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

Seán G. Roberts and Stephen C. Levinson

Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

sean.roberts@mpi.nl, stephen.levinson@mpi.nl

Abstract

This paper outlines a first attempt to model the special constraints that arise in language processing in conversation, and to explore the implications such functional considerations may have on language typology and language change. In particular, we focus on processing pressures imposed by conversational turn-taking and their consequences for the cultural evolution of the structural properties of language. We present an agent-based model of cultural evolution where agents take turns at talk in conversation. When the start of planning for the next turn is constrained by the position of the verb the stable distribution of word orders evolves to match the actual distribution reasonably well. We suggest that the interface of cognition and interaction should be a more central part of the story of language evolution.

Keywords: Turn taking; Pragmatics; Typology; Word Order; Cultural Evolution.

Bio

Seán G. Roberts studies cultural evolution at the Max Planck Institute for Psycholinguistics. He is interested in whether differences between languages are the product of adaptation.

Stephen C. Levinson is the director of the Language and Cognition group at the Max Planck Institute for Psycholinguistics. His work focusses on language diversity and its implications for theories of human cognition. He uses methods ranging from fieldwork and conversation analysis to neuroimaging studies to explore the role that language plays in our everyday cognition.

1 Introduction

The evolution of linguistic structure is constrained by various cognitive pressures. For example, studies have argued that basic word order (the dominant order of Subject, Verb and Object in a transitive clause) is adapted to pressures on efficient storage or processing (e.g. Hawkins, 1994; Ferrer-i Cancho, 2008) or the effectiveness of conveying semantic information (e.g. Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014).

While these effects are no doubt part of the story, we suggest that the greatest functional pressures on language structure are likely to come from the very special circumstances in which it is primarily used. That special niche is conversation, or more generally, face to face interaction. This is where language is learnt, and most heavily deployed: we each produce something like 15,000+ words a day in some 1200 turns at talk (Levinson 2006, 2016). Therefore, understanding the constraints and affordances of conversation is crucial for understanding the selective pressures on language use. As Schegloff, one of the founders of the field of Conversation Analysis, put it:

"What is the primordial natural environment of language use, within which the shape of linguistic structures such as grammar, have been shaped? Transparently, the natural environment of language is talk-in-interaction, and originally ordinary conversation. The natural home environment of clauses and sentences is turns-at-talk. Must we not understand the structures of grammar to be in some important respects adaptations to the turn-at-talk in a conversational turn-taking system with its interactional contingencies?" (Schegloff, 1989, p. 143-144)

As we will explain below, the interactional uses of language are cognitively intensive, due to the high speed of the expected response being right at the limits of human performance. The demands of interactive conversation should therefore impose selective pressures on linguistic structures. If there is variation in how effective different structures are in conversation, and if more effective structures are more likely to 'replicate' and be used again, then this suggests that such structures should be under selection over time by the forces of cultural evolution (Croft, 2000).

An example of this process links constraints from pragmatics to predictions about typology. Thompson (1998) points out that interrogative structures make turn transition relevant: a question demands an answer. Thompson argues that, in order to be effective, interrogatives should generally apply to prosodic units, and therefore appear at turn boundaries, rather than in the middle of turns. If interrogatives are morphologically bound to the verb, this constraint leads to a specific prediction: languages that place the verb at the end of a sentence should have interrogative suffixes (so that the interrogative appears after the verb at the boundary), while languages with verbs at the beginning should have prefixes. We tested this statistically by looking at the probability of interrogative suffixes for different word orders in a sample of the world's languages, controlling for historical influence. Indeed, we find that suffixes are much more likely than prefixes in verb-final languages (460 languages taken from Dryer, 2013b and Dryer, 2013a, mixed effects model controlling for language family, log likelihood difference = 12.27, $\chi^2 = 24.5$, df = 2, $p < 0.0001$, see supporting materials). This is a well-known pattern in typology, but we suggest that part of the pressure that leads to the emergence of this pattern could be motivated by the pragmatic – and more specifically interactional - pressures on structures of this kind.

In this article, we consider a specific aspect of conversation - turn taking - and how the tight processing constraints it entails may lead to the selection of specific grammatical structures within a cultural evolution framework. While the work is preliminary, we hope to demonstrate the possibility and promise of linking domains that are not usually considered together: language structure, conversation, cognition and cultural evolution.

1.1 A cognitive pressure derived from turn taking

In a conversation, speakers take turns at talking and try to minimise the amount of gap or overlap between the turns (Sacks et al., 1974). When talking in groups, there is competition for who speaks next (Levinson, 1983), and a delay in response is pragmatically marked, for instance, it can be interpreted as unwillingness (Kendrick and Torreira, 2015; Bögels, Kendrick & Levinson, 2015; Roberts, Margutti & Takano, 2011). This puts speakers under pressure to respond quickly in conversation.

Indeed, the average gap between questions and answers is around 200ms (Stivers et al., 2009). What makes this surprising is that the time to plan and begin executing a *single word* is at least 600ms (Indefrey, 2011). Even though speech planning is incremental (speech may start before the whole sentence is planned, Levelt, Roelofs & Meyer, 1999), this implies that at some point we must be predicting the course of the incoming turn, extracting its action or speech act, and preparing our response in advance of the other speaker coming to a conclusion (Levinson, 2016). This imposes a kind of 'crunch zone' in which production and comprehension must overlap in time (see figure 1).

