Opinion

Turn-taking in Human Communication – Origins and Implications for Language Processing

Stephen C. Levinson1,2,*

Most language usage is interactive, involving rapid turn-taking. The turn-taking system has a number of striking properties: turns are short and responses are remarkably rapid, but turns are of varying length and often of very complex construction such that the underlying cognitive processing is highly compressed. Although neglected in cognitive science, the system has deep implications for language processing and acquisition that are only now becoming clear. Appearing earlier in ontogeny than linguistic competence, it is also found across all the major primate clades. This suggests a possible phylogenetic continuity, which may provide key insights into language evolution.

Turn-Taking – Part of Universal Infrastructure for Language

Languages differ at every level of construction, from the sounds, to syntax, to meaning [1]. However, there is a striking uniformity in the way language is predominantly used across every language examined – the rapid exchange of short turns (see Glossary) at talking [2]. Although unremarkable in character at first sight, the turn-taking system turns out to shed real insight into language processing, and moreover goes some way to explain why language has the character that it does, organized into short phrase or clause-like units with an overall prosodic envelope. In addition, in contrast to the diversity of languages, the universal character of turn-taking, its early onset in ontogeny, and its continuity with other primate communication systems suggest an interesting phylogenetic story in which vocal turn-taking preceded language and provided a frame for its development. Although well explored in the branch of sociology termed conversation analysis [3], the human system has been until recently largely ignored in the cognitive sciences.

The great bulk of human language usage is interactive or conversational usage, which also forms the context of language acquisition. The basic properties of the conversational turn-taking system are as follows [3,4], with relatively small differences across languages [2]. Turns are of no fixed size, but tend to be short, about 2 s in length on average, although bids can be made for longer turns, as required for example to tell a story. The turn-taking system organizes speakers so as to minimize overlap, and is highly flexible with regard to the number of speakers or the length of turns. The system is highly efficient: less than 5% of the speech stream involves two or more simultaneous speakers (the modal overlap is less than 100 ms long), the modal gap between turns is only around 200 ms, and it works with equal efficiency without visual contact [4]. The dominant view [3] is that the system is organized around rights to minimal turns, the first responder gaining such rights, and relinquishing them upon turn-completion. Turns are built out of syntactic units, further individuated prosodically such that participants can predict upcoming

---

1Max Planck Institute for Psycholinguistics, Wundtlaan 1, NL-6525 XD Nijmegen, The Netherlands
2Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
*Correspondence: stephen.levinson@mpi.nl (S.C. Levinson)

---

http://dx.doi.org/10.1016/j.tics.2015.10.010
turn-completion. Some [5] have emphasized a turn-end signaling component, but this comes too late for the initiation of response planning, although it may well act as a launch signal for a pre-prepared turn [4,6]. As far as we know, the overall system employed in conversation is strongly universal, with only slight variations in timing [2], and it contrasts with other more specialized speech exchange systems such as those employed in classrooms, courtrooms, presidential press briefings, etc., which tend to be culture-specific.

The Cognitive Challenge of Turn-Taking

To appreciate the cognitive consequences of the turn-taking system, consider the following findings. Across languages, the modal response time (gaps between turns) is around 200 ms [2,4,5], the average duration of a single syllable. This is at the limit of human performance for a simple start signal with a single possible response (cf. a starting pistol beginning a race); reaction time systematically slows with the number of choices between response types (Hick’s Law), and languages have vocabularies of 50,000 words or more. Moreover, the language production system is notoriously slow – preparation before output begins takes 600 ms for a single word if primed [7,8], approximately 1000 ms if not [9], and around 1500 ms for a short clause [10]. Much of this latency is caused by the slow encoding of phonological forms and articulatory gestures (for a range of factors influencing latency of response see [11]). It follows that responses must be planned in the middle of the incoming turn which is being responded to (average turn duration is around 2 s) [4].

The implication of the slow production system is that, in interactive language use, comprehension and production overlap – one must plan while still listening and predicting what the rest of the incoming turn will contain. Let us take the point of view of the addressee B listening to an incoming turn from A, as in Figure 1 (Key Figure) [4]. Beyond simply comprehending the signal as it comes in, the preconditions for B making a sensible response on time (approximately 200 ms after the end of A’s turn) are the following: (i) B must attempt to predict the speech act (detect whether A’s utterance is a question, offer, request, etc.) as early as possible [12], because this is what B will respond to; (ii) B should at once begin to formulate a response, going through all the stages of conceptualization, word retrieval, syntactic construction, phonological encoding, articulation [13]; (iii) meanwhile, B should use the unfolding syntax and semantics of A’s turn to estimate its likely duration, listening for prosodic cues to closure; (iv) as soon as those cues are detected B should launch the response.

