CORAL: Using Natural Language Generation for Navigational Assistance

Robert Dale **Sabine Geldof** **Jean-Philippe Prost**

Center for Language Technology
Macquarie University
Sydney, Australia
Email: {rdale|sabine|jpprost}@ics.mq.edu.au

Abstract

In this paper we tackle the problem of generating natural route descriptions on the basis of input obtained from a commercially available way-finding system. Our framework and architecture incorporates the use of generic natural language generation techniques. Through examples we demonstrate that it is possible to bridge the gap between underlying representation and natural sounding descriptions. The work presented contributes both to the area of natural language generation and to the improvement of way-finding system interfaces.

Keywords: Natural language generation, way-finding systems, micro-planning

1 Introduction

Natural language generation (NLG) is concerned with the production of natural language output, whether written or spoken, from some underlying non-linguistic data source. The technology has been used in a wide variety of systems and contexts; see (Reiter & Dale 2000) for an overview. In this paper, we describe our current work in using NLG to provide fluent and natural navigational assistance in the context of driving directions.

There are now many web-based services which offer the automatic generation of driving directions. *MapBlast*, *MapPoint* and *MapQuest* are three major US providers of this functionality; in Australia, *WhereIs* provides the same kind of information.[1] Although there are interesting differences of detail in the user interfaces provided by each, all these systems are similar in concept and content: the user specifies a start address and a target address, and the system plans a route between these two points, possibly taking into account specific constraints such as a desire to use freeways or to avoid toll bridges. The output of each of these systems is in the form of 'turn by turn' instructions; an example from *WhereIs* is shown in Figure 1.

There may be some advantage to displaying this kind of information in a tabulated form like this: for example, the consistent row-by-row format may make it easier to quickly determine what is involved in the route. Nonetheless, when compared to a human-authored description for the same route, as in Figure 2, several differences become apparent:[2]

- Humans often omit steps that the automated systems include, typically because they are deemed unimportant or obvious; the automated system is not capable of making these assessments.
- Humans typically use landmarks and visible features of the environment to identify turning points, whereas the automated systems generally describe these points by distances or times of travel from previous decision points.
- Humans typically produce complex clause structures, gathering together related information into single sentences, whereas the automated systems produce what are in effect one-sentence-per-step mappings.

In this paper, we explore how natural language generation techniques can be exploited to imbue automatically generated route descriptions with some of these properties.

Of course, there is no prima facie reason why we should want an automated system to emulate what people do. There is no guarantee that a human-produced description is necessarily a good one, and it is clearly possible that the tabulated form of instructions that is common to existing applications is actually an improvement on what people do. There is some evidence, however, that route descriptions closer to those produced by humans are preferable: work on the graphical display of routes, for example by Agrawala and Stolte (2001), has suggested that users prefer modes of delivery which do not give equal status to all parts of the route description, and experiments have demonstrated that describing points by means of salient features of the environment results in route descriptions that are much easier to follow than those couched in terms of distances and travel times, which humans find difficult to estimate and keep track of (Streeter, Vitello & Wonsievicz 1985, Denis, Pazzaglia, C.Cornoldi & Bertolo 1999, Burnett 2000).

Our current work is concerned with the development of a route description system that uses the same underlying Geographical Information Systems (GIS) datasets as the commercially available web-based systems, but which incorporates techniques from natural language generation (NLG) research to produce more natural-sounding descriptions. In this paper, we focus on three aspects of our generation process:

- the use of discourse structure to facilitate understanding of the structure of a route;
- the use of aggregation techniques to combine information into fluent and coherent multiclausal sentences; and

[1] See www.mapblast.com, www.mappoint.com, www.mapquest.com and www.whereis.com.au respectively.

[2] All our human-authored examples are drawn from a corpus of real route descriptions, described later. Our examples of publicly available, web-delivered directions are obtained from www.whereis.com.au.

	Instruction	Street Address	Suburb	Distance	Est. Time
S	depart	RICHMOND ST	DENISTONE EAST	239 m	1 Min
↱	right	LOVELL RD	DENISTONE EAST	37 m	1 Min
↰	left	NORTH RD	EASTWOOD	360 m	1 Min
↱	right	DONOVAN ST	EASTWOOD	390 m	1 Min
↰	left	HERRING RD	EASTWOOD	1.88 km	3 Mins
↰	left	UNIVERSITY AV	MACQUARIE PARK	8 m	1 Min
F	arrive	UNIVERSITY AV	MACQUARIE PARK	Total 2.91 km	Total 8 Mins

Click on instruction icon for detailed map

Figure 1: An automatically generated route description

Leave the house and drive towards the Midway shops, at the end of the street turn right and then left at the roundabout. Drive along North road and take the third right turn, just after the first hump in the road. Go to the end of that road and then go straight ahead at the roundabout, there's a church on your left. Now go straight along Herring road for quite a way until you hit the main road (Epping Rd), go straight across at the lights and continue on until you get to the next set of lights. Turn right here into the university.

Figure 2: A human generated route description for the route in Figure 1

- the use of referring expression generation techniques to produce user-oriented descriptions of key elements in routes.

The techniques we use are an attempt to balance generalisability and domain dependence, in order to provide a practical solution.

The remainder of this paper is as follows. Section 2 sketches some background to the work described here. Section 3 describes the architecture of our system and outlines the approach we take to the problem in general; and Section 4 explores our use of discourse structuring and NLG techniques for referring expression generation and aggregation, along with example outputs that demonstrate the current capabilities of our system. Section 5 provides some brief comments on evaluation, and Section 6 draws some conclusions and points to ways forward.

