Reviewers' comments:

Reviewer #1: There is an enormous unrealized potential for obtaining new insights by applying ERP methods to the complexities of the cognitive literature on lexical access so it is always a pleasure to see efforts at doing so.  This experiment examined the N400 effects arising from associative and semantic similarity priming in the context of letter search and lexical decision tasks.  While the experimenters made some good choices regarding the basic manipulations and the articles that they referred to, there are some serious concerns with the reasoning behind the experiment and the manner in which it was carried out.  My general impression was that of talented but green researchers who can hopefully benefit from the following comments and go on to make important contributions to this field.

Major Issues

p. 6 "The next step in this line of research would be to investigate the effect, if any, of the type of relatedness shared between the prime and target words."

This transition sentence illustrates a general absence of a theoretical argument.  Why this particular direction rather than any of a million other directions?  There is an infinity of possible parameters to look at.  There is indeed a deep and long-standing line of research on the issue of associative vs. semantic relatedness but it arises from fundamental questions about the nature of semantic representations in semantic memory rather than an arbitrary question about the effects of primes on targets in lexical decision tasks.  The introduction needs to be entirely rewritten to describe the research program in terms of theoretical questions to be answered rather than just being the investigation of arbitrary manipulations.  Research like this is conducted to answer basic theoretical questions.  Nobody cares about the infinite varieties of lexical decision paradigms per se or what manipulations affect the N400, except insofar as it sheds light on these basic questions.

For example, the following sentence:

p. 7 "Both Lucas (2000) and Hutchison (2003) have examined "semantic" priming studies and found that many experimental stimuli are a mix of both semantic and associative relatedness."

is an example of how this manuscript entirely misses the conceptual aspects of this literature.  These two reviews did not simply find "that many experimental stimuli are a mix of both semantic and associative relatedness," rather Lucas (2000) argued that relations are primarily based on semantics with just a boosting effect for associations whereas Hutchison (2003) disagreed and argued instead that there is indeed evidence for associative priming.  There is an argument going on between these two authors and in order to contribute to this literature, the present manuscript needs to address these controversies and seek to shed some light on them.  There is such an opportunity here to contribute to this literature, it's a shame to waste it!

p. 7 "Maki, McKinley, and Thompson (2004) took the online dictionary, WordNet (Fellbaum, 1998), and used software by Patwardhan and Pederson (2003) to create a database of words displaying the semantic distance between individual words."

These kinds of latent semantic distance measures may not be the same as semantic feature overlap.  For example, Maki and colleagues subsequently suggested that all three types are tapping different types of relatedness:

Maki, W. S., & Buchanan, E. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. Psychon Bull Rev, 15(3), 598-603.

At any rate, as part of presenting a theoretically driven argument, there would be a need to justify a given measure as being an appropriate operationalization of the construct of interest and to justify not using the alternatives.  It's not enough to say that a measure exists and so you're going to use it.

p. 10 "They were found to be the same across data type using a mixed ANOVA, F(15, 1310) = 1.217, p = .251."

It's not clear to me how this was done from just this ANOVA. The way this is normally done is you compare the related to the unrelated items and do separate t-tests on each parameter and then (ideally) you report that each such test was non-significant (best to be non-sigificant at p > .10 so they're non-significant even for a one-tailed test).  Also, technically you can't conclude from a statistical test that two things are the same, you can only conclude that they weren't reliably different.  Also, it's best to use an approach where there are counterbalance groups such that initially all the items are related pairs and then for a given counterbalance group half the pairs are repaired so that they are now in unrelated pairs.  This strategy ensures that there are no confounds overall between the related and unrelated items as all the words end up being in both the related and the unrelated conditions once you've pooled all the counterbalanced subjects together.  As it is,

this experiment has not ruled out the possibility of confounds, although the list of parameters tested for was a good start.

p. 11 The methods section needs to provide the full description of the task.  It's not enough to refer the reader to the 2001 paper, especially since it's described as being a "modified" version.  First of all, it's not appropriate to require the reader to have to dig up the other paper just to find out what the experiment was about.  Second, the reader can't evaluate the study without knowing what the critical parameters were.  For example, without being sure what the SOA was between the prime and the target, I can't evaluate the likely effect of the priming manipulation.

p. 12 The predictions are cursory and poorly motivated.  That related words produce smaller N400 effects than unrelated words is very very well-known.  Why is the effect of associated pairs not predicted whereas the effect of semantically related pairs is predicted?  There is plenty in the literature to form such predictions with in both cases, starting with the ones cited in the introduction.  There have also been prior efforts to contrast associative vs. semantic similarity priming in the literature and so a first step would be to find them using pubmed searches.  Deacon for example.

