Annotating Chinese Noun Phrases Based on Semantic Dependency Graph

Yimeng Li,   Yanqiu Shao
Beijing Language and Culture University

2016.11.21

CONTENTS

CONTENTS

Annotation Schemes of NomBank, ChinesNombank and Sinica Treebank

■**Semantic Role Labeling (SRL)**

e.g. "近年来 (in recent years), [ARG1中韩两国之间 (China and South Korea) 的 (de) 经贸 (economic and trade) 往来 (exchanges)] [REL发展(developed)] 迅速 (rapidly)"
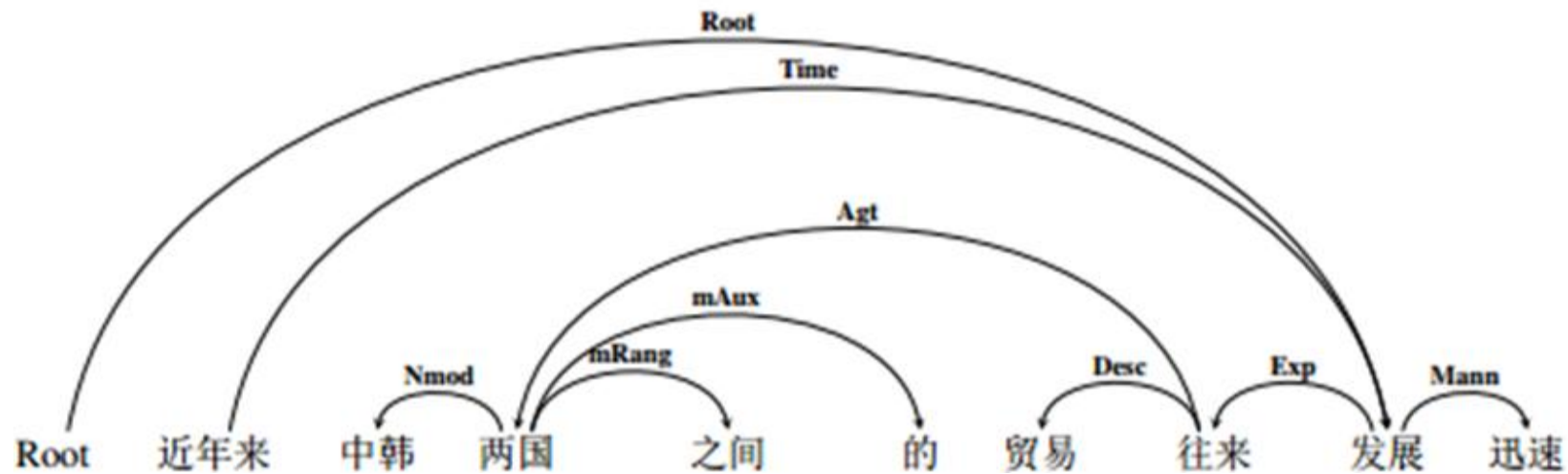
■ **Semantic Dependency Parsing**

Fig. 1 Example of SDG

■ **Semantic Dependency Graph can describe the syntax structure of a sentence.**

e.g. "这一地区 (this region) 成为 (became) [ARG0 海峡两岸 (straits two sides)] [ARG1科技 (scientific and technological)、经贸 (economic and trade)] [REL合作 (cooperation)] 的 (de) 最佳 (best) 地带 (place)"
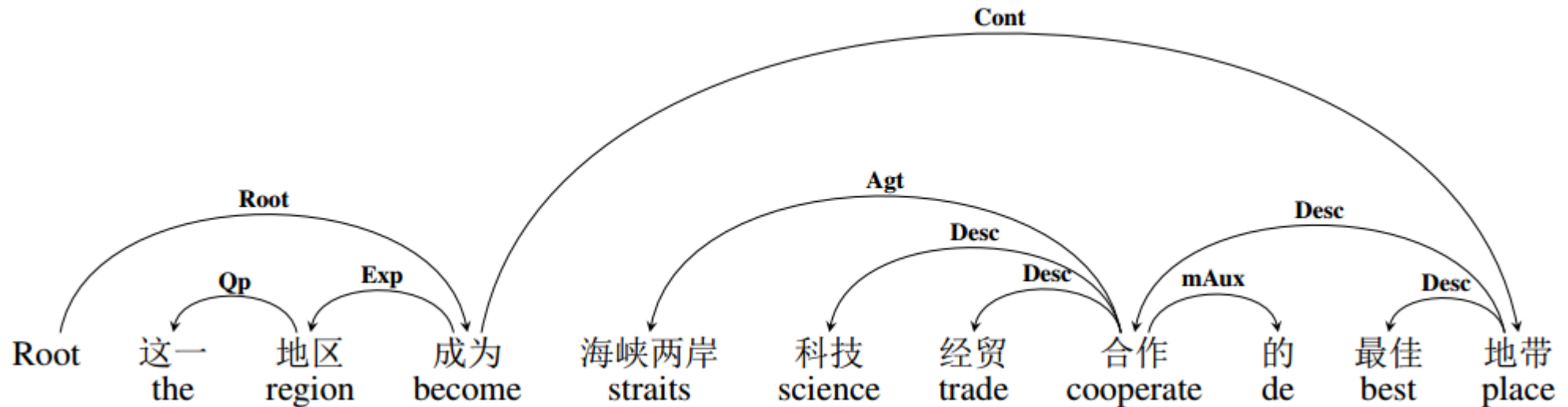