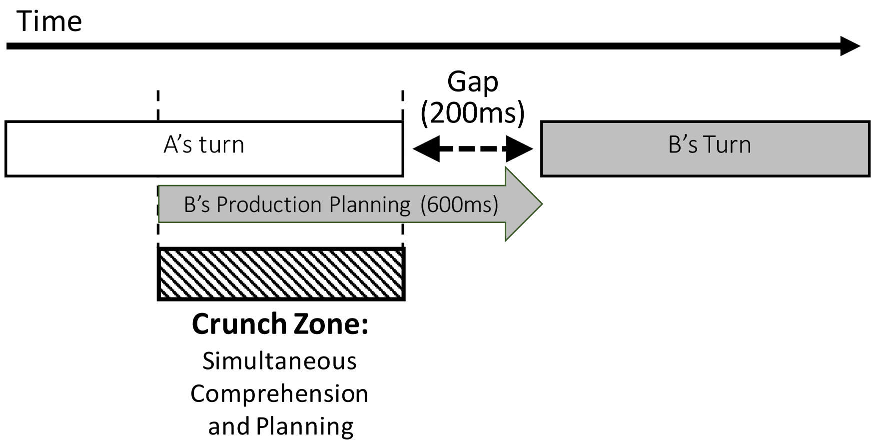

Figure 1: A schematic representation of turn taking.

This is a highly demanding ecology for rapid language use. The timing is remarkable – even in a non-linguistic context, 200ms is the normal minimum reaction time for a pre-prepared single response choice, and response times increase logarithmically in relation to the number of choice that have to be made ('Hick's Law', Hick, 1952, discovered first by Donders, 1868). Language speakers have vocabularies of 30,000 or more from which to begin a response.

This ecology puts a premium on speed with the single highest human skill, language. For example, if a recipient finds the incoming turn at talk unintelligible or hard to comprehend, he or she should respond with a request for repair (e.g. *"Huh?", "Who?", "Did I buy what?"*) before someone else continues because repair is hard to achieve beyond the immediate locale in which it occurs – it is only slight delayed to allow the speaker to do self-repair (Schegloff, Jefferson & Sacks, 1977; Kendrick, 2015). The repair system has adapted to this niche by an ordered preference for repair: self-repair is preferred over other-initiated repair, and specific repair initiators (*Who?; Which bottle?*) over general ones (*Huh?*, see also Dingemanse et al., 2015*)*, thus expediting repair.

We suspect that there are a large variety of adaptations to this niche in the interactive system itself (as just illustrated), but also in language structure, and indeed the cognitive skills that make it all possible. But here we focus on basic word order as an illustration of how language structures might adapt to the constraints of turn taking.

1.2 Linking processing and pragmatics

We could go further in linking pragmatics and typology by integrating constraints from online processing of interactive language use into a model of the cultural evolution of language. 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. It is not only the evanescent speech signal, but also the temporal pace of conversation that makes the cognitive pressures on normal language use so intensive. Therefore, one would expect the structure of language to adapt to this ecology, and we should be able to see signs of these adaptations in today's languages. For example, one possible locus of adaptation would be the order that information is presented in a turn. Information presented to a listener later is more likely to occur inside the crunch zone, and therefore present a greater challenge to producing the next turn on time.

Let us consider the implications for basic word order - that is, the order of the subject, object and verb in a canonical transitive clause. Through its lexically-specified argument structure, the verb provides the syntactic frame for a sentence and provides crucial semantic information about the action reported. Hence its position in the sentence might adapt to several processing pressures. Predictions here are complicated by the fact that the functional adaptation of a sentence structure to its interactive use must be viewed from two perspectives: the point of view of the speaker, and the point of view of the recipient or comprehender. As has been noted in previous pragmatic work, what is good for the speaker may be bad for the recipient, and vice versa. Consider, for example, the structure of the lexicon: making many semantic distinctions may be helpful for the recipient trying to recover the speaker's intended referent, but force the speaker to make careful choices between many alternatives (Zipf, 1949; Horn, 1984). In a similar way, verbs in final position may give speakers more time to plan the most complex component of the turn. On the other hand, verbs in initial position allow listeners to anticipate the unfolding of the incoming turn, using the predictive possibilities offered by the verb's argument structure, and thus start planning their own response much earlier. Here there is again a zero-sum type of situation: 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).

Notice that a mixed strategy will not help: if I put my verb at the end, it falls in your 'crunch zone', and it will be therefore especially difficult for you to put your verb at the beginning – you will not have had time to formulate the response. However, if you put your verb at the end too, then you will have most of the duration of the turn to plan the verb, the complex frame for the sentence (Figure 2). Alternatively, suppose I am considerate to you the recipient, then I could begin my turn with a verb, well clear of your crunch zone, and now aided by my co-operative gesture and the following more predictable components of the turn you will have time to compose your verb also in initial position, so returning the favour (see Figure 2). Both strategies will get the maximal distance between predicates, which is what will aid processing. 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.

Note however that the co-operative verb-initial solution is vulnerable, like all co-operation, 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.