Minor Issues

p. 5 "The N400 has been described as a "contextual integration process", in which meanings of words are integrated together (Silva-Pereya, Rivera-Gaxiola, Aubert, Bosch, Galan, & Salazar, 2003)."

This is an odd choice for a citation regarding the nature of the process reflected by the N400.  This paper itself cites Kutas and Federmeier (2000) for this view.  Silva-Pereya and colleagues themselves suggest that the "N400" reflects the summation of what are essentially different ERP components each reflecting different cognitive processes (such as expectancy, ASA, and post-lexical semantic matching), so these authors are actually opponents of the unitary N400 semantic integration camp, although they soft-pedal their position.

p. 6 "Rolke, Heil, Streb, and Hennghausen (2001) used the attention blink rapid serial visual presentation (RSVP) paradigm, which can be compared to masked priming, and showed automatic activation of semantic information even when targets were missed or "blinked"."

Should put this into a separate paragraph and then explain what the attentional blink paradigm is.

p. 6 "Stolz and Besner (1999) stipulate that this eliminated or reduced priming indicates non-automatic semantic priming."

Any discussion of the letter search task effects should also note the opposing position of Neely and associates, as in:

Tse, C. S., & Neely, J. H. (2007). Semantic priming from letter-searched primes occurs for low- but not high-frequency targets: automatic semantic access may not be a myth. J Exp Psychol Learn Mem Cogn, 33(6), 1143-1161.

p. 8 mean age?

p. 8 compensation?

p. 8 which University?

p. 8 It's not clear to me what EEG system was used.  There is a reference to using sponge electrodes but the only such commercial system that I am aware of is the EGI system but that's not typically used with Neuroscan hardward and software and the description of the EEG application is not consistent with EGI systems so I'm very confused.

pp. 8-9 please keep consistent past tense

p. 9 Need to compute the relatedness proportion and the non-word ratios.  For clear instructions on how to do so, see:

Dien, J., Franklin, M., & May, C. (2006). Is "blank" a suitable neutral prime for event-related potential experiments? Brain and Language, 97, 91-101.

Note that even an RP of .25 can still result in expectancy priming:

Hutchison, K. A., Neely, J. H., & Johnson, J. D. (2001). With great expectations, can two "wrongs" prime a "right"? Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(6), 1451-1463.

p. 10 "For non-word pairs, the target word had a letter changed so that it was no longer a real word, but the structure was left intact to require that the participant process the word cognitively (PUND)."

It's not clear to me what this sentence means.  I'm guessing the authors are trying to say that the non-words were orthographically regular pronounceable non-words?

"The program automatically scanned for and removed muscular artifacts caused by eye blinks."

Eyeblinks don't cause muscular artifacts.  The artifact created by eyeblinks is of a different sort entirely (created by the movement of the eyelids across the surface of the eye, producing charge separation much like that from rubbing a balloon across car fur).

p. 12 What about baseline correction?  Low-pass filtering?

Copy Editing

pp. 4-5 "Lexical decision tasks have been criticized for their inability to distinguish between automatic and controlled processing, so both SINGLE PRESENTATION LEXICAL DECISION TASKS AND MASKED PRIMING MANIPULATIONS have been introduced to negate controlled processing (Ford, 1983)."

p. 6 "Letter search tasks have also been usedto reduce" is missing a space.

p. 8 "It was expected that the N400 modulation might vary from the different types of word relation, which would indicate differences cognitive processing and word organization."

not grammatical

p. 10 The description of the procedure is way too long.  Everything other than the description of the number of blocks and the site of the reference electrodes should be deleted.

p. 10 " Next two ground sensors (baseline scalp electroconductivity without underlying brain activity) were placed on the right and left mastoid bones, or the slightly protruding bones just behind each ear."

These are "reference channels" not ground, which is something else entirely.  Also, reference channels define zero voltage, not " baseline scalp electroconductivity without underlying brain activity".  See:

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. Behavior Research Methods, Instruments, and Computers, 30(1), 34-43.

p. 11 It's not enough to provide the size of the stimuli.  It's also necessary to know the distance of the viewer from the screen so that the visual angle can be computed.  After all, a word presented in 16 poin Arial will be very easy to read from one foot away and be impossible to read from one mile away.  Best to just compute the visual angle and present it rather than to make the reader compute it though.

p. 12 " FZ, FCZ, CZ, CPZ, and CZ were used from the midline"

CZ is repeated.  I think you meant Pz?  Also, the "z" is normally noted in lowercase, as in "Cz"

Reviewer #2: Reviewer #2: This paper addresses the question of whether there is a difference in N400 amplitudes between semantic and associative word pairs in lexical decision and letter search tasks. The authors present an experiment in which subjects were asked to make either an orthographic (repeated letters vs. no repeated letters) or a lexical (word vs. non-word) judgment. Some words were preceded by unrelated words, by non-words or by words which were related either semantically or associatively.