Fig. 2 Example of Chinese NomBank

■ Sinica Treebank provides five kinds of semantic roles to specifically label noun phrases: **apposition, possessor, predication, property, and quantifier.**

e.g. NP (quantifier: DM: 这座| property: NP (property: Nca: 罗亚尔河| Head: Ncda:畔)| predication: VP的 (head: VP (time: Nddc: 最后|Head: VA12: 诞生) |Head: DE: 的)| property: Nad: 文艺复兴 |Head: Nab: 城堡). (The Renaissance castle that was born latest on the Royall River)
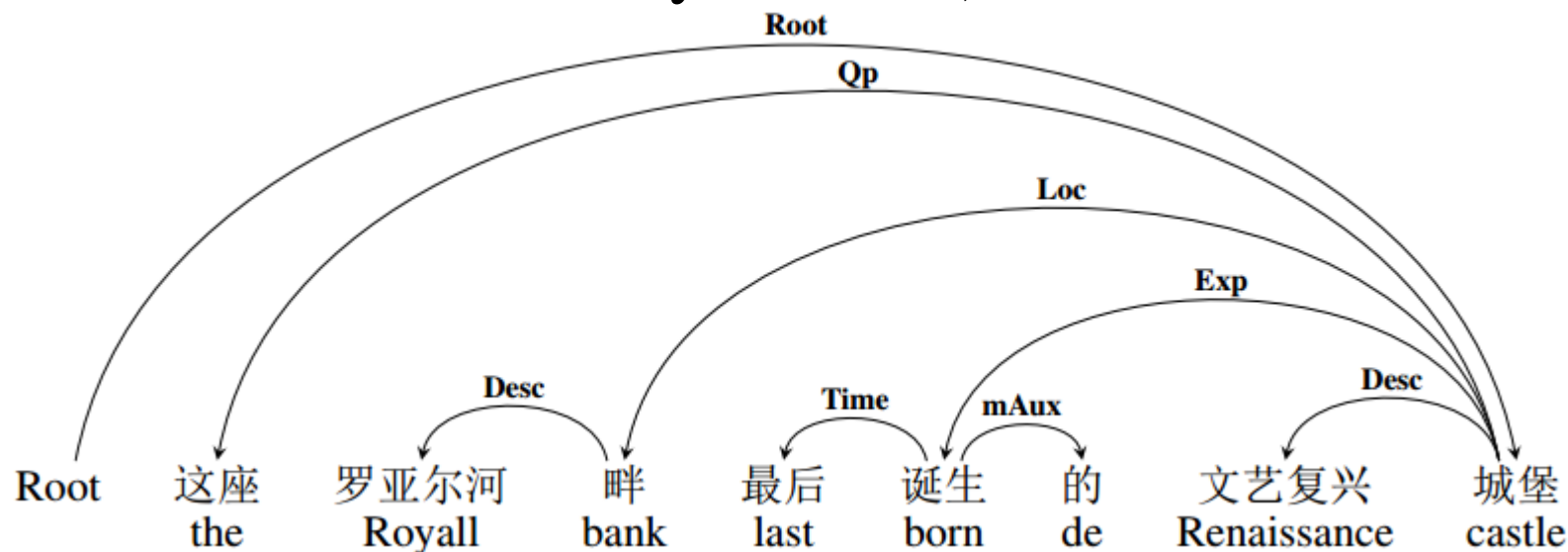

