**Extracting Spatial Relations From Place Descriptions**

Arbaz Khan Department of Computer Science and Engineering Indian Institute of Technology Kanpur

Uttar Pradesh, India

[arbazk@iitk.ac.in](mailto:arbazk@iitk.ac.in)

Maria Vasardani Department of Infrastructure Engineering

University of Melbourne Victoria, 3010 mvasardani

@unimelb.edu.au

Stephan Winter Department of Infrastructure Engineering

University of Melbourne Victoria, 3010 [winter@unimelb.edu.au](mailto:winter@unimelb.edu.au)

**ABSTRACT**

Computational model of understanding place descriptions is a cardinal issue in multiple disciplines and provides critical applications esp. in dialog-driven geolocation services. This research targets the automated extraction of spatial triplets to represent qualitative spatial relations between recognized places from natural language place-descriptions via a simple class of locative expressions. We attempt to produce triplets, informative and convenient enough as a medium to convert verbal descriptions to graphical depictions. We present a reasoning approach devoid of any external resources (maps, path geometries or robotic vision) for understanding place descriptions. We then apply our methodologies to situated place descriptions and study the results, its errors and im- plied future research.

**Categories and Subject Descriptors**

H.3.1 [Information Storage and Retrieval]: Computa- tional Linguistics; I.2.7 [Artificial Intelligence]: Natural Language Processing

**General Terms**

Verification, performance

**Keywords**

Place descriptions, spatial role labelling, spatial language understanding

**1. INTRODUCTION**

A place description provides spatial information in terms of spatial relations between objects in the environment. Place descriptions are used in everyday activities in spatial-learning and problem-solving in the absence of maps or compass in large-scale environments. Most descriptions possess a dy- namic nature where the human provider uses a tour route for communicating spatial information [14]. So place de- scriptions can also be thought of to encompass path descrip- tions and to understand a place description would mean to

infer the relative/absolute location of the objects and the static or dynamic spatial relations between them by follow- ing any route within. To add to the complexity, the ob- jects in the descriptions can be simply objects (wall, picture, house) or place names (vernacular or from the gazetteers). Its required to resolve the vernacular names and potential places from the objects. Nonetheless, this challenging task of understanding the natural language place descriptions pro- vides critical applications in numerous domains including in- struction following robots, description based localization in query/dialog based navigation services and automated geo- tagging of text. Also, having inferred the spatial relations from such a description eases the translation of abstract ver- bal descriptions to the more iconic medium of conveying in- formation i.e pictures and sketch maps.

This research targets the automated extraction of spatial triplets to represent spatial relations between recognized ob- jects from natural language place descriptions via a degen- erate form of locative expressions. Locative expression as described in [6], is an expression containing a preposition, its object and whatever the prepositional phrase modifies. However, a degenerate locative expression can be described as an expression containing a preposition and its object, or simply a noun place (no prepositions). The extraction pro- cess is accompanied with a trained learning model ([15]) which predicts the beginning and the span of a degenerate locative expression in a sentence using CRFs along-with the place names indicators (gazetteers and manual annotations). It gives a good indication on the spatial relations in NL text and can be used to extract spatial relationships in the form of triplets in the format of an object of interest, a reference object and a prepositional relation that connects the two . This representation is derived from the technique of spatial role labelling (SpRL) [12], where the preposition serves as a spatial indicator connecting a trajector (object of inter- est) with a landmark (reference object). The end goal of this effort is to ease the translation of verbal descriptions to graphical depictions. So the aim is to make the triplets as informative and convenient as possible for such a transla- tion. By being informative, we mean to exploit the spatial information in the three components of the triplets. Also, the triplet output format becomes convenient enough to a translator if it cuts down the linguistic complexities.

The introduction of the problem leads to the fundamental question - Given a degenerate locative expression, is it pos- sible to extract the spatial relations without introducing any

additional ambiguities? If yes, how effective it is to use de- generate locative expressions(DLEs) to identify the under- lying spatial relationships between the places. In this paper we present a technique to identify the spatial relationships between recognized objects and place names with the help of DLEs. We claim our identified spatial relations to be qualitative and informative enough to assist in sketch-map generation and graphical depictions. However for the pur- pose of extraction of triplets, we would be targeting only the static relations in the descriptions. We would then examine the complexities of resolving frame of reference in a descrip- tion and would end by pointing out the challenges for deal- ing with the descriptions involving indirect references and spatial motion.

