To appear in Morsella, E. (ed.), *Expressing oneself / expressing one's self: Communication, language, cognition, and identity*. London: Taylor and Francis. 11/09/07

MIND-MERGING[1]

David McNeill, Susan Duncan, Amy Franklin,[2] James Goss, Irene Kimbara,[3] Fey Parrill,[4] Haleema Welji,[5] Lei Chen,[6] Mary Harper,[5] Francis Quek,[7] Travis Rose,[8] and Ronald Tuttle[9]

University of Chicago

Abstract

We focus on highly multimodal depictions of multi-party meetings (US Air Force war gaming sessions). The data we code are eclectic—chains of linguistic co-reference, gaze deployments, 'F-formations' (a category from Kendon), and parses of turn management, floor control, coalition formation and conflict. The concept of a hyperphrase ties all the threads together; the idea of a growth point is the ultimate theoretical unit to which it is all tethered, with the hyperphrase describing various surface ways that growth points, the cognitive units communication aims to share, materialize during ongoing interactions. The project is relevant to a comment Krauss & Pardo (2004) made regarding a BBS contribution by Pickering & Garrod (2004), namely, that while communication plausibly involves the alignment of speakers and their cognitive states, a move to reduce this to mechanistic priming excludes the reflective processes in dialogue.

Our emphasis in this paper is on 'floor control' in multiparty discourse. The approach is broadly psycholinguistic, a perspective that includes turn management, turn exchange and coordination; how to recognize the dominant speaker even when he or she is not speaking; and a theory of all this. The data to be examined comprise multimodal depictions of 5-party meetings (US Air Force war gaming sessions).

Multiparty discourse can be studied in various ways, e.g., as signals of turn taking intentions, marking the next 'projected' turn unit and its content, and still others. We adopt a perspective that emphasizes how speakers coordinate their *individual cognitive states* as they exchange turns while acknowledging and maintaining *the dominant speaker's status*. This goal is similar to Pickering & Garrod's interactive alignment account of dialogue (2004), but we add gesture, gaze, posture, F-formations (Kendon

[1] This research was supported by the Advanced Research and Development Activity (ARDA), Video Analysis and Content Extraction VACE II grant #665661 (entitled From Video to Information: Cross-Modal Analysis of Planning Meetings), Francis Quek, Mary Harper and David McNeill, principle investigators.

[2] Now at Department of Linguistics, Rice University, Houston, TX.
[3] Now at Kushiro Public University, Kushiro, Japan.
[4] Now at Department of Cognitive Science, Case Western University, Cleveland, OH.
[5] Now in the Peace Corps, serving in Jordan.
[6] At Speech & Language Processing Lab, School of Electrical & Computer Engineering, Purdue University West Lafayette, IN.
[7] At Center for Human-Computer Interaction, Virginia Tech, Blacksburg, VA.
[8] At National Institute of Standards and Technology, Gaithersburg, MD
[9] At Air Force Institute of Technology, Dayton, OH.

1990) and several levels of coreferential chains—all to be explained below. We adopt a theoretical position agreeing with their portrayal of dialogue as 'alignment' and of this alignment as automatic, in the sense of not draining mental resources, but not the type of 'mechanistic' (priming) account of it they advocate (cf. Krauss & Pardo 2004 for other qualms). The theory we are following is described in the next section. Alignment in this theory is non-mechanistic, does not single out priming, and regards conversational signaling (cf. papers in Ochs et al. 1996) as providing a synchrony of individual cognitive states, or 'growth points.' The role of gesture in all this is to produce converging imagery as well as speech as the means of socially constructed mutual cognition. Our aim is to analyze this cognitive 'reflective' process (from Krauss & Pardo 2004).

Theoretical background

The growth point. A growth point (GP) is a mental package that combines both linguistic categorial and imagistic components. Combining such semiotic opposites, the GP is inherently multimodal, and creates a condition of instability, the resolution of which propels thought and speech forward. The GP concept, while theoretical, is empirically grounded. GPs are inferred from the totality of communication events with special focus on speech-gesture synchrony and co-expressivity (cf. McNeill 2005 for extensive discussion). It is called a growth point because it is meant to be the initial pulse of thinking for and while speaking, out of which a dynamic process of organization emerges. It is thinking for speaking in the sense of Dan Slobin—the adjustment of one's thought to fit the affordances provided by the language one is using; it is thinking while speaking in a dynamic sense, that thought and language, as they unfold, are inseparable processes. Growth points are brief such dynamic processes, during which idea units take form. If two individuals share GPs, they can be said to 'inhabit' the same state of cognitive being and this, in the theoretical picture being considered, is what communication aims to achieve, at least in part. The concept of inhabitance was expressed by Merleau-Ponty (1962) in the following way: "Language certainly has inner content, but this is not self-subsistent and self-conscious thought. What then does language express, if it does not express thoughts? It presents or rather it *is* the subject's taking up of a position in the world of his meanings" (p. 193; emphasis in the original). The GP is a unit of this process of 'taking up a position in the world of meanings.' On this model, an analysis of conversation should bring out how alignments of inhabitance come about and, as this is taking place, how the overall conversational milieu is maintained by the participants.

The hyperphrase. A second theoretical idea—the 'hyperphrase'—is crucial for analyzing these alignments and maintenances, and how they are attained in complex multi-party meetings. A hyperphrase is a nexus of converging, interweaving processes that cannot be totally untangled. We approach the hyperphrase through a multi-modal structure comprising verbal and non-verbal (gaze, gesture) data.

To illustrate the concept, we shall examine one such phrase from a study carried out jointly by Francis Quek, Mary Harper and David McNeill (the 'Wombats study'). This hyperphrase implies a communicative pulse structured on the verbal, gestural, and gaze levels simultaneously. The hyperphrase began part way into the verbal text (# is an audible breath pause, / is a silent pause, * is a self-interruption):

we're gonna go over to # thirty-five 'cause / they're ah* / they're
from the neigh borhood they know what's going on #".

The critical aspect indicating a hyperphrase is that gaze turned to the listener in the middle of a linguistic clause and remained there over the rest of the selection (Table 1). This stretch of speech was also accompanied by multiple occurrences of a single gesture type whereby the right hand with its fingers spread moved up and down over the deictic zero point of the spatialized content of speech.

Table 1. Illustrating a 'Hyperphrase.'

F0 1	*F0 2* →		*F0 3*	*F0 4*	*F0 5* →				*F0 6*
we're gonna go over to	thirty-five	'cause	they're ah	they're	from the nei	gh	borhood	they know what's	going on
Gaze 1	*Gaze 2* →								
Gesture 1 →	*Gesture 2 repeated* →								
	hyperphrase →								

The speaker and listener in this experiment had before them a model village with actual objects on a surface; as the speaker described what was to be done, her right hand constantly shifted locations, hovering over different spots on the board—in the example, it was over two houses with readable numbers 35 and 36. Considering the two non-verbal features, gaze and gesture, together with the lexical content of the speech, this stretch of speech is a *single production pulse* organized thematically around the idea unit, 'the people from the neighborhood in thirty-five.' This would plausibly be the unpacking of a growth point. Such a hyperphrase brings together several linguistic clauses. It spans a self-interruption and repair, and spans 5 F_0 groups. The F_0 groups subdivide the thematic cohesion of the hyperphrase, but the recurrence of similar gesture strokes compensates for this oversegmentation. For example, the F_0 break between "what's" and "going on" is spanned by a single gesture down stroke. It is unlikely that a topic shift occurred within this gesture. Thus, the hyperphrase is a production domain in which linguistic clauses, prosody and speech repair all play out, each on its own time-scale, and are held together as the hyperphrase nexus.