Some information about each of these stages has recently become available, with electroencephalography (EEG) providing good time-resolution of some of the processes involved. (i) Speech-act recognition is non-trivial because there is no one-to-one mapping from form to function [12]: “I have a car” could function as an answer to a question, a prelude to an offer to give a ride, or a declining of an offer of a ride, all depending on context (e.g., respectively, ‘Do you go by train?’, ‘I’ve just missed the last train’, ‘Do you need a ride?’). Nevertheless, in this kind of constraining context, speech-act recognition has been shown using EEG to be very fast, within the first 400 ms of the turn-begning [14]. (ii) As soon as comprehension identifies the function of an incoming turn, response preparation can begin: in an interactive task using EEG it was found that production processes kick in within 500 ms of sufficient information becoming available – the signal can be traced to language-encoding areas [15], but see [16]. (iii) The temporal estimation of turn duration can use the lexical, semantic, and syntactic structure to predict, in favorable cases, about half way through the turn the likely point of completion [17,18], even guessing likely upcoming words [19]. Manipulations show that semantics plays a large role in this predictive capacity [20]. (iv) Prosodic cues such as lengthened syllables often occur at the end of turns, and can be shown to be used by listeners [8] – they may provide the ‘Go’ signal for production of the response. This would account for the 200 ms modal gap – close to the basic human minimal response time. Preparation for the launch of speech triggered by such cues can...
(vowels and consonants), especially intonation, tone, stress, and rhythm.

**Semiotics**: the study of communication and sign systems in the broadest sense, for example beyond language.

**Speech act**: the point or intention behind an utterance (e.g., a request vs an offer, a statement vs a question), sometimes termed illocutionary force.

**Speech exchange system**: conversational turn-taking offers a basis for the elaboration of special turn-taking systems wherein, for example, a chairman controls bids to talk in a committee meeting, or questions may only be asked by one specific party of another (as in courtroom cross-examination).

**Turn**: the unit of conversational communication, expressing a speech act, averaging around 2 s in duration but highly variable; in spoken language, typically a phrase or clause grammatically and prosodically complete and pragmatically sufficient.

---

**Key Figure**

The Cognitive Challenge of Turn-Taking

(A) Responses in conversation are fast

(B) Latencies in production are threefold or more longer than the modal gap

(C) Production of response must therefore overlap with comprehension of the incoming turn

---

Figure 1. (A) Switching of speakers is rapid, with a typical gap or offset of 200 ms. Inset is a histogram of response times with 200 ms mode (0 is the end of the prior turn, with overlaps to the left, gaps to the right; from [4]). (B) Response latencies for the production of single words, as measured in primed picture-naming tasks, require ~600 ms (after Indefrey [8]). (C) The slow production mechanism may be compensated for by predicting the continuation and termination of the incoming turn, and launching production early.
be seen in the breathing signal using plethysmography [21], and is also reflected in the eye movements of onlookers [22]. There is more controversy about the role of pitch; filtering pitch out does little to diminish response times [23], but other measures demonstrate its use [24–26].

Human turn-taking thus involves multi-tasking comprehension and production, but multi-tasking in the same modality is notoriously difficult [27,28], and in this case involves using large parts of the same neural substrate [29]. Presumably this can only be achieved by rapid time-sharing of cognitive resources. This overlap of comprehension and production raises problems with current psycholinguistic theory: for example, there are proposals that comprehension intrinsically uses the production system to predict what is upcoming, but if the production system is already involved in planning output it would scarcely be available to aid comprehension except in the early stages of a turn [18,30].

Participants are hurried on by the fact that slow responses carry semiotic significance – typically conveying reluctance to comply with the expected response [31,32], an inference best avoided by maintaining normal pacing (in addition, processing bottlenecks favor moving as fast as possible [33]). Conversational turn-taking is thus very cognitively demanding, using prediction and early preparation of complex turns to achieve turn-transitions close to the minimal reaction time to a starting gun.

**Turn-Taking Partially Constrains Linguistic Diversity**

Such hungry cognitive processing might be expected to leave a significant imprint on the structure of languages, and in some respects it does. The fact that all languages organize their syntax around the clause, the minimal structure expressing a speech act and proposition, is likely an adjustment to the small turn units licensed by the turn-taking system [34]. Similarly, the pressure on response speed and the slow nature of sound encoding put a high premium on information compression – the solution is to use pragmatic heuristics that inferentially enrich the message [35,36]. Less obviously, there is pressure that speech acts (e.g., questions, requests, offers) should be recognizable early in the turn such that response preparation can begin long before the end. Despite the fact that many languages appear to ignore this pressure, putting speech-act encoding particles at the end of turns, they tend to have early signals too: for example, in a sample of 10 languages from around the world, speakers of all the languages used a boosted initial pitch in questions, with a further boost for special uses of questions to accuse, challenge, mock, or the like [37].

