

# **Collaborative Construction of Multimodal Utterances**

**Edwin Hutchins<sup>1</sup>**

**Saeko Nomura<sup>2</sup>**

Abstract:

The papers in this volume demonstrate the pervasiveness of multimodal utterances. The collaborative construction of utterances is also well known. In this chapter we explore utterances that are both multimodal and collaboratively constructed; in particular, utterances in which the gesture of one participant stands in a relation of mutual elaboration with the talk of another participant. Drawing on interactions between Japanese airline pilots and an American flight instructor, we show how when multiple participants talk and gesture simultaneously, relations among semiotic resources proliferate. Pilots use their speech and their bodies to jointly construct multimodal representations of the objects, the events, and the actions that constitute their profession.

---

<sup>1</sup> *Department of Cognitive Science, University of California San Diego*

<sup>2</sup> *Department of Communication, Cornell University, Ithaca, NY*

## *Introduction*

The production of collaboratively constructed utterances is well-known (*Goodwin, 1979; Goodwin, Goodwin, & Yaeger-Dror, 2002; Jacoby & Ochs, 1995*). In the most frequently studied type of collaboratively constructed utterance, one speaker begins an utterance in a way that projects possible completions. Another speaker then contributes utterance elements that are incorporated into a jointly produced utterance. The acceptance by participants of a collaboratively constructed utterance is strong evidence for the establishment of common ground understanding (*Clark, 1992*).

Multimodal utterances (*Goodwin, 2006*) contain both verbal (speech) and non-verbal (gesture) elements. Of course, virtually all verbal utterances are multimodal in the sense that they are produced in concert with coordinated modulation of body posture, facial expression and eye gaze. In this chapter, we will reserve the multimodal label for utterances in which the verbal and non-verbal elements mutually elaborate one another. Multimodal utterances are also extremely common.

In this paper we examine the intersection of the set of collaboratively constructed utterances with the set of multimodal utterances. We are especially interested in cases where the multimodal nature of the utterance intersects with the process of collaborative construction such that a gesture or other non-verbal element produced by one participant stands in a relation of mutual elaboration with a spoken element produced by another participant.

Mutual elaboration is a complex relationship. Goodwin speaks of it as existing among elements of a meaning making event. When the meaning of each of two or more elements is constrained, altered, or enriched by the meanings of the other elements, the elements can be said to mutually elaborate one another. In order for an analyst to claim that such a relationship exists, the analyst must know the concepts that constitute the domain of discourse and must be able to provide ethnographic warrants for claims about the meanings of the semiotic resources.

In the analysis that follows, we will focus on two kinds of correspondence among elements of the semiotic field.

- 1) Semantic correspondence, in which two or more elements in the active semiotic field refer to the related conceptual elements, even if they are not produced or processed simultaneously.

- 2) Temporal correspondence, in which two or more utterance elements are produced close in time so that they afford processing together, even if they do not refer to the same conceptual elements.

We hypothesize that human minds are always looking for these kinds of correspondences. Each of these kinds of correspondence probably recruits a different kind of processing underlying the mutual elaboration of the semiotic resources. This is a topic for subsequent experimental investigation.

Schegloff (1984) introduced the concept of “lexical affiliate” to address semantic correspondence between a gesture and a spoken element. He identified the lexical affiliate as “the word or words that correspond most closely to a gesture in meaning.” Kendon argued that this notion is problematic because not all gestures have lexical affiliates. It is also problematic because semantic relations are complex and it is not clear what measure of semantic distance is implied by the notion of close correspondence in meaning. This is problematic in a third way because neither gestures nor words have meanings that are independent of the context of their production. If a gesture and an element of speech are construed to be related in a meaningful way, then they probably mutually elaborate each other’s meanings. Finally, this is problematic because while the definition seems to want to be about the relation between talk and gesture, the label “lexical affiliate” highlights just one element of a complex relation.

No two representations ever refer to exactly the same concept. But conceptual objects as created in human activity are complex and have parts. We can say that the referents of two or more representations may sometimes be captured by a particular element of a conceptual object. When that happens we will say they bear a congruent semantic relation to one another. When the referents of two or more representations are captured by different elements of a single complex conceptual object then we say that the representations bear a complementary semantic relation to one another. Some common forms of complementary relations occur when one representation refers to a cause and the other refers to the effect of that cause (metonymy), and when one representation refers to a whole and the other representation refers to a part of that whole (synecdoche).

With respect to temporal correspondence, Schegloff (1984), studying single speaker production, noted that gestures tend to precede their lexical affiliates. Because we are interested in cases that involve two or more participants, one might assume that spoken elements produced

by one participant will normally precede and serve as cues or triggers for gesture elements produced by another participant. This does happen, but it does not appear to be the most frequent case. In the analysis presented below we identify seven instances of collaboratively constructed multimodal utterances.

McNeill (2005) used the phrase “co-expressive speech” to designate spoken language that co-occurs in time with gestures such that the speech and gesture have related referents. Like Schegloff’s term “lexical affiliate”, the phrase “co-expressive speech” is intended to describe a relation, but it highlights just one element of the relation. This class of relations is interesting because of the productivity of the emergent properties of the conceptual integrations. Of course, it is possible that the referents of two simultaneously produced representations are unrelated. We observed this in our corpus when an instructor scratched his elbow while describing a flying procedure. We use our ethnographic grounding to establish the presence or absence of relations among representations. It is hard to say that any two things are really unrelatable. Human imagination is a powerful constructor of relations.

No two representations ever occur at exactly the same time. The perception of simultaneity is an interesting problem in psychophysics that we cannot treat here. However, people judge that some things occur close enough in time to be taken as having been simultaneous. Other pairs of representations occur at perceptibly different points in time, yet are still taken to be part of a single larger act of meaning making. Still further separated in time, representations may be so remote that they are not construed as being part of the same meaning making activity. The threshold here is not simply a function of the passage of time, but depends on the understood temporal structure of the activity. That is, temporal relevance is a negotiated aspect of the interaction.

McNeill (2005) addressed both semantic and temporal relations when he distinguished the lexical affiliate from co-expressive speech. “A gesture, including the stroke, may anticipate its lexical affiliate but, at the same time, be synchronized with its co-expressive speech segment.” (2005, p. 37). Following this classification scheme analysis proceeds by first finding a gesture and then locating the gesture’s lexical affiliate (if any) and the gesture’s co-expressive speech (if any). We will take a different approach.

In our analysis we will examine the relations among the semiotic resources that are recruited by the activity of collaboratively constructing conceptual representations. Focusing on

the properties of relations rather than on the properties of the elements solves the problems noted above with the use of the terms “lexical affiliate” and “co-expressive speech”. The term “lexical affiliate” denotes the spoken element of a gesture-speech relation in which the two elements are semantically congruent, whether or not they occur at the same time. The term “co-expressive speech” denotes the spoken element of a gesture-speech relation in which the two elements occur at the same time, and are either semantically congruent or complementary. The table below maps the intersection of semantic relations and temporal relations and shows where the phenomena denoted by the terms “lexical affiliate” and “co-expressive speech” lie in our proposed new classification.

|                   |               | Temporal Relation    |        |           |
|-------------------|---------------|----------------------|--------|-----------|
|                   |               | Concurrent           | Offset | Unrelated |
| Semantic Relation | Congruent     | Lexical Affiliate    |        |           |
|                   | Complementary | Co-expressive Speech |        |           |
|                   | Unrelated     |                      |        |           |

This diagram highlights the fact that the existing categories focus on the properties of particular representations rather than on relations among representations. The movement from theories that focus on properties of elements to relations among elements is underway in many parts of science, including cognitive science (*Hutchins, in press*).

