Identifying Prosodic Indicators of Dialogue Structure: Some Methodological and Theoretical Considerations

Ilana MUSHIN
University of Melbourne / Latrobe University
Parkville
3051 Victoria, Australia,
i.mushin@linguistics.unimelb.edu.au

Lesley STIRLING
University of Melbourne
Parkville
3051 Victoria, Australia,
l.stirling@linguistics.unimelb.edu.au

Janet FLETCHER
University of Melbourne
Parkville
3051 Victoria, Australia,
j.fletcher@linguistics.unimelb.edu.au

Roger WALES
Latrobe University
Bundoora
3083 Victoria, Australia
dean.humanities@latrobe.edu.au

Abstract

This paper presents an empirical analysis of prosodic phenomena (intonation and timing) in 'common ground units' (Nakatani & Traum 1999). The analysis is used to address questions of the role of prosody in dialogue while taking into account the complexities of multispeaker discourse. We address some methodological concerns of how best to carry out a study of this kind as well as our theoretical questions about the formal identification of dialogue structures at levels higher than the micro-level of 'dialogue act', or 'move'.

Introduction

This paper reports some of the results from our research into the relationship between prosodic structure and discourse structure in dialogue. One of our particular interests is how to identify and analyse relevant prosodic parameters in multi-speaker discourse. We have been examining the kinds of dialogue structure frameworks that best account for patterns of prosodic phenomena; and conversely, the types of dialogue structure that exhibit prosodic regularities. Our research domain is human-human natural dialogue settings but our questions are equally relevant to researchers working on systems for more naturalistic human-computer interfaces, as well as those developing better automated systems for annotating large speech corpora.

The main methodological considerations associated with our current work are:

a) how natural dialogues can be reliably annotated to allow independent comparisons and correlations of prosodic and structural features,

b) the identification and classification of units of dialogue that reflect the 'joint action' feature of interactive discourse (ie. that both participants in a dialogue contribute to dialogue structure (Clark 1992, 1996), an aspect of dialogue that fundamentally differentiates it from monologic discourse).

These issues will be addressed in this paper using data from a corpus of naturally produced spoken dialogue taken from the Australian Map Task corpus (Millar et al 1994). Here we have focussed on the process of *grounding*, the assignment of utterances to 'common ground units' (CGUs - Nakatani & Traum 1999) and the internal structure of these units, as a means of illustrating some of the problems that both methodological issues raise. We show how

some of these problems might be overcome by focussing on sequences of initiating and responding (typically grounding) contributions within CGUs, as a site for prosodic analysis, rather than on the boundaries of the units as a whole (cf. Stirling et al 2000a). This approach thus preserves the notion that one can identify 'chunks' of dialogue in which particular types of information are acknowledged as being in the common ground of both participants, while remaining true to the dynamic nature of the grounding negotiation.

1. Background

1.1. Prosody and Discourse Structure

Most empirical work examining prosody in discourse has focussed on its function in monologue (eg. Swerts 1997, Nakatani et al 1995, Hirchberg & Nakatani 1996). These studies have found that a range of acoustic parameters associated with prosody, such as final lengthening and type of boundary tone, are good indicators of the boundaries between different discourse units at micro and macro levels of discourse structure.

More recently, there has been an interest in examining how prosody may be used in dialogue to signal discourse structure in that domain. Shriberg et al (1998) showed that various prosodic cues (duration, F0, pause length and speaking rate) were relevant for the automatic classification of dialogue 'acts'. Stirling et al (2000b) similarly showed strong correspondences between the boundaries of dialogue acts and prosodic phenomena such as pitch reset and intonational phrase boundaries (represented as ToBI 'break indices'). But dialogue acts are the 'parts' of dialogue most akin with structural elements of monologic discourse, since each 'act' can be analysed as independent utterances by a single speaker. Higher levels of dialogue structure necessarily involve some interactive 'chunk' of the discourse to which both participants in the dialogue contribute some speech.

