



Standardized ILR-Based and Task-Based Speech-to-Speech MT Evaluation

Douglas Jones

LREC Machine Translation Evaluation Workshop

26 MAY 2014

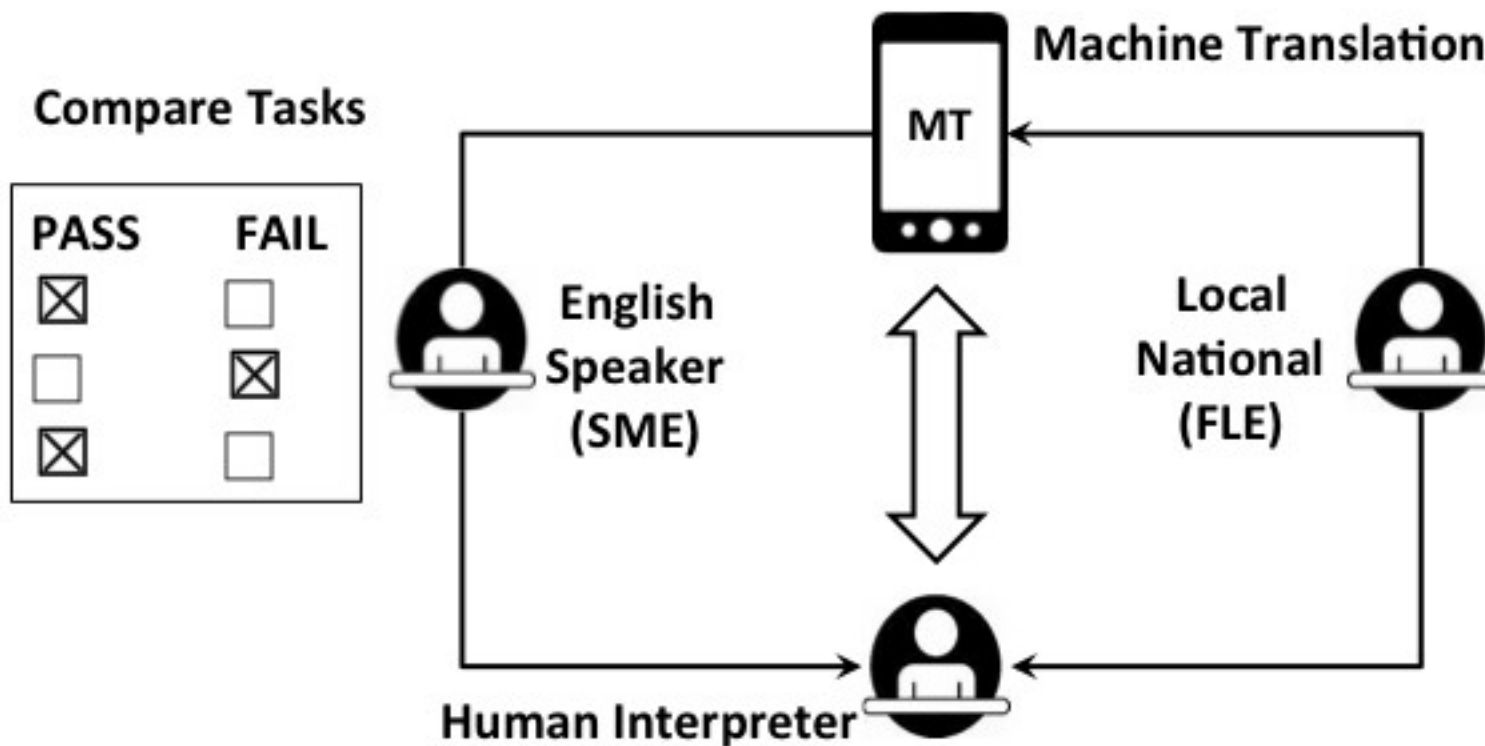


LINCOLN LABORATORY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY



Evaluation Concept





Sample Scenario

- **SHARED_CONTEXT:** A US soldier is talking to a doctor. This Area of Operations (AO) is safe and secure.
- **SME_KNOWS:** You are a Civil Affairs Soldier assigned to a civil-military operations center. You know how to conduct a Local Medical Health Assessment. Today you are concentrating on collecting human health information only.
- **SME_GOAL:** Follow the procedures for Conducting a Local Human Health Assessment. Collect health information including local facility names, the number of health workers and the nearest pharmacy. Determine the size of the population. Identify any endemic diseases and the leading causes of death.
- **FLE_KNOWS:** You are the only doctor here with responsibility for the two villages and the surrounding area. Your office is the only clinic and pharmacy. You have a meager stock of only the most basic medications. There are about 200 extended families. Twenty percent of the population is over 65 years old, or about 250 people. Children under the age of 12 number about 400. Cholera is the leading cause of death. There are scattered cases of hepatitis B. You have also seen a rise in cases of brucellosis.
- **FLE_GOAL:** Describe the medical situation of the population in your area.



Goals for Sample Scenario

Sample Soldier SME Goals

- 1 Collect health information.
- 2 Ask about the local facilities.
- 3 Ask about the local staff.
- 4 Ask about the nearest pharmacy.
- 5 Ask about the size of the population.
- 6 Ask about any endemic diseases.
- 7 Ask about the leading cause of death.

Sample FLE Goals

- 1 Describe the local facilities.
- 2 Describe the local staff.
- 3 Describe the nearest pharmacy.
- 4 Convey data on the population.
- 5 Describe any endemic diseases.
- 6 Describe the leading cause of death.
- 7 Convey health information.



Sample Transcripts with Back Translations

Line	Role Player	Back-translation from MT Output Used by Role Players for Communication	“Practice” Back-translation MT Output not used by Role Players for Communication
