

# Real-time Path Descriptions Grounded with GPS Tracks: a preliminary report

Nate Blaylock and James Allen

Institute for Human and Machine Cognition (IHMC)  
Pensacola, Florida, USA  
{blaylock,jallen}@ihmc.us

## Abstract

In this paper, we present a novel method for data collection which produces aligned real-time speech recordings of path descriptions and the corresponding GPS track of that path in the real world. We give a preliminary report on the pilot corpus we have gathered using this method and our initial annotation plans for references to location, orientation and movement within the speech. Using the GPS track of the real path and a GIS database, we plan to annotate spatial references in the corpus with the “ground truth” of objects in the GIS database, or lat/lon coordinates corresponding to referent object or location. This annotated data will provide a set of natural language descriptions of paths, locations, orientation and movement which can be used for training/testing algorithms for understanding spatial language situated in the real world and aided by a GIS database. We also describe an initial annotation tool we are building that uses Google Earth for visualizing and annotating the corpus.

## 1. Introduction

We are interested in building algorithms that understand natural language (NL) descriptions of spatial locations, orientation, movement and paths that are grounded in the real world. In particular, we are interested in algorithms that can ground these NL descriptions in real-world coordinates and entities by leveraging geographical information system (GIS) databases. Such algorithms would enable a number of applications, including automated geotagging of text and speech, robots that can follow human route instructions, and location pinpointing without the use of GPS.

To aide development of our NL understanding system, we are developing a corpus of natural language path descriptions (as recorded from humans in real-time while they travel the path) where each NL description of location, orientation, movement, and the overall path, are annotated with the actual entity (from a GIS database) or location (lat/lon) that is referred to by that description. This “ground truth” annotation will provide a set of NL descriptions that are marked with the “correct answer”, on which we plan to use to test our own GIS-informed NL understanding algorithms.

In this paper, we present a novel method for speech data collection in which subjects describe their path in real time (i.e., while they are traveling it) and a GPS receiver simultaneously records their actual paths. These GPS tracks of the actual path can aide the annotator in determining what GIS entities/locations were meant by each spatial reference. To be clear, the GPS tracks are to help a human annotator with the task of creating the annotated “ground truth” test corpus. GPS information would not help the eventual NL understanding system (which would be meant to geotag spatial references using just the natural language and a GIS database).

Although corpora exist for studying NL path descriptions, we are not aware of any that are bundled with the corresponding GPS track for the paths. In addition, many corpora are not in domains where real-world GIS databases would be useful for NL understanding. For example, in the Map Task Corpus (Anderson et al., 1991), paths described

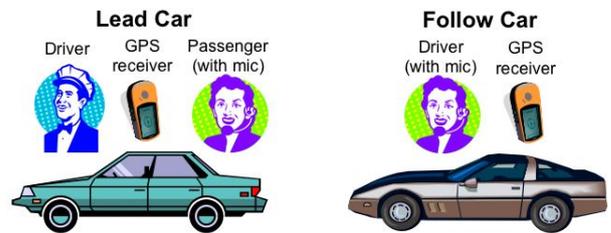


Figure 1: Experiment Setup

were drawn on 2-D maps of a fictitious world with relatively few landmarks and no streets. The MARCO corpus (MacMahon et al., 2006) describes paths through a 3-D virtual world of indoor corridors. The IBL corpus (Bugmann et al., 2004) contains path descriptions of robot movement through a miniature (fictitious) town model. None of these are directly applicable to GIS databases since each is in a fictitious environment and, with the exception of Map Task, movement on each is on a small scale. The smallest objects in our GIS database (as we will outline below) are at the scale of buildings—thus the scale of the path needs to be on the order of hundreds of meters so that multiple GIS objects might be referenced.

In the remaining sections of this paper, we first describe the pilot corpus and methodology. We then outline our plans for annotating it using a GIS database and Google Earth. Finally, we conclude and discuss future work and directions.

## 2. Pilot Corpus and Methodology

In this section, we describe our initial pilot data collection. We will first describe the general setup of the experiment and then the data we have collected thus far.

### 2.1. Setup

The setup of a typical data session is shown in Figure 1. Each session consisted of a lead car and a follow car. The driver of the lead car was instructed to drive wherever he

wanted for an approximate amount of time (around 15 minutes). The driver of the follow car was instructed to follow the lead car. One person in the lead car (usually a passenger) and one person in the follow car (usually the driver) were given close-speaking headset microphones and instructed to describe, during the ride, where the lead car was going, as if they were speaking to someone in a remote location who was trying to follow the car on a map. The speakers were also instructed to try to be verbose, and that they did not need to restrict themselves to street names—they could use businesses, landmarks, or whatever was natural. Both speakers' speech was recorded during the session. In addition, a GPS receiver was placed in each car and the GPS track was recorded at a high sampling rate (usually once per second).

