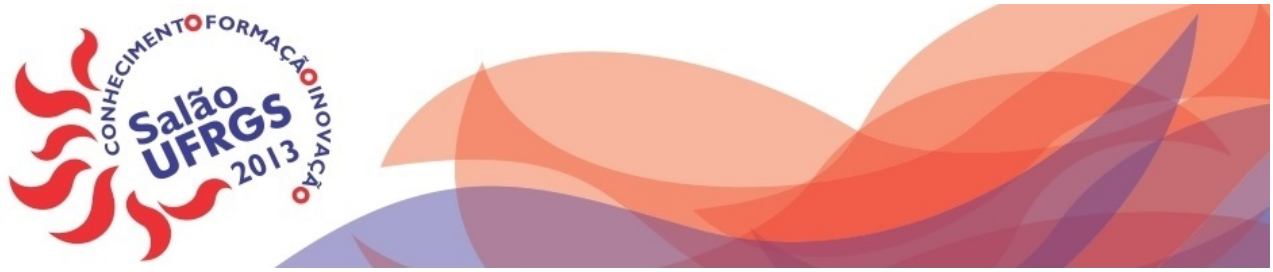

Evento	Salão UFRGS 2013: SIC - XXV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2013
Local	Porto Alegre - RS
Título	Anotação automática do corpus do VARSUL
Autor	MÔNICA RIGO AYRES
Orientador	GABRIEL DE AVILA OTHERO

Nosso trabalho pretende contribuir com a melhoria do etiquetador automático morfossintático Aelius, que é um etiquetador criado pelo professor Leonel Alencar da Universidade Federal do Ceará. Um etiquetador morfológico automático nos permite submeter um texto e conseguir, automaticamente, que as palavras sejam uma a uma etiquetadas conforme sua categoria, por exemplo, verbo, artigo, adjetivo, etc. As etiquetas utilizadas no Aelius são as mesmas do corpus de português histórico Tycho-Brahe (corpus eletrônico já anotado, composto de textos em português escritos por autores nascidos entre 1380 e 1845). Esse Corpus é desenvolvido junto ao projeto temático Padrões Rítmicos, Fixação de Parâmetros & Mudança Linguística da Universidade Estadual de Campinas.

Os textos que utilizamos para a anotação são do banco do projeto Varsul, que estuda a variação linguística na região Sul do Brasil e conta com a parceria de quatro universidades brasileiras, a saber: Universidade Federal do Rio Grande do Sul, Pontifícia Universidade Católica do Rio Grande do Sul, Universidade Federal de Santa Catarina e Universidade Federal do Paraná. Inicialmente, foram etiquetados automaticamente oito textos. A partir da etiquetagem feita pelo Aelius, corrigimos o que o etiquetador ainda não conseguiu detectar corretamente. A partir dos erros do etiquetador, buscamos depreender certos padrões de anotação para superar limitações apresentadas pelo programa.

Um dos impasses principais nessa tarefa é que o etiquetador Aelius foi programado para etiquetar textos escritos, e os textos do projeto Varsul são de língua falada. Sendo assim, há bastante limitação quando o anotador encontra alguns tipos de construções comuns na língua falada, como marcadores discursivos, nomes próprios compostos, hesitações, derivação imprópria, ambiguidades lexicais, etc.

Além de contribuirmos com uma maior eficiência do etiquetador Aelius, pretendemos também propor à equipe do Varsul um etiquetador automático de qualidade, tendo em vista que os textos anotados automaticamente podem auxiliar várias pesquisas linguísticas.