2 Background

There already exists a considerable body of work in the generation of route descriptions. Pattabhiraman and Cercone (1990) pointed to the importance of salience and relevance in content selection, the first sub-process in NLG. The domain of route descriptions illustrates their point clearly because of the inherent coupling of domain and linguistic knowledge. The notion of salience is further specified as a gradual value by Lapalme *et al* (1998) ; their system produces variants of subway route directions by mapping the relative importance of information entities onto syntactico-semantic features. While these approaches are all concerned with establishing a link between GIS knowledge on the one hand and linguistic realization principles on the other, Moulin and Kettani (1999) take a radically different approach. They advocate the encoding of geographical information centred around those elements that are believed to be crucial in the description of routes, thus conceiving the generation task as a straightforward mapping from the underlying data. Like Lapalme and his colleagues, Höök (1991) also aimed at generating different route descriptions for one particular route, but from a human-computer interaction (HCI) perspective; her focus was the matching of observed differences in navigation style. Finally, the route descriptions generated by Maaß and colleagues (Maaß, Baus & Paul 1995) are based on the integration of cognitive and perceptual information processing.

From our perspective, this earlier work suffers from two drawbacks.

- For the most part, earlier systems have not made use of real GIS data, but have relied on handcrafted knowledge sources to support the generation process. While this strategy allows exploration of desirable outputs in a way that might inform subsequent GIS data development exercises, it does not provide a solution to the limitations of existing GIS-based systems.
- The techniques used in these systems have tended to be somewhat ad hoc, in that they have not attempted to capitalise on more general techniques and approaches developed in the field of NLG.

Our own system, Coral, has evolved over the last few years through a range of quite different instantiations. Our earlier work addressed the provision of route descriptions within a University department (Williams & Watson 1999), providing multi-modal (text, graphics and speech) descriptions via the web; more recently, we have explored how higher-level segmentation of a route description may contribute to its ease of use, especially when delivered via a mobile device (Geldof & Dale 2002).

Our current work represents an attempt to address both of the problems identified above. We use as input precisely the same GIS data that is available to existing commercial web-based systems; and at the same time, we attempt to apply more general principles of natural language generation (see, for example, Reiter and Dale (2000)) to the production of the resulting output texts. To support this work, we have carried out an analysis of several specially collected corpora of human-produced route descriptions. Our corpora differ with respect to mode of navigation, means of communication, and type of environment: our first corpus consists of 49 spoken route descriptions (7 subjects $\times$ 7 routes) within our university department; another corpus consists of 30 written route descriptions (10 subjects $\times$ 3 routes) within the university campus. Of particular relevance to the work described here, we also collected a corpus of 20 written directions within the urban road network: 9 subjects were asked to describe the route from their homes to the university to a visitor and to a neighbour, as well as a route from the university to a fixed, known destination. Whereas the architecture of our system is applicable to the domains explored in each of these corpora, the strategies described in this paper are based on the last corpus; given the variety of parameters that influence the formulation of route descriptions, it was important to reduce our scope to a single mode of transportation and environment type. The familiarity with the environment and the fixed destinations in this corpus allowed us to constrain and control the variations in expression used by our subjects. Our approach to corpus analysis and its application to other corpora are the subject of another publication (in preparation).

3 The Coral Architecture

3.1 The Input Representation

The GIS datasets used in existing systems represent the world in terms of nodes (points in space), arcs (directed links that connect two nodes), and polygons (sequences of arcs that form bounded spaces). Nodes typically represent junctions or decision points in a road network; arcs are the travelable paths between points in that network; and polygons are used to represent areas such as parks or railway stations. A GIS system typically also provides, in additional data sources, information about these entities: street name, length of the path, category of point of interest, and so on. It will be the task of the natural language generation system to include the information deemed useful in a description of a particular route.

The construction of a route plan thus consists in determining a path between two specified nodes; the result of route planning is a sequence of arcs that form a path between these nodes. A number of constraints may be taken into account in planning this path: for example, some systems offer the user a choice of the fastest or the shortest route (not necessarily the same), or of routes that avoid toll bridges. Local constraints such as whether a segment of road is one-way must also, of course, be taken into account.

Before such a plan can be used to produce an output description, it typically undergoes a process of **arc aggregation**. This is our term for the process of merging together those contiguous arcs that are all part of the same road: since an arc joins two junctions, the path between each two intersections along a road constitutes a separate arc, and so an instruction like *Follow Epping Road for 10km* may in fact correspond to several arcs in the underlying representation. Arc aggregation thus turns a raw arc-based plan into what we call a **path-based plan**. From here, it is a fairly simple process to map the route into a sequence of turn-by-turn instructions as in Figure 1. Our interest, however, is in further manipulating the data to produce more fluent and natural output.

3.2 Levels of Representation

In line with current thinking in NLG research, we view the generation process as consisting of three distinct stages: **text planning**, **micro-planning**, and **linguistic realisation**. For our current purposes, text planning consists in taking a path-based route plan, and deriving from this a set of **messages** that are to be conveyed to the user. These can be thought of as the separate chunks of meaning that have to be conveyed. The micro-planning stage then uses these messages to build a sequence of sentence plans that determine the content to be realised in each sentence; this may involve combining clauses to build complex sentences, and also working out what information should be used to describe locations that are mentioned. Finally, the realisation stage maps these sentence plans, which are still in the form of semantic specifications, into the appropriate lexico-syntactic material of the target natural language. This architecture is shown in Figure 3.