**2. RELEVANT WORKS**

[22] worked on the cognitive linguistic research in spatial se- mantics and provided basic theoretical concepts for under- standing it. Ever since then, there has been growing trend to understand natural language (NL) descriptions. The re- search has moved towards understanding unrestricted NL in current works ([12, 18, 9]). However, most of these works make use of map-assisted learning and/or vision based un- derstanding ([8, 13, 16]), our methodology makes use of no external resources but the machine learning tools [15] to identify place names and the associated prepositional phrases (DLE). The spatial triplets are then identified by a case-based reasoning method, in whole which takes away the need of an annotated training dataset (which are usually noisy or biased) and/or a detailed map.

**3. THEORY**

**3.1 Definitions**

We start by comprehensively defining the major terms that

appear in this paper viz spatial triplets (or simply triplets)

and degenerate locative expressions (DLE).

*3.1.1 Spatial Triplets*

A spatial triplet is a triple of a locatum (LOC ), reference ob- ject (RO ) and a spatial relation (r ). Here, RO is the object that defines the reference to the location of LOC using the prepositional relation r. For example, <Melbourne Hospi- tal diagonally across Peter Doherty Institute> references Melbourne Hospital as the LOC in terms of Peter Doherty Institute as RO by the relation diagonally across as r. [?] provides more information on the theory and the structure of triplets. This method of defining spatial relation is derived from the technique of spatial role labelling(SpRL) which first appeared in [12]. In SpRL, one assigns spatial role labels

to the words or phrases in sentences from the set {trajector (tr), spatial Indicator (si), landmark (l), none }. For example, such a labelled sentence looks like: “[We ]tr are sitting [in ]si [the Baretto’s ]l [in ]si [the Alan Gilbert Building ]l ”. How- ever, spatial triplets that could be extracted out of such a sentence look like :

<We in the Baretto’s>

<We in the Alan Gilbert Building>

Observably, a spatial triplet relation provides a better rep- resentation of spatial information than the role labels, and provides a suitable medium to translate the verbal descrip- tion to a pictorial representation.

*3.1.2 Degenerate Locative Expressions (DLE)*

A locative expression is a spatial referring expression which is defined as an expression containing a preposition, its ob- ject and whatever the prepositional phrase modifies [6]. These expressions have been adopted as a standard for spatial se- mantics and cognition and widely referred ([19],[22]). How- ever, in informal communication or a situated dialog, peo- ple tend to disregard the subject and describe the location in terms of just the preposition and its object (e.g. [I am] at the Deakins university, ). We term such locative expressions as degenerate locative expresssions (DLE). While crowdsourc- ing for place descriptions, people described their location mostly in terms of DLEs1 (e.g. at the University of Mel- bourne, in my house ). Hence, an attempt was made to iden- tify the DLEs and train a model on the available dataset to provide predictions on any test corpus.

*Classification of Degenerate Locative Expressions*

From the definition of a degenerate locative expression (DLE), one can think of constructions like “of the train-station”, “from the University” and “to the Bourke Street”. Such ex- amples do follow from the definition but convey no spatial information independently. Such DLEs require special treat- ment before the corresponding triplet relation could be ex- tracted. We term such DLEs as directional DLE s. These locative expressions start with a prepositions of direction, from the set {from, of, to, by, towards, past, after, before} which may or may not be used in a directional sense in the text . The rest of the DLEs, which with their subjects give

a sense of spatial relation are termed as locative DLE s. The procedures for triplet extraction described in the subsequent sections are described first for locative DLEs and then for directional DLEs. Some directional DLEs can be converted to locative DLEs by completing the prepositional clause to form a locative phrase (e.g. “from the University” can com- plete to “at the 3rd house from the University” and “of the train-staion” to “in the north of train-station”). Such DLEs are converted so and thenceforth handled like locative DLEs.

