EJN European Journal of Neuroscience
FENS Federation of European Neuroscience Societies

Binding language: Structuring sentences through precisely timed oscillatory mechanisms

Katrien Segaert, Ali Mazaheri & Peter Hagoort

Review timeline:

Submission date:	31 July 2017
Editorial Decision:	23 October 2017
Revision received:	06 December 2017
Accepted:	14 December 2017

Editor: Heleen Slagter

1st Editorial Decision 23 October 2017

Dear Dr. Segaert,

First of all, apologies for the amount of time it has taken to deal with your manuscript. This was because of the difficulty in engaging reviewers. It has now been reviewed by two expert external reviewers as well as by the Section Editor, Dr. Heleen Slagter, and ourselves.

As you can see, both reviewers thought the findings reported in your manuscript were interesting and that the study methods and analyses are sound. However, they raise several issues that need to be clarified/resolved before we can consider your manuscript further for publication in EJN. Although reviewer 1 was mostly satisfied with the paper, reviewer 2 raised several more major comments. Addressing these points will improve the manuscript. In particular, it is important to further examine to what extent the beta effects reported might reflect evoked (ERP) activity, and reviewer 2 had some helpful suggestions in this regard. Reviewer 2 also rightly questioned the extent to which binding was really not possible in the no-pronoun condition and the interpretation of the alpha effect. This, and each of the other points raised by the reviewers, should be addressed in the revised version.

We also noted the following points that need to be addressed.
- Reference list is not quite in EJN style and too many authors named in the in-text citations.
- Figures are embedded in the text but should be submitted as separate files.
- Data statement needs to be relocated to the end of the manuscript
- Please ensure that the description of the statistical methods and reporting of statistical data adheres to EJN guidelines (see Author Guidelines).

If you are able to respond fully to the points raised, we would be pleased to receive a revision of your paper within 12 weeks.

Thank you for submitting your work to EJN.

Kind regards,

Paul Bolam & John Foxe
co-Editors in Chief, EJN

Reviews:

Reviewer: 1 (Andrew Dimitrijevic, Sunnybrook Health Sciences Centre, Canada)

Comments to the Author
The manuscript “Binding language: Structuring sentences through precisely timed oscillatory mechanisms” describes a study that examines oscillatory EEG activity to syntactically binding and non-binding word pairs. The main finding is that when words paired such that they are meaningful elicited differences in alpha and beta compared non meaningful word pairs.

The paper is well written and easy to follow. It would be of interest to the readership EJN.

The EEG analysis methods are thorough.

Very minor comments below:

I found the introductory comments phase-locked vs time-locked descriptions confusing. A phase locked event is the same as a time locked event. For example, page 5-6:

“A caveat of the event-related averaging approach is that it overlooks activity that is time-locked, yet not phase-locked to the experimental event.”

Don’t the authors mean:

“A caveat of the event-related averaging approach is that it looks at activity that is time-locked to the event and averages out the non phase-locked neural activity to the experimental event.”

.. or am I missing something. It might be semantics given that different fields often describe the same thing using different terms.

The authors describe pink noise and reserved speech … yet I don’t see any results of these. Please clarify.

Also … why did the authors choose such an odd word for the English example “dotch” .. I suggest the authors look up dotch in the Urban dictionary (http://www.urbandictionary.com/define.php?term=dotch).

Reviewer: 2 (Lars Meyer, Max Planck Institute for Human Cognitive and Brain Sciences, Germany)

Comments to the Author
SUMMMARY

the authors report on an n = 20 time-frequency EEG study on the role of neural oscillations in the binding of words into larger units (e.g., phrases) on the sentence level. the manuscript is interesting (given the diverse results in the prior literature) and reads very well.

i have two major concerns that do not question the validity of the findings and leave most of the statistical analyses untouched; i think addressing these comments would benefit the balancedness and strength of the findings, though.

MAJOR

1. experimental paradigm an alternative interpretations: i was slightly irritated by the experimental design; in particular, the authors suggest that in the no-pronoun experimental condition, binding is entirely impossible. i am not entirely convinced that this is the case, as i discuss in detail below (e.g., in prior work on minimal binding on the two-words level, pseudowords in combination with inflectional morphology were used to study syntactic binding, removing semantic binding). i think that it would be fine with saying that binding is harder / less expected in the no-pronoun condition, hence, in the most extreme case, this comment would just affect the interpretation and explanation. see some suggestions with my specific comments below:

3: "pob dotches" how is it clear that binding cannot occur here? the suffix -es may signal either a verb in simple present singular or a plural noun, both of which could bind with a preceding noun "pob" (first case) or a preceding adjective "pob" (second case). In fact, paradigms using pseudowords, keeping only inflectional morphology, have been often used to get rid of word semantics (e.g., Zaccarella et al., 2015) while still allowing for syntactic binding (in isolation). it would be helpful if the authors could comment on this.

3: unclear whether alpha-power increase indicates binding; the following interpretation might also explain the finding: in the pseudoword condition, binding is possible, but less strongly marked by cues in the input. essentially, these stimuli are syntactically ambiguous (see my last comment), and the alpha effect could relate to ambiguity / working memory demands. please see meyer et al.'s (2013) work on alpha power increases during working memory intensive sentence comprehension (i.e., long-distance dependencies) to consider this alternative interpretation.

6. "cil terst" see my above comment; couldn't "cil" be a proper noun or a name that could well form a phrase with a suffixed verb "tersten" inflected for simple present singular? the authors mention this correctly; maybe it is my mistake, but so far, i do not fully understand how the pseudoverb condition avoid

syntactic binding.

9. i still do not fully understand this paradigm; why do pseusoverbs, which have inflectional morphology, do not allow for syntactic binding to occur? it would be good to spend a few more words in setting your rationale straight.

