

CoGrOO¹ – Um Corretor Gramatical para a língua portuguesa, acoplável ao *OpenOffice*

Jorge Kinoshita

Universidade de São Paulo (USP), Escola Politécnica
São Paulo – SP – Brasil
jorge.kinoshita@poli.usp.br

Laís do Nascimento Salvador

Univ. Cruzeiro do Sul, Centro de Ciências Exatas e Tecnológicas
São Paulo – SP – Brasil
lais.salvador@poli.usp.br

Carlos E. Dantas de Menezes

Faculdades SENAC – Departamento de Ciências Exatas
São Paulo, SP – Brasil
carlos.edmenezes@sp.senac.br

Abstract

This paper describes an ongoing Portuguese grammar checker project for OpenOffice, called CoGrOO. Two of its features are highlighted: - hybrid architecture, mixing rules and statistics; - free software project. Therefore, this project has a social appeal too, because citizens and corporations can save money on software licenses.

Keywords: natural language processing, machine learning, statistical processing of language, free software.

Resumo

Este artigo descreve a construção em andamento de um corretor gramatical para a língua portuguesa, acoplável ao pacote de escritório OpenOffice. Entre os diferenciais deste projeto, podem ser citados: - arquitetura híbrida, ou seja, alguns módulos funcionam através de estatísticas e outros são baseados em regras; - projeto de software livre. Assim, este projeto propicia a difusão de conhecimento técnico, além do fato de possuir um forte apelo social, pois possibilita a economia com licenças de software.

Palavras chave: processamento linguagens naturais, aprendizado automático, processamento estatístico da linguagem, software livre.

1. INTRODUÇÃO

A ferramenta *OpenOffice* vem sendo amplamente adotada por um número crescente de usuários pessoais e corporativos. Trata-se de um pacote para escritório, composto de: processador de texto, planilha de cálculos, editor HTML e editor de apresentação. O *OpenOffice* é um projeto de código aberto e multi-plataforma, ou seja, pode ser executado sobre diferentes sistemas operacionais, como Windows, Linux, Solaris, etc. As funcionalidades do OpenOffice são comparáveis às de alguns pacotes de escritório não livres (ou proprietários) bem conhecidos, como o *Microsoft Office*. A popularização do uso desta ferramenta deve-se a dois fatores principais: o fato de ser *software* livre, e sua qualidade técnica. Porém, mesmo com uma boa aceitação, os usuários reclamam da falta de um módulo de correção gramatical para a língua portuguesa acoplado à ferramenta. Este recurso já existe nos equivalentes proprietários, e é de grande valia

¹ Este projeto é patrocinado pela FINEP, Chamada Pública MCT/FINEP/Software Livre CT-INFO/FINEP-01/2003.

na construção de textos das mais diversas áreas de conhecimento.

Em linhas gerais, os textos escritos estão sujeitos a diversos tipos de erros [5], entre os quais:

Erros Ortográficos: quando o usuário digita uma palavra com letras trocadas ou invertidas, por exemplo: sugeito (forma sugerida pelo `br.ispell` [7]: `sujeito`).

Erros Gramaticais: ocorrem quando a estrutura gramatical da sentença não segue as regras gramaticais vigentes, por exemplo “Nós vai andar a cavalo”. Envolve erros de concordância nominal e verbal, crase, regência nominal e verbal, etc.

Erros de Estilo: em determinados contextos, o uso de estruturas sintáticas complexas e palavras incomuns causa sérios prejuízos ao entendimento da sentença, por exemplo “Ouviram do Ipiranga as margens plácidas/De um povo heróico o brado retumbante,/E o sol da Liberdade, em raios fúlgidos,/Brilhou no céu da Pátria nesse instante”. Nesses versos do Hino Nacional Brasileiro há algumas palavras incomuns para os nossos dias, como `plácidas`, `fúlgidos` e `retumbante`, além de diversas inversões na ordem natural (sujeito-verbo-predicado). Este estilo seria inapropriado para a escrita técnica.