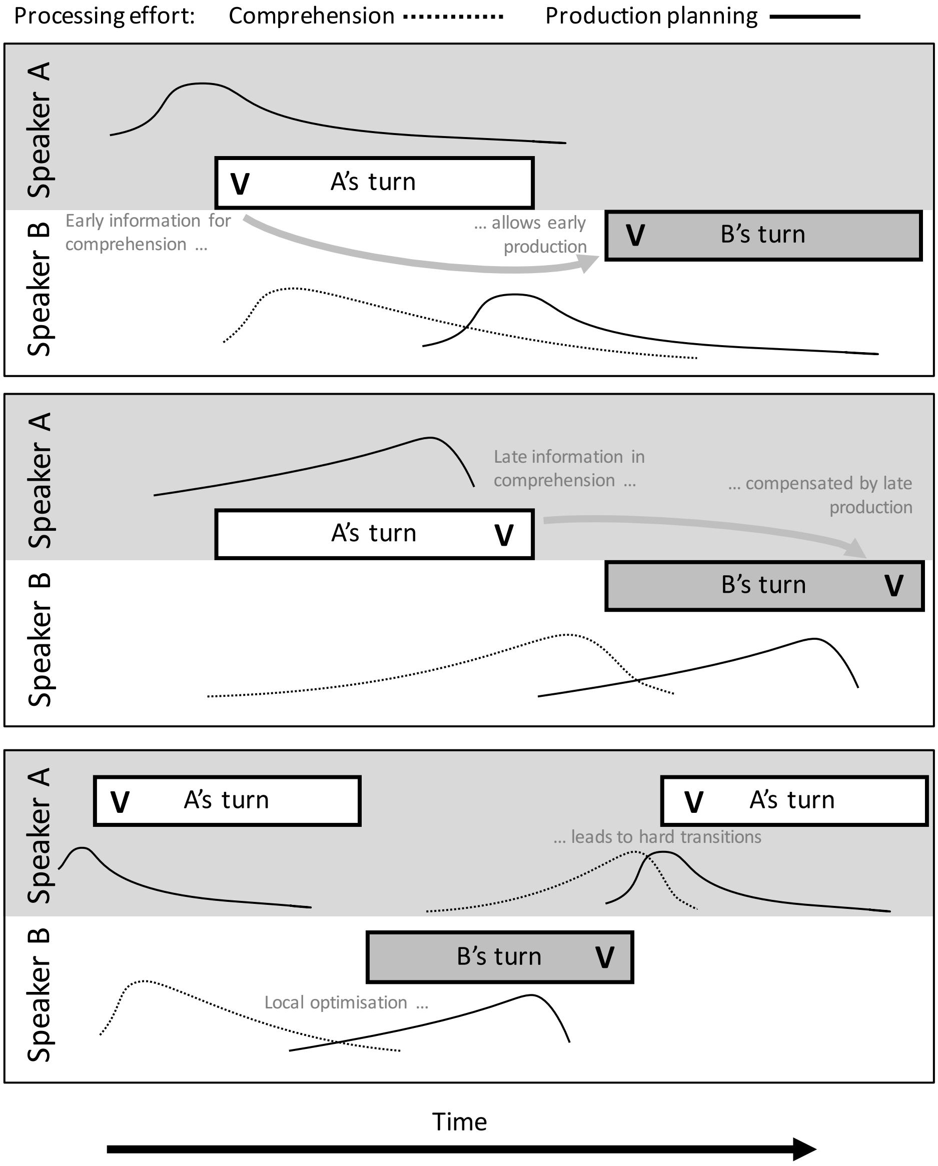

Figure 2: A schematic representation of the timeline of turn taking and the processing effort for comprehension and production. Speaker A and B take turns at speaking, placing the crucial information – the verb – at different points in the turn. Curves show the processing effort for comprehending their interlocutor’s turn and planning their own turn. Top: Verb-initial order provides information for the listener early in the sentence, allowing them to begin planning earlier. Middle: Verb final order provides information late, meaning that planning must start later, but this can be compensated by leaving the planning of the production of the verb until later. Bottom: Speakers could maximize the distance between verbs locally, optimizing the spread of processing that B has to do. However, this leads to a difficult subsequent transition for A, who has simulataneous high comprehension costs and high production planning costs.

Another solution might be to put the crucial verbal or predicate information in the middle of the utterance. This balances the distance from the crunch point for both comprehension and planning. This has the added bonus of preserving crucial information from overlap – the tendency for a small percentage of turns to be just slightly mistimed, with a second speaker coming in a bit early. This looks like a good compromise solution, again keeping maximal distance between successive predicates. In all cases, we see that the structure of A's turn has a knock-on effect on B's turn structure. Any strategy can facilitate turn taking, as long as everyone is using the same strategy.

We should note here that these considerations obviously oversimplify conversational exchanges which are often elliptical, but the point is that where full clauses are involved, they should be subject to constraints of this kind. These could – indeed should – have implications for how languages change over historical time, that is the cultural evolution of linguistic structure. We would predict that a language would be more likely to change to facilitate better turn taking than in the opposite direction. 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.

This can be tested in the following way. First, we identify a constraint that turn taking makes on a particular linguistic structure. That should lead to some predictions about the distribution of that structure we should see in the world's languages. We can then test whether the prediction can be observed in real data.

However, this involves two challenges. First, the precise interactions between conversation, cognition and cultural evolution are not easy to predict, since they form a complex system. In order to generate predictions, we implement a simple agent-based model of turn taking. Computational agents are simple computer programs whose behaviour we can specify. By placing many agents together in a model, we can see how they interact. In other words, the model helps us to generate predictions from our assumptions. In the sections below, we define and explore such an agent based model of cultural evolution through conversation.

The second challenge is testing whether the predictions from the model fit data in the real world. This is also not straightforward because the actual distribution of linguistic structures in the world are complicated by historical factors (for example, the colonizing success of particular social groups). In the next section, we explain this further and estimate the target phenomena which should emerge in the model.