Fig. 3 Example of Sinica

o2

CONTENTS

- **The Corpus for Present Study : BH-SDPB**
- **Annotation of NPs with Predicate**
- **Annotation of NPs without Predicate**
- **The Proportion of Different Noun Phrases**
- **Conclusion**

- **BLCU-HIT Semantic Dependency Graph Bank**

  Built by BLCU and Harbin Institute of Technology
   Source : Primary School Texts, Spoken Language and Machine Translation
Corpus.
   Size : 30, 000 sentences , 750,000 words.
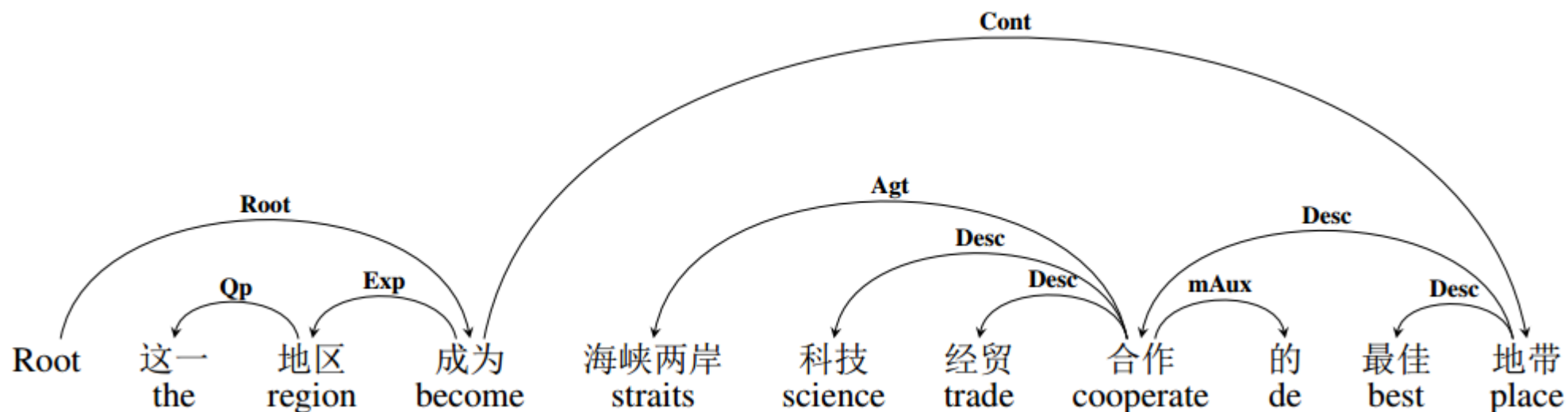- **Achievements:** The training corpus of SemEval of 2012 and 2016
(Semar

Fig.4 Example of SDGB

■**Data Extraction**

- Automatic Machine Extraction: 2860
- Manual Verification: 1830 NPs without the conjunction "的"（de）
  Noun Phrases with Predicate: 718 as shown in Fig. 5
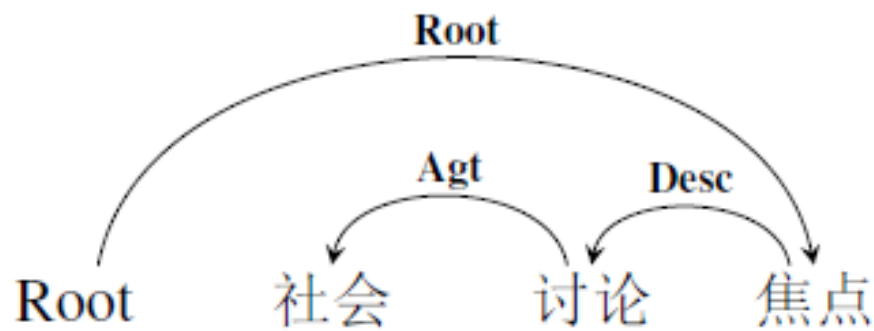 Noun Phrases without Predicate : 1112 as shown in Fig.6
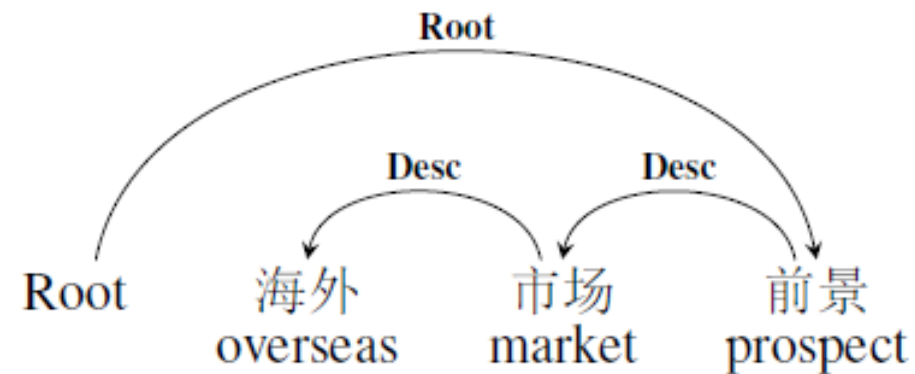

