

Understanding Speaker-Listener Interactions

Dirk Heylen¹

¹Department of Computer Science, University of Twente, Enschede, The Netherlands

heylen@ewi.utwente.nl

Abstract

We provide an eclectic generic framework to understand the back and forth interactions between participants in a conversation highlighting the complexity of the actions that listeners are engaged in. Communicative actions of one participant implicate the “other” in many ways. In this paper, we try to enumerate some essential relevant dimensions of this reciprocal dependence.

Index Terms: conversation, listening, backchannels

1. Introduction

In many books and papers, the process of communication is schematically depicted with a speaker who is active in the speech process and the listener who is involved in passively perceiving and understanding the speech (Fig. 1).

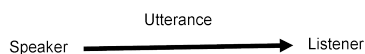


Figure 1: Picturing Conversation as an arrow

According to Bakhtin [1] linguistic notions such as “the ‘listener’ and ‘understander’ (partners of the ‘speaker’)” are fictions which produce a “distorted idea” of the process of speech communication. *The fact is that when the listener perceives and understands the meaning (the language meaning) of speech, he simultaneously takes an active, responsive attitude toward it. He either agrees or disagrees with it (completely or partially), augments it, applies it, prepares for its execution, and so on. And the listener adopts his responsive attitude for the entire duration of the process of listening and understanding, from the very beginning - sometimes literally from the speaker’s first word. [...] Any understanding is imbued with responsive and necessarily elicits it in one form or another: the listener becomes a speaker. [...] More over, Bakhtin claims, any speaker is in a sense also a respondent.*

In order to create agents that can listen to the speech of the humans they interact with, we need to have a proper understanding of what constitutes listening behaviour and how communication in general proceeds. We will introduce the major terms and concepts that are relevant for understanding what listeners do.

2. The organisation of conversational interaction

Bakhtin is not the only one who makes the point that listeners are not just passive recipients of messages emitted by a speaker. Conversation has been characterised as a collaborative activity, an interactional achievement or a joint activity by researchers such as Gumperz [2], Schegloff [3] and Clark [4]. By using the

term ‘interactional achievement’ Schegloff highlights the fact that conversations are incrementally accomplished and they involve dependency of the actions of one participant on the actions of the other and vice versa. The term joint activity is used by Clark to emphasise that it is only when the participatory actions of the different participants are seen *together* that one can talk about a conversation. In this paper, we try to point out the major aspects of this interrelation.

Communicative actions of one participant implicate the others in many ways. A typical communicative action is normally produced with the intention that one or more other participants (the addressees, the audience, the ‘listeners’) attend to them, are able to perceive them, recognize the behaviour as an instance of a communicative action, try to understand them and possibly act upon them in one way or another; preferably with the effect that the producer of the communicative action had intended to achieve. If these conditions are not met the action will fail to be ‘happy’ in Austin’s term [5] or will not be ‘felicitous’ (Searle, [6]).

The success of a communicative action thus depends on the states of mind and the behaviours of the other participants during the preparation and execution and ending of the communicative behaviours. As Schegloff and others have pointed out, the behaviours of the other participants do not only determine success but they may also influence and change the execution of the communicative actions *as they are being produced*, because the producer of the action will take notice of how the audience receives and processes the actions and also of the other reactions they invoke. A nice example is provided by Goodwin [7] who defines as a principal rule in face-to-face conversation that “When a speaker gazes at a recipient that recipient should be gazing at him. When speakers gaze at nongazing recipients, and thus locate violations of the rule, they frequently produce phrasal breaks, such as restarts and pauses, in their talk.” (Goodwin, [7, p. 230]).

We can picture the interaction between actions of the participants in conversation in a first, simple diagram (Figure 2) which is only slightly more complicated than the fictions Bakhtin was referring to but it tries to show something more of the dialogical nature of conversation.

