ELSEVIER

Speech Communication 36 (2002) 5–13

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Speaking in time ☆

Herbert H. Clark

Department of Psychology, Stanford University, Building 420, Jordan Hall, Stanford, CA 94305-2130, USA

Received 28 January 2000; received in revised form 28 September 2000; accepted 28 September 2000

Abstract

Most spoken disfluencies, it is argued, are not problems in speaking, but the *solutions* to problems in speaking. Speakers design most forms of disfluencies as signals, communicative acts, for coordinating with their addressees on certain of their speech actions. At the lowest level, speakers try to synchronize their vocalizations with their addressees' attention. At the next level up, they try to synchronize, or pace, the presentation of each expression with their addressees' analysis of those expressions. Speakers have a variety of strategies for achieving synchronization, and many of these lead to the common forms of disfluencies. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Disfluencies; Spontaneous speech; Timing; Pauses; Production

1. Introduction

Why are speakers disfluent? One common answer is that they have trouble planning their utterances – selecting what to say and how to express what they say. But is this the right way to put it? Take Reynard's utterance in (1): [1]

(1) 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).

This utterance is full of disfluencies – repeats (''if u:h if''), repairs (''Mallet said Mallet was''), fillers (''uh''), prolonged syllables (''u:h''), and editing expressions (''I mean'', ''you know''). And these appear to reflect Reynard's difficulties in deciding what to say and how to say it. Still, Reynard's speech reflects more than just those difficulties. He designs each disfluency so that Peter, his addressee, can eventually identify what he really wanted to say, namely this:

(1′) Reynard well, Mallet said he felt it would be a good thing if Oscar went

Although Reynard took his time in getting the utterance out, he planned it perfectly well.

Reynard's problem is not in planning *per se*, but in planning *in time*: Speakers are disfluent when they cannot plan within the time limits expected of them. Why is time so important? The ultimate answer, I will argue, is that speakers must *synchronize* certain of their own processes with those of their addressees, or communication

☆ This research was supported by NSF Grants SBR-9309612 and IRI-9314967. I thank Marc G.J. Swerts and Jacques M.B. Terken for constructive comments on the paper.

E-mail address: herb@psych.stanford.edu (H.H. Clark).

[1] All of the examples (but one) are taken from the London–Lund corpus (Svartvik and Quirk, 1980). Brief pauses are marked with a period (.), pauses of one stressed syllable by a dash (–), prolonged vowels with a colon (:), the ends of the tone units with a comma (,), and overlapping speech with paired pairs of asterisks (∗) in adjacent turns. The examples are numbered by the conversation (1.2) and line (370) in the corpus.

PII: S0167-6393(01)00022-X

will fail. To do that, they must attend closely to the timing of their own and their partner's speech and deal with timing when it goes awry. That makes timing central to spontaneous speech, and it leads to the special design of disfluencies.

Most spoken disfluencies can be partitioned into two parts: a problem and its solution. Take Reynard's "it would be a good thing if u:h . if Oscar went". By the time Reynard reached *thing*, he apparently had a problem: He did not know quite what he wanted to say next, or perhaps how to express what he wanted to say next. Although Peter, his addressee, could infer that problem, the problem itself was hidden – present but behind the scenes. What Peter actually heard was Reynard's *solution* to the problem. Although Reynard could simply have suspended speaking after *thing* and resumed when he was ready again, he did not. Instead, he did four things: (1) before suspending his speech, he produced *if* to commit himself to producing a conditional clause; (2) he produced *uh* to signal that he was delaying the resumption of his speech; (3) he prolonged *uh* to signal that he was continuing an ongoing delay; and (4) on resuming speech, he repeated *if* to restore continuity to the conditional clause. He took each of these actions, I will argue, to help Peter deal with the delay with the least effort. Most disfluencies are like Reynard's: The problems are hidden, and the forms we actually see are the solutions to those problems.

Using language is a joint activity. Like waltzing, playing a duet, or shaking hands, it requires people to coordinate their individual actions in order to succeed (Clark, 1996). Suppose Ann is speaking to Ben. They have to coordinate their actions at at least these four levels:

- *Level 1*: Ben must be attending to Ann's voice precisely as she is vocalizing.
- *Level 2*: Ben must be trying to identify the expressions that Ann is presenting while she is presenting them.
- *Level 3*: Ben must be trying to understand what Ann means as she speaks.
- *Level 4*: Ben must be considering the joint projects she is proposing with her utterance as she is proposing them.