So while it is clear that prosody serves to delimit dialogue acts, and to some extent distinguish between them (eg. Shriberg et al 1998, Koiso et al 1998, Stirling et al 2000b), the question remains whether prosody is also a reliable indicator of dialogue structure at higher levels (analogous with the higher levels of monologic discourse structure described in Swerts (1997), Nakatani et al. (1995), and others) and to what uses are prosodic phenomena put in the context of higher levels of dialogue structure.

1.2 Grounding and Common Ground Units

Grounding is the process by which information contributed by participants in interaction is taken to have entered the 'common ground', or mutual knowledge of the participants (Clark & Schaefer 1989, Clark 1996, Traum 1994). The process of grounding requires that one participant contributes something to the discourse (minimally, a dialogue act), and that the other participant make some indication that the contribution has been heard and accepted as a contribution (though not necessarily understood). This 'indication' may be a verbal acknowledgment (or some other kind of verbal response) or it may be some kind of non-verbal communicative act (like head nods, facial expression and other gestures).

Traum (1998) and Nakatani & Traum (1999) have recently proposed taking grounding as the basic principle behind the structuring of dialogue at levels higher than the dialogue act. Minimal units of acknowledged common ground have been considered as the building blocks of higher level dialogue structures based on intentional or informational content (eg. 'Common Ground Units', or 'CGUs' Nakatani & Traum (1999)). CGUs, which represent grounding at the 'illocutionary level' (Clark 1996), have been proposed as a meso-level dialogue structure - roughly the same level that dialogue games (Carletta et al, 1997) or adjacency pairs (eg.

Sinclair & Coulthard 1975) occupy in their dialogue structure frameworks.

The appeal of taking units based on grounding as the level of dialogue structure above the microlevel of 'act' (as argued in Nakatani & Traum 1999) lies in its prioritization of mutual understanding as a central component of dialogue, regardless of the type of initiation and response. In the 'CGU' framework, some responses themselves get grounded so that the result is a complex configuration of overlapping and embedded units of information entered into the common ground of the participants. This approach thus acknowledges importance of the contributions by both participants in the grounding process. It highlights the 'joint action' aspect of dialogic communication.

Evaluation of the coding of CGUs in dialogue by the Discourse Resource Initiative (Core et al 1999) showed a low degree of intercoder reliability, especially for those coding the HCRC Map Task corpus (Anderson et al 1991). Some of the inconsistencies across coders were attributed to "intonation and timing" (p. 61), as well as to difficulties in coding different types of acknowledgments. Some proposals for the classification of acknowledgments were made (Core et al 1999) and Stirling et al (2000a) have noted some parameters along which CGUs might be further classified.

The ways that this classification has been refined and utilised for the current paper are described in section 2 (methods), where we also describe our system of annotation for both prosodic and CGU properties of the dialogues, and explain our method for utilising this annotation for the current analysis. In section 3 we present some results of our investigation of prosody and grounding in the light of the methodological issues listed above. These results will be used in section 4 to address the following theoretical questions, as well as to set the agenda for future research into prosodic correlates of dialogue structure:

a) can prosody be used as a heuristic for identifying dialogue structure above the level of the 'dialogue act'?

b) does the process of 'grounding' have any formal basis, prosodic or otherwise, independent of the identification of dialogue acts?

2. Methods

2.1 Annotation Methods

Our corpus consists of 4 dialogues from the MAP TASK section of the Australian National Database of Spoken Language - ANDOSL (Millar et al, 1994). This corpus is closely modelled on the HCRC Map Task (Anderson et al 1991). Participants worked in pairs, each with a map in front of them that the other could not see. One participant (the 'instruction-giver' IG) had a route marked on their map and was required by the task to instruct the other (the 'instruction-follower' IF) in drawing the correct route onto their own map. The maps were similar, but differed in the presence, position and names of certain of the landmarks. Each pair of participants participated in two dialogues, swapping roles of instruction-giver and instruction-follower, and thus producing a first time and second time attempt at the task.

