### RESULTS 1Q94 ARPA EVALUATION OF MACHINE TRANSLATION

March 17, 1994



### Part 3 - System Test Results

- Computation Methods
- · Comprehension, Fluency, Adequacy
- Comparison with 2Q93
- FAMT systems by language

### Time Computation

Time axis = Translation time /AVG. (manual 93 time + manual 94 time)

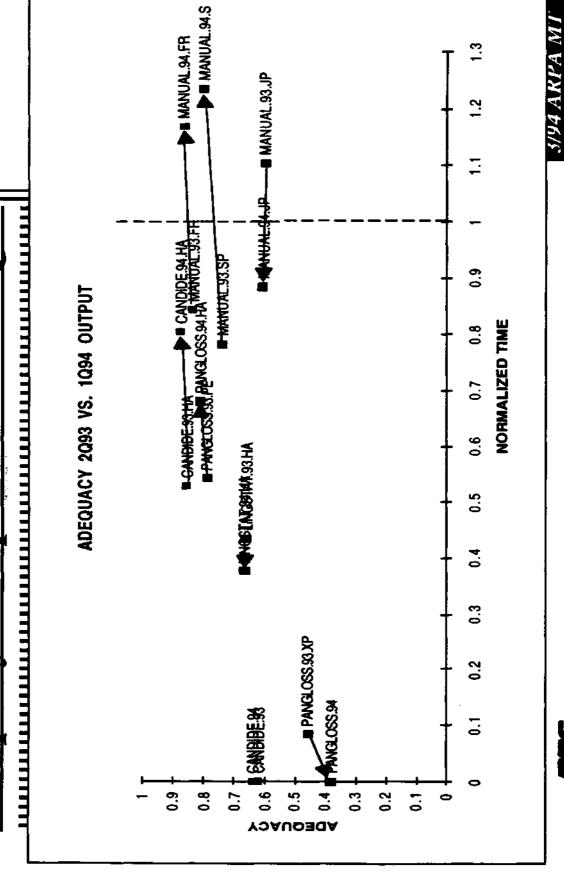
Mean of operator time, manual time for each system

Presumes indirect comparability

### 3/94 ARPA MT Workshop

### Comprehension, Fluency and Adequacy Computation

- Passage Score between 0 and 1
- Comprehension<sub>(P)</sub> = #Correct/6
- Fluency<sub>(P)</sub> =  $(\sum ((Judgment point 1)/(5-1)))$ /#sent. in passage
- Adequacy<sub>(P)</sub> =  $(\sum ((Judgment point 1)/(5-1)))$ /#frags. in passage
- System Score:
- Mean and STD DEV calculated for passage score over 20 passages
- STD DEV calculated over all system scores
- F-Ratio as ((VAR of System Means / mean of System VARS)