Thus, we have two major theoretical ideas with which to approach the topic of multiparty discourse—the growth point and the hyperphrase. The GP is the theoretical unit of the speaker's state of cognitive being. The hyperphrase is a package of multimodal information that presents a GP. Through hyperphrases, GPs can be shared. Multiple speakers can contribute to the same hyperphrases and growth points. Speaker 2 synchronizes growth points with Speaker 1 by utilizing various turn-taking 'signals' to achieve synchrony. This hypothesis assumes that conversationalists align GPs—Speaker 2 emits signals in a hyperphrase until he/she senses alignment, and then allows an exchange of the speaking turn. The signals can be seen as bringing one state of cognitive being into alignment with another, with the hyperphrase package managing the coordination. We do not suppose that all turn exchanges are so organized, but we see evidence, in multiparty discourse, that much of it is.

The repeated gesture in the example seems to serve as a metric to "regiment the flow" of the discourse in a cohesive/coherent text-segment, to use a concept introduced

by Michael Silverstein (1993). A "configuration of indexicals" across modalities in the hyperphrase shows mutual indexicality that is pointing towards one GP. The hyperphrase is an "indexical structure" of a "textual event" in Silverstein's parlance.

The VACE project

The aim of our research project under the VACE program is to understand, across a wide multimodal front, interpersonal interactions during meetings of c. 5~6 individuals, US Air Force officers taking part in military gaming exercises at the Air Force Institute of Technology (AFIT), at the Wright Patterson Air Force Base, in Dayton, OH. The participants represent various military specialties. The commanding officer for the gaming session is always in position E. The task of this particular meeting was to figure out how a captured 'alien missile head' (which in fact looked rather like a coffee thermos with fins) functioned. The session lasted approximately 42 minutes. The examples to be studied are extracted from the latter half of this period. Figure 1 shows the meeting room and camera configuration.

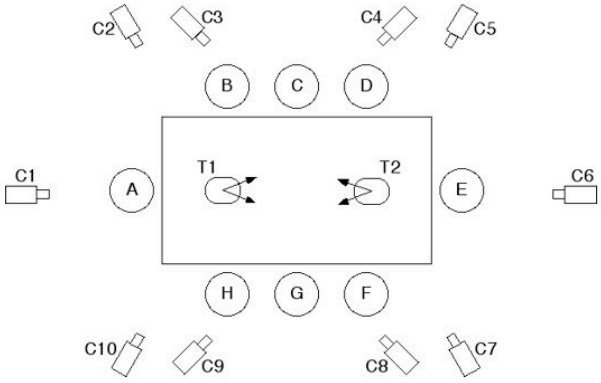

Fig. 1. Layout of the testing room. The participants were in positions C, D, E, F and G (positions A, B and H were vacant). Illustrations in later figures are from Camera 1's vantage point

We shall give some general statistics for gesture (pointing) and gaze during the entire meeting, including notes on some coding difficulties in the case of gaze, and then analyze two focus segments, concentrating on how the dominant participant (E) maintains his position, despite multiple shifts of speaker. We will also analyze the unique way the sole female participant seizes a speaking turn (participant C, who although of the same military rank as the others shows traits of marginalization in the group).

Pointing. The dominant participant, E, is the chief source of pointing but is the least frequent target of pointing by others. C and D are the least likely to point at anyone but are the most likely to be pointed at by others (D is notably passive in the group). So this pattern—rarely the source of pointing, often the target—may signal marginality, actual or felt, in a group setting. Table 2 summarizes the pointing patterns.

Table 2. Pointing Patterns in the Meeting

	Source C	Source D	Source E	Source F	Source G	Total
Target C	3	2	17	8	10	40
Target D	1	4	21	11	3	40
Target E	4	0	5	2	0	11
Target F	3	2	13	0	2	20
Target G	4	4	8	7	0	23
Target others	*12*	*10*	*59*	*28*	*15*	
Target All	0	0	5	0	0	5
Target Some	1	2	10	2	0	15
Target Obj	3	6	20	12	24	65
Target Abstract	5	11	8	1	1	26
Total	24	31	107	43	40	245

(Note: 'target others' excludes self-pointing)

Fig. 2.1. E (head of table) points with right hand and gaze at C (left front). Hand pointing is difficult to see in a still shot: it was accomplished dynamically, by rotating the hand toward C. Participants are festooned with motion tracking (VICON) jewelry.
Fig. 2.2. F (right rear) points at G with origo shift toward E.

Figures 2.1 and 2.2 illustrate two pointing events, the first showing E with his right hand rising from rest on the table to point minimally at C (and thereby authorizing—weakly—her participation as speaker; notice also how E's gaze reinforces the deictic action); the second is F pointing at G but in a curious way that shifts the origo or perspective base of the gesture to a locus in front of his own location, a maneuver that may unconsciously reflect the 'gravitational pull' of E on his right.

Gaze. Table 3 summarizes the distribution of gazes during the entire meeting. Again, as in pointing, E's dominant status is registered by an asymmetry, but now with reverse polarity: he is the most frequent gaze target but the least frequent gaze source. C, the sole female present, is the least frequent gaze target but the most frequent gaze source—a pattern also seen in a NIST interaction analyzed previously (unpublished data; NIST is the National Institute for Science and Technology) again involving a female participant, although not the sole female in that case, but again seemingly the marginal participant in the group.

Table 3. Frequency of gaze during the meeting.

	C Source	D Source	E Source	F Source	G Source	Total
C Target	X	38	45	59	67	209
D Target	70	X	83	112	94	359
E Target	212	136	X	144	149	641
F Target	150	107	98	X	116	471
G Target	75	52	63	68	X	258
Total	507	333	289	383	426	1938

However, gaze *duration* by E is longer—duration and shift of gaze may perform distinct functions in this tradeoff. Table 4 compares the frequency and duration of gazes by E to G vs. those of G to E. Indeed, E looks with longer durations at G than G does at E, but this asymmetry does not hold for gazes at neutral space, the object, or papers—at these targets G's gazes are actually longer (G is one of the most active gaze shifters). E's fewer, longer gazes at people but not at objects can be explained if he uses gaze to *manage* the situation—showing attentiveness (hence longer) but feeling no pressure to seek permission to speak (therefore fewer). Such fewer, longer gazes at people (but not at objects) are recognizably properties of a dominant speaker.

