Some tentative conclusions...

of five hard-working days

0. Some general remarks

- The use of the taxonomy is focused on some specific problems, some parts being neglected
  - groups do not always provide specific definitions of users and tasks
  - but G5 and G6 shows how useful this can be to justify the evaluation

0. Some general remarks (2)

- Interesting cross-domain links
  - MT and TM (G6)
  - MT and IE (G4)
  - - MT and terminology (G1)

- But... Nothing much about comparisons across language pairs or across text types

0. Some general remarks (3)

- Emergence of an interesting topic:
  - the translation of noun phrases (G1, G3, G4)

  - strong links with terminology

  - makes MT evaluation a more tractable matter?

1. The focus on the hard problem

- Central interest: evaluation of the « quality » of the output
- Intuitive, traditional, open question
- Difficult to define (but see taxonomy) and difficult to measure

- Various (complemetary) approaches

1.1 Method one

- Find a « formal » measure (G1, G3, G7)
  - analytic
  - local
  - intrinsic
  - quantitative
  - automatable?
- E.g., for noun phrases (structure, compunds, syntactic proximity)

1.2 Method two

- Use human judgments (G2, G5, G3?)
  - scoring of qualities or features
  - « cognitive » definition of features
  - more global than « method one »
  - sentences, texts

1.3 Method three

- Use external tasks (G4 but negative, G6)
  - evaluation of the MT contribution to a task
  - MT is a module
  - compare MT vs. HT texts

- Would require some more theoretical work
  - what exactly is evaluated
  - what comparisons can be made

2. Relation to taxonomy

- Method one:
  - 2.2.1.2 Accuracy
  - 2.2.1.2.2 Individual sentence level
  - 2..2.1.2.3 Types of errors
- Method two:
  - 2.2.1.2 Accuracy
  - 2.2.1.2.1 Text as a whole

2. Relation to taxonomy (2)

- Method three:
  - 1.2 The object of evaluation
  - 1.3 Characteristics of the translation tasks

Discussion

- About the previous points

- About the other questions in MT Evaluation

- Follow-ups:
  - NAACL workshop (June 3)
  - MT Summit workshop (September)