

THEMATIC PROJECTS

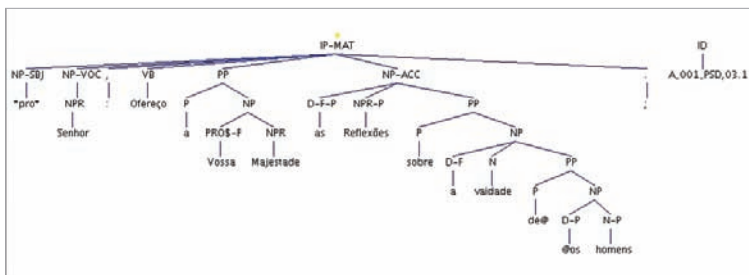
RHYTHMIC PATTERNS, PARAMETER SETTING AND LANGUAGE CHANGE

Charlotte Marie Chambelland Galves

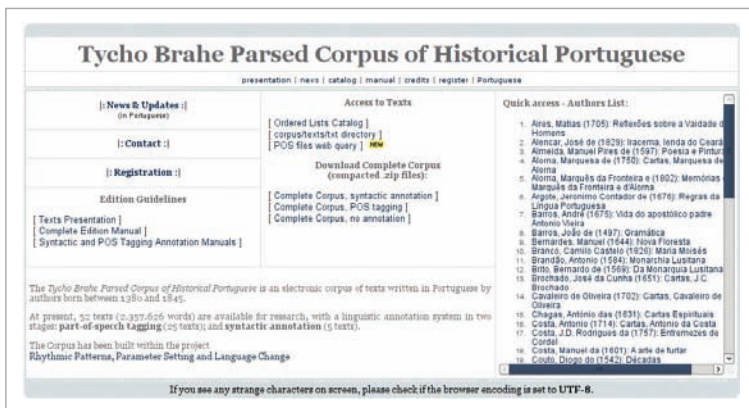
Institute of Language Studies / State University of Campinas (Unicamp)

FAPESP Process 2004/03643-0 | Term: Jan 2005 to Feb 2010

Main researcher: Maria Filomena Spatti Sandalo



Syntactic annotation of the first sentence of « Reflexões sobre a vaidade dos homens », by Matias Aires (born in 1705)



Front page of the site of the Tycho Brahe Corpus

The project “Rhythmic patterns, parameter setting and language change, Phase II” is the continuation of the homonymous project (<http://www.ime.usp.br/~tycho>). Its first goal is to broaden and consolidate the “Tycho Brahe Parsed Corpus of Historical Portuguese” (<http://www.ime.usp.br/~tycho/corpus>), developed during the former project, in the following directions:

- Diversification of the types, periods and spaces of production of the texts. The extended Corpus will contain texts from Portuguese authors born during the 14th and 15th centuries, texts whose authors are uncertain or unknown, non-literary texts, and texts produced in Brazil.

- Syntactic parsing of part of the texts.
- Restructuring of the Corpus in XML language, according to the international patterns of Corpus codification.

The second goal of the project is to use the revised Corpus to deepen the study of Middle Portuguese, the intermediary phase between Old Portuguese and the modern variants of the language, with the following questions in mind:

- What are the grammatical features of Middle Portuguese?
- What is its trajectory in time?
- What is the role of prosodic change in the emergence of Modern European Portuguese?

The project is anchored in one of the main research lines of modern linguistics with the goal of understanding what yields linguistic change and how this change proceeds in time. As in the previous project, the poorly addressed issue of the interaction between rhythm and syntax in change is emphasized. Moreover, we face the methodological challenge of the detection of prosodic patterns in written texts. Finally, by describing and analyzing the language state that gave birth to the two main modern variants of Portuguese, we establish the basis for a comparative history of Brazilian and Modern European Portuguese.

SUMMARY OF RESULTS TO DATE AND PERSPECTIVES

The Tycho Brahe Corpus is currently composed of 52 texts, for a total of 2,357,626 words, written in Portuguese by authors born between 1380 and 1845. It is freely accessible on the worldwide web. All of the texts are formatted in XML, with 25 of them available with a corrected tagging for Part of Speech (POS) and 5 syntactically parsed. The project has developed a morphological system for POS annotation and a syntactic parsing scheme, which is currently being applied to additional texts. Initially restricted to authors born in Portugal from the 16th century on, the Corpus now includes texts by Brazilian authors and Portuguese authors born before 1500.