Table 4. Comparison of E's gaze duration (fewest shifts) to G's (more shifts)

	E's gaze Number	(fewest shifts) Av. Duration secs	G's gaze Number	(more shifts) Av. Duration secs
At C	45	5.1	67	1.1
At D	82	4.0	93	2.6
At E	-	-	149	1.9
At F	98	3.9	116	1.6
At G	63	3.1	-	-
Neutral space	150	1.0	292	1.5
At object	58	1.7	42	2.8
At papers	33	3.2	18	8.2
Others	4	2.4	8	1.9
Average	67	3.0	98	2.7

To summarize dominance and marginality: Both pointing and gaze correlate with the social dimension of dominance, but in opposite directions:

> In *pointing*, the gesture has an active function—selecting a target; it is thus correlated positively with dominance and negatively with marginality. Marginal members may frequently be pointing targets as part of recruiting efforts.
>
> In *gaze*, the action has a passive or perceptual function—locating the source of information or influence; it is accordingly correlated negatively with dominance and positively with marginality, especially when brief.
>
> But in E's case, *gaze* is also active, not passive, and this is reflected in longer durations only at people, combined with fewer shifts of gaze overall; duration thus correlates with dominance positively.

Coding issues. Inferring gaze from video poses difficulties of coding, and it is well to say something about this. The following comments are based on notes by the coder (co-author Welji): F and G wear glasses, making it difficult to see where their eyes and even sometimes whether the eyes are open. Often it is necessary to look for a slight

movement of the eye or eyelid, which can be difficult to spot. Also, neutral space can coincide with the location of the object on the table and sometimes it is difficult to distinguish if the object is the target of gaze. A third difficulty is that at some orientations it is hard to get a good view of the eyes. Finally, when coding in slow motion a blink and a short glance away may be indistinguishable. Is perhaps reassuring, given the uncertainties, that no more than 8% of the gaze judgments for the be-glassed participants and less than 3% for the best participant were deemed tentative.

Focus segments

Two segments were selected for detailed analysis. Both came from the second half of the 42-minute session.

Focus 1. The first segment highlights turn taking exchanges in which hyperphrases carry multiple functions. The speech is as follows:

1. E: "okay. u-"
2. G: "So it's going to make it a little tough."
3. F: "It was my understanding that the- the whole head pivoted to provide the aerodynamic uh moment. But uh I could be wrong on. That uh …"
4. G: "that would be a different design from-"
5. F: "From what-"
6. G: "from- from the way we do it."
7. F: "Okay."
8. E: "Okay so if we-"
9. G: "But we can look into that."
10. E: "If we're making that assumption ((unintel.)) as a high fidelity test"
11. F: "Yeah."

Turn taking at momentary overlap of GPs. An obvious case of a GP starting with one speaker and passing to the next speaker appears at 5, where F says "from what" and G, at 6, takes over with "from- from the way we do it". The hyperphrase package of the joint inhabitance is seen in the deployment of gaze and gesture:

> F begins with a glance at E, then gestures interactively toward G, followed immediately by gaze at G and an iconic gesture depicting the alien coffee mug (see Figure 3).

The hyperphrase here is a multimodal unit within which dimensions of gesture and gaze exchange places in creating a GP comprised of imagery that depicts the object and a linguistic component asserting that this procedure is 'the way we do it' vis-à-vis the object. We also see a hyperphrase being constructed by F that includes social information: E's standing as dominant speaker, in the quick glance at him at the start; G's status as current speaker, in the interactive gesture to him; and the ongoing role of the 'thermos' as the discourse theme.

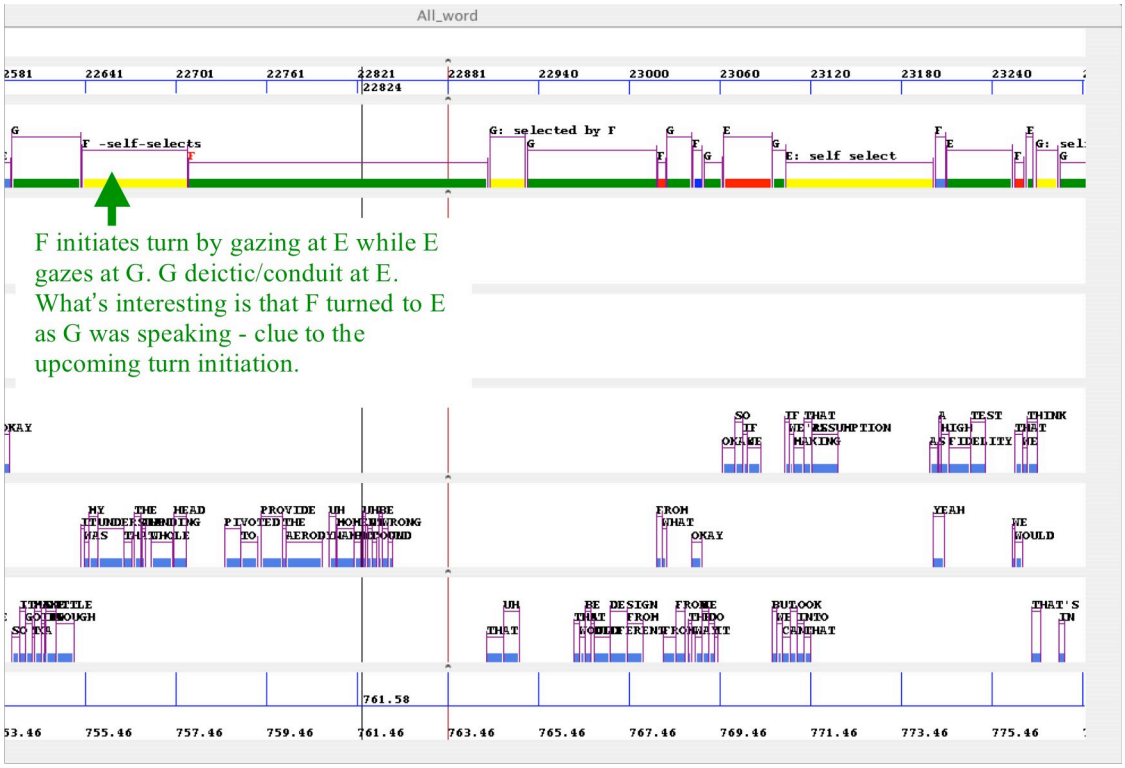

Figure 3. MacVissta screenshot of turn taking in Focus 1. Notes added on how turn taking correlated with gaze and gesture (see Chen et al 2006 for details on the MacVissta interface).

Figure 4 displays how gesture was recruited at the onset of the new turn—a further component of the hyperphrase at this moment.

F-formation analysis. The concept of an F-formation was introduced by Adam Kendon, who said, "An F-formation arises when two or more people cooperate together to maintain a space between them to which they all have direct and exclusive [equal] access." (Kendon 1990, p. 209). As an operationalization, an F-formation is discovered by tracking gaze direction in a social group; such direction can identify a shared focus of attention. The concept, however, is not just about gaze and shared attention toward a spatial locus. Crucially, space has an associated meaning, reveals a common ground, and helps us, the analysts, find the units of thematic content in the conversation. Figure 5 shows the F-formations so operationalized in Focus 1. Tracking the appearance of the same colors (shades of gray here) across participants identifies each F-formation, defined as a shared focus of attention on that individual. In the focus segment, an F-formation defined by shared gaze at F (lightest gray) is replaced by one defined by gaze at G (4th darkest gray). Interestingly, there is a brief transition or disintegration, with gaze either at E or at non-person objects—acknowledgement of E's status as dominant. But the main inference from the F-formation analysis is that speaker F was recognized as the next speaker *before* he began to speak, and this recognition was timed exactly with *his* brief gaze at Ec—a further signal of E's dominance. This gaze created a short F-formation

with G, since both then looked at E. This in effect signaled the turn exchange, and is another component of the hyperphrase at this moment, ushering in a joint growth point.

