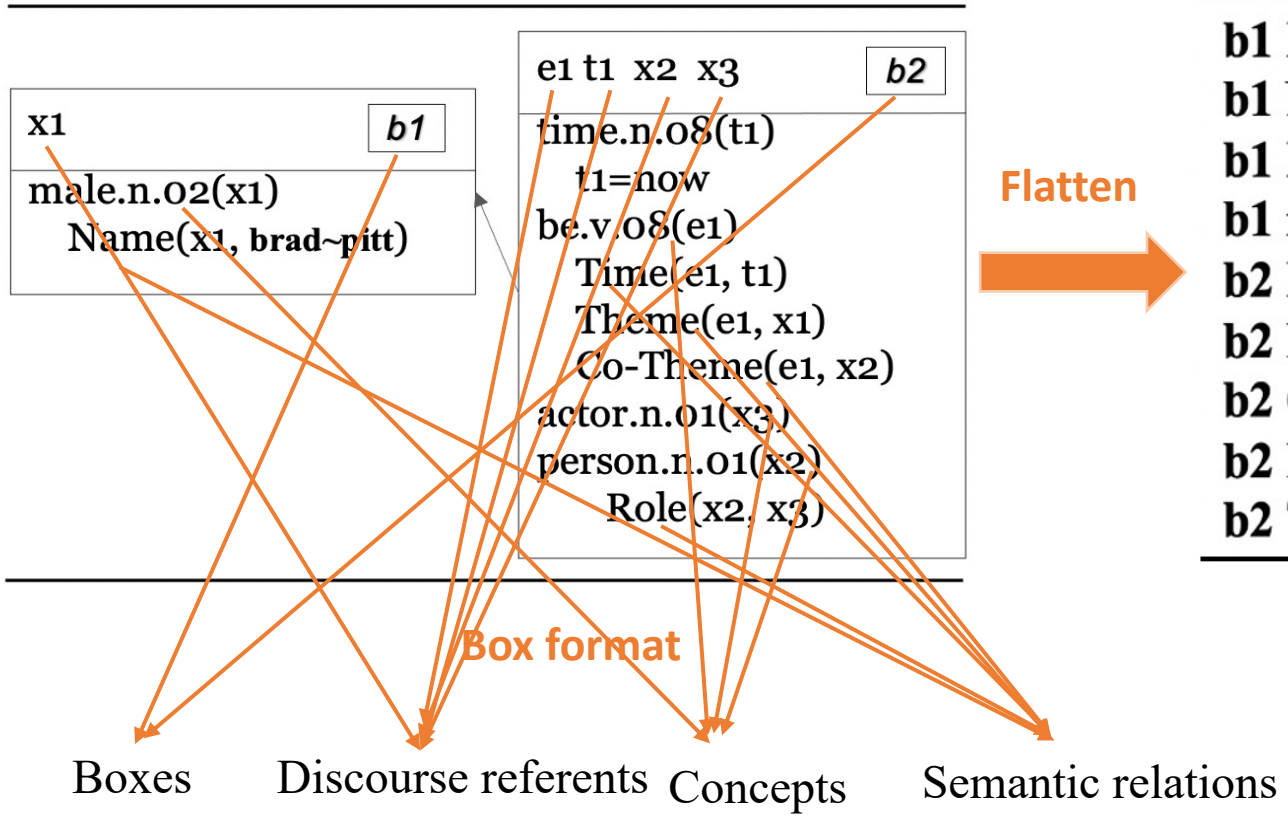
Input Representations for Parsing Discourse Representation Structures: Comparing English with Chinese

Chunliu Wang Rik van Noord Arianna Bisazza Johan Bos

Discourse Representation Structures

Example: Brad Pitt is an actor.

