



**NNU** · 南京师范大学  
NANJING NORMAL UNIVERSITY



**HARVARD**  
UNIVERSITY



**Brandeis University**

# Construction and Analysis of the Chinese **A**bstract **M**eaning **R**epresentation Corpus

## 中文**AMR**语料库的构建与分析

**Bin Li** 李斌

School of Chinese Language and Literature, Nanjing Normal University, China

Institute for Quantitative Social Science, Harvard University, USA

Email: libin.njnu@gmail.com

<http://www.cs.brandeis.edu/~clp/camr/camr.html>





# Chinese Processing Group at NJNU



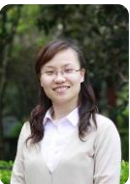
Founder Prof. Xiaohe Chen



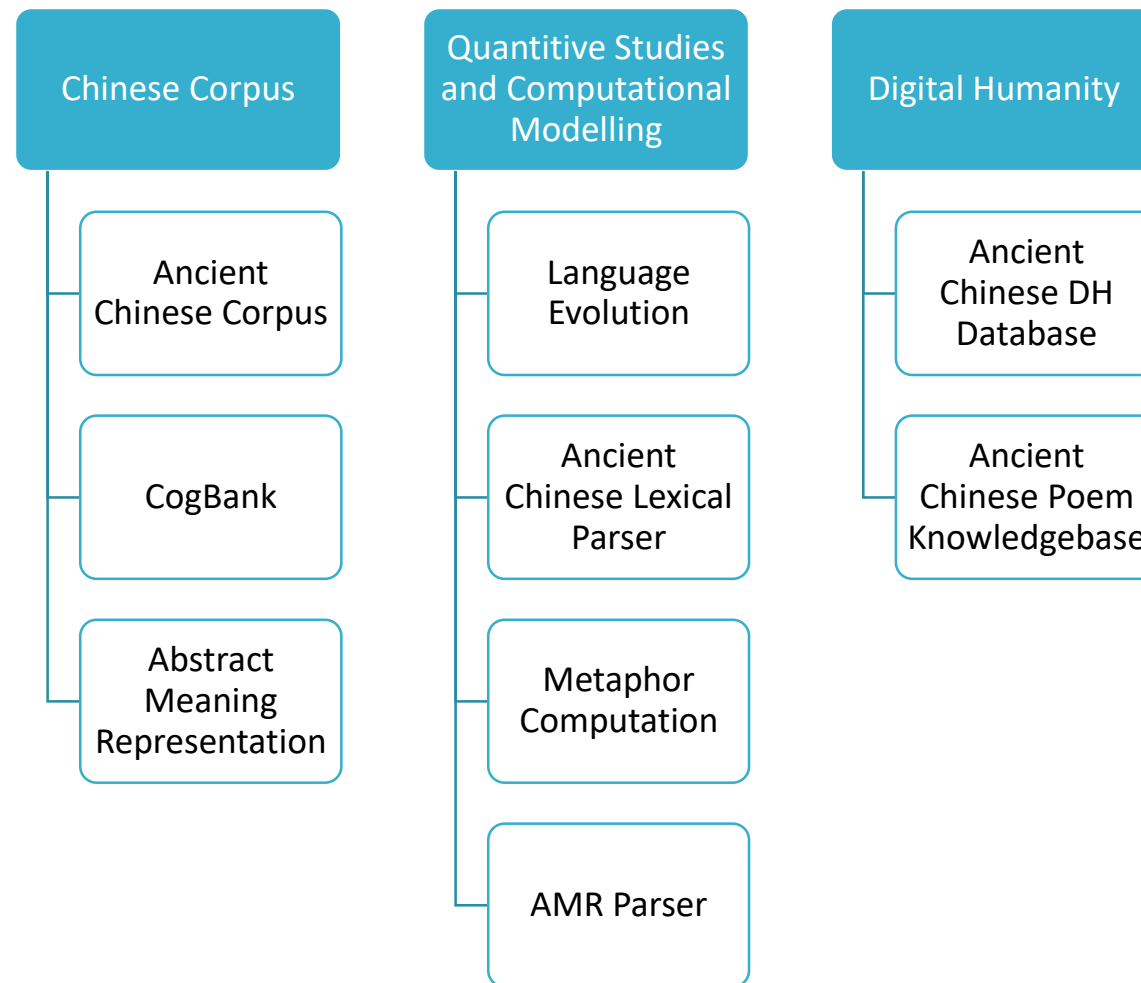
Prof. Weiguang Qu



Associate Prof. Bin Li



Associate Prof. Minxuan Feng



ABOUT

MEMBERS

COMMUNICATIONS

LANGUAGE RESOURCES ▾

Data ▾

Obtaining Data

Catalog

By Year

Top Ten Corpora

Projects

Search

Memberships

LDC Online

Data Scholarships

Tools >

Papers >

LR Wiki

DATA MANAGEMENT

COLLABORATIONS

[Home](#) > [Language Resources](#) > [Data](#)

## Ancient Chinese Corpus

<i>Item Name:</i>	Ancient Chinese Corpus
<i>Author(s):</i>	Xiaohe Chen, Bin Li, Minxuan Feng, Chao Xu, Runhua Xu, Min Shi, Lili Yu, Lei Xiao, Qingqing Wang
<i>LDC Catalog No.:</i>	LDC2017T14
<i>ISBN:</i>	1-58563-816-1
<i>ISLRN:</i>	924-985-704-453-5

### Introduction

Ancient Chinese Corpus was developed at [Nanjing Normal University](#). It contains word-segmented and part-of-speech tagged text from *Zuozhuan*, an ancient Chinese work believed to date from the Warring States Period (475-221 BC). *Zuozhuan* is a commentary on the *Chunqui*, a history of the Chinese Spring and Autumn period (770-476 BC). This release is part of a continuing project to develop a large, part-of-speech tagged ancient Chinese corpus.

### Data

Ancient Chinese Corpus consists of 180,000 Chinese characters and 195,000 segment units (including words and punctuation). The part-of-speech tag set was developed by Nanjing Normal University and contains 17 tags.

This release contains two text files: 268 paragraphs and 10,560 lines. A line is one sentence; paragraphs are separated by one empty line. Each word is tagged with its part-of-speech and separated by a space.

The files are presented in UTF-8 plain text files using traditional Chinese script.





ABOUT

MEMBERS

COMMUNICATIONS

LANGUAGE RESOURCES

Data

Obtaining Data

Catalog

By Year

Top Ten Corpora

Projects

Search

Memberships

Data Scholarships

Tools

Papers

LR Wiki

DATA MANAGEMENT

COLLABORATIONS

[Home](#) › [Language Resources](#) › [Data](#)

## Chinese Abstract Meaning Representation 1.0

<i>Item Name:</i>	Chinese Abstract Meaning Representation 1.0
<i>Author(s):</i>	Bin Li, Yuan Wen, Li Song, Rubing Dai, Weiguang Qu, Nianwen Xue
<i>LDC Catalog No.:</i>	LDC2019T07
<i>ISBN:</i>	1-58563-880-3
<i>ISLRN:</i>	376-537-072-369-4
<i>Release Date:</i>	April 15, 2019
<i>Member Year(s):</i>	2019
<i>DCMI Type(s):</i>	Text
<i>Data Source(s):</i>	weblogs, discussion forum
<i>Project(s):</i>	ACE
<i>Application(s):</i>	parsing, syntactic parsing, semantic role labelling
<i>Language(s):</i>	Mandarin Chinese

# Outlines



What is AMR?



Guidelines & Modifications



Corpus Construction



Statistics & Analysis



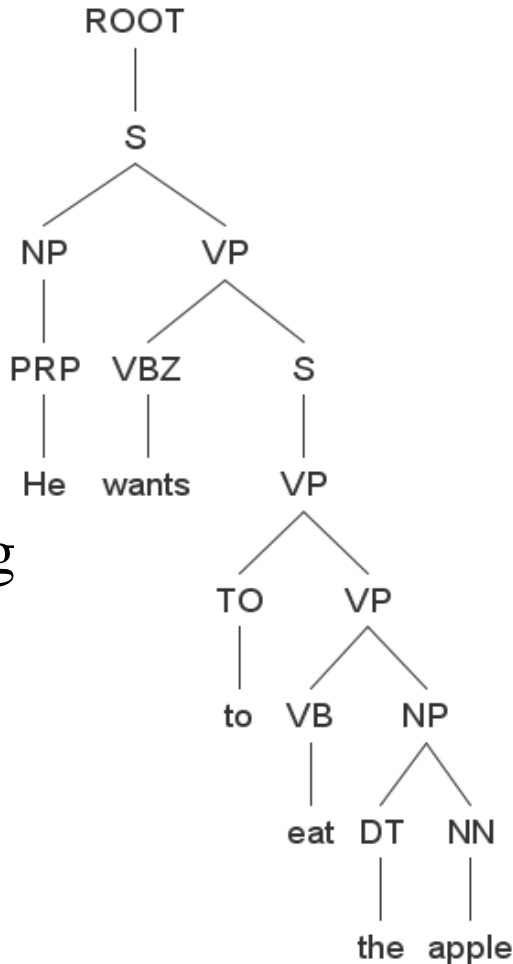
Discussion

# Main Questions

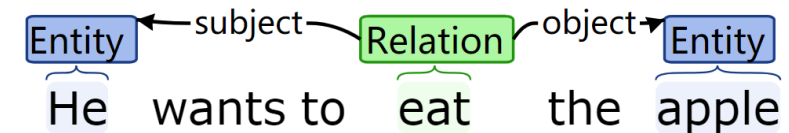
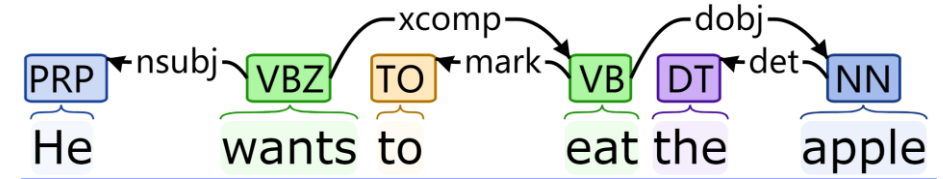
- What is AMR?
- Why graphs instead of trees?
- Could the concepts and relations on an AMR graph be aligned back to original words?
- Is AMR universal? Could it be applied to other languages?
- Is there something new discovered by Chinese AMR?

