Accessing Previously Shared Interaction States through Natural Language

Arthi Murugesan[1], Derek Brock[2], Wende K. Frost[2], and Dennis Perzanowski[2]

[1] NRC/NRL Postdoctoral Fellow
[2] Naval Research Laboratory, 4555 Overlook Ave. S.W.,
Washington, DC 20375 USA
{Arthi.Murugesan.ctr,Derek.Brock,Wende.Frost,
Dennis.Perzanowski}@nrl.navy.mil

Abstract. An important ambition of human-computer interaction research is to develop interfaces that employ natural language in ways that meet users' expectations. Drawing on elements of Clark's account of language use, this idealized form of human-computer interaction can be viewed as a coordination problem involving agents who work together to convey and thus coordinate their interaction goals. In the modeling work presented here, a sequence of interrelated modules developed in the Polyscheme cognitive architecture is used to implement several stages of reasoning the user of a simple video application would expect an addressee—ultimately, the application—to work through, were the interaction goal to locate a scene they had previously viewed together.

Keywords: natural language; coordination problem; common ground; salience; solvability; cognitive modeling; Polyscheme.

1 Introduction

An important ambition of human-computer interaction research is to develop interfaces that employ natural language in ways that meet users' expectations. Drawing on elements of Clark's [1] account of language use, this idealized form of human-computer interaction can be viewed as a particular type of coordination problem [2] involving agents who work together to convey and thus coordinate their interaction goals. People rely on a procedural convention for doing this that can be summarized as follows: in posing a coordination problem for an addressee to solve, one is expected to set things up so that the effort needed to work out the intended solution is minimized. "Setting things up" entails the use of a system of straight-forward practices that includes 1) making the focus of the coordination problem explicit or salient and 2) posing a problem one expects the addressee can solve.

In the modeling work presented here, a sequence of interrelated modules are developed with the Polyscheme cognitive architecture [3]. These modules simulate the stages of reasoning a user of a simple video application [4] might expect a real person standing in for the application to execute, if the goal were to find a frame or scene in the video they had both seen at an earlier time. The user's verbal description of the

C. Stephanidis (Ed.): Posters, Part I, HCII 2011, CCIS 173, pp. 590–594, 2011.

scene is treated as a coordination problem that the application must try to solve. Accordingly, each word and the higher-order semantics of the description—its conceptual references to objects, places, and events—are matched against a body of domain-specific lexical and schematic representations held by the application. Barring any failures at this level, the product of the preceding stages is then used to infer which scene, among those the user has previously inspected, is the one the user intends for the application to find.

2 Coordinating the Expression and Interpretation of Intentions

Natural language can be viewed as a collaborative means for expressing and interpreting intensions [5] through a body of widely shared conventions [2]. Clark [1] characterizes the challenge of conveying an intention from one agent to another—for example, from a speaker to an addressee—as a coordination problem that participants must work together to solve. To get to the intended solution, or a solution that will do, individuals routinely proceed in a conventional collaborative way. In particular, they rely on certain heuristic presumptions regarding a set of actions they expect to carry out together, which includes posing and grasping the problem and working out and acting on the result. Instantiations of three of these presumptions are modeled in this work from the point of view of the addressee. They are: 1) *common ground*—it is expected that the speaker has taken into account knowledge believed to be shared with the addressee as a basis for the words and actions that are used to convey the intention; 2) *salience*—the words and actions the speaker uses are expected to make identification of the intention and its implications prominent or obvious on the basis of common ground; and 3) *solvability*—the speaker is expected to have an intended result in mind and to have framed the expression of the intention so the addressee can readily "solve" or work out what the intended result must be and act on it.

3 Application Setup and Tools Used

The three heuristics cited above are implemented to work with a simple application that is loosely based on an experimental test bed known as InterTrack [4] used in the authors' lab for research on advanced traffic monitoring and user-interaction techniques. In this variant of InterTrack, a scene of interest is "shared" with the application by clicking somewhere in a video frame, which is then referred to as a "card." Automated identification of objects and events in shared scenes has not been implemented, so what the application "knows" about a specific scene is currently coded by hand. Once a shared card has been created, the user can then access it at a later time with a written sentence.