1	SME:	I need your help	
2	FLE:	زه نه شم کولای (<i>I can not</i>)	
3	SME:		Please help and take me to the explosion
4	SME:		Please help me gets to the explosion
5	SME:	Please help me go to the explosion	
6	FLE:		ما ته محفوظ ځای وو خان دی (<i>I was khan is safe</i>)
7	FLE:	زه وېرېږم (<i>I am afraid</i>)	
8	SME:		You will be okay come with me
9	SME:		You will not gephardt come with me
10	SME:	You will not be hurt come with me	
11	FLE:		صحیح ده (<i>Okay</i>)
12	FLE:	صحیح ده مننه (<i>Okay thank you</i>)	
13	SME:		Can you tell these people to stay far away
14	SME:		Tell all these people to stay back
15	SME:		Gets the people to back up
16	SME:		Tell the people to go away
17	SME:		Do you tell the people to go away
18	SME:	Tell the people to go away	
19	FLE:	ما سره مرسته وکړئ چې زه بې ته ولاړ شم (<i>Help me so that I can go</i>)	



Task Inventory

Task ID	TRADOC Task		ILR Speaking Skill Level
<i>Civil Affairs</i>			
CA2	331-38B-2020	Conduct a Local Medical Health Assessment	L2
CA4	331-38B-3015	Coordinate Handling of Supplies	L1+/2
CA5	331-38B-3033	Conduct Support to Civil Administration Operations	L2+
<i>Situational Awareness</i>			
SA1	301-35M-1200	Implement Approach Strategies	L3
SA2	301-35M-1250	Assess Source for Truthfulness and Accuracy	L2
SA3	191-376-5126	Conduct Interviews	high L1
<i>Checkpoint Operations</i>			
CP2	171-137-0001	Search Vehicles in a Tactical Environment	L1
CP6	191-376-5151	Control Access to a Military Installation	high L1
<i>Base Security</i>			
BS1	191-376-4130	Operate a Roadblock as a Member of a Team	L1
BS4	191-376-5154	Respond to a Crisis Incident	L0+



Sample Transcripts from Role Plays

Human Interpreter

Civil Affairs Scenario CA12, Team 6

SME: how many men and trucks can you provide ?
FLE: two cars four people four guns .
SME: do you know the roads very well ?
FLE: yes I know the roads very well .
SME: can you make sure nobody steals the supplies ?
FLE: okay . I'm trying my best .
SME: do you know the safest way to get there ?
FLE: yes I know a safe way .
SME: how many men can you provide ?
FLE: I can provide you with four men .
SME: how many trucks ?
FLE: two trucks .
SME: sounds good . I like the plan .
FLE: no problem . I'm thankful to you .