Erros Semânticos: tais erros são fortemente dependentes do contexto e ocorrem quando a sentença, mesmo correta do ponto de vista ortográfico, gramatical e de estilo, não apresenta uma idéia coerente, por exemplo “O caminhão come bananas”.

O objetivo do projeto CoGrOO é construir um protótipo de corretor gramatical acoplável ao *OpenOffice*, capaz de detectar erros no texto digitado, tais como erros de concordância nominal e verbal, crase, regência nominal e verbal, e outros erros comuns na língua portuguesa falada no Brasil (por exemplo, o uso do “mal”/“mau”, a confusão entre o uso de “a”/“há”), e apresentar alternativas de revisão dos mesmos.

Este artigo descreve a construção deste corretor gramatical para a língua portuguesa. Além do diferencial de ser um projeto de software livre, este projeto também apresenta contribuições nas abordagens técnicas implementadas. O texto é organizado da seguinte forma: a próxima seção fornece uma visão geral do corretor, sua arquitetura e seu funcionamento; a seção de conclusão apresenta as contribuições deste trabalho, como também uma análise comparativa com trabalhos correlatos.

2. ARQUITETURA

O sistema CoGrOO é composto por módulos responsáveis pela análise da sentença e por módulos de detecção de erros gramaticais. O primeiro módulo do corretor gramatical é o Separador de Sentenças: ele recebe o texto do *OpenOffice*, separa e armazena as sentenças contidas no texto. O segundo módulo, o Etiquetador Morfológico, recebe como entrada cada uma das sentenças e atribui etiquetas morfológicas aos seus itens lexicais (palavras e pontuações).