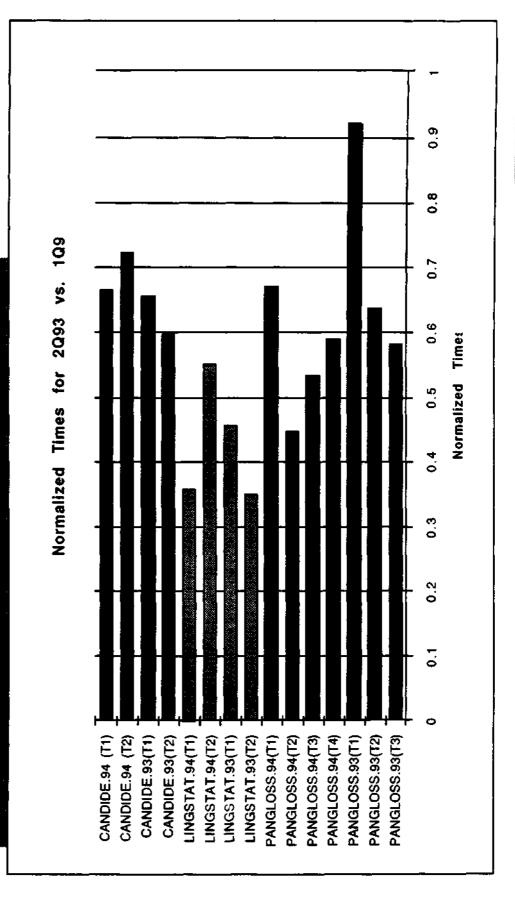
Workshop

3

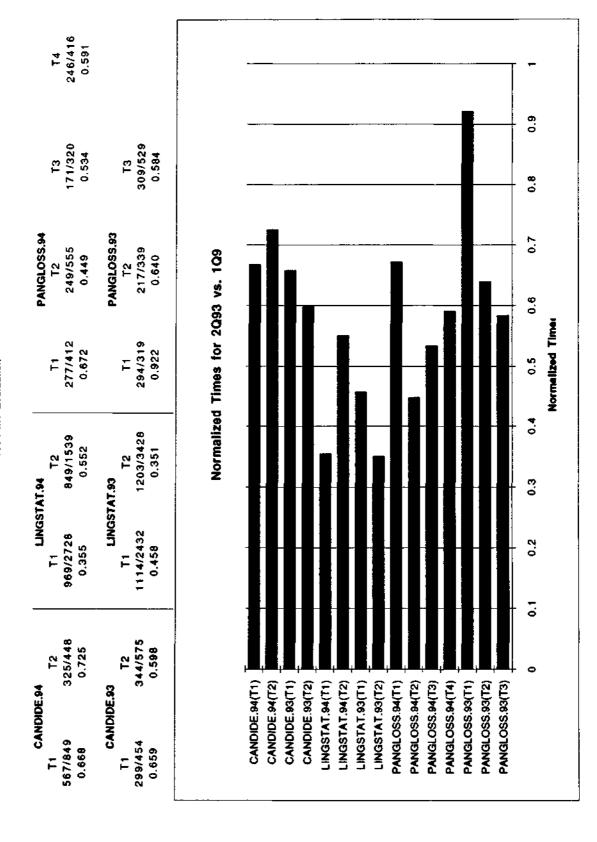
ADEQUACY 2080 VS. 1094 OUTPUT - 1094 MT EVALUATION

