Results and analyses

Tyler Peckenpaugh

April 7, 2019

Contents

This section reports various analyses of the recordings obtained. There is a total of 32 participants included, 48 total items, 16 of which are experimental items, with each item read twice by each participant. There were 4 versions of the experiment. Some items are missing due to technical issues during data collection. Other data were excluded due to factors explained later. Ultimately, 978 recordings are analyzed (489 pairs).

1 Prosody

This subsection presents the results of the independent rater's judgements regarding prosodic boundaries. Data are excluded if either recording from a pair of recordings (reading 1 or reading 2) is missing (there were 9 incomplete pairs excluded) because without the other member, it is difficult to determine the extent to which the participant followed instructions (i.e. did not study the sentence

before reading 1, did study the sentence before reading 2). Recordings with no reported breaks are excluded from the prosodic analyses, since they are unlikely to represent the reader's intended prosodic structure (there were 5 such recordings). The paired readings of these are not excluded, and IRTs extracted from them are included in the timing analyses, since delay information is available for both readings.

Table 1: N by condition

	D	Q
-GP	246	247
+GP	245	240

Table 1 shows the number of recordings included in analyses for each condition. If no items had been excluded, the expected number per bin would be 256[1].

The sentences can be informally broken into regions as in **??**.

(1) [$_{SUBJ}$ She] [$_{VERBS}$ had intended to set] [$_{OBJ}$ the clothes] $_{OBJ}$ [$_{PP1}$ in the hamper] $_{PP1}$ [$_{PP2}$ onto the dresser].

In what follows, I will discuss prosodic breaks at the points directly after the direct object (referred to as OBJ) and directly after PP1 (referred to as PP1).

1.1 Breaks after PP1 and direct object

The PP1 break is expected when PP2, attaches high; i.e., the garden path condition, and is not expected in the non-garden paths, when PP2 attaches low. The OBJ break does not correspond to any expected syntactic structure, and likely represents a breath group or some other not-syntactically-motivated break reason.

Table 2: Percentage of recordings containing PP1/OBJ break by condition

	PP1 Break		OBJ Break	
	D	Q	D	Q
+GP	168 (68.3%)	167 (67.6%)	199 (80.9%)	184 (74.5%)
-GP	243 (99.2%)	236 (98.3%)	160 (65.3%)	157 (65.4%)

Table 2 shows the presence of object and PP1 breaks by condition.

1.2 The two readings

Because "natural prosody" is likely to be less often produced in reading 1 than in a reading 2, it's important to consider the prosodic patterns for the two readings independently.

[1] There are 32 participants, 16 items, and 2 readings, divided across 4 bins.

The prevalence of object breaks and PP1 breaks for each condition in reading 2 recordings is described in table 3.

Table 3: Reading 2 recordings containing PP1/OBJ break by condition

	PP1 Break		OBJ Break	
	D	Q	D	Q
-GP	84 (68.3%)	83 (66.9%)	104 (84.6%)	90 (72.6%)
+GP	121 (99.2%)	117 (97.5%)	89 (73.0%)	89 (74.2%)

For comparison, 4 is the same table for reading 1 recordings.

Table 4: Reading 1 recordings containing PP1/OBJ break by condition

	PP1 Break		OBJ Break	
	D	Q	D	Q
-GP	84 (68.3%)	84 (68.3%)	95 (77.2%)	94 (76.4%)
+GP	122 (99.2%)	119 (99.2%)	71 (57.7%)	68 (56.7%)

1.3 PP1 and object breaks combined

Less than 1% of the data had neither break. Those data are excluded.

Table 5: Combined breaks by reading

Combined breaks	Percent of all data	Cold readings	Previewed readings
BOTH	55.6	51.0	60.2
OBJ	16.3	16.2	16.5
PP1	28.1	32.8	23.3

The distrbution of the combined breaks for both readings and for only previewed readings is shown in table 6.

1.4 PP1 and object breaks and their relative prominence

Table 7 presents data similar to those in the previous section, but it incorporates the rater's judgement of the relative prominence of the breaks. The > symbol indicates that the rater found the break on the left of that symbol to be stronger, or more prominent, than the break on the left. When no symbol is shown between the two breaks, the rather found them to be of equal prominence. Please be aware that inter-rater reliability for relative prominence was not good, as discussed in section **??**.

Table 6: Combined breaks by condition for previewed readings

	Both readings				Previewed readings			
	Declarative		Interrogative		Declarative		Interrogative	
	-GP	+GP	-GP	+GP	-GP	+GP	-GP	+GP
BOTH	49.8	64.5	44.4	63.7	54.1	72.1	43.0	71.7
OBJ	31.4	0.8	31.3	1.7	31.1	0.8	31.4	2.5
PP1	18.8	34.7	24.3	34.6	14.8	27.0	25.6	25.8

Table 7: Prosodic pattern by reading

Pattern	Percent of all data	Cold readings	Previewed readings
OBJ	16.2	16.2	16.3
OBJ > PP1	21.5	18.2	24.7
OBJ PP1	3.7	3.7	3.7
PP1	27.7	32.6	22.9
PP1 > OBJ	30.8	29.3	32.4

1.5 Prosody models

Models predicting the PP1 break and the object break are presented in table 8. No random slopes were included, as those models did not converge.

Table 8: Logistic regression models of prosody

	PP1	Object
Intercept	1.061 **	1.430 ***
	(0.329)	(0.310)
GP	4.526 ***	-0.958 ***
	(0.735)	(0.227)
Q	-0.045	-0.369
	(0.213)	(0.234)
GP:Q Interaction	-0.681	0.463
	(0.896)	(0.315)
Preview	-0.122	0.610 ***
	(0.206)	(0.157)
Participant	0.717	0.632
	(NA)	(NA)
Item	0.951	0.882
	(NA)	(NA)
N	1010	1010
logLik	-317.612	-530.263
AIC	649.224	1074.526

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

2 Fluency

This section presents the counts of recordings where the reader struggled to maintain fluency. Table 9 shows the overall pattern across reading.

