Expertise or Expert-ese? The Emergence of Task-Oriented Sub-Languages

Patrick G.T. Healey
ATR Media Integration & Communications Laboratories
Hikaridai, Seika-cho,
Kyoto,
Japan
ph@mic.atr.co.jp

Abstract

This paper reports an experiment which demonstrates the emergence of group-specific sublanguages or 'expert-ese' within groups engaged in a series of task-oriented dialogues. Extending the findings of Garrod and Doherty (1994), it is argued that neither simple appeal to task expertise nor the collaborative establishment of mutual beliefs can adequately account for these results. An alternative proposal, that identifes repair as the critical locus of semantic coordination is sketched.

Linguistic (Sub)Communities

One means of characterising sub-communities is by appeal to what is taken as common ground for its individual members. A wide range of factors are of potential relevance. Clark (1996) suggests, amongst others; folk psychological assumptions, lexical and syntactic conventions, and (sub)cultural norms, facts and procedures. Recognising what is held in common in a particular community is critical to determining what sorts of utterances are likely to be understood and has clear consequences for the conduct of communication amongst its members. The linguistic conventions themselves form a critical part of what characterises a community; dialects provide perhaps the canonical example of communal common ground.

Intuitively, interlocutors may be sensitive to a range of cues, e.g.; physical appearance, accent and the content of utterances in making an initial judgment about what linguistic (sub)community they may belong to (e.g.; Clark, 1996; Isaacs & Clark, 1987; Sacks, 1992). Once the extent of the initial common ground is estimated, interlocutors can tailor their utterances, using dialect or jargon terms accordingly.

The question addressed in this paper is what are the processes by which the linguistic conventions specific to a particular (sub)community become established? One, influential, answer to this question is provided by the collaborative model of dialogue, hereafter CM, developed by Clark and coworkers (e.g., Clark and Schaefer, 1989; Clark and Wilkes-Gibbs, 1986). The central commitment of the CM is that, in order for an utterance to accrue to common ground, the parties to a dialogue must establish the mutual-belief that it has been understood as intended. This is achieved through a cycle of presentation and acceptance where both parties can explicitly signal whether they accept an utterance as a contribution to the unfolding conversation. New elements can accrue to the personal common ground of the parties to the conversation through this collaborative process of presentation and acceptance, providing a means by which e.g., a novel name for an object may be added to the linguistic resources available to interlocutors.

Prima facie, this account faces difficulties in accounting for data reported by Garrod and Doherty (1994) who demonstrated the emergence of a group-specific reference scheme in the maze game, a cooperative dialogue task (further details below). The general characteristics of the task are that in order to complete each trial, participants, working in pairs (dyads) must produce and exchange a series of spatial descriptions. A range of alternative description types are possible. However, where individuals perform the task over a number of trials they tend to converge on a single description type. Moreover, this degree of convergence is not achieved through explicit negotiation of description schemes (Garrod & Anderson, 1987).

For current purposes, the important contrast from Garrod and Doherty (1994) is between two conditions; a community group and a control, non-community, group. In both these conditions each dyad, on each trial, was made up of participants who had not performed the task together before. In the community group, the dyads were composed of individuals drawn from the same small pool or sub-group of participants. That is, participants in the community condition would only carry out the task with other participants from within the same subgroup. This constraint was imposed in order to ensure that, over trials, a 'common interaction history' would develop in the community group. Although all dyads were composed of individuals who had not directly performed the task together before, on later trials, the number of indirect links between members of a dyad would increase. For example, on trial three a dyad might be composed of participants who had both previously been paired with a third participant during the preceding trials (one on trial one, the other on trial two). In contrast to this, the non-community group drew on a much larger pool of participants and dyad compostition was manipulated in order to ensure that no common interaction history developed over trials.

Garrod and Doherty (1994) found that, in contrast to the non-community, control group, the community group converge rapidly and strongly on a specific type of spatial description scheme. Importantly, the observed convergence in the community group occurs when each of the participants in

that group have, on average, encountered less than half the communitie's members. The emergence of a common description scheme in the community group is thus too rapid for the pairwise, collaborative, establishment of mutual belief. More specifically, the use of a particular description scheme by one pair in the community group also predicts it's use by other pairs in the group but this prediction apparently cannot be derived from a consideration of the mechanisms of the CM alone.