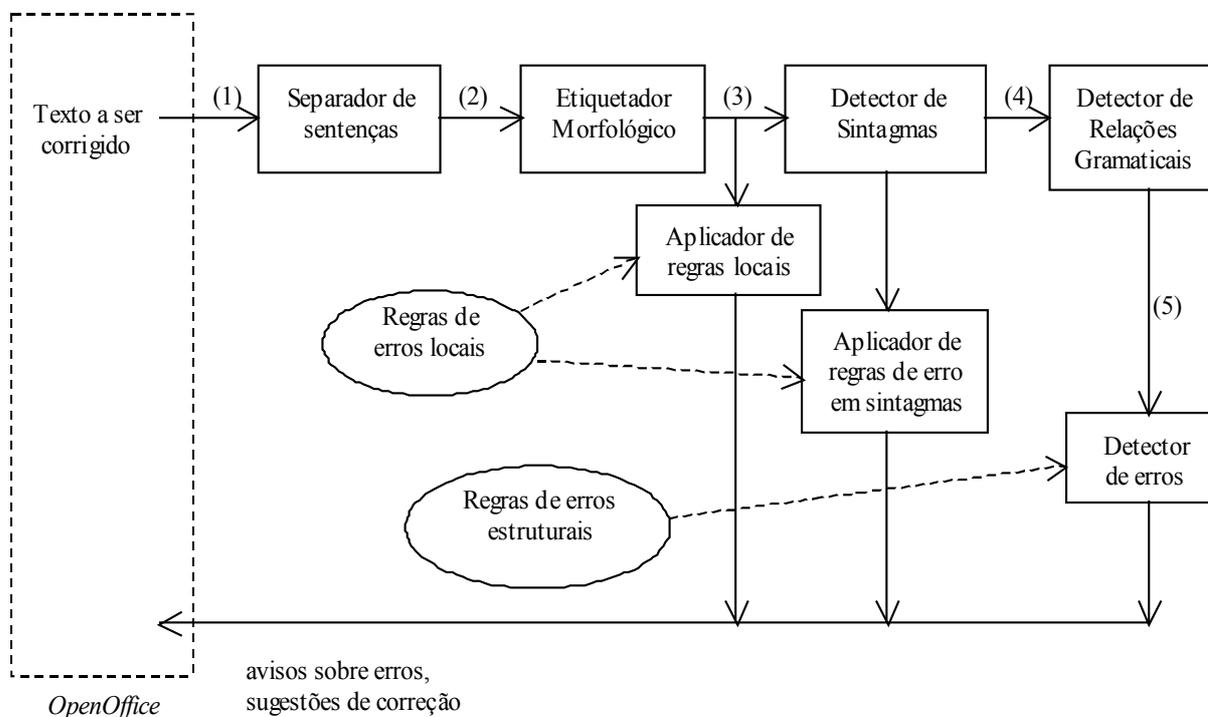


Figura 1 – Arquitetura do sistema CoGrOO

Após a classificação morfológica, submete-se o texto a outros dois módulos: o Detector de Sintagmas e o Detector de Relações Gramaticais; o primeiro agrupa seqüências de nomes/artigos/adjetivos (sintagmas nominais) ou seqüência de verbos (sintagma verbal); o segundo atribui relações gramaticais a estes agrupamentos, estabelecendo os papéis gramaticais (sujeito, verbo, predicado), quando possível, de cada agrupamento.

A arquitetura do CoGrOO com os seus principais módulos está representada na Figura 1. Como se observa nesta, existe um módulo de análise de erro para cada etapa da análise da sentença. A primeira fase do processo de correção gramatical se aplica a uma seqüência curta de palavras e/ou etiquetas morfológicas com base em regras de erros, chamadas regras locais. Para o tratamento de erros mais complexos como erros de concordância verbal, na segunda fase é implementada uma abordagem combinada de detecção de relações gramaticais e de aplicação de regras de erros estruturais. Nas regras estruturais são usadas seqüências de palavras e/ou etiquetas gramaticais associadas ao texto analisado. O processo de revisão de erros gramaticais é acionado inicialmente após a etiquetagem morfológica e é também realizado concomitantemente ao trabalho de detecção de sintagmas e de atribuição de relações gramaticais. Nos itens seguintes, será descrito mais detalhadamente cada módulo do sistema.

2.1. Separador de sentenças

A função do separador de sentenças é preparar o texto para o etiquetador morfológico. Ele recebe como entrada o texto a ser analisado, marca o texto com sinais especiais, levando em consideração os casos de abreviação. Este módulo consulta um dicionário de abreviações da língua portuguesa e sua saída é uma estrutura de dados para cada sentença encontrada no texto.

Segue um exemplo de texto de entrada para este módulo e a respectiva saída produzida:

Entrada (1)²: Ele foi procurar uma casa. Ele vai se casar com a Sra. Maria.
Saída (2): Ele foi procurar uma casa . (1^a. sentença)
 Ele vai se casar com a Sra. Maria . (2^a. sentença)

Pode-se observar que o ponto em “Sra.” não foi confundido com um ponto final, pois se trata de uma abreviação.

2.2. Etiquetador Morfológico

O papel do etiquetador é associar uma etiqueta morfológica para cada palavra da sentença. Para a atribuição das etiquetas morfológicas, o etiquetador executa as seguintes tarefas, nesta ordem: (1) associa etiquetas possíveis às palavras da sentença; (2) avalia a probabilidade de cada etiqueta; (3) define a etiqueta mais provável para cada palavra da sentença (examina o contexto).

A tarefa de se associar etiquetas possíveis às palavras das sentenças é realizada através do acesso aos bancos de dados léxicos (chamaremos “dicionários”), que são gerados pelo processamento de um *corpus* anotado [3], procedendo assim uma tarefa de aprendizado automático. Dessa forma, tenta-se associar a cada palavra uma ou mais etiquetas.