# 0. The structure of a sentence

- Syntax Tree
  - Chomsky 1957
  - Dependency 1959
- FrameNet
  - Fillmore 1977
  - Semantic Role Labelling



He wants to eat the apple





# FrameNet

## want.v

Frame Element	Core Type
Degree	Peripheral
Duration	Peripheral
Event	Core
Experiencer	Core
Explanation	Extra-Thematic
Focal_participant	Core
Location_of_event	Core
Manner	Peripheral
Place	Peripheral
Purpose_of_event	Peripheral
Role_of_focal_participant	Peripheral
Time	Peripheral
Time_of_event	Peripheral

## eat.v

Frame Element	Core Type
Degree	Peripheral
Duration	Peripheral
Ingestibles	Core
Ingestor	Core
Instrument	Peripheral
Manner	Peripheral
Means	Peripheral
Place	Peripheral
Purpose	Peripheral
Source	Peripheral
Time	Peripheral

He wants to eat the apple

Want. Experiencer

Eat. Ingestor

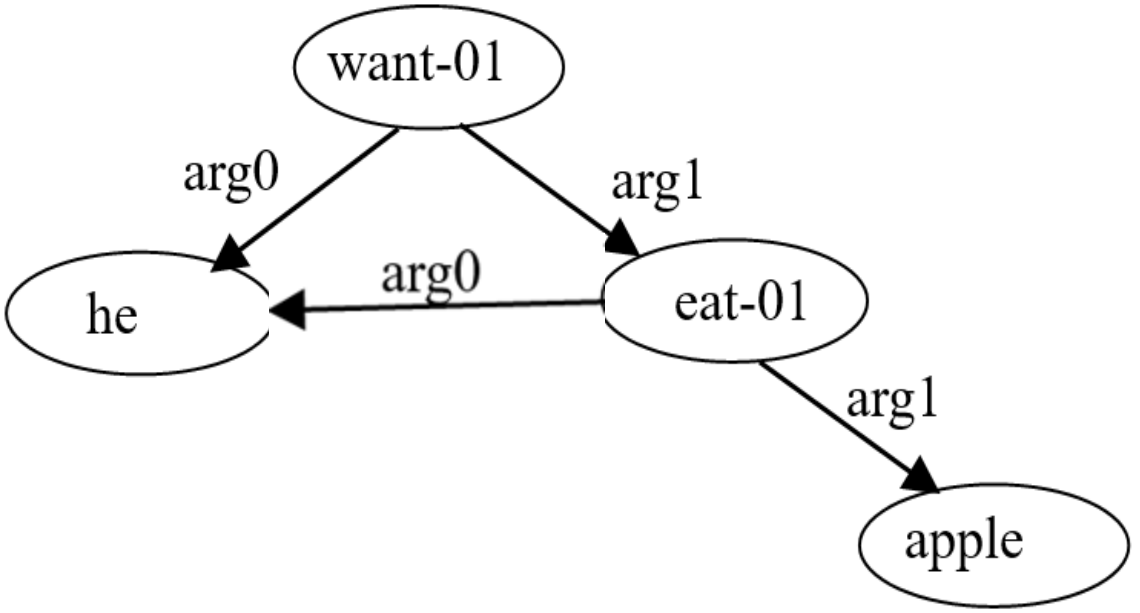
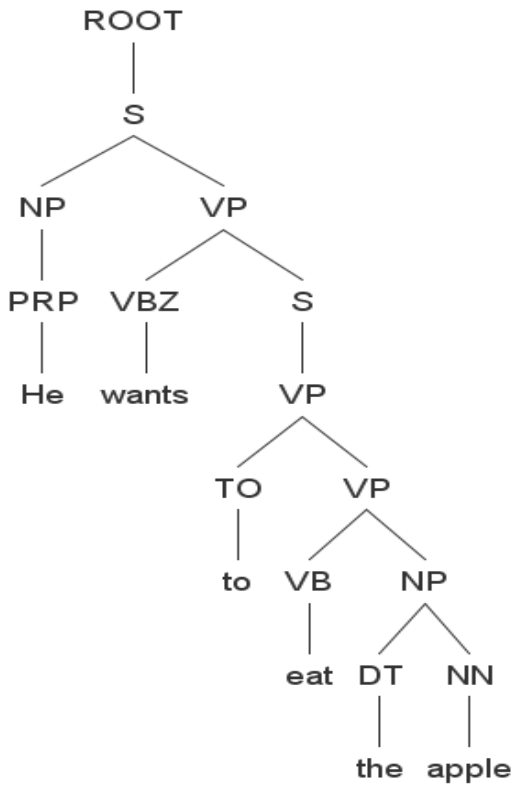


<https://amr.isi.edu/>

# 1. What is AMR?

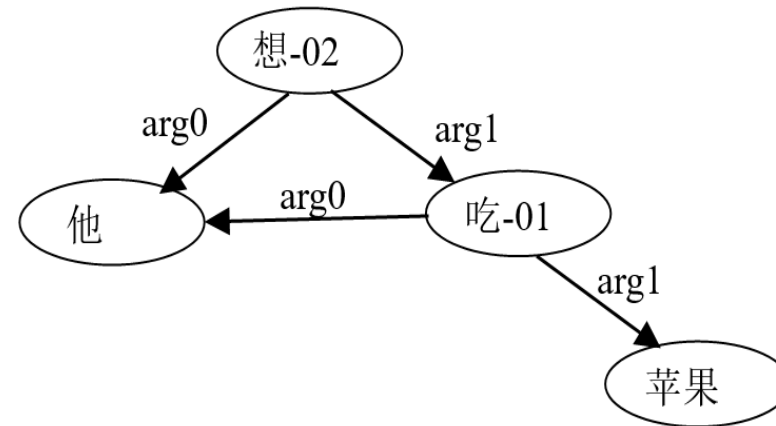
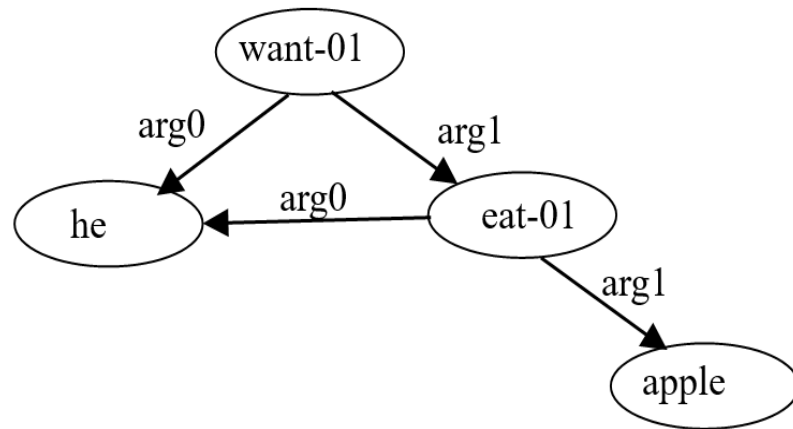
- According to Banarescu et al.(2013), the **A**bstract **M**eaning **R**epresentation of a sentence is a **rooted, directed, acyclic graph** with labels on arcs(relations) and nodes(concepts).

**He wants to eat the apple**



## 1.2. Major Advantages of AMR Abstraction

- Use graph structure to solve argument sharing
  - He wants to eat the apple
  - 他 想 吃 苹果



It works on Chinese!

# Re-analyze and adding concepts

- The dancer has gone.

The dancer has gone  
g/go-01  
:arg0 p/person  
:arg0-of d/dance-01

- 跳舞 的 走 了
- Dance DE(of) go PERFECT.ASP

跳舞的走了  
x/走-02  
:arg0 p/person  
:arg0-of x1/跳舞-01

- Good solution for DE-Structure

# 1.3. Main Questions

- ✓ What is AMR?
- ✓ Why graphs instead of trees?
- Could the concepts and relations on an AMR graph be aligned back to original words?
- Is AMR universal? Could it be applied to other languages?
- Is there something new discovered by Chinese AMR?

# 1.4. Major Problems of AMR

- Automatic Parsing is rather hard, only about 64%
  - SemEval 2016, 2017
- Alignment Problem
  - The word-concept alignment accuracy is about 90%
- Lacking corpus of other languages
  - English AMR corpus (40k sentences, <http://amr.isi.edu/>)
  - Chinese AMR corpus of *the little Prince* (Li et al., 2016)
  - Chinese Treebank, PropBank, now AMR.



# Pros and Cons of AMR



## Pros

Argument sharing

Add/delete/replace concepts

Inner structure



## Cons

No alignment

Tense/aspect/functional  
words

Discourse relations

## 2. Annotation Guidelines of Chinese AMR

- Following AMR's specification
  - Arg0-Arg5, relations, entities, etc.
- Improvement for Chinese
  - + Concept-to-word Alignment
  - + Relation-to-word Alignment
  - + 4 Non-core Relations
  - + Discourse Relations
  - + Specifications for Chinese
    - Reduplications
    - Headless Relative Construction
    - Verb-Complement Construction
    - Split Verb Construction

(Li et al., 2016)

# 2.1. Concept-to-word Alignment

- ✓ No alignment
- ✓ Typos and errs

- Word index – concept ID

- 谁<sup>1</sup> 帮<sup>2</sup> 了<sup>3</sup> 窝(我)<sup>4</sup> 这么<sup>5</sup> 大<sup>6</sup> 的<sup>7</sup> 忙<sup>8</sup> ?<sup>9</sup>
- Who help ASP nest (me) such big DE business
- Who helped me so much?

