Moving targets: human references to unstable landmarks

Adriana Baltaretu
TiCC
Tilburg Univeristy
a.a.baltaretu@uvt.nl

Emiel Krahmer
TiCC
Tilburg University
e.j.krahmer@uvt.nl

Alfons Maes
TiCC
Tilburg University
maes@uvt.nl

Abstract

In the present study, we investigate if speakers refer to moving entities in route directions (RDs) and how listeners evaluate these references. There is a general agreement that landmarks should be perceptually salient and stable objects. Animated movement attracts visual attention, making entities salient. We ask speakers to watch videos of crossroads and give RDs to listeners, who in turn have to choose a street on which to continue (Experiment 1) or choose the best instruction among three RDs (Experiment 2). Our results show that speakers mention moving entities, especially when their movement is informative for the navigation task (Experiment 1). Listeners understand and use moving landmarks (Experiment 1), yet appreciate stable landmarks more (Experiment 2).

1 Introduction

One of the applications of Natural Language Generation (Reiter et al., 2000) is the automatic generation of route directions, e.g., Roth and Frank (2009); Dale et al., (2005). These instruction typically involve Referring Expressions Generation (REG), (Krahmer and Van Deemter, 2012), for the generation of references to landmarks. Until recently, REG for landmarks and studies on human navigation have focussed exclusively on references to stable entities; in fact, to the best of our knowledge moving targets have never been studied before. Emerging technology (e.g., Google Glass) allows systems to include all relevant visual information in RDs. This raises the question whether references to moving landmarks actually occur.

With support from wearable technology, navigation systems could become spatially aware. For example, navigation systems could produce more human-like instructions by making use of the visual information captured by devices that incorporate video cameras. A navigation system could ground actions in space by referring to both stable ("the tall building") and moving ("the cyclist going left") information. However, we know little about how the dynamic character of the environment influences referential behaviour. We address this issue by analysing if moving entities in the environment affect route direction (RD) production and evaluation.

RDs are instructions guiding a user on how to incrementally go from one location to another (Richter and Klippel, 2005). These instructions contain numerous references to entities in the environment (henceforth landmarks). Traditionally, landmarks have been defined as route-relevant stable entities (such as buildings) that function as points of reference (Allen, 2000). One likely reason for which unstable entities are underrepresented in most standard navigation studies, is that the set-up of these studies often implies some kind of (temporal and / or spatial) asymmetry between the speaker and addressee perspectives, which makes moving entities unreliable reference points. For example, instructions are communicated over distance (e.g., telephone) or asynchronously (e.g., after travelling the route or on the basis of maps). In contrast, in this study we synchronize the two perspectives and focus on in-situ turn-by-turn RDs, where the request for assistance is formulated and followed on the spot. While having access to a shared dynamic environment, speakers can refer to any entity that could improve the instruction. We analyse if speakers refer to moving entities in RDs and asses listeners preference for such references.

Among other aspects, perceptual salience has been theorized to be an important quality of landmarks (Sorrows and Hirtle, 1999) and movement is known to contribute to the perceptual salience

Proceedings of the 15th European Workshop on Natural Language Generation (ENLG), pages 48–51,
Brighton, September 2015.

of objects. Movement is processed effortlessly by the visual system and attracts attention when informative about the location of a target (Hillstrom and Yantis, 1994). In this study, we focus on animated motion. In general, animate entities influence visual attention and reference production (Downing et al., 2004); (Prat-Sala and Branigan, 2000). Moreover, animated movement in itself (automatically) captures visual attention (Pratt et al., 2010). We hypothesize that if entities grab attention, then speakers would mention them and that listeners would prefer these RDs positively, especially when their motion is task-informative.

2 Experiment 1 - Production

2.1 Methods

2.1.1 Participants

56 dyads of native Dutch-speaking students of Tilburg University (50 women, 21.2 mean age) participated in exchange for partial course credits. Participants were randomly assigned to the speaker role (35 women). All participants gave consent to the use of their data.

2.1.2 Materials

144 street view HD videos were recorded in 72 intersections of Rotterdam. The experimental videos depicted 36 low traffic, +- shaped intersections. Each intersection was recorded three times illustrating a different movement manipulation (see Figure1): (a) no pedestrians / cyclists moving in the intersection (no movement condition (NM), 36 videos); (b) a person walking / cycling towards the intersection (irrelevant movement condition (IM), 36 videos); (c) the same person recorded some seconds later, while taking a turn in the required direction (relevant movement condition (RM), 36 videos). The people recorded were naive pedestrians casually walking / cycling down the street, without paying attention to the camera. In addition, each intersection had other stable object that could be referred to. The filler videos (36 videos) depict a different set of crowded and complex shaped intersections. In addition, two paper booklets with line drawing maps of the intersections were prepared (the speaker booklet included an arrow showing the direction to be taken).

