16 Pragmatics of Language Performance

HERBERT H. CLARK

1 Introduction

Language seems orderly when it is found in novels, plays, and news broadcasts, but much less so when it is heard in cafés, classrooms, and offices. Take this exchange between two British academics:

(1) Peter: And he's going to go to the top, is he?
Reynard: Well, Mallet said he felt it would be a good thing if Oscar went.

This is an example worthy of a playwright, but what Peter and Reynard actually produced was this:[1]

(2) Peter: and he's going to . go to the top, is he?
Reynard: well, . I mean this . uh Mallet said Mallet was uh said something about uh you know he felt it would be a good thing if u:h . if Oscar went, (1.2.370)

In his answer, Reynard decides what to say as he goes along. He takes first one direction ("Mallet said something about") and then another ("he felt it . . ."). Along the way he replaces phrases, makes clarifications (with *I mean* and *you know*), and introduces delays (with *uh*). Reynard's utterance looks anything but orderly, and yet he succeeds in coordinating with Peter on what he wanted to say. How do they manage?

Pragmatics traditionally has focused on the PREPLANNED, NON-INTERACTIVE language of novels, plays, and news broadcasts. The pioneers in the field (e.g. Austin 1962, Lewis 1969, Searle 1969, 1975a, Bach and Harnish 1979, Sperber and Wilson 1986a, Grice 1989) all worked from clean, invented examples.[2] But if pragmatics is the study of language in use, it must also account for the SPONTANEOUS, INTERACTIVE LANGUAGE of cafés, classrooms, and offices. Language evolved, after all, before people could read or write, attend plays, or

watch television. Even today, the primary setting for language use is conversation. Accounting for the features of (2) that are absent from (1) will require principles beyond those needed for preplanned, non-interactive language.

This chapter is about communicative acts that are needed in the performance of language. The argument is this (H. Clark 1996, 1999): Spontaneous, interactive language has its origins in joint activities. When people do things together in cafés, classrooms, and offices, they need to coordinate their individual actions, and they use a variety of communicative acts to achieve that coordination. These constitute the PRIMARY SYSTEM of communication – the official business of their discourse. But communicative acts are themselves joint actions that require coordinating, and people have a special class of communicative acts for this coordination – including many of Reynard's actions in (2). These constitute the COLLATERAL SYSTEM of communication. The goal here is to characterize the collateral system and the pragmatic principles by which it works.

2 Saying and Displaying

In using language, speakers make communicative choices of many types. Consider an exchange between two academics in a British university common room:

(3) Nancy: I acquired an absolutely magnificent sewing-machine, by foul means, did I tell you about that,
Julia: no.

Nancy utters the sentence *Did I tell you about that?* as a signal to Julia. A SIGNAL here is any action by which one person MEANS something for another person in the sense of Grice (1989). Nancy performs her utterance to ask Julia a question and, in turn, to gain her consent to tell her a story.

Signals, however, are Janus-faced objects. One face is CONTENT, the choice of *what* to signal. Nancy, for example, chooses to seek Julia's permission to tell a story; she chooses to do that by asking a question; to ask the question, she selects the English sentence *Did I tell you about that?*; and so on. The other face is PERFORMANCE, the choice of *how* to realize the signal. Nancy, for example, chooses to direct her voice, face, and gestures at Julia. She does this to designate herself as speaker, and Julia as addressee. She chooses to initiate her utterance at that precise moment – not earlier or later – to designate the now of her signal (e.g. for the interpretation of *did*) and its relation to her previous phrases (e.g. for the interpretation of *that*). Nancy's very realization of the sentence *Did I tell you about that?* designates the content of her signal.

Speakers make choices not only in WHAT they say, but in HOW they say it. They perform what they say IN A PARTICULAR TIME, PLACE, AND MANNER – at

the right moment, for the right duration, originating from and directed to the right locations, at the right amplitude, with the right gestures. In the terminology adopted here, they DISPLAY their signals to others in order to designate such things as the speaker, addressee, time, place, and content of their signals. What speakers mean by a signal, then, is determined by their choice of both content and display. Schematically:

signal = content + display

Both parts are necessary to the whole. Speakers cannot express the content of their signal without displaying it, nor can they form a display without content to display. And both parts require interpretation *à la Grice*.

2.1 Displays as indicative acts

Displays, in this view, are communicative acts of INDICATING. The prototype of indicating is pointing. When June asks, "Which car is yours?" and David points at a nearby Honda, David indicates the Honda as "his car." His act of pointing is an *index* to the car. An index, according to C. S. Peirce (Buchler 1940), signifies its object, its referent, by means of an INTRINSIC CONNECTION – a spatial or causal connection – between the index and the object. David's pointed finger has a spatial connection to the car. The problem is that his pointed finger also has intrinsic connections to his fingernail, his right arm, and the left front door of the Honda, and so brute pointing isn't enough. June is to recognize the connection David intends by inferring his purpose against their current common ground. As a response to her question, he must be indicating the Honda as his car.

