On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement

Darren Tanner

PII: S0010-9452(14)00122-1

DOI: 10.1016/j.cortex.2014.04.007

Reference: CORTEX 1189

To appear in: *Cortex*

Received Date: 16 April 2014

Accepted Date: 24 April 2014

On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement

Darren Tanner
Department of Linguistics
University of Illinois at Urbana-Champaign

Address for correspondence:

Darren Tanner
Department of Linguistics
4080 Foreign Languages Building, MC 168
707 S Mathews Ave
Urbana, IL 61801
dstanner@gmail.com
+1-217-244-5841

In a review of event-related potential (ERP) studies, Molinaro and colleagues (Molinaro, Barber, & Carreiras, 2011) lay out a theory of the temporal and neural dynamics of grammatical agreement comprehension. As is clear from their review, ERPs are an excellent tool for studying agreement processing because they have a temporal resolution high enough to detect transient events in the brain, such as those that characterize language comprehension, and their multidimensional nature allows one to make inferences about qualitatively dissociable cognitive processes engaged during real-time processing (Otten & Rugg, 2005; Rugg & Coles, 1995). In particular, their theory accounts for findings across domains of linguistic agreement (e.g., determiner-noun, subject-verb) and agreement feature types (e.g., person, number, gender) and proposes three separable sub-stages in the processing of sentence-embedded agreement anomalies. While the latter two stages in their theory are indexed by a large positive-going wave in the anomalous relative to well-formed condition (the P600 effect), it is the ERP index of the first stage that is of concern here. According to their theory, this first stage occurs between approximately 300 and 500 ms after presentation of an anomalous word, and reflects a syntactic analysis where a morphosyntactic violation is detected based on a mismatch with predicted features. The ERP index of this stage is a negative-going wave in the anomalous compared to well-formed condition, most prominent over left anterior electrodes: the left anterior negativity, or LAN.

Whereas the P600 effect is found nearly uniformly across languages, experimental tasks, and presentation modalities for agreement-violating words, the LAN shows considerable cross-study variability in its scalp topography, or even whether it is elicited at all (see Tanner & van Hell, 2014, for a review). Nonetheless, Molinaro and colleagues maintain that the LAN is one of the defining electrophysiological markers of morphosyntactic agreement processing, and discuss some issues that may relate to its presence or absence across studies. In addition to outlining several linguistic criteria that need to be met in order for a LAN to be elicited, they address two other issues that could potentially impact the presence or absence of a LAN: the reference electrode(s) chosen during data processing and the role that individual differences in ERP responses could play in giving the illusion of a LAN effect in the grand mean. These two issues will be addressed in this commentary.

Regarding the first issue, Molinaro and colleagues argue that LAN effects are most likely to be found in studies in which scalp electrodes are referenced to the averaged activity over left and right mastoid electrodes, rather than when only the left mastoid is used as a reference. While a full discussion of referencing in electrophysiological data processing is beyond the scope of this paper (see Luck, 2005), the logic behind this argument is as follows. Since reported voltages reflect the difference in electrical potential between the active and reference electrodes, references placed over the left hemisphere might disproportionately subtract away effects that are left hemisphere-dominant. Averaged mastoid references, on the other hand, will not preferentially subtract away activity over a single hemisphere. According to Molinaro and colleagues, LAN effects should therefore most readily surface with a hemisphere-neutral reference, such as the averaged mastoid references. As evidence for this claim, their review shows that the majority of studies using a left mastoid reference do not report a LAN (4 of 11), whereas the majority of studies using an averaged mastoid reference do report a LAN (8 of 10).

Molinaro and colleagues raise the second issue regarding the possible impact of individual differences in response to a caution made by Osterhout and colleagues (Osterhout, McLaughlin, Kim, Greewald, & Inoue, 2004) in interpreting biphasic waveforms. In particular, Osterhout et al. suggest that in some cases, biphasic responses can result from averaging across individuals who differ in brain responses to linguistic anomalies (cf. Osterhout, 1997). That is, if some individuals show a negative-going response (e.g., an N400 effect between 300 and 500 ms poststimulus) and others show a positive-going response (e.g., a P600 after 500 ms poststimulus), the result in the grand mean ERP will be the average of the two effects (a biphasic negative-positive experimental effect). However, Molinaro and colleagues argue that, if individual difference could indeed drive biphasic LAN-P600 effects frequently seen to agreement violations, based on sampling variability from the general population, some studies should report only P600 effects, others only N400 effects, and others something in between (p. 923). Indeed, even a cursory review of ERP studies of agreement processing (and morphosyntactic processing more generally) shows that there are numerous papers reporting only P600 effects or biphasic effects, but nearly no studies showing a N400 only to agreement violations (though see Severens, Jansma, & Hartsuiker, 2008). Again, this view of the literature is consistent with Molinaro and colleagues' argument that individual differences are unlikely to account for variability in LAN effects across studies.

However, both of these claims are largely theoretical, and moreover, they are falsifiable hypotheses, which can be empirically verified. That is the goal of this commentary. With respect to the referencing issue, because referencing is a linear transformation applied to EEG/ERP data involving the simple subtraction of voltage values from the reference from all other electrodes, data can be re-referenced offline an infinite number of times. One can verify the role that choice of reference electrode (left versus averaged mastoid) might play in distorting or eliminating experimental effects like the LAN. It is therefore possible to test the claim that choosing a left mastoid reference over an averaged mastoid reference will remove left hemisphere-dominant experimental effects while having little impact on other effects.

