

# Regras sintáticas livres de contexto na correção automática de Unidades de Leitura

Giancarlo Dondoni Salton

Universidade Federal da Fronteira Sul  
Campus Chapecó - SC  
+55 049 2049 1401  
giancds@gmail.com

Cláudia Finger-Kratochvil

Universidade Federal da Fronteira Sul  
Campus Chapecó - SC  
+55 049 2049 1401  
cf-k@uffs.edu.br

Carlos Andrei Carniel

Universidade Federal da Fronteira Sul  
Campus Chapecó - SC  
+55 049 2049 1401  
carlos.andreic@gmail.com

Braulio Adriano de Mello

Universidade Federal da Fronteira Sul  
Campus Chapecó - SC  
+55 049 2049 1401  
braulio@uffs.edu.br

## RESUMO

A avaliação da compreensão leitora em seus diferentes níveis de proficiência tem utilizado tarefas que possibilitam a utilização de respostas abertas. No entanto, este tipo de tarefa requer esforço maior especialmente na etapa de correção das respostas, exigindo inclusive mais de um avaliador capacitado para a tarefa. Essa demanda, por sua vez, aponta para a necessidade do desenvolvimento e uso de instrumentos que permitam automatizar o processo de correção, desde que níveis aceitáveis de acurácia sejam alcançados. Este trabalho apresenta a especificação e implantação de uma ferramenta para correção automática de questões abertas, no contexto da avaliação de proficiência leitora, observadas as recomendações de padrões internacionais e a composição de Unidades de Leitura, utilizando regras livres de contexto para a reconhecimento das estruturas lexicais e sintáticas das respostas e subsequente pontuação.

## ABSTRACT

Some kinds of assessments, such as reading proficiency, needs essay questions. However, this kind of assignment needs higher efforts on the evaluation step, characteristic who justifies the development and use of tools to automatize this task, since some levels of accuracy could be obtained. This work presents the specification and implementation of an automated scoring software to essay questions, on the context of proficiency in reading, observed international standards and the Reading Units technique, using context-free grammars to recognize structures and subsequent scoring.

## Categories and Subject Descriptors

I.2.7 Natural Language Processing  
K.3.1 Computer Uses in Education

## General Terms

Algorithms, Design, Languages, Verification.

## Keywords

Reading comprehension, Essay Question, Automated scoring, Context-free grammar, Reading Assessment.

## 1. INTRODUÇÃO

O uso de questões discursivas ou abertas é indispensável para muitos tipos de avaliação, como por exemplo, da proficiência em leitura. Em contrapartida, considerando a alta demanda de atenção e cuidado na correção desse tipo de questões, o esforço humano necessário pode inviabilizar testes baseados em questões abertas, especialmente, quando se intenciona a mensuração de um conjunto de competências e se amplia o contingente de participantes a ser testado e avaliado, buscando a confiabilidade da leitura dos dados, para o estabelecimento de metas e planejamento de ações, tendência das políticas educacionais nacionais e internacionais. A dificuldade se instaura diante da necessidade de equipes muito bem treinadas para a realização de correções com o máximo de consonância no processo de avaliação dos dados, implicando em elevados esforços de manutenção desse conjunto de profissionais.

Considerando que o uso de referenciais para padronização de critérios de correção ou avaliação favorece a construção e uso de regras para correção, este trabalho apresenta uma abordagem para a avaliação automática de respostas discursivas baseadas no conceito de linguagens livres de contexto. No escopo do trabalho foi construída uma ferramenta que, a partir de uma gramática livre de contexto específica para cada questão, realiza mapeamento de palavras-chave de cada resposta e atribui conceitos a partir de respostas de referência padronizadas e previamente validadas. O software que executa a correção de respostas utiliza regras descritas em gramática livre de contexto, específicas para cada questão. Os experimentos realizados com respostas que foram submetidas à correção humana demonstraram níveis de acerto acima de 96%.