Nancy's display of "Did I tell you about that?" has intrinsic connections to a number of SITUATION INDIVIDUALS, and these allow Nancy to use her display to indicate, or point to, these individuals. So in the very realization of her display, Nancy creates a set of PERFORMANCE INDEXES:

- PRODUCER (abbreviated *p*): Nancy uses the index *p* to indicate herself as producer of the signal. She creates *p* through the source and distinctive quality of her voice. Let us denote Nancy's index as *p* ("did I tell you about that"), which is to be read: the index to the producer of the display of "did I tell you about that."
- RECIPIENT (*r*): Nancy uses index *r* to indicate Julia as the recipient of the signal. She creates *r* by gazing at Julia and directing her voice in Julia's direction at an amplitude appropriate for a person 1.5 meters away.
- TIME (*t*): Nancy uses index *t* to indicate the current moment as the now of the signal. She creates *t* by producing her utterance over that precise interval.
- LOCATION (*l*): Nancy uses index *l* to indicate the current location as where she is posing her question. She creates this index by the source of her voice and gestures and by the placement of her body.

- CONTENT (*c*): Nancy uses index *c* to indicate what she is realizing as the content of her signal to Julia.

These are not the only indexes Nancy can create, and each can be analyzed further.

Content *c*, in particular, is a complex index. In traditional accounts, the content of a signal has several levels (Austin 1962, Searle 1969, Bach and Harnish 1979). When Nancy produces "Did I tell you about that," she performs a PHONETIC ACT (producing certain speech sounds), an ILLOCUTIONARY ACT (asking a question), and a PERLOCUTIONARY ACT (getting Julia to agree to answer it), among others.

These accounts, however, assume that speakers act independently of their addressees, and vice versa, and the assumption is clearly false (H. Clark 1996). In their dialogue, Nancy and Julia work together to assure (1) that Julia attends to Nancy's vocalizations and gestures; (2) that Julia identifies Nancy's words and gestures; (3) that Julia figures out what Nancy means by her words and gestures; and (4) that Julia considers what Nancy is proposing. The two of them engage in joint actions at the four levels shown in the table in (4), with A as speaker and B as addressee. These form a ladder of actions that run from Level 1 to Level 4. A and B perform their joint actions at Level 2 by means of their joint actions at Level 1, and so on up the ladder (see Goldman 1970).

(4) Four levels of joint action in communicative acts

Level	**Speaker A's action**	**Addressee B's action**
1.	A makes sounds, gestures for B	B attends to A's sounds, gestures
2.	A presents a signal for B	B identifies what A's signal is
3.	A means something for B	B understands what A means
4.	A proposes a joint project to B	B considers A's proposal

Nancy's display of "Did I tell you about that?" then indexes at least four levels of content: Nancy's sounds and movements; Nancy's phrases and gestures; what Nancy means; and Nancy's proposal. We can denote the index to what she means (Level 3), for example, as *c3* ("Did I tell you about that?"). When Nancy indicates what she means with this index, she expects Julia to identify the referent of the index – namely, what she means by "Did I tell you about that?" If Julia were to say, "What did you mean by that?" she would be using *that* to refer to that content.

In this scheme, therefore, speakers use the display of a signal – the time, place, and manner of its performance – to indicate situational individuals that are essential to the interpretation of the signal. It is as if the producer were saying to the recipient, "In displaying this signal to you, I hereby indicate myself as producer, you as recipient, now as the time of the signal, here as my location, and this sentence, among other things, as the content of the signal."

Without these indexes, Nancy has no way of establishing that she is addressing Julia there and then with the content of her utterance.

2.2 Uses of performance indexes

The most obvious use of displays is to fix the referents of INDEXICAL EXPRESSIONS such as *I*, *you*, *here*, *now*, *this*, and *that*. For Nancy to ask Julia, "Did I tell you about that?" she must specify the individual people, objects, and times she is referring to with *I*, *you*, *that*, and *did*. She does this by indicating them with the display of her utterance – with performance indexes *p*, *r*, *l*, and *t*.

Consider the word *I*, which means "the person producing this word." This meaning doesn't by itself specify the referent. It takes *p* ("I") to complete the specification. In (3), then, Nancy's use of *I* indexes *p* ("I"), and it is *p* ("I") that indexes Nancy herself. Even with the display, it isn't always easy for speakers to get addressees to identify the producer. When a student in a large class yells, "I do" in answer to the question, "Does anyone need a syllabus?" the professor may need to ask, "Who said that?" to fix the student's identity. The same holds for other indexical expressions.

Performance indexes have other purposes, too. When June sees David across the street and yells "Hey," she indicates him as his addressee by means of *r* ("hey"), based on the amplitude and direction of her voice and the direction of her gaze and waving hand. When June asks David in a dark room, "Where are you?" and he answers, "Yo," he uses *l* ("yo") to indicate his location at the moment indicated by *t* ("yo") . When a race official says "Ready . . . set . . . go!" he uses *t* ("go") as the starting time of the race. June, David, and the race official use these indexes to fix speakers, addressees, times, locations, and content even when there are no indexical expressions.

