Semantic Annotation of Street-level Geospatial Entities

Nate Blaylock
Florida Institute for Human and Machine Cognition (IHMC)
Ocala, Florida, USA
Email: blaylock@ihmc.us

***Abstract*—In this paper, we describe the PURSUIT Corpus—an annotated corpus of geospatial path descriptions in spoken natural language. PURSUIT includes the spoken path descriptions along with a synchronized GPS track of the path actually taken. Additionally, we have manually annotated geospatial entity mentions in PURSUIT, mapping them onto point entries in several geographic information system databases. PURSUIT has been made freely available for download.**

I. 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 geographic 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 aid the development and evaluation of our geospatial language understanding system [1], we developed the PURSUIT Corpus, which consists of 13 audio recordings of spoken path descriptions that were made in "realtime" as the path was driven in an urban area. Additionally, the corpus includes corresponding synchronized GPS tracks for each recording, which represent the "ground truth" of the path actually traversed.

The corpus has been manually transcribed, segmented, and annotated with geospatial entity references. The synchronized combination of information from the speech and GPS "modalities" has allowed us in a fairly reliable way to manually identify the intended real-world geospatial entities referred to by mentions in the speech. Previous work has dealt with the issue of annotating geospatial entities at a city/state/country level and tying them to real-world entities (e.g., [2]). However, very little attention has been paid to geospatial entities at the *street level* — the granularity of streets, buildings, etc., which is a much more open set than cities, states, and countries.

In the next sections, we describe the data collection method for the corpus and the various annotations performed on it. In particular, we describe our strategy for semantic annotation of entities based on a combination of name, address, and latitude/longitude (lat/lon) coordinates. We then describe related work and conclude by describing future planned work on the corpus. The PURSUIT Corpus is freely available for download at http://www.cs.rochester.edu/research/speech/pursuit/.

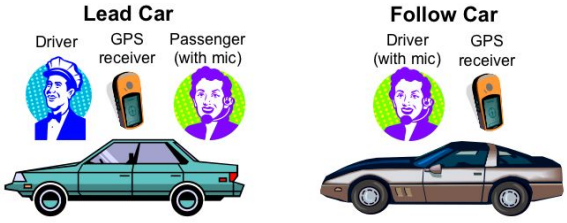

Fig. 1. Data Collection Setup

II. Corpus Data Collection

Our data collection methodology for the corpus is detailed in [3]. For convenience, we summarize it in this section.

A. Setup

Figure 1 shows an example of the data collection setup for the corpus collection. Each session consisted of a lead car and a follow car in downtown Pensacola, Florida. 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.

B. Data

The corpus contains 13 audio recordings of seven paths along with two corresponding GPS tracks from the cars.[1] The average session length was just over 18 minutes, and overall

[1]In one session only one audio recording was made.

	Corpus	Ave. per Session
Length	3h55m	18m
Utterances	3,155	243
Annotated Refs	1,649	127

TABLE I
PURSUIT CORPUS STATISTICS

1,649 geospatial references were annotated. Table I shows various information about the corpus size.

The corpus is rich with references to geospatial entities. Some sample utterances from the corpus are given below:

- *...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 ...*

C. Synchronization

The resulting 2 audio and 2 GPS track files for each session were synchronized by hand to start and end at the same point in time. As the recording on each device was started separately from the others, this lead to special challenges in synchronization. Using the TESLA annotation and visualization tool for this corpus [6], the annotator adjusted audio and GPS length and starting time by hand until the audio descriptions and GPS tracks seemed to be in concordance.

III. ANNOTATION

The corpus has been manually annotated with transcription, utterance, and location reference information. Before describing these, however, we first describe the annotation format of the corpus.

A. Annotation Format

We use the NITE XML Toolkit (NXT) data model [4] for storing both the corpus and annotations on it. NXT is a general XML data model for multimodal and heavily cross-annotated corpora. In the data model, a corpus is represented as a list of observations, which contain the data for a single session. An observation contains a set of synchronized signals, which are typically audio or video streams associated with the observation, although NXT is broad enough that a signal may be any timestamped stream of data (like our GPS tracks). Annotations are represented as a multi-rooted tree structure, where leaves are segments that are time-aligned with an underlying signal. This allows disparate annotations to be made on and saved with the same corpus.

B. Transcription

Transcription of the audio signal was done manually using the Transcriber tool [5]. The resulting transcription included not only words, but also preliminary utterance breaks that were useful to the transcriber.

Transcription rules were that no punctuation was to be transcribed, except in phrases requiring a hyphen, periods in names with abbreviations, and apostrophes. Proper nouns were capitalized, but the beginnings of utterances were not. Internet resources such as Google Local Search were used to verify canonical spellings of proper nouns such as business or street names. Numbered street names were spelled out (e.g., *Seventeenth Avenue*). In cases where the correct transcription could not be determined, the token `[unintelligible]` was inserted as a word.

