様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ６ 月 １４ 日現在

機関番号：３７５０３
研究種目：若手研究
研究期間：2018～2020
課題番号：１８Ｋ１２４４０

研究課題名（和文）中国語学習を支援するためのデータベースの構築

研究課題名（英文）Building a Database for Chinese Learners

研究代表者
周　振（Zhou, Zhen）

立命館アジア太平洋大学・言語教育センター・講師

研究者番号：００７９２９３８

交付決定額（研究期間全体）：（直接経費）　　1,700,000円

研究成果の概要（和文）：研究期間中は主に中国語データベースの構築および外国語教育の視点から中国語の各構文に関する再考察を行った。これまでのアノテーションマニュアルをさらに進化させ、中国語データの解析・修正を続けてきた。構築されたコーパスの中には中国語文に関する統語・意味解析情報が含まれている。外国語教育の視点から中国語の各構文に関する再考察は、各文法項目間の関連性や中国語全体の特徴に十分な注意を払った上で行った。なお、考察を進めていくうちに、基本の文法用語に対する従来の定義を再検討する必要性も見えてきた。さらに、考察結果を踏まえて、研究用データから教育用データへと変換するためのプログラムも作成した。

研究成果の学術的意義や社会的意義
本研究は中国語学と中国語教育の両面において意義があると考えられる。確立できた中国語各構文に関する解析法やコーパスを構築する際に蓄積してきたノーハウは、中国語学やコーパス言語学の研究に役立つと考えられる。また、データベースには中国語の統語・意味情報が多く含まれており、開発した変換プログラムを使うとこれらの研究用データをより学習者に親しみやすい形に変換することもできる。外国語教育の視点から中国語の各構文に関する再考察を通して積み重ねてきた研究成果は、中国語教育の現場だけではなく、新たな中国語教材の開発にも有益な示唆を与えられるものであると考えられる。

研究成果の概要（英文）：During the research period, I focused mainly on building a Chinese database and reconsidering each Chinese construction from the perspective of foreign language education. According to the annotation manual which is used in this study and constantly evolving, I have analyzed and modified a lot of Chinese data, from example sentences in grammar books to articles on Wikipedia. The Chinese database contains a wealth of deeply analyzed syntactic and semantic information. I also reconsidered the main Chinese constructions from the perspective of foreign language education, paying close attention to the relationships between each grammatical category and the characteristics of the Chinese language. As the discussion progressed, it became clear that some basic grammatical terms needed to be redefined. Based on the results of the discussion, I have developed a script which is able to convert our data into a form that is more relevant to language education.

研究分野：言語学

キーワード：統語解析　意味解析　コーパス　中国語学　中国語教育　場所性　アスペクト　存在表現

１．研究開始当初の背景

近年中国語学習のニーズが高まり学習者の人数が増えつつある。このような背景の中、中国語学習者の学習をサポートできる言語資源の需要は急激に増しているが、その整備は十分整っていないというのが現状である。特に、学習者がその学習に際して気軽に参考にできるデータベースはこれまでどこにも存在しなかった。言語学者たちにより開発された中国語のコーパスは幾つか存在しているが、それらをそのまま中国語教育に応用しようとすると、データの質および使いやすさの面において問題がある。

２．研究の目的

このような背景の中、本研究はこれまで蓄積してきた中国語学上の知見及び言語処理学上の技術を積極的に中国語教育に活用し、中国語学習者の習得を深いレベルで支援できるデータベースおよび当該データベースを効率よく使うためのユーザーインターフェイスを開発することを目標とした。このデータベースには、語の品詞情報の他、良質な構文情報も豊富に含まれている。学習者はその中から学習に必要な情報を過不足なくピンポイントで抽出し自分なりに活用することが可能である。