The point is especially clear with interjections (see Wilkins 1992). When June calls David on the telephone and says, "Hello," she is performing an illocutionary act whose basic meaning, crudely put, is "I hereby greet you here now." Unlike the paraphrase, June's utterance ("Hello") makes no explicit reference to the speaker, addressee, time, location, or content of that utterance. She indicates these individuals through her display of *hello*, creating *p*, *r*, *t*, *l*, and *c* to bind the arguments associated with *greet*. These is nothing special about interjections. Speakers create the same indexes whenever they display signals to other people.

In all these examples, performance indexes are being used in the primary system of communicative acts – for the official business of the discourse. They are also essential to the collateral system, the topic we turn to next.

3 Coordinating on the Use of Language

Language is ordinarily used for coordinating people's participation in joint activities. Consider Alan and Barbara assembling a TV stand from its parts, as

captured on videotape.[3] The two of them proceed by agreeing on which pieces to connect at which moments, how to orient each piece, who will hold a piece while the other attaches the screws, and so on. They reach these agreements, or JOINT COMMITMENTS, by means of language and other signals.

What are these agreements about? To engage in any joint activity – from assembling TV stands to negotiating contracts – people must become jointly committed, explicitly or tacitly, to certain REQUISITES, including these:

- Participants: What individuals are to participate in the joint activity?
- Roles: In what roles?
- Content: What actions are they to perform, and what conditions are they to adopt?
- Timing: When are the actions to take place and the conditions to take effect?
- Location: And where?

In assembling the TV stand, Alan and Barbara begin by agreeing to be the participants, as co-builders, in assembling a TV stand at that time and location. Later, as they go along, they agree to more specific content, timing, and locations.

A common way to reach joint commitments is with what I will call PROJECTIVE PAIRS. A projective pair consists of two actions, by two people, in which (a) the first person PROPOSES a joint project to the second, and (b) the second person TAKES UP that proposal in some way. The classic form is the ADJACENCY PAIR (Schegloff and Sacks 1973), as in this exchange in assembling the TV stand:

(5) Alan: Now let's do this one [picking up the top-piece]
Barbara: Okay.

In turn 1, Alan makes a proposal to Barbara, and in turn 2, she takes it up, establishing the joint commitment to do the top-piece next. In adjacency pairs, however, both parts must be spoken actions, and in many situations, one or both actions are non-linguistic, as in this example:

(6) Barbara: [Extends hand with screw] So you want to stick the screws in?
Alan: [Extends hand to take screw]

Here Alan takes up Barbara's request, but with a gestural signal. The term *projective pair* is intended to cover both types of pairs.

Using language, however, is itself a joint activity, which requires its own coordination. As noted earlier, communicative acts consist of joint actions at four levels, so in dialogue the participants have to manage who talks when, whether they are attending to, hearing, and understanding each other as intended, and so on. If so, then the participants should have to agree on the

same five requisites for each signal. For Alan to suggest to Barbara, "Now let's do this one," they must agree on the participants (Alan and Barbara), their roles (speaker and addressee), the timing of the signal (starting with Alan's "now"), the location (there), and the content of his signal (all four levels). These, of course, are just the elements indicated by the five performance indexes of Alan's display – producer, recipient, time, place, and four levels of content.

Using language, in short, requires speakers and addressees to work together to establish the intended producer, recipient, time, place, and four levels of content.

3.1 Grounding

To communicate is, etymologically, to "make common" – to establish something as part of common ground. But what is it that speakers try to make common? The obvious answer is what they are saying – for example, Alan's suggestion in (5) that he and Barbara do the top-piece next. But to do that, they have to make common all five requisites – not only content, but also participants, roles, timing, and place.

The process of establishing something as common ground is called GROUNDING (H. Clark and Marshall 1981, H. Clark and Wilkes-Gibbs 1986, H. Clark and Schaefer 1989, H. Clark and Brennan 1991). Take Alan's suggestion in (5). To succeed at Level 3, the two of them must establish the mutual belief that Barbara has understood what Alan means by "Now let's do this one." They don't need to establish this mutual belief for certain, but only WELL ENOUGH FOR CURRENT PURPOSES. To ground something, therefore, is to establish it as part of common ground well enough for current purposes.

In dialogue, people ground many signals in the very course of their official business. Consider this exchange:

(7) Kenneth: how how was the wedding, –
Fran: oh it was it was really good, (7.31.1441)

In the first turn, Kenneth invites Fran to tell him about a wedding. He does this by getting her attention, getting her to identify the sentence *how was the wedding*, getting her to understand what he means, and getting her to consider his invitation. But is he successful? Fran gives him evidence that he is by responding with an appropriate reply. With it, she shows that she has attended, identified his sentence, grasped his invitation, and considered telling him about the wedding. That evidence is just what they need to reach the mutual belief that they have established all four levels of content well enough for current purposes. It is also evidence for the mutual belief that they have established the speaker, addressee, timing, and location as well.

