# Incremental Generation of Multimodal Route Instructions

**Christopher Habel**

Department for Informatics (AB WSV), University of Hamburg
Vogt-Kölln-Str. 30, D-22527 Hamburg
habel@informatik.uni-hamburg.de

**Abstract**

Human produced route instructions are usually conveyed either verbally in spoken discourse or by written texts, or by graphical means, i.e. by illustrating the route in a map or by drawing sketch-maps, or by combining both kinds of external representations. Whereas verbal route instructions focus on the actions to be performed and take the spatial environment only as the frame for these actions, maps and other pictorial representations foreground the spatial environment without possessing adequate means for representing the actions.

Today, in the time of World Wide Web and Geographical Information Systems, way-finding queries can be given to systems, which provide 'driving directions' containing canned text as well as different types of maps. In this paper I describe the principles and the architecture underlying a system for generating multimodal route instructions combining natural language route descriptions and visualizations of the route to follow, such that the strengths of both means for communication route knowledge are brought together.

## Introduction

When humans have to solve the problem 'How to get from A to B' in an unknown environment, there are several ways to solve this spatial task by using *external representations*. On the one hand, there are prefabricated multipurpose means for spatial problem solving as maps: A city map is an external representation to help the user in finding a way from an origin X to a destination Y, where X and Y span up a variety of potential way finding problems. Even more specialized sketch maps like those designed for finding the way to a specific shopping mall or a chosen hotel are not entirely determined on an individual way finding process: While they are fixed with respect to the destination, they usually make this destination accessible from a (limited) number of origins. On the other hand, a successful strategy to get information about the individual problem 'finding the way from A to B' is to ask for a route instruction. In most cases such query leads to verbal descriptions of the route to follow—in face-to-face communication often accompanied by gestures, which give supplementary information. Another important way to transfer additional information is the production of a sketch map.

Today, in the time of World Wide Web and Geographical Information Systems, way-finding queries can be given to systems, for example to MapQuest (MapQuest.com, Inc.), which provide 'driving directions' containing canned text as well as different types of maps, as overview maps or turn-by-turn maps.

In this paper I describe the principles and the architecture underlying a system for generating multimodal route instructions combining natural language route descriptions and visualizations of the route to follow.[1] The aim of the architecture is to allow the generation of instructions of different types and on different levels of explicitness. This flexibility is necessary to provide the human user with adequate information in an instruction dialogue.

The organization of the paper is as follows: first, I give a short overview on the task of way-finding and discuss the two fundamental types of route instructions, namely verbal descriptions and map-like depictions, as well as their combinations in multimodal instructions. Second, I present , the principles of INC, an incremental conceptualizer, a module that roughly corresponds to the 'what-to-say' task of NLG systems, which is the basis for the multimodal route instruction system, presented in the final section.

## Way-finding and Route Instructions

In this paper, I focus on route instructions given with the communicative goal to assist the addressee in way-finding. [2] In principle, there are two types of route instructions, on the one hand, *in advance* route instructions, which are given before the way-finding task starts, and on the other hand, *on line* route instructions, which assist the addressee in navigating.

---

[1] Currently, the system is partially realized only. Some modules have been developed as part of different research projects. Other modules are currently under construction.

[2] In the following I use 'way-finding' on the strategic level and 'navigation' on the tactic level of solving the task to *go from A to B*. As the articles published in Golledge (1999), which distinguish between 'way-finding' and 'navigation' dependent on the specific focus taken by the respective authors, demonstrate, there is no consensus about these notions in the scientific community.

A property characteristic for *in advance* route instructions is, that neither the *instructor* nor the *instructee* perceives the relevant environment, the critical landmarks, or the tracks,[3] e.g., the roads, completely and directly. Cases of such *in-advance route instructions* are very common (cf. Tschander et al. 2002): they are given in face-to-face communication on the street or in dialogues at the telephone. Furthermore, *computer generated driving directions* provided in the WWW are also instructions of this type. During the *instruction phase* an instructor, who possesses knowledge about the environment in question, produces a route instruction. In comprehending the instruction, the instructee builds up conceptual, mental representations of the route. These representations, which contain spatial information about the route and the sequence of actions to be performed, have to be stored in memory. Later, in the *navigation phase*, the instructed navigator has to match the internal representations against the perceived scenes. This process involves the recognition of spatial configurations of landmarks, tracks, and positions in accordance with the spatial relations specified in the instruction.

