

# Annotating Chinese Word Senses with English WordNet: A Practice on OntoNotes Chinese Sense Inventories

Hongzhi Xu<sup>1</sup>, Jingxia Lin<sup>2</sup>, Pradhan Sameer<sup>3</sup>, Mitch Marcus<sup>4</sup>, Ming Liu<sup>5</sup>

<sup>1</sup>Shanghai International Studies University

<sup>2</sup>Nanyang Technological University

<sup>3</sup>Linguistic Data Consortium

<sup>4</sup>University of Pennsylvania

<sup>5</sup>The Hong Kong Polytechnic University

# Outline

- Introduction
- Annotating Chinese Word Senses with English WordNet
- Annotation Result Analysis
- Conclusion and Future Work

# Outline

- Introduction
- Annotating Chinese Word Senses with English WordNet
- Annotation Result Analysis
- Conclusion and Future Work

# Introduction

- The goal of this study
  - The possibility of linking word senses across languages
  - Exposing the difficulties of the task
- Different from previous wordnets linking
  - Examples-based linking
  - Annotating Chinese senses with concepts (represented as WordNet synsets)
- Data: OntoNotes 5.0
  - Chinese

# Introduction

- Three languages in OntoNotes 5.0
  - English
  - Chinese
  - Arabic
- Annotations in OntoNotes 5.0
  - Word Senses
  - Part of speech
  - Named entities
  - Syntax (from Treebanks)
  - Propositions and Argument Structures

# Introduction

- Problems existing in OntoNotes
  - Different languages use different sense inventories (no linked)
  - Sense definitions in Chinese sense inventories are sometimes inaccurate (direct sense linking is risky)
- Contributions of this study
  - Providing a link from Chinese word senses to English WordNet synsets
    - Promoting the value of OntoNotes
  - A new methodology of sense/synset linking
  - Exposing interesting linguistic phenomena in lexical semantics
    - Accounting for the difficulties of the sense linking task

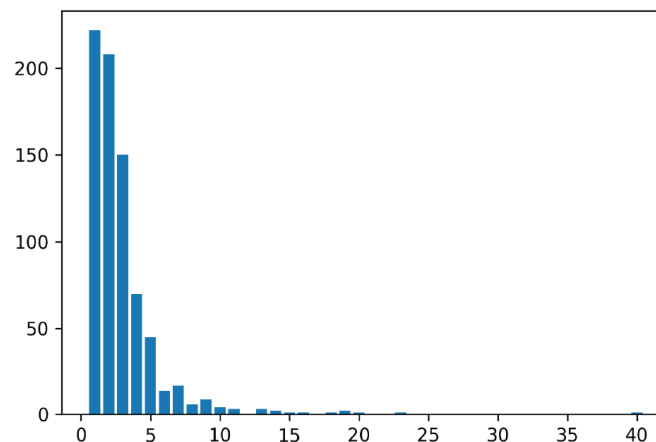
# Outline

- Introduction
- **Annotating Chinese Word Senses with English WordNet**
- Annotation Result Analysis
- Conclusion and Future Work

# Annotating Chinese Word Senses with English WordNet

- Data: Examples from OntoNotes Chinese sense inventories
  - 762 lemmas
  - 2235 senses
  - 2240 examples
    - Multiple usage example for one single sense
    - Empty senses with no examples

- No. of Senses Distribution



#Senses	1	2	3	4	5	$\geq 6$
#Lemmas	222	208	150	70	45	66

# Annotating Chinese Word Senses with English WordNet

- Annotation Procedure

- Only examples with the target words are exposed
  - No original sense information, no pos categories
- Double annotation
- Annotators:
  - Try different keywords for searching
  - Choose the most appropriate synset from WordNet
  - Refer to hypernyms when there are multiple possibilities

意大利 人 更 [爱] 为 足球 而 激动  
yi-da-li ren geng [ai] wei zu-qiu er ji-dong  
Italian people more [love] for football then excite  
'Italians are more excited about football.'



Searching WordNet 3.1  
and find the best synset



Possible keywords:  
Love  
Like  
Tend  
...

# Annotating Chinese Word Senses with English WordNet

- Annotation Result:
  - Overall Agreement on examples

Agreed	867 (38.7%)
Disagreed	1373 (61.3%)
Disagreed POS	244
Total	2240

# Annotating Chinese Word Senses with English WordNet

- Agreement on POS categories
  - 244 disagreed (10.9%)

	adj	adv	noun	verb
adj	105	47	24	81
adv	0	40	5	40
noun	0	0	116	47
verb	0	0	0	868

Confusion matrix of POS identification

颜色	和	气质	[接近]
yanse	he	qizhi	[jiejin]
color	and	temperament	close
'The color and temperament are close.'			

