

When to Start Speaking, When to Stop, and How

Herbert H. Clark

Department of Psychology
Stanford University, Stanford CA, USA
herb@psych.Stanford.edu

Abstract

Timing is essential in dialogue. When people speak, they choose not only *what* to say, but *when* to say what they say. They cannot speak until they have something to say—until they have formulated a word, phrase, or sentence. Nor can they speak beyond what they have formulated. Still, once they have a bit formulated, they can produce it whenever they want—within limits. And they can suspend speaking whenever they want—also within limits. The proposal is that speakers often mean things by their choice of timing.

1. Introduction

Timing is essential in dialogue. When people speak, they choose not only *what* to say, but *when* to say what they say. Their timing is determined partly by processing constraints. Speakers cannot speak until they have something to say—until they have formulated a word, phrase, or sentence. Nor can they speak beyond what they have formulated. Still, once they have a bit formulated, they can produce it whenever they want—within limits. And they can suspend speaking whenever they want—also within limits. The proposal is that speakers often mean things by their choice of timing. Although there is a well established study of what speakers mean by their choice of words, phrases, and sentences, there is an almost complete neglect of what speakers mean by their choice of timing.

Timing is often treated as solely a product of processing constraints. In the following example, Sam is telling Ted what to do next in building a Lego model (with pauses marked in seconds):

- (1) Sam Kay now get (1.03) a-uh eight piece green, (1.58) and join the two (.23) so it's all symmetric, yeah right in the center (Clark & Krych, 2004)

According to most models of speaking, Sam plans his utterance in increments. Once he has his first increment formulated, he begins “Kay now get.” After *get*, he apparently has nothing ready, so he pauses 1.03 sec, which allows him to finish formulating the next increment and start producing it. He does this twice more, with pauses of 1.58 sec and .23 sec. In this view, speakers initiate each increment as soon as it is formulated. And when they delay, it isn't that they choose to, but that they are forced to.

Sam's timing, however, is determined not just by processing constraints, but by his changing beliefs about Ted. This is clear from a close look at Sam and Ted's exchange on videotape. Sam does indeed appear to have trouble formulating “an eight piece green,” which would account for

his first 1.03 sec delay. But he is ready with “and join the two” long before he produces it. He doesn't start it until he sees Ted retrieve the right block. That is, he initiates “and join the two” not simply when *he* is ready to produce it, but when he believes *Ted* is ready to attend to and understand it.

The aim of this talk is to characterize some of people's choices about when to start and stop speaking. The broader aim is to characterize how speakers communicate not only with *what* they say, but with *how* they say it.

2. Displays

Speakers make choices about how to *display* their utterances—about the time, place, and manner in which they perform them (Clark, 2003). Let me consider three aspects of such displays—parcels, temporal placement, and temporal indexes.

Spontaneous utterances get realized in PARCELS. A parcel is what I call a bounded strip of continuous speech within a single intonation unit. Consider this exchange, in which the intonation units are marked off with commas (from Svartvik & Quirk, 1980):

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

There are four parcels here, which run the gamut in size from a full sentence (*did I tell you about that?*), to a prepositional phrase (*by foul means*), to a single word (*no*). All of these parcels are complete, intact intonation units. They are STANDARD PARCELS. In 1, in contrast, there are six parcels, many of which are non-standard.

Speakers display parcels at particular moments in time. I will call the placement of a parcel in time its TEMPORAL PLACEMENT. When I place a book on a table, a pan in an oven, or a coat on a hook, I place one object *with respect to* another—a FIGURE with respect to a GROUND. Temporal placement is no different. When Nancy displays the parcel “by foul means,” she places it temporally with respect to her previous phrase, “... sewing-machine.” She treats “by foul means” as figure and “... sewing-machine” as ground, and places the figure after the ground. And just as we can say that I parked my car six feet from a tree, we can say that Nancy placed “by foul means” 200 msec (say) after “... sewing-machine.”

What speakers mean by the placement of a parcel depends on what is figure and what is ground. When an official at a track

meet starts a ten-thousand kilometer race by saying, “Ready ... set ... go!” he uses the realization of *go* to indicate the beginning of the race. To do that, he places “go” (the figure) *with respect to* the “ready ... set” (the ground). Similarly, Sam places “and join the two” (the figure) right after Ted’s retrieval of the right Lego block (the ground) to signal that the content of the phrase “and join the two” is to be taken as the next step after Ted’s action. If Sam had placed the same phrase before Ted had retrieved the block, he might have been taken to mean something different.