With regard to the role of individual differences, recent data reported by Tanner and van Hell (2014; see also Osterhout, 1997) show that individual variability can indeed be a contributor to the appearance of a LAN effect in response to subject-verb agreement violations. In that study, grand mean ERP waveforms elicited by subject-verb agreement violations showed a statistically reliable left hemisphere negativity (with a scalp topography falling in the range of previous LAN reports) followed by a P600. However, subsequent analyses showed that this biphasic LAN-P600 responses was a result of averaging across individuals who differed in the quality of ERP effects: some individuals showed a primarily centrally-distributed N400 effect, whereas others showed primarily P600 effect with a right hemisphere preponderance (and no LAN), and moreover, individuals fell along a continuum between these N400- and P600-dominant responses. The left hemisphere scalp topography of the negativity in the grand mean was therefore a consequence of spatiotemporal overlap between the centrally-distributed N400 and right hemisphere-dominant P600 seen across individuals. Note also that this study used an average mastoid reference, such that we should have been able to detect a LAN effect co-occurring with the P600 effect were it present. However, despite these results showing that individual differences can give the illusion of a LAN effect in a grand mean analysis, Molinaro and colleagues' concern about the lack of reported grand mean N400 (with no accompanying P600) effects still stands. Since Tanner and van Hell identified that a subset of the population shows N400s to subject-verb agreement violations, it is unclear why, across studies, no samples showing only N400 effects are randomly drawn.

Both of these issues raised by Molinaro and colleagues will be addressed here. By using data reported in Tanner and van Hell (2014), this commentary will investigate the role that reference choice has on scalp the scalp topography of early negativities elicited by morphosyntactic anomalies, as well the likelihood of randomly drawing a sample from the population showing only an N400 and no P600 effect. The data from Tanner and van Hell are particularly well-suited to this second goal because 1) the report by Tanner and van Hell established that a sub-set of participants showed an N400 and no P600, and 2) the sample size in that study (n = 40) is approximately twice the size of most studies reviewed by Molinaro and colleagues and thus provides a better estimate of variability in the general population than many published ERP studies. As such, it is possible to use resampling statistics to investigate the likelihood of drawing any given experimental effect, as well as the effect of sample size on the distribution of ERP effects. The data reported below are drawn from the subject-verb agreement condition reported by Tanner and van Hell (2014). See that report for recording and data processing parameters.

1 Referencing and the LAN

Molinaro and colleagues propose that one source of cross-study variability in reported LAN effects reflects variability in choice of reference site, and moreover, that LAN effects are most likely to surface when the reference is not biased to subtract out activity over the left hemisphere. While this claim is based on a review of studies, it is in principle quite easy to test the contribution of reference site to the scalp topography of a given ERP effect, since data can be re-referenced offline using simple arithmetic. Although most ERP studies of language comprehension use either mastoid or earlobe electrodes as reference sites, very few reports actually show the waveforms from these sites. Studies using the left mastoid as a reference often report recording ERPs over the right mastoid in order to determine that experimental effects are not detectable over the mastoids, but generally do not show these waveforms (e.g., Kim & Osterhout, 2005; Osterhout & Mobley, 1995), making it impossible to know exactly how using an averaged mastoid reference might alter the presence of topographically isolated effects like the LAN.

Here we reanalyze data from Tanner and van Hell (2014) to investigate the impact of reference choice on the LAN. Figure 1 depicts data from the subject-verb agreement condition reported by Tanner and van Hell from three representative scalp electrodes (left frontal (F7), right frontal (F8), midline parietal (Pz)), as well as both the left (M1) and right (M2) mastoid electrodes using an averaged mastoid reference (Panel A) and a left mastoid reference (Panel B). As can be seen, Panel A shows a small negativity between approximately 300 and 450 ms poststimulus present over left, but not right hemisphere sites (a LAN) followed by a large positivity over Pz (P600). This is consistent with the grand mean analysis reported by Tanner and van Hell, which showed a reliable left hemisphere negativity followed by a P600. Here the mastoid electrodes show negligible but reciprocal deviations from zero. Panel B shows a similar state of affairs. Note that the right mastoid (M2) in Panel B shows only small deviations from zero in both conditions. Importantly, there were no experimental effects detectable over this electrode in any time window, nor did the amplitude in either condition deviate reliably from zero in any time window. This suggests that there was virtually no differential activity over the two mastoids, at least in the experimental conditions implemented by Tanner and van Hell, and that the choice of left or averaged mastoid reference would have little impact on detectability of topographically isolated effects, such as the LAN.

[FIGURE 1 ABOUT HERE]

Figure 2 depicts ungrammatical minus grammatical difference waves comparing experimental effects for grand means using the averaged and left mastoid references. As is evident, the effect of reference choice had nearly no impact on the experimental effects over any of the electrodes. Using the left mastoid reference, the experimental effects were inconsequentially more positive going at each of the electrode sites. Importantly, changing the reference site did not differently impact one hemisphere over another. This follows from the basic logic of referencing: any effects detectable in the reference electrode will be subtracted *from all scalp sites equally*. To the extent that using a left mastoid electrode reduces a negativity over left frontal sites, it will reduce negativities (or introduce positivities) at all other sites, with no hemispheric differences in how these referencing effects are distributed. Although it is quite clear that choice of reference can significantly impact ERP waveform morphology (see Luck, 2005, for clear demonstrations of this), the choice of left versus averaged mastoid references had little impact on experimental effects related to processing subject-verb agreement violations in the English monolinguals studied by Tanner and van Hell (2014).