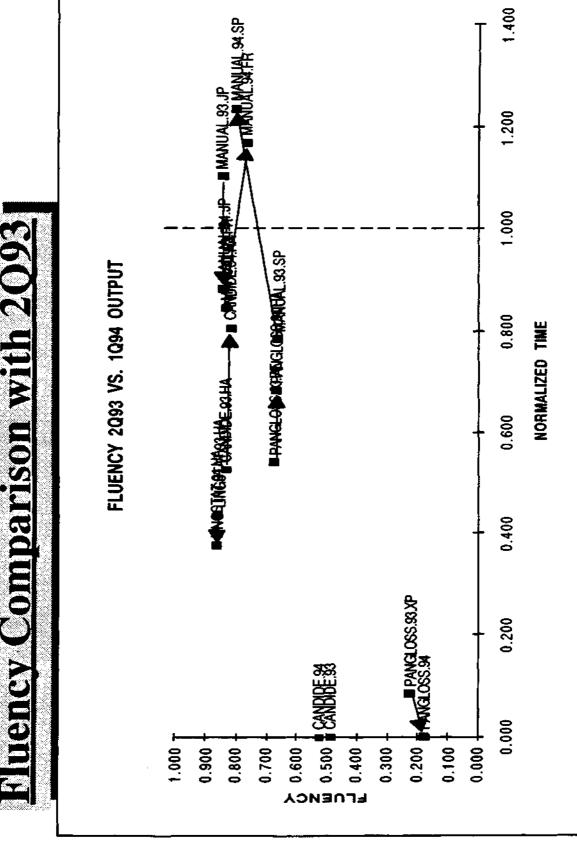
	CANDIDE:94	CANDIDE:94 CANDIDE:94.HA	₹		PANGLOSS:94.HA LINGSTAT.94.HA MANUAL.94.FR MANUAL.94.JP MANUAL.94.SP	MANUAL 94.FR A	MNUAL.94.JP I	WANUAL.94.SP	Ž.	STODEV	VARIANCE
(Pasa.Scorea/5)/ Num of Sentences	0.638	0.875	0.377	0.810	0.663	0.863	0.604	0.797	0.703	0,157	0.025
Var of 20 pts std dev of 20 pts	0.031	0.006	0.032	0.019	0.025 0.158	0.007	0.038 0.195	0.020	0.023		Total F Ratio 4.892
Norm Time Var of 20 pts and dev of 20 pts	0.000	0.805 0.101 0.318	0.000	0.682 0.037 0.194	0.377 0.005 0.072	0.212	0.885 0.105 0.325	1.237 0.113 0.336	0.645 0.072 0.213	0.449	AVG STD DEV 0.033
(Pass.Scores/5)/ Num of Sentences	CANDIDE.93 0.619	CANDIDE.99.HA. 0.854	PANGLOSS:80.XI 0.456	CANDIDE:90 CANDIDE:89.HAPANGLOSS:80.XP PANGLOSS:30.PE LINGSTAT:90.HA MANUAL:93.FR MANUAL:93.JP MANUAL:93.SP 0.619 0.854 0.456 0.787 0.660 0.836 0.595 0.739	LINGSTAT.99.HA 0.660	MANUAL.93.FR N	AANUAL.93.JP 1 0.595	MANUAL.93.SP 0.739	MEAN 0.693	STD DEV 0.127	VARIANCE 0.016
Var of 22 pts skd dev of 22 pts	0.017	0.019 0.140	0.031	0.015	0.021	0.011	0.034	0.017	0.021		Total F Ratio 3.664
Norm Time Var of 22 times aid dev of 22 times	0.000	0.528 0.028 0.168	0.084 0.018 0.133	0.542 0.026 0.162	0.437 0.003 0.050	0.845 0.058 0.242	1.105 0.067 0.259	0.784 0.056 0.256	0.540	0.350	AVG STD DEV 0.050
ADEQUACY  L Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q		E.B.B.B.P.M.GLOSS:80.XP	*	ADEQUACY 2098  ESANDHDESSTIN  PANGLUSSSSS  MANGEONICOUNTRY  AND  TO STAND THE STAND TH	3 VS. 10	MANUALSONA SS4. MANUALSONA MANUALSOSP	4	MANUAL 93.JP	UALS4.5F		
		5.0	0.3	. • • • • • • • • • • • • • • • • • • •	0.7 LCZED TBMES	6.8		. 2:	. <del>.</del> .		
										$\neg$	

