Revisions to Conversation, cognition and cultural evolution: a model of the cultural evolution of word order through pressures imposed from turn taking in conversation

We thank the reviewers for their thoughtful comments. We have revised the manuscript following their suggestions. The major change is an extended discussion of how our hypothesis relates to more well-established theories, and some justification for our various assumptions (including 45 new references in the bibliography).

In the text below, reviewers' comments appear in normal font and
>> Our responses appear with double arrows at the beginning of the line.

Reviewers' comments:

Reviewer #1: (the authors refer in various places to the supporting online information but I do not have access to it; thus my judgments are based on the main document)

The current article is another example of article that challenges the common view of cognitive biases as secondary in word order evolution (e.g., Gell-Mann and Ruhlen 2011 paper in PNAS). The authors also put forward a new track to understand pressures on word order: turn-taking (cognition in interaction) as a fundamental pressure. I am evaluating this article as a promising hypothesis, focusing on its evaluation, domains of applicability and connections with competing/complementary hypotheses.

The authors cite Ferrer-i-Cancho (2008). An updated and expanded version of the arguments is

Ferrer-i-Cancho, R. (2015). The placement of the head that minimizes online memory. A complex systems approach. Language Dynamics and Change 5 (1), 114-137.

>> We've added this citation

p. 3 "suffixes are much more likely than prefixes": the previous context suggests that the suffixes are interrogative suffixes but the wording suggest that the authors refers to suffixes in general. If they are considering suffixes in general, they should try to justify the generalization. Verbal affixes are perhaps a reasonable generalization between interrogative affixes and any affix.

>> It's true that the theoretical claim is about interrogative affixes and the statistical test is looks at affixing in general. However, Thompson's theory extends to affixing in general (see Thompson, 1998), and the statistical test here is simply an attempt to re-do her analysis with controls for historical inheritance. However, another author has also suggested that this test is a side-issue, and so we have moved it to the supporting materials.

The conclusion in p.3 does not imply a pragmatic explanation. Interrogative prefixes will

lengthen the dependencies between the subject and the other elements within a sentence. Thus the principle of dependency length minimization could explain the results. The conclusion is straightforward applying Ferrer-i-Cancho's (2008, 2015) mathematical theory of word order and treating suffixes as independent words. Indeed, the question that the authors are addressing resembles the problem of the optimal relative placement of verbal auxiliaries in SOV languages. Similar arguments apply to p. 16.

p. 16. Sentence final particles as buffers. What is their impact in terms of time?
A syllable takes about 300ms. A particle may take even less for not being lexical for conveying little information. The gain in dependency length minimization seems more important than the gain the in the buffering effect.

>> This is a good point. In the model we effectively treat them as half the length of an ordinary word. Yes, the dependency length minimisation would gain more time, but it's not mutually exclusive from our model. We've added this to the description of sentence final particles: "While sentence final particles are usually quite short, we assume that any extra time is beneficial and may lead to selection in the long term."

l. 121-123 "We argue that languages do not adapt just to our individual cognition (cf. Christiansen & Chater, 2008), but to the way we actually deploy the cognition in interaction". Btw: Ferrer-i-Cancho has also argued that individual cognition does not explain anything in word order. For instance, the relative placement of adjectives with respect to nouns in SVO languages is hard to explain in terms of individual biases. Instead, it could reflect adaptations that prevent regression to SOV (the typical predecessor of SVO), see Ferrer-i-Cancho 2015.

>> Our aim was to contrast pressures coming from individuals to pressures coming from the interaction between multiple interacting individuals, not to contrast it with evolutionary trajectories.

p. 6-7: conflicts between hearer and speaker needs in the ordering of subject, verb and object. Ferrer-i-Cancho has discussed conflicts between principles of word order (predictability maximization vs dependency length minimization):

Ferrer-i-Cancho, R. (2014). Why might SOV be initially preferred and then lost or recovered? A theoretical framework. In: THE EVOLUTION OF LANGUAGE - Proceedings of the 10th International Conference (EVOLANG10), Cartmill, E. A., Roberts, S., Lyn, H. & Cornish, H. (eds.). Evolution of Language Conference (Evolang 2014). Vienna, Austria, April 14-17. pp. 66-73.