Machine Translation

Base Security Scenario BS45, Team 6

SME: I need your help .
FLE: I can not
SME: please help me go to the explosion .
FLE: I am afraid
SME: you will not be hurt . come with me .
FLE: Okay thank you
SME: tell the people to go away .
FLE: Help me so that I can go
SME: move the people .
FLE: I did not understand
SME: help me move people back .
FLE: I am afraid I can not help you
SME: then move back so you do not get hurt .
FLE: I know where do I have time I did not help



Working Through Garbled MT

Civil Affairs Scenario CA54, Team 9

SME: thank you for meeting with me .

FLE: Not coming on us

SME: what did you say ?

FLE: All the last

SME: what's wrong with your courts ?

FLE: Every work is security

SME: why are trials not happening ?

FLE: They are hidden nervous

SME: why are you nervous ?

FLE: Put under the work

SME: your public needs to trust you and be safe .

FLE: He can be

SME: what did you say ?

FLE: Our security officials work

SME: will you let us help you ?

FLE: It is

SME: by having your public trust your court
your town will be more functional .

FLE: We have

SME: what do you need from us to help you ?

FLE: In our security is

SME: is all you need security ?

FLE: We have

SME: what did you say ?

FLE: The security is for the country

SME: I'm concerned about your court system
not your country .

FLE: If we have security

SME: to confirm all you need is security from me .

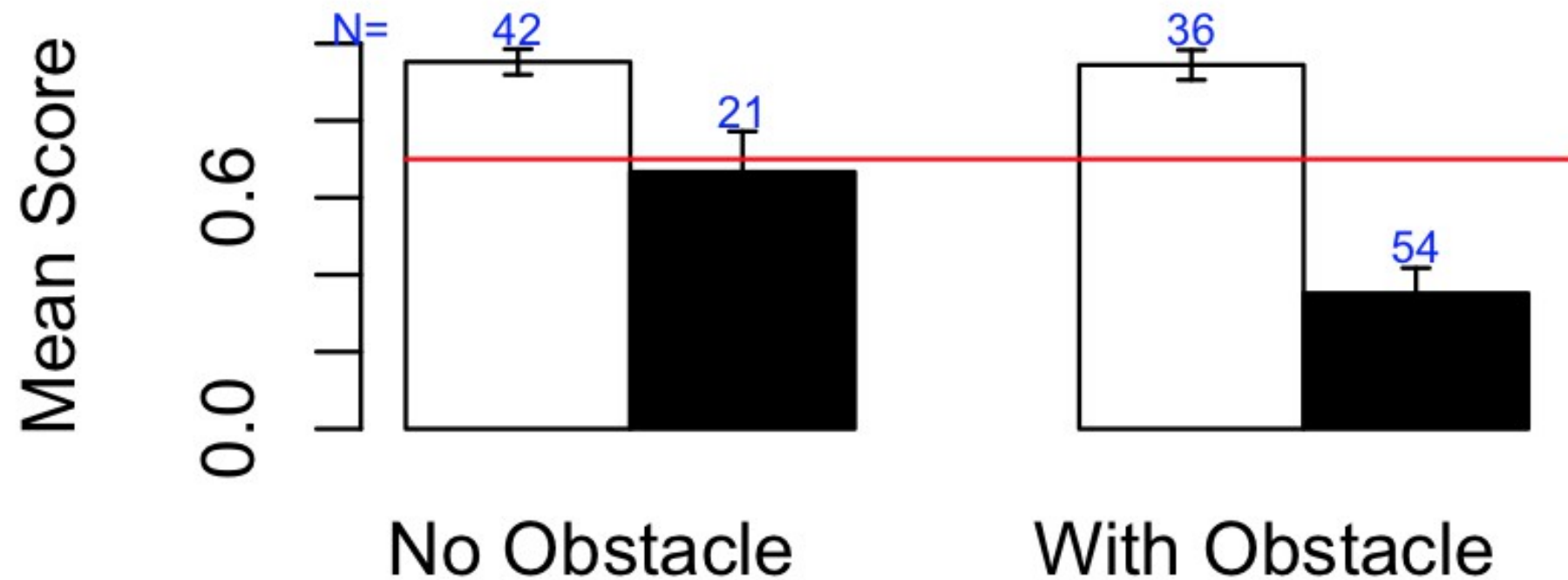
FLE: We were

SME: okay . we will send people over tomorrow .

