Electrophysiological evidence for top-down influences on early speech perception: Supplemental online material

Laura M. Getz & Joseph C. Toscano
Villanova University, Dept. of Psychological and Brain Sciences

Sequential Semantic Priming Word Lists

We began with a list of 470 potential primes, and selected 51 word pairs that had >85% agreement across subjects to serve as potential Association pairs. From this list, seven were removed because the target included a consonant cluster at onset (*e.g.*, magnifying *glass*, oak *tree*), two were removed for rhymes (*e.g.*, fender *bender*) or onomatopoeia (*e.g.*, ding *dong*), and three were removed for repeated target words (*e.g.*, grizzly *bear* had lower agreement than teddy *bear*). Eleven additional words were eliminated in order to balance the mean log frequency of voiced ($M = 1.59$) and voiceless ($M = 1.24$) targets ($t(26) = 1.64$, $p = 0.113$), as measured from the MRC Database (`websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm`) using the Kučera-Francis written frequencies (Kučera & Francis, 1967, three words are excluded from this measurement because they did not have K-F frequencies). This left us with 28 Association prime pairings, with 15 targets forming half of a minimal pair (*e.g.*, amusement *park*) and 13 forming a Ganong pair (*i.e.*, word/non-word pair, as in Ganong, 1980, *e.g.*, marching *band*).[1]

Eighteen of the Neutral primes had a most frequently chosen target word with <15% agreement across subjects (*e.g.*, only 14.8% of participants agreed on what word should follow the prime 'TURKEY'). Because these words were poor sequential semantic primes, they were paired with the same targets as the Association pairs, balancing the proportion with alliteration between the prime and target ($N = 6$; *e.g.*, bumble *bee* vs. book *bear*) and the presence of the target phoneme elsewhere in the prime ($N = 10$; *e.g.*, ballpoint *pen* vs. smoke *cube*). An additional ten Neutral primes were added in order to balance the number of proper nouns between the Neutral and Association prime pairs; they were chosen to be stand-alone primes without obvious sequential targets.[2] Details of the full word lists are displayed in Table S1.

[1] Two Ganong pairs (*i.e.*, band–pand and tower–dower) and one minimal pair (doll–tall) could potentially be considered members of the opposite group; however, a re-analysis of the RT data with these three pairs switched led to an identical pattern of results.

[2] Although proper and common nouns have different lexical properties (Semenza & Zettin, 1988), the critical factor in selecting the prime-target pairs for the current study was that they served as excellent sequential pairs; proper nouns are particularly strong in this regard and were therefore included in the experiments. Analyses of the RTs in Experiments 1 and 2 found no main effect of noun type (proper vs. common) and no interaction between noun type and prime type.

Corresponding author: Laura M. Getz (`laura.getz@villanova.edu`), Villanova University Dept. of Psychological and Brain Sciences, 800 E. Lancaster Ave. Villanova, PA 19085.