O artigo apresenta, inicialmente, os fundamentos e características do conceito de unidade de leitura para avaliação de proficiência leitora (compreensão leitora), a partir de critérios empregados em FINGER-KRATOCHVIL (2010) [6], baseado no Programa Internacional de Avaliação de Estudantes – PISA, OCDE (2002, 2010) [3][4]. Na seção 3, discute o uso de linguagens livres de contexto como instrumento base para correção de questões discursivas. Na seção 4, apresenta a arquitetura e implementação do software para correção automática de ULs desenvolvimento do trabalho. A última seção 5 apresenta um experimento com o uso da aplicação. Seguem, ainda, algumas considerações finais e

desdobramentos futuros desta pesquisa e referências bibliográficas.

## 2. AVALIAÇÃO DA COMPETÊNCIA LEITORA POR MEIO DE UNIDADES DE LEITURA: COMPLEXIDADE E DESAFIOS

Considerando os resultados da participação do Brasil, desde a primeira edição do PISA, em 2000, os resultados das avaliações nacionais a respeito da leitura e a crescente necessidade de entender melhor a construção da competência leitora ao longo da escolarização e seus reflexos diretos na construção de conhecimentos, incluindo-se o âmbito universitário, desde 2005, temos dedicado atenção ao desenvolvimento de estudos e pesquisas que possibilitem ampliar a compreensão dos processos envolvidos em torno do tema. O presente trabalho é decorrente da compreensão interdisciplinar da área de estudos da leitura e da necessidade de criar caminhos que permitam entrelaçar, no Brasil, cada vez mais, áreas afins, visando a avanços que possam ser traduzidos e transpostos de forma efetiva ao espaço acadêmico, independente do nível de escolarização, ou seja, da educação básica à universidade.

Inicialmente, em 2004, conhecer a proficiência em leitura do aluno universitário ingressante a fim de melhor trabalhar o seu desenvolvimento era uma das metas dos trabalhos de pesquisa. Na ocasião, a primeira lacuna detectada foi a ausência de um teste que pudesse ser aplicado para isso. Utilizamos as questões da amostra do próprio PISA, cedidas pelo INEP, para a primeira testagem. Contudo, entendemos ser necessário restringir os gêneros para uma compreensão mais focada dos textos que circulam no início da graduação, principalmente o informativo. Embasados nos princípios do PISA, partimos, então, para a criação de Unidades de Leitura, doravante ULs, sua testagem piloto, até a validação do instrumento por meio de análise estatística [6]. Esse processo revelou-se extremamente laborioso. Eram necessários textos com grau de legibilidade semelhante (readability) [10], e detectou-se reduzida pesquisa na área em língua portuguesa. Além disso, a elaboração de questões relativas a esses textos que, por sua vez, comporiam as tarefas nas ULs, demandou a apropriação de um vasto referencial teórico sobre a complexidade dos domínios e habilidades envolvidos no processamento em leitura e a transformação desse conhecimento em tarefas de acordo com a maturidade dos participantes e respectiva escolarização, *i.e.*, acadêmicos iniciantes. Assim, optamos por trabalhar com foco em três dimensões e seus respectivos níveis que os padrões internacionais têm proposto para a avaliação da competência leitora, ou seja, "Recuperação da Informação"–RI, "Interpretação da Informação"– II e "Reflexão e Avaliação da Informação" –RA. Elaboramos um instrumento composto de três ULs, com três textos e 18 questões com respostas abertas (nove) e fechadas (nove), nos diferentes domínios–RI, II, RA. Esse instrumento passou por teste piloto e avaliação de especialistas para a sua calibragem e composição de orientações para sua correção, visando à possível aplicação posterior por outros pesquisadores, mantendo-se a confiabilidade e a memória.

