Processing Learner Texts: from Annotation to . . .

Weiwei Sun

Wangxuan Institute of Computer Technology
Peking University

@BCLU 2020
give a topic and then discussion about it

Is it a good English sentence?

Can you guess the meaning?
English as a Second Language (ESL)

L1 speakers 29%
L2 speakers 71%

from Ethnologue (2019, 23rd edition);

898.4 million ESL speakers!
Learner texts are everywhere . . .

Language Tests

IELTS

TOEFL

Pearson PTE Academic

ACT

GRE

GMAT

SAT
Learner texts are everywhere . . .

Social Network

Our paper "Semantic Role Labeling for Learner Chinese: the Importance of Syntactic Parsing and the L2-L1 Parallel Data" has been accepted to @emnlp2018. Thanks for my supervisor Weiwei Sun and all the co-authors! See you in Brussels!

1:25 PM - 10 Aug 2018
Semantic Role Labeling for Learner Chinese: the Importance of Syntactic Parsing and L2-L1 Parallel Data

Zi Lin$^{1,2,3}$, Yuguang Duan$^3$, Yuanyuan Zhao$^{1,2,5}$, Weiwei Sun$^{1,2,4}$ and Xiaojun Wan$^{1,2}$

$^1$Institute of Computer Science and Technology, Peking University
$^2$The MOE Key Laboratory of Computational Linguistics, Peking University
$^3$Department of Chinese Language and Literature, Peking University
$^4$Center for Chinese Linguistics, Peking University
$^5$Academy for Advanced Interdisciplinary Studies, Peking University

{zi.lin, ariaduan, ws, wanxiaojun}@pku.edu.cn, zhaoyy1461@gmail.com
Learner texts are everywhere . . .

Number of Submissions per Country/Region (Contact Author)


and perhaps yours
First languages, second languages, cross-lingual transfer

L1 has an influence on L2
Something like

JAPANESE native

CHINESE to learn

L2-CHI, L1-JPN

ENGLISH native

CHINESE to learn

L2-CHI, L1-ENG

learn
Universals

Noun Phrase Accessibility Hierarchy (?, ?)
Subject ≻ direct object ≻ indirect object ≻ oblique ≻ genitive ≻ object of comparison
If a language can relativize on a position on the hierarchy, then any other higher position can also be relativized on.

(1) a. the man who I am taller than △ object of comparison
   b. the man whose father I know △ genitive

For example, if a language allows (1a), then it allows (1b).

A universal of SLA
L2 learners find relative clauses higher on the hierarchy easier to acquire.
There is naturally a need to automatically annotate second language data with rich lexical, syntactic, semantic and even pragmatic information.
There is naturally a need to automatically annotate second language data with rich lexical, syntactic, semantic and even pragmatic information.

High-performance automatic annotation,
- from an engineering perspective, enables deriving high-quality information by structuring this specific type of data, and
- from a scientific perspective, enables quantitative studies for Second Language Acquisition, which is complementary to hands-on experiences in interpreting second language phenomena.
There is naturally a need to automatically annotate second language data with rich lexical, syntactic, semantic and even pragmatic information.

High-performance automatic annotation,

- from an engineering perspective, enables deriving high-quality information by structuring this specific type of data, and
- from a scientific perspective, enables quantitative studies for Second Language Acquisition, which is complementary to hands-on experiences in interpreting second language phenomena.

Is this talk about annotating grammatical errors? Not really.
**Data: Reddit (https://www.reddit.com)**

Large-scale L2 texts are available!

250M native and non-native English sentences (3.8B tokens), covering over 45K authors from 50 countries (Rabinovich et al., 2018)

<table>
<thead>
<tr>
<th>L1</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td><em>I have to go to the Dr. to do a rapid check on my heart stability.</em></td>
</tr>
<tr>
<td>French</td>
<td><em>Maybe put every name through a manual approbation pipeline so it ensures quality.</em></td>
</tr>
<tr>
<td>French</td>
<td><em>Polls have shown public approbation for this law is somewhere between 58% and 65%, and it has been a strong promise during the presidential campaign.</em></td>
</tr>
<tr>
<td>Italian</td>
<td><em>The event was even more shocking because the precedent evening he wasn’t sick at all.</em></td>
</tr>
</tbody>
</table>
(Automatic) Annotation for learner languages

**POS tags**
- Syntactic dependencies (Ragheb & Dickinson, 2012)
- (Dickinson & Ragheb, 2013)
- (Ragheb & Dickinson, 2014)

**UD for learner Chinese**
- (Lee et al., 2017)

**Phrase Structure**
- (Nagata & Sakaguchi, 2016)

**Universal Dependencies**
- (Berzak et al., 2016)

Annotated L2 texts are available!

