JOHN   B.   HAVILAND

## "Nojon beel xal, xk'oon lok'el ts'in at'ele"
## (Me voy satisfecho, así llego bien en el trabajo):
## multimodal polyphonies, phatic communion,
## and metaconversation in a Tseltal encounter

Aurore Monod Becquelin, whom we honor in the present collection, I first met in Mexico sometime in the 1980s when we were both doing field research in closely related communities, hers in the Tseltal-speaking region in the northeastern part of the central Chiapas highlands, and mine to the southwest among Tzotzil speakers. Elegant, erudite, and endlessly generous with her time and ideas, Aurore came to represent for me a prototypical French scholar, a linguistic anthropologist with wide ranging tastes and predilections. When I was fortunate enough to visit her in Nanterre multiple times over the following two decades, she came also to epitomize Paris itself, sharing her home, her kitchen, and her ideas in the marvelous ancient house just around the corner from the Place de la Contrescarpe. (That was coincidentally where, more than two decades before, I had as a youthful banjo-player taken refuge from street busking briefly to join the entourage of New Zealand-born French folksinger Graeme Allwright.[1]) Aurore and Pierre's home was a world of its own, crammed with books in many languages about Brazil, Mesoamerica and beyond, and offering access to a social and intellectual life now, for me, indelibly linked to the flavors of the surrounding Paris streets.

Tseltal (Mayan) in Chiapas was already associated with a large corpus of ethnography, including then fashionable ethnobotany and ethnoscience, that dwelled on the sorts of categorial systems embodied in the lexicon of the language, and also on aspects of its characteristic conversational styles in the ongoing work with Tenejapa women by Penelope Brown. Aurore's own early interests in the language connected closely with and, indeed, inspired my own research on Zinacantec Tzotzil, especially her work on ritual language and "parallelism." But in those early years accessible digital technology did not yet allow serious attention to the embodiment of language in ordinary talk: we were only beginning

---

1. HAVILAND 1967.

to be able to videorecord interactions, and serious analysis of the results was still in its infancy. When invited to contribute to this volume, I thought it fitting to use empirical material that might resonate with Aurore and choose as a foil for consideration a fragment of conversation in Tseltal—one of her languages but not one in which I myself can claim to be competent—to test out some theoretical approaches, many of which she and I have conversed about in later encounters over the years. The fragment of embodied talk involved exemplifies the limits of textuality and what I have recently come to call "co-expressivity": how multiple simultaneous modalities of expression and levels of interactional structure conspire to produce utterances and actions that may lean upon but ultimately transcend words.

### *A Tseltal conversation*

On April 13, 2001, in a rural village in the Tseltal-speaking municipality of Oxchuk, in Chiapas, a young man here called "V" (for visitor) arrived for a visit at the house of Andrés ("A"), a respected former schoolteacher then turned relatively well-off corn and coffee farmer. There ensued a conversation of more than an hour, in which the visitor sought the older man's continued advice and help in a land matter. The conversation was videotaped by Mtro. José Daniel Ochoa Nájera, now of the *Centro Estatal de Lenguas, Arte y Literatura Indígena* in Chiapas and then a Master's student at CIESAS-Sureste. He was collaborating with several Tseltal-speaking families in different parts of the township of Oxchuk to trace phases of the ongoing processes of what appeared to be language shift between Tseltal and Spanish. Daniel shared the video with me, as part of an early web project to document the most widely-spoken indigenous languages of the state, overoptimistically dubbed *El Archivo de los Idiomas Indígenas de Chiapas (*or *AIIC)* whose main goal, never brought to completion, was to construct a corpus of linguistic materials that included videotaped conversations not easily "reduced" to text.

The video was thus a kind of "found object," beyond the scope of my own expertise, but nonetheless one where my own work on Tzotzil conversational interaction could at least educate my eyes, if only by contrast. Brown (2011:37) gives the following succinct characterization of Tseltal interaction (in a study of interactive patterns and the organization of attention in language socialization): "Adult interaction follows norms of restraint, nondemonstrativeness, and avoidance of eye contact."

Although we often describe the contexts we study as "face-to-face" interaction, it has been argued that it might be more appropriate to characterize Tseltal interaction instead as "side-by-side." For example, the best developed presentation of embodied Tseltal conversation appears in one of the few cross-cultural studies of gaze in talk, by Rossano, Brown, and Levinson (2009:192) who characterize the Tseltal "response system" for displaying conversational recipiency as follows: "Tzeltal[2] recipients appear to avoid gaze and tend to produce verbal responses, repeating part of the prior speaker's utterance" as a conventionalized way of providing conversational back-channel. They go on to illustrate, in an excerpt from Tenejapa Tseltal, that two sisters—"sitting, side by side on a bench, with their backs against a wall, approximately two meters away from each other"—"can produce entire [question-answer[3]] sequences without ever looking at each other" (*ibid.*: 200). They go on to characterize the Tenejapa Tseltal speech community as "relatively gaze aversive" (p. 231), although they are careful to stipulate that things might look very different in other Tseltal-speaking communities.

Relevant to the material I present here, Rossano *et al.* make several further observations about gaze in Tenejapa Tseltal. First, they question a very old speculation about gaze as a universal way interlocutors show attention to speakers. If in Tenejapa there is no general expectation that one gaze at a speaker to show one is listening,[4] it may be that "by not looking at anything else [e.g. by gazing into empty space], Tzeltal recipients might simply be displaying full commitment to attending to the conversation" *(ibid.)*. Second, the authors speculate that conventions about using space, notably how and where to sit "together," will partly determine favored patterns of joint attention in talk. They note that "Tzeltal speakers of Tenejapa prefer to sit side by side or at an angle," and that since "[t]he Tenejapan home is relatively private, and visitors would normally be seated on benches or small chairs on a patio

---

2. Conventional spelling of Chiapas Mayan language names has recently opted for the *ts* digraph instead of the previously preferred *tz*.

3. Importantly, Rossano *et al.*'s entire comparative study (2009) is based on a selected sample of conversational sequences, involving only questions-answer pairs, so the results can only be expanded to wider sorts of speech practices in a speculative way. For example, they argue (p. 231): "If one probes further for why Tenejapan Tzeltal speakers are relatively gaze aversive, there is little doubt that it has to do with politeness and decorum: Tzeltal

interactors will even turn their backs, or hide behind a structure, if the conversational matter is in any way face-threatening (e.g., in the case of substantial requests)." It may also be that gendered conventions that involve more embodied forms of protection against face threats (literally covering one's face with a shawl, for example) may be in play as well.