The words level in NXT requires not only the list of transcribed words, but also timing information on the start and end time of each word. This was estimated by using the rough Transcriber utterance boundaries for the start and end time of each rough utterance and equally dividing the utterance time into chucks for each word within it.

As an aside, the timing information in the corpus is quite important in this domain, as it places a constraint on possible distances moved in the given time. For example, if a speaker mentions a park in one utterance and then 10 seconds later mentions an intersection, we can assume that the car cannot have moved 5 miles during that time.

C. Utterance Segmentation

Utterance segmentation was done manually using the TESLA tool [6]. Utterance segments of spoken monologue are admittedly somewhat arbitrary, but annotators were instructed to use cues such as pauses and grammar to help determine natural utterance breaks.

D. Geospatial Reference Annotation

References to certain types of locations were segmented and annotated by hand with information about each referent using the TESLA tool.

The high-level classes annotated were:

- *Streets*: references to a given street, for example "Garden Street" or "a divided road"
- *Intersections*: references to street intersections, for example "the corner of 9th and Cervantes" or "the next intersection"
- *Addresses*: references to street address, for example "401 East Chase Street" or even "712" (when referring to the address by just the street number)
- *Other Locations*: this class is a grab bag for all other location types that we annotated, consisting of such data as businesses, parks, bridges, bodies of water, etc.

This classification was chosen because that was the separation of data types in our GIS databases, not for deep ontological reasons. We do believe, however, that the standardization of a geospatial entity ontology and the representation of entities

	Reference Type		
	Named	**Category**	**Total**
Street	77.2%	22.8%	**48.5%**
Intersection	45.5%	54.5%	**6.8%**
Address	100.0%	0.0%	**0.8%**
Other Loc	67.7%	32.3%	**43.9%**
Total	**71.1%**	**28.9%**	**100%**

TABLE II
BREAKDOWN OF GEOSPATIAL ENTITY REFERENCE ANNOTATIONS IN THE PURSUIT CORPUS

in this format is quite needed. (We mention this more below in future work.)

Note that not all geospatial entity references have been annotated in PURSUIT—just those types that are accessible in our GIS databases. Examples of entities referred to in the corpus but were not annotated are fields, parking lots, sidewalks, railroad tracks, neighborhoods, and fire hydrants. These were not annotated only because we did not have access to data about those entities. However, there is nothing inherent in our approach to path understanding which would prohibit the use of those classes of entities, if data were available for them. Indeed, we believe that much more GIS data will be made available in the not-too-distant future through the release of private or government databases and advanced mining techniques.

Although not all *classes* of entities were annotated, within those classes that were annotated, *all* references to entities of interest were annotated, whether or not they were actually found in the GIS databases. Additionally, all references to entities were annotated, including category and pronoun references. Thus "Garden Street", "a divided road", or even "it" were annotated if they referred to a geospatial entity of interest. Each entity was also annotations with whether an entity reference was *named* (i.e., contained at least part of the proper name of the entity, such as "the Music House" and "the intersection at Cervantes") or *category* (description did not include a name, such as "the street", "a Mexican restaurant", and "it").

Annotators were instructed to bracket the entire referring phrase (as opposed to e.g., just the headword as is done in SpatialML [2]). One reason for this is that it allowed the annotation to reflect embedded references. For example, many references to intersections also mention streets. The phrase "the corner of 9th and Cervantes" contains references to an intersection and two streets. Although it would be possible to just annotated the headwords (e.g., *corner*, *9th*, and *Cervantes*), that annotation loses the information that, indeed, the intersection is at these two roads.

In total, 1,649 geospatial entity references were annotated in the corpus. The breakdown of categories is shown in Table II.

E. Grounding Geospatial References

Although the manual bracketing of references is relatively easy, deciding which real-world entities they correspond to is not, in general. Additionally, as the set of geospatial entities at the street-level is not closed, there is a question as to how to represent the real-world entities as semantic individuals.

We treat these issues in turn and then describe the GIS databases we used for reference.

1) Semantic Representation: In an ideal world, we would have access to a single knowledge base (in e.g., OWL) which contained all possible geospatial entities with unique IDs that we could use to ground geospatial references. Our reality is that we had two GIS point databases with varying coverage, and, in the case of Google, with no direct access to the underlying dataset. In fact, out of 724 *other_loc* references, 25.7% were in both databases, 16.7% were only in TerraFly, 40.1% were in only Google, and 17.5% were in neither.

