Supplementary Material

Abstract Linguistic Structure Correlates with Temporal Activity during Naturalistic Comprehension

J. Brennan, E. P. Stabler, S. Van Wagenen, W.-M. Luh, & J. T. Hale

S.1. Modeling and ROI details

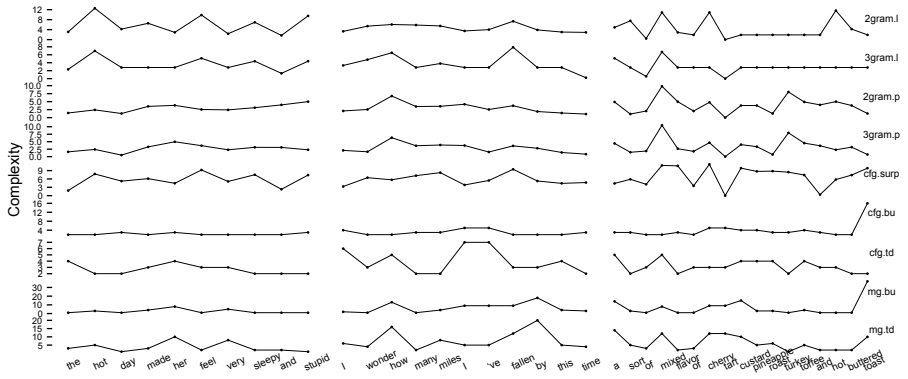

Figure S1: Example complexity counts from the nine syntactic models for three sentences. To link these values with the BOLD signal, they are time-aligned with the auditory offset of each word and convolved with the hemodynamic response function to yield estimates of the BOLD signal, as illustrated in Figure 2.

	LATL		RATL		LIFG		LPTL		LIPL		LPreM	
ID#	*x*, *y*, *z*	*t*	*x*, *y*, *z*	*t*	*x*, *y*, *z*	*t*	*x*, *y*, *z*	*t*	*x*, *y*, *z*	*t*	*x*, *y*, *z*	*t*
18	-42, 18, -30	2.30	48, 16, -26	2.49	-52, 4, 12	1.97	-56, -40, -4	2.32	-46, -68, 40	3.82	-38, 22, 54	3.00
22	-52, 6, -40	3.84	54, 18, -28	3.15	-58, 14, 20	3.07	-62, -34, 2	6.40	-52, -60, 12	5.30	-46, 0, 56	3.84
23	-58, -14,-16	2.87	58, -6, -32	2.91	-54, 10, 26	4.30	-52, -44, -2	4.12	-46, -60, 36	3.97	-42, 8, 54	3.44
24	-42, 22, -28	2.15	62, 0, -6	4.14	-50, 16, 8	1.76	-62, -38, 2	5.61	-54, -62, 14	4.55	-52, -4, 42	3.05
26	-54, 10, -24	3.30	50, 20, -24	3.93	-56, 16, 30	3.01	-48, -38, 2	4.28	-58, -48, 40	2.87	-48, 6, 54	4.05
28	-38, 10, -32	3.00	58, -2, -34	3.1 0	-60, 10, 6	2.88	-66, -34, 6	4.50	-46, -58, 30	3.70	-44, 0, 58	2.42
30	-50, 16, -34	4.22	48, 14, -36	3.46	-50, 14, 30	2.72	-66, -26, -2	4.60	-54, -44, 10	3.34	-36, 2, 64	3.30
31	-54, 8, -24	4.36	52, 10, -32	3.51	-56, 16, 2	2.18	-68, -20, -6	4.52	-46, -64, 24	3.20	-54, 6, 44	3.75
33	No ROI activation peaks met inclusion criteria											
35	-50, 4, -22	4.92	48, 8, -26	3.92	-52, 16, 10	3.57	-50, -32, 0	4.85	-62, -52, 12	4.99	-42, 2, 50	5.46
36	-54, 12, -16	4.44	50, 14, -42	2.17	-56, 10, 26	2.50	-66, -22, 0	5.36	-48, -58, 20	3.67	-38, -2, 48	3.47
37	-48, 0, -48	3.68	44, 6, -42	2.19	-56, 14, 28	3.38	-68, -32, 0	4.69	-54, -56, 24	4.83	-48, 6, 56	3.87
38	-58, 0, -34	4.20	52, -2, -48	2.38	-50, 16, 12	3.74	-66, -44, -8	6.35	-42, -58, 18	5.91	-44, 8, 44	4.07
39	-48, -2, -40	1.85	64, -12,-20	2.04	-54, 16, 24	1.91	-44, -42, 2	2.52	-38, -60, 16	3.70	-40, 0, 46	3.63
40	-46, 18, -28	3.06	52, 14, -32	0.06	-58, 10, 20	2.18	-54, -40, -2	3.71	-42, -64, 18	2.52	-40, -6, 44	1.55
41	-48, 12, -42	3.71	58, 10, -34	3.52	-56, 24, 22	2.43	-48, -28, -2	3.59	-36, -58, 18	2.3 0	-42, 6, 60	2.93
42	-54, 2, -34	2.32	62, 8, -28	1.77	-44, 26, 24	1.37	-46, -38, 2	2.75	-42, -54, 20	2.79	-50, 8, 50	2.75
43	-44, 14, -36	4.07	48, 0, -46	2.20	-54, 28, 4	2.62	-48, -34, -4	3.20	-46, -50, 24	4.04	-40, 22, 40	2.99
44	-52, 12, -26	3.32	50, 18, -20	4.05	-58, 26, 10	3.19	-66, -34, -8	5.21	-40, -56, 16	3.39	-40, 18, 58	2.25
45	-48, 0, -40	2.74	62, 2, -28	3.36	-48, 10, 8	2.74	-62, -42, 8	4.82	-52, -68, 20	4.66	-34, 10, 48	4.39
46	Excluded due to sub-threshold behavioral performance											
47	-42, -2, -38	3.00	54, 8, -36	2.27	-54, 18, 4	2.03	-62, -36, 2	5.39	-48, -54, 30	3.75	-42, 16, 48	3.28
48	-50, 16, -18	4.17	48, 4, -42	3.46	-56, 18, 16	3.72	-48, -38, -4	4.55	-40, -60, 20	5.03	-34, 6, 40	3.67
49	-42, 14, -24	0.91	46, 20, -26	1.08	-58, 18, 26	2.32	-46, -34, 8	1.96	-42, -56, 14	1.82	-38, 6, 58	2.57
50	-64, -2, -8	2.86	62, 2, -8	3.25	-56, 32, 8	3.31	-48, -28, -6	3.21	-46, -58, 18	3.98	-34, 6, 44	3.25
51	-48, 4, -34	3.40	48, 0, -34	3.41	-52, 18, 16	2.20	-60, -38, 2	5.42	-42, -60, 16	5.17	-42, 10, 52	5.21
52	Excluded due to sub-threshold behavioral performance											
53	-42, 12, -44	2.26	46, -8, -36	2.92	-54, 10, 8	2.01	-60, -38, -6	3.40	-50, -56, 34	3.20	-42, 12, 58	4.15
54	Excluded due to supra-threshold head movement											

