Quantified Sentences as a Window into Prediction and Priming: An ERP Study

Aniello De Santo Jonathan Rawski Amanda M. Yazdani John E. Drury
Stony Brook University

1 Introduction

This study used event related potentials (ERPs) to examine the processing of quantified sentences in an auditory/visual truth value judgment task. The primary aim of this pilot was to probe truth value and quantifier type influences on the N400 ERP response and, secondarily, to examine potential ERP markers of quantifier complexity.

The N400 is a well known ERP response thought to index processes involved in semantic long term memory (Kutas & Hillyard 1980; Lau *et al.* 2008). This effect has been argued to be a marker of access/retrieval (or binding) of conceptual semantic information stored in long term memory (Kutas & Federmeier 2011). It has also been suggested to index semantic integration (Hagoort 2003; Baggio & Hagoort 2011; Steinhauer *et al.* 2017) or inhibition (Debruille 2007). Though the initial discovery of the N400 arose in the context of reading experiments (Kutas & Hillyard 1980), subsequent research has revealed a wider family of effects that can be elicited across modalities and can be triggered in connection with a wide variety of different types of stimulus manipulations across cognitive domains (Kutas & Federmeier 2011; Cummings *et al.* 2006).

Understanding the etiology of this ERP response profile also requires examining which semantic manipulations the N400 may *not* be responsive too. In the context of language processing, the N400 has been argued to not be sensitive to all semantic content. For example, work going back to the early 1980s demonstrated that the N400 is insensitive to manipulations of truth value and negation in verification paradigms (Fischler *et al.* 1983; Kounios & Holcomb 1992). These studies, however, clearly show the N400 is modulated by subject/predicate relatedness (in 2x2 comparisons such as "*A robin IS/IS-NOT a ROCK/BIRD*", semantic priming effects in the form of N400 main effects only with amplitudes for ROCK (unrelated to the subject) > BIRD (related)). So in sentence contexts manipulating truth value, we find N400 modulations similar to what we would expect to find in simple word pair priming (e.g., ROBIN priming BIRD but not ROCK). Similar findings emerge when truth value depends on quantifiers (Kounios & Holcomb 1992), or when truth value of auditorily presented negative/affirmative statements is determined by visually presented picture stimuli (Lüdtke *et al.* 2008).

In contrast, Nieuwland & Kuperberg (2008) pointed out that the use of negation in the previously studied cases are pragmatically unnatural since they involve

*To Appear in the *Proceedings of CLS 54*

denial of statements that could not possibly be true (it would be a bizarre context indeed that would render an assertion that robins are not rocks pragmatically informative (Johnson-Laird & Tridgell 1972; Wason 1965; Glenberg *et al.* 1999)). Their study made an effort to control for this factor in an experiment crossing the contextual "un/naturalness" of negation with truth value and negative/affirmative manipulations (a 2x2x2 design). Their results suggest that in cases where negation is pragmatically natural, N400 amplitudes can indeed be modulated by truth value, with larger amplitudes for False than True sentences. However, in their pragmatically "unnatural" sentences, they failed to find a clear version of the semantic relatedness (priming) effects reported in previous work.

In the present pilot study targeting natural language quantifiers, questions about pragmatic unnaturalness do not arise. We used simple auditorily presented quantified sentences alongside visual presentation of arrays of colored geometric shapes where the sentences were either true or false of the co-presented picture (see Figure 1). In this way, our experiment pairs up the kind of study of quantifiers in Kounios & Holcomb (1992) with the cross-modal sentence/picture verification approach employed by Lüdtke *et al.* (2008). Thus we present a more straightforward way of examining the issue raised by Nieuwland & Kuperberg. If pragmatically natural uses of negation results in larger N400 responses for False than True sentences, we would clearly expect to find that result here, not the subject/predicate relatedness (priming) effects shown in all other studies.

Concerning our second aim, previous studies (McMillan *et al.* 2005; Troiani *et al.* 2009; Heim *et al.* 2012; Heim *et al.* 2016) demonstrated that additional working memory resources are recruited in the processing of proportional quantifiers (MOST) which, unlike the other quantifiers we test here (ALL/NONE/SOME) requires maintenance and comparisons of the cardinalities of sets to evaluate truth conditions (Pietroski *et al.* 2009; Lidz *et al.* 2011). However, the online time-course of complexity effects associated with MOST has not previously been investigated using ERPs, nor is it clear what ERP markers of sentence level language processing may be sensitive to such distinctions.