[FIGURE 2 ABOUT HERE]

A further consideration regarding how referencing may impact the presence of a LAN effect involves constraints on the source configuration of the cortical source(s) of the LAN. In order for a left mastoid reference to eliminate a LAN effect that would have been present using an averaged mastoid reference, the dipole generator(s) of the effect would have to be located and oriented in such a way so as to produce nearly identical negative signals over both left frontal and mastoid electrodes, a positive signal of equal strength over the right mastoid, and presumably neutral effects at all other scalp sites. In this scenario, using a left mastoid reference would indeed eliminate the LAN; however, it would also introduce a positivity of double the amplitude over the right mastoid. In such a case an experimental effect would certainly be detectable, counter to most authors' reports of no experimental effects over the right mastoid. Moreover, given properties of electrical dipoles and volume conduction in the brain, cortical electrical activity that is sufficiently synchronous to be detectable at the scalp will have at least some impact on electrical potentials at all recording sites – not only those adjacent to the equivalent current dipole created by the source generator(s). It therefore seems unlikely that the particular source configuration needed to give rise to such a scenario where a left, but not averaged mastoid reference selectively removes a LAN (and no other) effect exists. Thus, it is not necessarily the case the choice of left versus averaged mastoid references can account for much of the cross-study differences in presence of LAN effects, as suggested by Molinaro and colleagues.

2 Individual differences and population sampling

A second major issue raised by Molinaro and colleagues relates to the possible role of individual differences in giving rise to the illusion of LAN effects in grand mean waveforms. As suggested by Osterhout et al. (2004), biphasic LAN-P600 effects in grand mean ERPs could be a spurious result of averaging across individuals who show either primarily negative- or positive-going ERP effects in different time windows. Tanner and van Hell (2014) systematically investigated this, and showed that the scenario described by Osterhout and colleagues can indeed spuriously give rise to a grand mean LAN-P600 effect (see also Osterhout, 1997). Whereas some agreement violations elicited a centrally-distributed N400 effect in some individuals, violations elicited a right-hemisphere dominant P600 in others. By averaging across these individuals, the spatiotemporal overlap of these two ERP effects gave rise to the illusion of a grand mean biphasic LAN-P600 response (see Figures 3 and 4). However, if such variability in brain responses exists robustly in the general population, Molinaro and colleagues rightly ask why – by random sampling probabilities – some studies do not occasionally draw samples showing only N400 effects.

[FIGURE 3 ABOUT HERE]

[FIGURE 4 ABOUT HERE]

It is possible to address this question via resampling statistics, such as Monte Carlo simulations. By performing Monte Carlo simulations on the data reported by Tanner and van Hell, one can estimate the likelihood of drawing an N400-only sample, a biphasic sample, and a P600-only sample, and moreover, investigate how the sample size impacts these likelihoods. As previously mentioned, the dataset from Tanner and van Hell is particularly well-suited for this purpose because it is large relative to many sample sizes in ERP studies, and contains a subset of participants who have been identified as showing classic N400 effects in response to morphosyntactic agreement violations.

Figure 4 shows the distribution of individuals' ERP effects from Tanner and van Hell, and this data served as the input for the Monte Carlo simulations. 100,000 random samples of individuals were drawn from the Tanner and van Hell data for each sample size between 5 and 40 individuals. In order to maximize variability and approximate the general population, sampling was performed with replacement. Sampling with replacement should additionally increase the likelihood of drawing N400-only samples. For each iteration of the simulation at each sample size, the sample's mean N400 and P600 effect magnitudes were computed, as were one-sample t-tests to test whether the sample's N400 and P600 effects differed reliably from zero. The overall proportion of significant samples for both the N400 and P600 time windows for each sample size is depicted in Figure 5. The upper panel shows the proportion of significant samples in the P600 time window (# significant samples/100,000). 90% power to detect a significant P600 effect was reached at a sample size of 10, and 99% power was reached at a sample size of 15 (solid regression line). The dashed regression line shows that in approximately 5% of samples, a P600 effect was detected in conjunction with an effect in the earlier 300-500 ms time window. In some cases this effect was positive-going, such that the sample showed an extended positivity; in other cases, this effect was negative-going, indicating a biphasic negative-positive response in the sample (see lower panel).

The lower panel depicts the proportion of significant effects in the 300-500 ms time window. As can be seen, across sample sizes, there was only approximately a 5% chance of detecting a significant effect (either positive- or negative-going). This falls in line with the nominal alpha level for statistical reliability used in most ERP studies (e.g., $p < .05$), such that reliable effects in this time window might be considered statistical noise. Most relevant to the current argument is the proportion of samples showing only an N400 effect (and no P600 effect, depicted by the long dashed regression line with crossed markers). Maximum power was reached at a sample size of 6, with a 2.66% chance of detecting an effect. Power dropped below 1% at a sample size of 11, and below .1% at a sample size of 18. At a fairly typical sample size of 24, power was only .005%.