2 Identifying the target phenomenon

We would like to account for two basic phenomena in word order patterns. First, for the vast majority of language communities, speakers use the same basic word order for expressing the same kinds of meanings. There is certainly optionality within languages, and individual variation. For the most part, however, speakers do not use completely random word orders. Dryer (2013a) notes that under 14% of languages can be said to have no dominant word order, but we speculate that in conversation these too will mostly have a statistically dominant pattern . That is, basic word order is nearly always coordinated within a language community.

The second phenomenon is that some basic word orders are more frequent than others. For example, if we count the raw number of basic word orders, then the pattern we see is that SOV and SVO more frequent that VSO order. However, this does not take into account the historical relations between languages. For example, many Celtic languages are VSO, just as nearly all Dravidian languages are SOV, but the Celtic languages are all related historically, so it would bias the sample to count each as an independent data point (see Roberts and Winters, 2013; Dunn et al., 2011).

In this study we will use Harald Hammarstrom's estimation of word order types in language isolates, that is, languages that are not known to be historically related to any others, and thus approximate to fully independent data points.[1] This also happens to be close to other estimates based on using non-isolates and controlling for historical relations. This turns out to be 11% VSO, 16% SVO, 66% SOV and other orders account for 7%. That is, the further from the start of the sentence the verb is, the more frequent that word order type turns out to be. The majority of the world's languages place the subject before the object in canonical transitive sentences, so we focus on those, but the model below does not actually distinguish between subjects and objects - only the position of the verb is important in the models below.

In later sections, we also look at the interaction between basic word order and other typological variables. In this case, we use data from the World Atlas of Language Structures (Haspelmath et al., 2008) in a mixed effects model. We use this to estimate the relationship between basic word order and other typological features while taking into account historical relations. See the supporting information for details and results.

[1] This is an approximation because with further study some isolates may prove to be actually distantly related to known languages families, and indeed ultimately, all languages may be historically related. What is likely though is that isolates have gone their separate ways in cultural evolution over millennia.

3 A computational agent based model of turn taking

We model a conversation as an interaction between two computational agents A and B. Agent A produces a turn at talk which consists of three abstract elements - a verb, a subject and an object. There are three turn types of word order in the model - VSO, SVO and SOV. The agents do not understand these elements, and there is no meaning associated with the elements – the model simply captures the idea that in each turn there is some linear order, with some elements (e.g. the verb) being more crucial than others.

Each agent has an exemplar memory which stores all the turns it has heard. When agents produce a turn at talk, they select one turn from their memory at random to be the template for their utterance.

Once A has produced a turn, agent B now has to decide how to respond by choosing a template turn from their own memory. However, we constrain the probability of choosing different turn types according to the distance between the verbs in the sequence. For example, if A produces a VSO turn, then then B has more time to process this information and so is more likely to be able to produce a verb at the start of their sentence. If A produces an SVO turn, then this verb is closer to the crunch zone and B is less able to produce a verb-initial turn. If A produces an SOV turn, then the verb is in the crunch zone and so B is very unlikely to be able to produce a verb-initial sentence in time, and quite unlikely to be able to produce a verb-medial sentence in time.

To model this, each item in the agent's memory is given a weight which affects its probability of being chosen. If A produces a turn T1 which has the verb at position V_1 (start = 0, middle = 1, end = 2) and a length L_1 (at this stage, all turns have a length of 3), then a responding turn by B, T2, which has the verb at position V_2 is given the following weight,

$W_{T2} = ((L_1 - V_1) + V_2)^{\alpha}$

where α is a parameter which controls the strength of the effect. When $\alpha = 1$, then the weight increases linearly as the distance between the two verbs increases. The probability of choosing item *i* from a memory which contains M items is then directly proportional to its weight.

$$P_i = \frac{W_i}{\sum_{x=1}^{M} W_x}$$

Put another way, agents are less likely to choose turn structures which involve more verb processing in the crunch zone. The α parameter, then, controls how quickly the processing cost increases with time. This mechanism captures the basic idea that the location of crucial information in A's utterance has a knock-on effect for the structure of B's turn. The constraint on B's choices are greatest when A produces a turn with the verb at the end.

Conversations proceed in the following way. A produces a first turn by selecting randomly from her memory. B then produces a turn, drawing from his memory according to the weight function above. Then A produces a third turn, weighting her selection by the turn type that B produced. Then B responds, and so on.

Conversations are independent from each other, and always start with an un-weighted selection. Therefore, we can manipulate the strength of the effect from turn taking. For example, agents can have one conversation of three turns, which imposes a constraint after each turn, or three conversations of a single turn, in which case the turn taking constraints have no effect. The greater the number of turns in a conversation, the greater the knock-on effect of the crunch zone. In each generation (see below), agents will have $N_{conversations}$ conversations with N_{turn} turns each.

We also model a small amount of noise in communication. With a small probability β, an agent produces a random turn type from all possible turn types.

3.1 Cultural evolution

Now we need a model of cultural evolution. We start with a small population of N_{agents} 'adult' agents. Each agent is initialised with a random selection of turn types in their memory. This means that populations are initialised with no bias in their word order preferences. Each agent is randomly paired with another agent and they have a conversation with N_{turn} turns. This repeats until they have had $N_{conversation}$ conversations. This results in a series of turns and conversations, and we can measure the frequency of each turn structure.

At the same time, there is a second population of 'child' agents listening to the conversations of the adult population and 'learning' from them by adding their turn structures to their exemplar memory. That is, generation 2 are like children acquiring language. When the adult generation are done with their conversations, they are removed from the population and the child generation 'grows up' and become adults. This new generation starts having conversations in the same way as the first generation, while a new child generation (generation 3) listen and learn.