Another possibility, also raised by Wilkes-Gibbs and Clark (1992), is that communal common ground may develop on the basis of common, but independently obtained, expertise with a particular problem. Thus, if each member of a community independently arrives at a particular solution to a problem this may become part of the initial common ground for that community.

This reasoning suggests two broad explanations for the community group convergence reported by Garrod and Doherty (1994):

a) 'Expertise': convergence is the result of independently developed expertise at performing the task in a group (any group). i.e., there is an (more or less) ideal linguistic solution to the recurrent problem of communicating maze locations to a range of different individuals
b) 'Expertese': convergence is the result of a group-specific mechanism of linguistic coordination.

What distinguishes these possibilities is that in the case of a) convergence is understood to derive from recurrent, general features of the task whereas in b) it depends on the group-specific emergence of linguistic subcommunities. With respect to Garrod's and Doherty's (1994) findings, it might be argued that the convergence observed within their community group is the result of each member, independently arriving at the same solution to the problem of communicting locations in the Maze task (possiblity a) rather than the emergence of a group-dependent sub-language (possiblity b).

Methods

In order to provide a strong test of whether convergence on description schemes reflects the emergence of group-specific sublanguages, an experiment using a modified version of the maze task was carried out.

Design

The basic rationale for the study was to encourage, over the course of a number of trials at a collaborative task, the emergence of a number of 'community' groups similar to the single community group in Garrod and Doherty (1994). Thus for the first five trials, each participant performs the task with a new partner drawn from the same pool or subgroup of participants. This first phase provides the opportunity for each sub-group to develop into a community of the kind reported in Garrod and Doherty (1994). The second phase, in which the experimental manipulation is made, occurs in an additional, sixth trial. This time half the pairs are again composed of individuals from the same subgroup (the *Homogenous* condition) and half are composed of subjects drawn from different sub-groups (the *Mixed* condition). Assuming that task experience is adequately controlled across groups then an expertise based explanation predicts that there should be no difference between the Mixed and Homogenous pairs; each participant, in each condition has equivalent experience at the task, performing the same number of trials with a new partner on each occasion. Similarly, no difference is predicted between the two conditions solely on the basis of the need to establish mutual-beliefs; this is identical in both conditions. However, if there is a sub-group specific mechanism of linguistic coordination then transfer outside the original subgroup, as in the mixed condition, should interfere with performance. Thus, in contrast to Garrod and Doherty (1994) this study employed a number of 'community' or subgroups and compared performance of the maze task within an established sub-group with performance of the maze task between sub-groups.

In order to provide an adequate number of subgroups and to provide equal numbers of mixed and homogenous dyads in trail 6, the resulting design used multiples of 24 participants, randomly assigned into three sub-groups of eight. For the first five trials each pair was composed of individuals drawn from a single sub-group of eight. On the sixth trial half the pairs were composed of individuals drawn, as before, from within a single sub-group and the other half were composed of individuals drawn from different sub-groups. Composition of the Mixed dyads was manipulated to ensure that all combinations of sub-groups were equally represented. Materials were assigned to pairs according to a latin-square which ensured that no individual participant encountered the same materials twice and that the materials were also counterbalanced across subgroups and conditions.

Materials

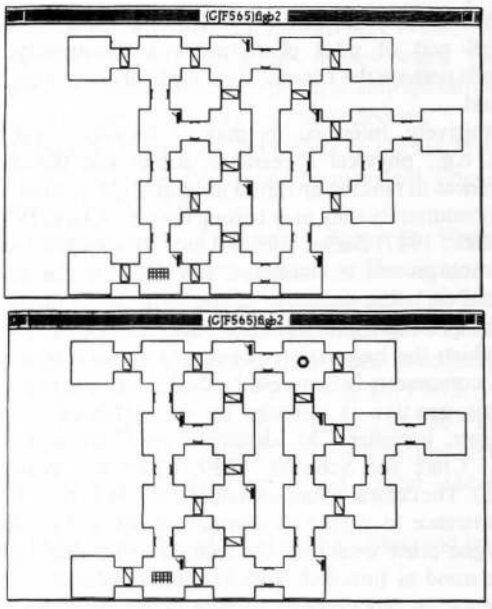

Figure 1: Example Pair of Maze Configurations

The orignal maze task was performed online (see Garrod & Anderson, 1987 for a full description). In this experiment a

modified, paper version, of the maze task was employed which was designed to preserve the requirement for generating spatial descriptions.