Table S1: MNI coordinates and activation *t* statistic for the center of each of six regions of interest from all participants.

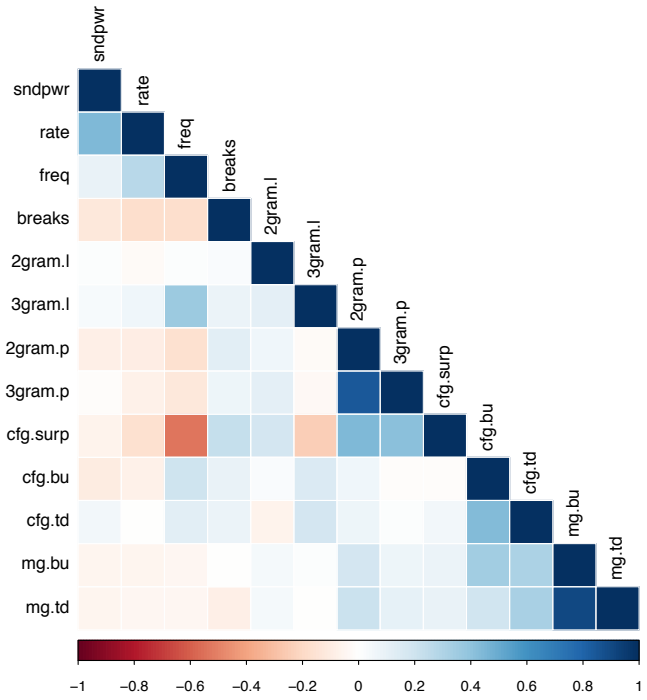

Figure S2: Correlation matrix showing Pearson's *r*-values between each predictor entered into the statistical models. "Breaks" describes a predictor based on prosodic annotations. "Freq" is a predictor based on word-frequency estimates from the HAL corpus, log transformed. "Rate" tracks the speech rate using a uniform impulse at the offset of each word.