The most familiar relation is semantically congruent and temporally concurrent. This is the case when gesture and talk are produced simultaneously and refer to the same conceptual element(s). This is probably the most frequently produced type of relation between speech and gesture in single-speaker utterances. It is probably not the most frequently produced type of relation in collaboratively constructed utterances. This possibility highlights the fact that these relations may arise in three different configurations of socio-cultural space: 1) relations among elements produced by a single speaker, 2) relations among elements produced by more than one speaker, 3) relations among elements that are produced by a speaker and representations in material media. One can imagine constructing a table for each configuration. We expect the relative frequencies of events to be different in the three configurations. This is a topic for further investigation.

## **Methods**

Since 2005, under a research agreement with Boeing's Flight Deck Concepts Center, we have been conducting a worldwide investigation of the roles of language and culture in commercial airline flight deck (cockpit) operations. Our ethnographic data collection procedures include the observation of airline pilots in revenue flight and in high fidelity simulators, and interviews with pilots and other airline personnel. From the observer's seat in the flight deck, we take extensive written notes, capture digital still images, and collect copies of all of the flight paperwork. These data are subsequently integrated into hyperlinked field notes. Video data from the flight simulator are transcribed and the micro-scale language and culture practices are documented.

### **Cognitive ethnography**

In experimental studies, the researcher's knowledge of the stimulus conditions and the organization of the activity in the experimental trials provide the warrant for interpretations of the observed responses. This is especially true when the interpretations involve claims about the meanings of the observed behavior. The researcher assumes that the meanings of the subjects' behaviors are knowable, and that they can be recovered not only by the researcher, but by readers who are first informed of the nature of the materials and the experimental tasks performed by the subjects. McNeill (2005, pp. 259 - 60) points out that some knowledge of this sort is needed to solve what he called the "circularity problem." Without an independent way to establish meanings, gestures could only be interpreted as having the same meaning as the talk with which they co-occur. One could never distinguish temporally offset congruent relations from temporally concurrent complementary relations. When we do cognitive research in the real world, we have the same need for warrants to support interpretations of the meanings of observed behaviors. But as cognitive ethnographers, we make no attempt to control the observed activities. When behavior is observed in naturally occurring culturally organized activities, an ethnographic study of the activity system takes the place of the laboratory researcher's knowledge of the experimental conditions. In either case, the interpretation of the significance of observed behaviors relies on knowledge of the conditions of its production. The interpretation of the behavior of any airline pilot requires a wealth of technical knowledge about aircraft and airline operations. Understanding and interpreting patterns of behavior of pilots from other

cultures requires a deep knowledge of the language and culture involved. Fortunately we have been able to assemble a research team that includes technical pilots and human factors specialists from Boeing in addition to a cognitive anthropologist. Our work with Japanese airlines has included an expert on Japanese language and culture. In the discussion below, we will refer to training documents and documented practices of professional pilot culture as well as native language competence in the languages spoken as sources of warrants for claims about the probable meanings of semiotic resources.

### **Data collection**

So far we have collected data in five nations. We have ridden in the flight deck with the crews, observing 64 pilots as they flew 70 segments of revenue flight. At various training centers we have made video recordings of 26 pilots as they engaged in more than 50 hours of simulator flying and approximately 30 hours of pre- and post-simulator session briefings. In addition to audio and video recordings, we also collected paperwork used in training (lesson plans, flight profiles) and electronic copies of the flight crew training manual and operating manual for the airplane being learned. Twenty three hours of training were recorded in Japan, where the training was conducted in Japanese. Twelve hours of training were recorded in Brazil, where the sessions were conducted in Brazilian Portuguese. We made video and audio recordings of instructor/pilot interactions both in the simulator and in the briefing room before and after the simulator session. A total of 37 hours of training for three Japanese pilots were recorded in Seattle Washington. This training was conducted in English. We base the analysis presented in this paper on just one brief clip from this Seattle corpus. The clip records 2 minutes and thirty seven seconds of interaction between an American instructor and two senior Japanese pilots engaged in a pre-simulator-session briefing. Both of the pilots were already qualified to fly as captains in different models of Boeing airplane. In this course, they were transitioning to the Boeing 737NG<sup>1</sup>.

### **Analysis method**

We have chosen a brief interaction to illustrate the phenomena of collaborative construction of multimodal utterances. The interaction creates conceptual objects. David McNeill (2005) claims that in complex acts of meaning making, the parts get their meanings from the whole, rather than the other way around. Following this insight, we first identify the

conceptual projects of each participant. The claims we make concerning conceptual structures are based on our extensive ethnographic documentation of the activity of commercial airline operations and training<sup>2</sup>. We infer from ongoing talk and gesture the structure of a conceptual object that the participants construct in interaction. Once the development of the conceptual objects of interest have been described, we examine the ways that verbal and non-verbal utterance elements instantiate various parts of the conceptual object of interest. By this we mean that we can examine the relations of mutual elaboration among the semiotic resources. We can see where gesture and talk refer to the same aspect of a conceptual object of interest, and where they refer to different aspects of that conceptual object. For each observed gesture, we identify the viewpoint implied by the production of the gesture in context (*McNeill, 1992; Kendon, 2004*). We also examine the timing of the production or highlighting of conceptual elements. Doing this allows us to examine the relations between verbal and non-verbal elements as they are mediated by the developing conceptual object. We identify both the temporal and semantic relations among the semiotic resources incorporated in the representations. Where the elements are temporally offset, we also note which element (gesture or speech) anticipates the production of the other. We code the semantic relations between semiotic resources as congruent or complementary.

### **Flight training activity**

Flying is an embodied activity. Even in the age of computerized autoflight systems that are capable of landing an airplane without the pilot touching the controls, pilots still learn how to hand-fly all maneuvers. Flying requires complex coordination skills. In an airplane with conventional controls, roll and pitch are controlled by the yoke, and yaw is controlled using rudder pedals. Thrust is controlled by thrust levers. Many maneuvers require the simultaneous coordinated manipulation of all of these controls.

High-fidelity flight simulators are powerful training tools because they allow pilots to practice flying skills, and especially to practice responding to events that would be dangerous to practice in an actual airplane. The pilots in a pre-simulator briefing typically imagine the actions they will take when they encounter specific circumstances of flight in the simulator. Since flying an airplane is partly a matter of using the body to manipulate controls, pilots often use their bodies to imagine or pre-enact the actions they expect to take in the simulator. Similarly, in

post-simulator debriefings, pilots often re-enact the actions they took, or should have taken, during the simulator session as a way of refining their motor representation of the actions.

All modern airliners are operated by a two-pilot flight crew. On each flight segment, one pilot serves as Pilot Flying (PF) and is responsible for controlling the aircraft and supervising its navigation. The other pilot serves as Pilot Monitoring (PM), and is responsible for communicating with air traffic services, operating the airplane's systems, reading checklists, and backing up the PF when needed. The coordination of crew activity under the regime of these roles is called Crew Resource Management, and is a part of all airline flight training.