**Unaligned**

x1/帮忙-01 help ~2!!  
:arg0 a/amr-unknown ~1  
:arg1 x2/我 me ~4  
:degree x3/大 big ~6  
:degree x4/这么such ~5  
:mode i/interrogative

**Aligned**

x2\_x8/帮忙-01 help  
:aspect x3/了 Complete  
:arg0 x1/ amr-unknown  
:arg1 x4/我 me  
:degree x6/大 big  
:degree x5/这么 such  
:mode x9/interrogative

- Internal analysis of words

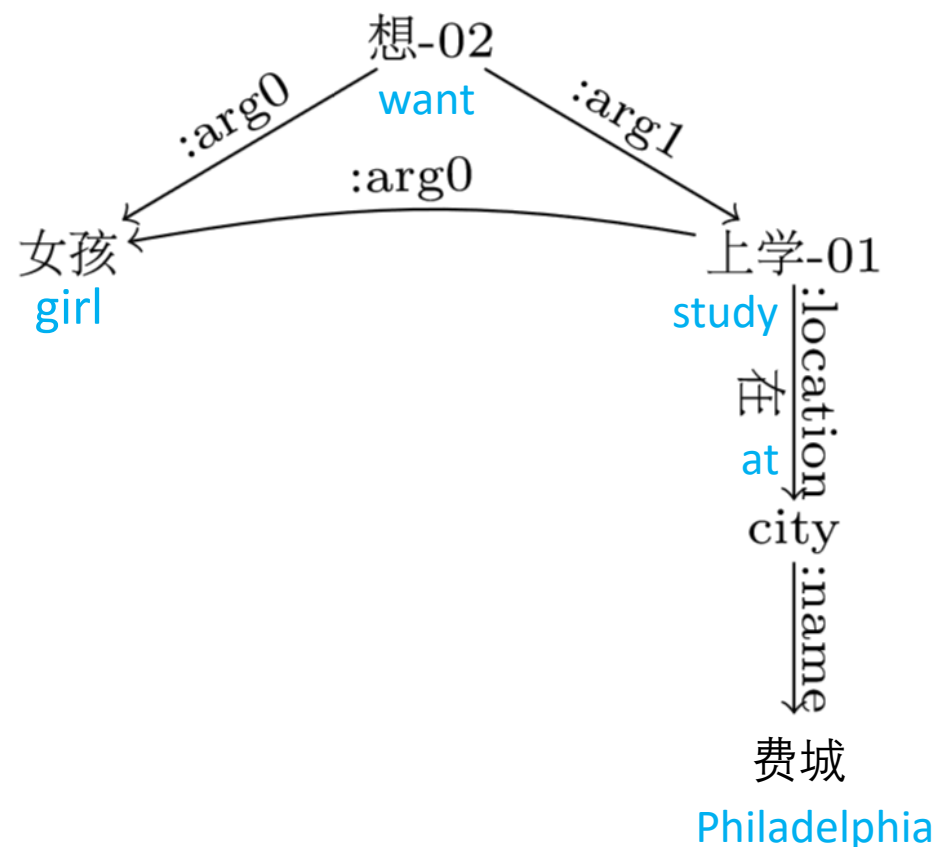
- 土地<sup>1</sup> 拥有者<sup>2</sup>
- Land possess-er

x2\_3/者  
:arg0-of x2\_1\_2/拥有-01  
:arg1 x1/土地

## 2.2. Relation-to-word Alignment

- ✓ No alignment
- ✓ Typos and errs
- ✓ Functional words

- 女孩<sup>1</sup> 想<sup>2</sup> 在<sup>3</sup> 费城<sup>4</sup> 上学<sup>5</sup>
- Girl want at Philadelphia study
- The girl wants to study at Philadelphia
- x2/想 want-02
  - :arg0 x1/女孩 girl
  - :arg1 x5/上学 study-01
  - :location(x3/在 at) x6/city
  - :name x4/费城 Philadelphia



## 2.3. Add 4 non-core relations

- tense (时)
  - 将, 曾
- aspect (体)
  - 着、了、过、起来、下去
- cunit (classifier)
  - 个、只、张
- perspective
  - 他经济独立了
  - He is financially independent.

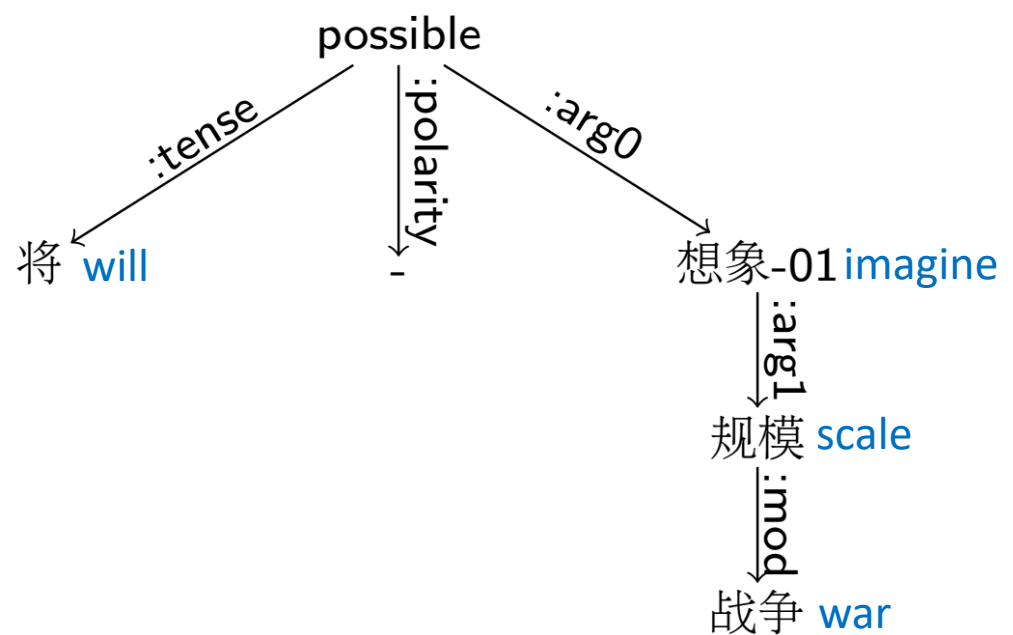
- ✓ No alignment
- ✓ Typos and errs
- ✓ Functional words
- ✓ Tense & aspect
- ✓ Classifier

# Tense

战争<sup>1</sup> 规模<sup>2</sup> 将<sup>3</sup> 无法<sup>4</sup> 想象<sup>5</sup>  
war scale will unable imagine

“The scale of the war will be unimaginable.”

x8/possible  
:polarity x4/-  
:tense x3/将  
:arg0 x5/想象-01  
:arg1 x2/规模  
:mod x1/战争





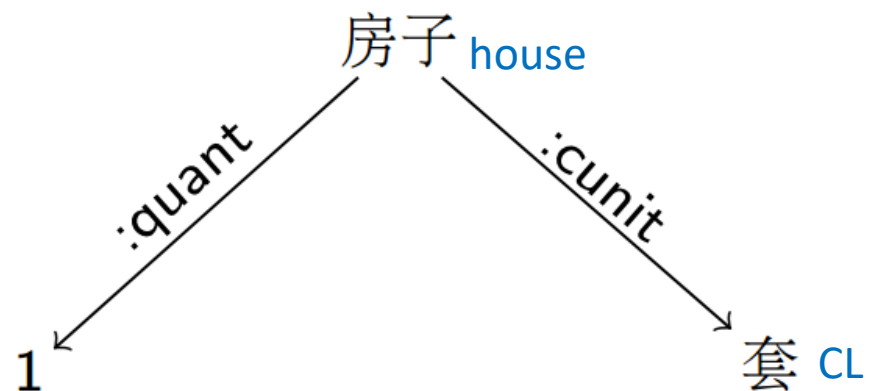
# CUNIT (classifier)

(11) 一<sup>1</sup> 套<sup>2</sup> 房子<sup>3</sup>  
a CL house  
“A house”

x3/房子

:quant x1/1

:cunit x2/套



## 2.4. Relation of Compound Sentences

- ✓ No alignment
- ✓ Typos and errs
- ✓ Functional words
- ✓ Tense & aspect
- ✓ Classifier
- ✓ Discourse relations

- Setting up 10 **concepts** to represent the discourse relations between clauses

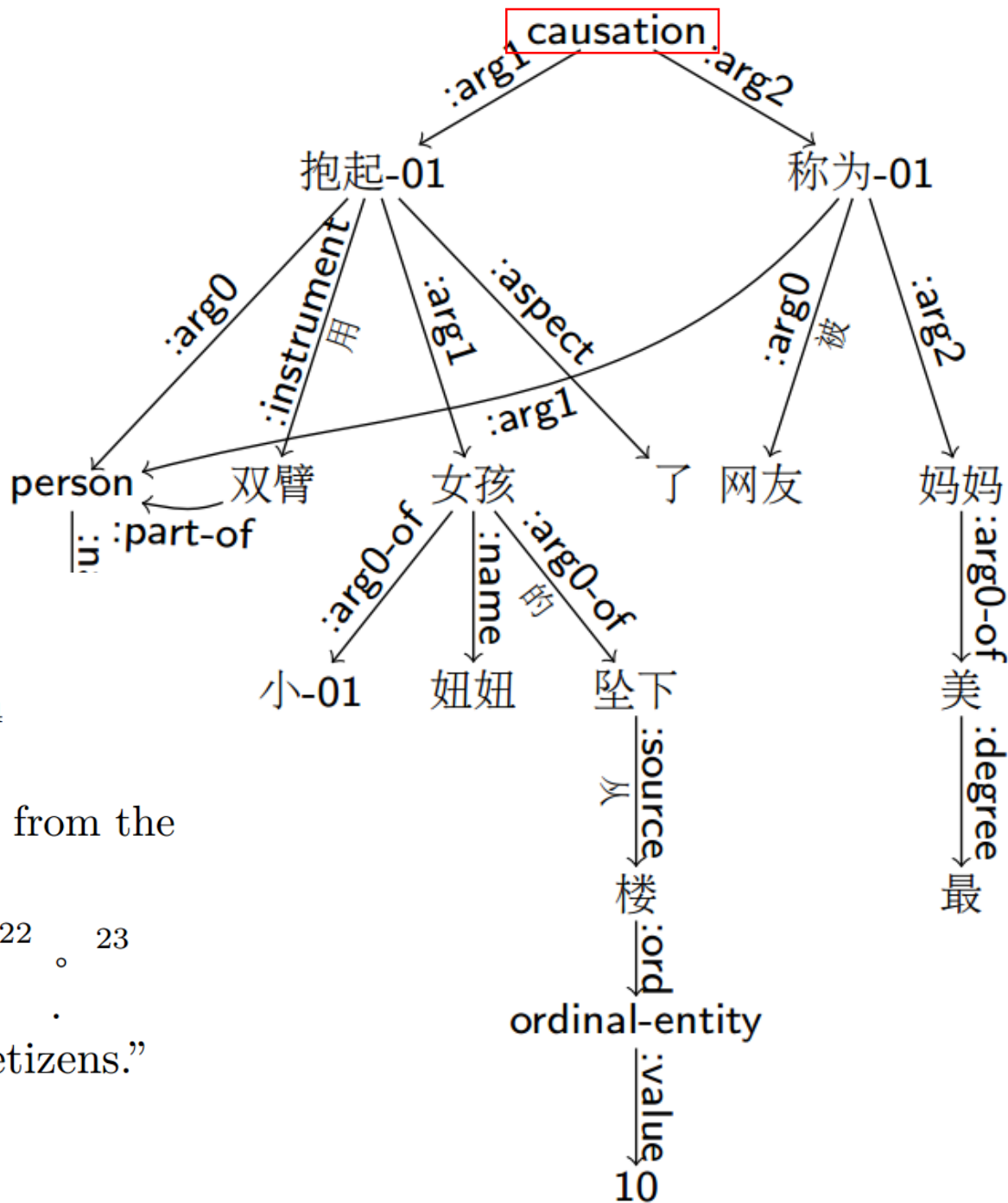
• and	并列	• or	选择
• causation	因果	• concession	让步
• condition	条件	• purpose	目的
• Contrast	转折	• <b>progression</b>	<b>递进</b>
• <b>temporal</b>	<b>时序</b>	• expansion	解说

