A Qualitative Model for Natural Language Communication about Vehicle Traffic

H. Joe Steinhauer

Department of Computer Science, Linköping University
S-58-183 Linköping, Sweden
joest@ida.liu.se

Abstract

In this paper we describe a qualitative approach for natural language communication about vehicle traffic. It is an intuitive and simple model that can be used as the basis for defining more detailed position descriptions and transitions. It can also function as a framework for relating different aggregation levels. We apply a diagrammatic abstraction of traffic that mirrors the different possible interpretations of it and with this the different mental abstractions that humans might make. The abstractions are kept in parallel and according to the communicative context it will be switched to the corresponding interpretation.

Motivation

Natural language communication is getting more and more important for artificially intelligent systems. Imagine for example an autonomous helicopter (e.g. the one in the WITAS project (Doherty et al. 2000)) patrolling over the rush-hour traffic and reporting the ongoing to the police headquarters. Further you can think of a driver support system in your car that not only gives you advice where to drive but as well interprets the traffic around you so that it can warn you of dangerous situations. It even can help you to avoid those by predicting the possible alternatives of maneuvers for the other traffic participants.
What we would like to achieve here is that the system talks to us in natural language, using our vocabulary that we are used to when we talk about traffic. The helicopter should say for example: ”The red Porsche is driving along Main Street, now passing the church and will soon turn right into Park Avenue.” The driver support system you might like to express something like: “Be careful, you should not overtake here, the car in front might turn left soon, its winker is probably broken.” To be able to speak like this the system does not only have to have the right interpretations of the information that it gets form its sensors, it must also be able to express them in the way that it states exactly the information that is important to us.

> Natural-language and multimedia dialogue with an autonomous robot is a challenging research problem which introduces several important issues that are not present in, for example, dialogue with a database or a service provider such as an automated travel agency (Sandewall et al., 2003, page 55).

We aim to develop a qualitative approach whereby traffic maneuvers can be represented in an artificially intelligent system. Natural language itself helps to find out what the important moments that classify a traffic maneuver might be. If we describe a maneuver to somebody else we probably mention just the parts that are important to us.
This leads to an approach where we use a very intuitive and simple model to describe relative positions of traffic participants to each other. These relative positions and the changes between them are sufficient to describe traffic maneuvers. Within the model objects can be grouped in a way that an analysis of the maneuvers on different levels of abstraction is possible. These different abstraction levels correspond to possible different interpretations and thereby to different mental abstraction levels of one situation. You may for example talk about a queue, a convoy or a crowd; each of these composite objects contains several individuals from which you abstract. For a queue you may further talk about its tail and its head and refer to a position as in the middle of the queue. All those parts are more or less abstract unless you are forced to describe them in detail and you really have to think about where the tail of the queue starts or which specific vehicle really is in the middle of the queue.

Relation to the Symposium Topic

The project is closely related to the symposium topic as it first converts external diagrams into mental diagrams, which will then be transformed according to different purposes into different mental abstractions that satisfy the communicative context of the situation. Changes in the communication lead to adjustment of the abstractions.
Regard for example the autonomous helicopter that reports the ongoing traffic. It has a camera attached and knows, after all image processing and object recognition, what objects are in the part of the world that it is observing and where they are. For the communication purpose this data is abstracted into a two-dimensional diagram that mirrors a possible mental structure of a human. This abstraction

allows further possible abstractions that humans might use while interpreting traffic. These correspond to the different human views of the traffic situation. Further the user of the system gives instructions in natural language. These will often contain references to his or her own mental abstraction. The system has to pick the right interpretation to be able to follow the communication and to fulfill the instructions. Actions of the observing system can lead to a different view of the situation; either because the system is active taking part in the traffic or due to changes in the observation focus. The diagram in figure 1 gives an overview about the different interpretation levels of the traffic situation and how they are connected within our approach.