10. "presence v. the absence" i am not enirrely convinced that simply the absence of a pronoun made binding impossible. i agree that any kind of prediction coudl not have been made, but binding should still be possible / attempted once the inflected verb is reached. i think the authors need to discuss this and tone down their introduction of their main contrast as well as the descriptions here.

10. "no agreement" in the extreme case, in a language that does not have agreement at all, binding will still occur, no (e.g., englich 'i run' versus 'we run', where there is no overt agreement, yet, still binding will occur, just based on the knowledge that the verb form has certain characteristics and word order is particular (e.g., 'donk runs' versus *'donk run').

22. "general cognitive functions such as" this is not entirely clear to me; binding is possible in the no-pronoun condition, but might be harder / more working memory demanding as not clear bindee (i.e., subject argument for the pseudoverb) is available in the buffer. this possibility should be discussed.

22. "this alpha increase" again, i think that this is not the only interpretation of the current results; hence, this very strong conclusion should be toned down.

2. beta effect and ERPs: in the first analysis, the beta band effect is very short-lived, hence, it might reflect an underlying ERP. this is also not fully ruled out by the second analysis approach (i.e., here, the series of short-lived beta band effects apparent from the spectrograms could be strong enough to carry significance in a more extended time window). i am afraid that the supplementary ERP analysis cannot fully rule out the beta-as-ERP interpretation, as it is restricted to the P600 time window. specifically, Wang et al. (2012) link beta to the N400, a relationship that cannot be assessed with the restricted time window used here.

14. "low-beta" this effect is very short-lived (100 ms, corresponding to only two or so beta cycles). are the authors sure that this is an oscillatory effect? it could also be an evoked response / event-related potential, no? maybe the authors should double-check whether there is an ERP whithin this analysis time window. my suspicion also is caused by the power spectra in the second figure, where the beta effects are short-lived bursts (which could be signatures of ERPs).

17. the signficant longer-lived beta effect in the permutation statistics do not fully rule out my beta-as-ERP suggestion (the spectrum is the same for this analysis). i think it would be nice to also look at the ERP for this contrast.

19. as the ERP analysis is restricted to the P600 time windod, the claim that "the differences [...] reflect timelocked [...], but not phase-locked processes" is maybe overly strong. i am sure that the authors are aware of the possibility to employ the cluster permutation approach to ERPs as well, and i encourage the authors to try this to rule out ERPs in the time windows of the oscillatory effects, in particular for the beta band.

MINOR

4. for merge, i think that the recent work by Zaccarella et al. (2015) should be cited.

4f. i think that it would be fair to stress / flesh out a little more the proposed differences between semantic and syntactic binding; some authors (including the group at NYU; also Wilson et al., 2014) have discussed the aTL effects in terms of semantic binding, whereas syntactic binding is more solidly related to the left IFG (e.g., Zaccarella et al., 2015, and many other papers correctly cited by the authors).

6: it is not entirely clear to me why the authors mention phonological binding here and further above; the authors do not introduce this in much detail

7. "that it overlooks" should be "that it mostly overlooks" (i.e., this is a matter of degree; decreasing phase-locking might also only result in a more sluggish ERP (e.g., a P600), but not necessarily in the full absence of an ERP).

7. the authors might consider here introducing beta in terms of top-down predictions (e.g., the Lewis & Bastiaansen work).

7. for the role of alpha in sentence processing, please also consider Meyer et al. (2013) here.

7. "novel utilization" please tone down, this technique is mostly the standard for time-frequency analysis these days.

7. statistical dualism unclear; why highlight the strengths of the cluster-permutation approach when frequency bands were defined from the literature?

8. "the present study" i am not sure the manuscript needs this in-between summary, the introduction so far was quite clear and concise; i suggest to remove this paragraph to shorten the manuscript.

9. please provide handedness information; i assume that the participants did not have any visual or neurological / psychological impairments, it would be nice to state this here.

9. "we cut pink noise segments" why was this done? unclear at this point in the text.

10. not fully clear to me why reverse speech "only occurred on filler trials". while i see that the experimental factor is not affected by this, it would still be good to have a few words on this design decision.

10. why was the presentation rate so slow? this is rather unnatural; could the authors provide some more motivation for this design decision?

11. "impedances were kept" i am not sure i know the measurement "K" for impedances; is this a typo or specific to the active electrodes? please explain.

11. "non-biological signal artifacts" what are these and how were these detected (e.g., waveform, power spectrum, topography)? please specify.

12. for the permutation statistics, please add detail on (1) number of neighboring sensors for cluster detection, (2) alpha for cluster entry (the authors specified the cluster-level threshold, but the cluster-entry threshold is a parameter of the test as well).

12. "we are possibility" should be "we are possibly".

13. "in a second analysis approach" was this decision made a priori or post-hoc? please make explicit. both would be fine, but i think the current dualism ("we are possibility") is not clearly motivated so far.

13. "complementary" this i don't fully understand. if the two approaches "largely converged", i think the authors should use the a-priori-only approach; if this approach failed to show a result, then the posthoc approach should be introduced as a cure and marked as posthoc in the text.

14. "in the power of theta" i know this is a null finding, but given recent reports of a modulation of theta-band responses by the preceding context (i.e., retrieval may be aided by predictability; Rommers et al., 2016), the discrepancy with the prior literature should be discussed and interpreted.

20. "induced by the onset of words" first, this is causal language and maybe not entirely adequate; in particular, the effect in the beta band could be related to prediction formation, which is an internal cognitive process rather than something triggered by a word onset. i suggest the authors rephrase this sentence.

20. "our second analysis" the authors are right in the frequency band is similar to the first analysis; yet, the time window is drastically different, which should be noted here.