３．研究の方法

本研究で開発するデータベースは、語の品詞情報はもちろん、語と語の間の係り受け関係に関する情報（統語情報）や動詞の格フレームに関する情報（意味情報）も比較的に単純な形式で提供することができる。データベースの構築作業のプロセスは図 1 の示す通りである。無制約の中国語テクストは、まず形態素・統語解析器にかけられ、初歩の統語解析結果が得られる。しかし、これは精度が低いものであり、人手によって修正を行う必要がある。修正されたもの（統語解析情報 1）は、さらに筆者が開発するツリーの構造変換プログラムに入れられる。構造変換プログラムは、言語学と外国語教育との間の差異を補完するためのものであり、これによりツリーは単なる言語学研究のためのものからより外国語教育・学習という目的を重んじるもの（統語解析情報 2）への変形が実現できる。得られた統語解析情報 2 はスコープ制御理論（Butler 2010, 2015）を実装したシステムに入力され、意味処理を行うことによって、文の論理意味表示が得られる。最後に、統語解析情報 2 と論理意味表示を共に統合プログラムに入れれば、目標のデータを得ることができる。一方、ユーザーインターフェイスの開発は、NPCMJ（NINJAL Parsed Corpus of Modern Japanese）の開発チームの協力を得て行った。

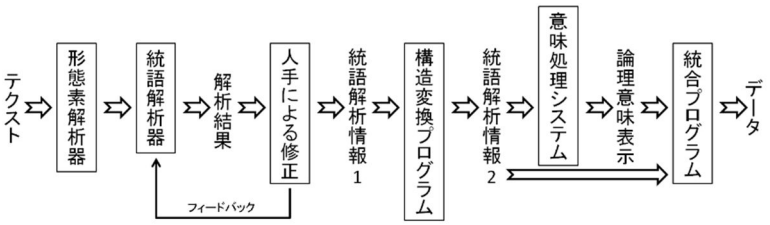

図 1 構築のプロセス

４．研究成果

中国語の句、節、構文の解析方針を検討しつつ、中国語データに対するアノテーション作業を進めてきた。統語解析を行う際に基本的にはペン通時コーパス式の解析規約（Santorini 2010）に従うが、中国語の実情および意味処理からの要請を考慮し、当解析規約に対する改善にも引き続き取り組んできた。例えば、中国語の助詞の文法的役割の多様化に対応するために、助詞句 PP に関する機能情報の付与を本研究の提案した手法の範囲で容認した。また、文の構成要素の間の同一指示関係の同定や否定の作用域の制御などを文の表層的な統語情報のみでより一般的に行うために、スコープの階層という概念を導入し、節レベルの要素に関する暫定的なスコープの階層を決めておき、そして意味処理システムにおいて従属節に関するデフォルトの解釈をインプリメントした。さらに、中国語に見られた主題重視という特徴から、主語と明確に区別できるような形で主題を定義し、それを実際に解析システムの中に導入した。

これらにより、研究の基盤が固められより巨視的で中国語に特化したアノテーションができるようになった。これはデータの一貫性の向上にもつながっていると考えられる。それだけではなく、本研究が従うペン通時コーパス式の規約はこれまで世界中の多くの言語（英語、フランス語、ポルトガル語、アイルランド語、ギリシア語、日本語など）に適用されてきたが、本研究はそれを孤立語の特徴が著しく見られる中国語に応用しようとするものである。系統の異なる言語を対象とするこのような試みは、各国の研究者に当統語解析規約の長所と短所をもう一度検討・認識させ、より完全な解析規約の誕生に有益な示唆を与えることができると考えられる。

解析規約を進化させながら、解析済みのデータに対する修正および新しいデータのアノテー

ションを進めてきた。データには、中国語の文法書（ケンブリッジ出版の*A Reference Grammar of Chinese*など）に載せてある例文、ほかのコーパスの中に収録された中国語文または英語や日本語の文が中国語に訳されたもの（例えば、田中コーパスの中にある日本語・英語の文を中国語に訳してアノテーションをした）、小説（宮沢賢治の『注文の多い料理店』の中国語訳本など）、Wikipediaの条目（“上海”など）、などがある。