This repeats for $N_{generations}$ generations. For each generation, we can track how the proportions of each type of sentence change.

3.2 Sentence particles

We can expand the model again to explore more complicated interactions between grammar and turn taking, for example the role of sentence final particles. Tanaka (2000; 2005) notes that the grammar of Japanese limits the projectability of turns. The predicate comes at the end of the sentence, and the sentence can be widely transformed by elements that come after the predicate. This appears to work against rapid turn taking. However, sentence final particles can potentially act as a 'buffer' which push crucial information away from the crunch zone and allow more time for the next speaker to plan their turn (this insight from Kobin Kendrick, 2010, see figure 3).

In the example of Japanese conversation in figure 4, we see that the sentence final particle is appearing constantly in overlap. This suggests that they can be treated as non-crucial elements of the turn (the overlap in the example can be partly attributed to the general projectability of the sentence in which the two speakers are agreeing with each other, but in general particles are not overlapped). A theory based on ease of production or perception which does not consider relationships between turns would have a hard time explaining why speakers bother to include these.

In this case, turn final particles seem to aid turn-transition in this verb-final language. However, the general prediction about which word order would benefit from final or initial particles is difficult to make. If a language is verb-initial, should sentence particles come at the start of the turn, or the end of the previous turn? At the beginning they would help to buffer the production by the speaker, while at the end they would serve to buffer the next speaker's production problems. Both would be logically helpful, but which are more likely to emerge? Are there some word orders which are less likely to need particles at all? It is difficult to work out the logical implications in a cultural evolutionary system, but this is precisely what the model is for. We can use it as a kind of transparent thought experiment.

Sentence particles were included in the model as follows. As well as the three basic word order types, agents could also produce versions with a sentence final or sentence initial particle (thus 9 combinations of types to choose from). Turn types with particles were less likely to be picked for production, since they are slightly longer (agents prefer to produce shorter turns). The relative length of particles to other words (verb, subject and object) could be manipulated via a parameter *p*. From the examples in Japanese, we would expect particles to be shorter than most words. The inclusion of a particle which added distance between verbs in a turn boosted the possibility that the verb can come earlier in a following sentence.

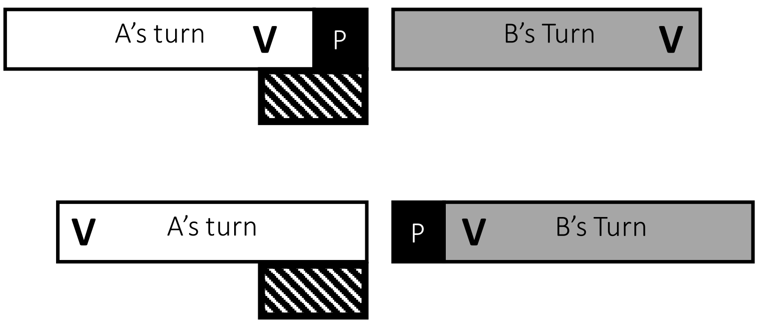

Figure 3: Sentence particles 'P' can act as a 'buffer' between turns, taking the crucial information away from the crunch zone.

```
W: 'N:  soo [ne
   yeah so  [FP
  "Yeah isn't it?"

G:          [Sore wa  aru   deshoo[: ne
            [that TOP exist COP   [  FP
            ["That's quite plausible, isn't it"

W:                                 [Soo na n de[shoo ne
                                   [so COP N  C[OP   FP
                                   ["That's probably right, isn't it?"

G:                                             ['N ...
                                               [yeah ...
```

Figure 4: A conversation in Japanese. Square brackets indicate where the next speaker overlaps with the previous one. The utterance final particles are in bold. Adapted from Tanaka (2000), Tokyo 7, p.26.

3.3 Summary of assumptions

Here we summarise the basic assumptions and simplifications of the model:

- All turns contain verbs
- We do not model semantics or detailed syntax/morphology
- Speakers must minimise gaps and overlaps
- Planning crucial elements is increasingly difficult as they approach the 'crunch zone'
- Verbs are crucial elements (they are hard to plan)
- The production cost of sentence is related to sentence length (though in the main model all sentences have the same length)
- In cultural evolution, agents learn by observing others and storing examples of behaviour

• Generations are discrete

Clearly, these assumptions are idealisations, and the actual factors are much more complex than this. As noted earlier, the assumption that all turns contain verbs is clearly counterfactual, given the elliptical nature of many responses. However, as a starting point, we think that this model captures some of the crucial constraints on interactive language use under temporal pressure. We are attempting to construct the simplest model which will help us think about the intricate inter- relationships between conversation, cognition and cultural evolution. One way to construe the model is that it captures only some conversations, not every interaction between agents, and that the selective pressure only applies in turns which match the conditions above.

4 Results

Figure 5 shows, as an example of the kinds of results obtained, three independent runs of the model with a population of 10 agents taking 2 conversations of 10 turns each (noise level $\beta = 0.01$). Along the horizontal axis we see generations and each line represents how the frequency of each type of basic word-order (or major sentence type) changes over time. We see that in the first generation, agents are equally likely to use any of the three types, but that the use of VSO rapidly declines. 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.