<table>
<thead>
<tr>
<th>Corpus</th>
<th>L2</th>
<th>L1</th>
<th>#sent</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLE</td>
<td>English</td>
<td>Multiple</td>
<td>5,124</td>
<td>Universal Dependency</td>
</tr>
<tr>
<td>Konan-JIEM</td>
<td>English</td>
<td>Japanese</td>
<td>3,260</td>
<td>Phrase Structure</td>
</tr>
<tr>
<td>ICNALE</td>
<td>English</td>
<td>10 Asian countries</td>
<td>1,930</td>
<td>Phrase Structure</td>
</tr>
<tr>
<td>Chinese-CFL</td>
<td>Chinese</td>
<td>Multiple</td>
<td>451</td>
<td>Universal Dependency</td>
</tr>
</tbody>
</table>
Data: Lang-8 (http://lang-8.com)

Large-scale L2-L1 parallel texts are available!
6.8M English sentence pairs and 720K Chinese sentence pairs (Mizumoto et al., 2011; Y. Zhao et al., 2018)

| L2 speaker | 城市里的人能度过多方面的生活。 | 城市里的人能过丰富多彩的生活。 |
| L2 speaker corrected | You know what should I done. | You know what I should have done. |

Post in the language that you are learning. Native speakers correct your writing!
Patterns of cross-lingual transfer

L2-English

Standard

**NP**

```plaintext
(x0:NN x1:NN) → x1 x0
```

**VP**

```plaintext
(x0:VB x1:ADVP x2:NP) → x1 x0 x2
```

Mandarin

面

包

香

气

French

faire

souvent

du

sport

English

smell of bread

L2-English

Standard

**NP**

**VP**

sports

often

play

...
Using patterns

Learner Texts

Tree-String Patterns → Vectors → Phylogenetic Structure

Hindi, German, Persian, Italian, Portuguese (Brazilian), French, Spanish, Ukrainian, Polish, Russian

Zhao et al. (2020); arXiv:2007.09076
(Automatic) Annotation for learner languages

- POS tags
  - Syntactic dependencies
    - (Ragheb & Dickinson, 2012)
    - (Dickinson & Ragheb, 2013)
    - (Ragheb & Dickinson, 2014)

- UD for learner Chinese
  - (Lee et al., 2017)

- English Resource Semantics
  - (Y. Zhao et al., 2020)
  - Negation Scope Resolution
    - (undersubmission)

- Phrase Structure
  - (Nagata & Sakaguchi, 2016)
  - Universal Dependencies
    - (Berzak et al., 2016)

- Semantic Roles
  - (Lin et al., 2018)

- POS tags
  - (Díaz-Negrillo et al., 2010)

**capture “meanings”**
Can human understand interlanguage robustly?

😊 It is difficult to define the syntactic formalism of learner language.

Grammaticality judgement

😊 But sometimes we can understand what they mean . . .
Can human understand interlanguage robustly?

😊 It is difficult to define the syntactic formalism of learner language.

Grammaticality judgement

😊 But sometimes we can understand what they mean …

I HAVE 27 YEARS

DO YOU MEAN "I AM 27 YEARS OLD"?
Research questions

- How can we capture meanings of L2s?
  How can we annotate L2 texts?
  Are there many differences from annotating L1 texts?

- How *badly* does an L1 data-trained semantic parser perform?
  Can state-of-the-art grammatical error correction systems help?

- What role does syntactic parsing play in processing L2 texts?
  What role does cross-lingual transfer play?
Research questions

▶ How can we capture meanings of L2s? How can we annotate L2 texts? Are there many differences from annotating L1 texts?