The four dialogues used here were from two pairs of participants: the two dialogues from a pair 'Known' to one another, differing in which participant took the role of instruction-giver, and the two dialogues from a pair 'Unknown' to one another, also differing in which participant took the role of instruction-giver. In each case, the pairs were mixed-gender. The dialogues were chosen randomly and the speakers belonged to the General Australian English dialectal grouping. The pre-recorded dialogues were copied from CD and digitised for analysis at 22 kHz using Entropic's ESPS/ Waves + speech analysis software running on a Sun workstation in the Phonetics Laboratory of the University of Melbourne. A complete orthographic transcription of the dialogues was carried out.

The prosodic features of the dialogues were labelled on separate tiers using the ToBI (Tone and Break Indices) prosodic transcription conventions for Australian English, detailed in Fletcher & Harrington (1996), closely modelled on the criteria developed for American English intonation (Beckman & Ayers Elam 1994/1997). The dialogues were prosodically annotated for break indices (degree of juncture) and phrase and boundary tones according to ToBI conventions using the Xwaves label function. The dialogues were also annotated on separate tiers in Xwaves for the timing features of pause location and duration (in ms), and overlap location and duration (in ms)[i]. These features were noted with respect to preceding and following talk.

The features of break index, phrase and boundary tones and pause and overlap phenomena were selected to address the 'intonation and timing' problems that were raised by the Discourse Resource Initiative (Core et al 1999), and because related work in conversation analysis has demonstrated the importance of intonation (especially final contours) to being able to account for the meaning of acknowledgments in interaction (eg. Müller 1996, Gardner 1998).

Break Indices (BI) were labelled as follows: 4 (full intonation phrase boundary); 3 (intermediate intonation phrase boundary); 3p (disfluent intermediate intonation phrase boundary). Boundary tones were labelled as follows: H-H% for a high rising tone (BI 4); L-L% for a falling or low tone (BI 4); L-H% for a low rising tone (BI 4); H-L% for a mid-level tone (BI 4); H- for a high intermediate phrase boundary (BI 3), L- for a low intermediate phrase boundary (BI 3).

The pause and overlap labelling allowed us to extract information about the timing of each speaker contribution with respect to the previous and following contributions to the talk. With respect to the preceding talk, contributions were analysed as having a pause before (pb), partially overlapping with the preceding contribution (olb), completely overlapping (ol), no pause or overlap with the preceding unit (or 'latch' - lb), or 'continued' (if the previous contribution was by the same speaker without an interceding pause (c)). With respect to the following talk, contributions were analysed as having a pause after (pa), partially overlapping with the following contribution (ola), completely overlapping (ol), no pause or overlap with following contribution (or latch - la), or 'continued' (if the next contribution was by the same speaker without an interceding pause (c)). Table 1 provides a summary of the labels used for prosodic and timing phenomena.

Table 1: Prosodic and Timing Labels

Break Index	Phrase and Boundary Tones	Timing (previous talk)	Timing (following talk)
4	H-H%	pb	pa
3	L-L%	olb	ola
3p	L-H%	ol	ol
	H-L%	lb	la
	H-	c	c
	L-		

Since the major goal of our research was to investigate associations between prosodic phenomena and discourse structure, we coded discourse categories independently from dividing the speech signal into prosodic units and the coding of prosodic phenomena (these parameters were coded by different researchers).

The dialogues were annotated for Common Ground Units, following Nakatani & Traum (1999).[ii] The coding was carried out by two researchers independently who then collaborated on a consensual version for each dialogue. The codes were entered on separate tiers in ESPS/ Waves +, separated according to which speaker gave which contribution. This meant that initiations and endpoints of CGUs were numbered and aligned with the speaker who began and ended each unit. Within each CGU we also coded the first 'response' by the other participant, which typically established either that the initial information was entered into the common ground, or that further negotiation was required in order to ground the information. In

this paper, we only considered CGUs that contained fully grounded information at their close (ie. we did not include abandoned or discontinuous CGUs - see Nakatani & Traum 1999). In Stirling et al (2000a) we reported on some of the prosodic characteristics of the final unit in CGUs (break indices and turn boundaries only). Here we report on the prosodic characteristics of the initiation move of the CGU and the prosodic profile of this first responding contribution.

