MT Evaluation
The DARPA measures
and
MT Proficiency Scale

The DARPA Series (1992-94)

- The DARPA MT Program
  - Radical approaches to MT
  - Heterogeneity w.r.t language, theory, maturity level
- Evaluation
  - Must accommodate the heterogeneity, yet
  - have some basis of measuring progress (and best generic approaches)
  - seems to necessitate black-box
- Evolution
  - toward "core" translation engines
  - toward higher validity

The DARPA MTE Method

- 3 Evaluation Approaches
  - avoids > 1 occurrence of a text
  - avoids > 1 occurrence of a system
  - avoids repetition of sequences
- Adequacy
  - Determine whether content is conveyed
  - Subjects respond to fragments on 1-5 scale
- Fluency
  - Determine how "English-like"
  - Subjects respond to sentences on 1-5 scale
- Informativeness
  - ability to gather essential information
  - Subjects answer multiple choice questions

The DARPA Method

[Funeral Service for Michael Jordan's Father]	**Funeral service for the father of Michael Jordan**
[Family and close friends of American basketball star Michael Jordan gathered together on Sunday]	The family and the near the star of the American basket-ball Michael Jordan develop are gathered Sunday for a funeral service to the memory of his/her/its father James.
[for a memorial service for Jordan's father, James.]	
[There was considerable security,]	The security was important and the press had been put secluded of the church Methodist épiscopalienne African (African Methodist Episcopal Church) placed nearly Teachey (Caroline of the north), where took place the service.
[and the press had been kept away from the African Methodist Episcopal Church near Teachey, North Carolina, where the service was held.]	
[Reporters had received a program of the service,]	The journalists had receipt a program of the ceremony, that comprehended a message of the widow of James Jordan, Deloris, and of the his/her/their five children Michael, James Ronald, Deloris, Larry and Roslyn.
[which included a message from James Jordan's widow, Deloris,]	
[and from his five children, Michael, James Ronald, Deloris, Larry, and Roslyn.]	

Fluency Sample

Test Number:

FLU 2N

Saigon and Montesque Comment on the Meeting with Buwaiz and French encouragement in conducting Elections.

Your Rating: [▼]

The president of the French National Assembly, Philip Saigon, and the vice president of the parliamentary Council for Foreign Affairs of State-owned corporations, Do Montesque, expect that there will be an absence of tangible results for several weeks from the policy of Benyamin Netanyahu towards the peace plan and they called for not abandoning Lebanon and not leaving the US to resolve all of the issues.

Your Rating: [▼]

The National News Agency reports in its report from Paris that Saigon was questioned by the Foreign Minister Faris Buwaiz during his most recent visit to Paris on some of the details linking current preparations in Lebanon for parliamentary elections and their promoting their conduct and the discussion of the results of the Israeli elections.
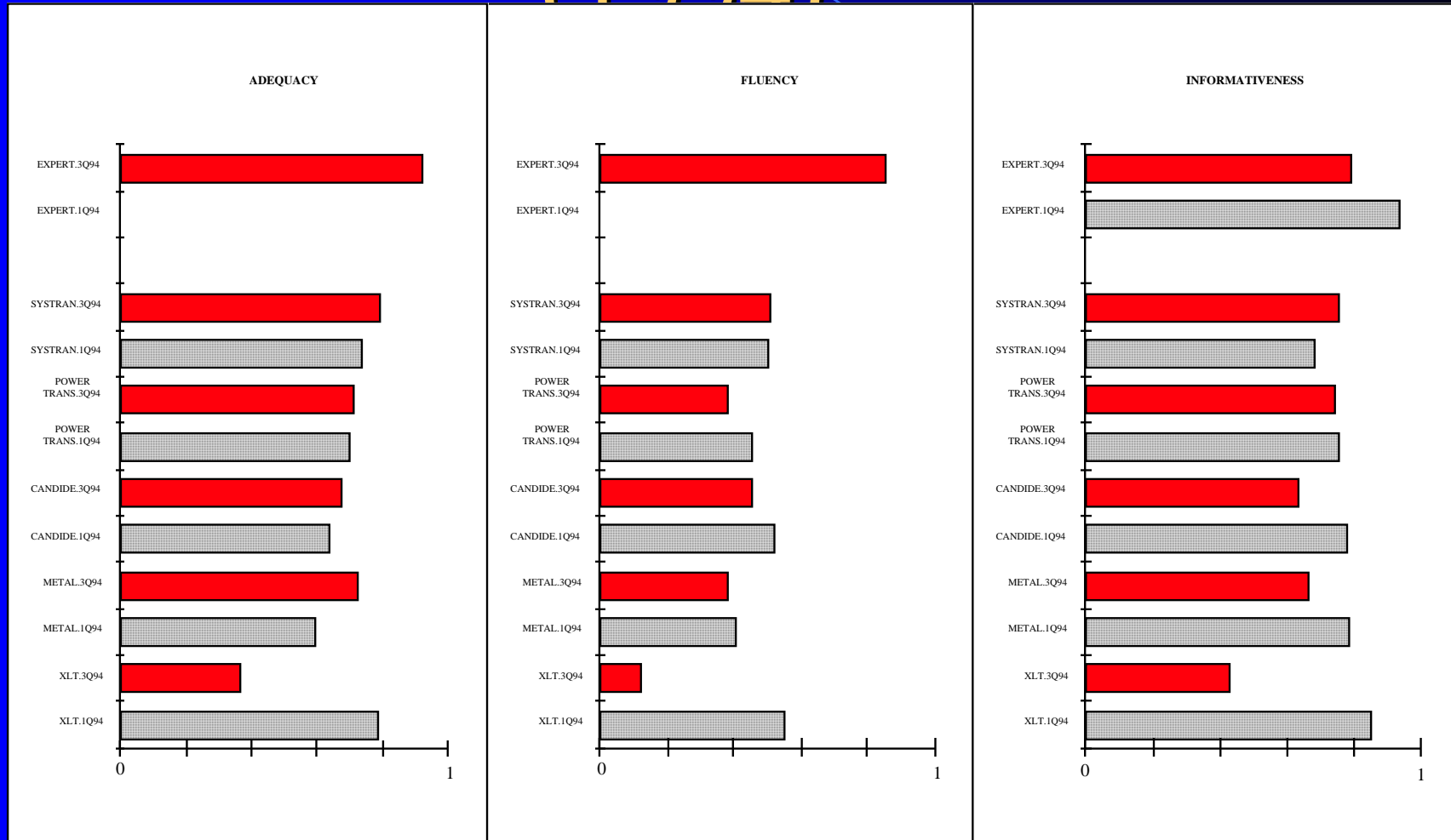
Your Rating: [▼]