Caso não seja possível encontrar a palavra no dicionário, tenta-se descobrir a sua etiqueta por checar as últimas letras dessa palavra (sufixo) num o dicionário de sufixos, também gerado pelo processamento de *corpus* anotado. Caso exista este sufixo, adota-se esta etiqueta. Caso não exista, a palavra é etiquetada como substantivo singular, visto que esta é a etiqueta mais comum em nossa língua.

Como através da consulta aos dicionários é possível a associação de mais de uma etiqueta à mesma palavra, é necessária a aplicação uma técnica para resolução de ambigüidades. Neste trabalho é usada a técnica de trigramas com base em cálculos de probabilidades:

- monta-se um banco de dados com todos os trios de etiquetas morfológicas presentes no *corpus* anotado, com suas respectivas freqüências;
- a heurística envolve levar em conta tanto a freqüência de cada etiqueta associada àquele item lexical quanto a freqüência dos trigramas possíveis que incorporam esta etiqueta. Da combinação mais provável é escolhida a etiqueta para a classificação deste item lexical.

Para ilustrar o trabalho do etiquetador, segue a simulação para a 1^a. sentença citada na seção anterior:

Entrada (2): Ele foi procurar uma casa .

A tabela 1 mostra as etiquetas morfológicas e suas respectivas notas (obtidas pela heurística comentada) associadas a cada item lexical. Percebe-se que, a menos do caso do ponto final, sempre houve necessidade de lidar-se com situações de ambigüidade.

² Os números entre parêntesis referem-se à Figura 1.

Palavras	Etiquetas Morfológicas
Ele	PERS_M_3S_NOM_(16740) PERS_M_3S_NOM/PIV_(62) N_M_S_(6)
foi	{ir}_V_PS_3S_IND_VFIN_(2753) {ser}_V_PS_3S_IND_VFIN_(55500) ADV_(138)
procurar	{procurar}_V_FUT_1/3S_SUBJ_VFIN_(1) {procurar}_V_INF_(759) {procurar}_V_INF_0/1/3S_(10) {procurar}_V_INF_3S_(33) {procurar} V_FUT_3S_SUBJ_VFIN_(15)
Uma	NUM_F_S_(12037) DET_F_S_(96697) NUM_M/F_P_(32) NUM_M_P_(2) NUM_F_P_(16)
Casa	{casar}_V_PR_3S_IND_VFIN_(114) N_F_S_(6034) {casar}_V_IMP_2S_VFIN_(1)
\$.	-PNT_NS(1108953)

Tabela 1 – Saída do Etiquetador Morfológico (3)

A palavra “casa”, por exemplo, pode ser um substantivo feminino singular (N_F_S), um verbo no presente (V_PR) ou um verbo no imperativo (V_IMP). Através das notas associadas a essas etiquetas, escolhe-se N_F_S como a mais provável para “casa”.

2.2.1. Aplicação de regras locais

A partir do texto etiquetado morfológicamente é aplicada a primeira bateria de regras de erros gramaticais. Estas regras são denominadas locais, pois envolvem um contexto muito curto, de poucas palavras ou etiquetas à esquerda ou à direita, sem levar em conta a estrutura da sentença. Este processo de revisão gramatical aponta alguns erros comparando em tempo real a seqüência de etiquetas morfológicas (substantivo, verbo, numeral, artigo etc.) das palavras digitadas com padrões armazenados nas regras locais. Caso a seqüência apresente uma ordem incomum, o sistema procura e apresenta a ordem correta, ou a mais comum.

Exemplos de erros detectados nesta fase: uso de crase antes de palavra masculina ou de verbo, o provável uso flexionado do advérbio meio, etc.

Uma entrada no arquivo de regras locais possui 4 componentes:

- a regra de erro, isto é, uma expressão regular com palavras e etiquetas morfológicas que especifica um padrão de erro gramatical;
- a mensagem explicativa sobre o diagnóstico do erro;
- um exemplo do uso adequado do padrão gramatical;
- um exemplo do uso não adequado do padrão gramatical.

