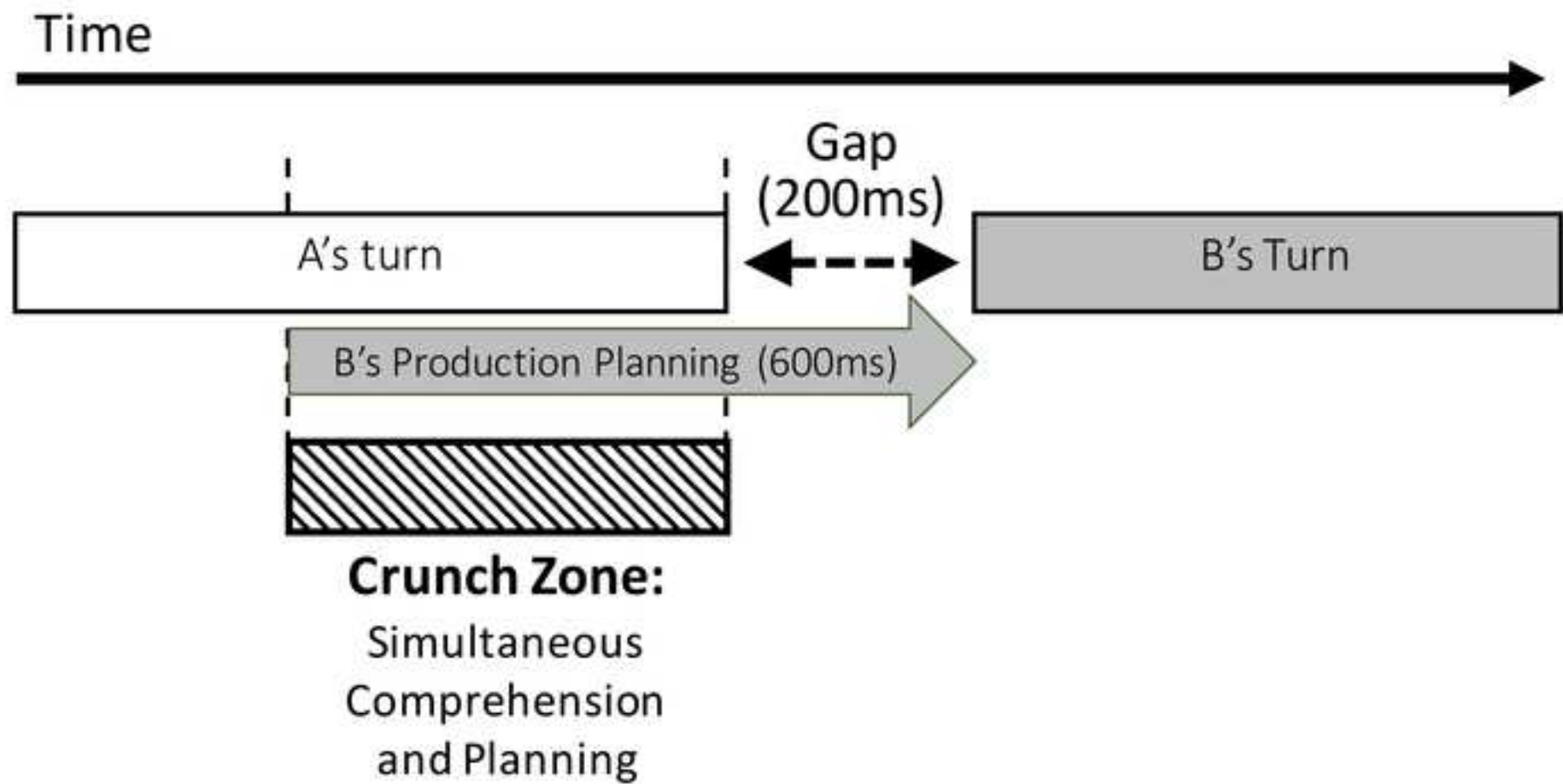# Interaction Studies

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

Manuscript Number:	IS-16017R2
Full Title:	Conversation, cognition and cultural evolution: a model of the cultural evolution of word order through pressures imposed from turn taking in conversation
Short Title:	Conversation, cognition and cultural evolution
Article Type:	Original Article
First Author:	Sean G Roberts
Other Authors:	Stephen C Levinson
Corresponding Author:	Sean G Roberts Max Planck Institute for Psycholinguistics Nijmegen, Gelderland NETHERLANDS

Section/Category:	Special issue: "Interaction and Iconicity in the Evolution of Language"
Keywords:	Turn taking; pragmatics; Typology; Word Order; Cultural Evolution
Manuscript Classifications:	20.09: Modelling; 10.05.04: Pragmatics; 10.05.06: Psycholinguistics

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 dominant word orders across languages 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.

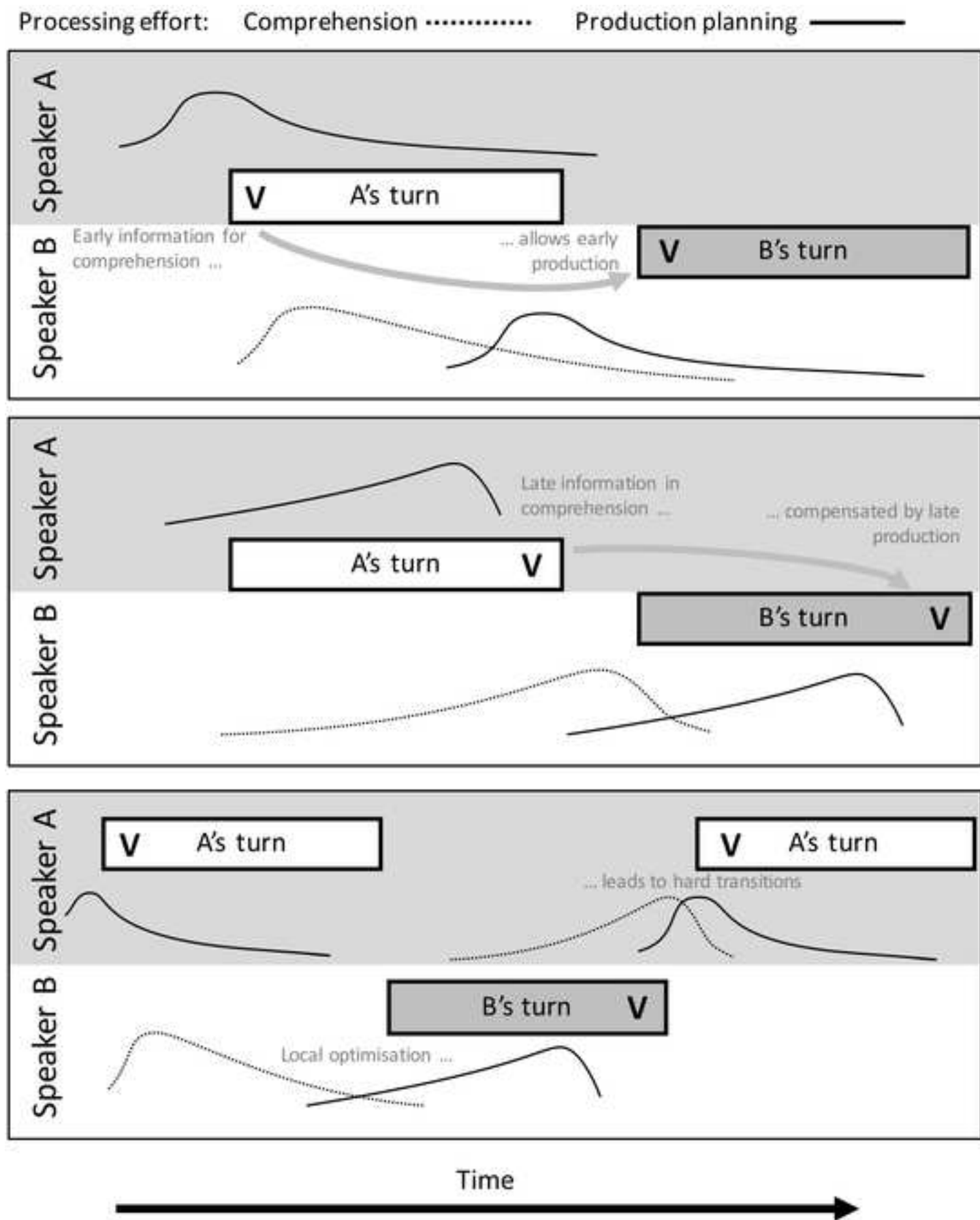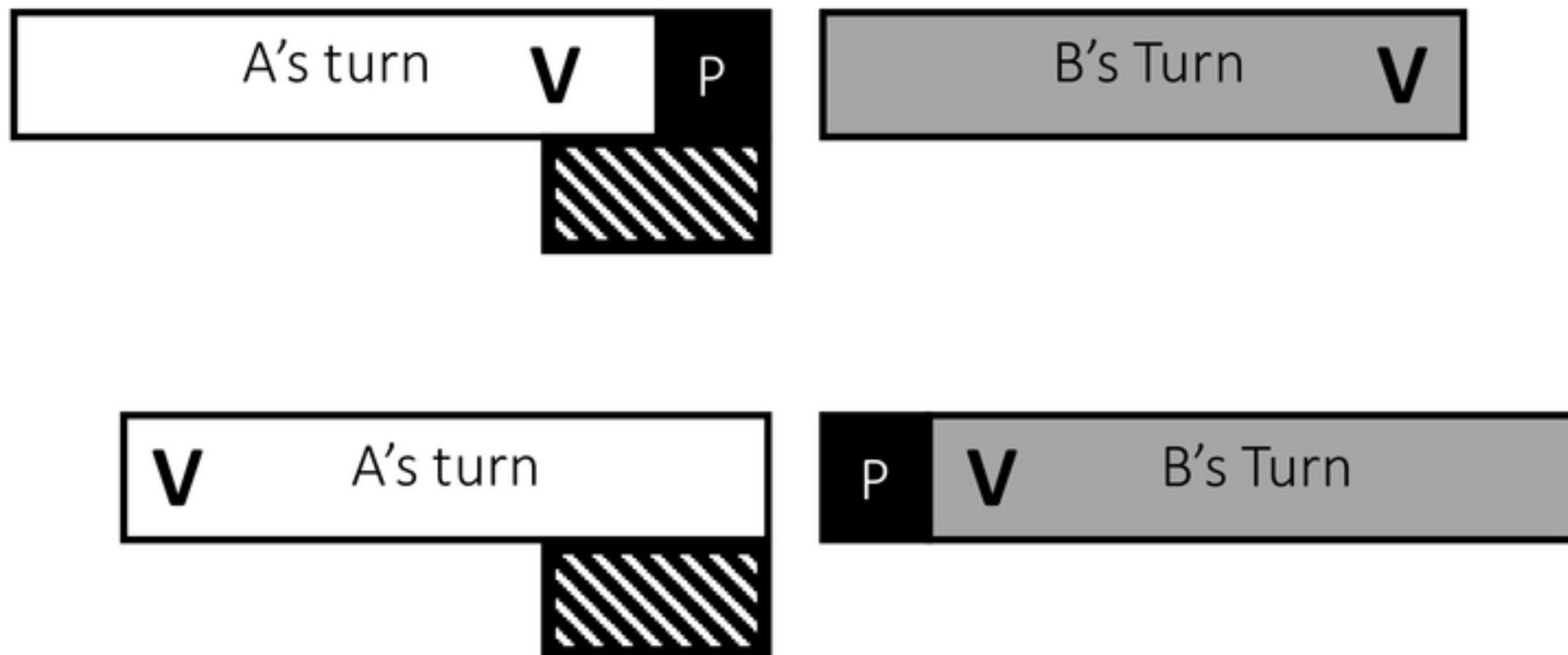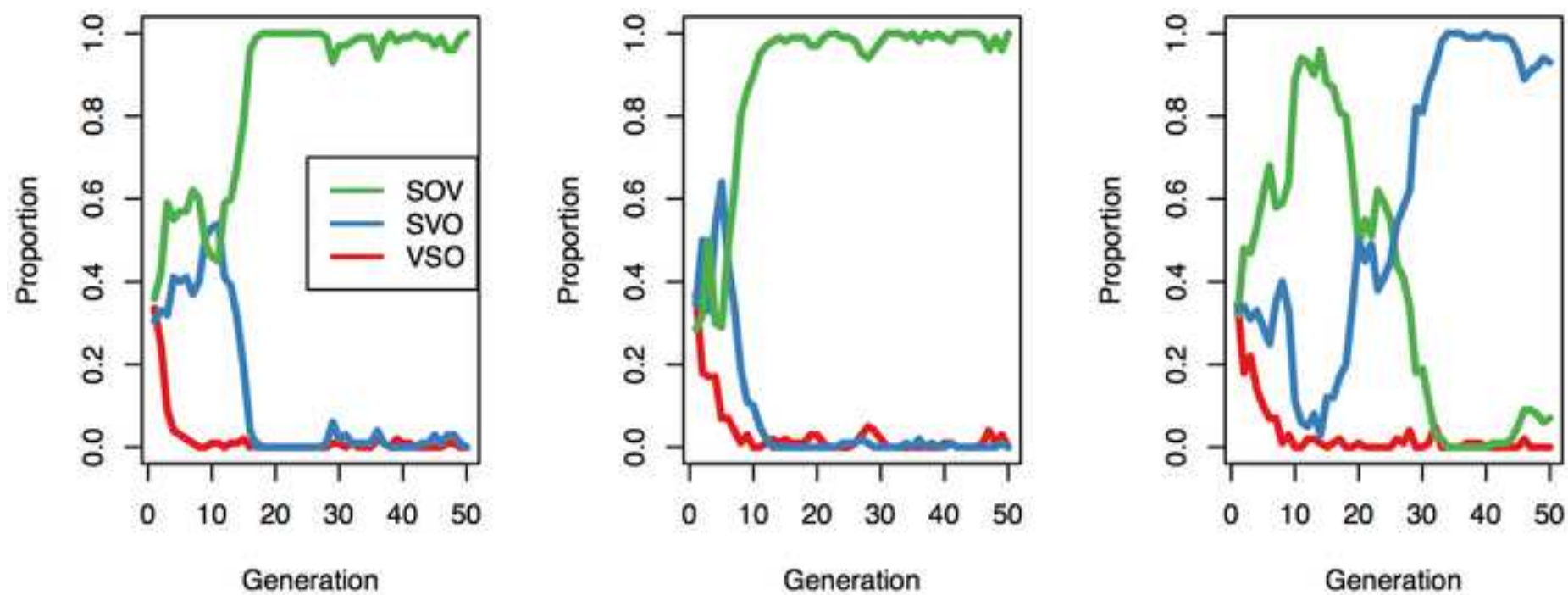|	
Author Comments:	To the editors, We submit a paper to the special issue on interaction and iconicity in language evolution. We presented this work at the ICLC conference panel which motivated this special issue. Our paper attempts to link cognitive processing during interaction to cultural evolutionary pressures on languages. As far as we know, this is the first computational model which tries to capture the links between conversation, cognition and cultural evolution. Best, Sean Roberts and Stephen Levinson
Order of Authors Secondary Information:	