▶ How *badly* does an L1 data-trained semantic parser perform? Can state-of-the-art grammatical error correction systems help?

▶ What role does syntactic parsing play in processing L2 texts? What role does cross-lingual transfer play?
Shallow and not-that-shallow meaning representations

Semantic Role Labeling

ARG0
Some boys want to go.

ARG1

Bi-lexical Semantic Dependency

ARG1
BV
ARG1
ARG2
Some boys want to go.

Conceptual Graphs

_some_q

_want_v_to

_boyn_1

_go_v_1
Language structures involving an interface between different language modules, like *syntax-semantics interface* and *semantics-pragmatics interface*, are less likely to be acquired completely than structures that do not involve this interface; see e.g. Sorace (2011).
Literal meaning versus intended meaning

Literal Meaning

- conventional meaning
- sentence meaning
- linguistic code features

Intended Meaning

- speaker meaning
- interpretation
- author’s intention

often ≠

give a topic and then discussion about it.

Give a topic and then discuss it.
SemBanking in Natural Language Processing

- Compositionally:
  - manually-annotated
  - PropBank (Kingsbury & Palmer, 2002)
  - FrameNet (Baker et al., 1998)
  - Redwoods Treebank (Oepen et al., 2004)
  - TREPIL (Rosén et al., 2005)
  - Groningen Meaning Bank (Basile et al., 2012)
  - Abstract Meaning Representation (Banarescu et al., 2013)

- Non-compositionally:
  - comprehensiveness
  - consistency
  - scalability

Two languages, three tasks

Chinese as a Second Language
- Semantic Role Labeling
- Negation Scope Resolution

English as a Second Language
- Compositional Semantics
- Semantic Graph Parsing
SemBanking in Natural Language Processing

Compositionally

- manually-annotated
- grammar-based

- PropBank
  (Kingsbury & Palmer, 2002)
- FrameNet
  (Baker et al., 1998)

Non-compositionally

- Redwoods Treebank
  (Oepen et al., 2004)
- TREPIL
  (Rosén et al., 2005)
- Groningen Meaning Bank
  (Basile et al., 2012)

- Abstract Meaning Representation
  (Banarescu et al., 2013)

Semantic Role Labeling

- **Argument (AN):** Who did what to whom?
- **Adjunct (AM):** When, where, why and how?

Our work

Data source

Initial collection: 1,108,907 pairs
Clean up: 717,241 pairs
Manual selection: relatively sized
Annotation: typologically different mother tongues

Languages included:
- Chinese
- Russian
- Arabic
- Japanese
- English
- Sino-Tibetan
- Slavic
- Semitic
- Germanic

Legend:
- Russian
- Arabic
- Japanese
- English
Inter-annotator agreement

- **Annotator:** two students majoring in linguistics
- **The first 50-sentence trial set:** adapting and refining the **Chinese PropBank** specification
- **The rest 100-sentence set:** reporting the inter-annotator agreement

![Bar chart showing inter-annotator agreement for different languages: ENG, JPN, RUS, ARA. The values range from 90 to 100, with ENG at 94, JPN at 98, RUS at 96, and ARA at 98. The chart indicates high agreement.]
Negation Scope Resolution

- **Negation cue**: linguistic unit that expresses negation.
- **Negation event**: the event related to a cue.
- **Negation scope**: the maximum part(s) of the sentence that are influenced or negated by negation cue.

(2) a. We **needs** actions and **not** thoughts.
    b. He **failed** to catch the first train.
    c. This is an **un**clean desk.
    d. 换言之，没有 宗教生活与日常生活差距。
    e. 换言之， 宗教生活与日常生活之间 **没有** 距离。

**Our work**

M. Zhang, W. Wang, Y. Zhao, S. Sun, W. Sun and X. Wan. Negation Scope Resolution for Chinese as a Second Language. (under submission)
### Inter-annotator agreement