Dessa forma, após a construção das ULs, um dos aspectos que demandou grande atenção e cuidado foi a correção das questões abertas. Mais de um momento de análise foi realizado. Especialistas em leitura fizeram a análise do instrumento e dos critérios de correção. Posteriormente, fizeram a correção

individualmente e discutiram dúvidas e suas possíveis respostas a fim de elencar, para além dos critérios de correção, exemplos de respostas modeladas a partir das respostas apresentadas pelos próprios participantes. Esse processo resultou na percepção de palavras-chave e sintagmas comuns entre as respostas que, por sua vez, deveriam estar presentes a fim de que a tarefa fosse avaliada como adequada. A metodologia mostrou-se eficaz e eficiente; contudo, sua aplicação, restrita a grupos pequenos de participantes, devido à demanda de trabalho para a análise e correção que cada testagem gera, sendo também necessário a permanência de uma equipe de pesquisadores treinada e atuando em parceria – para a averiguação das respostas, por exemplo –, fatores que podem vir a limitar a continuidade e desdobramentos dos estudos (ampliação do banco de itens, criação de novas ULs a partir de outros gêneros, aplicação e correção das ULs por professores sem equipe especializada para suporte, elaboração de novos instrumentos para outras esferas da educação, especialmente, a educação básica, entre outros).

A partir dessa experiência, ficou-nos evidente a necessidade de somar a construção de um instrumento paralelo que permitisse a aplicação a um quantitativo maior de participantes, inicialmente, e que também viesse, no futuro, a permitir a autonomia da aplicação das ULs, em outros espaços acadêmicos, como um recurso de sondagem e planejamento, visando ao desenvolvimento das competências e habilidades em torno do complexo processo da leitura. Dessa forma, a partir do banco de dados composto ao longo da validação do instrumento, isto é, as ULs com seus textos e tarefas, passamos a estudar as possibilidades de construção de uma interface que permitisse a correção das atividades propostas, com base em possíveis padrões de resposta e elementos sintagmáticos com saliência perceptual factíveis nas tarefas e nos exemplos de respostas.

Tratando de questões abertas, nas três dimensões que os padrões internacionais têm proposto para a avaliação, ou seja, RI, II e RA, e os diferentes níveis de complexidade no âmbito de cada um desses domínios (veja [3] OCDE, 2002 e [6] FINGER-KRATOCHVIL, 2010) e considerando a formulação das respostas modelo, entendemos que quanto mais textual e direto for o trabalho com a informação, mais plausível são as possibilidades de elaboração de regras que permitam a avaliação automática das tarefas. Consequentemente, a dificuldade se coloca proporcional ao trabalharmos com tarefas que envolvem a dimensão "Reflexão e Avaliação da informação", em níveis mais altos, pois a previsibilidade das respostas diminui em função da mobilização de conhecimento prévio do leitor (*i.e.*, conhecimento de mundo, conhecimento linguístico, conhecimento textual), oriundo de fontes anteriores e externas ao texto e que devem implicar em diversas justificativas de ponto de vista (OCDE, 2002).

Por esas razões, escolhemos para o início dos trabalhos, a Q3 da UL03, instanciada no domínio "Recuperação da Informação", o mais dependente de informações contidas na própria tarefa (*i.e.*, texto e questão), aspectos que ampliam a automatização da correção. Contudo, a Q3 possui duas divisões em sua pontuação para a sua avaliação, dependendo dos elementos apresentados na resposta formulada pelo participante, e encontra-se nos níveis 4 e 5, os mais complexos para qualquer um dos três domínios–RI, II e RA. Em seu escopo, a Q3 demanda a integração de diferentes partes do texto, para inferir qual(is) a(s) informação(ões) é(são) mais importante(s) para a resolução da tarefa. Seu detalhamento encontra-se na seção 5 quando discutiremos a aplicação do corretor às respostas do banco de dados.

### 3. USOS DE RECONHECEDORES SINTÁTICOS BASEADOS EM LINGUAGENS LIVRES DE CONTEXTO

Na hierarquia de linguagens de Chomsky [8] as linguagens ditas livres de contexto são utilizadas para representar características sintáticas de sentenças. Essas características podem ser descritas na forma de regras gramaticais que, então, são utilizadas nas atividades de geração ou de reconhecimento [7] de sentenças. Os reconhecedores verificam se uma determinada sentença atende as regras sintáticas, descritas na forma de gramáticas, da linguagem a que pertencem. Os geradores, a partir das regras das linguagens, são capazes de gerar sentenças válidas.