Other forms of grounding work by a different logic. Fran and Kenneth continue the dialogue in (7) as follows:

(8) Fran: it was uh it was a lovely day,
Kenneth: yes,
Fran: and . it was a super place, . to have it . of course,
Kenneth: yes, –

The first two lines may appear to form a standard adjacency pair, but they do not. In line 2, Kenneth does not mean "Yes, it was a lovely day." He has no idea whether the day was lovely or not. What he means is, "Yes, I *understand*, or *see*, what you mean by 'it was uh it was a lovely day.'" And in line 4, he means, "Yes, I *understand*, or *see*, what you mean by 'and it was a super place to have it of course.'" His *yes*'s are not about the content itself, but about his IDENTIFICATION of that content. The first projective pair is really this:

(8′) Fran: Have you identified *c* ("it was uh it was a lovely day")?
Kenneth: yes,

The *yes*'s in (8) are ACKNOWLEDGMENTS.[4] They are also called CONTINUERS. As Schegloff (1982) noted, when Kenneth says "yes" in (8), he is signaling Fran to continue, as if he were saying, "Go on." The BASIC MEANING of *yes* in these positions is "Yes, I understand (well enough for current purposes)," and speakers can use that basic meaning via Grice's Maxim of Relevance to implicate "Please continue." In English, the commonest acknowledgments are *uh-huh*, *yes*, *yeah*, *m-hm*, *m*, and head nods, which are all possible *yes* answers to yes/no questions. *Yes* answers are also used as acknowledgments in languages ranging from German (*ja* and *m-hm*) and French (*oui*) to Japanese (*hai*) and Mandarin (*shì*). All this is evidence that *yes* is an acknowledgment first ("Yes, I hear or understand") and a continuer by implicature.

Speakers and addressees, however, often run into problems establishing performance indexes, and when they do, they exploit a variety of techniques. One such technique is the SIDE SEQUENCE (Jefferson 1972, Schegloff 1972). In the following exchange, Maggie and Julie have a problem establishing index *c3*:

(9) Maggie: you fancy it yourself do you? –
Julia: what, the men's doubles?
Maggie: yeah,
Julia: well more than the singles, yes, – (7.3e.278)

Julia seems unsure of what Maggie meant by "it," so she initiates a side sequence to ground *c3* ("it"). In turn 2, Julia asks, in effect, "By c3 ('it'), do you mean c3 ('the men's doubles')?" and in turn 3, Maggie answers "yeah." With that collateral issue resolved, they return to the official business of their conversation. Side sequences can be used to establish all levels of content as well as other performance indexes. A potential addressee, for example, could initiate a side sequence about index *r* with "Are you talking to me?"

3.2 Performance indexes in dialogue

In dialogues, people use performance indexes not only to INTRODUCE new producers, recipients, times, locations, and content, but to MAINTAIN the current ones. In ongoing discourse, it is often easy to specify *p* and *r* because the potential producers and recipients are limited. When there are two participants, *p* predictably becomes *r*, and *r* becomes *p*, with every change in speaker. At the start-up of a discourse, people need to establish the initial set of participants, perhaps using a summons (e.g. "Hey, Barbara") or ringing someone on the telephone (Schegloff 1968, 1979).

Introducing new bindings for displays can occur even within an utterance. In the following example, John produces and utterance in three increments, each with a different addressee (Goodwin 1981: 160):

(10) John: I gave, I gave up [gazing at Don] smoking cigarettes:
Don: Yeah,
John: I-uh: [turning gaze to Beth] one-one week ago toda:y. [turning gaze to Ann] actually,

As John speaks, he moves his gaze from Don to Beth to Ann. He designs the first increment ("I gave up smoking cigarettes") as new information for Don. But since his own wife Beth already knows this, he designs the second increment ("one week ago today") to remind her of the one-week anniversary of his achievement. He designs the third increment ("actually") for Don's wife to acknowledge that all this is new to her. John designs each increment for each addressee, indicating these bindings with his display.

Coordinating on the use of language, in short, requires coordinating on each signal's performance indexes – producer, recipient, timing, location, and content. Speakers and their addressees, indeed, work to make this information common through the process of grounding.

4 Collateral Signals

Coordinating on performance indexes cannot always succeed with PRIMARY SIGNALS alone, that is, with signals that refer to the official business of the discourse. In cafés, classrooms, and offices, people often start before they know what they want to say and then change their minds. They normally start speaking before they have selected every word or phrase and are therefore often delayed in coming up with later words and phrases (Levelt 1989).

