Working results

Tyler Peckenpaugh

4/3/2019

Contents

Inter-item timing

Subjects were asked to read each sentence twice, once with no preview at all, and then again after unlimited preview. Inter-reading time (IRT) is a measure of the amount of time between when a subject stops speaking after a cold reading and when they begin speaking for a previewed reading.

> IRT = delay after the end of a cold reading and before the start of a previewed reading

Practically, this was done over 1548 recordings (33 participants, 48 items = 1584 pairs, with 36 missing data). This was measured using Google's WebRTC Voice Activity Detection (VAD) over .wav files that had been subjected to a high-pass filter with a low threshold of 0 to 500Hz[1] using the highest aggressiveness that yielded good results, depending on the noise level in the recording.

[1] a low hum in the room needed to be accounted for; the exact algorithm is available at github (URL: bit.ly/2uMrcrG)

Description and cleanup

The following section details the IRT data and the outlier removal and resulting participant attrition.

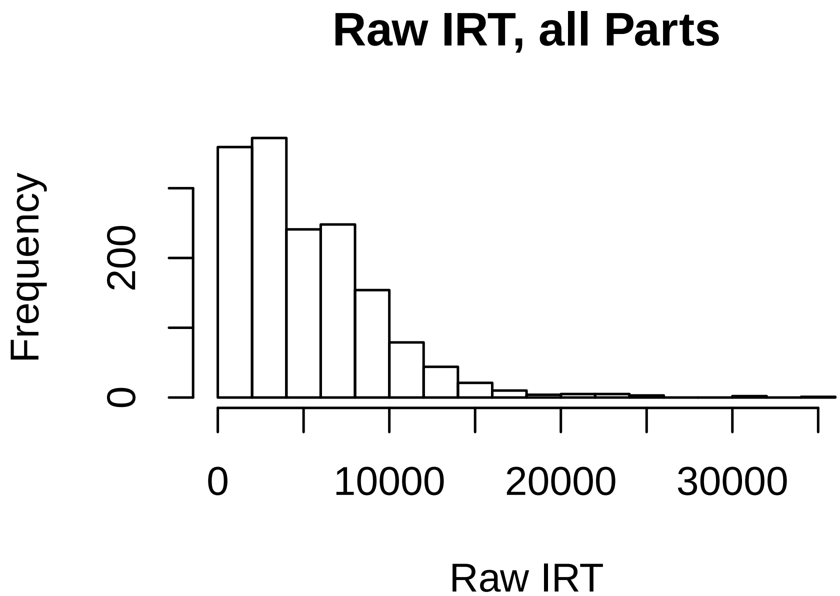
Distribution of IRTs, all participants

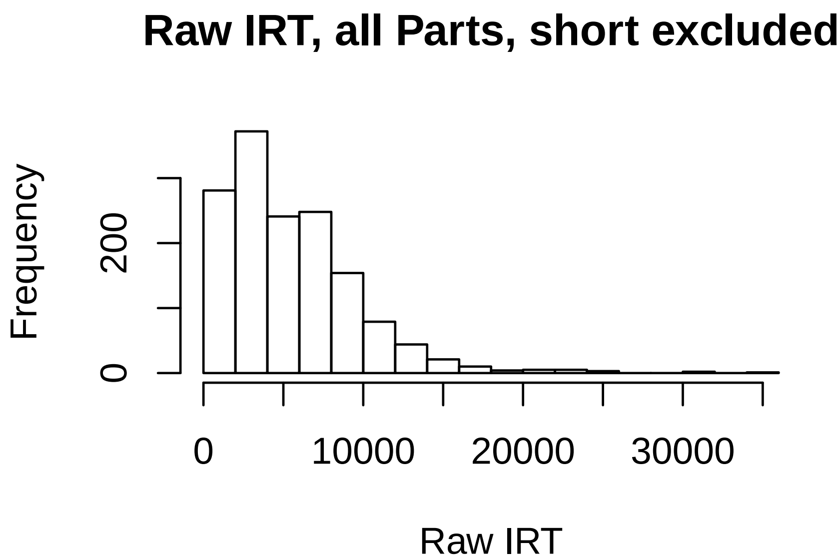
The overal mean IRT for all participants, all items (including fillers), and all conditions is 5317.99ms (sd = 4115.16). The highest IRT was 35762.85ms.