Prime Characteristics				Target Characteristics					Prime-Target Pair Characteristics		
Association Prime	Assoc Agree	Neutral Prime	Neutral Agree	Target Word	Log Freq	Pair Type	Voicing	Place of Articulation	Proper Noun	Alliteration	Phoneme Repetition
teddy	100.0	book	13.0	bear	1.76	minimal	voiced	bilabial		N	N
Notre	98.1	Sacramento	-	Dame	0.85	minimal	voiced	alveolar	A, N		
bunk	96.3	warrior	7.4	beds	2.10	Ganong	voiced	bilabial		A	A
chewing	96.3	Google	-	gum	1.15	minimal	voiced	velar	N	N	N
Sahara	94.4	December	-	Desert	1.32	Ganong	voiced	alveolar	A, N	N	N
wrecking	94.4	life	11.1	ball	2.04	minimal	voiced	bilabial			
Dunkin'	92.6	European	-	Donuts	-	Ganong	voiced	alveolar	A, N	A	A
Webster's	92.6	Charlotte	-	Dictionary	1.76	Ganong	voiced	alveolar	A, N		
Barbie	90.7	Wyoming	-	Doll	1.00	minimal	voiced	alveolar	A, N		
Nerf	90.7	lodge	11.1	Gun	2.07	Ganong	voiced	velar	A		
atomic	88.9	butter	11.1	bomb	1.56	minimal	voiced	bilabial		N	N
marching	88.9	evening	11.1	band	1.72	Ganong	voiced	bilabial			
bumble	87.0	moss	5.6	bee	1.04	minimal	voiced	bilabial		A	A
Gossip	85.2	NASA	-	Girl	2.34	minimal	voiced	velar	A, N	A	A
M(voiced)	**92.6**		**10.1**		**1.59**						
amusement	100.0	finger	13.0	park	1.97	minimal	voiceless	bilabial			
Eiffel	100.0	turkey	14.8	Tower	1.11	Ganong	voiceless	alveolar	A	N	N
Rubix's	100.0	smoke	13.0	Cube	0.00	Ganong	voiceless	velar	A		A, N
mashed	98.1	face	11.1	potatoes	1.18	Ganong	voiceless	velar			
pepperoni	98.1	castle	11.1	pizza	-	Ganong	voiceless	bilabial		A	A
guinea	96.3	paper	11.1	pig	0.9	minimal	voiceless	bilabial		N	N
tater	94.4	squeeze	9.3	tots	-	minimal	voiceless	alveolar		A	A
umbilical	94.4	freeze	9.3	cord	0.78	minimal	voiceless	velar			A
ballpoint	92.6	self	7.4	pen	1.26	minimal	voiceless	bilabial			A
remote	92.6	Denmark	-	control	2.35	Ganong	voiceless	velar	N		N
frying	90.7	space	14.8	pan	1.20	minimal	voiceless	bilabial			N
Shirley	90.7	Internet	-	Temple	1.58	Ganong	voiceless	alveolar	A, N		N
duct	88.9	Wednesday	-	tape	1.54	Ganong	voiceless	alveolar	N		A
stubbed	88.9	summary	11.1	toe	0.95	minimal	voiceless	alveolar			
M(voiceless)	**94.7**		**11.4**		**1.24**						

Table S1

Prime, Target, and Prime-Target Pair characteristics for the word lists used in Experiments 1 and 2. Note: Loq Freq = log frequency; A = Association; N = Neutral. See text for full explanation.

Experiment 2 Supplemental Results

Reaction Times

Participants' RTs in Experiment 2 showed a three-way interaction between target VOT, expected voicing, and prime type, collapsing across Ganong pairs and minimal pairs.[3] These RT differences could correspond to effects of lexical status for Ganong pairs across all prime types (*i.e.*, where the opposite VOT endpoint was a non-word) as well as the expectation elicited by the semantic prime in the Association condition. Here, we investigate the contributions of lexical status by examining differences in RT between Ganong pairs and minimal pairs.

We first transformed the variables target VOT and expected voicing into a single 'Voicing' variable with levels Expected (*i.e.*, the VOT presented on that trial matches the listeners' expectation; for example, the target BEAR is presented with a short VOT or the target PARK is presented with a long VOT), Ambiguous (*i.e.*, the VOT presented is intermediate between the voiced and voiceless endpoints), and Opposite (*i.e.*, the VOT presented on that trial is the opposite of the listeners' expectation; for example, the target BEAR is presented with a long VOT or the target PARK is presented with a short VOT). As in the main text, note that we use *expected* for consistency, even though only the Association condition leads to the expectation of a particular word to follow. Participants' RTs were submitted to 3×3 repeated-measures ANOVAs with prime type (Association vs. Neutral vs. Mask) and Voicing (Expected vs. Ambiguous vs. Opposite) as factors and logRT as the dependent measure separately for Ganong pairs and minimal pairs.

