Communicative Goals

Rich Thomason, Jerry R. Hobbs, and Johanna Moore

1 The Relationship between Interpretation and Generation

There are several phenomena that point to the close connection betweeen interpretation and generation. One of these is the "co-construction" of single sentences by more than one speaker. In

[A: So that leaves only ten minutes for your talk.
[B: no time

B is interpreting for the phrase "so that leaves". He then generates a continuation of that himself. It is not exactly the way A continues the utterance, but in the context it comes from, it carries an equivalent sense. After his contribution, he reverts to interpretation mode for the rest of A's utterance. This rapid alternation between interpretation and generation suggests a close relationship between the two processes.

The second phenomenon is what Dubois (1996) has called "dialogic syntax". He notes that whereas modern language text books decry the artificiality of reciting paradigms, in fact much discourse is exactly like this. He gives the following example, in which A is a woman, B is her male significant other, and C is a male friend of B's.

A: I went dancing last night, and ...
B: You went dancing last night?
C: She went dancing last night.

Such repetitions are very common in casual discourse. This "reuse" of the same phrases suggests that structures activated during interpretation are employed again in generation.

The third phenomenon is quantity implicatures. Consider

A: Are you 21?
B: Yes.
A: How old are you?
B: 28.

The first two turns indicates that in some contexts "21" can mean "over 21", say, when uttered by a bartender wanting to know if a customer is legally allowed to drink. The more common case is illustrated by the last two turns. B is probably exactly 28 (unless intending to lie about his age). In most contexts, if B were over 28, he would have said so. But notice that in reasoning thus, we are interpreting an utterance by considering the other ways the speaker could have conveyed or generated the same meaning. If we must reason about generation to interpret utterances, there must be a close relationship between the two.

2 Abduction

Abduction is a general method of inference from observations to underlying explanations. In recent years, it has been recognized as one of the central processes in cognition, as more and more cognitive processes are seen to be instances of it. We believe that it provides a unified framework for viewing the interpretation and generation of natural language.

Let us look first at the interpretation and generation of events in the physical world. In the classic example of abduction, a person goes outside in the morning and sees that his lawn is wet. There are two possible explanations—it rained last night, and the sprinkler was on. Suppose he knows for independent reasons that it is unlikely that it rained last night. Then he will assume that the sprinkler was on.

This is illustrated in Figure 1. The person has made the most reasonable assumption that would explain the observable facts. This assumption corresponds to new information he now has.

[FIG 1: Abduction in interpreting the world]

But notice that this same representation of possible causality can be used in changing the world as well, in generating actions. Suppose the person wants to have the lawn wet. He knows from the causal knowledge represented in this figure that there are two ways this could be brought about. He could make it rain. Or he could cause the sprinkler to be on. The first is not very feasible, so he decides to do the second. His task is to make the top-level goal true, and to do that he examines its causal structure, Performs the equivalent of making the assumptions that make it true on the basis of what is most feasible, and decides to cause that event to occur. The role of assumptions in this case is that they are the actions to execute.

In both the interpretation and generation problems, the underlying causal structure is the same. The propositions or events that are labelled as assumptions are those that are not known to be true that play an essential role in the causal network. In interpretation the assumptions are hypothesized to be true and constitute the new information. In generation the assumptions are to be made true by the agent.

In Hobbs et al. (1993) a very general framework for the interpretation of discourse based on abduction was presented. It subsumed a large number of interpretive processes under a single process, thereby making it easier to see how various factors could interact in determining the interpretation of texts. Fig 2 illustrates this for the sentence

> The Boston office called.

[FIG 2: Abduction in Interpreting Text - 1]

The aim of the abduction process is to prove that the string of words is a coherent, grammatical utterance conveying an eventuality e. This is done by proving that it has the right sort of syntactic structure, proving that this structure is one that will convey a particular logical form, and proving the logical form abductively. The "processes" that are subsumed under abduction in this case include syntactic composition, semantic interpretation, metonymy resolution, coreference resolution, and the interpretation of the vague relation conveyed by the compound nominal.

In this paper we will not address all of these complexities. Rather we will focus on a much simpler example.

> John called.

Figure 3 illustrated the interpretation.