<table>
<thead>
<tr>
<th></th>
<th>C. F₁</th>
<th>S. F₁</th>
<th>T. F₁</th>
<th>C. Kappa</th>
<th>S. Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD-SCO (2012)</td>
<td>94.88</td>
<td>85.04</td>
<td>91.53</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SFU Review (2012)</td>
<td>92.79</td>
<td>81.88</td>
<td>-</td>
<td>92.70</td>
<td>87.20</td>
</tr>
<tr>
<td>BioScope (2008)</td>
<td>98.65</td>
<td>95.91</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CNeSp (2015)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>95.00</td>
<td>93.00</td>
</tr>
<tr>
<td>L2-CHI, L1-ENG</td>
<td>100.00</td>
<td>92.55</td>
<td>97.12</td>
<td>100.00</td>
<td>96.13</td>
</tr>
<tr>
<td>CHI₄L₂⇒L₁, L1-ENG</td>
<td>100.00</td>
<td>92.55</td>
<td>97.71</td>
<td>100.00</td>
<td>95.65</td>
</tr>
<tr>
<td>L2-CHI, L1-JPN</td>
<td>100.00</td>
<td>90.09</td>
<td>94.62</td>
<td>100.00</td>
<td>92.15</td>
</tr>
<tr>
<td>CHI₄L₂⇒L₁, L1-JPN</td>
<td>100.00</td>
<td>90.09</td>
<td>94.35</td>
<td>100.00</td>
<td>92.32</td>
</tr>
</tbody>
</table>

- Inter-annotator agreement w.r.t. negation cue and scope.
- Previous negation corpora reported cue-level $F_1$ (C. $F_1$) at 91%-95%, scope-level $F_1$ (S. $F_1$) at 76%-85%, token-level $F_1$ (T. $F_1$) at 88%-92%, and kappa at 87%-91%.
SemBanking in Natural Language Processing

- Compositionally:
  - PropBank (Kingsbury & Palmer, 2002)
  - FrameNet (Baker et al., 1998)

- Manually-annotated

- Grammar-based:
  - Redwoods Treebank (Oepen et al., 2004)
  - TREPIL (Rosén et al., 2005)
  - Groningen Meaning Bank (Basile et al., 2012)

- Non-compositionally:
  - Abstract Meaning Representation (Banarescu et al., 2013)

Treebank of Learner English (Berzak et al., 2016)

TLE (http://esltreebank.org/)
- a collection of 5,124 ESL sentences
- manually annotated with POS tags and UD trees
- in original and error corrected forms.

Instructions
Search for sequences of words, universal/PTB POS tags and relation labels. Regular expressions are supported for searching words.

Examples
- *see it* matches the string "see it"
- *see DET NOUN* matches "see that show", "see the sign", etc.
- *\w+ing something* matches "seeing something", "seeking something", etc.
- *amod NNS* matches adjectival modifier followed by a plural noun, such as "best cakes", "bigger halls", etc.

ESL filters and highlighting
Filter query results to sentences with a specific grammatical error and/or specific native language.
An empty query will retrieve all the sentences that correspond to the specified filters. Highlight grammatical errors and show annotations of sentence corrections using the checkboxes.

Corpus (UD v2.3)
- *ESL* is the Treebank of Learner English
- *English* is the EWT UD corpus
SemBanking by integrating TLE and ERG

Input Sentence → ACE/PET parser (Packard, 2013) → Meaning Representation 1
Meaning Representation 2...
Meaning Representation \( K \) → Reranker → Meaning Representation
SemBanking by integrating TLE and ERG

English Resource Grammar (ERG; Flickinger (1999))

Hand-crafted computational grammar;
HPSG-based;
25+ person years

Input Sentence

ACE/PET parser (Packard, 2013)

Meaning Representation 1
Meaning Representation 2
...
Meaning Representation $K$

Reranker

Meaning Representation
SemBanking by integrating TLE and ERG

Hand-crafted computational grammar;
HPSG-based;
25+ person years

English Resource Grammar (ERG; Flickinger (1999))

Input Sentence

ACE/PET parser (Packard, 2013)

Meaning Representation 1
Meaning Representation 2
... 
Meaning Representation $K$

