Representation of the interlocutor’s mind during conversation

Marjorie Barker and T. Givón

Linguistics Department

and

Institute of Cognitive and Decision Sciences

University of Oregon
Representation of the interlocutor’s mind during conversation

1. Background

During natural face-to-face communication, people engage simultaneously in a great number of social, communicative, and cognitive tasks. Among these, the two most prominent ones are the interaction between the interlocutors and the management of the information flow. In other words, people engaged in conversation are simultaneously tracking both the social collaboration that allows their interchange of ideas, and the gradual development of a body of information as it flows back and forth between them. It seems obvious that the two tasks cannot be totally divorced from each other, since the interpersonal dynamics serve as the matrix within which the information flow takes place. This matrix involves not only the situational aspects of the interaction – speaker and hearer, time and place – but also the current states of knowledge and intention in the mind of each interlocutor at any given moment (Givón, 2001b; 2005; Grice, 1968/1975).

In both these aspects of human communication, grammar is used systematically as a context-cuing, perspective-shifting device (Givón, 2001b; 2002; 2005). Grammar indexes the relative importance of conversational referents, places them in space and time, and conveys the speaker’s opinion about the reliability and desirability of the information being conveyed, all of which change continually as conversation proceeds. As speakers use grammar to communicate, they rely, constantly and systematically, on their own constantly shifting mental models of the hearer’s belief and intention states. At the same time, they manipulate the hearer’s construction of their constantly shifting mental representations.

There is ample evidence in natural language data that tracking the listener’s physical and social position, as well as the listener’s perspective on information being communicated, is a
prime concern of speakers; entire portions of the grammar of any language are devoted to these things (Givón, 2001a). The features that demonstrate this concern in language are ubiquitous, but subtle. One such field is *deixis* – referential (*I, you, this, that*), spatial (*here, there*), temporal (*now, then, today, yesterday*), and social (*Mr. Smith vs. John vs. honey*) identification of the current speech situation. Such reference shifts constantly with the shift of conversational turns and speech situation: *I* to one speaker does not have the same referent as *I* to the other speaker; *this* and *that* may identify the same referent depending on the perspective taken.

The existence of mental models of the interlocutor’s knowledge is also the only explanation for the linguistic phenomenon of definiteness: the choice between ‘a woman’ and ‘the woman’ is based on the speaker’s perception of a difference in the *listener’s* mind. If I think my listener will be able to identify the person I am talking about, I use a definite article: ‘the woman;’ otherwise, I will bring *the same referent* into the conversation as ‘a woman.’ To argue that, for example, speakers use ‘a’ when first introducing a word and ‘the’ subsequently, at no point considering the hearer’s familiarity with the term, does not explain why some referents are introduced initially as known.

An equally conspicuous example of implicit mental models of the interlocutor’s epistemic and intentional states involves the grammar of speech-acts in ordinary conversation (Givón, 2001b). To use an interrogative appropriately, for example, the speaker must believe that the hearer might know the answer, and be willing to share the answer, or at least speculate about it. The appropriate use of a declarative involves the speaker’s belief that the hearer may not be aware of what the speaker is about to say and would welcome knowing it. Indeed, to initiate conversation at all, speakers make judgments about the other person’s willingness and ability to participate in the desired interaction. These mental models about the hearer’s presumed belief
and intention states must shift constantly as conversation proceeds, as new information is exchanged. And the most straightforward way to account for the felicitous use of the grammar of speech-acts during conversation is to assume the existence of such mental models.

Constantly shifting, and largely implicit, mental models of the interlocutor’s current belief and intention states are characteristic of both conversational and narrative discourse. Furthermore, the speaker’s shifting perspective about the hearer’s mental representation does not depend on shifting conversational turns. The status of referents within a single turn changes as the speaker continually revises her estimation of what the hearer knows. As a simple illustration, consider one speaker’s use of the grammar of referent tracking over the course of one conversational turn:

Vicki: OK. And after that she made a fire … umm she picked up a little, bad old tin pail, and she went over next to the lean-to, the white bucket of water, and she poured the water into the pail and then took the pail back and put it on the fire.