The relationship between these principles and speaker and hearer needs could help to improve the arguments.
In particular the principle of maximization of the head (the verb in this context) seems to favour the hearer while the principle of dependency length minimization seems to concern both speaker and hearer. "what is good for the speaker (verbs at the end) is bad for the recipient, and what is good for the recipient (verbs at the beginning) is bad for the speaker (who must plan the whole sentence up front)."

Because of the research referred above, my intuition is rather different. Verbs are heavy components in many languages (and reading the article the authors seem to agree with that). Postponing them facilitates its processing by the receiver. Putting the verb 1st could be an advantage for the listener if the complements are hard to predict/process. Perhaps the latter is secondary.

>> We thank the reviewer for making this point. We've made different assumptions, and can't address the differences directly in this paper without analyses of real data, which is beyond our scope. However, we discuss the point above in the new final section of the paper.

p. 6-7. "Thus we conclude that coordination of verb-placement, either at the end or at the beginning, is strongly favoured by processing under rapid turn-taking, arguing that languages reported to have no or free word order (like many Australian languages) are actually likely to have a statistically predominant single word order in conversation."

I do not understand the implication for languages lacking a dominant order. For instance, it could be that in those languages the most frequent order is one of the two where the verb located at the center (SVO or OVS). The authors need some statistical support for that claim.

>> There are two issues here. First, we are at fault for concatenating two different claims and making the latter appears as an implication of the former.

>> Secondly, we realise that there is a need to substantiate our claim about free word order languages, although the reviewer will appreciate that good data on word order in conversation of free word order languages is difficult to obtain. Our approach has been to give one example of the kind of bias we predicted, and then have a broader discussion of free word order languages in the conclusion. (As a side-note, in regards to the question of languages that alternate between SVO and OVS, the WALS chapter on languages with two dominant word orders lists 67 languages, none of which alternate between SVO and OVS.)

>> We have changed the quoted text above to the following:

>> "Thus we conclude that coordination of verb-placement, either at the end or at the beginning, is strongly favoured by processing under rapid turn-taking. Even in languages with flexible word order, we suspect that there are biases towards a particular word order in everyday conversation (e.g. Samoan, Duranti, 1981, p. 171; Ochs, 1982, p. 661, see discussion)."

>> And in the discussion we include this:

>> Another issue is the model's predictions about freedom of word order. Populations in the model effectively start as free word order, but then converge on a single dominant order. In reality, many languages have flexible word order. However, even these languages usually only use a sub-set of possible orders frequently (see Austin, 2001; Hale, 1992). For example, while many orders are possible in Samoan, during conversation between 70% and 86% of clauses with an overt subject, object and verb are verb-initial (Ochs, 1982, p. 661; Duranti, 1981, p. 171), in line with our model. However, many languages also go against our predictions. In Dryer (2013a), 79% of languages coded as having two dominant word orders involve a change to the position of the verb (though none alternate between verb final and verb initial). Warlpiri typically has the order topic, verb phrase, comment, with verb-medial constructions being most frequent of clauses with an agent, patient and verb (from written text, Swartz, 1987). Indeed, word order in free word order languages is often determined by pragmatic (information-structure) factors such as 'newsworthy' or prominent items appearing first (Givón, 1983b; Swartz, 1987; Mithun, 1992). This goes against our specific hypothesis about the position of verbs (and the predictions of the uniform information density hypothesis, see below). However, it is compatible with the general idea of consistently keeping elements which require more effort to comprehend in the same relative position in order to facilitate turn taking. Modelling this might require utterances to be sensitive to information structure or considerations of processing dependencies between the different constituents (see Ferrer-i-Cancho, 2016).

>> (we also discuss the Bardi language later on)

p. 7 "Note however that the co-operative verb-initial solution is vulnerable, like all cooperation, to a selfish move: you could always suit yourself and return a verb-final turn. These considerations suggest that while both solutions are viable, the verb-final solution might predominate in cultural evolution."

Other factors could be breaking the symmetry between verb-initial and verb-final languages: the higher processing difficulty/learnability of verbs with respect to complements when deciding whether to maximize the predictability of the verb or that of its complements. They authors add an equivalent point in l. 364.