Figure 1

Figure 2

Figure 3

Figure 4

```
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 5

Bio

Seán G. Roberts studies cultural evolution at the University of Bristol as part of the group *Exploring the evolution of cultural diversity*. 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.
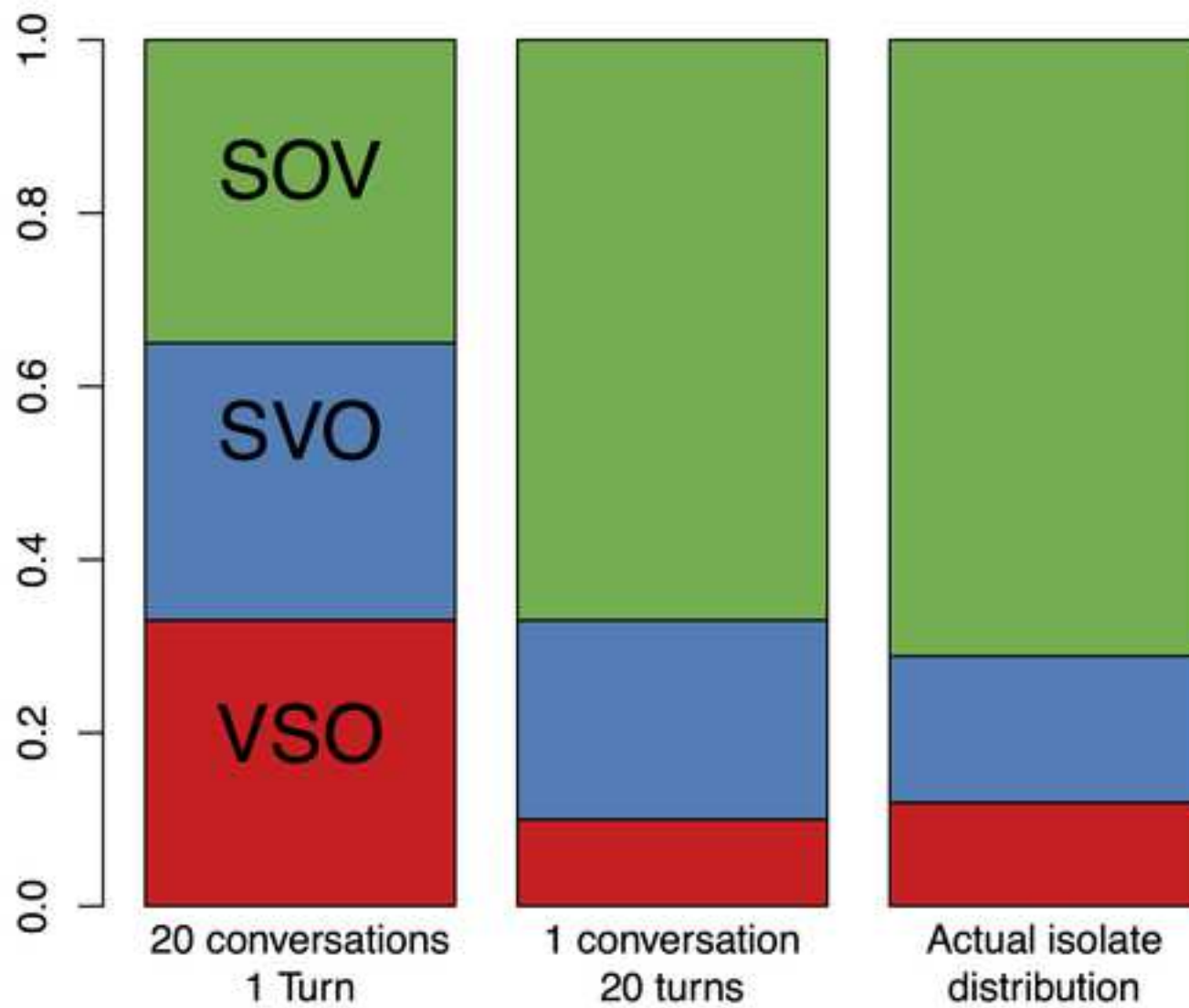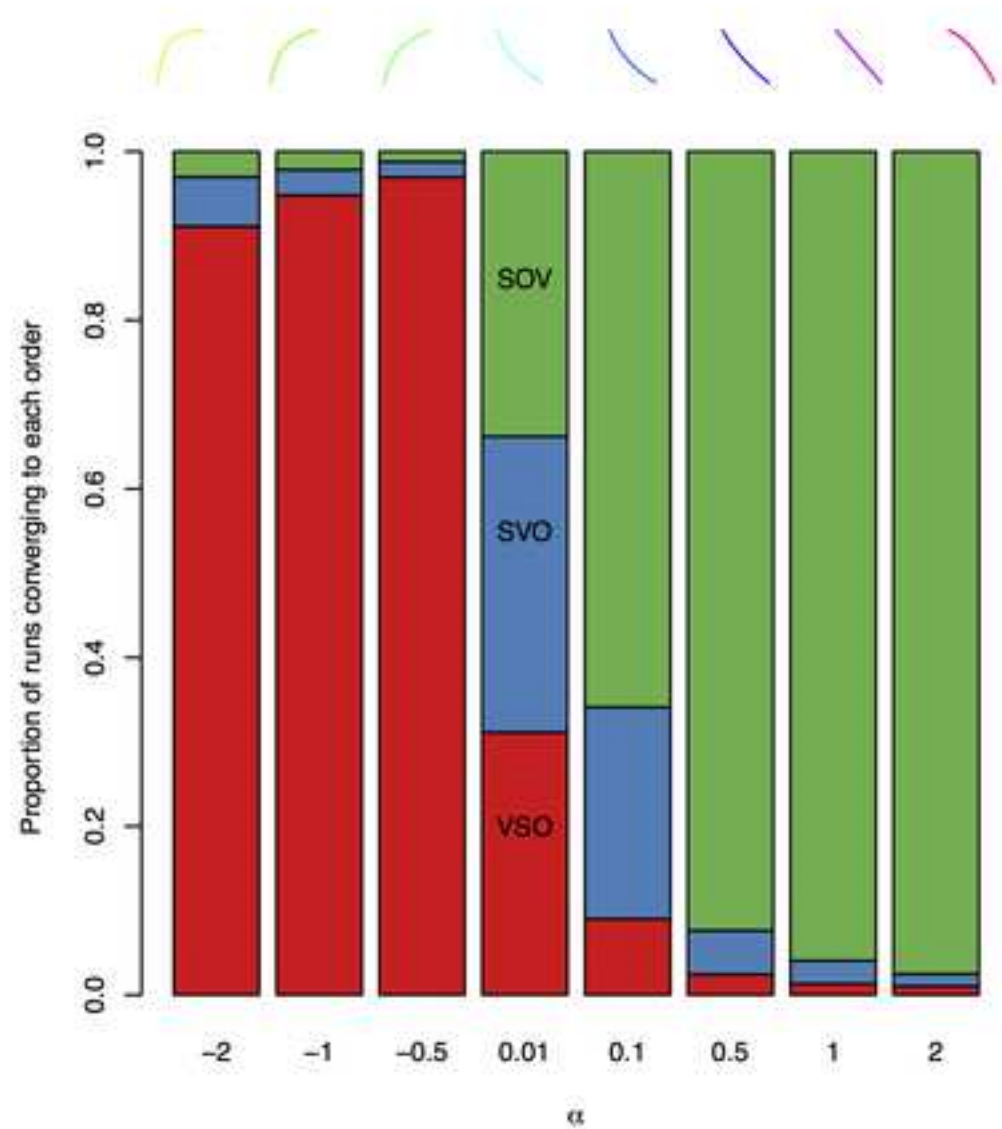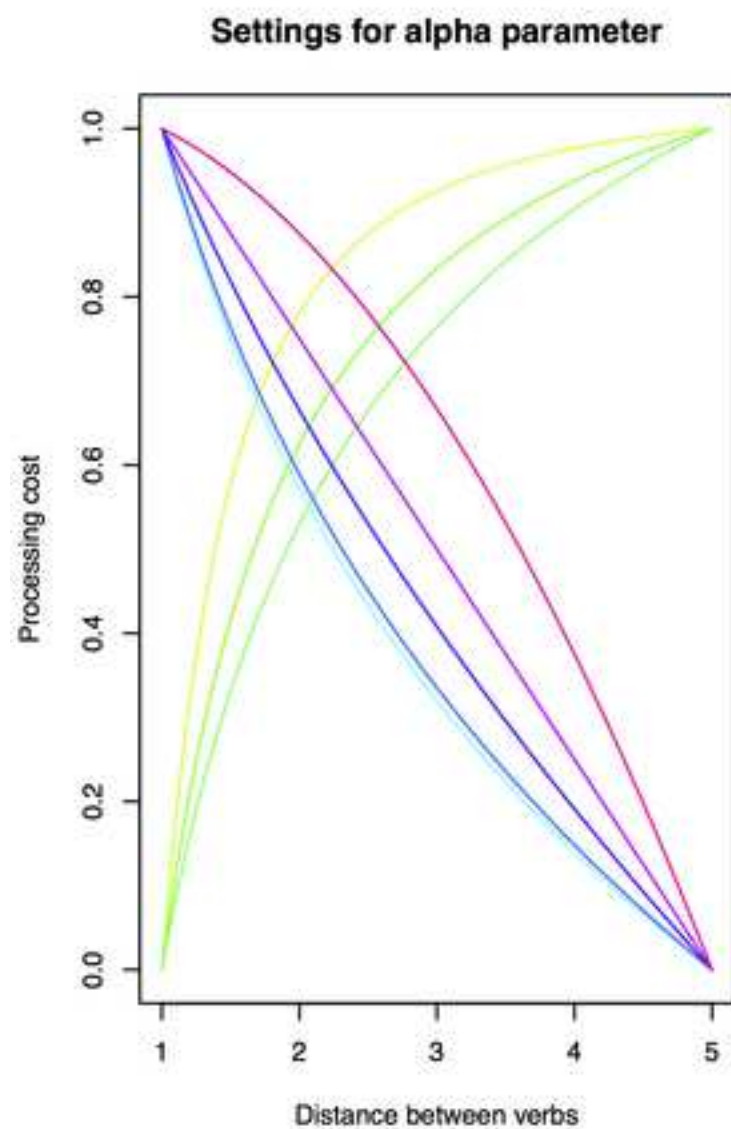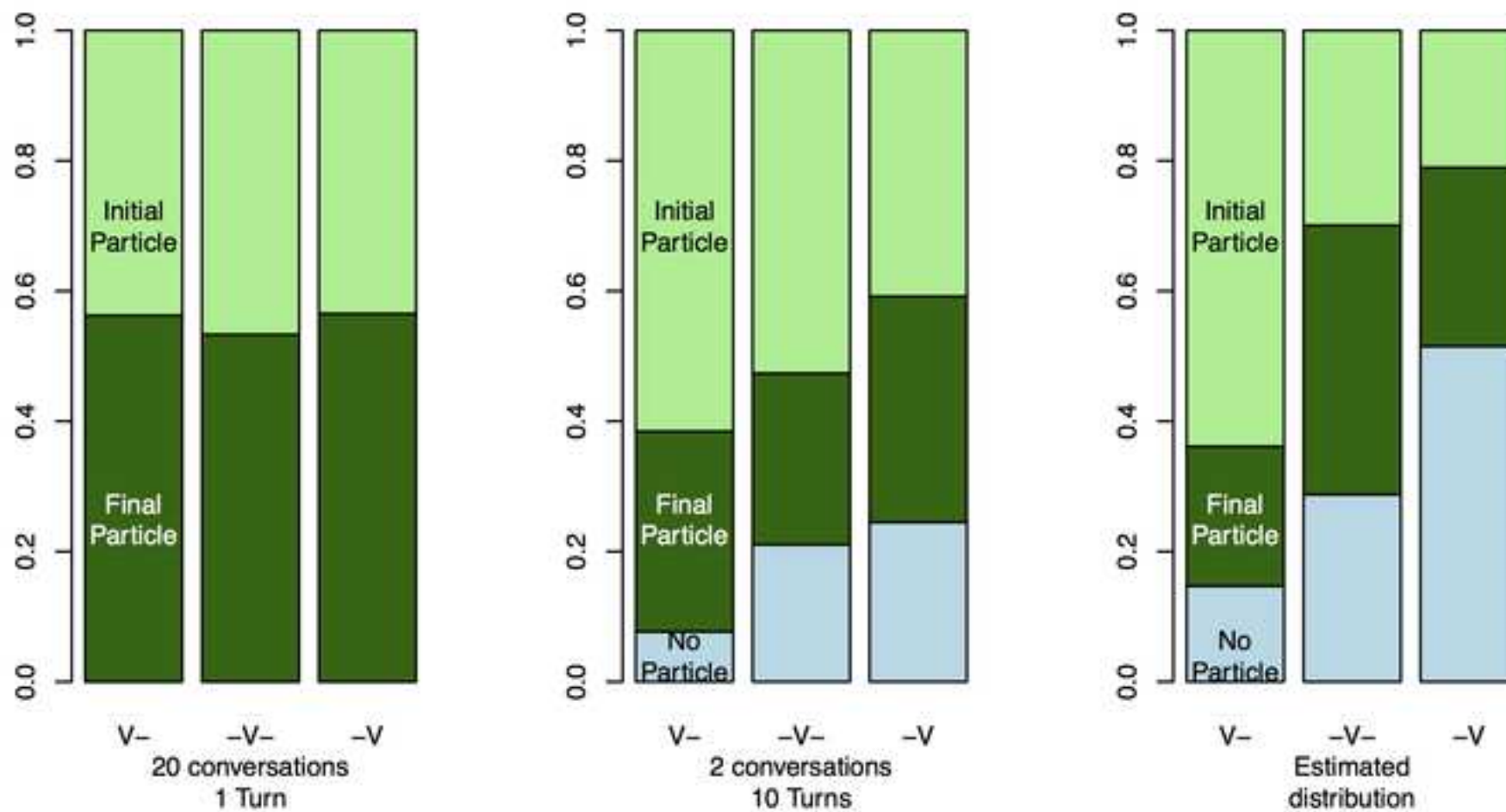
Figure 6

Figure 7

Figure 8

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[1,2] and Stephen C. Levinson[2,3]

1. Department of Archaeology and Anthropology, University of Bristol, 43 Woodland Road, BS8 1UU Bristol, United Kingdom

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

3. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN 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 dominant word orders across languages 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 University of Bristol as part of the group *Exploring the evolution of cultural diversity*. 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 including: efficient storage or processing (e.g. Krupa, 1982; Hawkins, 1994; Ferrer-i-Cancho, 2008; Ferrer-i-Cancho, 2015); the effectiveness of conveying semantic information (e.g. Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014; Gibson et al., 2013); semantic and syntactic restrictions (e.g. Tomlin, 1986; Christensen, Fusaroli & Tylén, 2016); acquisition (Lupyan & Christiansen, 2002); and information structure (e.g. Mithun, 1992).

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 (see also Givòn, 1983a; Ochs, Schegloff, & Thompson, 1996; Enfield, 2008). 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 (see below and Levinson, 2016). The demands of interactive conversation should therefore impose selective pressures on linguistic structures. If there is variation

65 in how effective different structures are in conversation, and if more effective structures

66 are more likely to 'replicate' and be used again, then this suggests that such structures

67 should be under selection over time by the forces of cultural evolution (Croft, 2000).

68 In other words, languages should change over time to better serve turn taking.

69 An example of this process links constraints from pragmatics to predictions about

70 typology. Thompson (1998) points out that interrogative structures make turn transition

71 relevant: a question demands an answer. Thompson argues that, in order to be effective,

72 interrogatives should generally apply to prosodic units, and therefore appear at turn

73 boundaries, rather than in the middle of turns. If interrogatives are morphologically

74 bound to the verb, this constraint leads to a specific prediction: languages that place the

75 verb at the end of a sentence should have interrogative suffixes (so that the interrogative

76 appears after the verb at the boundary), while languages with verbs at the beginning

77 should have prefixes (see supporting materials for an updated statistical test of this

78 claim). This is a well-known pattern in typology, but we suggest that part of the

79 pressure that leads to the emergence of this pattern could be motivated by the pragmatic

80 – and more specifically interactional - pressures on structures of this kind.

81 In this article, we consider a specific aspect of conversation - turn taking - and how the

82 tight processing constraints it entails may lead to the selection of specific grammatical

83 structures within a cultural evolution framework. While the work is preliminary, we

84 hope to demonstrate the possibility and promise of linking domains that are not usually

85 considered together: language structure, conversation, cognition and cultural evolution.

86 The paper is organized as follows. First, we review the literature on turn taking and

87 how it links to processing in conversation. Section 2 includes a brief review of the

88 literature on the typological distribution of word orders. Section 3 outlines a

89 computational model of the cultural evolution of word order under pressures from turn

90 taking. Section 4 shows the results of the model and section 5 discusses them. We

91 leave the relationship between our approach and others until the end when our position

92 is clearer.

93 **1.1 A cognitive pressure derived from turn taking**

3

94 In a conversation, speakers take turns at talking and try to minimise the amount of gap
95 or overlap between the turns (Sacks et al., 1974). When talking in groups, there is
96 competition for who speaks next (Levinson, 1983), and a delay in response is
97 pragmatically marked, for instance, it can be interpreted as unwillingness (Kendrick and
98 Torreira, 2015; Bögels, Kendrick & Levinson, 2015; Roberts, Margutti & Takano,
99 2011). This puts speakers under pressure to respond quickly in conversation.

100 Indeed, the average gap between questions and answers is around 200ms (Stivers et al.,
101 2009). What makes this surprising is that the time to plan and begin executing a *single*
102 *word* is at least 600ms (Indefrey, 2011). Even though speech planning is incremental
103 (speech may start before the whole sentence is planned, Levelt, Roelofs & Meyer,
104 1999), this implies that at some point we must be predicting the course of the incoming
105 turn, extracting its action or speech act, and preparing our response in advance of the
106 other speaker coming to a conclusion (Levinson, 2016). This imposes a kind of 'crunch
107 zone' in which production and comprehension must overlap in time (see figure 1).

Time

Gap
(200ms)

A's Turn

B's Turn

B's Production Planning (600ms)

Crunch Zone:
Simultaneous
Comprehension
and Planning

108
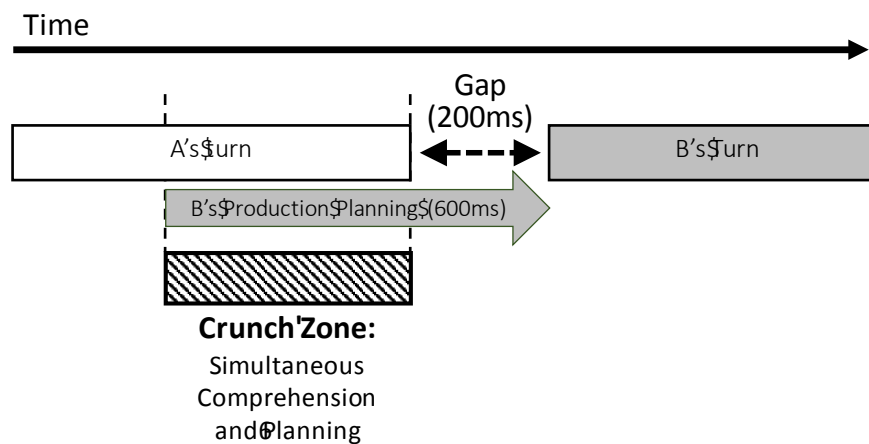
109 **Figure 1: A schematic representation of turn taking.**

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

4

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

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

129 **1.2 Linking processing and pragmatics**

130 We could go further in linking pragmatics and typology by integrating constraints from
131 online processing of interactive language use into a model of the cultural evolution of
132 language. We argue that languages do not adapt just to our individual cognition (cf.
133 Christiansen & Chater, 2008), but to the way we actually deploy the cognition in
134 interaction. It is not only the evanescent speech signal, but also the temporal pace of
135 conversation that makes the cognitive pressures on normal language use so intensive.
136 Therefore, one would expect the structure of language to adapt to this ecology, and we
137 should be able to see signs of these adaptations in today's languages. For example, one
138 possible locus of adaptation would be the order that information is presented in a turn.
139 Information presented to a listener later is more likely to occur inside the crunch zone,
140 and therefore present a greater challenge to producing the next turn on time.

141 Let us consider the implications for basic word order - that is, the order of the subject,
142 object and verb in a canonical transitive clause. Through its lexically-specified
143 argument structure, the verb provides the syntactic frame for a sentence and provides