2 Methods

2.1 Overview

We presented quantified sentences auditorily to participants which simultaneously viewed arrays of colored shapes (Figure 1). Shape/color combinations were used with four quantifier types (ALL/NONE/MOST/SOME) to yield eight conditions varying quantifier/truth value. Visual stimulus items each consisted of fourteen colored shapes against a grey backdrop, with an even contrast ratio for ALL/NONE (7 yellow-circles/7 blue-squares) and opposing 2:5/5:2 ratios for MOST/SOME (e.g., 2 yellow-/5 blue-circles and 5 blue-/2 yellow-squares). False conditions used color/shape predicates which were not present in the images (unprimed — see below). These visual/auditory-pairs were presented in a pilot study to adult, native English-speakers (N=8) who provided (mis)match judgments following each trial. EEG was recorded continuously to examine a range of ERP measures time-locked both the onset of the predicate segments in these sentences and, as we outline below,

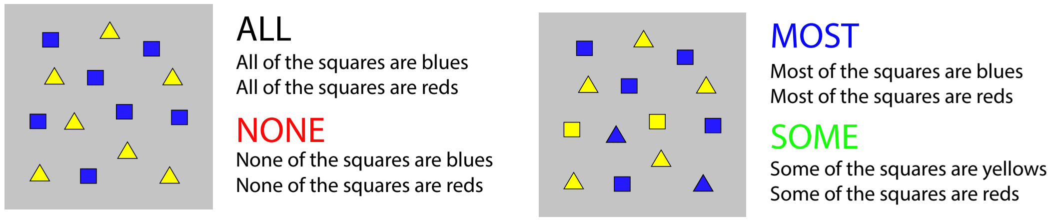

Figure 1: Experimental design and example stimuli. Sentences were presented auditorily while the pictures were displayed. Each visual item consisted of fourteen colored shapes against a grey backdrop, with an even contrast ratio for ALL/NONE (7 yellow-circles/7 blue-squares) and opposing 2:5/5:2 ratios for MOST/SOME (2 yellow-/5 blue-circles and 5 blue-/2 yellow-squares).

to the onset of the quantified subject segments as well.

2.2 Participants

Ten right handed (Oldfield 1971) adult native English speakers (5 male, 5 female; ages 18-26) with no history of cognitive impairment or neurological injury participated in this study after giving informed consent. Two participants' data had to be discarded due to insufficient number of trials remaining after artifact rejection procedures. The data reported below are thus based on 8 participants (4 male, 4 female).

2.3 Materials & Procedure.

The raw material for the auditory stimuli was recorded by the third author using a Shure SM-54 microphone and a Marantz digital recorder at 44000 Hertz. Our speaker was an adult male native English speaker (last author). Individual target sentences like those illustrated in Figure 1 were created by splicing out subject and predicate segments from the continuous recording of all sentences read in a prosodically and emotionally neutral tone of voice using Audacity(R) and Pratt (Boersma & Van Heuven 2001). First, a rough cut of each section, Subject and Predicate, was taken. Onset of the first portion of the Subject section was the beginning of periodicity or significant turbulence, and the offset was the end of the turbulent portion of the final "s". For the Predicate portion, onset was the beginning of periodicity of the vowel "a", and the offset was end of the turbulent portion of the final "s".

For each of four quantifiers (ALL, NONE, MOST, SOME), tokens were isolated using four colors (e.g., *All/None/Most/Some of the blues, reds, greens, yellows*) and four shapes (e.g., *All/None/Most/Some of the circles, squares, triangles, stars*). The same was done for the predicates (e.g., "... *are blues/greens/yellows/reds/circles/squares/triangles/stars*"). Using these four colors and four shapes across both subject and predicate segments yielded 32 unique sentences per quantifier (128 sentences total).

The visual portion of each trial was created in MS Powerpoint. The visual stimuli each involved 14 total colored shapes presented in a gray box in each picture, against a black background. Using a 75 by 75 pixel square as the guide, horizontally symmetrical shapes (circles, squares, triangles, stars) were created. Each had a black border which was 2 pixels thick and were created in 4 colors (blue, yellow, red, and green). Stimuli always involved two shapes, and each shape was always of a uniform color (e.g., 7 blue squares and 7 yellow circles; Figure 1).

Using the combination of pictures and auditory stimuli, each unique auditory sentence item appeared in either a True or False condition. All possible subject/predicate segments were rotated through all conditions, so that identical auditory stimuli entered into the averages for all conditions we examined.

Participants were instructed to verify whether the visual image rendered the sentence true or false by button press. The stimuli were presented in a sound attenuated booth, in 8 blocks. The Subject and Predicate audio files were presented with a 200 ms pause between them which provided naturalistic presentation. The visual stimuli disappeared with auditory stimulus offset, and participants were asked to respond with their truth value judgment during this interval. A fixation cross appeared for 1500 ms after each trial. The participants were not given feedback on their responses, and were given breaks after each block.

2.4 EEG recording and data preprocessing

EEG was continuously recorded via 32 cap mounted (active) electrodes (Biosemi Active 2 Amplifiers) with a 512 hz sampling rate and online bandpass filtering at 0-128 hz. Horizontal and vertical EOG was acquired via separate electrodes placed above and below the right eye and at the right/left canthi, and both left and right mastoid signals were also recorded. Raw EEG data were imported into the Matlab based platform EEGLAB (Delorme & Makeig 2004) with offline referencing to the average of the left and right mastoids. The data were then subjected to offline filtering (0.1 to 30 hz bandpass) before trigger codes were assigned to condition coded bins for averaging. Measurement epochs of 1200 ms (-200 ms to 0 ms baseline interval) time-locked to the onset of the sound files containing the critical predicate segments were extracted from the data. Automatic artifact rejection procedures eliminated trials with evidence of blinks, horizontal eye-movements, or muscle noise, resulting in removal of approximately 11% of the data. Remaining trials were averaged within our eight experimental conditions to create individual subject ERP average files, which were then subjected to statistical analysis. Grand average ERPs were generated and low-pass filtered at 7 hz for visualization purposes only (analyses were carried out over the unfiltered data).