Fig. 5

Fig. 6

■ **The Annotation Procedure of NPs with Predicate:**

    1. Determine the root of the phrase;

    2. Determine the predicate and find out its arguments such as Agt, Pat, etc.

    3. Annotate the modifiers of predicate such as Loc, Time, Mann.etc.

    4. Annotate the modifiers of nouns such as Desc, Poss, Quan.etc.
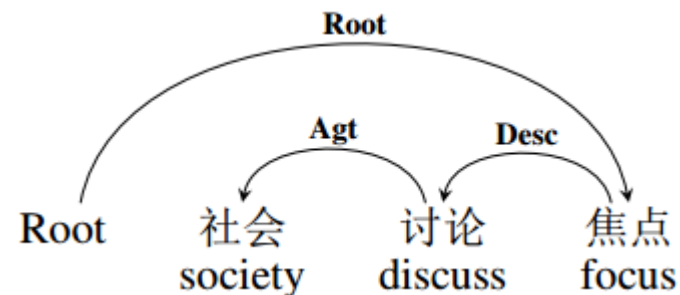
Fig.7 Example of Procedure

■**Type of DPM:** Type **Division** with **Predicate** in the **Middle**. The words on both sides of the predicate have direct semantic relations with the predicate. Fig. 8

■ **Type of IPM:** Type **Integration** with **Predicate** in the **Middle**. All the words before directly modify the root. Fig. 9
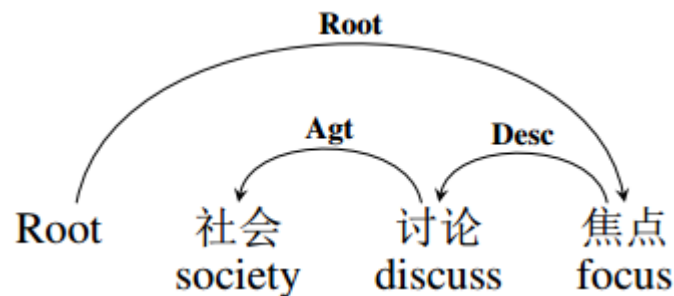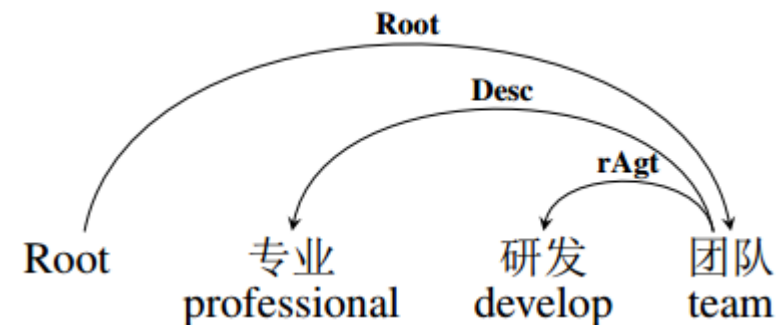

Fig. 8 : DPM

Fig. 9 : IPM

■ **Type of DPE:** Type of **Division** with **Predicate** in the **End**. The component$_1$ is a modifier of the component$_2$, which tends to be an argument of the predicate, i.e., the root. Fig. 10