4. "In the Tzeltal case, gaze is not a good indicator of active recipiency, and it plays no role as a predictor of whether a response will be forthcoming" (p. 225).

or enclosed public area" (p. 226), the kinds of interaction the researchers filmed (all outdoors in a patio) may have been skewed toward a quite specific distribution of side-by-side seating arrangements. Of course, patterns of sitting are related to patterns of "visiting" or congregating more broadly—and thus, by implication, they are also sensitive to conversational activities, the nature of relationships between participants and their identities (that is, to fully fledged social indexicality), and to genres of talk, in ways to which I turn at the end of this essay. Rossano *et al.* also suggest that "[t]he low-key role that gaze plays in Tzeltal interaction implies … other effective displays of recipiency … [R]epetitions can signal recipiency and understanding of what has been asked in a … precise way" (p. 230). The Tseltal propensity for extended sequences of selective conversational repetition (Brown 1998) may thus be a compensatory display, in words rather than gaze, that one is following and understanding what an interlocutor is saying.

This scholarship on Tseltal gaze partly explains why Mtro. Ochoa's videotape so drew my attention. My research at the time concentrated on speakers' gestures and their co-expressive roles in utterance. But if Tseltal speakers routinely avoid looking at each other when they talk, how can gesture be coexpressive? There is no obvious gestural parallel to the cyclical partial repetition of talk reported throughout the Mayan area. But the videotape of the visitor and the old man in Oxchuk is crammed with periodic gestural performances, despite the fact that—true to form—the two men arrange themselves side by side and do not appear frequently to look at one other, talking while mostly gazing elsewhere. Intrigued, I requested help from a native Oxchuk Tseltal speaker, María de Jesús Gómez Sánchez, one of the small army of Tseltales from various *municipios* working with Gilles Polian on a comprehensive documentation project (see Polian 2013). For no reason other than the striking visual features of the interaction, I asked her to transcribe the Tseltal and gloss it into Spanish, at which point I put the film away and (almost) forgot about it for almost two decades.

*Co-expressivity and synchrony in a conversational preamble*

Let's dive into the video, much as I did when I rediscovered it, to see what it reveals about spatial arrangements, gaze, gesture, talk, genre, phatic communion, politeness, and co-expressivity—all parts of the constellation of features that, following Aurore, we could group together under the rubric of polyphony, and what recent scholarship
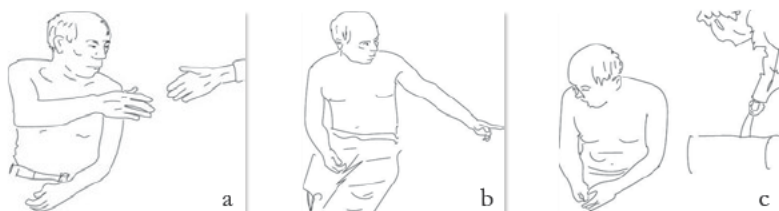
FIG 12.1 — *Opening: a handshake;*
*A pointing to a chair; V placing his machete*

has characterized as co-expressive or conjoint features of "composite utterances" (Enfield 2009).

The film begins (Fig. 1) with A seated on a chair in his yard. As V approaches from the old man's left, the host offers his hand, points to a nearby chair, and waits while V places his machete against a log. During the 14 seconds it takes for this to happen, A only once, at the very start, glances up, possibly at the visitor's face, otherwise keeping his eyes fixed on the proffered hand, the chair, and, at the end, a puppy sitting on the ground to his right.

During this initial episode, however, he does speak and gesture to the newcomer, who on arrival has initiated interaction with a greeting and a standard polite opening ("Are you still alive?"). His host A responds with an equally polite and noncommittal response—"So I am. In fact, I have just finished with my damned food"—then reaching up to shake V's hand (with a formulaic *la` me, mam* 'please come, Sir'), then pointing to the log where the visitor can prop his machete, saying in Spanish "Put it over there," and switching back to Tseltal to point at a chair ("Sit, take a seat") as V greets the lady of the house. All of this is, of course, routine and formulaic—an exchange of greetings and a courteous opening to an interaction that choreograph word, body, gaze, and the physical setting of the immediate scene where the ensuing conversation will take place.

Figure 2 diagrams the close synchrony in this simple opening sequence between what A says in Tseltal, where he gazes, and how he gestures, most particularly with his hands.[5] Each participant in each

---

5. These synchronized timelines are loosely inspired by conventions proposed by Lorenza MONDADA (2016). They were prepared in part using the annotation software ELAN (2022—see SLOETJES and WITTENBERG 2008).
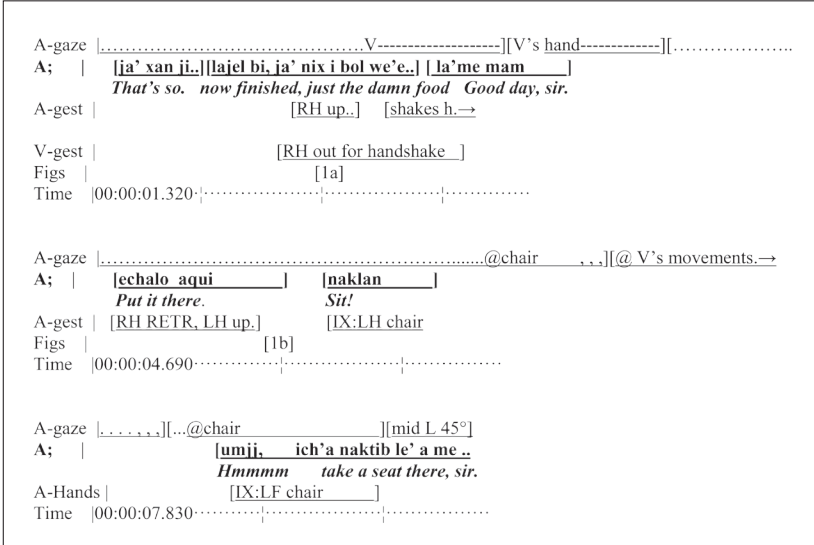
```
A-gaze |......................................V--------------------][V's hand------------][..................
A;    |    |ja' xan ji..||lajel bi, ja' nix i bol we'e..| [ la'me mam_____|
           That's so.   now finished, just the damn food   Good day, sir.
A-gest |                        [RH up..]    [shakes h.→

V-gest |                        [RH out for handshake   ]
Figs   |                             [1a]
Time   |00:00:01.320·|··················|··················|··············


A-gaze |.........................................….......@chair         , , ,][@ V's movements.→
A;    |    |echalo  aqui_____|    |naklan_____|
           Put it there.            Sit!
A-gest |  [RH RETR, LH up.]       [IX:LH chair
Figs   |             [1b]
Time   |00:00:04.690·············|··················|··············


A-gaze |. . . . , , ,][...@chair            ][mid L 45°]
A;    |               [umjj,     ich'a naktib le' a me ..
           Hmmmm       take a seat there, sir.
A-Hands |             [IX:LF chair         ]
Time   |00:00:07.830··········|··················|················
```