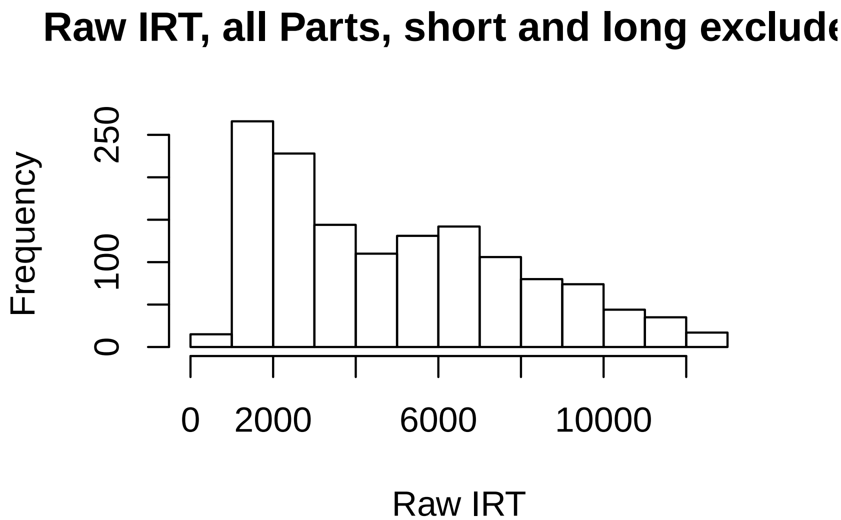
The following histograms show the distribution of IRT across all items and all participants. In the second graph, overly short IRTs (shorter than 899.109ms; 78[2] such data) are excluded. In the third, overly long (longer than 12734.687; 78 such data) and overly short IRTs are excluded.

[2] This is 5% of the 1548 total data

The third graph represent what I will call data that has undergone "basic outlier removal."

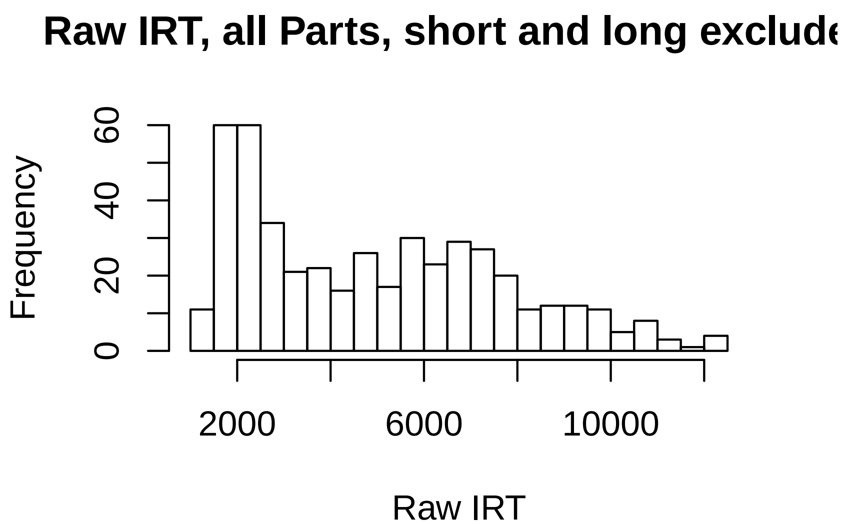

Raw IRT, all Parts
Frequency
0
200
0
10000
20000
30000
Raw IRT


Raw IRT, all Parts, short excluded
Frequency
0
200
0
10000
20000
30000
Raw IRT



IRTs were finally winsorized to lessen the impact of outliers.

A question for DCB: should the IRTs be winsorized by a participant's mean/sd for all items (including fillers), or only by experimental item mean/sd? I assume the former in this document. Should this be done before or after basic outlier removal? I assume after in this document.



Missing data and attrition

Due to noise in recordings and/or technical difficulties during data collection, a number of IRTs are missing for experimental items in the data. The following table shows which participants are missing how many IRTs; ideally each would have 48 IRTs and 16 experimental IRTs.