■ **Type of IPE:** Type of **Integration** with **Predicate** in the **End**. All the words before directly modify the root. Fig. 11
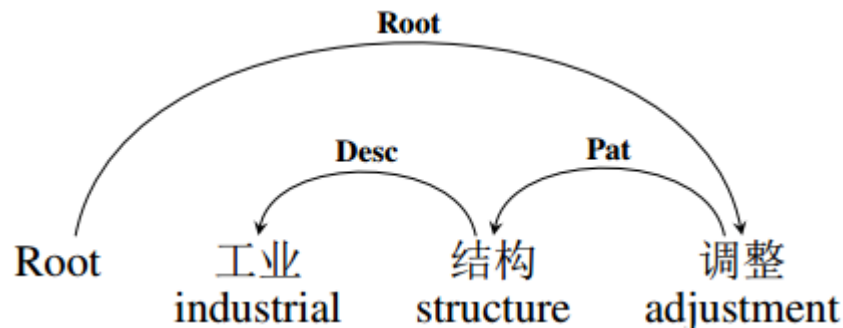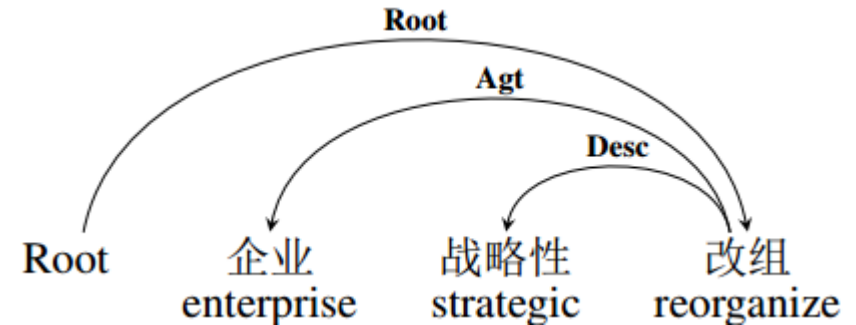

Fig. 10 DPE

Fig. 11 IPE

■ **Type of DPB :** Type of **Division** with **Predicate** at the **Beginning**. The predicate and the component after it are more related semantically and they modify the root together. Fig. 12.

■ **Type of IPB: Type** of **Integration** with **Predicate** at the **Beginning.** The words before directly modify the root. Fig. 13.

■ **Type of APB:** The predicate and the components after it constitute an **Argument-Predicate** structure that modifies the root. Fig. 14.
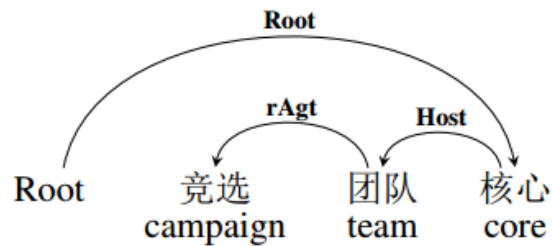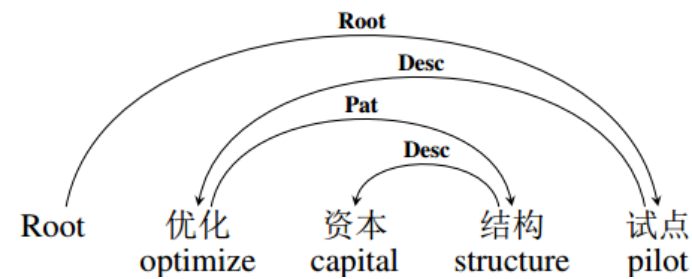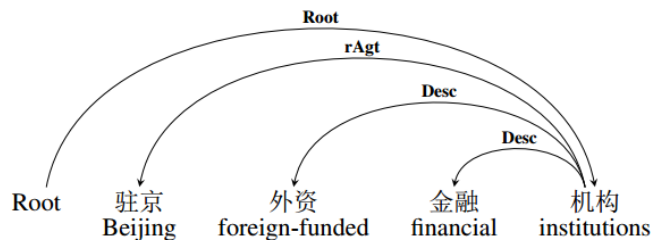

Fig. 12 DPB

Fig. 14 IPD

Fig. 13 APB

■ **Type of DN:** Type of **Division** of **Noun.** All the words before directly modify the root.