To deal with these problems, speakers and addressees often deploy COLLATERAL SIGNALS, this is, signals that refer to the *local, ongoing performance* of those primary signals.[5] The distinction between primary and collateral signals is nicely illustrated, in different terminology, in Goffman's (1981) analysis of radio talk. As he noted, radio announcers are expected "to produce the effect

of a spontaneous, fluent flow of words – if not a forceful, pleasing personality – under conditions that lay speakers would be unable to manage" (p. 198). As good as they are, they still run into problems, and when they do, they often add parenthetical asides to correct, poke fun at, apologize for, or otherwise explain the problems. Take the following example:

(11) Announcer: Seventy-two degrees Celsius. I beg your pardon. Seventeen degrees Celsius. Seventy-two would be a little warm. [Continues]

The announcer's official business is to report the weather, which he does with "seventy-two degrees Celsius" and "Seventeen degrees Celsius." But to present himself as knowing and attentive, he inserts an apology and brief joke as commentaries on his performance. These we recognize to be his personal actions and distinct from his official business, the weather report.

Collateral signals take many forms besides parenthetical asides. They fall into four main categories, each with its own properties:

- Inserts: signals that are achieved by INTERPOSING linguistic expressions between two parts of a primary utterance or between two primary utterances.
- Modifications: signals that are achieved by MODIFYING part or all of the display of a primary utterance.
- Juxtapositions: signals that are achieved by displaying one primary utterance JUXTAPOSED against another.
- Concomitants: signals that are displayed at the same time as, but separate from, a primary signal.

These signals differ in form, but they all refer to the local, ongoing performance, and that makes them collateral signals.

4.1 Inserts

Inserts are parenthetical comments or exchanges interposed within the official business of a discourse. They may be side exchanges, side moves, or simple asides.

Side exchanges. These are projective pairs interposed within or between utterances to deal with local, ongoing performance. The most common are side sequences. In (9), for example, Maggie and Julia interpose an exchange between Julia's question and Maggie's answer:

(12) Julia: what, the men's doubles?
Maggie: yeah,

Julia asks "what, the men's doubles?" to clarify what Maggie meant, and Maggie takes Julia up with her answer "yeah." The entire exchange is interposed within their primary talk.

Side moves. These consist of the interposition of the second part of a projective pair within or between utterances. Consider the acknowledgment in (8′), repeated here:

(13) Fran: it was uh it was a lovely day,
Kenneth: yes,

Fran refers to the content of what she is saying with c ("it was uh it was a lovely day"), which Kenneth takes up with "yes [I understand *c*]." The second part gets inserted within the ongoing talk as a side move.[6] Side moves can also originate much deeper within an utterance, as here (see also (17) below):

(14) Susan: they still talk about rubbish tins, which is the American the Australian
Jean: m yeah,
Susan: expression, . for that thing you put all the . stuff in at the back gate, you know? (1.10.388)

When Susan corrects "the American" to "the Australian," Jean acknowledges the correction with "m" and verifies it with "yeah" before Susan goes on. The projective pair "the Australian" plus "m yeah" occurs not merely mid-utterance, but mid-noun phrase (*the Australian expression*).

Asides. These are expressions that speakers interpose within their own utterances to comment on their local, ongoing performance.

Speakers tend to produce utterances one parcel of speaking at a time. By PARCEL OF SPEAKING, I mean a continuous fluent stretch of an intonation unit in the target utterance. Consider Reynard's utterance in (2). In the following, it is annotated with left curly brackets to mark points of suspension of fluent speech and right curly brackets to mark points of resumption of fluent speech. Each pair of brackets encloses a HIATUS, which may contain other speech, a pause, a gesture, or nothing. The hiatuses are lined up on the right:

(15) well,	{. I mean}
this	{. uh}
Mallet said	{ }
Mallet was	{uh}
said something about	{uh you know}
he felt it would be a good thing if	{u:h .}
if Oscar went,	

Here Reynard produces thirteen parcels. Seven belong to his primary utterance, and six come in hiatuses: *I mean*, *you know*, and four instances of *uh*. These six are all asides.

Asides are inserted to help explain features of local performance. In line 5, Reynard introduces *you know* to say he is adding a clarification to what he had

already started to say. It is as if he had said, "Instead of 'Mallet said something about' what I mean more specifically is 'he felt it would be . . .'" *I mean* and *you know* are both clarifying asides, but they contrast in meaning (Levelt 1983, 1989, Erman 1987, Fox Tree and Schrock 2002).

The commonest asides are the FILLERS *uh* and *um*. Speakers use *uh* and *um* to announce "I am initiating at *t* (filler) what I expect to be a minor or major delay in speaking" (Fox Tree 2001; H. Clark and Fox Tree 2002). They use *uh* to announce minor delays, and *um*, major delays. By introducing these at the right moments, speakers can also create implicatures, via Grice's Maxim of Relevance, that range from "I am now searching for a word" to "I now invite you to speak." All these signals are addressed to issues of performance.

