MT EVALUATION WORKSHOP

WARMUP EXERCISES

These exercises are designed to familiarize you with some evaluation techniques and let you try your hand at applying them. This packet contains output from the three language combinations being addressed in the DARPA machine translation initiative: French-English, Japanese-English, and Spanish-English. The output includes English translations done by humans, by machines, and by human-aided machines. The evaluation techniques are taken from the recent DARPA evaluation methodology[1] and from the ALPAC report.[2]

The purpose of these exercises is not to judge the MT systems that participated in the DARPA evaluation, nor to promote a specific evaluation methodology. Rather, we are trying to give everyone a feel for what is involved in evaluating the quality of translation output. We also hope to gather some data about the techniques themselves.

In your packet you will find a summary of DARPA's machine translation evaluation methodology and a copy of the "Quality Panel Evaluation Criteria" and a "Quality Panel Evaluation Tally Sheet," to be used for the quality evaluation. In addition, you will find a copy of Appendix 10 of the ALPAC report which contains the "Scale of Intelligibility" and the "Scale of Informativeness," to be used to measure intelligibility and fidelity (I&F). You will want to familiarize yourself with this material before proceeding with the exercises.

Don't try to do all the exercises. Please select your strongest language combination. English-speakers who are not proficient in any of the source languages can do only the intelligibility and comprehension exercises. We would like for you to complete one quality evaluation, one comprehension test, one I&F rating sheet, and the participant profile. You are welcome to do more exercises if you want.

Please bring your rating sheets with you to registration. The results will be tabulated and presented on the second day of the Workshop. If you will be arriving late, please fax your forms to AMTA at 202/667-8808 before 30 October.

We hope you will find time to do your warmup exercises. Your reward will be a more productive workshop!

---

[1] Report of the DARPA Machine Translation Program System Evaluation, July-August, 1992.

[2] Automatic Language Processing Advisory Committee. 1966. Language and Machines: Computers in Translation and Linguistics: A Report by the Automatic Language Processing Advisory Committee (ALPAC). Washington, D.C.: National Academy of Sciences, Division of Behavioral Sciences. National Research Council Publication 1416.

PARTICIPANT PROFILE

Name: _____

Profession: _____

Mother tongue: _____

Language skill levels: (See descriptions on the back of this page)

Language	Level					
	0	1	2	3	4	5
English						
French						
Japanese						
Spanish						

Comments on the Quality Evaluation Exercise:

Comments on the Comprehension Test:

Comments on the Intelligibility and Fidelity Evaluation Exercise:

General comments: