Universals and cultural variation in turn-taking in conversation

Tanya Stivers[a,1], N. J. Enfield[a], Penelope Brown[a], Christina Englert[b], Makoto Hayashi[c], Trine Heinemann[d], Gertie Hoymann[a], Federico Rossano[a], Jan Peter de Ruiter[a,e], Kyung-Eun Yoon[f], and Stephen C. Levinson[a]

[a]Language and Cognition Group, Max Planck Institute for Psycholinguistics, 6525XD Nijmegen, The Netherlands; [b]Center for Language and Cognition, University of Groningen, 9172TS Groningen, The Netherlands; [c]Department of East Asian Languages and Cultures, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [d]Sønderborg Participatory Innovation Research Center and The Institute of Business Communication and Information Science, University of Southern Denmark, 6400 Sønderborg, Denmark; [e]Faculty for Linguistics and Literary Sciences, Bielefeld University, D-33501 Bielefeld, Germany; and [f]Department of African and Asian Languages and Literatures, University of Florida, Gainesville, FL 32611

Edited by Paul Kay, International Computer Science Institute, Berkeley, CA, and approved April 28, 2009 (received for review April 2, 2009)

Informal verbal interaction is the core matrix for human social life. A mechanism for coordinating this basic mode of interaction is a system of turn-taking that regulates who is to speak and when. Yet relatively little is known about how this system varies across cultures. The anthropological literature reports significant cultural differences in the timing of turn-taking in ordinary conversation. We test these claims and show that in fact there are striking universals in the underlying pattern of response latency in conversation. Using a worldwide sample of 10 languages drawn from traditional indigenous communities to major world languages, we show that all of the languages tested provide clear evidence for a general avoidance of overlapping talk and a minimization of silence between conversational turns. In addition, all of the languages show the same factors explaining within-language variation in speed of response. We do, however, find differences across the languages in the average gap between turns, within a range of 250 ms from the cross-language mean. We believe that a natural sensitivity to these tempo differences leads to a subjective perception of dramatic or even fundamental differences as offered in ethnographic reports of conversational style. Our empirical evidence suggests robust human universals in this domain, where local variations are quantitative only, pointing to a single shared infrastructure for language use with likely ethological foundations.

cooperation | response speed | social interaction

Crucial to understanding the nature and origins of human language, perhaps our most distinctive trait, is understanding the social-interactional matrix in which it is used. Informal conversation is where language is learned and where most of the business of social life is conducted. A fundamental part of the infrastructure for conversation is turn-taking, or the apportioning of who is to speak next and when (1). Previous research on turn-taking has examined cues used in recognizing opportunities for turn transition (1–4), the time course of a turn in an exchange (5), and the timing of turn transitions (1, 6–10). In English conversation speakers do not wait for pauses to begin their turn but avoid gaps and overlaps. To achieve this they use grammar, prosody, and pragmatics to project when they can start a next turn, suggesting that turn-taking is specifically organized to achieve this close timing. Here, we consider whether this organization varies across human cultures or is reflective of a universal system of rules for turn-taking in conversation. To our knowledge, no previous study has set out to test the robustness of a turn-taking system for informal interaction across the diversity of human cultures.

In the anthropological literature there are frequent claims that cultures differ radically in the timing of conversational turn-taking, and thus that the findings for English are culture-specific. Nordic cultures, for example, are said to relish long delays between one turn and the next. As the report goes, "Two brothers of Häme (Finland) were on their way to work in the morning. One says, 'It is here that I lost my knife'. Coming back home in the evening, the other asks, 'Your knife, did you say?'" (11). Or receiving visitors in the North of Sweden: "We would offer coffee. After several minutes of silence the offer would be accepted. We would tentatively ask a question. More silence, then a 'yes' or a 'no'" (12). Compare this preference for silence between turns with the reported "fast rate of turn-taking" and "preference for simultaneous speech" in New York Jewish conversation (13) or the "anarchic" conversation of an Antiguan village, in which there is said to be "no regular requirement for 2 or more voices not to be going on at the same time" (12). Although there are many such claims in the anthropological literature of cultures where substantial overlap is the norm (14–16) or where long silences are said to be the rule (11, 12, 17), no broad-ranging, quantitative comparison has been made. These claims suggest that there are culturally variable turn-taking systems.

In contrast to these claims of diversity, there are arguments in favor of a universal system for turn-taking, that, as in English, follows a norm of "minimal-gap minimal-overlap" (18). First, there is a functional basis for turns to be immediately adjacent (rather than overlapping or overly separated): a timely response makes clear its link to another speaker's prior utterance (19), displaying that it is directly contingent on that utterance (20), and showing how the prior utterance was understood, allowing rapid correction if necessary (1, 21, 22). Second, there is evidence for a human ethological basis for adjacent sequences of communicative action and response, for example in very early "proto-conversation" between newborns and caregivers (23–26). Systems in which turn transitions occur with minimal delay or overlap have been described for several languages (1, 8, 27, 28), but no systematic cross-linguistic comparison has been undertaken.