S.2. *Full regression results and model comparison tables*

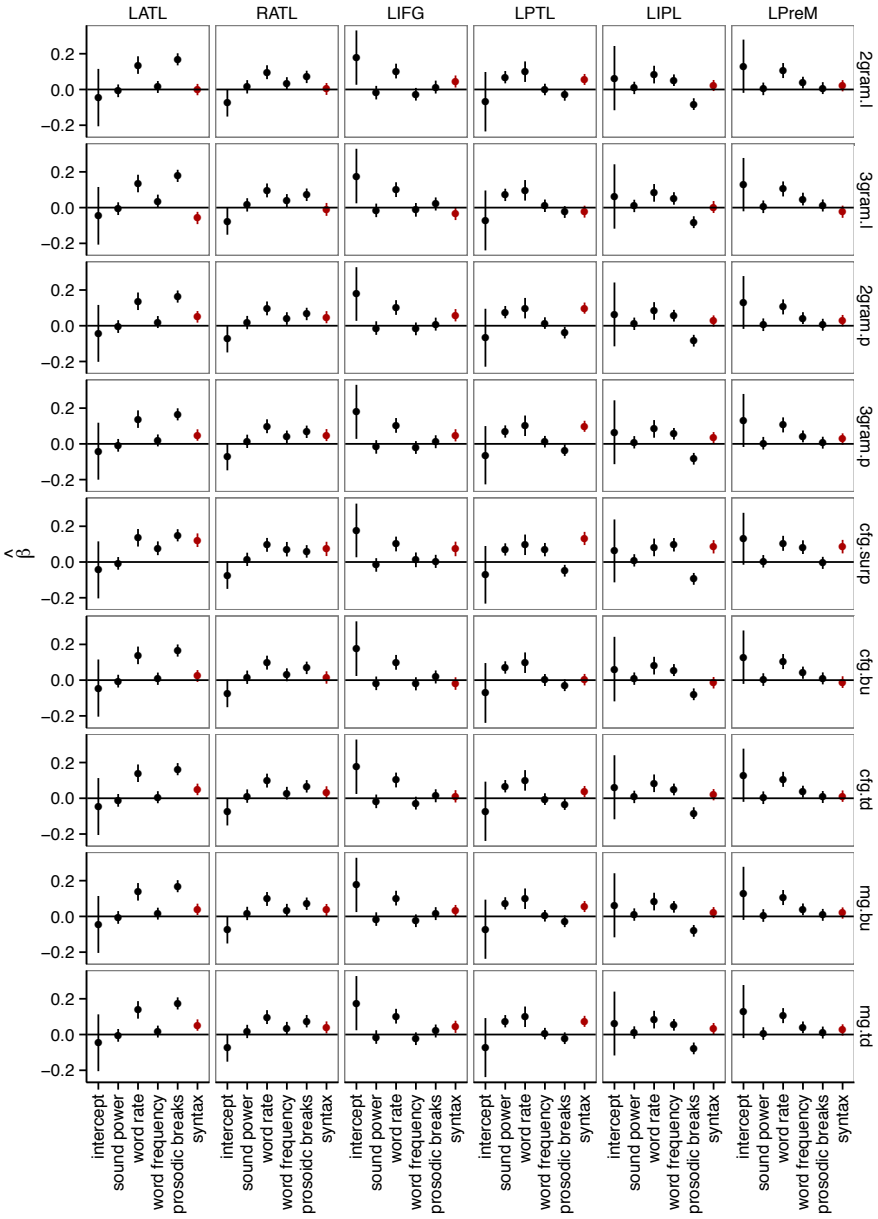

Figure S3: Fitted coefficients and 95% CIs for all linguistic predictors from nine models (rows) fitted against six ROIs (columns). Coefficients for head-movement parameters are excluded. Red points indicate the "Syntax" predictor which varied for each model.

LATL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-11549.51		
A	*2gram.l*	16	-11549.51	0.00	0.9632
B	*3gram.l*	17	-11536.63	25.77	<**0.001**
C	*2gram.p*	18	-11525.62	22.02	<**0.001**
D	*3gram.p*	19	-11524.88	1.48	0.2245
E	*cfg.surp*	20	-11493.29	63.18	<**0.001**
F	*cfg.bu*	21	-11492.14	2.29	0.1304
G	*cfg.td*	22	-11485.57	13.16	<**0.001**
H	*mg.bu*	23	-11483.02	5.10	0.0239
I	*mg.td*	24	-11479.47	7.09	**0.0077**

Number of timecourses: 23

RATL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-11217.40		
A	*2gram.l*	16	-11217.36	0.07	0.7941
B	*3gram.l*	17	-11216.98	0.75	0.3851
C	*2gram.p*	18	-11208.04	17.88	<**0.001**
D	*3gram.p*	19	-11206.91	2.27	0.1319
E	*cfg.surp*	20	-11199.04	15.74	<**0.001**
F	*cfg.bu*	21	-11198.70	0.68	0.4108
G	*cfg.td*	22	-11196.42	4.57	0.0326
H	*mg.bu*	23	-11193.75	5.33	0.0209
I	*mg.td*	24	-11193.57	0.37	0.5432

Number of timecourses: 22

LIFG

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-10664.70		
A	*2gram.l*	16	-10656.88	15.63	<**0.001**
B	*3gram.l*	17	-10651.79	10.18	**0.0014**
C	*2gram.p*	18	-10639.76	24.06	<**0.001**
D	*3gram.p*	19	-10639.55	0.43	0.5122
E	*cfg.surp*	20	-10636.61	5.88	0.0153
F	*cfg.bu*	21	-10634.13	4.95	0.0260
G	*cfg.td*	22	-10632.27	3.73	0.0535
H	*mg.bu*	23	-10629.85	4.84	0.0278
I	*mg.td*	24	-10628.73	2.23	0.1356

Number of timecourses: 21

LPTL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-11978.19		
A	*2gram.l*	16	-11963.26	29.86	<**0.001**
B	*3gram.l*	17	-11959.65	7.23	**0.0072**
C	*2gram.p*	18	-11918.04	83.21	<**0.001**
D	*3gram.p*	19	-11915.34	5.40	0.0201
E	*cfg.surp*	20	-11897.54	35.59	<**0.001**
F	*cfg.bu*	21	-11897.33	0.42	0.5189
G	*cfg.td*	22	-11890.97	12.72	<**0.001**
H	*mg.bu*	23	-11883.32	15.30	<**0.001**
I	*mg.td*	24	-11878.17	10.29	**0.0013**

Number of timecourses: 24

LIPL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-12012.55		
A	*2gram.l*	16	-12010.24	4.62	0.0316
B	*3gram.l*	17	-12010.24	0.00	0.9651
C	*2gram.p*	18	-12007.42	5.64	0.0176
D	*3gram.p*	19	-12005.27	4.30	0.0381
E	*cfg.surp*	20	-11987.19	36.17	<**0.001**
F	*cfg.bu*	21	-11985.43	3.52	0.0606
G	*cfg.td*	22	-11983.27	4.32	0.0376
H	*mg.bu*	23	-11980.58	5.37	0.0204
I	*mg.td*	24	-11978.81	3.54	0.0598

Number of timecourses: 24

LPreM

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-12144.83		
A	*2gram.l*	16	-12142.85	3.96	0.0466
B	*3gram.l*	17	-12140.26	5.18	0.0229
C	*2gram.p*	18	-12137.00	6.52	0.0107
D	*3gram.p*	19	-12136.93	0.13	0.7163
E	*cfg.surp*	20	-12120.51	32.84	<**0.001**
F	*cfg.bu*	21	-12119.23	2.57	0.1091
G	*cfg.td*	22	-12118.42	1.62	0.2033
H	*mg.bu*	23	-12116.75	3.34	0.0676
I	*mg.td*	24	-12116.51	0.48	0.4904

Number of timecourses: 24

Table S2: Step-wise model comparison results for all regions of interest. Statistical significance, indicated in **boldface**, is evaluated against an alpha-level of $0.008\bar{3}$ to correct for multiple comparisons across regions.