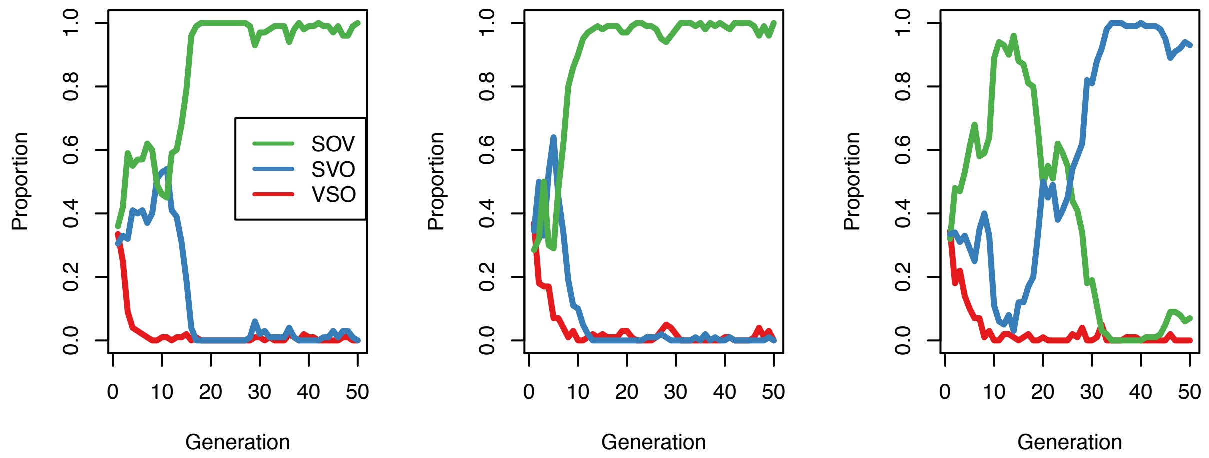

Figure 5: Proportions of each turn type used at each generation for three independent runs of the main model.

In fact, we ran the model 1000 times and measured the proportion of runs that converge to each word-order type on each run. In every simulation, the population converged on a single word order type within 100 generations, itself an interesting result. Figure 6 shows the resulting proportions of word orders in two different conditions (α = 0.1). When agents only have conversations with 1 turn (no constraints from turn taking), then each word order type is equally likely to win. When turns follow each other within a conversation, the proportions look very close to the actual ‘natural’ distribution of word orders we see in real languages, as measured by the proportions of word-orders in the language isolates of the world, where SOV is most frequent followed by SVO and VSO.

Essentially, the turn taking constraints impose a bias for pushing the verb out of the crunch zone to the end of the sentence. However, one crucial result is that although there is a small proportion of populations with VSO order, within those populations all agents are using VSO order. That is, the model is producing the two target phenomena: convergence within populations and a bias for verb-later orders across populations.

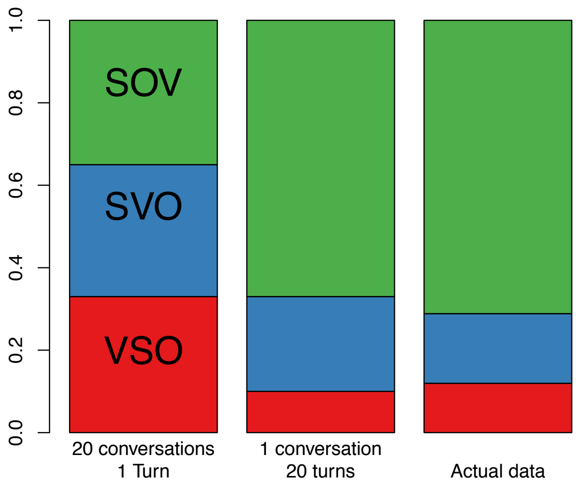

Figure 6: Proportions of each turn type that 1000 generations converge to in: a model without pressures for turn taking (left); a model with turn taking constraints (middle); and actual language data from the world's isolates (right).

The results in figure 6 fit the data qualitatively, but also quantitatively (the proportions as well as the ranks are quite close to the real ones). The quantitative fit depends on the parameters of the model. Figure 7 shows how the distribution of word order types varies with the α parameter, which controls how the distance between verbs relates to the processing cost. When α is close to 0, there is little difference between each of the sentence types in any context, and roughly the same proportion of each sentence type emerges. When α is positive, reflecting greater processing cost as the verbs enter the crunch zone, then the SOV advantage appears. If processing cost scales linearly (α = 1), then the model predicts that almost all languages should show SOV order. With negative values of α, where cost decreases as the verb enters the crunch zone, we see a preference for VSO languages. This suggests that the best fitting assumption would be for a positive, convex function: the cost is large for verbs inside the crunch zone, but rapidly declines as the verb moves further away.