The race official used the moment of realization of *go* as a *temporal index* to that very moment. Let us call this index *t*, written *t*(“go”), which is to be read, “the time of the performance or realization of *go*.” What the official meant by uttering *go* (in part) is this: “I hereby declare the race to begin at *t*(“go”).” Most parcels have several potential temporal indexes associated with them. Nancy’s “by foul means” has at least three:

t-beg(“by foul means”)	the moment at the <i>beginning</i> of /b/
t-end(“by foul means”)	the moment at the <i>end</i> of /z/ in <i>means</i>
t-con(“by foul means”)	the <i>continuous</i> interval of the phrase

(I will drop the suffix when it doesn’t matter which index is intended.) Nancy creates all three indexes with the realization of her parcel, and she can use them very differently. In effect, she can treat the left edge, right edge, or body of the parcel as figure and place that figure with respect to a ground.

Temporal indexes are used in three broad categories of timing. (1) EXTERNAL TIMING. This use is for marking a time with respect to external events, as in establishing the meaning of *now*, *today*, *yesterday*, *next year*, *the other day*, *did*, *had*, *has had*, etc. What I mean by “tomorrow,” for example, is “the day after the day that contains *t*(‘tomorrow’).” The parcel containing “tomorrow” is figure, and the external calendrical day is ground. (2) INTERNAL TIMING. This use of temporal indexes is for marking a time with respect to events within the speaker’s own speech or actions. (3) CROSS TIMING. This use is for marking a time with respect to one’s partner’s actions. We are now in a position to examine when, why, and how people start and stop speaking. Let us begin at the beginning.

3. When to start speaking

Speakers are dogged by insurmountable processing constraints. I will take four for granted. First, speakers try to *plan* utterances one intonation unit unit at a time—the *unit* (“one unit at a time”) constraint. Second, they try to *formulate* these major units in bite-sized increments—the *info* (“incremental formulation”) constraint. Third, they cannot start producing the first increment *any earlier than* they have formulated it—the *near* (“no earlier than”) constraint. And fourth, they can speak no further than the increment they have formulated—the *far* constraint. These limit how early and how long they speak.

But within these limits, when do speakers *actually* start speaking? The argument is this: Speakers recognize that most parcels have standard, expectable placements. In true Gricean fashion, speakers therefore try to choose the standard placement unless they want to implicate something more by placing it elsewhere.

3.1. Cross timing

In dialogue, people place their own speech with respect to their partners’ speech. Traditionally, that is a matter of taking turns, about which there are two standards (Sacks, Schegloff, and Jefferson, 1974). Suppose Ann is talking to Ben and Charles. If she selects Ben as the next speaker by, say, asking him a question, Ben is obligated to start speaking at the point when she finishes. Otherwise, whenever she completes a possible turn, the next person to speak (Ben or Charles) is granted the next turn. So we have two standards:

- (1) *Uptake standard*. When addressed with a “current-speaker-selects-next” technique, a person is expected to initiate his or her uptake at the end of that turn.
- (2) *Next contribution standard*. When a participant wants the next turn, he or she is expected to initiate the turn at the end of the current speaker’s turn.

By standard 2, if Ben wants a turn, he is under pressure to speak as soon after Ann has finished as possible, and that requires him to project *t-end*(Ann’s turn). Ben isn’t perfect, so he will sometimes mis-project and start his turn with a slight overlap or slight delay.

Placing a parcel by one of these standards leads to a standard interpretation. Take 3, which is part of 2:

- | | | |
|-----|-------|----------------------------|
| (3) | Nancy | did I tell you about that, |
| | Julia | no, |

By asking the question, Nancy selects Julia as the next speaker, and Julia initiates her answer (“no”) in the standard location—immediately after Nancy has finished. By placing “no” according to the uptake standard, she signals that it is an uptake to the question. Or consider 4:

- | | | |
|-----|--------|--|
| (4) | Adam: | I’ve only just discovered that, uh a week ago, |
| | Brian: | we used to have that in the war, |

By standard 2, Brian places his contribution at the end of Adam’s turn, signaling that it is to be taken as a next contribution.