20. "signifying" should be something like "potentially indicating" or so; given the potential rather gradual nature of the main contrast (see my above hunch that some binding may still occur in the non-pronoun condition), there are alternative interpretations (e.g., working memory demands, ambiguity); thus, the binding interpretation is not the only possible option.

21. for the beta effect before the target word and the associated prediction interpretation, the authors should cite the line of review papers and original studies by Lewis & Bastiaansen (2015, 2016) and Lewis et al. (2015, 2016).

21. for the interpretation of the alpha effect, the authors should consider discussing the work of Meyer et al. (2012) on long-distance dependency processing.

21. a more general point is on the relationship between prediction and binding; according to the proposal in the Rommers et al. (2016) paper, retrieval could be modulated by predictions. the same thing could be suspected for binding; i think that this manuscript would benefit in scope from an analysis on the

relationship between the pre-target prediction and binding at the target word (e.g., correlation analysis, etc.).

Authors' Response 06 December 2017

We would like to thank both reviewers for their constructive comments and suggestions, which we feel have greatly improved our manuscript. Please find our point-by-point reply below, with changes to the text copied in red.

Reviewer: 1

Comments to the Author: The manuscript "Binding language: Structuring sentences through precisely timed oscillatory mechanisms" describes a study that examines oscillatory EEG activity to syntactically binding and non-binding word pairs. The main finding is that when words paired such that they are meaningful elicited differences in alpha and beta compared non meaningful word pairs.
The paper is well written and easy to follow. It would be of interest to the readership EJN.
The EEG analysis methods are thorough.

Very minor comments below:
I found the introductory comments phase-locked vs time-locked descriptions confusing. A phase locked event is the same as a time locked event. For example, page 5-6: "A caveat of the event-related averaging approach is that it overlooks activity that is time-locked, yet not phase-locked to the experimental event." Don't the authors mean: "A caveat of the event-related averaging approach is that it looks at activity that is time-locked to the event and averages out the non phase-locked neural activity to the experimental event." .. or am I missing something. It might be semantics given that different fields often describe the same thing using different terms.

Thank you, we have adapted our introduction as follows: A caveat of the event-related averaging approach is that it looks at activity that is time-locked and phase-locked to the event, while averaging out the non-phase-locked neural activity to the experimental event. (p. 5-6)

The authors describe pink noise and reserved speech … yet I don't see any results of these. Please clarify.
We have included more information on the motivation for our filler trials.
(p. 9) "We also used the following trials as fillers. First, we included these *pink noise trials* for variation. A *pink noise trial* consisted of two segments of pink noise which was matched in length with a pronoun and a verb, or with two verbs (120 instances). Second, we also had *reversed speech trials*, which were included to create a task for the participants (see below). In reversed speech trials, a segment was played in reverse (90 instances; of these, 30 contained a pronoun and a reversed pseudoverb, 30 contained a reversed pseudoverb followed by a pseudoverb, 15 contained a reversed pseudoverb and pink noise, 15 contained pink noise and a reversed pseudoverb). Lastly, there were 30 instances of a *minimal sentence - mismatch condition,* in which there was no agreement in person and number between the pronoun and pseudoverb. These trials were inserted to ensure some continuity and similarity with the stimuli from the behavioural pretest experiment (see below)."

Also … why did the authors choose such an odd word for the English example "dotch" .. I suggest the authors look up dotch in the Urban dictionary.
Please do note that our stimuli were in Dutch. The English examples we mentioned in the manuscript were not used as stimuli, but were there to illustrate the paradigm to readers who do not understand Dutch. We have now chosen a different English example ('to grush').

Reviewer: 2

Comments to the Author
SUMMMARY: The authors report on an n = 20 time-frequency EEG study on the role of neural oscillations in the binding of words into larger units (e.g., phrases) on the sentence level. The manuscript is interesting (given the diverse results in the prior literature) and reads very well. I have two major concerns that do not question the validity of the findings and leave most of the

statistical analyses untouched; I think addressing these comments would benefit the balancedness and strength of the findings, though.
The thorough comments and suggestions for changes have substantially improved our manuscript and indeed made it more balanced. More information on our changes can be found below.