Table 1: Missing data, by participant

	Missing IRTs	Available % of IRTs	Missing experimental IRTs	Available % of experimental IRTs
1	0	100%	0	100%
2	0	100%	0	100%
3	0	100%	0	100%
4	0	100%	1	93.75%
5	23	52.08%	11	31.25%
6	0	100%	10	37.5%
7	0	100%	0	100%
8	1	97.92%	1	93.75%
9	0	100%	0	100%
10	0	100%	1	93.75%
11	12	75%	5	68.75%
12	0	100%	1	93.75%
13	0	100%	1	93.75%
14	0	100%	6	62.5%
15	0	100%	1	93.75%
16	0	100%	2	87.5%
17	0	100%	3	81.25%
19	0	100%	8	50%
20	0	100%	1	93.75%
21	0	100%	1	93.75%
22	0	100%	1	93.75%
201	0	100%	4	75%
203	0	100%	2	87.5%
204	0	100%	1	93.75%
205	0	100%	1	93.75%
206	0	100%	0	100%
207	0	100%	0	100%
208	0	100%	0	100%
209	0	100%	2	87.5%
210	0	100%	0	100%
212	0	100%	0	100%
214	0	100%	1	93.75%
215	0	100%	0	100%

The 12 participants missing more than 3 experimental IRTs (5, 6, 11, 14, 19, 201) are excluded.

Subjects with overall mean IRTs that are very short (< 2200) or very long (> 10000) are also excluded (20, 203, 204, 208)

Group sizes after attrition

The following table[3] shows how the participants are distributed across groupsafter attrition. Ideally, there would be 4 per group-order cell, but because of attrition the cells are uneven. Because regression is able to account for uneven groups, this defect will hopefully not play an important role in the analyses that follow.

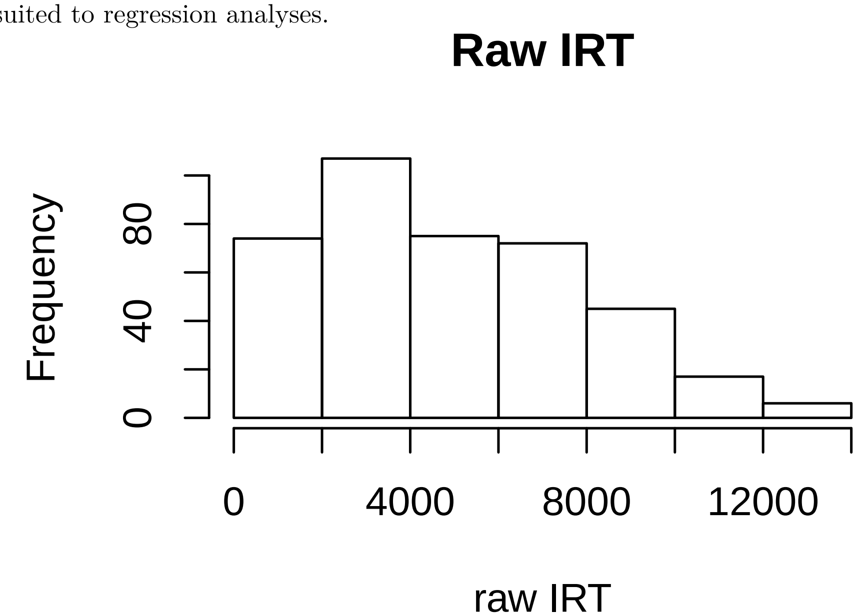
[3] There are 5 particpiants in Group 1, Split BA because I ran four participants per group-order, and then one extra who happened to be assigned to group 1, split BA; and by happenstance, none of the participants from that cell needed to be excluded.

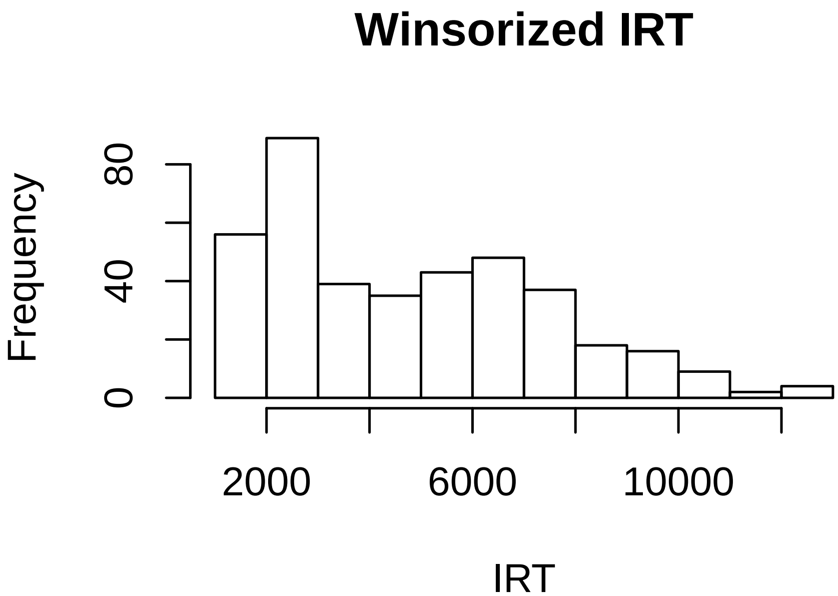
Table 2: Group/order totals after attrition