Speakers often introduce asides to comment on *why* they are doing what they are doing locally. It is common to comment on self-repairs (Schegloff et al. 1977; Levelt 1983), as in this example:

(16) Robert: or do they only know about thi:y {.} practical, {. excuse me} experimental aspects, of reading, (2.4.736)

Robert begins with *practical*, thinks better of it, and replaces it with *experimental*. He introduces *excuse me* to apologize for misleading his addressee, if only for one word.

Expressions such as *I mean*, *you know*, *excuse me*, *well*, *oh*, *like*, *ah*, *now*, *uh*, and *um* have been classified under many names. These include editing expressions (Levelt 1983), discourse markers (Schiffrin 1987), discourse particles (Schourup 1985), pragmatic expressions (Erman 1987), disjunct markers (Jefferson 1978), discourse operators (Polanyi 1985, Redeker 1986, 1990), clue words (Reichman 1978), cue words (Grosz and Sidner 1986), and cue phrases (Hirschberg and Litman 1987). Some of these schemes, such as Schiffrin's discourse markers, classify a wide range of expressions, whereas others, such as Levelt's editing expressions, are restricted to one or two functions. It remains to be seen how these schemes deal with the contrast between the primary and collateral systems.

4.2 Modifications

Another way to signal a local problem in performance is to MODIFY a current word or phrase. Among the common modifications in English are try markers, non-reduced vowels, and prolongations, but there are many others.

Try markers. When speakers present a name or description they aren't sure is correct or comprehensible, they can mark it with what Sacks and Schegloff (1979) have called a TRY MARKER, a rising intonation followed by a slight pause. With this modification, speakers initiate a projective pair to get their partners to confirm or correct the constituent before the speakers go on, as in (17):

(17) Alan: so I wrote off to {.} Bill, {. uh} who had presumably disappeared by this time, certainly, a man called Annegra? {–}
Barbara: {yeah, Allegra}
Alan: Allegra, {uh} replied, {. uh} and I {.} put {.} two other people, who'd been in for {.} the BBST job {.} with me . . . (3.2a.59)

Alan apparently is uncertain about the name *Annegra*, so he presents it with rising intonation and a slight pause. Barbara responds "yeah" to confirm she knows who he is referring to, then corrects the name to *Allegra*. Alan accepts the correction by repeating *Allegra* and continuing. All of this happens mid-sentence.

The try marker in (17) is an INTONATIONAL modification of the word *Annegra* that points to, or flags, a problem associated with the word itself, roughly, "Do you understand who I'm referring to with the word *Annegra*?" It also projects an immediate response. So Alan's try marker functions as the first part of a projective pair, which Barbara takes up with the side move "yeah, Allegra."

Non-reduced vowels. Speakers ordinarily pronounce *the*, *a*, and *to* with reduced vowels (schwas) which I will write *thuh monastery*, *uh gravel company*, and *tuh regulate*. On occasion, speakers produce these vowels in a marked NON-REDUCED form, which I will write *thiy*, *ei*, and *tuw* (rhyming with *see*, *day*, and *glue*). Here is an example for *thiy*:[7]

(18) Kate: it's thiy {.} thuh monastery, {– you know} thuh very Gothic monastery, with all thi:y {–} wedding-cake, {– –} I – it's a special kind of Gothic architecture, which is even more decorated than Decorated. (2.13.664)

Speakers often use *thiy* in place of *thuh* to signal that they are suspending speech immediately after the word *the*, as Kate does twice in (18) (Fox Tree and H. Clark 1997). In signaling a suspension, they implicate that they are having a problem that prevents them from continuing immediately. We find similar examples with *a* and *to* in (19) and (20):

(19) Roger: we are ei {um – –} ei gravel company (2.11b.1063)

(20) Albert Gore: not be tuw um regulate (televised debate, October 17, 2000, St. Louis, Missouri).

Non-reduced vowels can also combine with *uh* or *um* to form complex modification-plus-asides (H. Clark and Fox Tree 2002). Gore's *to um* in (20) was pronounced *tu.wum*, a trochee with the syllable boundary (marked by the period) before the *w* in *tuw*. Likewise, *the uh* is often pronounced *thi.yuh*, and *a uh* as *ei.yuh* – trochees with syllable boundaries before the *y* in *thiy* and *ei*. So when Gore uttered "not be tu.wum regulate," he was signaling not just a suspension at the end of *to*, but the initiation of a delay at the beginning of *um*.

Prolongations. Speakers often prolong words mid-utterance as they plan what to say next. Take this excerpt from a narrative about picking pears (Chafe 1980: 308):[8]

(21) A—nd u—m [3.35] he's just picking them, he comes off of the ladder, [.35] a—nd he— u—h [.3] puts his pears into the basket.

In (21), the narrator prolongs *and*, *um*, *and*, *he*, and *uh*. With these prolongations, he slows down the initial parts of two clauses, but then produces the remaining parts fluently.

