

9 | Utterances

Utterances are the most tangible products of language use. They have traditionally been treated as autonomous acts by speakers, but that isn't right. Although speakers may assume the major responsibility, they cannot present utterances without the coordination of their addressees. When Connie presents an utterance to Duncan, she is trying to get him to identify her words, constructions, and gestures, and that takes his actions too. Getting utterances attended to and identified is just as much a joint action as getting them understood.

We also tend to think of utterances as single, linear threads of talk. Connie begins with an intention, formulates an utterance piece by piece, and presents each piece to Duncan as she completes it. In reality, utterances are nonlinear and have more than one track. Connie may produce an expression, change her mind, and start over. She may make a mistake and repair it. She may realize she didn't have Duncan's attention and repeat what she said. She may interrupt herself to explain a pause or disfluency, or to apologize for a gaffe. Duncan may add his own bits to her utterance. These phenomena aren't rare. In spontaneous utterances, speakers and addressees have many intricate issues to manage.

In the last two chapters we looked at how people carry out joint projects and establish what the speaker is to be taken as meaning, the top two rungs of this ladder (Chapter 5):

Speaker A's actions	Addressee B's actions
4 A is proposing joint project w to B	B is considering A's proposal of w
3 A is signaling that p for B	B is recognizing that p from A
2 A is presenting signal s to B	B is identifying signal s from A
1 A is executing behavior t for B	B is attending to behavior t from A

In this chapter we will see how people manage utterances, and for that we will look at the bottom two rungs – level 2, presentation and identification, and level 1, execution and attention.

Presentations

In every communicative act, speakers present signals for their addressees to identify. Roger, for example, presents the utterance “now, - um do you and your husband have a j- car” in order to ask Nina whether she and her husband have a car. To complete his question (at level 3), he has to get Nina to identify his presentation (at level 2), but doesn’t quite manage the first time around:

Roger: now, - um do you and your husband have a j- car
 Nina: - have a car?
 Roger: yeah
 Nina: no -

Apparently, Nina isn’t sure of the last phrase, “have a car,” so she asks Roger to confirm it. Only once that identification is complete (at level 2) does she understand the question (level 3) and take it up with “no” (level 4). The joint process of presenting and identifying signals is essential to communicative acts.

Every use of a word, phrase, or sentence has an *ideal delivery* – a flawless presentation in the given situation (Clark and Clark, 1977). It is flawless in that it is fluent, and the pronunciation, intonation, speed, and volume are appropriate to the circumstances. It is the delivery speakers would make if they had formulated what they were going to say before speaking and could follow through on that plan. Something close to an ideal delivery is produced by radio and television announcers and stage actors. But as Roger’s utterance illustrates, speakers in conversation achieve ideal deliveries only in spurts. As hard as they try, they cannot sustain fluency for any length of time.

The problem is human limitations. Roger commits to speaking with “now,” but then has to pause to formulate the next word. Later, he begins a word “j-” but cuts himself off and uses the word “car” instead. Speakers have trouble deciding on, formulating, and articulating what they want to say and that interferes with their ideal delivery. To account for these problems and their repair, we must return to the notion of tracks of talk.

PRIMARY AND SECONDARY PRESENTATIONS

As we saw in Chapter 8, signals divide into tracks 1 and 2. Signals in track 1 are addressed to official business, and those in track 2, to the communicative acts themselves. The side sequence in Roger and Nina's exchange was an example of a collateral project:

Nina: - have a car?

Roger: yeah

Nina's question, "Was the phrase you just presented 'have a car'?" is about Roger's communicative act, not about the official business of their conversation, so this exchange belongs to track 2.

Roger's gross presentation "now, - um do you and your husband have a j- car" also divides into two tracks. Roger's "now, do you and your husband have a j- car?" is in track 1 because it is "official" or "for the record." The status of Roger's "j-", however, changes midutterance when he replaces it by "car" and then considers it "*no longer* for the record." In contrast, Roger's "um" is never "official" or "for the record" in this sense. It deals with the communicative act, so it belongs to track 2. And, finally, there are incidental elements in the gross presentation that are not intended. One may be the pause after "well," since Roger deals with it by saying "um."

We can therefore partition a gross presentation into four distinct classes of elements, as follows:

Primary presentation. Those parts of the gross presentation that the speaker intends as signals about the official business of the discourse (e.g., Roger's "now, do you and your husband have a j- car"). These belong to track 1.

Secondary presentation. Those parts of the gross presentation that the speaker intends as signals about the communicative acts themselves (e.g., Roger's "um"). These belong to track 2.

Ex-official elements. Those elements of the primary presentation that the speaker later intends to be preempted by other elements (e.g., Roger's "j-"). These were part of track 1 when they were produced.

Incidental elements. Those elements of the gross presentation that the speaker doesn't intend to be part of any communicative acts (e.g., Roger's pause after "well").

Speakers design presentations so their addressees can distinguish among these four types of elements. If they didn't, their addressees couldn't understand.

GESTURES

Gross presentations contain nonlinguistic as well as linguistic signals (Chapter 6). Speakers gesture with their voice, hands, arms, face, eyes, and body in order to describe, indicate, and demonstrate an indefinitely wide range of objects, events, states, and properties. Gestures divide into primary, secondary, and ex-official elements as well.

The commonest gestures in track 1 are indicative gestures, iconic gestures, and beats. Many demonstrative references aren't complete without indicative gestures, as in this example (Schegloff, 1984, p. 280):

Frank: why:nchu put that 't the end uh the ta:ble there

Frank refers to the dish "that" in front of Marge without a gesture, but refers to the end of the table "there" by pointing at it. Without the gesture, he couldn't get Marge to recognize where precisely on the table he meant. Indicative gestures are parts of other communicative acts too, as here (Schegloff, 1984, p. 284):

Linda: en I'm getting a sun tan

As Linda says this, she points first at her left cheek and then at her right, asserting she is getting a tan on her cheeks even though she never mentions her cheeks verbally. These, too, are part of her primary presentation.

So are most iconic gestures and beats. Recall the joke Fran told about the movie *Some Like it Hot* (Chapter 6). In the course of that joke, she says:

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

In the pause before "cake," Fran produces an iconic gesture, depicting the cake as large and round. Her addressees would have missed part of what she meant if they didn't understand that. Beats, in their purest form, reinforce major stresses in a presentation, as here (Schegloff, 1984, p. 273):

Sam: I mean it's like Eddie says, (1.0 sec) as time goes on it gets worse 'n worse 'n worse 'n worse

Sam makes a thrusting gesture on the word *time* and on each of the four instances of *worse*. He is apparently emphasizing these five words to help mark them as new information, making the beats part of track 1.

Speakers use descriptive, indicative, and iconic gestures in track 2 as well (see Chapter 8). They use indicative gestures to designate their current addressee. The commonest gestures are eye gaze or body orientation, as Charles Goodwin (1981) noted in this utterance:

Elsie: See first we were gonna have [gazing at Ann] Teema, Carrie and Clara, (0.2) a::nd myself. [gazing at Bessie] The four of us. The four [gazing at Clara] children. But then – uh:: I said how is that gonna look.

As Elsie gazes successively at Ann, Bessie, and Clara, she repeats and expands on her references, designing “Teema, Carrie, and Clara,” for Ann, “the four of us” for Bessie, and “the four children” for Clara. She tailors each reference for the particular woman she is addressing, indicating that woman by her gaze. Gestures like these are indispensable to indicating the addressees for particular utterances (see also Goodwin, 1986b, 1987).

For an example of iconic gestures in track 2, consider what Marjorie Harness Goodwin and Charles Goodwin (1986; Goodwin, 1987) have called the *thinking face* in this presentation:

A: He pu:t uhm, (0.7) tch! Put crabmeat on th'bo:dum.