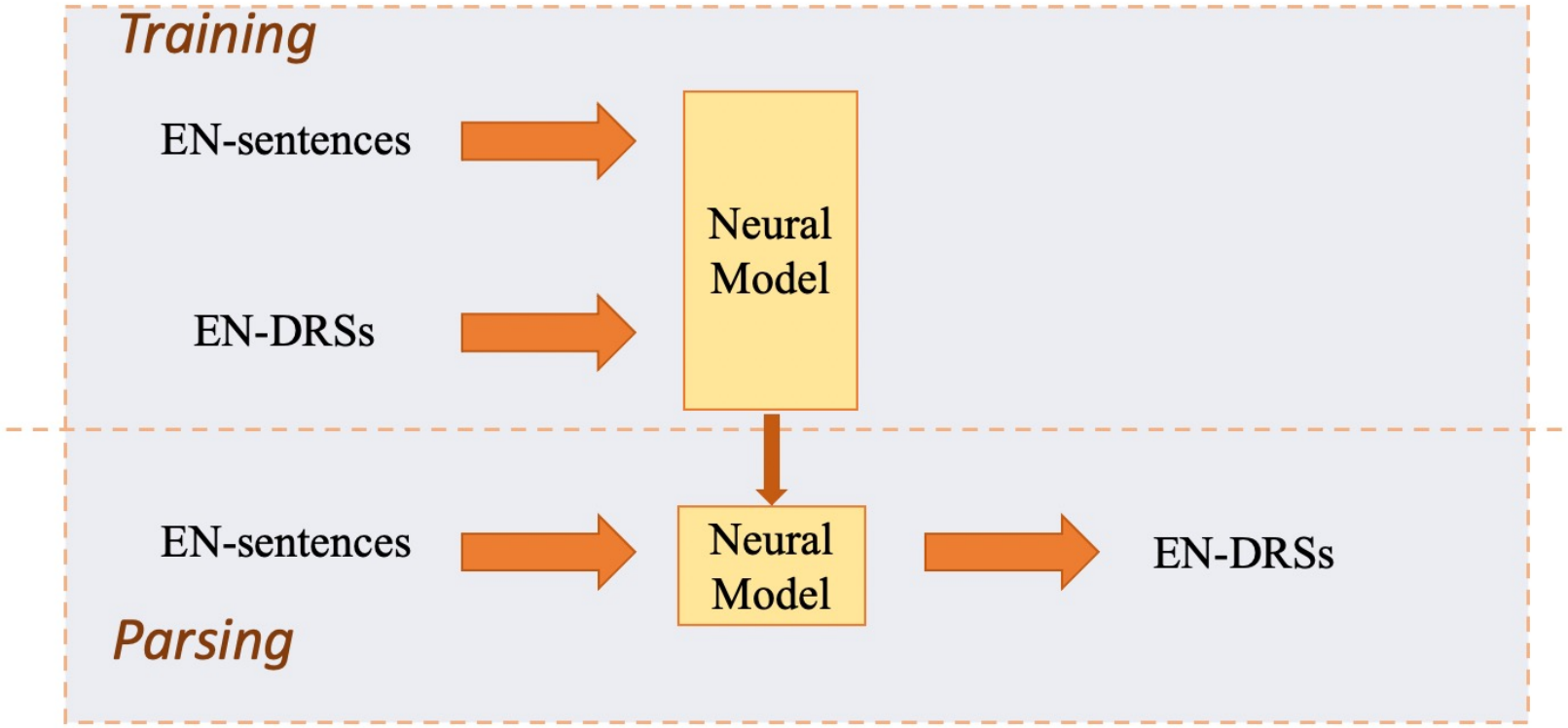


Flatten

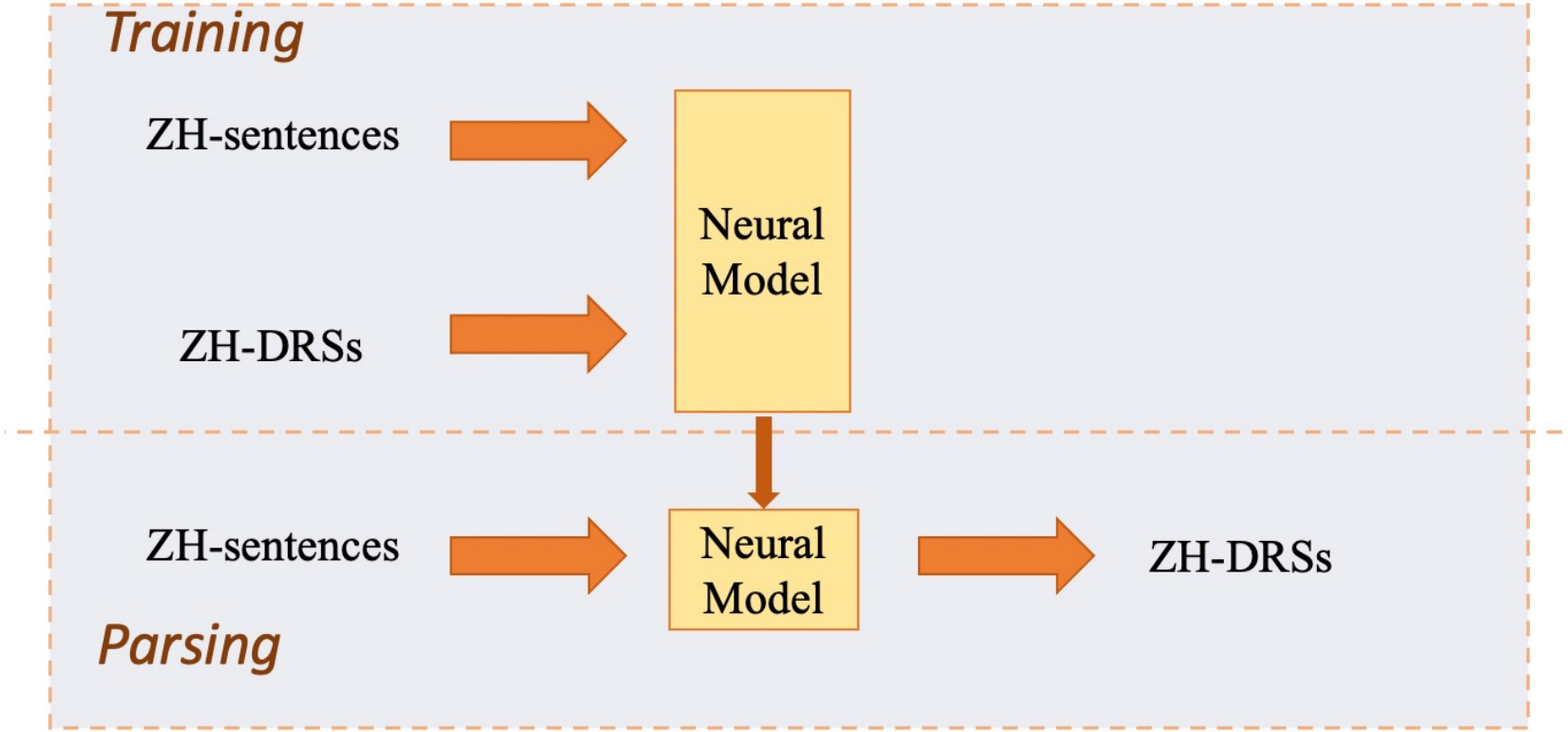
b1 REF **x1**
b1 Name **x1** **"brad~pitt"**
b1 PRESUPPOSITION **b2**
b1 male "n.02" **x1**
b2 REF **e1**
b2 REF **t1**
b2 Co-Theme **e1** **x2**
b2 EQU **t1** "now"
b2 Theme **e1** **x1**
b2 Time **e1** **t1**
b2 be "v.08" **e1**
b2 time "n.08" **t1**
b2 REF **x2**
b2 REF **x3**
b2 Role **x2** **x3**
b2 actor "n.01" **x3**
b2 person "n.01" **x2**