2.5 ERP Analyses

ERP analyses were carried out for three consecutive 100 ms time-windows to examine N400 responses (250-350, 350-450, and 450-550 ms) with mean amplitude as the dependent measure. Given the appearance of a systematic unexpected early negative peak difference around 200 ms for some of our comparisons, we also conducted analyses for 50 ms windows centered around this peak (175-225 ms). Three later 100 ms time windows were also examined in the MOST/SOME conditions to probe late P600-like effects (650-750, 750-850, 850-950 ms). Repeated measures ANOVAs were conducted for each of these time-windows for four midline electrodes (Fz, Cz, Pz, and Oz). Supplementary analyses including broader averaged regions of interest (ROIs) were also carried out involving lateral recording sites in addition to the midline, but none of these differed qualitatively from the planned midline comparisons (below we simply report the midline). While the foregoing analyses were time-locked to the onset of the critical predicate segment, we also examined mean amplitudes for 100 ms time-windows between 350-950 ms time-locked to the onset of the quantified subject segment, in order to probe possible complexity effects related to an initial encoding stage of processing. These analyses compared visually matched MOST/SOME cases, and collapsed over True/False conditions (as this condition difference is only apparent downstream when the predicate is encountered).

In general we pursued a top-down analysis of variance, separately for conditions with qualitatively similar shape/color ratios in the visual stimuli. This yielded two sets of first level 2x2x2 ANOVAs with the factors truth value (TV; True vs. False) and electrode position (Anterior-Posterior/AP; Fz/Cz/Pz/Oz), one where the two levels of the Quantifier factor (Q) were ALL/NONE, and a second set where the levels of Quantifier (Q) were MOST/SOME. Interactions of Quantifier and truth value in these analyses were followed up with second level analyses within the particular Quantifier conditions separately.

3 Results and Discussion

3.1 Behavioral responses

Response accuracy on the behavioral task was at ceiling, with participants rejecting the False cases 100% of the time (no errors in any of the 8 participants), and only misjudging very few of the True cases (94% acceptance overall). Perfect performance on the False cases was unsurprising given that these conditions always mentioned a shape/color which was not represented in the visual display at all. Participants thus performed the task as expected, and the data show that this task was quite easy for participants. As our main interest was in the ERP patterns, and since there appeared to be no variability of interest in the accuracy data, we will not discuss the behavioral results further.

3.2 Event related potentials

Grand average ERPs for midline recording sites are shown in Figure 2, separately for each of the four quantifiers we tested. Difference waves at midline electrode Cz (False minus True) are superimposed for all four quantifiers in Figure 3. Note that the ALL and NONE N400 responses are in opposite directions (Figure 2 and 3). Finally, Figure 4 shows ERPs for all four quantifiers at electrode Pz, time-locked to the onset of the quantified subject segment. Again, note that these averages collapse over True/False conditions, as this distinction is only apparent once the subsequent predicate segment is identified.

3.2.1 *All* versus *None*

Two prominent patterns emerge in the ALL/NONE x T/F comparison: a prominent negativity for False relative to True sentences in the ALL comparison only (absent for NONE) peaking around 200 ms for the target predicate segments (black vs red traces in Figure 2 difference waves at Cz), and opposite direction N400 responses

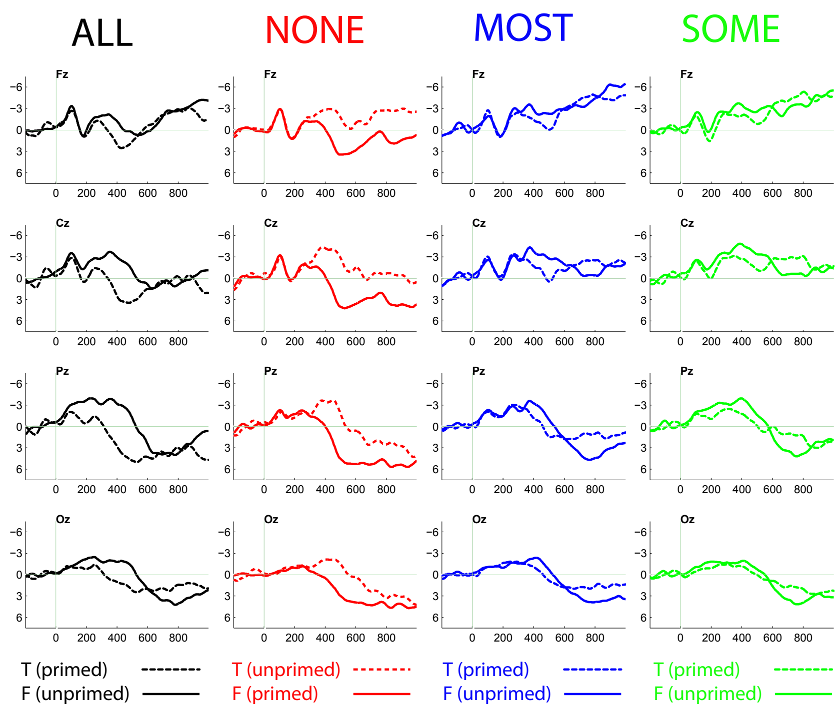

Figure 2: Grand average ERPs at midline recording sites for all conditions, time-locked to predicate onset.