A message is, effectively, a piece of semantic content that can be realised linguistically. As argued in Reiter and Dale (Reiter & Dale 2000, Section 3.4.2), the appropriate inventory of message types and their optimal granularity depends on specific characteristics of the application: the general idea is to view messages as data objects corresponding to the largest distinct linguistic fragments we need in order to generate the variety of texts we are interested in. Our analysis of human-produced route descriptions leads

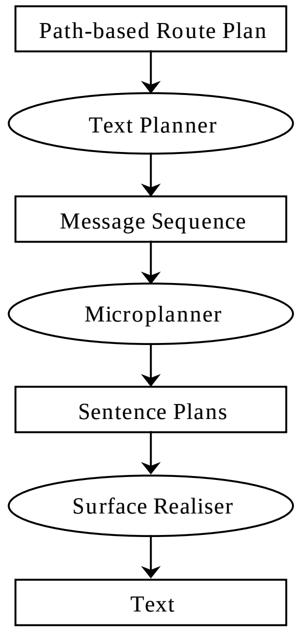


Figure 3: Coral's architecture

us to favour a message level that distinguishes three message types that may be combined in a variety of ways:

Points: Although descriptions of points rarely appear in the route descriptions produced by commercial systems, they are common in human-produced descriptions, where they often serve as a means of checking the user's position. These can either appear as parts of instructions, or in separate sentences whose sole function is to state position.

> Follow the road until *the traffic lights next to 'The Ranch' restaurant.*
> Take a right turn, just after *the Macquarie Center.*
> Turn right at *the first roundabout.*
> There's *a church on your left.*
> You'll go over *two bridges.*

Directions: These correspond to turns that are made at decision points in a route plan.

Paths: These correspond to continuous movements along parts of the road network.

In these terms, the instructions in commercial systems typically consist of a combination of a PATH message and a DIRECTION message; as noted, POINT messages typically do not occur at all.

Given a path-based route plan as introduced in the previous section, we build from this a **text plan** that consists of an alternating sequence of POINT, DIRECTION and PATH messages, terminating in a POINT message that corresponds to the target location. Each message contains information that can be used in describing that message; Figure 4 shows the content of typical POINT and PATH messages. A POINT message includes a list of the identifiers of points of interest (POIs) that are associated with that point and which can therefore be used in describing the point; a PATH

```
[type: point
 nodeID: n21330
 pointtype: start
 address: 'Herring Road'
 poi-list: [n18921]]

[type: path
 distance: [unit: meter
            count: 800]
 street: [name: 'George Street'
          level: 3]
 elements: [a30,n18978,a26,n19002,a21]]
```

Figure 4: Example point and path messages

message contains its level in the road status hierarchy (here, 3 means that this is a main road), the distance to be travelled along this path, and the constituent arcs and nodes that make up the path (these are the elements combined in arc aggregation).

This text plan then serves as the input to our micro-planning process, which is faced with two tasks:

- It must decide how to cluster together the POINTs, DIRECTIONs and PATHs into clause-sized units; and
- it must decide how to refer to each of these elements.

The first of these is a linguistic aggregation task (Dalianis 1999), while the second is an application of referring expression generation (Dale 1992, Dale & Reiter 1995). We describe our approach to each of these tasks in Sections 4.4 and 4.5 below, but first turn to the higher-level generation task concerned with determining the structure of the route description and its correlation to ease of task execution.

4 Applying NLG Techniques

4.1 Discourse Structure

Höppner (1995) formulated requirements for route descriptions in general: a route description needs to be both *recognizable* and *rememberable*. Our view is that these cognitive requirements can be addressed to some extent by introducing *segmentation* and *structuring* into the flat sequences of instructions provided by existing systems. Given a flat sequence of instructions of the kind delivered by a typical navigational assistance system, our approach is to segment this sequence of instructions in a meaningful way, and to generate a summary for each resulting segment. This hierarchical approach reduces the cognitive load on the user and enhances the rememberability of the route description.

Our solution is based on two elements. First, the route to be described needs to be segmented and summarized in a meaningful way. In an ideal world this might correspond to the top-down structure developed in a hierarchical planner; however, existing systems do not make use of or provide such structures, and so we have explored the use of bottom-up heuristics for the identification of appropriate segmentations.

Then, we need techniques that support flexible interaction with the segmented route in conjunction with task execution. For example, we exploit the hierarchical structure resulting from segmentation to present route descriptions on a small device (in particular, a Palm hand-held computer). This presentation mode allows step-by-step exploration of the description as the user performs the navigation task (see Figure 6).

Subsections 4.2 and 4.3 further explain each of these elements.

4.2 Segmentation

As noted above, existing route planning systems provide flat sequences of instructions, consisting of alternating paths and turns, rather than hierarchical structures. The process of segmentation therefore consists in grouping these path and turn instructions into higher level entities that we call *segments*. The notion of segmentation we are working with here bears some relation to the notion of a discourse segment as discussed by Grosz and Sidner (1986): elements that are more related are seen as aggregating together to form segments within a larger structure, and in theory this analysis applies recursively to produce a hierarchy.[3]

The concept of hierarchy in way-finding is not, of course, new. The process of human spatial knowledge acquisition is often assumed to result in a hierarchical structure, referred to as the cognitive map by Kuipers (1978); and Pailhous's observation of way-finding behaviour by experts (i.e. taxi drivers in Paris) confirmed the hypothesis of the existence of a hierarchical strategy, where first a route between regions is constructed at a higher level before being refined into concrete path components (Pailhous 1970).