The use of the indefinite noun phrases ‘a fire,’ ‘a …pail,’ and ‘water’ presupposes that the hearer has no prior episodic representation of these three referents. But the subsequent use of ‘the water,’ ‘the pail,’ and ‘the fire’ assumes that the hearer now has such mental representation.

The very act of speaking assumes an ability on the part of the hearer to develop a mental model of the speaker’s perspective – and this assumption is, in turn, a theory about the listener’s mind. The grammatical features described above, and others like them, serve as linguistic evidence that modeling the interlocutor’s mind is important to speakers and hearers. The prevalence of these indicators suggests that it should be possible to elicit empirically some acknowledgement by conversational participants of the mental modeling taking place during conversation.
Though the two main aspects of communication – interaction and information – cannot be totally separated, their study has generally been split into two methodological paradigms. The study of the information flow of communication has focused almost exclusively on either the produced text (in linguistics; e.g. Chafe, 1994; DuBois, 1987; Givón, 1994; Tomlin, 1991) or on text comprehension (in psychology; e.g. Anderson, Garrod, & Sanford, 1983; Ericsson & Kintsch, 1995; Gernsbacher, 1990; Trabasso, Suh, & Payton, 1995). The study of conversational interaction, on the other hand, began with a focus on the social aspect of the interaction, primarily the turn-taking system (e.g. Goodwin, 1988; Sachs, Schegloff, & Jefferson, 1974). This methodological cleavage has persisted in spite of numerous studies in both the discourse and acquisitional literature that suggest a strong interaction between the interactional and information-flow aspects of communication (e.g., Chafe, 1997; Coates, 1997; Ervin-Tripp & Kuntay, 1997; Goodwin, 1995: Wilkes-Gibbs & Clark, 1992).

An absolute separation between the social and informational study of natural communication is theoretically indefensible and methodologically unwise. The information flow and the speech situation are inextricably bound together. What is more, both are equally ‘cognitive;’ what is relevant in both is not some ‘objective’ reality but rather the conversational participants’ mental representation of whatever that reality may be. Any objective reality, be it situational or textual, that does not attain some type of mental representation is irrelevant to the process of communication. And the very same mental representation systems must be involved in the processing of both aspects of communication, since the task of language processing is at heart a matter of interpreting the mind and intention of the interlocutor.

The experiment described here was an attempt to tease out people’s mental representation of their interlocutor’s perspective during conversational interaction. In this way, the study probes
the representation in episodic memory of the interrelated factors we suspect to be equally salient during conversation: both the communicated contents and the communicative interaction. In order to keep the informational content of the conversations fairly constant, we asked participants to discuss a video they both had just finished watching. We predicted that asking participants to find differences in what they had just viewed would cause speakers to pay closer attention to both the information provided by their partner and to their partner’s mental attitude toward this information.

2. Methodology

2.1 Material and Participants

Participants \( (N = 14 \) undergraduate students) watched a 6.25-minute videotape, ‘The Chicken Story,’ developed and used for two other experimental projects (Dickinson & Givón, 1997; Givón, 1991). The film shows a man and a woman, working separately and interacting, in a rural setting. The video was filmed in Eugene, Oregon and the actors speak Swahili. There are no subtitles, allowing viewers to develop their own interpretations of the content of the actors’ interaction. Fifty-eight states/events in the video were established by the earlier experiments as baseline clauses for recall, because they were mentioned by at least seven out of ten participants who described the action aloud while watching it.

Participants watched the video at the same time, but separated from one another, on individual monitors. Prior to viewing, they were told to pay close attention and memorize the story. After watching the film, they were brought together and told (incorrectly) that the films they had just seen were similar in broad outline but not completely identical. They were then asked to talk with each other about the film in order to find out as much as they could about the video the other person had seen. These conversations about the film were recorded. Participants
were then debriefed individually and asked to recall in as much detail as possible the conversation they had just had. These recall monologues were also recorded.

2.2 Analysis of transcripts

The experiment produced two sets of transcripts, one set of seven dyadic conversations and the other of fourteen individual recalls of the conversations.