FLE: Thank you very much it

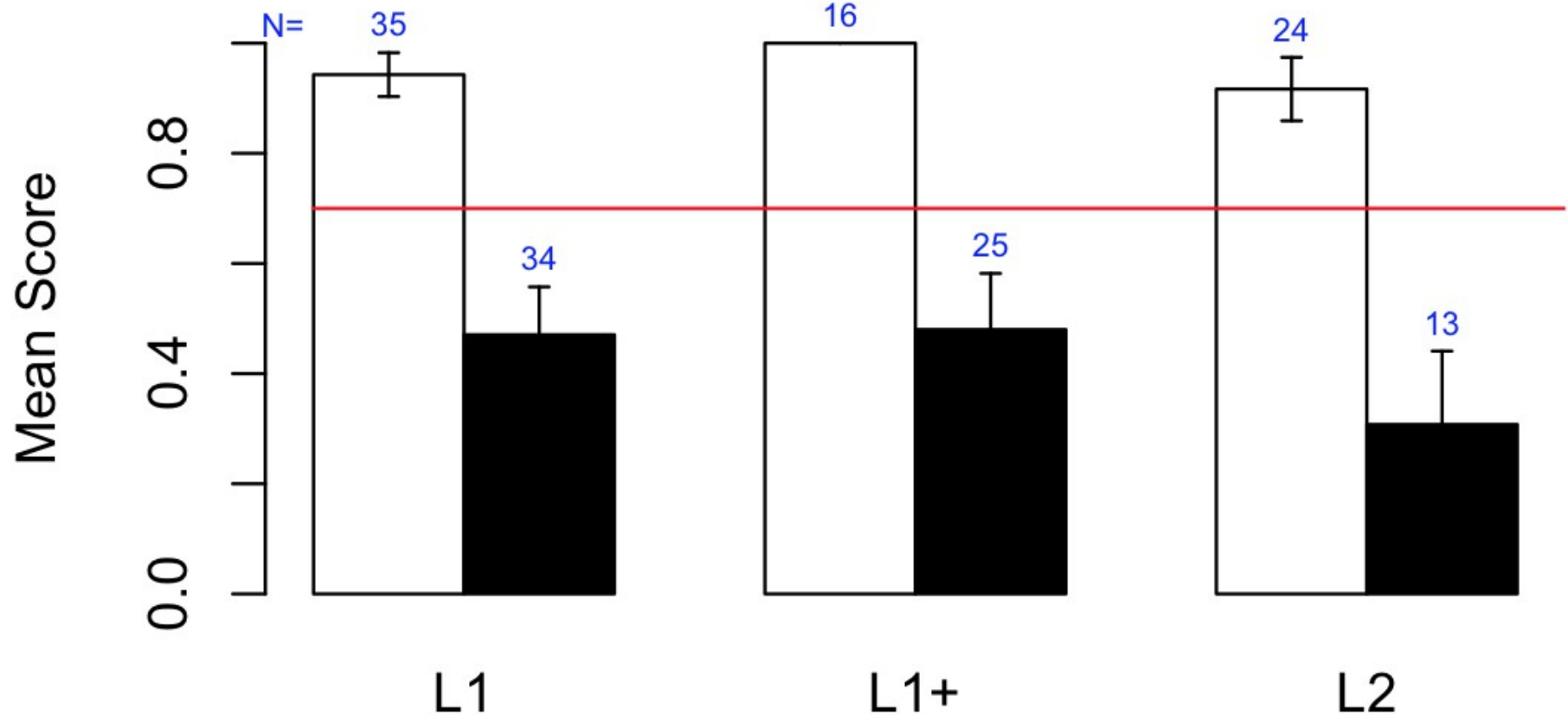


PASS/FAIL Results with Task Obstacles



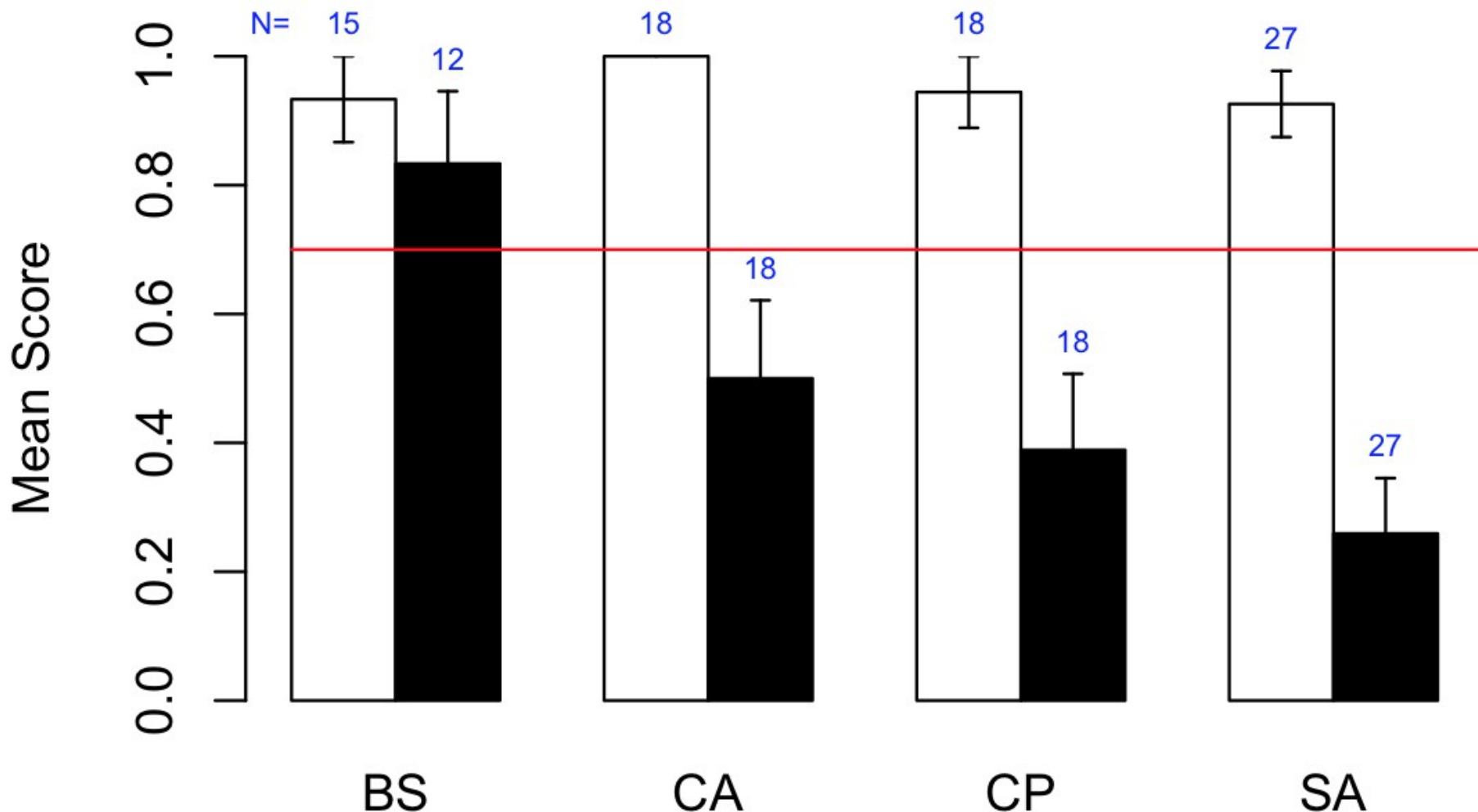


Scenario Outcomes by ILR Score



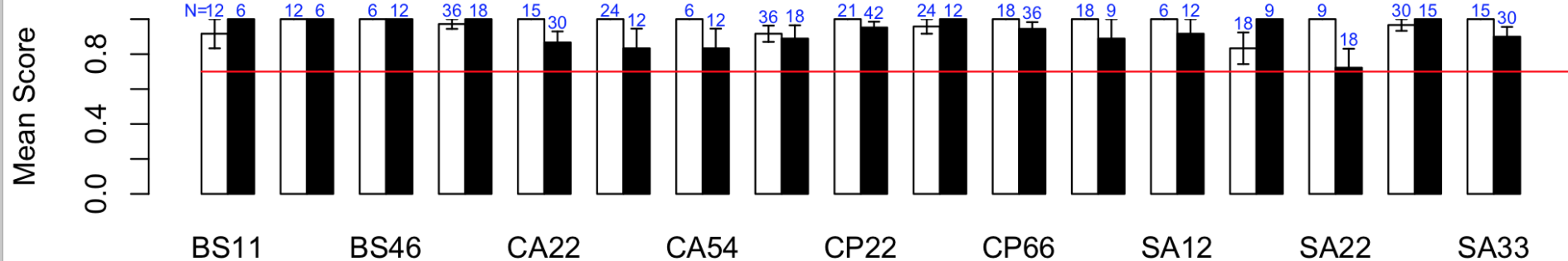


PASS/FAIL Results Across Domains



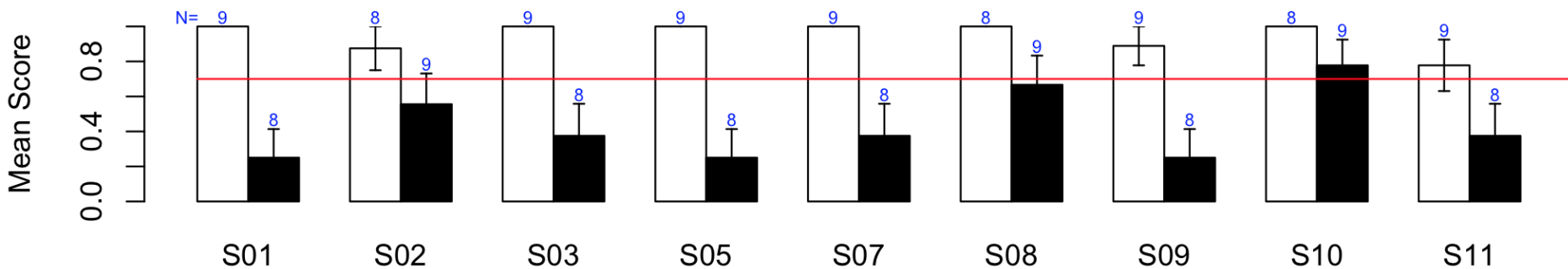


Communication Initiative Scores for Scenarios



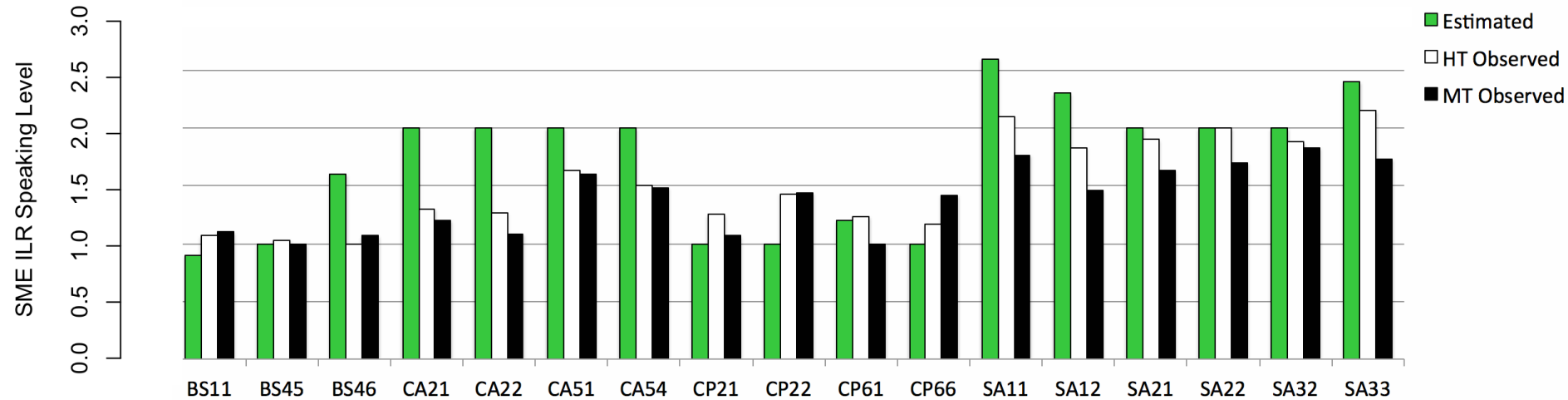


