(e2, _) ∧ arg1(e2) = x1 ∧ topic(e2) = x1)

(62) では，e2“发表 _ 了”の主語項の位置は空白になっており，x1 と e2 との意味関係は，arg1 および topic という二つの述語によって表出されている。

4.3.4 主語しか含まれていない文

この種の文には，主語に当たる要素は存在するが，主題は含まれていない。

(63) 进来 了 一 个 人。
入る 完了 一 個 人
一人の人が入った。 (Li and Thompson 1981)

Li and Thompson (1981) は，(63) のような文を導入文 (presentative sentence) と称し，その役割は不定名詞句をディスコースの中に導入することだと説明している。能動文である (63) では，“一个人”は“进来”との間に「〜がする」という関係が成り立っているので，文の主語である。しかし，中国語の場合，数助詞句修飾部だけを持つ名詞句は通常不定とされるため，“一个人”は動詞“进来”の前に来ることができない。つまり，“一个人”は文の主語にはなれるが，主題にはなれない。(63) の統語解析および意味処理の結果をそれぞれ (64) と (65) に示す。

(64) (IP-MAT (VB 进来)
(AS 了)
(NP-SBJ (NUMCLP (CARDP (CARD 一))
(NUMCL 个))
(N 人))
(PU 。))

(65) ∃ x1 e2 (人(x1) ∧ 一 个(x1) ∧ 进来_了(e2, x1))

4.3.5 主題も主語も含まれていない文

この種の文は，一般的には，文頭に来る主語（この場合は主題でもある）が省略された時に見られる。中国語の場合，主節レベルの主語の省略が可能なケースは二つある。まずは，談話の文脈から主語の予測が確実にできるのであれば，主語の省略が認められる。これは質問文の応答に生じることが多い。主語の省略が可能なもう一つのケースは命令文である。命令文の場合，主語を言い出さなくても，それは一般的には聞き手を指すということが自明である。（66）は命令文の一例である。

（66）吃　　药！
　　食べる　薬
　　薬を飲め。

（66）では，述語“吃”との間に「〜がする」の関係が成り立つような要素が現れない。また，“药”と“吃”との間に「〜がされる」という関係が成立しているが，（66）は明らかに受身文ではないため，“药”は文の主語ではない。このように，命令文の実際の主語は通常明示的に現れないのである。（66）の統語構造を（67）のように付与する。

（67）(IP-IMP (NP-SBJ *hearer*)
　　(VB 吃)
　　(NP-OB1 (N 药))
　　(PU !))

他の文と区別するために，（67）では主節のタグは通常の IP-MAT ではなく，IP-IMP（命令節）となっている。また，主語のゼロ代名詞として追加されたのは「NP-SBJ *hearer*」であり，これは明示的に現れない主語が聞き手であるということを表す。意味処理の結果は(68)の通りである。

（68）∃ z3 (z3 = hearer ∧ IMPERATIVE(∃ x1 e2 (药(x1) ∧ 吃(e2, z3, x1))))

統語解析の段階で主節は IP-IMP というタグが付与されたため，（68）においては，e2 全体は IMPERATIVE の中に置かれるようになっている。

5. 終わりに

本研究の枠組みの中で中国語名詞句の構造を解析し，主題と主語に関連する問題を取り上げて名詞句の統語的役割の区別を行った。名詞句の解析を通して，PHC 式の解析規約の特徴がよく観察できる。実用性重視という原則を徹底した当解析規約では，中間レベルの構造はいかなる場合も明示的にタグ付けしないことや，一部のヘッド（句を投射しても常に二股分岐しないようなもの）および前置単一語修飾部に対して句への投射を要求しないことなど，句の構造を極力シンプルなままに保つための一連の工夫を施した。その結果，句の内部階層が激減し，句の構造は平坦で複数の枝別れノードを持つようになった。名詞句は，主要部名詞の修飾部が多く，しかも前

置単一語修飾部がその中で大きな割合を占めている。そのため，PHC 式の解析規約を用いて名詞句を解析すると，元々複雑だった名詞句の構造を最大限に単純化することが可能である。