Beginning at “uhm” and ending at “tch!” A turns away from the addressee with a distant look in his eyes in a stereotyped facial gesture of someone thinking hard. Speakers use the thinking face to signal that they are doing a word search and to account for why they aren’t proceeding with their utterance. In this example, A ended his thinking face when he was able to proceed to the word *crabmeat*. Other times, speakers end the thinking face by turning to their addressees for help in finding the word they were searching for.

The commonest gestures in track 2, according to Bavelas (1994; Bavelas et al., 1992; see also Cassell and McNeill, 1991), fall into these main categories:¹

¹ Bavelas and her colleagues called these *interactive gestures*, which contrast with *topical gestures* (part of the primary presentation).

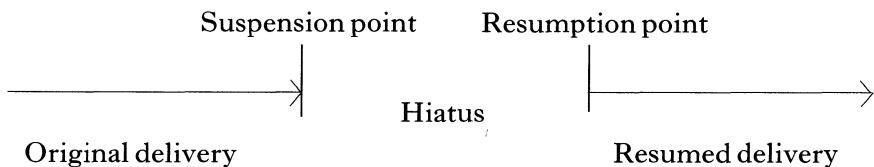
Type of gesture	Function of gesture	Example
Delivery gesture	to refer to the delivery of information by speaker to addressee	speaker "hands over" to addressee new information relevant to main point
Citing gesture	to refer to a previous contribution by addressee	speaker points at addressee to indicate "as you said earlier"
Seeking gesture	to elicit a specific response from addressee	speaker looks at addressee as if to say "Can you give me the word for...?"
Turn gesture	to deal with issues around the speaking turn	speaker "hands over" the turn to addressee

Among gestures in track 2, we find symbols (e.g., head nods, head shakes), indexes (eye gaze, pointing, nodding), and icons (thinking face, smiling, frowning). The full range of gestures is exploited in both primary and secondary presentations.

Gross presentations, then, divide into primary and secondary signals. To see how and why, we must examine how speakers and addressees manage presentations that are less than ideal.

Disruptions

When speakers cannot manage an ideal delivery, they usually *disrupt* their presentations. Most disruptions can be divided into three intervals whose boundaries are fixed by two points (Levelt; 1983, 1989; Shriberg, 1994). I will call this the *disruption schema*:



The five parts of the schema are defined as follows (illustrated for Roger's "j-"):

1. *Original delivery*. This is the smooth delivery before the disruption (here, "have a j-").
2. *Suspension*. This is the first outward sign of the disruption, the point at which the speaker suspends the original delivery (here, in the middle of the word beginning "j"). I will denote the suspension point with a left curly bracket {.

3. *Hiatus.* This is the interval from the point of suspension to the point at which the speaker resumes the primary presentation, here the interval between “j-” and “car.” In some cases, the time between suspension and resumption is zero (Blackmer and Mitton, 1991). Other times, it may contain such elements as “uh” or “I mean.” Whatever its length, I will call it a hiatus.²
4. *Resumption.* This is the point at which the speaker resumes the official presentation, here the beginning of the word “car.” I will denote the resumption point with a right curly bracket }.
5. *Resumed delivery.* This is the delivery after the resumption.

In my notation, Roger’s presentation is this: “do you and your husband have a j- {} car.”

Although disruptions are usually associated with disfluencies, they need not be. Speakers can suspend their delivery for any purpose whatever – to laugh, to take a sip of coffee, to clear one’s throat. Here is a good example (1.3.1):

Nancy: I'm not {- oh, thanks, .} not really comfortable, {} like this

When Nancy’s companion offers her tea, she suspends her current utterance to accept it.

Disruptions often come in spates. Here are two spontaneous utterances marked with the notation just introduced (1.2.370; 1.2.787):

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

well, they've accepted th- {uh} this as a commitment {-} because it's a {} it's a {} it's a m {} he's done a {} he's on a Ph.D. {- you see}

In the first example, the seven words after “well” encompass two disruptions, each with several parts. The hiatus in “well {. I mean} this” contains a pause and an editing phrase (“I mean”). The hiatus in “this {. uh} Mallet” contains a pause and a filler (“uh”). Despite the disruptions, these two examples posed no problems for the addressees. The reason lies in the way the speakers and addressees managed them. In each disruption, the speakers added signals in track 2 telling the addressees exactly what they were doing and why. To see how, let us examine the three parts of the disruption schema – the suspension, the hiatus, and the resumption.

SUSPENSION

The point of suspension is usually clear. Speakers cease their presentation at a point where, in the ideal delivery, they should not be ceasing.

² Shriberg (1994) calls this interval the *interregnum*.

This is really what defines the point of suspension. The cessation of speech is often accompanied by one or more *suspension devices*. Here are five common devices (1.1.819; 8.2a.335; 1.12.661; 1.2.229; 1.1.1164):

Suspension device	Example
Pause	there's another {-} fight I've got on my hands
Word cut-off	have a j- {} car
Elongation	and thuh: {} the parish church is beginning to go boing boing,
Nonreduction	come to look at thiy {.} the literature
Filler	and then you'll get {uh} difficulties about marking

1. *Pause*. The simplest accompaniment to cessation is a pause. It is often the only sign indicating the point of suspension.

2. *Word cut-off*. Speakers can deliberately suspend their delivery in the middle of a word. They often do this with an audible closure of the glottis, so that listeners hear not only the cessation of speech, but the glottal stop.

3. *Syllable elongation*. Sometimes speakers elongate the syllable just before the suspension, as in this example:

and **thuh:** {} the parish church is beginning to go boing boing,

In the ideal this speaker would have delivered “and the parish church is beginning ...” producing *the* only once with the reduced vowel schwa. Instead, he pronounced *the* with a lengthened schwa (as marked by the colon), suspended his delivery, and, on resuming, repeated *the* with a normal length. In repeated “the”s, the first token tends to be longer than normal, and the second, normal (Shriberg, 1994).

4. *Nonreduced vowel*. In the ideal delivery, many unaccented syllables are pronounced with reduced vowels. If you deliver “The dog found a bone” fluently, you will produce both *the* and *a* with the reduced schwa “uh.” But speakers sometimes mark suspensions by producing vowels *without* reducing them. They pronounce *the* and *a* as “thiy” and “ai” (rhyming with *me* and *day*), as in this example:

come to look at **thiy** {.} the literature

In the London–Lund corpus, 81 percent of the tokens of “thiy” were at points of suspension, whereas only 7 percent of the tokens of “thuh” were. Clearly, “thiy” can be used for signaling points of suspension (Fox Tree and Clark, 1994). Likewise, speakers may produce *to*, ordinarily pronounced “tuh,” as a nonreduced “too,” and the same for other

schwas. Note that producing a nonreduced vowel (“thiy”) is not the same as lengthening a vowel (“thuh:”).

5. *Fillers*. Another way to mark suspensions is by placing a vocal or visual element in the hiatus. Two common elements in English hiatuses are “uh” and “um,” but there may also be throat clearing, laughter, and gestures.

Suspension devices aren’t produced accidentally. They are the result of the speakers’ own actions – they are *self*-suspensions – and are signs in Peirce’s sense (Chapter 6). But Peirce’s signs can be either signals (actions by which speakers mean something for their addressees) or symptoms (such as spots for measles). At least some suspension devices are signals.

Take “thiy.” Whenever speakers use “thiy,” they choose it over “thuh.” They cannot pronounce *the* as “thiy” simply by lengthening the vowel, as if they were accidentally slowed down in speaking. Their choice of “thiy” over “thuh” is like any other lexical choice – for example, *the* over *a*, or *in* over *at*. It is part of the signal they have planned. What they signal with their choice of “thiy” is not official business (track 1), but a suspension of the presentation of the utterance (track 2). With it they mean, “I am suspending the primary presentation at this point for a moment.” That makes the vowel choice in “thiy” a collateral signal.