144 crucial semantic information about the action reported. Hence its position in the
145 sentence might adapt to several processing pressures (see the final section for a

5

146 discussion of this assumption). Predictions here are complicated by the fact that the

147 functional adaptation of a sentence structure to its interactive use must be viewed from

148 two perspectives: the point of view of the speaker, and the point of view of the recipient

149 or comprehender. As has been noted in previous pragmatic work, what is good for the

150 speaker may be bad for the recipient, and vice versa (e.g. Hawkins, 2004; Langus &

151 Nespor, 2010; Jaeger, 2010; Piantadosi & Gibson, 2011; Fedzechkina, Jaeger &

152 Newport, 2012; Ferrer-i-Cancho, 2014;). Consider, for example, the structure of the

153 lexicon: making many semantic distinctions may be helpful for the recipient trying to

154 recover the speaker's intended referent, but force the speaker to make careful choices

155 between many alternatives (Zipf, 1949; Horn, 1984). In a similar way, verbs in final

156 position may give speakers more time to plan the most complex component of the turn.

157 On the other hand, verbs in initial position allow listeners to anticipate the unfolding of

158 the incoming turn, using the predictive possibilities offered by the verb's argument

159 structure, and thus start planning their own response much earlier. Here there is again a

160 zero-sum type of situation: what is good for the speaker (verbs at the end) is bad for the

161 recipient, and what is good for the recipient (verbs at the beginning) is bad for the

162 speaker (who must plan the whole sentence up front).

163 Notice that a mixed strategy will not help: if I put my verb at the end, it falls in your

164 'crunch zone', and it will be therefore especially difficult for you to put your verb at the

165 beginning – you will not have had time to formulate the response. However, if you put

166 your verb at the end too, then you will have most of the duration of the turn to plan the

167 verb, the complex frame for the sentence (Figure 2). Alternatively, suppose I am

168 considerate to you the recipient, then I could begin my turn with a verb, well clear of

169 your crunch zone, and now aided by my co-operative gesture and the following more

170 predictable components of the turn you will have time to compose your verb also in

171 initial position, so returning the favour (see Figure 2). Both strategies will get the

172 maximal distance between predicates, which is what will aid processing. Thus we

173 conclude that *coordination* of verb-placement, either at the end or at the beginning, is

174 strongly favoured by processing under rapid turn-taking. Even in languages with

175 flexible word order, we suspect that there are biases towards a particular word order in

176 everyday conversation (e.g. Samoan, Duranti, 1981, p. 171; Ochs, 1982, p. 661, see

177 discussion).

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

179 operation, to a selfish move: you could always suit yourself and return a verb-final turn.

180 These considerations suggest that while both solutions are viable, the verb-final solution

181 might predominate in cultural evolution.

182

Processing'effort:"   Comprehension ············   Production'planning ─────

Speaker'A

Speaker'B

Early'information' for'
comprehension' …

...'allows'early'
production

V   A's' turn

V   B's'turn

Speaker'A

Late'information' in'
comprehension' …

...'compensated'by'late'
production

A's' turn   **V**

B's'turn   **V**

Speaker'B

Speaker'A

V   A's' turn

V   A's' turn

...'leads'to'hard'transitions

V   B's'turn

Local' optimisation …
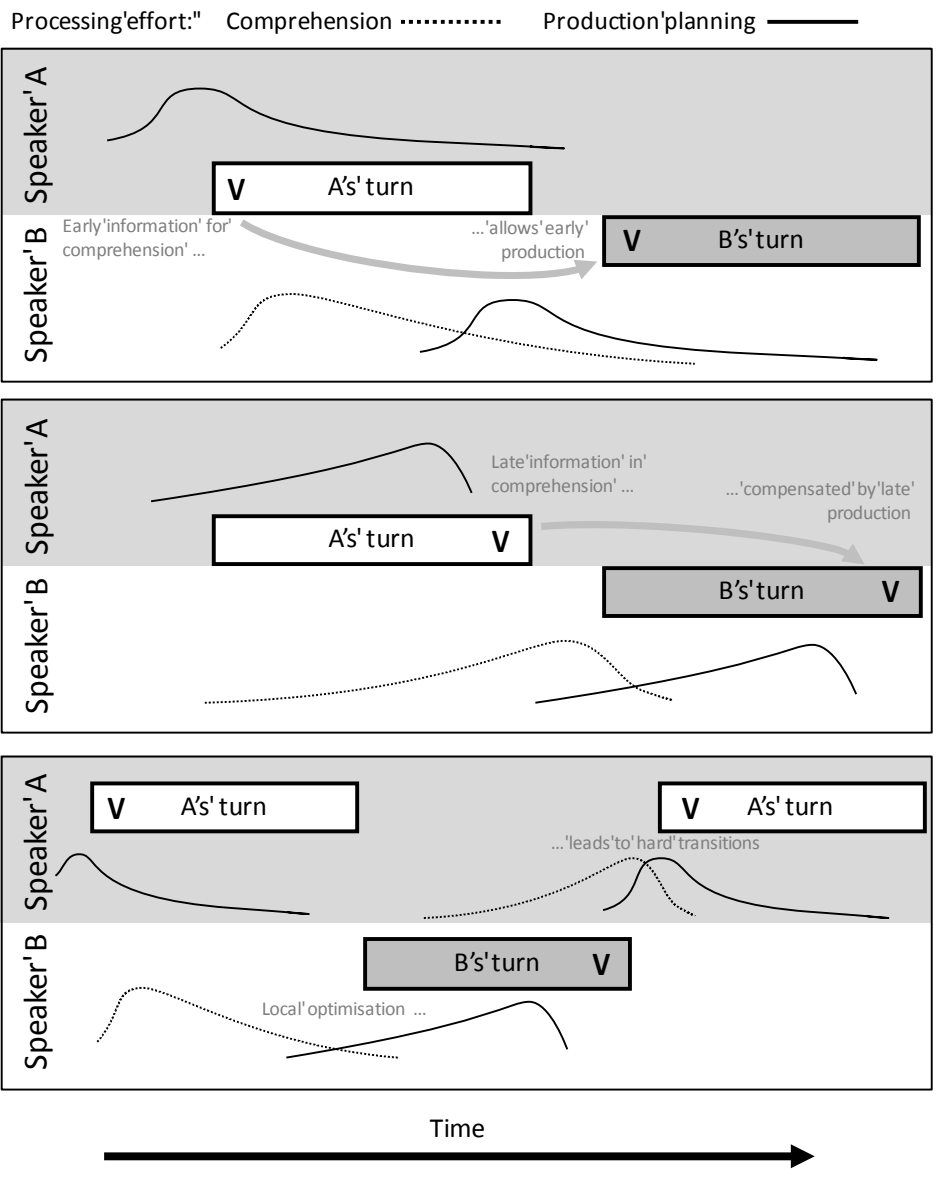
Speaker'B

Time

183

184 Figure 2: A schematic representation of the timeline of turn taking and the processing

185 effort for comprehension and production. Speaker A and B take turns at speaking,

186 placing the crucial information – the verb – at different points in the turn. Curves show

187 the processing effort for comprehending their interlocutor's turn and planning their own

7

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 simultaneous high comprehension costs and high production planning costs.

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

205 We should note here that these considerations obviously oversimplify conversational
206 exchanges which are often elliptical, but the point is that where full clauses are
207 involved, they should be subject to constraints of this kind. These could – indeed should
208 – have implications for how languages change over historical time, that is the cultural
209 evolution of linguistic structure. We would predict that a language would be more likely
210 to change to facilitate better turn taking than in the opposite direction. This suggests that
211 the proportion of languages that facilitate turn taking (e.g. by having fixed word orders
212 ensuring coordination) should increase over time, while the proportion of languages that
213 make turn taking less efficient should decrease[1].

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

---

[1] 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; 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.

217 test whether the prediction can be observed in real data.

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

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

2 Identifying the target phenomenon

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

240 The second phenomenon is that some basic word orders are more frequent than others.
241 For example, if we count the raw number of basic word orders, then the pattern we see
242 is that SOV and SVO are more frequent that VSO order. However, this does not take
243 into account the historical relations between languages. For example, many Celtic
244 languages are VSO, just as nearly all Dravidian languages are SOV, but the Celtic
245 languages are all related historically, so it would bias the sample to count each as an

10

246 independent data point (see Roberts and Winters, 2013; Dunn et al., 2011).

247 In this study we will use Harald Hammarstrom's estimation of word order types in
248 language isolates, that is, languages that are not known to be historically related to any
249 others, and thus approximate to fully independent data points.[2] This also happens to be
250 close to other estimates based on using non-isolates and controlling for historical
251 relations. This turns out to be 11% VSO, 16% SVO, 66% SOV and other orders account
252 for 7%. That is, the further from the start of the sentence the verb is, the more frequent
253 that word order type turns out to be (note some other approaches use number of
254 speakers, e.g. Bentz & Christiansen, 2010, but we are more concerned with the number
255 of communities). The majority of the world's languages place the subject before the
256 object in canonical transitive sentences, so we focus on those, but the model below does
257 not actually distinguish between subjects and objects - only the position of the verb is
258 important in the models below.

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

264

---

[2] 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