Lat/lon coordinates alone are also not a viable way to uniquely identify a geospatial entity. First, lat/lon coordinates, represented as decimals, have arbitrary precision. One dataset may represent lat/lons to the thousandths place, whereas another to the hundred-thousandths, making it impossible to know if two lat/lons refer to the same entity (remember, our goal is to ground *entities*, not locations). Second, although represented as point data, most geospatial entity data is actually 2-dimensional—a business may refer to the lot it is on, instead of an arbitrary point on that property. Third, many databases are geolocated automatically based on street address (where available). In our experience, many times the given lat/lon can be up to 200 meters from the actual location. It can be worse in rural areas. Different datasets may have different geolocations, and there is no trivial way to determine if two lat/lons refer to the same entity. Lastly, consider the case where several businesses reside at a single address, such as in a shopping mall. A dataset may have several entities for a single lat/lon.

Using the entity name as a unique ID is also problematic, as an entity may have several aliases or may be referred to in different ways — for example *IBM*, *I.B.M.*, *IBM Corp.*, *International Business Machines*, etc.

Although we do not have a perfect solution, we outline the approximation we have taken for the different types of entities.

a) Other Locs: Our catch-all *other_loc* class contains entities such as businesses, parks, bodies of water, etc. Each is annotated minimally with a canonical name (from the GIS database, if available, or chosen by the annotator based on internet searches) and a lat/lon (in the GIS, or manually chosen by the annotator on Google Earth). For entities which have an address, this is added as well.

b) Streets: Streets are actually an interesting case. GIS databases represent streets as raster data of a set of street segments (a street segment is roughly a section of a street between two intersections). To complicate matters, some street segments may have additional names (such as county road designators), or more precisely, multiple logical streets may overlap in certain segments.

Our solution is to represent both street segments and logical streets. Street segments are principally identified by the lat/lon coordinates of both of their endpoints. They are also labeled with the set of street names applied to them. We have also created a database of streets, which are identified by name, as

well as the set of street segments that comprise them. As a practical matter, in the corpus, we annotate a street reference with the street segment closest to the speaker, since this allows us to treat these references as points.

c) Intersections: Intersections are typically not represented explicitly in publicly available GIS databases. However, they are referred to quite often as street-level entities. For our project, a database of intersections was created within TerraFly using street raster data.

An intersection is principally identified by the set of streets that intersect at it. Note that the size of this set is sometimes greater than two due to: (1) more than two roads intersecting (e.g., 6-point intersections); (2) two physical roads intersecting and the roads have different names on either side of the intersection; or (3) roads with multiple designations intersecting. Secondarily, we use a lat/lon to uniquely identify an intersection, as some pairs of roads intersect in more than one location.

d) Addresses: Addresses are identified by their full street address.

2) Grounding: As mentioned above, we developed a tool for annotation which aided in the grounding of entity references [6]. The tool allows the replay of a speech signal synchronized with a moving car icon located at the car's location from the GPS track at that instant. Furthermore, it supports keyword search centered on that location to both the Google Maps and TerraFly databases, showing matches at their lat/lons in Google Earth. This interface allowed an annotator to determine the correct entity for the annotated reference.

In cases where the entity was not found in the databases, annotators had good local knowledge of the area, and with this, in many cases were able to know the intended referent. If that was not enough, annotators would sometimes use Google Street view to look at images at the location. In a small amount of cases, however, the annotator had to physically travel to the given location in order to understand and get the entity information.

In cases where the entity was not in the databases, the human annotator searched for the missing data by hand using various resources, including, for several, retracing the driven route to find the intended referent.

3) Source Database Information: As noted above, several sources were used to search for geospatial entity information for annotation. The data sources are also noted in the annotation on each reference. The two main data sources used are TerraFly and Google Local (which we will describe in more detail below). Entities which were not available in either data source are marked correspondingly. Overall, 92.2% of geospatial entity references of interest were in either or both of the GIS databases used.

We now describe the two databases.

a) TerraFly: A primary source used was a custom-made subset of the TerraFly GIS database [7]. The custom database was made by compiling data from a number of datasets, including NAVTEQ NAVSTREETS and POI, Yellow Pages Business Information, Info Business Database, Geographic Names Information System (GNIS), and US Census Data. Additionally, because of the widespread reference to intersections in the corpus, the TerraFly team created a dataset of all intersections in the US (derived from NAVSTREETS data), indexed by the streets intersecting.

b) Google Maps: Google Maps[2] provides a service for searching for businesses near a location. Note that this database only includes point data, and not streets or intersections.

IV. Related Work

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 [8], paths described were drawn on 2-D maps of a fictitious world with relatively few landmarks and no streets. The MARCO corpus [9] describes paths through a 3-D virtual world of indoor corridors. The IBL corpus [10] 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.

On the other hand, work has been done in the domain of geospatial referent resolution for toponyms (e.g., [11], [2]). However, this line of work is concerned with the resolution of referents at the city/state/country level, whereas we are interested in the *sub*-city level. Indeed, we believe that reference at the sub-city level is more difficult, due to the large number of diverse entities as well as the tendency to refer to them by category or only partial name.