Reranker

Meaning Representation
SemBanking by integrating TLE and ERG

- Hand-crafted computational grammar; HPSG-based; $25^+$ person years
- English Resource Grammar (ERG; Flickinger (1999))
- Input Sentence
  - ACE/PET parser (Packard, 2013)
  - Meaning Representation 1
  - Meaning Representation 2
  - ... Meaning Representation $K$
  - Reranker
  - Meaning Representation
SemBanking by integrating TLE and ERG

Hand-crafted computational grammar; HPSG-based; 25+ person years

English Resource Grammar (ERG; Flickinger (1999))

ACE/PET parser (Packard, 2013)

Meaning Representation 1
Meaning Representation 2
...
Meaning Representation $K$

Reranker

Manually annotated

Universal Dependencies (UD)

Meaning Representation
SemBanking by integrating TLE and ERG

Hand-crafted computational grammar; HPSG-based; 25+ person years

English Resource Grammar (ERG; Flickinger (1999))

ACE/PET parser (Packard, 2013)

Input Sentence

Meaning Representation 1
Meaning Representation 2
... 
Meaning Representation K

Reranker

Universal Dependencies (UD)

manually annotated

K candidate graphs $G_1, G_2, ..., G_K$ and gold UD tree $T$:

$$\hat{G} = \arg\max_{1 \leq i \leq K} \text{SCORE}(G_i, T)$$

$$\text{SCORE}(G_i, T) = W^T F(f_{G_i}, f_T)$$
SemBanking by integrating TLE and ERG

Hand-crafted computational grammar;
HPSG-based;
\(25^{+}\) person years

English Resource Grammar (ERG; Flickinger (1999))

Input Sentence

ACE/PET parser (Packard, 2013)

Meaning Representation 1
Meaning Representation 2
... 
Meaning Representation \(K\)

Reranker

Universal Dependencies (UD)

\(K\) candidate graphs \(G_1, G_2, \ldots, G_K\)

\[ \hat{G} = \arg \max_{1 \leq i \leq K} \text{score}(G_i, T) \]

Target Representation: Elementary Dependency Structures (EDS; Oepen and Lønning (2006)) and several others

\[ \text{score}(G_i, T) = W^T \mathcal{F}(f_{G_i}, f_T) \]
Evaluation on DeepBank

Inter-Annotator Agreement (IAA) of EDM is reported in Bender et al. (2015).
Research questions

- How can we capture meanings of L2s? How can we annotate L2 texts? Are there many differences from annotating L1 texts?

- How **badly** does an L1 data-trained semantic parser perform? Can state-of-the-art grammatical error correction systems help?

- What role does syntactic parsing play in processing L2 texts? What role does cross-lingual transfer play?
Two languages, three tasks

Chinese as a Second Language
- Semantic Role Labeling
- Negation Scope Resolution

English as a Second Language
- Compositional Semantics
  - Semantic Graph Parsing

```
_some_q
  BV
    ARG1
      _boy_n_1
        ARG1
          _go_v_1
        ARG2
  _want_v_to
```
Evaluation – **SMATCH** (Cai & Knight, 2013)

**Gold Graph**
- **ARG0**: a: WANT
- **ARG0**: b: BOY
- **ARG0**: c: GO

**Predicted Graph**
- **ARG0**: a': WANT
- **ARG0**: c': GO
- **ARG0**: b': BOY
Evaluation – SMATCH (Cai & Knight, 2013)

![Graph](image)

**Gold Graph**

1: (a, b, ARG0)  
2: (a, c, ARG1)  
3: (c, b, ARG0)

**Predicted Graph**

1: (a', b', ARG0)  
2: (a', c', ARG1)  
3: (c', b', ARG0)

\[ v_{xx'} = 1, x \in \{a, b, c\} \]
\[ t_{11} = 1 \]
\[ t_{22} = 1 \]
- Most of the structure is **good**.
- We need to focus on **errors**!
- Most of the structure is **good**.
- We need to focus on **errors**!  

Add weights
We went this discuss.
We went to this discussion.

Statistical Machine Translation (SMT) discuss ↔ discussion

Paraphrase Table

Large-scale L2-L1 data exists: LANG-8
We went this discuss.

We went to this discussion.

