Handling Non-Sentential Utterances in a Continuous Understanding Framework

Carlos Gómez Gallo
University of Rochester, Rochester, NY, USA
cgomez@cs.rochester.edu

Abstract

The goal of my research is to understand speech input in a continuous manner by treating the input stream as fragmental utterances. This allows us to use various approaches to predict what comes downstream. Possible interpretations are trimmed by such predictions which in turn also allow us to complete information not readily available in the fragmental utterance. Semantic frames can encode all possible arguments for domain actions. As utterances are processed continuously, appropriate frames can be activated so that fragment interpretations can fill, correct or extend frames under consideration. In turn, feedback can be provided to the parser as the frames are manipulated possibly based on the completeness of the semantic frame construction.

1. Problem Statement

My current research interests are focused on understanding speech input in a continuous manner by treating the input stream as fragmental utterances. Following the work on continuous understanding (Stoness, 2004; Aist et al., 2006, Gómez Gallo et al., 2007), such an approach implies three things. First the processing of utterances is incremental as the speech stream arrives. This means there is no need to wait until the end of the speaker's turn to find an interpretation. Second, the modules in the conversation agent architecture operate asynchronously. For example, the pragmatic module (e.g., intention recognition) can start producing an interpretation for the utterance before it has received the entire output from the parser. And third, knowledge can be shared among the different modules simultaneously as opposed to a pipeline architecture flow. For instance, the parser may take as input interpretations from the intention recognition module as they become available, and vice versa.

Processing sentences in this way has a number of advantages. First of all, it allows us to predict what the following utterance fragments will be. For instance, when processing a single word that can be interpreted as an action (e.g., a move action), we can activate the knowledge we have about such an action. This knowledge encodes the verb and argument structure and can have the form of a semantic frame. As processing of new fragments from the utterances proceeds, new slots in the frame can be filled, replaced or extended. Using these frames as the main structure for representing discourse, we can smoothly incorporate partial information from fragmental input and

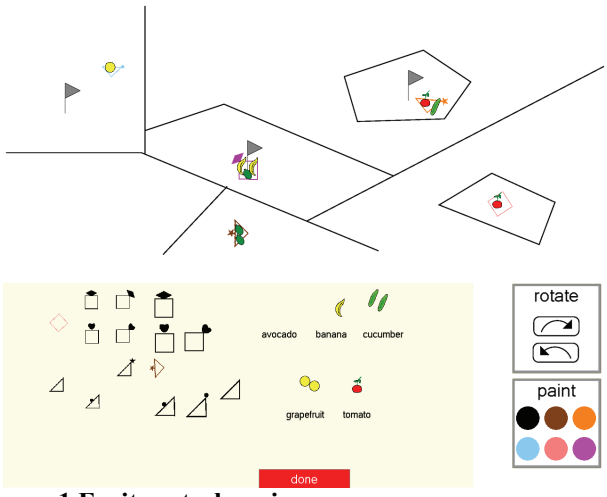

Figure 1 Fruit carts domain

also compute some type of measure of completeness (or appropriateness) as feedback to the parser. This measure will rate how a new constituent under consideration fits the expectations of the discourse analysis at that point. We are then in a better position to boost the constituent probability with knowledge from actions in the domain, previously understood utterances, and the way people signal information repairs.

The data I am going to focus on in my thesis comes from types of non-sentential utterances that arise in dialogue when highly defined domains exist in which users can rely on context to be understood. This is the case of the Fruit Carts domain, which is designed to investigate issues in continuous or incremental understanding. In this experiment, subjects are given a map similar to Figure 1 which describes a set of fruits and geometric figures, with different color, size and angle configurations. The subject's task is to replicate the map by giving commands to another human user. An excerpt from a simple dialogue of the data we collected is in Table 1. Here we can see a number of non-sentential utterances such as "to the left" or "keep going".

USR> Um, let's see
USR> Then in Morningside there needs to be a triangle with a star on its hypotenuse
USR> Right there and then it needs to be rotated um
USR> to the left
USR> keep going
USR> right there

Table 1. Example dialogue from Fruit Carts Domain

Sentences are typically encoded by grammars as containing a noun phrase and verb phrase. From this

dialogue we can see that even though these utterances do not achieve a full sentential realization from a syntactic point of view, they do carry propositional content. Other examples include simple answers to questions (e.g., when a person is asked "where is your car?", a proper answer can be "over there"). Non-sentential utterances are enough for us to reconstruct the propositional content intended by the speaker. In the next section, I describe the completed and future work of my research.