MAJOR
1. experimental paradigm an alternative interpretations: i was slightly irritated by the experimental design; in particular, the authors suggest that in the no-pronoun experimental condition, binding is entirely impossible. i am not entirely convinced that this is the case, as i discuss in detail below (e.g., in prior work on minimal binding on the two-words level, pseudowords in combination with inflectional morphology were used to study syntactic binding, removing semantic binding). i think that it would be fine with saying that binding is harder / less expected in the no-pronoun condition, hence, in the most extreme case, this comment would just affect the interpretation and explanation. see some suggestions with my specific comments below:
3: "pob dotches" how is it clear that binding cannot occur here? the suffix -es may signal either a verb in simple present singular or a plural noun, both of which could bind with a preceding noun "pob" (first case) or a preceding adjective "pob" (second case). In fact, paradigms using pseudowords, keeping only inflectional morphology, have been often used to get rid of word semantics (e.g., Zaccarella et al., 2015) while still allowing for syntactic binding (in isolation). it would be helpful if the authors could comment on this.
In the present version of our manuscript, we have included additional details on our stimulus materials and pre-test experiment performance results, which the reviewer will find helpful.
First, let us consider the case where 'pob' could be a noun and 'dotches' a pseudoverb. In the section on our pre-test materials, we have added:
"The results of our pre-test experiment also suggest that it is unlikely that participants are performing syntactic binding in the wordlist condition. One could be concerned that participants attempt binding in specific instances of wordlists, such as 'pob grushes' (in Dutch: 'cil terst'), if the participants were to consider 'pob' as a singular noun and 'grushes' as a pseudoverb. Note however that in the wordlist condition, the second pseudoverb was presented in the 1st, 2nd or 3rd person singular or plural. This means that we have instances in the wordlist condition such as 'cil ters' ('ters' in 1st person singular, which happens in 1/6 cases) and 'cil tersen' ('tersen' in 1st, 2nd or 3rd person plural, which happens in 3/6 cases). An English equivalent would be 'pob grush'. If participants were to indeed bind these wordlists instances, they would identify 66% of the trials in the wordlist condition as having an agreement mistake in number and person between pseudonoun and pseudoverb. Given that performance accuracy was 97.4% for correctly saying that there was no mistake in the wordlist condition, it is thus unlikely that participants were performing syntactic binding in the wordlist conditions."
Next, let us consider the possibility where 'pob' could be an adjective. Zaccarella and Friederici (2015) have demonstrated that adjective determiners can be merged with pseudowords that are nouns. However, in our experiment the pseudowords are all verbs and introduced as such to the participants (we included: Participants were told that the second word would be a pseudoverb and were given examples and practice trials.) So again, this would suggest that if the participants perform binding in the wordlist condition (based on 'pob' being an adjective), they would identify the wordlist trials as syntactically wrong (adjective+pseudoverb), which the pre-test performance results show was not the case.
Note also that the same participants who do our EEG study first did the behavioural pre-test experiment, for which we are reporting the results.
We hope that these added explanations will clarify to the reviewer why we consider our paradigm to be similar to Zaccarella, Meyer et al. (2017) who compared "prepositional phrases and sentences—both involving merge—to word lists—not involving merge" and similar to Zaccarella and Friederici (2015) who were "using two-word sequences either allowing hierarchical syntactic binding to apply (phrase trials) or not (list trials)". Analogous to these established publications, we would pose that our wordlist condition does not involve syntactic binding either.

3: unclear whether alpha-power increase indicates binding; the following interpretation might also explain the finding: in the pseudoword condition, binding is possible, but less strongly marked by cues in the input. essentially, these stimuli are syntactically ambiguous (see my last comment), and the alpha effect could relate to ambiguity / working memory demands. please see meyer et al.'s (2013) work on alpha power increases during working memory intensive sentence comprehension (i.e., long-distance dependencies) to consider this alternative interpretation.

Above we have addressed the concern that our wordlist condition may involve binding. Our behavioural performance results speak against this. This suggests that the cognitive load of our wordlist condition is therefore minimal. Our arguments are similar to the arguments of others who have published papers using a minimal phrase paradigm, for example Zaccarella and Friederici (2015): “the amount of cognitive load required to process such small constructions is very limited”. We do agree that it would be good to point out to the reader that alpha power increases have been found for working memory in the context of sentence comprehension, so we have included a section on p. 6:
“Alpha power increases have also been linked to the involvement of working memory in sentence comprehension (Meyer et al., 2013), but this would mostly play a role in working memory intensive sentence processing such as long-distance dependencies, rather than the processing of two-word sentences.”

6. "cil terst" see my above comment; couldn't "cil" be a proper noun or a name that could well form a phrase with a suffixed verb "tersten" inflected for simple present singular? the authors mention this correctly; maybe it is my mistake, but so far, i do not fully understand how the pseudoverb condition avoid syntactic binding.
See above. We have included information on our behavioural performance results, which speak against the possibility that participants are syntactically binding these wordlists. Cil could be a proper noun, but our participants seem to not be interpreting it as such. If participants would indeed bind these wordlists, they would indicate that they are incorrect in 66% of the instances and thus the performance % in the wordlist condition would be around 33%. Instead our results indicate that in fact it is 97.4%.

9. i still do not fully understand this paradigm; why do pseusoverbs, which have inflectional morphology, do not allow for syntactic binding to occur? it would be good to spend a few more words in setting your rationale straight.
We have included more examples in this section. Also we have included an additional paragraph in the section on the behavioural pre-test results (see comments above). Pseudoverbs could be bound with more than just pronouns. But our pre-test results strongly suggest that in our paradigm, participants are not binding the lists of pseudoverbs.

10. "presence v. the absence" i am not enirrely convinced that simply the absence of a pronoun made binding impossible. i agree that any kind of prediction could not have been made, but binding should still be possible / attempted once the inflected verb is reached. i think the authors need to discuss this and tone down their introduction of their main contrast as well as the descriptions here.
We have clarified above why we believe there is strong evidence that participants are not performing binding in the wordlists condition of our experiment. However we are happy to tone down statements such as “presence vs. absence” and have changed these to “conditions of interest thus only differing in the extent to which the binding process can occur” (p. 7) and “(i.e. a comparison of the extent to which syntactic binding occurs)” (p. 9).

10. "no agreement" in the extreme case, in a language that does not have agreement at all, binding will still occur, no (e.g., englich 'i run' versus 'we run', where there is no overt agreement, yet, still binding will occur, just based on the knowledge that the verb form has certain characteristics and word order is particular (e.g., 'donk runs' versus *'donk run').
We apologize for the confusion caused. In the “minimal sentence - mismatch condition” it is not the case that there is no overt agreement. Instead, there is an *agreement error* in person and number between the pronoun and the pseudoverb. This condition is a crucial condition in the behavioural pretest experiment. We have clarified this: “Lastly, there were 30 instances of a *minimal sentence - mismatch condition,* in which there was an agreement mistake in person and number between the pronoun and pseudoverb. These trials were inserted to ensure some continuity and similarity with the stimuli from the behavioural pretest experiment (see below).”
We have furthermore expanded on our report of the behavioural pre-test experiment results also to further clarify (see above).