Statistical Machine Translation (SMT)

Paraphrase Table

Large-scale L2-L1 data exists: LANG-8
Semantic parsers (Koller et al., 2019)

Modeling syntactico-semantic derivation/composition vs derived/composed structures
The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined.

B. Partee

brown + cow = brown cow

Fodor and Lepore (2002)
and then

and then

discussion about it
Hyperedge Replacement Grammar-based parser

and then

and then

discussion about it
and then discussion about it

and
then

NP

arg1
arg2

BV

NP

discussion

pronoun

R-INDEX

CONJ

and
discussion
about
pronoun

BV

NP

arg1
arg2

ndef
and then discussion about it
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure

Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure

Systems

CoNLL Shared Task on Cross-Framework Meaning Representation Parsing.
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure

Systems
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure

\[ \text{arg max}_{\text{discussion}} \]

**Biaffine Score**: \( \text{ScoreEdge}(c_4, n_1, c_6) \)

**Systems**

Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure
Factorization-based parser

- Inspired by graph-based dependency parsers
- Explicitly models the target structure

\[ \text{SCORE}_\text{Edge}(\text{and}_c \rightarrow \text{discussion}_n.1) \]

Systems

Results — Parsing to literal meanings

<table>
<thead>
<tr>
<th>Model Type</th>
<th>DeepBank</th>
<th>L1</th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composition-Based NO</td>
<td>89.04%</td>
<td>82.14%</td>
<td>85.75%</td>
</tr>
<tr>
<td>Factorization-Based NO</td>
<td>90.96%</td>
<td>84.48%</td>
<td>87.86%</td>
</tr>
<tr>
<td>composition-based YES</td>
<td>71.46%</td>
<td>79.77%</td>
<td>76.66%</td>
</tr>
<tr>
<td>factorization-based YES</td>
<td>73.55%</td>
<td>80.27%</td>
<td>77.75%</td>
</tr>
</tbody>
</table>

- DeepBank: Model for deep knowledge representation.
- L1: Lower level of analysis.
- L2: Higher level of analysis.

Legend:
- Composition-Based
- Factorization-Based
# Results — Parsing to literal meanings

## Composition-Based Factors

<table>
<thead>
<tr>
<th>Model</th>
<th>Error-oriented</th>
<th>Node</th>
<th>Edge</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
<td>89.04</td>
<td>82.14</td>
<td>85.75</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>71.46</td>
<td>79.77</td>
<td>76.66</td>
</tr>
</tbody>
</table>

## Factorization-Based Factors

<table>
<thead>
<tr>
<th>Model</th>
<th>Error-oriented</th>
<th>Node</th>
<th>Edge</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
<td>90.96</td>
<td>84.48</td>
<td>87.86</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>73.55</td>
<td>80.27</td>
<td>77.75</td>
</tr>
</tbody>
</table>
ArtOrDet:
It is obvious to see that (internet → the internet) saves people time and also connects people globally.

WOinc:
(Someone having what kind of disease → What kind of disease someone has) is a matter of their own privacy.
manually annotated

Give a topic and then discuss it.

Original Sentence

grammatical error correction (GEC)

Zhao et al. (2019)
give a topic and then discuss it.

Chollampatt and Ng (2018)
give a topic and then discuss it.

Semantic Parser

given a topic and then discuss it.
Results — Parsing to intended meanings

<table>
<thead>
<tr>
<th>Model</th>
<th>$F_{0.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chollampatt and Ng (2018)</td>
<td>45.36</td>
</tr>
<tr>
<td>W. Zhao et al. (2019)</td>
<td>61.15</td>
</tr>
</tbody>
</table>
How can we capture meanings of L2s?
How can we annotate L2 texts?
Are there many differences from annotating L1 texts?

How *badly* does an L1 data-trained semantic parser perform?
Can state-of-the-art grammatical error correction systems help?