Here, we test these opposing hypotheses: (*i*) a universal system hypothesis, by which turn-taking is a universal system with minimal cultural variability, and (*ii*) a cultural variability hypothesis, by which turn-taking is language and culture dependent. The universal system hypothesis predicts a unimodal distribution of turn transitions with most transitions occurring ≈ 0 in all languages, whereas the cultural variability hypothesis predicts that overlap is more common in some languages and gaps more common in others.

If a community of speakers shows a highly regular target for the timing of turn transition, deviations will come to have a natural communicative significance (e.g., delays implying prob-

Author contributions: T.S. and N.J.E. designed research; T.S., N.J.E., P.B., C.E., M.H., T.H., G.H., F.R., J.P.d.R., K.-E.Y., and S.C.L. performed research; T.S. and N.J.E. analyzed data; and T.S., N.J.E., and S.C.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: tanya.stivers@mpi.nl.

This article contains supporting information online at www.pnas.org/cgi/content/full/0903616106/DCSupplemental.

lems with the prior utterance), so giving rise to implicit norms of timely response that will be maintained to avoid such added implications (29). Research on questions in English conversation has shown that speakers display inhibition in producing responses that in some way fail to conform with the terms of the question or with the questioner's agenda: thus, responses are often delayed by up to 1 s if, for example, they do not answer the question (e.g., *I don't know* or *I can't remember*) (30, 31) or if they give a response that runs against the bias of the question (e.g., A: *Is that your car*? B: *No*) (32, 33).

Two further explanations for variation in turn transition speed are associated with nonverbal behavior such as head movements (e.g., nodding) and gaze. Although the rules for turn-taking may discourage overlap in the vocal channel, they may nevertheless leave other channels exempt. If nonverbal signals are viewed as less intrusive upon speech, they may come earlier than purely verbal responses. Additionally, if questioners fix their eye gaze on their addressees, this may be expected to elicit faster responses. Research on conversation in European languages suggests that a speaker's gaze toward a listener may increase the pressure to respond and to respond quickly: eye gaze does this by indicating who is addressed (1), by providing early possible cues that the speaker's turn is now coming to an end (4, 6, 34) and signaling the speaker's heightened expectancy for a response (35). However, gaze behavior may show substantial cultural variation (36).

With respect to these 4 accounts for delayed turn transition (nonanswering responses, disconfirmations, vocal-only responses, and nongazing questions), the 2 hypotheses make different predictions. The universal system hypothesis predicts that the languages will all show the same pattern of slower turn transitions when these factors are present. By contrast, the cultural variability hypothesis predicts that delayed turn transition will be explained by different factors in different languages and that the 4 factors just mentioned are unlikely to account for variation in the same way cross-linguistically.

To test these competing hypotheses, we compared data from video recordings of informal natural conversation in 10 languages from 5 continents, e.g., from Southeast Asia, Mexico, Namibia, and Papua New Guinea (see Table S1). The languages vary fundamentally in type (e.g., in word order, sound structure, grammatical options) and are drawn from cultures of quite different kinds (from hunter–gatherer groups to peasant societies to large-scale postindustrial nations). To achieve a natural control over the discourse environment to be compared, we took advantage of a universal context for turn transition, namely that between questions and their responses. For optimal comparability we restricted the comparison to polar questions (questions that expect a yes or no answer). These are the most common type of questions in 9 of the 10 languages (67% of total questions in our 10-language sample were of this type), and they are also logically the simplest type: unlike responses to WH- questions (see Table S2), the desired response to a polar question comes from a small, closed set, usually yes or no. Although not all languages have precise equivalents of English yes and no, they all do have ways of asking polar questions and ways of conveying the basic functions of yes and no. For example, yes can be conveyed by repeating the key information in the question [e.g., Q: *Is John going*?, A: *He's going* (= yes)] or the use of nonstandard expressions like *uh huh* or *yep*. To determine whether question–response sequences are representative of turn-taking in general, we examined a corpus of Dutch conversation (8) for timing across all types of turns and responses and found no difference between response times after questions and nonquestions (see Fig. S1). This suggests that the use of question–answer sequences is a reasonable proxy for turn-taking more generally.

Results

Distribution of Turn Transitions. The temporal relation between a turn and its response we will call the response offset, measured in milliseconds, when there is a gap we have a positive offset, when there is an overlap we have a negative offset. As Fig. 1 shows, we find that the response timings for each language, although slightly skewed to the right, have a unimodal distribution with a mode offset for each language between 0 and +200 ms, and an overall mode of 0 ms (see Fig. 1 and Table S3). The medians are also quite uniform, ranging from 0 ms (English, Japanese, Tzeltal, and Yélî-Dnye) to +300 ms (Danish, ǂĀkhoe Hai‖om, Lao) (overall cross-linguistic median +100 ms).