Overall, I believe that the issues addressed in this present paper are one of relevance to researchers interested in contextual language. It has the potential to address the boundaries of "context"-how much do idiosyncratic associations (such as ALIEN to PREDATOR)--which share no traditional semantic relatedness yet are pervasive in culture-bias speaker judgment? However, I have a number of concerns about this paper which prevent me from accepting this paper in its present form. I will outline these concerns in the following paragraphs.

First, the introduction is cursory and incomplete. While I do feel I understand the nature of associative vs. semantic relationships, I do not understand from this paper why an ERP paradigm is the best means of addressing this concern. Their treatment of the nature of semantic spreading (especially from the computational literature) as well as their rationale for using the N400 as a dependent measure is absent or unclear. This is not to say that they cannot be found, merely that works cited are dated and nonspecific and papers of the last decade -e.g. Kutas & Federmeier, 2000--- which address factors affected the N400 in paradigms outside of traditional sentence-based paradigms are absent.  Without proper context, any findings are harder to place in the literature.

Method

Although all information about procedure, tasks, dependent measure etc. is somewhere in the method section, I found it hard to keep up with all the details. Excessive time was given to universal details of EEG procedure, while descriptions of the task itself (e.g. interstimulus intervals, etc.) appeared not to be present. A figure which displays a full block for the letter-search task and the lexical decision task is needed. Also, more details about trial analyses are needed. For instance, were ERPs time-locked to the beginning or end of the word? Further, it appears that the design only allowed for 30 trials of interest in each of the semantic and associative conditions. The authors said that 90 percent of the data was retained after artifact correction; however this number rather high (loss rates of 20-30 percent being typical with older EEG systems). It seems as though many of the nonsignificant effects the authors found may be due to issues of low power.

Results

The fundamental hypotheses upon which analyses were undertaken were too scattered and poorly justified as to be intellectually satisfying. The authors assert that N400 attenuation reflects automatic processing, while more negative-going amplitudes reflect a need to search the mental lexicon. No justification for these assumptions is given. This is the heart of the argument, and if the authors do not provide evidence as to why this core assumption should be believed, the results themselves are suspect. The literature suggests that in lexical decisions tasks, unrelated words show a greater N400 than related words. This predicts to the author that both semantic and associated primes would show attenuated amplitude. However, it is unclear why this reflects "automated" processing vs. "controlled lexical search". Further, it appears the authors did not directly test amplitude differences between semantic and associative primes. Also, no methodological details were given about how

N400 amplitude was determined (e.g. by taking isolated peak value, by an average of nearest points, etc.).

Discussion

The authors again used "controlled' and "automatic" without specification.

Further, the authors state that gender differences for males were "focused" in traditional Broca's areas. This is an exaggerated claim given the types of analyses done by the authors.

Again, the authors show a rather limited knowledge of the N400 literature.

I regret that I cannot recommend this paper for publication at this time. While the subject matter is of interest, the paper itself does not adequately address the problems they themselves have laid out.

Reviewer #3: The authors present a study using ERPs to look at priming for semantic and associative relationships in a lexical decision and  letter search task.  Few studies have directly compared these types of word level relationships with ERPs, and none to my knowledge have done so across different tasks.  Thus the data from this study can make a useful contribution to the ERP literature on N400 sensitivities.  However, there are a number of problems with the current manuscript, which I first summarize and then describe in more detail.  At the highest level, it is not clear exactly what cognitive or neural question the authors are trying to answer and in the end they provide neither a cognitive nor a neural account of what their data say.  Most of their literature review seems focused on automaticity, but the relationship between the N400 and automaticity (and, indeed, the definition of automatic

itself) is controversial (the authors might wish to take a look at the review by Deacon & Shelley-Tremblay 2000), and the authors' manipulations and findings do not address the heart of the controversy and thus cannot

contribute notably to that debate.  Instead, the authors' data seems better placed to speak to similarities and differences in processing types of semantic relations and their generality across task conditions, but the authors do not provide much background related to these questions.  Indeed, overall, the manuscript is not well-integrated with the current literature on the N400, including higher-level points such as what the N400 is thought to index and more specific details of what a fairly extensive literature on  word priming has already revealed about semantic and associative priming.  In addition, the manuscript fails to provide appropriate details about ERP methods.  Finally, because the authors do not really talk about the brain at all, it is not clear that this manuscript in its current form is at all appropriate for Brain and Language.