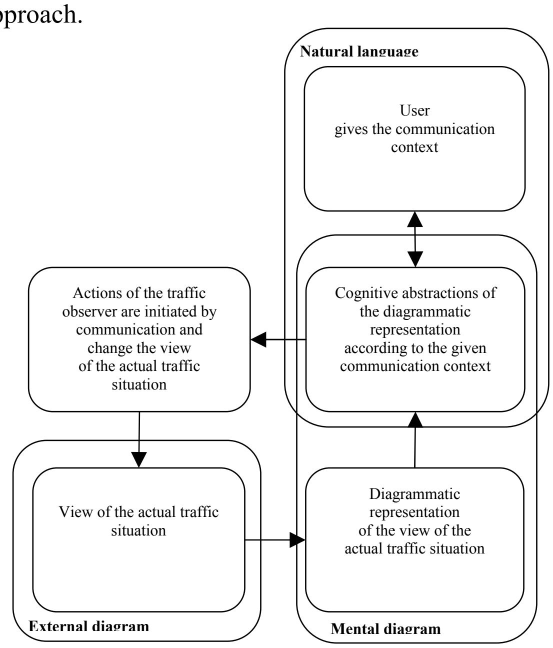

Fig. 1. Diagram of the different interpretations of the actual traffic situations that takes part in different parts of the system.

The Directed Moving Object Model

When we describe the position or the maneuver of a vehicle we need a reference object and we give the vehicle's relative position or its actions in relation to that reference object. For example: "The red car is driving behind the truck", where the truck is the reference object. "It is driving in the street", where the street is the reference object. " It is driving through the tunnel and across the bridge" where the tunnel respectively the bridge is the reference object.

Our approach to describe traffic maneuvers uses a very simple and intuitive model that can be applied as the basis for defining more detailed position descriptions and transitions. It can also be used as a framework for relating different aggregation levels. In the model, shown in figure 2, the plane around a two-dimensional directed object (the reference object) is divided into nine qualitative two-dimensional regions including the area where the reference object itself is situated. The resulting regions are named from the intrinsic perspective of the reference object. The names are the same as in Freksa's double cross calculus for oriented objects (Freksa, 1992). Freksa uses a neighborhood-oriented representation to reason about spatial direction information for one-dimensional objects.

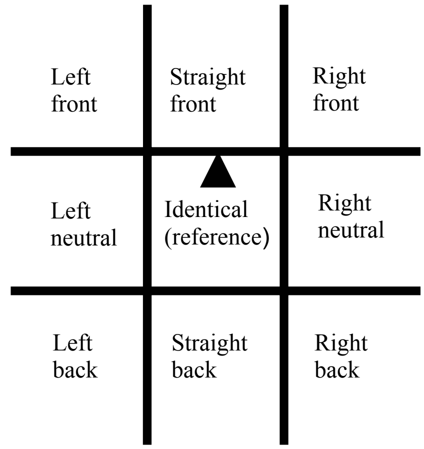

Fig. 2. The directed moving object model gives the relative positions of an object to a reference object.

Several other approaches exist that use this intuitive division of two-dimensional space for several interpretation purposes. Mukerjee and Joe (1990) use a two-dimensional model for calculating position relations of objects. The direction relation matrix is used by Goyal and Egenhofer (2000) to calculate distances in similarity between spatial scenes. One approach for tracking traffic, from a video input taken by a stationary camera, has been done by Fernyhough et al. (2000). This approach even includes learning traffic maneuvers from the observed information.

If the reference object has been chosen we can state which relative positions the vehicle-of-interest has according to this reference object. We call the vehicle that we are interested in the vehicle in focus, which often will be abbreviated to vif in the following text. If for example the vif is in the straight front region it would be appropriate to say: "The vehicle in focus is driving in front of the reference object." As long as the vif is within the same qualitative region there is no need to give further information about its position or movement. When it changes from one qualitative area to another an important change has occurred that may have to be stated.