Using screen dumps from the original task, a set of 168 different maze diagrams were generated, each based on a 6-by-4 grid, with different configurations of boxes and features (see the example in Figure 1). These were made into pairs of 14 page booklets such that for any given page, the configuration of boxes on those pages would be identical but one booklet would also have a target location, marked by a thick circle, in one of the boxes (see Figure 1).

Procedure

The experiment was run in two parts with a total of 48 participants. These were recruited from students studying a secondary level psychology course at two colleges of further education in Edinburgh. They consisted of 9 males and 39 females ranging in age from 16 to 54 years with an average age of 33.

Participants were each randomly assigned a number that was used on all six trials for pairing them into dyads. On each trial, dyads were seated opposite each other at a desk with a low partition between them that obscured their view of each other's booklets whilst permitting eye contact. Members of a dyad each received one of a pair of matching booklets. The task was that for each item (corresponding pair of pages) the participant with the target location marked should communicate, without pointing or showing, its location to their partner so that she/he could mark where they thought it was on their own copy. Pairs had two minutes, start and finish indicated by the experimenter, in which to work through as many items as possible while preserving accuracy.

It was made clear that all the interactions would be recorded and transcribed (anonymously) and that they were free to withdraw if this presented any problem. No indication was given to participants, either in the instructions or the materials that they were divided into sub-groups.

Results

Two measures of task expertise were chosen; number of errors (misidentifications of target location) and number of items completed. An analysis of variance was performed on the proportion of items that received incorrect identifications of the target location by each dyad with trial number as a between subjects factor. This showed a reliable decrease in the number of errors across trials; Omnibus $F(4,138)=1.057$, $p=0.386$, Linear trend: $t(138)=1.930$, p(one-tailed)=0.027, falling from approximately 30% to 20%. Conversely, an analysis of variance on the average number of items completed by each dyad showed a reliable increase across trials; Omnibus $F(4,138)=17.71$, $p=0.000$, Linear trend: $t(138)=9.03$, p(one-tailed)=0.000, rising from 2.4 in trial one to 8.1 in trial six.

Two analyses of variance were also performed using the same dependent variables with the experimental manipulation of group composition (Homogenous vs. Mixed) as a between subjects factor. No reliable difference was found; Number of items attempted: $F(1,22)=0.04$, $p=0.843$; Proportion of Errors: $F(1,22)=0.006$, $p=0.936$.

To summarise, although the decline in errors and increase in number of items completed across trials provided reliable measures of increasing expertise with the task, they were not sensitive to the experimental manipulation of dyad composition. Transfer outside a sub-group did not appear to interfere with participants accumulated expertise with the task.

Transcripts

Two measures of linguistic coordination were chosen; type of description scheme used and incidence of repair of each description. A total of 144 dialogues were transcribed and coded for the occurrence of description scheme types.

Description Types: Garrod and Anderson (1987) identify four broad categories of description type that are commonly used in the maze task to identify target locations; Figural, Path, Line and Matrix. The basic characteristics of each category are summarised below together with an example description drawn from the corpus.

- Figural: relatively concrete and tied to specific elements of the particular configuration of boxes and features e.g., "right there's two double boxes on the right okay and they've both got kind of two triangles"
- Path: describes a route traversed from some origin through the specific configuration of boxes and passages between them e.g.; "take the bottom left hand corner, up one box, right one box, up one box".
- Line: abstracts across specific instances of the maze to impose an ordered set of rows or columns onto the underlying grid. Positions are specified in terms of some line, followed by specification of a position on the line e.g., "Right, second row down and fourth from the left"
- Matrix: imposes an abstract coordinate system on instances of the maze with a specified origin and numbers and/or letters for each axis e.g., "two, four".

Overall the transcripts generated a corpus of 975 descriptions each of which was classified into one of the above description types following the criteria in Garrod and Anderson (1987)[1].

