Generating and Evaluating Landmark-based Navigation Instructions in Virtual Environments

Amanda Cercas Curry
School of Mathematical and
Computer Sciences
Heriot-Watt University
Edinburgh
ac293@hw.ac.uk

Dimitra Gkatzia
School of Mathematical and
Computer Sciences
Heriot-Watt University
Edinburgh
d.gkatzia@hw.ac.uk

Verena Rieser
School of Mathematical and
Computer Sciences
Heriot-Watt University
Edinburgh
v.t.rieser@hw.ac.uk

Abstract

Referring to landmarks has been identified to lead to improved navigation instructions. However, a previous corpus study suggests that human "wizards" also choose to refer to street names and generate user-centric instructions. In this paper, we conduct a task-based evaluation of two systems reflecting the wizards' behaviours and compare them against an improved version of previous landmark-based systems, which resorts to user-centric descriptions if the landmark is estimated to be invisible. We use the GRUVE virtual interactive environment for evaluation. We find that the improved system, which takes visibility into account, outperforms the corpus-based wizard strategies, however not significantly. We also show a significant effect of prior user knowledge, which suggests the usefulness of a user modelling approach.

1 Introduction

The task of generating successful navigation instructions has recently attracted increased attention from the dialogue and Natural Language Generation (NLG) communities, e.g. (Byron et al., 2007; Dethlefs and Cuayáhuitl, 2011; Janarthanam et al., 2012; Dräger and Koller, 2012) etc. Previous research suggests that landmark-based route instructions (e.g. "*Walk towards the Castle*") are in general preferable because they are easy to understand, e.g. (Millonig and Schechtner, 2007; Chan et al., 2012; Elias and Brenner, 2004; Hansen et al., 2006; Dräger and Koller, 2012). However, landmarks might not always be visible to the user. A recent corpus study by Cercas and Rieser (2014) on the MapTask and two Wizard-of-

Oz corpora, Spacebook1 and Spacebook2,[1] empirically investigated the type of reference objects human instruction givers tend to choose under different viewpoints. It was found that human "wizards" do not always generate instructions based on landmarks, but also choose to refer to street names or generate user-centric instructions, such as "*Continue straight*".

This paper compares three alternative generation strategies for choosing possible reference objects: one system reflecting an improved version of a landmark-based policy, which will resort to a user-centric description if the landmark is not visible; and two systems reflecting the wizards' behaviours in Spacebook1 and Spacebook2. We hypothesise the first system will outperform the other two in terms of human-likeness and naturalness, as defined in Section 3. We use the GRUVE (Giving Route Instructions in Uncertain Virtual Environments) system (Janarthanam et al., 2012) to evaluate these alternatives.

2 Methodology

We designed two corpus-based strategies (System **B**, **C**) and one rule-based system based on a heuristic landmark selection algorithm (**A**). Also see examples in Table 1. Strategies for systems B and C aim to emulate the wizards' strategies dependent on different viewpoints: System B uses data from Spacebook1, where the wizard follows the user around, and thus, shares the viewpoint of the user. System C uses data from Spacebook2, where the wizard follows the user remotely on GoogleMaps via GPS tracking, and thus, street names are visible to the wizard, but only the approximate location is known.

- **System A: Landmark and User-centric** strategy reflects an improved version over

---

[1] The Spacebook data is freely available here: http://www.macs.hw.ac.uk/ilabarchive/ spacebook/login.php

Systems	Output
System A	"Keep going straight towards Farmfoods Ltd." (***landmark***)
System B	"Continue straight" (***user-centric***)
System C	"Keep walking along Nicholson Street"(***street name***)

Table 1: Example of user view on GoogleStreetMaps (left) and system outputs (right).

previous work, in that it mainly produces landmark-based instructions, but resorts to user-centric instructions when no landmarks are available (also see our landmark selection algorithm as described below). We also call this the *visibility* strategy.

- **System B: Spacebook1-based** strategy produces instructions using street names, landmarks and user-centric references in the same proportions as the wizards in Spacebook1. We also call this the *shared viewpoint* strategy.

- **System C: Spacebook2-based** strategy produces landmark-based and street name-based instructions as in Spacebook2. A landmark or a street name is selected based on a threshold on the landmark's salience (determined through trial and error). We also call this the *birds-eye* strategy.