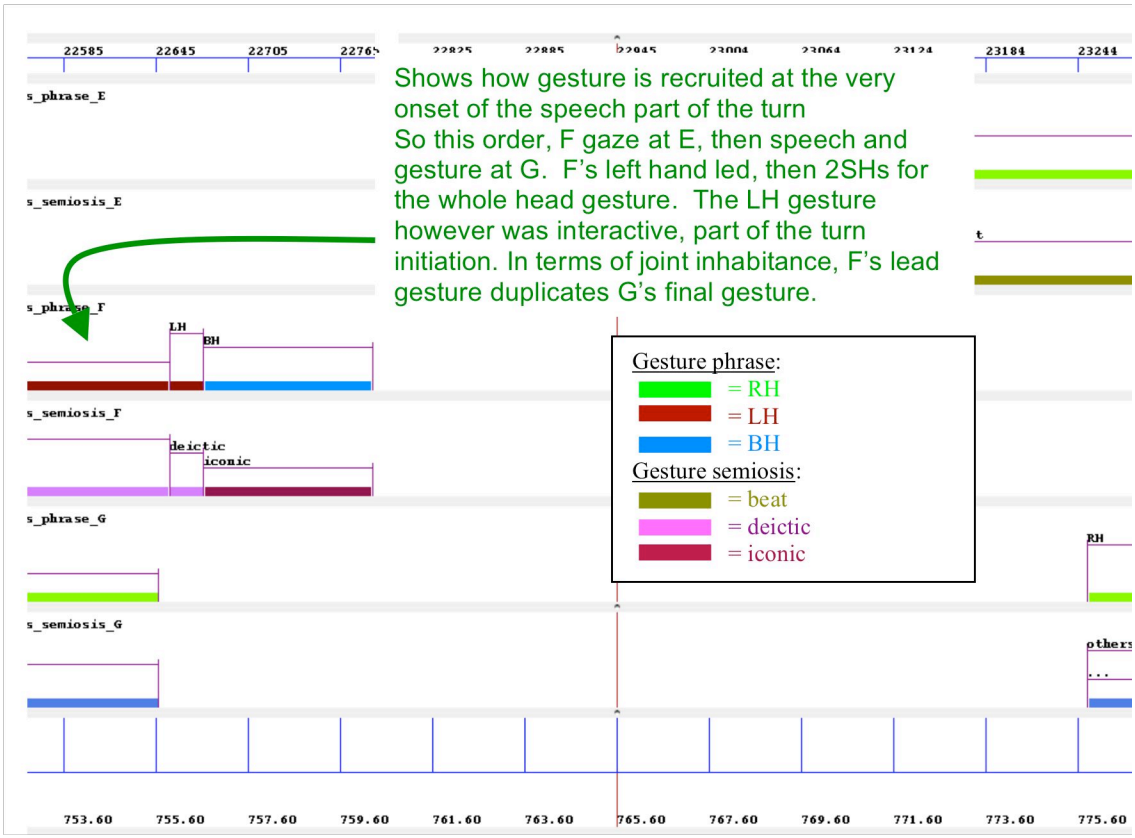

Fig. 4. MacVissta screenshot of gesture in Focus 1. Notes added on how gesture correlated with gaze and turn taking.

Back to momentary sharing of GPs. Thus, at the turn exchange, there was a synchronizing of inhabitance (the nonrepresentational mode of significancation defined by Merleau-Ponty) by F (the next speaker) with G (the current speaker) via their joint F-formation with E the target. F's hyperphrase (a bundle of multimodal features) encompassed all these features. F's GP included the idea of his collaboration with G, their joint clearance with E, and with this he could lock-step their current cognitive states. F's first GP was in fact a continuation of G's. The details appear in how gaze and gesture deployed around the table:

> Dominant E continues to gaze at designated speaker G when G gestures at object and others apparently look at the object.
>
> G gazes at the dominant participant, and makes deictic/conduit gestures in his direction (cf. McNeill 1992 for these terms). G then shifts his gaze to the object, and then quickly shifts back to E. Nonspeaker D doesn't shift to E when G shifts but keeps gaze at G—suggesting that what we see is the speaker affirming the dominant status of E, but the overhearers are free to respond to the speaker's new turn.

Also, when F takes turn from G he waits until G finishes his ongoing sentence, but first turns to look at E in the middle of the sentence, and then starts his turn while still looking at E (only after this shifting to G).

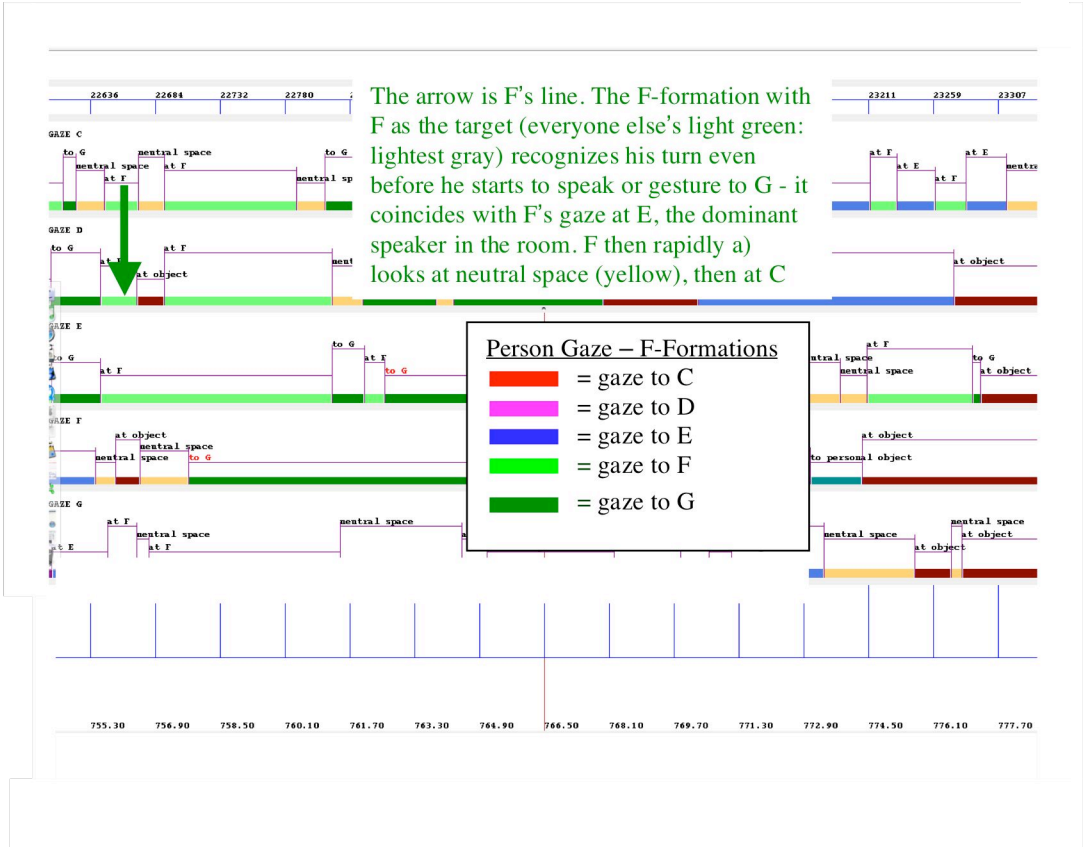

Fig. 5. MacVissta screenshot of F-formations in Focus 1. Notes added on how F-formations correlated with gesture, gaze and turn taking.