In contrast, *on line* route instructions, for example those produced by navigation systems installed in cars, focus mostly on small-scale environment, which is perceivable by the addressee. The actual action that is produced to instruct the navigator has mostly to be performed in the currently perceived part of the environment.

With respect to way finding and navigation the two types of route instruction, *in advance* and *on line*, have different advantages and disadvantages, which I exemplify with the case of way finding and navigation in *driving from A to B*. In advance instructions assist you to build up a large-scale spatial mental model on the level of survey knowledge. But such instructions mostly give you only a limited amount of detailed information about the actions to be performed during navigation before you start; in other words, they provide primarily coarse-grained route knowledge. On the other hand, on line instructions, e.g. those given by a navigation system, help you to solve current navigation problems during driving, i.e. they provide pieces of route knowledge.[4] But, car navigation systems posses—what automotive engineers as P. Green (1997) call—'crash-inducing potential', that is related to 'eyes-off-the-road' and 'mind-off-the-road' phenomena (Cf. Green, 2000). From the perspective of spatial cognition, survey knowledge as well as coarse-grained route knowledge has the promise to reduce the risk potential, since the can decrease eyes-off-the-road as well as mind-off-the-road distractions. [It is this deficit of automated systems for route instructions, which is intended to overcome by the system proposed in the present paper.]

Current computer generated driving directions, as MapQuest, which belong to the class of advance route instructions, are monological. If the instruction is given in a dialogue then instructed people can immediately try to solve ambiguities or other types of difficulties in comprehending, i.e. building up a spatial model, by querying the instructor, Especially, they have the possibility to react on informational units that they fear to be unsatisfactory or inadequate in later navigation.

## Requirements on Route Instructions

The overall criterion for the adequacy of a route instruction is whether it enables navigators to find their way. Thus, adequacy depends on a wide spectrum of parameters. For example, epistemological parameters, such as the knowledge of the participants (the instructor and the instructee), or perceptual parameters, which concern the navigator's perception of the environment and the perceptual salience of landmarks, can influence the performance of the navigator. Since not all objects or spatial configurations that the navigator will perceive on the route can be specified in the instruction, the type and amount of information, e.g. concerning landmarks, provided by the instructor is crucial for successful route instructions (Fraczak, Lapalme and Zock 1998, Tversky & Lee 1999).

Generating good route instructions implies to give adequate information about

- those *actions*, i.e., locomotion acts to be performed by the instructee, that get the instructee to the intended goal.
- the *spatial environment* in which the intended locomotion of the instructee will take place.

The instructor's primary task to be solved is choosing a good combination of communicational means to transfer the relevant information to the instructee. Combining different means is advantageous, since each of them seems to be

[3] 'Track' is used here as a generalizing term, which subsumes roads, hiking paths, sidewalks, etc., i.e., distinguished real world entities, which are used for moving on. I propose to distinguish two roles of tracks: on the one hand, they are guiding structures for navigation, namely they provide the way we have to travel on, on the other hand, they can function as landmarks to find the way we have to go (cf. Tschander et al. 2002). This contrasts to some authors, who count tracks—due to their function in identifying a position on the route—among the landmarks (for example, Denis 1997, or Lovelace, Hegarty & Montello 1999).

[4] Montello (1998) discusses different types and levels of spatial knowledge, as route knowledge and survey knowledge, as well as a framework for describing the acquisition of spatial knowledge.

optimal only with respect to some aspects, which I exemplify in the following:[5]

- *declarative vs. imperative mood (in verbal instructions)*
  In describing the actions to be performed, imperatives, as “Turn left”, are standard, but declaratives are also used, e.g. “On Route 30 you will cross Route 320, …”. Furthermore, declaratives are used to describe the goal condition, e.g. “The gym is on the right at the end of the parking area”.
  Using the declarative mood characteristic for survey knowledge level descriptions is primarily used in human in advance instructions, whereas the imperative mood is more frequent in online instructions.
- *prescribing the action vs. describing the environment*
  Imperatives and descriptions of future actions are the means used in *verbal instructions* to foreground the actions to be performed. In contrast, the description of the environment, for example mentioning of landmarks, can be seen as providing the background. To build up survey knowledge in advance, the descriptive parts of an utterance are more important than the instructive parts.
  On the other hand, maps and other static pictorial representations place the spatial environment in the foreground and keep the action in the background. Figure 1 exemplifies how the action to be performed—verbally expressed by two instructions (left column)—can graphically be represented by depicting a route.