Table 9: Difficulty in reading by reading type

	Cold reading	Previewed reading
Fluent	439	465
Struggle	50	24

Table 10 shows the pattern across reading for just +GP items.

Table 10: Difficulty in reading by reading type for garden paths

	Cold reading	Previewed reading
Fluent	215	229
Struggle	28	13

2.1 Fluency models

Table 11 shows a number of logistic regression models. Models with complex random slope structures failed to converge, and models with random slopes for just *Previewed Reading* were worse than those without random slopes, so the models presented have no random slopes for any predictors. None of the models differ in statistically significant ways, except that the model with no random effects is significantly worse than the others (e.g. no random effects (AIC=575) compared to full model (AIC=547), $X^2(3)=3$, $p < 0.001$).

Cold vs. previewed reading is a statistically significant predictor in every case.

Table 11: Logistic regression models of disfluency

	Full	No interaction	Reading only
Intercept	-2.872 ***	-2.858 ***	-2.516 ***
	(0.360)	(0.322)	(0.250)
+GP	0.281	0.256	
	(0.383)	(0.252)	
Q	0.410	0.386	
	(0.376)	(0.253)	
+GP Q	-0.044		
	(0.508)		
Previewed reading	-0.897 ***	-0.896 ***	-0.893 ***
	(0.266)	(0.266)	(0.265)
Participant	0.927	0.927	0.921
	(NA)	(NA)	(NA)
Item	0.094	0.094	0.113
	(NA)	(NA)	(NA)
N	1018	1018	1018
logLik	-248.638	-248.642	-250.327
AIC	511.276	509.284	508.655

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

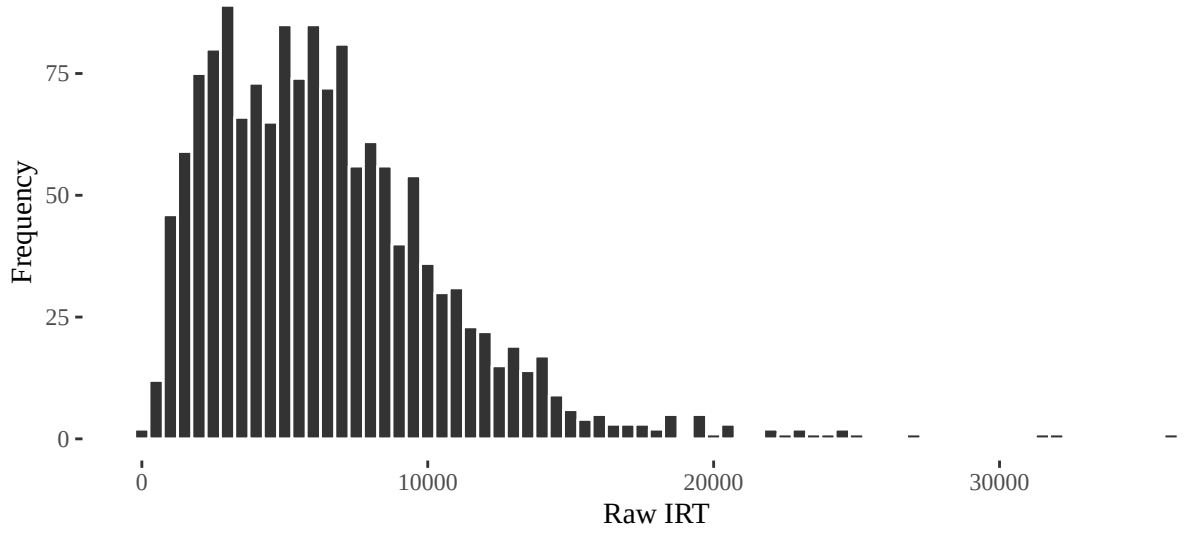


Figure 1: Distribution of raw IRT

3 Inter-reading time (IRT)

This document examines the inter-reading time (IRT) from the study. IRT was measured over 489 recordings: 32 participants, 48 items = 512 recording pairs (reading 1 and reading 2), with 23 missing pairs. The missing data are a result of one or both recordings from a pair being unusable due to technical issues (e.g. a failure of recording equipment, or participant error).

3.1 Distribution of IRT

The raw IRTs including fillers and before any outliers are trimmed are distributed as shown in Figure 1. Overall mean IRT of these data (n = 489), is 6.6s. The longest is 23.4s and the shortest 682ms. Median IRT is 6.1s.

IRTs below 250ms (2) and above 25s (5) are (assumed to be implausible) omitted. Experimental data were then Winsorized by participant to bring data in the 2.5th and 97.5th percentile of data to the value at those tresholds. The resulting measure is referred to as wIRT and is distribued as shown in Figure 2 (n = 489). Overall mean for wIRT is 6.5s. The longest IRT is 22.2s and the shortest is 709ms. Median wIRT is 6.1s.

For the purposes of regression analysis, a common log transformation reduces the skew in the data. This distribution is seen in Figure 3.