The next example however displays a very different form of turn exchange, one based on *non*-joint inhabitance of a hyperphrase.

Focus 2. For reasons not entirely clear but possibly connected to the fact that, although of equal military rank, C was the sole female present, this speaker does not create a series of moves designed to synchronize idea units with any current speaker. She appears instead to wait until there is no current state of joint inhabitance, and then embarks on a turn. In other words, C exploits the phenomena that we have seen but in reverse: she waits until a break in hyperphrasing; when it appears she plunges in. Focus 2 begins as F signaled the end of his turn and E's gaze briefly left the interaction space: C then quickly moved to speak. The speech is the following, but to understand the action requires a multimodal picture:

F: "to get it right the first time. So I appreciate that."

F relinquishes turn—intonation declines.

E gazes straight down table (no target?), setting stage for next step.

C intervenes with ferret-like quickness:

C: "I'm thinking graduation exercise kind of thing. You know we might actually blow something up. Obviously we don't want to".

E (not F, the previous turn-holder) acknowledges C's turn with gesture and gaze, but in a manner that suggests surprise—further confirming that C's strategy was to wait for a general lapse of inhabitance before starting to speak.

Figure 6.1 shows the moment C spots her chance to speak (the first line above). Figure 6.2 depicts 9 frames (0.3 s) later. Note how all the participants, in unison, are shifting their gaze to C and forming in this way a multiparty F-formation and hyperphrase with C the focal point.

Fig. 6.1. C leaps in. Gaze around the table is generally unfocused.
Fig. 6.2. 9 frames (0.3 s) later, gaze generally shifts to C and E points at C.

One has to ponder the effects of a strategy like C's that avoids shared hyperphrasing and transitional GPs. C's experience of the interaction dynamics is seemingly quite different from the others and theirs equally from hers. Whether this is due to 'marginality' (as evident in pointing and gaze, Tables 1 and 2) or is a personal trait, is unclear. An all-female meeting would be of great interest, but we have not managed to assemble one to date.

Comparison of Focus 1 and Focus 2

In contrast to Focus 1, where we saw an intricate buildup of a hyperphrase out of gaze and gesture, in Focus 2 C gazes at E (even though she is following G), and E provides authorizing back channels in the form of gaze and pointing, and this is the total exchange; there is no real hyperphrase or possibility of a shared transitional GP.

Taking the two focus segments together, it seems clear that speaker status can be allotted, negotiated, or seized in very short time sequences, but dominant speaker status is ascribed and changes slowly if at all.

Coreference, F-formations, and gaze

The way in which discourse coheres—how segments beyond individual utterances take form—can be observed in various ways, but we have found tracking coreferential chains in speech to be highly useful. A 'reference' is an object or other meaning entity nominated in speech; a coreferential chain is a set (not necessarily consecutive) of linguistic nominations of the same referent. As a whole, the chain comprises a 'topic' in the conversation. A coreferential chain links extended text stretches and by its nature is interpretable on the level of meaning and can be the basis of hyperphrases. An important insight is that coreferential chains also can span different speakers, and so can tie together multiparty hyperphrases and shared growth points in dialogues.

Coreferential chains thread across different levels in the structure of discourse. A given chain might track over each of the following:

> **Object level**: cohesion through references to object world; e.g., "a confirming design".
>
> **Meta level**: cohesion through references to the discourse itself; e.g., "I propose assuming a US design".
>
> **Para level**: cohesion through references that include individual participants; e.g., "I agree with the assumption".

In Figure 7, a hyperphrase builds up between participants over each the above levels. In so doing it unites references to the alien object by tying them to the theme of how it is designed and what should initially be assumed about this design, each contribution from a different speaker and on a different level.

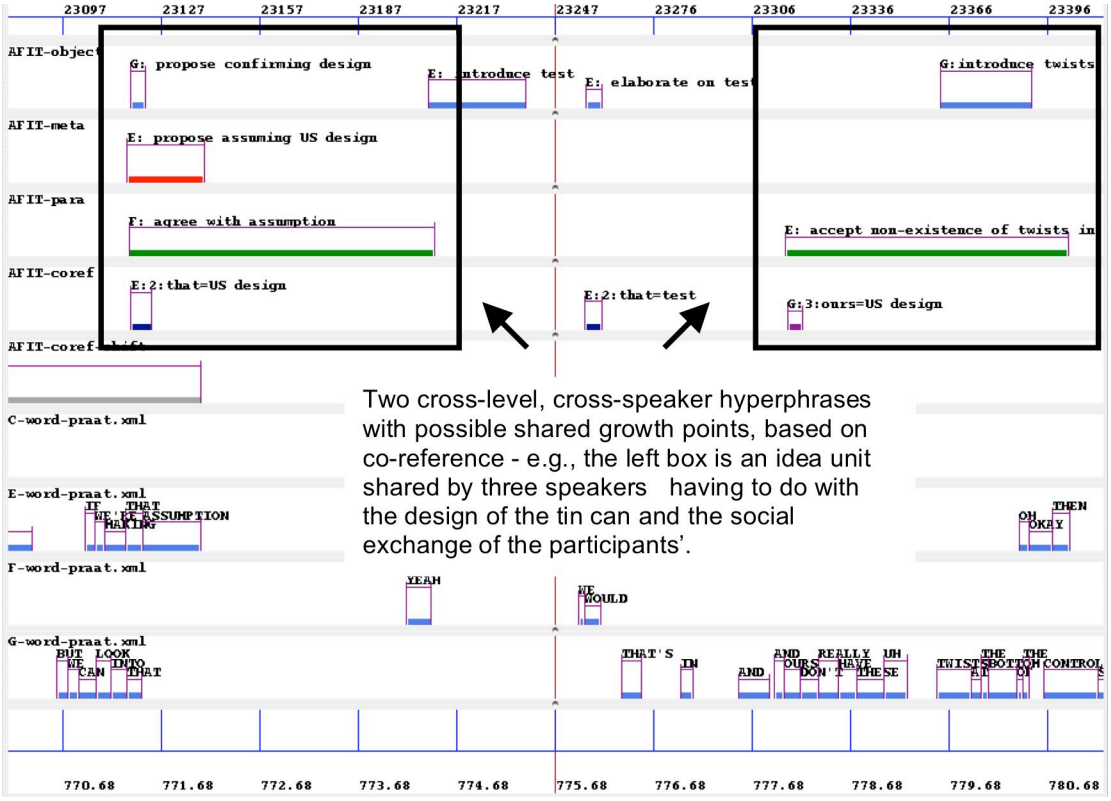

Fig. 7. MacVissta screenshot of coreference threads across multiple speakers creating F-formations

Coreferences also provide an overall profile of thematic content within a conversation. Figure 8 shows the cumulative distribution of coreferences over the total 42 minutes of the AFIT session. A small number of references account for the vast bulk of cohesion in this discourse. The curve can be read from left to right as listing the dominant topics and then less dominant topics—'FME people' (those who work on foreign material exploitation), operators of Air Force systems, and so forth, with the bulk of references on the long tail of single mentions.