## HAMT/Manual by Translate



1994 MT Evaluation



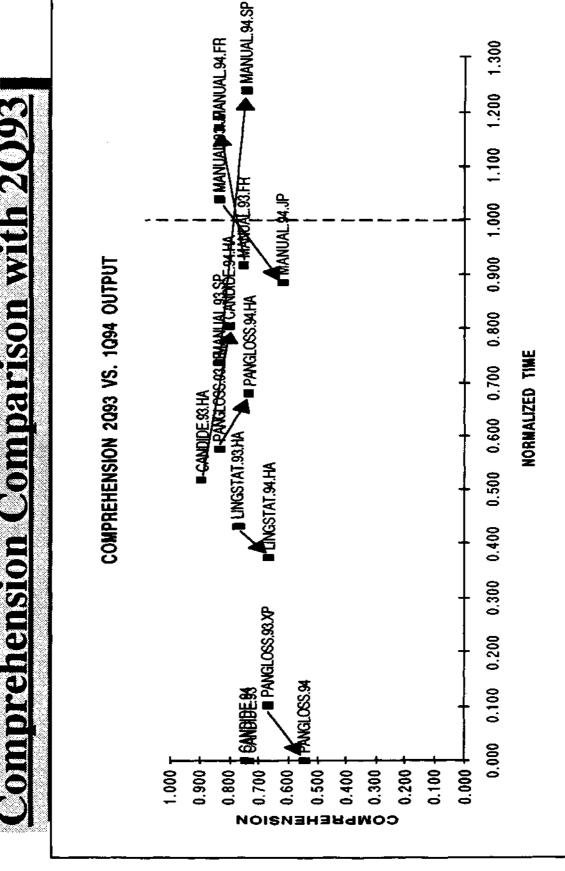


3/44 AKTA MI

Workshop

FLUENCY 2093 VS. 1094 OUTPUT - 1094 MT EVALUATION

0.051 0.021 0.019 0.065 0.062 0.764 0.841 0.797 0.051 0.025 0.018 0.029 0.018 0.029 0.018 0.029 0.018 0.029 0.019 0.029 0.018 0.029 0.019 0.029 0.019 0.029 0.019 0.029 0.017 0.029 0.017 0.020 0.020 0.029									THE PERSON NAMED IN			
0.005 0.005 0.009 0.009 0.016 0.002 0.018 0.029 0.016 0.002 0.018 0.029 0.015 0.029 0.029 0.016 0.002 0.018 0.0191 0.0194 0.0127 0.0178 0.0198 0.0171 0.000	Num of Sentences	0.524	0.017	0.176	0.665	0.062		0.841	0.797	0.681	0.217	0.047
0.000 0.806 0.000 0.682 0.377 1.171 0.885 1.237 0.000	Var of 20 pls	0.051	0.021	0.019	0.039	0.016	0.032	0.018	0.029	0.028		Total F Ratio
0.000 0.805 0.000 0.682 0.377 1.171 0.885 1.237 0.000 0.000 0.037 0.005 0.212 0.105 0.113 0.000 0.0318 0.000 0.037 0.005 0.212 0.105 0.113 0.000 0.0318 0.022 0.000 0.0318 0.022 0.000 0.0318 0.022 0.0018 0.015 0.000 0.0218 0.014 0.016 0.015 0.000 0.022 0.0014 0.015 0.000 0.022 0.0014 0.015 0.000 0.022 0.0014 0.015 0.015 0.000 0.025 0.014 0.016 0.015 0.020 0.025 0.025 0.025 0.025 0.000 0.025 0.0018 0.022 0.000 0.025 0.000 0.000 0.025 0.000 0.000 0.025 0.000 0.000 0.025 0.000 0.	aid dev of 20 pts	0.225	0.145	0.139	0.198	0.127	0.178	0.135	0.171	0.165		
0.000 0.101 0.000 0.037 0.005 0.113 0.105 0.113 0.005 0.222 0.460 0.325 0.336 0.100 0.037 0.005 0.007 0.005 0.325 0.336 0.223 0.223 0.670 0.685 0.890 0.0142 0.680 0.025 0.014 0.016 0.015 0.009 0.127 0.114 0.142 0.000 0.028 0.0018 0.028 0.005 0.0124 0.009 0.127 0.114 0.142 0.142 0.000 0.028 0.018 0.026 0.003 0.0127 0.114 0.142 0.056 0.000 0.028 0.018 0.026 0.003 0.027 0.056 0.005 0	Norm Time	0.000	0.805	0.000	0.682	0.377	1.171	0.885	1.237	0.645	0.449	AVG STD DEV