266 We model a conversation as an interaction between two computational agents A and B.

267 Agent A produces a turn at talk which consists of three abstract elements - a verb, a

268 subject and an object. There are three turn types of word order in the model - VSO,

269 SVO and SOV. The agents do not understand these elements, and there is no meaning

270 associated with the elements – the model simply captures the idea that in each turn there

271 is some linear order, with some elements (e.g. the verb) being more crucial than others.

272 Each agent has an exemplar memory which stores all the turns it has heard. When

273 agents produce a turn at talk, they select one turn from their memory at random to be

274 the template for their utterance.

275 Once A has produced a turn, agent B now has to decide how to respond by choosing a

276 template turn from its own memory. We constrain the probability of choosing different

277 turn types according to the distance between the verbs in the sequence. For example, if

278 A produces a VSO turn, then B has more time to process this information and so is

279 more likely to be able to produce a verb at the start of their turn. If A produces an SVO

280 turn, then this verb is closer to the crunch zone and B is less able to produce a verb-

281 initial turn. If A produces an SOV turn, then the verb is in the crunch zone and so B is

282 very unlikely to be able to produce a verb-initial sentence in time, and quite unlikely to

283 be able to produce a verb-medial sentence in time.

284 To model this, each item in the agent's memory is given a weight which affects its

285 probability of being chosen. If A produces a turn T1 which has the verb at position

286 $V_{initiate}$ (start = 0, middle = 1, end = 2) and a length L_1 (at this stage, all turns have a

287 length of 3), then a responding turn by B, T2, which has the verb at position $V_{respond}$ is

288 given the following weight,

289 $$W_{T2} = ((L_1 - V_{initiate}) + V_{respond})^{\alpha}$$

290 where α is a parameter which controls the strength of the effect. When $\alpha = 1$, then the

291 weight increases linearly as the distance between the two verbs increases. The

292 probability of choosing item i from a memory which contains M items is then directly

293 proportional to its weight.

12

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

294

295   Put another way, agents are less likely to choose turn structures which involve more

296   verb processing in the crunch zone. The α parameter, then, controls how quickly the

297   processing cost increases with time. This mechanism captures the basic idea that the

298   location of crucial information in A's utterance has a knock-on effect for the structure

299   of B's turn. The constraint on B's choices are greatest when A produces a turn with the

300   verb at the end.

301   Conversations proceed in the following way. A produces a first turn by selecting

302   randomly from her memory. B then produces a turn, drawing from his memory

303   according to the weight function above. Then A produces a third turn, weighting her

304   selection by the turn type that B produced. Then B responds, and so on.

305   Conversations are independent from each other, and always start with an un-weighted

306   selection. Therefore, we can manipulate the strength of the effect from turn taking. For

307   example, agents can have one conversation of three turns, which imposes a constraint

308   after each turn, or three conversations of a single turn, in which case the turn taking

309   constraints have no effect. The greater the number of turns in a conversation, the greater

310   the knock-on effect of the crunch zone. In each generation (see below), agents will have

311    $N_{conversations}$ conversations with N_{turn} turns each.

312   We also model a small amount of noise in communication. With a small probability β,

313   an agent produces a random turn type from all possible turn types.

314   **3.1 Cultural evolution**

315   Now we need a model of cultural evolution. We start with a small population of N_{agents}

316   'adult' agents. Each agent is initialised with a random selection of turn types in their

317   memory. This means that populations are initialised with no bias in their word order

318   preferences (see the SI for different starting conditions). Each agent is randomly paired

319   with another agent and they have a conversation with N_{turn} turns. This repeats until they

320   have had $N_{conversation}$ conversations. This results in a series of turns and conversations,

321 and we can measure the frequency of each turn structure.

322 At the same time, there is a second population of 'child' agents listening to the
323 conversations of the adult population and 'learning' from them by adding their turn
324 structures to their exemplar memory. That is, generation 2 are like children acquiring
325 language. When the adult generation are done with their conversations, they are
326 removed from the population and the child generation 'grows up' and become adults.
327 This new generation starts having conversations in the same way as the first generation,
328 while a new child generation (generation 3) listen and learn (so called "iterated
329 learning", see Kirby, Griffiths & Smith, 2014).

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

3.2 Sentence particles

333 We can expand the model again to explore more complicated interactions between
334 grammar and turn taking, for example the role of utterance final particles. Tanaka
335 (2000; 2005) notes that the grammar of Japanese limits the projectability of turns. The
336 predicate comes at the end of the sentence, and the sentence can be widely transformed
337 by elements that come after the predicate. This appears to work against rapid turn
338 taking. However, final particles can potentially act as a 'buffer' which push crucial
339 information away from the crunch zone and allow more time for the next speaker to
340 plan their turn (this insight from Kobin Kendrick, 2012, see figure 3). While sentence
341 final particles are usually quite short, we assume that any extra time is beneficial and
342 may lead to selection in the long term.

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

14

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

360 Sentence particles were included in the model as follows. As well as the three basic
361 word order types without particles, agents could also produce versions with a sentence
362 final or sentence initial particle (thus 9 combinations of types to choose from). Turn
363 types with particles were less likely to be picked for production, since they are slightly
364 longer (agents prefer to produce shorter turns). The relative length of particles to other
365 words (verb, subject and object) could be manipulated via a parameter p. From the
366 examples in Japanese, we would expect particles to be shorter than most words. The
367 inclusion of a particle which added distance between verbs in a turn boosted the
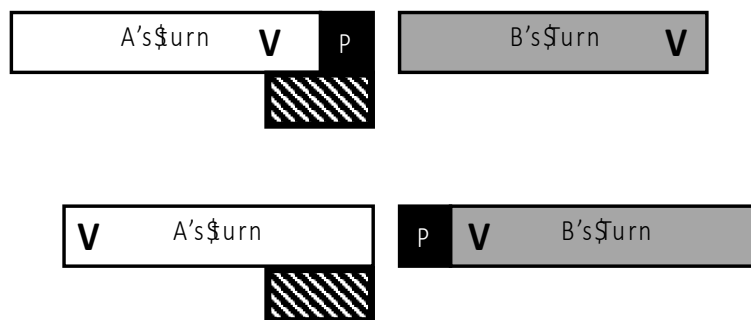368 possibility that the verb can come earlier in a following sentence.

369

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

372

15