<p>"à" N_M_S_</p> <p>mensagem: O sinal indicativo de crase indica que temos “a + a” expressos em um só “à”. Somente ocorre crase quando há encontro de preposição a com artigo ou pronome demonstrativo a/as.</p> <p>correto: Refiro-me ao trabalho remunerado./Refiro-me a trabalho remunerado.</p> <p>incorreto: Refiro-me à trabalho remunerado.</p>
<p>"(meia meias)" _ADJ</p> <p>mensagem: A palavra “meio” usada no sentido de “um pouco” é advérbio, portanto invariável. Exceção: A palavra “meia” usada como substantivo e portanto concorda com o adjetivo.</p> <p>correto: A conclusão está meio confusa/ As conclusões estão meio confusas.</p> <p>incorreto: A conclusão está meia confusa/ As conclusões estão meias confusas.</p>

Tabela 2 – Exemplos de regras locais (vide Figura 1)

Na tabela 2 são apresentados dois exemplos de regras locais: a regra para detecção do uso de crase antes de palavra masculina singular, cuja etiqueta morfológica é N_M_S_; e a regra para detecção do uso do advérbio meio antes de um adjetivo, cuja etiqueta morfológica é _ADJ.

2.3. Detector de sintagmas

O objetivo deste módulo é localizar pequenos sintagmas nominais ou verbais, também chamados *noun phrase* (NP) e *verbal phrase* (VP). Um sintagma é uma unidade da análise sintática composta por um núcleo e outros termos associados, formando uma locução – conjunto de palavras que equivalem a um só vocábulo – que faz parte da estrutura da sentença. Os sintagmas se classificam de acordo com os seus elementos nucleares, neste trabalho detecta-se o sintagma nominal (NP), quando o núcleo do sintagma é um nome; e o sintagma verbal (VP), quando o núcleo do sintagma é um verbo. A implementação do detector de sintagmas é baseada numa máquina de estados que varre a sentença etiquetada morfológicamente na busca de seqüências de artigos, substantivos e adjetivos, no caso do sintagma nominal; ou seqüências de verbos e advérbios no caso dos sintagmas verbais.

Voltando ao exemplo apresentado nas seções anteriores:

Entrada (3):

Ele:[PERS_M_3S_NOM_] foi:[{ir}_V_PS_3S_IND_VFIN_] procurar:[{procurar}_V_INF_] uma:[DET_F_S_] casa:[N_F_S_] \$,-[PNT_NS]

Saída (4):

Palavras	Etiquetas Sintáticas
Ele	NP_3S_M_0_
foi	VP_3S_0_
procurar	
uma	NP_3S_M_1_
casa	-PNT_NS
.	

Tabela 3 – Saída do Detector de Sintagmas (4)

Esta saída (tabela 3) indica que foram detectados três sintagmas na sentença: dois nominais (NP) e um verbal (VP). Além desta saída, este módulo fornece a informação sobre a posição das palavras aglutinadas. A etiqueta de um sintagma nominal/verbal apresenta as seguintes informações adicionais:

- a pessoa associada ao sintagma – no exemplo, os três sintagmas estão na 3ª pessoa do singular (3S);
- o gênero associado ao sintagma, informação restrita aos sintagmas nominais – no exemplo, os dois sintagmas nominais são do gênero masculino (M);
- a última parte da etiqueta (os números 0 e 1) representa uma numeração dada aos sintagmas dentro da sentença.

2.3.1. Aplicação de regras de erro em sintagmas

Na detecção de um sintagma nominal são aplicadas regras de erros locais, como por exemplo: erros de concordância entre o artigo e substantivo, entre substantivo e adjetivo.