PASS/FAIL Scores for Scenarios



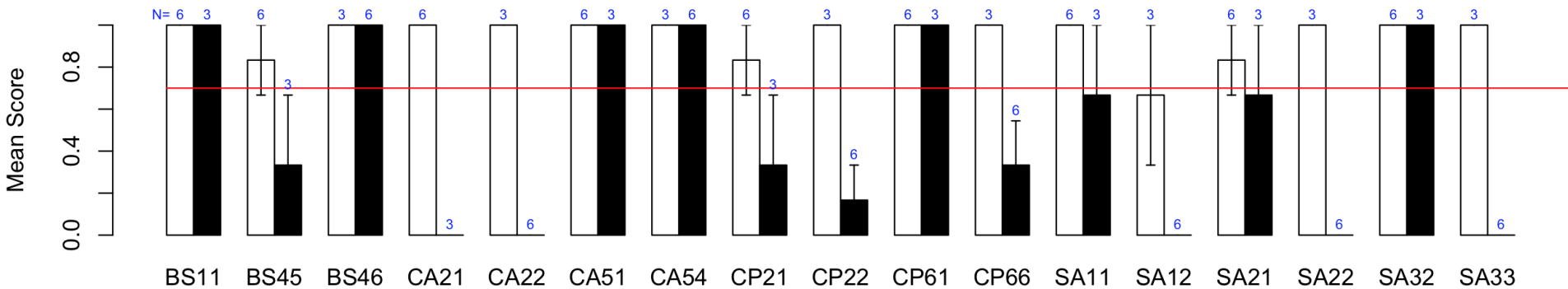


ILR Speaking Skills Estimated and Observed for Scenarios





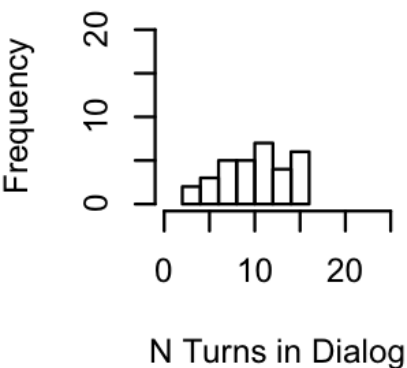
PASS/FAIL Scores by Subject Team



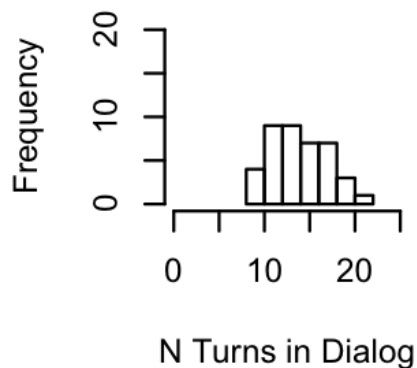


Length of Dialogs and PASS/FAIL Scores

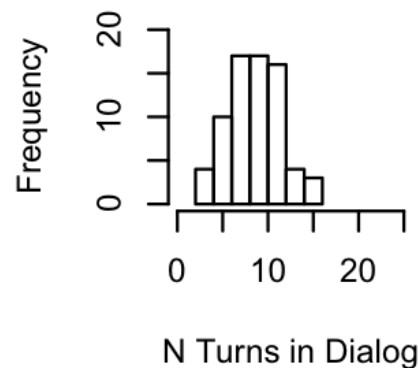
MT / PASS



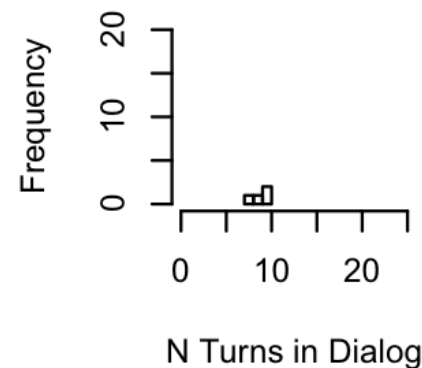
MT / FAIL



HT / PASS