**3.2 Extracting Spatial Triplets**

To deal with the task of spatial information extraction, we

set ourselves the aim of identifying ‘informative’ spatial triplet relations. The problem can be defined similar to a spatial role labelling task(SpRL), the difference being that we at- tempt not to miss any information from the place names and the spatial relation. For example, in “two women in the front row with their heads turned around”, SpRL marks the ref- erence object as ‘row’ to eventually give the triplet “women in row”. However, we are interested in including any infor- mative meaning inherent in the description and thus output “women in front row”. This also becomes essential in the long term goal of translating verbal descriptions to graph- ical depictions, where the graph/sketch drawing algorithm could be vitally benefitted by localizing the object to ‘front row’ rather than just any ‘row’ (from the above example). There exist datasets for identifying spatial expressions ([3,

5, 11]) but none are characteristic enough of a place descrip- tion. For instance, place descriptions in practical are more descriptive than mere spatial expressions and the spatial re- lations can be fairly complex and difficult to formalize (e.g

1 from the dataset taken from tell us where mobile game -

<http://telluswhere.net/>

Table 1: Extending the prepositional relations of direction using stanford dependencies

|  |  |  |
| --- | --- | --- |
| Governor | Action on relation | Example |
| noun | extend with the noun and its modifiers | prep from(house, station)  extends from as ”3rd house from” |
| adjective | extend with the adjective | prep to(next, station)  extends to as ”next to” |
| verb | extend with the verb’s modifiers\* | prep from(living, station)  adds modifiers from npadvmod(living, minutes)  extends from as ”2 minutes from” |

\*-preference to verb’s direct object (if any)

‘between’,‘across from’). Due to the goals set thereafter and the fact that there is no available dataset for training on such special and essential requirements, we chose a reason- ing approach rather than a machine-learning approach to extract triplet relations. This approach would help identify the structure of problem of understanding place descriptions and thus eventually aid the machine learners in identifying the better features and deciding the distribution of weights to these features.

Extracting triplet relations poses varying challenges owing to the complex nature of place descriptions. The spatial re- lations can be defined either simply using position descrip- tors (e.g on, at, in) or using concepts of motion and frame of reference ([22]). For example, in ”From here, you can go left towads X and then head north to reach Y”. The phrase ”From here” carries context(indirect reference) and the current context is shown spatially related to X and Y via motion indicators. For the purpose of this paper we limit ourselves from the detection of spatial motion and recogni- tion of frame of reference and instead focus on the structure of static place-descriptions. We deal with place descriptions by classifying the expressions in place descriptions on the ba- sis of whether they carry any context or motion. The process of identification goes through the two types of DLEs to ex- tract triplets assuming all descriptions are static i.e without any context or motion. We start by describing the process of triplet extraction for locative DLEs.

*Locative DLE*

Given the DLEs, to identify the triplet relation, one needs to find the subject of DLE. For example, in “There is a con- struction site in the alley.”, the DLE is ‘in the alley’ and the triplet relation extracted should be <construction site in the alley>. To identify the subject, the dependency be- tween the preposition in and the two noun places ‘construc- tion site’ and ‘the alley’ needs to be identified. Luckily, from the Stanford typed dependencies parser [4], one can extract the dependencies between the words of a sentence. Using the collapsed dependencies representation, prep relation reports a prepositional phrase and the verb/noun/adjective modi- fied by the prepositional phrase. The former is called the dependent and the latter is called as governor. As for this case, the dependency reported is prep in (site,alley) where

‘site’ is the governor and ‘alley’ is dependent. When the relation is between two nouns, the triplet relation is direct. However, in the case of verb/adjective, we further explore the typed dependencies to find the nominal subject or the

noun modified by the corresponding verb/adjective. The subject finding task (for verb/adjective/noun) then recur- sively checks for the modifiers of the subject (if any) by searching through the set of modifier relations provided by typed dependencies. The final triplet relation is reported thereafter. The steps can be explained by a simple exam- ple in Figure-1. In descriptions where a person describes location using reference to pronouns, for e.g. You can also find NAB and Commonwealth bank branches near the Union House, it’s the direct object of the verb that gets linked to its prepositional relation (near ). So, in this sentence, sub- ject of the verb find is You, but the direct objects are NAB and Commonwealth Bank, and thus the triplet should be

<NAB near Union House> (and not You near Union House ). Thus, the heuristic that works here is if the there is a direct object of the verb, report the direct object as the locatum of the triplet. The heuristic can be further strengthened by using the fact that in such cases, the nominal subject of the verb is a pronoun.