Speech is evanescent. At level 1, if Ben is not attending precisely as Ann is speaking, he has no chance of recovering her vocalizations. At level 2, if Ben is not identifying Ann's expressions as she is presenting them, he will permanently lose their content. At level 3, Ben needs to understand what Ann means so far in order to help him identify what she is about to present. And, at level 4, Ben must be considering what Ann is proposing in order to plan his response by the end of her utterance. Ben needs to monitor all these levels if he is to ask for a repair whenever he does not hear, identify, or understand what she is saying.

Ben cannot succeed at these four levels without Ann's collusion. Ann has her *primary* signals, the linguistic devices by which she accomplishes the official business of the discourse. But for coordination, she also needs what I have called *collateral signals*. These are lexical, syntactic, prosodic, and gestural devices that help coordinate her primary signals. She needs to let Ben know at what moment she expects to vocalize, what she is about to present, what she intends to revise or abandon, and much much more. She lets him know all these things, I argue, by *telling* him with collateral signals. In this paper, I will take up just a few of these collateral signals.

2. Synchronizing voice and ear

One of the oldest principles in perception (see Woodworth, 1938) is the *readiness principle*: People can identify an object more quickly and accurately when they are *ready* for it. One factor in this readiness is time: People will be quicker and more accurate when the object appears at the precise moment they expect it to appear. Suppose, in a series of trials, you see a warning light, and then you hear a word that you must repeat aloud as quickly as possible. You will be faster and more accurate if the words appear at short, predictable intervals after the warning light than if they appear at longer, less predictable intervals. It is a matter of attention. You are temporarily confused by surprises, so you want to know when to attend. Let me call this the *on-time principle*.

How does the on-time principle play out in speaking? When Ann produces an expression, Ben will recognize it more quickly, and with greater certainty, if he can predict precisely when it will arrive. Most theories of production assume that one of Ann's implicit goals is to speak clearly enough for Ben. But if she is to meet that criterion, she will need more than clear articulation. She will need to coordinate her vocalization with Ben's attention. She will need to *synchronize* voice and ear. Speakers like Ann have at least four general strategies for achieving this

- *Strategy 1*: Signal the precise moment at which you intend to initiate an utterance.
- *Strategy 2*: In producing an utterance, produce each of its pieces at the precise moment they would be expected in an *ideal delivery* of that utterance.
- *Strategy 3*: If you must suspend speaking for any reason, signal your intent to suspend speaking.
- *Strategy 4*: If you need to delay vocalizing for any reason, signal your intention to delay and, if possible, how long you expect to delay.

There are many techniques for realizing each of these strategies. I will take up only some of them.

2.1. Strategy 1: Signal the initiation of speaking

In face-to-face conversation, speakers generally try to establish that their addressees are gazing at them before they initiate their utterances (Goodwin, 1981). Addressees, in turn, use gaze to signal that they are currently attending to the speakers. Once speakers have established mutual gaze, they can be confident that the addressees are attending as they initiate their speech. Speakers have a range of linguistic and non-linguistic techniques for getting their addressees' eye gaze in preparation for speaking (see Goodwin, 1981).

Speakers can also signal their intention to initiate speaking by using introductory *orienting expressions*. The idea is to *request* the addressees' attention and *signal* the intention to proceed. One type of orienting expression are conjunctions such as *and*, *but* and *so* (see Chafe, 1980). These are often articulated in separate intonation units (tone units in the London–Lund corpus), showing that they are to be treated as introductory signals for attention (see example 7, discussed later). Another type are discourse markers such as *well*, *oh*, *now* and *anyway*, as here:

(2) Bess	do you recognize this, . that's . Velasquez', Pope Innocent the fourth, . a copy of, –
Alan	are these *copies?*
Bess	**well* that's a copy, (1.4.431)

Bess introduces "well" to signal the start of her utterance even before Alan is finished. These, too, are often separated prosodically from the rest of the utterance.

Introductory orienting expressions and discourse markers do two things at once. Bess's "well", for example, is a primary signal that marks a change of stance in her contribution to the discourse. But its temporal placement in the conversation is a collateral signal by which Bess makes a commitment to initiate speaking *at that moment*. As it happens, "well" and their kind carry light enough content that they can be heard and understood without much attention. Alan could have missed the primary content of Bess's "well" without serious consequences, but not its collateral content. For strategy 1, it is the mere production of "well" at that moment that constitutes the collateral signal.