>> True, and we now discuss this broadly in the discussion section. Again, our aim is not to present an opposing theory, but a complementary one.

l. 218-225 The authors predict that languages with a dominant order should be more

abundant that languages that lack it. How do languages exhibiting a couple of primarily alternating word orders fit into their theoretical framework? e.g.,

Ferrer-i-Cancho R. (2016). Kauffman's adjacent possible in word order evolution. In S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher & T. Verhoef (eds.). The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11). New Orleans, USA, March 21-24.

Can they predict more than one dominant order? If not they should present it as limitation of their model.

>> True. We now cite this as a limitation of our model: "Another issue is the model's predictions about freedom of word order. All runs of the model converge on a single dominant order, while there are many languages which have flexible word order or use a sub-set of orders rather than a single type (see Dryer, 2013). Accounting for this might require utterances to be sensitive to information structure or considerations of processing dependencies between the different constituents (see Ferrer-i-Cancho, 2016)."

I think that the authors should clarify that their model is a model for the dominant order of a language, not a model for the actual distribution of word orders within a language.

>> Changed the abstract to "the stable distribution of dominant word orders across languages evolves". And we have changed the example description to:
"In the first two runs, both SVO and SOV are used for some time, but after about 15 generations, all agents are using SOV all the time (with some small deviations due to noise). So, we can classify the language of these agents as SOV. In the third run, enough agents selected SVO by chance that the conventional pressure pushed the frequency up. Eventually, the third population converges on SVO order. That is, a dominant word order emerges, and we are not concerned with the distribution of word orders within a language."
(It is possible for a population in this model to use more than one word order, but that does not happen due to alignment)

l. 228. "more frequent" -> "are more frequent"

>> Fixed

l. 260. "their own memory" -> "its own memory"? (or perhaps I misunderstood memory in the model; the current wording suggests that agents have a common memory)

>> (singular 'they' changed to 'its')

l. 328-330. A theory of individual costs may not explain why they exist but can explain why they appear after the verb (and not before) as I have indicated above.

l. 434-443. It is not clear if agents can choose to have particles or not, or if instead, they can only decide on their location. If location is the only choice, then a competing explanation has been given by Ferrer-i-Cancho using dependency length minimization (see comments above). One advantage of dependency length minimization is its predictive power (and the corresponding parsimony of the whole theory). See for a review of its predictions and the challenge of parsimony see

Ferrer-i-Cancho, R. & Gomez-Rodriguez, C. (2016). Liberating language research from dogmas of the 20th century. Glottometrics 33, 33-34 (available also on arxiv.org).

Having said that, it is possible that the author's proposal of a turn-taking approach has less predictive power simply due to its early stage compared to the theory of dependency length minimization.

>> Agents can choose to have particles or not, as well as decide their location if they do use them. We have clarified the description of this: "As well as the three basic word order types without particles, agents could also produce versions with a sentence final or sentence initial particle (thus 9 combinations of types to choose from)". One of the interesting parts of the model is that agents bother using particles at all, since it increases processing cost for the speaker. We feel this is one aspect that an interactive explanation addresses nicely.

The authors have managed to reproduce the frequencies of the (dominant) orderings of subject, verb and object but how can we be more certain that the essence of the authors' model is really what happens? The authors depart from an initial condition where all verbal placements are equally likely. Is there any evidence that this is also the case of in real languages? Historical data is crucial. An initial or early stage in word order evolution is SOV (see articles by Ferrer-i-Cancho above for a summary of sources of evidence). If the simulations contained only SOV initially (or SOV with a large proportion), would the authors still be able to reproduce the statistics of word order?

The author's model does not seem to be adequate to explain the initial stage SOV as simulations do not produce SOV consistently but only SOV as the most frequent variant. This seems to be incompatible with the real evidence in support of SOV as the dominant order.

Historical data supports the following path of evolution SOV->SVO->VSO/VOS (Gell-Mann and Ruhlen 2011). Could the model be able to reproduce that path? The current dominant word order frequencies have been generated following precisely that path (assuming that what Gell-Mann and Ruhlen report is really true).
When the authors write "there is a bias for cultures to evolve towards pushing the verb further back in the turn" (l. 475-476), word order evolution seems to suggest that when a language changes, it does it against the authors argument, fronting the verb (moving it towards the beginning of the sentence). In other words, the authors model is strong concerning its capacity to predict a preference for SOV or to reproduce the actual distribution of word orders but does not seem adequate to explain who languages have got to the

current distribution of word orders. Or am I wrong? Looking at Figure 5 I have realized that the dominance of SVO is preceded by a dominance of SOV. Could the authors check if the when the simulations yield VSO/VSO the preceding dominant languages are first SOV and then SVO? Similarly, if the simulations yield SVO, is it preceded by a dominance of SOV? If that is the case, the support for the model would increase dramatically.

