

Basic Rules of Engagement

- ◆ Desired output of workshop teams
- ◆ Scores
- ◆ Justifications
- ◆ Computational procedures
- ◆ Pitfalls

Desired Output of Teams

- ◆ On last two days
 - You will be presenting your evaluations
 - ◆ Methodology
 - ◆ Basis for methodology
 - ◆ Results
 - ◆ Analysis of results
 - Doing a short write-up of results
- ◆ MT-Summit Workshop Strong-Arming

Winning Scores

- ◆ Really useful evaluations
 - Are objective
 - Are replicable
 - ◆ Not resource intensive
 - ◆ Scalability is kept
 - Are meaningful
 - ◆ To MT providers / developers
 - ◆ To MT users
 - ◆ To linguists / NLP researchers
 - Fit into the framework or inform changes
 - Can be applied across language pairs / families

Justifications

- ◆ Relationship to ISLE framework
 - Or reasonable argument for changing it
- ◆ Relationship to linguistic / NLP research
- ◆ Relationship to “real world”
 - Task-based evaluation

Computational Procedures

- ◆ Should speak to automating (as much as possible) evaluation process
- ◆ Should be mathematically rigorous
- ◆ Must reflect available data tools where possible
 - Caveats – bringing your own data
 - Request for sensitivity of current data sets

Pitfalls

- ◆ We only have a week
- ◆ This is a workshop not a holy war
- ◆ MT will not take the jobs of translators
 - Evaluation with respect to goal of output use, not the holy grail
 - Evaluation with respect to continuum of use (which pieces are best machine / human)