The conversations were not the prime target of this investigation; our interest was in the participants’ mental representation of the interaction: could they report on the nature of the interaction as well as the information involved? For this reason, the focus of the analyses was to compare the recall transcripts to the original conversations and assess participants’ accuracy in their attribution of who said what, how confident they were in saying it, and differences of opinion – in other words, their accuracy of understanding the other person’s perspective. For this reason, the recall transcripts were more extensively analyzed than the conversation transcripts.

Conversation transcripts were analyzed to assess the number of baseline events mentioned by each pair during the course of conversation and to determine who mentioned each event, data that served as criteria for the recall analysis.

The text of individual recall transcripts was divided into (a) recall of the contents of the video and (b) recall referring to the dynamics of the conversation itself. The following brief passage from a representative recall transcript illustrates this separation. Portions of the text referring to the conversational interaction are boldfaced and those referring to the video are in plain type.

… the conversation I had with Vicky. First of all she started out by saying, *saying* what the man was wearing, *and that* he was carrying three farming utensils plus a hatchet, *and, actually, no, she started out asking me what* he
was wearing. **And I said that** he was wearing red shorts and a white T-shirt and no thongs or anything. **And she said that** he was wearing red shorts and a white T-shirt, but **that** he had flip-flops on.

The proportion of clauses in each recall transcript referring to video contents vs. to conversational interaction was established. Recall transcripts were also analyzed to determine the number of video baseline events mentioned in the subjects’ recollection of the conversation. In addition, these transcripts were compared, for each pair, to the conversation they were attempting to recall. We computed accuracy of recall for (1) which speaker produced a particular clause in the original conversation and (2) the speech-act modality used, divided broadly into ‘certain’ vs. ‘uncertain.’

Uncertain modality found in the conversation transcripts was marked in five different ways: as interrogative speech-acts (*what was the guy wearing?*); nonfactive or negative cognition verbs (*I think* mine had two; *I don’t remember* seeing her fill it); nonfactive or negative perception verbs (*it looked like* cheese; *I didn’t see* her put anything in the water); nonspecific reference (*whatever* she had wrapped up, in the towel, she put that down; and brings back *some, um, something*); and uncertain quantifiers or adverbs (*there’re maybe* seven or eight of them; *she kinda* looks disgusted).

3. Results

3.1. Conversations

During conversation, subjects collaboratively mentioned on average 37 (64%) of the 58 baseline events of the video. Averaging over all pairs, each partner produced about a third of the baseline clauses mentioned in conversation, with the remaining third spoken by both partners. There is wide variation, though, in the proportion of baseline material handled by any individual
speaker, ranging from 13% to 55% (SD = 13.2%). Interestingly, and perhaps predictably, the first speaker of each pair tends to handle significantly more of the informational content of conversation ($M = 39\%$) than the second speaker ($M = 27\%$); $t=3.75, df=12, p < .01$. The difference between male and female speakers was not significant.

3.2 Recall

In recalling the conversations they just participated in, subjects produced speech that referred to the collaborative aspect of their discussion as well as speech recounting the information they had exchanged. On average, about one fourth of recall clauses referred to the circumstances of interaction. The three most common features of interaction mentioned in the recall transcripts were identification of the speaker (*she* said, *I* said) epistemic/modal qualification of the recalled information (‘*I think* that she said,’ ‘*and* maybe,’ ‘*we couldn’t remember*’), and identification of the speech-act or modality used in the conversation (‘*she started by asking,*’ ‘*we kind of disagreed,*’ ‘*I thought I saw*’). We explored the extent and accuracy of each of these features.

Interestingly, participants had no apparent trouble attributing every chunk of remembered content of their conversation to either themselves or their interlocutor. Our initial assumption was that this attribution need not be accurate, because the two people saw the same video and agreed on most of the details of what they had seen. Their conversation was thus extremely collaborative, and it seemed there would be no point in mentally representing two separate points of view. Nonetheless, speakers were 86% accurate overall in remembering who said what in the original conversation. (Only two speakers had a significant amount of the hedged attribution ‘*we*’; these instances were counted as neither accurate nor inaccurate.) This result indicates that
speakers have an interest in distinguishing speaker identity irrespective of agreement about the conversational content.