FIG 12.2 — *Time-line of speech, gaze, and gesture on A's part*
*(with V's hand making a cameo appearance)*

subsection of this "transcript" is annotated by a group of lines or "tiers." Following early suggestions by Charles Goodwin (1981), above the line for a given speaker's words (if any), there appears a "Gaze tier" which shows where the speaker in question is looking. There follow two lines to transcribe (and gloss) any accompanying speech. These speech lines may, in turn, be followed by one or more further synchronized timelines to show the precise timing and usually an abbreviated[6] characterization of that participant's gestures or bodily movements, with more or less detail, depending on the purposes of the annotation.[7] Such

6. In the gaze line I use the number '2' as a mnemonic for "to" to identify the apparent target of gaze: e.g., "2 V = turning the eyes to gaze at V. I also, sadly somewhat inconsistently, use the abbreviation "@ A" to mean "hold gaze on A." Throughout this paper A identifies the older host participant, and V the younger visitor. Because these Tseltal speakers tend to use a kind of mid-range blank stare, I have coined the abbreviation NW 'nowhere' to indicate such an indeterminate gaze at nothing in particular. In the gesture lines the following abbreviations appear: H 'hand'; F 'finger'; P

'palm'; L 'left'; R 'right'; e.g., RH 'right hand' and LH 'left hand'; BH 'both hands'; R, L, U, D for 'right, left, up, down'; O for 'outward' (i.e., away from the body); RETR for 'retraction' of a gesture; IX for index, or pointing gesture, usually followed by a colon to introduce the pointing articulator, e.g., IX: LF 'left finger pointing index', and then followed by the putative target or referent of a pointing gesture.

7. When it is important, for example, both gaze and gesture tiers can show more precisely the dynamics and synchronization of movements, using conventions introduced by

FIG 12.3 — *Sitting side-by-side but looking away*

annotations allow precise tracking of the dynamics of visual and other bodily movement and their timing with respect to one another and to accompanying speech. There is also a separate tier to identify the rough placement in the accompanying video of the still frames illustrated in the Figures, for example, A's reaching out to take V's hand at Figure 1a. All lines are synchronized to a graduated timeline following each group of simultaneous tiers.

The two interactants then seated themselves, just as the literature on Tseltal might suggest, not face-to-face but rather side-by side, on chairs set a little more than a meter apart. They rarely looked directly at one another as they carried on the theme that A had introduced—his recent meal—but instead mostly gazed in "no particular direction" (Figure 3).

### Conversational genres: Phatic communion or "smalltalk"

As I have argued elsewhere (Haviland 2017), Mayan conversation characteristically combines what might be thought of as quite different speech "genres" within most varieties of ordinary talk, partitioning conversation into distinct phases with different kinds of structure. Particularly in a formal visit, *Tseltales* from Oxchuk seem to resemble *Tzotziles* from Zinacantán: they rarely raise the important matters at hand without first

KENDON (2004)—for example, when the gaze begins to move towards a specific target, or when a hand begins to move towards a specific locus where it may achieve a desired hand-shape (here, somewhat inconsistently marked by a sequence of periods or full stops .....); it may show that an articulator halts or pauses or holds its position (notated via a sequence of dashes -----) or produces what Kendon termed the "stroke" of a fully formed gestural hand configuration (which he indicated with an exclamation point '!'); and, finally, it may indicate when gaze moves away from a target or a gesturing hand retracts toward a rest position (with a sequence of commas ,,,,,).
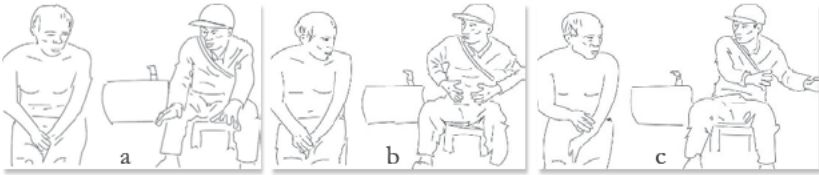
FIG 12.4 — *V glances at A, otherwise occupied*



FIG 12.5 — *"They give me food, I'm full, and I set out to work"*



FIG 12.6 — *Annotated timeline of the first part of this "smalltalk" conversation*

8

exchanging stylized pleasantries, what elsewhere might be called "small-talk," itself a characteristic example of what Malinowski (1936) long ago dubbed "phatic communion." Of course, Malinowski somewhat over-played his hand by suggesting that such "speech in which ties of union are created by a mere exchange of words…fulfill[s] a social function… to establish bonds of personal union between people…and does not serve any purpose of communicating ideas" (pp. 314-316). But why can't it often—if not always—do both things at once, a perfect case of "co-expressivity"?

Before they, as it were, get down to business, A and his visitor carry on several minutes of conversation, starting with the topic A himself has introduced, his recent meal. They thus exemplify a different kind of generic polyphony—not just the multimodal tension and comple-mentarity between, say, words and other visible bodily expressions, but rather the complex transition between a sociable preamble to the con-versation and the principal consequential matters of the entire visit. It is on this latter tension, and the native Tseltal theorizing about it evinced in the later conversation, that I concentrate in the last part of this essay.

So why did A call what he had just managed to eat "the damned food"? As V remarked, "That's what makes one tired." The older man went on to agree: "Of course, one can't suppose that food is free! You eat only if you work, and if you don't want to work neither will you see anything to eat." As A reached down to pick something up from the ground, V allowed himself a quick glance in the other man's direction (Figure 4), then quickly looked away again, almost as if to avoid any sug-gestion that he himself might be subject to the suspicion of being lazy.

"Where," asked A, "will one be able to eat? God has made it clear. One who is lazy will not eat. What can you do? Where can you go? You might just once have someone offer you a tortilla, but only once. But early the next day you'll have nothing." While he narrated this entire hypothetical miniature scenario, A never once actually looked directly at his visitor.

At the end of this little just so story, however, V looked up directly at A, repeated the older man's central dictum—"but only once" will you be given a free meal—and began his own more elaborated version of a comparable moral tale, which took a slightly different perspective, that of a hired worker. From his performance I have drawn the title for this essay. V began, "You're, right, that's happened to me many times. Only after you have worked, you eat. 'Eat tortillas! Drink sweet coffee!'" V gestured [Figure 5a] to show how food was laid out before him. "Then I'm full [Figure 5b], and I set out happily to do my little work." He held

his hands out to his left, depicting how he set out to work with a full belly [Figure 5c].