As so far described, segmentation can be viewed as a way of coherently organising and structuring information. However, it can also be seen as addressing a key question in the provision of information in dialogic contexts: how do we convey information in installments so that the course of information exchange approximates the way humans interact? The segmentation of information in human dialogue responds to the need to decrease the cognitive effort required from the interlocutor (Clark & Schaefer 1989).

Of course, only a subset of all (mathematically) possible segmentations of a stream of information is meaningful. Thus a key task is to determine which segmentations are valuable. We have explored two alternative strategies: one determines optimal break points in the sequence of paths that make up the route, and the other aggregates several paths into a higher level structure on the basis of properties of the constituent elements. These strategies have been applied to the output of existing route description systems.

4.2.1 Landmark-based segmentation

Our first strategy relies on the experimentally verified idea that landmarks at decision points constitute useful cognitive entities that improve the effectiveness of route descriptions (Lovelace, Hegarty & Montello 1999, Denis et al. 1999, Burnett 2000). Although what constitutes a landmark remains vague and ill-defined, attempts have been made to distinguish different categories of landmarks. Sorrows and Hirtle (1999), for example, identify visual landmarks (objects such as churches and towers which are clearly distinguishable from their environment by virtue of salient visual features), cognitive landmarks (for example, the desk of a receptionist, which may be significant because it has a particular function for a user), and structural landmarks (entities such as Trafalgar

[3] In practice, we have so far only found need for one level of hierarchy in our structures.