S.3. Sensitivity to ROI analysis parameters

We conducted several auxiliary analyses to evaluate whether results were sensitive to the specific details of how we defined our ROIs and to the ordering of our step-wise model comparison. Specifically, we: (A1) varied the ordering by complexity metric used for model comparison ($td \prec bu \prec surp$ rather than $surp \prec bu \prec td$), (A2-A3) varied the activation inclusion threshold used ($t \geq 2.5$ or $t \geq 3.0$, rather than $t \geq 2.0$), and (A4) varied the size of the ROIs (6 mm, rather than 10 mm).

Extreme caution must be used when testing multiple reasonable values for any parameter of the analysis, as exploration may invalidate key statistical assumptions (Simmons et al., 2011; Carp, 2012). To ensure that these additional analyses did not violate our statistical assumptions, they were conducted blind to the results of the primary analysis, our interpretation and discussion are guided by patterns that held up across all analyses, and results from all of the analyses that we conducted are reported in these Supplementary Materials in Tables S3–S6.

LATL results were quite consistent across analyses; *n*-gram, CFG, and MG results obtained significantly or marginally (corrected $p \leq 0.054$) in all cases. In one analysis (A3), significant improvements were found for *mg.bu*, rather than *mg.td*, but note that these two predictors are highly correlated with each other ($r = 0.9$, see Supplementary Figure S2). Consistent results also obtained for the RATL where the *2gram.p* and *cfg.surp* results found in the primary analysis were also observed in all auxiliary analyses. In addition, *cfg.td* node counts appeared to be significant in the RATL when statistical models were evaluated in the reverse order (A1) and when 6 mm ROIs were used (A4). Further consistent results were observed in the LPTL: there were improvements in model fit for *n*-gram models, CFG surprisal and top-down node counts, as well as for MG node counts in all auxiliary analyses. There was some variation as to whether only top-down MG node counts or both top-down and bottom-up node counts reliably increased fits in this region.

Results from the IFG showed some sensitivity to analysis parameters. While *n*-gram results were consistent, effects for CFG predictors varied such that we observed significantly improved model fits in the four auxiliary analyses but not in the primary analysis. However, there was variation in whether significant fits were observed for surprisal (*cfg.surp*) or node counts (*cfg.bu*). Thus, we do not find consistent effects for any single CFG model in the IFG.

Results from the LIPL were consistent across analyses for *cfg.surp*, but all of the auxiliary analyses contrasted with the primary analysis in showing a significant effect for *mg.td* node counts. In addition, when 6 mm ROIs were analyzed (A4), the *3gram.p* model was reliable; no other analyses showed an effect in this region for the *n*-gram models.

Finally, the effect for *cfg.surp* was consistent in all analyses in the LPreM ROI. When more rigorous inclusion criteria (A2, A3) or smaller ROIs (A3) were analyzed, there were also reliable effects for the *2gram.p* model.

In summary, LATL and LPTL showed sensitivity to even the most abstract hierarchical models that was consistent across analysis parameters. Also consistent were the effects for CFG-based surprisal in almost all ROIs and effects for string-level *n*-gram surprisal in the LATL, RATL, LPTL and LIFG. In contrast, we saw variation when

parameters were adjusted in two important respects. First, the LIFG showed sensitivity to either CFG surprisal or node counts in the auxiliary analyses, but the specific model that showed improvements was not consistent. Secondly, the LIPL ROI showed a sensitivity to abstract MG node counts in the auxiliary analyses, but not the primary analysis. To be conservative our we will focus our interpretation or those patterns that were robust across all analyses.

LATL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-11549.51		
A	*2gram.l*	16	-11549.51	0.00	0.9632
B	*3gram.l*	17	-11536.63	25.77	<**0.001**
C	*2gram.p*	18	-11525.62	22.02	<**0.001**
D	*3gram.p*	19	-11524.88	1.48	0.2245
E	*cfg.td*	20	-11511.90	25.96	<**0.001**
F	*cfg.bu*	21	-11511.88	0.04	0.8494
G	*cfg.surp*	22	-11485.57	52.62	<**0.001**
H	*mg.td*	23	-11480.45	10.22	**0.0014**
I	*mg.bu*	24	-11479.47	1.97	0.1606

Number of timecourses: 23

RATL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-11217.40		
A	*2gram.l*	16	-11217.36	0.07	0.7941
B	*3gram.l*	17	-11216.98	0.75	0.3851
C	*2gram.p*	18	-11208.04	17.88	<**0.001**
D	*3gram.p*	19	-11206.91	2.27	0.1319
E	*cfg.td*	20	-11202.77	8.29	**0.0040**
F	*cfg.bu*	21	-11202.77	0.00	0.9805
G	*cfg.surp*	22	-11196.42	12.70	<**0.001**
H	*mg.td*	23	-11193.68	5.49	0.0192
I	*mg.bu*	24	-11193.57	0.22	0.6405

Number of timecourses: 22

LIFG

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-10664.70		
A	*2gram.l*	16	-10656.88	15.63	<**0.001**
B	*3gram.l*	17	-10651.79	10.18	**0.0014**
C	*2gram.p*	18	-10639.76	24.06	<**0.001**
D	*3gram.p*	19	-10639.55	0.43	0.5122
E	*cfg.td*	20	-10638.76	1.58	0.2084
F	*cfg.bu*	21	-10635.02	7.48	**0.0062**
G	*cfg.surp*	22	-10632.27	5.50	0.0190
H	*mg.td*	23	-10628.81	6.93	0.0085
I	*mg.bu*	24	-10628.73	0.14	0.7070