When speakers manifestly depart from these standards for non-processing reasons, they use their timing for marked interpretations. One way to depart is to *overlap* with the previous turn. Here is an example of a STRATEGIC INTERRUPTION:

- | | | |
|-----|--------|---|
| (5) | Maggie | There doesn’t really seem *anything* |
| | Ken | *but how* long do you think it’ll take them to finish? |

Ken signals that he is requesting the floor by manifestly initiating his turn in the middle of Maggie’s—a non-standard

placement. For her part, Maggie signals that she is granting his request by DISCONTINUING her turn—that is, by terminating it prematurely. Ken could have waited until Maggie had finished, and Maggie could have continued her turn. By departing from these standards, they each signaled something special. Two other types of overlap are RECYCLED TURN BEGINNINGS (Schegloff, 1987) and RECYCLED ASSESSMENTS (Clark & Schaefer, 1987), which have other marked interpretations. Speakers can also depart from the standards with longer-than-expected delays, and these lead to still other marked interpretations.

There is also a standard placement for acknowledgments like *uh-huh* and assessments like *terrific*. Consider the acknowledgments *yea* and *m* in this exchange:

- (6) Ann um well I hadn't done any English at all, you know, since O-level,
 Ben **yea** .
 Ann and I went to some second year seminars, where there are only about half a dozen people,
 Ben ***m***
 Ann ***and*** they discussed what a word was,

Here Ben times “yea” and “m” to abut the end of certain of Ann’s intonation units. Why? To signal that they are acknowledgements to *those* units. Ann in turn signals that she is continuing her extended turn by proceeding *in overlap with* the acknowledgments. Assessments are treated differently, as next speakers try *not* to place their speech in overlap with them (Goodwin, 1986).

Speakers can depart from these standards. With a PREMATURE ACKNOWLEDGMENT, they place an acknowledgment *overlapping* the end of the current intonation unit. With it, they can signal that they understand that unit without hearing all of it (Jefferson, 1973). With a RECYCLED ACKNOWLEDGMENT, speakers can signal strength of agreement or enthusiasm with the current speaker’s intonation unit. And there are others.

3.2. Internal timing

Spontaneous speakers are rarely fluent, and yet they recognize fluency as the standard against which to measure timing. We can divide fluency into two standards:

(3) *Continuity standard*. Once speakers have initiated an intonation unit, they are to speak continuously until they complete it.

(4) *Immediacy standard*. If speakers intend to initiate a second intonation unit within a turn, they are to initiate it within roughly 1 second of the end of their previous one.

Although speakers may fail to meet these standards for processing reasons, they can also manifestly depart from them to signal other things. Let us begin with continuity.

Speakers may suspend speaking in the middle of an intonation unit for at least three reasons: (1) they are delayed in formulating the next increment; (2) they intend to reformulate part or all of what they have already produced; or

(3) they are abandoning their current turn altogether. Ordinarily, speakers want to make sure that silences caused by the first two reasons are not misconstrued as lapses caused by the third. How do they guard against such misconstruals?

One strategy is to add FILLERS to announce that they are initiating a delay—and, by implication, that they are *not* abandoning their turn (Clark & Fox Tree, 2002). Take this example from a student who has been asked about “recent novels” he has read:

- (7) Alan I've **u:m** recently read **u:m** . oh, . Lord of the Flies, (3.5a.110)

Alan adds a prolonged *um* once after the parcel “I’ve” and again after the parcel “recently read.” He is delayed at both points apparently because he is having trouble coming up with a novel he has read recently. He uses *um* to *announce* that he is initiating a delay and not abandoning his answer. Speakers use *uh* to announce minor delays and *um* major ones. The timing of these announcements signals the start of the delay and often, by implication, why the speaker is delaying.

Another strategy for guarding against such misconstruals is the PRELIMINARY COMMITMENT (Clark & Wasow, 1998). In 7, Alan initiates his turn even though he surely recognizes, even before starting, that he will discontinue it after the single word *I’ve*. Why would he do that? By placing “I’ve” according to standard 1, he (a) makes a preliminary commitment to producing a clause that begins with *I’ve* and, by implication, (b) signals that he has initiated his answer to the question. This is in contrast with a unit initial *uh*, as here:

- (8) Sam but but also if anyone wants to raise anything else about the college, - . **u:h** do please do so, (3.4.21)

To meet standard 4, Sam uses *u:h* to commit himself to continuing his turn. Preliminary commitments are a common source of repeated words (Clark & Wasow, 1998), as in 9:

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

Reynard makes a preliminary commitment with the first token of *I* (to meet standard 1), but he then restarts the clause at *I* to restore fluency to his utterance (for standard 3). The result is a repeated *I*.