For the sake of simplicity, assume that a conversation takes place between two persons (x and y). Given that some conversational action (CA1) is performed by one of them, (say x), as indicated by the top left corner (A) of this diagram, the other person (y) is supposed to perceive and interpret this action, as indicated by the top right corner (B). We will summarise the various actions that this involves using the term “perceive”, which is taken from the classical notion in Artificial Intelligence that an intelligent agent is involved in Perception-Decision-Action loops. This may prompt this person (y) (i.e. lead y to decide) to produce certain actions (CA2 in the bottom right corner, D). These actions in turn can communicate something to the pro-

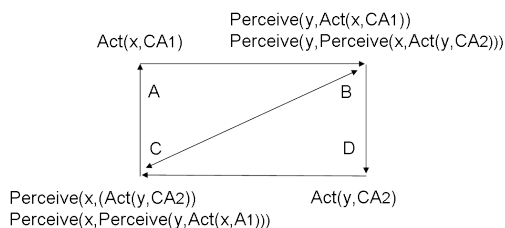


Figure 2: Picturing Conversation as an Interactional Achievement

ducer of CA1 (x) about the reception and up-take of the production of CA1 by y (bottom left corner, C) which may either change the execution of action CA1 or prompt a new action. The behaviours that make up the act of perception of CA1 by y (B) may themselves be observable to x who is monitoring them, hence the arrow connecting corner B with C. Vice versa, the actions that go into the perception of CA2 by x may also be observable to y . Of course, mutual gaze, is the typical instance enabling this connection. Actions by one thus elicit actions by the other in reply.

So far, only general terms such as ‘communicative action’, ‘producer’ and ‘recipient’ and ‘perceiver’ were used because any action could enter these perception-action loops. Therefore, also the time scale was left unspecified. The diagram can be instantiated in many different ways and the rest of this paper is dedicated to articulate the most prominent of these. For instance, the communicative action CA1 by x could be the utterance of a statement, which makes x a *speaker* during which y , the *listener*, attending to the speech, shows a puzzled face (CA2) accompanied by a vocalisation “oh” with a rising intonation. This verbal and nonverbal *feedback* in the *backchannel*, which is monitored by the speaker x may prompt x to enter into reformulation mode or to speak up. All of this can happen almost instantaneously.

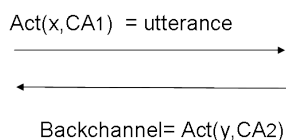


Figure 3: Simultaneous Elicitation - Response

At any given time, there will be multiple instantiations of the schema active as participants can communicate with different modalities in parallel or because one can view the process as operating on different levels as will be pointed out below.

Another common instantiation is the case where someone (x) produces a speech act (CA1), which is attended to and interpreted by y who decides to offer a speech act (CA2) in reply, after which x responds by producing a new speech act (CA1'). The two participants take alternating turns and each next utterance is a reply to the previous one forming adjacency pairs as they are commonly called in the tradition ([8]) of conversation analysis¹.

A third common instantiation has been labelled *interactional synchrony*. It was first described by Condon and Ogston

¹In [9], Goffman provides a very insightful analysis of this process of replies and responses.

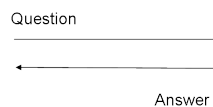


Figure 4: Sequential Implicativeness and adjacency pairs

[10] and an episode in a conversation was analysed in detail by Kendon [11]. The term refers to the case where the flow of movements of the listener are rhythmically coordinated with those of the speaker. Other forms of coordination have been called mimicry [12], mirroring [13, 14]. Hadar and colleagues [15] report that approximately a quarter of all the head movements of the listeners in the conversations they looked at occurred in sync with the speech of the interlocutor. Interestingly, McClave [16] notes that (many of) these kinds of movements may be elicited by the speaker.

Microanalysis of speaker head movements in relation to listener head movements reveals that what were heretofore presumed to be spontaneous, internally motivated, listener responses are actually responses to the speaker's nonverbal requests for feedback. These requests are in the form of up-and-down nods, and listeners recognize and respond to such requests in a fraction of a second.

In our corpus we regularly found a similar pattern with small shakes.

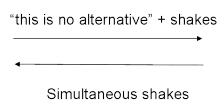


Figure 5: Synchronous behavior

Again, this shows the dependence of an action by one participant on the action of another, the back-and-forth of eliciting actions and responses.

We turn to some fundamental notions in linguistics that provide more insight or at least terms, related to this back-and-forth mechanism.