Speakers were also surprisingly certain in recalling the speech-acts that occurred in the course of their conversation. Epistemic and modal markers of uncertainty in connection with conversational recall were rare. There were 557 instances of quotative verbs in the recall transcripts (‘I said,’ ‘she agreed,’ ‘he asked,’); of these, only 56 (10%) were explicitly qualified by some lower-certainty marking (‘I think he said,’ ‘I don’t remember if we talked about that,’ ‘she might have asked’). Thus, regardless of their actual memory accuracy, subjects were – at least if one judges by their overt verbal use of modality – overall quite confident in their recall.

It is more difficult to assess the accuracy of subjects’ recall of the specific speech-act modality used in their original conversation, for several reasons. First, there are virtually no quotative speech-act verbs in the conversations themselves. People simply do not say things like, “I’m asking you what he’s wearing” or “I say there was some kind of noise in the background.” Rather, they say, “What was he wearing?” and “There was some kind of noise in the background.” Also, multiple quotative expressions in the recall texts – say, talk, mention, describe, tell, remember, comment, recall – may be used to refer to the very same declarative speech-act of saying. And while the predominant quotative verb in the recall transcripts is ‘say’ (281/557 = 50%), there is no principled way of ruling ‘say’ as the ‘correct’ quotative usage or preferring it to any of the other verbs used.

It is obviously problematic to determine the accuracy of conversationalists’ recall for speech-act modality if we need to make judgments about whether the original speech-act should be considered saying rather than talking, commenting, or any of the other expressions used by speakers to recall conversational speech. One possible line of analysis to would be to lump
together all the declarative speech-acts in the conversations, contrast them with the interrogative ones, and then compare all the uses of ‘ask’ in the recall transcripts to see if they match the use of interrogatives in the respective conversations. But to separate out only interrogatives would reduce our sample to less than 5% of the total use of quotative verbs (27 instances of questions out of 557 total recalled speech acts).

As an alternative, we elected to assess a somewhat rougher speech-act variable, that of the epistemic modality of speech-acts. We divided epistemic modality of conversational speech-acts into certain (reals) vs. uncertain (irreals), taking certainty to be the default case (i.e., if not specifically marked by uncertain grammar as described in §2.2, utterances were considered to be expressed as certain).

There are two kinds of errors: (a) an instance of uncertain modality in conversation that was later recalled as certain and (b) an instance of certain modality that was recalled as uncertain. There were only five exemplars of (b) in the recall transcripts, three produced by a single speaker and the other two by two participants recalling the same utterance. All five concerned interchanges where there was a difference of opinion in the original conversation as to the video content. An example is the following exchange, where G’s original statement in the conversation was certain, but later recalled as uncertain by both participants:

Conversation:

G: since she built the fire?
D: yeah …
G: then she put it out, for some reason …
D: Oh no, she put, mine, with mine she put, uh… she put water on there and had it boil
G: oh, ah …
D: boiling …

G: that’s what she’ws doing

Recall, separately by both participants:

G: I thought she put the fire out, but apparently she was boiling water to cook chicken

D: he thought that it was put out, and I thought that it was still going

The other three exemplars of certain utterances later recalled as uncertain were all produced by one participant, who differed with her partner’s interpretation of the video in each instance in the original conversation:

Conversation:                         Recall:

L: it was a bag                       V: she thought it was a bag
L: she made two cuts                  V: she made two cuts, she thought
L: she was bringing him lunch        V: she thought she was bringing him lunch

In each of the cases above, the recall that the original speaker ‘thought’ something happened is used to highlight a contrast, either between the two conversational partners’ memory of the video or between the speaker’s original interpretation and a change in that interpretation during the course of interaction. In each case, there is an identification of what the person said with what he or she thought – of speech as evidence of mental representation.