In contrast to previous studies with the maze task, relatively few Matrix description types were identified in the corpus; 1.2% (11/975) compared with 23.4% in Garrod and Anderson (1987) and approximately 40% in Garrod and Doherty (1994). Rather than discard them from the analysis these were combined with Line descriptions into a composite category: Abstract. The rationale being that, relative to Path and Figural descriptions, both effectively preserve invariant information about the overall configuration of the mazes and rely less on the specific layout of any given instance of the maze.

In the corpus as a whole, the proportions of description types in each category were; Figural: 25%, Path: 53% and Abstract: 22%. Although the proportion of Path descriptions remained relatively constant across trials, the relative distribution of Figural and Abstract descriptions altered with a fall in the proportion of Figural shadowed by a rise in the proportion of Abstract. The distribution of the three description types is illustrated in Figure 2 with trials 1 & 2, 3

[1] Garrod and Doherty (1994) report 93% agreement between two coders using this scheme.

& 4 and 5 & 6 pooled to give three levels of experience; Low, Medium and High.

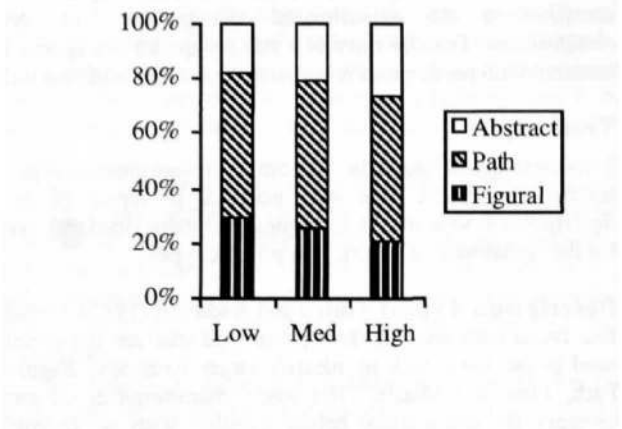

Figure 2: Proportion of Description Types Across Trials

Calculation of Pearson's product-moment correlation showed a strong negative relationship between the Figural and Abstract description types; r=-0.9. The reliability of this pattern was assessed firstly by calculating an omnibus Chi^2 for the raw frequencies of Figural and Abstract descriptions across all six trials. This showed a reliable change in their relative distribution; $Chi^2(5)=13.28$, $p=0.020$. Secondly, a focused comparison of the frequencies of Figural and Abstract descriptions on Trial 1 vs. Trial 6 also proved reliable; $Chi^2(1)=10.66$, $p=0.001$. The overall pattern with increasing experience at the task was thus reflected by a shift from the relatively concrete Figural descriptions to the more idealised Abstract descriptions.

Comparison of the distribution of description types in the Homogenous and Mixed dyads on trial six (illustrated in Figure 3) suggested a substantial difference between the two conditions. This was analyzed by calculating Chi^2 for the frequencies of Figural and Abstract description types in the Homogenous and Mixed dyads on trial 6. These were reliably different; $Chi^2(1)=11.00$, $p=0.000$.

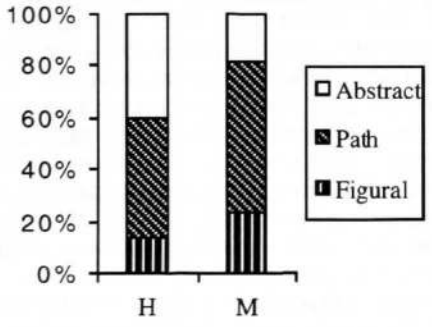

Figure 3: Distribution of Description Types in Homogenous (H) and Mixed (M) Dyads

A further comparison was made of the frequencies of Figural and Abstract descriptions in the Mixed Dyads on Trial 6 and all dyads on Trial 1. This indicated no reliable difference; $Chi^2(1)=0.944$, $p=0.331$.

Repair: Each description was additionally coded for whether or not it was repaired before the members of a dyad were satisfied they had completed the item; typically signaled by an "okay" or "yeah" or by moving onto the next page. Descriptions subject to either reformulation, where the location would be described again in different terms, or clarification, where a query such as "counting from the top or the bottom?" occurred, were classified as repaired. Simple recycling due to utterance overlap was not counted as an instance of repair for these purposes nor were 'echoes' of the original description that provoked no additional turns or clarifications.[2] Descriptions that were not repaired but led to incorrect identifications of the target location (effectively false positives) were not included in the analysis.