Thus there are 9 relative positions and 24 transitions between those positions (if we assume a four-neighborhood of the areas) that we need to be able to express. These all can be organized in the conceptual neighborhood graph shown in figure 3.

“It is driving left behind the black car”, “It is catching up to the black car, driving beside it, passing it, driving left in front of it, is sheering in in front of it ” and finally “The white car is driving in front of the black car” are generated. This is a maneuver description where every change of relative position was interpreted and mentioned. In a more concise communication the whole chain of elementary actions can be grouped together and stated as one overtaking maneuver: “The white car is overtaking the black one.” In the same way several overtaking maneuvers in a row can be group to an overtaking of a queue.

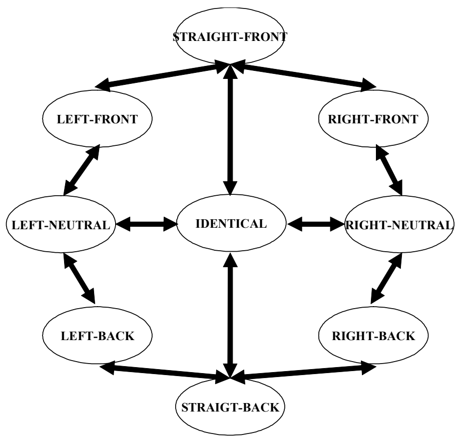

Fig. 3. The conceptual neighborhood graph

For the traffic domain where movement is normally continuous it is sufficient to only concentrate on continuous processes.

That means that all movement can be described along the transitions of the conceptual neighborhood graph. A real maneuver can then be seen as an actual path through that graph. A couple of maneuver descriptions in the conceptual neighborhood graph are shown in figure 4.

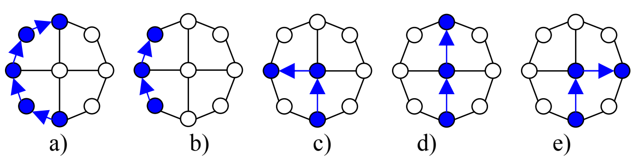

Fig. 4. Traffic maneuvers of one vehicle described as states and transitions in the conceptual neighborhood graph. a) overtaking, b) passing, which is as well a part of the overtaking maneuver, c) turning left, d) crossing, e) turning right.

Example: Description of an Overtaking Maneuver

The input in the system can be a sequence of snapshots of the scene of interest wherein all objects have been identified. It can also be a sequence of messages about state transitions. As an example we give the description of an overtaking maneuver grounded on the snapshots shown in figure 5. We concentrate on the white car, which becomes the vehicle in focus and we are going to describe its actions. As reference object we choose the black car in front of it and establish the qualitative regions around it. While formulating the natural language expressions we will take into account what kind of objects participate in the maneuver and adjust the terminology in the way humans would do. That means that for the first snapshot we would say: “The white car is driving behind the black car” instead of: “It is driving in the straight back region of the black car.” For the following snapshots the statements “The white car is sheering out from behind the black car”,