What role does syntactic parsing play in processing L2 texts?
What role does cross-lingual transfer play?
Two languages, three tasks

Chinese as a Second Language
- Semantic Role Labeling
- Negation Scope Resolution

English as a Second Language
- Compositional Semantics
- Semantic Graph Parsing

用 汉语 也 说话 快 对 我 来 说 很 难
Three SRL systems

**Parsers**
- Berkeley parser
- Minimal span-based parser

**Systems**
- PCFGLA-parser-based SRL system
- Neural-parser-based SRL system
- Neural syntax-agnostic SRL system

- Trained on Chinese TreeBank that has SRL in CPB
- Trained on Chinese PropBank (CPB)
用汉语也说话快，对我来说来说很难
用汉语也说话快对我来说很难
Syntax-agnostic SRL

用汉语也说话快对我来说说话很难
用汉语 using Chinese

也 also

说话快 speaking quickly

对我来说 for me

很 very

难 hard
Syntax-based SRL

用汉语
using Chinese

也
also

说话快
speaking quickly

也
also

对我来说
for me

对 me

很
very

REL
也
also

difficult

hard

very

to me

for me

speaking quickly

also
difficult
Speaking quickly for me also very.
Syntax-based SRL

```

CP
  IP
    VP
      VP
        VP
          VP
            VP
              VP
                VP
                  VP
                    VP
                      VP
                        VP
                          VP
                            VP
                                VP
                                    VP
                                        VP
                                            VP
                                                VP
                                                    VP
                                                        VP
                                                            VP
                                                                VP
                                                                    VP
                                                                        VP
                                                                            VP
                                                                                VP
                                                                                    VP
                                                                                        VP
                                                                                            VP
                                                                                                VP
                                                                                                    VP
                                                                                                        VP
                                                                                                            VP
                                                                                                                VP
                                                                                                                    VP
                                                                                                                        VP
                                                                                                                            VP
                                                                                                                                VP
                                                                                                                                    VP
                                                                                                                                        VP
                                                                                                                                            VP
                                                                                                                                                VP
                                                                                                                                                    VP
                                                                                                                                                            VP
                                                                                                                                                                VP
                                                                                                                                                                    VP
                                                                                                                                                                        VP
                                                                                                                                                                            VP
                                                                                                                                                                                VP
                                                                                                                                                                                    VP
                                                                                                                                                                                        VP
                                                                                                                                                                                            VP
                                                                                                                                                                                                VP
                                                                                                                                                                                                    VP
                                                                                                                                                                                                        VP
                                                                                                                                                                                                            VP
                                                                                                                                                    CP
                                                                                                                                                        IP
                                                                                                                                                            SP
                                                                                                                                                                PU
                                                                                                                                                    ah
                                                                                                                                                    MOD
                                                                                                                                                    .

A0
AM
NULL
```

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
very

快
quickly

用
using

汉
Chinese

对我说
for me

也
also

很
Syntax-based SRL

CP

IP

VP

null

VP

AM

ADVP

very

VP

REL

A0

AM

null

AM

Rel
Syntax-based SRL

Ah, very fast speaking also using Chinese. For me.
Performance on L1

- The syntax-based systems are more robust when handling learner texts.

---

**Comparison of Performance Across Languages:**

- **ENG:**
  - PCFGLA-parser-based-L1: [65]  
  - Neural-parser-based-L1: [70]  
  - Neural syntax-agnostic-L1: [75]

- **JPN:**
  - PCFGLA-parser-based-L1: [70]  
  - Neural-parser-based-L1: [75]  
  - Neural syntax-agnostic-L1: [80]

- **RUS:**
  - PCFGLA-parser-based-L1: [75]  
  - Neural-parser-based-L1: [80]  
  - Neural syntax-agnostic-L1: [85]

- **ARA:**
  - PCFGLA-parser-based-L1: [70]  
  - Neural-parser-based-L1: [75]  
  - Neural syntax-agnostic-L1: [80]

- **ALL:**
  - PCFGLA-parser-based-L1: [70]  
  - Neural-parser-based-L1: [75]  
  - Neural syntax-agnostic-L1: [80]
Evaluation and findings

Performance on L1 & L2

The syntax-based systems are more robust when handling learner texts.
The syntax-based systems are more robust when handling learner texts.
Why syntactic analysis is important?

It is very hard for me to speak Chinese quickly.
Why syntactic analysis is important?

The whole structure is bad, but some parts may be good.
Why syntactic analysis is important?