>> This question involves the transitions between dominant word orders within a single run of the model. Gell-Mann & Ruhlen (2011) review historical changes to word order and estimate transitions between languages. They find that word order tends to change from SOV > SVO > VSO.

>> We ran many sequences with the standard parameters, observing the dominant order at each generation, where the dominant order was the order accounting for more than 50% of the distribution, or the previous dominant order (this was done to smooth out random fluctuations). We then explored the probability of transitions between each word order.

>> The most common pattern is to stay with the existing dominant variant, accounting for about 98% of transitions. SOV is about twice as likely to transition to SVO than VSO, in line with Gell-Mann & Ruhlen's order. However, SVO is about twice as likely to transition to SOV than to VSO. Also, VSO is about equally likely to transition to either of the other types. In summary, the model transitions do not fit well with those outlined in the linguistic literature on historical change.

>> However, the mechanism of change discussed in Gell-Mann & Ruhlen is grammaticalisation, rather than drift, and we don't have any model of interactions between grammatical structures. They also suggest that word order distributions have not reached a stable equilibrium, while in our study we assume that they have. Our model is more focussed on the stage of language evolution that Kirby & Scott-Phillips (2010) call "cultural evolution" - the initial emergence of salient features of language such as dominant word order, rather than "historical change" which describes more recent changes between dominant states. In this case, an alternative view of our model is as an explanation of the starting conditions which feed into the historical changes described in Gell-Mann & Ruhlen. That is, languages are likely to go from no fixed word order to having SOV order. Indeed, Our model does not have any additional mechanisms such as grammaticalisation which would cause transitions between dominant word orders after the initial convergence.

>> We have added the results and discussion above to the supporting information.

The authors should clarify the purpose of their model and the phenomena that the model suits: dominant order vs distribution of word orders, initial order/word order changes vs current distribution of dominant orders, ...

>> See above.

I appreciate the authors' list of assumptions in p. 18. I have realized that the authors are making an important assumption when evaluating their model: they are assuming that statistics of the languages reflects an stationary state. This implies that they assume that the current frequencies will not change in any direction (various researchers have argued that it is not the case). In contrast, their model reaches a stationary state (they authors say that it converges quickly in "some" generations).

>> This is true, but it's an assumption of the target data rather than of the mechanics of the model (so we don't include stationary distributions as an assumption in this list). The final section now discusses the difference between the current distribution of word order and language change. We also note that changes to current languages often occur due to grammaticalisation or contact, neither of which are modelled. We have changed "some generations" to "In every simulation, the population converged on a single word order type within 100 generations".

Reading p. 16 one realizes that another assumption the model is making is that word order variation is unconstrained in the sense that speaker can decide freely, e.g., a speaker can decide easily between an initial and a final placement of the verb. In contrast, this involves deciding between word orders that are at distance 2 or 3 in the permutation ring that is hypothesized to constrain word order variation (see Ferrer-i-Cancho 2016).

>> It's true that there are no processing dependency costs in the model. We've added this to the list of assumptions.

Italics is not used in all mathematical symbols: e.g., Eq. in l. 273 vs Eq. in l. 278. "i" versus "M" in l. 276.
>> Fixed

Reviewer #2: The authors investigate how conversational interaction might determine the structural features of languages, and in particular canonical word order. This is clearly and methodically motivated in this paper. Ideas from Conversation Analysis are brought into the contemporary debate over the role of functional pressures in the cultural evolution of linguistic structure. Although I broadly recommend the paper for publication, I have a number of comments which I would like to see addressed. These are split between theoretical/empirical points, some more model-centred ones, and some more minor ones.