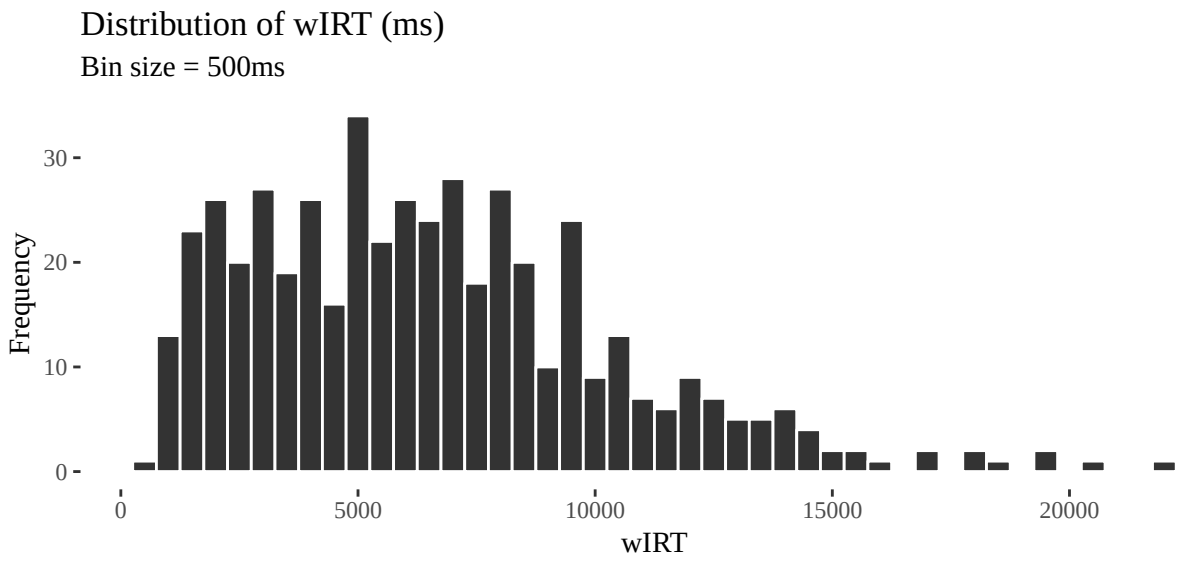


Figure 2: Distribution of wIRT

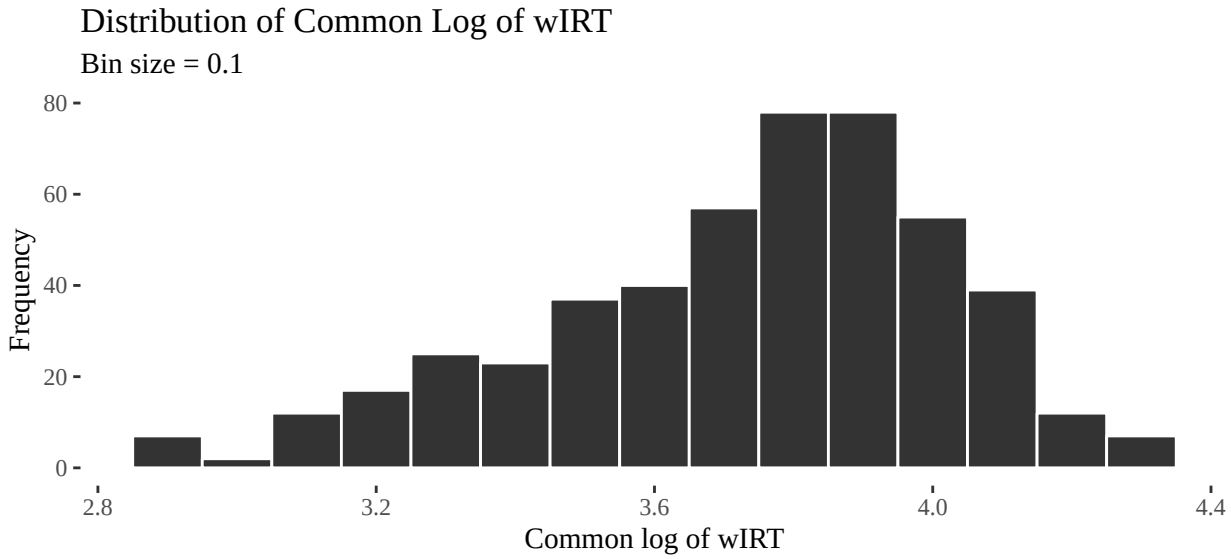


Figure 3: Common log of wIRT

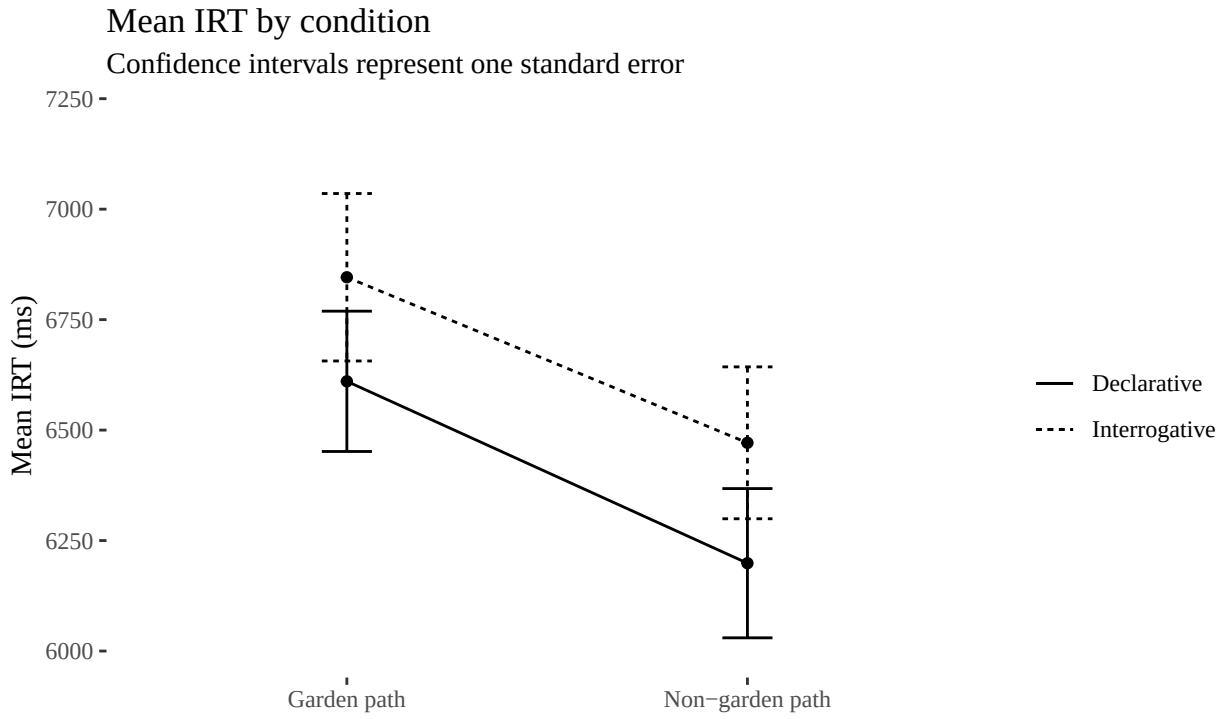


Figure 4: Mean IRT by condition

3.2 Means by condition

Table 12 shows the mean wIRT by experimental condition. The top left cell represents the mean wIRT for the declaritive controls ("D -GP"). The bottom row shows the increase in IRT across the garden path condition.

Table 12: Means (s) by condition

Condition	D	Q
-GP	6.20	6.47
+GP	6.61	6.85
Increase	0.41	0.37

The difference in the effect of ±GP across ±Q is 0.04s. That is, the mean amount that IRT increased for a garden path declarative compared to a non-garden path declarative is 0.04s more than the amount that IRT increased for a garden path interrogative compared to a non-garden path interrogative.

3.3 Regression models of IRT

The models with random slopes for participant and item did not converge, so the tables in this section show models with no random slopes.

For the first model, fixed effects of ±GP and ±Q as well is the interaction between them were included, along with random effects of participant and item. The second model removes the interaction, but keeps both main effects.

Table 13: Models

	Full	No interaction	No random effects	No fixed effects
+GP	0.056	0.041	0.057	
	(0.023)	(0.016)	(0.037)	
Q	0.030	0.016	0.030	
	(0.023)	(0.016)	(0.037)	
+GP Q	-0.029		-0.036	
	(0.033)		(0.053)	
N	496	496	496	496
logLik	89.853	89.449	-95.352	85.816
AIC	-165.706	-166.898	200.704	-163.632

A model with no fixed effects and one with no random effects were also run. The estimates from these models can be seen in table 13.

Table 14: Full model vs. no fixed effects

	Df	AIC	BIC	logLik	deviance	χ^2	Df(χ)	Pr(>χ^2)
No fixed effects	4	-163.6318	-146.8055	85.81592	-171.6318	NA	NA	NA
Full	7	-165.7062	-136.2602	89.85311	-179.7062	8.07438	3	0.0444993

Table 15: Full model vs. no interaction

	Df	AIC	BIC	logLik	deviance	χ^2	Df(χ)	Pr(>χ^2)
No interaction	6	-166.8984	-141.6589	89.44919	-178.8984	NA	NA	NA
Full	7	-165.7062	-136.2602	89.85311	-179.7062	0.8078459	1	0.3687579

Several model comparisons were made, seen in tables 14, 15, and 16.

