Analysis and Repair of Name Tagger Errors
Heng Ji and Ralph Grishman (New York University)

Outline
- What is behind “F-Measure”? – Decompose name errors into identification error (missing, spurious, boundary) and classification error
- Repair errors by global re-ranking of N-Best Hypotheses
- Check remaining errors by comparing with human

Re-Ranking Features from Joint Inference

Test on 100 ACE texts (2813 names: 1126 Persons, 712 GPEs, 785 Organizations and 190 Locations)

Performance on Chinese Name Tagging

Conclusion and Remaining Error Types
- Global re-ranking features help to reduce classification and boundary errors
- Main human annotation errors were the names overlooked by one annotator
- Gain from events was limited by event pattern frequency → extend event
- 70% of remaining spurious errors were ‘other’ name types → explicit model
- Unsupervised learning to boost cross-doc coref re-ranking