1. PUBLISHABLE SUMMARY

The project DynNetLAc was conceived with the aim of creating dynamic networks in order to help users to overcome the tip of the tongue problem, a kind of anomia, blocking momentarily the access to a desired word.

To achieve this ambitious goal, several sub-goals had to be achieved. They were the focus of DynNetLAc. More precisely, the subgoals are the following:

1. Building lexical networks capable to simulate:
    a. part of the human memory (storage, access and organization of the mental lexicon),
    b. the cognitive process of lexical insertion in language production,
    c. the flexibility of the human mind when modifying and improving the organization and access to the data.
2. Contribute to the theory of man-machine interaction, with the development of hybrid interactive systems where humans and computers must collaborate to reach the goal.
3. Continue the line of research initiated by WordNet to design lexical networks coming closer to what we call the *mental lexicon* (semantic maps representing words and their organization).
4. Evaluate the possibilities to model navigation in complex networks by taking methods from DNA regulatory networks, and study possible methodological interactions between biology and mathematical system to converge towards a linguistic model.

Overall, these objectives have been achieved. We have built networks that have been able to simulate (generate similar results than) humans when associating words. WordNet has played an important role in the research, providing the idea of linking words via a set of associations yielding a network. Moreover, some techniques of complex networks have been taken to model simple navigation algorithms. The less developed objective has been the design of man-machine interaction systems. The reason has been the poor compatibility between Python and dynamic webs, an issue that has being addressed by a number of researchers to be (probably) solved in the near future.

There are two lines of development in the project: the scientific contribution and the personal training and expertise. First I will describe the work and results achieved in the former part, concerning the contents of the research.

One of the main goals of the project was using graph-based techniques to approach problems related to lexical storage and organization. Contrary to our initial idea, a non-labeled graph has been used in order to simulate the working of human connections with words. The Wikipedia abstracts and the British National Corpus have been used as corpora to design the network. A co-occurrence network generates a lot of noise with the function words (hence, secondary information) as main hubs. This has led us to design graphs containing only (Verbs), Nouns, and Adjectives, and to consider only the shortest distance, 1, immediate neighbors.

Next, some navigation and dynamization techniques have been implemented, using mainly neighborhood algorithms, and very simple implementation of the rules. Starting from some clues, the user should be able to find a target word (ie, a word that he cannot access). The weights of the arcs have been taken into account in order to rank the resultant nodes.

We have tested the system comparing our best-ranked words with the primary responses retrieved by humans in the Edinburgh Association Thesaurus (EAT). Even though the results have been quite good, they could be improved with respect to matching our first word with the one given by the EAT.

In order to improve the performance of the system, a clustering algorithm has been developed, based on the WordNet synsets. Even though, at the time of closing the project, the results with clustered graphs are in progress and have not been published yet.

Unfortunately I was not able to reach the 'online' parameter of the last goal of the proposal: the design of an online interactive system to search the graphs with given clues. The reason for this lie in the impossibility to integrate Python in the dynamic web design. Also, I did not have the needed time to implement a Perl program able to do the same as the Python original.

Concerning the training objective of the project, a series of concrete goals were established in the proposal, which are the following:

1. Broadening the scope of my research.
2. Acquiring new methodologies of research: use of corpus and data bases, application of statistical methods, testing and training the systems, etc.
3. Developing new skills on the computational implementation of formal and mathematical linguistic theories.
4. Starting new ways of collaboration with centers explicitly working in applied natural language processing.

Concerning the training objectives, DynNetLAc has been a very successful project. As I expected, I acquired a lot of new research methodologies, going from a very theoretical framework to being able to process corpora, using statistical methods and designing applications for users. My new skills also include the capacity to study lexical graphs and to apply statistical analysis to data obtained from corpora. So, the scope of my research has been broadened and my capacity of interaction with the Natural Language Processing community has improved significantly.

The training achieved during these two years has been a good opportunity to increase the collaboration between mathematics – especially statistics – computer science and linguistics. With this background I will be able to make new interdisciplinary contributions to the understanding of natural language generation.

In general, the results of the project are very positive, not only in the training and expertise, as described above, but also from the scientific point of view. I think this project opens a new line of research, the use of automatically-generated non-annotated resources to simulate human cognitive capabilities. More research is needed in the area, but we have demonstrated that, with very simple techniques, for example, statistically-based functions, the main processes of natural language generation can be described, especially the tip-of-the-tongue problem, which is an excellent example of the difficulties humans have to access lexical items stored in their minds.