2018-06-23

# Perception & Perspective: An Analysis of Discourse and Situational Factors in Reference Frame Selection

Robert Ross
*Technological University Dublin*, robert.ross@tudublin.ie

Kavita E. Thomas
*University of Gavle*, kavita.thomas@hig.se



## Recommended Citation
Ross, R. & Thomas, K.E. (2018). Perception & Perspective: An Analysis of Discourse and Situational Factors in Reference Frame Selection. *Workshop on Dialogue and Perception 2018*, Gothenburg, Sweden, 14-15 June. doi:10.21427/q7r2-rs61

# Perception & Perspective: An Analysis of Discourse and Situational Factors in Reference Frame Selection

**Robert Ross**
School of Computer Science
Dublin Institute of Technology
Ireland
robert.ross@dit.ie

**Kavita E. Thomas**
Department of Humanities
University of Gävle
Sweden
email@domain

## Abstract

To integrate perception into dialogue, it is necessary to bind spatial language descriptions to reference frame use. To this end, we present an analysis of discourse and situational factors that may influence reference frame choice in dialogues. We show that factors including spatial orientation, task, self and other alignment, and dyad have an influence on reference frame use. We further show that a computational model to estimate reference frame based on these features provides results greater than both random and greedy reference frame selection strategies.

## 1 Introduction

Perception, unlike static spatial modeling, is anchored with respect to a spatial perspective. Agents perceive their environment from a given perspective, and the spatial language they use to construe their environment is often constructed with respect to a specific perspective or reference frame. Reference frame choices are fortunately relatively simple, but our understanding of how to use reference frames is particular contexts is a very real challenge.

In previously published work (**?**) we briefly looked at the issue of spatial elements in influencing perspective choice in a human-human navigation corpus. In this paper we take our previous analysis further by analyzing a wider range of predictive factors more closely. We begin in Section 2 by providing a brief background on perspective selection. Then in Section 3 we review the details of our data collection. Section 4 provides a summary of our analysis, before we present conclusions in Section 5.

## 2 Perspective & Reference Frame

Levinson (1996) describes three reference frames that are used for static relation description, i.e., the intrinsic, relative, and absolute reference frames. For the case of dynamic prepositions as used in action descriptions similar to those analyzed in this paper, two other reference frames are proposed (Klatzky, 1998). The *route*, or *egocentric* perspective, tied to the intrinsic reference frame, is defined by a trajectory created by the direction of movement of an object. *Survey*, or *allocentric* perspectives on the other hand are related to absolute reference frames in that they are defined by virtue of global rather then mover properties. These various perspective uses have been discussed and illustrated in detail elsewhere (Tenbrink et al., 2010).

The diversity of perspective and reference system choices for a given situation introduces significant complication in mapping between descriptive language and space. Unfortunately speakers are not consistent with regard to perspective use within a single task. For example (Taylor and Tversky, 1996) found that despite a perceived wisdom that coherence maxims would favor the retention of a single perspective throughout a task, speakers frequently switched between so-called survey and route perspectives.

Taylor and Tversky's experiments, like most cognitive and linguistic experiments on verbal route instructions, focused on the case of monologic instructions provided by route givers to route followers prior to the route follower's movement. In terms of computer-mediated communication focusing on spatial tasks between humans, Lawson et al. (2008)'s findings suggest considerable flexibility in perspective choice in dialogue. Such flexibility is reflected in the findings of (Goschler et al., 2008) who found considerable mixing of survey and route perspective. More recently, Thomas and Andonova (2012) show that speakers' perceptions of addressees' level of understanding based on addressees' clarification requests can affect speakers' perspective choice in dialogue.

## 3 Data Collection & Annotation

To examine the relationship between perspective use and contextual factors, we conducted an analysis based on an existing human-human corpus of action oriented dialogues (Ross and Thomas, 2010; Tenbrink et al., 2010). Here we briefly summarize key points with respect to the corpus and the subsequent analysis that we performed.

The corpus consists of 15 recorded dyads where each dyad performed a route instruction task up to 11 times. In all 15 dyads participants played the same role (either route giver or follower) throughout the 11 trials they participated in. In each trial, the route giver had to direct the route follower to the goal which only the route giver could see. Participants could neither see nor hear one another and both participants interacted via chat boxes below the indoor map of a schematized office environment shown on their screens. During a given dyad, both participants saw the same map except that only the giver's map showed the goal location. Both participants could see the avatar which the route follower moved via joystick. Individual tasks were rnadomised between dyads to minimize the influence of learning effects.

The resultant corpus consisted of 1108 utterances, of which the majority (50.2%) lack perspective, 31.7% have route (i.e., egocentric) perspective, 7.5% have survey (i.e., allocentric) perspective, 1.0% have mixed perspective, and 8.1% have conflated perspective (i.e., orientation of the avatar was facing up, so descriptions in route and survey perspectives were indistinguishable). The corpus is unbalanced in terms of speaker participation and initiative, with 88.5% of utterances spoken by the route giver and 11.5% spoken by the follower.

