from relation detection and 27% (237/864) of errors results from relation characterization, among which 17.8% (154/864) of errors are from misclassification across relation types and 9.6% (83/864)

of errors are from misclassification of relation subtypes inside the same relation types. This suggests that relation detection is critical for relation extraction.
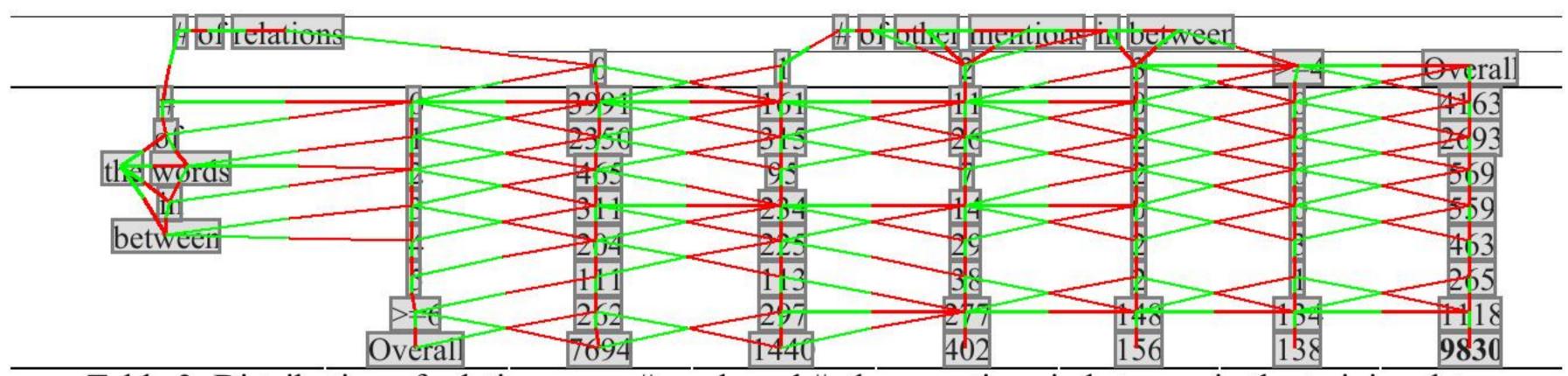

# of the words in between	# of other mentions in between					Overall
	0	1	2	3	>4	
0	3994	161	11			4163
1	2259	375	26			2693
2	455	92	7			559
3	311	234	11			559
4	234	225	20			463
5	11	113	38			265
>6	262	207	277	146	164	1118
Overall	7694	1440	402	156	138	**9830**

Table 3: Distribution of relations over #words and #other mentions in between in the training data

Type	Subtype	#Testing Instances	#Correct	#Error	P	R	F
AT		**392**	**224**	**105**	**68.1**	**57.1**	**62.1**
	Based-In	85	39	10	79.6	45.9	58.2
	Located	241	132	120	52.4	54.8	53.5
	Residence	66	19	9	67.9	28.8	40.4
NEAR		**35**	**8**	**1**	**88.9**	**22.9**	**36.4**
	Relative-Location	35	8	1	88.9	22.9	36.4
PART		**164**	**106**	**39**	**73.1**	**64.6**	**68.6**
	Part-Of	136	76	32	70.4	55.9	62.3
	Subsidiary	27	14	23	37.8	51.9	43.8
ROLE		**699**	**443**	**82**	**84.4**	**63.4**	**72.4**
	Citizen-Of	36	25	8	75.8	69.4	72.6
	General-Staff	201	108	46	71.1	53.7	62.3
	Management	165	106	72	59.6	64.2	61.8
	Member	224	104	36	74.3	46.4	57.1
SOCIAL		**95**	**60**	**21**	**74.1**	**63.2**	**68.5**
	Other-Professional	29	16	32	33.3	55.2	41.6
	Parent	25	17	0	100	68.0	81.0

Table 4: Performance of different relation types and major subtypes in the test data

System	Relation Detection			RDC on Types			RDC on Subtypes		
	P	R	F	P	R	F	P	R	F
Ours: feature-based	**84.8**	**66.7**	**74.7**	**77.2**	**60.7**	**68.0**	63.1	**49.5**	**55.5**
Kambhatla (2004):feature-based	-	-	-	-	-	-	**63.5**	45.2	52.8
Culotta et al (2004):tree kernel	81.2	51.8	63.2	67.1	35.0	45.8	-	-	-

Table 5: Comparison of our system with other best-reported systems on the ACE corpus

Error Type		#Errors
Detection Error	False Negative	462
	False Positive	165
Characterization Error	Cross Type Error	154
	Inside Type Error	83

Table 6: Distribution of errors

6 Discussion and Conclusion

In this paper, we have presented a feature-based approach for relation extraction where diverse lexical, syntactic and semantic knowledge are employed. Instead of exploring the full parse tree information directly as previous related work, we incorporate the base phrase chunking information

first. Evaluation on the ACE corpus shows that base phrase chunking contributes to most of the performance improvement from syntactic aspect while further incorporation of the parse tree and dependence tree information only slightly improves the performance. This may be due to three reasons: First, most of relations defined in ACE have two mentions being close to each other. While short-distance relations dominate and can be resolved by simple features such as word and chunking features, the further dependency tree and parse tree features can only take effect in the remaining much less and more difficult long-distance relations. Second, it is well known that full parsing