Figure 6 transcribes this little sequence with more synchronous detail. V was doing the talking here, elaborating on A's earlier aphorisms. He allowed himself to look directly at the older man, who somewhat indirectly returned the gaze, clearly watching V's bodily performance as it developed. With the gesture at 5a, V conjured an imaginary scene in which he, an industrious worker, had been served food, and his "quoted" words personified his imaginary host's invitation to eat and drink. In response, A—inserting himself pronominally into the imagined scene—commented "Let's eat!" V presented an embodied image of a satisfied, full belly (at 5a), and then moved it off stage left (5c).

Conjuring a less happy subsequent possibility, V then moved from the scene of the contented belly moving off stage by raising his hands demonstratively with a shrug to suggest "who knows" or perhaps "after that, nothing left." The synchronized timeline in Figure 8 clearly demonstrates that V's "turn" here (at 7) was gestured rather than audible;



FIG 12.7 — *"Nothing"*



FIG 12.8 — *Timeline for "who knows if there will be any food left?"*

and that it preceded and refigured A's following comment about having to wait for tomorrow: tomorrow there might be NO food. That is, A's first comment *ja' to-ix pajel a bi* "no more until tomorrow" in its second repetition seemed to suggest "no more *at least* until tomorrow, *if then*." V then echoed that sentiment in words.

"Slapping" or wiping of his empty palms together, one up, one down, in a conventional gesture for "leave suddenly," V concluded: "If there's nothing in the house [Figure 9a], then we'll have to depart again to look for more [Figure 9b] because there will be nothing to eat." Here, too, it was the hand gesture (at 9b), coming at the end of the spoken turn, that suggested that once again, to eat one has to (go off to) work. (See the last detailed transcript shown at Figure 10.)

A agreed, "True, life is hard." V did not, however, abandon the theme of the daily grind. His parents, he said, would give him tortillas (Figure 11a) if he made an effort and looked for work. If he shared with them what money he earned, they would in turn share their coffee with him (Figure 11b).



FIG 12.9 — *Nothing in the house, we have to go find more (food)*



FIG 12.10 — *Nothing in the house, have to leave again*

FIG 12.11 — *Share tortillas, share coffee*
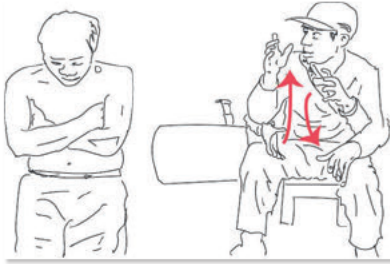


FIG 12.12 — *That's how we pass the days*

And that was how things would go on: *jich nan-ix kuxulotik a, jich yak k'axel, k'axel k'aal ts'in bi* "so that's how we live, that's the way the days keep passing" (Figure 12). V alternately raised and dropped his two hands opposite one another to depict the continual alternation of rhythms of work and domestic life.

For the rest of what I have called the "smalltalk" segment that began this interaction, the two men exchanged conventional wisdom about a series of related topics: not just how one eats only if one works, but also how one has only one life to live; how one cannot expect to become a millionaire, but instead must be content to remain covered in mud, only needing to take a steam bath to get clean again; how good health is the most important thing; and so on—all good Malinowskian "empty" topics.

### Turns, gaze, and gesture

As I have mentioned, past characterizations of Tseltal conversation suggest that interlocutors—at least in the community of Tenejapa—are "relatively gaze averse" and do not, in general, look at one another as

they talk. In this little conversation, the obvious and deliberate use of iconic, indexical, and at least apparently partly conventionalized bodily movements or "gestures" thus raises, at least in principle, a conundrum. How are such meaningful visible aspects of talk meant to be perceived and incorporated into interaction? How expressive can gestures be if people don't look at each other when they talk?[8] One of the aspects of this videotaped conversation that most caught my eye was its apparent exuberance and elaborate gestural expressivity, along with the fact that the two interlocutors subjectively at least often seemed to avoid looking at one another directly. Using admittedly crude quantitative measures, which I had applied before to Tzotzil interaction (Haviland 2017), I decided to explore somewhat more carefully how gaze and gesture worked together in A's conversation with his visitor in Oxchuk. After carefully annotating selected parts of the videotape, I applied three different measures to the material. First, as an approximate measure of each participant's spoken contribution to the conversation, I counted the numbers of spoken syllables each produced. For reasons I will explain in the next section, I divided the entire conversation into different sections, the first of which (Part 1) was roughly the first two minutes of the interaction. Figure 13 shows a simple graph of the number of syllables in each turn (measured as a stretch of speech undivided by the turn of another), where the speech of A, the old man, is graphed in light gray, and that of V the visitor in darker gray. A quick inspection of the graph will show that the turns were more or less evenly distributed between the two men, although there were "sections" where V spoke more than A, and others where the opposite was true; also that V's long turns slightly outweighed those of A. Thus, in general the visitor spoke slightly more than his older host, although on a couple of occasions A's speech was ascendent.

Second, as a way of characterizing the apparent lack of mutual (or even unilateral) gaze, I applied a different sort of gaze judgment to each utterance. Here the notion of 'utterance' is slightly more delicate than that of the (somewhat mechanical) notion of 'turn' used in the previous graph. 'Turns' were there used as a crude measure of raw speech quantity; each 'turn' represents a stretch of uninterrupted talk by a single speaker, ended only when his interlocutor instead took the floor. On the other hand, for the purpose of estimating gaze I have defined 'utterances' using intonational and syntactic criteria to divide the speech

---

8. The conundrum has pushed a theoretical divide in gesture studies, put crudely, between those theorists who imagine that gestures are "designed" primarily for speakers or for hearers, an issue I will try to sidestep with the data provided here.

Part 1, turns
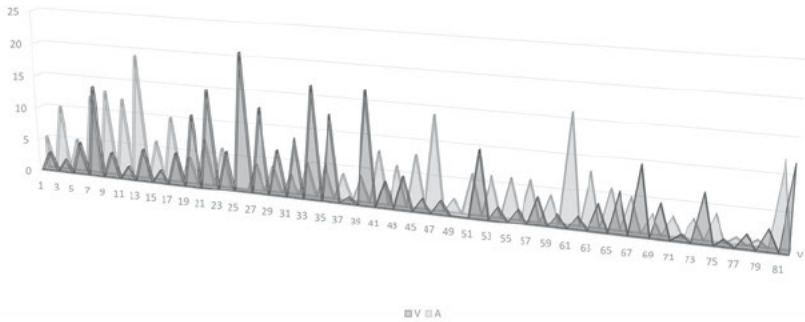


FIG 12.13 — *Syllables per turn by participant*

flow up into more natural speech units. As a result, the total number of individual 'utterances' is much larger than the number of aggregate 'turns' graphed in Figure 13. (Similarly, by this measure, V has more total 'utterances' than A.)