2. Research Status

I am currently extending the TRIPS architecture (Ferguson and Allen, 1998) to interpret fragmental utterances. The idea is to have incremental input going into the parser and constituents being fed forward as soon as they are built. The module will then construct a discourse history that is composed of the active set of semantic frames under consideration. As a first step, probabilistic feedback to the parser has been implemented and is explained next.

2.1 Completed Work

Parser feedback has been implemented as a first step to include domain specific knowledge into the analysis of constituents. This work, presented in Aist et al. (2006), improved parser efficiency as measured by the number of constituents built. Even though the Fruit Carts experiment allows users to use freestyle language, the set of actions that can be performed on objects provide us with a well-defined construction that we can exploit. Table 2 summarizes the library of actions along with all thematic roles expressed that were seen in the data.

Action	Thematic Roles
Move	verb, object, distance, heading, loc
Rotate	verb, object, angle, heading
Select	verb, object
Paint	verb, object, color

Table 2. Actions and their prototypical arguments.

Due to elliptical constructions, not all arguments are realized in every speech act. Certain argument realization patterns are more common than others. The parser can use this information to adjust what arguments it expects to see and prioritize ones commonly seen in the corpus. This helps the parser arrive at a more accurate analysis with less effort.

To this end, an initial set of six dialogues were manually annotated with verb and verb argument type labels. Then statistics that measured how often a verb argument appears given the verb were collected. For example, the most likely MOVE action is performed by giving the verb, object and location. This is intuitively correct, but occurs only 66% of the time. MOVE actions are also specified by stating a location only; in such cases the object is presumably apparent from the context, perhaps by a previous SELECT action. This is the case of object elision.

The mechanism works as follows. The TRIPS parser is domain independent. The logical form (LF) it constructs, however, is translated to our domain specific semantics. When the parser is constructing a VP, it asks the VP advisor how likely the construction under consideration is in this domain. The VP adviser acts on the translated domain-specific LF. Therefore, we can think of the advisor as encoding semantic restrictions for each verb. The parser then modifies the probability of the constituent in the chart and puts it back into the agenda.

Experimental results show that on average the number of constituents built by the parser decreases with the VP advice. The best result can be seen on sentences as complicated as the following: “take the box in Morningside and put it into pine tree mountain on the bottom of the flag”; here, the number of constituents were decreased by as much as 19%. On less complex sentences such as "and then change it to brown" there is no difference in the number of constituents, since the parser already finds a spanning parse efficiently.

2.2 Future Work

An algorithm for fragment processing will have as input partial analysis from the parser. As these come along they will trigger semantic frames available from the action library. How many frames are needed and how to rank them needs to be explored. The type of fragment will reveal information on which frames can be brought to context. For example, a fragment may be a location, or an object, or an adjective. This information can be used to know which new frames to activate, or, on the other hand, which already existing frames to extend.

I am currently working on annotation of a gold standard corpus which will relate thematic roles with their respective actions and specify the current state of the world at that moment (Gómez Gallo et al., 2007). I will also need to find a set of repair signals for clarifying a previously stated object or undoing a previous action. Finally, I intended to explore the automatic construction of semantic frames from the action library.

3. References

Aist, G., Allen, J., Campana, E., Galescu, L., Gómez Gallo, C., Stoness, S., Swift, M., and Tanenhaus, M. 2006. Software architectures for incremental understanding of human speech. Interspeech 2006.

Ferguson, G. and Allen, J. 1998. TRIPS: An Integrated Intelligent Problem-Solving Assistant. AAAI: pages 567-572.

Gómez Gallo, C., Allen, J., Swift M., Coria, S., Pardal, J., de Beaumont, W., Gegg-Harrison, W., and Aist G. Annotating Continuous Understanding in a Multimodal Dialogue Corpus. Decalog 2007.

Stoness, S., Tetreault, J., and Allen, J. 2004. Incremental Parsing with Reference Interaction. ACL 2004.