Another form of strategy 1 is to produce a pre-utterance filler – *uh* or *um* in English – to signal a delay in initiating the primary utterance. And another is to produce the first word of the utterance, stop, and repeat it. I will return to these strategies later.

2.2. Strategy 2: Pursue the ideal delivery

Speakers, as a rule, try to speak fluently. They try to produce each utterance with an *ideal delivery*, the way they would have wanted to produce it if they had no problems (Clark and Clark, 1977). The ideal delivery is what is characterized in standard prosodic theories.

Why pursue the ideal delivery? One reason is to help synchronize voice and ear. Consider this exchange:

(3) Roger:	now, - um do you and your husband have a *j- car*
Nina	- have a car?
Roger:	yeah
Nina:	no - (8.2a.335)

While Roger is producing "do you and your husband have a car", he realizes that Nina is expecting a nominal after "a" with 0 ms delay. She is expecting an ideal delivery unless he tells her otherwise. So when he completes the utterance with a word fragment plus nominal ("j- car"), he delays the nominal, throws off her timing, and makes *car* harder to hear and identify. The interference and delay force her to ask for a repeat – "have a car?" – which disrupts the conversation even further. So speakers are best off producing each word at its expected time – trying to achieve the ideal delivery – and the conventions of language use have evolved to reflect that advantage.

Speakers will try to produce the constituents of an utterance fluently even when they cannot produce the entire utterance fluently. Take this utterance:

(4) Kate	they had . *they shortlisted* five people, – including me, (1.3.255)

Kate suspends speaking after *had*, pauses, and then revises what she is about to say from (perhaps) "had five people on the shortlist" to "shortlisted five people". She could have made the repair with one less word, as in "they had . shortlisted five people", but that would have left a mangled clause. Instead, she restarts the clause and produces it fluently, as "they shortlisted five people". Speakers regularly restart constituents after speech disruptions, as in repairs (e.g., Levelt, 1983) and repeats (Clark and Wasow, 1998). Wasow and I have called the preference for producing constituents fluently the *continuity principle*, and there is broad evidence for it.

The logic behind strategy 2, therefore, is this. (1) Listeners must attend to what speakers say if they are to identify it. Identification requires attention. (2) Listeners can identify an expression more quickly and accurately when it arrives at the expected moment. This is the on-time principle. (3) So speakers should try to produce utterances, or at least constituents, with a predictable prosody – the ideal delivery. This is the continuity principle.

2.3. Strategy 3: Signal your intention to suspend speaking.

Although speakers pursue the ideal delivery, they rarely achieve it. Often, they are forced to suspend speaking in the middle of a constituent, as Kenneth does here in speaking to Ned:

(5) Kenneth	I would have to go down to *dhi* – film school, and talk to some of the people there, (1.10.1145.A)

Kenneth suspends speaking after "the" (here pronounced [dhi], rhyming with *tea*), pauses, and then resumes with "film school". By the on-time principle, suspensions like this should disrupt the synchronization of voice and ear. So it would be a great help if Kenneth could signal Ned that he intended to suspend speaking.

One common signal for suspending speech is the nonreduced vowel in function words, as in [ei] for *a*, [dhi] for *the*, and [tuw] for *to* (Fox Tree and Clark, 1997). The point is illustrated in (5). In an ideal delivery, Kenneth would have cliticized *the* onto *film* and produced "thuh.film" [2] (Selkirk, 1996). Instead, he produced "dhi" *not* cliticized onto "film", namely "dhi – film". He produced a *nonreduced* vowel where his addressee would expect a *reduced* one. As Fox Tree and I have argued, speakers use nonreduced vowels like this to signal that they are immediately suspending speech to deal with a production problem. Kenneth apparently recognized in advance that he was having trouble accessing the nominal *film school*, so he signaled a suspension after *the* by pronouncing it as [dhi]. In the London–Lund corpus, the full [dhi] [3] was followed by a suspension of speaking 81% of the time, whereas the reduced *the* was followed by a suspension only 6% of the time.

[2] I will use *uh* for schwas and the period as a syllable boundary. Otherwise, I will use ordinary orthography.

