Fightin’ Words

Points for discussion

Points for review

Looking for a report structure?
System Assessment

• What metrics are suitable for assessing what system characteristics?
• What system characteristics reflect what user needs?
• Is there a radical difference between evaluation focusing on research or development needs and evaluation focusing on end-user needs?
• When should real-world data be used?
  ♦ What is the impact of using it?
Metric Choice

• What constitutes a valid metric
  ♦ How can you demonstrate that a metric is valid?
• What metrics can be automated
• What are the advantages/disadvantages of specific metrics?
• For the metric(s) selected, what are the difficulties in applying them?
Metric Choice - II

• For a given metric, what variations in scores are typically produced?
  ◦ What are the statistical error variances?

• For a given metric, what are the score ranges for ‘good’ and ‘bad’ systems

• Are there metrics which correlate with one another?
  ◦ Are there metrics which indicate an overall quality score?

• Are there metrics which work better for specific language pairs?
Where to from here?

• MTE Primer?
  - Book with papers, and commentary fitting them into the right context (including the taxonomy?)
  - Exercises the reader can do
  - Showing how to fit new ideas into taxonomy
  - Contribution to the language community in general

• MT Summit Workshop
• Other workshops?
BIG HUG
Web Sites

• Classification (taxonomy)
  ◦ Open to everyone

• Working area
  ◦ Possibly password protected
  ◦ Rite of initiation?
  ◦ Data for people - at least until next workshop

• URL’s