We limited further analysis to instances where the speakers expressed uncertainty. If listeners are tracking their partner’s mental state during the course of conversation, they should be paying particular attention to explicit expressions of uncertainty and might be expected to recall these as less confident.

We next matched the original conversation with transcripts of its recall by both speakers involved and determined how accurately each instance of uncertainty was recalled. Instances of
uncertainty in the conversation that were not mentioned by either partner in recall were excluded from analysis. The remaining instances were grouped in four categories, as shown in the examples below.

1. Correctly recalled as uncertain by the original speaker

   Conversation:
   
   M: and then a woman came, tall and uh … white blouse and pink … skirt
   
   D: yeah
   
   M: and … came and … seems to be yelling at him about something

   Recall:
   
   M: I had mentioned that she seemed to be yelling

2. Correctly recalled as uncertain by the original listener

   Conversation:
   
   L: she was trying … trying to, y’know, slit the chicken’s throat …
   
   V: yes, she was, the chicken was just a wild little guy …
   
   L: yeah, I don’t think she really had the heart to do it

   Recall:
   
   V: and Lori commented that she didn’t really think that the woman would do it

3. Incorrectly recalled as certain by speaker

   Conversation:
   
   L: OK of course she … she walked over to where the fire pit was
   
   V: uh-huh
   
   L: she put wood into the fire pit … and then she uh … she started the uh … fire … and she put … I thought she put a little bit of kindling in to get it going?
Recall:

L: then I said that she … she took the wood. I said went over to the fire and started it, and went to get, she, she lit it and *then she put* a little more brush on it to get it going

4. Incorrectly recalled as certain by listener

Conversation:

L: OK.

T: uh … sh… she seems to use a … she *seems to* have a little … uh big butcher knife

L: yeah

Recall:

L: he said something about the knife, she *had* this big butcher knife is what he said

The modality of uncertain speech-acts in the conversation was recalled correctly by the original speaker 94% of the time (73 out of 78 instances), and by the listener 85% of the time (64 out of 75 instances). These results confirm, albeit in a limited way, that both participants are accurately tracking the speech-act modality of conversation.

4. Discussion

We have shown that subjects recalled two features of their conversation with a high degree of accuracy – attribution of quoted information to the correct speaker (86%) and epistemic modality of the speech-act (90%). The first is somewhat surprising because the conversations we provoked were extremely cooperative and speakers had little disagreement about the contents of the video. It might be expected, therefore, that it would hardly matter who
said what, and thus this information would not be easily remembered. The second finding is surprising, too, because the distinction between certain vs. uncertain modality is often subtle; and again, it is not easy to see why it would be important in the present experimental context. That both aspects were accurately recalled is evidence that both the informational content and the interactional aspect of face-to-face communication receive consistent conscious episodic representation.

Further, we have shown that the mental representation of conversation is integrated – memory for the contents and memory for the interaction itself are closely interlaced. These findings in themselves are not too surprising and may have been predicted from either general consideration or anecdotal evidence.

What is more interesting is the striking contrast between the type of interactional information that consistently achieves conscious representation and the type that just as consistently does not. A whole range of interactional information, described earlier as necessarily activated – readily, copiously, and perhaps automatically – during face-to-face conversation, never achieves the conscious representation necessary for verbalization, and is entirely absent from the recall transcripts in this experiment. This is the information associated with the use of grammatical constructions and morphemes whose felicitous deployment is indispensable to human communication as we know it, be it face-to-face or narrative – information like deixis, definiteness, and speech-act grammar. Here again, the explicit mention of this kind of information simply does not occur in ordinary conversation – people don’t report their reasons for using the rather than a (‘I told him about a woman in the video, not the woman, because I wasn’t sure whether he had seen her) or their motivation for initiating specific interactions (‘I notice you don’t seem to be doing anything right now, and you made eye contact
with me, and I’m assuming, since we are both in the United States, that you understand English, so I’m going to try to initiate an interaction …’) Nonetheless, the consistent use in conversation of these grammatical features argues that the mental modeling in spoken interaction responsible for their application is continually taking place. The fact that it is never verbally reported by our participants in the transcripts of conversational recall suggests that it takes place automatically and unconsciously.