# causation

- From relation to concept

- :cause
- Causation
  - Arg1, Arg2

(9) 吴菊萍<sup>1</sup> 用<sup>2</sup> 双臂<sup>3</sup> 抱起<sup>4</sup> 了<sup>5</sup>  
 Juping Wu with arms pick up ASP  
 从<sup>6</sup> 十<sup>7</sup> 楼<sup>8</sup> 坠下<sup>9</sup> 的<sup>10</sup> 小<sup>11</sup> 女孩<sup>12</sup> 妞妞<sup>13</sup> ,<sup>14</sup>  
 from tenth floor fall DE little girl Niuniu ,  
 “Juping Wu picked up the little girl Niuniu who fell from the  
 tenth floor with her arms,”  
 被<sup>15</sup> 网友<sup>16</sup> 称为<sup>17</sup> “<sup>18</sup> 最<sup>19</sup> 美<sup>20</sup> 妈妈<sup>21</sup> ”<sup>22</sup> 。<sup>23</sup>  
 by netizens call “ most beautiful mother ” .  
 “and was called ‘the most beautiful mother’ by netizens.”



## 2.5. Specifications for Chinese

Better representation

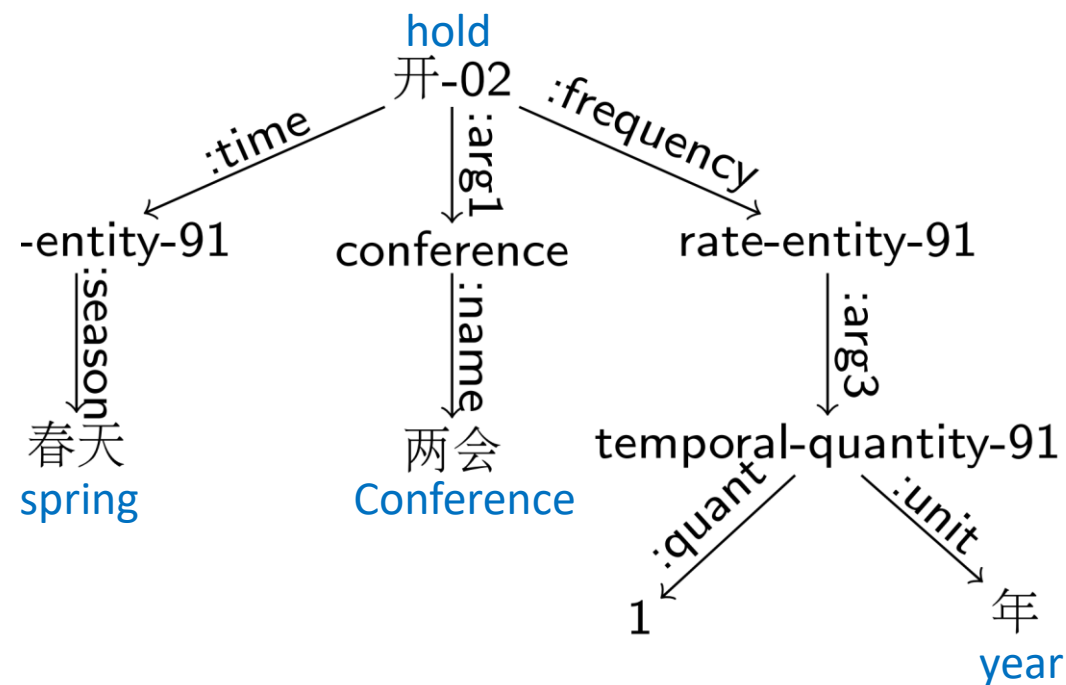
- Reduplications
  - 看看→看; 开开心心→开心; 打扫打扫→打扫
  - 年年→每年
- Split verb construction
  - 帮忙(help) → 帮了一个忙
- Serial verbs structure
  - He **wants** to **eat** the apple
- Verb-complement structure
  - 走不了→不可能走; 唱哭→ the result of “唱” is “哭”

# Reduplications

(17) a. 说 说 → 说  
say say say

b. 干 干 净 净 → 干净  
dry dry clean clean clean

c. 年 年 → every 年  
year year every year



# Serial verbs structure

(12) 放<sup>1</sup> 着<sup>2</sup> 好<sup>3</sup> 日子<sup>4</sup> 不<sup>5</sup> 过<sup>6</sup>  
leave ASP good life not live

“Do not want to settle with living a good life”

x8/and

:op1 x1/放-01

:aspect x2/着

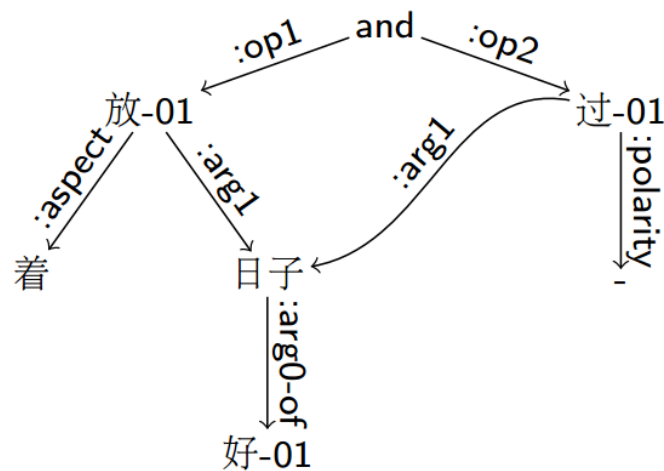
:arg1 x4/日子

:arg0-of x3/好-01

:op2 x6/过-01

:arg1 x4

:polarity x5/-





# Verb-complement structure

- 走 不 了 → 不 可能 走
- Go not ASP          not possible go
  
- 他 唱 哭 了 观众 → the result of “唱” is “哭”
- He sing cry ASP audience
- He sings, causing the audience to cry.

# 3. Corpus Construction

- We annotated 10,149 sentences
  - selected from the Chinese TreeBank 8.0
  - predicate frames were extracted from Chinese PropBank 3.0
  - followed the new CAMR specifications
- Two linguistic undergraduate students
- The inter-agreement Smatch (Cai and Knight, 2013) score between the two annotators was 0.83

# 3.1. Predicate Dictionary

- predicate frames were extracted from the Chinese PropBank 3.0

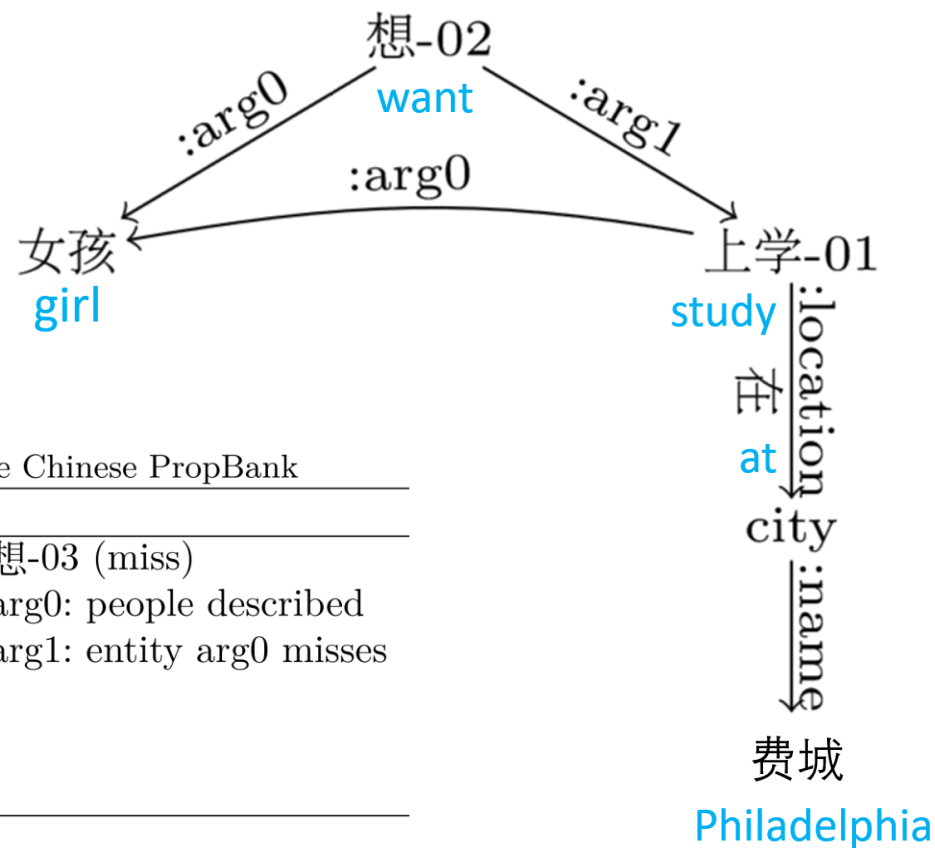


TABLE 1 Semantic roles for core arguments of *want* in PropBank and 想 in the Chinese PropBank

want	想		
want-01	想-01 (think)	想-02 (want)	想-03 (miss)
arg0: wanter	arg0: people described	arg0: people described	arg0: people described
arg1: thing wanted	arg1: thoughts of arg0	arg1: thing arg0 wants	arg1: entity arg0 misses
arg2: beneficiary			
arg3: in-exchange-for			
arg4: from			

