"Well, that's *one* way": Interactivity in parsing and production

Christine Howes, Patrick G. T. Healey, Arash Eshghi, and Julian Hough

Queen Mary University of London, Cognitive Science Research Group, School of Electronic Engineering and Computer Science, London E1 4NS, United Kingdom.

c.howes@qmul.ac.uk   ph@eecs.qmul.ac.uk
arash@eecs.qmul.ac.uk   julian.hough@eecs.qmul.ac.uk
http://www.eecs.qmul.ac.uk/~chrizba/

Abstract: We present empirical evidence from dialogue that challenges some of the key assumptions in the Pickering & Garrod (P&G) model of speaker-hearer coordination in dialogue. The P&G model also invokes an unnecessarily complex set of mechanisms. We show that a computational implementation, currently in development and based on a simpler model, can account for more of this type of dialogue data.

Pickering & Garrod's (P&G's) programmatic aim is to develop an integrated model of production and comprehension that can explain intra-individual and inter-individual language processing (Pickering & Garrod 2004; 2007). The mechanism they propose, built on an analogy to neuro-computational theories of hand movements, involves producing and comparing two representations of each utterance; a full one one containing all the structure necessary to produce the utterance and an "impoverished" efference copy that can predict the approximate shape the utterance should have.

Although not our central concern, there is a tension between endowing the efference copy with enough structure to be able to predict semantic, syntactic, and phonetic features of an utterance and nonetheless making it reduced enough that it can be produced ahead of the utterance itself. To avoid a situation in which the "impoverishment" proposed for the efference copy is just those things not required to fit the data, we need independently motivated constraints on its structure.

Neuro-computational considerations might provide such constraints, but there are dis-analogies with the models of motor control P&G use as motivation. Efferent copies were originally proposed to enable rapid cancellation of self-produced sensory feedback, for example, to maintain a stable retinal image by cancelling out changes due to eye-movements. However, the claim that we use an analogous mechanism to predict, and correct, linguistic structure before an utterance is produced involves something conceptually different. The awkwardness of phrases such as "semantic percept" highlight this difference; until the utterance is actually produced there is nothing to generate the appropriate sensory percept. Conversely, if the "percept" is internal we are still in the cognitive sandwich.

These points aside, the target article provides a valuable overview of the evidence that language production and comprehension are tightly interwoven. P&G's main target, the "traditional model", treats whole sentences, "messages" or utterances as the basic unit of production and comprehension. However, there is evidence from cognitive psycholinguistics and neuroscience to show that language processing is tightly interleaved around smaller units. The close interconnections between production and comprehension are especially clear in dialogue where fragmentary utterances are commonplace and people often actively collaborate with each other in the production of each turn (Goodwin 1979).

It is unclear if the interleaving of production and comprehension requires internally structured predictive models. Recent progress on incremental models of dialogue suggest a more parsimonious approach. In our computational implementation based on Dynamic Syntax (Purver et al. 2006; 2011; Hough 2011), the burden of predicting full utterances does not need to be employed in parsing, as speakers and hearers have incremental access to representations of utterances as these emerge. Contrarily, P&G's approach to self-repairs is analogous to Skantze and

Hjalmarsson's (2010), which compares string-based plans and computes the difference between the input speech plan and the current state of realisation. In our model, instead of having to regenerate a new speech plan from scratch, we can repair the necessary increments, reusing representations already built up in context, which are accessible to both speaker and hearer. Currently, it is difficult to distinguish empirically between a dual-path model with predictions and a single-path incremental model because both combine production and comprehension.

As the paper highlights, the "vertical" issue of interleaving production and comprehension is independent from the "horizontal" problem of accounting for how language use is coordinated in dialogue. Nonetheless, this article extends previous Pickering and Garrod work (2004; 2007) in claiming that the model of intra-individual processing can be extended to inter-individual language processing (conversation). Unlike previous work, the new model operates in different ways for speakers and hearers, and the potential for differences between people's dialogue contexts is acknowledged (although not directly modelled).

The problem with this generalisation is that in dialogue we do not just predict what people are going to say, we also respond. Even if I could predict what question you are about to ask, this does not determine my answer (although it might allow me to respond more quickly). In terms of turn structure, all a prediction can do is make it easier for me to repeat you. Repetition does occur in dialogue but is rare and limited to special contexts. Corpus studies (Healey et al. 2010) indicate that we repeat few words (less than 4%) and little more syntactic structure (less than 1%) than would be expected by chance. Crudely, a cross-person prediction model of production-comprehension cannot explain 96% of what is actually said in ordinary conversation.

One conversational context that seems to depend on the ability to make online predictions about what someone is about to say is compound contributions, in which one dialogue contribution continues another, as in this excerpt from Lerner (1991):

> Daughter: Oh here Dad, one way to get those corners out
> Father: is to stick your fingers inside
> Daughter: Well, that's *one* way.

Although it is unclear whether a predictive model better accounts for the father's continuation than one in which he is building a response based on his partial parse of the linguistic input, the daughter's response seems to be based on the mismatch between what was said and what she had planned to say. Although possible she was predicting he would say what she herself had planned to, there is no need for this additional assumption. Many cases of other-repair (Schegloff 1992) such as clarification requests asking what was meant by what was said (e.g., "what?") also seem to require that any predictability used is impoverished at precisely the level it might be useful.

In a study on responses to incomplete utterances in dialogue (Howes et al. 2012), increased syntactic predictability led to more clarification requests. Although participants made use of different types of predictability in producing continuations, predictability was neither necessary nor sufficient to prompt completion, and, in extremely predictable cases, participants did not complete the utterance, responding as if the predictable elements had been produced. Our assumption is that it is the things we cannot predict that are the most important parts of conversation. Otherwise, it is hard to see why we should speak at all.

Seeking predictions from a predictive framework

T. Florian Jaeger[a,b] and Victor Ferreira[c]

[a]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627-0268; [b]Department of Computer Science, University of