NAACL Workshop on MTE

3rd in a Series of MTE adventures
3 June, 2001
Structure of Workshop

• One-day workshop
• Constrained by this – less flexibility in choice
• First part of day – review of MTE and ISLE
• Second part of day – hands-on exercise
  • Much more constrained than this workshop
First Exercise

• Information Extraction and MT
  ● Why named entity scoring alone is not sufficient
  ● What else might go into the scoring process

• Users will be given
  ● Data set (pre-translated)
  ● Templates to fill out
  ● Different source texts from different engines
  ● Score the ability to populate templates
  ● Mark where information came from
Second Exercise

• Question-Answering as a follow-on to MT process:
  • Given a set of questions and related documents
  • Answer the questions
    ● Scoring ability of system to support task
    ● Marking text as to where information came from
    ● Looking at the types of information (where versus when) preserved / lost

• Inspired by the TREC Q/A tasks
Analysis of Exercises

• Given number of participants can:
  ● Look for inner-annotator agreement
  ● Look for stronger correlation of scores
  ● Look at specific linguistic constructs

• Expanding / comments on ISLE
  ● According to experiment results
Some Specifics

• Language pairs
  ● Chinese > English
  ● Arabic > English
  ● Spanish > English
• Follow-up reporting for later