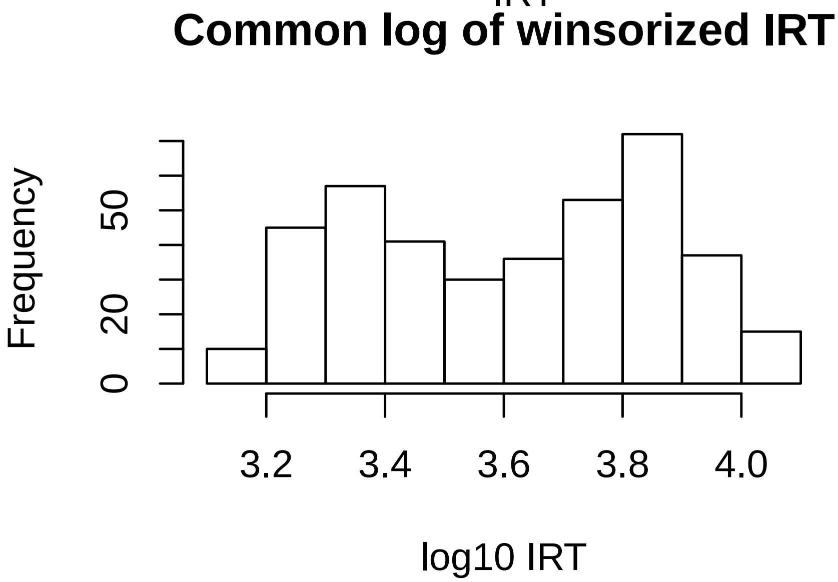
	Split AB	Split BA	Group Total
Group 1	3	5	8
Group 2	3	4	7
Group 3	2	3	5
Group 4	3	3	6
Split Total	11	15	26

Distribution of experimental item IRT after attrition

The following histograms show the distribution of experimental item IRTs after attrition, and then the Winsorized IRTs, and finally the common log of winsorized IRTs, which are the shape of the data most suited to regression analyses.







Mean and SD of winsorized IRT by condition

If we assume that interrogative PP-attachment garden paths are easier to process as an interrogative than in the declarative, and that IRT represents how difficult a sentence is to process, we would expect the difference in mean IRT to be larger for declarative garden paths compared to declarative controls than for the same comparison of interrogatives.

	Mean	SD

Table 3: Condition means

	Mean	SD
-Q -GP	4742.339	2586.327
-Q +GP	4873.071	2578.032
+Q -GP	4990.714	2625.889
+Q +GP	4685.307	2813.493

The means of the Winsorized IRT by condition indeed show this pattern.