[FIG 3: Abduction in Interpreting Text]

The sentence is segmented into the subject and the verb phrase. An axiom of the form

$$call'(e,j) \wedge person(j) \wedge Occ(\text{``called''}) \supset Syn(\text{``}called'', e, j, \ldots)$$

says that if e is a calling event by j where j is a person, and the word "called" occurred, then the word "called" could be being used to convey the eventuality e if a subject referring to j is found. The lexical axiom for

"John" is similar. The occurrence of "John" and "called" are established because it is known, say, that A uttered these words to B. The proposition $person(j)$ follows from the resolution of j to an entity such that $John(j)$ holds. The proposition $call'(e, j)$ is assumed to be true and constitutes the new information.

Now consider the generation of the same sentence. The proof is exactly the same, except for what needs to be assumed. The new information, $call'(e, j)$, is already known by the speaker. What is not known is that the "speaker" B has uttered the words to the "hearer" A, because in fact he hasn't yet. So these propositions are assumed, and in generation these assumptions correspond to the actions to execute. Thus, B utters to A the words "John" and "called".

[FIG 4: Abduction in Generating Text]

Now consider a third case. A begins the sentence, "John ..." B already knows this information, so he completes it, "*ldots* called." The state of B's knowledge can again be represented by the same causal structure, illustrated in Figure 5.

[FIG 5: Abduction in Co-Construction of Text]

Since B already knows $call'(e, j)$ and is able to predict the entire sentence, the proof is already entirely present to him after the utterance of the first word. The only thing that can't be proved is that the word "called" occurred—Occ("called")—since it hasn't yet. B can however make this true by uttering the word "called" himself. So he does.

We thus see the proof that in our abductive approach constitutes the interpretation functions as well in the generation and the co-construction of the text. All that differs from one case to the next are the assumptions that must be made for the proof to go through, and whether these assumptions are then treated as new information or as actions to be executed.

This view of interpreting and generating text treats the text as though it were entirely separate from the world. But in fact the occurrence of the text is itself an event in the world and thus an observable to be explained. It is thus desirable to subsume our picture of interpreting and generating texts under the previous picture of abduction in interpreting and changing the world. Figure 6 illustrates how this is done.

[FIG 6: Embedding the Text in the World]

A common explanation for the utterance of a string of words by someone is that the string constitutes a grammatical sentence conveying the eventuality e and the speaker has the goal of having the hearer believe e. This can be expressed in the following axiom:

$$Syn(w,e,\ldots)\$goal(a,believe(b,e)) \supset utter(a,b,w)$$

Proving the Syn predication takes us into analyzing the syntactic structure of the sentence as before, and then via the lexical axioms into reasoning about world knowledge. The *goal* predication leads us to reason about the speaker's goals in making this utterance, and this generally leads to reasoning about world knowledge as well. The interaction between the syntax and semantic interpretation on the one hand and the reasoning about the speaker's goals on the other is captured by the fact that the two sides of the proof graph share variables and are thus mutually constraining.

This picutre needs to be made more sophisticated in several ways. First of all, the speaker's goal is generally not merely to cause the speaker to believe a proposition, but rather to make it mutually believed among the two of them. The speaker is trying not oly to get the hearer to believe e but also to believe that the speaker believes e and so on.

The second was that the picture needs to become more sophisticated is that we must replace "believe" with something more like "entertain" (Walker, 19??). We do not come to believe something merely because we hear it said; persuasion would not make sense in that case. But we do come to entertain the thought. Moreover, we mutually entertain the thought, in that we believe the speaker is entertaining the thought, and that the speaker believes we are entertaining the thought, and so on. Some things are not believed right away, and some things are not even intended to be believed, such as metaphor, irony, jokes, and hypotheticals. Entertaining is similar to belief in that inferences can be drawn, but it differs in that we do not act on the propositions we entertain, whereas we are likely to act on the propositions we believe. Belief revision, the process of incorporating the new information we learn into our system of beliefs, can be viewed as a separate and logically subsequent process.

These considerations lead to the notion of the "Conversational Record" (Thomason, 19??). The Conversational Record is the set of propositions that are mutually intertained by the participants in the conversation. The Conversational Record is a part of the world, just as the conversation is and just as the beliefs of the participants are, and hence it is something that part of the conversation can be about. Correcting misunderstanding, for example, is conversation about the Conversational Record.