*Directional DLE*

The directional DLEs convey no spatial information inde- pendently but they form an important medium for infering spatial relationships in the place description with an under- lying tour-route structure. As mentioned earlier, directional DLEs can be converted to locative DLEs in some sentences and can be directly used to output triplets. In other cases, a directional DLE can be made informative and meaning- ful by just extending its prepositional clause. For example, consider the sentence “I am 300 meters far from the auburn train-station” with the directional DLE as “from the auburn train-station”. Clearly, by including “300 meters far”, we get a more informative triplet <I 300 meters f ar f rom the train station>. Thus, the aim is to capture this infor- mation from the typed dependencies. Table-1 provides the exhaustive rules to extract this information from the typed dependencies. By doing this, we exploit the locative purpose served by a directional DLE. Thereafter, we proceed similar to a locative DLE to find subject of the preposition.

**4. IMPLEMENTATION**

The methodology mentioned in the previous section works

on the natural language via the results of a parser which is trained to identify degenerate locative expressions in infor- mal text[15]. The parser builds a conditional random field (CRF) machine learning model on the manually annotated corpus of place descriptions[1] and attempts to identify both informal and formal place references. To enhance its predic-



Figure 1: Sample run of the identification process for locative DLE in the sentence “I am on the Docklands Esplanade”

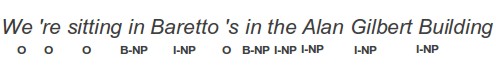
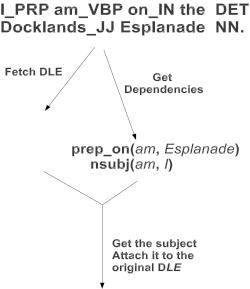


Figure 2: IOB-encoded identification of degenerate locative expressions

tion scores, it exploits the external resources like gazetteers and dictionaries. The output provided by the parser is in the form of IOB encoding as marked in Figure-2. A B-NP tag denotes the beginning of DLE, I-NP tags indicate the span of a DLE. All words with O tags are outside of any DLE. Hence in the example shown in Figure-2 identifies “in Baretto” and “in Alan Gilbert Building” as DLEs in the sen- tence. The DLEs output by the parser are either simply place names (e.g. Chetwynd Place) or place names with prepositional attachments (e.g. of the train station).

The parser is written in Python and our program works at the top of it. The input requirements to the parser are met by chunking the POS-tagged sentences. The parser can also take the manual annotations of place names as input, in the format provided by [20]. The manual annotations are optional for the working of parser, but enhance the DLE prediction scores. For each DLE extracted from the parser, our program identifies its subject (after extending prepo- sitional clauses for directional DLEs) using the output of Stanford parser [7] for POS tags and dependencies on the original sentences.

**5. EXPERIMENTAL STUDY**

In this section, we present the experimental analysis of our methodology. We start by describing the dataset used for testing out methodology. Next, we give an insight into the performance of our approach in extraction of triplets from the corpora. We then provide an analysis on the errors pro- duced in extracting the triplets and the sources behind and conclude by a remark on the triplets missed by the approach.

**5.1 Dataset Description**

For our experimental study, we used the Tell Us Where data[1] as one of the test corpora. The dataset comes from a location based mobile game which locates its users and

asks to answer the question, “tell us where you are?”. Origi- nally, the corpus consisted of 1,858 place descriptions but for the purpose of our research, we filtered out the descriptions containing less than 20 words. This gives 54 distinct situ- ated place descriptions, each of which describes relationships between more than two places. We also use a set of 4 NL de- scriptions of the campus of University of Melbourne for our studies. These were submitted by graduate students with varying extents of familiarities to the campus. Also, there were no instructions given to the students while submitting the descriptions which yields a set of unrestricted NL place descriptions. With such dataset, our testing model is set up to be descriptive enough and more natural in relation to and characteristic of place descriptions.