The logic here is based on a *principle of choice*: Whenever speakers have more than one option for part of a signal and choose one of the options, they must mean something by that choice, and the choice is a signal. By this logic, the word cut-off is a signal: Speakers could have chosen to complete the word as formulated. To cut it off is to signal they have changed their minds about it. The same argument applies to elongation, nonreduction, and suspensions by fillers.

What about mere cessation, as in this example:

there's another {-} fight I've got on my hands

On one view, the cessation after “another” and the pause “-” aren’t signals, but symptoms – *unintended* consequences of not yet having formulated the next piece of the utterance. On another view, this speaker chose to begin “there’s another” *before* he had formulated that next piece. He could have waited until he had formulated the entire clause, but he chose not to. At some level of planning, he intended to suspend his delivery after “another” and to pause until he could continue. By this logic, cessations and pauses are signals. It is as if speakers were saying: “I am

suspending the primary presentation now because I haven't yet formulated the next piece of the utterance." There are strengths and weaknesses in both of these views. For now, I will be conservative and treat mere cessation as a symptom – an unintended sign. Still, I will place pauses in track 2, for whatever their status, they do not belong to the primary presentation.

HIATUS

The hiatus in fluent speech – the interval between the suspension and resumption points – is often filled with more than silence. Here are six common types of content (8.2a. 335; 1.1.819; 1.1.1164; 1.3.253; 2.7.692; Goodwin and Goodwin, 1986):

Contents	Example
No pause	have a j- {} car
Pause	there's another {-} fight I've got on my hands
Filler	and then you'll get {uh} difficulties about marking
Editing expression	I'd {I mean} I haven't had any {.} results
Elongation	I think he th {o:} he'd realized then
Iconic gesture	He pu:t {uhm (0.7) [thinking face] tch} put crabmeat on th'bo:dum

Hiatuses may contain nothing, silence, sound, or gestures. They often contain more than one type of content – pauses, fillers, editing phrases, elongations, gestures. Which of these elements are signals, and if they are signals, what do speakers mean by them?

Editing expressions are clearly signals (DuBois, 1974; Erman, 1987; Hockett, 1967; Holmes, 1986; James, 1972, 1973; Levelt, 1983, 1989; Schiffrin, 1987). Here are a few examples with interpretations of some of their uses:

Editing expression	Interpretation of some uses
I mean	"Instead of that I mean this"
you know	"I am about to tell you something you already know"
sorry	"I'm sorry for that error"
oh	"I have just thought of a way of expressing this"
like	"I mean this only approximately"
no	"That was in error"
that is	"I am specifying that further in this way"

All these expressions make reference to the communicative acts they are placed within. Take the example with "I mean":

I'd {I mean} I haven't had any { } results,

What the speaker means by "I mean" is roughly this: "Instead of what I would have meant from 'I'd' onward, I mean 'I haven't had any results.'" In each of these interpretations, the terms "that" and "this" refer to particular preceding and following acts. So editing expressions are genuine signals in track 2.

Fillers are also genuine signals in track 2. In British and North American English, speakers have a choice between at least *uh* and *um*, and even between normal and elongated versions of these expressions. These also contrast with the simple pause. What are they used for? In a study by Smith and Clark (1993), respondents were interviewed in a conversational setting and asked a series of factual questions that each had one-word answers. Here is a typical exchange (the numbers in parentheses denote pause lengths in seconds):

Experimenter: In which sport is the Stanley Cup awarded?

Respondent: {(1.4) um (1.0)} hockey

When the respondents didn't have the answer on the tip of their tongue, they would delay, sometimes introducing *uh* or *um* into the hiatus. The average time to answer was 2.23 seconds when they didn't use a filler, 2.65 seconds when they used *uh*, and 8.83 seconds when they used *um*. The respondents used *uh* to signal that they were having a minor problem and *um* to signal they were having a major one. Speakers recorded in the London–Lund corpus also distinguished between *uh* and *um*.³ They were much more likely to pause after *um* than after *uh* – 55 percent to 20 percent of the time. They too used *uh* to signal a minor break and *um* a major one.

By the principle of choice, elongated vowels are also signals in track 2. Consider this example:

I think he th{o:} he'd realized then

Here the speaker began saying "he thought" but suspended his delivery midword and restarted again with "he'd realized." In suspending his delivery, the speaker had these choices, among others:

1. full word + pause: thought {-}
2. word fragment + glottal stop + pause: tho- {-}
3. word fragment + elongated vowel + no pause: th{o:}

³ H. H. Clark and J. E. Fox Tree, unpublished study.

In choosing the elongated vowel, the speaker therefore meant something – roughly, “I am suspending the primary presentation for a brief time while I formulate a repair for this word.”

Hiatuses aren’t just idle moments in speaking. They may contain a variety of actions, many of which are signals in track 2.

RESUMPTION

When speakers resume speaking after the hiatus, they present further official content of their utterances, more signals in track 1. But what is the relation of the resumed delivery to the original delivery? Here are the five basic patterns (1.1.819; 1.2.1304; 1.2.80; 1.2.231; 1.1.748):

Type		Example
Continuation	there's another	{-} fight I've got on my hands
Repetition	surely, <u>your</u>	{ } your committee is not going to come to that conclusion
Substitution	what <u>is</u>	{ } has happened since then
Deletion	I don't think they' <u>ve</u>	{ } they ever in fact embodied
Addition	one of the things that	{- uh} one of the many things

In the first pattern, the speaker continues the utterance he had suspended. In the next three, the speakers repeat, substitute for, or delete one or more of the original words, and in the fifth, the speaker adds a word to the original. The last three patterns are often called *self-repairs*.

All these resumptions reflect a single operation I will call *replacement*: The resumed delivery is intended to replace – take precedence over – a continuous portion of what has been produced up to the point of resumption. Here is what got replaced in the five examples:

Type	Elements replaced	Resumed delivery
Continuation	{-}	fight I've got on my hands
Repetition	your{ }	your committee is not going to come to that conclusion
Substitution	is{.}	has happened since then
Deletion	they've{.}	they ever in fact embodied
Addition	one of the things that{- uh}	one of the many things

A continuation replaces only the hiatus; the rest replace the hiatus plus some final portion of the original delivery. Note that to delete a word or other element (e.g., *'ve*), speakers must return to an earlier point (*they*) to show what is to be deleted. And to insert a word (e.g., *many*), they must

show where it is to be inserted (before *things*). Replacement is a single method by which speakers can perform all five functions.

Resumptions, therefore, work at the level of presentation and identification (level 2). Replacements are simply *re-presentations* of what the speakers, on reconsideration, would prefer to have presented in the first place.⁴ Take an utterance examined earlier:

Type of resumption	Example
repetition	well, they've accepted th- {uh}
continuation	this as a commitment {-}
repetition	because it's a {}
repetition	it's a {}
substitution	it's a m {}
substitution	he's done a {}
	he's on a Ph.D. {-you see}

Although Sam changed his mind six times and for three different reasons (repetition, continuation, substitution), he signaled each change in the same way: He returned to a previous point (the onset of boldface elements) and started over. If we delete the boldface elements, we get what Sam, afterward, would prefer to have presented: “Well, they’ve accepted this as a commitment because he’s on a Ph.D.” In most resumptions, the continuity is not just in wording, but in prosody. If we excise the boldface elements in a *recording* of Sam’s utterance, we get an utterance that is prosodically well formed, that sounds natural (Fox Tree, 1995; Levelt, 1984).

Replacements, especially with prosodic continuity, are optimal for both speakers and addressees. When speakers resume speaking, they must indicate what they are doing, and replacements offer a simple, uniform method. It is easy to return to a previous point in the formulation and start again. As for addressees, they must be able quickly to identify what is being repaired. Replacements allow them to match the syntax and prosody of the resumed delivery against a portion of the original delivery with great precision. Not all repairs are achieved by replacement (Van Wijk and Kempen, 1987), but almost all resumptions are.