Turn RIGHT onto
LAGUNA ST.
Turn LEFT onto
MARINA BLVD.

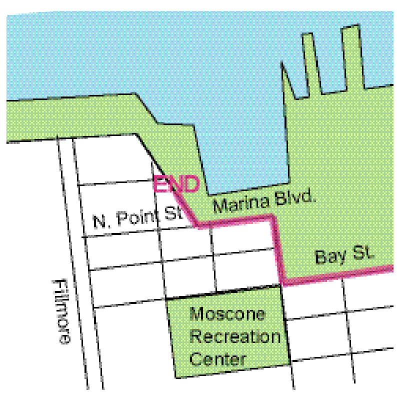


Figure 1: Prescribing locomotion vs. depicting the route to be followed[6]

- *static vs. dynamic aspects of the way-finding task*
  Way-finding and way-following can be characterized as *sequences* of actions, i.e., as actions in a *temporal order*. Natural language possesses explicit means for describing time and event ordering, e.g. the use of *temporal connectives,* as in “After passing through the Fort McHenry Tunnel (Baltimore), take exit 52”. Furthermore, the linear order of language—explicitly temporal in speech, and implicitly temporal in written language—induces temporal ordering of the events or actions mentioned in a text. Thus, a sequence of instructions, as in Figure 1, will mostly be interpreted as temporally ordered (cf. ‘the order of mention contract’, Clark and Clark, 1977).
  The spatial counterparts to the sequences of actions discussed above are routes to be followed in way-finding, or focusing on the action to be performed, the trajectories of the intended, future motion event. Thus, from a formal point of view, the representing entities in the map (or other pictorial representations) should be *oriented curves*, i.e. curves that possess a starting point and an end point (Cf. Eschenbach, Habel and Kulik 1999). How to depict oriented curves in an adequate manner is controversial. There are two standard solutions for this task—both non-optimal—first, the use of arrows, inducing orientation, second, ordering labels, as used in MapQuests turn-to-turn maps (see figure 2.).

13. Turn RIGHT onto
    LAGUNA ST.
14. Turn LEFT onto
    MARINA BLVD.

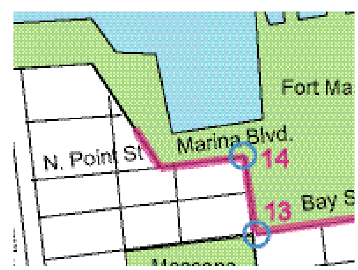


Figure 2: Representing ordering of events: Language vs. Depiction

The route, as trajectory of a motion event, plays a core role in the production of route instruction. As Klein (1982) argues, performing a ‘*virtual journey*’ is the instructor’s primary organizing principle for determining the content to be communicated to the instructee. Thus, I propose the *virtual journey*, formally constituted by combining a route with the route-following actions, as the informational input of the language generation system described in the next section.

Let us consider an intermediate result: The pros and cons of verbal descriptions and maps consider different aspects of information to be conveyed in route instructions. Whereas the dynamic aspects of the route following are adequately reflected on in verbal instructions, pictorial representations are advantageous in communicating global as well as specific types of local spatial knowledge about the environment. Thus

---

[5] All linguistic examples are collected from the WWW.

[6] The verbal descriptions of figure 1 and 2 are generated by MapQuest.com. The depictions are designed by the author following the layout principles of the sketch map generated by MapQuest.com.

combinations of maps and texts are—in principle—a promising way of multimodal route instructions. On the other hand, the current state of the art is far from being satisfactory (cf. the clumsy combination exemplified with figure 2.) To overcome the disadvantages discussed above, I propose in the final section of the present paper the use of '*dynamic maps*'.[7] In contrast to the multimodal instructions discussed above, the combination of spoken instructions and dynamic maps seems to enable an elegant and easy to comprehend connection between verbally and pictorially communicated content.