■ **Type of IN:** Type of **Integration** of **Noun**. All the words directly modify the root.
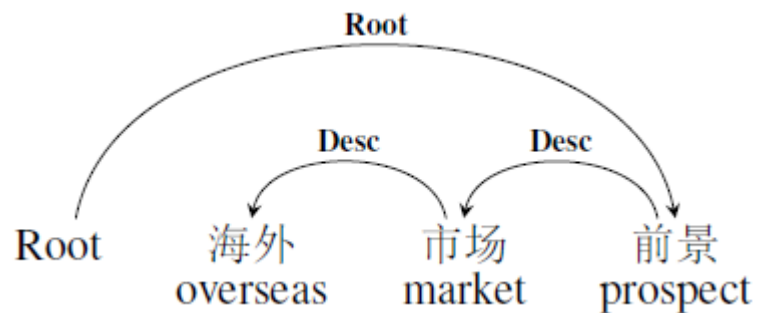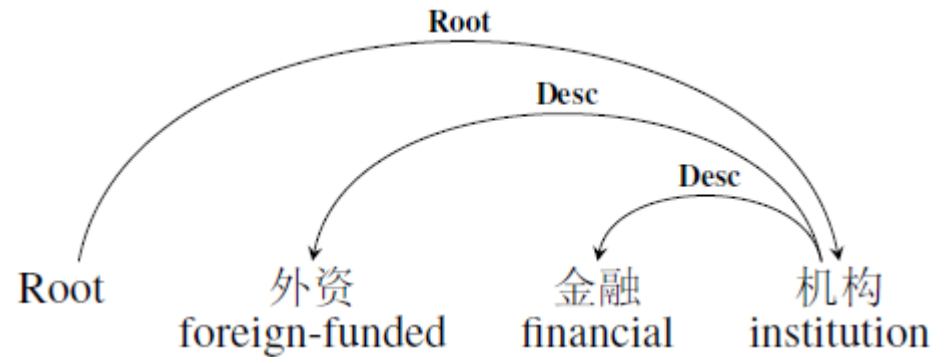

Fig. 15 DN

Fig. 16 IN

Table 1: Proportion of NPs with Predicate

Table 2: Proportion of NPs without Predicate

NPs with Predicate	Quantity	Proportion
DPM	418	58%
IPM	63	9%
DPE	123	17%
IPE	61	9%
DPB	26	4%
IPB	17	2%
APB	10	1%
Total	718	

NP without Predicate	Quantity	Proportion
DN	455	40%
IN	657	60%
Total	1112	

■ **Conclusion**

• Design Annotation Procedure and Propose an Annotation Scheme for Chinese Noun Phrases.
• Divide Noun Phrases into 9 Types According to the Semantic Relation.
• The number of Type DPM (Fig. 17) and IN (Fig. 18) is more than others respectively accounting for 58% and 60%.
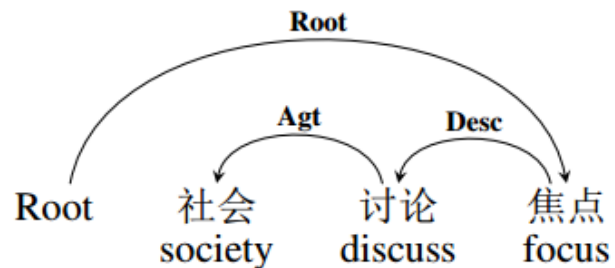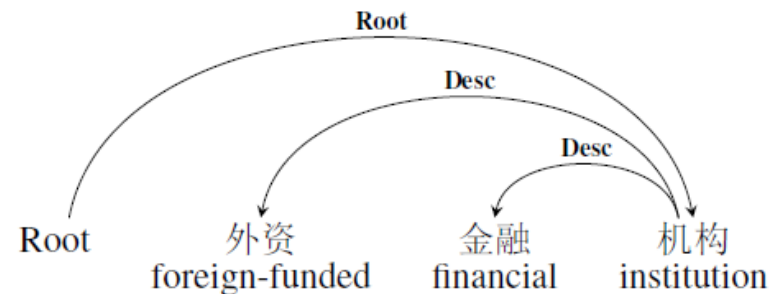

Fig. 17 DPM

Fig. 18 IN

■**Application**

The labeled corpus can provide materials for machine learning and it can be useful in other fields such as nominal predicate recognition, information extraction and machine translation, etc.

■**Expectation**

The following job will be to investigate the semantic roles of noun phrase, hoping to provide linguistic support for the development of semantic analyzer.

Thanks

Beijing Language and Culture University

2016.11.21