Theoretical and Empirical comments

1) A major premise of this paper is that verbs are more intensive to process than the other main constituents of a sentence. This may well be the case, but there is no direct empirical evidence for this referenced in the paper: is there any psycho-linguistic work which points towards this? If not, I can think of a 'harmonic' account of processing which would (I think) have most of the same implications as this. For example, if speakers tried to maximise the distance between all similar constituents we would probably see the same thing.

>> This is a fair point. The reviewer will appreciate that there are few quantitative studies that look at the timecourse of processing during interactive conversation that specifically addresses our assumption. However, we discuss this now in the final section (see the final section of the paper, pages 26-28). Also, reviewer 1 has focussed on the alternative theory of minimising the distance between constituents, and we have responded to that above.

2) Related to this point, is there any evidence that speakers of, say, SOV languages interrupt each other less than SVO and VSO? Or maybe VSO speakers are more willing to interrupt each other?

>> The study of gap duration in many languages by Stivers et al. (2009) suggests that languages do not differ in their overlap. Indeed, an informal test on these languages and on English data from the Switchboard corpus (not shown here) did not find any differences by word order or the position of the word in the sentence. However, we're not sure that the prediction is relevant. Our overall theory is that languages adapt to the constraints of turn taking, so that a language with an 'unhelpful' word order might compensate in other ways by having e.g. more morphology to help prediction or sentence particles to act as buffers etc. That is, we do not predict that different word orders will interrupt each other to different degrees.

3) You bring up the issue of sentence particles in section 3.2, which is nice. Your argument that the particles act as a sort of 'buffer zone' seems reasonable, but you could also make the opposite argument for SOV languages with particularly rich suffixing verbal morphology: there's a lot of information packed into the end of the turn, which seem anti-functional somehow?

>> This is a puzzle, as we point out for Japanese. However, we also suggest that putting the hard stuff at the end of the sentence can limit the amount of planning a

responder needs to do before starting their own turn. We also note that languages with rich verbal morphology also tend to have rich agreement, which helps project the content of verbs.

4) I think this work interacts with, for example, Piantadosi's ideas about predictability in sentence processing, so I'd like to see how the authors see their work fitting in with that.

>> We have added a section at the end of the paper that discusses the relationship between our hypothesis and other work, including Piantadosi's.

Model-related comments

It feels like the design of this model is such that SOV languages are privileged from the outset. Assuming an initially uniform distribution of word-order exemplars and alpha is set to 1, imagine V0 is selected. The response weights for V0, V1, V3 respectively are {3, 4, 5}. For V1 first they are {2, 3, 4}, for V0 they are {1, 2, 3}. In all cases, this puts an initial bias on SOV, and this bias is enough to send it in that direction.
I have the feeling that your result is something like the stationary distribution of a Markov Chain whose initial transition probabilities are set by the weights matrix and alpha. This won't completely be the case (because exemplar models have memory, and the details of agent interaction and transmission etc), but I've written a short script which mostly replicates your results. Would it be possible to control for this effect, by changing the initial distribution of exemplars? Even if (as I suspect) the result is still robust, could you press back against my understanding here, that SOV is just naturally favoured by this model?

>> Yes. The weights bias agents towards SOV, and this is what drives the convergence to SOV. The abstract characterisation of the model as a Markov chain is accurate. And we agree that SOV order is favoured by the model. We also agree that it's easy to see that the biases from interaction will bias the model towards SOV order. In fact, that's the point of our paper.

>> We've added the following to the paper:
>> "This result essentially derives from the model favouring SOV word order. Indeed, it would be possible to generate similar results to the current ones with a simpler model. For example, a Markov chain with a bias towards SOV, without any of the details about turn taking. However, this would be a phenomenological model which captures the target distribution without specifying the underlying mechanism. In this paper, we are interested in articulating a possible mechanism and investigating whether it does in fact lead to the right kind of prediction. In our case, assumptions about what constrains responding turns lead to an emergent bias towards SOV. As a consequence, when the number of turns in a conversation (Nturns) is low so that there are few responding turns, the proportion of populations with dominant SOV order is reduced, and in the extreme case populations are equally likely to converge on any word order (column 1 of Figure 6), which shows that the model is not biased towards SOV, except when the constraints of turn taking are applied."

>> However, perhaps the point above is more about initial conditions affecting the trajectory. In this case, the reviewer can see our addition of a test of different initial conditions (see reply to reviewer 1 above, or supporting information section 4).