A tabela 4 apresenta dois exemplos de regras de erro aplicadas nesta fase:

<p>N_F_S_ ADJ_F_P_ mensagem: O adjetivo concorda em gênero (masculino ou feminino) e número (singular ou plural) com o substantivo a que se refere. correto: Faces rosadas da criança ou a face rosada da criança. incorreto: Face rosadas da criança.</p>
<p>DET_M_S_ (N_M_P_ N_F_S_ N_F_P_) mensagem: Os artigos definidos (o, a, os, as) concordam em número (singular ou plural) e em gênero (masculino ou feminino) com o substantivo a que se referem. correto: Os copos caíram da mesa. A roupa está no guarda-roupa. As roupas estão no guarda-roupa. incorreto: O copos caiu da mesa. O roupa está no guarda-roupa. O roupas estão no guarda-roupa.</p>

Tabela 4 – Exemplos de regras de erro em sintagmas (vide Figura 1)

2.4. Detector de relações gramaticais

O objetivo deste módulo é levantar relações gramaticais como sujeito-verbo, verbo-objeto, verbo-preposição, na sentença a ser analisada. Na fase atual do projeto está sendo implementada a detecção da relação sujeito-verbo. Para isso, foi implementada a seguinte estratégia: no texto analisado é detectada uma seqüência de etiquetas e sintagmas que se encaixam com um padrão que indica que um determinado sintagma nominal é sujeito de um sintagma verbal. Um exemplo de padrão sujeito-verbo é onde "!" significa começo de sentença. Nesse caso o corretor verifica a concordância verbal entre NP e VP.

Os padrões de relações sujeito-verbo foram extraídos do *corpus* anotado com informações morfo-sintáticas. Se esse padrão se repetir muitas vezes é uma indicação de uma relação gramatical válida. No *corpus* observa-se que o para o padrão : "! NP VP", NP é sujeito de VP em 85% dos casos. Devido à esta alta freqüência adota-se "! NP VP" como um padrão para se detectar sujeito nas sentenças a serem analisadas pelo corretor gramatical. Um padrão corresponde a:

<elemento> NP <elemento>* VP , ordem direta (sujeito antes do verbo)

VP <elemento>* NP <elemento> , ordem inversa

,onde <elemento> = ! ou etiqueta ou sintagma

Os padrões encontrados no *corpus* servem de base para uma máquina de estados que é usada pelo corretor gramatical na detecção de erros de concordância entre sujeito e verbo. Uma vez que um padrão foi encontrado na sentença analisada é produzida uma etiqueta sintática [SUBJ>] associada ao NP e a etiqueta sintática (M)VP associada a VP.

Voltando ao exemplo em questão:

(2) Ele foi procurar uma casa.

O detector de relações gramaticais produz a seguinte saída (5):

```
| - [ [SUBJ>]_M_3S_ ]
| - [ (SJ>)SN_M_3S_0_ ]
| - [ SN_M_3S_0_ ]
| - PERS_M_3S_NOM_ --> Ele
| - [ (M)VP_3S_0_ ]
| - {ser}_V_PS_3S_IND_VFIN_ --> foi
| - {procurar}_V_INF_ --> procurar
| - [ SN_F_3S_1_ ]
| - DET_F_S_ --> uma
| - N_F_S_ --> casa
| - -PNT_ABS --> $.
```

O corretor detectou o sintagma nominal “Ele” como sujeito da sentença e produziu a etiqueta sintática [SUBJ>]_M_3S_ (sujeito masculino na 3ª. pessoa do singular). Foi também detectado o sintagma verbal “foi procurar” e foi produzida a etiqueta sintática (M)VP_3S_0_ (M, verbo principal, na 3ª. pessoa do singular, relacionado ao sujeito).

2.4.1. Processo de Correção Gramatical

Durante a fase de detecção de relações gramaticais, o processo de correção gramatical é implementado de duas formas:

- através da verificação de informações contidas nas etiquetas sintáticas;
- baseada em regras de erros estruturais.

Um tipo de erro detectado com base na inspeção das etiquetas sintáticas é o erro de concordância verbal. O processo é bastante simples: com as informações contidas nas etiquetas sintáticas, verifica-se a correspondência entre a pessoa do sujeito e a do verbo a ele associado.