### **Approach to Stall Recovery**

One of the maneuvers practiced by the pilots is called an "Approach to Stall Recovery."<sup>3</sup> As we will see, there is a difference between the way Boeing teaches this maneuver and the way it is practiced at the airline for which the pilots work. A pilot can approach a stall by holding back-pressure on the yoke as the airplane decelerates. To recover from a stall approached this way, a pilot adds power and then simply relaxes the back-pressure on the yoke. This is how the maneuver is taught by the airline for which the pilots work. Another way to approach a stall is to use stabilizer trim to neutralize control pressures while decelerating. To recover from a stall approached this way, the pilot adds power and must now push the yoke forward to restore an angle of attack suitable for the target speed. This can require quite a lot of force. Boeing teaches the maneuver using this second, more difficult approach to stall technique. The pilots refer to this as the "Boeing way." The procedure shown in the Flight Crew Operating Manual (FCOM) (Figure 1) is a generic procedure that can be used to recover from an approach to a stall in any configuration (setting of flaps and landing gear). The pilots are preparing to practice an approach to a particular kind of stall event known as a departure stall. This is flown with the landing gear retracted, the flaps extended at 5°, and with 20° bank attitude (Figure 2). The flap setting is a key element because it determines the speed at which the maneuver is begun and the target speed for its completion.

|                             |                    |
|-----------------------------|--------------------|
| <b>Maneuvers</b>            | <b>Chapter MAN</b> |
| <b>Non-Normal Maneuvers</b> | <b>Section 1</b>   |

**Approach to Stall Recovery**

The following is immediately accomplished at the first indication of stall buffet or stick shaker.

| Pilot Flying   | Pilot Monitoring   |
|--|--|
| <ul style="list-style-type: none"> <li>• Advance thrust levers to maximum thrust.</li> <li>• Smoothly adjusting pitch attitude* to avoid ground contact or obstacles.</li> <li>• Level the wings (do not change flaps or landing gear configuration).</li> <li>• Retract the speedbrakes.</li> </ul> | <ul style="list-style-type: none"> <li>• Verify maximum thrust.</li> <li>• Monitor altitude and airspeed.</li> <li>• Call out any trend toward terrain contact.</li> </ul> |
| <p>When ground contact is no longer a factor:</p> <ul style="list-style-type: none"> <li>• Adjust pitch attitude to accelerate while minimizing altitude loss.</li> <li>• Return to speed appropriate for the configuration.</li> </ul>  |  |

**Note:** \*At high altitudes it may be necessary to decrease pitch attitude below the horizon to achieve acceleration.

Figure 1: “Approach to stall recovery” in the Flight Crew Operating Manual (FCOM)

The computer displays the specific procedure for practicing a recovery from a departure stall in the simulator (see figure 2). In the figure and the following discussion, “stick shaker” refers to a vibration that is felt in the control yoke. This is one of the indications that the airplane is on the verge of a stall<sup>4</sup>.

**#1: FLAPS 5**

**20° Bank**

- Start: Flaps 5. **FLAPS 5 speed**
- N1 45% (approx 1 knot/sec decel)
- Establish 20° bank (PF check VSI, ALT, PLI)
- Stick shaker
  - Smoothly apply MAX Power (PM adjust to GA)
  - Level wings, do not change config, retract SB

- Power comes up, apply nose down trim
- Airspeed increases, lower pitch to 5 - 6°
- Approaching **Flaps 5 speed** - 65%
- Finish: **FLAPS 5 speed**

Figure 2: The Departure Stall practice procedure.

## **Analysis**

Our example is a video clip which is 2 minutes and thirty-seven seconds in length. Three pilots are seated at a table (See the figure accompanying case 3 below). On the far side of the table, facing the camera is an instructor pilot. The instructor makes use of diagrams, lesson plans, etc. both on paper-based materials and on a computer display. He has a laptop computer in front of him that he uses to control the display screen placed at the left end of the table. On the near side of the table with their backs to the camera are the two Japanese pilots (Pilot Flying (PF) on the left and Pilot Monitoring (PM) on the right). The students have their own materials which they can annotate. On the table in front of the pilots are notebooks. In this excerpt, the instructor placed the FCOM on the table in front of the pilots so it was right side up for them.

## **Overview of the clip**

The instructor began by reading the “Approach to Stall Recovery” procedure from the FCOM. Reading text while tracing the words with one’s finger is a practice that supports comprehension by non-native English speakers (*Hutchins, Nomura, & Holder, 2006*), and is a good example of the coordination of action with an artifact. He used gesture to elaborate his speech by modeling the attitude of the airplane and the crew’s manipulation of the controls. The instructor provided commentary on elements of the procedure as he read them, and exemplified some of the procedural steps by role playing, miming the actions of a pilot recovering from an approach to stall. PM noticed the difference between the Boeing technique for practicing this maneuver and the one used at his airline. This gave rise to a discussion of the differences between the techniques and what the pilots would have to do to fly the maneuver the Boeing way. In the last 25 seconds the instructor drove home a point about the Boeing technique by linking it to a dynamical property of the airplane. With engines mounted under the wings, the airplane will

tend to pitch up when power is increased suddenly. The pilots knew this dynamical principle and PF anticipated the punch line of the instructor's story.

Throughout the clip the pilots collaborated with the instructor producing what would conventionally be called verbal and gestural backchannel behavior<sup>5</sup> (*Yngve, 1970*). At times, one or another of the pilots became the most active speaker, and when they did they used their words and their bodies together to create multimodal utterances. When pilots spoke, the instructor produced conventional verbal and gestural backchannel behavior. However, in this clip we observe many instances in which multiple speakers are simultaneously active producing representational elements in different modalities. We call the utterances produced in this way, collaboratively constructed multimodal utterances.

The pilots produce a few single-author utterances without any visible gestural accompaniment. While all verbal utterances in this setting are accompanied by coordinated facial expression, body posture, eye gaze, and so on, we will treat as multimodal only those utterances that incorporate both spoken language and meaningful gesture. All collaboratively constructed utterances in this clip were also multimodal in the sense that the participants produced coordinated talk and meaningful content-bearing gesture.

### ***Multimodal utterances***

Since coherent meaning structures are created by multiple utterances we organize the presentation by cases, rather than by utterances. A single conceptual object is created in each case. Each case is given a number and a brief descriptive title. Following the case title, we show, in brackets, the time boundaries of the case in the clip and note whether it primarily contains single-author multimodal utterances, denoted by the letter S, or collaboratively constructed multimodal utterances, denoted by the letter C. We then give a concise description of the conceptual object that is constructed in the case. Excerpts from the transcript are provided with each case<sup>6</sup>.

#### ***Case 1: Enacting the procedure as read [00:03 – 00:13; S]***

Conceptual object: A specific sequence of actions to be performed by PF presented in the imperative mood. The instructor read the steps shown in the FCOM.

## *CollaborativeMultimodal*

I: Level the wings ↑,

*[models airplane roll attitude with right hand palm down]*

PF: Hmmm

I: right,

I: don't change your flap or landing gear configuration↑,<sup>7</sup>

*[raises fingers of right hand and wags to right]*

PF: Hmmm, hmmm

I: right↑?

*[drops hand to table. Pilots nod]*

Retract speedbrake.

*[right hand models pushing speedbrake lever forward and down]*

Hopefully that's not (0.1), not gonna be a problem. (1)

*[raises right hand wags to right, then drops to table]*

*[PF shakes his head side to side]*

While reading “level the wings” the instructor used the standard hand shape for aircraft attitudes, palm down, fingers slightly spread. This is one of many conventional gestural forms shared by pilots around the world. Note that there is no need to describe the control input that will level the wings.

We believe that the hand wag to the right produced in synchrony with the word “configuration” was a metaphorical echo of the negation in the word “don’t.” In the procedure as it was read, syntactic constraints kept the negation near the verb “change” and far from the noun “configuration.” The gesture brought that negation into temporal proximity with the noun. This illustrates how gesture, operating under different constraints from speech, can produce multimodal utterances with semantic juxtapositions that are not possible in speech alone. The gesture has two meaningful relations to elements in the speech stream. The gesture is congruent with and temporally offset from the negation in the word “don’t”. At the same time, it is produced concurrently with and has a complementary semantic relation to the word “configuration”.

The gesture that the instructor produced with the words “retract speedbrake” modeled the manipulation of a control, rather than an airplane attitude. This was an iconic character viewpoint gesture produced concurrently with and is semantically congruent with the phrase “retract speedbrake”. This gesture seems idiosyncratic and demonstrates the productivity of pilots using speech and gesture to imagine interaction with their familiar flight deck environment. The pilots’ response to the utterance, “Hopefully, that’s not (0.1) not gonna be a problem. (1)” was deeply

embedded in the setting. The pilots know that when practiced in a simulator, the recovery from departure stall maneuver is normally entered with the speedbrake already retracted. Of course, if a stall were approached accidentally in some other phase of actual flight, the speedbrake might be extended. When the instructor commented on the procedural element, “Retract the speedbrake”, saying “Hopefully, that’s not (0.1) not gonna be a problem. (1)” the nodding pilots indicated that they understood that they would not have to retract the speedbrake when they practice the departure stall maneuver in the simulator (and this was the case).