All three strategies select landmarks based on landmark salience, following Götze and Boye (2013), using a heuristic based on (also see Figure 1): the distance between the landmark and the user, the distance between the user and the target, the angle formed by these two lines, the type of landmark and whether the landmark has a name. We adjusted this heuristic to match our system.
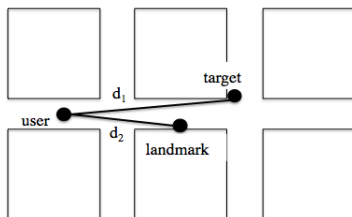

Figure 1: Spatial features used by landmark heuristic.

Note that GRUVE only provides information on static landmarks, e.g. shops, restaurants, banks, etc., available from GoogleMaps and Open Street Map. It does not identify moving objects, such as cars, as potential landmarks. In current work (Gkatzia et al., 2015), we investigate how to generate landmarks based on noisy output from object recognition systems.

3 Evaluation

3.1 Experimental Setup

We used the GRUVE virtual environment for evaluation. GRUVE uses Google StreetView to simulate instruction giving and following in an interactive, virtual environment, also see Table 1. We recruited 16 subjects, with an even split amongst males and females and age ranges between 20 and 56. Six users were not native English speakers.

Before the experiment, users were asked about their previous experience. After the experiment we asked them to rate all systems on a 4-point Likert scale regarding human-likeness and naturalness (where 1 was "Agree" and 4 was "Disagree"). *Human-likeness* is defined as an instruction that could have been produced by a human. *Naturalness* is defined as being easily understood by a human. The order of systems was randomised.

4 Results

In total we gathered 1071 navigation instructions. For evaluation, we compared a number of objective and subjective measures. The results are summarised below (also see Table 2) :

- **Task Completion:** The overall task completion rate (binary encoding) was 68.1%. System A was slightly more successful with a task completion rate of 80% compared to 62.5% for systems B and C, but this difference was

not statistically significant (χ^2 test, p=.574).[2] However, a planned comparison for task completion time showed that users take longer when using System A compared to the two other systems, but again the difference between the systems was not found to be statistically significant (Mann-Whitney U-Test[3], p=.739 for System A vs. B, p=.283 for A vs. C, and p=.159 for C vs. B).

- **Human-likeness and Naturalness:** Furthermore, users tend to rate System A higher for human-likeness (χ^2, p=.185) and for naturalness (χ^2, p=.093) than system B and C, but again the difference was not statistically significant. We also observed the following mixed effects: Users tend to report the system to be more natural and human-like if they had managed to complete the task (χ^2, p=.002 and p=.000, respectively). This could be a reflection of user frustration, where users report the system to be less human-like if they are dissatisfied with the instructions provided.

- **Familiarity Effects:** We also observed the following effects of prior user knowledge: Ten users reported they were familiar with the location before the experiment. These users were significantly more likely to report that the instructions were accurate and of the correct length (χ^2, p=.037).

In addition, users familiar with Google StreetView found the instructions to be significantly easier to follow (χ^2, p=.003), more accurate and more natural and human (χ^2, p=.021) compared to those with little or no experience. Only two users reported having no experience with Google StreetView, eight reported having a little experience and six reported being very familiar with it. These familiarity effects of prior knowledge suggest a user-modelling approach.

4.1 Discussion

The data shows an indication that System A is able to better support task completion, while being perceived more natural than Systems B and C. However, this trend is not significant. Table 3 shows an analysis of how often each system chose

Measure	objective		subjective	
Metric	compl. rate	compl. time	natural-ness	human-likeness
Scale	binary	seconds	4-point Likert	
A	0.80	900.06	1.0	1.0
B	0.63	799.75	2.0	2.0
C	0.63	883.31	1.0	2.0

Table 2: Average results for objective (mean) and subjective (mode) measures.

a reference object in our experiments. System A produces significantly more landmark-based descriptions than B and C (Mann-Whitney U-test for nonparametric data, p=.003 and p=.041 respectively). These results seem to confirm claims by prior work that landmark-based route instructions are in general preferable. In future work, we will compare our improved version, which also uses user-centric descriptions, with a vanilla landmark-based strategy in order to determine the added value of taking visibility into account.

System	landmark	user-centric	street name
System A	66.22	33.78	0
System B	61.54	23.50	14.96
System C	56.05	0	43.95

Table 3: Frequencies of reference objects chosen by each system.

4.2 User Comments and Qualitative Data

Users were asked to provide some additional comments at the end of the questionnaire. Overall, the subjects reported liking the use of landmarks like shops and restaurants. Users not familiar with the location found this less useful, particularly when the system referred to buildings that were not labelled on StreetView. For example, the location natives can easily identify the Surgeon's Hall in Edinburgh, but for those who are unfamiliar with the neighbourhood, the building is not so easily identifiable. Users also reported liking user-centric instructions as they are simple and concise, such as "Turn left". Some users reported they would like to know how far away they are from their destination. A few users also commented that the instructions could be repetitive along long routes.