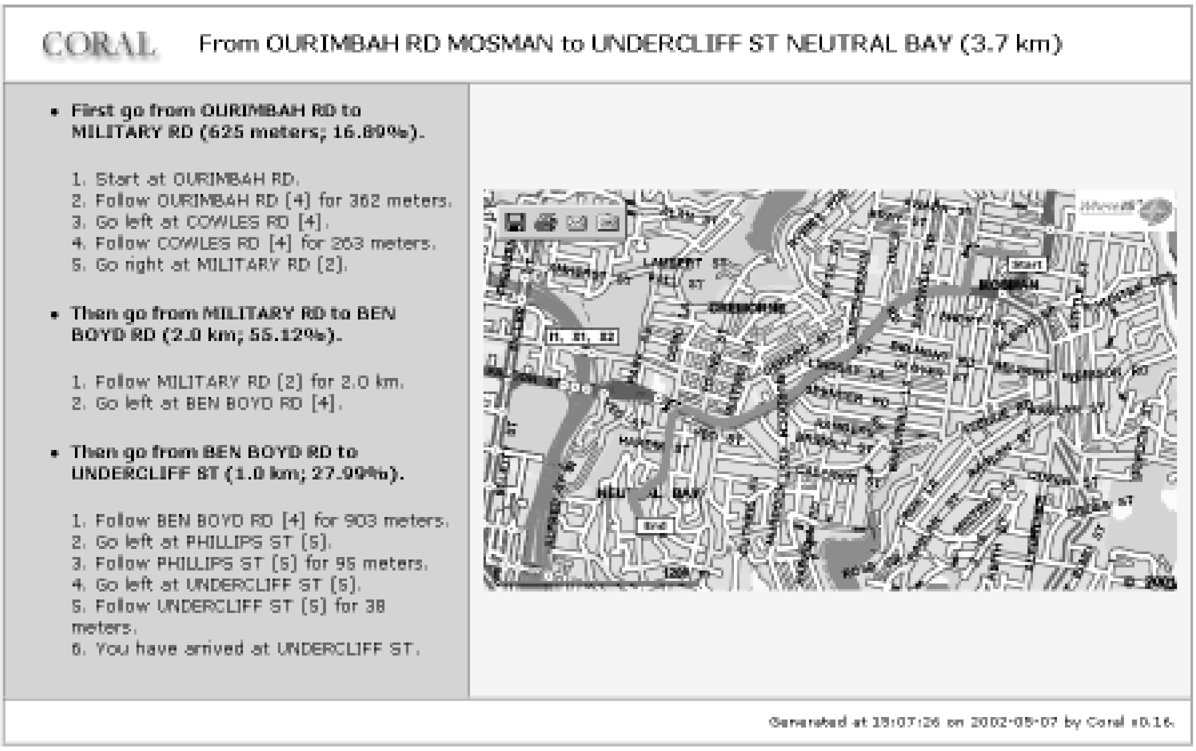


Figure 5: Example of a segmented route presentation via the Web

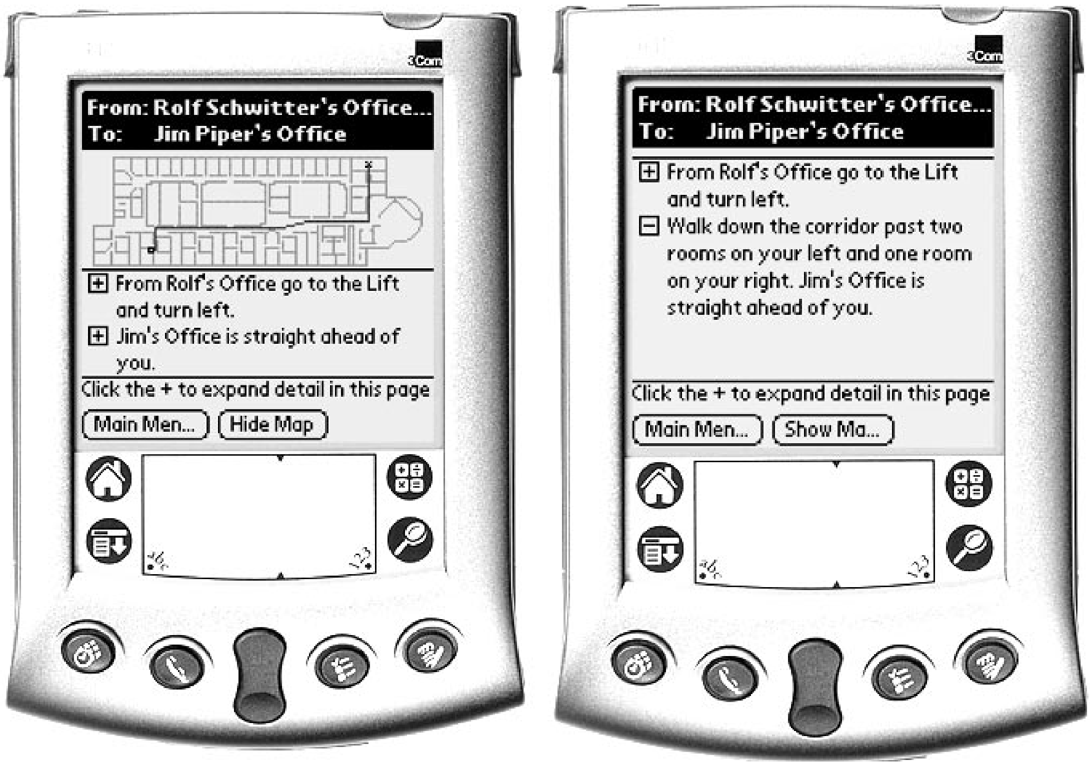


Figure 6: Example of a segmented route presentation via the Palm

Square in London, which assists in structuring a spatial environment). Raubal and Winter (2002) developed measures to formally specify the saliency of various features in view of landmark selection. However, these require the availability of rich datasets.

We explored the hypothesis that the global saliency of landmarks can be exploited to structure route descriptions. A landmark at a decision point delimits a part of the route to be followed, so the navigator will be aware whether she has reached that point in the route and will thus know how far she has progressed in the navigation task at hand.

We have applied this idea to an earlier version of Coral which provided indoor route descriptions for our department (Williams 1998). The knowledge representation used in that system includes landmarks as domain objects, and these are included in the intermediate representation from which the textual route description is generated. The route plan representation consists of a sequence of alternating path and turn specifications as shown in the following example, which underlies part of the route presented in Figure 6:

⟨start(r333), via(⟨⟩), end(p333)⟩,
turn(lhs),
...
⟨start(p2), via(⟨c1, pass(lhs:[4,room]),
pass(rhs:[1,room]), final(rhs:lift1)⟩),
end(lift1)⟩
turn(lhs),
...
⟨start(p362), via(⟨⟩), end(r362)⟩

Our segmentation strategy makes use of a separate knowledge source that indicates which domain objects are plausible landmarks; in the present case, the lift is one such object. Since this appears at a decision point (just before the final left-hand turn in the fragment above), it is selected as a segment border and included in the summary for this segment. Consequently, the route is decomposed into one segment leading to the lift and a second segment from this landmark to the destination.

The intuition behind this approach to segmentation is quite straightforward: if the user is familiar with the environment, she will recognize the landmark that terminates the segment and realise that she does not need the detailed instructions for that segment. It is also easy for the user to keep this landmark in mind as an intermediary target and to remember that, once she has reached it, she should revert back to the instructions.

There are, however, limitations to this strategy, since it depends on the presence of landmarks at appropriate locations along the route. Applied blindly, it can lead to segments of significantly varying lengths, which can be confusing. Overall, then, whereas a landmark-based segmentation might be feasible for route descriptions on a small dataset (such as an indoor area), where it is relatively easy to determine which objects of the domain constitute landmarks, it becomes more difficult to apply on a larger scale, and this is particularly the case with currently available GIS data.

4.2.2 Path-based segmentation

Another approach to segmentation is to investigate characteristics of the constituent paths of the route to determine whether they belong to a meaningful higher-level entity. Other work (see, for example, Höök (91) has explored the hypothesis that recurring higher-level patterns can be found in route descriptions. A frequently occurring pattern consists of three segments corresponding to the beginning, middle and end of a route; typically these involve, respectively, getting onto a main thoroughfare or higher-level road, travelling along that road, and then leaving that road to reach the destination via a number of lower-level roads. We refer to this route pattern as 'BME'. For example:

- How do I get from Macquarie University to the Queen Victoria Building, in the City?
- [Well, first you get onto Epping Road $_B$], [then you continue ahead via the freeway, following signs to the City $_M$]. [Exit at Druit Street, then the QVB is not far from there $_E$].

Given a flat sequence of paths and turns, we need to determine how these constituents are allocated to segments within such a structure. Our analysis of a small corpus of human-generated routes led us to formulate the hypothesis that three features of paths and turns play a role in this segmentation:

Road status hierarchy: Routes often involve travelling on roads of different status within the road network, from freeways down through main roads to side roads. Our analysis demonstrated that a series of consecutive paths of the same or similar road status is likely to be perceived as constituting a higher-level entity.

Path length: For some routes, segmentation on the basis of road status alone can result in a large number of segments. In such cases, the total length of a segment can help to decide which one of the segments is the stable middle segment.

Turn typology: A turn that is very salient (for example, a T-junction) or that requires careful navigation (for example, a right turn in a drive-on-the-left road context) is a likely segment border candidate.

These principles are very prominent in the prototypical BME route, as demonstrated in the example above: the middle segment consists of a long stretch of one or more steps on higher level roads, and the absence of explicit or difficult turns along this middle segment reinforces the perception of a stable section in the route. However, when examining a larger number of routes, it becomes clear that many variants on this pattern exist, and that these three features interact in a complex manner.