There are clearly main effects of both ±GP and ±Q: table 14 shows that the full model is significantly better than the one with no fixed effects.

The interaction between main effects, though, is not able to be confirmed (in fact, table 15 shows the non-interaction to be better, but not to a stastically significance degree).

Table 16 shows that the random effects of participant and item improve the model in a stastically signifcant way.

Table 16: Full model vs. no random effects

	Df	AIC	BIC	logLik	deviance	χ^2	Df(χ)	Pr(>χ^2)
No random effects	5	200.7040	221.7368	-95.35198	190.7040	NA	NA	NA
Full	7	-165.7062	-136.2602	89.85311	-179.7062	370.4102	2	0

3.4 Effect of verb on IRT

If we consider the mean wIRT by which of the 4 verbs occured in the target sentence (*cram*, *set*, *put*, or *stick*), we see that there was some difference in how the two experimental manipulations (±GP = garden path status and ±Q = interrogative vs. declarative status) effected wIRT. The means (and standard deviation) across the conditions are reported for each verb in table 17.

Table 17: Mean (sd) wIRT by condition and verb in seconds

Verb	D -GP	D +GP	Q -GP	Q +GP
cram	6.4 (3.9)	6.9 (4)	5.9 (3.1)	6.5 (3.8)
put	6.5 (3.8)	6.8 (3.7)	6 (3.3)	7 (5.2)
set	6.1 (3.2)	6.3 (3.2)	7.4 (4.5)	6.6 (3.6)
stick	5.8 (4.2)	6.5 (3.2)	6.7 (4.1)	7.4 (4.2)

To isolate the effect of the garden path for interrogatives as compared to declartives, we can subtract the mean for declarative non-garden paths from the mean for declarative garden paths, and then do the same for interrogatives. These measures are referred to as the "declarative GP effect" and "interrogative GP effect" in **??**. The difference between the declarative GP effect from the interrogative GP effect is labeled the "difference in effect." We can see that for 2 of the 4 verbs (*set* and *stick*), there was a positive difference in effect, while for the other two we find a negative one.

Table 18: Effect of GP on wIRT by verb

Verb	Declarative GP effect	Interrogative GP effect	Difference in effect
cram	0.5	0.6	-0.1
put	0.3	1.0	-0.8
set	0.2	-0.9	1.0
stick	0.7	0.7	0.0

3.5 Individual variation in IRT

Individuals vary with regard to the effect of the garden path condition on IRT. For 17 of 32, the increase in IRT for garden paths is greater for interrogatives than it is for declaratives.