## 3.2. Abstract Concepts

Type	Abstract concepts	Num
*discourse	and, or, (*causation, *condition, *contrast, *temporal, *concession, *progression, *purpose, *expansion), multi-sentence	11
subjectivity	-(polarity), +(polite), possible	3
mode	interrogative, expressive, imperative	3
unknown	amr-unknown	1
quantity	monetary-quantity, distance-quantity, area-quantity, volume-quantity, temporal-quantity, frequency-quantity, speed-quantity, acceleration-quantity, mass-quantity, force-quantity, pressure-quantity, energy-quantity, power-quantity, voltage-quantity, charge-quantity, potential-quantity, resistance-quantity, inductance-quantity, magnetic-field-quantity, magnetic-flux-quantity, radiation-quantity, concentration-quantity, temperature-quantity, score-quantity, fuel-consumption-quantity, seismic-quantity	26
91 concept	have-concession, have-condition, be-destined-for, have-frequency, have-instrument, be-located-at, have-manner, have-mod, have-name, have-part, have-polarity, have-purpose, have-quant, be-from, have-subevent, include, be-temporally-at, rate-entity	18
Total		179

\* marks new concepts added to CAMR

TABLE 2 List of abstract concepts used in CAMR

Type	Abstract concepts	Num
	thing	1
	person, family, animal, language, nationality, ethnic-group, regional-group, religious-group	8
	organization, company, government-organization, military, criminal-organization, political-party, school, university, research-institute, team, league	11
	location, city, city-district, county, local-region, state, province, country, country-region, world-region, continent, ocean, sea, lake, river, gulf, bay, strait, canal, peninsula, mountain, volcano, valley, canyon, island, desert, forest, moon, planet, star, constellation	29
Named Entity (108)	facility, airport, station, port, tunnel, bridge, road, railway-line, canal, building, theater, museum, palace, hotel, worship-place, market, sports-facility, park, zoo, amusement-park	20
	event, incident, natural-disaster, earthquake, war, conference, game, festival	8
	product, vehicle, ship, aircraft, aircraft-type, spaceship, car-make, work-of-art, picture, music, show, broadcast-program	12
	publication, book, newspaper, magazine, journal	5
	naturalobject	1
	molecular-physical-entity, small-molecule, protein, protein-segment, amino-acid, macro-molecular-complex, enzyme, rna, pathway, gene, dna-sequence, cell, cell-line, organism, disease	15
	law, treaty, award, food-dish, dynasty	5

## 3.3. Relations

TABLE 3 The full set of semantic relations used in CAMR

---

:accompanier, :age, \*:aspect, :beneficiary, :cause , :compared-to,  
:consist-of, :cost , \*:cunit, :degree, :destination, :direction,  
:domain, :duration, :example, :extent, :frequency, :instrument,  
:li, :location, :manner, :medium, :mod, :mode, :name, :ord,  
:part, :path, \*:perspective, :polarity, :polite, :poss, :purpose,  
:quant, :range, :source, :subevent, :subset, :superset, \*:tense,  
:time, :topic, :unit, :value \*:dcopy

---

\* marked the new relations added to CAMR

# 4. Statistics and Analysis

- Basic information
- Graph ratio
- Details of Concepts and Relations
- Non-projective Radio

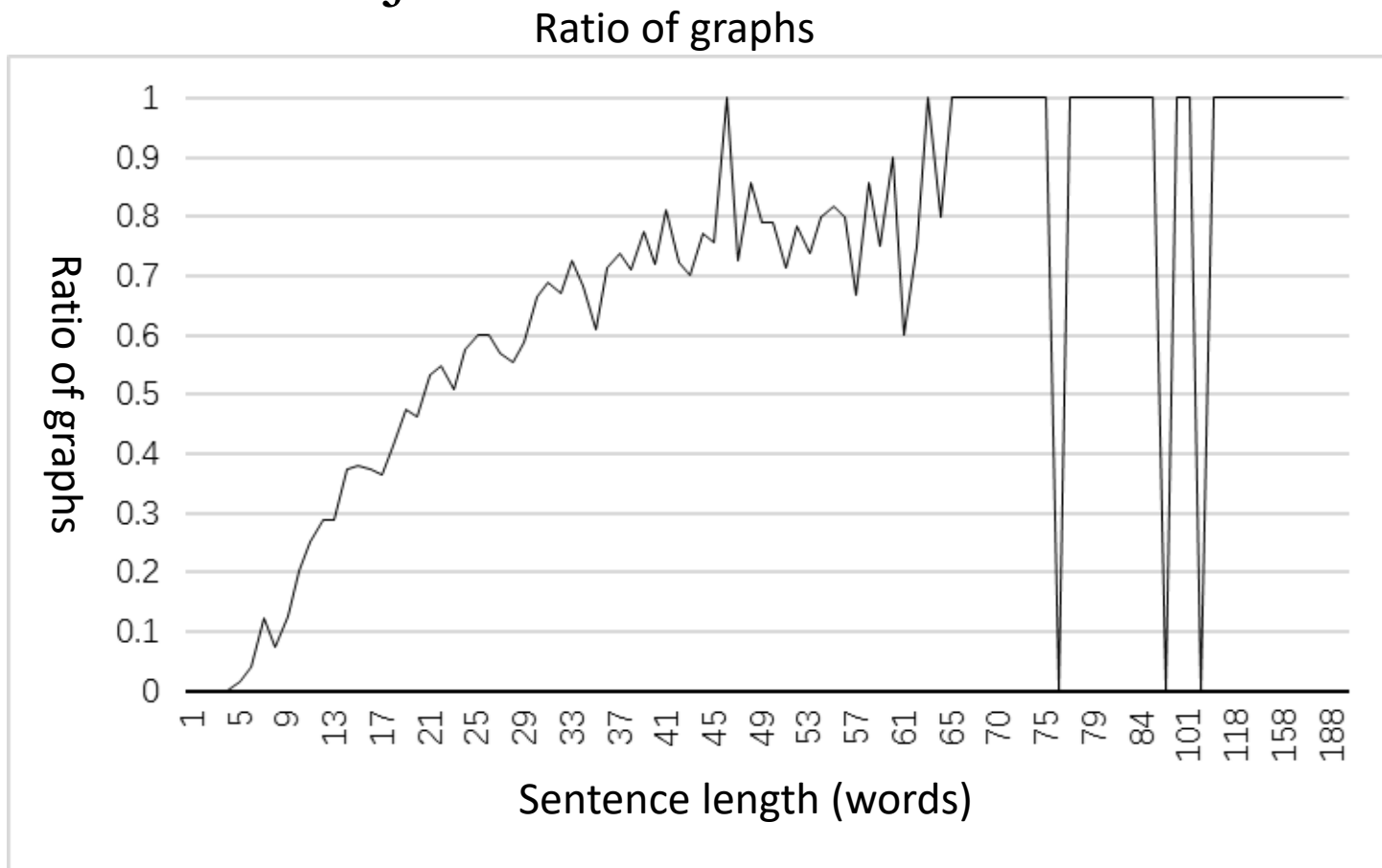
# 4.1. Basic Statistics

AMR Corpus	sentences	graphs	% graphs
Eng_bolt	1,062	722	0.68
Eng_dfa	1,703	898	0.53
Eng_mt09sdl	204	137	0.67
Eng_proxy	6,603	2,954	0.45
Eng_xinhua	741	423	0.57
Eng_little prince	1,562	663	0.43
Eng_total	11,875	5,797	0.49
Chs_little prince	1,562	576	0.36
<b>Chs_CT B1</b>	<b>6,923</b>	<b>3,360</b>	<b>0.48</b>
<b>Chs_CT B2</b>	<b>10,149</b>	<b>4,741</b>	<b>0.47</b>



## 4.2. Graphs

- 4,741 sentences(47%) are graphs
- Others sentences are just trees



## 4.3. Details of Concepts and Relations

TABLE 6 Basic statistics of the CAMR corpus

sentences	words	concepts	relations	re-entrancies	non-tree graphs	abstract concepts	sentences with abstract concepts
10,149	227,661	195,282	228,410	9,449	4,741 (46.71%)	26,269	9,039 (88.95%)
characters	words per sent	concepts per sent	relations per sent	re-entrancies per sent	concrete concepts per sent	abstract concepts per sent	words aligned to relations
347,750	22.43	19.24	22.50	1.99	16.35	2.88	29,533 (12.97%)

# The relations marked by prepositions

P	:arg1	:location	:arg2	:source	:arg0	:time	:direction	:arg3	:beneficiary	:instrument	:purpose
在at	204	431	67	1	20	64	18	2	0	0	2
把take	591	0	10	0	28	0	0	0	0	1	0
从from	28	5	25	153	7	8	0	1	77	1	45
给give	45	0	68	0	8	0	1	0	0	0	6
对to	39	0	12	0	0	0	5	101	11	0	1
向toward	44	0	34	0	1	0	62	3	16	0	1
被by	35	1	3	0	113	0	0	2	0	0	0
到to	48	13	26	0	0	0	0	10	0	0	0
用using	6	0	9	0	8	0	0	0	0	93	0
为for	24	0	18	0	2	0	0	0	41	0	8

# 4.4. Non-Projective Trees

- Non-projective Trees 3,208 (31.6%)

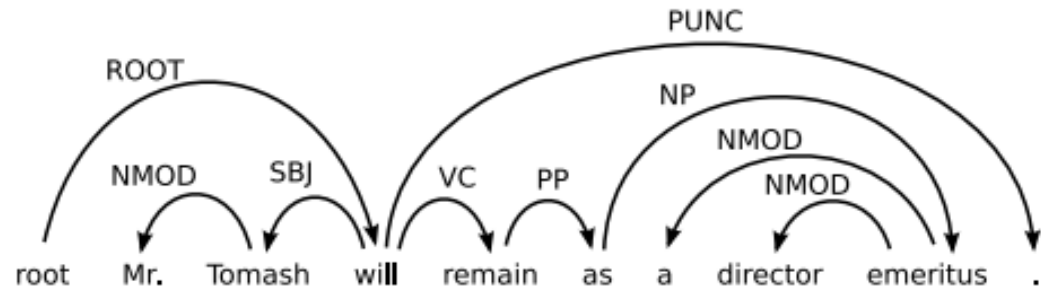


Figure 1: A projective dependency graph.