O uso de recursos geradores e reconhecedores remete, principalmente, para a área de linguagens formais que trata das teorias fundamentais para o tratamento de linguagens de programação. No entanto, os conceitos e especificações de hierarquia de linguagens têm sido utilizados para uma vasta variedade de aplicações. Entre elas, tratamento de processamento de dicionários e tradução automática [12], seqüenciamento de DNA [11] e processamento de linguagem natural [2].

O processamento de linguagem natural, tema de interesse deste trabalho, esta fundamentado na especificação de estratégias para interpretar diferentes estruturas de sentenças. Essas estruturas, nas linguagens naturais, não seguem formalismos rigorosos, característica que aumenta complexidade para a construção de analisadores, ou corretores, automatizados baseados em formalismos de linguagens.

Nesse trabalho, as regras sintáticas utilizadas para a correção de respostas discursivas são construídas a partir de respostas de referencia. A partir destas respostas, as regras precisam ser especificadas de modo que o mapeamento das respostas, realizado pelo instrumento de correção, alcance um estado final de acordo com a pontuação adequada para a resposta. Um estado final pode ser interpretado como um estágio de correção, após a leitura da resposta inteira, que atribui uma pontuação ou um erro para a resposta.

Nesse modo de uso o formalismo não permite processar, por exemplo, concordância de gênero e número. No entanto, para os propósitos de correção de respostas, conforme objetivo do trabalho, esta restrição não interfere nas etapas necessárias para atribuir pontuação às respostas.

### 4. ARQUITETURA E IMPLEMENTAÇÃO DO SOFTWARE DE CORREÇÃO AUTOMÁTICA

Em processadores sintáticos de linguagens, o reconhecimento de cadeias válidas, ou sentenças, é processado através do mapeamento (transição de um estado origem para um estado destino) dos *tokens* (palavras) conforme regras sintáticas descritas em uma tabela de parsing. Esta tabela é construída a partir da gramática gerativa livre de contexto. O mapeamento da tabela de parsing requer procedimento algorítmico específico de acordo com a técnica de reconhecimento utilizada, considerando as seguintes possibilidades: LSR, LS e LR (incluir referencia/Braulio). Na construção do corretor automático de respostas para as ULs, apresentado neste trabalho, foram utilizadas técnicas alternativas de mapeamento da GLC (Gramática Livre de Contexto) com o objetivo de flexibilizar o

uso dos mesmos módulos de software para aplicar as regras de diferentes gramáticas. A necessidade de uma GLC específica para tratamento de cada questão das ULs, resultando em dois princípios básicos para a continuidade e desenvolvimento de trabalhos futuros, tem sido a principal justificativa para esta decisão de projeto nessa primeira versão do instrumento de reconhecimento.

O processo de correção ou reconhecimento de respostas é exemplificado na Figura 1. Nele, a aplicação, inicialmente, obtém as respostas e as produções da gramática. Em seguida, a aplicação chama uma instância do objeto Processador, enviando a ele as respostas e produções obtidas. O objeto processador desmembra a resposta em seus tokens, valida as respostas e as repassa, junto com as produções da gramática, para uma instância do objeto Pontuador. Este objeto, então, realiza as comparações e atribui as pontuações para cada resposta.

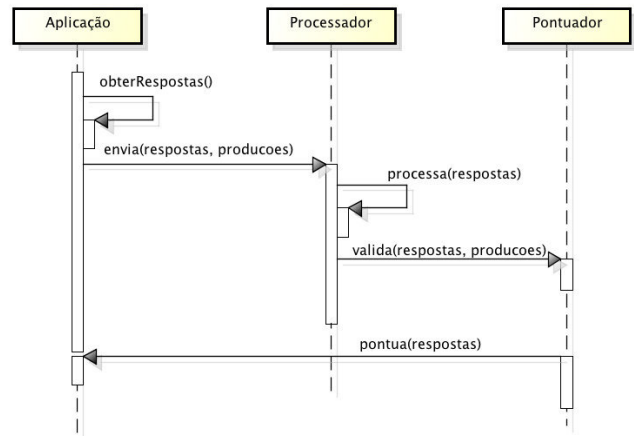


Figura 1. Procedimento para reconhecimento de respostas

A linguagem de programação Scala, utilizada na implementação, possui um recurso denominado “Case Classes” (identificado como feature). Este recurso permite que módulos (ou classes) da aplicação sejam estendidos e posteriormente utilizados como comparação em cláusulas, do tipo *match*, recurso similar ao recurso de “condições aninhadas” usado em programação. Essas cláusulas permitem verificar a ocorrência das palavras chave nas respostas, durante o processamento da correção, sem estabelecer requisitos de ordenação na ocorrência.