Minor Comments

1) P3: instead of interrogative 'suffixes', how about 'elements' (as we see all sorts of clitics and particles as well as suffixes.)

>> This prediction and test is specifically about interrogative suffixes ("If interrogatives are morphologically bound to the verb, this constraint leads to a specific prediction"). Many other forms exist and the prediction might extend to them, but it we focus on suffixes here following Thompson's theory.

2) P3: Nice stats, but the postposition/head-final pattern is so well known that this leaves you open to criticism on this point. Is there any evidence of, say, interrogatives patterning more robustly than other postpositional features?

>> We wanted to cite Thompson's claim, which also extends to affixing more generally (see Thompson,1998). Thompson's evidence does not control for historical relatedness, so we wanted to fact-check this part of the claim. However, another reviewer also suggested that the stats are a side-issue which distracts from the main argument, which is about the theory linking interaction and processing. Therefore, we've cut out the details of the test and put them into the supporting information.

3) I like your statement on p10: "This suggests that the number of languages that facilitate turn taking (e.g. by having fixed word orders ensuring coordination) should increase over time, while the number of languages that make turn taking less efficient should decrease." Is there any way you can bring this forward in the paper, as I feel it is a more concrete statement than the ones you have at the moment.

>> We have added to the introduction "In other words, languages should change over time to better serve turn taking.". Also, we realise that the quote on page 10 is not quite right. We're not predicting anything about direct competition between languages, so we should have said "the *proportion* of languages". .

4) Your description of the weighting function on p14 confused me for a moment: it's stated for the particular case V1 and V2, but it would be clearer (for me) if it used something like V_sent and V_respond, making the generality clearer.

>> Changed to V_initiate and V_respond

5) On p15, this looks a lot like what Alex Mesoudi calls 'vertical learning' and Simon Kirby calls 'iterated learning'. Is it worth bringing this up?

>> Yes, citation added.

6) The graph on p22 confuse me a bit too: as far as I can tell, there should never be a distance of more than 5 (in this version of the model), but the x-axis marks distances up to 6: is this a misunderstanding on my behalf? Also, this graph is quite hard to read anyway.

>> The distance scale issue was an error in plotting the curves in R (not in the agent based model code). The range should indeed be 1 to 5, and this has been changed. After discussion with some colleagues, we have inverted the alpha curves to reflect processing cost rather than production ease. This now matches the description in the text better. We have also labelled the graph more clearly and added more explanation to the text.

7) P29 references line 546: two references have been mixed up.

>> FIXED

Reviewer #3: COMMENTS TO THE AUTHOR.
Please answer these questions (they can be copy/pasted into the window):

1) Identify the type of work:
__ agent-based model

2) Identify the nature of the work:
__ Specialist work but results very relevant to interdisciplinary readership

3) Does the paper contain sufficient and appropriate references?
__ References are sufficient and appropriate

4) If the paper was submitted as a "research report", please tick one:

5) What are the main contributions of this paper?
See below.

6) Please give explanations of your decisions above, add suggestions to improve the quality of the manuscript, point out additional references etc.

Recommendation: Accept, pending major revisions.

The present article is interesting, and is a good fit for the journal. Still, I have a number of major comments. If the authors can amend for these, or if they can successfully refute them, I would recommend it for publication. In addition, I have also added some minor comments and questions that would need to be addressed.

Major comments

Comment 1

The fact that language users align their language use with others (p.7-10), and that they prefer to use the same variant (e.g. an SOV word order) within a single conversation is not in need of an explanation specific to word order. Language users do this when confronted with any kind of language variation. The cause of this forms the foundation of usage-based theory (entrenchment of frequent patterns) and the basis of sociolinguistics in general (language users align with peers of the same social group). Work by a.o. Szmrecsnayi (2005) on what he calls persistence, also shows that this happens in short term, e.g. within one and the same conversation. As far as I know, agents aligning themselves with their peers is also the basis of any agent-based model of language. Unless the authors have a solid reason to assume why these more general and simpler explanations cannot hold for word order, and we would therefore need another explanation specific to word order, this alignment cannot form an explanandum for the simulation. At the most, it can serve as a reality check for the model. If the agents would not align with each other, there would be something seriously wrong.