“To replace,” however, doesn’t mean “to obliterate.” When the one speaker says “what is {} has happened since then,” he intends “has happened...” merely to *take precedence* over “is” in the original delivery. The point is illustrated here (1.3.305):

⁴ Although not always. See Polanyi (1978).

Nancy: {um ---} the interview, was {-} it was all right

After replacement, Nancy says “it was all right.” But “it” is “the interview.” To see this, Nancy’s addressees must appeal to the reference in what she has just replaced. So replacements take precedence over the elements they replace, but without obliterating them.

Resumptions, therefore, don’t substitute, delete, or add elements per se. They *re-present* elements that speakers, on second thought, would prefer to have presented in the first place. It just happens that re-presenting these elements has the effect of continuing, repeating, substituting for, deleting, or adding to portions of the original delivery.

Strategies of presentation

Primary presentations should and do have a privileged status. The most striking evidence of this is the ubiquity of replacement. Almost all presentations are continuous and well-formed once we do all the replacements. From “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,” we get “well, Mallet felt it would be a good thing if Oscar went.” We have seen some of the devices for creating such presentations, but when and why do speakers choose the devices they do? The answer lies in the joint nature of language use.

Speakers’ actions in talk aren’t independent of their addressees’ actions, or vice versa, and that goes for their problems as well. When speakers need extra time to plan an utterance, that isn’t their problem alone. The time they need belongs to them and their addressees together, so they have to coordinate with their addressees on the use of that time. Or when speakers present the wrong word or phrase, that problem isn’t theirs alone either. It belongs to them both, and it takes the two of them together to fix it. Most problems in using language are *joint* problems, and dealing with them requires *joint* management (Clark, 1994).

TWO IMPERATIVES

Time is inexorable, irreversible. We cannot turn the clock back, erasing a time interval and claiming that it didn’t occur. Whatever we do or don’t do in a time interval is irrevocable, and we must deal with the consequences. In joint actions, time is doubly important. When Roger and Nina talk to each other, what he does in his time is also done in her time, and vice versa. That places them under the following constraint:

The temporal imperative. In a joint action, the participants must provide a public account for the passage of time in their individual parts of that action.

Because of this imperative, participants try to act in a timely fashion. They are also subject to a second imperative:

The formulation imperative. Speakers cannot present an expression before they have formulated it.

Although the formulation imperative is a truism, it is an important truism. In order to present a signal of any kind, speakers must formulate it first, and that may prevent them from acting in a timely fashion.

Speakers are continuously pressed by these two imperatives. So long as they speak with an ideal delivery, they are providing the best possible account of what they are doing with their time: They are doing their part in the joint action of presenting and identifying an utterance. They should strive for the ideal delivery for this reason alone. On the other hand, formulating a presentation is a fitful, cyclic process, and speakers rarely have a presentation entirely formulated before they start speaking. They generally have only a vague plan and formulate one short phrase at a time (Bock and Levelt, 1994; Levelt, 1989).

Once speakers embark on a presentation, they place themselves on a tight schedule. At each point they have two choices: (1) continue speaking, or (2) stop speaking. The ideal is to continue—*prima facie* evidence they are doing their part. The non-ideal choice is to pause. When they pause, they no longer have a public justification for their actions. They may still be busy formulating; they may have aborted their presentation; they may be reconsidering what they have just presented; they may have been distracted. Stopping midutterance is *not* part of the joint action. In short, continuing to speak doesn't need an explanation, but stopping does. The preference for continuing over stopping leads to three broad strategies.

STOP-AND-CONTINUE STRATEGIES

Speakers can pause in an ideal delivery, but only briefly at certain phrase boundaries. One unit of presentation is the *intonation unit*, a stretch of talk spanned by a single prosodic contour.⁵ Each line from this narrative by Nancy is an intonation unit (1.3.100):

⁵ These units have been variously named *tone groups* (Halliday, 1967), *tone units* (Crystal, 1969, Crystal and Davy, 1975; Svartvik and Quirk, 1980), *intonation groups* (Cruttenden, 1986), *intonation units* (Chafe, 1992), *information blocks* (Grimes, 1975), *idea units* (Chafe, 1979, 1980), and *lines* (Gee, 1986).

well when I was { } doing freelance advertising,
 - thiy: {} advertising agency,
 that I { } sometimes did some work for,
 . rang me,
 and said {um - } we've got a client,
 who wants {um - - } a leaflet designed,
 . to go to s- {uh} instructions how to use a sewing-machine,

There are pauses before three intonation units, yet they are heard not as disruptions, but as parts of Nancy's official presentation. In narratives studied by Wallace Chafe (1979, 1980, 1992), there were pauses before 88 percent of the intonation units, and they averaged about 1 second in length.⁶ Speakers apparently pause at these boundaries to help formulate the next intonation unit.

The six hesitations that arise mid-unit in Nancy's narrative, however, *are* heard as disruptions. The first five of these are *pure hesitations*: They are followed by continuations, the simplest form of resumption. Hesitations take two main forms – pauses (as in four of Nancy's disruptions) and elongated syllables (as in Nancy's "thiy:"). Both give speakers extra time mid-unit to formulate what to say next. Elongations give the added illusion of fluency, as if the speakers weren't truly disrupted. Speakers may choose elongations over pauses when they think the disruption will be brief.

There is a limit to how long speakers can pause without discomfort. According to Gail Jefferson (1989), who examined more than a thousand examples, speakers tend to limit initial pauses midutterance to about one second. They tend to resume talking, produce a filler, or clear their throat – something to signal they will continue – after only one second of silence. It isn't surprising that there is such a limit if speakers are pressed by the temporal imperative. What is remarkable how little silence is tolerated. Let us call this the *one-second limit*.⁷

If midutterance silences over a second are a problem, speakers must monitor for this possibility. Nancy's first three hesitations don't cause problems because she resolves them within one second, but the next three do. Apparently, anticipating a hiatus longer than one second, she produces

⁶ Compare Boomer's (1965) average for juncture pauses of 1.03 sec.

⁷ Jefferson (1989) called this the "standard maximum tolerance." Other cultures may tolerate much longer silences or only shorter silences. Jefferson's data seem to come from urban North American and British speakers of English.

“um.” She does so not to help her formulate the following words more quickly, but to tell her audience what she is doing and help them deal with the disruption (Brennan and Williams, 1995; Smith and Clark, 1993).

Even between intonation units, silences become a problem when they grow too large. Take this exchange:

- Reynard: i-{} is {} is it this year, that {u:h} Nightingale goes
 Sam: {}-u:h} no next year,
 Reynard: {}-u:m . -} sixty-f-{}
 Sam: sixty-five
 Reynard: -four {} sixty-five
 Sam: yeah

When Sam is asked “Is it this year that Nightingale goes?” he must respond. If he delays too long, he may be misconstrued as unable or unwilling to answer, as waiting for more of the question, or as distracted. To show he is preparing a response, he cuts off the silence with “u:h” (in track 2). Reynard does much the same with “u:m.”

In the stop-and-continue strategy, therefore, speakers stop until they have formulated the next element and then continue. Pure hesitations are by far the commonest disfluency. But if their hesitations are too long, speakers need to justify them with fillers, editing expressions, and other signals.

COMMIT-AND-REPEAT STRATEGY

Another way speakers can justify a stoppage is to present the first word or words of the phrase they are formulating even though they cannot present the rest of it. Suppose Duncan is trying to formulate “he may be qualified.” Once he decides it begins with “he,” he can present “he” and stop. This way he shows he is committed to presenting a phrase beginning with *he*. But that raises a secondary problem. If he simply continues after the interruption, as in “he {[pause]} may be qualified,” the phrase won’t be fluent or easy to identify. The solution is to repeat “he,” as in this example (1.2a.985):

- Duncan: {u:m .} and that he {} he {uh} he may be qualified to {} to be recognized
 as a teacher of French

The *commit-and-repeat strategy* accomplishes two things. The first “he” provides advanced evidence that Duncan is committed to a constituent beginning with “he.” The last “he” allows for the fluent delivery of the phrase “he may be qualified.”