CANDICE 20 CANDICE 20 IAPANGLOSS 20.7P PANGLOSS 20.7P CANDICE 20 IAPANGLOSS 20 IAPANGLOS 20 IAPANGLOSS 20 IAPANGLOSS 20 IAPANGLOSS 20 IAPANGLOSS 20 IAPAN	Var of 20 pis	0.000	0.101	0.00	0.037	0.005	0.212	0.105	0.113	0.072		0.037
CANDIDE 50 CANDIDE 50 HAPANGLOSS 30.0P PANGLOSS 30.PE LINGSTAT 30.HA  0.486 0.833 0.223 0.670 0.656 0.027 0.014 0.016 0.015 0.009 0.165 0.118 0.124 0.093 0.000 0.528 0.084 0.542 0.437 0.000 0.028 0.018 0.056 0.003 0.000 0.028 0.018 0.056 0.003 0.000 0.168 0.133 0.162 0.050  FILIENCY 2093 VS. 1094  ENVELOSS 30.XP	std dev of 20 pts	0.000	0.318	0.000	0.194	0.072	0.460	0.325	0.336	0.213		
0.486 0.833 0.223 0.670 0.856 0.830 0 0.027 0.014 0.016 0.015 0.009 0.016 0 0.165 0.118 0.128 0.124 0.093 0.127 0 0.000 0.528 0.084 0.542 0.437 0.845 1 0.000 0.028 0.018 0.026 0.003 0.058 0 0.000 0.168 0.133 0.162 0.050 0.242 0  FLUENCY 2083 VS. 1094  FLUENCY 2083 VS. 1094  FLUENCY 2083 VS. 1094		CANDIDE:89	CANDIDE 93. HAP	WANGLOSS 393.XF	PANCLOSS.90.PE		MANUAL.93,FR	MANUAL B3.JP	MANUAL 93.SP	MEAN	STDDEV	VARIANCE
0.165 0.118 0.128 0.124 0.099 0.127 0 0.105 0.108 0.124 0.093 0.127 0 0.000 0.528 0.084 0.542 0.437 0.845 1 0.000 0.028 0.018 0.026 0.003 0.058 0 0.000 0.168 0.133 0.162 0.050 0.242 0  FLUENCY 2083 VS. 1094  FLUENCY 2083 VS. 1094  FLUENCY 2083 VS. 1094  FLUENCY 2083 VS. 1094	(Pass.Scores/5)/ Nam of Sentences	0.486	0.833	0.223	0.670	0.856	0.830	0.843	0.660	0.675	0.209	0.044
0.027 0.014 0.016 0.015 0.009 0.016 0.016 0.124 0.093 0.127 0.128 0.124 0.093 0.127 0.016 0.000 0.028 0.128 0.124 0.093 0.127 0.084 0.500 0.000 0.028 0.0018 0.026 0.000 0.058 0.000 0.000 0.018 0.026 0.000 0.050 0.242 0.000 0.168 0.133 0.162 0.050 0.242 0.000 0.168 0.133 0.162 0.050 0.242 0.000 0.168 0.133 0.162 0.050 0.242 0.000 0.168 0.133 0.162 0.050 0.242 0.000 0.168 0.133 0.162 0.050 0.242 0.000 0.168 0.133 0.162 0.050 0.242 0.000 0.168 0.000 0.168 0.050 0.162 0.050 0.162 0.000 0.168 0.000 0.168 0.000 0.168 0.000 0.168 0.000 0.168 0.000 0.168 0.000 0.168 0.000 0.162 0.00												Total F Ratio