2.1.3 Procedure

The speaker's task was to provide route instructions based on the map and on the video. The

Figure 1: Experimental trials: an intersection with no movement, with a cyclist going towards the intersection, and with a cyclist taking a turn.

listener had to mark in his booklet the indicated street. The listener was allowed to ask questions only if the instructions were unclear. Each video lasted about three seconds and was projected on a white wall (size: 170 x 120 cm). The videos could not be replayed, but the last frame was displayed until the listener announced he is finished. Pointing was discouraged by installing a screen between participants up to shoulder level. Each intersection was shown only once to each dyad. Participants were randomly assigned to one of the three presentation lists. The task started with two warm-up trials followed by 72 video trials (36 experimental trials). There were no time constraints.

2.1.4 Design and statistical analysis

This study had Movement Type (levels: no movement, irrelevant movement, relevant movement) as within participants factor and Presentation List (levels: 1, 2 and 3) as between participants factor. We analysed the type of landmark mentioned by the producer in the first instruction (moving man / stable objects) using logit mixed model analysis with Movement Type and Presentation List as fixed factors; participants and item pictures as ran-

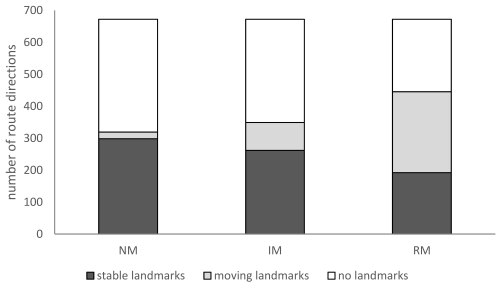

Figure 2: For each condition, number of route directions with different types of landmarks

dom factors; p–values were estimated via parametric bootstrapping. The factors were centred to reduce colinearity. The first converging model is reported. This included random intercepts for participants and videos and random slope for Movement Type in videos. Only significant results are reported. Next we analysed if moving entities are mentioned together with stable ones, clarification questions and listener error rates.

3 Results

2016 RDs (56 speakers * 36 videos) were produced in this experiment. Across the three conditions, participants mentioned both stable (N = 752) and moving entities (N = 361) (see Figure2).

In the NM condition, participants rarely referred to moving people (M = 0.03). Statistical analysis was performed only on the data from the IM and RM conditions. There was no significant effect of Presentation List ($p > .05$). There was a main effect of Movement Type ($\beta = 1.913$; SE = 0.27; $p < .001$). In the RM condition participants referred more often to the moving person taking a turn (M = 0.37), than in the IM condition (M = 0.13).

Few cases (0.02 %) of RDs included both the moving and the stable landmarks (3 cases in NM; 18 cases in IM, and 22 cases in RM).

In general, the task was easy: there were 80 questions asked by listeners and no signals of major communication breakdowns. The questions were asked when the speaker did not refer to landmarks in his initial instruction (55%), when the speaker referred to a stable landmark (31.25%), when the speaker referred to a moving landmark (13.75%). The most frequent type of question was the one in which listeners introduced (new) stable landmarks.

When choosing the street, listeners made few errors (11 cases of incorrectly marked streets and 8 cases in which the first choice was corrected).

4 Experiment 2 - RD evaluation

4.1 Participants

32 native Dutch-speaking students of Tilburg University (12 women, 20.7 mean age) participated in exchange for partial course credits. All participants gave consent to the use of their data.

4.2 Materials

The materials consisted of 72 videos (the experimental trials from the IM and RM condition used in Experiment 1). Overlaid on the videos, a semi-transparent red arrow depicted the route and the direction to be followed.

Based on the production data, for each video a set of three route directions was created as follows: a route direction without landmarks (e.g., turn left); a route direction with a stable landmark (e.g., turn left at Hema); a route direction with a moving landmark (e.g., turn left where that man / woman / cyclist is going). The stable landmarks used in these RDs were the most often mentioned objects in Experiment 1. The moving landmarks were referred to as *the man / woman / cyclist*.

