**Comments to author(s)**

The paper is basically well written but I'd recommend a minor english review. For instance:

\* "... section, which not higher ..."

\* "... because the did not constitute ..."

The paper also could go through some structural changes. For instance, the concepts of two zoom levels could be better understood if the authors were able to provide some examples. Specially the part where they say that the description “the pub at Cowgate” is underspecified at a higher zoom level and ok at a lower zoom level.

I feel that if you could provide some examples, it would be easier for the reader to grasp what are your arguments. For instance, in Section 3.5, the authors describe the Data Collection strategy used and say that 36% of the descriptions collected was minimally distinguishing. That concept was never explained before. The same thing applies for the concepts of overspecified and underspecified. In the case of underspecification, the example given comes only at Section 3.6.

Also the evaluation could be given more space, for instance, only at the final remarks the authors explain that the same evaluation approach used with GRE3D3 and other present higher accuracy. I think these data are important and should be given more detail.

Summary of review

The authors present the procedure for building a corpus for REG in portuguese and spanish. They provide an evaluation using SVM classifiers.

Although, in my opinion, the authors fail to compare their approach with others listed on the paper. The paper also needs some minor english and structural revision in order to improve its readability (see comments to the authors section).

**Comments to author(s)**

This paper deals with the construction and (partial) evaluation of a new data set for the use of natural language description of map locations. The paper is very easy to follow and does not contain significant mistakes. Considering, however, that a typical AI -paper should start with identifying a problem, then touching upon one or more solutions proposed solutions proposed and an evaluation of their contribution to a solution of that problem, I could not figure out what exactly the paper is aiming at or what it should contribute to.

A large part of the paper is devoted to a description of the construction the corpus with very little motivation: I'm motivated to learn how an experiment has been conducted if I'm interested in hypotheses that are tested and I'm told which ones. In this case I could not find any hypothesis proposed. The only motivation for constructing a new corpus is an extended set of features. But it do what?

Then imo not very well motivated, the authors start a discussion about an ML-approach to learn descriptions. I cannot judge the ultimate value of ML approaches in this domain, but I found the current analysis quite unsatisfactory: it is well-known that SVM is superior to a large class of other classifiers in this case.

Summary of review

I don't know what the contribution of this paper is. It mainly consists of the description of the experiment done to collect data for a new natural language corpus.

**Comments to author(s)**

Summary of review

The paper provides a corpus of natural language descriptions in Spanish and Portuguese languages. The paper is clearly motivated and well illustrated with examples and empirical evaluations.

this paper could benefit from some additional clarification and evaluation as described below.

In section 3.3 (Procedure), the paper mentioned that "Subjects were required to select the language of the experiment (Portuguese, Spanish or English)" .

The English data is also collected? if yes , any commonality or dissimilarity among these three data set? any empirical findings on this?

In section 3.4 (Materials), how you pick up the singular and plural targets to describe? is that a random choice? what the proportion of the singular and plural targets respectively?

In section 3.6 (Discussion), the paper mentioned that all descriptions containing only the basic type (e.g., “the church”) is removed, the reason for this removal is not totally clear to me.

In general, the descriptions containing only basic type is sometimes discriminating enough to uniquely identify the target object.

Since the Zoom corpus has two different levels of map details, it is interesting to see how the REG approach performs on the two different zoom levels separately. This will give us insight on how context with different granularity affects REG algorithms.