As an alternative, the authors could choose to focus more on the finding that a positive convex function is needed to obtain the results we observe in reality (p. 22). This is an interesting finding, which does not seem to follow trivially from the model and for which - as far as I am aware, but my knowledge here is rather limited - we have no previous evidence.

>> Indeed, populations in our model always converge on a single dominant order, even if the constraints of turn taking do not apply, because of the usage-based memory. We may have phrased this result too strongly - our concern was only that the readers understand the following: it's not that each population converges on using VSO 10% of the time, but that 10% of populations converge on using VSO all the time.

>> We have clarified the above in one place and removed the reference to the result in the conclusion (we agree that it doesn't actually add much). We have also highlighted the point about the convex function in the conclusion.

Comment 2

Right now, it is not entirely clear which assumptions are strictly needed to obtain which results of the model. Any assumptions that are not strictly necessary - even if they make the model more realistic - should absolutely be removed from it. Occam's razor trumps realism. Don't be afraid to boast about your model's simplicity, instead of apologizing for it. If any of my fellow reviewers criticize the model for being too simple, I recommend the editors to back the authors herein.

>> We agree with this approach: our aim was to give a full list of assumptions of the model (which many papers avoid doing). We've clarified in the list which ones are necessary and which ones are simplifying.

The following lists a number of assumptions for which I am not sure if they are strictly necessary. I am probably still missing a few.

* Do generations need to be discrete? Is it impossible to obtain the same results if agents are simply taken out and introduced one by one into the model, and if so, why?

>> Interestingly we actually used discrete generations as a simplification over rolling populations. Adding rolling turnover to a population involves including extra parameters such as the rate and consistency of turnover, which agents are selected for turnover, the possibility that some agents will be 'older' and therefore have more consolidated usage, and we may need to think about limiting an agent's upper lifespan limit. Since we are not measuring convergence times, merely eventual convergence, we leave this possibility to future studies.

* Do we need a separate 'child' generation that does not produce language but only

listens? Is it impossible to obtain the results if new agents immediately start speaking the moment they are capable of it, and if so, why?

>> This is also linked to the issue of discrete generations. There needs to be some mechanism of vertical cultural transmission, and in our model it is done effectively discretely. An alternative would be to have agents interact as they are learning. This would be implemented by having a population of half adults and half children, with the adult half being replaced by new children (and the current children becoming adults) at each generation. That is, a kind of rolling turnover.

* Do we need noise?

>> We have added a short analysis of running the model without noise (in the supporting information). In this case, the model is less likely to transition to more rare states (since trends accumulate faster and there is no change after convergence). The quantitative fit to the data improves with noise, but the qualitative result is the same as without noise. Noise is a reasonable assumption, and we maintain it as the default in the main body of the paper.

Comment 3

On p. 10, the authors argue that in historical language change, language strategies that encumber turn taking (e.g. free word order) should disappear, while language strategies that facilitate turn taking (e.g. fixed word order) should increase. This reasoning is flawed. What would be the reason for languages to develop strategies that encumber turn taking in the first place, especially if they might as well immediately develop superior strategies?
There might have been some reason for this of course. For example, a free word order might have some other benefits that a fixed word order does not have. Because of some change in the language's habitat, e.g. the development of a larger community of speakers, an influx of second language learners, or some language internal change, these benefits may become less pervasive, paving the way for the language to change to a different strategy that does facilitate turn taking. Of course, a change is the opposite direction would be equally possible. There is no a priori reason to assume that changes in either direction would be more frequent. Such an argumentation is proposed - in different ways - in for instance Lupyan & Dale (2010), van Trijp (2013) and Bentz & Winter (2013) when explaining why many Germanic and Romance languages recently seem to be switching from a case system to a system of fixed word order and prepositions.
However, turn taking itself does not seemed to have changed fundamentally in the last millennia, as also argued by the authors themselves in the introduction. As such, the reason for a language to develop a system A that encumbers turn taking, and then later to change to system B that facilitates turn taking, cannot lie in the domain of turn taking itself.

>> This is a good question. It is indeed the position of the authors that the pressures of turn taking have been imposed very early in the evolution of language. We have three responses.