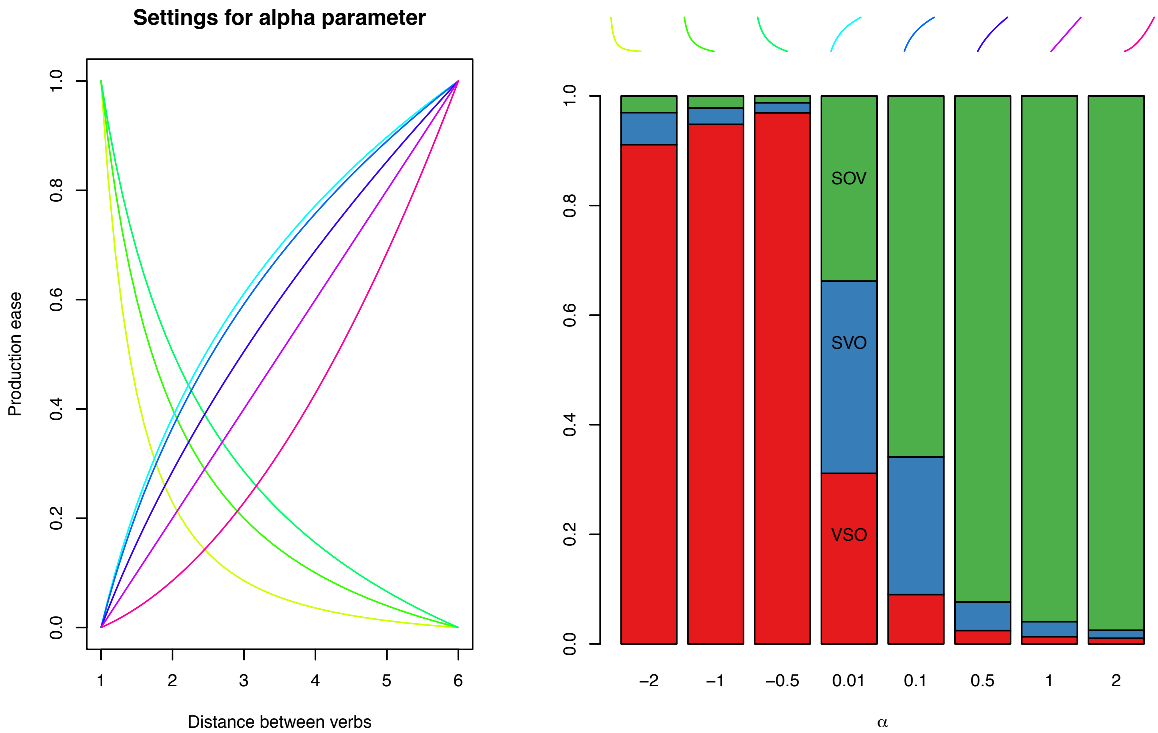

Figure 7: Right: how the α parameter affects the function which relates the distance between verbs in neighbouring turns and the ease of producing the subsequent turn on time. Left: how the proportions of different word-order types varies with the α parameter.

The supporting information shows that the model results are robust to settings of various parameters, including N_{agents}, $N_{conversations}$, N_{turns} and β.

4.2 Sentence final particles

Figure 8 shows some results for sentence final particles ($\alpha = 0.1$, $\beta = 0$, $p = 0.5$, $N_{agents} = 10$, comparing 20 conversations of 1 turn with 10 conversations of 2 turns). The model without turn taking constraints predicts that languages are similarly likely to have initial or final sentences regardless of verb position. However, with the constraint we see two things: Initial particles are more likely than final particles for verb initial languages, and that, for verb final languages, final particles are proportionately more likely. That is, if a language happens to settle on verb final structures, it is also more likely to develop sentence final particles. This prediction also matches the real data quite well (data from position of polar question particles, Dryer, 2013b, see SI). Interestingly, it also predicts that verb final languages should be less likely to have particles at all.

However, this result was not robust to changes in parameters. The fit to the data was better when noise level was low, and in addition the inclusion of a question particle in a buffer zone had a big effect. This is a reasonable result, given that the first model predicted that the processing cost declines rapidly as the verb moves away from the crunch zone. Outside of a narrow window around the parameters above, the predictions range from no effect to the opposite of the effect we see in the data (final particles more likely for verb-final languages). This suggests that the use of particles to buffer interactive language use emerges only under specific conditions.

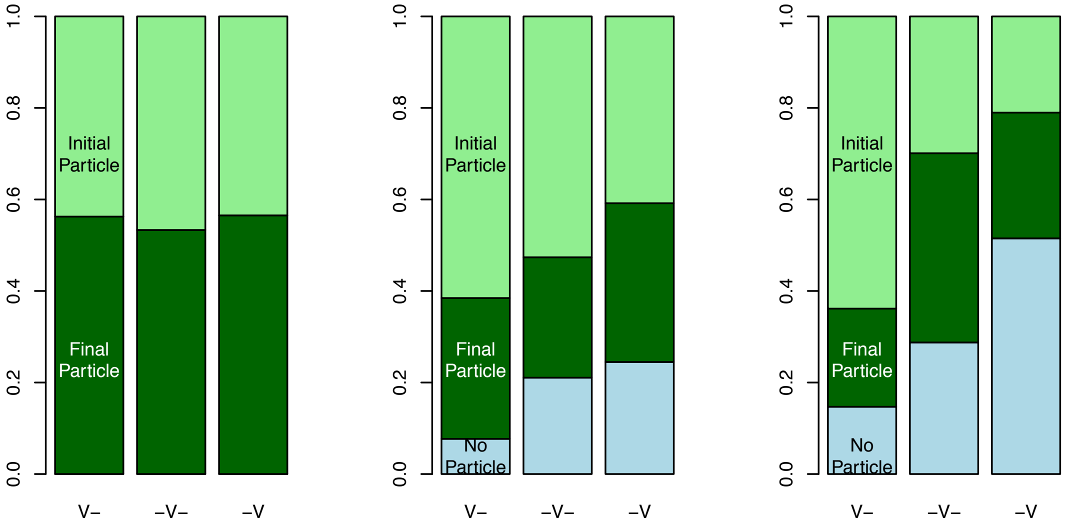

Figure 8: Distribution of word order types and the presence of absence of sentence particles.