**5.2 Extraction Of Triplets**

The task of extraction of ‘spatially informative’ triplet rela- tions from NL descriptions is challenging enough to directly deal with using simply a language parser. The DLE parser helps us here to seperate out the place names. We employ the setting of using the manual annotations with the DLE parser2 . This subsection focusses on the triplets extraction study pertaining to static place descriptions. The triplets output in such cases had a fair similarity with the expected triplets. Below are some of the example cases representative of our results.

Example 1. I am [at a 3 stories town house], number 7 [Chetwynd Place]. [The house] is located [in a small alley] [behind a row of town house] [along Chetwynd Street], [near the corner] [between Chetwynd Street and Queensberry Street]. There is [a construction site] [in the alley]. Its the 3rd house from the head [of the alley].

This example is taken from Tell Us Where corpus. The bracketed texts correspond to the DLEs identified by the parser. The triplet relations output by the program are pre- sented in Table-2. This is a characteristic example of a static place description without a frame of reference, except in the last sentence where ‘it’ is used to symbolise the house of the subject and thus no triplet relation was extracted. If the sen- tence is rephrased as “I am in the 3rd house from the head of the alley ”, the triplet extracted is, <I in 3rd house f rom the head of the alley>3 . Nonetheless, from the output of the program, the triplets produced seem promising and interpretable for an algorithm for graphical depiction.

The ‘between’ relation output here is more informative but not syntactically a triplet as it has 2 reference objects con- nected to a landmark. This disturbs the computational interpretation of triplets. [21] puts up the semantics of triplet relation and proposes ‘between’ relation to be bro- ken into two halves which can be done here by identify- ing the conjuncts of the dependent (‘Chetwynd Street’ and

‘Queensberry street’). The conjuncts were not broken in this case because the DLE parser identifies Chetwynd Street

2 The results do not differ considerably with the set-up of not using the manual annotations except for the errors described in 5.3.3

3 See how the preposition ‘from’ gets extended to ‘3rd house from’, and the clause related to the preposition ‘of ’ gets attached into the reference object as ‘head of the alley’

|  |
| --- |
| <I at a 3 stories town house> |
| <The house in a small alley> |
| <a small alley behind a row of town house> |
| <The house along Chetwynd Street> |
| <The house near the corner> |
| <the corner between Chetwynd Street  and Queensberry Street> |
| <a construction site in the alley |

Table 2: Program output for spatial triplet relations in Example 1

and Queensberry street as one single place name and thus leading to a bias in the subject finding methodology. In a different case, where such a restriction does not hold, triplets produced for a between relation are shown below (Example

2). The output comes out as desired in [21].

Example 2. Between this building and [the campus], [on Swanston St], is the major public transport hub for [the Uni- versity ].

Example 3. We’re sitting [in Baretto]’s [ in the Alan Gilbert Building], [across Grattan street] is one of the [ medi- cal buildings ]. [Down the hill] [along Grattan street] the [new building] being constructed is the [Peter Doherty Institute] and diagonally across the road ([Royal Parade]) is [Melbourne Hospital]. In the other direction the open area is [University square] (there is a [carpark] underneath). At the city end [of University square] is [the law building]. [Across Grat- tan street] [from University square] there is [an entrance] [to the campus], straight ahead is [an overpass building] and to the right are the various [Engineering buildings]. The road goes in a big loop around [the campus] you can either go left [towards the Medical buildings] or right passed [the En- gineering buildings] then head North (away from University square). One other area you may want to explore is [South Lawn] which you can get to by going [underneath the over- pass building] directly in front of you when you enter [the campus].

This is an example taken from the campus descriptions. The triplets output by the program are shown in Table-34 . We observe that the static spatial relations come up as desired. For instance, the program infers the complex triplet relation:

<an entrance across Grattan Street

from University Square>

with its complete inherent information using the DLEs re- lated to it viz. [Across Grattan street], [from University square] and [an entrance]. Again for formalizing triplets, the ‘across from’ relation can be broken analogously to that for ‘between’ relation in the previous example. However, it can be seen that not all triplet relations were extracted by the program. There are indirect references in the phrases “straight ahead .. Engineering buildings ” and “In the other

4 The first and the third sentence were split at clause bound- aries to eliminate Stanford parser errors. See Section 5.3.2

Table 3: Program output for spatial triplet relations in Example 2

|  |
| --- |
| <We in Baretto> |
| <We in Alan Gilbert Building> |
| <one of medical buildings across Grattan street> |
| <the Peter Doherty Institute Down the hill> |
| <the hill along Grattan street> |
| <the law building at the city end of  University square> |
| <an entrance across Grattan street  from University square> |

direction .. ”, and spatial motion in the sentence, “The road goes .. ”.