Computational implementation of natural language interactions is a complex undertaking that requires both cognitive modeling and linguistic tools. The Polyscheme cognitive architecture [3] is used in the present effort because of its role as the substrate for recent modeling work in sentence comprehension [6] and intention-based reasoning [7], which are both needed to model linguistic communication as a collaborative activity. Head-driven phrase structure grammar (HPSG) [8] is also used in the

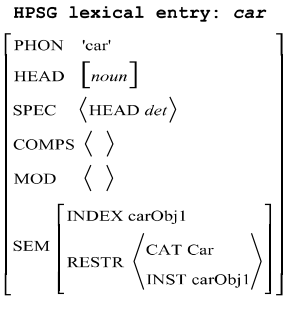
modeling work because of its lexical integration of syntax and semantic constraints and the computational advantages of its framework.

4 Modeling Expectations in Language Use

The models described in this section are conceived as a set of interrelated modules. Reasoning about the user's sentence input is performed at different levels in stages that roughly correspond to the heuristics outlined above. Although the presumption of common ground is addressed in grossly simplified terms, the salience and solvability heuristics are applied, respectively, to the utterance level of the user's sentence and, in two stages, to its linguistic and practical implications.

4.1 Common Ground

Because of the complexity in modeling common ground, we proceed on two simplifying assumptions, leaving certain issues for future investigation. First, the application interacts with only one user, and second, we only model the listener's interpretation, and do not address issues concerning a speaker's generation of language. In this simplified model, knowledge is limited to a subset of actions and objects in a vehicular traffic domain. HPSG feature structures of words are encoded in Polyscheme (Fig. 1).




```
<constraint>
IsA(?word,WordUtteranceEvent, E, ?w) ^
Phonology(?word, car, E, ?w)
      ==>
Head(?word, Noun, E, ?w) ^
Specifier(?word, ?x, E, ?w) ^
Head(?x, Determiner, E, ?w) ^
-CompleteSpecifier(?word, E, ?w) ^
CompleteComplement(?word, E, ?w) ^
CompleteModifier(?word, E, ?w) ^
LinguisticReferent(?word, ?carObject, E, ?w) ^
IsA(?carObject, Car, E, ?w)
</constraint>
```

Fig. 1. The HPSG feature structure of the word "car" is shown on the left. On the right is its Polyscheme constraint in XML. Propositions on the right of the arrow ==> (consequents) are inferred from propositions on the left (antecedents). Note, "?" used as prefix denotes a variable.

4.2 Salience

Clark's principle of salience [1] suggests, roughly, that the ideal solution to a coordination problem is one that is most prominent between the agents with respect to their common ground. Thus, for example, when the user enters "…the red car…," it is expected that these words are intended to make objects tied to this phrase more prominent than other objects in the knowledge and experiences the user shares with the addressee. This effect is achieved with a reasoning "specialist" in Polyscheme.

4.3 Solvability

Stage 1—Natural language understanding. Although an addressee's syntactic, semantic, and pragmatic processing may overlap in real-world collaborations, these levels are currently staged in separate models. Basic syntax and semantics are thus handled first, producing a set of "atoms" such as those shown in Fig. 2.

```
IsA(car1, Car)
Color(car1, red)
IsA(car2, Car)
Color(car2, black)
IsA(passingEvent280, PassingEvent)
Agent(passingEvent280, car1)
Theme(passingEvent280, car2)
IsA(personListener242, Person)
ListenerOfUtterance(personListener242)
IsA(mePerson1, Person)
SpeakerOfUtterance(mePerson1)
IsA(displayEvent239,
DisplayEvent)
Agent(displayEvent239,
personListener242)
Beneficiary(displayEvent239,
mePerson2)
Theme(displayEvent239,
passingEvent280)
Mode(displayEvent239,Directive)
```

Fig. 2. Semantic output (simplified) after parsing "Show me the red car passing the black car"

This output is then processed by a second model in this stage that applies Polyscheme "constraints" representing common sense and domain knowledge to achieve a higher-level understanding of the user's sentence. The example in Fig. 3 shows how this model reasons about the higher-level implications of the atoms in Fig. 2.