We have said that the interpretation of a text or sequences of utterances can be represented by a proof graph. In an ideal conversation, the speaker and hearer have identical proof graphs for the discourse that has occurred so far. Generation is thus is not merely the process of planning utterances, but rather planning utterances in a way that identical interpretations or proof graphs will result. We plan interpretations rather than planning utterances.

3 Context and the Selection of Interpretations

Hasida et al. (1996) introduces the notion of "communication game" to account for how utterances and meanings or contents are associated with each other in different contexts. Consider the situation illustrated in Figure 7.

[FIG 7: Communication Games]

Here there are two possible contents—$call(J_1)$ (P_1) and $call(J_2)$ (P_2). There are three possible utterances—"John called" (U_0), "John Black called" (U_1), and "John Green called" (U_2). Each of the utterances has an associated utility, which we can take to be inversely proportional to its length. Each of the contents has an associated probability, the probability of its being the message that will be conveyed at this moment. The problem is to construct a mapping between contents and utterances that is in some sense optimal. The suggestion is that it should maximize expected utility:

$\sum_{i=0}^{2} \sum_{j=1}^{2} U_i P_j$

This picture allows us to see graphically how one must sometimes reason about generation in the course of interpretation. The hearer of the sentence "John called" could reason that A said "John called", it could mean either that J1 called or that J2 called, but if A had meant that J2 called, he would have said "John Green called." Therefore he must have meant that J1 called. This is essentially the structure of reasoning that must take place in quantity implicatures.

One problem with this kind of reasoning is that it seems to involve excessive complexity. Instead of merely computing the link between P1 and U0, the hearer must also compute the link between P2 and U0 and the link betweeen P2 and U2, in order to determine the correct interpretation. How is such complexity possible?

We can begin to answer this question by relating the communication game framework to the logical level as captured by the abduction framework. Each of the arrows in Figure 7 in fact expands at the logical level into an entire proof, as shown in Figure 8.

[FIG 8: The Logical Level]

The utterance U1 is represented here as the predication $Syn($"John Black called"$,\ldots)$, and similarly for U0 and U2. Content P1 consists of the propositions $John_1(x)$ and $call(x)$, and content P2 of $John_2(x)$ and $call(x)$. The arrows in Figure 7 correspond to the proofs shown in Figure 8. What we see is that the mappings do no consist of discrete arrows but of highly overlapping proofs. This picture makes it much more plausible that in working out one Content-Utterance link, the language user explores some adjacent nodes and is able to rule out other possibilities, essentially by reasoning that, as mappings, they would not be as good.

What emerges is that what are distinct mappings at the level of the Communication Game, are highly overlapping proof graphs at the Logical level. Utilities at the level of the Communication Game are related to the size of the proof graph at the Logical level. Specifically, it is reasonable to identify them as a measure of that part of the reasoning that must be done in the syntactic part, involving the predicate Syn. Utility is thus not merely a measure of the difficulty of uttering something, but the corresponding difficulty in interpreting it.

At the level of the Communication Game, context can be seen as a matter of differing probabilities on the messages. If John Black is a co-worker and John Green your wife's brother, then in the office, "John called" may be the optimal way of saying that John Black called, and at home the optimal way of saying John Green called.

There are several ways that notions of context can be captured at the Logical level. These are spelled out in various places in Hobbs et al. (1993), but it is useful to review them here. In weighted abduction, propositions that are mutually entertained, whether because they are mutually believed or because prior discourse has placed them in the Conversational Record, are available for free. Thus, what has happened in the previous discourse or is otherwise salient can influence what interpretations are chosen.

Moreover, whole topics, or clusters of axioms, can be made available by paying the price for it once. Suppose we have a number of axioms of the form

$$
\begin{aligned}
&cluster_1 \wedge p_1(x) \supset q_1(x)\\
&cluster_1 \wedge p_2(x) \supset q_2(x)\\
&cluster_1 \wedge p_3(x) \supset q_3(x)
\end{aligned}
$$

where $cluster_1$ is a proposition that cannot be proved but must be assumed for some cost. Then it will cost to use any of these axioms, but using one once and paying the cost will "release" all of the other axioms in the cluster for free use. The proposition $cluster_1$ can be thought of as saying something like "This cluster of facts is relevant to the discourse at hand".

We thus see that the effects of context can be modelled at the level of the Communication Game as differeing probabilities and utilities on the utterances and their possible contents. At the Logical level, they are approximated by the weights and costs in weighted abduction.