[3] This is *not* merely the reduced "the" before vowels.

Another common suspension signal is syllable prolongation. Note that nonreduced vowels are easy to prolong, as in [dhi:] or [tu:w]. So are other syllables, both in function and content words. Speakers appear to use such prolongations to mark suspensions.

2.4. Strategy 4: Signal your intention to delay and, if possible, for how long.

Speakers may or may not delay after they have suspended speaking. Consider (6):

(6) Desmond Hamlet um – – – starts, . uh as a noble soul, th- there's no doubt that . that Hamlet has got this nobility of soul, (3.5a.315)

Desmond suspends speaking after "Hamlet", "starts", "th-" and "that", but the delays after the suspensions range from nil after "th-" to quite long after "Hamlet". Although the suspensions alone should cause attention problems for Desmond's addressee, the variable delays should exacerbate those problems. By the on-time principle, they should desynchronize voice and ear. To reduce these problems, speakers ought to try to signal the anticipated delays and, if possible, their anticipated lengths.

The best known delay signals are *uh* and *um*, as in (6). What is less often appreciated is that *uh* and *um* signal contrasting delays (Clark, 1994; Clark and Fox Tree, in preparation). As Fox Tree and I have found, speakers introduce *uh* when they expect a brief delay, and *um* when they expect a longer one (as illustrated in (6)). We argue that they use *uh* to signal a minor delay in speaking, and *um* to signal a major delay.

Fillers can also be used as suspension signals, often in combination with nonreduced vowels, prolonged syllables, or both (Clark and Fox Tree, in preparation; Clark and Wasow, 1998). The technique is to cliticize *uh* or *um* onto the previous word, as in this recorded example from a radio interview:

(7) Robert th- there is a (0.2) *a uh* (0.5) a potential problem,

Although Robert pronounces the first and third "a" as schwas (as in "sofa"), he pronounces the second one as "e.yuh", cliticizing "uh" onto "ei" (rhyming with "day") and adjusting the syllables appropriately into "e" and "yuh". That is, Robert pronounces "a uh" as a single prosodic word. In spoken corpora, Fox Tree and I have found that cliticization like this is especially common on the introductory conjunctions *and*, *but* and *so*, but it is also found on function and content words mid-utterance (e.g. "vie.wuh" and "air.plane.suh").

Are "uh" and "um" truly cliticized onto the previous words? The best evidence comes from their combination with *a*, *the* and *to*. For *uh* to be cliticized onto *a*, *a* must be pronounced not as a schwa, but as [ei], and that leads to the prosodic word "e.yuh". Likewise, *the* plus *uh* goes to "dhi.yuh", and *to* plus *uh* goes to "tu.wuh". These forms are common in spontaneous conversation (Clark and Fox Tree, in preparation). The cliticization of *uh* and *um* onto prior words is genuine.

Are cliticized fillers truly *signals* of suspension? Yes. Robert cannot produce "e.yuh" in (7) without formulating the entire prosodic word as a piece. He cannot formulate "a", and then, once he has begun to execute it, decide to interject "uh". That would lead to two separate prosodic words, "a" followed by "uh", which I will write "a uh". Indeed, some speakers produce both the cliticized and the non-cliticized forms (e.g. "e.yuh" and "a uh") in the same conversation. So Robert must have planned to produce "e.yuh" and not merely "a". With that selection, he signals not only that he is producing an indefinite noun phrase (the primary meaning of *a*), but that he is immediately suspending his speech (the collateral signal).

We therefore have two distinct forms, "a uh" and "e.yuh", and speakers apparently use them in contrast. They use the *uh* in "a uh" to signal a delay, but they use the *uh* in "e.yuh" to signal both a suspension at the end of *a* and a delay. So "e.yuh" is much like "dhi". In their primary function, they mean "a" and "the". But at the same time, their collateral function is to signal an immediate suspension of speech, in which they contrast with "a uh" and "thuh uh".

Another delay signal is the mid-word cut-off, as in Robert's "th- there" in (6). Many mid-word cut-offs are accomplished with a glottal closure or voice creak, so they too are distinctive. As it happens, most suspensions in speaking occur between words (e.g. Levelt, 1983), so to suspend speaking mid-word is highly marked. It also happens that suspensions between words are far more likely to be followed by a pause or filler than are suspensions mid-word (Clark and Wasow, 1998). That is, speakers use mid-word cut-offs only when they intend to resume speaking without a delay. The hypothesis is that they use them to *signal* that they intend (a) to suspend speaking and (b) then to resume without delay.