## Incremental Generation of Text

Reiter and Dale (2000) characterize the first phase in NL generation as the solving of two tasks, *content determination* and *document structuring*. The system has to decide what information has to be communicated and how this content to be communicated is organized. From a psycholinguistic perspective on language production dealing these tasks is located in the *conceptualizer*, the first module of Levelt's (1989) cognitive architecture of the human language production system.

Systems that generate natural language descriptions of what happens in a dynamically changing world can be improved substantially by working incrementally. Incrementality enhances the overall quality of the systems for—at least—four reasons: (1) The dynamic nature of a sequential stream of input information can be handled more directly and, therefore, easier. (2) Incremental systems are capable of producing fluent speech, i.e. speech without artificial auditory gaps. (3) Parallelism that comes with incrementality makes better use of the available resources. (4) In incremental architectures are extremely suitable for multimodal systems since they support fusion and fision between the modalities.[8]

As preparation of section 3, I give an overview of the Incremental Conceptualizer INC, which is the core of a multimodal language/map generation system currently under development.

### INC – An Incremental Conceptualizer

The primary research topic of the ConcEv project is the *what-to-say* component, in which the content of utterances is planned (cf. *Document planning* in Reiter & Dale 2000). We use the terminology of Levelt (1989), who calls the first component the *conceptualizer*, the second the *formulator*. These modules interact via *preverbal messages*, which are propositional, non-verbal representations of the utterance built up by the conceptualizer, which is the topic of the present section.

There are some incremental models for that part of the language production faculty that does the linguistic encoding, i.e. the modules that solve the *how to say* problem, cf. Kempen & Hoenkamp (1987), De Smedt, Horacek & Zock (1996). INC (*incremental conceptualizer*) models this first stage of the language production process from a stream of perceptual or conceptual input. Up to now, INC has been developed for and tested on producing preverbal messages for the on-line description of events (cf. Guhe, Habel and Tappe 2000, Guhe and Habel 2001, Guhe, Habel and Tschander 2003).

The incremental conceptualizer INC, that corresponds to the first component of Levelt's (1989) language production model, is located between the pre-processing units that provide the basic conceptual representations and the formulator in which the linguistic encoding is done. It consists of four main processes:

1. *construction* reads the output of the pre-processing unit and builds up internal conceptual representations of the current state of affairs, the so-called current conceptual representation, or CCR for short.[9]
2. *selection* selects situations for verbalization from the CCR.
3. *linearization* brings the selected situations into an appropriate order.
4. *PVM-generation* incrementally generates preverbal messages for each selected situation and sends them to the formulator.

These four processes work incrementally. That means, (1) all processes run in parallel and (2) they are arranged in a fixed

[7] Up to now we investigated the comprehension of two types of dynamic maps using verbalization studies. In a first study by Tappe & Habel (1998) we presented *dynamic sketch maps*, which are free-hand sketches of routes, which emerge dynamically on a computer screen. Klippel, Tappe and Habel (2002) discuss the chunking behavior for so-called *moving-dot maps,* i.e. animate depictions, in which a route is represented by a dot moving on a map presented on a computer screen. Our experiments give evidence that dynamic maps enhance memory for some spatial information relevant in following route instructions, e.g., at point where the instructee has to change direction.

[8] In particular, in the case of multimodal generation / production of utterances or messages, the content to be communicated has to be distributed to different modalities, e.g. to language and gestures. In the present paper I focus on the integrated production of language and sketch maps.

[9] It is called current conceptual representation, because the representation changes over time. That means that all processing takes place on the current state of the representation.

sequence so that the output of one process is the input to its successor. To take just one example, a situation can only be selected for verbalization after it has been inserted into the CCR by construction.

In INC conceptual representations are realized using *referential nets* (Habel 1986, 1987). The referential net approach, which is kindred to discourse representation theory (Kamp & Reyle 1993), has been developed to model cognition-motivated linguistic processes, especially, representations that change over time. In referential nets, all information about entities is associated with referential objects (refOs), which can be connected via *relations*, so that a network structure arises. The basic conceptual entities provided by the pre-processing components already contain some information about what attributes (e.g. which sort) have to be ascribed to a refO. In the following we use symbolic constants to refer to refOs. These are just arbitrary labels; the important point is that the refOs can be related to suitable refOs of subsequent processes, which, for example, stand for lexical items.