DOMAIN KNOWLEDGE

```
<constraint>
IsA(?p,PassingEvent,
E,?w) ^
Agent(?p,?pOb,E,?w)
  (.9)>
InMotion(?pOb,?t,?w)
</constraint>
```

COMMONSENSE KNOWLEDGE

```
<constraint>
IsA(?u,Person,E,?w)
^ListenerOfUtterance(
?u,E,?w)
  ==>
LiteralReferenceOfMeta-
phor(?u,
IntertrackApplica-
tion,E,?w)
</constraint>
```

INFERRED OUTPUT

```
Direction(car1,dir1)
Direction(car2,dir1)
InMotion(car1)
InMotion(car2)
IsA(user1OfIntertrack,
Person)
SpeakerOfUtterance(
user1OfIntertrack)
LiteralReferenceOfMeta-
phor(listener242,
IntertrackApplication)
```

Fig. 3. Examples of domain and commonsense knowledge constraints: (left) the agent of a passing event is in motion with a probability of 0.9; (middle) the application is the addressee (listener). The atoms on the right are inferred based on the semantics from Fig. 2.

Stage 2—Task recognition. When one agent's intentions must be grasped and acted upon by another, addressees presume the speaker has a practical task outcome in mind they can recognize and help achieve. Hence, in this stage, the model marks a recognized directive or a question as an expected task (Fig. 4 left), and, barring any problems, executes the appropriate application command (Fig. 4 right).

```
<constraint>
Mode(?entity, Directive, E, ?w) ^
Agent(?entity, ?entsAgent, E, ?w)
^ LiteralReferenceOfMetaphor(
?entsAgent,IntertrackApplication,
E, ?w)
  ==>
ExpectedTask(?entity, E, ?w)
</constraint>
```

```
<constraint>
ExpectedTask(?entity, E, ?w) ^
IsA(?entity, DisplayEvent, E, ?w)^
Ability(IntertrackApplication,
DisplayEvent, E, ?w) ^
Theme(?entity,  sceneToBeDisplayed,
E, ?w)
  ==>
Perform-
Task(MatchCardWithAnswerWorld,
?sceneToBeDisplayed, E, R)
</constraint>
```

Fig. 4. Example of task identification (left) and application processing capability (right)

5 Conclusions and Future Directions

Modeling coordinated activity between agents has the potential to offer more flexibility to users than other types of interfaces, e.g., menu-driven systems. The models outlined here focus on an addressee's heuristic expectations of a speaker's use of common ground, salience, and solvability in the coordination of meaning and understanding. Related heuristics that remain to be modeled are those of sufficiency and immediacy [1]. An advantage of modeling these stages of an addressee's processing is the ability to identify the precise nature of the problem when coordination failures requiring repairs arise. In future work, models of repair will be pursued by allowing the user to correct or clarify various types of misunderstandings such as the use of unfamiliar words or the inability to perform a particular task, etc.

References

1. Clark, H.H.: Using Language. Cambridge University Press, Cambridge (1996)
2. Lewis, D.K.: Convention. Harvard University Press, Cambridge, MA (1969)
3. Cassimatis, N.L.: A Cognitive Substrate for Achieving Human-Level Intelligence. AI Magazine 27(2), 45–56 (2006)
4. Pless, R., Jacobs, N., Dixon, M., Hartley, R., Baker, P., Brock, D., Cassimatis, N., Perzanowski, D.: Persistence and Tracking: Putting Vehicles and Trajectories in Context. In: 38th IEEE Applied Imagery Pattern Recognition Workshop, Washington, DC (2009)
5. Allen, J., Perrault, R.: Analyzing Intentions in Utterances. Artificial Intelligence 15(3), 143–178 (1980)
6. Murugesan, A., Cassimatis, N.L.: A Model of Syntactic Parsing Based on Domain-General Cognitive Mechanisms. In: Proceedings of the 28th Annual Conference of Cognitive Science Society, Vancouver, pp. 1850–1855 (2006)
7. Bello, P., Cassimatis, N.L.: Understanding Other Minds: A Cognitive Modeling Approach. In: Proceedings of the 7th International Conference on Cognitive Modeling, Trieste (2006)
8. Sag, I.A., Wasow, T., Bender, E.: Syntactic Theory: A Formal Introduction, 2nd edn. University of Chicago Press, Chicago (2003)