[FIGURE 5 ABOUT HERE]

Thus, the hypothesis that some grand mean LAN effects are the result of component overlap between an N400 and P600 across individuals seems to be fully consistent with the lack of grand mean N400 effects seen across studies of agreement processing. Even though some individuals do show primarily N400 effects to subject-verb agreement violations, it is exceedingly difficult with any reasonable sample size to draw a group of individuals where the grand mean shows only an N400 effect, even when a known subsample of the population shows N400 effects with no P600 effects. Assuming that the variability present in the Tanner and van Hell data set is representative of variability in the broader population (or at least the population typically sampled in most published ERP studies, i.e., literate university undergraduates), the dominant pattern across most individuals is to show positivity-dominant effect (i.e., P600s). The dominance of P600 effects in most individuals, which can show an onset in some individuals in the same time window as the N400 effect, will swamp the central negativity seen in some individuals. This would lead to a pattern where the residual negativity is only apparent over scalp regions where the N400 and P600 effects show the weakest spatiotemporal intersection. As P600 effects tend to be posteriorly distributed (sometimes with a right hemisphere dominance), the residual negativity would only remain over left frontal electrodes. This would give the appearance of a LAN.

3 Summary and Discussion

Many ERP studies of agreement processing have reported a biphasic LAN-P600 response to violations of grammatical agreement. Because of this, the LAN features prominently in many neurocognitive theories of grammatical comprehension as an early index of syntactic processing, much like in the theory proposed by Molinaro and colleagues (e.g., Batterink & Neville, 2013; Friederici & Weissenborn, 2007; Friederici, 2002, 2011; Hagoort, Wassenaar, & Brown, 2003; Hagoort, 2003; Pakulak & Neville, 2010; Ullman, 2004). However, there is reason to be cautious in placing substantial theoretical weight on LAN effects. LAN effects in response to agreement violations tend to be elusive and not particularly replicable across studies, and P600 effects in response to morphosyntactic violations are often found without any preceding negativity. This makes it unlikely that the LAN indexes an early stage of anomaly detection (which would then give rise to a later stage of reanalysis, indexed by the P600), as held by these models. Importantly, the data presented here empirically show that 1) choice of left versus averaged mastoid references likely cannot solely be responsible for detectability of LAN effects, thus removing one putative source of variability in the LAN across studies, 2) that individual differences in brain responses between N400 and P600 effects can indeed give the illusion of LAN effects in the grand mean analysis, and 3) that given the dominance of P600 effects in response to morphosyntactic violations across individuals, it is highly unlikely to randomly draw a sample of individuals where only a reliable N400 would be found, with no following P600 – even though some individuals show negativity-dominant brain responses to morphosyntactic violations. Thus, an interesting degree of variability across studies in the presence, magnitude and scalp topography of negativities preceding the P600 will be at least partly a function of sampling variability when drawing from a population with individuals showing qualitatively different brain responses to morphosyntactic violations.

This last point must additionally be viewed in the context of component overlap in ERP studies more generally. Although the Tanner and van Hell data show that illusory LAN effects can arise by averaging across individuals showing qualitative differences in brain responses, it is also conceivable that biphasic LAN-P600 effects seen within single individuals may similarly reflect component overlap between a centralized N400 effect and the onset of a right-lateralized P600 effect, which onsets in the same time window. This possibility for the LAN in response to agreement violations was raised as early as Osterhout & Mobley (1995). In such cases of within-subject component overlap it may be impossible to know if the LAN is simply a residual N400, or if it reflects a functionally unique component. Even blind source separation techniques like Independent Components Analysis (ICA) may not be able to solve this problem, as ICA can misallocate variance and fail to correctly identify independent components whose time course and topography overlap (Makeig, Jung, Ghahremani, & Sejnowski, 2000). As late positivities can show extended temporal overlap with prolonged negativities when seen across individuals (Kos, van den Brink, & Hagoort, 2012; Tanner & van Hell, 2014), there is no way at the present time to recover the underlying component structures when these two effects co-occur within individuals.

Finally, it is worth pointing out that the above arguments should not be taken to claim that the LAN does not exist as a functional ERP component. Indeed, the most compelling arguments for the presence of a LAN effect seem to come from studies reporting a LAN with no following P600 in response to agreement violations (e.g., Coulson, King, & Kutas, 1998; Kim & Sikos, 2011; O'Rourke & van Petten, 2011; Ojima, Nakata, & Kakigi, 2005), or when processing long-distance dependencies that require storage of linguistic material in working memory (e.g., Fiebach, Schlesewsky, & Friederici, 2002; King & Kutas, 1995; Kluender & Kutas, 1993). The evidence and arguments presented here, however, should be taken as a cautionary note when interpreting biphasic ERP waveforms in response to linguistic violations. The evidence supporting the LAN as indexing a critical phase in a serial processing architecture proposed by Molinaro and colleagues (i.e., LAN ➔ early P600 ➔ late P600) requires further scrutiny. Future research taking into account individual differences – and component overlap, more generally – will surely help to disentangle some of these factors and further refine our understanding of the neural dynamics underlying processing of grammatical agreement.

Acknowledgments

I would like to thank Janet van Hell for her collaboration and Lee Osterhout for his comments on an earlier draft of this paper. Data collection was supported by NSF OISE-0968369.

References

Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience*, *33*, 8528–8533. doi:10.1523/JNEUROSCI.0618-13.2013

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *Journal of Memory and Language*, *47*, 250–272. doi:10.1016/S0749-596X(02)00004-9

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*, 78–84.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*, 1357–1392. doi:10.1152/physrev.00006.2011

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research*, *1146*, 50–58. doi:10.1016/j.brainres.2006.08.038

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage*, *20*, 18–29. doi:10.1016/j.neuroimage.2003.1796

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, *16*, 38–50.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225. doi:10.1016/j.jml.2004.10.002

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, *118*, 15–22. doi:10.1016/j.bandl.2011.03.002

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 376–395.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*, 196–214.