Para atender ao modelo das *Case Classes*, foi criada uma classe abstrata denominada “Producao”, que não possui nenhum campo variável ou métodos. Esta classe foi estendida em duas novas classes, a “ProducaoTerminal” e a “ProducaoNaoTerminal”, respectivamente. No reconhecimento das respostas, a primeira permite o tratamento dos *tokens* (palavras) e a segunda permite o tratamento dos estados (estágios de correção). Os estados são utilizados para identificar a pontuação de cada resposta após a aplicação das regras descritas na GLC.

A classe “ProducaoTerminal” controla os terminais da gramática. Ela é composta por um atributo chamado código, no formato *String*, que representa um símbolo não terminal e por um atributo no formato lista de *Strings*, que representa os *tokens* da gramática. A classe “ProducaoNaoTerminal” preocupa-se em gerar os estados da gramática para processamento da correção. Ela é

composta por um código no formato *String* que representa o símbolo não-terminal ao qual ela corresponde e por uma lista de “Producoes”. As produções de uma GLC descrevem as regras de correção.

Uma classe chamada “Resposta” foi desenvolvida para modelar os dados, com as respostas dos participantes, em transcrição *ipsis litteris*, constituintes do banco de dados sobre o desempenho de estudantes universitários na área de leitura. Essa classe possui três campos: um identificador numérico, um campo *String* que armazena a resposta a ser corrigida e outro campo numérico que indica a pontuação dada na correção de referência (resultado da correção manual para comparação). Esta classe também possui um campo que recebe o resultado da correção realizada pela aplicação. A classe “Resposta” possui um procedimento que atribui a pontuação resultante da correção automática realizada pela aplicação. O resultado é mantido em um módulo denominado “Respostas”.

A aplicação possui dois módulos principais, o “Processador” e “Aplicacao”. O primeiro realiza as ações de correção e o segundo exibe os resultados realizando o comparativo com o resultado da correção de referência de cada resposta. Neste trabalho de comparação, a aplicação identifica se a resposta possui, em qualquer ordem, os *tokens* definidos pelas regras descritas na gramática. O atendimento integral das regras gera pontuação “2”, o atendimento parcial pontuação “1”, e o atendimento insuficiente pontuação “0”. Os critérios de correção são totalmente controlados nas regras da gramática. Esta característica é fundamental para que a mesma aplicação possa ser utilizada, de forma flexível, no reconhecimento das respostas de quaisquer outras questões bastando, para isso, construir as regras sintáticas específicas para reconhecimento das respostas de cada questão.

## 5. APLICAÇÃO DO CORRETOR AUTOMÁTICO EM QUESTÕES DE DIMENSÃO DOIS

Conforme mencionamos, anteriormente, o corretor automático de ULs foi aplicado às respostas coletadas na aplicação da Questão03 [6], UL03. A questão, “Porque os cogumelos serão mais conhecidos no futuro do que os são hoje? Justifique sua resposta com dados do texto, se possível”, considera em sua pontuação máxima respostas que localizam e inferenciam qual a informação textual é relevante para a resolução da tarefa e fazem referência ao fato de que os cogumelos são alvo de pesquisa no mundo todo, desde 2003, apontando, ao menos, duas razões diferentes, para a ampliação do conhecimento sobre o tema. As respostas poderão conter citações diretas do texto, desde que essa não seja a única forma de resposta. Por exemplo: os alunos poderão mencionar as pesquisas relativas às propriedades nutricionais e/ou farmacológicas; as pesquisas que demonstram alguns resultados tangíveis com pacientes cancerosos; as publicações de estudos feitos com pacientes e/ou na catalogação de diferentes espécies – divulgação dos estudos; a formação de grupo de estudos multidisciplinar para preservação e uso das espécies estudadas; os encontros de pesquisadores; fabricação de medicamentos a partir de cogumelos; os avanços biotecnológicos nos estudos.