3. Synchronizing diction and analysis

People like Ann and Ben must do more than synchronize voice and ear (level 1). Ann must present each expression in a way that allows Ben time to identify that expression as they go along (level 2). The two of them must, as I will put it, *synchronize diction and analysis*. The synchronization of diction and analysis has more leeway than the synchronization of voice and ear, but it is no less important. With diction and analysis, the major issue is *pacing*: Ann must engineer the content and timing of her presentation to allow Ben to keep up with his analysis. How do the two of them do that?

Again, we can turn to the readiness principle, but in another form: People can identify an object more quickly and accurately the more they know about it in advance – again, see (Woodworth, 1938). If you are asked to name an object as quickly as possible, you will be faster and more accurate the more you know about it ahead of time – for example, its category membership or its orientation. You will be slower and less accurate the more you are *misled* about its properties. Let me call this the *expected object principle*. The principle applies to the synchronization of diction and analysis in at least two ways. First, when Ann presents an expression, Ben should be faster at identifying it the more information she has correctly projected about it in advance. And second, whenever Ann decides to revise or abandon an expression she has already presented, Ben should be faster at reanalysis the sooner he knows (a) that she is revising or abandoning an expression and (b) which expression that is. Roughly speaking, Ben will be slower and less accurate the longer he spends analyzing *incorrect* information.

The expected object principle suggests at least three general strategies for synchronizing diction and analysis:

- *Strategy 5*: Try to produce utterances with the ideal delivery (because that helps project what you intend to present next).
- *Strategy 6*: When you suspend speaking, signal (if possible) what you intend to present on resuming speaking.
- *Strategy 7*: If you intend to revise or abandon an expression you have already presented, signal that intention as early as possible.

There are many techniques for realizing strategies 5 through 7. Here I will take up only a few of them.

3.1. Strategy 5: Pursue the ideal delivery

By strategy 5, speakers should pursue the ideal delivery, not merely to synchronize voice and ear (strategy 2), but to synchronize diction and analysis. The idea can be illustrated with Roger's utterance in (3), repeated here:

(3) Roger: now, - um do you and your husband have a *j- car*

When Roger finishes "do you and your husband have a", he projects not simply the arrival time of the next word, but its category: It is a nominal. By the expected object principle, Nina should try to use that information and prepare for the next element to be a nominal. So when Roger continues with "j-", he confounds Nina's preparation. Indeed, "j-" is *misinformation* about the nominal he finally produces, and that slows her down enough that she has to ask for a repair ("have a car?"). Roger's less than ideal delivery has cost the two of them extra time and effort, and that is something he will try to avoid.

Even when speakers cannot achieve an ideal delivery for an entire utterance, as I noted earlier, they try to achieve the ideal for its constituents. That should help in synchronizing diction and analysis as well. One technique for dealing with a suspension, for example, is to *restart* a constituent that contains the point of suspension, as here:

(8) Ian well I think that u:h a com- *a commitment* to a united Ireland, ... (2.8a.494)

Once Ian suspends speaking in the middle of "commitment", he has at least five options:

- *Option 1*. Simply proceed: "com- *mitment* to a united Ireland"
- *Option 2*. Restart the word: "com- *commitment* to a united Ireland"
- *Option 3*. Restart the noun phrase: "com- *a commitment* to a united Ireland"
- *Option 4*. Restart the clause: "com- *that a commitment* to a united Ireland"
- *Option 5*. Restart the sentence: "com- *I think that a commitment* to a united Ireland"

Options 2–5 are consistent with the continuity principle, restarting at the beginning of a constituent, whereas option 1 is not. In the London–Lund corpus, mid-word cut-off's like Ian's "com-" are followed most often by option 2 and sometimes by option 3 or higher (as in Ian's case). They are almost never followed by option 1, which does not restart at the beginning of a constituent. The argument, however, goes beyond mid-word cut-offs, for it applies to suspensions in the middle of any constituent. Almost all replacements, like Kate's replacement in 4, restart at constituent boundaries (Levelt, 1983). So speakers pursue an ideal delivery even when they do not quite succeed, and that should help synchronization of diction and analysis.