Number of timecourses: 21

LPTL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-11978.19		
A	*2gram.l*	16	-11963.26	29.86	<**0.001**
B	*3gram.l*	17	-11959.65	7.23	**0.0072**
C	*2gram.p*	18	-11918.04	83.21	<**0.001**
D	*3gram.p*	19	-11915.34	5.40	0.0201
E	*cfg.td*	20	-11907.73	15.22	<**0.0011**
F	*cfg.bu*	21	-11906.06	3.33	0.0680
G	*cfg.surp*	22	-11890.97	30.18	<**0.001**
H	*mg.td*	23	-11878.85	24.24	<**0.0010**
I	*mg.bu*	24	-11878.17	1.36	0.2435

Number of timecourses: 24

LIPL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-12012.55		
A	*2gram.l*	16	-12010.24	4.62	0.0316
B	*3gram.l*	17	-12010.24	0.00	0.9651
C	*2gram.p*	18	-12007.42	5.64	0.0176
D	*3gram.p*	19	-12005.27	4.30	0.0381
E	*cfg.td*	20	-12003.11	4.33	0.0375
F	*cfg.bu*	21	-12000.33	5.56	0.0184
G	*cfg.surp*	22	-11983.27	34.13	<**0.001**
H	*mg.td*	23	-11979.04	8.46	**0.0036**
I	*mg.bu*	24	-11978.81	0.45	0.5001

Number of timecourses: 24

LPreM

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-12144.83		
A	*2gram.l*	16	-12142.85	3.96	0.0466
B	*3gram.l*	17	-12140.26	5.18	0.0229
C	*2gram.p*	18	-12137.00	6.52	0.0107
D	*3gram.p*	19	-12136.93	0.13	0.7163
E	*cfg.td*	20	-12135.93	2.00	0.1575
F	*cfg.bu*	21	-12134.33	3.22	0.0729
G	*cfg.surp*	22	-12118.42	31.81	<**0.001**
H	*mg.td*	23	-12116.53	3.78	0.0518
I	*mg.bu*	24	-12116.51	0.03	0.8555

Number of timecourses: 24

Table S3: Step-wise model comparison results under alternative model ordering ($bu \prec td \prec surp$) for all regions of interest. Statistical significance, indicated in **boldface**, is evaluated against an alpha-level of $0.008\bar{3}$ to correct for multiple comparisons across regions.

LATL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-9482.42		
A	*2gram.l*	16	-9482.08	0.67	0.4136
B	*3gram.l*	17	-9470.42	23.32	<**0.001**
C	*2gram.p*	18	-9455.16	30.52	<**0.001**
D	*3gram.p*	19	-9455.02	0.28	0.5964
E	*cfg.surp*	20	-9419.22	71.60	<**0.001**
F	*cfg.bu*	21	-9417.08	4.28	0.0386
G	*cfg.td*	22	-9409.07	16.02	<**0.001**
H	*mg.bu*	23	-9406.01	6.12	0.0133
I	*mg.td*	24	-9402.59	6.83	0.0090

Number of timecourses: 19

RATL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-8156.34		
A	*2gram.l*	16	-8156.27	0.14	0.7094
B	*3gram.l*	17	-8155.96	0.61	0.4330
C	*2gram.p*	18	-8147.82	16.28	<**0.001**
D	*3gram.p*	19	-8147.31	1.01	0.3151
E	*cfg.surp*	20	-8143.55	7.53	**0.0061**
F	*cfg.bu*	21	-8143.45	0.20	0.6586
G	*cfg.td*	22	-8141.45	4.00	0.0455
H	*mg.bu*	23	-8141.41	0.08	0.7799
I	*mg.td*	24	-8141.40	0.02	0.8803

Number of timecourses: 16

LIFG

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-7075.73		
A	*2gram.l*	16	-7072.50	6.47	0.0110
B	*3gram.l*	17	-7066.86	11.26	<**0.001**
C	*2gram.p*	18	-7059.42	14.89	<**0.001**
D	*3gram.p*	19	-7059.37	0.09	0.7622
E	*cfg.surp*	20	-7053.69	11.37	<**0.001**
F	*cfg.bu*	21	-7053.25	0.88	0.3490
G	*cfg.td*	22	-7053.09	0.32	0.5742
H	*mg.bu*	23	-7052.52	1.14	0.2862
I	*mg.td*	24	-7050.43	4.18	0.0408

Number of timecourses: 14

LPTL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-11461.71		
A	*2gram.l*	16	-11447.12	29.18	<**0.001**
B	*3gram.l*	17	-11442.95	8.35	**0.0039**
C	*2gram.p*	18	-11402.11	81.68	<**0.001**
D	*3gram.p*	19	-11399.92	4.37	0.0365
E	*cfg.surp*	20	-11380.96	37.92	<**0.001**
F	*cfg.bu*	21	-11380.64	0.64	0.4230
G	*cfg.td*	22	-11373.29	14.70	<**0.001**
H	*mg.bu*	23	-11367.34	11.90	<**0.001**
I	*mg.td*	24	-11362.88	8.91	**0.0028**

Number of timecourses: 23

LIPL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-11533.91		
A	*2gram.l*	16	-11531.59	4.63	0.0314
B	*3gram.l*	17	-11531.59	0.00	0.9489
C	*2gram.p*	18	-11529.62	3.94	0.0472
D	*3gram.p*	19	-11527.37	4.51	0.0337
E	*cfg.surp*	20	-11504.89	44.95	<**0.001**
F	*cfg.bu*	21	-11503.52	2.74	0.0980
G	*cfg.td*	22	-11500.86	5.34	0.0209
H	*mg.bu*	23	-11496.57	8.57	**0.0034**
I	*mg.td*	24	-11494.89	3.35	0.0670