Nevertheless, languages show surprising diversity, to the point that it is actually hard to specify universal properties that all languages share [1]. Languages differ in the predictive parsing they offer – if they are right-branching in structure, with for example initial verbs (as in Welsh), prediction is facilitated, but if they are left-branching, with for example verbs at the end (as in Japanese), prediction is difficult [38,39] (see branching structure in the Glossary). However, the turn-taking system relies on prediction. Languages also differ in the size of the units or increments that must be planned in advance of beginning to speak – these are large if the language is verb-initial, but small if the language is subject-initial [40]. However, the turn-taking system puts a premium on early response. These systematic mismatches between language structure and optimal design for turn-taking suggest a degree of modularity of language with respect to turn-taking, and modularity (as with modularity of the senses) is often suggestive of distinct evolutionary heritage. Setting aside some dialect differences in songbirds, the contrast with other animal communication systems is striking – why do we not also only have a single communication system across all social groups? The obvious suggestion is that the complexities of individual languages are largely cultural [41]; it is as if we have an innate basis for vocal imitation and turn-taking, but have out-sourced the grammatical complexities to cultural evolution.
Origins of the Turn-Taking System

What then is the origin of the turn-taking system in humans? It might be thought that it is an obvious adaptation to two communicators using a single auditory channel. However, one voice is poor masking for another [42], and in fact the study of heated speech shows that people can respond in overlap to the utterance they are hearing [43]. Most telling, however, is turn-taking in sign languages of the deaf – when due allowance is made for preparatory and held movements, sign languages seem to conform almost perfectly to the turn-taking system of spoken languages [44]. Another functional argument would be that a system of short turns and responses makes immediately evident whether interlocutors have understood one another, and affords the chance for quick repair [45]. However, in that case participants might be expected to respond as soon as they have understood, and thus substantially overlap each other, especially because speech-act recognition seems to be early, whereas in fact where overlaps occur the modal overlap is less than 100 ms in length [4].

Functional explanations for turn-taking may then not be sufficient. There are three reasons to think that turn-taking has in fact deeper roots in human nature. The first we have already reviewed: in contrast to the diversity of languages, turn-taking exhibits strong universality – informal communication in all cultures seems to be based on the same exchange principles. In fact, turn-taking seems to belong to a package of underlying propensities in human communication, including the face to face character that affords the use of gesture and gaze, and the motivation and interest in other minds, which I have dubbed ‘the interaction engine’ [46,47]. These propensities generate a large number of universals of language use, including principles of pragmatic inference [36] and repair [45]. The large proportion of waking hours spent in such communication is also remarkable (we tend to spend a couple of hours a day, producing about 1500 turns, extrapolating from a cross-cultural study [48]). Although there are cultural and individual variations and constraints in all such matters, the whole interaction system looks pan-human in character.

A second reason to think that turn-taking is simply part of our ethology is the proto-conversation evidenced in early infancy [49], where infants participate in structured exchange with caretakers (at least in Western languages) long before they understand much about language [50]. Interestingly, the timing of turn-taking of these non-linguistic vocalizations in the first 6 months approximates the timing of adult spoken conversation, although with greater overlap. Later, from around 9 months the responses of infants actually become slower, while overlap reduces [51]. This slowing down corresponds to the ‘nine-month revolution’ [52] when the infant begins to grasp the significance of intentional communication and can follow pointing. Interestingly, the response times remain slow (about double adult latencies) well into middle childhood, presumably because, as more and more language is acquired, the challenge of cramming even more complex linguistic material into brief turns only increases. By contrast, prediction of turn-endings is fast even at age 1 year [25]. Turn-taking would thus seem to have an instinctive basis but also to involve a large learned component.

A third argument for the biological nature of human turn-taking comes from comparative primate evidence (Figure 2). The vocal systems of the ~300 primate species remain understudied, but we have detailed reports of vocal turn-taking or alternating duetting from all the major branches of the family: (i) from the lemurs, Lepilemur edwardsi [53], (ii) from New World monkeys the common marmoset Callithrix jacchus [54,55], the pygmy marmoset Cebuella pygmaea [56], the coppery titi Callicebus cupreus [57], and squirrel monkeys of the Saimiri genus [58]; (iii) from the Old World monkeys Campbell’s monkey Cercopithecus campbelli [59], and (iv) from the lesser apes, siamangs Hylobates syndactylus [60,61]. One can expect that many other cases are yet to be reported. Exactly as with human infants, this behavior seems to be partly instinctive and partly learned [54,55].
While it remains possible that these convergences are analogies (by parallel evolution) rather than homologies (by shared inheritance) [62], it also seems entirely possible that vocal turn-taking is ancestral in origin in the primate order. A puzzle, however, is that vocal turn-taking is not reported from the other great apes, who prioritize gestural communication systems [63];