No exemplo anterior tem-se uma correspondência entre a pessoa do sujeito e a do verbo, no caso, 3ª. pessoa do singular, e não é acusado erro de concordância verbal. No caso da frase “Os corretores foi procurar uma casa”, será produzida a seguinte seqüência de etiquetas e o seguinte aviso de erro gramatical:

Sentença: Os corretores foi procurar uma casa.

```
| - [ [SUBJ>]_M_3P_ ]
  | - [ (SJ>)SN_M_3P_0_ ]
    | - [ SN_M_3P_0_ ]
      | - DET_M_P_ --> Os
      | - N_M_P_ --> corretores
| - [ (M)VP_3S_0_ ]
  | - {ser}_V_PS_3S_IND_VFIN_ --> foi
  | - {procurar}_V_INF_ --> procurar
| - [ SN_F_3S_1_ ]
  | - DET_F_S_ --> uma
  | - N_F_S_ --> casa
| - -PNT_ABS_ --> $.
```

Mensagem de erro: Verificou-se erro de concordância entre o verbo e o sujeito.

Neste exemplo, o sintagma “Os corretores” recebeu a etiqueta sintática [SUBJ>]_M_3P, que não concorda em número com a etiqueta sintática (M)VP_3S, associada ao sintagma “foi procurar”.

Outros erros, que envolvem relações gramaticais, estão relacionados ao mau uso da língua em situações bem específicas. Nestes casos são usadas as chamadas Regras Estruturais.

Uma regra estrutural de erro é uma expressão regular que envolve palavras e etiquetas morfológicas, como no caso das regras locais, porém, também engloba informações sintáticas. Um exemplo é o uso do verbo “fazer” na indicação de tempo: foi criada uma regra estrutural que detecta um erro na expressão “fazem 20 anos”. Esta regra só é aplicada desde que a palavra “fazem” não seja um verbo associado a um sujeito na terceira pessoa do plural. Logo, numa frase deste tipo “Fazem 20 anos que não o vejo”, é detectado um erro gramatical através da aplicação da regra estrutural acima. E numa frase como “Hoje eles fazem 20 anos de casados”, não é detectado erro gramatical. Dessa forma pretende-se reduzir as situações de falsos positivos, que uma abordagem com apenas regras de erros locais pode incorrer.

3. RESULTADOS

Embora o projeto CoGrOO não esteja concluído, alguns resultados já podem ser apresentados: O Etiquetador Morfológico foi implementado e possui uma taxa de acerto de aproximadamente 95%. Este etiquetador foi construído a partir do zero, visto que não havia nenhum para a língua portuguesa em código aberto. Na literatura encontram-se vários trabalhos relacionados a etiquetadores [1], [2] e [4]; tais sistemas utilizam um *corpus* manualmente etiquetado para extrair as principais regras e generalizá-las para outros textos não etiquetados, conseguindo altas taxas de acerto. Para o treinamento deste módulo, foi utilizado o *Corpus* CETENFOLHA, disponível em [3].

A primeira versão do aplicador de regras já funciona, apenas para regras locais, ou seja, aquelas que se aplicam somente a uma vizinhança de 2 ou 3 itens lexicais, sem usar a estrutura da frase. Tanto o detector de sintagmas quanto o detector de relações gramaticais estão funcionando parcialmente.

Até o presente momento foram implementadas regras que cobrem parcialmente os seguintes fenômenos lingüísticos:

- crase (ignorando a análise da regência dos verbos)
- casos de concordância nominal (adjetivo com substantivo na ordem direta, tanto na função de adjunto adnominal como o de predicativo de sujeito)
- erros no emprego da expressão invariável "em anexo"
- "meio" no sentido de "um pouco" é advérbio, portanto invariável.
- erros de concordância de gênero do adjetivo de adjunto adnominal
- verbo fazer indicando tempo
- verbo haver indicando tempo e indicando existência
- concordância de artigo com substantivo
- uso dos pronomes mim e ti
- uso de mal e mau
- regência do verbo “preferir”

O corretor gramatical está sendo implementado na linguagem Perl e a interface com o OpenOffice, em Java e C++. Um protótipo já em funcionamento, pode ser acessado no portal do projeto em <http://cogroo.incubadora.fapesp.br>.