The goal of this setup was to yield interesting data for our research goals mentioned above. The paths taken are in the real world and because the movement is in cars, it is on a scale that makes frequent references to entities in the GIS database. Additionally, the recorded GPS tracks provide us with the ground truth about the path taken against which we can compare the results of path description understanding algorithms. Finally, having the recordings of two speakers in two separate cars gives us essentially two different descriptions of the same path at the same time.

## 2.2. Data

In this pilot study, we conducted seven data collection sessions in the downtown Pensacola area with IHMC staff and associated researchers. This yielded 13 audio recordings<sup>1</sup> of seven paths along with corresponding GPS tracks. The average session length was 19 minutes.

The data is still in the process of being transcribed, but the following gives some examples of the contents:

- *...and we're going under the I-110 overpass I believe and the Civic Center is on the right side on the corner of Alcaniz and East Gregory Street where we are going to be taking a left turn...*
- *... he's going to turn left right here by the UWF Small Business Development Center heading toward Gulf Power ...*
- *... we've stopped at a red light at Tarragona Street okay we're going now across Tarragona passing the Music House ...*
- *... we're at the intersection of East Gregory and 9th near a restaurant called Carrabas I think and a Shell station just a little south of the railway crossing ...*

## 3. Planned Annotation

In this section we discuss our initial ideas for annotating the corpus. These ideas will be refined as annotation proceeds. With a GPS track and access to a GIS database, we can annotate the corpus with "ground truth" as to the spatial references which are made as well as the events in the path which are described.

We have created an initial visualization tool in Google Earth (shown in Figure 2) which allows an annotator to see the lead car's path and replay its progress along the track in real time. This is more than a video, however, as the annotator can freely zoom and scroll in Google Earth during the track replay. This, synchronized with the speech track, shows the annotator where the car was located when certain utterances were made, giving the annotator a much better context.

The GPS track gives us ground truth for the path described, but it does not give ground truth for most spatial location references used in describing the path (such as businesses, streets, bodies of water, etc.) These we will annotate with objects in a GIS database, instead of just lat/lon coordinates. We have obtained access to TerraFly (Rishe et al., 2005), an industrial-strength GIS system developed by the High-Performance Database Research Center/NASA Regional Applications Center at Florida International University (FIU). TerraFly contains a large aggregation of GIS data from major distributors, including NavTeq and Tiger streets and roads, 12 million U.S. businesses through Yellow Pages, and property lines for the state of Florida. Property lines are useful, as they enable a more accurate placement (geocoding) of the location of street addresses (e.g., from the Yellow Pages). We plan to hierarchically bracket and annotate each spatial location reference as well as movement and orientation events in the corpus. Spatial entities will be annotated with the corresponding object in the TerraFly database.

Figure 3 shows an example annotation for the first utterance shown in the examples above. Here, each segment is assigned a GIS.ID corresponding to the object in the TerraFly database. For example, it is obvious from the GPS track that the phrase "the Civic Center" corresponds to the "Pensacola Civic Center", which is a business in the TerraFly database. The annotation of this phrase is shown in Figure 3a. Similarly, the streets "Alcaniz" and "East Gregory" are found and annotated with corresponding TerraFly objects.

Composite objects not found in the database (such as the intersection in Figure 3e) are built up out of smaller objects (the streets). The two events in this utterance, passing the Civic Center and then turning left, are annotated as shown in Figures 3b and 3f, respectively. Each makes reference to the previous spatial locations as arguments.

## 4. Conclusion

We have presented a methodology and pilot data collection which aligns real-time speech recordings of path descriptions with a GPS track of that path. We also presented an initial tool in Google Earth for replaying that track in real time and our initial plans for data annotation using Google Earth and a GIS database.

For future work, we plan to collect and annotate more data using a similar methodology (as informed by our pilot experience). In addition, we plan to gather data by replaying these tracks using our Google Earth tool described above and having subjects describe the path as it unfolds from an aerial view. In addition to giving us more descriptions of the same path, we believe this will provide interesting com-

<sup>1</sup>In one session, there was no speaker in the lead car.