2.2. Classification of CGUs

We go a step further than Nakatani & Traum (1999) and Stirling et al (2000a) in further classifying CGUs in terms of the internal characteristics of the grounding process - whether they consisted of a 'simple' exchange of initiating move and responding move, or whether their structure involved more contributions (by either speaker) before the CGU in question was considered to be complete (ie. the information was acknowledged as entered into the common ground). The former type we called 'Simple' CGUs, while the latter type we called 'Complex' CGUs. Complex CGUs were further classified into four types based on pragmatic and sequential criteria, as follows:

a) Overlapping CGUs, where the grounding element of one unit was itself grounded with some verbal acknowledgment in the next CGU, as in (1) below.
b) CGUs which consisted of further acknowledgment(s) by the other speaker subsequent to the first grounding element, as in (2) below.
c) CGUs which contained more than one acknowledgment by the same speaker (as in (3) below)
d) CGUs which negotiated information at different levels of communication lower than the 'illocutionary' level (eg. the level of 'presentation' or 'locution' (Clark 1996)), and were therefore considered part of a larger CGU which was grounded at the level of illocution, as in (4) below.

(1)
IF: am I to the left-hand side or the right-hand side
of the d~
of your Galah Open-cut Mine
IG: **looking at it you're on the left**
(finishes one CGU and starts another)
IF: okay

(2)
IG: oh you've got Whispering Pine have you
IF: **yes**
IG: **right**

(3)
IG: that is uh as a point of looking at the Consumer Trade Affair
[right]
IF: **[okay] yeh**

(4)
IG: so you're] swee[ping east]
IF: [so am I sweep]ing right around [am I going east]
IG: [you're sweeping] east
IF: yeh okay well don't~~ yeh okay allright I'm going east

All four types of complex CGUs represent some kind of 'expansion' of the canonical CGU exchange. In overlapping CGUs, the second 'response' functions to ground the first response. In multiple acknowledgment CGUs, the second response is not analysed as 'grounding', but is nevertheless some kind of further acknowledgment (either by the same speaker or by the other speaker). In the final case of complex CGUs, the first response is a signal that perhaps some part of the initial contribution had not entered into the common ground and that further collaboration was required.[iii]

These categories are not claimed to represent the only way that CGUs can be classified (cf. Core et al 1999). Nevertheless they provide us with a useful basis for addressing differences in the role of prosody in CGUs.

Altogether there were 419 CGUs in the corpus. Of these, 241 were 'simple' CGUs and 178 were 'complex'. Of the 178 complex CGUs, 47 (26%) were 'overlapping', 59 (33%) contained multiple acknowledgment by both participants, 52 (29%) contained multiple acknowledgements by the same speaker, and 20 (11%) contained more complex negotiations of understanding until acknowledgment of mutual understanding of the initial dialogue act was reached (the level at which CGUs were delineated).

We hypothesised that since, on our definition, complex CGUs differed from the simple type in terms of the structure of their response 'phase', rather than in terms of their initiation phase, there should be no difference between simple and complex CGUs with respect to the timing and prosodic profile of initiations. However, we predicted that there might be quantitatively recognisable properties of the first response in complex CGUs which motivated their expansion, and that these would be different from those found in simple CGUs. While we expected that these 'recognisable' properties would involve a combination of grammatical, pragmatic and prosodic properties, here we were only interested in the extent to which simple and complex CGUs could be differentiated on prosodic grounds.

In order to check these hypotheses, we pooled the four dialogues and compared initiation and first response contributions in simple and complex CGUs for each of the four parameters identified above (BI, boundary tone, timing (before) and timing (after)). Chi square tests were carried out on each of these comparisons to determine which sets of parameters displayed significant variation between simple CGUs and complex CGUs. The results are presented in the following section.