```
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. There are no processing costs related to syntactic dependencies in the model

• 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 (not necessary, but a simplifying assumption)

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. Despite this, as a starting

16

393 point, we think that this model captures some of the crucial constraints on interactive

394 language use under temporal pressure. We are attempting to construct the simplest

395 model which will help us think about the intricate inter- relationships between

396 conversation, cognition and cultural evolution. One way to construe the model is that it

397 captures only some conversations, not every interaction between agents, and that the

398 selective pressure only applies in turns which match the conditions above.

399

400   4 Results

401 Figure 5 shows, as an example of the kinds of results obtained, three independent runs

402 of the model with a population of 10 agents taking 2 conversations of 10 turns each.

403 Along the horizontal axis we see generations and each line represents how the

404 frequency of each type of basic word-order (or major sentence type) changes over time.

405 We see that in the first generation, agents are equally likely to use any of the three

406 types, but that the use of VSO rapidly declines. In the first two runs, both SVO and

407 SOV are used for some time, but after about 15 generations, all agents are using SOV

408 all the time (with some small deviations due to noise). So, we can classify the language

409 of these agents as SOV. In the third run, enough agents selected SVO by chance that the

410 conventional pressure pushed the frequency up. Eventually, the third population

411 converges on SVO order. That is, a dominant word order emerges, and we are not

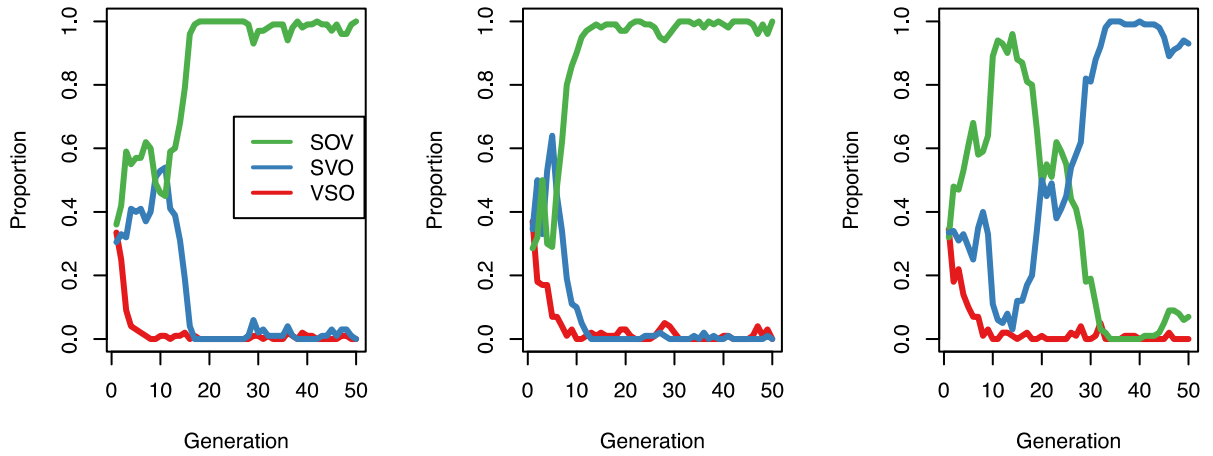412 concerned with the distribution of word orders within a language.