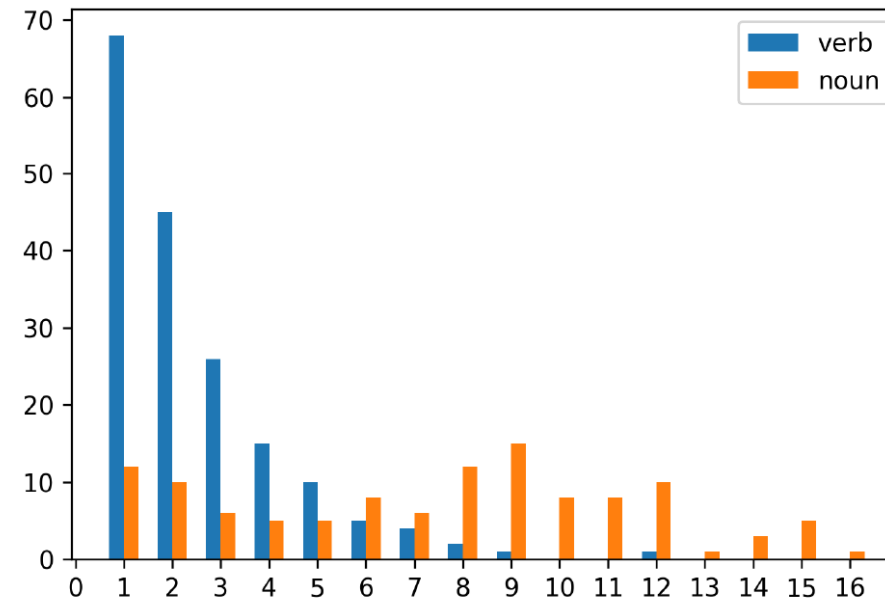
An example of confusion between adj vs. verb

# Outline

- Introduction
- Annotating Chinese Word Senses with English WordNet
- **Annotation Result Analysis**
- Conclusion and Future Work

# Annotation Result Analysis

- Example Level Disagreement
  - How the two selected senses are related in WordNet
  - Three types
    - Hypernym-hyponym
    - Sharing hypernym
    - Unconnected
- Distance Distribution
  - Maximum: 16