Table 19: Mean wIRT (ms) by condition and participant

Participant	D -GP	D +GP	Q -GP	Q +GP
10	9072.500	9416.155	8734.438	9630.157
5	13082.620	10694.565	9437.685	12188.125
1	2723.095	2764.875	3165.468	3061.438
2	6077.595	7088.435	4410.845	6398.435
3	5655.472	8400.845	6056.405	7267.347

Table 19: Mean wIRT (ms) by condition and participant *(continued)*

Participant	D -GP	D +GP	Q -GP	Q +GP
4	7993.717	5457.905	10609.750	7010.720
6	12848.093	10828.407	13018.030	14695.872
7	5706.287	5864.283	6291.778	4918.595
9	6234.815	6632.685	6386.938	6484.312
11	8635.833	9030.440	5688.040	12075.837
12	5436.092	5236.717	6146.935	7347.062
13	10779.875	11638.595	9410.875	10931.093
14	7180.315	8851.722	8443.062	12052.435
15	9673.310	7771.158	9463.500	5764.438
16	12583.907	9725.440	9352.470	15201.440
17	3725.878	6283.250	5736.560	5128.127
19	10846.250	13891.583	14803.670	NA
20	9963.565	9211.125	12225.282	10081.030
21	5035.065	5748.627	6117.688	7204.748
22	5066.752	4134.533	5006.845	7975.283
201	1396.033	4249.562	2533.628	2206.812
203	1451.470	1501.438	1909.753	2472.343
204	3164.688	3173.847	2792.125	1676.000
205	6456.967	11925.468	7450.278	11345.407
206	3736.810	3707.875	3223.593	4985.440
207	5083.280	6428.315	6700.748	5199.312
208	2612.250	3449.810	2482.940	3263.970
209	2126.185	3625.500	2563.718	2967.907
210	5690.440	6105.592	6664.688	5098.160
212	4699.658	5192.250	4647.190	5065.035
214	3628.378	4118.707	5185.467	2026.030
215	4432.130	4132.312	5207.998	5075.938

3.6 Interrogative processing cost is represented in IRT

Interrogatives appear to have a computational processing cost when compared to interrogatives (cf. Lehiste, 1973). The filler sentences in this study were designed so as to provide a diagnostic of the interrogative effect on IRT.

	Only Interogitivity	Full	No interaction
(Intercept)	6236.06	6261.55	6226.62
	(554.38)	(573.11)	(565.57)
Q	379.21	311.62	379.40
	(155.31)	(236.77)	(155.32)
+GP		-49.00	18.69
		(284.55)	(221.71)
Q * +GP		135.53	
		(357.36)	
N	998	998	998
logLik	-9278.27	-9278.20	-9278.27
AIC	18566.54	18570.39	18568.54

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

The mean IRT for interrogative fillers was 6.2s; for declarative, 6.2s. Interrogatives elicited a mean IRT of 372ms longer than declaratives. Half of the fillers had a string of two PPs at the end. This did not impact IRT to the same degree (a difference of 2ms across ±PP).

A mixed effects regression model (no random slope structure, due to convergence errors) found a significant main effect of ±Q (estimate = 379ms, t = 2.24, $p < 0.05$)

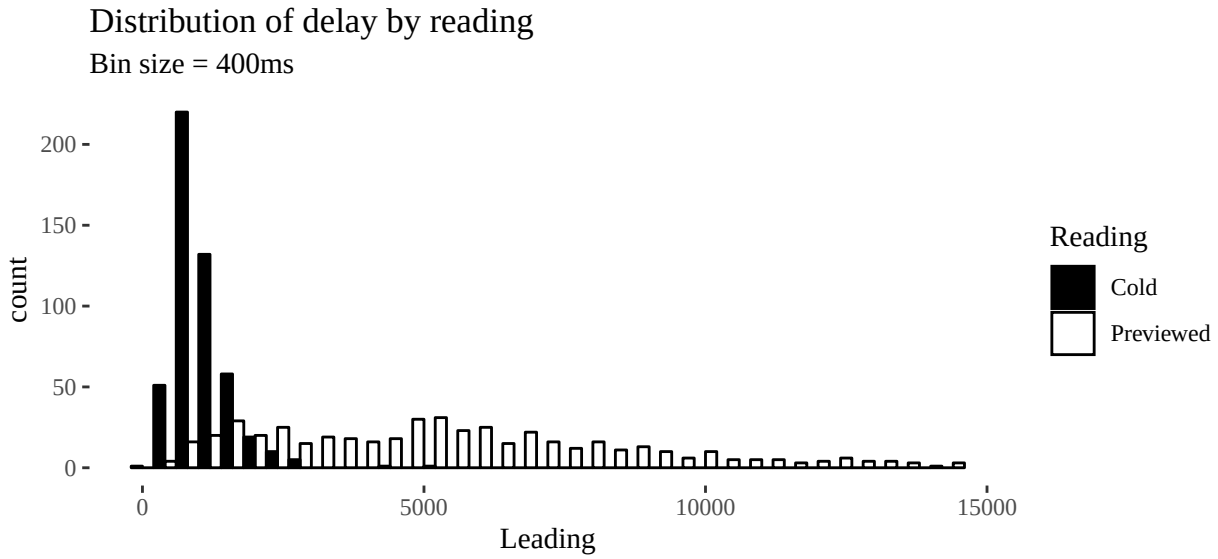


Figure 5: Delay comparison across reading

4 Delay comparison for cold vs. previewed readings

A comparison of the delay for cold readings compared with that of previewed readings can lend insight into the extent to which subjects followed task instructions.

"Delay" here is the amount of time after the start of a recording until the beginning of phonation of the target sentence. Cold readings are also called "reading 1", while previewed readings are the same as "reading 2". Implausible delays of >15s are excluded in the data shown here.