Prolongations are used to signal the CONTINUATION of a delay in progress (H. Clark and Fox Tree 2002). They contrast with *uh* and *um*, which mark the INITIATION of a delay. Also, prolonged words belong to the primary utterance, whereas *uh* and *um* do not. Speakers often want to advance the primary utterance – or give the appearance of doing so – even when they cannot really go on. They can do that with prolongations, but not with *uh* or *um*.

4.3 Juxtapositions

Juxtapositions are signals achieved by juxtaposing the display of one primary utterance against the display of another. The two displays may be produced by the same or different speakers. Here are three techniques that exploit juxtaposition.

Replacement. The commonest way to repair a word or phrase is to juxtapose a new display directly against what it is to replace (see Levelt 1983). Consider (22):

(22) Peter: there was {u:m .} w- {} what is {.} has happened since then, is that there has been another meeting of the executive committee, (1.2.80)

Treated literally, Peter's utterance is ungrammatical (even ignoring *um*): *there was w-what is has happened since then* . . . Peter surely isn't trying to be ungrammatical, and yet he produces each word deliberately, not in error (except perhaps *w-*), as a part of what he wants to say at that moment. It is just that he changes his mind several times, and he uses juxtapositions to let his addressee know *how* he is changing his mind.

In (22), Peter replaces (1) *there was* with *what has happened since then*, (2) *w-* with *what has happened since then*, and (3) *is* with *has*. He signals each replacement by displaying the replacing words adjacent to, or JUXTAPOSED AGAINST, his display of the original words. What Peter means by these juxtapositions is this: "Give the juxtaposed display PRECEDENCE over a final continuous part of the original display." This way speakers can use replacements to substitute new elements for old ones, as in (22), or to add or delete elements (see H. Clark 1996: 264).

Repetition. In spontaneous speech, people often repeat words like *the*, *in*, *if*, and *I*, as in (23):

(23) Reynard: well, I {uh} I wouldn't be surprised at that, (1.1.278)

Treated literally, Reynard's utterance, like Peter's in (22), is ungrammatical: *I I wouldn't be surprised at that*. And yet he produces the two tokens of *I* as part of what he wanted to say at the moment. Repeating *I* is a violation of Grice's Maxim of Manner, namely "Be brief (avoid unnecessary prolixity)," so why does Reynard do it?

The two tokens of repeated words serve distinct purposes (H. Clark and Wasow 1998). The first token is often, though not always, used to make a PRELIMINARY COMMITMENT to what speakers are about to say but cannot yet produce. Reynard produces the first *I* to signal that he intends to produce a clause that begins with *I*. The second token is used to RESTORE CONTINUITY, or fluency, to the current phrase or clause. With the second *I*, Reynard restores continuity to the clause he has just committed himself to (*I wouldn't be surprised at that*). So the second *I* is a type of repair, but what Reynard is repairing is the continuity of his display of the clause.

Overlap. Speakers can do a variety of collateral chores by producing a display juxtaposed against what another speaker is currently displaying (see H. Clark 1996: 278–82). Consider the overlap in the last line of this exchange:

(24) Wendy: and as long as I'm in my own {–} little nit, and nobody's telling me what to do,
Ken: yes,
Wendy: there doesn't really seem *anything*
Ken: *but how* long do you think it'll take then, {–} to finish?

Ken times his display "but how long . . ." to begin before Wendy is finished, juxtaposing his primary utterance against hers. He does this, apparently, to ask Wendy for the floor, and one word later she accedes to his request. That is, he uses the juxtaposition of his utterance against hers as a STRATEGIC INTERRUPTION, a collateral signal for requesting the floor. Another type of juxtaposition, the RECYCLED TURN BEGINNING, is used by a second speaker to claim the next turn (Schegloff 1987).

4.4 Concomitants

When people talk face to face, they rely not only on speech, but on gestures – manual, facial, ocular, postural, and vocal gestures. When June asked David, "Where is your car?" and he pointed at the nearby Honda, he used his gesture as a primary signal – as part of their official business. But speakers also use gestures for collateral signals, performing them at the same time as, or in parallel with, the primary utterances. In this case, gesture and speech are

concurrent, even coordinated, but not composite signals. Collateral signals of this type will be called CONCOMITANTS.

Gestures that depict what they refer to are called ICONIC GESTURES (see, e.g., Schegloff 1984, McNeill 1992). Most are parts of composite signals in primary utterances. In the following example (from Kendon 1980), Fran is telling a friend about a scene from the Billy Wilder movie *Some Like it Hot*:

(25) Fran: they wheel a big table in, with a big with a big [1.08 sec] cake on it, and the girl, jumps up.

Fran uses one gesture to depict the height and movement of the table, the gesture peaking at the word *table*. She uses a second gesture to depict the size and shape of the cake during "[1.08 sec]," and she uses a third to depict the vertical movement of the girl jumping during *girl*. Her phrase *a big cake* plus her circular gesture depicting the cake form a composite signal – a unified description of a cake of a particular size and shape.