Para a pontuação parcial, as expectativas são próximas, pois também são consideradas respostas que localizam e inferenciam qual a informação textual é relevante para a resolução da tarefa e fazem referência ao fato de que os cogumelos são alvo de pesquisa no mundo todo, desde 2003; contudo, satisfaz a questão

o apontar de **uma razão apenas**, para a ampliação do conhecimento sobre o tema. A descrição completa da UL e as respostas de referência utilizadas na construção das regras da GLC utilizada na automação do processo de correção de respostas à Questão03 são apresentadas em [6].

A gramática, apresentada na Figura 2, foi elaborada a partir das palavras (*tokens*) esperadas em respostas corretas para a Questão03. As palavras que apresentam maior relevância estão presentes na maioria das respostas que obtiveram pontuação máxima. As palavras de relevância intermediária para a correção das respostas, ou de ocorrência menos significativa, foram utilizadas para complementar respostas de pontuação máxima ou qualificar respostas de pontuação média. Algumas produções da gramática possuem o prefixo da palavra – de acordo com Formalismos da Linguagem –, desde que a mudança de sufixo não comprometa o processo de reconhecimento para pontuação. Por exemplo, a regra A possui as palavras ‘estud’ e ‘pesq’ que permitem o reconhecimento de estudos, estudiosos, pesquisa, pesquisador. O *software* de reconhecimento realiza o tratamento de sufixo variável. A implementação de estratégias para tratamento de palavras, considerando a variabilidade de sufixo e a probabilidade de ocorrência, coloca-se entre as perspectivas de continuidade do trabalho.

A correção é realizada através da verificação da ocorrência de símbolos terminais (*tokens*) da gramática nas respostas dadas e o estado final em que se encontra a gramática ao fim da verificação. Os símbolos não terminais que dão nome às regras (A,B,C,D,Y,X e Z) são identificados pelo reconhecedor automático como “estados”. Os estados identificados como estados finais, ou estados que coincidem com a leitura do último *token* da resposta corrigida, são utilizados para atribuir pontuação para a resposta corrigida.

```
A::= estud | pesq | sabemos | culinaria | prevencao | descobertas
B::= remedios | medic | funcionamento | qualidades
C::= principios | saude | medic | beneficios | conhec | saber
C::= bem | mundo | recuperacao | terapeutica | farmacologicos
C::= preservacao | descobertas
D::= doencas | cogumelos | recentes | beneficios
Y::= D
X::= A | Y | B
Z::= C | A
```

Figura 2. Descrição das regras de correção da Q03 da UL03 em Gramática Livre de Contexto

Os estados finais são Z e X. O estado final Z indica que a resposta deve conter pelo menos um símbolo de cada um dos estados C e A. O estado final X indica que a resposta deve conter pelo menos um símbolo de cada um dos estados A, Y, e B. Se, ao fim da verificação o estado final for Z, a pontuação conferida à resposta será “1”. Se o estado final for X, a pontuação será “2”. Não tendo sido alcançado nenhum estado final ao fim correção da sentença (resposta) fornecida, a pontuação será “0”.

Respostas que apresentam partes literais do texto deveriam ser desconsideradas e receber pontuação 0 (zero), mas a atual implementação ainda não oferece recursos para desqualificar estas

respostas, visto que elas podem apresentar palavras relevantes que fazem parte da gramática, e assim, são pontuadas pelo corretor. A identificação de respostas com cópia literal do texto, sem indicador de citação, também faz parte das perspectivas para continuidade do trabalho.