In the following I give a more detailed overview on the four, cascaded processes that constitute the 'heart' of the conceptualizer exemplified with the production of a route instruction:

- The *construction* process takes the route, i.e. the trajectory of the *virtual* journey through the *virtual* environment as input and builds up a hierarchical event representation in the CCR. In the case of incremental generation of route instructions, we focus currently on two different types of input into construction. First, an incremental planer provides the segments of the route in question incrementally via the pre-processing unit to the *construction* unit. Second, the planer produces a data stream as used in the dynamic map of 'moving dot' type mentioned in the last section.]

  In the domain or route instructions *turn*-events are highly relevant. The representation of such an event is depicted in figure 3 (cf. Guhe, Habel & Tschander, 2003, on generating verbal descriptions of motion events). On the layer of trajectories this *turning* contains three paths, namely r3, a straight path, r5, a completely curved path, and r7 the concatenation of these paths, which is partially curved. Furthermore, the spatial configuration contains a location r6, which is the transition point between r3 and r5. On the layer of events, there are to moving events of the type chpos, which are summarized to event r9.
- The *selection* process chooses the nodes that will be verbalized. Following the *virtual journey* strategy of route descriptions, the verbalization focuses on the *motion events*. Thus refOs of this sort are linked together to form a (partial) path—the *traverse*—through the network. We use the notion *traverse* in two readings: (i) it denotes the *resulting* path, i.e. the concatenation of all nodes that are actually chosen and verbalized; (ii) it means the *current* traverse that contains a limited number of selected elements that are not yet taken by the PVM-generation at a given point of time. Elements that are contained in the current traverse may be removed and are not contained in the resulting traverse anymore.[10]

  The selection process can manipulate the traverse by means of two operations: *appending* and *replacing* elements. While the first one is more fundamental, the second especially serves the purpose to substitute already chosen nodes by better candidates (for details cf. Guhe, Habel & Tappe 2000).

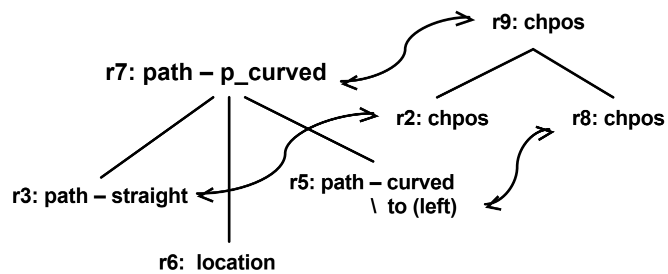

Figure 3: Conceptual structure: turn to the left at location r6

- The *linearization* process has the function of bringing the selected nodes into an appropriate order. Linearization is necessary in a general psycholinguistic model, for example, to move the more prominent (more informative) items to the beginning of a sentence or more complex dialogue unit. Since this is not the case in describing simple motion events as well as in generating route instructions following the *virtual journey* strategy, I do not discuss this module here.
- *PVM-generation* observes the traverse and generates preverbal messages from the selected nodes. There is a latency time between the time when a node is selected and the moment when the PVM-generation uses it as preverbal message, i.e. passes it on to the subsequent process of the formulator (cf. Guhe & Habel 2001).

  In INC we use a algorithm for the generation of incremental preverbal messages that bears some resemblance to the incremental algorithm proposed by Dale & Reiter (1995). In contrast to them, we are concerned with the generation of the preverbal messages

[10] The limited capacity of the current traverse can be varied for simulations of cognitive behavior with different processing (memory) capacities, leading to different verbalizations.

which may contain referring expressions and not with the generation of referring expressions by the formulator.

Deciding upon content to be verbalised is always a two-step process in INC. In the first step the conceptualizer decides to verbalise a refO. Usually, refOs contain lots of descriptions, of which some refer to other refOs. Thus, in the second step descriptions from the refO and further refOs (with further descriptions with further refOs and so on) are selected. The large number of descriptions makes constraints for the selection of descriptions indispensable. Additionally, not all information given by the descriptions is needed in a verbalization.