The means display somewhat more variation, as shown in Fig. 2. Danish has the slowest response time on average (+469 ms) and Japanese has the fastest (+7 ms). The mean response offset for the full dataset is +208 ms, and the language-specific means fall within ≈250 ms either side of this cross-language mean, approximately the length of time it takes to produce a single English syllable (37).

The Implications of Turn Delay. ***Answering vs. not.*** Speakers of all of the languages provide answers significantly faster than nonanswer responses to questions (Fig. 3). In all of the languages we also found a greater proportion of answers than nonanswer responses (ranging from 64% of all responses in Korean to 87% in Dutch and Yélî-Dnye) (see Table S4).

Confirming vs. disconfirming. Within the set of answers, those that are confirmations are delivered faster than disconfirmations in all languages, between 100 and 500 ms faster on average (see Fig. 4). This difference reaches significance in 7/10 languages. In all of the languages, we also found a greater proportion of confirmations than disconfirmations (ranging from 70% of all answers in Danish to 89% of all answers in Yélî-Dnye; see Table S4). This advantage for affirmation also holds, incidentally, even if the affirming response is negative in form (as in "You're not coming?" and "No, I'm not"), showing that it is not simply a side-effect of the greater processing costs of negative responses (38) (confirmations using no are not significantly slower than confirmations using yes; 90 vs. 36 ms; $t[693] = -1.1$).

The Implications of Nonverbal Channels. ***Visible responses vs. vocal-only responses.*** Visible responses were most commonly head nods, but we also found shrugs and head shakes, and in some languages like Yélî-Dnye conventionalized extended blinks and eyebrow flashes in response to questions. When visible responses occurred in response to a question, they were faster than speech in every language (see Fig. 5). This reached significance in 7/10 of the languages even though there was substantial variation in how frequently visible responses were included in a response (from 21% of responses including a visible component in ǂĀkhoe Hai‖om to 60% in Italian) (see Table S4).

Questioner gaze vs. no gaze. We found in 9 of the 10 languages that responses were delivered earlier if the speaker was looking at the recipient while the question was asked (Fig. 6). The differences reach statistical significance in only 5 languages. That Danish shows the opposite timing trend, although nonsignificant, combined with known differences in reliance on interactional gaze in different languages, suggests that gaze may be more culturally variable than other behaviors (36). This is also supported by the range of frequencies of gaze to addressee (from 21% in ǂĀkhoe Hai‖om to 88% in Japanese) (see Table S4). This is incidentally not the expectation in the literature, where addressee gaze rather than speaker gaze has been argued to be the norm (4, 6, 39).

Multivariate Results. The results so far show broadly similar patterns of response timing across the languages, with the same 4 factors each independently accounting for faster or slower than