![Diagram of primate species with known vocal turn-taking.](figure2)

**Figure 2.** Some Primate Species with Known Vocal Turn-Taking (Red Star) Although vocal turn-taking is not clearly present in the non-human great apes, at least two species (orangutans and bonobos) are gestural turn-takers (blue stars [64,65]). Pictures (from left to right) by Frank Vassen, Raimond Spekking, Malene Thyssen, Davidwfx, Steve Wilson, Badgernet, Suneko, Eleifert, Roger Luijten, Thomas Lersch, Lisa DeBruine and Benedict Jones, used under creative commons license.

While it remains possible that these convergences are analogies (by parallel evolution) rather than homologies (by shared inheritance) [62], it also seems entirely possible that vocal turn-taking is ancestral in origin in the primate order. A puzzle, however, is that vocal turn-taking is not reported from the other great apes, who prioritize gestural communication systems [63];

![Diagram illustrating the timeline of hominin evolution.](figure3)

**Figure 3.** The Argument for Gestural before Elaborate Vocal Turn-Taking. Diagram and details from [66,68]; for the lack of breath control in Homo ergaster see [67]. The vertical scale is in million years before the present.
nevertheless, systematic turn-taking does take place here in the gestural modality [64,65], exactly as it does in human sign languages [44]. If human turn-taking is homologous to that of other primates, this would suggest a stratified evolution of human communication along the lines sketched in Figure 3 [66]. The African variety of *Homo erectus* (ca 1.6 My) appears to have lacked the breath control necessary for modern speech [67], but may (as have the other great apes) have had a developed gesture system that is still visible in human communication [68]. Somewhere before the common ancestor of modern humans and Neandertals (600 000 years ago) all the genetic and physiological prerequisites for speech seem to have been in place [68]. During the intervening million years, simple vocal turn-taking may have provided the framework for an evolving linguistic complexity, exactly as it does with infants today. The temporal properties of turn-taking may have remained fixed as ever more complex linguistic material was progressively packed within turns, with language diversity now being driven by cultural evolution. This would go some way to explaining how the modern system evolved with the intensive processing forced by rapid production and response of brief vocal turns.

**Concluding Remarks**

This article has advanced five propositions, each with substantial empirical backing, which together suggest a sixth more speculative one:

*Proposition 1.* Turn-taking among humans is universal, although languages are culture-specific.

*Proposition 2.* Turn-taking is at the limits of human performance, involving the rapid encoding of complex structures in small chunks and the anticipation of incoming content.

*Proposition 3.* Languages are surprisingly free to vary despite these functional pressures.

*Proposition 4.* Turn-taking precedes language in ontogeny, but when language is acquired children struggle for years to squeeze complex language into the short turn sizes within adult response times.

*Proposition 5.* Turn-taking is evidenced across all the major branches of the primate order. Taken together, these five propositions suggest a sixth, more speculative proposition:

*Proposition 6.* Turn-taking was prior to language in phylogeny, a proposition that would help to explain propositions 1–5.

For all these reasons, the study of turn-taking promises new insights into the foundations of human communication, while raising many questions for future research (see Outstanding Questions).

**Acknowledgments**

The work reported in this article was funded by the European Research Council (Advanced Grant INTERACT 269484) and the Max Planck Gesellschaft. The author would like to thank the many members of the Max Planck Institute for Psycholinguistics, especially in his department, who have contributed to the research that underlies this paper; and Sara Bögels, Penelope Brown, Elma Hilbrink, Judith Holler, Kobin Kendrick and Francisco Torreira for corrections on a draft.

**References**


**Outstanding Questions**

The study of turn-taking in the cognitive sciences is still in its infancy, and gives rise to the following questions and challenges:

The rapid exchange of communicative acts in conversation is the elite capacity of our species – what special adaptations make it possible?

How can psycholinguistics investigate language in its native dialog habitat? The challenge is to find experimental paradigms that retain sufficient control while not losing the essential phenomenon of interlocked comprehension and production.

Crucial for rapid response is early speech-act prediction or recognition, but how is this achieved? What in general is the time-course of the apparent overlap of production and comprehension processes?

What is the systematic imprint on language structure of the intense cognitive processing involved in turn-taking? What, for example, are the different costs and benefits of different word-orders in different languages?

How exactly does turn-taking capacity develop through infancy and childhood? Most of the work on infant and child turn-taking was performed in the 1970s; we need more research using modern methods.

Is the turn-taking in some of the other primates, in both gestural and vocal modalities, an evolutionary analogy or homology?