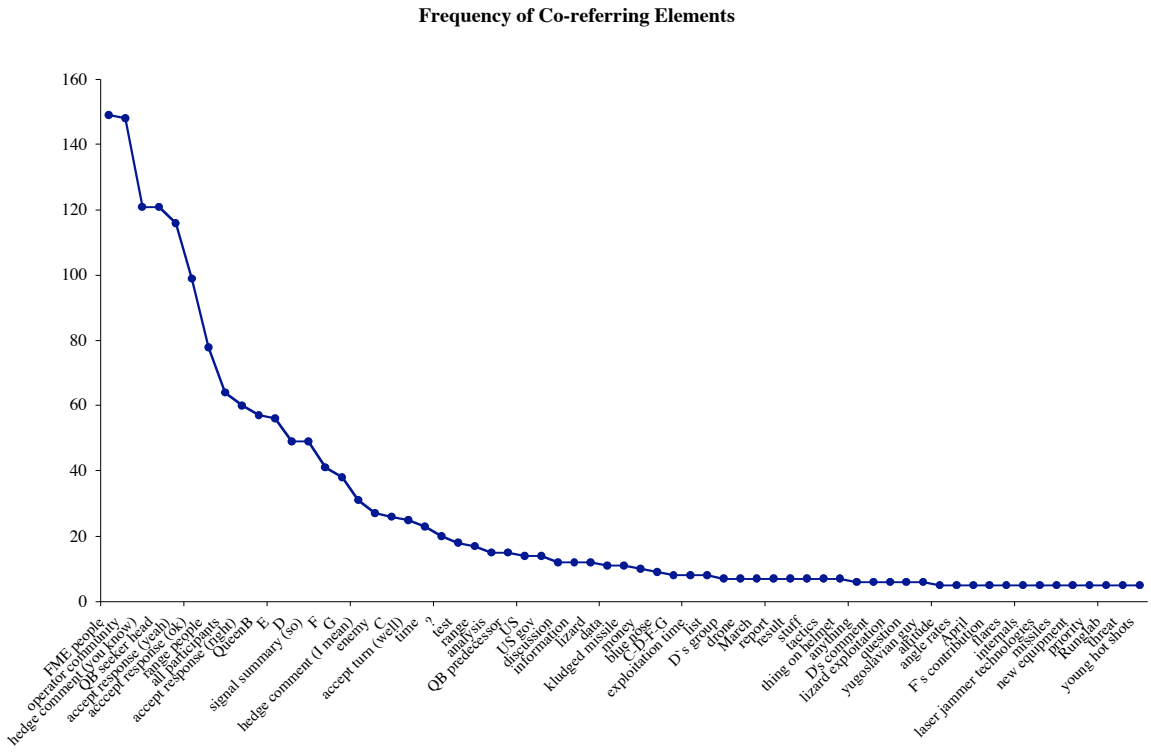

Fig. 8. Distribution of coreferences in the 42 minute session.

Elaborations on the F-formation. In the discourse situations we observe, we see two types of F-formation:

1) *social*, in which the elements are other individuals, and

2) *instrumental*, in which two or more people gaze at a common event or object in space.

The elaborations identify different kinds of social interactive configurations that can be seen in conversations that involve both participants and physical displays of objects (projection screens, the alien object of the AFIT session, etc.). Social F-formations are accompanied by significant shifts of the discourse levels of coreferential chains (object, meta, and para in various permutations); instrumental F-formations tend to stay on the same level (usually but not necessarily the object level). Table 5 shows the difference between social and instrumental F-formations in earlier data (a 4-party roundtable interaction recorded at NIST).

Table 5. Gaze and Level Shift

	Shift	Not Shift	N
Instrumental Gaze	44%	66%	32
Social Gaze	67%	33%	15

As hyperphrases, social F-formations thus open up a variety of trading relations with which to engender growth points during interactions. This richer variety is of course significant in itself. It makes sense in terms of the stimulus value of another person in a social context. The discovery is that social gaze has an immediate effect on the cohesive structure of discourse with coreference shifts strapped together into hyperphrases by gaze.

Coalitions

To illustrate coalition formation, we draw on another AFIT session in which one military officer and 4 civilian Institute instructors wrestled over (fictional) students for scholarship assistance, a process rife with coalition formation, as anyone who has served on academic admissions committees knows. We focus on manifestations of coalition formation, and how to interpret these in the light of mind-merging.

Recognizing coalitions

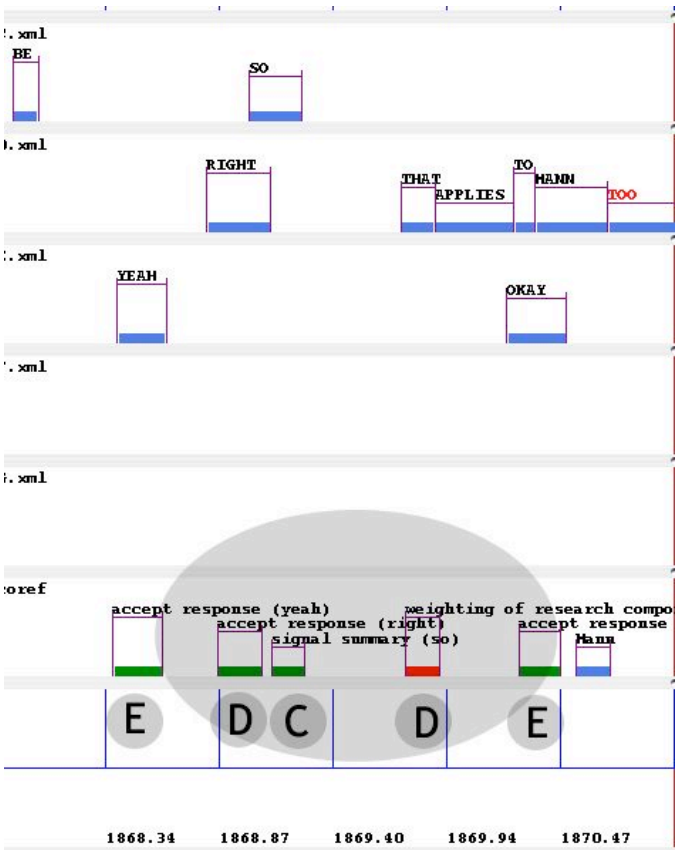

Fig. 9. MacVissta screenshot of a coreferential thread tying together a coalition.