Specific comments:

The authors write "The N400 has been described as a "contextual integration process", in which meanings of words are integrated together (Silva-Pereya, Rivera-Gaxiola, Aubert, Bosch, Galan, & Salazar, 2003).

When presented with related words, there is an attenuation of the N400; meaning a more positive waveform when compared to unrelated word presentation. This difference in waveforms

indicates a lessened contextual integration process because word meanings are already activated."

However, there are actually a number of theories of what the N400 indexes, which would say different things about what the facilitation in semantic priming indexes.  In particular, the view that the authors describe

later "Semantic word pairs will have significantly positive values because priming will decrease the need to search the mental lexicon." seems more consistent with the (fairly prevalent) view that the N400 indexes initial contact with semantic memory (see, e.g., Kutas and Federmeier, 2011).  Regardless of the authors' particular take on the literature, they should make clear what the major views are and how each might explain/predict their data patterns.

The authors do not discuss ERP literature looking at associative versus semantic priming.  A number of different types of priming paradigms have been used across studies, some concentrating on primarily

associative relations and others on primarily semantic.  Few have directly contrasted the two in the same study -- making this a novel contribution -- but it would seem appropriate for the authors

to discuss this literature and what the cross-study comparisons would suggest.  Moreover, some studies HAVE contrasted the two, including a study by Deacon (2004) that claims that there are hemispheric

differences in lexico-semantic representation revealed by comparing associative and semantic priming in a visual half-field design.  This background would provide stronger links between the authors' data

and the brain, since, as already mentioned, in the current paper the contribution of the data to our understanding of brain processes seems limited to a degree that raises the question of whether Brain and Language is the appropriate journal.

Authors should consult the guidelines for publishing ERP papers ( to see what information is critical to include (including electrode arrangement, filter settings during and after recording, sampling rate, etc.):  Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria  [208 KB]

T.W. Picton, S. Bentin, P. Berg, E. Donchin, S.A. Hillyard, R. Johnson, Jr., G.A. Miller, W. Ritter, D.S. Ruchkin, M.D. Rugg, and M.J. Taylor

Psychophysiology, 37 (2000), 127-152.

Some descriptions of methods suggest poor understanding of ERP methods.  For example, it is not the case that electrodes measure only electrical activity "just below the scalp" as electrical signals from distant locations are volume conducted (indeed, a primary source for the N400 measured at, e.g., Cz, comes from temporal cortical areas that are notably distant from the sensors).  Similarly, reference (not ground) electrodes placed on the mastoid -- or indeed anywhere on the head -- DO measure brain activity, which is subtracted.  The authors may wish to consult Luck's excellent book on ERP methods to make sure they are describing the method

accurately.

The authors write that "Five sites were chosen to examine priming for nonwords, associative and semantic word pairs based on a survey of the literature. FZ, FCZ, CZ, CPZ, and CZ".  What literature was surveyed and how does it point to these sites? N400s are not usually prominent at Fz.  The electrode choice needs to be better clarified.

The authors predictions need to be better rooted in the literature and to take into account theories about the N400.  For example, the authors state:

"Non-word pairs may show significantly more negative waveforms (more negative area) due to the need to search the lexicon before a decision can be made."

There are many studies comparing unrelated words to nonwords with ERPs.  Many find no differences in amplitude (when neighborhood size is controlled).  The authors should bring that background into the paper.

"Semantic word pairs will have significantly positive values because priming will decrease the need to search the mental lexicon."

The proposed mechanism at work here is unclear.

"Associative word pairs may have significantly different values from unrelated word pairs, but a direction is not predicted. More positive values would indicate automatic activation

similar to semantics, while more negative values would indicate a need to search the mental lexicon."

Again, there are already many studies looking at associatively related words, including some that have controlled for semantic relatedness (e.g., work by Kandhadai and Federmeier).  These studies

have already shown that associative priming decreases N400 amplitudes, and does so across a variety of tasks, including lexical decision.

Are ERPs only to correct trials?

Line types in figures are very difficult to differentiate and the channel arrangement is somewhat confusing as it bears no relationship to actual scalp placement (e.g., more frontal electrodes consistently more anterior in the figure)

Graphs should show the prestimulus baseline period (especially important here, since in some cases the ERPs do not seem to be well-aligned at zero -- it is important to ensure that there is appropriate baselining)

Gender differences should be dropped.  There are no analyses -- only visual inspection -- and the meaning of topographic differences on the N400 is very unclear, especially across different groups that vary along factors such as head size, skull thickness, etc.

What were the actual reaction times for each condition in each task?  This is important since it bears on how much decision making/response related activity one might see in the ERP waveforms across tasks and conditions.