For cold readings, n = 498 and for previewed, n = 483.

4.1 Reading 1 delays by participant

```
## Linear mixed model fit by maximum likelihood . t-tests use
##   Satterthwaite's method [lmerModLmerTest]
## Formula: wirt.log10 ~ Condition_GP + (1 | Participant) + (1 | Item)
##    Data: irt_data
##
##      AIC      BIC   logLik deviance df.resid
##   -175.2   -154.2     92.6   -185.2      484
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7052 -0.4813  0.0324  0.5571  3.0934
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  Participant (Intercept) 5.408e-02 0.232543
```

```
##  Item        (Intercept) 8.269e-05 0.009093
##  Residual                3.230e-02 0.179710
## Number of obs: 489, groups:  Participant, 32; Item, 16
##
## Fixed effects:
##                   Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)        3.72166    0.04274  34.32751   87.07   <2e-16 ***
## Condition_GPTRUE   0.03678    0.01627 443.40043    2.26   0.0243 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## Cndt_GPTRUE -0.187
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: PP1 ~ Condition_Q * Condition_GP * r1DelCat + (1 | IID) + (1 |
##     SID)
##    Data: mmdata
##
##      AIC      BIC   logLik deviance df.resid
##    178.8    214.8    -79.4    158.8      260
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.9948  0.0000  0.1210  0.3451  1.0628
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  SID    (Intercept) 0.4388   0.6625
##  IID    (Intercept) 0.9424   0.9708
## Number of obs: 270, groups:  SID, 17; IID, 16
##
## Fixed effects:
##                                                Estimate Std. Error z value
## (Intercept)                                      0.8724     0.4845   1.801
## Condition_QTRUE                                  1.2096     0.6047   2.000
## Condition_GPTRUE                                19.9629  4654.3443   0.004
## r1DelCatSLOW                                     1.0182     0.7361   1.383
## Condition_QTRUE:Condition_GPTRUE                -17.7420  4654.3444  -0.004
## Condition_QTRUE:r1DelCatSLOW                    -2.2289     0.9341  -2.386
## Condition_GPTRUE:r1DelCatSLOW                   -0.6049  7785.8297   0.000
## Condition_QTRUE:Condition_GPTRUE:r1DelCatSLOW    1.3990  7785.8299   0.000
```

```
##                                                Pr(>|z|)
## (Intercept)                                      0.0717 .
## Condition_QTRUE                                  0.0455 *
## Condition_GPTRUE                                 0.9966
## r1DelCatSLOW                                     0.1666
## Condition_QTRUE:Condition_GPTRUE                 0.9970
## Condition_QTRUE:r1DelCatSLOW                     0.0170 *
## Condition_GPTRUE:r1DelCatSLOW                    0.9999
## Condition_QTRUE:Condition_GPTRUE:r1DelCatSLOW   0.9999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##                    (Intr) Cn_QTRUE Cn_GPTRUE r1DCSL Cn_QTRUE:C_GPTRUE
## Cndtn_QTRUE        -0.449
## Cndt_GPTRUE         0.000  0.000
## r1DelCtSLOW        -0.489  0.305    0.000
## Cn_QTRUE:C_GPTRUE   0.000  0.000   -1.000     0.000
## C_QTRUE:1DC         0.301 -0.664    0.000    -0.645  0.000
## C_GPTRUE:1D         0.000  0.000   -0.598     0.000  0.598
## C_QTRUE:C_GPTRUE:   0.000  0.000    0.598     0.000 -0.598
##                    C_QTRUE:1 C_GPTRUE:
## Cndtn_QTRUE
## Cndt_GPTRUE
## r1DelCtSLOW
## Cn_QTRUE:C_GPTRUE
## C_QTRUE:1DC
## C_GPTRUE:1D         0.000
## C_QTRUE:C_GPTRUE:   0.000    -1.000
## convergence code: 0
## unable to evaluate scaled gradient
##  Hessian is numerically singular: parameters are not uniquely determined
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: OBJ ~ Condition_Q * Condition_GP * r1DelCat + (1 | IID) + (1 |
##     SID)
##    Data: mmdata
##
##      AIC      BIC   logLik deviance df.resid
##    282.7    318.7   -131.4    262.7      260
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -6.2475  0.1443  0.3474  0.4847  1.4098
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  SID    (Intercept) 0.9216   0.9600
##  IID    (Intercept) 0.6754   0.8218
## Number of obs: 270, groups:  SID, 17; IID, 16
##
## Fixed effects:
##                                              Estimate Std. Error z value
## (Intercept)                                    1.5923     0.5649   2.819
## Condition_QTRUE                               -0.1312     0.5795  -0.226
## Condition_GPTRUE                              -0.7720     0.5666  -1.362
## r1DelCatSLOW                                   1.6449     1.0003   1.644
## Condition_QTRUE:Condition_GPTRUE                0.7165     0.8151   0.879
## Condition_QTRUE:r1DelCatSLOW                  -1.4460     1.0910  -1.325
## Condition_GPTRUE:r1DelCatSLOW                 -0.9320     1.0802  -0.863
## Condition_QTRUE:Condition_GPTRUE:r1DelCatSLOW   1.2526     1.4447   0.867
##                                              Pr(>|z|)
## (Intercept)                                   0.00482 **
## Condition_QTRUE                               0.82092
## Condition_GPTRUE                              0.17307
## r1DelCatSLOW                                  0.10009
## Condition_QTRUE:Condition_GPTRUE               0.37936
## Condition_QTRUE:r1DelCatSLOW                  0.18503
## Condition_GPTRUE:r1DelCatSLOW                 0.38823
## Condition_QTRUE:Condition_GPTRUE:r1DelCatSLOW  0.38592
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##                    (Intr) Cn_QTRUE Cn_GPTRUE r1DCSL Cn_QTRUE:C_GPTRUE
## Cndtn_QTRUE        -0.530
## Cndt_GPTRUE        -0.556  0.535
## r1DelCtSLOW        -0.455  0.303    0.296
## Cn_QTRUE:C_GPTRUE   0.395 -0.721   -0.709    -0.213
## C_QTRUE:1DC         0.267 -0.539   -0.281    -0.706  0.387
## C_GPTRUE:1D         0.275 -0.290   -0.518    -0.710  0.373
## C_QTRUE:C_GPTRUE: -0.212  0.420    0.401     0.537 -0.575
##                    C_QTRUE:1 C_GPTRUE:
## Cndtn_QTRUE
## Cndt_GPTRUE
## r1DelCtSLOW
## Cn_QTRUE:C_GPTRUE
## C_QTRUE:1DC
```