3. Results

Chi square tests on the combined totals of all CGUs showed highly significant differences ($p<0.001$) between initiations and responses for all four prosodic and timing parameters. This result is consistent with those reported in Stirling et al (2000b), where it was shown that initiation type dialogue acts could be prosodically differentiated from response type dialogue acts. Here we were more interested in the question of whether the complexity of the CGU was also reflected in initiation and response contributions. The results reported in the following subsections therefore refer to the percentage number of each prosodic and timing parameter as a proportion of the CGU type (percentages are in boldface) - the total number of simple CGUs or the total number of complex CGUs - for initiations and responses independently.

3.1 Correspondences with Break Indices

Table 2: Initiations (CGU type x BI)

	BI4	**BI3**	**BI3p**	**No BI**	**Total**
Simple	227	7	5	2	241
CGUs	**94.2**	**2.9**	**2.1**	**0.8**	**100**
Complex	165	4	9	0	178
CGUs	**92.7**	**2.3**	**5.0**	**0**	**100**
	392	11	14	2	419
Total	**93.6**	**2.6**	**3.3**	**0.5**	**100**

χ^2 (df=3, N=419) = 2.81, NS

Table 3: Responses (CGU type x BI)

	BI4	**BI3**	**BI3p**	**No BI**	**Total**
Simple	161	53	2	25	241
CGUs	**66.8**	**22.0**	**0.8**	**10.4**	**100**
Complex	145	18	2	13	178
CGUs	**81.5**	**10.1**	**1.1**	**7.3**	**100**
Total	306	71	4	38	419
	73.0	**16.9**	**1.0**	**9.1**	**100**

χ^2 (df=3, N=419) = 12.69, $p<0.01$

As expected, there was no significant difference between simple and complex CGUs with respect to their initiation contribution break indices. Examination of the first response of these units did show a significant difference between simple

and complex CGUs with respect to break indices. In particular, the first responses for simple CGUs had a high proportion of BI3 (corresponding with an intermediate phrase boundary) compared with complex CGUs and a relatively low proportion of BI4 (corresponding with a full intonation phrase). That is, the first response contribution of a complex CGU was more likely to end with a full intonational phrase boundary (BI4) than the first response contribution of a simple CGU.

3.2 Correspondences with Boundary Tones

Table 4: Initiations (CGU type x tones)

	H-H%	L-L%	L-H%	H-L%	H-	L-	No	Tot
Simple CGUs	82 **43.0**	71 **29.5**	72 **29.9**	4 **1.7**	4 **1.7**	3 **1.2**	5 **2.0**	241 **100**
Compl CGUs	53 **29.8**	75 **42.1**	33 **18.5**	2 **1.1**	2 **1.1**	4 **2.3**	9 **5.1**	178 **100**
Total	135 **32.2**	146 **34.9**	105 **25.1**	6 **1.4**	6 **1.4**	7 **1.7**	14 **3.3**	419 **100**

χ^2 (df=6, N=419) = 18.00, p<0.01

Table 5: Responses (CGU type x tones)

	H-H%	L-L%	L-H%	H-L%	H-	L-	No	Tot
Simple CGUs	51 **21.2**	42 **17.5**	55 **22.8**	13 **5.4**	21 **8.7**	28 **11**	31 **12**	241 **100**
ComplCGUs	40 **22.5**	53 **30.3**	39 **21.9**	13 **7.3**	7 **3.9**	12 **6.8**	13 **7.3**	178 **100**
Total	91 **21.7**	96 **22.9**	94 **22.4**	26 **6.2**	28 **6.7**	40 **9.6**	44 **10**	419 **100**

χ^2 (df=6, N=419) = 17.23, p<0.01

In contrast with the results for BI reported above, there were significant differences found between simple and complex CGUs for boundary tones in both initiation and response contributions. With respect to initiating contributions, a higher proportion of low falling (L-L%) boundary tones (42.1% vs. 29.5%) and a lower proportion of low rising (L-H%) boundary tones (18.5% vs. 29.9%) were found in the complex CGUs, compared with simple CGUs. Proportions of high rising (H-H%) boundary tones were not significantly different and there were proportionally few instances of other types of contours.