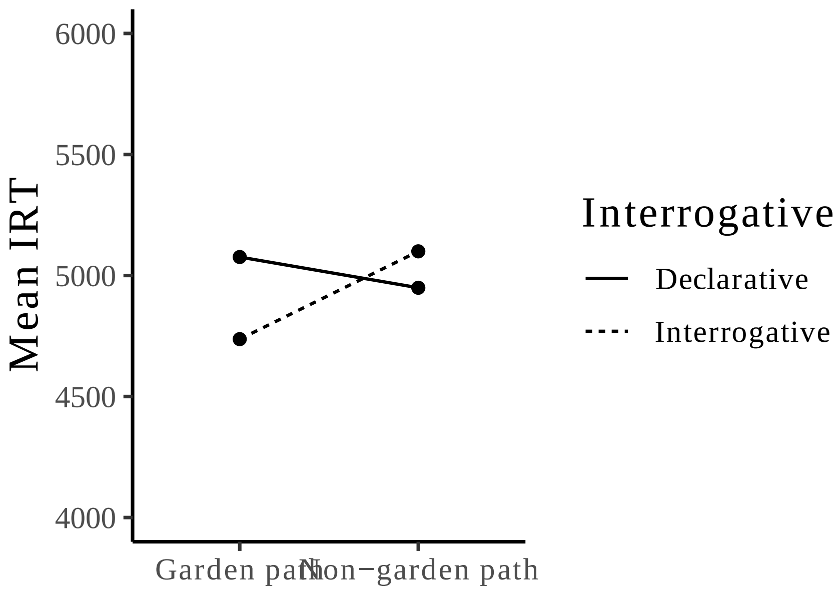


Figure 1: Mean experimental IRT by condition

The difference in mean IRT acriss ± for declaratives is 130.73; for interrogatives, it's -305.41. This is a difference of 436.14, representing the impact of ±GP for +Q compared to -Q. This supports the hypothesis that garden paths are easier to comprehend when presented as interrogative. It is strange that the garden-path interrogatives appear to be comprehended more quickly than the non-garden path interrogatives. A possible explanation will be explored in the discussion section.

Item and subject variation

There is variation across participants in terms of whether or not they show this pattern.

Number of participants who show predicted pattern

In the analyzed data,14 of 26 participants show the expected pattern.

Table 4: Mean IRT by condition and participant

participant	-Q -GP	-Q +GP	+Q -GP	+Q +GP	pattern
1	2321.508	2607.383	2713.935	2573.972	TRUE
2	5952.547	6220.012	5214.915	6914.200	TRUE
3	5940.587	6621.430	6401.395	5796.435	TRUE
4	7634.454	6360.074	8667.562	7112.926	FALSE
7	5338.530	5580.860	6119.302	6089.198	TRUE
8	6556.561	5071.637	3102.182	5963.673	FALSE
9	6167.350	6326.610	6236.807	6210.695	TRUE
10	5106.720	5072.834	6514.931	5575.137	FALSE
12	7334.830	7284.235	8722.731	8161.461	FALSE
13	9652.891	10771.226	9303.601	10676.079	TRUE
15	5060.193	3760.873	6510.255	4491.407	FALSE
16	2464.049	2154.396	2268.823	1987.429	FALSE
17	3097.555	6647.079	5259.298	3439.485	TRUE
20	10013.548	9315.809	10908.555	9796.060	FALSE
21	5108.422	5684.605	5783.965	6292.400	TRUE
203	1832.505	1741.404	1692.527	1952.706	FALSE
204	2052.124	1729.273	2021.573	1719.567	FALSE
205	5356.396	6593.677	5642.820	3400.438	TRUE
206	3715.318	3407.920	3291.235	3073.832	FALSE
207	4545.120	4991.815	4846.675	3648.440	TRUE
208	1788.898	2283.023	1930.420	2261.302	TRUE
209	2438.793	3100.746	2493.816	2637.569	TRUE
210	3897.847	4643.092	5679.913	2205.215	TRUE
212	3456.225	3264.745	3453.795	4052.240	FALSE
214	2835.318	2625.828	2466.315	1640.227	FALSE
215	3029.610	3712.920	3590.963	3689.343	TRUE

Number of items that show predicted pattern

For items, 7 of 16 show the pattern.

Table 5: Mean IRT by condition and item

item	-Q -GP	-Q +GP	+Q -GP	+Q +GP	pattern
1	4742.312	4042.900	4894.544	2858.503	FALSE
2	5865.177	4925.174	4731.989	4624.811	FALSE
3	5455.618	3718.398	5433.181	5385.808	FALSE

item	-Q -GP	-Q +GP	+Q -GP	+Q +GP	pattern
4	4340.826	5103.008	5578.320	4297.122	TRUE
5	4366.925	4301.081	5285.380	3391.629	FALSE
6	4291.746	4285.568	5260.358	5488.981	FALSE
7	3926.198	6590.852	5485.221	4936.701	TRUE
8	4869.698	5523.300	4784.719	5185.386	TRUE
9	4532.066	5741.277	4964.692	4428.341	TRUE
10	5162.286	4631.096	6339.241	3709.438	FALSE
11	3487.558	6099.569	4776.422	5533.677	TRUE
12	4948.887	3909.282	4606.605	3265.813	FALSE
13	5229.858	5554.131	3485.350	5376.064	TRUE
14	3627.347	4707.184	4978.228	4423.826	TRUE
15	5312.378	3912.743	4821.377	4974.233	FALSE
16	4946.281	4635.837	3963.138	6024.165	FALSE

Analyses

The following models explore the effect of garden path (±GP) and interrogativeness (±Q) on IRT.