Although the cognitively motivated architecture of INC is based on simple design principles it leads to a model that exhibits behavior very similar to that observed in verbalizations of humans when they give on-line descriptions of events (cf. Guhe & Habel 2001): By varying three resource parameters—two of them, namely *Latency Time* (LT) and *Length of Traverse Buffer* (LoTB) I mentioned above—it was possible to identify those *parameter settings* that lead to verbalizations typical our verbalization corpus,. Furthermore, INC produced all types of verbalizations of this corpus. Furthermore, INC does not need explicit instructions about what degree of detail or what level in the event hierarchy is used for the verbalization. Instead, by assigning different parameter settings—using the three parameters motivated by cognitive considerations—different types of output are produced.[11]

To conclude this section I describe the interplay of *selection* and *PVM-generation* by an informal example: In verbalizing the *turn*-event whose conceptual representation is depicted in figure 3, the conceptualizer has—dependent on the setting of parameters—different options to select refOs and descriptions:

- One event, r9, is verbalized. Further specification of the direction is 'inherited' from the to left description of r5 via r9. Result: "Go to the left."
- One event, r9, is verbalized. The spatial details the trajectory with respect to the description chpos, namely that it is concatenation of r3 and r5, which hold a specific spatial relationship, leads to an additional second, more specific descriptions of r9, namely, turn. Result: "Turn to the left."
- Two events are verbalized, namely r2 and r8. Than the specification of r2 requires information from r3, for example about the final point of r3, which is identical with r3 (on details of conceptual representations of motion events in the referential net approach, cf. Guhe, Habel & Tschander 2003). A possible result: "Stay straight on Van Ness Ave. Turn left onto Bay St."

## Towards Multimodal Route Instructions: Dynamic Maps Augmented by Speech

In the following I present the combination of *moving-dot maps* and *spoken route instructions* as an exemplary case of incremental multimodal route instructions.[12] This combination of modalities is chosen, since it can prevent the cognitive overloading of the instructed people.[13]

I use, as described above, the *virtual journey*, formally constituted by combining a route with the route-following actions, as input for both types of determining content and structuring the multimodal document. Furthermore, I assume that the first three processes, namely *construction, selection* and *linearization*, undergo only minor changes in case of multimodal communication. Thus I will go on with the fourth module, only mentioning shortly what is different in the first three modules.

As discussed above, the major deficit of maps as means for route instruction is, that they do not focus on the sequence of actions to be performed but on the spatial environment. Or from another perspective, the linear objects representing tracks are static and not oriented. Schweizer, Herrmann, Janzen & Katz (1998) report that the direction of the route as experienced during acquiring route knowledge enhances later usage of this knowledge. Moving-dot maps provide ordering information during presentation. Thus they overcome the disadvantage of maps—especially, the deficit in expressing temporal information—since they present the motion to be performed later by the instructee in an analog manner. But there other types of relevant relations between events, which can easily be expressed in verbal descriptions, for example purpose, as in "Keep left at the fork in the ramp, since you have to merge onto the interstate". To generate this, the

[11] This can also be seen as evidence that also humans do not deliberately choose one specific level, but instead, that they make unconscious—or sometimes even conscious—decisions about the resources they use in language production.

[12] Since a thorough analysis of language-depiction multimodality is out of the focus of the present paper, I omit here the discussion of research in this area completely, even if there are some approaches highly relevant for the subjects presented here, cf, for example, Wahlster et al. (1993), Towns, Callaway & Lester 1998, André (2000).

[13] This does not mean that combining speech and visualization guarantees decrease of cognitive load. Furthermore, it is currently not clear whether decreasing cognitive load is due to a modality effect or to avoidance of the split-attention effect (Cf. Guan, 2002).

conceptualizer described in the last section has to be extended by adding 'purpose links' between event refOs, for example by mechanisms developed in Rhetorical Structure Theory (cf. Marcu1996).