To illustrate how 'turns' differ from 'utterances,' consider the short sequence with which Figure 6 above begins. A standard transcript of the talk, with each line representing a single 'utterance,' rendered in Tseltal with morpheme-by-morpheme glosses and a free English translation is presented in Figure 14.[9]

In the first 4 lines of Figure 14, the visitor is speaking with no interruption from A. For the purposes of the syllable counts showing raw volume of speech, this all counts as one long 'turn,' which ends at line 5 when A speaks, overlapping the last part of V's ongoing speech at line 4. However, there is internal structure to the series of 'utterances' in V's long 'turn' from lines 1-4, and each has a slightly different pattern of mutual gaze between the two men, which is part of what Figure 6 is meant to show.

9. The abbreviations used in the rough Tseltal morpheme-by-morpheme glosses are: 1AIN 1st person inclusive Absolutive enclitic; 1As 1st person singular absolutive enclitic; 2sA 2nd person singular Absolutive enclitic; 2sSUBJ 2nd person singular Subjunctive enclitic; 3E 3rd person Ergative proclitic; ASP aspect marker; BEN benefactive suffix; CL (en)clitic, DIR directional particle; EVID evidential particle; EXIST substantive of existence; IMP 2nd person imperative; INT interrogative particle; NEG negative prefix; SUBJ subjunctive affix. The open square bracket ([) marks the beginning of overlap between two adjoining lines.

1. V: ja`     to          y-al-b-at          we'-an
   EMPH   CL(still)   3E-say-BEN-2sA    eat-2sSUBJ
   *Only then will they say to you, 'Eat!'*

2.    we'-a   me   lek   waj,   uch'-a   me   lek   ch'in   kafé
      eat-IMP   CL   well   tortilla   drink-IMP   CL   well   little   coffee
      *"Eat your fill of tortillas, drink you little coffee."*

3.    noj-on   be'el   xal,   x-k'o-on   lok'el   ts'in   at'el-e
      full-1As   DIR   CL   ASP-go-1As   DIR   CL   work-CL
      *I'll be satisfied when I leave, I will go out for work.*

4.    ay          to        waj      xal        ts'in
      EXIST   CL   tortilla   CL   CL
      *If there are still tortillas left.*
                     [
5. A:              we'-k-otik        xan
                   eat-SUBJ-1AIN   CL
                   **We can still eat.**

6. V: mok    m-ay-uk-
      INT    NEG-EXIST-SUBJ
      *But if there aren't-*

7. A: ja'     to-     ix      pajel     a       bi
      EMPH   Cl-CL   tomorrow   CL      EVID
      *It won't be until tomorrow, indeed, (that we will eat).*
[
8. V:                  ja'        to-      ix        pajel        aw-ay   xal   ts'in bi
                    EMPH     CL-CL   tomorrow   2E-understand   CL      CL   EVID
                    *It won't be until tomorrow, you see, indeed.*
                              [
9. A:                            ja'      to-ix      pajel       a    bi
                             EMPH   Cl-CL   tomorrow   CL   EVID
                             *It won't be until tomorrow, indeed.*

FIG 12.14 — *A standard written transcript of the small section of talk with which Figure 6 (above) begins.*

For each utterance four possible gaze values were possible: if both speaker and hearer appeared to be gazing at each other mutually during the course of the utterance, it was scored 3. If the speaker was gazing at his interlocutor, it was scored 2; if only the hearer was gazing at the speaker, it was scored 1; and if neither man was gazing at the other, the score was 0.

By these criteria the utterance shown in line 1 of Figure 14 would receive the gaze value 2, since at the start of that utterance the visitor trained his gaze on the old man A, at least briefly, although A was looking into empty space. During the utterance in line 2, the two men looked directly at each other for almost the entire time the visitor was talking, yielding a mutual gaze value of 3. After a brief pause in which neither man gazed at the other, V uttered line 3, illustrating his full belly and his departure from the hypothetical scene with double hand gestures. The visitor looked at A, but A looked elsewhere (partly at the younger man's gesturing hands)—a score of 2. Lines 4 and 6 in Figure 14 represent what I recorded as a single utterance on intonational and syntactic grounds: the visitor said, without pause, "*if* there are still tortillas, because if not-." He cut off his full thought, apparently because he was overlapped by A in line 5, who said "[So] we can eat." For V's utterance (shown as lines 4 and 6), then, the gaze score was 2: he had lost A's gaze but continued to look at his interlocutor. The old man A had dropped his gaze from the visitor, so the score for his own utterance at line 5
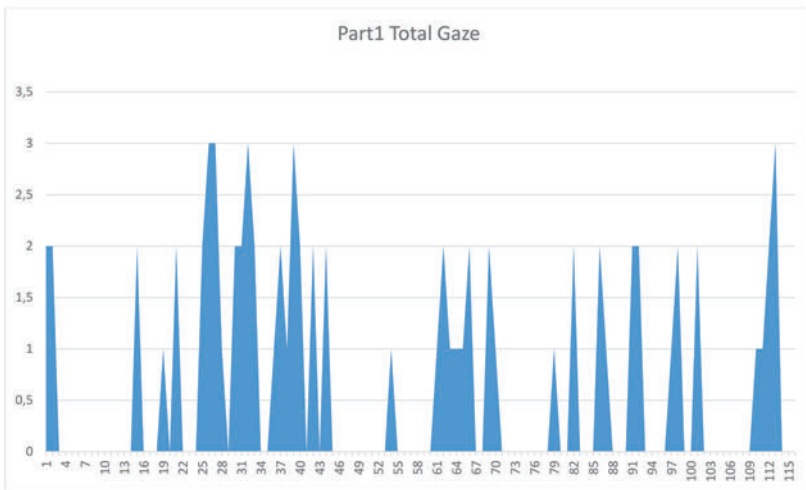


FIG 12.15 — *Mutual and partial gaze in Part 1*

was 1 since his interlocutor's eyes were still on him. A did not gaze at the visitor for the rest of the short sequence shown in Figure 14. At lines 7 and 9 he repetitively seemed to finish V's cut-off thought (from lines 4 and 6) for him: [if there are NO tortillas left] then "we won't eat until tomorrow." Those utterances accordingly receive scores of 1, since only the interlocutor was gazing at the speaker. Finally, V's overlapping partial echo of the older man's conclusion (at line 8) is scored as 2, since V was then staring directly at A, who did not reciprocate.

