Project 1
ReadMe

Mike Roylance

roylance@uw.edu

Language coded	Ruby (1.93 and 1.85)	
Approach	I created three classes to do the heavy lifting in this project.	
	ptb.rb	This class was in charge of parsing the raw text and turning it into a usable structure.
	constituent.rb	This class was the basic model of a constituent. This class has part of speech, value and children properties.
	countPtb.rb	This class was in charge of counting the constituents for the S, NP, VP, Ditransitive and Intransitive patterns
	tests.rb	This file was my testing framework. I wrote three tests to help me validate my parsing approach.
	main.rb	This is the entry point for the application

Results Table:

File Path	S	NP	VP	Di	In
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1400.prd	5	23	7	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1484.prd	12	49	19	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1498.prd	2	9	3	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1450.prd	100	328	163	1	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1486.prd	1	28	7	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1444.prd	1	48	2	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1447.prd	48	135	82	0	3
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1432.prd	132	307	231	0	5
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1409.prd	50	93	80	2	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1452.prd	7	14	10	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1406.prd	5	11	6	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1446.prd	132	295	218	0	5
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1449.prd	50	173	81	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1431.prd	8	30	8	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1421.prd	95	158	151	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1434.prd	6	319	16	0	0

/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1494.prd	12	35	16	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1427.prd	3	17	3	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1445.prd	66	233	113	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1479.prd	8	41	12	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1475.prd	11	30	14	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1451.prd	9	26	16	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1478.prd	20	61	34	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1469.prd	124	315	216	1	3
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1496.prd	11	28	18	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1458.prd	16	81	30	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1411.prd	218	572	373	5	6
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1420.prd	3	6	4	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1492.prd	34	95	55	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1417.prd	6	20	9	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1410.prd	79	205	127	1	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1465.prd	15	62	33	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1466.prd	13	50	28	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1477.prd	10	24	12	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1430.prd	2	6	4	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1464.prd	172	353	277	1	15
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1437.prd	6	29	12	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1490.prd	123	383	221	2	8
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1428.prd	87	227	151	1	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1462.prd	20	48	38	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1487.prd	14	23	23	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1443.prd	19	64	34	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1497.prd	12	40	18	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1457.prd	143	316	249	0	4
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1424.prd	113	370	199	1	5
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1413.prd	37	76	72	3	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1481.prd	16	35	25	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1404.prd	47	141	87	2	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1440.prd	129	346	196	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1401.prd	2	17	6	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1441.prd	21	72	41	1	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1483.prd	6	24	13	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1453.prd	124	328	183	0	4
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1471.prd	23	72	53	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1493.prd	96	280	171	0	6
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1454.prd	11	31	18	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1436.prd	95	330	158	2	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1442.prd	142	362	221	0	7
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1439.prd	18	78	28	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1429.prd	61	118	98	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1467.prd	106	271	198	1	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1448.prd	64	234	125	1	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1416.prd	53	195	79	1	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1468.prd	65	157	120	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1415.prd	54	126	96	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1412.prd	31	66	60	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1402.prd	4	19	5	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1425.prd	4	17	6	2	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1459.prd	8	21	11	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1489.prd	13	31	24	0	0

/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1476.prd	9	40	15	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1480.prd	164	395	248	1	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1408.prd	4	11	7	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1419.prd	80	192	145	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1414.prd	9	26	11	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1460.prd	20	61	44	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1407.prd	28	89	45	1	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1426.prd	8	28	12	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1418.prd	69	185	118	0	3
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1463.prd	38	164	61	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1435.prd	99	217	174	2	5
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1403.prd	7	30	10	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1488.prd	9	29	11	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1422.prd	51	131	106	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1438.prd	21	58	27	1	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1455.prd	121	377	218	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1499.prd	91	197	143	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1495.prd	66	170	110	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1433.prd	10	20	17	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1472.prd	17	43	27	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1474.prd	115	285	196	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1405.prd	68	199	121	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1470.prd	93	239	189	1	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1461.prd	17	50	30	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1485.prd	10	36	15	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1423.prd	74	177	121	0	2
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1482.prd	2	167	4	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1456.prd	20	64	32	0	1
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1473.prd	90	268	137	0	4
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14/wsj_1491.prd	7	46	9	0	0
/corpora/LDC/LDC99T42/RAW/parsed/prd/wsj/14	4670	13221	7920	34	123