(False > True for ALL and True > False for NONE; Figure 2). Due to the False minus True direction of subtraction in the Figure 3 difference waves, these N400 responses show up as mirror images (precisely the usual priming pattern shown in most previous research).

These patterns, evident from visual inspection of the data, were confirmed by our statistical analyses. First, analysis of the apparent N200 effect (175-225ms) demonstrated a significant three-way QUANTIFIER (Q) x TRUTH VALUE (TV) x ANTERIOR/POSTERIOR (AP) interaction $[F(4,28) = 2.84, p < 0.05]$, corresponding to the presence of a negative-going effect for False over True cases for ALL and the absence of TV effects for NONE in this latency range (visible in Figure 3; black vs. red difference waves).

As expected, follow-up analysis of TV effects for ALL revealed a significant TV x AP interaction $[F(3,21) = 5.76, p < 0.01]$, consistent with larger amplitude differences over more posterior midline sites, in particular Cz and Pz. There were no significant effects for NONE $[F's < 1]$.

In the N400 time-windows we also found significant Q x TV interactions between 350-550 ms $[350\text{-}450 : F(3,21) = 4.35, p = 0.02; 450\text{-}550 : F(3,21) = 11.44, p < 0.0001]$, though these interactions were not yet present in the first N400 time-window we analyzed $[250\text{-}350 : F(3,21) = 1.07, p = 0.38]$. Follow-up anal-

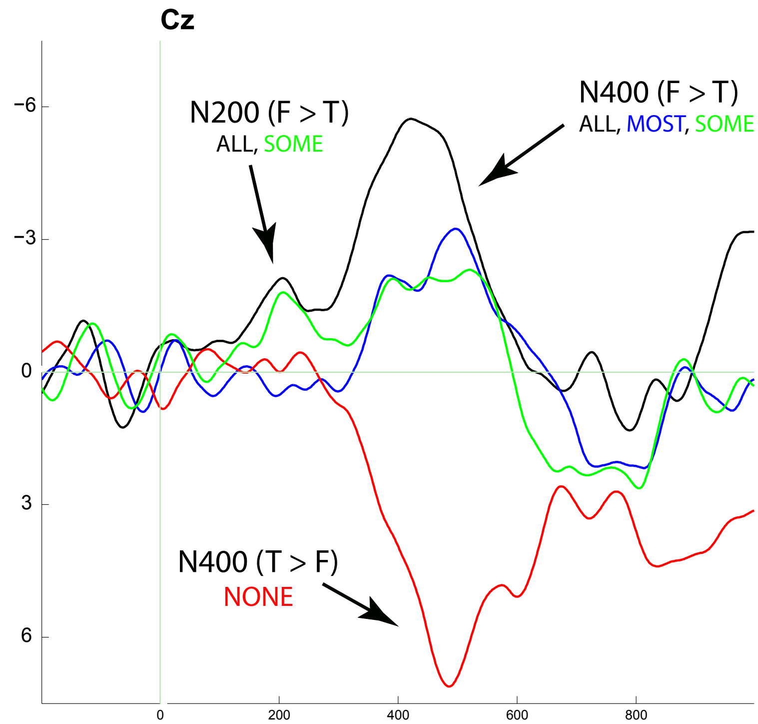

Figure 3: (False - True) difference waves at predicate onset

yses for ALL and NONE separately demonstrated significant relative negativity for False ALL sentences relative to True in the form of TV main effects in both time-windows $[350\text{-}450 : F(1,7) = 14.53, p = 0.01; 450\text{-}550 : F(1,7) = 9.35, p = 0.02]$ and TV x AP interactions $[350\text{-}450 : F(3,21) = 4.64, p = 0.01; 450\text{-}550 : F(3,21) = 9.89, p < 0.001]$, consistent with the usual N400 effect scalp distributions (largest at Cz and Pz). For the NONE cases, the opposite direction N400 response (True $>$ False) yielded a main effect of TV in the $350\text{-}450$ ms interval $[F(1,7) = 21.94, p < 0.01]$ while the subsequent interval showed a significant effect of TV $[450\text{-}550 : F(1,7) = 16.80, p < 0.01]$ and a TV x AP interaction $[450\text{-}550 : F(3,21) = 3.26, p < 0.05]$.

Also evident in the difference waves plotted in Figure 3 is an apparent peak latency difference in the N400 (priming) response between the ALL and NONE cases. Collapsing peak latencies in the difference waves across all midline electrodes we find that the ALL effect reached its maximum earlier than NONE by roughly 30 ms $[F(1,7) = 8.51, p = 0.02]$.

