U.PORTO

Investigação na Pré-Graduação
Education and language in Memories of Labour
Projecto IPG 118
Final Report

Coordinator: Nelma Moreira

October 2008

The project was developed between July 2007 and July 2008, and the research team included different science domains: Linguistics (Joaquim Barbosa - FLUP) , Education Sciences (Maria Teresa Medina, João Caramelo - FPCEUP), (Cristina Nogueira , Silvestre Lacerda - CIIE), History (Manuel Loff - FLUP), and Computer Science (Nelma Moreira (coordinator), Rogério Reis, Norberto Lopes - FCUP & LIACC). Several students were also part of the team: Mara Matias (Mestrado em Ciência de Computadores, FCUP), Bruno Monteiro (Mestrado em Sociologia, FLUP), Adriana Marcela Veiga Pinho Ferreira, Carla Patrícia Figueiredo, Marlene Sousa Carvalho, Pedro Afonso Lebre da Rocha, Raquel Rodrigues Monteiro (Licenciatura de Ciências da Educação, FPCEUP), Carla Susana Fernandes Silva, Carlos Manuel Rodrigues da Silva, Ivânia Magalhães da Silva Ribeiro, João Emanuel Leitão

Ramalho, Joana Marcela Ribeiro de Sousa Barbosa da Rocha, Juliana Silva da Cunha (Licenciatura de Língua e Literaturas Modernas, FLUP).

As stated in the proposal, the project aimed to contribute, in a multi-referential perspective, to understand the processes of interaction between work, personal formation and professional identity of workers, in Porto, during the second half of the 20th century, through the analysis of biographical narratives that are being collected by the "Centro de Documentação e Informação sobre o Movimento Operário e Popular do Porto" (CDI) of Universidade Popular do Porto [Uni]. The project had two main research axes that aim to understand the processes of interaction between work, personal formation and professional identity. The first of these axes focus on a linguistic and social-linguistic analysis, the second one focus on the education/formation processes, both of them interconnected with research in computational linguistics and semi-structured data processing.

The CDI has a corpus of about 80 interviews with workers of different professions and with different social experiences . All biographical narratives have been recorded in audio and video, and a transcription of the interview was produced (in a total of more than 140 hours of audio recordings and 3500 pages). Both groups of students used this corpus for their research.

The work of Linguistics and Computer Science students developed essentially according to the task I of the proposal. The Computer Science student begun by analyze some of the existing software tools for natural language processing and specially for Portuguese. For corpora tagging, a tree tagger tunned with special Portuguese rules was chosen for this work. Several lexical statistics analysis where implemented based on a compact storage of the tagged text: word occurrences, syntactic categories occurrences and occurrences of syntactic patterns identified in this project by the linguistics students and reported below. Because the corpus of this project is based in XML documents with different specifications (abstracts, interviewers information, transcripts, etc), XML APIs were also studied and used to extract the relevant information and to built new XML documents.

The Linguistics students analyzed several syntactic patterns, namely: coordination, cases of subordination (temporal, concessive and completive clauses) and passive constructions. Based on the usual grammatical definitions and on examples they extracted several simple rules (patterns) that allowed (some) automation. This work was not trivial as these students do

not have any formal or computational training during their degree, and so it was also innovative and a challenge. It is relevant to notice that the interplay of both linguistics' and computer science students was not an easy task, as their technical languages and tools were different. The working meetings (also with members of the research team) allowed the necessary approximation of methodologies and linguistic codes. Based on the linguistic patterns, a first set of software tools were developed for their detection and filtering. The information collected from the corpus allowed a first statistical analysis which results were partially reported in Silva *et al* [SSR$^+$07].

For a deep analysis of the semantic values and uses of the several linguistic constructions it was necessary to have better software search tools that allowed the selection of relevant parts of speech. The classification of this information was specified in the generic categorical hierarchy XML schema developed by the project team and described in [LLMR08]. For each transcription, an XML file conforming this schema was produced. Those files can be visualized and edited in a dedicated editor for content analysis manipulation that was also developed in the scope of the project (see below).

Four narratives were selected for analysis and the preliminary results were reported in Silva *et al* [SSR$^+$08]. From the initial set of linguistic constructions, the passive was not considered because no relevant information was available.

Finally, a dynamic website was developed for interfacing with some of the lexical analysis, as well as for statistical analysis and syntactic patterns identification.