A second promising type of language augmentations concerns foregrounding of landmarks, facilitating extracting information from the map and making implicit information explicit. Since—especially in the case of dynamic maps—the map comprehender's primary problem is to distinguish between task-relevant and task-irrelevant tokens on the map, speech can aid in map comprehension. The verbal description "Follow Brigton Avenue (Route 25) for approximately 3 miles until the stop light at a five way intersection (Deering Street/Falmouth Street/Brighton Street intersection)" exemplifies all three types mentioned above: *the five way intersection* is foregrounded, the distance of *approximately 3 miles* is made explicit verbally, and the names of streets constituting the intersection, are acoustically easier to comprehend than by reading the map.

The conceptualization process—as part of natural language generation—is based on the same data as the moving-dot map visualization, in other words, both systems, the NL generating system and the map-animation system, get the same *virtual journey* as input, the synchronization of both subsystems of the multimodal route instructing system is strongly supported by the NLG's incremental character. Figure 3 depicts the links between paths and events; similar links hold between the other entities, as landmarks etc., which are conceptually connected to the events in question (cf. Guhe, Habel & Tschander 2003). Thus the paths-refOs—and their associated start-points, end-points and transition-points—function as the synchronizing units in the multimodal generation process.

## Conclusion

In this paper I presented the principles and the architecture underlying the incremental conceptualizer INC, which is a 'what to say'-component for verbalizing continuous streams of perceptual or conceptual input. INC can be augmented to a 'what to communicate'-component of a *multimodal route instruction system* combining natural language route descriptions and visualizations of the route to follow, such that the strengths of both means for communication route knowledge are brought together.

While the basic principles for incremental multimodal route instruction are understood now, the realization and empirical testing of different types of such systems will be subject of future research.

The second major line of our future research is the extension to dialogical route instructions. The following topics will be stay in the foreground:

- Multimodal reaction by the instructee, i.e., the instructee will have the choice between spoken, graphical and other haptical response.
- In general, we will investigate *cross-modal* turns. This includes the question, whether and in which way the distribution of content to different modalities influences the responding participant in a multimodal dialogue to use one modality or the other.

**Acknowledgements** The research reported in this paper was partially supported by grants (Ha 1237/7 [ARK] and Ha 1237/10 [ConcEv]) of the Deutsche Forschungsgemeinschaft (DFG) in the priority programs on 'Spatial Cognition' and 'Language Production', respectively. I thank Markus Guhe, Heike Tappe and Ladina Tschander for their collaboration in designing and realizing INC. I am indebted Carola Eschenbach for many fruitful discussions on Spatial Language and Cognition in general, and for her help in axiomatizing spatial concepts in particular.

## References

André, Elisabeth: The Generation of Multimedia Documents. (2000). In R. Dale, H. Moisl and H. Somers (eds.). *A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text*. (pp. 305-327). Marcel Dekker Inc.

Clark, H. H. & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Dale, R. & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. Cognitive Science, 18 , 233–263.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive 16*. 409–458.

De Smedt, K.; Horacek, H. & Zock, M. (1996) Architectures for natural language generation. In G. Adorni & M. Zock, (eds.), *Trends in natural language generation*, (pp. 17–46). Berlin: Springer.

Eschenbach, C., Habel, Ch. & Kulik, L. (1999). Representing Simple Trajectories as Oriented Curves. In A.N. Kumar & I.Russell (eds.), *FLAIRS-99, Proceedings of the 12 th International Florida AI Research Society Conference* (pp. 431–436), Orlando, Florida.

Fraczak, L.; Lapalme, G.; and Zock; M. (1998). Automatic generation of subway directions: Salience gradation as a factor for determining message and form. In *Proceedings of the Ninth International Workshop on Natural Language Generation, Niagara at the Lake, Ontario, Canada*. (pp. 58–67).

Golledge, R. G. (ed.) (1999). *Wayfinding behavior. Cognitive mapping and other spatial processes*. Baltimore, MD: John Hopkins University Press.

Green, P. (1997). Potential Safety Impacts of Automotive Navigation Systems. Paper presented at the Automotive Land Navigation Conference, June 18, 1997, Detroit, MI.

Green, P. (2000). Crashes Induced by Driver Information Systems and What Can Be Done to Reduce Them (SAE paper 2000-01-C008). Convergence 2000 Conference Proceedings (SAE publication P-360). Warrendale, PA: Society of Automotive Engineers, 26-36.