4. CONCLUSÕES

O projeto CoGrOO usa uma abordagem híbrida: estatística (por exemplo, o etiquetador morfológico é totalmente estatístico) e baseada em regras. Um dos diferenciais do trabalho apresentado é a abordagem usada na detecção de relações gramaticais, ou seja, a busca por padrões sintáticos no *corpus* anotado. Não foi encontrada abordagem similar na literatura.

A proposta e a estrutura de regras de erros locais usadas neste projeto são baseadas no proposto por [5], contudo este trabalho não trata de questões relacionadas à análise gramatical da sentença, o que o impede de detectar de maneira plena alguns tipos de erros, como os de concordância e regência verbais.

O projeto mais importante de corretor gramatical da língua portuguesa é o ReGra [6], usado no Microsoft Office. Como o projeto CoGrOO ainda está em andamento, não é possível compará-los quanto à taxa de acertos, porcentagem de falsos positivos e falsos negativos.

O cronograma do projeto prevê a disponibilização do código para dezembro de 2005, conforme acordado com o órgão de fomento, mas o protótipo já pode ser testado através do portal criado (<http://cogroo.incubadora.fapesp.br>).

Agradecimentos

Devem-se profundos agradecimentos a todos os membros da equipe CoGrOO, que muito têm trabalhado para a viabilização do mesmo, entre os quais, Edgard Lemos Jr., Fábio Wang Gusukuma, Marcelo Suzumura, Marcos Yoshio Okamura Oku, Sueli Caramello Uliano e William Daniel Colen M. Silva.

Agradece-se também à FAPESP, Fundação de Amparo à Pesquisa do Estado de São Paulo, pela hospedagem de nosso projeto em sua incubadora (<http://incubadora.fapesp.br>), e à FINEP, Financiadora de Estudos e Projetos, pelo patrocínio via Edital de Software Livre.

Referências

- [1] Brill, E. (1992) “A Simple Rule-Based Part Of Speech Tagger”, Proceedings of ANLP-92, 3rd Conference of Applied Natural Language Processing, Trento, Italy.
- [2] Daelemans, W., Zavrel, J., Berck, P. and Gillis, S. (1996) “MBT: A Memory-Based Part of Speech Tagger-Generator”, In: E. Ejerhed and I. Dagan (eds.) Proceedings of the Fourth Workshop on Very Large Corpora, Copenhagen, Denmark, 14-27.
- [3] Linguateca (2005) “CETENFolha”, <http://www.linguateca.pt/CETENFolha/>, acessado em 15/03/2005.
- [4] Menezes, C. (2000) “Um método para a construção de analisadores morfológicos, aplicado à língua portuguesa, baseado em autômatos adaptativos”, Dissertação de Mestrado, Universidade de São Paulo, Brazil.
- [5] Naber, D. (2003) “A Rule-Based Style and Grammar Checker”, Diplomarbeit Technis Fakultät, Universität Bielefeld, Germany.
- [6] Nunes, M.G.V.; Oliveira Jr., O.N. (2000) “O processo de desenvolvimento do Revisor Gramatical ReGra”, Anais do XXVII SEMISH (XX Congresso Nacional da Sociedade Brasileira de Computação), Volume 1, p.6 (resumo). Artigo Completo na Versão em CD-Rom. PUC-PR, Curitiba, Brazil.
- [7] Ueda, R. (2002) “Dicionário br.ispell”, <http://www.ime.usp.br/~ueda/br.ispell/>, acessado em 29/03/2005.