Here is a tentative explanation for why the vast reservoir of mental models that speakers construct about their interlocutor’s mind is implicit rather than conscious, verbal and explicit. First, grammar is an automated speech processing device, sensitive to extremely local communicative contexts (Blumstein, Milberg, & Shrier, 1982; Givón, 1979; 1989; 1992; 1994; Kintsch, 1992; Neville, Mills, & Lawson, 1992). Second, the relevant mental states – of both speaker and hearer – shift constantly, from one clause or even one word to the next. Third, preserving a longer-term – global – conscious episodic representation of such context-specific mental states is adaptively useless, since their relevance decays rapidly. And finally, preserving a long-term mental trace of these transitory, local mental states may indeed be adaptively damaging, since they would then interfere with the more relevant mental models of the hearer’s newer, current mental state.

The dichotomy between conscious (attended) and implicit (automated) information processing, and thus mental representation, has been a major theme in cognitive psychology and neuroscience for over three decades (e.g., Kahneman & Treisman, 1984; Posner & Snyder, 1974; Schneider & Shiffrin, 1977). This dichotomy has been demonstrated extensively in visual information processing (Fernández-Duque, 1999; Posner & Peterson, 1990; Treisman & DeSchepper, 1996), but also in the processing of written words (Sieroff & Posner, 1998) and in
the recall of declarative verbal information (Nissen & Bulleme, 1986). What is more, attentional activation – working memory – can itself be either conscious or implicit (Gathercole & Baddeley, 1993; Shallice, 1988).

This interpretation of our results may seem to stand in contrast to the claims of Keysar and others that the perspective of the interlocutor is not initially considered in formulating speech, and only later conscious revision permits speakers to tailor their utterances to a particular listener (e.g., Barr & Keysar, this volume; Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998), but there are at least two factors to consider in comparing the different findings. First, the contrasting studies are focused on different features of language. The phenomenon of interest in our experiment was the rapid deployment of grammar such as definiteness, deixis, modality or voice during the course of interaction. The logic of these grammatical phenomena is not transparent to language users; it is only apparent after linguistic analysis. By contrast, people in Keysar’s studies are dealing with lexical, not grammatical, items; comparing, for example, a ‘large’ and ‘small’ candle. Such items have meaning that is consciously accessible to users, and it is not surprising that their use is consciously evaluated. Second, participants in the experiments conducted by Keysar and colleagues interact in a situation that is not wholly shared. The settings for the experiments have features that are not accessible to both participants; speakers are aware that what they see is not the same as what their conversational partner sees. When people do not share context in a communicative situation, it can be expected that they will need to take more conscious effort to make their perspective available to, or understand the perspective of, their conversational partner. This phenomenon of more effortful framing due to features of unshared context is one of the features characteristic of written as opposed to oral language register.
Though the exchange is oral, Keysar and colleagues’ experimental paradigm moves participants’ linguistic production closer to the written end of the oral–written continuum. Our experiment went a certain distance toward demonstrating that some features of the conversational interaction receive systematic, reliable conscious episodic representation. But our methodology, dependent as it is on conscious verbal recall, is in principle unsuited for teasing out the speaker’s implicit, subconscious mental models of the hearer’s constantly shifting belief and intentional states. If grammar is a highly automated processing device, then the models of the interlocutor’s mind associated with each grammatical device are constructed automatically and thus remain largely implicit. As philosophers, linguists, or discourse analysts, we may argue – given suggestive data that cannot be explained otherwise – that such mental models must exist. But the modal force of that ‘must’ remains that of a hypothesis (Hanson, 1968).

Teasing out implicit mental models of whatever kind probably cannot be accomplished with conscious, verbal recall. A less direct methodology, relying on more subtle retrieval clues – such as semantic priming – may be necessary. The phenomenon we seek is as elusive as it is pervasive. The mental representation of other minds in a social, communicating species is perhaps like the air we breathe – we hardly notice it, in spite of its extreme adaptive value, because it is so hopelessly ubiquitous.
References