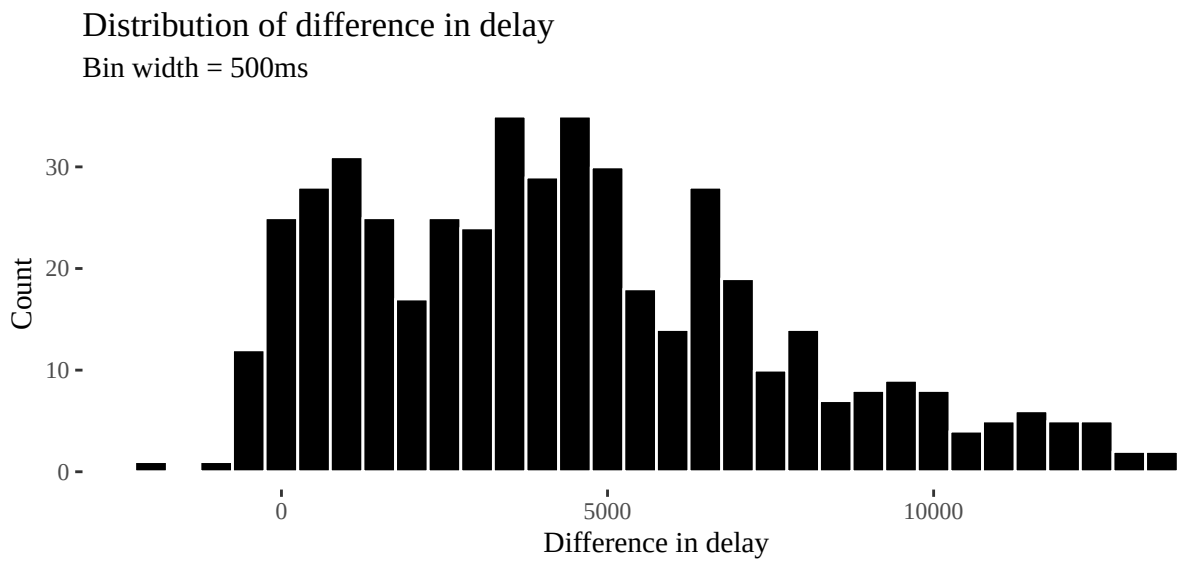


Figure 6: Distribution of DelDif

```
## C_GPTRUE:1D         0.656
## C_QTRUE:C_GPTRUE: -0.766    -0.760
```

4.2 Difference in delay across paired readings

Overall, each recording pair (n = 482) has a mean difference in delay (DelDif = previewed delay - cold delay) of 4.4s (sd = 3.2s), with a minumum of -1.9s and a max of 13.4s. The median DelDif is 4.1s. The distribution DelDif is shown in Figure 6.

If we calculate the mean delay difference by participant, we find a mean participant DelDef of 4.5s. Each participant's DelDif is ≤ 386ms and ≥ 10.2s, with a median of 3.98s. Table 20 shows these values.

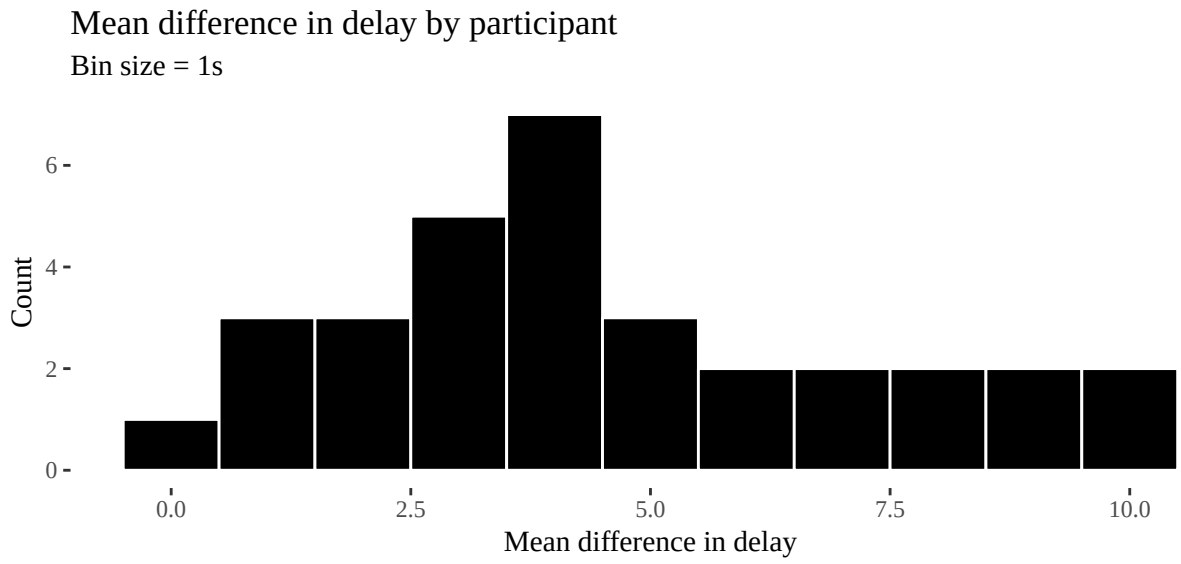
Table 20: Delay differences by participant