If one considers the entire initial 2-minute sequence (Part 1), the results of this admittedly crude gaze measure are shown in Figure 15. (The short example sequence we have just examined in detail is represented on the graph at lines 37-41.) The graph seems to confirm that there was relatively *very little* mutual gaze in this initial part of the conversation—only about five utterances in the whole section seemed to achieve at least fleeting mutual visual attention. Although there is a slight tendency for speakers to look more at their interlocutors than vice versa (which might be another way of saying that these two men frequently avoid looking at their partner when he is speaking), generally the preference seems for these two interactants *not* to look at one another when they talk—a confirmation of at least one sort of "gaze aversion." There also seems to be an asymmetry between these two individual speakers. Comparing utterances where A is speaking (Figure 16) versus those where V is speaking (Figure 17), it appears that the older man A looks *less* at his younger interlocutor when speaking than the younger man looks at him while speaking, especially as Part 1 of the conversation proceeds. From such a rough and ready sample, it is hard to justify more than mere speculation about, for example, a possible relationship between status, age, and respect and the willingness of interlocutors to engage in mutual gaze in these circumstances—at which Rossano *et al.*'s 2009 study also hints, as mentioned above.

Somewhat more telling is the relationship between mutual gaze and these two men's production of "gestures," taken as visible bodily actions apparently designed directly for the composite utterances they help constitute. I have selected another crude measure (all based on my own subjective judgements) of such co-expressivity as applied to motions almost exclusively of the hands. 0 means no such motion; 0.5 means a motion apparently rhythmically tied to accompanying speech (what McNeill [1992] dubbed a "beat"); 2 means an indexical gesture toward an apparent referent or direction; 2.5 means such an indexical element that seems also to have iconic properties; 3 means a clearly depictive or iconic motion, and 4 means an apparently conventionalized gestural

emblem. The scale is meant to approximate how easily one might argue that the "gesture" was intended to contribute something substantive to the reading of the whole utterance of which it is part. The resulting graphs are most revealing, for present purposes, when paired with the gaze diagrams in Figure 16 and Figure 17. There are two: one corresponding to visible aspects of A's utterances (Figure 18), and the other to those of V's utterances (Figure 19), and I have arrayed them against
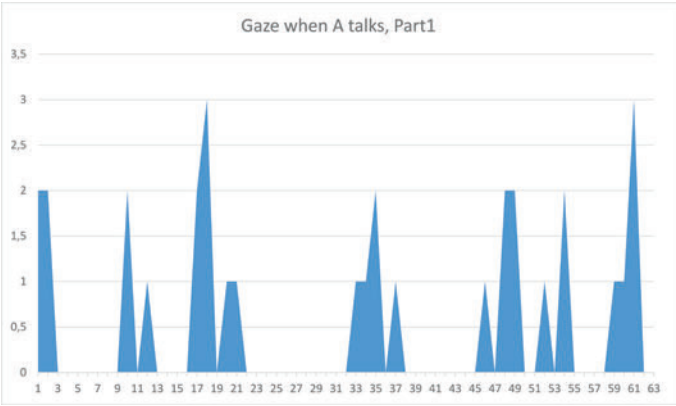


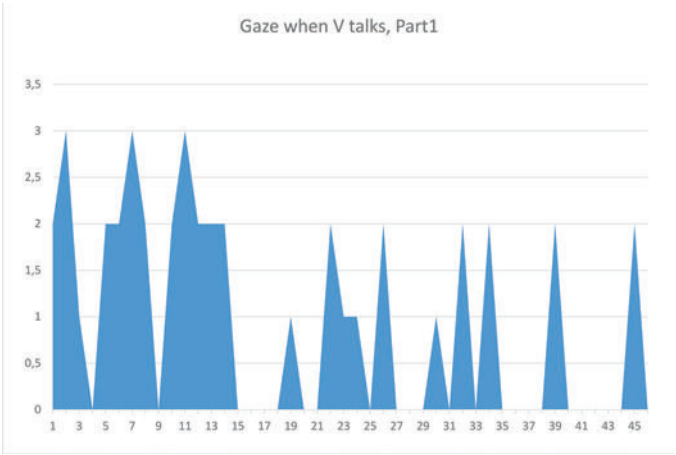FIG 12.16 — *The older man A's gaze*
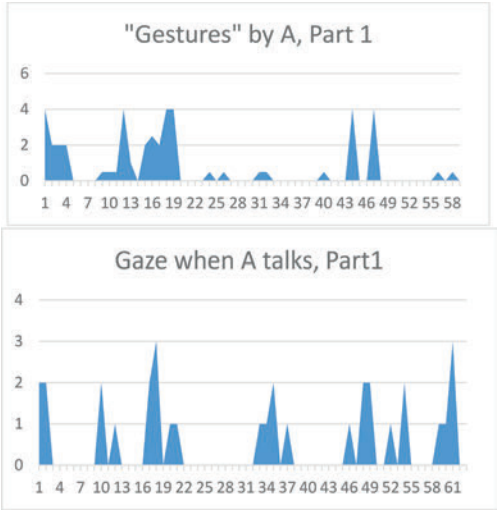


FIG 12.17 — *The younger V's gaze*

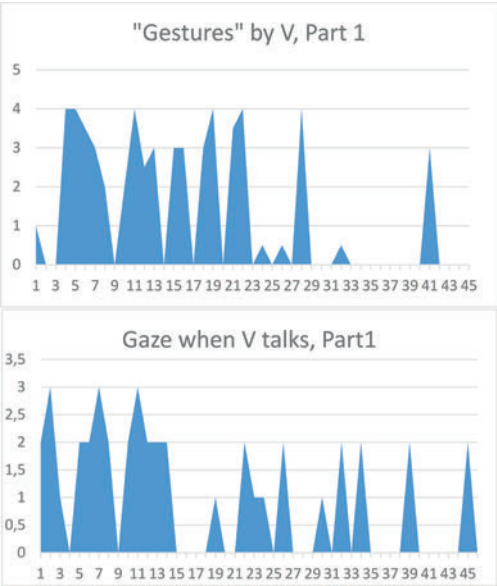FIG 12.18 — *A's apparent gestures (compared to gaze at him by V)*



FIG 12.19 — *V's apparent gestures (as arrayed against A's gaze at him)*

smaller versions of the previous gaze tables to facilitate comparison. The results are less conclusive in the case of the older man, who gestured less than his visitor. But there is still a suggestion that—gaze aversive or not—demonstrative use of gesture does attract at least some visual attention, although that cannot be the whole story.