0.165 0.118 0.128 0.124 0.093 0.127 0 0.000 0.528 0.084 0.542 0.437 0.845 11 0.000 0.028 0.018 0.026 0.005 0.058 0 0.000 0.168 0.133 0.162 0.050 0.242 0 FLUENCY 2093 VS. 1094  FLUENCY 2093 VS. 1094	Ver of 22 pts	0.027	0.014	0.016	0.015	0.009	0.016	0.013	0.020	0.016		12,584
0.000 0.528 0.084 0.542 0.437 0.845 11 0.000 0.028 0.018 0.026 0.003 0.058 0 0.000 0.188 0.133 0.162 0.050 0.242 0 0.000 0.188 0.133 0.162 0.050 0.242 0  FLUENCY 2093 VS. 1094	atd dev of 22 pts	0.165	0.118	0.128	0.124	0.093	0.127	0.114	0.142	0.126		
0.000 0.028 0.018 0.026 0.003 0.058 0 0.000 0.168 0.133 0.162 0.050 0.242 0  FLUENCY 2093 VS. 1094	Norm Time	0.000	0.528	0.084	0.542	0.437	0.845	1.105	0.784	0.540	0.350	0.350 AVG STD DEV
0.000 0.168 0.133 0.162 0.050 0.242 0 FLUENCY 2Q83 VS. 1Q94	Var of 22 times	0.000	0.028	0.018	0.026	0.003	0.058	0.067	990'0			0.045
1.000 0.900 0.800 0.500 0.400 0.300 0.300 0.300	d dev of 22 times	0.000	0.168	0.133	0.162	0.050	0.242	0.259	0.256			
1.000 0.900 0.900 0.500 0.500 0.400 0.300 0.300					FLUENCY 20	283 VS. 1094						
0.900 0.900 0.900 0.900 0.900 0.900 0.900 0.900 0.900 0.900 0.900	1.000	<b>-</b>					<u>.</u> .					
0.500 CANDESS 99.XP	000						<b>-</b> -					
0.500 CANDEESS 0.400	0.80				STATE SOLDE OF THE	-	Charles of the Control of the Contro	- MAANUAL.83.JI				
0.600 - 0.500 EMBESS 0.400 - 0.300 - 0	008.0						  - 	SON STATE	かま正なり			
0.500 EXXEDE 83 0.400 0.300 0.300	0.700				MPANOLOGICAL	PROGRAMME O	386					
0.300					•							
0.300	0.500	EXABLESS STATES										
	0.300											
b	0.200	VC OSS	IGLOSS,88,XP									
0.100	0.100						. <del></del>					
0000	0.000					+	1					
0,000 0,200 0,400 0,600 0,800 1,000 1,200 1,400	9.0	90	0.200	0.400	0.600	0.600	1.000	1.200	1.400			
NOBMALIZED THE												