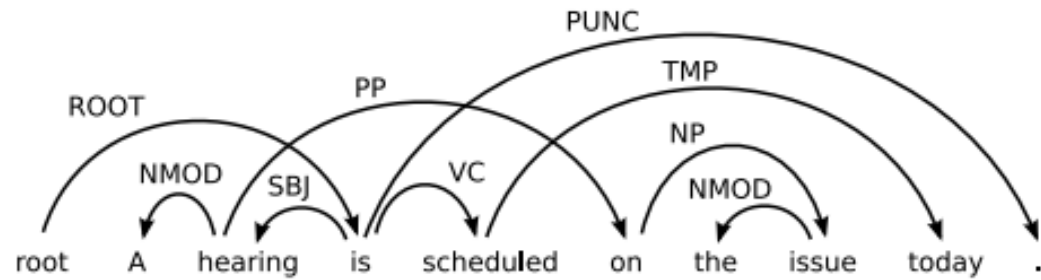
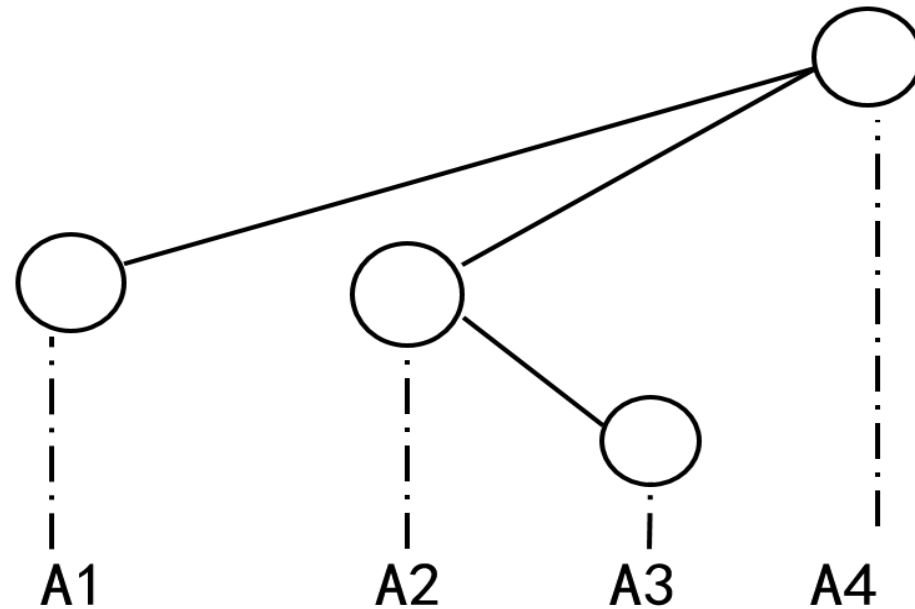


Figure 2: Non-projective dependency graph.

Type	%
Modal word	52.37
Split word	28.49
Topicalization	13.34
Movement	5.14
Other	5.33

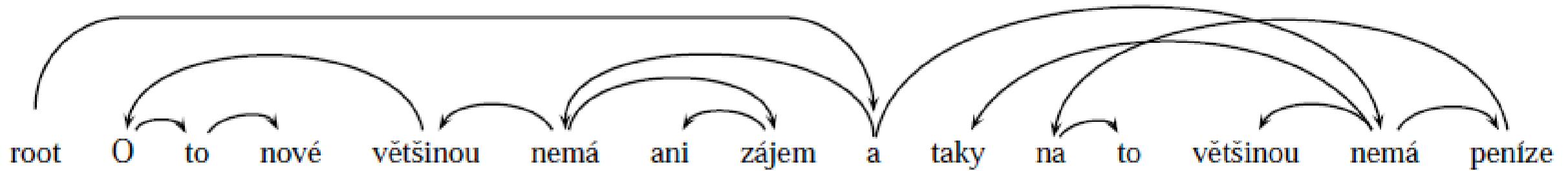
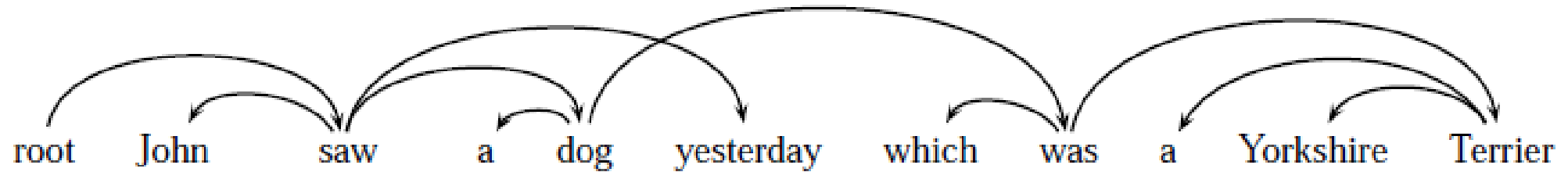
# History

- Tesnière(1959)
- Hays(1964), Robinson(1970): Projective





Hajičová et al. (2004) Prague Treebank, 7,308 sentences,  
23.2% are non-projective.



*He is mostly not even interested in the new things and in most cases, he has no money for it either.*



**Havelka (2007)**  
**investigates 12**  
**languages**



**McDonald (2005)**  
**Non-projective**  
**Dependency Parsing**  
**using Spanning Tree**  
**Algorithms**

Language	Non-Prj Ratio
Spanish	1.72
Japanese	5.29
Bulgarian	5.38
Swedish	9.77
Arabic	11.16
Turkish	11.6
Danish	15.63
Poruguese	18.94
Slovene	22.16
Czech	23.15
German	27.75
Dutch	36.44



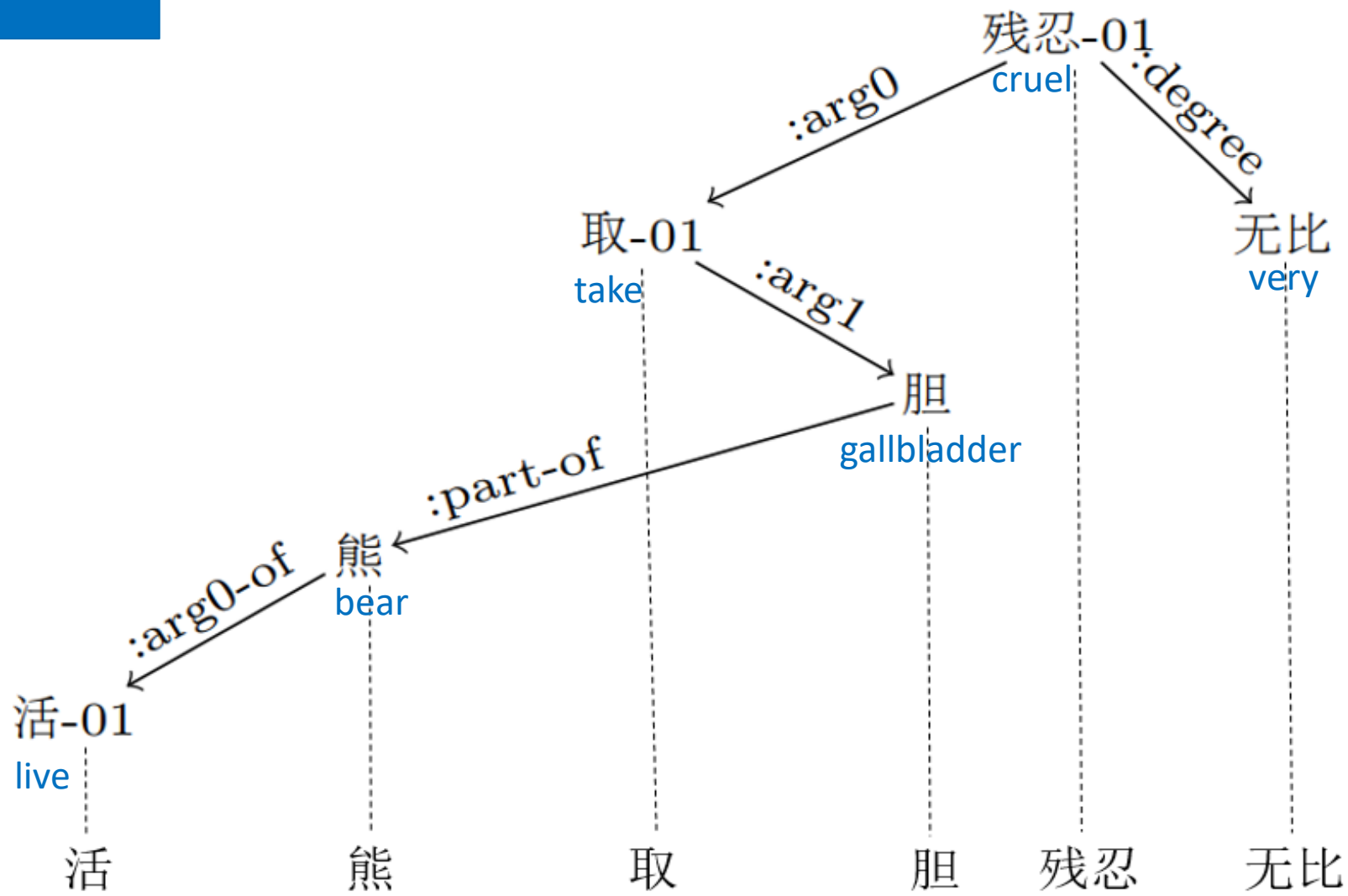
## Zeman (2014) 29 languages, Non-proj arc ratio

Language	% NonPrj. Arc	Language	% NonPrj. Arc	Language	% NonPrj. Arc
Arabic	0.37	Finnish	0.51	Portuguese	1.31
Basque	1.27	German	2.33	Romanian	0.00
Bengali	1.08	Greek(el)	1.17	Russian	0.83
Bulgarian	0.38	<b>Greek</b>	<b>19.58</b>	Slovene	1.92
Catalan	0.00	Hindi	1.12	Spanish	0.00
Czech	1.91	Hungarian	2.90	Swedish	0.98
Danish	0.99	Italian	0.46	Tamil	0.16
Dutch	5.41	Japanese	1.10	Telugu	0.23
English	0.33	Latin	7.61	Turkish	5.33
Estonian	0.07	Persian	1.77		