Figure 5: Proportions of each turn type used at each generation for three independent runs of the main model (N_{Agents}=10, N_{Turns}=10, $N_{Conversations}$=2, β =0.01, α =0.1).

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. This is not surprising, since any set of communicating agents will tend to converge on a common set of variants, as has been shown in a variety of models (e.g. Steels & Belpaeme, 2005; Nowak & Baggio, 2016) and experiments (e.g. Garrod & Pickering, 2009).

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 against having a verb in initial position. VSO is an unstable word order due to what we call the *first turn push*. The first turn in a conversation is unconstrained by turn-taking pressures - the first speaker is

18

free to choose any order in their memory. If they choose a verb-initial order, the choice of order in the 2nd turn is not affected much. However, choosing an order with the verb in a later position will bias the 2nd turn to also place their verb later, which will bias the 3rd turn to also place their verb later, and so on. SOV is a more stable order for the same reason.

Note, however, that the pressure from turn taking is not so strong as to make convergence on verb-initial order impossible. To be clear, although there is a small proportion of populations with VSO order in the model, within those few 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.
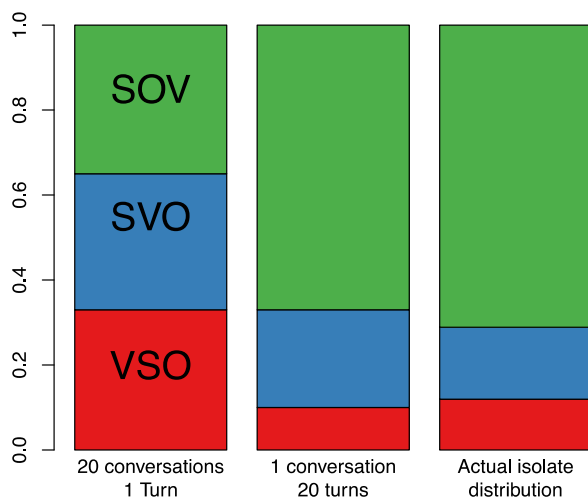

Figure 6: Proportions of each turn type that 1000 generations converge to in: Left: a model without pressures for turn taking (N_{Agents}=10, β =0, α =0.1); Middle: a model with turn taking constraints; and Right: actual language data from the world's isolates (right).

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). This quantitative fit depends on the parameters of the model. Figure 7 shows how the distribution of word order types varies

452 with the α parameter, which controls how the distance between verbs relates to the

453 processing cost by weighting the effect. When α is close to 0, there is little difference

454 between each of the sentence types in any context, and roughly the same proportion of

455 each sentence type emerges. When α is positive, reflecting greater processing cost as the

456 verbs enter the crunch zone, then the SOV advantage appears. If processing cost scales

457 linearly (α = 1), then the model predicts that almost all populations should converge on

458 SOV order. With negative values of α, where cost decreases as the verb enters the

459 crunch zone, we see a preference for VSO languages. This suggests that the best fitting

460 assumption would be for a positive, convex function: the cost is large for verbs inside

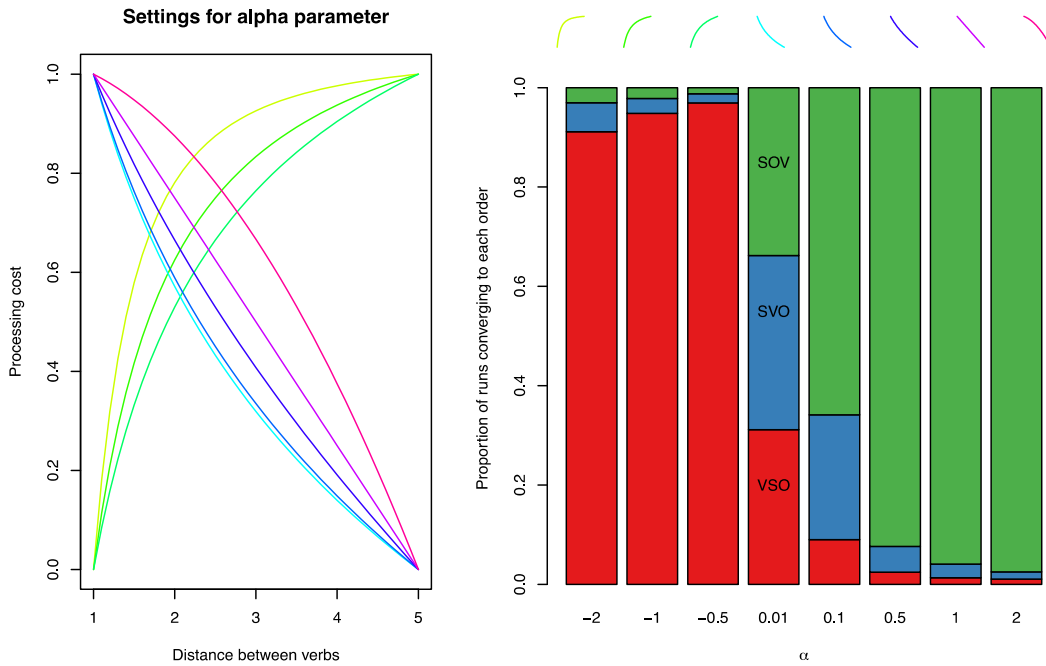461 the crunch zone, but rapidly declines as the verb moves further away.

462

Figure 7: Left: how the α parameter affects the function which relates the distance between verbs in adjacent turns and the cost of processing for the speaker of the 2nd turn. Right: how the proportions of different word-order types varies with the α parameter. For example, the extreme negative curve (yellow) represents α = -2 and around 90% of runs converge on VSO, while the extreme exponential positive curve (pink) represents α = 2 and over 90% of runs converge on SOV. The best fit to the real world distribution happens when α is between 0 and 1, which creates a convex curve.

20

471 The supporting information shows that the model results are robust to settings of

472 various parameters, including N_{agents}, $N_{conversations}$, N_{turns}, β and initial conditions

473 (discussed below).

474

4.2 Sentence final particles

475

476   Figure 8 shows some results for sentence final particles ($\alpha = 0.1$, $\beta = 0$, $p = 0.5$, $N_{agents} =$

477   10, comparing 20 conversations of 1 turn with 2 conversations of 10 turns). The model

478   without turn taking constraints predicts that languages are similarly likely to have initial

479   or final sentences regardless of verb position. In contrast, with the constraint we see two

480   things. Initial particles are more likely than final particles for verb initial languages, and

481   final particles are proportionately more likely for verb final languages. That is, if a

482   language happens to settle on verb final structures, it is also more likely to develop

483   sentence final particles. This prediction also matches the real data quite well (data from

484   position of polar question particles, Dryer, 2013b, see figure 8 and SI). Interestingly, it

485   also predicts that verb final languages should be less likely to have particles at all. The

486   explanation may be the following. If the first turn in the conversation places the verb

487   later (the first turn push), the second turn now has two options to mediate the pressure:

488   either move the verb further back or add an initial particle to the 2nd turn. Therefore,

489   languages are more likely to gain initial particles than final particles. However, since

490   SOV order is more robust to the turn taking pressures and a more stable state in general

491   and, it is less likely to transition to using a particle at all.

492   This result was not robust to changes in parameters. The fit to the data was better when

493   noise level was low, and in addition the inclusion of a question particle in a buffer zone

494   had a big effect. This is a reasonable result, given that the first model predicted that the

495   processing cost declines rapidly as the verb moves away from the crunch zone. Outside

496   of a narrow window around the parameters above, the predictions range from no effect

497   to the opposite of the effect we see in the data (final particles more likely for verb-final

498   languages). This suggests that the use of particles to buffer interactive language use
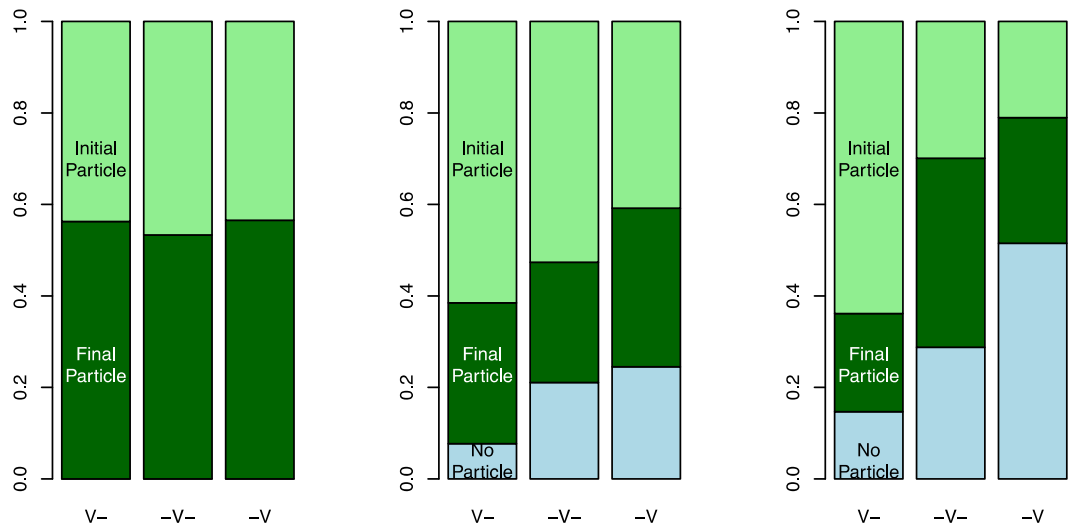
499   emerges only under specific conditions.

Figure 8: Distribution of word order types and the presence of absence or sentence particles in a model with: (Left) no turn taking constraints ($N_{conversations}$=20, N_{turns}=1, N_{agents} = 10, α = 0.1, β = 0, p = 0.5); (Middle) with turn taking constraints ($N_{conversations}$=2, N_{turns}=10); and (Right) the distribution in real languages (Dryer, 2013b).

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, 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.

This result essentially derives from the fact that the model tends to favour 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 fitting exercise 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 (N_{turns}) 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.

24

539 Given this approach, the model makes two interesting predictions that other theories
540 which do not take into account the need for rapid reactions between turns would find
541 hard to explain. First, the convex relationship between distance between verbs in two
542 adjacent turns and the ease of production. This could be empirically tested using an a
543 range of new experimental techniques (see De Ruiter, Mitterer & Enfield, 2006; Bögels,
544 & Levinson, 2016). Secondly, the presence of sentence particles at the ends of
545 utterances. These do not aid prediction (since they come last) and take effort to
546 produce, but do give some advantage to turn taking.