Respostas também podem fazer uso de transcrições e, contudo, não constituírem apenas cópia. Por exemplo, na resposta 45 do experimento, *“Porque a tendência é que sejam ingeridos por indicação médica. Tanto, que estima-se que 12mil espécies de fungos produzam corpos de frutificação para serem considerados cogumelos, sendo que 2mil são de grande valor para o consumo. No Paraná cerca de 1700 espécies já foram catalogadas, mas muitas outras ainda devem ser descobertas.”*, não existe relação entre a primeira frase e a segunda frase, considerando que a palavra 'tanto', um conectivo explicativo, deveria introduzir o o fato que serviria de argumento, estabelecendo a relação entre as sentenças. Além disso, foram utilizados muitos dados de partes diferentes do texto que, apesar de não constituírem totalmente cópia, não são objetivos a ponto de garantir correção da pontuação. O reconhecedor não possui mecanismos automatizados o suficiente para interpretar corretamente este tipo de construção. Como perspectiva de trabalho para tratamento de situações desse tipo, a primeira etapa é capacitar o reconhecedor a detectar ou identificar respostas desse tipo gerando, ainda, uma demanda para correção manual (professor).

```

GramaticaModel.scala
class GramaticaModel {
  // Gramática
  val a = new ProducaoTerminal("A", List("estud", "pesq", "sabemos", "culinaria", "prevenc
  val b = new ProducaoTerminal("B", List("remedios", "medicinas", "funcionamento", "qual
  val c = new ProducaoTerminal("C", List("principios", "saude", "medic", "beneficios", "co
  val d = new ProducaoTerminal("D", List("doencas", "cogumelos", "recentes", "beneficios"

  // Símbolos Não terminais
  val y = new ProducaoNaoTerminal("Y", List(d))
  val x = new ProducaoNaoTerminal("X", List(a, y, b))
  val z = new ProducaoNaoTerminal("Z", List(c, a))
  // Lista de símbolos da gramática
  def producoesDaGramatica: Producoes = List(a, b, c, d, x, y, z)

  // Lista de pontuações.
  def gramaticaPontuada = List((List(z), 1), (List(x), 2))
}

Console
<terminated> Application[Scala Application] C:\Program Files\Java\jre6\bin\javaw.exe (12/07/2013 21:22:34)
Original: 1 - Correcao: 1 - Resposta: 1-Pois com o avanco biotecnologico ja existe a via
Original: 0 - Correcao: 0 - Resposta: 2-Ainda que esse seja um objetivo atingido a longo
Original: 2 - Correcao: 2 - Resposta: 3-Pois descobriram que os cogumelos sao importante
Original: 2 - Correcao: 2 - Resposta: 4-Pois atualmente, iniciou-se um pequeno estudo so
Original: 1 - Correcao: 1 - Resposta: 5-A maioria da populacao, quando se refere ao cogu
Original: 1 - Correcao: 1 - Resposta: 6-Creio que o motivo dos cogumelos nao serem conh
Original: 0 - Correcao: 0 - Resposta: 7-Pela eficiencia ajudando o corpo, dando mais ati
Original: 0 - Correcao: 0 - Resposta: 8-Porto cresce o numero de pssoas que usarao, dev
Original: 1 - Correcao: 1 - Resposta: 9-Porto as pesquisas em relacao ao cogumelos vao
Original: 1 - Correcao: 1 - Resposta: 10-Com o desenvolvimento de uma pesquisa realiza
Original: 1 - Correcao: 1 - Resposta: 11-Pois cada vez mais os cogumelos estao presentes
Original: 0 - Correcao: 1 - Resposta: 12-Para o medico Ricardo Veronezi, o empenho dos p
Original: 1 - Correcao: 1 - Resposta: 13-O cogumelo ainda precisa ser pesquisado mais aft
Original: 1 - Correcao: 1 - Resposta: 14-Porto hoje sabemos muito pouco de seus benefi
Original: 1 - Correcao: 1 - Resposta: 15-Porto que estao surgindo algumas pesquisas cienti
Original: 1 - Correcao: 1 - Resposta: 16-Sim, pois hoje diferentes tipos desses organizm
Original: 2 - Correcao: 2 - Resposta: 17-Pois hoje um dos unicos cogumelos conhecido eh
Original: 0 - Correcao: 1 - Resposta: 18-"A tendencia eh que cada vez mais cogumelos sai
Original: 1 - Correcao: 1 - Resposta: 19-Porto sao feitos muitas pesquisas com cogumelo
Original: 1 - Correcao: 1 - Resposta: 20-Porto muitas das especies existentes ainda nao
Original: 1 - Correcao: 1 - Resposta: 21-Porto com varios estudos iremos poder saber ma
Original: 1 - Correcao: 1 - Resposta: 22-Porto eles estao deixando de ser apenas usados
    
```