The news agency reported Saigon as saying, "Lebanon decided to resume its operation after the last Israeli war despite the spiritual and material losses from the war while depending on the Lebanese people to promote unity and to join together the divisions and this in my opinion is necessary for the goal.

Your Rating: [▼]

5 = Excellent   4 = Good   3 = Fair   2 = Poor   1 = Very Poor

French to English Results (1994)

Correlation of measures (DARPA series)

	Adequacy	Fluency	Inform.
Adequacy	1 . 0	.8860	.9 6 5 4
Fluency	.8860	1 . 0	.7 6 2 5
Informativeness	.9 6 5 4	.7 6 2 5	1 . 0

High correlation between Adequacy and Informativeness
and **between Adequacy and Fluency**

The MT Proficiency Scale

The development of the measure involves four principal steps:

Identifying text-handling tasks

what text-handling tasks do users perform with translated material as input?

Discovering task tolerance order

how good must a translation be to be useful for a particular task?

Analyzing translation problems

what linguistic and non-linguistic translation problems occur in the corpus?

Developing source language patterns

which patterns correspond to diagnostic target phenomena?

Task-Oriented Exercises

TEXT-HANDLING TASKS

Publication quality output
Gisting
Extraction
    Deep extraction
    Intermediate extraction
    Shallow extraction
Triage
Detection
Filtering

The analyst's ability to perform the exercise with each MT sample is scored and reported, using a metric appropriate to the task.

- *For each task in the text-handling task inventory, an exercise is developed that is close to a participating analyst's task.*