Clause format

English DRS Parsing ➢ Chinese DRS Parsing



English DRS Parsing ➢ Chinese DRS Parsing



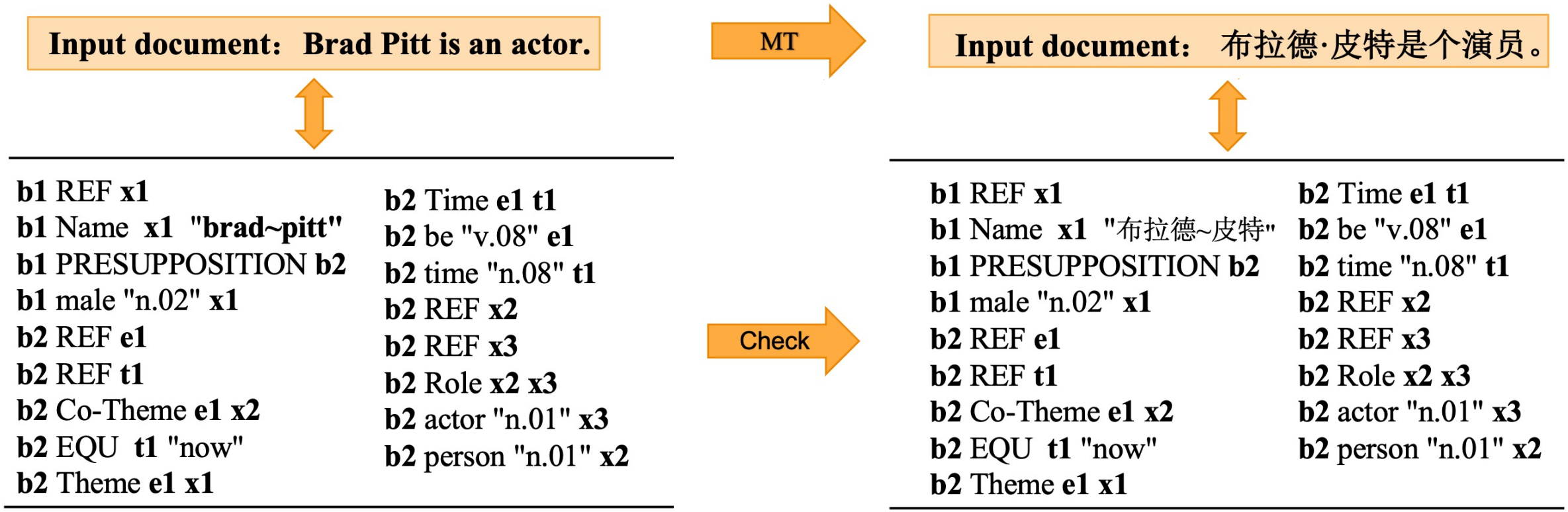
Questions

Q1: Can existing DRS parsing models achieve good results for Chinese?

Q2: Given the different writing systems used for English and Chinese, which input granularity is best for either language?

Q3: Is rule-based word segmentation beneficial for Chinese DRS parsing?

How to get Chinese DRSs data?



Methodology: Input representations

Type	English input representation	Chinese input representation
Char (raw)	^ b r a d | ^ p i t t | i s | a n | a c t o r .	布 拉 德 · 皮 特 是 个 演 员 。
Char (continuous)	^ b r a d ^ p i t t i s a n a c t o r .	
Char (tokenized)	^ b r a d | ^ p i t t | i s | a n | a c t o r | .	布 拉 德 | · | 皮 特 | 是 | 个 | 演 员 | 。
Word	brad pitt is an actor .	布拉德 · 皮特 是 个 演员 。
BPE (5k)	^ b@ ra@ d ^ p@ it@ t is an ac@ tor@ .	布 拉 德 · 皮 特 是个 演 员 。

Word boundary

Uppercased characters

No word boundary

Raw and continuous char-level are identical in Chinese

Methodology: Output representation

Input sentence

汤姆提着一桶水。

Output DRS

```
b1 REF x1
b1 Name x1 "汤姆"
b1 PRESUPPOSITION b2
b1 male "n.02" x1
b2 REF t1
b2 TPR t1 "now"
b2 Time e1 t1
b2 time "n.08" t1
b2 REF e1
b2 Agent e1 x1
b2 Theme e1 x2
b2 carry "v.01" e1
b2 REF x2
b2 bucket "n.01" x2
b2 Content x2 x3
b2 REF x3
b2 water "n.06" x3
```