Using coreferential chains. As mentioned earlier, coreferential chains shift levels (e.g., from meta to para), and do so particularly when the focus includes other participants. Such shifts mark the negotiation of reference, and often signal the formation of a coalition. Coalitions seem to show up as *clusters of paranarrative references surrounding one or more metanarrative references*. Such a sandwich-like structure suggests that coalitions between participants end with statements related to the individual participant's roles, but also orient to focus on one or more statements whose significance relates to the discourse itself, devoid of personal identifications. In terms of 'participation frameworks' (Cassell & McNeill 1991), a coalition is a package of two such frameworks: one is the activity of the task on which the coalition focuses, and the other is the social group joining it—the meta and para levels, respectively. By looking at who is speaking, we can detect membership in coalition. Figure 9 illustrates such a case.

In this instance, D is seeking acceptance (*you know*—paranarrative), C provides it (*yeah*—paranarrative), and both refer at the metanarrative level to D's previous speech (*that*). The meta comment (*weighting of research components)* is sandwiched between paras, which reflects several important qualities of these coalitions—the coalition is brief, and the coalition launches from a level with some personal identification (agreement, opinion) but then fixes on a level of impersonal focus on the structure of the discourse task itself.

Identifying conflicts. Conflicts in a group, as opposed to coalitions, can be identified as *changes across coalitions.* In the following, participant D disagrees with the system used to rank candidates. One clue to this disagreement is that he drops out of the ongoing coalition in which he had been participating (see Fig. 10):

Coalition 1: C, D, E

Coalition 2: C, D

Coalition 3: C, D, E

Coalition 4: C, E

Thus, by the end, a new C, E coalition has formed. In terms of hyperphrases, the structure changes with the absence of D, so that any further sharing of growth points would be defined in fields to which he would not contribute. Therefore, temporarily, he would also not influence the discourse.

Summary so far. Coalitions are marked by para-level comments bracketing at least one meta-level comment. This makes sense in that participants in a new coalition first indicate their allegiance the theme (para-level) and then indicate the significance of the theme to the overall discourse (meta-level). It is possible that such a pattern is a signature of these kinds of temporary coalitions that form around specific discourse themes.

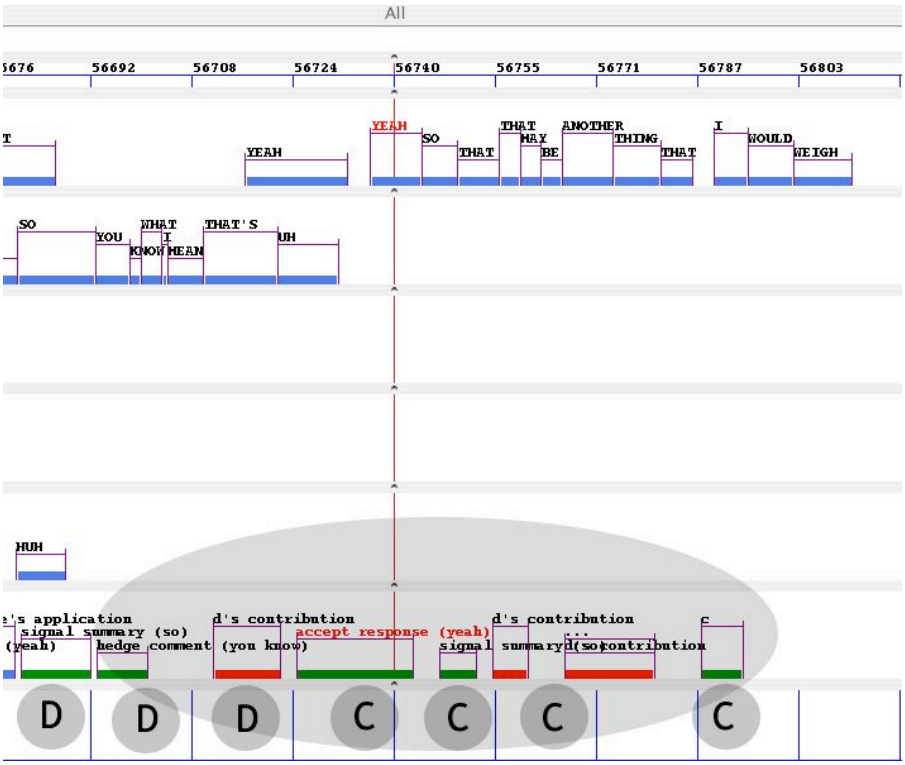

Fig. 10. MacVissta screenshot of a conflict and disappearance of participant from a coalition.

Sequence of gaze—another clue. Attempts to secure mutual gaze show HOW coalitions are built. During a segment identified (via coreference) as a coalition, C's gaze pattern shows how the coalition was created: this function of gaze includes deixis, C pointing (with gaze) at G, and so forth.

C tries to secure G, but his gaze is not reciprocated, then he secures D, and finally E (arrows show direction of gazes). See Figures 11 and 12.

1. C→G
2. C↔D
3. C↔E

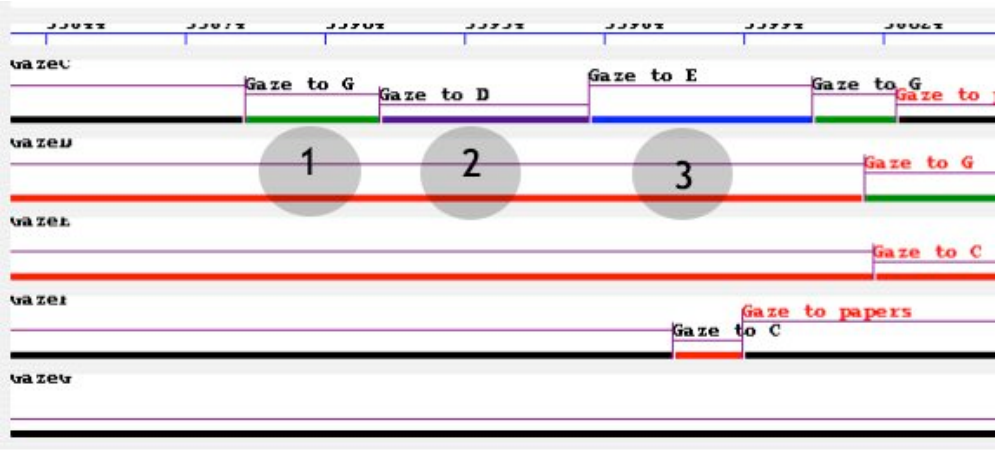

Fig. 11. MacVista screenshot of gaze recruiting members of a coalition. See Fig. 12.