**Case 2: Thrust is set [00:13 - 00:16; C]**

Conceptual object: A specific sequence of actions to be performed by PM presented in the imperative mood.

The instructor continued reading from the FCOM on the table in front of the pilots. At this point he was reading descriptions of actions to be accomplished by the PM. All the while, the practice procedure was shown on the computer display. (See figures 1 and 2.)

I: (reading) PM verifies thrust is set,  
monitors altitude and airspeed,  
[PM raises his left hand and spreads his fingers and thumb]  
calls out any trends towards terrain.

The instructor began his utterance with the acronym used to designate Pilot Monitoring, PM. He then slightly paraphrased the instructions transforming “Verify maximum thrust” into “verifies thrust is set”. We do not know why the important conceptual element “maximum” was omitted from this representation. Of course the pilots know that maximum thrust is the appropriate setting and they can read the words in the printed procedure highlighted by the instructor’s finger as he paraphrased the procedure. At this moment, three representations of the same action were present in the setting. The text in the FCOM read “Verify maximum thrust”, the instructor has said “verifies thrust is set”, and the procedure shown on the computer screen said “Smoothly apply MAX power (PM adjust to GA)”. While the instructor spoke the words “monitors altitude,” PM made a gesture that modeled the manipulation of the thrust levers. PM raised his left hand and spread his fingers and thumbs into the hand shape characteristic of

grasping the two thrust levers<sup>8</sup>. There was a slight downward jerk of the hand before it was smoothly retracted to PM's lap.

It is clear that the gesture does not enter a relation of mutual elaboration with the words about monitoring altitude that were being spoken when it was produced. If we were using the old categories, we would say that there was simultaneous speech and gesture, but the speech was not "co-expressive". The gesture that PM produced as the instructor read from the FCOM the words, "monitors altitude" might however enter into a relation of mutual elaboration with any or all of the three representations of the action to be taken by the PM with respect to the thrust.

With respect to the temporal relationships, the gesture was clearly offset in time from the spoken words, "verifies thrust is set". The temporal relationships between the gesture and the two printed representations of the action are more difficult to assess. Since we cannot see the pilot's eyes we do not know when or if he looked at the printed representations.

The pragmatic relation of the gesture to the spoken words "verify thrust is set" is complementary. Specifically, it is a synecdoche because moving the thrust levers (enacted in gesture) is part of the complex perceptual/motor process of verifying thrust (described in speech). Notice that this event is not captured by the traditional categories. The spoken words "verify thrust is set" are a poor example of "lexical affiliate" because they do not refer to the same concept as the gesture. Furthermore, because the speech and gesture do not co-occur in time, the words cannot be "co-expressive speech" with respect to the gesture. The gesture and talk are thus semantically complementary and offset in time.

The technique shown on the computer display described the same action by specifying that the PM should "adjust [thrust] to GA." This means to adjust the thrust to the Go-Around Thrust limit, which will usually require pulling the thrust levers back slightly from the full forward position that the pilot flying will have pushed them to. The slight downward jerk in PM's gesture matches the motion required to produce the anticipated minor reduction in thrust. The gesture bears an iconic relation to the word "adjust" shown in on the computer display. Thus, the words on the computer display are semantically congruent with the gesture and because the text on the screen is continuously available, they may be temporally concurrent as well.

The relation experienced by PM while performing the gesture might have been either, or even both, the semantically congruent (iconic) relation to the written word "adjust" and the

semantically complementary (synecdoche) temporally offset relation to the spoken word “verify.” On the basis of the data we have it is not possible to eliminate either of these hypotheses.

PM’s gesture was a demonstration of his understanding of the required action, and also a pre-enactment of the action he would take in the upcoming simulator session. It presupposed his role as Pilot Monitoring and the details of his planned method of verifying that maximum thrust was set. What accounts for the lag between the instructor’s verbal element “thrust is set” and PM’s gesture elaborating the same concept? PM began leaning back from the table to free the motion of his right hand as soon as the instructor said “set.” It appears that PM’s gesture was a direct response to the instructor’s words. The gesture followed the words, having been triggered by them. This sort of gestural “follow-on” indicates that the listener inhabits a conceptual world that is constructed in response to what the speaker has already said.

**Case 3: Flaps 5 speed [00:48 – 00:54; C]**

Conceptual object: Airplane dynamics; accelerate an airplane from stall speed to flaps 5 speed, constructed from the point of view of the crew<sup>9</sup>.

I: For us, if we start out flaps five (1) sp..

[looks to computer monitor and right index point to it]

Let’s go to flaps five.

[raises right arm and opens right and left palms to make “five”]

flaps five speed

[PF positions his hands as if to hold a yoke and pushes forward]

that’s what we are gonna go to, okay?

[shakes right and left hands rhythmically]

[scratches left elbow with right hand]



The instructor resumed reading the procedure from the FCOM, tracing the text with his left index finger as he read, “Return to speed appropriate for the configuration.” He then looked at the computer monitor and pointed to highlight the portion of the procedure described by the words “Finish: FLAPS 5 speed” (See figure 2). The instructor elaborated this part of the maneuver and as he withdrew his right hand from a full hand point to the procedure shown on the computer monitor, he said “flaps five speed.” Simultaneously, PF positioned his hands as if holding the control yoke and pushed them forward (see figure in transcript above). This gesture

enacted the control input needed to return to flaps 5 speed. Notice that the instructor’s utterance does not specify the sort of control input that will be needed to return to flaps five speed. The pilot knows that in order to accelerate, he will have to push the yoke forward. Thus, the gesture provides the cause that is not present in the verbal description of the effect “go to flaps five, flaps five speed.” The words provide the effect that is not present in the gesture. The two elements mutually elaborate each other as a metonymic cause and effect relationship. This is a collaboratively constructed multimodal utterance in which the instructor’s speech and the pilot’s gesture are temporally concurrent and semantically complementary.

What functional role did PF’s gesture play in the activity? At that moment, the instructor was using his body to highlight relevant information, and not to imagine going to flaps 5 speed. This may have created a conceptual void in the interaction that PF’s gesture filled. The cognitive ecology of the pre-simulator briefing suggests another cognitive function for this gesture. Since PF was representing a component of the procedure that he would later execute, it might also be a sort of pre-enactment that could facilitate memory for the procedure later. This effect also appeared in case 2.

**Case 4: Back pressure only [01:11 – 01:19; C]**

Conceptual object: To decelerate an airplane in level flight, reduce power and hold back pressure, constructed from the PF’s character viewpoint.

PM: Yes, I know difference between

Boeing and (Company X)’s procedure.

Our procedure just trim out at flap five speed (0.2)

I: and then \_\_\_\_\_

[makes two fists to represent holding yoke and pulls toward his chest]

PM: keep back pressure only,

[I continues holding his two fists near his chest]

not applying any more trim.



The instructor produced a verbal element “and then” that was a continuation of PM’s verbal utterance, “Our procedure just trim out at flap five speed.” By itself, the instructor’s continuation “and then” would create a collaboratively constructed verbal utterance. But the instructor also simultaneously gestured to model pulling back on the yoke (see figure in transcript above). The added gesture here fills in content for the projection of his own words

“and then” and projects a conceptual structure for PM to complete verbally. Every pilot knows that if you do not trim to decelerate, you must pull back on the yoke<sup>10</sup>. The instructor used that knowledge to anticipate the projection of PM’s words with his gesture. This projection was especially well marked as PM had stated that his company’s procedure is different from Boeing’s procedure, which they had discussed and which involves trimming to the stall speed. Thus, the instructor’s gesture is semantically congruent with and temporally anticipates PM’s spoken words “back pressure”. The gesture also has a relation of mutual elaboration to the concurrently produced words, “and then.” The semantic relation here is complementary (synecdoche) because the talk represents a sequence in which the back pressure enacted in gesture is a component action. The initiation of the instructor’s gesture was anticipatory, but he held it while PM continued speaking, saying “keep back pressure only.” By the end of this statement, the instructor’s gesture and PM’s speech were semantically congruent and temporally concurrent.