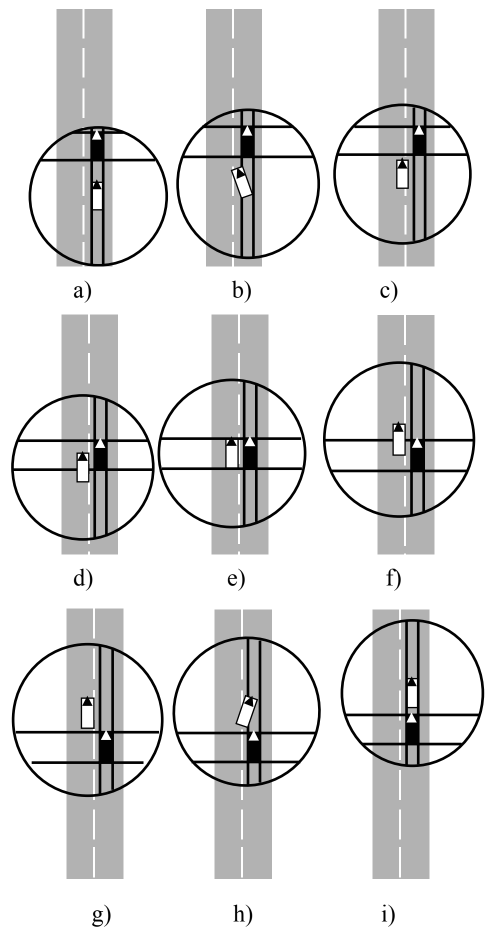

Fig. 5. Snapshots of the overtaking maneuver. The black car is the reference object; the white car is the vehicle in focus which actions are described according to its position relative to the reference object. a) driving behind, b) sheering out from behind, c) driving left behind, d) catching up, e) driving beside, f) passing, g) driving left in front, h) sheering in in front, and i) driving in front.

Parallel Use of Different Qualitative Interpretation Levels of Traffic Maneuvers

What you wish to describe from the traffic situation you observe depends on the context of the communication. According to that you interpret the ongoing scene differently. In a line of cars like the one in figure 6 you may regard each car for itself or you may like to group them all together to one queue, like it is done in figure 7, which you can refer to as one object. In the latter case you mentally abstract from details like the number of cars and their colors. Depending on how your conversation develops it may be necessary to switch between different interpretations, stating some actions of the vif with respect to the whole queue and others just to parts of it or to a single vehicle.

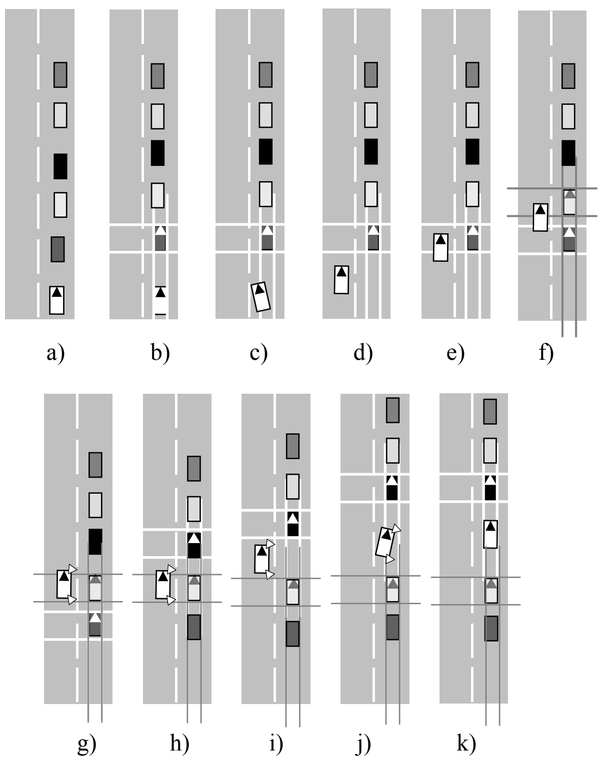

Fig. 6. A possible interpretation of the situation where the vehicle in focus is approaching a couple of vehicles from behind. Those vehicles are all traveling on the same lane and in the same direction as the vehicle in focus. This interpretation keeps the vehicles as single objects, some of them are chosen to function as reference objects in the appropriate situations. The vif is sheering out and overtaking two of the vehicles before it is sheering in into the right lane between the vehicle that it has just overtaken and the one in front of it.

In our approach it is possible to establish as many different interpretations of the situation as desired and keep them all in parallel. For each statement one of the interpretations can be chosen. Look for example at the snapshots in figure 6. In snapshot a) you see that the vif is approaching a couple of other vehicles, each of them can be used as a reference object. “The vif is driving behind the black vehicle,” is a correct conclusion of the situation even though a human probably wouldn’t use this description without at least mentioning that there are two other vehicles in between.