To allow for a systematic exploration of the space, we implemented a segmentation module that takes as input a route obtained from a route planning system available on the web, augmented with road status information derived from a widely used street directory. We used 23 routes of different length and in various suburbs of Sydney in our initial exploration.[4] Our main criterion for segmentation quality was approximation to the prototypical BME pattern. We experimented with various combinations of road status-based and length-based heuristics for segmentation; our conclusions from this study were that road status is a good indicator for segmentation (in 43% of the cases); in most other cases (another 34%), segmentation can be improved by augmenting this with heuristics that combine segments on the basis of path length. Our intuition is that turn type information could be used as an additional factor in determining segmentation borders but further experiments would be needed to confirm this.

[4] These routes can be inspected online. See `http://www.ics.mq.edu.au/~coral/Routes/Sydney/Segm/rte002.html` for an example.

```
<route-plan context="Sydney">
    <summary>
        <from>BAY RD ARCADIA</from>
        <to>UNIVERSITY AV MACQUARIE PARK</to>
        <distance>35.0 km</distance>
    </summary>
    <map url="http://www.ics.mq.edu.au/~coral/
             Routes/Sydney/map302.gif"/>
    &lt;segment sid="1"&gt;
        <summary>
            <string>First go from BAY RD to PACIFIC HWY.
            </string>
        </summary>
        <detail>
            <utterance uid="1">
                <string>Start at BAY RD.</string>
            </utterance>
            ...
```

Figure 7: A fragment of RPML

4.3 A Route Planning Markup Language

Our goal is to produce one route description that can be rendered via a variety of devices; in the first instance we have been exploring rendering via both standard desktop web browsers and via hand-held computers (specifically, the Palm), and we are also extending this to voice delivery via VoiceXML.

To support this variety of outputs, we have defined an intermediate, device-independent representation called RPML (for Route Planning Markup Language). Two principle features of this representation are that (a) it allows for the annotation of a route description with segmentation information that can be used for presentation by the rendering device; and (b) it allows for multi-modal content, such as links to graphical representations of the described route and to voice output. Using this representation, we use XSLT to produce web pages for pre-trip planning like those found at `http://www.ics.mq.edu.au/~coral/Routes/Sydney/Segm`, and the same input is used by a specially written renderer on the Palm that formats the output for interactive display to support incremental exploration of the route description while travelling. Figure 7 shows a fragment of RPML; this demonstrates how individual instructions can be provided both as canned output (*First go from BAY RD to PACIFIC HWY*) and as more abstracted specifications (as in the contents of the top level `<summary>` element) which the renderer can decide how to realise.

4.4 Aggregation

Aggregation is the process of building sentences which communicate several pieces of information at once. Although the messages in our text plan could be realised one-per-sentence, the result would be less than fluent, as exemplified in Figure 8.

Of course, there are many situations where one sentence will indeed be used to convey a single message. However, our examination of human-produced route descriptions has identified two specific aggregation strategies that people frequently pursue:

Path+Point: A common strategy is to fold a description of a point into the description of a path, in order to provide a more effective way of identifying the end of that path:

> Now go straight ahead along Herring Road for quite a way until you hit the main road (Epping Road).

Start at Liverpool Street.
Follow Liverpool Street for 86 meters.
You are at George Street.
Turn right.
Follow George Street for 230 meters.
You are at Bathurst Street.
Turn left.
Follow Bathurst Street for 8 meters.
You have arrived at your destination.

Figure 8: One message per sentence

Start at Liverpool Street.
Follow Liverpool Street for 86 meters.
Turn to the right at George Street.
Follow George Street for 230 meters until you reach Bathurst Street.
Turn left.
Follow Bathurst Street for 8 meters.
You have arrived at your destination.

Start at Liverpool Street.
Follow Liverpool Street for 86 meters until you reach George Street.
Turn right.
Follow George Street for 230 meters.
Turn to the left at Bathurst Street.
Follow Bathurst Street for 8 meters until you reach your destination.

Figure 9: Different aggregations

> Continue on until you get to the next set of lights.

Point+Direction: Very often, a turn direction is combined with a specification of the location where this instruction is to be executed:

> ... and take the third right turn, just after the first hump in the road.
> ... and then go straight ahead at the roundabout.
> ... at the end of the street turn right.

Note here that the point description can be realized either before or after the turn or follow instruction; we view this variation as a choice made in the realisation stage, so both variants involve the same aggregation strategy.

We also find sentences that combine all three of path, point and direction, as in *Go to the end of that street and then go straight ahead at the roundabout.* However, from our perspective this is the result of a clause combining process that takes effect once aggregation at the message level has been applied: in effect, aggregation determines the content of major clauses, which may then be realised as single-clause sentences, or combined to form conjoined sentences.

Clearly, applying different combinations of strategies to the same route plan will result in different ways of describing that plan. Currently, our Prolog implementation uses backtracking to produce all possible combinations of the applications of these strategies to a given text plan; Figure 9 shows some of the various realisations possible for the route shown in Figure 8. In future work, we aim to explore a scoring regime that ranks the various renderings.

4.5 Referring Expression Generation

Referring expression generation is the process of determining what semantic content should be used in

Start at Liverpool Street.
Follow Liverpool Street for 86 meters until you reach George Street.
Turn right.
Follow George Street for 230 meters.
After you pass Wilmot Street turn to the left at Bathurst Street.
Follow Bathurst Street until you reach *St. Andrew's Cathedral*.