その中では、特に小説やWikipediaのデータはいわゆる本物の中国語データであり、それらをアノテーションする意義は大きいと思われる。例えば、中国語版Wikipediaに載せてある“秦始皇「始皇帝」”という条目には707文（24, 505語）あまりの中国語データが入っており、そのほとんどは平均30語を超えた長文である。これらのデータには、“把”構文、動詞連続構文、存現文、各種の補語（結果補語、方向補語、数量補語、様態補語など）が含まれる文など従来中国語に関する研究・教育の関心を集めた項目が多数含まれている。本研究の解析規約の下でこれらの項目に対してその解析をそれぞれ決めていき、そしてハードルの高いアノテーションに実際に挑戦することによって、本研究で採用されている解析方針がさらに磨かれ、その実用性と信頼性がある程度検証できたと考えられる。また、小説は基本的には対訳のあるものに限定したため、今後NPCMJ (NINJAL Parsed Corpus of Modern Japanese）の中に収録されている日本語のデータを同時に考察することにより、多様な日中対照研究が行えると考えられる。

さらに、研究が進むにつれて、場合によっては本研究で取り扱っていない言語情報も必要になってくるということが分かってきた。その典型的な例は語彙レベルの情報と文脈レベルの情報である。例えば、部分コントロール（partial control）の問題については、本研究で採用されている方法（スコープ階層の設定とデフォルトの解釈の付与）だけではなかなか解決できない部分もあり、それを完全に解決するためには語彙レベルでのより詳しい情報が不可欠になる。また、意味処理を通して、指示詞の指示に関する問題や談話におけるテンス・アスペクト情報の形式化を完全に取り扱うためには、文の境を超えた文脈レベルのアノテーションが必要になると思われる。現段階ではこれらの情報は解析の対象ではないが、今後これらの情報に関するアノテーションを本格的に開始し、本研究の可能性をさらに広げていきたい。

問題点としては、研究代表者の研究期間中の最初の二年の身分およびその後の所属異動のため、アノテーターの安定的な雇用が実現できなかったことが挙げられる。研究代表者一人では実際のアノテーション作業に回せるエフォートに限りがあるので、当初予想したデータの量にはまだ至っていない。今後アノテーション作業を行い続けていき、できるだけ早い段階で予定の数のデータを確保できるように取り組みたい。

一方、外国語教育の視点からの中国語の各構文に関する再考察は、中国語の各構文を孤立的に捉えるのではなく、それらの間の関連性ないし中国語全体の特徴をも十分意識した上で行った。これはもちろんデータベースの構築を意識したものではあるが、中国語を体系的に習得するためには、このような中国語の言語データをより一般的に解釈できる枠組みがそもそも不可欠であると考えられる。以下幾つかの根本的な課題を取り上げ本研究の取り組みおよびその意義を紹介する。

孤立語である中国語の抱えている特徴の一つとして、数の限られた機能語の中に複数の文法的役割を担うものが数少なく存在するということが挙げられる。中国語教育の現場では、この問題についてはあまり触れないことが一般的であるが、関連の文法項目を学習者に確実に習得させるためには、機能語の担っている異なる文法的役割を明示的に提示し、それらを明確に区別する必要があると思われる。その典型的な例は中国語の“的”である。学習者は初級から上級まで“的”に関するさまざまな構文を学ぶことになるが、それを同じものとして教える・勉

強することがほとんどである。しかしながら、“的”は少なくとも構造助詞、文末助詞および補文標識の三つの使い方があるとされている。多機能を果たす“的”は、中国語では他の要素と一緒に様々な構造を形成し多くの構文を構築することができる。例えば、中国語の教科書では、“的”と“是”とが共起し“是……的。”構造を取る文が随所に見られているが、これらの文を同様に捉えて教えようとすると学習者に混乱を招いてしまう恐れがある。本研究は、先行研究を踏まえて“是……的。”構造を取る中国語の文をその役割の相違から単純なコピュラ文、焦点文および説明文に分類しそれぞれ考察を行った。その結果、“是”に続く部分については構造的な違いが観察され、単純なコピュラ文は実はさらに関係節の含まれるものと助詞句の含まれるものに分けられるべきであるということが分かった。