Kos, M., van den Brink, D., & Hagoort, P. (2012). Individual variation in the late positive complex to semantic anomalies. *Frontiers in Psychology*, *3*(318). doi:10.3389/fpsyg.2012.00318

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge: MIT Press.

Makeig, S., Jung, T.-P., Ghahremani, D., & Sejnowski, T. J. (2000). Independent component analysis of simulated ERP data. In T. Nakada (Ed.), *Integrated Human Brain Science* (pp. 123–146). Amsterdam: Elsevier Science.

Molinaro, N., Barber, H., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*, 908–930. doi:10.1016/j.cortex.2011.02.019

O'Rourke, P. L., & Van Petten, C. (2011). Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. *Brain Research*, *1392*, 62–79. doi:10.1016/j.brainres.2011.03.071

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of Proficiency. *Journal of Cognitive Neuroscience*, *17*, 1212–1228.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*, 494–522.

Osterhout, L., McLaughlin, J., Kim, A., Greewald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking-ERPs, and beyond* (pp. 271–308). New York: Psychology Press.

Osterhout, L., & Mobley, L. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739–773.

Otten, L. J., & Rugg, M. D. (2005). Interpreting Event-Related Brain Potentials. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 3–16). Cambridge, MA: MIT Press.

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*, 2728–2744.

Rugg, M. D., & Coles, M. G. H. (1995). The ERP and cognitive psychology: Conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 27–39). Oxford: Oxford Univ Press.

Severens, E., Jansma, B. M., & Hartsuiker, R. J. (2008). Morphophonological influences on the comprehension of subject-verb agreement: An ERP study. *Brain Research*, *1228*, 135–144. doi:10.1016/j.brainres.2008.05.092

Tanner, D., & van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289–301. doi:10.1016/j.neuropsychologia.2014.02.002

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.

Figure Captions

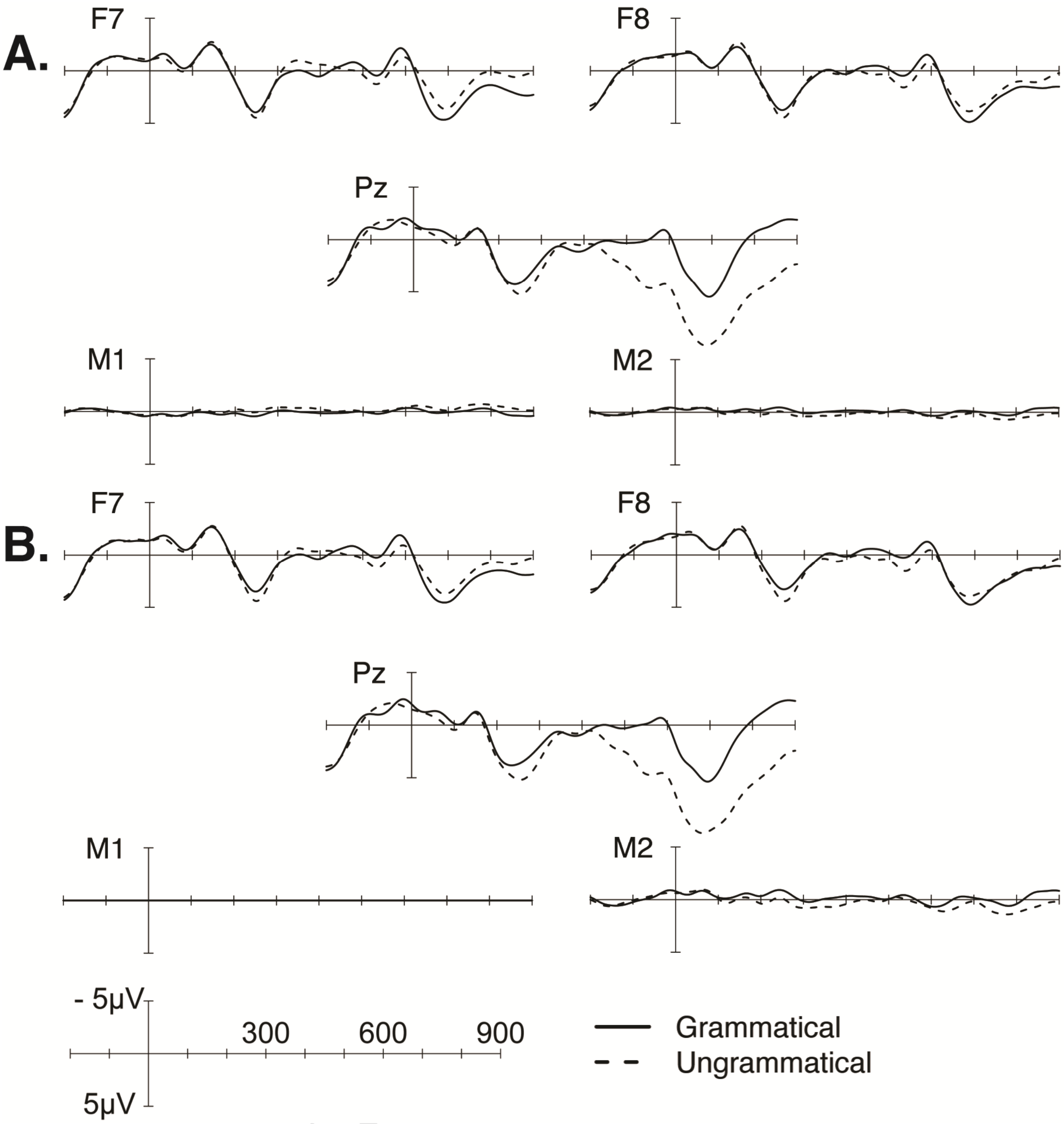
Figure 1. Grand mean ERPs depicting data in the subject-verb agreement condition from Tanner and van Hell (2014) using an averaged mastoid reference (Panel A) and a left mastoid reference (Panel B). Data are presented from left frontal (F7), right frontal (F8) and midline parietal (Pz) electrodes, as well as the left (M1) and right (M2) mastoids. The solid line depicts brain responses to grammatical verbs (*The winner of the big trophy is…*) and the dashed line depicts brain responses to verbs violating constraints on grammatical agreement (*The winner of the big trophy are…*). Onset of the verb is indicated by the vertical bar; ERPs were computed relative to a 200ms prestimulus baseline. The vertical calibration bar shows 5μV of activity; negative voltage is plotted up.