Number of timecourses: 23

LPreM

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-11116.33		
A	*2gram.l*	16	-11114.27	4.11	0.0427
B	*3gram.l*	17	-11110.85	6.84	0.0089
C	*2gram.p*	18	-11106.06	9.58	**0.0020**
D	*3gram.p*	19	-11105.93	0.25	0.6138
E	*cfg.surp*	20	-11092.87	26.13	<**0.001**
F	*cfg.bu*	21	-11090.92	3.89	0.0485
G	*cfg.td*	22	-11090.47	0.90	0.3439
H	*mg.bu*	23	-11087.68	5.59	0.0180
I	*mg.td*	24	-11087.15	1.05	0.3045

Number of timecourses: 22

Table S4: Step-wise model comparison results with alternative inclusion criteria of $t \geq 2.5$ for timecourses for all regions of interest. Statistical significance, indicated in **boldface**, is evaluated against an alpha-level of $0.008\overline{3}$ to correct for multiple comparisons across regions.

LATL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-7965.13		
A	*2gram.l*	16	-7965.08	0.10	0.7521
B	*3gram.l*	17	-7954.32	21.51	<**0.001**
C	*2gram.p*	18	-7938.89	30.87	<**0.001**
D	*3gram.p*	19	-7938.58	0.62	0.4321
E	*cfg.surp*	20	-7902.22	72.72	<**0.001**
F	*cfg.bu*	21	-7901.57	1.30	0.2545
G	*cfg.td*	22	-7894.20	14.73	<**0.001**
H	*mg.bu*	23	-7889.76	8.89	**0.0029**
I	*mg.td*	24	-7887.18	5.15	0.0233
			Number of timecourses: 16		

RATL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-7109.90		
A	*2gram.l*	16	-7109.88	0.05	0.8294
B	*3gram.l*	17	-7109.44	0.88	0.3475
C	*2gram.p*	18	-7097.97	22.92	<**0.001**
D	*3gram.p*	19	-7097.91	0.13	0.7211
E	*cfg.surp*	20	-7093.20	9.42	**0.0021**
F	*cfg.bu*	21	-7093.13	0.14	0.7068
G	*cfg.td*	22	-7089.76	6.73	0.0095
H	*mg.bu*	23	-7089.75	0.02	0.8933
I	*mg.td*	24	-7089.75	0.01	0.9225
			Number of timecourses: 14		

LIFG

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-4418.77		
A	*2gram.l*	16	-4414.92	7.69	**0.0056**
B	*3gram.l*	17	-4411.10	7.65	**0.0057**
C	*2gram.p*	18	-4401.09	20.02	<**0.001**
D	*3gram.p*	19	-4401.08	0.01	0.9350
E	*cfg.surp*	20	-4394.68	12.81	<**0.001**
F	*cfg.bu*	21	-4394.66	0.03	0.8627
G	*cfg.td*	22	-4394.57	0.18	0.6753
H	*mg.bu*	23	-4394.54	0.06	0.7999
I	*mg.td*	24	-4393.94	1.21	0.2709
			Number of timecourses: 9		

LPTL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-10457.05		
A	*2gram.l*	16	-10441.34	31.41	<**0.001**
B	*3gram.l*	17	-10437.54	7.61	**0.0058**
C	*2gram.p*	18	-10402.43	70.22	<**0.001**
D	*3gram.p*	19	-10401.39	2.08	0.1493
E	*cfg.surp*	20	-10384.83	33.13	<**0.001**
F	*cfg.bu*	21	-10384.81	0.04	0.8429
G	*cfg.td*	22	-10379.99	9.65	**0.0019**
H	*mg.bu*	23	-10375.05	9.88	**0.0017**
I	*mg.td*	24	-10370.34	9.41	**0.0022**
			Number of timecourses: 21		

LIPL

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-10041.72		
A	*2gram.l*	16	-10040.37	2.71	0.0995
B	*3gram.l*	17	-10040.36	0.02	0.8838
C	*2gram.p*	18	-10038.57	3.58	0.0586
D	*3gram.p*	19	-10037.32	2.50	0.1139
E	*cfg.surp*	20	-10019.28	36.07	<**0.001**
F	*cfg.bu*	21	-10017.62	3.33	0.0679
G	*cfg.td*	22	-10015.39	4.45	0.0349
H	*mg.bu*	23	-10010.87	9.04	**0.0026**
I	*mg.td*	24	-10009.45	2.84	0.0919
			Number of timecourses: 20		

LPreM

	Added Param.	df	LogLik	$\chi^2(1)$	p-value
Ø		15	-9101.37		
A	*2gram.l*	16	-9097.23	8.27	**0.0040**
B	*3gram.l*	17	-9094.72	5.01	0.0252
C	*2gram.p*	18	-9087.17	15.10	<**0.001**
D	*3gram.p*	19	-9086.74	0.86	0.3528
E	*cfg.surp*	20	-9069.43	34.63	<**0.001**
F	*cfg.bu*	21	-9067.52	3.82	0.0507
G	*cfg.td*	22	-9067.43	0.18	0.6673
H	*mg.bu*	23	-9065.04	4.77	0.0290
I	*mg.td*	24	-9064.74	0.61	0.4354
			Number of timecourses: 18		

Table S5: Step-wise model comparison results with alternative inclusion criteria of $t \geq 3.0$ for timecourses for all regions of interest. Statistical significance, indicated in **boldface**, is evaluated against an alpha-level of $0.008\overline{3}$ to correct for multiple comparisons across regions.