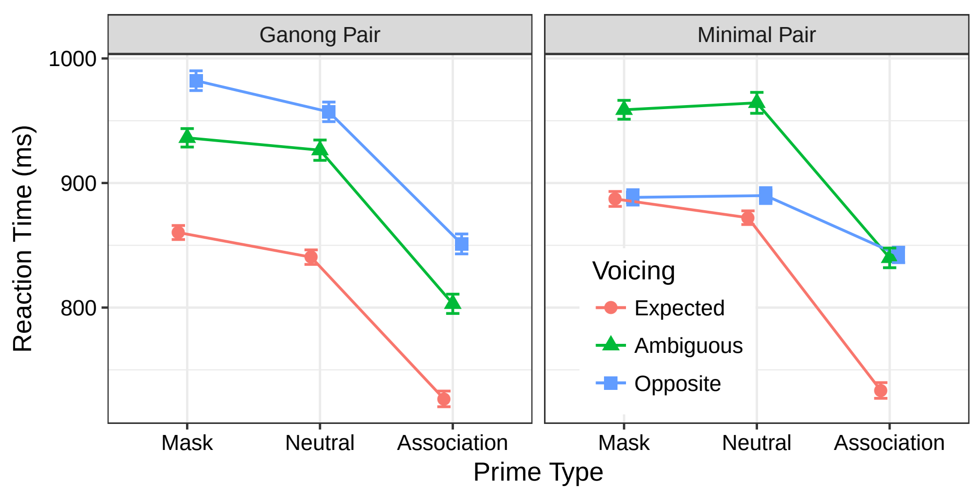

Figure S1. Experiment 2 reaction times by prime type, target voicing, and pair type. Error bars represent ±1 SE.

Ganong Pairs. For Ganong pairs, the Expected voicing condition is always a word whereas the Opposite voicing condition is a non-word; an effect of lexical status would therefore result from the slowest RTs in the Opposite condition and fastest RTs in the Expected condition, with the Ambiguous RTs falling in between. The ANOVA results support such an effect. There was a main effect of voicing [$F(2,56) = 72.47$, $p < 0.001$, $\eta^2_G = 0.14$], with follow-up analyses revealing that

[3] Note that here and in Experiment 1, we found no interactions with block (first half vs. second half of the experiment). Although participants generally respond faster over the course of the experiment, the Association condition already shows a sizable RT advantage in the first block. This suggests that any perceptual learning effects are small in comparison to effects driven by the pre-existing lexical connections between the semantic primes and the expected target words.

RTs were faster in the expected condition than ambiguous condition [$t(86) = -11.99$, $p < 0.001$] and faster in the ambiguous condition than the opposite condition [$t(86) = -5.93$, $p < 0.001$]. There was also a main effect of prime type [$F(2, 56) = 86.54$, $p < 0.001$, $\eta^2_G = 0.24$], with faster RTs in the Association condition [$t(86) = -14.86$, $p < 0.001$] than in the Neutral condition, and faster RTs in the Neutral than Mask condition [$t(86) = -3.75$, $p < 0.001$], demonstrating an effect of semantic priming. These results are displayed in the left panel of Figure S1.

Minimal Pairs. For minimal pairs, both the voiced and voiceless endpoints form words; we would therefore expect RTs to be slower in the Ambiguous condition than either the Expected or Opposite conditions in the absence of any expectation from the prime (*i.e.*, in the Neutral and Mask conditions). In the Association condition, the Expected condition should be faster than the Opposite condition given the expectation set up by the prime word. The ANOVA results show an interaction between voicing and prime type [$F(4, 112) = 35.92$, $p < 0.001$, $\eta^2_G = 0.05$], resulting from successful semantic priming in the Association condition (*i.e.*, faster RTs with the Expected compared to Opposite voicing [$t(28) = -9.16$, $p < 0.001$] and no RT difference between Opposite and Ambiguous voicing [$t(28) = -1.28$, $p = 0.211$]). This is in contrast to the Neutral and Mask conditions, where there was no RT difference between the Expected and Opposite voicing conditions [$t(28) = -1.77$, $p = 0.087$ and $t(28) = 0.06$, $p = 0.95$, respectively] and RTs were slower for the Ambiguous voicing condition [$t(28) = 5.44$, $p < 0.001$ and $t(28) = 6.28$, $p < 0.001$, respectively]. These results are displayed in the right panel of Figure S1.