**5.3 Investigating Errors**

From the results obtained on the test corpora, there was a clear indication to dependency of validity of the produced triplets on the accuracy of Stanford typed dependencies and the DLE parser. The invalid cases also correspond to the finding subject methodology for the DLEs mainly due to its inability to disambiguate spatial sense in the subject. In other cases, it was the informal or incorrect usage of gram- mar that led to the failure of triplet extraction. We discuss each of these in detail below.

*5.3.1 Invalid Subjects*

The errors corresponding to invalid subjects found can be prominently classified into two types - invalid place name and subjects with unwanted information. Since, there is no spatial sense disambiguation done for the noun phrases, some of the subjects reported turned out to be invalid. For instance, in example ex:halfway, the subject of the DLE along the street is Approximately halfway (a noun phrase) and is deemed correct by the program. Similarly in example

7, subject of at the gate and the house is identified as roses (instead of an arch ) which is not a place object and thus the corresponding triplet does not give any spatial information. However, such errors can be corrected by choosing subjects which are identified as place names by the DLE parser. But this leads to a side-effect of making the production through- put of triplets heavily dependent on the accuracy of DLE parser. Also, the approach of finding subject of a DLE in most informative way leads to extension with undesired in- formation. In the Example 4, we see that Wilson Hall gets undesirably extended to Wilson Hall a multi-function hall

Example 4. Next [to the sandstone core] is [Wilson Hall], a multi-function hall ..

*5.3.2 Limitations Of The Stanford Parser*

The errors corresponding to limitations of the Stanford parser include incorrect POS tagging and hence or otherwise, in- correct typed dependencies. Since, the backbone of subject finding task is the typed dependencies, the major errors that occur in the extraction process is when the dependencies are incorrect. For instance, in Example 5, the dependency parser failed to realize the dependency between private res- idence and along the street. Some of the errors originate from the incorrect POS tags like that in Example 3, where

diagonally is identified as a noun(instead of an adverb), and thus we get an invalid triplet.

Example 5. I am [at a private residence] located on the western side [of Barrington Avenue] [in Kew], approximately halfway [along the street].

Although, a study of the error of these types led to an in- teresting observation. The major cases where the Stanford parser fails to identify the dependencies are those with long sentences containing multiple clauses. This can be fairly jus- tified considering the inexhaustive training behind the Stan- ford parser. However,it was seen that, such long sentences had visible clause boundaries (usually seperated by ‘,’) and if the clause boundaries were detected and the sentence was split into its clauses, the dependencies of the same sentence turned correct. This was made use of wherever possible to study the triplet extraction to eliminate dependency errors. To correct the Stanford parser is clearly out of scope for this paper, but the hints to use the parser surely come out as retraining the parser to parse multi-clausal sentences or to preprocess the data to split the sentences at clause bound- aries before feeding it to the parser.

*5.3.3 Inaccuracy Of The DLE Parser*

Since the task of DLE parser is to identify preposition and the following place names, errors sourced by it are related to reported span of DLE and preposition’s spatial sense. Example 6 is a case of where the lack of disambiguation of preposition sense is exposed. The DLE parser identifies at night as a DLE although at here has no spatial sense, which thus results in an invalid triplet <constant construction at night>.

Example 6. There is [constant construction] which keeps all residents up [at night].

Next example (Example 7) highlights the case where an in- correct span of DLE is reported. The triplet thus reported is <roses at the gate and the house>.

Example 7. There is an [arch] with roses growing on it [at the gate and the house] is a double - [fronted Victorian house].