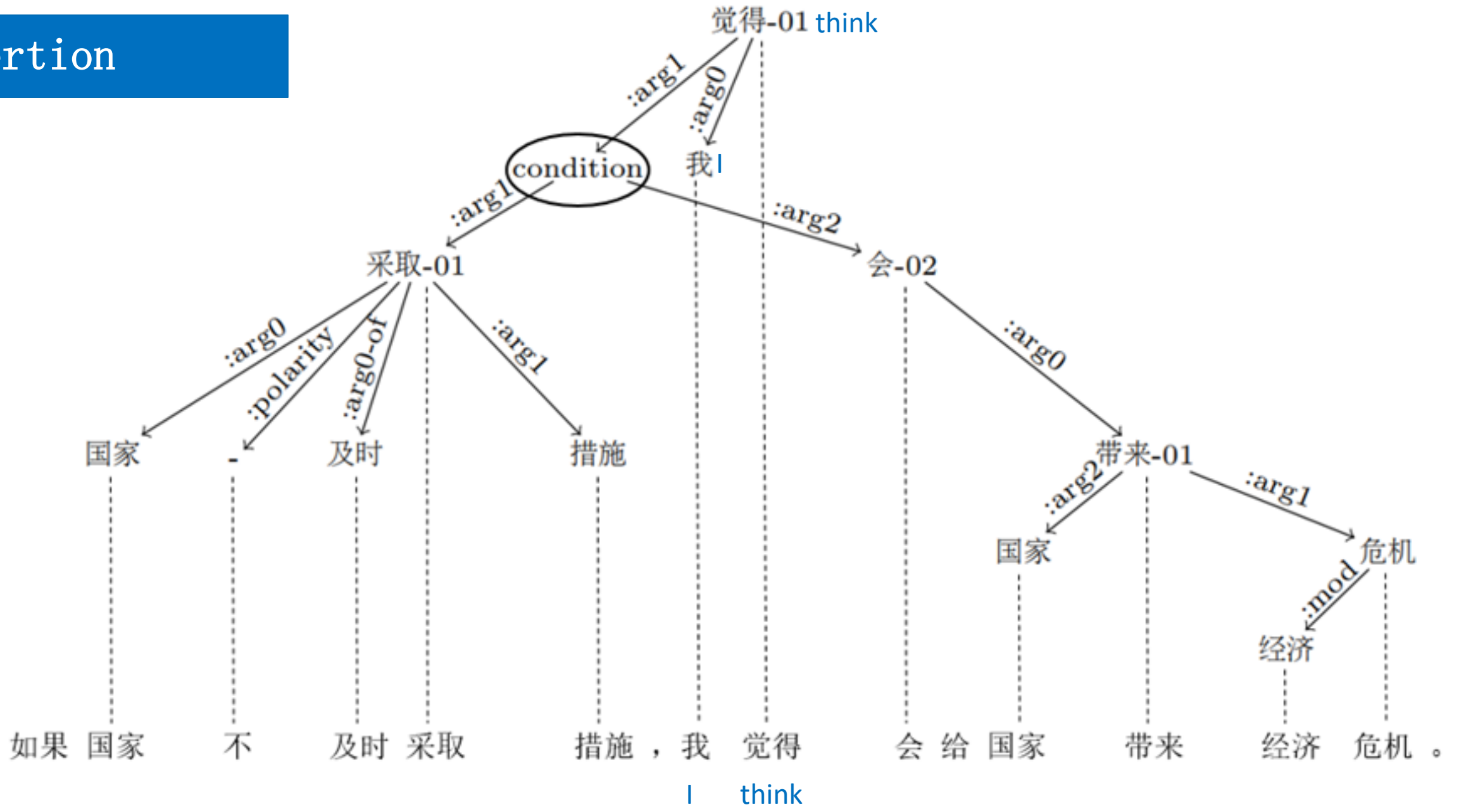
**Chinese ?**



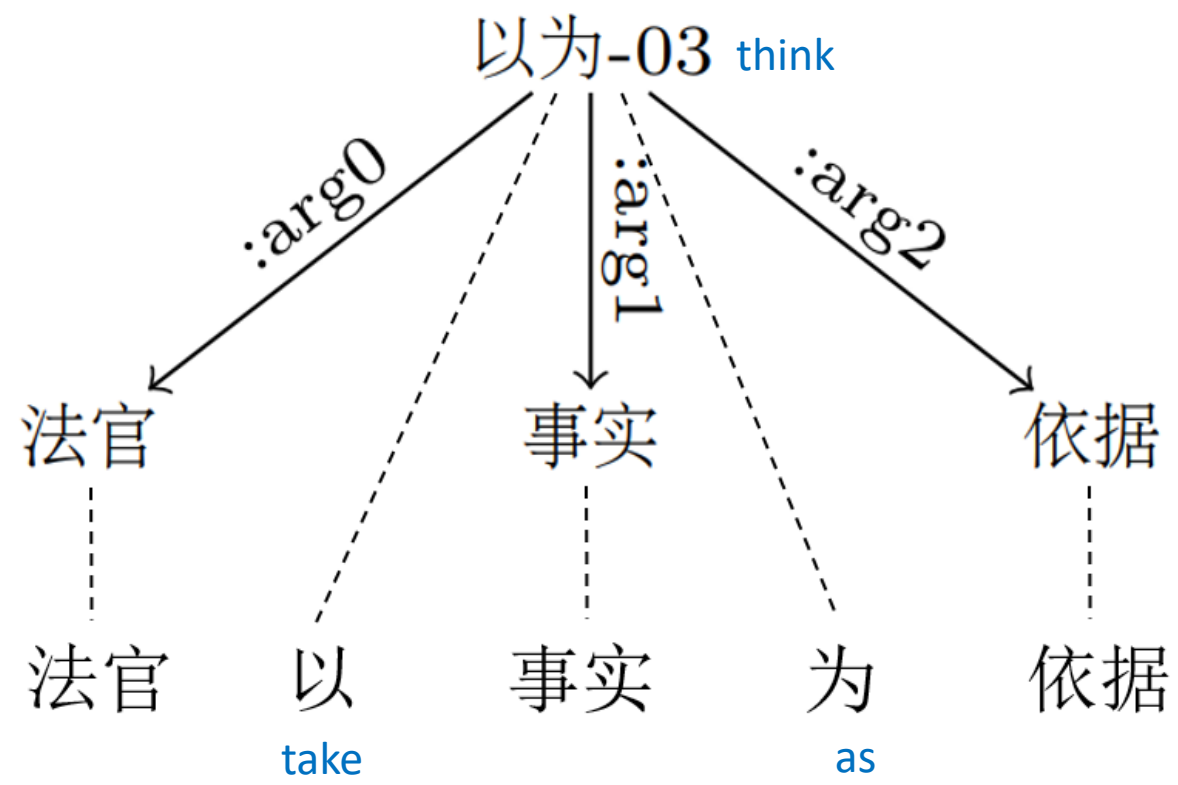
# Part-of



# Insertion



# Frame



# 4.5. Summary

- **Non-projective Ratio: 31.62%**
- **Simple or Complex?**
- **Projective Tree → Non-projective Tree → Graph → AMR**
- **Better description/representation for Chinese**
- **Parsing F-score**
  - **Chinese: Chuan et al.(2018) 0.587; Wu et al.(2019) 0.61**
  - **English: Lyu(2018) 0.74**
- **Linguistics & NLP**

# 5. Conclusions and Future Work

- **Conclusions**
  - **Why graphs instead of trees?**
    - **Argument sharing**
  - **Could the concepts and relations on an AMR graph be aligned back to original words?**
    - **Yes!**
  - **Is AMR universal? Not yet, but close.**
  - **Could it be applied to other languages?**
    - **Chinese is OK.**
  - **Is there something new discovered by Chinese AMR?**
    - **Relation-to-word alignment, Nonprojective,**

# 5.1. Additional Work

- **Ellipsis**
- **Predicate Dictionary**
- **Construct**
- **Discourse**

# Ellipsis

- :dcopy

- Same word, different concepts

(8) 你<sup>1</sup> 的<sup>2</sup> 收入<sup>3</sup> 比<sup>4</sup> 我<sup>5</sup> 高<sup>6</sup>  
you DE income than I high  
“Your income is higher than mine.”

