An Empirical Examination of Challenges in Chinese Parsing

Jonathan Kummerfeld
Dan Klein
UC Berkeley

Daniel Tse
James Curran
U Sydney
Parsing Errors

named chief executive officer of Google in 2001
chief executive officer named Google of in 2001
named chief executive officer of Google in 2001
Parsing Errors

VP
  S
    NP
      VBN
      NP
      IN
      PP
        NP
        NNP
        PP
          NP
          Google
          in 2001

named chief executive officer of Google in 2001
chief executive officer named Google in 2001
Parsing Errors

The diagram represents the sentence: "A chief executive officer named Google in 2001 was VP of Google in 2001."
chief executive officer named Google of in 2001
Manual Error Analysis
Manual Error Analysis

Too slow!
Automatic Error Analysis

Kummerfeld et al. (2012)
chief executive officer named Google in 2001
Automatic Error Analysis

Kummerfeld et al. (2012)
Automatic Error Analysis

Tree Transformations

Kummerfeld et al. (2012)
Automatic Error Analysis

List of Transformation Steps

Kummerfeld et al. (2012)
Automatic Error Analysis

List of Transformation Steps

Categorise

Kummerfeld et al. (2012)
Automatic Error Analysis

List of Transformation Steps

Categorise

List of Errors

Kummerfeld et al. (2012)
English and Chinese Errors

- PP Attachment
- Clause Attachment
- Coordination
- Right Span, Wrong Label
- Single Word Phrase
- NP Attachment
- Modifier Attachment
- Unary
- NP Internal Structure
- VP Attachment
- Other

Av. Incorrect Brackets per Sentence
English and Chinese Errors

Av. Incorrect Brackets per Sentence

- PP Attachment
- Clause Attachment
- Coordination
- Right Span, Wrong Label
- Single Word Phrase
- NP Attachment
- Modifier Attachment
- Unary
- NP Internal Structure
- VP Attachment
- Other

English
Chinese
English and Chinese Errors

Av. Incorrect Brackets per Sentence

- PP Attachment
- Clause Attachment
- Coordination
- Right Span, Wrong Label
- Single Word Phrase
- NP Attachment
- Modifier Attachment
- Unary
- NP Internal Structure
- VP Attachment
- Other

English
Chinese
Chinese Parsing Errors

请 request 专家 论证
Chinese Parsing Errors

request specialist proof

System
Chinese Parsing Errors

System

Gold

(request specialist proof)

(request a specialist to prove)
Chinese Parsing Errors

System

Gold

request specialist proof

request specialist proof

Chinese Parsing Errors

System

Gold
Chinese Parsing Errors

System

Gold

(request specialist proof)

(request specialist proof)
Chinese Parsing Errors

System

Gold

(request specialist proof)

(request a specialist to prove)
Chinese Parsing Errors

支撑  生产  增长
Chinese Parsing Errors

```
support increasing production

System
```
Chinese Parsing Errors

support increasing production

support production to increase

System

Gold
Automatic Error Analysis

List of Transformation Steps

List of Errors

Tree Transformations

Categorise
Automatic Error Analysis

List of Transformation Steps

Tree Transformations

List of Errors

Categorise
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Av. Incorrect Brackets per Sentence
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Av. Incorrect Brackets per Sentence

0 0.5 1 1.5 2 2.5 3 3.5 4
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Av. Incorrect Brackets per Sentence

- Berkeley
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Berkeley
Stanford

Av. Incorrect Brackets per Sentence

Av. Incorrect Brackets per Sentence
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Av. Incorrect Brackets per Sentence

Berkeley
ZPAR (Zhang & Clark 09)
Stanford
Chinese Parse Errors

Av. Incorrect Brackets per Sentence

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Berkeley
ZPAR (Zhang & Clark 09)
Stanford
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Berkeley

Av. Incorrect Brackets per Sentence

0  0.5  1  1.5  2  2.5  3  3.5  4
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Berkeley
Berkeley Product Parser

Av. Incorrect Brackets per Sentence
Chinese Parse Errors

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Av. Incorrect Brackets per Sentence

Berkeley
Berkeley Product Parser
Chinese Parse Errors

NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Noun Boundary
- Other

Av. Incorrect Brackets per Sentence

Berkeley
Berkeley Product Parser
Tagging Error Impact
Tagging Error Impact

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Other

Av. Incorrect Brackets per Sentence

No Gold Input
计划在星期六晚的阿拉法特、佩雷斯会晤...
Tagging Error Impact
Tagging Error Impact

- NP Internal
- Coordination
- Verb Arguments
- Unary
- Modifier Attachment
- Single Word Phrase
- Right Span, Wrong Label
- Sense Error Cascade
- Other

No Gold Input
Gold POS Input

Av. Incorrect Brackets per Sentence

0 0.5 1 1.5 2 2.5 3

No Gold Input
Gold POS Input
Tagging Error Impact

NP Internal
Coordination
Verb Arguments
Unary
Modifier Attachment
Single Word Phrase
Right Span, Wrong Label
Sense Error Cascade
Other

Av. Incorrect Brackets per Sentence

No Gold Input
Gold POS Input
Tagging Error Impact

NP Internal
Coordination
Verb Arguments
Unary
Modifier Attachment
Single Word Phrase
Right Span, Wrong Label
Sense Error Cascade
Other

No Gold Input
Gold POS Input

Av. Incorrect Brackets per Sentence

- No Gold Input
- Gold POS Input

0 0.5 1 1.5 2 2.5 3
Tagging Error Impact

计划 在 星期六 晚 的 阿拉法特 、 佩雷斯 会晤
Tagging Error Impact
Tagging Error Impact

NP Internal
Coordination
Verb Arguments
Unary
Modifier Attachment
Single Word Phrase
Right Span, Wrong Label
Sense Error Cascade
Other

Av. Incorrect Brackets per Sentence

No Gold Input
Gold POS Input
Noun / Verb Confusion
Conclusion
Conclusion

- Quantitative validation that Chinese and English parsers deal with different challenges
Conclusion

- Quantitative validation that Chinese and English parsers deal with different challenges
- Perfecting Chinese POS Tagging will only solve certain parsing challenges
Conclusion

- Quantitative validation that Chinese and English parsers deal with different challenges.
- Perfecting Chinese POS Tagging will only solve certain parsing challenges.
- Download our automatic analysis tool: nlp.cs.berkeley.edu/software
Conclusion

- Quantitative validation that Chinese and English parsers deal with different challenges
- Perfecting Chinese POS Tagging will only solve certain parsing challenges
- Download our automatic analysis tool: nlp.cs.berkeley.edu/software

Thank you!