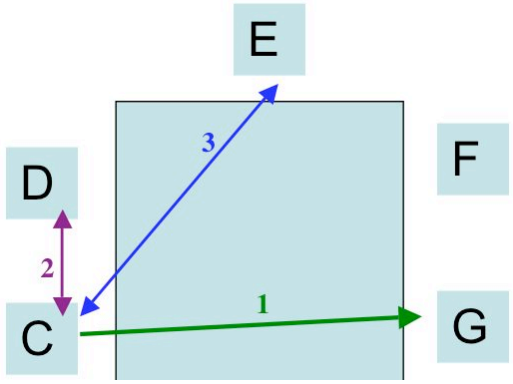

Fig. 12. Map showing sequence of gazes in Fig. 11.

A shared coreferential chain with C is secured when D and E return his gaze; G does not return gaze, and is not part of the coalition, and might now loom as a point of conflict. The important point is the alignment of the coalition with a shared field of oppositions, making possible merging growth points.

Garnering agreement. Within a third AFIT session devoted to brain-storming pollution damage to the Lincoln Memorial, behavioral indices comprised patterns according to the acceptance or rejection of proposed ideas; these patterns in turn may be functionally related to inducing acceptance or rejection or simply overlooking some proposed ideas. Both gaze and the quantity of speaking turns appear to be crucial:

- Mutual gaze triggers a secondary turn by a listener and a possible coalition for the proposal.
- Multiple turns take place when an idea is accepted (cf. the para- and meta-level coreferences sandwich when a collation forms).
- Multiple turns by a speaker supporting his own idea leads to acceptance. Thus, the cliché of the 'fast talker' has some reality.
- Rejected proposals, on the other hand, do not trigger mutual gaze between speakers and listeners, and thus do not initiate the above sequence.

It is noteworthy that the fate of a proposal—acceptance or rejection—is correlated with this pattern of information flow but not necessarily with the validity of the proposal itself.

In the session under study, the correct solution (in the sense that it was the one actually adopted by the Park Service) was rejected (it did not initiate mutual gaze).

The acceptance pattern can be explained in terms of the main theoretical concepts of this paper. Acceptance includes mind-merging, just as does turn exchange, and in fact relies on the latter as a vehicle. Thus both shared growth points and hyperphrases play a part in the acceptance of a proposal. Shared inhabitance is the key. There must be agreement, of course, but agreement without inhabitance is blind, so inhabitance is the key, as far as these observations go.

Conclusions

For communication studies, the implications of this research seem clear: a multimodal approach uncovers phenomena not seen otherwise. The concept of a hyperphrase, as a group of multimodal features in trading relationships, is particularly interesting from an instrumental viewpoint—how to pick up these interacting features? We focus on floor management: who is dominant, how are turns at speaking managed, what are the ways in which someone seizes a turn, and how does the alpha participant maintain control, etc.?, as well as the formation of coalitions, cleavages and coups, etc.

The psycholinguistic interest in these meetings lies in the apparent synchronizing of states of joint inhabitance that the turn taking process engages. However, we see a different mode of turn taking in Officer C's case, in which her procedure was not the synchronization of, but rather waiting for momentary lapses of joint inhabitance. While a single example cannot rule out individual style as the source of a pattern, it is the case that C's social isolation, as the sole female participant, is also a possible factor. While common ground (Clark 1996) seems indisputable in a general sense (the officers all knew, for example, they were in the US Air Force, were at AFIT, were taking part in a training exercise, had before them an alien object—in fact, assumed all the high frequency topics seen in Fig. 8), C jumped in precisely when she sensed a lapse in the common ground—F had just given up his turn, E was drifting into the ether, no one else was starting to speak, etc. It is therefore worth considering that common ground has two orientations: a general one, which, as Clark emphasized, is a precondition for all communication; and a local one, which is not a precondition but a *product* of the interaction, and is not a given in the conversation but is constantly unfolding. From this viewpoint, by interjecting, C created a new common ground. With the general-local common ground distinction, we can track the dynamics of the interaction.

From a psycholinguistic and social psychology viewpoint, the management of turn taking, floor control, and speaker dominance (even if not speaking) are crucial variables, and the prospect of instrumentally recording clues to these kinds of things has been the basis for valuable interdisciplinary work between psycholinguists and engineers (cf. Chen et al 2006). These descriptive features are the 'reality' of the meeting to which instrumental recording methods need to make reference. The automatic or semi-automatic monitoring of meetings needs to be related to the actual events taking place in the meeting at the social level, and our coding is designed to provide an analytic description of these events. The coding emphasizes the multimodal character of the meeting, attending equally to speech, nonverbal behavior and the use of space, and the

aim of the collaboration is to test which (if any) recoverable audio and video features provide clues to such events, thus warranting human inspection.

Coalition formation and maintenance follow the same meshing-of-mental-configurations principle. Like-minded partners take on demonstrably similar meanings in these materials. We can compare this interpretation to the mechanisms proposed by Pickering & Garrod (2004). We are in agreement with their overall theory that dialogue succeeds by 'aligning' participants on several levels, phonetic up to 'situation models'. Our difference lies in the 'mechanism'. While they look to an automatic process of priming, so that the subsequent dialogue is causally dependent on what has gone before, we are suggesting that the metapragmatic calibration is the hyperphrase, a 'reflective' process to use Krauss & Pardo's (2004) term, whereby the elements of dialogue that inhabit the same or similar growth points across participants are interchangeable and of equal potency. In this way, dialogue is not confined to overlaps with the immediate surround; on the contrary, the 'mechanism' in a hyperphrase is the *non*-overlaps that meet the condition of the co-inhabiting growth points, and is ultimately dependent on thought in context. We consider this hypothesis to have inherent plausibility in its freedom and openness, as opposed to the limits of priming repetition.

In overall conclusion we emphasize the importance of a truly multimodal perspective in uncovering interaction dynamics.

References

Cassell, Justine & McNeill, David. 2001. Gesture and the poetics of prose. *Poetics Today* 12: 375-404.

Chen, Lei, Rose, Travis, Parrill, Fey, Han, Xu, Tu, Jilin, Huang, Zhongqiang, Harper, Mary, Quek, Francis, McNeill, David, Tuttle, Ronald and Huang, Thomas. 2006. VACE Multimodal Meeting Corpus. In Steve Renals & Samy Bengio (eds.), *Machine Learning for Multimodal Interaction. Second International Workshop, MLMI 2005*, pp. 40-51. Berlin: Springer.

Clark, Herbert H. 1996.*Using Language*. Cambridge: Cambridge University Press.

Kendon, Adam 1990. *Conducting Interactions: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

Krauss, Robert M. and Pardo, Jennifer S. 2004. Is alignment always the result of automatic priming? *Behavioral and Brain Sciences* 27(02):203-204.

McNeill, David 1992. *Hand and Mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Merleau-Ponty, Maurice. 1962. *Phenomenology of Perception* (C. Smith, trans.). London: Routledge.

Ochs, Eleanor, Schegloff, Emanuel A, and Thompson, Sandra A. (Eds.) 1996. *Interaction and Grammar*. Cambridge: Cambridge University Press.

Pickering, Martin J. and Garrod, Simon 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(02):169-226.

Silverstein, Michael 1993. Metapragmatic discourse and metapragmatic function. In John A. Lucy (ed), *Reflexive Language: Reported speech and metapragmatics*, pp 33-58. Cambridge: Cambridge University Press.