Even if speakers avoid option 1 for mid-word cut-offs, they still have options 2–5. Why does Ian select option 3 whereas most speakers select option 2? Presumably Ian means something by his selection. He might be telling his addressee, for example, that he is working not on the noun alone ("commitment"), but on the entire noun phrase ("a commitment"). But without other evidence, we cannot say what speakers mean by their selections.

3.2. Strategy 6: Before you suspend speaking, signal (if possible) the type of expression you intend to present after the suspension.

Speakers often repeat the initial function word of major constituents, as Reynard does in 9:

(9) Reynard *I uh I* wouldn't be surprised at that, – – I really wouldn't, (1.1.278)

As Wasow and I (Clark and Wasow, 1998) have argued, repeats like "I uh I" reflect two strategies. The first is to make a commitment to the upcoming constituent by producing its first word ("I"). The second is to restore continuity to the disrupted constituent by restarting the constituent ("I would not be surprised at that"). Although the second strategy is yet another form of strategy 5, the first strategy is a form of strategy 6. Let us see how.

The first tokens of many repeats (as in 9) are *prosodic orphans*. (a) They are separate prosodic words (or phrases) when in the ideal delivery they would be tied prosodically to the next word (e.g., as a clitic). (b) They are often followed by a cliticized *uh* or *um*, as in "I.yuh". And (c) they are often produced in ignorance of the word to follow, as in "I uh I'm ..." or "a uh an apple". Speakers produce these prosodic orphans, Wasow and I have argued, in order to make *preliminary commitments* to the constituents they are about to initiate. In doing so, they accomplish two things. First, they signal a delay in initiating the full constituent (fitting strategies 3 and 4). And, second, they project useful information about the constituent they are about to initiate. So, when Reynard produces the first "I" in (9), he commits himself to initiating a clause beginning with *I*. By the expected object principle, that should help his addressee get prepared to identify such a clause. There is good evidence that it does (Fox Tree, 1995).

Speakers use prosodic orphans in other circumstances as well. Consider this example:

(10) Duncan *and u:m*, . I think it's better than Guinness myself, (1.7.306)

Duncan's "and u:m" is an orphan. In the ideal, *and* should have been produced as part of a fluent

clause "and I think it's better than Guinness myself". Instead, it is combined with "u:m" and set apart in its own intonation unit. Why does Duncan create such an orphan? It is to make a preliminary commitment to a clause beginning with *and*. Again, by the expected object principle, that should help his addressee identify what comes next. The same argument holds for *dhi*, as in example (5), repeated here:

(5) Kenneth I would have to go down to *dhi* – film school, and talk to some of the people there, (1.10.1145.A)

"Dhi" (rhyming with *tea*) is also a prosodic orphan. With it, Kenneth is committing himself to producing a definite noun phrase beginning with *the*. That should help Kenneth's addressee prepare for the following nominal and, therefore, identify it more quickly when it does come. If Kenneth had suspended speech right after *to*, his addressee would have had less information about the noun phrase to come and should have taken longer to identify it.

3.3. Strategy 7: Signal any expression you intend to revise or abandon it as early you can.

Speakers often present expressions prematurely. By this, I mean that they later change their minds and revise or abandon those expressions. Most premature presentations provide both valid and invalid information about what speakers eventually commit themselves to. By the expected object principle, speakers should flag the invalid information as early as possible. That will keep the addressees from creating misanalyses they later have to redo. It takes time to correct misanalyses, and that in turn disrupts the pacing.

Speakers have many techniques for marking items to be revised or abandoned. One is the use of editing expressions, as illustrated here:

(11) Duncan is there a doctrine about that, – – *I mean* a doctrine about u:h – disfavouring American applicants, (2.6.978)

In using "I mean", Duncan tells his addressee that he is about to present a clarification or qualification. But of what? Duncan indicates what by the form of his resumption. It begins "a doctrine about . . ." which indicates that it is a revision of a noun phrase of the same type, namely "a doctrine about that". Or take this example:

(12) Helen well I had intended to be, looking . for – uh *or rather* eliminating, people over thirty- . – three or four, (2.6.505)

In using "or rather", Helen tells her addressee that she is correcting an item of her presentation, and in presenting "eliminating", she indicates that she is correcting a verb in the progressive tense, namely "looking for". English has many editing expressions for this purpose, including *I mean*, *or rather*, *you know*, *pardon*, *no*, *sorry*, *well* and *oh*. All of these signal the *type* of repair, elaboration or qualification the speaker is about to present (see, e.g., Levelt, 1983).