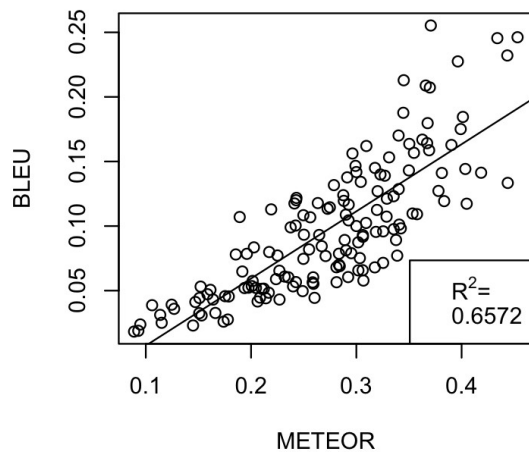
HT / FAIL



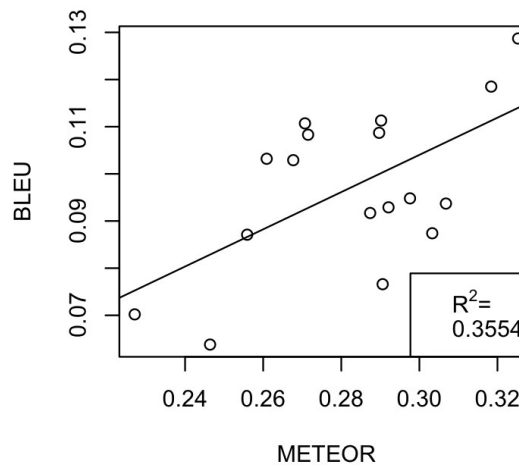


Automatic Scores

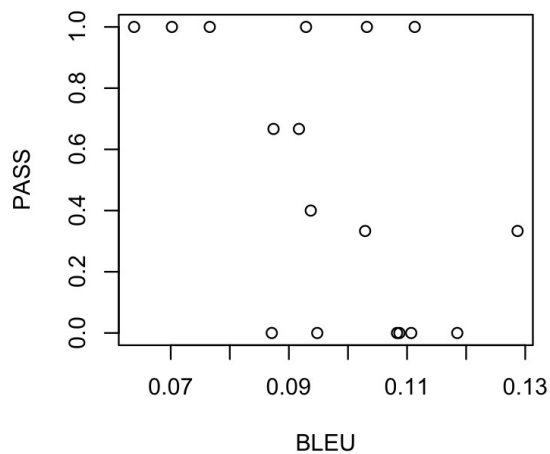
Mean Dialog Transcript Scores



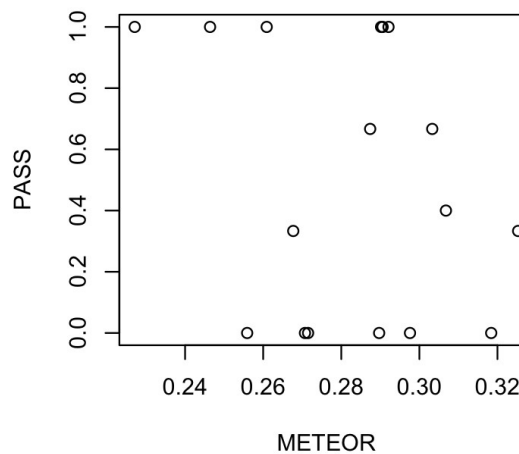
Mean Scenario Transcript Scores



Mean Scenario Transcript Scores



Mean Scenario Transcript Scores





Conclusions

- **The key lessons we learned in the experiment is that it is not enough for the role players to express themselves. What they say has to be understood by their conversational partner for the scenario to be completed successfully.**
- **Moreover, success depended as much on task simplicity as it did upon the translation condition, given that 67% of the base case scenarios were successfully completed using MT, but only 35% of the contrasting scenarios with even minor obstacles received passing scores.**
- **In other words, we observed that MT had the greatest chance of success when the task was simple and the language complexity needs were low.**
- **The implication of the performance variation shown in these results is that the technology should be tested in advance for specific situations in which it might be used, and not to assume that it just “works” for a particular broad domain.**