5.1 Relationship with other theories and future work

548 The distribution of basic word orders is one of the most scrutinised phenomena in
549 typology, and this first attempt at linking typology, processing and turn taking does not
550 aim to supplant any of the other theories. Indeed, a pressure from turn taking does not
551 exclude pressures from other domains, but here we consider how they might interact.

552 One domain that clearly has an impact on the explanandum is historical change. Gell-
553 Mann and Ruhlen (2011) review historical changes to basic word order, largely caused
554 by grammaticalisation and dependencies with other aspects of grammar, and estimate
555 transitions between orders. They find that word order tends to change from SOV to
556 SVO to VSO, and suggest that languages began as SOV. In opposition to this,
557 communities in our model tend to gravitate *towards* SOV (see also the SI). Gell-Mann
558 and Ruhlen also suggest that word order distributions have not reached a stable
559 equilibrium, while in our study we assume that the target distribution is stable. Perhaps
560 it is better to see our model as a model of transition to initial consensus within a
561 population, rather than historical change between established types. However, we find
562 that, when no pressures from turn taking apply, the only way to recover the target
563 distribution is to assume that populations begin with a dominant SOV order. When
564 pressures from turn taking do apply, the target distribution is achieved from a more
565 diverse range of initial conditions (see SI). In this sense, our hypothesis is more
566 agnostic to the initial conditions of word order, which might fit better with findings that
567 the evolutionary trajectory of word order can be different in different language families
568 (Dunn et al., 2011).

569 Another issue is the model's predictions about freedom of word order. Populations in
570 the model effectively start as free word order, but then converge on a single dominant
571 order. In reality, many languages have reported flexible word order (although one
572 should note that few reported word orders are based on conversational data), but even
573 these languages usually only use a sub-set of possible orders frequently (see Austin,
574 2001; Hale, 1992). For example, while many orders are possible in Samoan, during
575 conversation between 70% and 86% of clauses with an overt subject, object and verb
576 are verb-initial (Ochs, 1982, p. 661; Duranti, 1981, p. 171), in line with our model.
577 However, many languages also go against our predictions. In Dryer (2013a), 79% of
578 languages coded as having two dominant word orders involve a change to the position
579 of the verb (though none alternate between verb final and verb initial). Warlpiri
580 typically has the order topic, verb phrase, comment, with verb-medial constructions
581 being most frequent of clauses with an agent, patient and verb (from written text,
582 Swartz, 1987). Indeed, word order in free word order languages is often determined by
583 pragmatic (information-structure) factors such as 'newsworthy' or prominent items
584 appearing first (Givón, 1983b; Swartz, 1987; Mithun, 1992). This goes against our
585 specific hypothesis about the position of verbs (and the predictions of the uniform
586 information density hypothesis, see below), although it is compatible with the general
587 idea of consistently keeping elements which require more effort to comprehend in the
588 same relative position in order to facilitate turn taking. Modelling this might require
589 utterances to be sensitive to information structure or considerations of processing
590 dependencies between the different constituents (see Ferrer-i-Cancho, 2016).

591 One of the crucial assumptions of the model is that verbs require the most effort to
592 process as a listener and plan as a speaker. We assumed this since, for many
593 languages, the verb provides the syntactic frame for a sentence. In a conversational
594 discourse, topics tend to be 'given' information, while comments (predicates/verbs)
595 tend to convey the new information (e.g. Van Valin & LaPolla, 1997). Studies which
596 track the cognitive timecourse of comprehension and planning during interactive turn-
597 taking are only beginning to emerge (Gisladottir, Chwilla & Levinson, 2015; Bögels,
598 Magyari & Levinson, 2015), and do not directly address our assumption. However, we
599 note that some studies are compatible with our position. For example, several studies
600 using the visual world eye-tracking paradigm show that the listener integrates

26

constraints on the possible or likely upcoming elements when the verb appears (Altmann, 1999; Kamide, Altmann & Haywood, 2003; Altmann, 2004; see Kamide, 2008). The semantics of verbs help listeners predict upcoming referents in verb-initial languages (Sauppe, 2016). Corpus analyses of written language also suggests that verbs carry more information to the listener (easing subsequent processing) on average than nouns (surprisal measure from Piantadosi, Tily & Gibson, 2011 as calculated in Roberts, Torriera & Levinson, 2015), which is in line with our position. However, these results are inconsistent with others. For example, integration for prediction can occur when hearing constituents other than verbs depending on the structure of the utterance (see experiment 3 of Kamide et al. 2003; Knoeferle et al., 2005), and a study of child-directed speech found that objects convey more information than subjects or verbs (Maurits, Perfors & Navarro, 2010, see below). SOV order also emerged in an evolving population of neural-network agents when there was a selection pressure for predictability (Reali & Christiansen, 2009).

More generally, real conversations are more complex than simple 3-constituent constructions. For example, speakers use a range of strategies to defer the beginning of the content of their turn (e.g. turn-preserving placeholders such as "umm…"), which mitigates the need for rapid processing to some extent. Turns in conversations often do not have overt subjects, objects or verbs. For example, Bowern (2012) shows that in Bardi (a free word order language), clauses with an overt subject, object and verb are very rare in texts (less than 2%). Furthermore, morphological marking and word order itself can help listeners predict the upcoming verb, reducing the processing load at the verb itself (Lupyan & Christiansen, 2002). Indeed, Ferrer-i-Cancho (2015) argues that verbs at the end of a turn are better for the listener because the prior context helps predict it, in opposition to our prediction.

There is therefore no simple consensus about the difficulty of processing verbs during conversation. We note that the previous literature on word order and cognition tends to focus on semantic comprehension, while an important part of turn taking in conversation is the comprehension of pragmatic acts (Gisladotir, Chwilla & Levinson, 2015). In any case, questions about how verb position, context, semantic relations and pragmatic acts relate to planning and comprehension effort is at least empirically

27

632 testable with recent large-scale databases and new experimental methods (e.g. Roberts,

633 Torreira & Levinson, 2015; Barthel et al., 2016; Bögels & Levinson, 2016).

634 Even if the general assumption about verbs is correct, the model could be rooted in

635 more concrete measures of processing. For example, the work on uniform information

636 density suggests that languages are optimised for conveying information at a constant

637 rate, avoiding high information rates which are unreliable or low information rates

638 which are inefficient (e.g. Jaeger & Levy, 2006; Jaeger, 2010; Piantadosi, Tily &

639 Gibson, 2011). Relating to word order in particular, Maurits, Perfors & Navarro (2010)

640 analyse spoken conversations and measure the predictability of verbs from their subjects

641 and objects. They show that VSO and SVO orders provide more uniform information

642 density, and therefore might be more efficient orders, helping to explain the drift

643 towards them in Gell-Mann & Ruhlen's study.

644 One weakness is that the uniform information density accounts are motivated by the

645 rational strategy for successfully transmitting a single utterance in noisy conditions

646 (studies like Maurits et al., 2010 also assume that all words are the same length and that

647 previous utterances do not carry information about the current one). Furthermore, the

648 uniform information density accounts focus on the ease of decoding rather than the ease

649 of planning. We argue that real time conversation involves simultaneous encoding and

650 decoding at certain points in each turn, and so the ideal information profile may not be

651 uniform, but one of the skewed distributions discussed above. In general, however, the

652 findings may be compatible with our account. For example, according to the results in

653 Maurits et al., 2010, SVO order conveys the most uniform information rate. Yet

654 considering the three orders where the subject precedes the object, the last element in

655 the utterance contains more information (and therefore requires more cognitive

656 resources) as the verb moves away from the crunch zone at the end of the turn. That is,

657 SOV order is the best profile for a turn-taking listener, since they are already able to

658 predict the verb from the subject and the object, and are therefore able to dedicate more

659 resources to planning in the crunch zone. This is compatible with the result of our

660 model that turn taking imposes a pressure to push the verb further back in the sentence.

661 In another approach, Ferrer-i-Cancho (2015) argues that the length of syntactic

662 dependencies between the verb and its subject and object (within a turn) has a

28

663 considerable effect on short term memory load, and that planning effort is minimised

664 when placing the syntactic head in the center of the construction. This opposes a

665 pressure for predictability by the listener, which favours verb-final constructions.

666 Furthermore, historical changes between dominant word orders tends to proceed in

667 single steps between adjacent orders (see also, Ferrer-i-Cancho, 2016). These factors

668 combine to explain many phenomena such the prevalence of SVO order, optionality

669 between SOV and SVO order, historical movement towards SVO and OVS order being

670 rare since it is many changes away from the presumed initial SOV order. Currently, our

671 model is too abstract to integrate notions of syntactic dependency within a turn, and

672 transitions between any order to any other order occur (see SI).

673 The model presented here is not intended to supplant any of these other explanations

674 and, as many others have pointed out, several factors could be at play in this complex

675 system (Hawkins, 2004; Langus & Nespor, 2010; Ferrer-i-Cancho, 2015). There is

676 clearly work to be done to relate the different accounts to each other. For now, we point

677 out that the need for processing efficiency derives to some extent from the real-time

678 nature of natural conversation, and that all of the approaches above consider processing

679 within utterances or from the perspective of an isolated speaker or hearer, while we

680 have argued that there are cognitive constraints imposed from the relationship between

681 turns by multiple individuals.

682 To conclude, there are many issues to resolve. The model is extremely simple and

683 makes many assumptions that could be relaxed. The parameters also need to be tied to

684 specific cognitive mechanisms, rather than abstract notions of processing cost. Rules of

685 the sequential organisation of conversation could also be built into the model. The

686 general hypothesis also makes more general predictions about grammatical structures

687 within conversations which could be tested. For example, do speakers alter the

688 information structure of their turns to aid processing by local co-ordination? Finally, the

689 constraints from turn taking are just one domain from many that impact the evolution of

690 grammatical structure. Despite these limitations, we believe that the model provides a

691 useful tool for thinking about the relationship between conversation and cognition in a

692 cultural evolution framework. Our take-home message is that interactive turn-taking in

693 conversation must impose constraints on cognition, and that these may have

29

694 implications for the way in which languages change over time.

695 Acknowledgements

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

Altmann, G. T. M. (1999). Thematic role assignment in context. Journal of *Memory and Language*, 41.124–45.

Altmann, G. T. M. (2004) Language-mediated eye movements in the absence of a visual world: the 'blank screen paradigm'. *Cognition* 93.79–87

Austin, P. (2001). Word order in a free word order language: the case of Jiwarli. In J. Simpson, D. Nash, M. Laughren, P. Austin, B. Alpher, (eds), *Forty years on: Ken Hale and Australian languages*, 205-323. Canberra: Pacific Linguistics.

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, 7: 1858.

Bentz, C., & Christiansen, M. H. (2010). Linguistic adaptation at work? the change of word order and case system from Latin to the Romance languages. In A. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *Proceedings of the eighth international conference on the evolution of language* (p. 26-33). London: World Scientific.