2.1. Speech Acts

The crucial insight that Speech Act theory ([5], [6]) has emphasised is that “language is used for getting things done”. Typically, in the case of language, these things implicate the person or persons to which the utterance is being addressed. From a speech act perspective, any utterance is some kind of invitation to the addressees to participate in a particular configuration of actions: attend to what is being said, try to figure out what is meant and carry out what was intended by the speaker, which could range from updating a belief state, to feeling offended, or closing the window. Speech act theory focusses on the perspective of the speakers and their *intentions* which implicate the audience in that an utterance is primarily intended to get the audience to recognize the speaker’s meaning: “To say that a speaker meant something by X is to say that the speaker intended the utterance of X to produce some effect in the audience by means of the recognition of this intention.” This is essentially Grice’s definition [17]. Another way in which the perspective of the speaker comes to the fore is in the way that Grice [18] formulates his maxims of co-operative behaviour (be relevant, be conspicuous, etcetera) in terms of what the speaker

should and should not do. All of these maxims indirectly take listeners into account as they urge the speaker to keep them in mind for the sake of co-operation. They presuppose some kind of Theory of Mind that is capable to deal with the right amount of audience or recipient design.

As with any event, a speech event can be described in several ways. One might say that in describing a particular situation the speaker was “stuttering”, “trying to say something in English” “trying to propose”, “making a fool of himself”, etcetera. By using the word “stuttering” one is referring to an aspect of the production and vocalisation process. The second characterisation points out that the vocalisations were not random but attempts to construct an English sentence. The third describes the intention behind the action and the last the effect it may have achieved on the other participants, the observers or those that have heard about the event. Austin [5] proposed some different terms to distinguish the levels in the speech event. The uttering itself, he called the locutionary act. The act of getting the audience to recognize what is intended is called the illocutionary act (the speaker tries to make it clear that the utterance is intended as a promise, for example). The effects the execution of the speech act has on the audience are called the perlocutionary effects. The acts that caused these effects were the perlocutionary acts. Note that not all of the effects may have been intended. For instance, if the speaker is not aware that the action promised is not something the audience wants, than the promise may actually turn out to be a threat.

In Clark’s framework ([4]), a speaker acts on four levels (action ladders). (1) A speaker executes a behaviour for the addressee to attend to. This could be uttering a sentence but also holding up your empty glass in a bar (to signal to the waiter you want a refill) (2) The behaviour is presented as a signal that the addressee should identify as such. It should be clear to the waiter that you are holding up the glass to signal to him and not just because of some other reason. (3) The speaker signals something which the addressee should recognize. (4) The speaker proposes a project for the addressee to consider (believe what is being said, except the offer, execute the command, for instance). In this formulation of levels, every action by the speaker is matched by an action that the addressee is supposed to execute: attend to the behaviour, identify it as a signal, interpret it correctly and consider the request that is made. If one considers the diagram above, one could say that instead of one arrow going from A to B there are four. Also, the arrow should be considered both from the perspective of the speaker and the recipient.

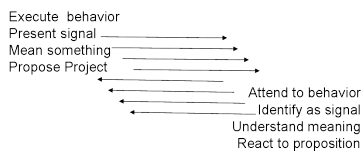


Figure 6: Action Ladders and Reciprocity

2.2. Monitoring and Feedback

The back and forth of speaker and listening activity is accounted for by the necessity of the speaker to monitor for success and the need of the listener to provide feedback. If we take the perspective of the listener, we can make a similar distinction in four levels on which the listener can provide feedback. Allwood ([19],

for example) put forward a distinction of the following four basic communicative functions on which the interlocutor can give feedback: *Contact* (i.e., whether the interlocutor is willing and able to continue the interaction), *Perception* (i.e., whether the interlocutor is willing and able to perceive the message), *Understanding* (i.e., whether the interlocutor is willing and able to understand the message), *Attitudinal reactions* (i.e. whether the interlocutor is willing and able to react and (adequately) respond to the message, specifically whether he/she accepts or rejects it). These levels repeat and complement the action ladders from Clark above.

