

句本位語法圖解析句系統的設計與實現¹

彭煒明*、何靜**、宋繼華***

摘要

黎錦熙先生首倡的句本位語法體系和圖解析句法以其形象直觀、易於理解的特點在漢語語法教學和研究中產生過重大影響。隨著漢語句法、語義等研究的深入，以及大規模語法樹庫構建實踐的開展，句本位語法在架構句子意義方面的若干優點正在引起語言學界，特別是計算語言學界的重新審視。

本研究遵循句本位語法的主要思路，以黎氏語法體系及其圖解析句法為原型，改造設計出一套形式化的句法分析框架，利用計算機視覺化技術構建了一個句本位語法圖解析句系統。該系統將手工圖解後的句子以 XML 格式存儲，實現了句子結構與圖解圖形的雙向映射，並提供基於字、詞、短語和句式結構的語料檢索和統計功能。利用該系統視覺化標註工具構建了包括《論語》、《唐宋八大家·45 篇》和《漢語語法教材》等 11.4 萬字規模的語法樹庫，進一步利用語料檢索和句式歸納工具，從語法樹庫中歸納了 16 種核心句式，驗證了黎氏語法圖解格局的有效性，並初略比較了古代漢語與現代漢語、先秦文言和唐宋仿古文言之間的句法異同。

關鍵字：黎氏語法、句本位語法、圖解析句、樹庫、句法分析

¹ 本文受北京師範大學跨學科沙龍“計算機技術與漢語言文字學的創新”資助，特此感謝。

* 北京大學計算語言學研究所講師，北京，100871，Email：athing@tju.edu.cn。

** 北京師範大學信息科學與技術學院博士生，北京，100875，E-mail：hejing8@mail.bnu.edu.cn。

*** 北京師範大學信息科學與技術學院教授，北京，100875，E-mail：sjh13@163.com。

The Design and Implement of Diagrammatical Sentence-based Grammar Parsing System

Wei-ming Peng^{*}, Jing He^{**}, Ji-hua Song^{***}

Abstract

Sentence-based Grammar and Diagram Method, which are initiated by Li Jinxi and characterized by its visual and direct observation as well as closeness to semantic understanding, have exerted a great influence on Chinese grammar teaching and research. Along with the research on Chinese grammar and semantics, construction of large-scale syntactic Treebank, some advantages of Sentence-based Grammar in representing sentence meaning have recently drawn re-survey from the linguistics circle, especially the computational linguistics circle.

According to the main thought of Sentence-based Grammar, this research designed the framework of formal syntactic parsing taking Li's Grammar System and its Diagram Method as prototypes, and implemented a diagrammatical Sentence-based Grammar parsing system by using computer visualization technology. This system saved the manually parsed sentence in XML format, realized the bidirectional mapping of sentence syntactic structure and its diagram, and provided the retrieval and statistical function on character, word, phrase and sentence pattern. A 114-thousand-Character Chinese Treebank from *The Analects, Essays of the Tang and Song Dynasties (45 chapters)* and *Chinese Grammar Teaching Material* is constructed utilizing the visualized parsing tool. Then 16 core Chinese sentence patterns are concluded from this Treebank, which verified the effectiveness of Li's Diagram Layout preliminarily. The paper compared the syntactic system between ancient

^{*} Lecturer, Institute of Computational Linguistics, Peking University 100871, China. E-mail: athing@tju.edu.cn.

^{**} Ph.D Candidate, College of Information Science and Technology, Beijing Normal University Beijing 100875, China.

E-mail: hejing8@mail.bnu.edu.cn.

^{***} Professor, College of Information Science and Technology, Beijing Normal University Beijing 100875, China.

E-mail:
sjh13@163.com.

Chinese and contemporary Chinese, Pre-Qin Chinese and Tang-Song Chinese.

Keywords: Li Jinxi's Grammar, Sentence-Based Grammar, Diagram Parsing, Treebank, Syntactic Parsing

2012

DADHIC