Bowern, C. (2012). *A grammar of Bardi*. Walter de Gruyter.

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.

Bögels, S., & Levinson, S. C. (2016). The Brain Behind the Response: Insights Into Turn-taking in Conversation From Neuroimaging. *Research on Language and Social Interaction*. Advance online publication.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. Scientific Reports, 5: 12881.

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146, 67-80.

31

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

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

733 Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S.,
734 Gisladottir, R. S., Kendrick, K. H., Levinson, S.C., Manrique, E., Rossi, G & Enfield,
735 N. J. (2015). Universal principles in the repair of communication problems. *PloS one*,
736 *10*(9), e0136100.

737 Donders, F. C. (1868). La vitesse des actes psychiques. Archives Néerlandaise, 3, 269–
738 317.

739 Dryer, M. S. (2013a) Order of Subject, Object and Verb. In: Dryer, Matthew S. &
740 Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. Leipzig:
741 Max Planck Institute for Evolutionary Anthropology. (Available online at
742 http://wals.info/chapter/81, Accessed on 2017-01-11.)

743 Dryer, M. S. (2013b) Position of Polar Question Particles. In: Dryer, Matthew S. &
744 Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. Leipzig:
745 Max Planck Institute for Evolutionary Anthropology. (Available online at
746 http://wals.info/chapter/92, Accessed on 2017-01-13.)

747 Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of
748 language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345),
749 79-82.

750 Duranti, A. (1981). The Samoan fono: A sociolinguistic study (No. 80). *Pacific*
751 *Linguistics Series B*, 80. Department of Linguistics, Research School of Pacific Studies,
752 Australian National University.

753 Enfield, N. J. (2008). Language as shaped by social interaction. *Behavioral and brain*
754 *sciences*, 31(05), 519-520.

755 Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure
756 their input to facilitate efficient communication. *Proceedings of the National Academy*
757 *of Sciences*, *109*(44), 17897-17902.

758   Ferrer-i Cancho, R. (2008). Some word order biases from limited brain resources: A

759   mathematical approach. Advances in Complex Systems, 11(03):393–414.

760   Ferrer-i-Cancho, R. (2014). Why might SOV be initially preferred and then lost or

761   recovered? A theoretical framework. In: THE EVOLUTION OF LANGUAGE -

762   Proceedings of the 10th International Conference (EVOLANG10), Cartmill, E. A.,

763   Roberts, S., Lyn, H. & Cornish, H. (eds.). Evolution of Language Conference (Evolang

764   2014). Vienna, Austria, April 14-17. pp. 66-73.

765   Ferrer-i-Cancho, R. (2015). The placement of the head that minimizes online memory.

766   A complex systems approach. *Language Dynamics and Change,* 5(1), 114-137.

767   Ferrer-i-Cancho, R. (2016) Kauffman's Adjacent Possible In Word Order Evolution. In

768   S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér & T. Verhoef

769   (eds.) *The Evolution of Language: Proceedings of the 11th International Conference*

770   *(EVOLANG11).* Available online: http://evolang.org/neworleans/papers/83.html

771   Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog.

772   *Topics in Cognitive Science*, *1*(2), 292-304.

773   Gell-Mann, M. and Ruhlen, M. (2011). The origin and evolution of word order.

774   P*roceedings of the National Academy of Sciences*, 108(42), 17290–17295.

775   Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A

776   noisy-channel account of crosslinguistic word-order variation. *Psychological science*,

777   0956797612463705.

778   Gisladottir, R. S., Chwilla, D., & Levinson, S. C. (2015). Conversation electrified: ERP

779   correlates of speech act recognition in underspecified utterances. PLoS One, 10(3):

780   e0120068.

781   Givón, T. (1983a) *Topic continuity in discourse: A quantitative cross-language study*.

782   Amsterdam: John Benjamins Publishing.

783   Givón, T. (1983b). Topic continuity in discourse: An Introduction. In T. Givón (ed)

784   *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John

785   Benjamins Publishing.

786   Goldin-Meadow, S., So, W. C., Özyürek, A., and Mylander, C. (2008). The natural

33

787 order of events: How speakers of different languages represent events non- verbally.

788 Proceedings of the National Academy of Sciences, 105(27):9163–9168.

789 Hale, K. (1992). Basic word order in two "free word order" languages. In D. Payne (ed),

790 *Pragmatics of word order flexibility*, 63-82. Amsterdam: John Benjamins.

791 Haspelmath, M., Dryer, M. S., Gil, D., and Comrie, B. (2008). World Atlas of

792 Language Structures. online at http://wals.info. Accessed 2013-04-18. Munich: Max

793 Planck Digital Library.

794 Hawkins, J. (1994). A performance theory of order and constituency. New

795 York: Cambridge University Press.

796 Hawkins, J. A. (2004). Efficiency and complexity in grammars. Oxford: Oxford

797 University Press.

798 Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of*

799 *Experimental Psychology*, *4*(1), 11-26.

800 Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-

801 based implicature. In D. Schiffrin (ed.) *Meaning, form, and use in context: Linguistic*

802 *applications*. Georgetown University Press, Washington DC. 11-42. Reprinted in

803 Kasher (ed., 1998), vol. IV: 389–418.

804 Indefrey, P. (2011). The spatial and temporal signatures of word production

805 components: a critical update. *Frontiers in psychology*, *2*(255), 1-16.

806 Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information

807 density. *Cognitive psychology*, *61*(1), 23-62.

808 Jaeger, T. F., Levy, R. P. (2006). Speakers optimize information density through

809 syntactic reduction. In P.B. Schölkopf and J.C. Platt and T. Hoffman (eds) *Advances in*

810 *neural information processing systems 19*, pp. 849-856.

811 Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and*

812 *Linguistics Compass*, *2*(4), 647-670.

813 Kamide, Y., Altmann, G. T., Haywood S. L. (2003) The time-course of prediction in

814 incremental sentence processing: Evidence from anticipatory eye-movements.

815 *Journal of Memory and Language,* 49 133 –59.

34

816   Kendrick, K. H. (2012) Particles at the transition space. Talk presented at the

817   Interactional Foundations of Language Meeting. Kreuth, Germany. 2012-11-01 - 2012-

818   11-02.

819   Kendrick, K. H. (2015). The intersection of turn-taking and repair: the timing of other-

820   initiations of repair in conversation. *Frontiers in psychology*, 6, 250.

821   Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: a

822   quantitative study. *Discourse Processes*, *52*(4), 255-289.

823   Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of

824   language. *Current opinion in neurobiology*, *28*, 108-114.

825   Knoeferle, P., Crocker, M. W., Scheepers, C. & Pickering, M. J. (2005) The influence

826   of the immediate visual context on in cremental thematic role assignment: evidence

827   from eye-movements in depicted events. *Cognition,* 95.95 –127

828   Krupa, V. (1982). Syntactic Typology and Linearization. *Language*, 58(3), 639–645.

829   Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order.

830   Cognitive Psychology, 60(4), 291-318.

831   Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech

832   production. *Behavioral and brain sciences*, *22*(01), 1-38.

833   Levinson, S. (2006). On the human interaction engine. In Enfield, N. and Levinson, S.,

834   editors, Roots of Human Sociality: Culture, Cognition and Human Interac- tion, pages

835   39–69. Oxford: Berg.

836   Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

837   Levinson, S. C. (2016). Turn-taking in Human Communication–Origins and

838   Implications for Language Processing. *Trends in cognitive sciences*, *20*(1), 6-14.

839   Lupyan, G., & Christiansen, M. H. (2002). Case, word order, and language learnability:

840   Insights from connectionist modeling. In *Proceedings of the 24th annual conference of*

841   *the cognitive science society* (p. 596-601). Mahwah, NJ: Lawrence Erlbaum.

842   Maurits, L., Perfors, A., and Navarro, D. (2010). Why are some word orders more

843   common than others? a uniform information density account. Advances in Neural

844   Information Processing Systems, 23:1585–1593.

35

845   Mithun, M. (1992). Is basic word order universal? In D. Payne (ed), *Pragmatics of word*
846   *order flexibility*, 63-82. Amsterdam: John Benjamins.

847   Nowak, I., & Baggio, G. (2016). The emergence of word order and morphology in
848   compositional languages via multigenerational signaling games. *Journal of Language*
849   *Evolution*, 1(2): 137-150.

850   Ochs, E. (1982). Ergativity and word order in Samoan child language. *Language*, 58(3),
851   646-671.

852   Ochs, E., Schegloff, E. A., & Thompson, S. A. (1996). *Interaction and grammar* (Vol.
853   13). Cambridge University Press.

854   Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for
855   efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9),
856   3526-3529.

857   Reali, F., & Christiansen, M. H. (2009). Sequential learning and the interaction
858   between biological and linguistic adaptation in language evolution. Interaction Studies,
859   10, 5-30.

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

863   Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and
864   sequence organisation on the timing of turn taking: A corpus study. *Frontiers in*
865   *Psychology*, 6: 509.

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

868   De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's
869   turn: A cognitive cornerstone of conversation. *Language*, 515-535.

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

872   Sauppe, S. (2016). Verbal semantics drives early anticipatory eye movements during the
873   comprehension of verb-initial sentences. *Frontiers in Psychology*, 7: 95.

874 Schegloff, E. A. (1989). Reflections on language, development, and the interactional

875 character of talk-in-interaction. Interaction in human development, pages 139–153.

876 Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in

877 the organization of repair in conversation. *Language*, 361-382.

878 Schouwstra, M. and de Swart, H. (2014). The semantic origins of word order.

879 Cognition, 131(3):431–436.

880 Steels, L. and Belpaeme, T. (2005). Coordinating perceptually grounded categories

881 through language: a case study for colour. Behavioral and Brain Sciences, 28(4):469–

882 489.

883 Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T.,

884 Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., and Levinson, S. C. (2009).

885 Universals and cultural variation in turn-taking in conversation. Proceedings of the

886 National Academy of Sciences, 106(26):10587–10592.

887 Swartz, S. (1987) Pragmatic structure and word order in Warlpiri. *Papers in Australian*

888 *linguistics*, 17, 151-166.

889 Tanaka, H. (2000). Turn projection in Japanese talk-in-interaction. *Research on*

890 *Language and Social Interaction*, 33(1):1–38.

891 Tanaka, H. (2005). Grammar and the "timing" of social action: Word order and

892 preference organization in Japanese. *Language in Society*, *34*(03), 389-430.

893 Thompson, S. A. (1998). A discourse explanation for the cross-linguistic differences in

894 the grammar of interrogation and negation. *Case, typology and grammar: In honor of*

895 *Barry J. Blake*, 309-341.

896 Tomlin, R. S. (1986). *Basic word order: functional principles*. Croom Helm.

897 Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*.

898 Cambridge University Press.

899 Zipf, G. K. (1949). Human behavior and the principle of least effort. Oxford, England:

900 Addison-Wesley Press.