Users reported the system used landmarks that

---

[2]Although the percentage difference seems large it is equivalent to only two subjects.

[3]We used the non-paramentric version of a t-test since the data was not normally distributed.

were not visible, whether because they were too far away or they were hidden by another building. There was no difference in the number of users reporting this for each system. This suggests the landmark-selection heuristic will require further adjustments, e.g adjusting the weights or limiting the search area. Users that were familiar with the location reported that although some of the landmarks presented were not visible, they were still helpful as the users knew where these landmarks were and could make their way to them. The use of landmarks that are not necessarily visible but are known to the instruction follower is common amongst human direction givers, using these landmarks as a starting point for further directions (Golledge, 1999). Again, these findings suggest the usefulness of a user modelling approach to landmark selection.

5 Conclusions and Future Work

This paper presented a task-driven evaluation of context-adaptive navigation instructions based on Wizard-of-Oz data. We found that a heuristic-based system, which uses landmarks and user-centric instructions dependent on estimated visibility, outperforms two corpus-based systems in terms of naturalness and task completion, however, these results were not significant. In future work, we hope to recruit more subjects in order to show statistical significance of this trend. Our results also show that there are significant familiarity effects based on prior user knowledge. This suggests a user modelling approach will be useful when it comes generating navigation instructions, e.g. following previous work on user modelling for NLG in interactive systems (Janarthanam and Lemon, 2014; Dethlefs et al., 2014). Finally, we hope to repeat this experiment under real-world conditions, rather than in a virtual setup in order to eliminate artefacts, such as the influence of technical problems.

References

Donna Byron, Alexander Koller, Jon Oberlander, Laura Stoia, and Kristina Striegnitz. 2007. Generating Instructions in Virtual Environments (GIVE): A Challenge and an Evaluation Testbed for NLG. In *Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation*.

Edgar Chan, Oliver Baumann, Mark a Bellgrove, and Jason B Mattingley. 2012. From objects to landmarks: the function of visual location information in spatial navigation. *Frontiers in psychology*, 3:304, January.

Nina Dethlefs and Heriberto Cuayáhuitl. 2011. Combining hierarchical reinforcement learning and bayesian networks for natural language generation in situated dialogue. In *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG)*.

Nina Dethlefs, Heriberto Cuayáhuitl, Helen Hastie, Verena Rieser, and Oliver Lemon. 2014. Cluster-based prediction of user ratings for stylistic surface realisation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 702–711. Association for Computational Linguistics.

Markus Dräger and Alexander Koller. 2012. Generation of landmark-based navigation instructions from open-source data. In *Proceedings of the Thirteenth Conference of the European Chapter of the ACL (EACL)*, Avignon.

Birgit Elias and Claus Brenner. 2004. Automatic Generation and Application of Landmarks in Navigation Data Sets. *Developments in Spatial Data Handling*, pages 469–480.

Dimitra Gkatzia, Verena Rieser, Phil Bartie, and William Mackaness. 2015. From the virtual to the real world: Referring to objects in spatial real-world images. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Lisbon. Forthcoming.

Reginald G Golledge. 1999. *Wayfinding behavior: cognitive mapping and other spatial processes*, volume 41.

Jana Götze and Johan Boye. 2013. Deriving Salience Models from Human Route Directions. In *Proceedings of the IWCS 2013 Workshop on Computational Models of Spatial Language Interpretation and Generation (CoSLI-3)*, pages 7–12, Potsdam, Germany. Association for Computational Linguistics.

Stefan Hansen, Kai-florian Richter, and Alexander Klippel. 2006. Landmarks in OpenLS - A Data Structure for Cognitive Ergonomic Route Directions. In *GIScience*, pages 128–144. Springer-Verlag, Berlin Heidelberg.

Srinivasan Janarthanam and Oliver Lemon. 2014. Adaptive Generation in Dialogue Systems Using Dynamic User Modeling. *Computational Linguistics*, 40(4):883–920.

Srinivasan Janarthanam, Oliver Lemon, and Xingkun Liu. 2012. A web-based evaluation framework for spatial instruction-giving systems. In *50th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Alexandra Millonig and Katja Schechtner. 2007. Developing Landmark-Based Pedestrian-Navigation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 8(1):43–49, March.

Verena Rieser and Amanda Cercas Curry. 2014. Towards Generating Route Instructions Uncer Uncertainty: A Corpus Study. In *SemDial'14*, Edinburgh.