With the commit-and-repeat strategy, repeats should be common on the first words of major constituents, and they are. Compare nominative versus accusative pronouns – e.g., *he* versus *him*. *He* is almost always the first element in a larger constituent, such as the clause “*he* may be qualified,” but *him* is usually its own constituent, as in “everyone saw *him*.” In a large sample of telephone conversations studied by Thomas Wasow and myself,⁸ nominative pronouns were repeated 5.2 percent of the time, and accusative pronouns, only 0.04 percent of the time – a difference of more than 100 to 1.

The constituent being planned can also be large or small. Consider the most inclusive phrase initiated by the noun phrases in these four roles:

Role of noun phrase	Example	Percentage of “the”’s repeated
Topic	the van that we’ve got the gentleman who owned it had died	7.9
Subject	the dog I have is a German shepherd	5.0
Object of verb	I managed to find the stereo I wanted in Austin	3.4
Object of preposition	my wife parked her car in the garage across from our house	1.8

On average, topics initiate the longest major constituent, subjects the next longest, objects of verbs the next, and objects of prepositions the shortest. In the same sample of telephone conversations, the longer this constituent, the more speakers repeated *the*, the initial word in the noun phrase. The noun phrase initiated by *the* may itself be simple or complex. For example, *the stereo I wanted* is complex because it has a modification after the head noun *stereo*, whereas *the stereo* is simple. In the same telephone conversations, speakers repeated *the* more often in complex than in simple noun phrases, 4.1 to 2.7 percent of the time. In general, constituents tend to begin with function words (pronouns, prepositions, articles, auxiliary verbs, etc.) and not content words (nouns, verbs, adjectives, adverbs), and function words were ten times as likely to be repeated as content words.⁹

The commit-and-repeat strategy, therefore, has two main advan-

⁸ H. H. Clark and T. Wasow, unpublished data.

⁹ H. H. Clark and T. Wasow, unpublished data.

tages. It provides early evidence of what speakers are doing, and it results in fluent constituents, and indeed, that helps addressees (Fox Tree, 1995). Except for pure hesitations, repeated words are by far the commonest disfluency in spontaneous speech.

COMMIT-AND-REPAIR STRATEGIES

Speakers often change their minds about what they are presenting. As Willem Levelt (1989) has argued, they monitor their presentations for these (among other) problems (1.2.33; 1.2.787; 1.13.246; 2.13.1204; 2.8.304):

Question	Example of repair
1. Is this the message or concept I want to express now?	we must ha- {} we're {} big enough to stand on our own feet now
2. Is this the way I want to say it?	he's done a {} he's on a Ph.D.
3. Am I making a lexical error?	if she'd been {} he'd been alive,
4. Are my syntax and my morphology all right?	he think E- {} thinks Ella's worried about something
5. Am I making a sound-form error?	everything is mitch {} much more complex

And when they find something they want to change, they repair it. The element to be repaired (marked here in boldface) is the *reparandum*. In principle, speakers could avoid repairs if they took enough time before speaking, but in practice, repairs are inevitable. In conversation at least, speakers can never anticipate everything they might change their mind about. They are forced to proceed by a *commit-and-repair* strategy.

Although speakers should be able simply to replace the reparanda and then proceed, they often replace entire stretches of the original presentation in addition to the reparanda. We can distinguish four types of resumption:

1. *Instant replacements*. In these, the speaker replaces the reparandum and nothing more, as here (1.14a.124):

have I ever **tel-** {} talked to you about Cookstown County Tyrone?

2. *Trailing replacements*. When speakers don't suspend their delivery immediately after the reparandum, they need to replace the reparandum plus the trailing elements, as here (3.4.707; 1.13.246; 1.2.29):

- to buy any more sites, {} **in the college**, {} for {} the college
- if **she'd been** {} he'd been alive
- **we're not prepared, to go on being part**, {} I'm not prepared to go on being part of Yiddish literature

In these examples, the trailing elements are “the college,” “’d been,” and “not prepared to go on being part.” In the operation of replacement, trailing replacements are obligatory.

3. *Anticipatory replacements*. Speakers often replace not only the reparandum, but elements before it as well. Sometimes the process is obligatory, as in these examples:

- he **think E-** {} thinks Ella's worried about something
- everything is **mitch** {} much more complex

The first speaker couldn’t add “-s” to “think” without re-presenting “thinks,” and the same goes for the second. More often, the process is optional, as here (1.1.750):

this is {} this is **one of the things that {- uh}** one of the many things, {- uh} in English structure, which is {- u:m -- } an item in a closed system

4. *Fresh starts*. Speakers sometimes abandon an entire presentation and make a fresh start, as here (1.2.33):

we must ha- {} we're {} big enough to stand on our own feet now

In fresh starts, it is hard to single out particular elements that are being repaired. The entire fragment is simply abandoned.

Each resumption is a signal in track 1, just as each original presentation is: It is part of the official presentation. But the speakers’ *choice* of resumption is also a signal in track 2: It tells addressees what the speakers are replacing and, often, why. Speakers indicate what they are replacing, as Levelt (1989) has argued, by two main strategies.

1. *Word-identity*. When the first word of the resumption is identical to a recent word in the original presentation, the resumption is to replace everything from that word on. This works for repeats and anticipatory replacements:

- and that **he {} he {uh}** he may be qualified to {} to be recognized as a teacher of French
- this is **one of the things that {- uh}** one of the many things, {- uh} in English structure, which is (etc.)

Addressees can identify what is to be replaced as beginning with “he” and “one of the.”

2. *Category-identity*. When the first word of the resumption is a member of the same category as a recent word in the original presentation, the resumption is to replace everything from that word on. This is illustrated in these earlier examples:

- if **she'd** been {} he'd been alive
- to buy any more sites, {} **in** the college, {} **for** {} the college

In the first example, “she” and “he” are both personal pronouns, so the resumption is to replace everything from “she” on. In the second example, “in” and “for” are both prepositions, with an analogous result. Although fresh starts don’t always adhere to these two strategies, most other replacements do. For speakers working under pressure, that is a remarkable feat.

EDITING EXPRESSIONS

By the temporal imperative, speakers in trouble need to account for their actions. When Roger makes an error, Nina is entitled to ask, “Why did he make an error? Does he realize he has made it? Is he going to repair it?” Even when he merely suspends his presentation, she can ask, “Why is he stopping? Will he resume and, if so, when? What should I expect next?” One way speakers provide these accounts is with editing expressions produced in track 2. We have already seen how *uh* and *um* are used to signal length of hiatus. Other expressions have more specific uses.

Editing expressions like *no*, *or rather*, *I mean*, and *that is* are used for characterizing the trouble a speaker is in and its relation to the repair. With *no*, speakers characterize the reparandum as “incorrect”, and with *or rather*, they compare it with the new expression to be offered. Each term is appropriate for a different type of trouble or repair. For example, the Dutch equivalents of *no* and *or rather*, which describe something as incorrect, are used for error repairs and not appropriateness repairs (Levett, 1984), and the English expression *you know* is reserved for appropriateness repairs (Clark and Gerrig, 1990).

Editing expressions like *well*, *oh*, *ah*, *aha*, and *let me see*, in contrast, comment on the *source* of the speaker’s troubles (Heritage, 1984; James, 1972, 1973; Schourup, 1982). With each, speakers disclose their current thinking about what they are about to say, as when Elizabeth is talking to Ned (1.6.232):

Elizabeth: he {} I think he thinks it's all a little bit {uh: - **well**,} stupid but{uh:, .}
[continues]

With “well,” Elizabeth tells Ned, in track 2, that she is consulting her thoughts about how best to express the next trait. She accounts for the hiatus and implies she considered other (perhaps weaker) adjectives.