22. "general cognitive functions such as" this is not entirely clear to me; binding is possible in the no-pronoun condition, but might be harder / more working memory demanding as not clear bindee (i.e., subject argument for the pseudoverb) is available in the buffer. this possibility should be discussed.

Above we have addresses the issue of binding in the wordlist (i.e. no pronoun) condition: there is no evidence that participant are performing syntactic binding in this condition. In line with previous studies using similar paradigms, we believe that with the two-word phrases the working memory load is minimal. For example, Zaccarella at el. posit: “Phrases of two-word length —like this ship— are the ideal level to investigate this most basic process of syntactic binding, as the amount of cognitive load required to process such small constructions is very limited.” (p.2 of Zaccarella and Friederici, 2015). Therefore, in Zaccarella at el. their manipulation is taken to be one of syntactic binding (establishing a neuroanatomical signature in lIFG for merge) and not a manipulation of working memory. Likewise, we believe that our manipulation is one of syntactic binding and not working memory.
We have further clarified our position on p. 18: “There is likely little contribution from general cognitive functions such as working memory (a concern that has been raised for previous studies on syntactic processing using more complex syntactic structures). Processing phrases or sentences with a length of two-words involves minimal cognitive load (Zaccarella & Friederici, 2015).”

22. "this alpha increase" again, i think that this is not the only interpretation of the current results; hence, this very strong conclusion should be toned down.
Throughout our manuscript, we have used less strong language when we are interpreting our findings, e.g. using ‘could be’ instead of ‘is’. p. 18 and elsewhere: “This alpha increase could be a neural signature for binding taking place.”

2. beta effect and ERPs: in the first analysis, the beta band effect is very short-lived, hence, it might reflect an underlying ERP. this is also not fully ruled out by the second analysis approach (i.e., here, the series of short-lived beta band effects apparent from the spectrograms could be strong enough to carry significance in a more extended time window). i am afraid that the supplementary ERP analysis cannot fully rule out the beta-as-ERP interpretation, as it is restricted to the P600 time window. specifically, Wang et al. (2012) link beta to the N400, a relationship that cannot be assessed with the restricted time window used here.
14. "low-beta" this effect is very short-lived (100 ms, corresponding to only two or so beta cycles). are the authors sure that this is an oscillatory effect? it could also be an evoked response / event-related potential, no? maybe the authors should double-check whether there is an ERP whithin this analysis time window. my suspicion also is caused by the power spectra in the second figure, where the beta effects are short-lived bursts (which could be signatures of ERPs).
17. the signficant longer-lived beta effect in the permutation statistics do not fully rule out my beta-as-ERP suggestion (the spectrum is the same for this analysis). i think it would be nice to also look at the ERP for this contrast.
19. as the ERP analysis is restricted to the P600 time window, the claim that "the differences [...] reflect timelocked [...], but not phase-locked processes" is maybe overly strong. i am sure that the authors are aware of the possibility to employ the cluster permutation approach to ERPs as well, and i encourage the authors to try this to rule out ERPs in the time windows of the oscillatory effects, in particular for the beta band.
Following the reviewer’s suggestion we used the cluster-permutation approach to assess any difference in ERPs at the time intervals of the oscillatory differences that we observe. We did not get anything significant. We then took this further and some performed additional checks as well. We performed tests on the time-frequency representation of the ERPs, and did not find any condition differences. Finally, we removed the spectral components of the averaged ERP (i.e. phase-locked oscillatory activity) from the total spectra of EEG and reran our cluster-based permutation tests. The removal of the ERP from the spectra did not alter our results. We have now included these analyses in the manuscript with the following additions:
“Time-frequency representations of power in an EEG epoch capture non-phase locked activity as well as the spectral representation of the ERP. It is therefore a possibility that the oscillatory power changes reported above (Figure 2-4) are driven by ERP effects, which are time- and phase-locked to target word onset. To investigate this possibility we used several different approaches.
First we used the cluster-permutation test to assess if there were any differences in ERP amplitudes between the two conditions (Figure 5A) in the P600 window, or at any of the time intervals in which significant oscillatory differences were observed. We did not find any differences in the amplitude of the ERPs between the two conditions (i.e. no cluster of electrodes surpassed the threshold of $p<0.05$).
Our next approach focused on the spectral representation of the ERPs in each condition. Figure 5B illustrates the difference between the time-frequency representation of the ERP for the minimal sentence condition and the time-frequency representation of the ERP for the wordlist condition (i.e. the ERPs depicted in Figure 5A). We performed t-tests on the time-frequency representation of these ERPs to determine if there was a difference between the conditions, and found that there were no differences in alpha (8-12Hz) in the time

EJN European Journal of Neuroscience FENS Federation of European Neuroscience Societies

interval -0.4 to 0 s preceding the second word (t(19)=1.36, p<0.12), in beta (15-20 Hz) in the time interval 0.25 to 0.15 s preceding the second word (t(19)=1.2989, p<0.2), and in alpha (8-12 Hz) in the time interval 0.05 to 0.35 s following the target word (t(19)=1.7, p<0.1). This suggests that the oscillatory differences we observed above (Figure 2-3) are not likely to be driven by phase-locked activity.
Lastly, we removed the spectral components of the averaged ERP from the total spectra (i.e. phase and non-phase locked) (Cacace & McFarland, 2003) and reran the cluster-based permutation tests as reported above. The significance levels of our results were unchanged. Based on these analyses, it is unlikely that changes in the ERP drive the oscillatory power changes we report above."