Clause boundary

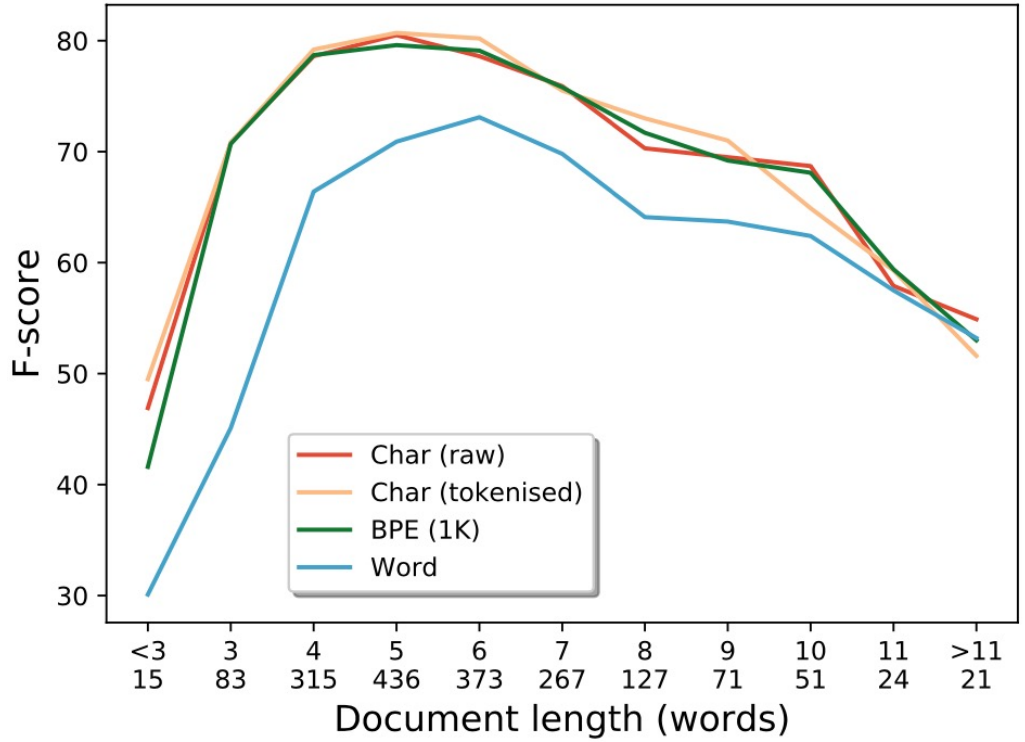
Indices,
instead of variables

DRS representation

```
$NEW ||| REF *** $0 ||| Name ||| @0 ||| " 汤 姆 " *** $0 ||| PRESUPPOSITION ||| $NEW *** $-1 ||| m a
l e ||| " n . 0 2 " ||| @0 *** $0 ||| REF *** $0 ||| TPR ||| @0 ||| "now" *** $0 ||| Time ||| @1 ||| @0 *** $0 |||
t i m e ||| " n . 0 8 " ||| @0 *** $0 ||| REF *** $0 ||| Agent ||| @0 ||| @-2 *** $0 ||| Theme ||| @0 ||| @1 ***
$0 ||| c a r r y ||| " v . 0 1 " ||| @0 *** $0 ||| REF *** $0 ||| b u c k e t ||| " n . 0 1 " ||| @0 *** $0 ||| Content
||| @0 ||| @1 *** $0 ||| REF *** $0 ||| w a t e r ||| " n . 0 6 " ||| @0
```