Still other editing expressions address interpersonal problems

caused by disfluencies. Speakers may apologize for an error, as when a radio announcer said (Goffman, 1981b, p. 291):

**And now, Van Cliburn playing Tchaikovsky's Piano Concerto Number One in
Blee Fat Minor...I beg your pardon, that should be Fee Blat Minor**

They may explain why they aren't as incompetent as they appear to be, as when another announcer said (Goffman, 1981, p. 294):

**Stay tuned for Aeolia where they will be reading – if you wait a moment I'll be
able to tell you...here it is...**

They may even make light of the error, as this disc jockey did (Goffman, 1981b, p. 299):

**We hear now a song from the new Columbia album featuring Very Jail...Oops, I
ought to be in jail for that slip...of course, I mean Jerry Vale!**

As Goffman suggested, pauses and errors can make speakers look incompetent, and speakers can use editing expressions to mitigate the damage.

Speakers work hard, then, to deal with the temporal and formulation imperatives. They cannot speak until they have something formulated, but they must also provide a public account of what they are doing. The optimal strategy is to produce a fluent, ideal delivery. When that isn't possible, speakers commonly resort to three strategies: stop and then continue (with an explanation, if necessary); commit to the beginning of a phrase and then repeat it; and commit to an expression and then repair it if necessary. Whenever speakers pause beyond the one-second limit, they say why they are pausing. And they use editing expressions to help addressees prepare for and identify the repairs they make. By explaining the trouble, they help addressees understand what they intend to say and mitigate any interpersonal damage they might have caused.

Execution and attention

In order to signal anything, speakers must execute audible and visible behaviors for addressees to attend to. In the action ladder, success at level 2, presentation and identification, depends on success at level 1, execution and attention, just as success at level 3, meaning and understanding, depends on success at level 2. What does it take to succeed at level 1? Roger may produce the sounds "now, - um do you and your husband have a j- car" yet not get Nina to attend to them. Roger and Nina face the *attention problem*: how to coordinate his articulating those sounds with her attending to them in the right way. Coordination of attention is taken

for granted in most theories of language use, but in fact isn't easy to achieve. People have a battery of techniques for achieving it, techniques that shape language use at its very core.

THE ATTENTION PROBLEM

Speech is evanescent – it fades immediately. If Roger is to succeed in telling Nina something, he must make sure she is trying to attend to his sounds *at the very instant* he is articulating them. Executing behaviors to be attended to and attending to those behaviors, then, are participatory acts: Roger cannot do his part without Nina doing hers, and vice versa.

What is attention? If we think of people as attending to a strand of events over a period of time, attention has three notable properties:

1. *Selectivity*. People can attend to only one level of one strand of events at a time.
2. *Redirectability*. People can redirect their attention to a second level or strand of events very quickly, often within milliseconds.
3. *Vulnerability*. Attention to one strand of events is fragile and easily captured by another strand of events.

At an orchestra concert, I can attend to the music as a whole, or to the violins alone, but not to both at once. Still, I can shift attention from one to the other so quickly that, if neither strand is too complicated, I can keep track of both at once. I can also redirect my attention from the music to what I am doing tomorrow. Yet, my attention to any of these strands is fragile. If my neighbor coughs, my attention may shift to the coughing despite my best efforts. Orienting reflexes shift our attention automatically, so certain events appear to capture our attention regardless of our intentions.

These properties are easy to demonstrate in listening to speech (Clark and Clark, 1977). If you listen through earphones to two people speaking at once, you can shadow – repeat immediately – what one of them is saying. But soon you cannot report whether the unattended voice is male or female, or speaking English or French, or even that it has repeated your own name twenty-five times. This is selective attention. Still, you can switch your attention to the other voice when you choose to. And if the voices are too similar – both male, both speaking English, both coming from right in front of you – it becomes more difficult to shadow either one of them, to maintain selective attention. When you are attending to one voice, it is also harder to carry out other tasks.

People implicitly take these properties into account in monitoring each other's states of attention. If Ann and Bob are the current speaker and addressee, Ann must monitor Bob's state of attention as she executes her utterance, and Bob must help her do that. They recognize both positive and negative evidence of B's attention to A:

Evidence that B is attending to A's execution now

B is gazing at A.

Evidence that B is not attending to A's execution now

B is doing something that takes too much competing attention.

B is making a primary presentation himself.

B is attending to another speaker.

B is trying to think about or do something else.

The situation may potentially interfere with B's attention.

B isn't in a position to hear or see A's execution.

A nearby event is so loud or bright that B cannot hear or see A's execution.

A nearby event is likely to capture too much of B's attention.

And this list is hardly complete.

Face to face, listeners generally signal attentiveness with eye gaze (see Argyle and Cook, 1976). While Bob is listening to Ann speak, he gazes at her much of the time, apparently to signal that he is attending to her (Kendon, 1967). She gazes back at him, but for much shorter periods, apparently to check whether he is already gazing – to acknowledge his signals (Goodwin, 1981; Kendon, 1967). As Goodwin (1981) argued, the preferred condition in gazing is this: "When speaker's gaze reaches a recipient, that recipient should be gazing at the speaker" (p. 76). These signals – gazing and acknowledging gazes – are collateral to the official business at hand, so they are in track 2.

Eye gaze ordinarily provides valid, economical, and timely evidence of a person's full attention (see Chapter 8). While Bob is gazing at Ann, he cannot be looking at something else, and if he is also silent, he isn't overtly engaged in a competing activity.¹⁰ Not only is his gaze valid evidence, but it is cheap to provide and available precisely when it is most useful to Ann. But Bob's unilateral gaze isn't enough – Ann must acknowledge it. Mutual gaze is just what the two of them need as a shared basis for the mutual belief that he is attending to her at that moment.

¹⁰ He could be thinking of something else, of course, but people appear to believe they can distinguish attentive from inattentive gazes.

ONE PRIMARY SPOKEN PRESENTATION AT A TIME

People in conversation do more than monitor each other's attention: They anticipate interference and try to work around it. If one of them coughs or pounds a hammer, the noise could interfere with hearing, so they time their actions – both their speech and the interfering actions – so as not to overlap (Jefferson, Sacks, and Schegloff, 1987). Likewise, public speakers time their oratory to be clear of their audience's applause or laughter, and audiences accommodate their applause and laughter to the speaker's talk (Heritage, 1984). Accidental overlaps are repaired too, as here (1.1.531):

Sam: well, {uh}*I put* {} I put the linguistic jargon in, Reynard
 Reynard: *(coughs)*

The ideal is to talk clear of serious auditory interference – to speak *in the clear*.

The commonest source of interference in conversation is other talk – overlapping speech. When Ann considers starting an utterance while Bob is speaking, she should consider these factors:

- | | |
|--------------------------|--|
| 1. Auditory interference | A's and B's utterances may obscure each other auditorily. |
| 2. Conflict of attention | A and B may be unable to give each other their full attention while they are deciding on, formulating, and presenting their own utterances. |
| 3. Distraction | A and B may be distracted by the other's utterance and make errors in their own presentations. |
| 4. Shift of attention | Still, A and B may each be able to shift their attention to the other's utterance – if one of the two presentations is short enough or easy enough to process. |

Ann should conclude that she can present an overlapping utterance if it takes Bob's attention only briefly. That would allow secondary presentations – head nods, gestures, “uh huh”s, utterance completions – but not, ordinarily, primary spoken presentations.

Speakers, therefore, generally observe a *one-primary-spoken-presentation-at-a-time* limit during conversations. To do this, they need to manage who speaks when, and the result is an emergent system of turn taking, which I will return to in Chapter 11. To avoid overlaps in primary speech, they try to project the end of the current speaker's presentation before starting their own, and they are extraordinarily precise in doing

this (Jefferson, 1973; Sacks, Schegloff, and Jefferson, 1974). For most turns, the current speaker begins his or her official presentation with (1) no overlap with the previous speaker, (2) overlaps of only one or two syllables, or (3) very brief pauses (Beattie, 1983; Beattie and Barnard, 1979; Schegloff et al., 1977). All three are heard as smooth transitions between speakers – as no pause or overlap. On the other hand, speakers are perfectly happy for secondary presentations such as “uh huh” to overlap with a primary spoken presentation. So overlap is common, but mostly of track 2 presentations with track 1 presentations.