Figure 2. Grand mean ERPs depicting difference waves in the ungrammatical minus grammatical subject-verb agreement conditions from Tanner and van Hell (2014). Difference waves are shown for left frontal (F7), right frontal (F8) and midline parietal (Pz) electrodes, as well as the left (M1) and right (M2) mastoids. The solid line depicts difference waves using an averaged mastoid reference, whereas the dashed line depicts difference waves using a left mastoid reference. Onset of the critical verb is indicated by the vertical bar; ERPs were computed relative to a 200ms prestimulus baseline. The vertical calibration bar shows 5μV of activity; negative voltage is plotted up.

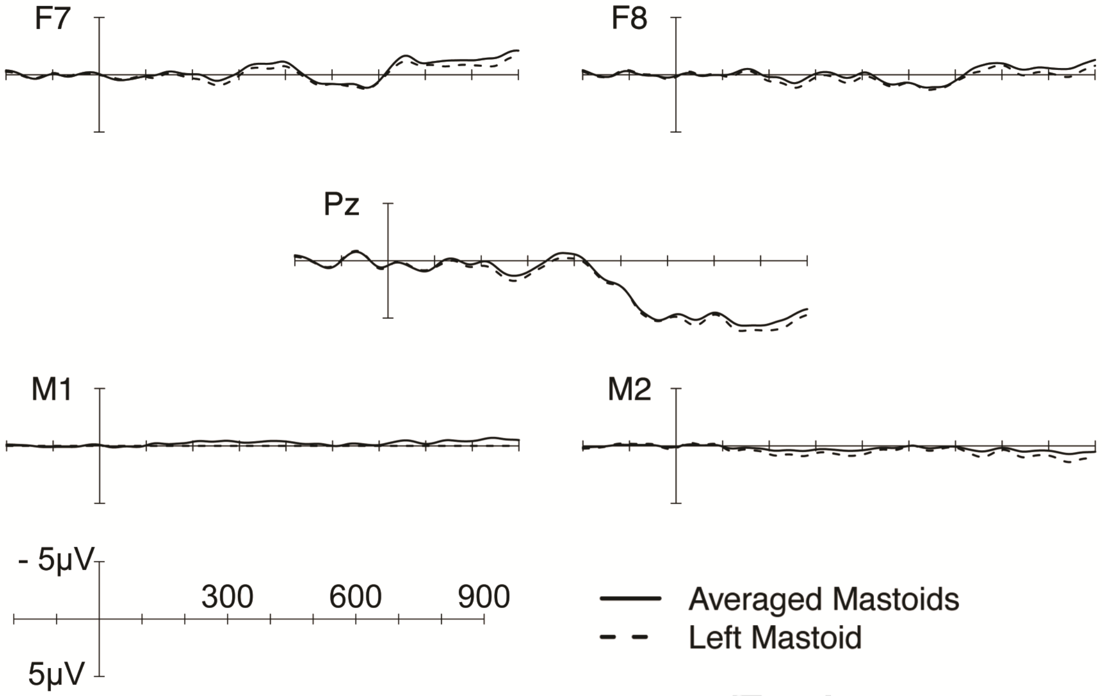
Figure 3. ERP effects from Tanner and van Hell (2014), depicting grand mean waveforms in the grammatical and ungrammatical subject-verb agreement conditions (left), ERPs from those showing negativity-dominant responses (center), and ERPs from those showing positivity-dominant responses (right). Midline vertex electrode Cz is shown. The solid line depicts brain responses to grammatical verbs (*The winner of the big trophy is…*) and the dashed line depicts brain responses to ungrammatical verbs (*The winner of the big trophy are…*). Onset of the critical verb is indicated by the vertical bar; ERPs were computed relative to a 200ms prestimulus baseline. The vertical calibration bar shows 5μV of activity; negative voltage is plotted up.