A more intuitive way is to concentrate on the most important one of those objects, which often happens to be the closest object to the vif, and take that one as a reference. The snapshots b) to e) have the last dark grey vehicle of the queue as reference object. The vif’s actions can be described as b) driving behind, c) sheering out from behind, d) driving left behind, and e) catching up to the dark grey vehicle. In snapshot f) the dark grey vehicle is not the only one that is closest to the vif. Here we could change the reference object to the light grey vehicle instead and say that the vif is now catching up to that one, as well as we can say at the same time that it is passing the dark grey vehicle. Both alternatives are absolutely correct and mentioning them both would give an even better overview of the situation. The same applies for the next couple of snapshots g) to k) where the vif first indicates, by setting the winker, that it wants to sheer in into the right lane between the light grey and the black vehicle and then is fulfilling this wish. The most appropriate way of describing the situation in snapshot j) would be to say that the vif is sheering in in front of the light grey vehicle and behind the black one.

The illustration in figure 6 is not the only way of interpreting the situation. When you look at snapshot a) again you may think of all the vehicles in front of the vif together as queue of vehicles. In this case you would like to refer to the queue as one composite object. The snapshots in figure 7 visualize this interpretation. Now the vif is a) driving behind the queue, b) sheering out from behind the queue, c) driving left beside the queue, d) catching up to the queue, and e) driving beside the queue. It doesn’t matter which vehicles of the queue it passes or it is catching up to. These details can easily be ignored as long as they are not important for what you want to say. In snapshot g) the vif is sheering in into the queue and in h) driving inside the queue.

Nevertheless there is no need to stick to only one interpretation once it has been established. Your interpretation changes whenever your focus of the situation changes. Figure 8 gives a third possible interpretation of the same maneuver. Here the interpretation of one queue is given up when the vif indicates the wish to sheer back in into the right lane. The queue is split into two parts, one part where the vif will be driving behind and one part where the vif will be driving in front of, after the maneuver is finished.

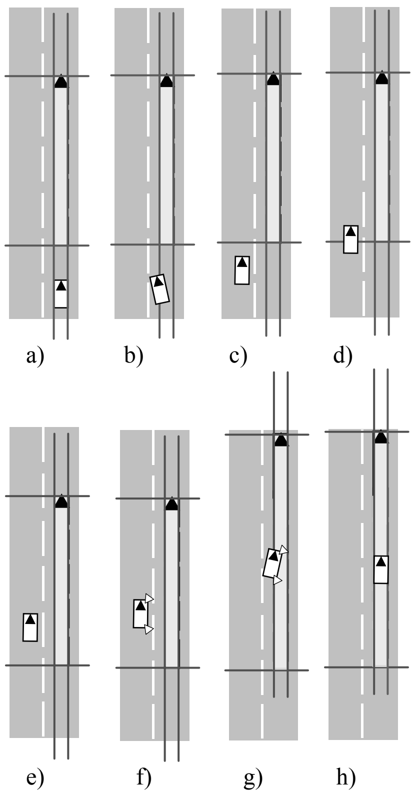

Fig. 7. The vehicles that are shown in figure 6 are now interpreted as one queue object. This queue becomes the reference object and the maneuvers of the vif are described in relation to the queue. a) approaching the queue, b) sheering out from behind the queue, c) driving left behind the queue, d) catching up to the queue, e) driving beside the queue, f) driving beside the queue and indicating the wish to sheer in into the queue by setting the winker, g) sheering in into the queue, and h) driving inside the queue.

The different interpretations can be used simultaneously. “The vif is driving between the dark grey and the light grey vehicle inside the queue where the part of the queue in front of it is as long as the part behind it.” In this description three different interpretations are used. 1st the interpretation of the single vehicles that are driving in front of and behind the vif without further details about the rest of the surroundings. 2nd the interpretation of the whole queue where the vif is driving inside and 3rd the interpretation that there are two parts of the queue which are separated from the vif, one in front of it and one behind it. Of course you can come up with a couple of other possible interpretations for this maneuver, all of them will use single objects or groups of them as reference objects. In our model it is possible to visualize whatever interpretation you may have.