Like initiation contributions, the results for response contributions also show a higher proportion of low falling tones in complex CGUs than in simple CGUs (30.3% cf. 17.5%). The proportions of both high rising and low rising tones appears stable across the types of CGUs however.

The numbers of response contributions which had other types of boundary tones (including no boundary tone) were much higher than those found for initiating units. This was expected, since responses to initiation units were typically acknowledgments of grounding (like 'okay' or 'yeh'), which were then followed by other talk by the same speaker continuing an intonational phrase. The higher proportion of such 'non-final' contours (H-L%, L-, H- and none) for simple CGUs than complex CGUs reflects a set of instances in which the respondent answers a yes-no question, and then makes a further response *in the same intonational phrase* that is new information which was itself grounded (and was therefore coded as a new CGU). These non-final contours typically also corresponded with break indices of less than 4, accounting for the relatively high proportion of BI3s in responses in complex CGUs (noted in the previous section).

3.3. Correspondences with timing

3.3.1. Timing relative to preceding stretch of talk

Table 6: Initiations (CGU type x timing before)

	pb	olb	ol	lb	c	Tot.
Simple CGUs	98 **40.7**	38 **15.8**	1 **0.4**	44 **18.2**	60 **24.9**	241 **100**
Complex CGUs	62 **34.8**	34 **19.1**	3 **1.7**	27 **15.2**	52 **29.2**	178 **100**
Total	160 **38.2**	72 **17.2**	4 **1.0**	71 **16.9**	112 **26.7**	419 **100**

χ^2 (df=4, N=419) = 4.59, NS

Table 7: Responses (CGU type x timing before)

	pb	olb	ol	lb	c	Tot.
Simple	100	18	27	94	2	241
CGUs	**41.5**	**7.5**	**11.2**	**39.0**	**0.8**	**100**
Complex	67	33	18	60	0	178
CGUs	**37.7**	**18.5**	**10.1**	**33.7**	**0**	**100**
	167	51	45	154	2	419
Total	**39.8**	**12.2**	**10.7**	**36.8**	**0.5**	**100**

χ^2 (df=4, N=419) = 13.06, p<0.01

Like the results for break indices, the results for the timing of units with respect to immediately prior talk showed no significant differences between simple and complex CGUs for initiation contributions, but did show significant differences with respect to first responses. In particular, there was a relatively high proportion of responses in complex CGUs whose onset occurred while the other speaker was still talking, resulting in an overlap (18.5% vs. 7.5%). When the first response in a CGU is at least partially overlapping with the initiation contribution, it creates an environment in which both participants must work harder (ie. make more contributions) in order for the acknowledgment of common groundedness to be clear. At least some of these overlaps resulted in the same speaker repeating an acknowledgment of grounding with no overlap, as in example (3) above. However, they also resulted in complex CGUs of the other kinds.

3.3.2. Correspondences with timing relative to following stretch of talk

Table 8: Initiations (CGU type x timing after)

	pa	ola	ol	la	c	Tot.
Simple	111	31	2	90	7	241
CGUs	**46.1**	**12.9**	**0.8**	**37.3**	**2.9**	**100**
Complex	69	40	2	60	7	178
CGUs	**38.8**	**22.5**	**1.1**	**33.7**	**3.9**	**100**
	180	71	4	150	14	419
Total	**43.0**	**16.9**	**1.0**	**35.8**	**3.3**	**100**

χ^2 (df=4, N=419) = 7.64, NS

Table 9: Responses (CGU type x timing after)

	pa	ola	ol	la	c	Tot.
Simple	55	18	36	29	103	241
CGUs	**22.8**	**7.5**	**14.9**	**12**	**42.8**	**100**
Complex	40	7	21	40	70	178
CGUs	**22.5**	**3.9**	11.8	**22.5**	**39.3**	**100**
	95	25	57	69	173	419
Total	**22.7**	**5.9**	13.6	**16.5**	**41.3**	b