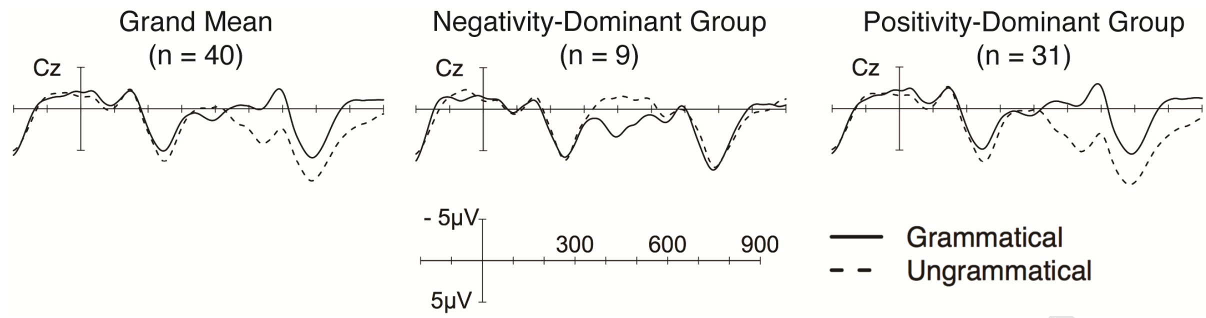
Figure 4. Distribution of N400 and P600 effects across individuals (n = 40) from the subject-verb agreement condition Tanner and van Hell (2014). N400 effect magnitudes are quantified as mean amplitude in the grammatical minus ungrammatical condition between 300 and 500 ms poststimulus over a centroparietal ROI (C3, Cz, C4, CP1, CP2, P3, Pz, P4); P600 effect magnitudes are quantified as mean amplitude in the ungrammatical minus grammatical condition between 500 and 800 ms poststimulus over the same ROI. N400 and P600 effect magnitudes are reliably negatively correlated across individuals ($r = -.589$, $p < .001$). The solid line depicts the best fit regression line; the dashed line depicts equal N400 and P600 effect sizes. Those above/to the left of the dashed line show primarily negative-going effects to agreement violations (n = 9); those below/to the right of the dashed line show primarily positive-going effects to agreement violations (n = 31).

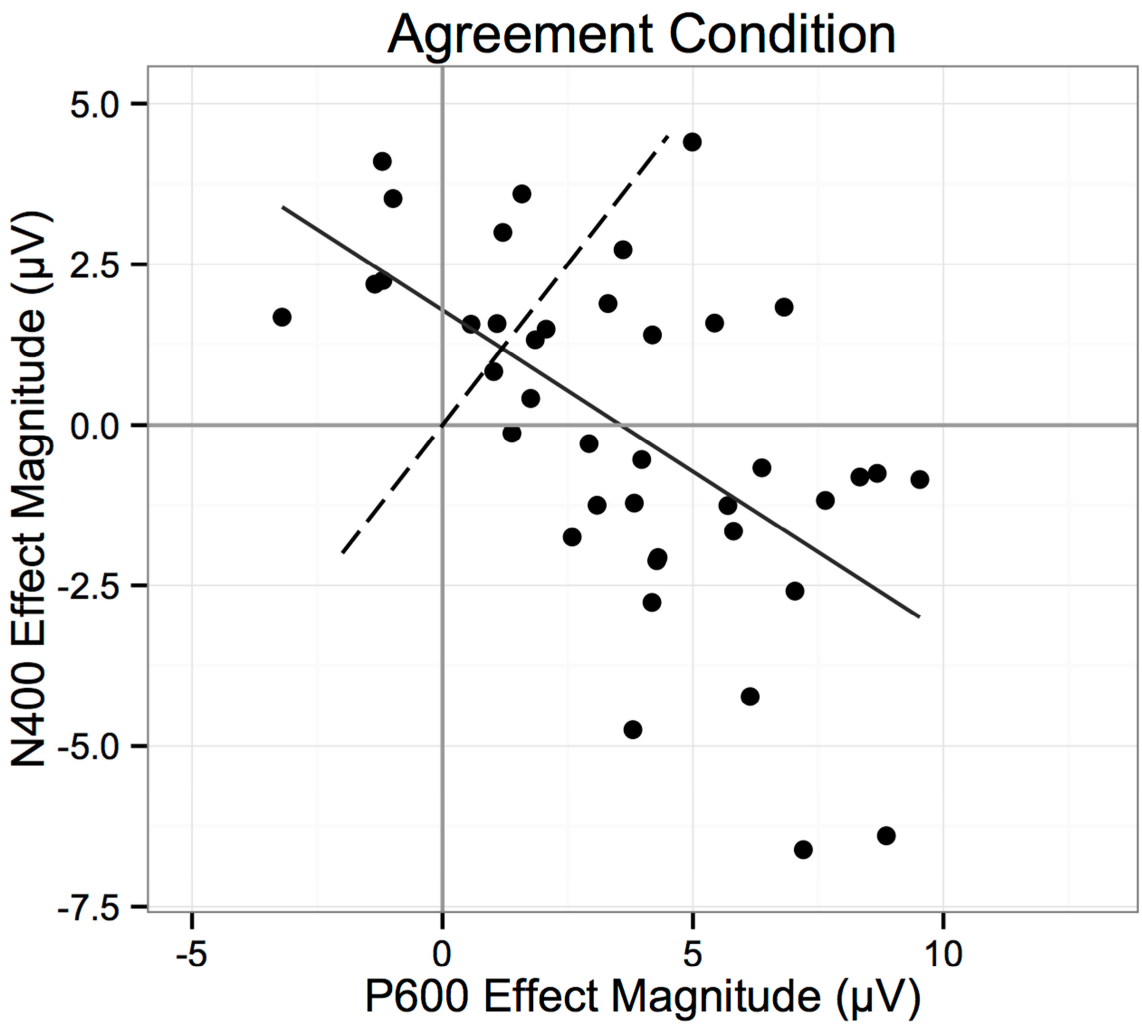
Figure 5. Results of Monte Carlo simulations depicting probability of detecting significant effects in the 300-500 ms time window (lower panel) and 500-800 ms time window (upper panel). Effects of sample size are shown along the x-axis. Data are derived from mean amplitude differences between the ungrammatical and grammatical subject-verb agreement conditions over a centro-parietal ROI (electrodes C3, Cz, C3, CP1, CP2, P3, Pz, P4), where N400 and P600 effects are largest. Smooth regression lines are fit by LOESS.