Figure 10: Applying referring expressions generation

describing an intended referent; the goal is to distinguish the intended referent from other entities with which it might be confused. So, for example, describing the location of a turn by referring to an object at the relevant intersection is only effective if that description does not also apply to other intermediate intersections on the way: an instruction like *Turn left at the traffic lights* may be a true description of the location of a turn, but it is not helpful if there are intermediate intersections that also have traffic lights.

In Dale (1992), the task of referring expression generation is characterised as being driven by three principles: **sensitivity** (the speaker must pay heed to what the hearer can be presumed to know), **adequacy** (the referring expression should identify the intended referent unambiguously), and **efficiency** (the referring expression should not contain more information than is required for the task at hand). Following Dale and Reiter (1995), we take the view that the best way to meet these requirements is to use a general purpose algorithm that is fed by a 'preference ranking' of domain properties and relations that can be used in building referring expressions; properties and relations from a predetermined list of types are added to the content of a description until enough information to identify the referent has been collected.

In the context of our current work, a number of distinct reference strategies are applied by the algorithm in turn. Again, on the basis of a first corpus analysis and the readily available GIS information, we have identified the following strategies for **referring to junction points**:

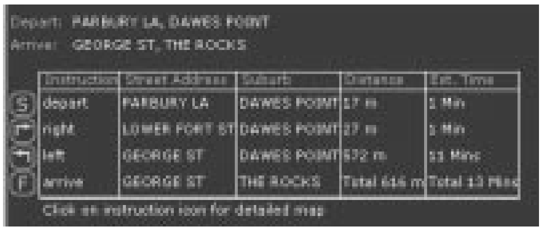
1. Use a landmark that is at, or close to, the junction.
2. Use the type of intersection (for example, roundabout, T-junction, or fork).
3. Use the name of the immediately preceding intersection.
4. Use the name of the intersecting street.

Thus, we use whatever information the underlying dataset makes available, and only fall back on the 'intersecting street name' strategy as a last resort. Examples of the third and the first strategy respectively are shown in Figure 10.

A similar range of reference strategies is used to provide appropriate **descriptions of paths**:

1. Mention street name and any landmarks that are passed on the path.
2. Mention street name and the distance to be travelled along the path.

The effectiveness of these strategies is determined by the richness of the underlying data set. For example, in the data we are currently using, there are no details of whether junctions have traffic lights, so a type of landmark that is commonly used by humans is not


Depart: PARBURY LA, DAWES POINT
Arrive: GEORGE ST, THE ROCKS

Instruction	Street Address	Suburb	Distance	Est. Time
depart	PARBURY LA	DAWES POINT	17 m	1 Min
right	LOWER FORT ST	DAWES POINT	27 m	1 Min
left	GEORGE ST	DAWES POINT	572 m	11 Mins
arrive	GEORGE ST	THE ROCKS	Total 616 m	Total 13 Mins

Click an instruction icon for detailed map

Start at Parbury Lane.
Follow Parbury Lane until you reach the end.
Take a right.
Follow Lower Fort Street for 30 meters.
Turn to the left at George Street.
Follow George Street until you reach your destination.

Figure 11: Route descriptions generated by Whereis and Coral compared

available to us. However there are a sufficient number of other kinds of landmarks encoded in the data to be able to provide useful descriptions.

4.6 An Example

Combined with the aggregation strategies described in the previous section, the application of these techniques allows us to generate route descriptions which are considerably more fluent than those in commercial systems. Figure 11 shows a route description provided by *WhereIs*, and the same route described by our system, making use of the aggregation and referring expression strategies described above.

5 Evaluation

Performing a rigorous and comprehensive evaluation of technology of this kind is an extrememly costly enterprise. Following the recommendations of Ross and Burnett (2001) for the evaluation of navigational systems, we have performed a small-scale expert evaluation in a task-based context, described in more detail in Geldof and Dale (2002). The aim of the experiment was to obtain feedback on the use of segmented route descriptions and their incremental presentation on a mobile device. The experiment was well-defined and carefully designed to minimize the many factors which influence the performance of a navigation task. Three teams, each consisting of a navigator and a driver, were asked to drive 2 routes each, one using a segmented and one using a non-segmented route description delivered via a Palm hand-held computer. An observer accompanied each team and feedback was collected via a form to be filled out during and just after the navigation task. A striking difference in navigation style was observed, between a navigator who relied heavily on distance information provided in the descriptions, and the two others used more general orientation and higher level information such as that provided through the summary of segments. While no noticeable problems occurred with respect to task performance, the first navigator expressed a preference for the non segmented presentation mode, whereas the other navigators indicated that they found the segmented presentation on the mobile devices more useful. Thus, although a larger scale experiment would need to be set up to further confirm this result, this pilot evaluation is indicative of the utility of this aspect of our approach.

The observed differences in navigation style point to one important aspect which has an impact on any evaluation task in this domain, while at the same time highlighting the potential benefit of natural language generation technology for tuning route descriptions to personal preferences. On the one hand, individual judgements of the quality of route descriptions will to a high extent depend on personal preferences. This increases the complexity of an evaluation set-up (as one needs to ensure that the results are not biased towards a particular subject—note that there is no clearly specified gold standard), but also creates a situation in which no objective measure is available. On the other hand, once elicited, the principles underlying these preferences may be embedded in a generation system that outputs route directions better adapted to individual needs.

6 Conclusions

In this paper, we have presented a framework and architecture for generating route descriptions that approximate the naturalness of human generated route descriptions. Unlike other attempts towards this goal, our approach allows us to take as input GIS data like that currently used by commercial systems, and uses generic natural language generation techniques in constructing the resulting textual output.