Other iconic gestures are used instead to refer to features of performance. One example is what Goodwin and Goodwin (1986; Goodwin 1987) called a THINKING FACE:

(26) Arthur: He pu:t uhm, (0.7) tch! Put *crab*meat on th'bo:dum

Arthur apparently has trouble finding the word *crabmeat*. Beginning at *uhm* and ending at *tch!* he turns from his addressee and, with a distant look in his eyes, puts on a stereotyped facial gesture of someone thinking hard – a thinking face. The thinking face is an iconic gesture speakers use (a) to assert that they are thinking hard, hence (b) to implicate that they are searching for the next word. In (26) Arthur ended the thinking face when he went on to say *crabmeat*. In other examples, speakers end the thinking face by turning to gaze at their addressees to request help in finding the elusive word. The thinking face and the return gaze are both concomitant signals.

Collateral gestures like this appear to fall into four broad classes (see also Cassell and McNeill 1991, Bavelas et al. 1992, Bavelas 1994):[9]

- DELIVERY GESTURES, as when speakers "hand over" to addressees new information relevant to their main point;
- CITING GESTURES, as when speakers point at addressees to indicate "as you said earlier;"
- SEEKING GESTURES, as when speakers gaze at addresses as if to say "Can you give me the word for . . . ?"
- TURN GESTURES, as when speakers "hand over" the turn to addressees.

These gestures take many forms. Some are so-called EMBLEMS (e.g. head nods, head shakes, thinking faces); others are indicative gestures (eye gaze, pointing); and still others are iconic gestures (smiles). What is common to them all

is that they are displayed concurrently with the primary utterances and deal with issues of performance.

5 Summary

People use language primarily to coordinate activities they are doing jointly – from assembling TV stands to negotiating contracts. They use both linguistic and non-linguistic signals to reach joint commitments on who is participating, in what roles, with what content, and at what time and location. These signals constitute the primary system of communication. But using language is itself a joint activity and requires the same joint commitments. The signals used in that process constitute the collateral system of communication.

The collateral system relies on two main devices. The first is displays – the time, place, and manner with which a signal is performed. Speakers design each display to indicate the producer, recipient, content, timing, and location of the signal. But for speakers to be sure their addressees have identified who and what they are indicating, they ordinarily work with their addressees to ground each signal. The second main device is collateral signals, which refer to the local, ongoing performance. In spontaneous speech, people add these to their primary signals in the form of inserts, modifications, juxtapositions, and concomitants. It is these that give the language of cafés, classrooms, and offices its spontaneous character.

What, then, about novels, plays, and news broadcasts? Writers, actors, and radio announcers do indeed display their utterances, but their displays are limited, and they have no way of grounding what they say. Nor do they have recourse to most collateral signals. They can only hope and pray that they will succeed. So, for language in general, we must go to cafés, classrooms, and offices, for it is there that we find the primary and collateral systems of communication in their fullest form.

ACKNOWLEDGMENTS

I am grateful to the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, for their hospitality while I worked on the issues in this chapter. I am especially indebted to David P. Wilkins for many discussions. I also thank Adrian Bangerter, Eve V. Clark, Laurence Horn, Teenie Matlock, and Gregory Ward for constructive comments.

NOTES

1 Most of the examples in this chapter are taken from the London–Lund corpus (Svartvik and Quirk 1980). In the notation retained here, a period (".") marks a brief pause, a dash ("–") a unit pause, a colon (":") a prolongation, a question mark ("?") an intonation unit ending in a rising intonation, and a comma (",") an intonation unit ending in any other intonation. The examples are numbered by conversation (e.g. 1.2) and line number (e.g. 370).

2 Many in pragmatics have since turned to spontaneous language for evidence, but others reject it as a proper subject for pragmatic theories.

3 I am indebted to Julie Heiser and Barbara Tversky for videotapes of this example.

4 They are often called BACK CHANNEL RESPONSES (Yngve 1970), but that term includes other phenomena too.

5 Most (but not all) collateral signals may be described as metalinguistic, though not all metalinguistic techniques are collateral signals.

6 The same goes for INSTALLMENT UTTERANCES, as in the transfer of telephone numbers, addresses, and instructions (Goldberg 1975, H. Clark and Schaefer 1987, 1989), and for certain LEFT DISLOCATIONS (Geluykens 1987, 1988, 1992).

7 In most dialects of English, when *the* is cliticized onto a word that begins with a vowel, it is an unstressed version of *thiy* with a reduced vowel. So *the egg* is pronounced *thi.yegg* (where the syllable boundary is marked by a period). The version of *thiy* at issue here has a non-reduced stressed vowel, as in the trochee *thi.yum* (Fox Tree and H. Clark 1997).

8 In Chafe's transcripts, prolongations are marked with dashes, and silences are given in seconds in parentheses.

9 Bavelas and her colleagues called these INTERACTIVE GESTURES, which they contrasted with TOPICAL GESTURES, which are composite parts of primary utterances.