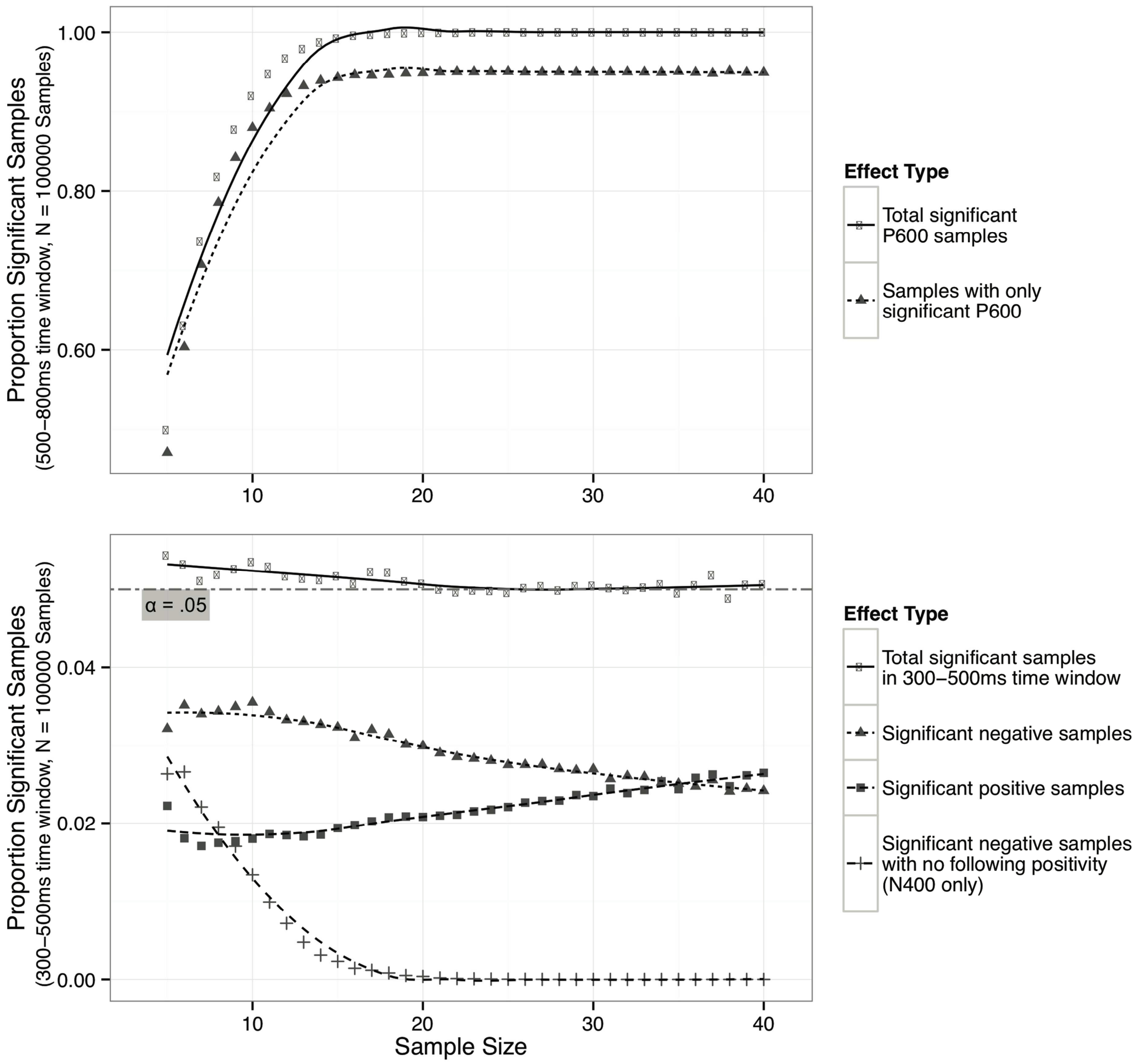
A.
F7
F8
Pz
M1
M2
B.
F7
F8
Pz
M1
M2
- 5μV
5μV
300
600
900
Grammatical
Ungrammatical


F7
F8
Pz
M1
M2
- 5μV
300
600
900
5μV
Averaged Mastoids
Left Mastoid


Grand Mean
(n = 40)
Cz
Negativity-Dominant Group
(n = 9)
Cz
Positivity-Dominant Group
(n = 31)
Cz
- 5μV
300
600
900
5μV
Grammatical
Ungrammatical



Agreement Condition
N400 Effect Magnitude (µV)
5.0
2.5
0.0
-2.5
-5.0
-7.5
-5
0
5
10
P600 Effect Magnitude (µV)


Proportion Significant Samples
(500–800ms time window, N = 100000 Samples)
1.00
0.80
0.60
10
20
30
40
Effect Type
Total significant
P600 samples
Samples with only
significant P600
Proportion Significant Samples
(300–500ms time window, N = 100000 Samples)
α = .05
0.04
0.02
0.00
10
20
30
40
Sample Size
Effect Type
Total significant samples
in 300–500ms time window
Significant negative samples
Significant positive samples
Significant negative samples
with no following positivity
(N400 only)