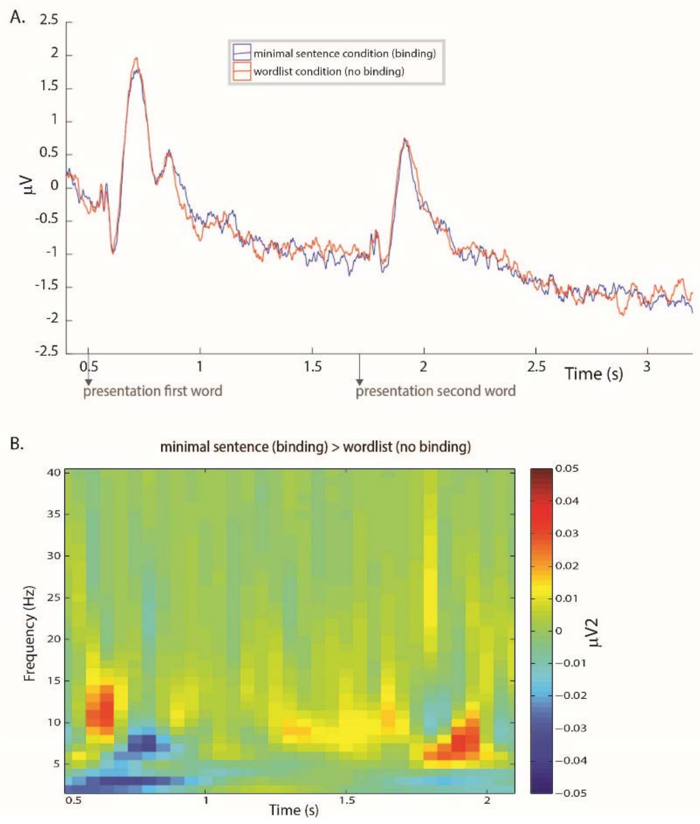

Figure 5. Phase-locked activity. A. The evoked response is illustrated for two central electrodes (corresponding to electrodes 1 and 2 on the equidistant electrode montage as displayed in Figure 1 of Simanova, van Gerven, Oostenveld and Hagoort (2010)), for the minimal sentence and the wordlist condition. B. The difference between the time-frequency representation of the ERP for the minimal sentence condition and the time-frequency representation of the ERP for the wordlist condition.

MINOR

4. for merge, i think that the recent work by Zaccarella et al. (2015) should be cited.

We added the following references: Zaccarella & Friederici, 2015; Zaccarella et al., 2017 on p. 2 where we mention *Merge*. Although we were already aware of Zaccarella, Meyer et al. 2017 (we had cited it in other parts of our manuscript), we had not read Zaccarella and Friederici (2015) yet. So thank you for pointing us in the direction of this paper. We included a more in depth discussion of Zaccarella and Friederici (2015) in different parts of our manuscript also, since their design is very relevant for our study.
p. 4: "Different is the fMRI study by Zaccarella and Friederici (2015), who used two-word phrases combining an adjective determiner with a noun. Crucially, the noun was a nonword. With this design, they investigated the neuroanatomical basis of syntactic binding in a context with minimal semantic information, and found the anterior section of the ventral left pars opercularis (a subregion of BA44) to be involved."
p. 5: "In the present study we use a minimal two-word sentence paradigm to target syntactic binding processes, while minimizing contributions of semantic and phonological binding as much as possible. This is unlike most previous studies using a minimal paradigm (although see Zaccarella and Friederici (2015) for an exception), which have in common that they all have manipulated binding at multiple levels at the same time."

4f. i think that it would be fair to stress / flesh out a little more the proposed differences between semantic and syntactic binding; some authors (including the group at NYU; also Wilson et al., 2014) have discussed the aTL effects in terms of semantic binding, whereas syntactic binding is more solidly related to the left IFG (e.g., Zaccarella et al., 2015, and many other papers correctly cited by the authors).

We feel this is a rather complicated issue and fleshing it out may not be in the scope of the present manuscript. We agree with you that there may indeed be differences in terms of the neuroanatomical architecture between semantic and syntactic binding, and many interesting fMRI findings are being

published on this issue (e.g. also Zaccarella, Meyer et al. 2017, which we have added to our introduction section in this version of the manuscript). We have published several fMRI papers on the issue ourselves as well (e.g. Menenti, Gierhan, Segaert and Hagoort, 2011, Segaert, Menenti, Weber, Petersson and Hagoort, 2012) which may not necessarily be fully in line with above however, so we feel that taking a stronger position on this issue may not be within the scope of this neural oscillations manuscript.
We are very careful in our links with fMRI work also in the discussion section, e.g. "We believe that these findings could be interpreted against the backdrop of theoretical frameworks proposing that a left-lateralized network of frontal-temporal areas is associated with syntactic binding … we found a power increase in left lateralized frontal-temporal electrodes in the alpha frequency band (8-12 Hz) following the onset of the target word."

6: it is not entirely clear to me why the authors mention phonological binding here and further above; the authors do not introduce this in much detail
For completeness when discussing binding, we want to mention that binding happens at the syntactic, semantic and also phonological level.
"The process of binding needs to happen at multiple levels. At the level of phonology, words are bound into intonational phrases. Another level of binding is semantic binding: the construction of complex meaning when words are combined into phrases and sentences. A third level is syntactic binding: the combination of words into larger structures, taking into account features that mark syntactic structure, tense, aspect and agreement."
However since it is not a topic of investigation in our study, we do not mention it in greater detail.

7. "that it overlooks" should be "that it mostly overlooks" (i.e., this is a matter of degree; decreasing phase-locking might also only result in a more sluggish ERP (e.g., a P600), but not necessarily in the full absence of an ERP).
We have changes this section (see also comments reviewer 1).