3/94 ARPA MT Workshop

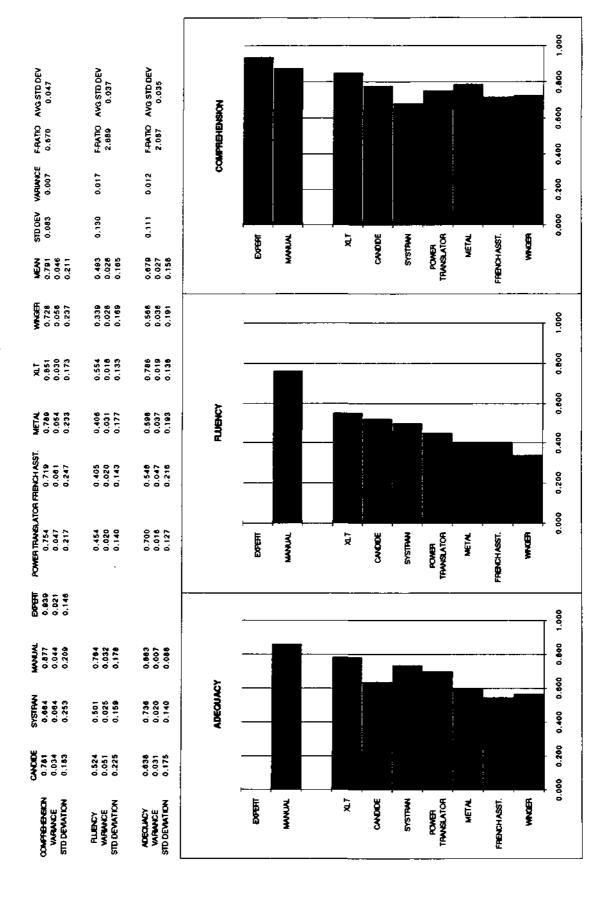
34

COMPREHENSION 2093 VS. 1094 OUTPUT - 1094 MT EVALUATION

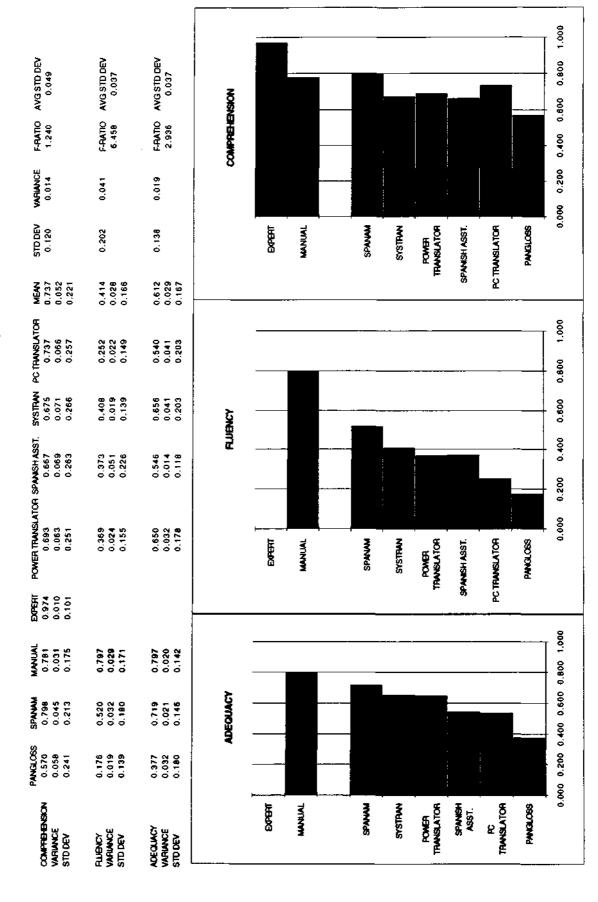
(Page Section)				1		MANNIAL OF CO MANNIAL OF 10 MIANILLA OF CO	MANITAL OF IV	77 VIII 101 117 VIII			COMPANY
Num of Questions		0.800	0.542	0.733	0.667		0.617	0.742	0.709	0.090	
Var of 20 pts std dev of 20 pts	0.032	0.066	0.055 0.235	0.048	0.053 0.230	0.042	0.059 0.242	0.029	0.048		Total F Ratio 0.756
Norm Time Var of 20 pts std dev of 20 pts	0.000 0.000 0.000	0.805 0.101 0.318	0.000	0.662 0.037 0.194	0.377 0.005 0.072	1.171 0.212 0.460	0.885 0.105 0.325	1.237 0.113 0.336	0.645 0.072 0.213	0.449	AVG STD DEV 0.049
(Pass.Scores/6)/ Num of Questions		CANDIDE,03,HAP 0,896	ANGLOSS 93.X0 0.667	CANDIDE.93 CANDIDE.92.HAPANGLOSS.93.XP PANGLOSS.93.PE 0.729 0.896 0.667 0.833	LINGSTAT.93.HA 0.771	MANUAL.93.FR MANUAL.93.JP MANUAL.93.SP 0.750 0.833 0.833	MANUAL.93.JP 1 0.833	MANUAL.93.SP 0.833	MEAN 0.789	STD DEV 0.069	VARIANCE 0.005
Var of 8 pts std dev of 8 pts	0.023	0.023	0.063	0.016	0.031	0.048	0.040	0.040	0.036 0.185		Total F Ratio 0.375
Norm Time Var of 8 pts std dev of 8 pts	0.000	0.519 0.049 0.220	0.104	0.577 0.025 0.157	0.433 0.000 0.017	0.919 0.092 0.304	1.038 0.074 0.273	0.738 0.053 0.230	0.541 0.049 0.151	0.340	AVG STD DEV 0.065
CONFRENCION	0.500	WGLOSS	<b>₩</b>	COMPREHEN  CANDID  CAN	COMPREHENSION 2083 VS. 1094  "CANDUE SAHA"  "RANTESAHA"  "NOSTAT.93.HA"  "PANCE OSS.94.HA"  "TAT.94.HA"  "AND TAT.94.HA"  "TAT.94.HA"  "TAT.94.HA"	Miles San Server Sasas HA	FEST HA STANKUA	MAAYLAL DO HRANUAL SA FR	NAL, 94, FR		