5 Discussion

In this article, we have suggested that turn-taking in conversation imposes constraints on the efficiency of different basic word orders in interactive language use. Languages should adapt to these constraints, and we should see evidence of this adaptation in the structures of the world's languages. Support for this idea can be found by identifying a set of constraints that conversation imposes, generating a prediction about the distribution of linguistic structures that should emerge from these constraints, and then testing this prediction against real data. We have suggested that the need for rapid turn-taking imposes a 'crunch zone' for online language processing around the ends of turns, and hypothesised that this might affect the optimal position of crucial elements in a clause. We presented an agent-based model to help generate predictions about how these constraints should affect the cultural evolution of language, then compared the results to real data. We found a reasonable qualitative and quantitative match between the output of the model and the distribution of basic word orders in the real world.

The model suggests that speakers within a culture will tend to co-ordinate their grammatical structures, such that genuine free word-order is unlikely to survive in a conversational context, and that any of the basic orders can become the conventional way of communicating. However, because the structure of a prior turn has knock-on effects for the production of the next turn, there is a bias for cultures to evolve towards pushing the verb further back in the turn. This leads to a distribution of basic word order which mirrors the distribution we see in the real world.

There are many issues to resolve. The model is extremely simple and makes many assumptions that could be relaxed. The parameters also need to be tied to specific cognitive mechanisms, rather than abstract notions of processing cost. Rules of the sequential organisation of conversation could also be built into the model. The model also makes more general predictions about grammatical structures within conversations which could be tested. For example, do speakers alter the information structure of their turns to aid processing through by local co-ordination? Finally, the constraints from turn taking are just one domain from many that impact the evolution of grammatical structure. Despite these limitations, we believe that the model provides a useful tool for thinking about the relationship between conversation and cognition in a cultural evolution framework.

Acknowledgements

This research was supported by a European Research Council Advanced Grant No. 269484 INTERACT to Stephen Levinson. We thank the Max Planck Institute for additional support.

References

Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N.P., Rai, I.P., Rai, M., Rai, N.K. and Stoll, S. (2007). Free prefix ordering in Chintang. *Language*, 83(1), 43-73.

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never Say No… How the Brain Interprets the Pregnant Pause in Conversation. *PloS one*, *10*(12), e0145474.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, *31*(05), 489-509.

Croft, W. (2000). Explaining language change: an evolutionary approach. Harlow, Essex: Longman.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S.C., Manrique, E., Rossi, G & Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PloS one*, *10*(9), e0136100.

Donders, F. C. (1868). La vitesse des actes psychiques. Archives Néerlandaise, 3, 269–317.

Dryer, M. S. (2013a). Order of Subject, Object and Verb. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Dryer, M. S. (2013b). Prefixing vs. Suffixing in Inflectional Morphology. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79-82.

Ferrer-i Cancho, R. (2008). Some word order biases from limited brain resources: A mathematical approach. Advances in Complex Systems, 11(03):393–414.

Goldin-Meadow, S., So, W. C., Özyürek, A., and Mylander, C. (2008). The natural order of events: How speakers of different languages represent events non- verbally. Proceedings of the National Academy of Sciences, 105(27):9163–9168.

Haspelmath, M., Dryer, M. S., Gil, D., and Comrie, B. (2008). World Atlas of Language Structures. online at http://wals.info. Accessed 2013-04-18. Munich: Max Planck Digital Library.

Hawkins, J. (1994). A performance theory of order and constituency, volume 73. Cambridge University Press.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11-26.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (ed.) *Meaning, form, and use in context: Linguistic applications*. Georgetown University Press, Washington DC. 11-42. Reprinted in Kasher (ed., 1998), vol. IV: 389–418.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology*, *2*(255), 1-16.

Kendrick, K. H. (2010). Epistemics and Action Formation in Mandarin Chinese. PhD Thesis, University of California, Santa Barbara

Kendrick, K. H. (2015). The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation. *Frontiers in psychology*, 6, 250.

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: a quantitative study. *Discourse Processes*, *52*(4), 255-289.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(01), 1-38.

Levinson, S. (2006). On the human interaction engine. In Enfield, N. and Levinson, S., editors, Roots of Human Sociality: Culture, Cognition and Human Interac- tion, pages 39–69. Oxford: Berg.

Levinson, S. C. (1983). Pragmatics. Cambridge University Press. Roberts, S. G. and Winters, J. (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. PLoS One, 8(8):e70902.

Levinson, S. C. (2016). Turn-taking in Human Communication–Origins and Implications for Language Processing. *Trends in cognitive sciences*, *20*(1), 6-14.

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, *48*(5), 331-354.

Roberts, S., & Winters, J. (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PloS one*, *8*(8), e70902.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 40(4):696–735.

Schegloff, E. A. (1989). Reflections on language, development, and the interactional character of talk-in-interaction. Interaction in human development, pages 139–153.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 361-382.

Schouwstra, M. and de Swart, H. (2014). The semantic origins of word order. Cognition, 131(3):431–436.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., and Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. Proceedings of the National Academy of Sciences, 106(26):10587–10592.

Tanaka, H. (2000). Turn projection in Japanese talk-in-interaction. *Research on Language and Social Interaction*, 33(1):1–38.

Tanaka, H. (2005). Grammar and the “timing” of social action: Word order and preference organization in Japanese. *Language in Society*, *34*(03), 389-430.

Thompson, S. A. (1998). A discourse explanation for the cross-linguistic differences in the grammar of interrogation and negation. *Case, typology and grammar: In honor of Barry J. Blake*, 309-341.

Zipf, G. K. (1949). Human behavior and the principle of least effort. Oxford, England: Addison-Wesley Press.