7. the authors might consider here introducing beta in terms of top-down predictions (e.g., the Lewis & Bastiaansen work).
We agree that this work should be discussed in relation to our findings but feel that the discussion would be a more natural place to do so than the introduction. We have therefore included it in the general discussion on p. 17: "This would be in line with the proposal of a role for beta oscillations in top-down predictions, stimuli expectation and maintenance of information (Weiss & Mueller, 2012; Lewis & Bastiaansen, 2015; Lewis et al., 2015; Lewis et al., 2016)."

7. for the role of alpha in sentence processing, please also consider Meyer et al. (2013) here.
We have added: "Alpha power increases have also been linked to the involvement of working memory in sentence comprehension (Meyer et al., 2013), but this would mostly play a role in working memory intensive sentence processing such as long-distance dependencies, rather than the processing of two-word sentences." (p.6)

7. "novel utilization" please tone down, this technique is mostly the standard for time-frequency analysis these days. 7. statistical dualism unclear; why highlight the strengths of the cluster-permutation approach when frequency bands were defined from the literature?
We re-wrote this section to take into account these comments and clarify what we mean. (p.6-7)
"We use two distinct but complementary analysis approaches. The first analysis approach is based on frequency bands as defined in previous literature, using a cluster-randomization test which circumvents multiple comparisons (Maris and Oostenveld, 2007). Guided by previous research on syntactic binding, we could expect oscillatory power changes associated with binding in the alpha, theta or beta frequency range. Given that there is some variability across the literature in what frequencies are included within a frequency band we also use a second analysis approach that does not require us to determine pre-set frequency bands, but rather is data-driven. The second analysis approach looks at the entire range of frequencies between 3-30 Hz. For this, we use the cluster-randomization test (Maris and Oostenveld, 2007) in a non-traditional way. Traditionally, when using the cluster-randomization test one collapses across time, frequency or channels; a pre-defined time window, frequency range or channels of interest is based on prior literature. We here decided to, in addition to the traditional approach, not collapse across any of these dimensions. Instead, we examined which frequency, time and electrode combination shows a condition difference (similar to the approach used in some ICOG studies, e.g. Bauer et al. (2013))."

8. "the present study" i am not sure the manuscript needs this in-between summary, the introduction so far was quite clear and concise; i suggest to remove this paragraph to shorten the manuscript.
We respectfully disagree with the reviewer on this issue. This paragraph is quite short and we would prefer to keep it in.

9. please provide handedness information; i assume that the participants did not have any visual or neurological / psychological impairments, it would be nice to state this here.
We have added this information.

9. "we cut pink noise segments" why was this done? unclear at this point in the text.
We added the following: The reversed speech and pink noise segments were used on filler trials (see below).

10. not fully clear to me why reverse speech "only occurred on filler trials". while i see that the experimental factor is not affected by this, it would still be good to have a few words on this design decision.
The reason for the reversed speech trials was to create a task on some of the trials, which is explained later. We are now clearer on this: "Second, we also had reversed speech trials, which were included to create a task for the participants (see below)." The task is then explained in detail in the next paragraph, including that this task is used to ensure participants pay attention: "The participants' task was to detect reversed speech (which only occurred on filler trials). With this task we ensured that participants paid close attention to the stimuli throughout the EEG measurements. Also, there was thus no difference in response decision processes between the crucial conditions of interest, i.e. the minimal sentence and the wordlist condition."

10. why was the presentation rate so slow? this is rather unnatural; could the authors provide some more motivation for this design decision?
The motivation for this design decision was to avoid spectral leakage by the stimulus evoked responses to the words. It should also be noted that the participants did not describe the task as unnatural. Because there is a behavioural pre-test experiment (which uses the same timing), the participants are well used to the timing of the design before the EEG measurements start.

11. "impedances were kept" i am not sure i know the measurement "K" for impedances; is this a typo or specific to the active electrodes? please explain.
The symbol got cut off, we have corrected it. Impedances were kept below 20 kΩ.

11. "non-biological signal artifacts" what are these and how were these detected (e.g., waveform, power spectrum, topography)? please specify.
All trials prior to being sorted into any conditions were visually inspected for non-biological signal artifacts (e.g. electrode jumps or gross movements by participants) based on visual inspection of the waveforms.

12. for the permutation statistics, please add detail on (1) number of neighboring sensors for cluster detection, (2) alpha for cluster entry (the authors specified the cluster-level threshold, but the cluster-entry threshold is a parameter of the test as well).
We have included additional info on p.11-12: "The power of the frequencies of interest, in each channel and time point within the interval 0.5-3.1s after the fixation cross (i.e. the period in between the onset of the first word and the onset of the response screen) was clustered depending on if it exceeded a dependent samples t-test threshold of $p<0.05$ (two-tailed). We considered a cluster to consist of at least 2 significant adjacent electrodes. Next, the Monte Carlo p values of each cluster obtained were calculated on 1000 random partitions in which the minimal sentence and wordlist condition labels were shuffled. At each shuffle the data was clustered again and the cluster with the largest sum of t-statistics entered the shuffling distribution. A p-value was derived by calculating the number of times the t-statistics in the shuffled distribution was higher than the original t-statistic we derived by contrasting our conditions. We considered the critical alpha-level here to be 0.05."

12. "we are possibility" should be "we are possibly".
This has been corrected.

13. "in a second analysis approach" was this decision made a priori or post-hoc? please make explicit. both would be fine, but i think the current dualism ("we are possibility") is not clearly motivated so far.