User exercise -- shallow extraction

Persons   **Organizations**   **Locations**   **Dates**   **Times**   **Money/Percent**
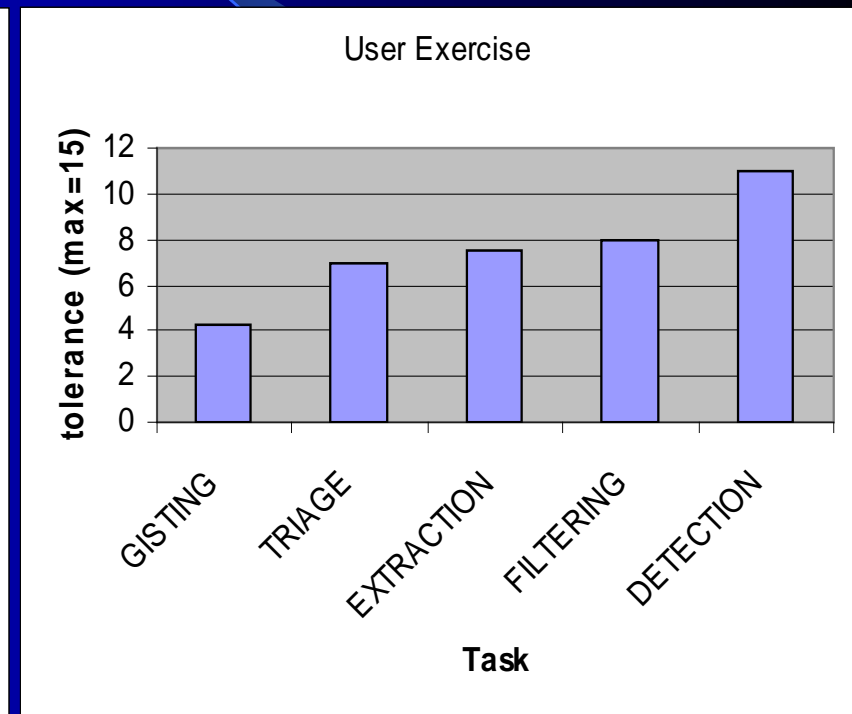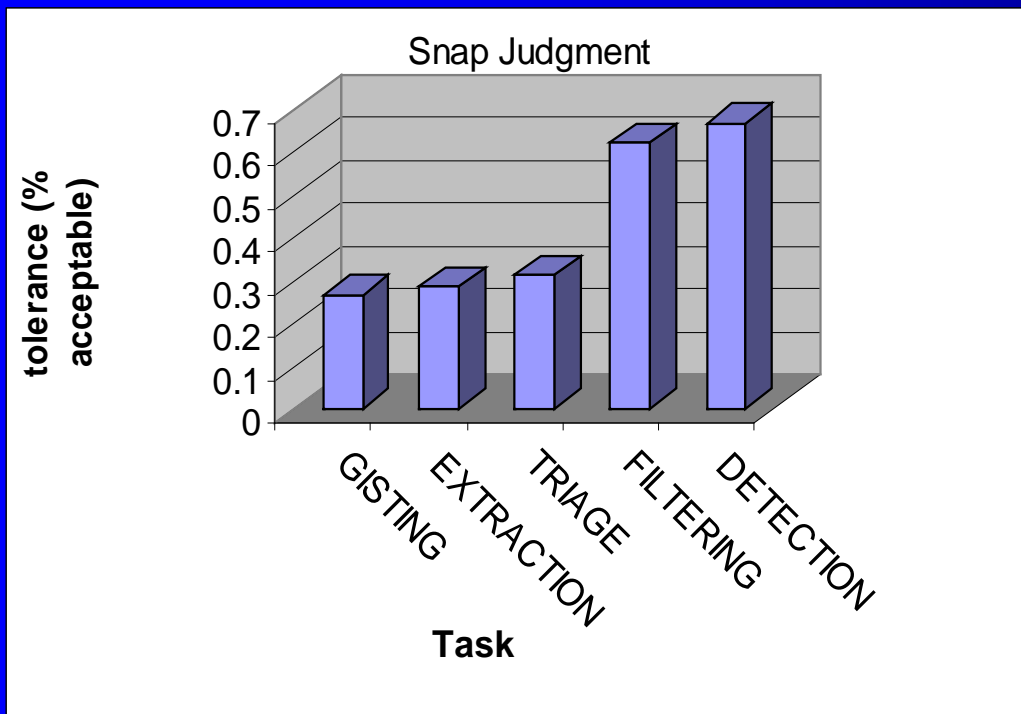
2050L

To the herding player taking part in the rice five rings to entrust to is the approval after the choumon is a general meeting

It is the Nancy kerigan player attack case of the United States of America woman skating, and the tenure herding player that the zenpu plural was arrested is a room with 12th and America Olympic committee (USOC), and it aitedorutte Do to entrust to, and for it drops suit that was causing, it concurred with to cause to appear Do player in rirehanmeru winter season five rings. The Harding player is the expectation that appears to five rings technical program of departure and 23th to Norway in the 15th.

The suit, the USOC, is the thing that called for compensation for damages of the 20 million dollars (date 2 billion 170 million yen) in the USOC with the decision suspending of situation that settled five rings appearing suspension of herding player. The opened of American Oregon state Portland city court of justice de, 11th, and oral argument is being being called for a consulting each other settlement with the bar foreign from judge, and both proxy negotiates with, and it reached in agreement.

The Patrick Galilee judge that is being in charge of the action is setting "it establishes a choumon of mangaichi and USOC, and the state of affairs that is appearing to be interpreted with unjust to herding player if happens, and it is diminishing five rings team Do player, and power that pays appearance of restoring order, courthouse, does saving".

Task Tolerance Levels

Can Intelligibility predict Fidelity?