Accidental overlaps in primary presentations are often considered problems, as in this example (1.3.222):

Nancy: I mean this whole system, of being invited somewhere for lunch, and
then for dinner, {-} and overnight, {} *and breakfast,*
Nigel: *oh you st- {}* you {} you did stay

Nigel projected Nancy’s utterance to end after “overnight” and started to present “oh you stayed.” Once he discovered his error, he stopped and, omitting “oh,” repeated his primary presentation “you {} you did stay” in the clear. Speakers use much the same technique when two people initiate turns simultaneously.

So speakers tacitly recognize how much they can overlap and still be attended to, identified, and understood. Here is a rough ordering of speech overlaps from least to most interfering:

1. There is no overlap; A's primary presentation is in the clear.
2. A's primary presentation is partly overlapped by B's secondary presentation.
3. The end of A's primary presentation is overlapped with pre-placed expressions like “well” and “in other words” of B's primary presentation (Schegloff, 1987).
4. A's primary presentation is partly or entirely overlapped by B's primary presentation.

Overlapping speech should be tolerated only so long as both parts can be attended to well enough for current purposes.

GAINING ATTENTION

At level 1, Ann and Bob must go through an attention cycle for each presentation. This cycle has the entry–body–exit format of all joint actions:

Entry. A must get B's attention in advance.

Body. A must hold B's attention throughout the presentation.

Exit. A may relinquish B's attention afterwards.

For entry into this cycle, Ann needs to get Bob's attention. She can get it by requesting it, by capturing it, or even by making it impossible for him not to attend. Here are five common techniques:

1. *Summons*. Ann can initiate a summons–answer exchange with Bob – a minimal joint project – in this way (Schegloff, 1968; Schegloff and Sacks, 1973):

Ann: Bob?

Bob: Yes?

With “Bob,” she designates Bob as the person whose attention she wants, and with her rising intonation, she asks him to respond. This way she requests his attention. Bob takes her up by responding “yes” and, with his rising intonation, turns the conversation back to her. Vocatives like “Bob” are useful when gaze is ineffective. When I enter a house, I can yell “Sam” to request Sam’s attention if he’s there. In a large lecture, I can address one of the students, “Mr. Kaplan,” to request his attention in particular. Summons can also take the form of telephone rings, doorbells, and whistles.

2. *Turn restarts*. When Ann starts speaking to Bob, she will ordinarily check to see whether he is gazing back. If he isn’t, she must secure his attention. One technique for doing that, as Goodwin (1981) argued, is the turn restart, as in this example (p. 61):

Track 1	Track 2
Lee:	Can you bring (0.2)
Ray:	*[starts to turn head]*
Lee: *Can* you bring me here that nylon	

Lee starts in on “Can you bring,” but merely to request Ray’s attention. Once Ray gazes back, Lee restarts his presentation to be sure Ray hears it from the beginning. On Goodwin’s videotape one can see Lee’s restart begin precisely as Ray begins turning his head. Lee’s initial fragment is a natural way of requesting Ray’s attention, yet it is up to Ray to signal that he is actually attending. Both signals are in track 2.

3. *Mid-turn delays*. Ann can also request Bob’s attention by delaying mid-presentation, as in this example from Goodwin’s videotapes (p. 76):

Track 1	Track 2
Barbara:	uh
Barbara: my kids	*(0.8 sec pause)*
Ethel:	*[starts to turn head]*
Barbara: had all these blankets, and quilts and sleeping bags.	

Barbara starts her presentation and then pauses, all to request Ethel's attention. By the start of the 0.8-second pause, Ethel starts turning her head toward Barbara, and by its end, she is gazing at Barbara. Unlike Lee, Barbara doesn't restart her presentation. Speakers tend to choose mid-turn delays when they haven't yet gazed at their addressees, but turn restarts when they have (Goodwin, 1981).

4. *Recycled turn beginnings.* Speakers can also request the other participants' attention before the previous speaker has completed his or her presentation. To do this, they need a special technique, as illustrated here (Schegloff, 1987, pp. 80-81):

- A: Yeah my mother asked me. I says I dunno. I haven't heard from her. I didn't know what days you had *classes or anything.*
- B: ***Yeah an I didn't know*** I didn't know when you were home or - I was gonna.

As Schegloff argued, even though B projects the end of A's turn, he starts his presentation early in order to claim the right to speak next. Since B realizes that he cannot be attended to fully – the one-primary-spoken-presentation-at-a-time limit – he recycles the beginning of his turn to present it in the clear. B is remarkably precise at starting his full recycle in the clear. Also, he drops “yeah” in the recycled presentation perhaps because it demands so little attention or carries so little information. So with recycled turn beginnings, speakers request their addressees' attention and yet articulate their entire presentation in the clear.

5. *Strategic interruptions.* Speakers can also exploit the one-primary-spoken-presentation-at-a-time limit with strategic interruptions, as here (1.9.804):

- 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 them to finish?

In the last line, Ken initiates a primary presentation – marked by “but how” – before Wendy can finish hers. Since the two of them cannot make primary presentations at the same time, one has to give way, and Wendy does. Indeed, this is the goal of Ken's strategy. He tacitly reasons: “She and I cannot make primary presentations at the same time. So by beginning mine in the middle of hers, I am signaling that she should stop because what I have to say is more important now than what she has to say.” Wendy accedes, though she needn't have.

RETAINING ATTENTION

In the second step in the attention cycle, speakers need to keep their addressees' attention on their execution. Ann must hold Bob's attention throughout her presentation, and Bob needs to reassure her of his continued attention. How do they do this?

Once Ann has got Bob's attention and begun a presentation, she can ordinarily assume he will continue to attend until she is finished. So it is important for her to keep him informed of her progress: Is she continuing, or is she done? We have already seen three broad strategies for letting him know: stop-and-continue, commit-and-repeat, and commit-and-repair. With all three, she signals that she is still working on a piece of her presentation and intends to continue. And Bob signals that he is still attending by (1) continuing to gaze back, (2) not initiating an official presentation, and (3) not performing an action that competes for his attention.

Still, these signals aren't always clear. One problem is projecting the end of a presentation. In an earlier example, Calvin misprojected the end of Nancy's utterance and began on his own presentation "oh you stayed." He resolved the competing attention by stopping. Other times the second speaker doesn't abandon his or her new presentation, and if the first speaker wants to continue, he or she has to take positive actions to retain the addressee's attention. Here, British prime minister Margaret Thatcher is interrupted by a television interviewer, Dennis Tuohy (Beattie, 1983, p. 137):

Thatcher: ...there are comparatively few people they could be measured in
thousands who wish to destroy the kind of society which you and I
value destroy the free society
***Please, please this is the most**
please this is the most
please this is*

Tuohy: ***You were talking about striking ambulance workers**
you were talking about ancillary workers in hospitals*

Thatcher: the most important point you have raised there are people in this
country who are the great destroyers.

Tuohy misprojected the end of Thatcher's response and initiated his next question before she was finished. But when Thatcher was interrupted, she recycled "please this is the most" three times before she got Tuohy's full attention again. Just as determinedly, Tuohy recycled his

utterance once before giving up. Both recognized that the other – and the viewing audience – couldn’t be attending fully to both primary presentations. They each recycled to force the other to abandon their presentation and to attend to the other.

