



Corretor Gramatical Para o Emacs

Thiago Maciel Batista
Professor Orientador: Marcelo Finger

Instituto de Matemática e Estatística da Universidade de São Paulo
thiguns@gmail.com



Resumo

O editor de texto emacs nunca teve um corretor gramatical para a língua portuguesa do Brasil. Para suprir essa necessidade, este trabalho teve como principal objetivo acoplar um corretor gramatical ao emacs. O corretor usado no acoplamento foi o cogroo. Na figura 1, podemos ver um trecho de texto no emacs, corrigido e com o erro destacado.

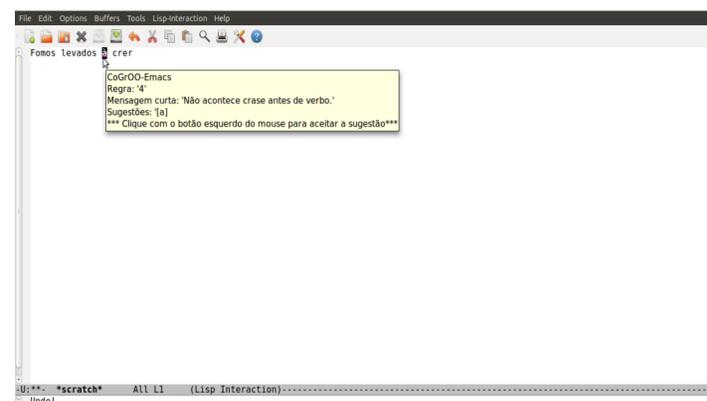


Figura 1: Texto com um erro, que está sendo destacado pelo emacs.

Também fez parte do trabalho o estudo de processamento de linguagem natural e o estudo do cogroo.

1. PLN Estatístico

PROCESSAMENTO estatístico de linguagem natural (PLN estatístico), emprega conceitos de estatística para processar uma linguagem. Para isso, usa-se um corpus para treinamento, modelos probabilísticos e técnicas de aprendizado de máquina.

2. CoGrOO

É um corretor gramatical da língua portuguesa do Brasil feito, a princípio, para o OpenOffice.org, que utiliza PLN estatístico.

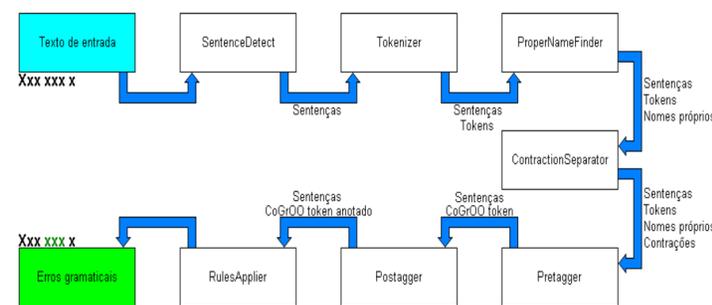


Figura 2: Principais módulos do cogroo e o fluxo de processamento de texto (Ilustração gentilmente cedida por Edwin Miguel)

- *SentenceDetect*: Separa o texto em sentenças.
- *Tokenizer*: É o módulo que separa uma sentença em tokens.
- *ProperNameFinder*: Identifica nomes próprios.
- *ContractionSeparator*: Responsável por identificar contrações.
- *Pretagger*: Prepara os tokens para receber informações morfológicas.
- *Postagger*: Coloca a classificação morfológica dos tokens.
- *Rulesapplier*: Aplica as regras gramaticais e gera uma lista com os erros.

3. Como Foi Feito o Acoplamento

O acoplamento foi feito por meio de interfaces no cogroo e no emacs. Essas interfaces se comunicam para executar a correção.

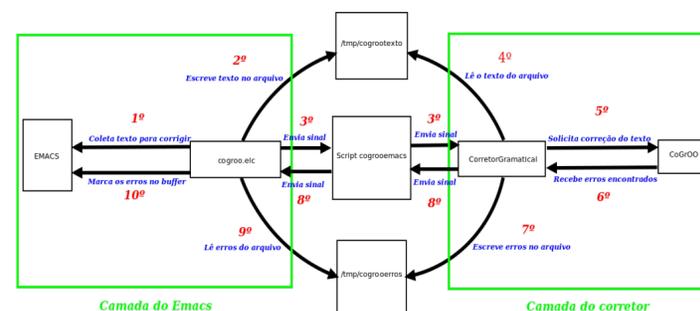


Figura 3: Modelo do acoplamento. Os números indicam a sequência de passos para corrigir um texto

4. Conclusões e Resultados

Conseguiu-se um resultado satisfatório, e o emacs já é capaz de usar o cogroo.

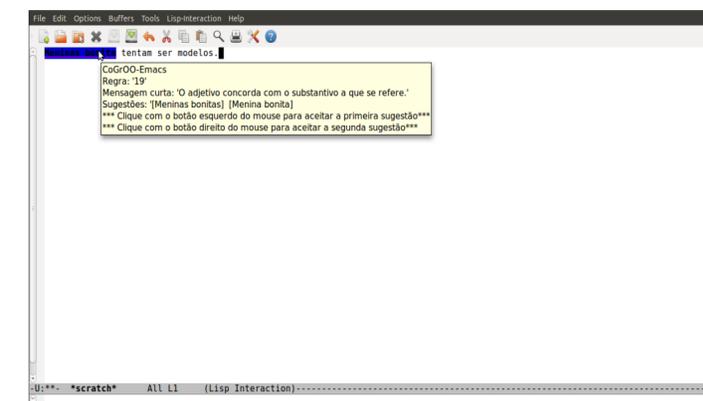


Figura 4: Emacs destacando erro no texto. Neste caso há duas sugestões de correção para o erro

Com isso, mostra-se a mobilidade do emacs e do cogroo e a importância do processamento de linguagem natural atualmente.

5. Trabalhos Futuros

- Disponibilizar para os usuários.
- Usar sugestões para melhorias.
- Aperfeiçoar correção.
- Reportagem automática de erros.
- Aprendizado de máquina com os erros reportados.

6. Referências

- [1] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT press, second edition, 1999.
- [2] GNU. Gnu emacs. Disponível em: <http://www.gnu.org/software/emacs/>. Acesso em: 2010.
- [3] William Colen. Cogroo - corretor gramatical acoplável ao openoffice.org. Disponível em: <http://cogroo.sourceforge.net/>, Acesso em: 2010.