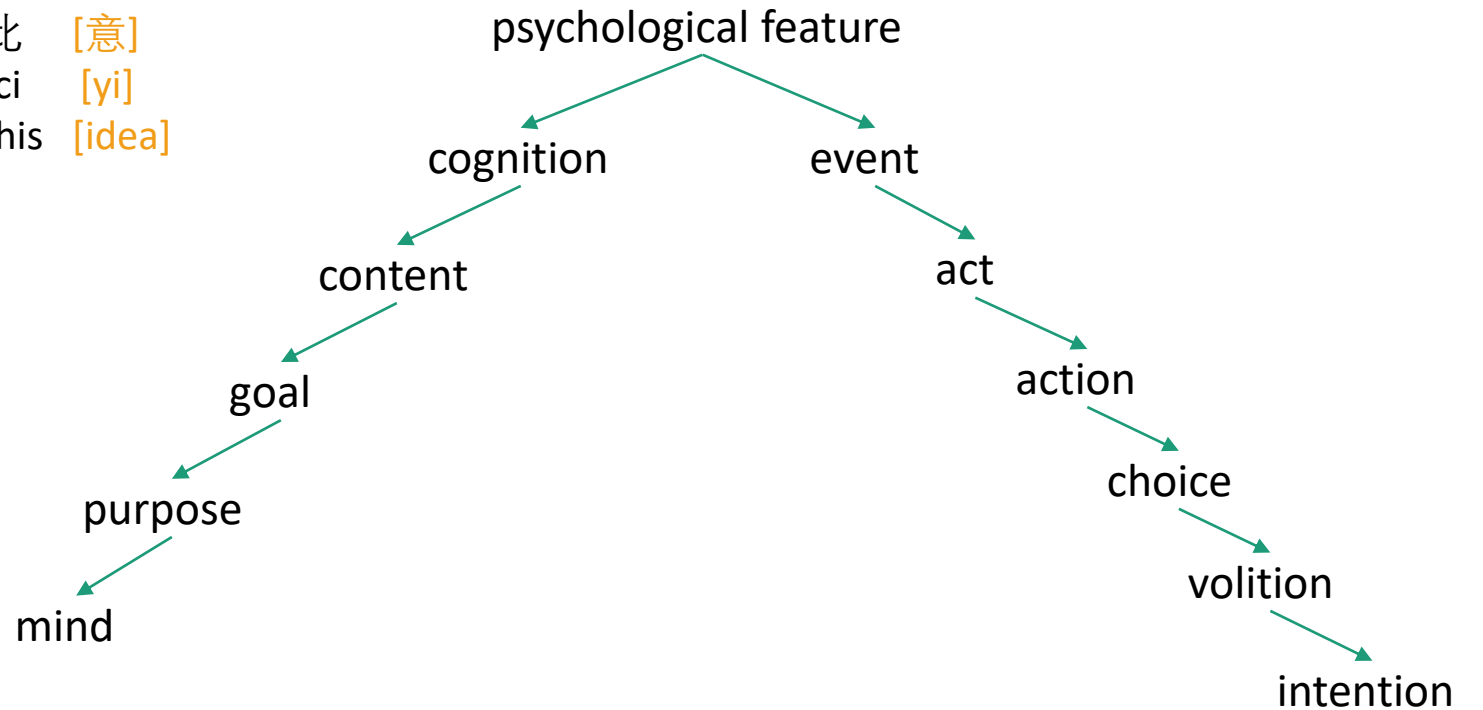


The distribution of path lengths between two selected synsets

# Annotation Result Analysis

- An example of disagreement

他 被 说服 而 打消 此 [意]  
ta bei shui-fu er da-xiao ci [yi]  
He BEI persuade then give-up this [idea]  
'He was persuaded to give up the idea.'



The path between the two selected synsets in WordNet

# Annotation Result Analysis

- Sense Level Disagreement

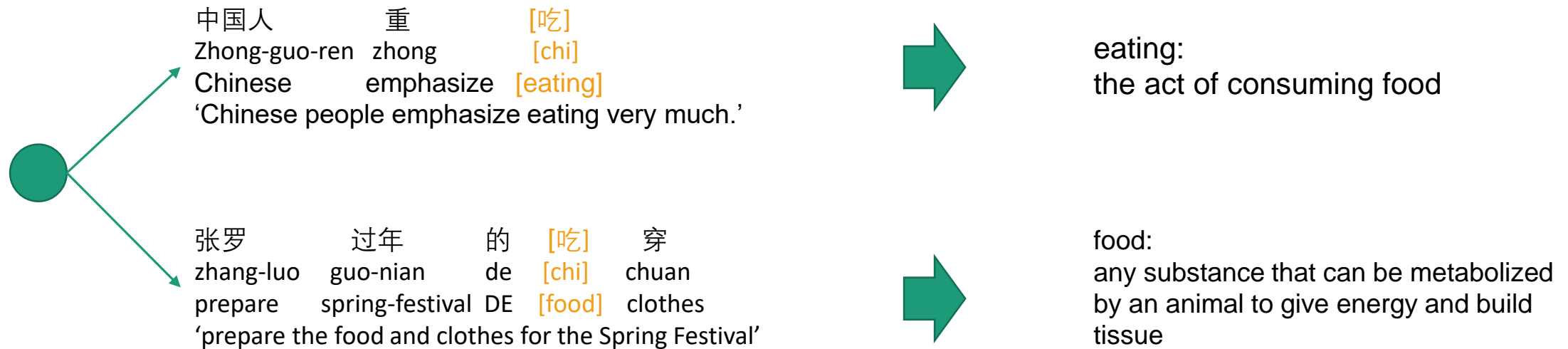
- One-to-many: usage examples of the same sense are annotated with different WN synsets
  - 343 out of 462 cases

#examples	#senses	agreed ratio
2	322	0.295
3	101	0.208
4	27	0.111
≥ 5	12	0.0
Total	462	0.258

- Reflecting interesting semantic relations of target synsets

# Annotating Chinese Word Senses with English WordNet

- An one-to-many example:
  - Two facets of the same semantic frame/event



# Annotation Result Analysis

- Many-to-one mappings
  - 1) Examples from different senses of **the same word** are mapped to the same WordNet synsets (sense merge)
  - 2) Examples from senses of **different words** are mapped to the same WordNet synsets (synonym)

# Annotation Result Analysis

- Many-to-one mappings (synonyms)
  - 爱 ‘ai’ and 好 ‘hao’ (to love/like)
  - 摆 ‘bai’ and 放 ‘fang’ (to place/arrange)
  - 包 ‘bao’ and 裹 ‘guo’ (to wrap)
- One type of Chinese word formation is to concatenate two synonymous characters (words) together, e.g.
  - 爱 ‘ai’ + 好 ‘hao’ = 爱好 ‘ai-hao’ (to love/like)
  - 摆 ‘bai’ + 放 ‘fang’ = 摆放 ‘bai-fang’ (to place/arrange)
  - 包 ‘bao’ + 裹 ‘guo’ = 包裹 ‘bao-guo’ (to wrap)

# Outline

- Introduction
- Annotating Chinese Word Senses with English WordNet
- Annotation Result Analysis
- **Conclusion and Future Work**

# Conclusion and Future Work

- Linking Wordnet synsets are challenging
  - Low agreement
- Example-based sense mapping
  - Avoiding inaccurate sense definitions
  - Allowing more detailed analysis of the disagreement
- Revealing the issues that have long been encountered in lexical semantic studies
  - Causative vs. inchoative
  - Facets of the semantic frames
  - Chinese word formation and POS identification

# Conclusion and Future Work

- Future work
  - Release the data with original annotations
    - <https://github.com/xuhongzhi/ontonotes-sense-mapping>
  - Empty senses in Chinese sense inventories
    - Adding more examples
  - Provide a suggested linking and integrate it in the next version of OntoNotes