N1 Amplitude

For targets that form minimal pairs, participants have no reason to favor one phonological endpoint over the other in the absence of an Association prime (*e.g.*, neither FINGER *park* nor FINGER *bark* sets up a semantic expectation, whereas AMUSEMENT *park* is a more likely sequential word pair than AMUSEMENT *bark*). Targets that form Ganong pairs, in contrast, have a built-in expectation based on which endpoint is a word, which affects both Neutral and Association primes (*e.g.*, FACE *potatoes* is more likely to occur in English than FACE *botatoes*, just as MASHED *potatoes* is more likely than MASHED *botatoes*); this effect of lexical status is reflected in listeners' RTs, as described above. Note, however, that because the prime-target pairs were fully randomized, and because the N1 response is early and reflects encoding of the initial consonant (*i.e.*, before listeners could determine the lexical status of a Ganong pair stimulus), we expect no difference in N1 amplitude based on word pair type, *per se*.

However, Ganong pairs with Neutral primes may provide some insight into potential effects of lexically-mediated perceptual learning (Norris, McQueen, & Cutler, 2003). Because Neutral primes were paired with the same target words (*e.g.*, *b/potatoes* always followed the prime FACE in the Neutral condition), it is possible that listeners learned this arbitrary association over the course of the experiment. Note though, that this could only produce top-down effects on perceptual encoding (*i.e.*, differences in N1 amplitude) if the prime is associated with a *single* lexical item. Such an effect is not possible for minimal pairs because any learned association between the Neutral prime and target word would lead to an equal expectation for voiced and voiceless words (*e.g.*, FINGER *bark* and FINGER *park* are equally likely). Thus, top-down effects driven by perceptual learning in the Neutral condition (if they occur) would only be expected for Ganong pairs (*e.g.*, where a learned association between FACE and its potential targets [*b/potatoes*] can only lead to activation of the lexical item *potatoes*). Though our study was not designed to investigate such effects and we lack the power to make any firm conclusions, the ERP responses in the Neutral condition may help address whether perceptual learning is the result of interactions between higher and lower-level representations (Kraljic & Samuel, 2005).

In order to investigate this, we examined N1 amplitude as a function of both prime type and word pair type (Ganong vs. minimal). Figure S2 shows mean N1 amplitude by pair type in addition to prime type (Association vs. Neutral vs. Mask), target VOT (short vs. intermediate vs. long), and expected target voicing (voiced vs. voiceless). As expected, N1 amplitude of the ambiguous VOT stimuli varied as a function of the expected target voicing category in the Association condition regardless of whether the targets came from a minimal pair or Ganong pair. In the Mask condition, the effect of target VOT again remains the same regardless of whether targets were minimal or Ganong pairs, with no effect of expected voicing (since the mask does not lead to such an expectation). In the Neutral condition, however, there is potential evidence to support an effect of lexically-mediated perceptual learning on N1 amplitude, with the pattern for Ganong pairs looking similar to the Association condition.

These observations were validated statistically by submitting mean N1 amplitudes to separate 3 (target VOT) $\times$ 2 (expected voicing) $\times$ 2 (word pair type) repeated measures ANOVAs for each of the three prime types. Most importantly, in the Association condition, we found an interaction between target VOT and expected target voicing [$F(2,56) = 4.55$, $p = 0.015$, $\eta^2_G = 0.01$] but no main effect of pair type [$F(1,28) = 0.05$, $p = 0.828$] and no three-way interaction between pair type, target VOT, and expected target voicing [$F(2,56) = 1.11$, $p = 0.338$]. This means that ambiguous VOTs were encoded similarly to the voicing endpoint elicited by the prime for both minimal pair and Ganong pair targets.