Thus for example the first prominent points at which the more gesturally voluble younger speaker V attracts the otherwise reluctant gaze of the older man A are those when V's gestures seem to help characterize what he is describing, as when his performance of having food placed before him, feeling full, and leaving for work in the scene we have already met (Figure 5) are depicted by a sequence of two handed gestures. More dramatically, even when A has been steadfastly looking away from his visitor, he turns to look directly at him (Figure 9b) when the younger man gestures "nothing to eat" by slapping his opposing palms together in a conventionalized "nothing left, all gone" gesture transcribed in Figure 10. (These turns appear at lines 7-8 and 11 in the following two graphs.)

A further glance at the configuration of gaze between the two men shows that although the older man does occasionally look (or half look) at the younger man's expressive gestures (e.g., in Figure 9), often he entirely looks away (e.g., in Figure 11 or Figure 12), even when the visitor's entire conversational contribution seems to consist of only a gesture (e.g. in Figure 7).

*Phases of interaction and meta-conversation*

Let me now turn to the last topic that this fortuitous Tseltal conversation brought to mind. It takes me in an unexpected direction, that nonetheless grows naturally out of this tiny encounter and its cultural ecology. I characterized the first couple of minutes of this conversation as belonging to a genre of "smalltalk," an at least partially phatic necessity in many contexts in highland Mayan communities in Chiapas as well, of course, as in many other places around the world. In Zinacantán, one never just "drops in" to visit someone. There always must be a "reason" for such a visit—whether to borrow a tool, ask for money, someone's services, or even a spouse—and sooner or later participants know that a request will be coming. However, one never, as it were, gets "straight to the point." "Smalltalk," that is, is not *simply* "phatic" but always instrumental, designed to break at least some sort of ice before getting down to the true business of one's errand.

FIG 12.20 — *Two yawns*

Figure 20 shows two illustrations drawn from the video, about 30 seconds apart, at the very end of the "smalltalk" section of this interaction. The drawings suggest that even gaze averse interlocutors should be able to tell, with or without looking directly at each other, and without explicit "gesture," that whatever the visitor's business is with A, he had better get on with it.

Indeed, as it happens, within about two seconds after the moment the second frame illustrates, V remarked, "I have come to see you." For almost the first time since he arrived, V suddenly had A's full attention. The older man whirled around to fix him with a stare (Figure 21). Clearly speaking of a presupposable mutually known 3rd person, A asks, "What did he say? Why didn't you come tell me?" The central land issue that V had come to discuss then began.

I will not dwell on the main body of the conversation itself. Suffice it to say that it lasted for more than an hour, and at some point A excused himself to consult in private with his wife about how to resolve various issues, returning after a pause to reach a final agreement. Some of the



FIG 12.21 — *The main business of the visit begins*

21

FIG 12.22 — *"I came for what we agreed on."*
*"The conversation will become long."*

measures I have used for parsing out segments of the initial "phatic" talk could diagnose different phases of the long interaction that follows. (For example, once talk shifted to the land issue itself, individual turns were much less equally distributed, and each speaker tended to dominate talk for longer periods of time. [See Haviland 2017]) Instead, let me conclude this essay with a short, striking fragment of what I am tempted to label "meta-conversation," that is, conversation about conversation. It relates directly to the preliminary phases of this interaction, and to what one might understand as a cultural ethic about ordinary conversation in these highland Maya communities.

Several minutes into his main narrative, V described a visit he had made to a brother-in-law to raise the land related issue at hand. He arrived in the other village, after a lengthy journey, to discover that a different event—a fiesta—involving his brother-in-law was already in progress. Although he had made the journey specifically to discuss his land issue with the other man, he never could find an appropriate opportunity to bring it up. Nor could he simply leave politely, whatever his more pressing business might have been. Thus, one of his excuses for not coming to visit A earlier was the resulting delay in resolving this related matter with the brother-in-law. "The conversation took a long time," said V. "That's how it is."

Having heard this account, A himself offered his own elaborate commentary, portraying himself in V's shoes talking to the brother-in-law with multiply embedded hypothetical "reported speech." "He's not going to mention it quickly, nor will I say 'I came for what we agreed upon'" he said (Figure 22a). One of the ways one signals "reported speech" in Tzotzil and Tseltal is to "re-enact" it. That is, one minimally revoices someone else's (at least imagined) words, as here when A used
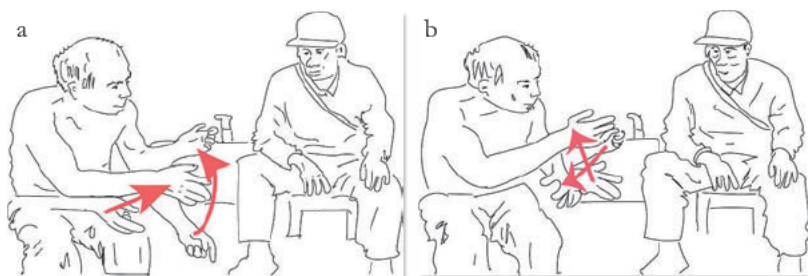
FIG 12.23 — *"I came for this"; then take off.*

the pronoun 'I' to put his own words into his co-present interlocutor's virtual mouth. But one also can re-depict or re-frame one's actions as those of another, by altering one's posture, heightening or exaggerating one's gestures, or even conjuring a scene markedly different from the present space: for example, by gazing "at" (and thus bringing onstage) different—here imagined—interlocutors. A did all of these things in Figure 22a as he glanced up (apparently at the imagined brother-in-law) and gestured energetically at him in the virtually created interactional space between them, thereby filling it with an invented (and inappropriately heightened) sense of urgency. V agreed, spatially but less demonstratively, "The conversation will take a long time" (Figure 22b).

Even more animatedly, A elaborated: "I'm not going to say, 'This is what I came for' (Figure 23a) and then just turn around and go home (Figure 23b)." He stuck his hands forward into the space separating the two depicted interactants, and waved them rapidly up and down, have turned directly to face his virtual interlocutor to demonstrate the behavior of a hypothetical, pushy, and intrusive speaker. Then (in Figure 23b), he performed another version of the miniature hand-slapping "take off" or "finished" gesture, which we have already met (Figure 9b), to caricature someone lacking sufficient social grace simply to wait for a proper moment to raise a serious (but in the circumstances impertinent) topic. His co-expressive gestural comportment was thus—via its exaggerated and almost caricatured parodic excess— itself meta-co-expressive,[10] a gestural comment on appropriate gestural co-expressivity as part of an extended and highly interactive conversational meditation on the topic.

10. A phrasing kindly suggested by Valentina Vapnarsky's perceptive reading of the example.

FIG I 2.24 — *One should come back another day*

The little meta-conversational sequence ended with both men agreeing (Figure 24) that business is not conducted in such an impatient, hasty way, and that in such circumstances one should instead excuse oneself for intruding with extraneous business and come back another day.