Though the whole structure is bad,
Why syntactic analysis is important?

Though the whole structure is *bad*, some parts may be *good*. 
Research questions

- How can we capture meanings of L2s?
  How can we annotate L2 texts?
  Are there many differences from annotating L1 texts?

- How *badly* does an L1 data-trained semantic parser perform?
  Can state-of-the-art grammatical error correction systems help?

- What role does syntactic parsing play in processing L2 texts?
  What role does cross-lingual transfer play?
Something like

- JAPANESE native
- CHINESE to learn
- ENGLISH native
- CHINESE to learn

L2-CHI, L1-JPN
L2-CHI, L1-ENG

learn
Two languages, three tasks

Chinese as a Second Language
- Semantic Role Labeling
- Negation Scope Resolution

English as a Second Language
- Compositional Semantics
  Semantic Graph Parsing

(3) a. We needs actions and not thoughts.

b. He failed to catch the first train.

c. This is an un clean desk.

d. 换言说，没有 宗教生活与日常生活差距。

e. 换言说， 宗教生活与日常生活之间 没有 距离。
L2-Japanese data is more useful

<table>
<thead>
<tr>
<th>Train Set</th>
<th>Test Set</th>
<th>L2-Chi, L1-Eng</th>
<th>Chi_{L2} \Rightarrow L1, L1-Eng</th>
<th>L2-Chi, L1-Jpn</th>
<th>Chi_{L2} \Rightarrow L1, L1-Jpn</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2-Chi, L1-Eng</td>
<td>73.4/67.8/61.3</td>
<td>73.7/69.4/62.9</td>
<td>71.6/64.9/56.3</td>
<td>71.1/64.5/56.9</td>
<td></td>
</tr>
<tr>
<td>Chi_{L2} \Rightarrow L1, L1-Eng</td>
<td>73.4/68.1/61.6</td>
<td>73.8/69.7/62.4</td>
<td>70.7/65.1/58.2</td>
<td>71.0/65.4/57.6</td>
<td></td>
</tr>
<tr>
<td>L2-Chi, L1-Jpn</td>
<td>73.2/65.6/57.3</td>
<td>74.7/68.0/60.7</td>
<td>75.2/68.5/62.8</td>
<td>74.2/68.4/61.5</td>
<td></td>
</tr>
<tr>
<td>Chi_{L2} \Rightarrow L1, L1-Jpn</td>
<td>72.9/65.1/58.0</td>
<td>74.0/68.3/60.8</td>
<td>74.6/68.4/59.8</td>
<td>74.1/68.5/60.7</td>
<td></td>
</tr>
</tbody>
</table>

Table: Recall scores of BERT-BiLSTM/ELMo-BiLSTM/Random-BiLSTM.
Some examples

(4) a. 所以大多数的人觉得 **受不了** 日本的夏天。

b. 我还 **没有** 上小学的时候。

c. 我们 **不** 应该恐怕说错 **还有** 不好意思的事。

▶ pro-drop
▶ relative clause
▶ coordination
Conclusion and future work

- How can we capture meanings of L2s?
  - How can we annotate L2 texts?
  - Are there many differences from annotating L1 texts?
- How *badly* does an L1 data-trained semantic parser perform?
  - Can state-of-the-art grammatical error correction systems help?
- What role does syntactic parsing play in processing L2 texts?
  - What role does cross-lingual transfer play?
Conclusion and future work

- How can we capture meanings of L2s? How can we annotate L2 texts? Are there many differences from annotating L1 texts?
- How badly does an L1 data-trained semantic parser perform? Can state-of-the-art grammatical error correction systems help?
- What role does syntactic parsing play in processing L2 texts? What role does cross-lingual transfer play?

- How can we effectively enlarge annotated corpora?
- What is the best practice to annotate syntactic structures of second languages?
- What types of computational analyses can we develop for second language acquisition?
- Can we access learners’ language capability by annotating their language outputs?
THANK YOU

Joint work with Yuanyuan Zhao, Mengyu Zhang, Weiqi Wang, Zi Lin, Yuguang Duan


References III


Zhao, W., Wang, L., Shen, K., Jia, R., & Liu, J. (2019). Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data.