4.3 Procedure

The participants' task was to watch the videos, read the RDs and choose the one that they liked most. Participants saw 36 trials as follows: first a fixation cross was displayed for 500ms, followed by the video and the three instructions placed below the video. The position on screen of the RDs was counterbalanced. Each intersection was shown only once, and participants were randomly assigned to one of the two presentation lists.

4.4 Design and statistical analysis

This study had Movement Type (levels: irrelevant movement, relevant movement) as within participants factor and Presentation List (levels: 1, 2) as between participants factor. The dependent variable was the type of RD chosen. Statistical analysis was performed as in Experiment 1. The model had Movement Type and Presentation List as fixed factors; subjects and videos as random factors.

5 Results

Out of 1152 cases (36 scenes x 32 participants), RDs with landmarks were chosen more often

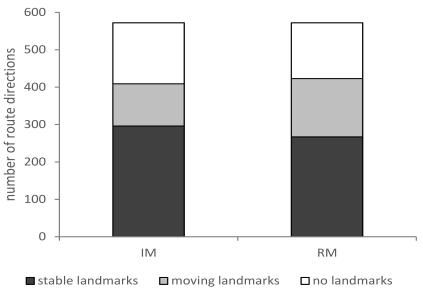

Figure 3: For each condition, types of landmarks chosen

(73% of the cases) than RDs without landmarks (see Figure3). To see if movement influenced the choice for a specific type of landmark, the statistical analysis was done on a data set consisting of the RDs with landmarks.

In general, participants chose more often stable landmarks (77.06% of the cases) than moving landmarks. There was a main effect of Movement Type (β = 1.211; SE = .265; p $<$ .001). This model included random intercepts for subjects and for videos.

For videos depicting irrelevant movement, participants chose more often instructions with stable landmarks (M = 0.85) than with moving landmarks (M = 0.15). For videos depicting relevant movement, the same pattern is observed though there was a slight increase in the preference for moving landmarks (stable landmarks M = 0.75; moving landmarks M = 0.25). There was no significant effect of Presentation List (p $>$ .05).

6 Conclusions

In conclusion, human speakers do use references to moving landmarks. Speakers referred to moving objects especially when their movement was informative. Listeners did not encounter difficulties understanding these instructions. Yet, they preferred instructions with stable landmarks. In the light of technological developments our results highlight that navigation systems should not only add landmarks to the instructions, but also adjust the type of landmarks. Speakers naturally refer to items with a relevant movement trajectory. Further work is needed to investigate if moving entities were mentioned because they were more salient than their stable counterparts and second, to validate the efficiency of such RDs for listeners. In future research, we hope to address the question how current REG algorithms can be adapted to generate references to moving targets.

References

Gary L Allen. 2000. Principles and practices for communicating route knowledge. *Applied Cognitive Psychology*, 14(4):333–359.

Robert Dale, Sabine Geldof, and Jean-Philippe Prost. 2005. Using natural language generation in automatic route. *Journal of Research and practice in Information Technology*, 37(1):89.

Paul E Downing, David Bray, Jack Rogers, and Claire Childs. 2004. Bodies capture attention when nothing is expected. *Cognition*, 93(1):27–38.

Anne P Hillstrom and Steven Yantis. 1994. Visual motion and attentional capture. *Perception & Psychophysics*, 55(4):399–411.

Emiel Krahmer and Kees Van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.

Mercè Prat-Sala and Holly P Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in english and spanish. *Journal of Memory and Language*, 42(2):168–182.

Jay Pratt, Petre V Radulescu, Ruo Mu Guo, and Richard A Abrams. 2010. Its alive! animate motion captures visual attention. *Psychological Science*, 21(11):1724–1730.

Ehud Reiter, Robert Dale, and Zhiwei Feng. 2000. *Building natural language generation systems*, volume 33. Cambridge University Press, Cambridge.

Kai-Florian Richter and Alexander Klippel. 2005. A model for context-specific route directions. In *Spatial cognition IV. Reasoning, action, interaction*, pages 58–78. Springer, Berlin.

Michael Roth and Anette Frank. 2009. A NLG-based application for walking directions. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics*, pages 37–40, Singapore.

Molly E Sorrows and Stephen C Hirtle. 1999. The nature of landmarks for real and electronic spaces. In *Spatial information theory. Cognitive and computational foundations of geographic information science*, pages 37–50. Springer, Berlin.

Acknowledgements

The first author received financial support from the Netherlands Organization for Scientific Research, via NWO Promoties in de Geesteswetenschappen (322-89-008), which is greatly acknowledged. Partial results of this study have been presented in CogSci 2015.