Across trials, the incidence of repair showed a steady decline illustrated in Figure 4. This pattern was reliable; Omnibus $F(5,137)=33.464$, $p=0.000$, Linear trend, $t(137)=3.669$, p(one-tailed)=0.000. Turning to the experimental comparison, the frequencies of repair in the Mixed and Homogenous dyads proved to be reliably different; $Chi^2(1)=4.449$, $p=0.034$. Homogenous dyads engaged in repair for only 21% of their descriptions whereas Mixed dyads repaired 51% of their descriptions.

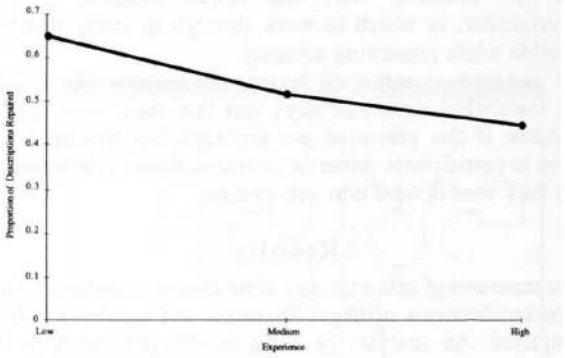

Figure 4: Fall in Incidence of Repair Across Trials

Discussion

Considering first the simple measures of task expertise, as individuals became more experienced at the task the number of items completed per trial increased and the number of errors identifying the target location decreased. In this study, no reliable difference was observed on these simple measures of task performance between the Mixed and Homogenous dyads.

[2] These commonly occured during participants attempts to interpret the previous description. They were spoken with a relatively reduced intensity and appeared to function as place-holders rather turns specifically directed at their partners (cf. Clark's and Schaefer's, 1989, notion of a 'display').

The language-specific or 'expertese' measures also showed a systematic change across trials. Increasing experience at the task was reflected in a shift from the relatively concrete Figural description type to more abstract Line and Matrix description types. Over trials, participants are exposed to increasing numbers of maze configurations that all conform to a basic underlying grid. This pattern is consistent with an expertise based explanation; Abstract description types effectively preserve information about what is invariant across different instances of the maze configuration. The proportion of descriptions subject to repair also showed a reliable decrease with experience, indicating that dyads became increasingly well coordinated in their production and interpretation of descriptions across trials.

However, in contrast to the simple measures of task expertise, the two measures of linguistic co-ordination were clearly sensitive to the experimental manipulation of dyad composition. In particular, it seems that transfer outside a subgroup interfered substantially with individuals' ability to manage their linguistic coordination with a new partner. Mixed dyads, composed of individuals from different sub-groups, both relied more on the relatively basic, Figural, description scheme and engaged in repair or renegotiation of descriptions approximately twice as often as Homogenous dyads. In fact, on these measures, Mixed dyads were comparable to the 'naïve' dyads performing the task for the first time.

Consequently, the observed differences between the Homogenous and Mixed dyads are not accounted for by simple appeal to task expertise. In both conditions individuals had essentially equivalent experience; they had to contend with new materials and a new partner on each trial. Furthermore, neither the initial division into sub-groups nor the subsequent crossing between sub-groups was indicated to participants and, on debriefing, none of the them reported detecting these aspects of the design. It is also worth noting that, in both parts of the experiment, each group of participants performed each trial synchronously in the same large room and were consequently exposed to identical treatment.

The difference between the subgroups also seems to elude explanation in terms of the CM. In both conditions the need to establish the relevant mutual beliefs about appropriate referring expressions (description types) was the same. In some sense, the observed difference resides in what initial linguistic common ground is available; dyads composed of individuals drawn from the same sub-group showing greater coordination than those from different sub-groups. However, the CM does not predict the group-specific nature of this initial common ground. The conditions for the establishment of mutual-belief, through explicit cycles of presentation and acceptance, are equivalent across all the sub-groups and both conditions and therefore fail to discriminate between them. More generally, the data reported here highlight the need for a more adequate account of how the personal common ground, accrued through direct interaction between individuals, can become established as part of the initial common ground for a particular community.