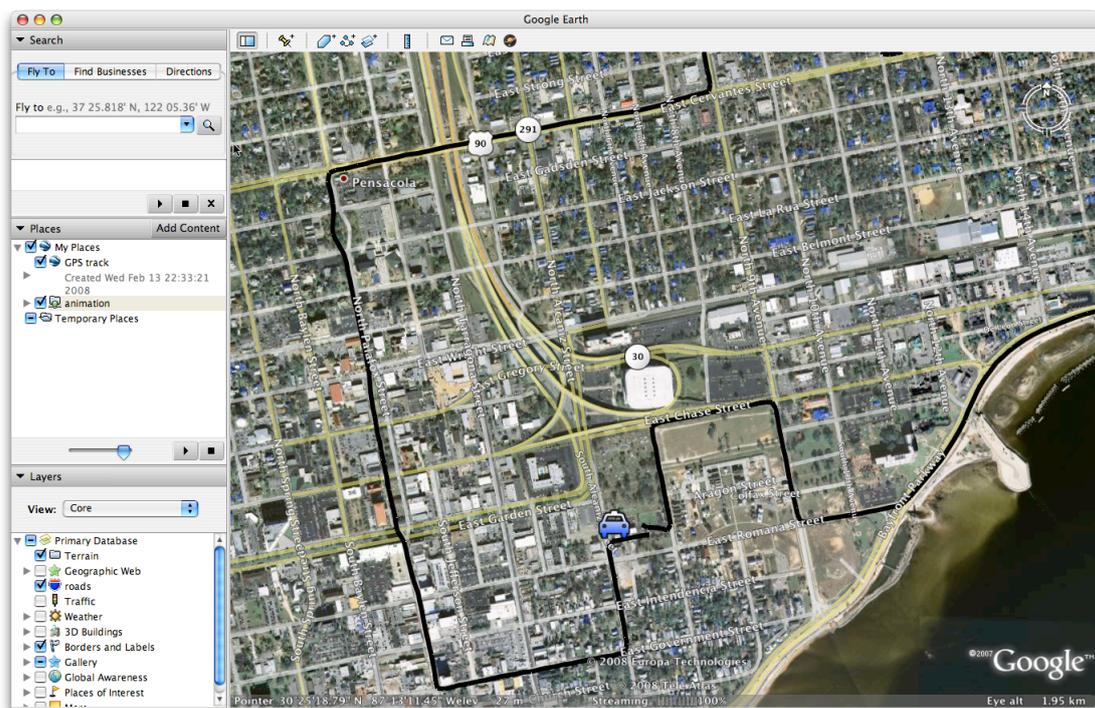


Figure 2: Screenshot of an overall path (in black) and the lead car’s “current” position in Google Earth

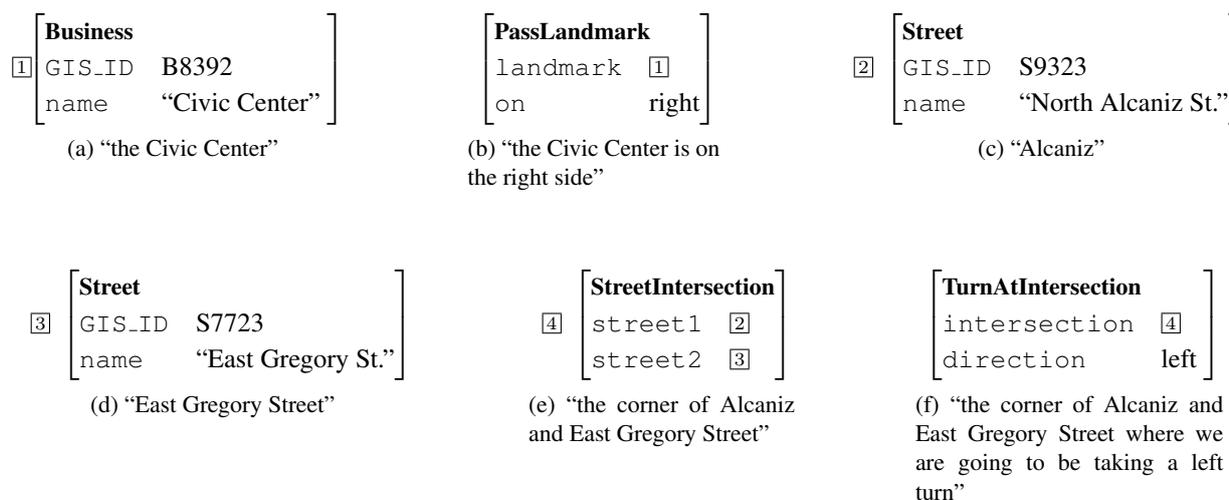


Figure 3: Example Hierarchical Annotation of the Entities and Events in the Utterance: “the Civic Center is on the right side on the corner of Alcaniz and East Gregory Street where we are going to be taking a left turn”

parison data as to how paths are described from a ground view (e.g., from the follow car) versus from an aerial view (e.g., from a following helicopter or UAV).

## 5. References

A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert. 1991. The HRCR map task corpus. *Language and Speech*, 34.

Guido Bugmann, Ewan Klein, Stanislaw Lauria, and Theocharis Kyriacou. 2004. Corpus-based robotics: A route instruction example. In *Proceedings of Intelligent Autonomous Systems (IAS-8)*, Amsterdam.

Matt MacMahon, Brian Stankiewicz, and Benjamin Kuipers. 2006. Walk the talk: Connecting language, knowledge, and action in route instructions. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*, pages 1475–1482, Boston.

N. Rishe, M. Gutierrez, A. Selivonenko, and S. Graham. 2005. TerraFly: A tool for visualizing and dispensing geospatial data. *Imaging Notes*, 20(2):22–23.