Important for all the parties in the cooperative endeavour that is conversation is to know that common ground has been established, that the addressee understands what the speaker intended with the talk produced and the speaker knows that the intentions were achieved. So the feedback that is voluntarily or involuntarily provided by listeners is monitored by the speakers in order to get closure on their actions i.e. in order to know to what degree the intended actions were successful. Goodwin’s rule - whenever a speaker looks at his audience, the audience should look at the speaker - provides a basic example of this need to check for contact and perception. By monitoring the behaviour of the other participants, a speaker can thus derive information about such elements as attention, perception, understanding, and the willingness to engage and accept or reject collaboration. Some of the information derives from the actions of listeners that go into perception of the signals (such as their gaze telling something about the focus of attention) but other behaviors may be explicit signals of understanding and agreement or lack thereof through facial expressions or small non-disruptive interjections.

Several conversational actions are conventionally dedicated to establish “grounding” (the mutual belief by the partners in conversation that they have understood what the contributor meant, Clark & Schaefer, [20]). In Clark & Schaefer, a discourse model is presented in which it is assumed that the presentation phase of the speaker is paralleled with an acceptance phase by the recipient which is essential for grounding. Either following, in the next moves or by behaviours during the production of communicative actions by the speaker. Obvious signs of neglect of attention, or signs of difficulty in understanding will yield reparative actions by the speaker. Positive signs indicating attention, perception, understanding, processing (understanding, agreement, willingness, etc.) will lead the speaker to assume the message has been grounded or successfully executed on all the relevant levels.

The acceptance phase itself consists of the presentation of a contribution to which the original presenter can react with an accepting contribution, illustrating another way to describe some of the loops presented in Figure 2.

One type of accepting contribution Clark & Schaefer call *acknowledgements*, which are “expressions such as *mhm*, *yes*, and *quite* that are spoken in the background, or gestures such as head nods and smiles.” These are commonly called *backchannels*².

²Yngve [21] is generally credited for having introduced the term. Some authors use other terms to refer to similar phenomena sometimes restricting the scope to a particular class of listener responses. Kendon [22] introduced the term accompaniment signals for “short utterances that the listener produces as an accompaniment to a speaker, when the speaker is speaking at length” which he divides into two groups: attention signals (in which one appears to signal no more than that one is attending) and assenting signals that express ‘point granted’ or ‘agreement’. Rosenfeld [23] uses the general term *listener response*. A related

The ‘speaker’ and ‘listener’ are both continuously in reception and production mode. Where actions of one are perceived and interpreted by the other and responded to.

3. Conclusions

The *interaction* between participants in a conversation comes about through the interplay of actions that elicit acknowledgement of reception and further response. With the acts of acknowledgement and the responses themselves eliciting further action. Figure 7 highlights this complexity of the layered interactions.

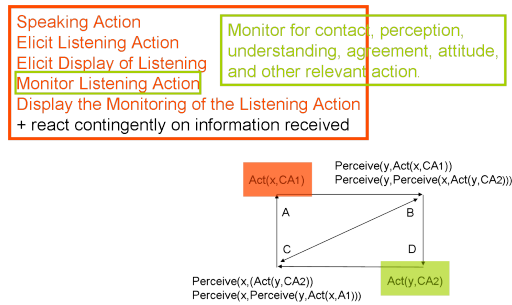


Figure 7: Picturing Conversation as an Interactional Achievement

A speaker producing an utterance (or any other communicative act) is also eliciting a listening action. For reasons of grounding, this also may involve not just an action of listening but also a “display” of listening. In this mode, the speaker as producer turns into a “recipient” monitoring the listening action - which can relate to the different levels of perception, understanding, agreement, etcetera. The speaker as listener might make sure that it is clear that he is taken the acknowledgements of the listener as producer of communicative actions into account. Speakers thus become listeners. Listeners become speakers. And yes, conversation is a tennis match, and a game of ping pong and of badminton, and of ... all at the same time.

4. Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 211486 (SEMAINE).

5. References

[1] M. Bakhtin, “The problem of speech genres,” in *The Discourse Reader*, A. Jaworski and N. Coupland, Eds. Routledge, 1999, pp. 98–107.

[2] J. Gumperz, *Discourse strategies*. Cambridge, England: Cambridge University Press, 1982.

