Referring Expressions: From the Good Enough to the Most Useful

Jette Viethen

A number of algorithms for referring expression generation (REG) exist that are in principle capable of producing useful descriptions of a variety of different types and in a variety of different domains. In many cases, a minimal description will do the job. Producing minimal, and non-redundant descriptions, as well as descriptions that contain a small amount of redundant information, appears to be fairly straightforward.

This brings the field of REG into a position where we can take the step up from the basic task of trying to produce a single description that is 'good enough' towards the more advanced task of finding all 'good' or 'useful' REs and possibly even 'the best'. This next step requires a definition of what it means for a referring expression (RE) to be good. We need to start thinking about the problem of how to distinguish a good description from one that is clearly bad; an ultimately, the aim must be to find the characteristics that make one or a number of descriptions the best' or most useful ones.

A variety of factors have a bearing on this sought after definition of goodness of REs. Among these are factors such as domain characteristics and discourse situation, as well as the perspective which we choose to take; that of the speaker or that of the hearer. If we are interested in cognitive modelling, both mimicking speaker behaviour and serving hearer needs are worthwhile endeavours. However, if our aim is to build systems to be deployed in actual real world situations, the hearer perspective should be our main focus. Such systems should produce REs that are easy to interpret for listeners and enable them to quickly identify the target referent.

There are two problems with using traditional collections of human-produced referring expressions as development or evaluation corpora for systems geared towards listener satisfaction. Data collections of this kind are necessarily speaker oriented. On the one hand, they might contain REs that are not particularly useful for a hearer, and on the other hand they might omit REs that are.

For a systematic approach to this problem, we need to start with a constructed set of all logically possible REs and work on reducing this set to those that are acceptable for the hearer and ultimately to those that are ideal. This reduction can be accomplished by applying constraints that blacklist REs based on certain characteristics such as length, number of related objects involved or types of properties used. Also based on these characteristics, we can apply rules to construct a ranked list of the remaining REs. We then need to evaluate the constraints and rules based on success rate and timing data from experiments involving human participants.