Participant	Mean difference in delay (ms)
1	386
203	938
208	1183
209	1290
204	1502
201	1584
214	1604
206	2582
215	2814
21	2908
17	3139

Table 20: Delay differences by participant *(continued)*

Participant	Mean difference in delay (ms)
212	3167
7	3682
9	3816
210	3908
22	3973
207	3988
12	4303
2	4369
4	4798
3	4841
11	4845
15	5961
10	6358
205	6858
14	7157
5	7838
13	8001
20	8685
16	9105
19	9537
6	10184

The distribution of the participants' DelDifs can be found in Figure **??**.



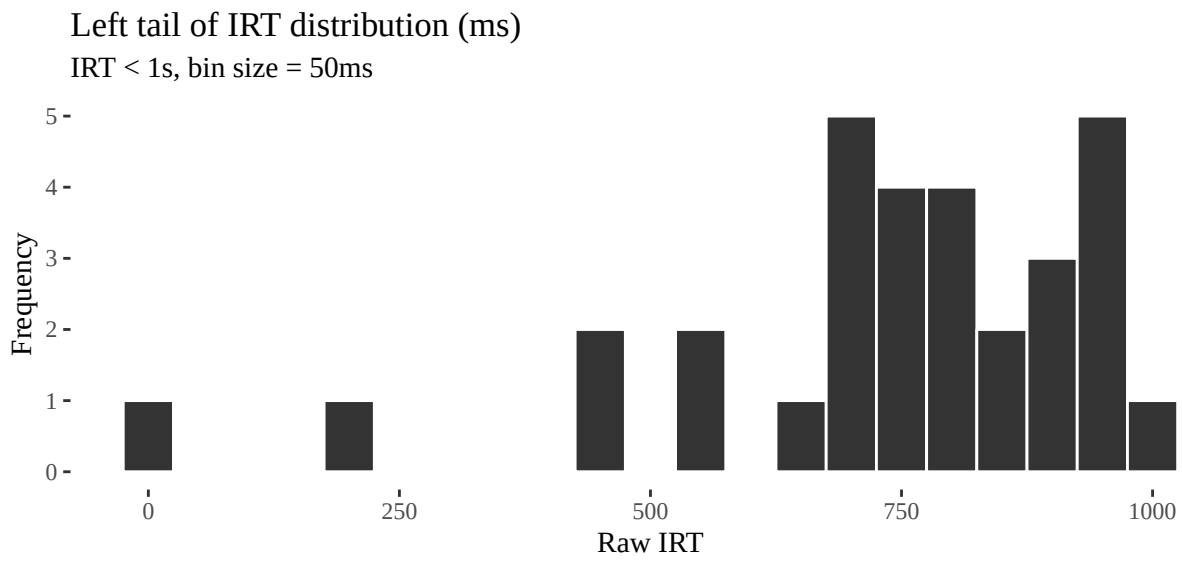


Figure 7: Left tail of raw IRT distribution

5 Additional tables and figures

Addtional figures and tables appear here without discussion, for the curious reader.

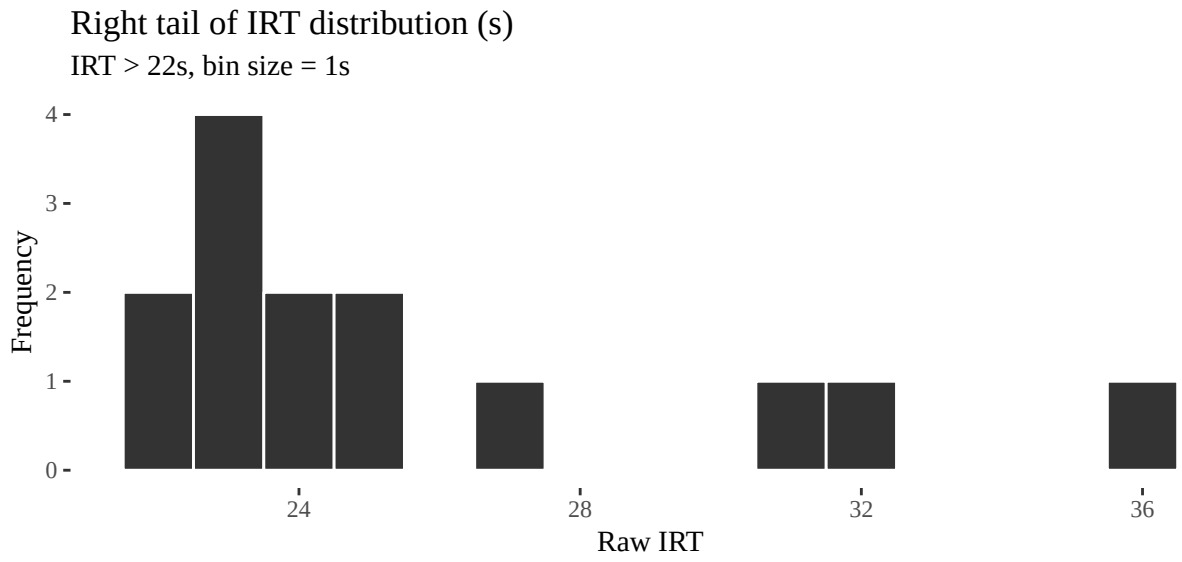


Figure 8: Right tail of raw IRT distribution

References

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, 53(1):380–380.