### Coming to an end

Tzotzil speakers from Zinacantán, not far from Oxchuk where A and V's conversation took place, often treat talk as a metonym for sociality in general. When an unreasonably troublesome person is bested in a dispute, for example, one can crow that one's adversary "*mu xa xtak'av*, couldn't answer back"—i.e., was reduced to vanquished silence (Haviland 2010). I was first struck, in this Tseltal conversation, that speakers of a language sometimes characterized as exhibiting little multimodal exuberance, and especially in this somewhat stiff and business-like circumstance, would engage each other in noticeably animated and gesturally enthusiastic interaction. Why, I wondered, would "gaze aversive" and corporeally restrained interlocutors motivate themselves to gesture at all, let alone weave movements of their bodies into the fabric of their interaction?

Even the first parts of Tseltal interaction captured on Mtro. Ochoa's film leave no doubt that—despite apparently strict regimentation of posture, position, and social distance—gesture and gaze are integral to the course of the talk, albeit in a quite local idiom. In the opening section of the interaction, when both interlocutors are feeling each other

24

out, exchanging preliminary pleasantries and platitudes, and waiting for an appropriate moment to get down to the main business that inevitably must follow, tracking attention, expressive bodily movement, posture, and talk reveals mutual dynamic synchronization and patterning. Even in the course of developing otherwise stylized, phatic "smalltalk" topics, the interlocutors coordinate with one another, both corporeally and thematically, and spoken utterances alternate with, or are replaced or refigured by kinesic elements.

In part legislated by my own incompetence in spoken Tseltal, I also applied a series of rough and ready quantitative measures to the introductory section of the two men's talk to gain a sense of the somewhat contradictory conjuncture of their supposed "gaze aversiveness" with the evident substantive integration between their talk and their gestured movements. A potential suggestion from this simple enquiry is that "gaze aversiveness" in Tseltal conversation is possibly as much a matter of courtesy or status as a general cultural style: perhaps older men avoid looking directly at their junior interlocutors; or vice versa, perhaps older, more powerful people command more visual attention than younger ones.

Finally, these two Oxchuk interlocutors indirectly suggest a useful moral about conversational form and an explicit theory of interactional ethics. Just as the men dutifully observe, in their own encounter, the convention that one allows conversational smalltalk to take its own sweet time before launching into matters of serious import—and despite evidence, as well, that in this case A was poised to pounce on the moment when his visitor shows signs of coming, finally, to the point (at Figure 21 above)—so, too, can such meta-conversational discourse be explicit. It can, moreover, itself be coexpressive, as is evident in the last examples from the second part of the Oxchuk film, where co-expressive parodic conversational gestures illustrate iconically some of the features of politeness that would be ruptured if one ignored proper standards of interaction and insisted overmuch on pushing one's conversational agenda in inappropriate ways.

I beg Aurore's indulgence for venturing into a language I do not properly command to explore the tiny vignette of linguistic and interactional mastery on display in both V and A's filmed performances. There is an advantage, perhaps, in trying to penetrate interaction without the benefit of detailed linguistic knowledge—although more of that could undoubtedly have helped me here—if one tries, as I have in this short essay, to convey an impression of the simultaneous co-expressivity of a series of different semiotic resources—words, to be sure, but also

25

bodies in action, eyes and faces, movements, positions, and even the socio-physical spaces they jointly constitute—that unavoidably figure in human interaction, from Chiapas to Paris and everywhere between and beyond.

## References

BROWN, Penelope

1998    Conversational structure and language acquisition: the role of repetition in Tzeltal adult and child speech, *Journal of Linguistic Anthropology,* 8, pp. 197-221.

2006    A sketch of the grammar of space in Tzeltal, *in* Stephen C. Levinson and David P. Wilkins (ed.), *Grammars of Space* (Cambridge University Press), pp. 230-272.

2011    The cultural organization of attention, *in* Alessandro Duranti, Elinor Ochs, and Bambi B. Schieffelin (ed.), *The Handbook of language socialization* (Amsterdam, John Benjamins B.V.), pp. 29-55.

BROWN, Penelope and Stephen C. LEVINSON

1993    'Uphill' and 'Downhill' in Tzeltal, *Journal of Linguistic Anthropology*, 3(1), pp. 46-74.

2000    Frames of spatial reference and their acquisition in Tenejapan Tzeltal, *in* Larry Nucci, Geoffrey B. Saxe, and Elliot Turiel (ed.), *Culture, Thought, and Development* (Mahwah, NJ, Lawrence Erlbaum), pp. 167-197.

ELAN (Version 6.4) [Computer software].

(2022)    Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan.

ENFIELD, Nicholas J.

2009    *The anatomy of meaning: Speech, gesture, and composite utterances* [No. 8] (Cambridge, Cambridge University Press).

GOODWIN, Charles

1981    *Conversational organization: Interaction between speakers and hearers* (New York, Academic Press).

HAVILAND, John. B.

1967    The modern troubador: streetsinging in Europe on no dollars a day, *in* John Wurster and John Andersen (ed.), *Let's Go II: The Student Guide to Adventure* (Cambridge, MA, Harvard Student Agencies, Inc), pp. 51-55.

2010    Mu xa xtak'av: "He doesn't answer", *Journal of Linguistic Anthropology*, 20(1), pp. 195-213

2017    Mayan Conversation and Interaction, *in* Judith Aissen, Nora C. England, and Roberto Zavala (ed.), *The Mayan Languages* (Oxford, Routledge) pp. 401-432.

KENDON, Adam
  2004  *Gesture: Visible action as utterance* (Cambridge, Cambridge University Press).

MALINOWSKI, Bronislaw
  [1923]  1936 The problem of meaning in primitive languages, *in* Charles Kay Ogden and Ivor Armstrong Richards, *The meaning of meaning* [supplement I] (London, Kegan Paul, Trench, Trübner & Co.), pp. 296-336.

MCNEILL, David
  1992  *Hand and mind: What gestures reveal about thought* (Chicago, University of Chicago Press).

MONDADA, Lorenza
  2016  Conventions for multimodal transcription. https://franz.unibas.ch/fileadmin/franz/userupload/redaktion/Mondadaconvmultimodality.pdf

POLIAN, Gilles
  2013  *Gramática del tseltal de Oxchuc* [tomo 1 y 2], (México, CIESAS).

ROSSANO, Federico, Penelope BROWN, and Stephen C. LEVINSON
  2009  Gaze and questioning and culture, *in* J. Sidnell (ed.), *Conversation Analysis: Comparative perspectives* (Cambridge, Cambridge University Press), pp. 187-249.

SLOETJES, H. and WITTENBURG, P.
  2008  Annotation by category - ELAN and ISO DCR, in *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008).