This decision was not made post-hoc. We have rewritten the introductory section because we agree that our explanation lacked clarity. Hopefully all is clear now. (See earlier comment of same reviewer – changes made on p.6-7)

13. "complementary" this i don't fully understand. if the two approaches "largely converged", i think the authors should use the a-priori-only approach; if this approach failed to show a result, then the posthoc approach should be introduced as a cure and marked as posthoc in the text.
We do believe the approaches can be complementary, also when the more traditional approach does not fail the show a result. We have hopefully clarified this with the changes we made on p.6-7 (see earlier comments). We highlight this in our discussion section also (p.18):
"We uncovered effects in the high-beta range in our second analyses approach, suggesting that a separation based on pre-defined frequency bands could in some cases lead to overlooking condition differences. With this, our paper offers a methodological advance which may be particularly useful in future investigations where researchers do not have a-priori hypotheses about particular frequency bands involved in a task. This may prove particularly useful in a field like psycholinguistics, for which (in comparison to other cognitive domains) relatively little previous literature on neural oscillations is available."

14. "in the power of theta" i know this is a null finding, but given recent reports of a modulation of theta-band responses by the preceding context (i.e., retrieval may be aided by predictability; Rommers et al., 2016), the discrepancy with the prior literature should be discussed and interpreted.
We agree that Rommers et al. 2016 is very interested and should be discussed in relation to our findings. We think though that the theta modulation in Rommers et al. cannot readily be related to our findings as Rommers et al. interpret this as being related to control processes that happen following the presentation of unexpected words (a prediction not being confirmed). However, we have included a note on the alpha finding in Rommers et al. on p. 17 of our manuscript: "Importantly, both Wang, Hagoort and Jensen (2017) and Rommers et al. (2017) report evidence for the role of alpha power modulations in predicting upcoming language input." Similar to our results, Rommers et al. find and alpha band modulation *prior* to critical word onset.

20. "induced by the onset of words" first, this is causal language and maybe not entirely adequate; in particular, the effect in the beta band could be related to prediction formation, which is an internal cognitive process rather than something triggered by a word onset. i suggest the authors rephrase this sentence.
In this case we were using the word "induced" in contrast with the word "evoked", "evoked" meaning EEG changes phase-locked to an event and "induced" meaning that it is time-locked not phase-locked to the onset of an event. We did not mean to imply causality. We have changed the wording to elicited to avoid confusion.

20. "our second analysis" the authors are right in the frequency band is similar to the first analysis; yet, the time window is drastically different, which should be noted here.
I think some confusion has arisen because for our second analysis approach we have averaged over a 0.5 s lasting timewindow. This is because looking at Figure 4A, it is clear that this is the window in which there are condition differences. Averaging over this long timewindow, the beta effect comes out significant, even though it is more short-lived than this, and in fact the timing seems very similar to that observed in Figure 2. We have clarified this by adding in the results section on p.14:
"We should note that even though for this analysis we have averaged over a 0.5 s lasting time-window, panel 4A shows that the beta effect is more short-lived than this."

20. "signifying" should be something like "potentially indicating" or so; given the potential rather gradual nature of the main contrast (see my above hunch that some binding may still occur in the non-pronoun condition), there are alternative interpretations (e.g., working memory demands, ambiguity); thus, the binding interpretation is not the only possible option.
We have changed the wording to 'indicative of' and acknowledge the potentially gradual nature of the paradigm (e.g. we removed the wording "presence vs. absence of binding" in several other places in the manuscript). However we do want to stress that our behavioural performance results speak against the interpretation that participants are attempting binding in the wordlist condition.

21. for the beta effect before the target word and the associated prediction interpretation, the authors should cite the line of review papers and original studies by Lewis & Bastiaansen (2015,

2016) and Lewis et al. (2015, 2016).
We agree. We have included on p. 17: "This would be in line with the proposal of a role for beta oscillations in top-down predictions, stimuli expectation and maintenance of information (Weiss & Mueller, 2012; Lewis & Bastiaansen, 2015; Lewis et al., 2015; Lewis et al., 2016)."

21. for the interpretation of the alpha effect, the authors should consider discussing the work of Meyer et al. (2012) on long-distance dependency processing.
We agree it should be discussed but feel it would be more appropriate in the introduction so we have included on p. 6: "Alpha power increases have also been linked to the involvement of working memory in sentence comprehension (Meyer et al., 2013), but this would mostly play a role in working memory intensive sentence processing such as long-distance dependencies, rather than the processing of two-word sentences."

21. a more general point is on the relationship between prediction and binding; according to the proposal in the Rommers et al. (2016) paper, retrieval could be modulated by predictions. the same thing could be suspected for binding; i think that this manuscript would benefit in scope from an analysis on the relationship between the pre-target prediction and binding at the target word (e.g., correlation analysis, etc.).
We have expanded our general discussion to make the relationship between prediction and binding more clear.
p. 16-17: "We also found increases in power in the alpha, low-beta and high-beta frequency bands immediately preceding the onset of the second word. This could be interpreted as neural responses for the expectation of binding needing to occur. In our paradigm, participants can expect that binding will need to occur when the first word of the two-word sentence is a pronoun. Likewise, when the first word is a pseudoverb, participants can expect that no binding will need to happen. This is analogous to how syntactic binding would occur in natural language processing. When comprehending sentences, words come in one at a time. We expect the sentence to unfold further and can anticipate that upcoming words will need to be bound to the words that are already presented to us. Bastiaansen, Magyari, & Hagoort (2010) observed a progressive increase in power in the theta and low-beta band as the sentence unfolded. Importantly, both Wang, Hagoort and Jensen (2017) and Rommers et al. (2017) report evidence for the role of alpha power modulations in predicting upcoming language input. Moreover, beta oscillations have been shown to have an important role in top-down predictions, stimuli expectation and maintenance of information (Weiss & Mueller, 2012; Lewis & Bastiaansen, 2015; Lewis et al., 2015; Lewis et al., 2016). Thus, we seem to have observed a signature of an expectation for the sentence to unfold further, and binding needing to take place."