また、学習者に中国語の文法構造を理解・習得させるためには、まず主語や目的語と呼ばれてきた幾つかの文法用語をきちんと定義しておく必要があると考えられる。しかし、形式上の明示的な特徴（即ち英語や日本語に見られる名詞や代名詞の格の表示など）が乏しいため、中国語にとってはこのような定義は決して簡単なものではない。何を基準にして中国語の主語や目的語を決定すべきかという問題がこれまで長く取り沙汰されてきた。このような研究の面における不足もあり、現存の中国語教科書の多くからは主語や目的語の乱用およびそれによってもたらされた矛盾が見られている。このような背景の中、本研究は中国語のデータを網羅的に考察し、外国語教育の視点から中国語の主語と目的語をその特徴に合わせた形でそれぞれ再定義した。主語に関する定義は比較的に単純であり、意味的な基準に従うとそれを必要十分条件的に行うことができる。これに対して、中国語の目的語をめぐっては幾つかの複雑な課題が残っており、本研究は形式と意味を組み合わせた基準の下でそれをプロトタイプ的に定義してみた。これによって、中国語における各種の構文をより体系的に分析することが可能になった。

以上のほかにも、本研究は中国語名詞の場所性に関する問題や中国語における存在と所有の境界に関する問題を検討した。これらの根本的な課題の解決を目指しながら本研究は中国語の各構文の文法構造を決定してきた。さらに、これまでの考察結果を踏まえて、本研究はtsurgeonツール（Levy and Andrew 2006）を用いて研究用データから教育用データへと変換するためのスクリプトを作成した。これにより、アノテーション作業の負担が大幅に減少しデータの効率的な利用が実現できた。

ユーザーインターフェイスは、研究代表者が共同研究員として関与しているNPCMJプロジェクトの検索エンジンを使用している。インターフェイスには文字列検索のほか、タグ・ブラウザー、語の依存関係、ツリー検索とテクスト解析、クエリ作成のツールも含まれている。なお、新型コロナウイルスの感染拡大による出張自粛が長期間続いている中、現段階では当時予定した一部の機能（例えば、本研究専用の統語解析器の埋め込みやファクターの自動的な特定）がまだ実現していない。今後引き続き開発を続けていきたい。

＜引用文献＞

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.

Levy, Roger and Galen Andrew.（2006）Tregex and tsurgeon: tools for querying and manipulating tree data structure. In *5th International Conference on Language Resources and Evaluation*.

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

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　2件）

１．著者名 周振、吉本啓	４．巻 17
２．論文標題 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	５．発行年 2019年
３．雑誌名 国立国語研究所論集	６．最初と最後の頁 35-65
掲載論文のDOI（デジタルオブジェクト識別子） 10.15084/00002223	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 周振、吉本啓	４．巻 26
２．論文標題 統語・意味情報付き中国語コーパスの構築：統語解析の詳細について	５．発行年 2020年
３．雑誌名 国際文化研究	６．最初と最後の頁 89-104
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計5件（うち招待講演　0件／うち国際学会　3件）

１．発表者名 周振、吉本啓
２．発表標題 中国語の存在表現に関する再考察及び教授法への提案
３．学会等名 中国語教育学会第16回全国大会
４．発表年 2018年

１．発表者名 周振、吉本啓
２．発表標題 中国語存現文の解析
３．学会等名 言語科学会第20回年次国際大会（国際学会）
４．発表年 2018年

１．発表者名
周振、吉本啓

２．発表標題
“是……的。”構造を取る中国語文の構造的曖昧性に関する考察

３．学会等名
言語科学会第21回年次国際大会（国際学会）

４．発表年
2019年

１．発表者名
周振、吉本啓

２．発表標題
中国語名詞の場所性について

３．学会等名
日本中国語学会第70回全国大会

４．発表年
2020年

１．発表者名
周振

２．発表標題
中国語存在表現のアスペクトに関する考察

３．学会等名
言語科学会第22回年次国際大会（国際学会）

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関