**CHAPTER 4**

**CONCLUSION**

In this report, we have presented and addressed the shallow parsing technique and the clause identification problems. Using this technique, the relevant information for each task can be determined. This approach will use sophisticated learning methods when an appropriate definition of the input and output vocabularies is provided. Moreover, this approach maintains the efficiency of the system throughout both the learning phase.

The specialization methods proposed are independent of the corpus and the language used. The lexicalization criteria presented provide sets of words that are very common, such as words that belong to closed categories or words that appear frequently in the corpus.

These selected words can also appear in other English corpora and, therefore, the chunking or clause identification problem could be successfully solved using this technique.

We think the method presented here can be improved in two aspects:

* The selection of the features that have to be included in the input and output vocabularies for each disambiguation task, and the selection of the words that really improve the performance of the system. To do this, it would be necessary to take into account not only statistical criteria, but linguistic criteria as well.
* Due to the fact that this technique does not need to change the learning and tagging processes, we think that the application of this technique using other taggers based on different paradigms could be of interest.