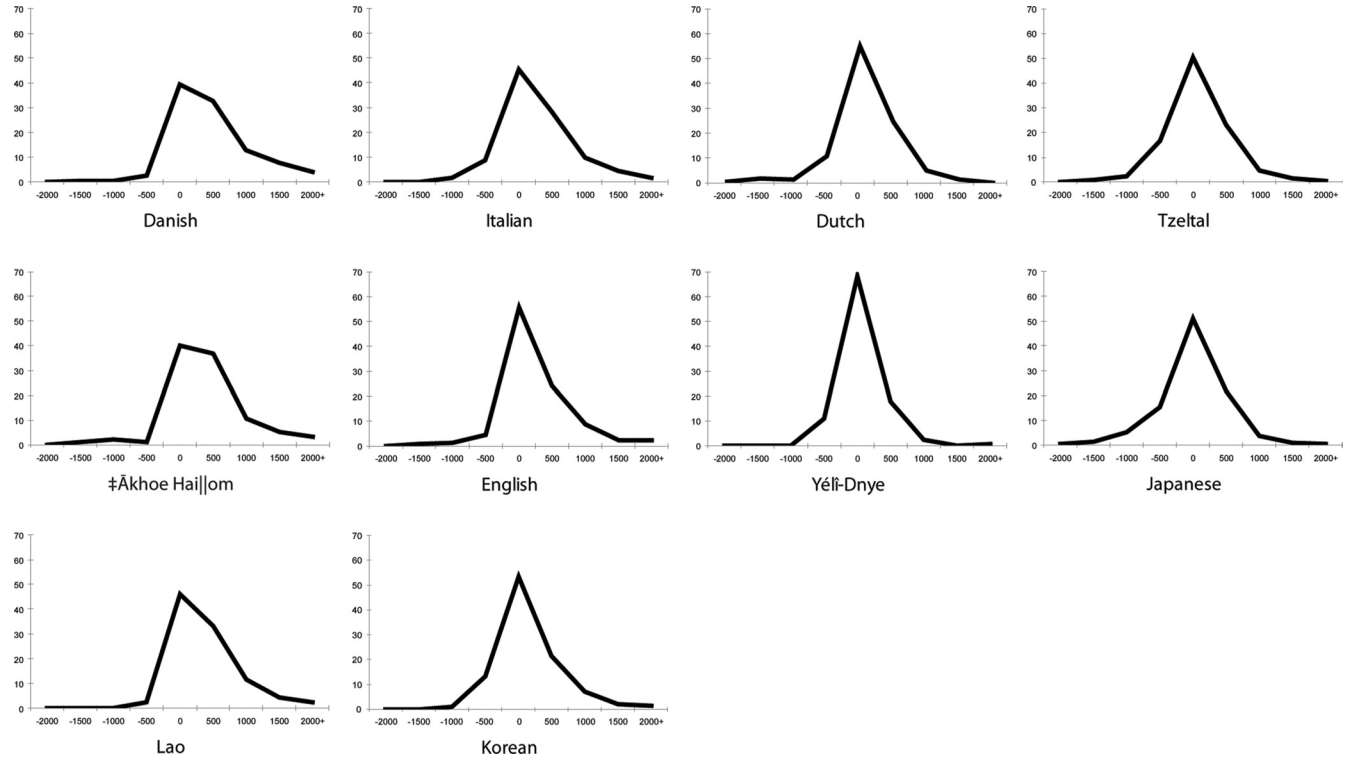

Fig. 1. The distribution of turn transitions for each language in the 10 sample languages. All distributions are unimodal with the highest number of transitions occurring between 0 and 200 ms. The percentage of turn transitions is shown on the *y* axis, and milliseconds of turn offset are shown on the *x* axis.

average responses Multivariate analysis confirms that these 4 are significant predictors of the speed of response in turn transition (see Table 1). Nonanswer responses are significantly slower than answer responses (positive value indicates longer turn transition time). Confirmation responses are faster than disconfirmation responses (negative value indicates shorter turn transition time); visible responses are faster than responses without a visible component; and questions delivered with questioner gaze are responded to more quickly than questions without questioner gaze. Information requests are slower than questions with other functions such as those initiating repair. These factors are significant predictors even when considered together. The model also shows that the conversation in which the question occurs and the language being spoken both further contribute to the variation we observed (see Fig. 2). However, because language spoken and source conversation were treated as different levels the 4 predictor variables are shown to be language-independent predictors.

Discussion

Our results provide substantial support for the universal system hypothesis. The findings suggest a strong universal basis for turn-taking behavior, in that all languages show a similar distribution of response offsets (unimodal peak of response within 200 ms of the end of the question). The distribution of response offsets in all languages reflects a target of minimal overlap and minimal gap between turns. These results also show that the same set of explanations for a delayed response apply across languages.

Amid a strong universal pattern, we do see measurable cultural differences. However, the range that we show, mean

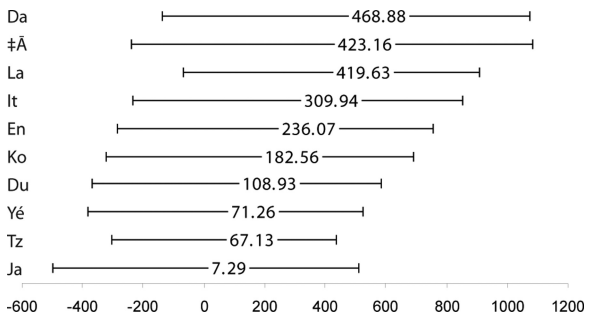

Fig. 2. The mean time (in ms) of turn transitions for each language (±1 SD) in the 10 sample languages shows that speakers of all languages have an average offset time that is within 500 ms. However, there is a continuum of faster to slower averages across the sample. Milliseconds are shown on the *x* axis. Languages are arrayed along the *y* axis. Da, Danish; ‡Ā, ǂĀkhoe Hai‖om; La, Lao; It, Italian; En, English; Ko, Korean; Du, Dutch; Yé, Yélî-Dnye; Tz, Tzeltal; Ja, Japanese.