χ^2 (df=4, N=419) = 9.95, p<0.01

The results for the timing of the unit following the target contribution also showed no significant difference for initiations between simple and complex CGUs, while the responses did show a difference of distribution. However, the asymmetry in overlapping found in the preceding set of the results did not occur when one considered the contribution following the first response. In this case, the category which showed the greatest degree of difference between simple and complex CGUs was the category of 'latch' (22.5% vs. 12%), where the second speaker timed his/her next turn to follow from the target contribution without any pause or overlap. Responses in complex CGUs showed a higher proportion of latching with the following contribution. This high proportion of latching is perhaps an indication of a smooth transition from first response to further grounding contributions within the CGU.

4. Discussion

The results show that for three of the parameters investigated (break index, timing before and after the target contribution), there was little or no variation according to the type of CGU that the contribution initiated (simple or complex). This finding was consistent with our hypothesis that if there were any measurable differences between these CGU types, they would be found in the first response contribution of the CGU, and not in its initiation contribution.

Also consistent with our hypothesis was the result that all four prosodic and timing

parameters displayed differences between simple and complex CGUs in the first response contribution. Complex CGUs displayed more overlapped first responses, fewer incomplete intonation phrases (BI of less than 4) and a higher proportion of low falling boundary tones than the response contributions of simple CGUs.

Explanations for some of these results were provided in the previous section, but further analysis is required before we can really gain an accurate picture of the formal profile of complex CGUs (cf. simple CGUs). In particular, we require an analysis of the different types of complex CGUs to determine whether they display formal regularities. It may be that only one or two types of complex CGUs are contributing to the patterns observed here. Such an analysis is planned in future work. We also plan to conduct a similar study that takes into account the dialogue act type of initiation and first responses to determine more extensively the full range of factors which influence intonation patterns and the timing of contributions in dialogue.[iv]

Our results do confirm the necessity of developing strategies for the analysis of prosody in dialogue that take into account the effects of sequence. Not only did initiation contributions overall behave differently to response contributions with respect to prosody and timing (a result which we did not address in detail here), but the responses themselves could also be differentiated based on whether they 'finished' a CGU, or whether they were followed by other contributions before grounding was achieved.

Results such as these suggest that with respect to the formal identification (cf. pragmatic identification) of levels of dialogue structure higher than the dialogue act, structures defined in terms of sequential position, as such as adjacency pairs, might be more compatible with patterns we have observed in our data. The low level of intercoder reliability for CGUs, as reported in Core et al (1999) also suggests the lack of readily identifiable formal properties. Even taking 'intonation and timing' into account, as we have done here, does not clearly point to a formal profile of CGUs independent of sequences of dialogue acts.

Empirical investigation\ of the role of prosody in multi-speaker discourse is still in its infancy, and we are still developing tools of analysis that free us from looking only at those aspects of dialogue that best mirror the structure of monologic discourse. We have found that our method of both prosodic and discourse segment annotation in ESPS/Waves+ has provided us with the power to develop this investigation further. In future work we will be able to easily incorporate additional variables, such as dialogue acts, complex CGU type, and a range of other prosodic phenomena into our analysis, while keeping track of the sequences in which these joint actions occur.

References

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. & Weinert, R. 1991. "The HCRC map task corpus." Language and Speech Vol. 34, pp. 351-366.

Beckman, M.E. & Ayers Elam, G. 1994/1997. "Guide to ToBI Labelling – Version 3.0", electronic text and accompanying audio example files available at http://ling.ohiostate.edu/Phonetics/E_ToBI/etobi_homepage.html.

Carletta, J., Dahlbäck, N., Reithinger, N. & Walker, M. (Eds.) 1997. Standards for dialogue coding in natural language processing. Schlöss Dagstuhl, 1997. Seminar report 167, also available at http://www.dfki.de/dri/.

Clark, H.H. 1992. Arenas of Language Use. Cambridge: CUP

Clark, H.H. 1996. Using Language. Cambridge: CUP

Clark, H. & Schaefer, E. (1989). Contributing to discourse. Cognitive Science,13, 259-294.