As a result of this large amount of data, it was possible to confirm the hypothesis of the existence of an intermediate grammar of Portuguese between Old Portuguese and Modern European Portuguese, called Middle Portuguese (MP). This grammar has V2 features and is strongly pro-drop. The exact nature of the formal properties of the categories responsible for the positions of the arguments in clauses is currently being investigated.

The project has produced several empirical pieces of evidence suggesting that the grammar of Modern European Portuguese (EP) begins to compete with MP in the texts written by individuals born at the beginning of the 18th century.

The detailed study of MP and how it changed to EP provides new tools for a comparative history of Portuguese in Portugal and in Brazil. In texts written in Brazil during the 19th century, we see the reflection of the change that occurred in Portugal 100 years earlier. The comparison of MP and EP also sheds new light on the much-debated aspects of the structure of the clause in EP.

The project has developed a methodology to detect the fingerprints of rhythm in written texts along two lines of research. First, on the basis of a sample of texts of the corpus and in collaboration with the Portuguese Free P Project, the researchers have produced a preliminary mapping of the phonological changes that affected EP between the 17th century and the 19th century. Second, using probabilistic tools, we have been able to discriminate modern Portuguese and Brazilian rhythms in written texts. We are currently applying these tools to historical texts. The junction of the two approaches will allow us to locate the rhythmic change in time and to empirically support the hypothesis that the phonological change preceded the syntactic change.

Finally, we propose an analysis of clitic placement that accounts for the way prosody and syntax interact and can be integrated into a model of language acquisition in which a change in the former can provoke a change in the latter.

MAIN PUBLICATIONS

Frota S, Galves C, Vigário M. 2008. Ler a fonologia: do português clássico ao português europeu moderno. In Frota S, Santos AL. (eds.) *Textos Seleccionados do XXIII Encontro Nacional da Associação Portuguesa de Linguística*: 193-206.

Galves A, Galves C, Garcia NL, Leonardi F. 2009. Context tree selection and linguistic rhythm retrieval from written texts. Can be retrieved from ArXiv:0902.3619.

Galves C. 2007. A língua das caravelas: periodização do português europeu e origem do português brasileiro. In De Castilho A, Torres MMAT, Lopes RV, Cyrino SL (eds.) *Descrição, história e aquisição do português brasileiro*, Campinas, Pontes: 513-528.

Galves C, Moraes MAT, Ribeiro I. 2005. Syntax and morphology in the placement of clitics in European and Brazilian Portuguese. *Journal of Portuguese Linguistics*. **4**: 143-177.

Galves C, Namiuti C, Paixão de Sousa MC. 2006. Novas perspectivas para antigas questões: revisitando a periodização da língua portuguesa. Endruschat A, Kemmler R, Chafer-Pries B. (eds.) *Grammatische strukturen des Europäischen Portugiesisch*, Tübingen: Calepinus Verlag: 45-74.

Galves C, Paixão de Sousa MC, Britto H. 2005. The change in clitic placement from classical to modern European Portuguese: Results from the Tycho Brahe Corpus. *Journal of Portuguese Linguistics*. **4**: 39-67.

Paixão de Sousa MC, Trippel T. 2006. Building a historical corpus for classical Portuguese: some technological aspects. Proceedings of the V International Conference on Language Resources and Evaluation (LREC 2006).

Sandalo F, Abaurre MB, Mandel A, Galves C. 2006. Secondary stress in two varieties of Portuguese and the Sotaq optimality based computer program. *Probus*. **18**: 97-125.

Charlotte Marie Chambelland Galves

Instituto de Estudos da Linguagem/Unicamp
Departamento de Linguística
Rua Sérgio Buarque de Holanda, 571
CEP 13083-859 – Campinas, SP – Brasil

+ 55-11-3521-1541
galvesc@unicamp.br