**Case 5: It’s realistic the Boeing way because [01:29 – 01:38; C]**

Conceptual object: A comparison of techniques, from two implied character viewpoints, PM (a pilot looking down on the stabilizer trim indicator) and instructor (pilot flying).

PM: But a, it’s realistic the Boeing way. (0.5)

Because, ah

[I makes trim gesture]

PF: Hmmm.

PM: We always manage to keep our trim...

[I nods continuously]

I: Uh, huh.

[nods continuously]

PM: ( ) you know, forward out of habit.

[I nods continuously]



By using the words, “the Boeing way,” PM refers to the earlier discussion which established the contrast between his company’s technique of entering a stall recovery maneuver using back pressure on the yoke only (no trim) and the “Boeing way” which involves trimming as the airplane slows on the stall entry. Framing the topic as “It’s realistic the Boeing way” constructs an implicit comparison between the techniques. The instructor knew this and the movement of his right thumb models the action that PM would take as a pilot flying when he

trims the airplane. The entire conceptual schema was clear at the pause before the word “Because”. Possible projections included elaborating on either the realistic or the not realistic method. Thus, the instructor’s gesture is an iconic representation of an anticipated spoken description of the realistic method. This case is interesting because the gesture seems to have a relationship to an anticipated spoken representation that never actually occurred. We could even say that the gesture is positioned and formed to facilitate the production of a verbal element with which it could be both temporally concurrent and semantically congruent.

The gesture also has a relationship of mutual elaboration to the concurrently produced spoken word “Because”. This relation is semantically complementary (metonymic) because the gesture represents a cause (trimming) for the effect (realism) that is the basis of the difference in the comparison schema. It soon became clear that trimming was not the aspect of the comparison schema that PM went on to elaborate, and the instructor quickly abandoned the trimming gesture.

This gestural mismatch may have happened for one of two reasons. First, PM’s projection of a reason for Boeing realism could have been illustrated with either a feature of the Boeing technique or by a feature of PM’s company’s technique. In choosing to model a feature of the Boeing technique, the instructor may have simply mistaken which continuation PM was projecting. However, the situation could be even more interactive. A second reason for the mismatch is that PM may have also been projecting a feature of the Boeing technique, but once this had been created by the instructor in the collaborative construction process, PM was free to provide the other meaningful completion. This interpretation relies on something like the Grice’s (1981) maxim of quantity. Since the instructor had already illustrated the distinctive feature of the Boeing way, PM could increase the informativeness of his contribution by describing the distinctive feature of his company’s procedure. “We always manage to keep our trim forward, you know, out of habit.”

PM can refer to this as keeping the trim “forward” because the trim indicator is mounted on a horizontal surface at either side of the center console. On that indicator, airplane nose down trim is forward, and nose up trim is aft (See figure 3).

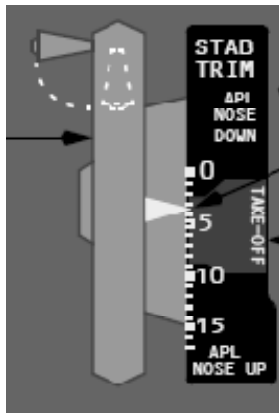


Figure 3: The 737NG trim wheel and stabilizer trim indicator as seen looking down from the vantage of the left pilot's seat. Up on the page is forward in the airplane.

Notice that when talking to PM, who would occupy the right seat in the simulator, the instructor modeled the trim action using his right thumb. The yoke-mounted trim switch is on the outboard horn of each control yoke. Thus, for a pilot in the right-hand seat (co-pilot's seat), the trim switch will be under the right thumb. Later in the same discussion, the instructor modeled pushing the thrust levers up with his right hand. This gave his gesture an implicit body location in the left seat (captain's seat) and his subsequent gestural reference to trim was made with the left thumb. This coherence of gesture indicates that the imagination of component actions, such as thrust changes and trim adjustments, involves the whole situation of the body in the flight deck, not just imagining the control that is to be manipulated.

**Case 6: You have to push [01:45 – 01:52; S]**

Conceptual object: To recover from stall attitude push the yoke forward to cause nose down pitch attitude. This was constructed from the PF character viewpoint.

The instructor created a role-playing narrative in which he modeled an inattentive pilot trimming into a stall. As the instructor finished his narrative, he continued to model the application of nose-up trim. PM began the following utterance over the end of the instructor's narrative.

## CollaborativeMultimodal

PM: It's very really difficult \_\_\_\_\_

[models pushing the yoke]

[I stops modeling nose up trim]

to get Ah, (0.5) back to normal (0.2)

[right hand offer shape]

nose down

[models pushing the yoke]

PF: To ah:::: to recover from

[looks toward and flicks right fist toward PM]

PM: because [ ] you have to push

[models pushing the yoke again]

[I nods and points at PM with his right index finger]



This complex example integrates seven gestures and five spoken elements. A full inventory of the relations among these elements is beyond the scope of this chapter<sup>11</sup>. All of the spoken elements and three of the gestures refer to the core conceptual object being constructed. Of the other four gestures, one refers to a previously developed conceptual object, one solicits agreement from another speaker and two provide assessments of other speaker's conceptual project.

Let's look first at the three push gestures produced by PM. While saying "It is very really difficult" PM modeled pushing the yoke forward. PM repeated the yoke pushing gesture while saying "nose down". Finally, he said "because you have to push" accompanied by a third pushing gesture. Each of the pushing gestures modeled pushing the yoke forward and all are semantically congruent with the spoken words, "to push" that occur at the end of the utterance. PM thus produced two anticipatory gestures followed by a third one that was produced simultaneously with the talk it elaborated. There are three content nodes represented in the speech stream. Each bears a different semantic relation to the conceptual content of the push gestures. The pilot action required to accomplish the recovery is represented by the spoken fragment, "you have to push". This spoken element bears a congruent relation to the push gestures. Pushing the yoke forward causes a nose-down pitch attitude represented in the spoken fragment "nose down". This spoken element bears a complementary (metonym) relation to the push gestures. The recovery itself is represented by three spoken fragments: "It's very really difficult", "back to normal", "To ah:::: to recover from". These spoken elements bear a semantically complementary (synecdoche) relation to the push gestures.

In the beginning of this case, while the instructor was producing an iconic gesture as follow-on to his previous narrative, PM changed the subject. The instructor stopped his trimming gesture after PM said “difficult.” At this point, he seemed to recognize the topic shift. The unexpected topic change created a relation of incongruence between gesture and the concurrent speech. This was not without meaning, however, because the alignment of conceptual projects is an indication of membership in a shared community of practice. This was the second time that the instructor had anticipated a projection of PM’s utterances that was not consummated (the first happened in case 5). The conceptual projects of instructor and PM seemed less well aligned than those of the instructor and PF. This sort of interaction pattern may give rise to a feeling of uncooperativeness for which it is difficult to articulate an explanation.

**Case 7: Under slung engines [02:19 – 02:27; C]**

Conceptual object: Airplane dynamics, airplanes with engines mounted under the wings tend to pitch up when thrust increases. The instructor's gestures were constructed from a character viewpoint taking the speaker's body to be the airplane. PF's gestures were constructed from an observer viewpoint above and behind the wings, facing forward.

I: once those engines,

[cuppéd hands at side below shoulders]

they are under slung engines,

[two beats with cupped hands at side below shoulders]

right?