Regression analyses

Regression models with fixed effects of ±GP and ±Q were run, one including the interaction of ±GP and ±Q and one without the interaction term. Both included random effects for item and participant.

Models with random slopes for GP, Q, and their interaction for both error terms fails to converge. A model with random slopes for just GP and Q ain effects likewise fails to converge. Models without random slopes of fixed effects were used.

The interaction model represents a better fit; the non-interaction model represents a singular fit that is worse overall ($X^2 = 4.695$, $p < 0.03$). This supports the hypothesis and the earlier observation over the means that garden paths are more difficult as declaratives than interrogatives.

The relevance of random effects were also tested, by comparing models that exclude each to the model with both random effects (I call this the "full model" in what follows).

Removing the random effect of item does not degrade the model in a stastically significant way (AIC~full model~ = -305; AIC~no item error~ = -307; -$X^2 = 0$, $p = 0.98$), but removing the random effect of participant does (AIC~full model~ = -305; AIC~no participant error~ = 47; $X^2 = 353.75$, $p = 0$). The model with no random effects is worse than both the full model and the model with only item re-

Table 6: Mixed effects model

	Dependent variable:	
	Common log of IRT	
	(1)	(2)
Garden path	0.019	−0.012
	(0.020)	(0.015)
Interrogative	0.028	−0.004
	(0.021)	(0.015)
Interaction	-0.063^{**}	
	(0.029)	
Constant	3.604^{***}	3.620^{***}
	(0.044)	(0.043)
Observations	396	396
Log Likelihood	159.327	156.980
Akaike Inf. Crit.	−304.655	−301.960
Bayesian Inf. Crit.	−276.785	−278.071
Note:	*p<0.1; **p<0.05; ***p<0.01	

moved. Ultimately, it's difficult to select between the full model and the "no item" model, as both offer strong fits with more or less the same outcome.

Breaks by condition

PP1 Break

Table 7: Both readings

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
No break	81	2	83	4
Break	181	262	181	255

Table 8: Reading 1

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
No break	39	1	40	1
Break	92	132	92	130

Table 9: Reading 2

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
No break	42	1	43	3
Break	89	130	89	125

OBJ Break

Table 10: Both readings

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
No break	52	96	69	86
Break	210	169	196	175

Table 11: Reading 1

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
No break	31	60	33	54

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
Break	100	73	99	77

Table 12: Reading 2

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
No break	21	36	36	32
Break	110	96	97	98

Breaks by ±Q

PP1 Break by ±Q

Table 13: Both readings

	-Q	+Q
No break	83	87
Break	443	436

Table 14: Reading 1

	-Q	+Q
No break	40	41
Break	224	222

Table 15: Reading 2

	-Q	+Q
No break	43	46
Break	219	214

OBJ Break by ±Q

Table 16: Both readings

	-Q	+Q
No break	148	155

	-Q	+Q
Break	379	371

Table 17: Reading 1

	-Q	+Q
No break	91	87
Break	173	176

Table 18: Reading 2

	-Q	+Q
No break	57	68
Break	206	195

Breaks by ±GP

PP1 Break by ±GP

Table 19: Both readings

	-GP	+GP
No break	164	6
Break	362	517

Table 20: Reading 1

	-GP	+GP
No break	79	2
Break	184	262

Table 21: Reading 2

	-GP	+GP
No break	85	4
Break	178	255

OBJ Break by ±GP

Table 22: Both readings

	-GP	+GP
No break	121	182
Break	406	344

Table 23: Reading 1

	-GP	+GP
No break	64	114
Break	199	150

Table 24: Reading 2

	-GP	+GP
No break	57	68
Break	207	194

3-level prosodic pattern

Table 25: Both readings

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
OBJ	70	3	74	6
OBJ > PP1	68	42	65	32
OBJ > V	7	0	3	0
OBJ > V PP1	4	6	2	2
OBJ PP1	12	12	7	7
PP1	44	79	55	79
PP1 > OBJ	44	102	40	122
PP1 > OBJ > V	0	0	0	1
PP1 > V	6	15	7	7
PP1 > V OBJ	2	5	5	4
V	0	0	3	0
V > OBJ	3	0	2	0
V > OBJ PP1	0	0	0	1
V > PP1	0	1	1	0