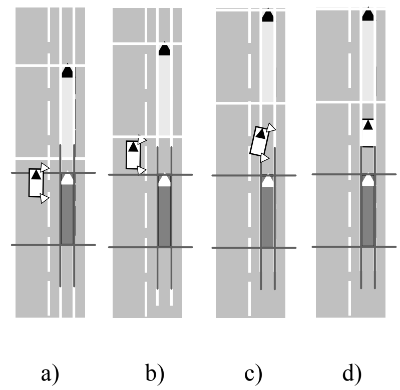

Fig. 8. After the vif has indicated the wish to sheer in into the queue the interpretation of a single queue as reference object is given up. Instead two parts of the queue are given: one that will be in front of the vif and the other that will be behind the vif when the maneuver is finished.

Summary and Future Work

We have introduced a qualitative approach for modeling traffic situations and traffic maneuvers in a diagrammatic way that mirrors human’s different mental interpretations and interpretation levels of traffic in order to communicate about it. Each single object can be taken as a reference object and several objects can be grouped into composite reference objects. The detailed information about the single objects within the composite objects is not lost; it is just not used when not necessary. Switching back and forth between different interpretations and interpretation levels is used to adjust the interpretation to the context of the communication.

The modeling system described here will be used as the basis for a modeling language. This modeling language must also capture the continuous processes as we see and interpret them in the form of states that have to be frequently updated. Furthermore we plan to collect a number of different traffic maneuvers for single vehicles as well as for composite objects. This includes maneuvers that are described in relation to one, two or even more reference objects, like “driving between”. Several reference objects are necessary as well to clarify some ambiguous situations that occur when you only regard the relative positions of the vehicles in snapshots that do not give you any information about which of the vehicles has changed its absolute position. When ongoing traffic can be described in natural language the way humans would describe it, the next step will be to further develop the system so that it can predict, or at least restrict the possible outcomes of a current situation.

References

Doherty, P., Granlund, G., Kuchcinski, K. Sandewall, E. Nordberg, K., Skarman, E and Wiklund, J (2000).The WITAS Unmanned Aerial Vehicle Project. In Proceedings of the 14th European Conference on Artificial Intelligence. IOS Press Amsterdam. 747-755.

Fernyhough, J., Cohn, A. G., and Hogg D. (2000). Constructing qualitative event models automatically from video input. In Image and Vision Computing, 18. 81-103.

Forbus, K. D. (1997). Qualitative Reasoning. In Allen B. Tucker (Ed.). The Computer Science and Engineering Handbook. CRC Press. 715-733.

Freksa, C. (1992). Using orientation information for qualitative spatial reasoning. In A. U. Frank, I. Campari, U. Formentini (Eds.). Theories and methods of spatio-temporal reasoning in geographic space. Springer, Berlin. 162-178.

Goyal, R., K Egenhofer M., J. (2001). Similarity of Cardinal Directions. In C. S. Jensen et al. (Eds.). Advances in Spatial and Temporal Databases. SSTD 2001, LNCS 2121. Berlin Heidelberg: Springer-Verlag. 36-55.

Mukerjee, A. and Joe, G. (1990). A qualitative model for space. In Proceedings of the AAAI-90. The MIT Press Boston. 721-727.

Sandewall, E., Doherty,P., Lemon, O., and Peters, S. (2003). Words at the Right Time: Real-Time Dialogues with the WITAS Unmanned Aerial Vehicle. In: Andreas Guenter et al (eds). KI 2003: Proceedings of the 26th Annual German Conference on AI. Springer.

WITAS homepage   http://www.ida.liu.se/ext/witas/