The contribution of Education Sciences and Sociology students was essentially according to the task II of the proposal. This work begun by editing and printing the transcriptions and a digital copy of the interviews (audio and video) was organized in order to constitute a permanent archive for the project. Through the first analysis of the transcriptions of the interviews (task II. 2), a synthesis was elaborated that establishes, for each interviewed, sociologic data, labour personal biography, political and trade unionism experiences and the participation in other associative movements, and sketches the formal and informal relationship nets, in which they had been involved.

In order to annotate and to classify the biographic narratives, all transcriptions were automatically converted from simple text (Microsoft Word)

to XML documents conforming a specific language developed by the project team (task II.3). A first version of a transcriptions' classifier was developed that allows the creation of categorical hierarchies and the association of text segments. The documents produced can be also exported to HTML formats.

Tasks II.4-8 were developed by the students through the identification of narratives topics, a deep analysis of a selected set of narratives, the identification and the construction of new categories, and the resulting analysis and its theoretical interpretation. Resulting from this work the student of Sociology presented two oral communications [Mon08, Mon]. The Education Sciences' students produced several academic works, namely the promotion of two training sessions under the topic *"Percursos de Vida e Formação"*, based on the analysis of some narratives and in the context of the course *Psicossociologia da Formação de Adultos*. Two students produced a final report for the course *Seminário de Iniciação à Mediação e à Formação* [dRF08].

Participation in other activities

The students participated in several seminars and a short course, within the activities of the research project *"Memórias do Trabalho - processos de construção de uma identidade operária no Porto"* (POCTI/CED/60786/2004), which allowed a better integration of the students with the multidisciplinary research team.

The main activities were the following[1]:

- *"Memórias do Trabalho"* (12 sessions, 24h), course promoted by the Universidade Popular do Porto, with the support of FPCE-UP and FCUP, 01-03/2008.

- Seminar *"Memórias do Trabalho - questões metodológicas e epistemológicas"*, 12/12/2007.

- Open working sessions: *"História, memória e cidadania europeia"*, with Giulia Strippoli, Museo Diffuso della Rezistencia, Turim and *"Transformações do mundo do trabalho e intervenção dos sindicatos"*, with Manuel Freitas and Gonçalo Duarte.

- *"Memórias e Vivências do 25 de Abril"*, 30/04/2008, round table with the presence of some of the interviewees which biographical narratives were collected by the CDI.

---

[1]More information can be found in the Web-site `http://memorias.ncc.up.pt`:

4

A digital copy of the reports and papers produced within this project can be found in the webpage `http://www.ncc.up.pt/ nam/elml`.

References

[dRF08]   Pedro Afonso Lebre da Rocha and Adriana Marcela Veiga Pinho Ferreira. Restituir à história os seus protagonistas: das memórias biográficas à aprendizagem experiencial. Technical report, Faculdade de Psicologia e Ciências da Educação, UP, 2008. Relatório final da unidade curricular Seminário de Iniciação à Mediação e à Formação, Projecto II.

[LLMR08] Silvestre Lacerda, Norberto Lopes, Nelma Moreira, and Rogério Reis. A toolkit for an oral history digital archive. In José Carlos Ramalho and João Correia Lopes, editors, *Actas XATA 2008, XML: aplicações e tecnologias associadas*, pages 40–51, Universidade de Évora, 2008.

[Mon]   Bruno Monteiro. Men, beasts and machines. on the vocabularies of insurgence and indignity among industrial workers. Session 18 The linguistic transformation of public spaces, (chair: Federico Farini), RC25 Language and Society. 6/09/2008.

[Mon08]   Bruno Monteiro. A incorporação da vocação militante. apontamentos sobre as lógicas da adesão e a geração de disposições políticas nas organizações operárias. Jornadas pela memória das lutas pela liberdade, organizadas pelo movimento cívico Não apaguem a memória! e pela Associação José Afonso, 2008. C. M. Matosinhos, 11 de Abril de 2008.

[SSR+07] Carla Silva, Carlos Silva, Ivânia Ribeiro, Joana Rocha, and Juliana Cunha. Relatório preliminar - ipg 118. Technical report, Faculdade de Letras, UP, 2007.

[SSR+08] Carla Silva, Carlos Silva, Ivânia Ribeiro, Joana Rocha, and Juliana Cunha. Relatório final - ipg 118. Technical report, Faculdade de Letras, UP, 2008.

[Uni]   Universidade Popular do Porto. Centro de Documentação e Informação sobre o Movimento Operário e Popular do Porto. `http://cdi.upp.pt/`.