SpatialML [2] is also focused on the city/state/country level of annotation. Additionally, it can represent certain spatial relations between geospatial entities, which our data does not annotate (although this is an area of future work as we mention below). SpatialML only annotates the headword of a referring expression, whereas in PURSUIT we annotate the entire referring expression. PURSUIT also allows overlapping annotations, which are especially critical for references to intersections, which may be composed of two (annotated) street names (e.g., *the intersection of Alcaniz and Romana*).

V. Conclusion

This document has described the PURSUIT Corpus, which is a corpus of spoken, realtime descriptions of paths, synchronized with "ground truth" GPS tracks. The corpus has been annotated with a number of geospatial referents and their corresponding information from several GIS databases. As far as we are aware, PURSUIT is the first corpus to cover reference at a street level in the geospatial domain.

[2] `http://maps.google.com`

In the future, we are interested in annotating movement events in the corpus along with their arguments. There seem to be many such event descriptions in the corpus, including turns, passing a landmark, and heading in a cardinal direction.

Additionally, as we mention above, there is a great need for semantically structured geospatial entity data. This would include an ontology of geospatial entity classes, along with canonical properties. There is also a large set of geospatial entity types that have not been annotated in the corpus because we did not have access to that type of GIS data. These include entities such as parking lots, wooded areas, neighborhoods, street signs, etc. We are currently annotating the full set of geospatial entities in the PURSUIT Corpus in order to analyze the types of missing entity types. We are studying various ways to make them available, such as mining web information and computer vision techniques.

Finally, the path descriptions in the PURSUIT Corpus were all done from a first-person, ground-level perspective. As TESLA allows us to replay the actual routes from GPS tracks within Google Earth, we believe we could use this tool to gather more spoken descriptions of the paths from an aerial perspective from different subjects. This would give us several more versions of descriptions of the same path and allow the comparison of descriptions from the two different perspectives.

ACKNOWLEDGEMENTS

We would like to thank the following: James Allen, who gave scientific oversight to this corpus development; Bradley Swain, who helped with the annotation and development with the TESLA annotation tool; and Dawn Miller, who also helped with the annotation.

REFERENCES

[1] N. Blaylock, B. Swain, and J. Allen, "Mining geospatial path data from natural language descriptions," in *Proceedings of the 1st ACM SIGSPATIAL GIS International Workshop on Querying and Mining Uncertain Spatio-Temporal Data*, Seattle, Washington, November 3 2009.

[2] I. Mani, J. Hitzeman, J. Richer, D. Harris, R. Quimby, and B. Wellner, "SpatialML: Annotation scheme, corpora, and tools," in *6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, May 2008.

[3] N. Blaylock and J. Allen, "Real-time path descriptions grounded with GPS tracks: a preliminary report," in *LREC Workshop on Methodologies and Resources for Processing Spatial Language*, Marrakech, Morocco, May 31 2008, pp. 25–27.

[4] J. Carletta, S. Evert, U. Heid, and J. Kilgour, "The NITE XML toolkit: data model and query language," *Language Resources and Evaluation Journal*, vol. 39, no. 4, pp. 313–334, 2005.

[5] C. Barras, E. Geoffrois, Z. Wu, and M. Liberman, "Transcriber: development and use of a tool for assisting speech corpora production," *Speech Communication special issue on Speech Annotation and Corpus Tools*, vol. 33, no. 1–2, January 2000.

[6] N. Blaylock, B. Swain, and J. Allen, "TESLA: A tool for annotating geospatial language corpora," in *Proceedings North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT) 2009 Conference*, Boulder, Colorado, May 31–June 5 2009.

[7] N. Rishe, M. Gutierrez, A. Selivonenko, and S. Graham, "TerraFly: A tool for visualizing and dispensing geospatial data," *Imaging Notes*, vol. 20, no. 2, pp. 22–23, 2005.

[8] 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, "The HRCR map task corpus," *Language and Speech*, vol. 34, 1991.

[9] M. MacMahon, B. Stankiewicz, and B. Kuipers, "Walk the talk: Connecting language, knowledge, and action in route instructions," in *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*. Boston, Massachusetts: AAAI Press, July 2006, pp. 1475–1482.

[10] G. Bugmann, E. Klein, S. Lauria, and T. Kyriacou, "Corpus-based robotics: A route instruction example," in *Proceedings of Intelligent Autonomous Systems (IAS-8)*, Amsterdam, March 10–13 2004.

[11] J. L. Leidner, "Towards a reference corpus for automatic toponym resolution evaluation," in *Workshop on Geographic Information Retrieval*, Sheffield, UK, 2004.