As compared to the F-score of DLE parser without the man- ual annotations (0.76), the F-score of the parser with anno- tations is very high (0.99)[15]. Though using manual an- notations increases F-score for DLE identification, but it was observed that the recall increases (in some descriptions) when not using the annotations at the stake of accuracy. With the manual annotations, the parser becomes strictly restricted to the annotated place objects and a DLE gets identified only if the associated place object is present in the annotations. But for large enough place descriptions, manual annotations can miss out marking place objects like house, street, wall.

*5.3.4 Incorrect Grammar Usage*

The informal/incorrect usage of grammar becomes a major obstruction while crowdsourcing for place descriptions. This also includes the concatenation of sentences without appro- priate punctuations. Unfortunately in such cases, neither the Stanford parser, nor the DLE parser is robust enough to correct the errors.

**5.4 Missing Triplets**

The results of triplet extraction had a good precision in static place descriptions which had no indirect/exphoric place references or spatial motion. But talking about the corpus on the whole, besides dependencies missed by the Stanford parser, the inability of the approach to handle the limit- ing cases results in a low recall in terms of total expected triplets. For example, in Example 3, though about 80% of static relations were extracted, only about 40% of the total expected triplets (i.e including those representing dynamic relations and indirect references) were reported. This is un- derstandable as parser has no access to motion indicators and no specific approach targeting reference resolution. We leave the rest of the discussion on spatial motion and refer- ence resolution for Section 6.

**6. CONCLUSIONS AND FUTURE WORK** This paper addresses the task of extracting spatial infor- mation by using triplet representation. We have come up with a simple yet effective reasoning approach to understand static place descriptions which does not make use of any ex- ternal resources(like maps, path geometry). The attempt to find informative and computationally convenient triplets sets up a good potential medium for translating verbal de- scriptions to graphical depictions. Our experiments indicate a good precision for extracting static spatial relationships. We have also provided an analysis on the use of DLE parser suggesting that learning prepositional sense can improve the accuracy of the parser to identify DLEs. We highlight that the use of DLE parser is well suited for understanding static place descriptions with no indirect place references. Fur- thermore, using Stanford dependencies parser with place de- scriptions can be fruitful if the parser is re-trained to work on sentences with multiple clauses. And it stays as a crit- ical tool to explore spatial language further by aiming at descriptions involving motion.

This attempt to understand place descriptions exposes sev- eral gaps in understanding place descriptions. One obvious direction of extending the research and making it more ro- bust is to migrate to a case based reasoning approach on a training dataset to build a natural mechanism that can be integrated into . Understanding natural language instruc- tions to robots and infering paths using the knowledge of maps and path geometries has been successfully dealt with [8] but the work on grounding spatial motion using the for- mal study of linguistics and spatial cognition (and no exter- nal resources) is still an open area of research. [17] provides the linguistic study on expression of motion. [10] provides an annotation scheme to study motion detection in NL text. [2] addresses the task of identifying spatial roles in such an annotation scheme and the results indicate a weaker perfor- mance in a situated description comprising complex text.

The other obstacle found in the performance of our approach

was its inability to resolve references. In a situated place de- scription, people frequently make use of indirect references. The indirect references can be exophoric as well as implicit. For example, “Near this cornering point, you occassionally find a crepes stand ” makes use of exophoric reference, but “To the West is the Baillieu Library ” makes use of an im- plicit reference. Thus, the task of resolving indirect refer- ences subsumes the task of exophoric reference resolution and the previous well-studied domain to resolve exophoric references can not deal completely with these indirect refer- ences. However, it was observed that the indirect references used in human descriptions go in a depth-first fashion esp. to describe descriptions involving path intersections. It is nat- ural to describe one path as far as you could and then come back to the others, rather than providing path descriptions in a parallel fashion. For example, in the desctiption , “[..] there are three main alternative paths that you can use to head north. The central one is up some stairs [..]. The path will take you to the [..] where there is [..] and a little to the east, the Union House. Union House is a large building con- taining [..]”. And then the reference ends and the describer switches back to the other branch to explore a new depth, “The second path [..] from the south entrance takes you a lit- tle bit to the west and then switches north [..]. Near this [..], you find a crepes stand. From this path [..]”. (Here, words like “this path” , “the road” represents the current reference). And then again it switches back, “The third path [..]”. How- ever, if there is a single path to describe, it comes out as a tree with single branch. Hence, it hints towards an idea of parsing for references in a depth-first fashion to resolve the demonstrative pronouns like “this building” where it would mean the current reference, i.e. the root of the subtree in progress. However, the complexity behind this approach is to infer the decision of whether moving further on the depth of the tree or to add a new branch to the existing depth.