4. When to stop speaking

Speakers also have choices about when to stop speaking. The overarching standard for stopping is the continuity standard. But that standard can be hard to meet for processing reasons. One is the *far* constraint: speakers cannot continue beyond the material already formulated. But even within this constraint, speakers have choices about when to suspend speaking. And they have other choices as well.

4.1. Internal timing

Suppose speakers are having trouble formulating a word or a constituent and are forced to stop speaking because of the far constraint. If they are monitoring their progress—and much evidence suggests that they do—they have choices about when and how to suspend speaking. As one example, let me describe the use of non-reduced vowels in articles and prepositions (Fox Tree & Clark, 1997). Consider this utterance:

- (10) Susan I would find **thi:y** um – the colour, not the theme so much, but the colour, (1.8.496)

Susan is having trouble deciding on what decoration she finds suitable for a room, so she is delayed in coming up with *colour* vs. *theme*. Realizing that she will have to delay in that choice, she suspends speaking after *the*. Furthermore, she delivers *the* with a non-reduced vowel, written *thi:y*, which she follows with *um*, a delay signal. The argument is that she is marking her suspension as a signal that she is having trouble in planning.

Speakers have contrasting ways of suspending speech mid-utterance. They can stop in the middle of a word, with or without a glottal closure, to signal that that word is incorrect (Levelt, 1983). They can stop between words but without a non-reduced vowel to mark a more minor problem in speaking (Fox Tree & Clark, 1997). They can also prolong a word to indicate an ongoing delay (Clark & Fox Tree, 2002). These are just a few of the reasons and manners in which speakers can stop speaking mid-utterance.

4.2. Cross timing

In cross timing, speakers can suspend, or refuse to suspend, their speech at non-standard moments for many reasons. In 5, recall that Ken initiated his utterance to request Maggie to stop speaking and let him speak. Maggie in turn discontinued her utterance after *anything* to signal that she was granting his request. In other examples, speakers refuse to grant such requests by not discontinuing their speech.

Another reason for suspending speech prematurely is to request help from addressees, for example, in finding the right word (Goodwin & Goodwin, 1986). Still another is to signal that what the speaker was about to say is to delicate or sensitive to say aloud. And yet another is to signal that the speaker is abandoning the remainder of the current unit because it would be otiose in the circumstances. Each of these reasons leads to a different manner of suspension.

Another class of reasons for premature suspensions is to deal with actions by the interlocutor. One example is illustrated in 1. There Sam waits 1.58 sec before continuing the next installment of his instruction while Ted finds the right Lego block. In other cases, speakers suspend speaking to comment on what their partner is doing at the moment. In the following example, Jane is telling her partner Ken where to put a Lego block:

- (11) Jane and put it on the right hand half of the- **yes** (0.3) of the green rectangle

Mid-utterance, Ken poises the block over the right location, at which point Jane suspends her speaking to say “yes” before continuing. This way she confirms his location as correct before he tries out another location.

5. Conclusions

Speakers can only speak when they have something to say—and when they have formulated how they are going to say that. Still, they start and stop speaking for a wide range of strategic reasons. Some of these reasons are to deal with their difficulties in planning, formulating, and executing an utterance. But many others are to manage other features of the dialogue. In many cases, speakers use the temporal placement of parcels of speech to signal things. The challenge is to say just how speakers use timing to do things.

6. References

- [1] Clark, H. H. (2003). Pragmatics of language performance. In L. R. Horn & G. Ward (Eds.), *Handbook of pragmatics*. Oxford: Blackwell Publishing.
- [2] Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73-111.
- [3] Clark, H. H., & Krych, M. A. (2003). Speaking while monitoring addressees for understanding. *Journal of Memory & Language*, in press.
- [4] Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, 26, 209-225.
- [5] Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201-242.
- [6] Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62(2), 151-167.
- [7] Goodwin, C. (1986). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, 9, 205-217.
- [8] Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62, 51-75.
- [9] Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
- [10] Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696-735.
- [11] Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In G. Button & J. R. E. Lee & et al. (Eds.), *Intercommunication series* (Vol. 1, pp. 70-85). Clevedon: Multilingual Matters Ltd.