Finally, note that while in principle we would be interested in later time-windows for these conditions to investigate the nature of late (P600-like) positivities potentially indexing processing of False vs. True sentences — which indeed seem apparent in the data (e.g., see Figure 2 posterior electrodes) — such comparisons are confounded given the opposite direction priming effects, and would be difficult to interpret without additional control conditions. Thus we do not pursue these downstream analyses in this investigation (but see below for MOST/SOME comparisons in these later latency ranges).

3.2.2 ***Most* versus *Some***

MOST/SOME X TRUE/FALSE comparisons revealed an early pattern similar to that just shown above for ALL/NONE, namely a relative negativity for False SOME sentences compared to True which was absent for MOST, evident in a Q X TV interaction $[175\text{-}225 : F(1,7) = 6.46, p < 0.05]$. Follow-up analyses within the quantifier types revealed a marginal effect of TV for SOME $[F(1,7) = 4.50, p = 0.07]$, and no TV effect for MOST $[F < 1]$.

In contrast to this early Q X TV interaction, the N400 time-windows demonstrated a shared pattern for MOST and SOME, similar to the response for ALL, with relative negativities for False $>$ True evident in main effects of TV which did not arise until the classic N400 latency ranges between 350-550 ms $[250\text{-}350 : F < 1; 350\text{-}450 : F(1,7) = 12,21, p = 0.01; 450\text{-}550 : F(1,7) = 12.24, p = 0.01]$.

Finally, to investigate apparent late P600-like effects (False $>$ True) evident for both MOST and SOME (Figure 2), we examined three 100 ms time-windows between 650-950 ms. These analyses revealed a shared pattern with no interactions involving the factor Quantifier between 650-850 ms $[Fs < 1]$, only TV main effects $[650\text{-}750 : F(1,7) = 7.33, p = 0.03; 750\text{-}850 : F(1,7) = 6.26, p = 0.04]$. However, the final time-window revealed a Q X TV interaction $[F(3,21) = 4.63, p = 0.01]$ corresponding to the fact that the P600 for the MOST comparison sustained into this time window [TV X AP: $F(3,21) = 5.53, p = 0.006$] whereas the correspond effect for SOME did not [all $F's < 1$].

3.2.3 ERPs time-locked to the quantified subject segments

Figure 4 shows grand average ERPs at Pz for all four quantifier conditions time-locked to the onset of the subject/quantifier segment. Visual inspection reveals that MOST departs from the other conditions around the 300 ms mark, and that this difference sustains throughout the epoch. Comparison of the visually matched MOST and SOME cases showed this to be the case, yielding a significant main effect of Quantifier $[F(1,7) = 6.26, p = 0.04]$.

3.3 Summary of Findings

An unexpected early negativity shared by ALL and SOME and absent for MOST/NONE was found peaking around 200 ms after predicate onsets. ALL/NONE comparisons revealed opposite direction N400 responses tracking priming at the crucial predicate segments, and not truth value. These priming effects exhibited a peak latency difference, with the N400 for ALL comparisons reaching its maximum ~30 ms earlier than the NONE case. Smaller amplitude N400 responses in the same direction as the ALL conditions (False $>$ True) arose in the MOST/SOME cases. Late P600-like positivities (False $>$ True) emerged between 650-850 ms for both the SOME and MOST conditions, but sustained a further 100 ms (850-950 ms) for the MOST condition only. Finally, examination of ERPs time-locked to the quantified subject segment onset revealed a relative positivity for MOST compared to the

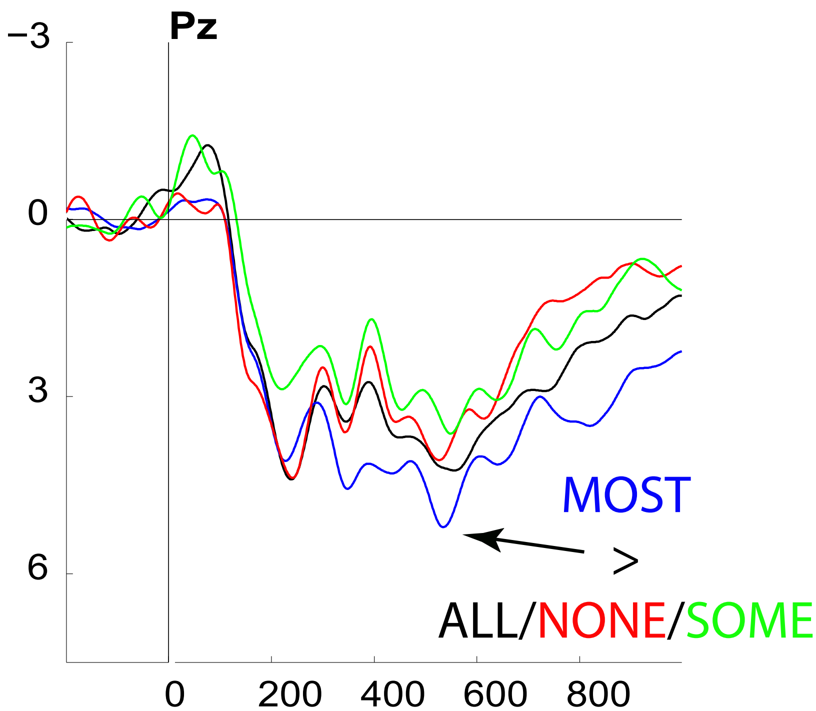

Figure 4: ERPs at quantifier onset

other conditions arising ~350 ms and sustaining throughout the measurement epoch (until at least 950 ms post-stimulus onset).