また，名詞句は文の中で多彩な統語的役割を果たしているが，それ故にそれらの区別が困難な場合もある。主題と主語はその典型的な一例である。中国語の場合，主題と主語は双方とも文を構築する根本的な要素であり，中国語の各構文の解析を決める際に，主題と主語に与える定義は判断の要になることが多い。本研究で行った主題と主語に関する定義は，個人的な解釈の排除に尽力したものであり，シンプルな割には実用面に優れているとはいえ，言語学の視点から見ると，必ずしも客観的に十全なものではない。しかしながら，本研究の目標は，主題や主語という統語的要素を従来の言語学的な手法でより一般的に定義することよりも，本研究で「主題」や「主語」と見なされるものを一貫してアノテーションしていくことにある。タグ付きコーパスは，開発者の個人的な解釈が多少含まれるという宿命的な弱点を持っている。それにもかかわらず，筆者たちの目指すコーパスの有効な利用方法は，目標のデータ（少なくとも，目標のデータに近いデータ）を正確に抽出し，そして自分の必要に応じてさらなる加工をしてから利用することである。この方針の下で考えると，コーパスを作る際には，統語的要素に関する扱い方がその客観性や正確性はもちろん一貫性も重視すべきだと考えられる。特に，形態的な手掛かりの少ない中国語に主題と主語という二つの常用概念をタグとして明示的に導入し，それらを一貫してアノテーションし統語・意味情報付きコーパスを構築しようとすることの意義は大きいと思われる。

参照文献

Butler, Alastair (2010) *The semantics of grammatical dependencies*. Bingley: Emerald.

Butler, Alastair (2015) *Linguistic expressions and semantic processing: A practical approach*. Switzerland: Springer International Publishing.

Chao, Yuen Ren (1968) *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chomsky, Noam (1981) *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam (1982) *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Greenberg, Joseph H. (1963) Some universals of grammar with particular reference to the order of meaningful elements. In: Joseph H. Greenberg (ed.) *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Li, Charles N. and Sandra A. Thompson (1976) Subject and topic: A new typology of language. In: Charles N. Li (ed.) *Subject and topic*, 457–489. New York: Academic Press.

Li, Charles N. and Sandra A. Thompson (1981) *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Santorini, Beatrice (2010) Annotation manual for the Penn Historical Corpora and the PCEEC (Release 2). Technical report, Department of Computer and Information Science, University of Pennsylvania.

Teng, Shou-Hsin (1974) Double nominatives in Chinese. *Language* 50: 455–473.

寺村秀夫（1975）「連体修飾のシンタクスと意味」『日本語・日本文化』4: 71–119.

寺村秀夫（1977a）「連体修飾のシンタクスと意味：その 2」『日本語・日本文化』5: 29–78.

寺村秀夫（1977b）「連体修飾のシンタクスと意味：その 3」『日本語・日本文化』6: 1–35.

Tsao, Feng-fu (1977) A functional study of topic in Chinese: The first step toward discourse analysis. Unpublished doctoral dissertation, University of Southern California.

Xue, Nianwen and Fei Xia (2000) The bracketing guidelines for the Penn Chinese Treebank (3.0). Technical report, Institute for Research in Cognitive Science, University of Pennsylvania.

周振・アラステア＝バトラー・吉本啓（2015）「意味処理を目的とする中国語構文解析の研究：動詞連続構文を中心に」*Studies in Language Sciences: Journal of the Japanese Society for Language Sciences* 14: 227–252.

Development of a Parsed Corpus: On the Analysis of Chinese Noun Phrases

ZHOU Zhen[a]　YOSHIMOTO Kei[b]

[a]Tohoku University / Project Collaborator, NINJAL
[b]Tohoku University / Invited Professor, Theory & Typology Division, Research Department, NINJAL

Abstract

This paper focuses on the annotation of Chinese noun phrases as part of a broader endeavor to develop a parsed corpus with both syntactic and semantic information for Chinese characters. For this purpose, the following two notable issues are considered: 1) clarifying the internal structure of noun phrases with a formal analysis; and 2) distinguishing the syntactic roles played by noun phrases. As a frequently used phrase type supporting various modifiers, it is not easy to conduct a consistent analysis on noun phrases. Under a detailed but practical classification of noun phrase modifiers, a comprehensive observation of noun phrase structures with a full range of modifiers is undertaken following a scheme adapted from the *Annotation manual for the Penn Historical Corpora and the PCEEC* (Santorini 2010). As an isolating language, Chinese lacks formal clues that can explicitly indicate the syntactic roles played by noun phrases. The distinction between topic and subject has been considered as an open issue. Following our aim of developing a parsed corpus, a precise definition of topic and subject will be provided. Based on these definitions, we clarify the criteria for determining the topic and subject by considering various examples. This lays an important foundation for further studies into the grammatical structure of Chinese sentences by establishing a solid foundation for exploring further construction types of Chinese within a framework that is instantiated by an actively queried parsed corpus.

Key words: corpus, Chinese, syntactic analysis, topic, subject