0.000 | 1.000 0.200 0.300 0.400 0.500 0.600 0.700 0.800 0.900 1.000 1.100 1.200 1.300 NORMALIZED TMAE

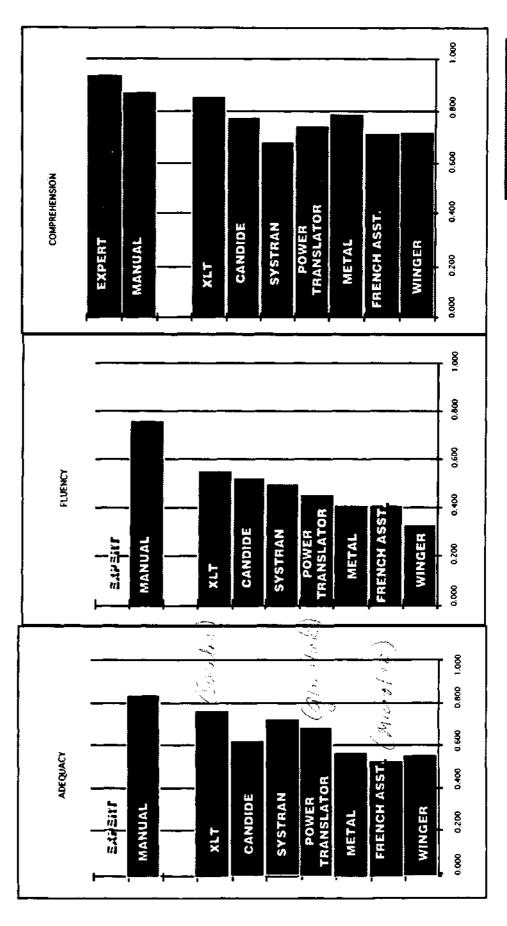
FAMT EVALUATION FOR FRENCH, 1094



FAMT EVALUATION FOR SPANISH, 1094

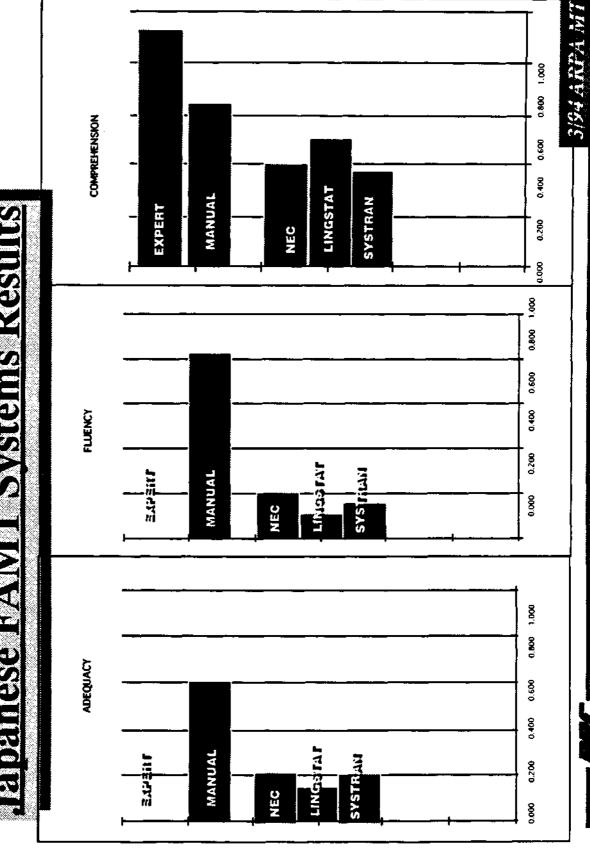


## French FAMT Systems Results



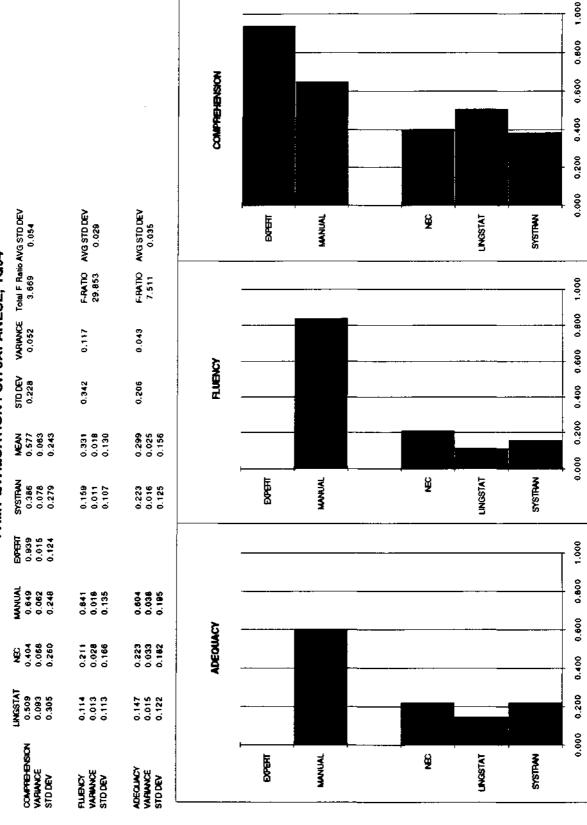
3/94 A RPA MIT Workshop

# Japanese FAMT Systems Results



37

FAMT EVALUATION FOR JAPANESE, 1094



### Summary

- All three evaluations have value
- Did calibrate underlying MT technologies
- But not intended as diagnostic of system value/effectiveness
- · HAMT .. NOT
- Valuable portability exercise
- Gracious participation by production systems
- A healthy technology evaluation paradigm for both ARPA and the MT industry