LATL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-11544.89		
A	*2gram.l*	16	-11544.86	0.05	0.8173
B	*3gram.l*	17	-11534.74	20.24	<**0.001**
C	*2gram.p*	18	-11517.36	34.76	<**0.001**
D	*3gram.p*	19	-11516.56	1.61	0.2043
E	*cfg.surp*	20	-11471.19	90.73	<**0.001**
F	*cfg.bu*	21	-11469.74	2.90	0.0888
G	*cfg.td*	22	-11460.19	19.10	<**0.001**
H	*mg.bu*	23	-11456.59	7.20	**0.0073**
I	*mg.td*	24	-11451.51	10.15	**0.0014**

Number of timecourses: 23

RATL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-11144.64		
A	*2gram.l*	16	-11144.64	0.00	0.9958
B	*3gram.l*	17	-11143.13	3.02	0.0825
C	*2gram.p*	18	-11134.38	17.51	<**0.001**
D	*3gram.p*	19	-11133.31	2.14	0.1438
E	*cfg.surp*	20	-11117.25	32.12	<**0.001**
F	*cfg.bu*	21	-11116.83	0.83	0.3637
G	*cfg.td*	22	-11113.02	7.63	**0.0058**
H	*mg.bu*	23	-11109.58	6.88	0.0087
I	*mg.td*	24	-11109.18	0.81	0.3684

Number of timecourses: 22

LIFG

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-10606.46		
A	*2gram.l*	16	-10598.55	15.82	<**0.001**
B	*3gram.l*	17	-10595.55	6.02	0.0142
C	*2gram.p*	18	-10580.14	30.81	<**0.001**
D	*3gram.p*	19	-10579.98	0.32	0.5736
E	*cfg.surp*	20	-10573.26	13.45	<**0.001**
F	*cfg.bu*	21	-10572.61	1.31	0.2529
G	*cfg.td*	22	-10568.33	8.55	**0.0035**
H	*mg.bu*	23	-10563.85	8.96	**0.0028**
I	*mg.td*	24	-10562.13	3.44	0.0637

Number of timecourses: 21

LPTL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-11964.99		
A	*2gram.l*	16	-11946.63	36.71	<**0.001**
B	*3gram.l*	17	-11941.48	10.31	**0.0013**
C	*2gram.p*	18	-11895.99	90.99	<**0.001**
D	*3gram.p*	19	-11893.62	4.73	0.0297
E	*cfg.surp*	20	-11871.55	44.15	<**0.001**
F	*cfg.bu*	21	-11871.32	0.45	0.5022
G	*cfg.td*	22	-11864.22	14.21	<**0.001**
H	*mg.bu*	23	-11857.32	13.80	<**0.001**
I	*mg.td*	24	-11851.27	12.09	<**0.001**

Number of timecourses: 24

LIPL

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-11969.32		
A	*2gram.l*	16	-11967.04	4.57	0.0326
B	*3gram.l*	17	-11966.64	0.79	0.3730
C	*2gram.p*	18	-11964.09	5.11	0.0238
D	*3gram.p*	19	-11960.60	6.98	**0.0082**
E	*cfg.surp*	20	-11934.20	52.79	<**0.001**
F	*cfg.bu*	21	-11932.87	2.66	0.1027
G	*cfg.td*	22	-11931.11	3.53	0.0602
H	*mg.bu*	23	-11927.55	7.12	**0.0076**
I	*mg.td*	24	-11926.04	3.02	0.0823

Number of timecourses: 24

LPreM

	Added Param.	df	LogLik	$\chi^2(1)$	*p*-value
Ø		15	-12121.20		
A	*2gram.l*	16	-12118.40	5.59	0.0180
B	*3gram.l*	17	-12116.25	4.31	0.0379
C	*2gram.p*	18	-12107.61	17.28	<**0.001**
D	*3gram.p*	19	-12107.19	0.84	0.3605
E	*cfg.surp*	20	-12089.51	35.36	<**0.001**
F	*cfg.bu*	21	-12087.01	5.01	0.0253
G	*cfg.td*	22	-12085.30	3.42	0.0646
H	*mg.bu*	23	-12081.95	6.70	0.0096
I	*mg.td*	24	-12081.37	1.15	0.2826

Number of timecourses: 24

Table S6: Step-wise model comparison results with alternative 6 mm diameter regions of interest. Statistical significance, indicated in **boldface**, is evaluated against an alpha-level of $0.008\overline{3}$ to correct for multiple comparisons across regions.