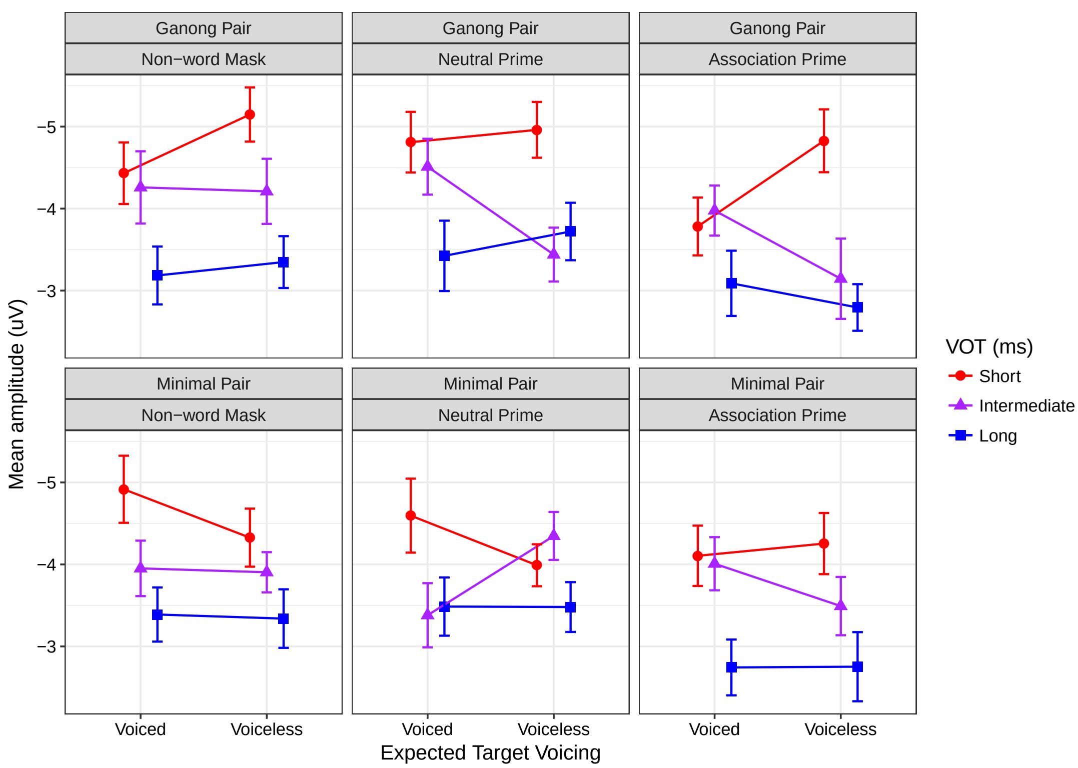

Figure S2. Experiment 2 mean N1 amplitude by prime type and target pair type. Error bars represent ±1 SE.

In the Mask condition, there was a main effect of target VOT [$F(2,56) = 15.59$, $p < 0.001$, $\eta^2_G = 0.03$], as in the analyses in the main text. There was no main effect of pair type and no interactions, meaning the VOT effect on N1 amplitude was similar regardless of whether the target formed a minimal pair or Ganong pair and whether the expected target was voiced or voiceless.

In the Neutral condition, there was a three-way interaction between expected target voicing, target VOT, and pair type [$F(2,56) = 4.33$, $p = 0.018$, $\eta^2_G = 0.01$], potentially indicating evidence for lexically-mediated perceptual learning for the Ganong pairs over the course of the experiment. Although promising, follow-up analyses revealed only a main effect of target VOT for the Ganong pairs [$F(2,56) = 8.19$, $p < 0.001$, $\eta^2_G = 0.03$] and no significant effects for the minimal pairs. Because the experimental design was not set up to test for these effects, we will leave firm conclusions about the top-down nature of lexically-mediated perceptual learning for future investigations. Note also that the same explanations presented in the main text for the source of semantic priming effects on perception (e.g., direct lexical-prelexical effects vs. a route via the speech production system) could apply here as well. The current results nonetheless suggest a useful paradigm for addressing this debate.

References

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110-125.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141-178.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204-238.

Semenza, C., & Zettin, M. (1988). Generating proper names: A case of selective inability. *Cognitive Neuropsychology*, *5*(6), 711-721.