Guan, Ying-Hua. (2002). Reexamining the modality effect from the perspective of Baddeley´s working memory model. In Ploetzner, R. (Ed.). *Proceedings of the International Workshop on Dynamic Visualizations and Learning*. Tübingen: Knowledge Media Research Center. http://www.iwmkmrc.de/workshops/visualization/proceedings.htm

Guhe, M., Habel, Ch. & Tappe, H. (2000) Incremental Event Conceptualization and Natural Language Generation in Monitoring Environments. *INLG-2000*, 85–92, Mitzpe Ramon, Israel. (Published by ACL).

Guhe, M. & Habel, Ch. (2001) The Influence of Resource Parameters on Incremental Conceptualization. In: E. M. Altmann, A. Cleeremans, C. D. Schunn & W. D. Gray (eds.) *Proceedings of the Fourth International Conference on Cognitive Modeling*: July 26-28, 2001, George Mason University, Fairfax, Virginia, USA. 103–108 Mahwah, NJ: Lawrence Erlbaum.

Guhe, M.; Habel, Ch. & Tschander, L. (2003). Describing motion events: Incremental representations for incremental processing. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, January 15_17, 2003, Tilburg, The Netherlands. (pp. 410–424).

Habel, Ch. (1986). *Prinzipien der Referentialität*. Berlin: Springer Verlag.

Habel, C. (1987). Cognitive linguistics: The processing of spatial concepts. *T. A. Informations (Bulletin semestriel de l'ATALA, Association pour le traitement automatique du langage)* 28. 21–56.

Jeung, H. J., Chandler, P. & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology,* 17, 329-343.

Kamp, H. & Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer.

Kempen, G. & Hoenkamp, E. (1987) An Incremental Procedural Grammar for Sentence Formulation. *Cognitive Science*, 11:2, 201–258.

Klein, W. (1982). Local deixis in route directions. In R. J. Jarvella & W. Klein (eds.), *Speech, place, and action*. (pp. 161–182). Chichester: Wiley.

Klippel, A.; Tappe, H. & Habel; C. (2002). Pictorial representations of routes: Chunking route segments during comprehension. In C. Freksa, W. Brauer, C. Habel and K. Wender (eds.), *Spatial Cognition III*. Berlin: Springer. Forthcoming.

Levelt, W.J.M. (1989) *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lovelace, K. L.; M. Hegarty; and D. R. Montello (1999). Elements of good route directions in familiar and unfamiliar environments. In C. Freksa and D.M. Mark (eds.), *Spatial Information Theory*. (pp. 65–82). Berlin: Springer.

Marcu, D. (1996). Building Up Rhetorical Structure Trees. *The Proceedings of the Thirteenth National Conference on Artificial Intelligence,* vol 2, 1069–1074.

Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In M. J. Egenhofer & R. G. Golledge (eds.), *Spatial and temporal reasoning in geographic information systems* (pp. 143-154). New York: Oxford University Press.

Reiter, E. & Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.

Schweizer, K.; Herrmann, T.; Janzen, G. & Katz, S. (1998). The Route Direction Effect and its Constraints. In C. Freksa, C. Habel & K.F. Wender (eds.), Spatial Cognition. (pp. 20–38). Berlin: Springer.

Tappe, H. & Habel, Ch. (1998). Verbalization of Dynamic Sketch Maps: Layers of Representation in the Conceptualization of Drawing Events. Poster at Cognitive Science Conference; Madison WI, August, 1.–4., 1998.

Tschander, L. B.; Schmidtke, H. R.; Habel, C.; Eschenbach, C.; and Kulik, L. (2002). A geometric agent following route instructions. In C. Freksa, W. Brauer, C. Habel and K. Wender (eds.), *Spatial Cognition III*. Berlin: Springer. Forthcoming.

Towns, S.; Callaway, C. & Lester, J. (1998). Generating Coordinated Natural Language and 3D Animations for Complex Spatial Explanations. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pp. 112-119, Madison, Wisconsin, July 1998.

Tversky, B. & P. U. Lee (1999). On pictorial and verbal tools for conveying routes. In C. Freksa & D.M. Mark (eds.), *Spatial Information Theory*. (pp. 51–64). Berlin: Springer.

Wahlster, W.; André, E.; Finkler, W.; Profitlich, H. -J. & Rist, Th. (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence, 63*. 387–427.