S.4. Detailed whole brain results

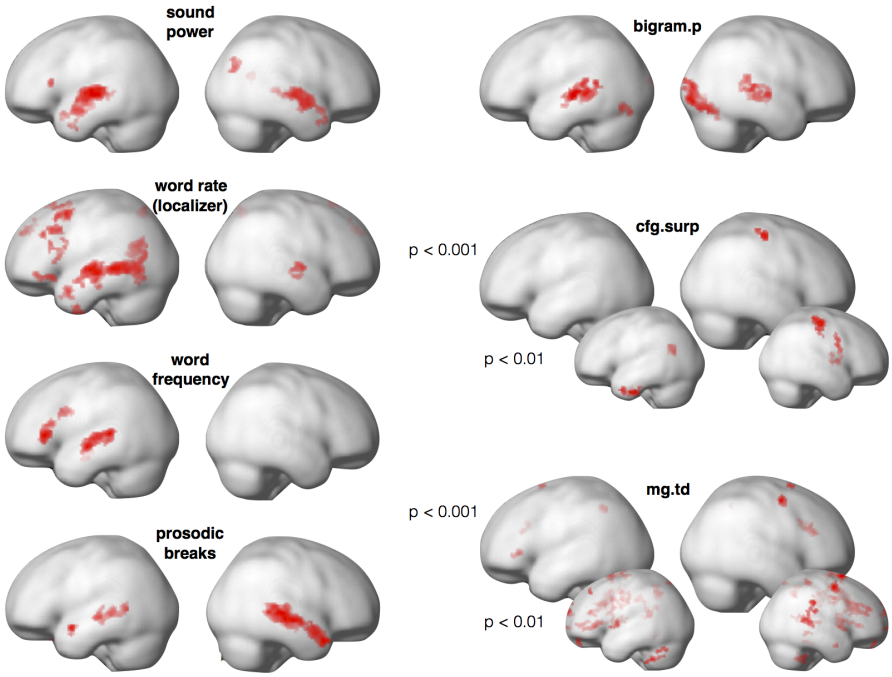

Figure S4: Statistical activation maps for the seven predictors used in the whole brain analysis; predictors for head movements are excluded. Positive correlations are shown in all maps except for word frequency, which plots anti-correlations (i.e. greater activation for lower-frequency words). All maps are projected on to the smoothed surface of the SPM8 template brain. Unless otherwise indicated, maps are thresholded at $p < 0.001$ with a cluster size of at least 50 voxels.

Predictor	MNI *x, y, z*	Label		# Vx	Peak *t*	Cluster *p*	
Sound Power	-66, -24, 6	L	Superior Temporal	1084	8.45	< 0.001	***
	64, -2, -6	R	Superior Temporal	787	8.37	< 0.001	***
	12, -56, 20	R	Precuneus	52	4.82	0.19	
	46, -74, 26	R	Middle Occipital	79	4.64	0.037	*
Word Rate	-64, -12, -2	L	Middle Temporal	2109	6.98	< 0.001	***
	-38, 8, 54	L	Middle Frontal	346	5	< 0.001	***
	-50, 34, 32	L	Middle Frontal	143	5.11	0.002	**
	66, -8, -2	R	Superior Temporal	130	5.09	0.004	**
	-48, 22, -12	L	Temporal Pole	101	4.24	0.017	*
	-56, 18, 8	L	Inferior Frontal, Pars Triangularis	96	4.36	0.023	*
	-10, 54, 34	L	Superior Frontal, Medial	85	4.3	0.04	*
	-54, 8, -24	L	Middel Temporal	82	4.78	0.047	*
	-8, -68, 56	L	Precuneus	62	4.75	0.14	
	-12, 22, 62	L	Supplementary Motor	62	4.33	0.14	
	-50, -4, -46	L	Inferior Temporal	58	4.34	0.175	
Word Frequency	-62, -12, -8	L	Middle Temporal	436	6.73	< 0.001	***
(negative effects)	-50, 30, 6	L	Inferior Frontal, Pars Triangularis	299	6.23	< 0.001	***
	-40, 4, 26	L	Inferior Frontal, Pars Opercularis	101	5.02	0.007	**
	-26, -14, -18	L	Hippocampus	62	4.8	0.154	
Prosodic Breaks	50, 12, -24	R	Middle Temporal Pole	1625	7.94	< 0.001	***
	-50, -36, -2	L	Middle Temporal Pole	281	6.52	< 0.001	***
	-54, 2, -16	L	Middle Temporal	59	5.47	0.175	
	-62, -22, -6	L	Middle Temporal	52	5.03	0.256	
2gram.p	-58, -22, -2	L	Middle Temporal	762	6.66	< 0.001	***
	40, -74, -16	R	Right Fusiform	679	5.86	< 0.001	***
	48, -28, -2	R	Superior Temporal	486	5.98	< 0.001	***
	-40, -70, -14	L	Left Fusiform	98	4.76	0.019	*
cfg.surp	56, -18, 54	R	Postcentral	58	6.1	0.205	
mg.td	42, -2, 54	R	Middle Frontal	59	5.06	0.188	
	30, 28, 22	R	Middle Frontal	68	4.93	0.116	
	-4, -52, 46	L	Precuneus	66	4.56	0.129	

Table S7: Whole brain results for all linguistic predictors showing positive activation clusters of at least 50 voxels exceeding $p < 0.001$. Results are ordered by cluster size. MNI coordinates indicate the location of the cluster maximum. Anatomical labels are given for the peak of voxel in each cluster following the parcellation of Tzourio-Mazoyer et al. (2002). *P*-values indicate the probability of observing a cluster of at least that size by chance given the estimated smoothness of the data, corrected for multiple comparisons according to Random Field Theory (Worsley et al., 1996).

MNI *x, y, z*	Label		# Voxels	Peak *t*	Cluster *p*
56, -18, 54	R	Postcentral	212	6.1	0.277
-48, 4, -34	L	Inferior Temporal	185	4.07	0.423
-38, -64, 18	L	Middle Occipital	153	3.78	0.656
50, 2, 4	R	Rolandic Operculum	124	3.49	0.869
62, 4, 34	R	Postcentral	71	4.23	> 0.9
-8, -4, -34	L	Parahippocampal	51	3.68	> 0.9

Table S8: Uncorrected whole brain results for *cfg.surp* reporting positive activation clusters of at least 50 voxels exceeding a liberal $p < 0.01$. Results are ordered by cluster size. MNI coordinates indicate the location of the cluster maximum. Anatomical labels are given for the peak of voxel in each cluster following the parcellation of Tzourio-Mazoyer et al. (2002). *P*-values which indicate the probability of observing a cluster of at least that size by chance given the estimated smoothness of the data, corrected for multiple comparisons according to Random Field Theory (Worsley et al., 1996). These results are included to complement the exploratory activations that are shown in green on Figure 5; they are not statistically reliable.

6. Supplementary References

Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, 6.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 135966.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15, 27389.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4, 5873.