Chinese Treebanking and parser evaluation issues

Nianwen Xue

Brandeis University

Feb. 24, 2009

BBN Technologies

CTB: Milestones

Version	Year	Quantity (words)	Source	Parallel	OntoNotes data
CTB1.0	2001	100K	Xinhua	yes (100K)	yes (100K)
CTB3.0	2003	250K	+HK News	no	no (250K)
CTB4.0	2004	400K	+Sinorama	yes (100K)	yes (150K)
CTB5.0	2005	500K	+Sinorama	yes (100K)	no (100K)
CTB6.0	2007	780K	+ BN	no	yes (300K)
CTB.xx	2008	950K	+BC	yes(100K)	yes (150K)
CTB.xx*	2009*	1.1M	+NG/WL	yes(100K)	yes (150K)

OntoNotes funded treebanking

The effect of Chinese word segmentation on parsing

- Three factors
  - Word segmentation quality
  - Parsing quality
  - Evaluation metrics
- Three sources of data
  - Xinhua
  - Sinorama
  - Broadcast news

Word segmentation

genre	precision	recall	F-score
Xinhua	96.80	96.25	96.52
Sinorama	93.18	92.46	92.81
Broadcast news	95.36	95.51	95.44

SOA in Sighan 2008 bakeoff: F1 = 95.89%

Sparseval metric

```
(IP (NP-SBJ (NP (NN 全球))
            (QP (OD 第五)
                (CLP (M 个)))
            (NP-PN (NR 迪斯尼)
                   (NN 乐园)))
    (VP (ADVP (AD 即将))
        (PP-LOC (P 在)
                (NP (PN 这里)))
        (PP-DIR (P 向)
                (NP (NN 公众)))
        (VP (VV 开放)))
    (PU 。))
```

```
(IP (NP-SBJ (NP (NN 全)(NN 球))
            (QP (OD 第五)
                (CLP (M 个)))
            (NP-PN (NR 迪斯尼)
                   (NN 乐园)))
    (VP (ADVP (AD 即将))
        (PP-LOC (P 在)
                (NP (PN 这里)))
        (PP-DIR (P 向)
                (NP (NN 公众)))
        (VP (VV 开放)))
    (PU 。))
```

The fifth Disney World around the globe is going to be open here to the public soon

SParseval: Precision = Recall = F1 = 11/13

CParseval metric: a proposal

A constituent is parsed correctly if it spans over the same string of characters as some constituent in the gold standard parse.

SParseval vs CParseval

Genre	Segmentation	SParseval	CParseval
Xinhua	gold	80.37	79.52
	auto	75.93	77.13
Sinorama	gold	71.37	70.20
	auto	63.41	65.96
BN	gold	79.39	78.51
	auto	74.56	76.44

Conclusion

- **By SParseval metric, parsing accuracy drops 4% to 8% when automatic word segmentation is used as input**
- **By "CParseval" metric, parsing accuracy drops by 2% to 4%**