It appears that some additional group-specific mechanism is required to account for the emergence of task based sub-languages or dialects. In particular, one that predicts both the degree of convergence within a group and the interference that results from transfer outside the group.

A full account of such a mechanism is beyond the scope of this paper however it is possible to construct an outline proposal. Following Lewis (1969), linguistic conventions are assumed to be arbitrary solutions to recurrent problems of coordination. The initial common ground for participants in this task can be idealised as the resources available to them through the conventions developed for everyday coordination problems. In the experimental setting, these resources are recruited to the problem of dealing with the maze task and are roughly characterised here by the Figural description type. This utilises only very general, and largely ungeneralisable, means of representing target locations which depend on the relative salience of specific elements in each item and refer to them in terms of familiar shapes or patterns such as "triangle", "dog leg" or "left-indicator". During the task, participants attempt to coordinate their behaviour according to the criterion of success set by the task; successful and timely identification of target locations. Where problems arise in interpretation they can engage in minimal, task-oriented repair until they are mutually satisfied that an appropriate (although not necessarily correct) resolution has been achieved.[3] What terms such as "row" or "column" mean in this context will become progressively more specific each time a problem with their interpretation is detected and repaired. This provides a basic mechanism by which the relatively vague linguistic conventions available initially can alter and become more specialised during performance of the task.

Because a common interaction history is progressively building up across a sub-group, such recurrent patterns of repair will, over trials, tend to promote a relative degree of convergence or 'regression to the mean' within that group. The experimental task itself places only very weak constraints on the nature of the possible linguistic conventions that may emerge in response to the coordination problems it poses. There is therefore no prior reason to expect that the conventions that emerge in any two sub-groups will be comparable. Put another way, there is no reason to expect that any two groups will develop the same 'expertese'. Transfer outside groups is therefore likely to interfere substantially with whatever group-based linguistic resources have been built up over trials.

An explanation along these lines has several advantages. It provides for the development of group-specific sub-languages without recourse to explicit negotiation. As noted, this mechanism does not appear to be supported empirically and ultimately faces a boot-strapping problem ("What do you mean by 'what do you mean'?"). It also provides for the switch, observed in the Mixed dyads from Abstract description schemes to Figural description schemes. In principle, they could have switched to an alternative but equally abstract scheme, a suggestion consistent with the expectations of an expertise based account. The direction of this switch is, however, predicted by this account. Once the linguistic

[3] Note: this need not amount to explicit negotiation of the meaning of particular terms or expressions. For example, it may involve a refusal to accept one description followed by a reformulation that is locally satisfactory.

conventions developed within, and specific to, their subgroup are unavailable participants must switch to those of the community they share with their new partner i.e., the same conventions that were available on the first trial at the task.

Acknowledgments

The author wishes to acknowledge the EPSRC, for a research studentship, and the European Community for the IRS Network (CHRX-CT93-0099) and an HCM Fellowship (ERBCHBGCT 9405610) which provided the financial support that made this work possible. Thanks also to Jimmy Cuthbert for technical assistance on the order of divine intervention and to R. McCabe and the two anonymous referees who reviewed this paper for valuable comments.

References

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

Clark, H.H. and Wilkes-Gibbs, D. (1986) Refering as a collaborative process. *Cognition*, 22, 1–39.

Clark, H.H. and Schaefer, E.F. (1989) Contributing to discourse. *Cognitive Science*, 13, 259–294.

Garrod, S.C. and Anderson, A. (1987) Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181-218.

Garrod, S.C. and Doherty, G. (1994) Conversation, coordination and convention: An emprirical investigation of how groups establish linguistic conventions. *Cognition*, 53 181–215

Isaacs, E.A. and Clark, H.H. (1987) References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1) 26–37.

Lewis, D. (1969) *Convention: A Philosophical Study.* Oxford: Basil Blackwell.

Sacks, H. (1992) *Lectures on Conversation: Volumes I and II.* Oxford: Basil Blackwell

Wilkes-Gibbs, D. and Clark, H.H. (1992) Coordinating beliefs in conversation. *Journal of Memory and* Language, 31, 183–194.