4 General Discussion

Concerning the primary aim of pilot study, we find that in contexts employing pragmatically natural uses of negation, true/false comparisons demonstrated that the N400 response at the critical predicate segment was driven by priming of the relevant color/shape displayed in the picture, relative to the targets picking out colors/shapes which were not present in the display. Moreover, the amplitude of the effects was not modulated by truth value, consistent with earlier findings in this domain (Fischler *et al.* 1983), including previous ERP reading experiments examining simple quantified sentences (Kounios & Holcomb 1992) and work using cross-modal paradigms targeting negation and truth value similar to the present experiment (Lüdtke *et al.* 2008). We also found an early response that has no correspondent in previous work, in the form of a relative negativity peaking ~200 ms for False (unprimed) ALL and SOME sentences and not MOST or NONE. We discuss these patterns (Section 4.1) before turning to our secondary aim, investigating quantifier complexity (Section 4.2). We believe these two parts of the story are related.

4.1 Priming and prediction in the processing of quantified sentences

It is tempting to relate the early (~200 ms) negativity for ALL/SOME to Phonological Mismatch Negativities (Connolly & Phillips 1994:PMMNs) or N200 effects (Van Den Brink *et al.* 2001). The early nature of these responses would then be consistent with some variety of mismatch between a potential phonological level prediction and the actual properties of the incoming speech stimuli. However, regardless of the specifics of how we might relate this finding to previously documented types of ERP response profiles, the question of why this effect should be shared by only the ALL/SOME cases raises some possibilities that we believe merit serious further inquiry. That this may be best viewed as a response tied to specific phonological level predictions, we believe, may fit together with other considerations about the cases we tested (see Nieuwland *et al.* (2018) for related discussion).

First consider the ALL/NONE comparison. At the point where participants have heard the end of the initial subject segment of our stimuli, it seems plausible that listeners are able to combine auditory and visual information in order to generate a unique prediction regarding the subsequent predicate segment. For example, considering our example stimuli in Figure 1, when listeners have heard "*All of the squares...*", the predicate segment is uniquely predicted to be "*...are blues*" given the visual array. In the corresponding negative (NONE) cases, a continuation with predicate segments rendering the sentence true could contain *any* of the other color words. Thus, the early negativity for the ALL cases could reflect a *phonological* mismatch in predicted (primed) versus unexpected false continuations. Since there is no such specific predicted continuation for the NONE cases, we might expect this response not to arise, though the ALL and NONE comparisons would still be expected to show opposing N400 effects reflecting *semantic* level processing. Interestingly, the N400 peak latency differences for ALL and NONE (~30 ms later for NONE) are consistent with this interpretation.

However, while this account provides a fairly sjtraightforward explanation for the ALL/NONE contrast, the question remains why this early effect should differentiate between the SOME and MOST cases. In order to account for this pattern, it would have to be that comprehension mechanisms deliver a *unique* prediction regarding the predicate segment for SOME, but not for MOST.

Consider again our sample stimuli in Figure 1. At the point where listeners have heard "*Some of the squares...*" where only 2 of the 7 squares are yellows, the question is why comprehension mechanisms would deliver a specific/unique prediction that the predicate segment should continue "*...are yellows*". Although the continuation "*...are blues*" in the case in Figure 1 would still make these SOME cases true, processing SOME might deliver a prediction based on the minimal set[1].

[1]Simply put, *yellow* — and not *blue* — gets uniquely predicted in 2-yellow/5-blue contexts because listeners reason that if MOST was true, the speaker would have explicitly said MOST, on *Gricean* grounds of informativeness (Grice 1975). Although previous results in this direction (Politzer-Ahles *et al.* 2013; Degen & Tanenhaus 2016) have focused on the distinction between *logical* SOME and *pragmatic* (*some-but-not-all*) SOME, the present findings seem to point toward a *some-but-not-most* implicature yielding sustained activation of the property corresponding to the minimal set. Thus, the rationale behind our suggestion is along well known work on scalar im-

However while this (post-hoc) explanation accounts for an early phonological prediction effect for SOME, it is not immediately obvious why this would not also hold for MOST. In fact, assuming as we did for the ALL case that comprehension mechanisms base predictions on the expectation of a continuation which renders the sentence True, we should expect the same effect form MOST, though in this case it would be the *majority* shape/color property represented in the picture driving the prediction.