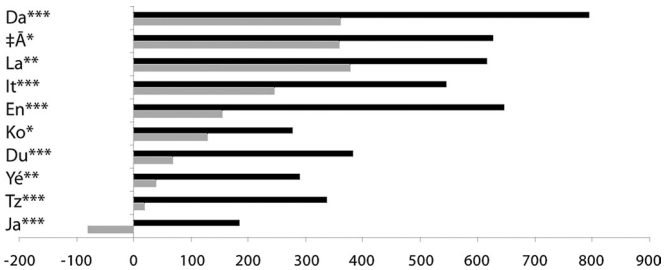

Fig. 3. The mean time of turn transition for responses coded as answers versus responses coded as nonanswer responses in each of the languages. Speakers of all languages produced answers (gray) faster, on average, than they produced nonanswer responses (black). *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$. Milliseconds are shown on the *x* axis. Languages are arrayed along the *y* axis. Da, Danish; ‡Ā, ǂĀkhoe Hai‖om; La, Lao; It, Italian; En, English; Ko, Korean; Du, Dutch; Yé, Yélî-Dnye; Tz, Tzeltal; Ja, Japanese.

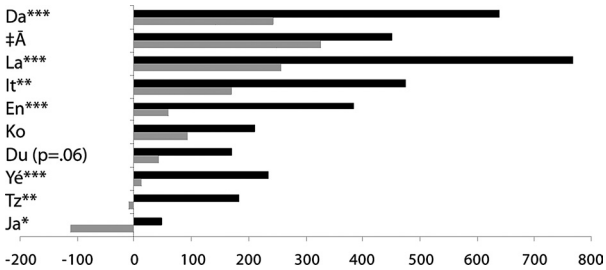

Fig. 4. The mean time of turn transition for responses coded as confirmations versus responses coded as disconfirmations in each of the languages. Speakers of all languages produced confirmations (gray) faster, on average, than they produced disconfirmations (black). *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$. Milliseconds are shown on the *x* axis. Languages are arrayed along the *y* axis. Da, Danish; ǂĀ, ǂĀkhoe Hai‖om; La, Lao; It, Italian; En, English; Ko, Korean; Du, Dutch; Yé, Yélî-Dnye; Tz, Tzeltal; Ja, Japanese.

offset of next turn in each language departing no more than a quarter-second from the overall mean, is not of the kind that would imply fundamentally different types of turn-taking systems in the different languages, as the cultural variability hypothesis would suggest.

Language structure does not explain the variance we observe. Languages that mark questions using a sentence-final marker might plausibly have been associated with slower responses because the fact that the utterance is a question may not be evident until the very end of the turn (28). However, Japanese, Korean, and Lao all use sentence-final marking for questions, yet they do not cluster together within the cross-language range of mean turn offsets (Fig. 2). A converse prediction, that languages like Danish, Dutch, and English, which tend to mark questions at the beginning of a turn, would allow faster responses, also turns out not to hold up. These 3 languages similarly do not cluster together (Fig. 2). Finally, note that this failure of Dutch, English, and Danish to cluster within the cross-language range of mean turn offsets is also evidence that linguistic and cultural kinship (in this case, West Germanic) does not predict interactional tempo.

We suggest that the differences involve a different cultural "calibration" of delay, thus constituting minor variation in the local implementation of a universal underlying turn-taking system, in which speakers aim to minimize the perceived gap before producing a following turn at talk. This target for ideal turn transition remains in a narrow window within each language, with each of 4 factors predisposing a response to be slower (or faster in the case of gaze) than the mean and having similar effects for all of the languages. These differences could either be

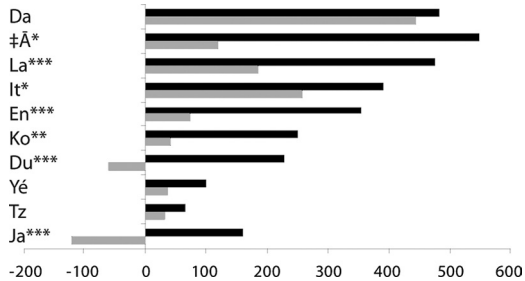

Fig. 5. The mean time of turn transition for responses coded as including a visible response versus responses coded as vocal only in each of the languages. Speakers of all languages produced responses with a visible component (gray) faster, on average, than they produced vocal only responses (black). *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$. Milliseconds are shown on the *x* axis. Languages are arrayed along the *y* axis. Da, Danish; ǂĀ, ǂĀkhoe Hai‖om; La, Lao; It, Italian; En, English; Ko, Korean; Du, Dutch; Yé, Yélî-Dnye; Tz, Tzeltal; Ja, Japanese.