x6/高-01

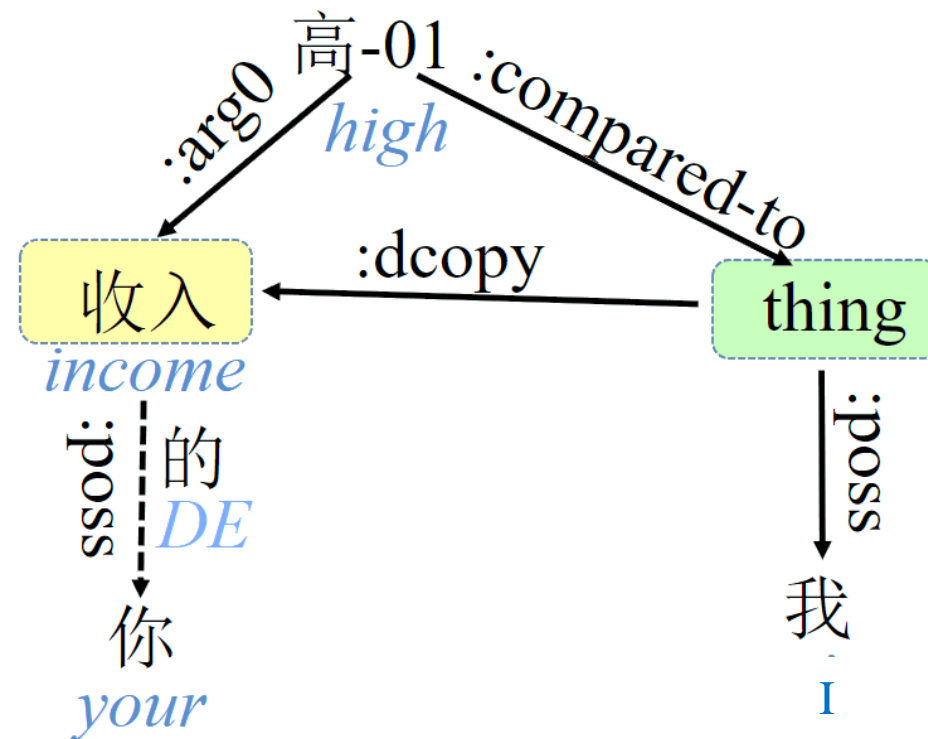
:arg0() x3/收入

:arg1(x2/的) x1/你

:compare-to(x4/比) x8/thing

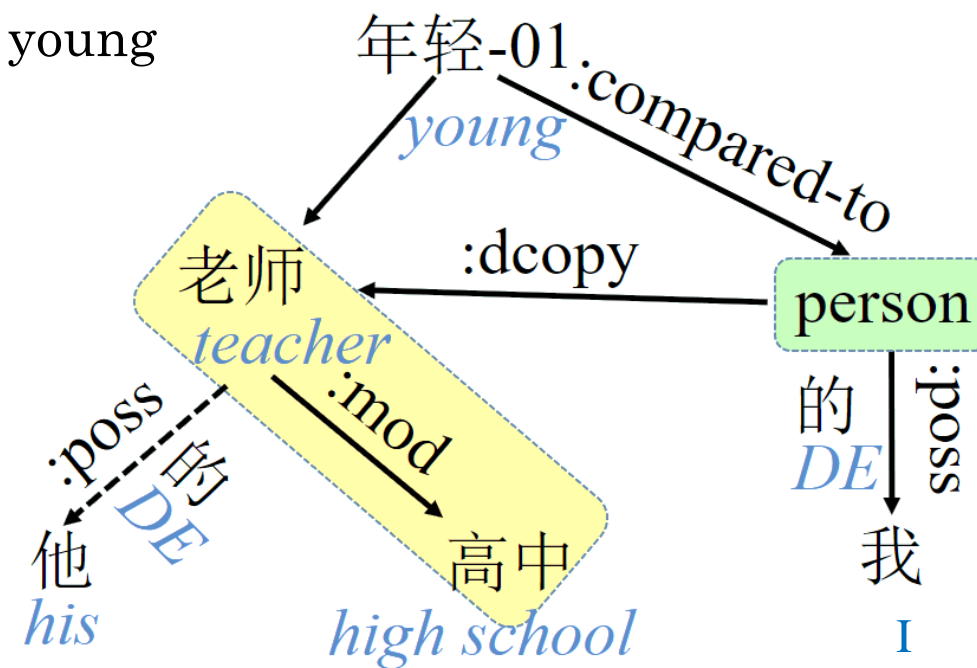
:poss() x5/我

:dcopy() x3\_s/收入



# Not root, not subtree, but part of a tree

- 他<sup>1</sup> 的<sup>2</sup> 高中<sup>3</sup> 老师<sup>4</sup> 比<sup>5</sup> 我<sup>6</sup> 的<sup>7</sup> 年轻<sup>8</sup> 。
- He DE high school teacher than I DE young 。
- His high school teacher is younger than mine.
- :dcopy x3\_x4/高中老师
- This is an issue for dependency grammar.





## 5.2. Rebuild the Predicate Frames

- CPB frame files
  - Only frames, no full senses
- Refined dictionary linked to HowNet
  - 8,470 words
  - 14,389 senses
  - 10,800 frames
  - 14,549 items
- New words(OOV)
  - 人艰不拆, 给力, 妥存
  - 排火车票

## 5.3. Discourse

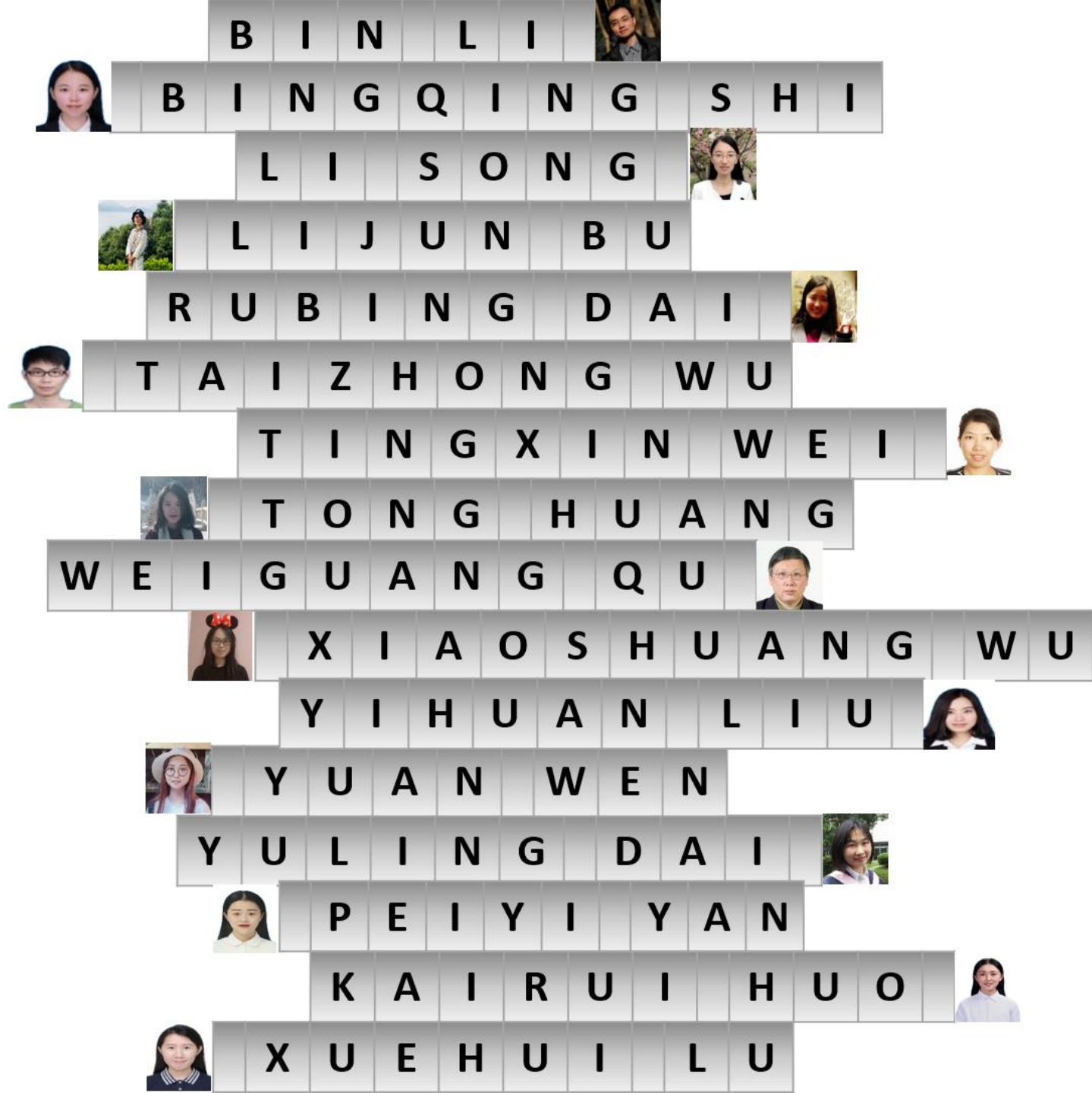
- Annotation Toolkit is hard to design.
- For news data, the sentences are too long.
- Dialogue is much more difficult
  - We tried some small dialogues
  - Ellipsis & Co-reference
  - Literal meaning VS actual meaning in context

## 5.4. Chinese Constructions

- Construction
  - 张三长, 李四 短
  - Jack long, Tom short.
  - Criticize Jack and Tom.
- The Chinese Construction Database developed by Peking University
  - 1,005 Constructions
- After AMR annotation, about 61% could be represented by AMR

# Future Work

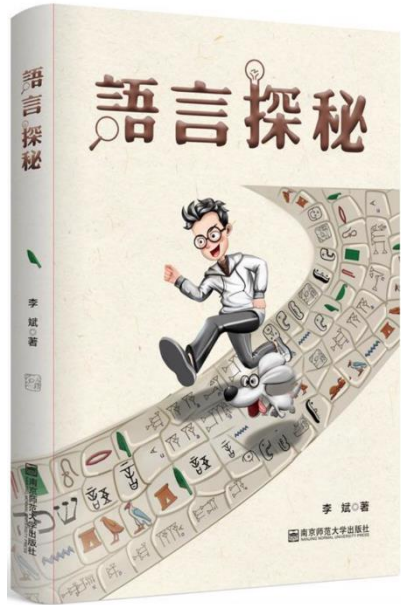
- More Chinese AMR Corpus (LDC 2019.04)
- A better theory of semantic representation
- Chinese AMR parser
- Better dictionary





NNU · 南京师范大学  
NANJING NORMAL UNIVERSITY

STAY  
CONNECTED



Thank You!

Bin Li

Email: [libin.njnu@gmail.com](mailto:libin.njnu@gmail.com)

*IQSS, Harvard University*

*School of Chinese Language & Literature*

*Nanjing Normal University*





# Discussion

- Limitation of dependency structure
  - Lack of phrase
- Discourse annotation
- UMR (Universal Meaning Representation)
- Annotation inter-agreement
- Parsing on graphs



- Bin Li, Yuan Wen, Li Song, Weiguang Qu, Nianwen Xue. Building a Chinese AMR Bank with Concept and Relation Alignments. *Linguistic issues in Language Technology*, Vol.18, 2019.(PDF)
- WU Taizhong, GU Min, ZHOU Junsheng, QU Weiguang, LI Bin, GU Yanhui. Chinese AMR Parsing using Transition-based Neural Network. *Journal of Chinese Information Processing*. 2019, 33(4): 1-11.
- Bin Li, YuanWen, Lijun Bu, Weiguang Qu, Nianwen Xue. Annotating the Little Prince with Chinese AMRs. *Proceedings of LAW X- The 10th Linguistic Annotation Workshop(LAW-2016)*, Aug 11, 2016, Berlin, Germany:7-15.
- LI Bin, WEN Yuan, SONG Li, BU Lijun, QU Weiguang, XUE Nianwen. Construction of Chinese Abstract Meaning Representation Corpus with Concept-to-word Alignment. In *Journal of Chinese Information Processing, China, 2017(6)*
- Yihuan Liu, Bin Li, Peiyi Yan, Li Song, Weiguang Qu. Ellipsis in Chinese AMR Corpus. *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 92–99. Florence, Italy, August 1st, 2019.
- DAI Yuling, LI Bin, DAI Rubin, FENG Minxuan, QU Weiguang. Representation and Analysis of Chinese Function Words Based on Abstract Meaning Representation with Relation Alignment. *CCL(2019)*
- Tong Huang, Bin Li, Peiyi Yan, Yuling Dai, Weiguang Qu. Chinese Constructions Annotation and Analysis Based on the Abstract Meaning Representation. *CCL(2019)*
- Li Song, Yuan Wen, Sijia Ge, Bin Li and Weiguang Qu. An Easier and Efficient Framework to Annotate Semantic Roles: Evidence from the Chinese AMR Corpus(*CLSW2019*)
- Bin Li, Junsheng Zhou, Weiguang Qu. Research on Non-projective Structure Based on the Chinese Abstract Meaning Representation Corpus. *Journal of Chinese Information Processing*, 2018, 32(12).
- CHENG Ning, LI Bin, GE Sijia, HAO Xingyue, FENG Minxuan. Integration of Automatic Sentence Segmentation and Lexical Analysis of Ancient Chinese based on BiLSTM-CRF Model. *CCL(2019)*
- Huidan Xu, Siyu Chen, Jingjing Cai, Lin Cao, Chen Wan, and Bin Li. The Construction and Statistical Analysis of Pre-Qin Ancient Chinese WordNet. (*CLSW2019*)