Speakers can also use prosody alone to mark the items they are revising or abandoning. Let us return to (4), repeated here:

(4) Kate they had . *they shortlisted* five people, - including me, (1.3.255)

Although Kate does not use an editing expression, she signals that she is making a repair by producing "they shortlisted" with the identical intonation of "they had". Such an intonation allows Kate to do two things. She exploits the fact that the intonation of "they shortlisted" is *inappropriate* as the continuation of "they had . . ." to signal that she is revising or abandoning part of what she has just presented. At the same time, she designs the intonation of "they shortlisted" to *match* the prosody of "they had", and not, for example, just "had". In this way she indicates the words "they had" as what she is revising or abandoning. Both of these collateral signals should help her addressee arrive more quickly at a proper analysis of what she is presenting.

4. Summary

Speakers, I have argued, have to synchronize their actions with those of their addressees if they

are to succeed. They have to synchronize voice and ear, making sure their addressees are ready to hear their vocalizations at the right moment. They have to synchronize diction and analysis, making sure their addressees are able to analyze their presentations correctly before they go on. And, though not discussed here, they also have to synchronize meaning and understanding, getting their addressees to understand what they mean before they go on. Speakers pursue a variety of strategies for achieving synchronization at these three levels, and these lead not only to attempts to be fluent, but also to the common forms of speech disfluencies – fillers, prolonged syllables, nonreduced vowels, preliminary commitments, restoration of continuity, editing expressions, and many others.

An alternative view of these forms of disfluency is that they are not communicative acts, but simply the by-products of problems with planning utterances. They are regular because the different forms of disfluencies are associated with different types of problems. Indeed, speakers do not seem to plan these forms consciously. This alternative view, however, has serious problems (see Clark and Fox Tree, in preparation; Clark and Wasow, 1998; Fox Tree and Clark, 1997). First, most of these forms are conventional. *Uh* and *um*, for example, are a conventional part of English, whereas French and Japanese have other conventional fillers. Planning may lead to much the same problems for speakers everywhere, but the solutions vary from language to language. Second, planning does not require awareness. We select *the* over *a*, *but* over *and*, and *since* over *because*, all without consciously doing so. Selecting one form of disfluency over another – for example, *uh* over silence – is no different. And, third, speakers have general control over their use of these forms of disfluency. Good public speakers avoid these forms in formal speeches, but exploit them in casual conversation.

Most forms of disfluency, therefore, should not be considered problems in speaking, but solutions to problems in speaking. They are genuine signals – collateral signals – that speakers design and produce with skill. Taken together, they constitute an elegant solution to a complicated problem: how speakers and addressees are to coordinate their communicative actions while carrying out the official business of their discourse.

References

Chafe, W.L. (Ed.), 1980. The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production. Ablex, Norwood, NJ.

Clark, H.H., 1994. Managing problems in speaking. Speech Communication 15, 243–250.

Clark, H.H., 1996. Using Language. Cambridge University Press, Cambridge.

Clark, H.H., Clark, E.V., 1977. Psychology and Language. Harcourt, Brace and Jovanovich, New York.

Clark, H.H., Fox Tree, J.E. Using uh and um in spontaneous speaking, in preparation.

Clark, H.H., Wasow, T., 1998. Repeating words in spontaneous speech. Cognitive Psychol. 37, 201–242.

Fox Tree, J.E., 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. J. Memory Language 34, 709–738.

Fox Tree, J.E., Clark, H.H., 1997. Pronouncing 'the' as 'thee' to signal problems in speaking. Cognition 62, 151–167.

Goodwin, C., 1981. Conversational Organization: Interaction between Speakers and Hearers. Academic Press, New York.

Levelt, W.J.M., 1983. Monitoring and self-repair in speech. Cognition 14, 41–104.

Selkirk, E., 1996. The prosodic structure of function words. In: Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition. Lawrence Erlbaum, Mahwah, NJ, pp. 187–213.

Svartvik, J., Quirk, R. (Eds.), 1980. A Corpus of English Conversation. C.W.K. Gleerup, Lund.

Woodworth, R.S., 1938. Experimental Psychology. Henry Holt, New York.