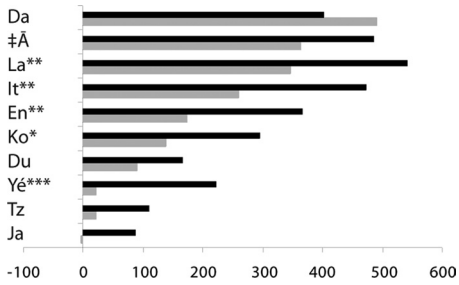

Fig. 6. The mean time of turn transition for questions coded as with speaker gaze versus questions coded as without speaker gaze in each of the languages. Speakers of 9/10 languages produced responses to questions with speaker gaze (gray) faster, on average, than they produced responses to questions without speaker gaze (black). *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$. Milliseconds are shown on the *x* axis. Languages are arrayed along the *y* axis. Da, Danish; ǂĀ, ǂĀkhoe Hai‖om; La, Lao; It, Italian; En, English; Ko, Korean; Du, Dutch; Yé, Yélî-Dnye; Tz, Tzeltal; Ja, Japanese.

a because of a specific cultural interactional pace or follow from more general differences in the overall tempo of social life (40). This would mean that speakers of all languages aim at minimizing significant delays relative to the specific rhythm of that language in conversation (e.g., ref. 41), a perspective that is supported by existing studies of some non Indo-European languages (27, 28, 42). To address this hypothesis we coded the offset of our responses for whether or not, when a relative subjective measure of the conversation's rhythm was taken into account, responses were coded as late versus on time. Mean response times for subjectively on-time responses are much longer in Danish and Lao (203 and 202 ms, respectively) than in Japanese and Tzeltal (36 and 83 ms, respectively) and comparing the 3 languages with longest response offsets to all others, the difference is significant [$t(847) = -10.97$, $P < 0.001$]. Thus, a silence of 200 ms, judged as a delay in most languages, was still considered on time. Such a silence is thus not phenomenologically salient within a speech community (but may be to an outside observer). In short, what constitutes a subjectively notable delay involves greater absolute duration in some languages than in others. This is consistent with the presence of a universal, stable system of turn taking avoiding overlap and

Table 1. Mixed-level multiple linear regression model predicting response time

Level 1 variables	Estimate	95% CI
Response variables		
Nonanswer response	131.78***	59.34, 204.23
Confirmation	−206.87***	−268.61, −145.12
Visible response component	−86.93***	−136.76, −37.10
Question variables		
Information request only	129.38***	79.30, 179.46
Questioner gaze	−69.28**	−123.48, −15.08
Context variables		
Level 2: Variance at language level	19555.05*	7342.23, 57304.20
Level 3: Variance at interaction level	14091.24***	7715.57, 25735.37

The mulitvariate model shows that nonanswer responses are slower than answer responses and responses to information questions are slower than responses to other sorts of questions. Confirmations, responses with a visible component, and responses to questions that are delivered with speaker gaze are shown to be faster than disconfirmations, vocal responses and other sorts of questions, respectively. Language (i.e., the language being spoken) and conversation (i.e., the conversation from which a data point was taken) were treated as levels and thus the results are language independent. *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$.

minimizing gaps, but where there are different local metrics for what counts as a delay in response (18).

The variation we found between mean response times in different languages does not coincide wholly with the ethnographic expectations reported in the literature (14, 17). On the basis of these reports, Italian speakers should be more tolerant of overlap, but we found a mean offset of +310 ms, indicating that they in fact tend to leave a slightly longer than average gap before producing a next turn. And only 17% of all responses overlap, not at all an unusual proportion. Similarly, Japanese speakers are said to leave substantial gaps of silence before responding, but our findings show that Japanese speakers are, on average, the earliest to respond of all of the languages in our sample. Danish speakers were more consistent with ethnographic reports for Nordic languages, showing the longest mean response time in our sample. However, note that the mode was still quite small (100 ms), suggesting that here too speakers target minimizing gaps and overlaps in response offset time. To put the Danish case in context, recall that the mean offset in Danish was less than a half-second in total, a quarter-second deviation from the cross-linguistic average, and thus far from the lengthy pauses measured in minutes or even hours suggested by the ethnographic reports mentioned above.

Conclusion

We have shown strong parallels in turn-taking behavior across 10 languages of varied type, geographical location, and cultural setting. All of the languages show on average a small positive offset in response time, i.e., responses tend to be neither in overlap nor delayed by more than a half-second. The factors that predict whether a response will be faster or slower within each language are identical across the languages. These results offer systematic cross-linguistic support for the view that turn-taking in informal conversation is universally organized so as to minimize gap and overlap, and that consequently, there is a universal semiotics of delayed response.

How then to account for the ethnographic reports of diversity in this domain and the phenomenology of significant cross-cultural differences in response timing? Our data suggest that the regimentation of tempo within a culture is tight, and we come to expect a particular interactional metabolism as it were, slight departures from which have the associated contexts and interactional significance we have established above. Speakers become hypersensitive to perturbations in timing of responses, measured in 100 ms or less. This sensitivity to subtle variation may be responsible for the subjective impression by outsiders of "huge silences" in the case of Nordic languages (insiders, of course, will be calibrated to a local norm, as shown above by different subjective measures of what counts as delay). The actual difference between the 10-language norm and the average turn transition in Danish is confined to the time it takes to utter a single syllable. Our findings are in line with other close examinations of ethnographic outliers; see ref. 27 on the alleged Antiguan preference for overlap (12).