Core, M., Ishizaki, M., Moore, J., Nakatani, C., Reithinger, N., Traum, D. & Tutiya, S. (1999). The report of the Third Workshop of the Discourse Resource Initiative, Chiba University and Kazusa Academia Hall.

Fletcher, J. & Harrington, J. (1996). Timing of intonational events in Australian English. In P. McCormack & A. Russell (Eds.), Proceedings of the Sixth Australian International Conference on Speech Science and Technology (pp. 611—615).

Gardner, R. 1998. Between listening and speaking: The vocalisation of understandings. Applied Linguistics, 19,

Hirschberg, J. & Nakatani, C. (1996). A prosodic analysis of discourse segments in direction-giving monologues. In Proceedings of the 34th Annual Meeting of the ACL, Santa Cruz (pp. 286—293).

Jurafsky, D., Schriberg, L. & Biasca, D. (1997). Switchboard SWBD-DAMSL Shallow-Discourse-Function-Annotation Coder's Manual, Draft 13. Technical Report TR 97-02, Institute for Cognitive Science, University of Colorado at Boulder. Also available from http://stripe.colorado.edu/~jurafsky/manual.august1.html.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. & Den, Y. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. Language and Speech, Special issue on Prosody and Conversation, Vol. 41 (3-4), pp. 295-322.

Millar, J., Vonwiller, J., Harrington, J. & Dermody, P. (1994). The Australian National Database of Spoken Language. In Proc. ICASSP-94 (pp. 197—100).

Müller, F. 1996. Affiliating and disaffiliating with continuers: Prosodic aspects of recipiency. In E. Couper-Kuhlen (ed). Prosody in Conversation: Interactional Studies. Cambridge: CUP.

Nakatani, C., Grosz, B., & Hirschberg, J. (1995). Discourse structure in spoken language: studies on speech corpora. In Proceedings of the AAAI-95 Spring Symposium on Empirical Methods in Discourse Interpretation and Generation.

Nakatani, C. & Traum, D. (1999). Coding discourse structure in dialogue (Version 1.0). University of Maryland Institute for Advanced Computer Studies Technical Report UMIACS-TR-99-03.

Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M. & Van Ess-Dykema, C. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? Language and Speech, Special issue on Prosody and Conversation, Vol. 41 (3-4), 443-492.

Sinclair, J.M. & Coulthard, R.M. (1975) Towards an analysis of discourse: the English used by teachers and pupils. Oxford: Oxford University Press

Stirling, L., Mushin, I, Fletcher, J. & Wales, R. (2000a) The nature of common ground units: an empirical analysis using map task dialogues. Gothenberg Papers in Computational Linguistics 00-5. 159-165

Stirling, L., Fletcher, J., Mushin, I. & Wales, R. (2000b.) Representational issues in annotation: using the Australian map task corpus to relate prosody and discourse structure. Speech Communication.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. Journal of the Acoustical Society of America 101(1), 514--521.

Traum, D. (1994). A computational theory of grounding in natural language conversation. PhD thesis, Department of Computer Science, University of Rochester. Also available as TR 545, Department of Computer Science, University of Rochester.

Traum, D. 1998. Notes on dialogue structure. Unpublished ms, available from http://www.cs.umd.edu/users/traum/DSD/.

---

[i] Overlap location and duration could be measured accurately because the dialogues were recorded on dual channels.

[ii] We also coded other features of dialogue structure such as dialogue acts. See Stirling et al (2000b) for more details of that annotation method.

[iii] The categories of complex CGUs were not all mutually exclusive. For example, a CGU could contain multiple acknowledgments by both speakers. Since these overlaps occurred infrequently, for the purposes of this paper we assigned each complex CGU to only one 'primary' classification.

[iv] Such an analysis might for example, help provide an explanation for why, contra to our expectations, the initiation phases of complex CGUs showed differences in boundary tone phenomena from the initiation phases of simple CGUs.