Here we offer a way for the two aims of this study to intersect. Among the quantifiers we tested, only MOST requires that the relevant shape/color properties of both the minority and majority sets in our visual stimuli be kept active in order to compare their respective cardinalities (Lidz *et al.* 2011; Shikhare *et al.* 2015). Thus, if the early ~200 ms effect seen for ALL and SOME is due to priming tied to the prediction of a unique continuation[2], we might expect this effect *not* to arise for MOST if both properties present in the visual display (*yellow* & *blue* in Figure 1) remain activated, so that the corresponding cardinalities can be compared to judge truth value (but see Urbach *et al.* (2015) on the time-course of quantifiers' interpretations).

However, one might also expect — consistent with other findings in the literature (McMillan *et al.* 2005) — that we should see some independent index of processing cost associated only with MOST, and not with the SOME conditions. We discuss this briefly in the next section.

4.2 Complexity Effects for MOST

The idea of additional processing costs (e.g. due to working memory demands) associated to proportional quantifiers is supported both by recent neuroimaging (McMillan *et al.* 2005; Hackl 2009; Heim *et al.* 2012; Troiani *et al.* 2009) and behavioral results (Szymanik 2016; Zajenkowski & Szymanik 2013; Zajenkowski *et al.* 2013). As previous studies have not investigated the precise time-course of these events, we might in principle expect such *complexity effects* to arise either at the moment in which the quantifier is first encountered (the *encoding* stage), or at the point in which truth value can be evaluated relative to contextual information (e.g. at the predicate segment).

Interestingly, our own data suggest that complexity effects arise in both positions: ERPs time-locked to the onset of the quantifiers revealed a positivity for MOST > ALL/NONE/SOME, beginning ~350-450 ms and sustaining for ~500 ms (cf. Figure 4). This early positivity is consistent with complexity effects associated with initial encoding of MOST, reflecting the need for continued maintenance of the cardinalities for the contrasting sets.

An explanation for these early effects could also be tied to the idea that the interpretation of proportional quantifiers can be carried on through different verification

---

plicatures — a mechanism for SOME requiring the minimal set being to evoke a gradient scale: SOME > MOST > ALL (Van Tiel *et al.* 2016; de Carvalho *et al.* 2016). While we are aware that this study was not designed to test these effects directly, upon future replications of our results this paradigm could offer a direct way to probe the time-course of the processing of scalar implicatures that involves intermediate positions on the scale.

[2] Perhaps at the level of phonological expectations, consistent with this being an early response independent of the N400.

strategies. In particular, Pietroski *et al.* (2009) suggest that MOST can either be described in terms of a comparison relation between the cardinalities of two sets (*approximation of cardinalities* strategy), or in terms of a correspondence relation between the individual elements of those sets (*one-to-one with remainder* strategy). In a follow up study, Lidz *et al.* (2011) argued that the semantic representation of a quantifier (e.g. its *canonical specification*) plays an essential role in identifying the corresponding verification procedure. In this sense, the early effects measured at the onset of MOST could be due to the need to keep two possible verification strategies active.

While a precise account of early effects for proportional quantifiers is beyond the scope of this paper, these ideas are consistent with the presence of early response profiles for MOST, which suggests that we should not expect specific (phonological) predictions to cue early mismatches, as we found instead for ALL and SOME.

5 Conclusion

We presented a pilot study examining the processing of quantified sentences in an auditory/visual verification task to probe (i) truth value/quantifier type influences on the N400 ERP response and (ii) ERP markers of quantifier complexity. In pragmatically natural contexts N400s were driven by priming of the expected auditory continuation and were not modulated by truth value, consistent with earlier findings (Fischler *et al.* 1983). We relate early (~200 ms) negativities for ALL/SOME to PMMNs modulated by anticipatory effects tied to the truth-conditional properties of the quantifiers. Finally, our data suggest that complexity effects for MOST may reflect initial encoding, in addition to arising downstream during verification. While the present work was framed as a pilot (N=8) and we are thus aware of the limits intrinsic to our data, the magnitude of the effects we found both in the confirmatory (i) and exploratory (ii) parts of this study makes our findings particularly encouraging. In particular, to the best of our knowledge, the time-course of complexity effects associated with MOST has not previously been investigated using ERPs. Future experiments grounded in these results could help understand the different ways prediction and priming modulate ERP effects, and could focus on ways to specifically disentangle quantifier encoding from verification effects.

References

Baggio, G., & P. Hagoort. 2011. The balance between memory and unification in semantics: A dynamic account of the n400. *Language and Cognitive Processes* 26.1338–1367.

Boersma, P., & V. Van Heuven. 2001. Speak and unspeak with praat. *Glot International* 5.341–347.

Connolly, J. F., & N. A. Phillips. 1994. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience* 6.256–266.

Cummings, A., R. Čeponienė, A. Koyama, A. P. Saygin, J. Townsend, & F. Dick. 2006. Auditory semantic networks for words and natural sounds. *Brain research* 1115.92–107.

de Carvalho, A., A. C. Reboul, J.-B. Van der Henst, A. Cheylus, & T. Nazir. 2016. Scalar implicatures: The psychological reality of scales. *Frontiers in Psychology* 7.1500.