The engines, these.

[arms extended out above previous engine gesture location]

PF: (unintelligible)

[fingertips of both hands rotate up quickly]

I: So, it's gonna .

[bends forward bringing wing gesture down]

PF: tend to, yeah

[flicks fingertips up again]

[I: entire body and arms come up]

I: hhh.

It's gonna tend to (0.3)

[bends at waist and lowers arms]

sling this airplane up.

[entire body and arms come up again]



Like the previous example, this one is so complex that a full analysis is not possible here. In this case, all but one of the spoken elements and all of the gestures participate in the construction of the conceptual object. We can simplify the discussion somewhat by noting that the conceptual object has two principal parts: the location of the engines under the wing, and the pitch-up moment created by increasing thrust on engines that are so located. The instructor constructs the engine location by himself. The resulting pitch-up moment is collaboratively constructed by PF and the instructor.

“Underslung” describes a relationship between engine and wing. To create the relationship, the instructor used his body to enact the key parts of the airplane. The instructor's first cupped-hands gesture was a relatively simple iconic representation produced concurrently

with the semantically congruent words “those engines”. The gesture was idiosyncratic and would have been quite ambiguous if taken in isolation. The words and gesture mutually elaborated each other. The words resolved the referent of the gesture (hands are engines), and the gesture contributed positioning information (the two engines are located in an imagined space here) that was not present in the words. With the engines now located in an imaginary body-based space, the instructor elaborated on their location, simultaneously emphasizing the cupped hand gestures while saying “they are under slung engines”. This gesture and spoken fragment have a concurrent complementary relation; the gesture anchored the engines in a space, and the words implied something else (a wing) that had not yet been explicitly represented. The instructor then extended his arms out to the sides of his body, giving explicit representation to the previously implied wing, and said, “the engines, these.” This gesture was positioned in space above the previously depicted location of the engines. While the space implied by the previous gestures and talk, was completely invisible and imaginary, it endured as a resource that could be exploited by subsequent meaning making activities. These words and gesture had a complementary semantic relation (gesture depicted the wing while speech referred to the engine) and were temporally concurrent. At this point, the construction of the location of the engines with respect to the wing was complete. The fact that the space that was constructed by earlier actions could later give meaning to new gestures demonstrates that this discussion of pair-wise relations is fundamentally incomplete. We have picked out what appear to us to be the most significant relations, but our description remains partial because all of the elements of this complex semiotic field have important semantic and temporal relations to one another.

In the context of the discussion of the need to apply maximum thrust, the instructor’s multimodal construction of the location of the under-wing location of the engines projected a pitch up in response to a rapid increase in thrust. PF used his two hands to model the rotation in the pitch axis caused by the increasing thrust on the two engines (see figure in transcript above). His enactment was quite specific, showing the two engines and the torque that they would apply to the wings of the airplane when thrust was increased. Simultaneously, he said something which we have not been able to reconstruct. PF’s gesture may have had congruent semantic relations with two spoken elements, one produced concurrently by PF himself, and the other anticipated in the speech of the instructor. The instructor continued to develop his narrative, saying “So, it’s gonna” while bending at the waist with his arms still extended to his sides. PF

seemed to recognize this as preparation for a full-body stroke. A moment later, as the instructor swept his body and arms upward, PF flicked his fingers up again and said, “tend to, yeah”. This gesture by PF’s is semantically congruent with an anticipated, but not yet produced, description by the instructor of the airplane pitching up. This gesture is temporally concurrent and semantically complementary (metonym) with PF’s own words “tend to”. PF performed this gesture in synchrony with the instructor’s full-body upward stroke. Thus in addition to relations with spoken elements that were produced before, concurrently with, and after the gesture, PF’s gesture also has a temporally concurrent and semantically congruent relation to the gesture produced by the instructor. Both gestures provided an iconic representation of the pitch up event, but they were rendered from slightly different viewpoints<sup>12</sup>. PF’s utterance fragment “tend to, yeah” has a temporally concurrent and semantically complementary (metonymic: cause and effect) relation to the instructor’s first full-body gesture. The instructor’s second sweeping full-body gesture was produced concurrently with his own, now eagerly anticipated, verbal description of the pitch up event, “sling this airplane up.”

It is evident that when multiple authors speak and gesture together, the relationships of mutual elaboration proliferate. The extent to which participants become conscious of this wealth of meaning is currently unknown. We suspect however, that the impression of complexity created by examining the relations among semiotic resources one relation at a time is somewhat misleading. From the participants’ point of view, a single conceptual object emerges and the many relations among the elements from which the object emerges fit naturally into the familiar structure of the conceptual object<sup>13</sup>.

## **Discussion**

The participants are engaged simultaneously in two kinds of projects: they are enacting conceptual objects of interest (what they are talking about), and they are conducting a social interaction. While these projects are analytically separable, in action, they are woven into the same fabric. This was evident in case 6 where three of seven gestures modeled conceptual content, while the other four gestures accomplished speaker positioning in the interaction.

Surely pilots can imagine their work without speaking or gesturing. However, when they speak and gesture, the process of imagination becomes observable. This is important for the

participants, because it allows them to collaboratively construct conceptual projects. It is critical for us as analysts because it enables us to record and analyze the process of conceptualization.

Gesture, talk, printed words and material objects all have different representational affordances. Imagining an activity by simultaneously talking and gesturing about it produces a richer representation than is produced by either talk or gesture alone. In addition to modeling specific actions, many of the observed gestures presupposed specific flight deck roles, the seat occupied while performing the imagined action, and the fine details of the bodily motions of the pilot. Such details are rarely represented linguistically in our data. The coherence of gestural enactments indicates that the imagination of component actions involves the whole situation of the body in the flight deck, not simply imagination of the control that is to be manipulated. The richness and specificity of the pilot's shared knowledge of the flight deck environment is evident in the rapid shifts in viewpoint implied by the gesture sequences. Pilots transition seamlessly from character viewpoint to observer viewpoint, and among multiple vantage points as observers.

One way to bring relations of mutual elaboration into focus is to notice what does NOT appear in talk. For example, the control yoke, the trim switch, and the thrust levers play central roles in the interaction, yet these controls were never mentioned in the verbal utterances produced by the instructor and students. The controls are brought forth as implied elements in an imagined world of culturally meaningful action. The words "you have to push," could apply to many controls in the flight deck. That these words describe an action taken on the control yoke is established by their relation of mutual elaboration with particular gestures.

Gestures may enter into relations of mutual elaboration with many other semiotic resources in the activity system; written materials, objects, bodies, talk and even other gestures. Gestures are complex movements. Which aspects of movement are taken to be relevant in the current moment of discourse depend on how the gesture is mutually elaborated by other semiotic resources. For example, recall the last gesture in case 1. The words "retract speedbrake" say nothing about how the retraction of the speedbrake is accomplished. The speedbrakes are panels on the wings. Where is the activating control? How is the control operated? The instructor held his right fist upright in front of his body at elbow level. As he moved his hand forward his wrist rotated down slightly. If this gesture were to occur alone, even in this context, its meaning would probably be misunderstood. Viewed without sound, the gesture could easily be seen as modeling a pilot in the right seat pushing the right horn of the control yoke forward. But the

gesture co-produced with the words “retract speedbrake” in this context brings forth an unambiguous whole. A pilot seated in the left seat of the flight deck uses his right hand to grasp the raised speedbrake handle and push it a few inches forward and down. Details of the motion that did not seem important when viewed without sound now jump out. The speedbrake handle rotates around a hinge at its base, and this detail is shown in the gesture as the slight rotation of the wrist. Furthermore, details of the gesture that should be ignored fade away. In the airplane, the speedbrake handle is adjacent to the pilot’s right thigh. A gesture that perfectly modeled speedbrake retraction would be performed below the pilot’s waist. But the surface of the table intervenes in the instructor’s local space preventing him from lowering his hand further. In mutual elaboration with the talk, the height of the gesture can be disregarded. This is a reminder that even seemingly simple gestures may be extremely complex. What is meaningful and what is not, what should be attended to and what should be disregarded as noise, depends on how the gesture is construed. And the level of detail that can be achieved in the construal depends on the depth of knowledge that the participants have about the domain of discourse. It is not just the words and gesture that mutually elaborate each other. The words and the gesture enact or bring forth a meaningful action in a known world. In the presence of the talk, an ambiguous body motion becomes a detailed model of a meaningful action. This example demonstrates the productivity of pilots using speech and gesture to imagine interaction with their familiar flight deck environment.