Based on our analysis of the existing literature, we hypothesized that 5 different factors would have an effect on perspective choice in an interaction. Firstly, *Orientation and Turn Direction* play a role, as relatively more survey perspective use should be produced by speakers when orientation is facing down and a movement with respect to the horizontal axis is under discussion. Secondly, *Dialogue Acts* influence perspective, as backward-looking signals of non-understanding by the route follower (i.e., not understanding the previous route instruction) should result in relatively more route perspective use in the next route giver turn, while forward-looking information requests by the route follower should result in relatively more survey perspective use in the next route giver turn. Thirdly, *Resultant Action* affects perspective use since Incorrect, i.e., misunderstood, movements by the route follower should result in relatively more route perspective use in the next route giver turn, while correct movements should result in the maintenance of the current perspective. Fourthly, *Alignment* affects perspective choice, since a weak effect for same- and cross-speaker alignment across turns has been found by Watson et al. (2004) and (Vorwerg, 2009). Finally,*Individual Differences* play a role, as participants may well differ in their perspective preferences; thus we expect significant differences across dyads in perspective use.

Based on these hypothesised factors, the corpus was annotated for a range of specific features. Perspective was coded as one of six types: route, survey, mixed, conflated, unclear or without. Dialogue Act was coded using a simplified version of the DAMSL annotation scheme (Allen and Core, 1997) which only allowed exclusively forward or backward looking acts to hold, not both. Orientation was manually annotated for the avatar when the interlocutor began typing the utterance into a four level category equivalent to up, down, left, right from a survey perspective. In addition to annotating orientation, the *intended direction* of a given turn was also annotated with a four level factor corresponding to up, down, left, right from a survey perspective.

Likewise, physical actions made by the avatar were annotated and aligned with the utterance which either immediately precedes with or overlaps with its beginning. Annotators also noted what the actions were (e.g., turn-left, turn-right, go-straight, turn-around, stop, etc.) and whether they followed the preceding instruction, followed an earlier instruction, misinterpreted the preceding instruction, were made on the route follower's own initiative as an "offer", i.e., guessing the direction to move in, or were moves made to correct an earlier incorrect move following the route giver's correction.

Part of the data-set was coded by a second annotator to assess the reliability of annotation. Co-

| Model | Type | Predictors | Accuracy | $\kappa$ |
|---|---|---|---|---|
| 1 | RF | Model 1 | 75.7 | 0.49 |
| 2 | RF | Model 2 | 81.2 | 0.62 |
| 3 | RF | Model **3** | 76.0 | 0.47 |

Table 1: Classification Results. Model 1 = Ori*Dir*DADir+PPSSST+PPSSAT +PPOS+Role+TN; Model 2 = Ori*Dir+Dyad; Model 3 = Ori*Dir+PPSSST. Note PPSS = Previous Perspective Same Speaker, PPOS = Previous Perspective Other Speaker; PPSSST = Previous Perspective Same Speaker Same Turn; PPSSAT = Previous Perspective Same Speaker Across Turn; TN = turn number

hen's Kappa scores of 0.77, 0.77, 0.86, 0.57, and 0.77 were found for the features dialogue act, perspective, orientation, instruction direction and avatar action respectively.

## 4 Results & Discussion

Using the annotated data, a series of classifiers were built to predict perspective use based on the features outlined in the previous section. The classifier was based on a RandomForest which is an ensemble model that is well recognized at providing state of the art results even for small datasets such as our own (Kelleher et al., 2015).

We took the corpus and first reduced it to all utterances that had an associated perspective. From this set we eliminated all cases of mixed and unclear perspectives, resulting in a data set consisting of 547 utterances. Of these perspective indicating utterances, 353 (64.54%) had a route perspective, 90 (16.45%) had a survey perspective and 104 (19.01%) had a conflated perspective.

A number of classifier variants using different features were trained through 10-fold cross validation. In all over 30 different variants were considered. Table 1 shows accuracy and Kappa scores calculated from a model using all features and the best performing model along with one variant on that model. The highest scoring model found is a function of the dyad and hence indicates an interdyad variability in perspective choice as predicted by the chi-square test results. Eliminating dyad as a predictor variable, orientation and intended direction together with previous perspective of the same speaker gave the model with the highest useful predictive power.

While the results were encouraging, they do clearly leave room for improvement. On that basis, in the following we provide a more fine grained analysis of the influence of individual factors on perspective choice.

We expect both orientation of the route follower and instruction direction to have a significant influence on perspective choice. Looking at these factors, we found that both factors were significant predictors of perspective use. Specifically, a chi-square test for independence showed that a null hypothesis assuming independence of perspective should be rejected at the 95% confidence threshold for orientation ($\chi^2(6, N = 547) = 194.86, p < 0.001$). Similarly we found that independence of perspective and orientation direction should also be rejected at the 95% confidence threshold ($\chi^2(8, N = 547) = 81.52, p < 0.001$).