Debruille, J. B. 2007. The n400 potential could index a semantic inhibition. *Brain Research Reviews* 56.472–477.
Degen, J., & M. K. Tanenhaus. 2016. Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science* 40.172–201.
Delorme, A., & S. Makeig. 2004. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods* 134.9–21.
Fischler, I., P. A. Bloom, D. G. Childers, S. E. Roucos, & N. W. Perry Jr. 1983. Brain potentials related to stages of sentence verification. *Psychophysiology* 20.400–409.
Glenberg, A. M., D. A. Robertson, J. L. Jansen, & M. C. Johnson-Glenberg. 1999. Not propositions. *Cognitive Systems Research* 1.19–33.
Grice, H. P. 1975. Logic and conversation. 41–58.
Hackl, M. 2009. On the grammar and processing of proportional quantifiers: most versus more than half. *Natural Language Semantics* 17.63–98.
Hagoort, P. 2003. Interplay between syntax and semantics during sentence comprehension: Erp effects of combining syntactic and semantic violations. *Journal of cognitive neuroscience* 15.883–899.
Heim, S., K. Amunts, D. Drai, S. Eickhoff, S. Hautvast, & Y. Grodzinsky. 2012. The language number interface in the brain: A complex parametric study of quantifiers and quantities. *Frontiers in Evolutionary Neuroscience* 4.4.
Heim, S., C. T. McMillan, R. Clark, L. Baehr, K. Ternes, C. Olm, N. E. Min, & M. Grossman. 2016. How the brain learns how few are "many": An fmri study of the flexibility of quantifier semantics. *NeuroImage* 125.45–52.
Johnson-Laird, P. N., & J. Tridgell. 1972. When negation is easier than affirmation. *The Quarterly Journal of Experimental Psychology* 24.87–91.
Kounios, J., & P. J. Holcomb. 1992. Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General* 121.459.
Kutas, M., & K. D. Federmeier. 2011. Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (erp). *Annual review of psychology* 62.621–647.
Kutas, M., & S. A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207.203–205.
Lau, E. F., C. Phillips, & D. Poeppel. 2008. A cortical network for semantics:(de) constructing the n400. *Nature Reviews Neuroscience* 9.920.
Lidz, J., P. Pietroski, J. Halberda, & T. Hunter. 2011. Interface transparency and the psychosemantics of most. *Natural Language Semantics* 19.227–256.
Lüdtke, J., C. K. Friedrich, M. De Filippis, & B. Kaup. 2008. Event-related potential correlates of negation in a sentence–picture verification paradigm. *Journal of cognitive neuroscience* 20.1355–1370.
McMillan, C. T., R. Clark, P. Moore, C. Devita, & M. Grossman. 2005. Neural basis for generalized quantifier comprehension. *Neuropsychologia* 43.1729–1737.
Nieuwland, M. S., & G. R. Kuperberg. 2008. When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science* 19.1213–1218.
Nieuwland, M. S., S. Politzer-Ahles, E. Heyselaar, K. Segaert, E. Darley, N. Kazanina, S. V. G. Zu Wolfsthurn, F. Bartolozzi, V. Kogan, A. Ito, *et al.*. 2018. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife* 7.e33468.
Oldfield, R. C. 1971. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9.97–113.
Pietroski, P., J. Lidz, T. Hunter, & J. Halberda. 2009. The meaning of most: Semantics, numerosity and psychology. *Mind & Language* 24.554–585.
Politzer-Ahles, S., R. Fiorentino, X. Jiang, & X. Zhou. 2013. Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Research* 1490.134 – 152.
Shikhare, S., S. Heim, E. Klein, S. Huber, & K. Willmes. 2015. Processing of numerical and proportional quantifiers. *Cognitive Science* 39.1504–1536.

Steinhauer, K., P. Royle, J. E. Drury, & L. A. Fromont. 2017. The priming of priming: Evidence that the n400 reflects context-dependent post-retrieval word integration in working memory. *Neuroscience letters* 651.192–197.

Szymanik, J. 2016. Computing simple quantifiers. In *Quantifiers and Cognition: Logical and Computational Perspectives*, 41–49. Springer International Publishing.

Troiani, V., J. E. Peelle, R. Clark, & M. Grossman. 2009. Is it logical to count on quantifiers? dissociable neural networks underlying numerical and logical quantifiers. *Neuropsychologia* 47.104–111.

Urbach, T. P., K. A. DeLong, & M. Kutas. 2015. Quantifiers are incrementally interpreted in context, more than less. *Journal of memory and language* 83.79–96.

Van Den Brink, D., C. M. Brown, & P. Hagoort. 2001. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus n400 effects. *Journal of cognitive neuroscience* 13.967–985.

Van Tiel, B., E. Van Miltenburg, N. Zevakhina, & B. Geurts. 2016. Scalar diversity. *Journal of Semantics* 33.137–175.

Wason, P. C. 1965. The contexts of plausible denial. *Journal of Memory and Language* 4.7.

Zajenkowski, M., & J. Szymanik. 2013. Most intelligent people are accurate and some fast people are intelligent.: Intelligence, working memory, and semantic processing of quantifiers from a computational perspective. *Intelligence* 41.456–466.

Zajenkowski, M., J. Szymanik, & M. Garraffa. 2013. Working memory mechanism in proportional quantifier verification. *Journal of Psycholinguistic Research* .