[3] E. A. Schegloff, “Discourse as interactional achievement: Some uses of “uh huh” and other things that come between sentences,”

concept is that of *acknowledgement token* as used by Jefferson [24] or *continuers* from Schegloff [3]. Schegloff reflects on the use of ‘uh-huh’ as a signal of attention. This attention-signalling function of an ‘uh-huh’ or a head nod becomes apparent only if it is in response to an extended gaze by the speaker or a rising intonation soliciting some sign of attention, interest or understanding ([3, p. 79]. In other cases, the term continuer may be appropriate, according to Schegloff.

in *Analyzing discourse, text, and talk*, D. Tannen, Ed. Washington, DC: Georgetown University Press, 1982, pp. 71–93.

[4] H. H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.

[5] J. A. Austin, *How to Do Things with Words*. London: Oxford University Press, 1962.

[6] J. R. Searle, *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press, 1969.

[7] C. Goodwin, “Notes on story structure and the organization of participation,” in *Structures of Social Action. Studies in Conversation Analysis*, M. J. Atkinson and J. Heritage, Eds. Cambridge University Press, 1984, pp. 225–246.

[8] E. Schegloff and H. Sacks, “Opening up closings,” *Semiotica*, vol. 8, pp. 289–327, 1973.

[9] E. Goffman, “Replies and responses,” *Language in Society*, vol. 5, no. 3, pp. 225–313, 1976.

[10] W. Condon and W. Ogston, “Sound film analysis of normal and pathological behavior patterns,” *Journal of Nervouse Disease*, vol. 143, no. 4, pp. 338–347, 1966.

[11] A. Kendon, “Movement coordination in social interaction: some examples described,” *Acta Psychologica*, vol. 32, pp. 100–125, 1970.

[12] T. Chartrand and J. A. Bargh, “The chameleon effect, the perception-behavior link and social interaction,” *Journal of Personality and Social Psychology*, vol. 76, no. 6, pp. 893–910, 1999.

[13] M. LaFrance, “Nonverbal synchrony and rapport: analysis by the cross-lag panel technique,” *Social Psychology Quarterly*, vol. 42, no. 1, pp. 66–70, 1979.

[14] M. Lafrance and W. Ickes, “Posture mirroring and interactional involvement: sex and sex typing effects,” *Journal of nonverbal behavior*, vol. 5, pp. 139–154, 1981.

[15] U. Hadar, T. Steiner, and C. F. Rose, “Head movement during listening turns in conversation,” *Journal of Nonverbal Behavior*, vol. 9, no. 4, pp. 214–228, 1985.

[16] E. Z. McClave, “Linguistic functions of head movements in the context of speech,” *Journal of Pragmatics*, vol. 32, pp. 855–878, 2000.

[17] H. P. Grice, “Meaning,” *The Philosophical Review*, vol. 66, no. 3, pp. 377–388, 1975.

[18] —, “Logic and conversation,” in *Syntax and Semantics: Vol. 3: Speech Acts*, P. Cole and J. L. Morgan, Eds. San Diego, CA: Academic Press, 1975, pp. 41–58.

[19] J. Allwood, “Feedback in second language acquisition,” in *Adult Language Acquisition. Cross Linguistic Perspectives*, C. Perdue, Ed. Cambridge, New York: Cambridge University Press, 1993, pp. 196–232.

[20] H. Clark and E. Schaefer, “Contributing to discourse,” *Cognitive Scienc*, vol. 13, pp. 259–294, 1991.

[21] V. Yngve, “On getting a word in edgewise,” in *Papers from the sixth regional meeting of the Chicago Linguistic Society*, Chicago: Chicago Linguistic Society, 1970, pp. 567–77.

[22] A. Kendon, “Some functions of gaze direction in social interaction,” *Acta Psychologica*, vol. 26, pp. 22–63, 1967.

[23] H. Rosenfeld, “Conversational control functions of nonverbal behavior,” in *Nonverbal behavior and communication*. Hillsdale NY: Lawrence Erlbaum Associates, 1987, p. 563601.

[24] G. Jefferson, “Notes on a systematic deployment of the acknowledgement tokens ‘yeah’ and ‘mm hm’,” *Papers in Linguistics*, vol. 17, pp. 197–206, 1984.