Generating More Natural Route Descriptions

Robert Dale, Sabine Geldof, Jean-Philippe Prost

Generating More Natural Route Descriptions

Robert Dale, Sabine Geldof and Jean-Philippe Prost
Centre for Language Technology
Macquarie University
Sydney, Australia
{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 wayfinding system. We adopt a standard pipelined natural language generation architecture, and focus in particular on the role of the generation subtasks of aggregation and referring expression generation in producing fluent output. Through examples we demonstrate that it is possible to bridge the gap between underlying representation and natural sounding descriptions. Further work along these lines will contribute both to the area of natural language generation and to the improvement of wayfinding system interfaces.
Keywords: Natural language generation, wayfinding systems, microplanning.

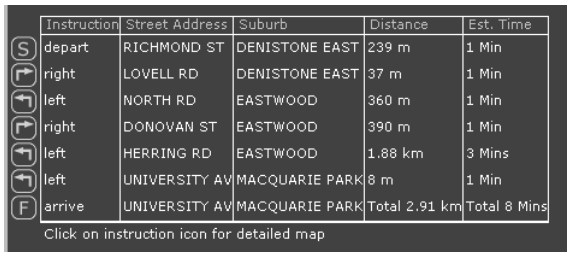
1 Introduction

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] Apart from interesting differences in the user interfaces, all these systems are similar in concept and content: the user specifies a start 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 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 in-

[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.



	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

formation into single sentences, whereas the automated systems produce what are in effect one-sentence-per-step mappings.

Of course, there is no prima facie reason why we should want an automated system to emulate what people do, especially in written output. 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 is actually an improvement on what people do. There is some evidence, however, that at least with respect to their contents, route descriptions closer to those produced by humans are prefered by users: work on the graphical display of routes, for example (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 et al., 1985; Denis et al., 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 particularly on approaches to **aggregation** and techniques for **referring expression generation**.

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 our approach in general; and Section 4 explores our use of NLG techniques for referring expression generation and aggregation, along with an example output that demonstrates the current capabilities of our system. Section 5 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) focused on the role of salience and relevance in content selection for NLG. Route description illustrate 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 derived relative importance of information onto syntactico-semantic features. 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 gen-

eration task as a straightforward mapping from the underlying data. 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 prototypical navigation styles. Finally, the route descriptions generated by Maaß and colleagues (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 hand-crafted 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 [3]. Our earlier work addressed the provision of route descriptions within a university department (Williams and 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 and Dale, 2002).

Our current work attempts 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 textual output. 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 indoor route descriptions (7 subjects $\times$ 7 routes); another corpus consists of 30 written route descriptions (10 $\times$ 3) 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, and 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, within a familiar environment. 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.

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 exam-

[3] For more information on the project see: http://www.ics.mq.edu.au/~coral.

ple, 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 what we call **arc aggregation**. Since an arc joins two junctions, the path between each two intersections along a same 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, and this is, effecively, what current systems do. 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; 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; and the realisation stage maps these sentence plans 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, 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.

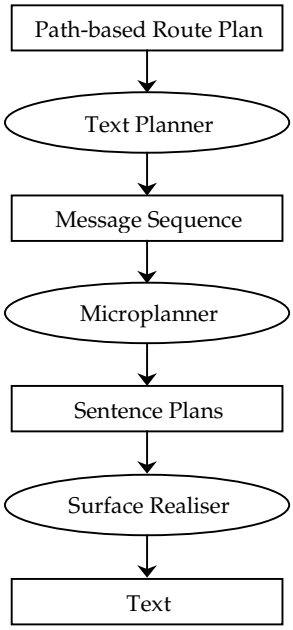


Figure 3: Coral's architecture

Our analysis of human-produced route descriptions leads 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*.

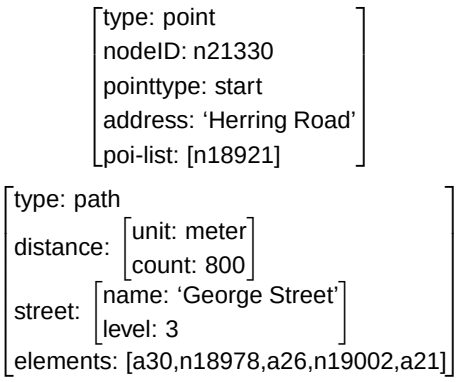


Figure 4: Example point and path messages

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 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 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 and Reiter, 1995). In abstract terms, the principles of aggregation and referring expression generation are generally considered quite domain-independent; however, these principles have to be instantiated with domain specific knowledge in order to be made workable. We describe our approach to each of these tasks below.