RELINQUISHING ATTENTION

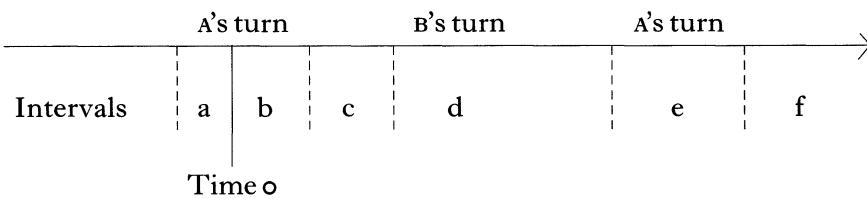
At the end of a presentation, speakers can relinquish their addressees' attention. But to complete the presentation phase of a contribution is to be prepared for the acceptance phase, which is ordinarily initiated by the addressees. So speakers often let their addressees know they are relinquishing their attention by turning their own attention to their addressees.

One technique is to use eye gaze. Recall Goodwin's preferred condition for gazing: "When speaker's gaze reaches a recipient, that recipient should be gazing at the speaker." When Ann wants to show Bob that she is done with her presentation and ready to attend to his, she can turn her gaze to him. This is what people generally do (Kendon, 1967). Other techniques are to use an intonation appropriate for the end of an utterance, or deliberately *not* to rush into the next utterance to signal an intention to continue. All these techniques make room for the next speakers to execute their presentations in the clear.

Executing a presentation and attending to that execution, then, takes continuous coordination. Ann and Bob need to coordinate their entry, continuation, and exit from these actions. For her part, Ann needs valid, economical, and timely evidence that Bob is attending to what she is producing, and he needs to attend to her and give her evidence that he is doing so. They manage this coordination by exchanging gazes, managing sources of interference, and making repairs when coordination goes wrong. Actions at the level of execution and attention are just as much joint actions as those at higher levels of talk.

MONITORING AND COMPLETION

When does an utterance become complete? That depends on the level. We can divide a conversation into intervals relative to the moment of speech (time 0). This diagram depicts turns by A, then B, then A, with six intervals labeled *a* through *f*:



At time o , during Ann's presentation, Ann is trying to get Bob to attend to her execution (level 1), identify her presentation (level 2), understand what she means (level 3), and consider her proposed joint project (level 4). To reach closure on each of these joint actions, the two of them need the right evidence, and that becomes available in different intervals for the four levels.

AVAILABILITY OF EVIDENCE

The evidence Ann and Bob need for closure can be divided initially into *self-* and *other-evidence*. Self-evidence comes from monitoring oneself – Ann monitoring her actions in speaking, and Bob monitoring his mental states in reception. Ann and Bob also monitor for other-evidence, and that takes coordination. Bob must present evidence about his mental states at just those moments he believes Ann is monitoring for it, and Ann, knowing this, should monitor him closely at these points.

At level 1, both self- and other-evidence tend to be available *continuously*. Ann listens to herself speak and sees herself gesture in interval b , precisely as Bob is attending to her behavior. Ann is able even in interval a to monitor a pre-spoken version of what she is about to deliver (Levelt, 1989). As for other-evidence, Bob shows his attention by gazing at Ann, and she returns his gaze, all in interval b . So at level 1 Ann and Bob ordinarily reach joint closure almost immediately.

At levels 2, 3, and 4, in contrast, both self- and other-evidence become available only *periodically*. At level 2, self-evidence is available in interval b , but only a word or phrase at a time. Bob can be certain of Ann's presentation only after major phrases, after all her replacements, and she cannot be certain of his identification until he has nodded, smiled, said "uh huh," or given other evidence, sometimes in interval b , but often in interval c or d . At levels 3 and 4, self-evidence is available only as Ann completes larger units in interval b – phrases or entire sentences. But other-evidence usually isn't available to Bob until interval c , or to Ann until interval d , when Bob takes up what Ann has proposed.

To gather up these points, self-evidence is generally available before other-evidence. And the higher the level, the later either type of evidence is available. Ann and Bob ordinarily reach joint closure at level 1 in interval *b*, at level 2 in interval *c* or *d*, at level 3 in interval *c* or *d*, and at level 4 in interval *d* but sometimes not until interval *e* or *f*. That is, the higher the level, the later the closure.

REPAIRS

According to the principle of opportunistic closure (Chapter 8), Ann and Bob should consider an action complete at the first opportunity – at the first evidence of completion. On the same grounds, they also should consider an action in need of repair at the first evidence of failure, as expressed in this principle:

Principle of repair. When agents detect a problem serious enough to warrant a repair, they try to initiate and repair the problem at the first opportunity after detecting it.

By this principle, repairs should be initiated and completed as soon as possible. Just how soon depends on the availability of evidence and the opportunities for initiating and for making a repair.

Most repairs are probably invisible. Ann can detect and correct a problem in her actions without Bob ever knowing; likewise, Bob can detect and correct a problem in his reception without *Ann* ever knowing. These problems are private, and so are the repairs. Although the repairs are invisible, they are in line with the principle of repair because they are initiated as soon as a problem is detected.

Whenever a problem becomes public, however, it becomes a *joint* problem. Public problems are joint problems for two reasons. First, it is often impossible to identify who in the conversation is responsible for them. The point is illustrated in the exchange between Roger and Nina:

Roger: now, - um do you and your husband have a j- car
 Nina: - have a car?
 Roger: yeah
 Nina: no -

Who is at fault for the problem repaired with “- have a car?” “yeah”? Is Roger to blame for a muddy pronunciation, or is Nina to blame for not listening closely enough? The source of the problem is often indeter-

minate.¹¹ Second, regardless of source, public problems require joint solutions. Nina and Roger have to act jointly even to repair problems that only one of them is really responsible for. Roger, for example, repairs the problem caused by the pause after “now” by using “um” to tell Nina that he is still formulating the next word. Public repairs are joint actions (Clark, 1994).

Public repairs are part and parcel of the process of joint closure. As discussed in Chapters 7, 8, and 9, they characteristically occur in different intervals for levels 1, 2, 3, and 4:

Intervals	a	b	c	d	e	f
Level 4				[— — —]		
Level 3			[— — —]			
Level 2		[— —]				
Level 1	[— —]					

Repairs like these are simply part of the joint actions Ann and Bob carry out at each of the four levels. It is impossible to reach joint closure without the availability of repair.

People in conversation, in sum, are opportunistic in trying to reach closure on their actions. They try to repair problems as quickly and as efficiently as possible. Doing this, however, isn’t easy. It requires the participants to monitor both themselves and their partners at all levels of action – from execution and attention upward – and to be prepared to initiate and make repairs at the first opportunity.

Conclusions

Utterances are often viewed as the prerogative of speakers – products that speakers formulate and produce on their own. Nothing could be further from the truth. Uttering things involves two levels of joint actions. At level 1, speakers execute certain behaviors – vocalizations and gesticulations – for addressees to attend to. At level 2, speakers present signals for

¹¹ According to Schegloff et al. (1977), repairs can be classified by who initiates them, as *self-* or *other-initiated* repairs, and by who makes them, as *self-* or *other-corrections*. For this classification to work, one must be able to identify the source of the problem, and for repairs like Nina’s and Roger’s, that seems impossible. If we attribute the problem to Roger, the repair is *other-initiated* and *self-repaired*. If we attribute it to Nina, it is *self-initiated* and *other-repaired*. If we view it as a joint problem, or as indeterminate, as it seems proper to do, the terms don’t apply at all.

addressees to identify. There can be no communication without tight coordination at both levels.

Speakers and addressees have a battery of strategies for coordinating at these two levels of action, strategies that exploit signals in track 2. At level 1, the main issue is how to establish, hold, and relinquish the addressees' attention to the speakers' vocalizations and gesticulations, and one of the most useful signals is eye gaze. At level 2, the main issue is how to deal with disruptions in the speakers' delivery, and speakers use special signals to mark suspensions in delivery, to account for the disruptions, and to indicate what is to be replaced. Utterances are truly products of speakers and addressees acting jointly.