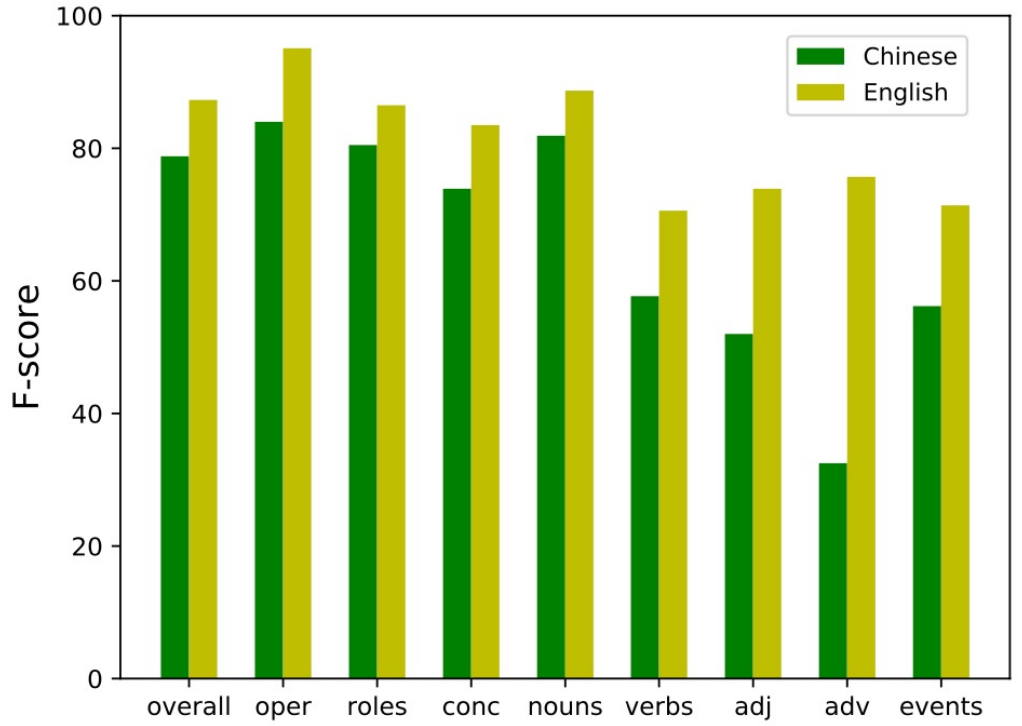
Evaluation: Bi-LSTM model

	English		Chinese	
Input type	**Dev**	**Test**	**Dev**	**Test**
Char (raw)	87.9	87.6	78.8	76.2
Char (continuous)	86.1	86.9		
Char (tokenised)	88.0	88.1	79.5	76.2
BPE (1k)	86.8	87.0	78.5	76.2
BPE (5k)	87.4	87.1	75.1	71.8
BPE (10k)	82.5	82.3	68.5	65.2
Word	84.5	83.2	74.7	71.6



Evaluation: Bi-LSTM model

	English		Chinese	
Input type	**Dev**	**Test**	**Dev**	**Test**
Char (raw)	87.9	87.6	} 78.8	76.2
Char (continuous)	86.1	86.9		
Char (tokenised)	88.0	88.1	79.5	76.2
BPE (1k)	86.8	87.0	78.5	76.2
BPE (5k)	87.4	87.1	75.1	71.8
BPE (10k)	82.5	82.3	68.5	65.2
Word	84.5	83.2	74.7	71.6



Conclusions

Q1: Can existing DRS parsing models achieve good results for Chinese?
A1: Projecting meaning representations from English gives remarkable performance

Q2: Which input granularity is best for Chinese?
A2: Characters are the preferred input representation for Chinese

Q3: Is rule-based word segmentation beneficial for Chinese DRS parsing?
A3: Tokenisation of the input offers a small advantage only for English

university of groningen

Parallel MEANING BANK

Input Representations for Parsing Discourse Representation Structures: Comparing English with Chinese

Chunliu Wang Rik van Noord Arianna Bisazza Johan Bos

Q1: Can existing DRS parsing models achieve good results for Chinese?
A1: Projecting meaning representations from English gives remarkable performance

Q2: Which input granularity is best for Chinese?
A2: Characters are the preferred input representation for Chinese

Q3: Is rule-based word segmentation beneficial for Chinese DRS parsing?
A3: Tokenisation of the input offers a small advantage only for English