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
V PP1	1	0	0	0

Table 26: Reading 1

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
OBJ	36	1	36	1
OBJ > PP1	29	15	30	12
OBJ > V	1	0	2	0
OBJ > V PP1	2	2	1	1
OBJ PP1	5	8	4	2
PP1	27	50	27	51
PP1 > OBJ	23	47	25	60
PP1 > OBJ > V	0	0	0	0
PP1 > V	3	10	3	3
PP1 > V OBJ	2	0	1	1
V	0	0	0	0
V > OBJ	2	0	1	0
V > OBJ PP1	0	0	0	0
V > PP1	0	0	1	0
V PP1	1	0	0	0

Table 27: Reading 2

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
OBJ	34	2	38	5
OBJ > PP1	39	27	35	20
OBJ > V	6	0	1	0
OBJ > V PP1	2	4	1	1
OBJ PP1	7	4	3	5
PP1	17	29	28	28
PP1 > OBJ	21	55	15	62
PP1 > OBJ > V	0	0	0	1
PP1 > V	3	5	4	4
PP1 > V OBJ	0	5	4	3
V	0	0	3	0
V > OBJ	1	0	1	0
V > OBJ PP1	0	0	0	1
V > PP1	0	1	0	0
V PP1	0	0	0	0

2-level prosodic pattern

Table 28: Both readings

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
OBJ	80	3	79	6
OBJ > PP1	72	48	67	34
OBJ PP1	12	12	7	8
PP1	51	95	63	86
PP1 > OBJ	46	107	45	127

Table 29: Reading 1

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
OBJ	39	1	39	1
OBJ > PP1	31	17	31	13
OBJ PP1	5	8	4	2
PP1	31	60	31	54
PP1 > OBJ	25	47	26	61

Table 30: Reading 2

	-Q -GP	-Q +GP	+Q -GP	+Q +GP
OBJ	41	2	40	5
OBJ > PP1	41	31	36	21
OBJ PP1	7	4	3	6
PP1	20	35	32	32
PP1 > OBJ	21	60	19	66

PP1 or PP1 > OBJ

Table 31: Both readings

	PP1 Not Dominant	PP1 Dominant
-Q -GP	167	97
-Q +GP	63	202
+Q -GP	159	108
+Q +GP	49	213

Table 32: Reading 1

	PP1 Not Dominant	PP1 Dominant
-Q -GP	76	56
-Q +GP	26	107
+Q -GP	76	57
+Q +GP	16	115

Table 33: Reading 2

	PP1 Not Dominant	PP1 Dominant
-Q -GP	91	41
-Q +GP	37	95
+Q -GP	83	51
+Q +GP	33	98

OBJ or OBJ > PP1

Table 34: Both readings

	OBJ Not Dominant	OBJ Dominant
-Q -GP	112	152
-Q +GP	214	51
+Q -GP	121	146
+Q +GP	222	40

Table 35: Reading 1

	OBJ Not Dominant	OBJ Dominant
-Q -GP	62	70
-Q +GP	115	18
+Q -GP	63	70
+Q +GP	117	14

Table 36: Reading 2

	OBJ Not Dominant	OBJ Dominant
-Q -GP	50	82

	OBJ Not Dominant	OBJ Dominant
-Q +GP	99	33
+Q -GP	58	76
+Q +GP	105	26

Logistic regression models

The interaction between ±GP and ±Q approaches significance ($p < 0.06$) as a predictor of the object break, but not the PP1 break.

	Dependent variable:	
	OBJ	PP1
	(1)	(2)
Condition_GP	−1.043***	4.531***
	(0.223)	(0.733)
Condition_Q	−0.392*	−0.050
	(0.229)	(0.209)
Reading2	0.615***	−0.166
	(0.154)	(0.202)
Condition_GPTRUE:Condition_Q	0.582*	−0.679
	(0.307)	(0.894)
Constant	1.432***	1.108***
	(0.309)	(0.326)
Observations	1,053	1,049
Log Likelihood	−554.313	−328.071
Akaike Inf. Crit.	1,122.626	670.142
Bayesian Inf. Crit.	1,157.342	704.831
Note:	*p<0.1; **p<0.05; ***p<0.01	