With respect to interpersonal issues, we first examined variation in perspective use with respect to participant role and the dialogue act associated with the utterance. With respect to participant role, the use of perspective-carrying utterances was almost exclusively seen in the route giver's language, with only 14 out of 533 perspective using utterances by the route follower (2.6%). No significant difference in proportional use of route versus survey perspective was seen across the two roles (i.e., 21.4% survey perspective use was by the route follower and 20.2% use of survey perspective was by the route giver); however, given the small amount of route follower perspective use in the corpus, this is only a tentative claim. With respect to dialogue act use, we found no significant difference in perspective use between backward-looking signal non-understanding acts (e.g., "huh?") or forward looking information requests. Thus our predictions regarding the influence of specific dialogue acts on perspective choice do not hold here at least for these particular dialogue acts. However, in a follow-up analysis we categorised all task utterances as either forward-looking or backward-looking dialogue acts. Analysis of these dialogue acts showed that in this case there was a significant though weak influence on perspective choice by dialogue act direction ($\chi^2(2, N = 547) = 8.949, p < 0.05$).

With respect to avatar movements' correctness we found that acceptances and offers combined resulted in 77% route, 21% survey and 2% mixed perspective use in route giver responses, while wrong moves (where participants misunderstood instructions) resulted in 80% route use and 20% survey use. Here we had sparse data for wrong moves, with only 8 cases of route and 2 of survey use. Fisher's Exact test gave a two-sided p-value of 1.000 and Pearson's Chi-Square test had a p value of 0.901 with a Chi-Square value of 0.207. What this shows is that the subsequent route giver utterance, which is usually a response to the wrong action, does not seem to involve switching of perspective, unlike what would be expected from the findings of Thomas and Andonova (2012). However, This may be because perspective was often initially ambiguous in these cases and caused the incorrect, misinterpreted moves, so route givers often used devices other than perspective to clarify their instructions (e.g., "opposite room", or "other way"). Alternatively they would make perspective explicit, which did not necessarily involve indicating spatial direction again (e.g., "from my perspective in the chair"), and so would have been classed as lacking perspective, as only directions indicating perspective were considered here.

Alignment of perspective choice is another feature we hypothesised would play a role in our data. As indicated earlier, we annotated the data to note whether perspective shifted either with respect to the speaker's perspective use in the same turn, or with respect to perspective use of the same speaker with respect to the previous turn. Speakers were found to align with their previously used perspective in the same turn ($\chi^2(12, N = 547) = 31.62, p < 0.01$). The Chi-Square test applied to alignment across speakers was however not significant (p=0.309; Chi-square=9.406).

## 5 Summary of Findings and Limitations

This work quantified the influence of a number of features on perspective choice which should be accounted for in computational models that bind perception and language in dialogue. However, our overall classifier results leave considerable grounds for improvement. Further analysis of our results demonstrates that imbalance due to a lack of survey targets in the training data may lead to poor performance. In future work we hope to overcome this limitation through further data collection and using upsampling to provide a more balanced dataset.

**Acknowledgement.** Robert Ross wishes to acknowledge the support of the ADAPT Research Centre. The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Funds.

## References

James Allen and Mark Core. 1997. Draft of DAMSL: Dialog act markup in several layers. Unpublished manuscript.

Juliana Goschler, Elena Andonova, and Robert J. Ross. 2008. Perspective use and perspective shift in spatial dialogue. In *Spatial Cognition VI: Learning, Reasoning and Talking about Space*. Springer.

John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. The MIT Press.

Roberta L. Klatzky. 1998. Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. *Lecture Notes in Computer Science*, 1404.

Shaun Lawson, Emile van der Zee, and Laura Daley. 2008. Spatial language in computer mediated communication. In Phil Turner and Susan Turner, editors, *Exploration of Space,Technology, and Spatiality: Interdisciplinary Perspectives*. Hershey, Pennsylvania: Information Science Reference.

S.C. Levinson. 1996. Language and space. *Annual Review of Anthropology*, 25:353–382.

Robert Ross and Kavita E. Thomas. 2010. An empirically-based model for perspective selection in route-finding dialogues. In *Spatial Cognition 2010*, Portland, OR.

Holly A. Taylor and Barbara Tversky. 1996. Perspective in spatial descriptions. *Journal of Memory and Language*, 35:371–391.

Thora Tenbrink, Robert Ross, Kavita E. Thomas, Nina Dethlefs, and Elena Andonova. 2010. Route instructions in map-based humanhuman and humancomputer dialogue: A comparative analysis. 21.

Kavita E. Thomas and Elena Andonova. 2012. Coordination of spatial perspectives in response to addressee feedback: Effects of perceived addressee understanding. *Pragmatics and Cognition*, 20(3).

Constanze Vorwerg. 2009. Consistency in Successive Spatial Utterances. In *Spatial Language and Dialogue*. Oxford Linguistics.

Matthew E. Watson, Martin J. Pickering, and Holly P. Branigan. 2004. Alignment of reference frames in dialogue. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, Mahwah, NJ. Lawrence Erlbaum Associates.