Our findings so far consist in a better understanding of the multiple aspects giving rise to variation in human route descriptions. We have unravelled the basic description components of route directions and identified the mechanisms that impact on their combination and refinement towards full-fledged semantic input structures. Further experimentation within this framework will allow us to focus on the interaction between the techniques we use for aggregation and referring expression generation: some route descriptions we produce can contain redundant information because these two processes work in a pipeline. Insights about this interaction should lead towards more general heuristics at the level of micro-planning in natural language generation.

A principled approach to route directions generation may also be valuable to two important issues in the domain of route guidance: customization to different navigation styles and inclusion of landmarks. The former consists in applying different strategies for generating referring expressions. The latter also relates to this topic, since the very notion of a landmark and the conditions that govern the choice of one over another can be viewed in terms of generating a referring expression for a decision point or path.

References

Agrawala, M. & Stolte, C. (2001), Rendering effective route maps: improving usability through generalization, *in* 'Proceedings of Siggraph 2001', Los Angeles, CA.

Burnett, G. (2000), "Turn right at the traffic lights' The requirements for landmarks in vehicle navigation systems', *The Journal of Navigation* **53**(3), 499–510.

Clark, H. & Schaefer, E. (1989), 'Contributing to discourse', *Cognitive Science* **13**, 259–294.

Dale, R. (1992), *Generating Referring Expressions*, MIT Press, Cambridge, MA.

Dale, R. & Reiter, E. (1995), 'Computational interpretations of gricean maxims in the generation of referring expressions', *Cognitive Science* **18**(2), 233–263.

Dalianis, H. (1999), 'Aggregation in natural language generation', *Journal of Computational Intelligence* **15**(4), 384–414.

Denis, M., Pazzaglia, F., C.Cornoldi & Bertolo, L. (1999), 'Spatial discourse and navigation: an analysis of route directions in the city of venice', *Applied Cognitive Psychology* **13**, 145–174.

Fraczak, L., G.Lapalme & M.Zock (1998), Automatic generation of subway directions: salience gradation as a factor for determining message and form, *in* 'Proceedings of the International Workshop on Natural Language Generation', Niagara-on-the-Lake, CA, pp. 58–67.

Geldof, S. & Dale, R. (2002), Improving route directions on mobile devices, *in* 'Proceedings of the ISCA workshop on Multi-Modal Dialogue in Mobile Environments', Kloster Irsee, Germany.

Grosz, B. & Sidner, C. (1986), 'Attention, intentions and the structure of discourse', *Computational Linguistics* **12**(3), 175–204.

Höök, K. (1991), 'An approach to a route guidance interface', Licentiate Thesis, Dept. of Computer and Systems Sciences, Stockholm University, ISSN 1101-8526.

Höppner, W. (1995), Modularity in NLG, *in* 'Proceedings of the 4th Intl. Colloquium on Cognitive Science', San Sebastian, pp. 117–121.

Kuipers, B., J. (1978), 'Modelling spatial knowledge', *Cognitive Science* **2**, 129–153.

Lovelace, K., Hegarty, M. & Montello, D. (1999), Elements of good route directions in familiar and unfamiliar environments, *in* C. Freksa & D. Mark, eds, 'Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science. Proceedings of the International Conference COSIT'99', Berlin: Springer Verlag, Stade, Germany, pp. 65–82.

Maaß, W., Baus, J. & Paul, J. (1995), Visual grounding of route descriptions in dynamic environments, *in* R. K. Srihari, ed., 'Proc. of the AAAI Fall Symposium on Computational Models for Integrating Language and Vision', Cambridge, MA.

Moulin, B. & Kettani, D. (1999), 'Route generation and description using the notions of object's influence area and spatial conceptual map', *Spatial Cognition and Computation* **1**, 227–259.

Pailhous, J. (1970), *La representation de l'Espace Urbain - l'example du chauffeur de Taxi*, Presses Universitaires de France.

Pattabhiraman, T. & Cercone, N. (1990), Selection: Salience, relevance and the coupling between domain-level tasks and text planning, *in* K. McKeown, J.Moore & S. Nirenburg, eds, 'Proceedings of the 5th International Workshop on Natural Langauge Generation', Dawson, Pennsylvania.

Raubal, M. & Winter, S. (2002), Enriching wayfinding instructions with local landmarks, *in* 'GIScience 2002', Lecture Notes in Computer Science, Springer.

Reiter, E. & Dale, R. (2000), *Building natural language generation systems*, Cambridge University Press.

Ross, T. & Burnett, G. (2001), 'Evaluating the human-machine interface to vehicle navigation systems as an example of ubiquitous computing', *Int. Journal of Human-Computer Studies* **55**, 661–674.

Sorrows, M. E. & Hirtle, S. C. (1999), The nature of landmarks for real and electronic spaces, *in* C. Freksa & D. M. Mark, eds, 'Spatial Information Theory', number 1661 *in* 'Lecture Notes in Computer Science', Springer.

Streeter, L., Vitello, D. & Wonsievicz, T. (1985), 'How to tell people where to go: comparing navigational aids', *International Journal of Man/Machine Systems* **22**(5), 549–5624.

Williams, S. (1998), Generating pitch accents in a concept-to-speech system using a knowledge base, *in* 'Proceedings of the 5th International Conference on Spoken Language Processing', Sydney, Australia.

Williams, S. & Watson, C. I. (1999), A profile of the discourse and intonational structures of route descriptions., *in* 'Proceedings of the 6th European Conference on Speech Communication and Technology, Eurospeech'99', Budapest, Hungary, pp. 1659–1662.