Abstracting from this fine-tuning of interactional tempo across cultures, our results point to robust universals in this domain. Strong universals of this kind seem to be much harder to find in the grammatical structure of languages than in the interactional systems that underlie their use (43). Our results argue for an interactional foundation for language that is relatively stable and relatively separable from the specific languages and cultural practices that instantiate it (23). Understanding this will be crucial for understanding the origin of language and the foundations of social life, because it is out of primordial interaction that languages and cultures are ultimately built.

Data and Methods

Data. All contributors collected videotaped interactions of maximally informal, spontaneous, naturally occurring conversations, each with 2–6 consenting participants. We confine ourselves here to informal conversation, in which turn-taking is essentially self-organizing. Other procedures hold in highly structured institutional interaction (e.g., courts of law, church services, news interviews), which are often subject to explicit rules for who may speak and when. Participants were often engaged in additional activities (e.g., eating, drinking, or stringing beads). As long as the task was not determining the direction or structure of the conversation this was considered acceptable. Each contributor identified 350 consecutive questions across 5–17 separate interactions (101 conversations in the total dataset). No interaction accounts for >3% of the dataset and very few individuals participated in >1 conversation. Both of these features minimized the influence of any 1 individual or interaction on the overall pattern.

Each question and response, where one occurred, was coded for its form and function. In coding question–response sequences we drew on conversation analytic research on social interaction (1). This study uses an applied version of this method to test findings comparatively. In this study, only functional yes–no questions were included. Codes that are relevant to this study are described in Table S2. Numerical overviews of the data coding by language are provided in Table S4.

Response time was defined as the time elapsed between the end of the question turn and the beginning of the response turn. The response turn was considered to begin if a vocal or gestural response was initiated. Absolute response time was coded auditorily and then the time from the end of the question to the beginning of the response was measured instrumentally by using annotation software ELAN (www.lat-mpi.eu/tools/elan) in 10-ms increments rounded to 100 ms. Subjective measures of whether a response was delivered on time were done auditorily for all responses not delivered in overlap. Coders were asked to consider the rhythm of the conversation leading up to the response and to judge whether it sounded delayed or not.

Analytic Methods. To examine the distribution of turn offsets by language we calculated mean, median, and mode information for each language (see Table S3) and plotted the distribution of turn offset times. To test whether the same factors account for faster and slower response times in different languages, 2 sample *t* tests were done across 8 categories of data reported in Figs. 3–6: answers vs. nonanswer responses, confirmations vs. disconfirmations, responses with visible components vs. vocal-only responses, and responses to questions with questioner gaze and responses to questions without gaze. Details of these t tests are in Tables S5 and S6. This information informed the design of multivariate analysis that was performed by using a multilevel mixed effects linear regression model in STATA. This takes into consideration that there is clustering in the data: responses to questions are clustered within 101 interactions. Interactions are clustered within 10 languages. This model tested for association and explanatory power at each of 3 levels.

ACKNOWLEDGMENTS. We thank Michael Dunn, John Heritage, and Asifa Majid for comments on earlier drafts of this article. This work was carried out in the Multimodal Interaction Project within the Language and Cognition Group at Max Planck Institute for Psycholinguistics, funded by the Max Planck Society.

1. Sacks H, Schegloff EA, Jefferson G (1974) A simplest systematics for the organization of turn taking for conversation. *Language* 50:696–735.
2. Stephens J, Beattie GW (1986) On judging the ends of speaker turns in conversation. *J Lang Soc Psychol* 5:119–134.
3. Clancy PM, Thompson SA, Suzuki R, Tao H (1996) The conversational use of reactive tokens in English, Japanese, and Mandarin. *J Pragmat* 26:355–387.
4. Kendon A (1967) Some functions of gaze direction in social interaction. *Acta Psychol* 26:22–63.
5. Chapple ED (1939) Quantitative analysis of the interaction of individuals. *Proc Natl Acad Sci USA* 25:58–67.
6. Duncan S, Jr, Fiske DW (1977) *Face-to-Face Interaction: Research, Methods, and Theory* (Wiley, New York).
7. Beattie GW, Cutler A, Pearson M (1982) Why is Mrs. Thatcher interrupted so often? *Nature* 300:744–747.
8. de Ruiter JP, Mitterer H, Enfield NJ (2006) Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82:515–535.
9. Wilson M, Wilson TP (2005) An oscillator model of the timing of turn-taking. *Psychonom Bull Rev* 12: 957–968.
10. Jefferson G (1973) A case of precision timing in ordinary conversation: overlapped tag-positioned address terms in closing sequences. *Semiotica* 9:47–96.

