Making a Contribution: Processing clarification requests in dialogue.

Patrick G. T. Healey, Arash Eshghi, Christine Howes and Matthew Purver
Queen Mary University Of London
Interaction, Media & Communication Research Group,
School of Electronic Engineering & Computer Science, England
{ph, arash, chrizba, mpurver}@eecs.qmul.ac.uk

Abstract: Clarification Requests (CR) provide a useful window on how contributions to dialogue are processed. We present chat-tool experiments that introduce CRs mid-turn into ongoing dialogue. The pattern of responses shows people are sensitive to both constituent structure at the interruption point and apparent origin of the CR: the conversational partner or a 'peripheral' chat-bot participant. The results highlight the need for context-sensitive, sub-sentential but structured and incremental processing models.

Introduction: People can collaborate to build up contributions to conversation:

A: They X-rayed me, and took a urine sample, took a blood sample. Er, the doctor …

B: Chorlton?

A: Chorlton, mhm, he examined me [. . .]

(From the British National Corpus)

B's clarification request (CR) offers a possible refinement of the reference of 'doctor' before A has completed their sentence. This highlights the fact that a) contributions to dialogue are routinely built by more than one participant (see Clark, (1996)) and b) this process is incremental. A CR can be posed, understood and responded to before its antecedent sentence is finished. This raises two basic questions about language processing in dialogue:

1. What is the structural organization of the increments to a contribution to dialogue?
2. How does collaboration with a conversational partner affect the processing of those increments?

Increments could be organized in a variety of ways and this affects how parts of a contribution can be accessed or amended. A word- or string-based processing model would predict that the distance between the target of a CR (A's "the doctor") and the interruption point (where B says "Chorlton?") would be a key factor in the processing of the CR. A flat semantic model, in which phrases introduce discourse referents into an unstructured record (e.g. Kamp & Reyle (1993)), might predict no effect of interruption point. On the other hand, a model based on constituent structure, either syntactic (e.g. Demberg & Keller (2008)) or semantic (e.g. Kempson et. al. (2001)), would predict effects of structural factors at the point of interruption.

The second question arises because collaboration with a conversational partner can affect language processing. Most work on the effects of interruption has focused on individuals completing a task on their own and focuses on factors such as memory load. However, collaborative tasks, like conversation, place additional constraints on processing (Chevalley and Bangerter, 2010). In the example above, A and B suspend construction of the current turn, engage in a clarification sub-dialogue, and resume the original turn. These transitions need to be coordinated, possibly requiring extra effort and affecting processing.

We report an experiment that addresses two specific questions:

1. Does constituent structure affect the ease with which a CR is processed (independently of linear string distance)?

2. Is a CR processed differently if it comes from an active conversational partner or a ratified but inactive participant?

Method:

We test these predictions experimentally by introducing artificial probe CRs into live text-based dyadic dialogues. Using a character-by-character version of the DiET chat-tool methodology (Healey et. al. 2003), we track the syntactic structure of turns in real time as they are typed. This enables us to select Noun Phrases (NPs) as potential targets for a CR. We then insert probe CRs – verbatim repeats of the target NP plus a question mark (Reprise Fragments) – into the dialogue. The probe CR interrupts the person typing, who responds (both probe and response are not relayed to the other participant) and then the dialogue resumes.

Two experimental manipulations are used:

1. Insertion Point: (Within-subjects) The CR target is always a prior, completed NP. The probe CR is inserted either a) within a subsequent incomplete constituent (e.g. after a determiner) or b) at the end of a subsequent complete constituent (e.g. after a completed NP). Note that the average string distance between target and probe increases a) to b).
2. Apparent Origin: (Between-subjects) The probe CR appears to originate either from the actual conversational partner or from a chatbot ('Q-Bot'). In both cases CRs are triggered in exactly the same way, as in 1.

34 pairs of participants carried out 'balloon task' debates using the DiET chat tool. Participants were exposed to an average of 3.7 spoof CRs.

To assess how the CR was processed by the recipient, three dependent measures were used: a) Speed of response onset, to index how easily the response is produced; b) Frequency of CR target reformulation, as a measure of semantic reprocessing; c) Restarts: the extent to which the response repeats *non*-target elements of the interrupted turn in a structure-preserving way, as a measure of the effort needed to deal with the interruption.

Results:

Response Onset*:* Insertion Point has no reliable effect on speed to respond (within constituent 7.0 seconds, between constituents 7.4 seconds ($F_{(1,262)} = 0.56$, $p = 0.45$), but people are quicker to respond to Q-Bot CRs, 6.4 seconds, than ones from their conversational partner, 8.0 seconds ($F_{(1,262)} = 10.72$, $p = 0.00$)

Reformulation*:* No reliable effect of Insertion Point on likelihood that the response reformulates the target ($\chi^2_{(2)} = 0.59$, $p=0.80$) but the response to the Q-Bot is more likely to reformulate the target fragment, Q-Bot: 53%, Other: 37% ($\chi^2_{(2)} = 6.67$, $p=0.01$).

Restart*:* People are more likely to restart if the CR occurs within-constituent ($\chi^2_{(1)} = 13.34$, $p=0.00$) (Within Constituent: 54%, Between Constituent: 32%) and if the CR appears to come from their conversational partner ($\chi^2_{(1)} = 6.89$, $p=0.01$, Q-Bot: 34%, Other: 51%)

Concluding Discussions:

CR Insertion Point does not affect the time taken to respond or the level of semantic processing used to produce the response. This suggests that the difficulty of processing the CR itself is not significantly altered by the structure interrupted by the insertion of the CR. However, the restart data indicates that Insertion Point does affect the effort needed to deal with the interruption and subsequently rejoin the conversation. Specifically, it appears to be more difficult to integrate the CR response into the resumption when the interruption is mid-constituent; despite the fact that CRs are systematically closer to their targets in this case. This provides evidence that structured syntactic/semantic constituents are integral to the organization of contributions and demonstrates the need for sub-sentential but structured, incremental processing (e.g. Kempson et. al. (2001)).

Apparent origin of CR, by contrast, does have significant effects on processing. People respond more quickly and also reformulate more when the question appears to come from the Q-Bot. This suggests that participants attend to the Q-Bot and answer the question. However, because the Q-Bot is not a full participant they do less to integrate their response, in terms of restarts, into the interrupted turn.

A second explanation is that people give the Q-Bot systematically different answers because they assume it does not have the same shared context. For example, the CRs might be more easily taken to be asking about the content or reference of the queried NP. If this requires a lower level of processing effort, then there should be less disruption, consistent with faster response onsets in the Q-Bot condition.

References:

Chevalley, E. and Bangerter, A. (2010). Suspending and Reinstating Joint Activities With Dialogue. *Discourse Processes.*

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

Demberg V. and Keller F. (2008) A Psycholinguistically Motivated Version of TAG. *In Proceedings of the International Workshop on Tree Adjoining Grammars.*

Healey, P. G. T., Purver, M., King, J., Ginzburg, J., and Mills, G. (2003). Experimenting with clarification in dialogue. In *Proceedings of the 25th annual Conference of the cognitive science society.*

Kamp, H. and Reyle, U. (1993). From Discourse to Logic. *Dordrecht: Kluwer.*

Kempson, R. Meyer-Viol, W. and Gabbay, D. M. (2001). Dynamic Syntax: The Flow of Language Understanding. *Wiley-Blackwell.*