In the domain of professional pilot training, the participants use gesture to represent activities, objects, and events with respect to which all of the participants have thousands of hours of experience. Extensive embodied experience results in rich representational potential. Representational potential is realized in the enactment of the concepts in word and deed. Some of the meaningful flight deck actions are enacted so often and so distinctively that the gestures derived from the actions attain the status of conventions in the community. The “retract speedbrake” gesture produced by the instructor is not so widely used as to be considered a convention. Conventional status depends on the specificity of the gesture and its relations to other forms in the eco-system (*Hutchins & Johnson, in press*). Nothing else that is done in the flight deck looks like holding two thrust levers (with a characteristic hand shape<sup>14</sup>) and pushing them forward in a vertical arc that models the arc of the thrust lever quadrant. This character viewpoint gesture contains elements of both path and manner. The specificity of the gesture also

depends on the standardization of the flight deck. Thrust levers and throttles are nearly universal in transport aircraft. Since virtually every airline pilot experiences the thrust levers in the same way, and since the bodily motions associated with manipulating the thrust levers are distinctive, this motion has gained the status of an iconic convention in the pilot community. Control yokes are not as widely distributed as thrust levers (having been replaced by side sticks in Airbus airplanes), but are still present in most airplanes and are understood by all pilots. The presence, position and activation of many other controls are more variable across the world's airplane fleets, and so, while the manipulation of these controls can be meaningfully enacted by pilots in context, they do not stand as interpretable context-independent iconic representations.

We have seen that the pilots use talk, gesture and other semiotic resources to imagine their domain of discourse. We do not think this is a matter of visual imagination followed by linguistic and motor activation. Rather, talking and gesturing *are* a means of imagining (*Alac & Hutchins, 2004*). It has long been known that the visual and motor systems co-activate each other (*Smith, 2005; Spivey, 2007*). One might say that they mutually elaborate each other.

Goodwin (*1994*) (*2007*) coined the phrase “environmentally coupled gesture” and showed how gestures can enter into relations of mutual elaboration with elements of a culturally meaningful physical world. Phenomena in the world are highlighted by and acquire meaning from gestures enacted in coordination with them. Simultaneously, gestures acquire meaning from the elements of the physical world with which they are coordinated. Of course, environmentally coupled gesture is pervasive when pilots work together in a flight deck (*Hutchins & Palen, 1997; Hutchins, Middleton, & Newsome, 2009*). Many of the gestures we observe in the pre-simulator briefing mutually elaborate physical elements of the briefing setting. But what of the gestures that refer to the absent flight deck? The fact that pilots have so much experience of this setting changes the dynamics of these processes. Once it has been invoked in speech or gesture, the entire flight deck becomes available (in imagination) as an environment to which subsequent gestures can be coupled. The same processes that are at work in meaning making with environmentally coupled gesture are at work here, except that these gestures both bring forth the imagined environment and are coupled to elements of that imagined environment. As we saw in the case of retracting the speedbrake, a gesture can selectively highlight elements of an imagined environment, while the imagined environment simultaneously draws attention to and gives meaning to subtle details of the gesture.

Gesture provides evidence that imagination can run ahead of talk (*Schegloff, 1984*). In case 6, PM made three yoke-pushing gestures, but did not verbally describe the push action until the third gesture, seven seconds after the first push gesture was produced. The first two push gestures anticipated the semantically congruent spoken words. They were produced concurrently with semantically complementary elements of a verbal preamble that contextualized the pilot's stance with respect to the recovery maneuver (it's difficult) and with respect to the effect of the push (nose down attitude). The third push gesture was produced concurrently with the words "to push." One consequence of repeating the gesture is that it kept the main point active while the verbal preamble was delivered. That is, the pilot was clearly imagining the push action seven seconds before he got around to describing it verbally. The syntactic constraints of language impose sequential order on the articulation of conceptual elements. Gesture that anticipates one's own talk is a constituent of this pre-articulatory imagination.

Before doing this analysis, we would have guessed that gestures that follow the production of semantically congruent words (as seen in case 2, for example) would be the most likely timing relation for collaboratively constructed multimodal utterances. Gestural follow-on assumes that the listener inhabits a conceptual world that is constructed *in response to* what the speaker has already said. Sometimes however, gestures in collaboratively constructed multimodal utterances occur concurrently with the words they elaborate. In case 7, PF and the instructor executed perfectly synchronized, but morphologically distinct, enactments of a sudden pitch up attitude. PF's gestures were performed in anticipation of the instructor's subsequent metaphorical description of the pitch up event. The cross-speaker production of such multimodal elements in precise temporal and conceptual alignment requires joint participation in the embodied construction of this key conceptual element. It is further evidence that the participants jointly inhabit the world they imagine in interaction. Simultaneity of cross-speaker gesture and talk in collaboratively constructed utterances is evidence of a shared activity and aligned expectations. This relation indicates that the speakers inhabit a shared conceptual world that is constructed in parallel. As in single-speaker utterances, gestures in collaboratively constructed multimodal utterances often precede the spoken elements to which they bear semantic relations. The production of utterances in which a one speaker's gestures anticipate the conceptual projections of another speaker's words provides strong evidence that speakers can inhabit a shared imagined world.

The details of such imagined worlds are built up incrementally as the semiotic resources of the setting are marshaled in interaction. When PM said (in case 4), “Our procedure just trim out at flap five speed” he evoked an imagined world of a pilot-flying preparing for the maneuver. The word “just” signals the absence of the further trimming below flaps five speed that the previous discussion led one to expect. The instructor’s next character viewpoint gesture showed that he had entered the imagined role of pilot flying created by PM. The instructor filled the projected conceptual hole by enacting the next part of the maneuver. He said, “and then” while modeling pulling back on the yoke.

Not all conversational projections are consummated. In case 5, PM introduced a conceptual scheme (a comparison) that could be developed in either of two ways. The instructor produced a gesture that committed to one projection of what PM had said. PM went on to articulate the other projection. We have no evidence concerning PM’s original intentions, but his action does suggest that in the dynamic process of co-authorship of ideas, participants make choices in real time based on the shifting direction of the development of the conceptual object.

The occurrence of collaboratively constructed multimodal utterances indicates that the pilots treat the development of the conceptual object as a shared project. The properties of this ecosystem create particular cognitive roles for gestures. In the cognitive ecology of flight training, some gestures seem to be pre-enactments of actions that will be taken later in flight. We stress that when a behavior has more than one function, it may be that many functions are served simultaneously.

The pilots’ bodies are a key resource in the process of conceptualizing their world and the actions they take in it. Conceptualization is not only multimodal, but may also be a collaborative project. The range of possible relations of mutual elaboration among semiotic resources is extremely rich in collaboratively constructed multimodal utterances produced by experts engaged in consequential activity in a culturally constructed setting. Meanings emerge from juxtapositions of gestures and words with material artifacts, with one’s own body and the bodies of others, with one’s own words and the words of others, and with one’s own gestures and the gestures produced by others.