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 5: One message per clause

4 Applying NLG Techniques

4.1 Aggregation

Aggregation is the process of building clauses which communicate several pieces of information at once. Although the messages in our text plan could be realised one-per-clause, the result would be less than fluent, as shown in Figure 5.

Of course, there are many situations where one clause 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).
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 in one sentence, 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 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 6 shows some of the various realisations possible for the route shown in Figure 5. In future work, we aim to explore a scoring regime that ranks the various renderings.

4.2 Referring Expression Generation

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

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 6: Different aggregations

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). Although the task can be characterised in a very domain-independent manner, as in (Dale, 1992), subsequent work (Dale and Reiter, 1995) has taken 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. Our work here suggests that even this process needs to be driven by higher-level strategies which are domain-

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 7: Applying referring expression gene ation

specific.

On the basis of a first corpus analysis an the readily available GIS information, we hav identified the following properties which ca be used for **referring to junction points** :

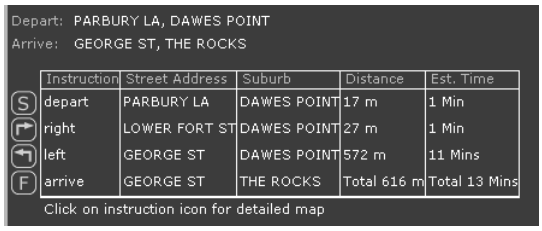
1. Use a landmark that is at, or close to, th junction.
2. Use the type of intersection (e.g. rounc about, T-junction, fork).
3. Use the name of the immediately preced- ing intersection.
4. Use the name of the intersecting street.

Thus, we use whatever information the un- derlying 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 7.

A similar range of properties is used to pro- vide 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 deter- mined by the richness of the underlying data set. In particular, to determine whether or not an entity counts as a landmark is a knowledge intensive question (Raubal and Winter, 2002).



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 8: Whereis compared to Coral generated route description

However, our corpus analysis has revealed that a more frequently used property is the intersection type (for example, whether a turn is at a roundabout, a T-junction, or an intersection with traffic lights). These data are more readily available to GIS systems. The preference ordering of the properties reflects this observation: if no obvious landmark is present, we use the intersection type.

4.3 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 8 shows a route description provided by *WhereIs*, along with the same route described by our system, making use of the aggregation and referring expression strategies described above.

5 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 takes as input GIS data currently used by a commercial system, and combines general principles

and concepts from natural language generation with domain knowledge acquired from corpora 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 conditions that govern the choice of one over another can be viewed in terms of generating a referring expression.

References

M. Agrawala and Ch.W. Stolte. 2001. Rendering effective route maps: improving usability through generalization. In *Proceedings of Siggraph 2001*, Los Angeles, CA.

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

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

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

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

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

L. Fraczak, G.Lapalme, and 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*, pages 58–67, Niagara-on-the-Lake, CA.

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

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

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

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

T. Pattabhiraman and N. Cercone. 1990. Selection: Salience, relevance and the coupling between domain-level tasks and text planning. In K. McKeown, J.Moore, and S. Nirenburg, editors, *Proceedings of the 5th International Workshop on Natural Langauge Generation*, Dawson, Pennsylvania.

M. Raubal and S. Winter. 2002. Enriching wayfinding instructions with local landmarks. In *GIScience 2002*, Lecture Notes in Computer Science. Springer.

E. Reiter and R. Dale. 2000. *Building natural language generation systems*. Cambridge University Press.

L.A. Streeter, D. Vitello, and T. Wonsievicz. 1985. How to tell people where to go: comparing navigational aids. *International Journal of Man/Machine Systems*, 22(5):549–5624.

S. Williams and C. I. Watson. 1999. A profile of the discourse and intonational structures of route descriptions. In *Proceedings of Eurospeech'99*, pages 1659–1662, Budapest, Hungary.