ATR Japanese Corpus (Spoken Language DataBase)

		# of Morphemes		# of Characters	
ATR Corpus	**# of Sent.**	**Range**	**Ave.**	**Range**	**Ave.**
Training set	10,361	1-34	6.69	2-58	12.57
Test set	545	1-22	6.36	2-42	12.03

Corpus	APB
SUSANNE	1.256
SEC (Spoken English Corpus)	1.239
ATR (character-base)	1.348

Number of parses $= APB^n$

where n is the number of words in a sentence: Average Parse Base (APB)

Experimental Results

Models	2-42 Characters (545 sentences)						
	PA	**LP**	**LR**	**BP**	**BR**	**0-CB**	**m-CB**
B&C	88.62 (58.1%)	97.72	97.50	98.48	98.05	93.94 (75.7%)	0.15
Two-leval PCFG	62.39 (87.3%)	96.28	95.32	98.61	97.38	95.23 (69.2%)	0.10
PCFG	53.03 (89.8%)	95.67	94.54	98.77	97.35	94.86 (71.4%)	0.08
PGLR	**95.23**	**99.08**	**98.50**	**99.54**	**98.76**	**98.53**	**0.03**

Models	15-42 Characters (160 sentences)						
	PA	**LP**	**LR**	**BP**	**BR**	**0-CB**	**m-CB**
B&C	73.75 (61.9%)	96.00	97.26	96.84	98.14	83.75 (76.9%)	0.44
Two-level PCFG	56.25 (77.1%)	97.44	97.31	98.90	98.76	93.13 (45.4%)	0.18
PCFG	35.62 (84.5%)	95.86	95.64	98.60	98.39	90.63 (60.0%)	0.17
PGLR	**90.00**	**98.98**	**98.99**	**99.49**	**99.50**	**96.25**	**0.08**

LALR and CLR table-based PGLR

Models	2-42 Characters (534 sentences)						
	PA	LP	LR	BP	BR	0-CB	m-CB
PGLR(CLR)	95.13	99.04	98.40	99.46	98.61	97.57	0.04
PGLR(LALR)	95.32	99.06	98.47	99.53	98.73	98.50	0.03

PGLR model-2 against PGLR and B&C, on an open test

Models	2-48 Characters (500 sentences)						
	PA	**LP**	**LR**	**BP**	**BR**	**0-CB**	**m-CB**
B&C	90.20	97.26	97.39	97.95	98.03	94.60	0.14
PGLR	94.00	98.37	97.86	98.96	98.30	97.20	0.06
PGLR model-2	**94.00**	**98.60**	**98.02**	**99.15**	**98.36**	**97.20**	**0.05**

PGLR model-2 against PGLR and B&C, on a closed test

Models	2-48 Characters (500 sentences)						
	PA	**LP**	**LR**	**BP**	**BR**	**0-CB**	**m-CB**
B&C	96.00	98.53	98.34	98.89	98.62	98.00	0.04
PGLR	96.00	98.77	98.35	99.17	98.62	98.20	0.03
PGLR model-2	**97.00**	**99.15**	**98.53**	**99.53**	**98.68**	**98.80**	**0.02**