## ***Acknowledgements***

Access to the field site was arranged through and data collection was supported by a contract with Boeing Flight Deck Concepts Center. Barbara Holder served as contract monitor. Whitney Friedman created the cartoon representations of the video frames. We are grateful to Charles Goodwin and Susan Goldin-Meadow for reading early drafts of the paper and providing expert advice. Any errors that remain are our own. Funding for the data analysis was provided by NSF award #0729013, “A multiscale framework for analyzing activity dynamics”, James Hollan, Edwin Hutchins, and Javier Movellan, principal investigators. Finally, we are especially grateful to the many pilots and instructors who have participated in our research.

## References

- Alac, M., & Hutchins, E. (2004). I see what you are saying: action as cognition in fMRI brain mapping practice. *Journal of cognition and culture* , 4 (3), 629-661.
- Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.
- Goodwin, C. (2007). Environmentally coupled gestures. In S. Duncan, J. Cassell, & E. Levy (Eds.), *Gesture and the dynamic dimensions of language* (pp. 195 - 212). Amsterdam/Philadelphia: John Benjamins.
- Goodwin, C. (2006). Human sociality as mutual orientation in a rich interactive environment: Multimodal utterances and pointing in aphasia. In N. Enfield, & S. Levinson (Eds.), *Roots of human sociality* (pp. 96-125). London: Berg Press.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist* , 96 (3), 606 - 633.
- Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: studies in ethnomethodology* (pp. 97 - 121). New York: Irvington.
- Goodwin, C., Goodwin, M., & Yaeger-Dror, M. (2002). Multimodality in girls' game disputes. *Journal of pragmatics* , 24 (10-11), 1621 - 49.
- Grice, P. (1981). Presupposition and conversational implicature. In P. Cole (Ed.), *Radical Pragmatics* (pp. 183 - 198). New York: Academic Press.
- Hutchins, E. (in press). Cognitive Ecology. *Topics in Cognitive Science* .
- Hutchins, E., & Johnson, C. (in press). Modeling the emergence of language as an embodied collective cognitive activity. *Topics in Cognitive Science* .
- Hutchins, E., & Palen, L. (1997). Constructing meaning from space, gesture, and speech. In L. Resnick, R. Saljo, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools, and reasoning: essays on situated cognition* (pp. 23-40). Heidelberg, Germany: Springer-Verlag.
- Hutchins, E., Middleton, C., & Newsome, W. (2009). Conceptualizing spatial relations in flight training. *Proceedings of the 15th International Symposium on Aviation Psychology*, (pp. 384 - 389). Dayton, OH.
- Hutchins, E., Nomura, S., & Holder, B. (2006). The ecology of language practices in worldwide airline flight deck operations: the case of Japanese airlines. *Proceedings of the international conference on human-computer interaction in aeronautics*, (pp. 90-96). Seattle, WA.

- Jacoby, S., & Ochs, E. (1995). Co-construction: an introduction. *Research on Language & Social Interaction* , 28 (3), 171 - 183.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. New York: Cambridge University Press.
- Langewiesche, W. (1990/1944). *Stick and Rudder: an explanation of the art of flying*. New York: McGraw-Hill.
- Maynard, S. (1986). On back-channel behavior in Japanese and English casual conversations. *Linguistics* , 24, 1079 - 1108.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- McNeill, D. (1992). *Hand and Mind: what gestures reveal about thought*. Chicago: University of Chicago Press.
- Schegloff, E. (1984). On some gestures' relation to talk. In J. Atkinson, & J. Heritage (Eds.), *Structures of Social Action* (pp. 266 - 298). Cambridge University Press.
- Smith, L. (2005). Action alters shape categories. *Cognitive Science* , 29, 665 - 679.
- Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.
- Yngve, V. (1970). On getting a word in edgewise. *Papers from the 6th Regional Meeting, Chicago Linguistics Society* (pp. 567 - 578). Chicago: Department of Linguistics: University of Chicago.

## Notes

---

<sup>1</sup> The NG (Next Generation) is an updated 737 model with new engines, wing profile, on-board systems, and flight deck displays.

<sup>2</sup> First author Hutchins has twenty years of experience studying commercial air operations worldwide. He holds a commercial pilot certificate with type ratings in an airliner and a business jet. Second author Nomura has four years of experience studying airline operations and training in Japan and Oceania.

<sup>3</sup> An airplane stalls when the flow of air over the wings separates from the surface of the wing. When this happens, the wing ceases to produce lift. Pilots never practice taking an airliner into a full stall. Instead, they practice response to the first indications of an impending stall; thus the approach to stall recovery.

<sup>4</sup> The feel of traditional controls in small airplanes provides a lot of information about the behavior of the airplane. Controls feel crisp and firm at high speeds and get “mushy” when the airplane is going slowly. When an airplane begins to stall, the airflow on the surface of the wings is disturbed. This causes the control surfaces to flutter, and this is felt as vibration in the control stick or yoke. When hydraulic devices position the control surfaces, however, these vibrations are not transmitted to the yoke, so the pilot loses an important source of tactile information about the behavior of the airplane. Modern airplanes partly compensate for this loss by adding a device, called a “stick shaker,” that vibrates the control yoke as the airplane approaches a stall. This is one of the few concessions to multimodal perception in contemporary flight decks which are otherwise dominated by visual perception. Modern airplanes also provide visual indications of approach to stall including angle-of-attack indicators and in the airplane treated in this article, a “pitch limit indicator.”

<sup>5</sup> We are aware that the conventions for producing back-channel behavior are different for Japanese speakers than they are for English speakers (*Maynard, 1986*). We do not think these differences affect the arguments we make in this paper.

<sup>6</sup> Following Goodwin, we use a modified form of the Jefferson conventions for transcription. The three speakers are identified as “I” instructor, “PF”, and “PM”. Punctuation is used to represent intonation: A period indicates falling pitch, a question mark rising pitch, and a comma falling contour, as would be found for example after a nonterminal item in a list. A colon indicates lengthening of the current sound. Numbers within single parentheses mark silences in seconds and tenths of a second. Words within parenthesis indicate uncertain transcription. Underlining denotes words that are spoken in synchrony with gestures. Where video frames are illustrated a line from the transcript to the illustration indicates the temporal location of the frame.

<sup>7</sup> It is interesting that the procedure specifies something to NOT do. The expectation that pilots might want to change configuration comes from a general piece of pilot knowledge that when recovering from a stall, it is good to increase lift or reduce drag and that is what changing configuration does. Many airplanes include configuration changes in stall recovery procedures, but the 737 does not.

<sup>8</sup> This gesture is very distinctive, and while it is not common, it cannot be mistaken for any other action in the flight deck.

<sup>9</sup> The use of first person plural pronouns is very common in flight deck conversation. It is a form of metonymy in which the crew stands for the airplane.

<sup>10</sup> For the curious reader, a very accessible description of the basics of airplane behavior and pilot technique is Wolfgang Langewiesche’s (1990/1944) *Stick and Rudder: An Explanation of the Art of Flying*.

<sup>11</sup> In a set of 12 elements (five spoken and seven gestures) there are 66 pair-wise relations. Working out which of these relations are actually experienced by any of the participants is a difficult methodological problem. It cannot be done using the kinds of data we have collected here. It may be possible to probe for this experience in experimental settings using brain imagining techniques.

<sup>12</sup> It could be argued that gestures that have the same referent, but are rendered from different actor viewpoints should be regarded as semantically complementary rather than congruent. At this time, we do not have a strong view on the matter. Simply posing the question highlights the possibility that semantic congruence is a continuous rather than discrete function.

<sup>13</sup> A metaphor may help to make this idea clear. When a point is added to a chart, it immediately acquires precise spatial relations to every other one of a potentially infinite number of points already on the chart. This explosion of relations does not pose any problem for the navigator because the new relations are now potential and available. They can be easily accessed, but there is no need to attend to any except the ones that are relevant to the task at hand.

<sup>14</sup> The heel of the hand against the knobs, three middle fingers over the knobs, thumb and pinky on the opposite ends of the row of knobs where the autothrottle disconnect buttons are located.