**7. REFERENCES**

[1] Tell Us Where. [http://telluswhere.net.](http://telluswhere.net/) Location based

mobile game.

[2] E. Bastianelli, D. Croce, R. Basili, and D. Nardi.

Unitor-hmm-tk: Structured kernel-based learning for spatial role labeling. In Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Association for Computational Linguistics, 2013.

[3] J. Bateman, T. Tenbrink, and S. Farrar. The role of conceptual and linguistic ontologies in interpreting spatial discourse. Discourse Processes, 44(3):175–212,

2007.

[4] M.-C. De Marneffe and C. D. Manning. Stanford

typed dependencies manual. URL [http://nlp.](http://nlp/) stanford. edu/software/dependencies manual. pdf, 2008.

[5] M. Grubinger, P. Clough, H. Mu¨ller, and T. Deselaers.

The iapr tc-12 benchmark: A new evaluation resource

for visual information systems. In International

Workshop OntoImage, pages 13–23, 2006.

[6] A. Herskovits. Semantics and pragmatics of locative expressions. Cognitive Science, 9(3):341–378, 1985.

[7] D. Klein and C. D. Manning. Accurate unlexicalized parsing. In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1, pages 423–430. Association for Computational

Linguistics, 2003.

[8] T. Kollar. Learning to understand spatial language for robotic navigation and mobile manipulation. PhD thesis, Cambridge, MA, USA, 2011. AAI0823797.

[9] T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on, pages

259–266, 2010.

[10] O. Kolomiyets, P. Kordjamshidi, S. Bethard, and M.-F. Moens. Semeval-2013 task 3: Spatial role labeling. In Proceedings of the 7th International Workshop on Semantic Evaluation, volume 2, pages

255–262, Atlanta, Georgia, USA, 2013. Association for

Computational Linguistics.

[11] P. Kordjamshidi, S. . M. Bethard, and Marie-Francine.

Semeval-2012 task 3: Spatial role labeling. In Proceedings of the SemEval-2012 Evaluation Exercises. Association for Computational Linguistics, 2012.

[12] P. Kordjamshidi, M. Van Otterlo, and M.-F. Moens.

Spatial role labeling: Towards extraction of spatial relations from natural language. ACM Trans. Speech Lang. Process., 8(3):4:1–4:36, Dec. 2011.

[13] M. Levit and D. Roy. Interpretation of spatial

language in a map navigation task. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 37(3):667–679, 2007.

[14] C. Linde and W. Labov. Spatial networks as a site for the study of language and thought. Language, pages

924–939, 1975.

[15] F. Liu. Automatic identification of locative expressions from informal text. Master’s thesis, Melbourne, Victoria, Australia, 2013.

[16] M. Macmahon. Walk the talk: Connecting language, knowledge, and action in route instructions. In In Proc. of the Nat. Conf. on Artificial Intelligence (AAAI, pages 1475–1482, 2006.

[17] I. Mani and J. Pustejovsky. Interpreting Motion: Grounded Representations for Spatial Language, volume 5. Oxford University Press, 2012.

[18] C. Matuszek, D. Fox, and K. Koscher. Following directions using statistical machine translation. In Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, pages

251–258. IEEE Press, 2010.

[19] P. Olivier and J.-i. Tsujii. A computational view of the cognitive semantics of spatial prepositions. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics, pages 303–309. Association for Computational Linguistics, 1994.

[20] I. Tytyk and T. Baldwin. Component-wise annotation and analysis of informal place descriptions. In International Workshop on Place-Related Knowledge Acquisition Research, 2012.

[21] M. Vasardani, S. Timpf, S. Winter, and M. Tomko.

From descriptions to depictions: a conceptual framework. Scarborough, North Yorkshire, UK, 2013. COSIT.

[22] J. Zlatev. Spatial semantics. Handbook of Cognitive

Linguistics, pages 318–350, 2007.