11. Lehtonen J, Sajavaara K (1985) in *Perspectives on Silence*, eds Tannen D, Saville-Troike M (Ablex, Norwood, NJ), p 198.
12. Reisman K (1974) in *Explorations in the Ethnography of Speaking*, eds Bauman R, Sherzer J (Cambridge Univ Press, Cambridge, UK), pp 110–124.
13. Tannen D (1985) in *Perspectives on Silence*, eds Tannen D, Saville-Troike M (Ablex, Norwood, NJ), pp 93–112.
14. Agliati A, Vescovo A, Anolli L (2005) in *The Hidden Structure of Interaction: From Neurons to Culture Patterns*, eds Anolli L, Duncan S, Jr, Magnusson MS, Riva G (IOS Press, Amsterdam), pp 223–235.
15. Sugawara K (1996) in *Afrcan Study Monographs* 22(Suppl):145–164..
16. Wieland M (1991) in *Pragmatics and Language Learning 2*, eds Bouton L, Kachru Y (University of Illinois Press, Urbana), pp 101–118.
17. Gudykunst WB, Nishida T (1994) *Bridging Japanese/North American Differences* (Sage Publications, Thousand Oaks, CA).
18. Schegloff EA (2006) in *Roots of Human Sociality: Culture, Cognition, and Interaction*, eds Enfield NJ, Levinson SC (Berg, Oxford), pp 70–96.
19. Schegloff EA (1968) Sequencing in conversational openings. *Am Anthropol* 70:1075–1095.
20. Gergely G, Nádasdy Z, Csibra G, Bíró S (1995) Taking the intentional stance at 12 months of age. *Cognition* 56:165–193.
21. Sacks H (1963) Sociological description. *Berkeley J Sociol* 8:1–16.
22. Schegloff EA, Jefferson G, Sacks H (1977) The preference for self-correction in the organization of repair in conversation. *Language* 53:361–382.
23. Levinson SC (2006) in *Roots of Human Sociality: Cognition, Culture, and Interaction*, eds Enfield NJ, Levinson SC (Berg, London), pp 39–69.
24. Murray L, Trevarthen C (1986) The infant's role in mother–infant communication. *J Child Lang* 13:15–29.
25. Meltzoff AN, Moore MK (1977) Imitation of facial and manual gestures by human neonates. *Science* 198:75–78.
26. Striano T, Henning A, Stahl D (2006) Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction Studies* 7:251–271.
27. Sidnell J (2001) Conversational turn-taking in a Caribbean English Creole. *J Pragmatics* 33:1263–1290.
28. Tanaka H (1999) *Turn-Taking in Japanese Conversation: A Study in Grammar and Interaction* (John Benjamins, Amsterdam).
29. Levinson SC (1983) *Pragmatics* (Cambridge Univ Press, Cambridge, UK).
30. Clayman S (2002) in *Advances in Group Processes: Group Cohesion, Trust and Solidarity*, eds Lawler EJ, Thye SR (Elsevier, Oxford), pp 229–253.
31. Stivers T, Robinson JD (2006) A preference for progressivity in interaction. *Lang Soc* 35:367–392.
32. Heritage J (1984) *Garfinkel and Ethnomethodology* (Polity Press, Cambridge, UK).
33. Pomerantz A (1984) in *Structures of Social Action: Studies in Conversation Analysis*, eds Atkinson JM, Heritage J (Cambridge Univ Press, Cambridge), pp 57–101.
34. Duncan S, Jr (1974) in *Nonverbal Communication*, ed Weitz S (Oxford Univ Press, New York), pp 298–311.
35. Stivers T, Rossano F (2009) Mobilizing response. *Res Lang Social Interaction*, in press.
36. Rossano F, Brown P, Levinson SC (2009) in *Conversation Analysis: Comparative Perspectives*, ed Sidnell J (Cambridge Univ Press, Cambridge), pp 187–249.
37. Greenberg S (1999) Speaking in shorthand: A syllable-centric perspective for understanding pronunciation variation. *Speech Commun* 29:159–176.
38. Clark HH (1976) *Semantics and Comprehension* (Mouton, The Hague).
39. Bavelas JB, Coates L, Johnson T (2002) Listener responses as a collaborative process: The role of gaze. *J Commun* 52:566–580.
40. Hall ET (1959) *The Silent Language* (Doubleday, New York).
41. Couper-Kuhlen E (1993) *English Speech Rhythm: Form and Function in Everyday Verbal Interaction* (John Benjamins, Amsterdam).
42. Moerman M (1988) *Talking Culture: Ethnography and Conversation Analysis* (Univ Pennsylvania Press, Philadelphia).
43. Evans N, Levinson SC (2009) The myth of language universals: Language diversity and its importance for cognitive science. *Behav Brain Sci*, in press.