>> One might wonder, assuming that the pressures from turn-taking were present at very early stages of language emergence (Levinson, 2006), why structures that go against this pressure would emerge at all. There are three responses to this. First, we assume that the pressure is weak bias rather than an absolute condition. Communicating in a variety of ways can be successful enough for everyday needs. Secondly, the pressure from turn taking comes from the interaction between two individuals, and may go against the selfish biases of individuals. At early stages, Individuals may be unlikely to innovate a solution that fits turn taking. Over time, however, the turn taking pressure may override the individual biases. This means that we assume random innovation and guided selection. There is some evidence for this in studies of iconicity in the lexicon, which may emerge over time and through interaction, rather than being present at the beginning (e.g. Verhoef et al., 2015; Tamariz et al., under review; Blasi et al., 2016). Finally, change is probabilistic rather than strictly directional. Adapting to turn taking has many solutions and interacts with pressures from other domains. Because a language changes piece by piece rather than by wholescale renovation, it is not guaranteed to reach an optimal turn-taking solution quickly, nor to remain there if it does reach it. However, modelling the interaction between turn taking and other processes such as grammaticalisation or contact is beyond the scope of the current model. Here, we ask simply how turn taking might influence the way that a conventional word order arises in a population.

>> We have added the text above as a footnote in the introduction. This point is also partially addressed by our new analysis of how the model behaves with different initial conditions.

>> Verhoef, T., Roberts, S. G., & Dingemanse, M. (2015). Emergence of systematic iconicity: Transmission, interaction and analogy. In D. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci 2015) (pp. 2481-2486). Austin, Tx: Cognitive Science Society.

>> Tamariz, M., Roberts, S., Martinez, I., Santiago, J. (under review) The interactive origin of iconicity. Cognition.

>> Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. Proceedings of the National Academy of Sciences. doi:10.1073/pnas.1605782113

Minor comments

* p.3: 'the interactional uses of language are (...) right at the limits of human performance'. Could you support this with a reference? Personally, I find that many of the conversations I have during the day follow recurrent templates, and I can definitely maintain

a conversation while driving, watching TV or making coffee. In fact, I feel like driving or making coffee require more focused attention than interactional language use.

>> This sentence starts with "As we will explain below", and in section 1.1 we explain what we mean (the gap between turns is smaller than the minimum time to plan and produce a word) and cite sources there. To this introductory sentence, we've added "(see below and Levinson, 2016)". Of course there are ritualised exchanges in conversation, and indeed we project what a speaker will say at all times, but our main point about turn-taking stands: we minimise gaps and overlaps with a precision that is at the limit of at least our physical reaction times. (We also note that many studies show that conversation puts a severe pressure on parallel tasks such as driving)

* p.4 what is meant by 'single "highest" human skill, language'. Are other human skills, say humor or persistence hunting, less "high"? Why?

>> Changed to "This ecology puts a premium on speed for the most complex human skill, language"

* p. 7:: 'as has been noted in previous pragmatic work'. Can you give some examples please? At the end of the next sentence, Zipf (1949) and Horn (1984) are mentioned, but although these represent very solid work, they are hardly state-of-the-art pragmatic research.

>> We have added some references to the more recent debate on the trade-off between length and information.

* p. 18-19: an explicit summary of the assumptions made by the model is very good practice, though some points could be more concrete. It is not necessary to mention that you 'do not model semantics and/or detailed syntax or morphology'. As mentioned earlier, you should take pride in making your model as simple as possible, and if you were to list all things that you are nót modelling, the list would become infinite.

>> This is a fair point, but we've found when giving talks about this that researchers who are not used to reading modelling papers (e.g. conversation analysts) often assume that our agents are producing utterances with semantic content. We included the explicit mention of this to make it clear to these readers.

* p. 23: why, in the model, do verb final particles appear in SOV runs, and verb-initial particles in VSO, exactly? This was not immediately clear from the text.

>> We have run some extra analyses and explain this in more detail in the results section. We've also expanded our explanation of the model without sentence initial particles.

* It is good practice to mention all exact parameter settings with each graph, either in a footnote or in a table in the appendix.

>> We have added the parameter settings to the graph titles.

* How did you choose the parameter settings shown in the graphs?

>> They represent our initial guesses about what the parameters should be. A more in-depth analysis of the range of parameters is in the supporting materials.