**Figura 3. Resultado parcial na execução de um processo de correção de respostas para a Q03 da UL03**

No experimento realizado no escopo do trabalho, 68 respostas foram submetidas ao corretor automático. As respostas utilizadas integram o banco de dados da testagem mencionada anteriormente. Esses dados, especificamente, advém do experimento piloto feito para a calibragem das questões e utilizado para criar as respostas de referência, assim, como a testagem do grupo de pesquisa, transcritos literalmente segundo cada participante. A partir desse conjunto de respostas, as regras da GLC foram especificadas. A Figura 3 apresenta a interface da

aplicação utilizada para o acompanhamento do resultado da correção. A primeira coluna mostra a resposta atribuída na correção manual pelo professor responsável pelo trabalho. A segunda coluna apresenta o resultado da correção automática realizada pelo corretor. Observa-se a correspondência entre ambas as correções, validando, neste experimento, as estratégias de correção que utilizam regras sintáticas livres de contexto para correção de unidades de leitura.

Analisemos alguns exemplos. Na resposta 4, *“Pois atualmente, iniciou-se um pequeno estudo sobre as qualidades e benefícios que os cogumelos podem trazer para o ser humano e suas doenças, principalmente o cancer. Afirma-se que eles serao mais conhecidos futuramente, porque, pesquisas e consequentemente, maiores resultados serao obtidos daqui para frente”*, identificamos a presença de vários dos símbolos esperados para uma resposta correta, além da ideia de que resultados serão obtidos no futuro (através de pesquisas). Os principais argumentos foram utilizados para justificar a resposta, sem a utilização de citações ou cópia do texto original. A palavra “estudo” se encontra entre as produções de A, “benefícios” entre as produções de Y e “qualidades” entre as produções de B. A resposta ainda conta com outros símbolos esperados, como “pesquisas”, “doenças” e “cogumelos”. Cada uma dessas palavras está em pelo menos um estado da gramática que leva ao estado final que determina a pontuação 2 (O estado final X).

Ao observarmos a resposta 15, *“Por que estao surgindo algumas pesquisas científicas que permitiraõ um conhecimento ainda maior”*, percebemos símbolos requeridos para uma resposta correta e a ideia de que as pesquisas avançarão ao longo do tempo. Porém, faltam argumentos como o de que os cogumelos poderão ser utilizados como medicamentos para a cura de várias doenças (o que os tornaria mais conhecidos). A palavra “pesquisas” está entre as produções de A e “conhecimento” está entre as produções de C. A e C são estados que levam ao estado final que determina a pontuação 1 (O estado final Z).

A análise da resposta 37, permite-nos detectar alguns símbolos da gramática. Entretanto, não apresenta nenhuma informação sobre como os cogumelos se tornarão conhecidos e os motivos que poderão levar a isso. Vejamos: *“Por que o conhecimento da populacao na sua grande maioria se reduz ao champignon, nao sao informados, ou nao buscam mais informacoes. Contudo, a visao dessas pessoas não sao alimentadas. Porem, com o tempo, atraves de divulgacos sobre os cogumelos, o aumento da procura e o nivel de aceitacao vai progredir”*. A palavra “conhecimento” está entre as produções de C, e “cogumelos” está entre as produções de D. Porém, símbolos encontrados apenas nesses dois estados não levam a um estado final. Respostas que terminam sua correção sem alcançar um estado final recebem pontuação 0.

Respostas que apresentam partes literais do texto deveriam ser desconsideradas e receber pontuação 0 (zero), mas a atual implementação ainda não oferece recursos para desqualificar estas respostas, visto que elas podem apresentar palavras relevantes que fazem parte da gramática, e assim, são pontuadas pelo corretor. Por exemplo, a resposta 12 possui algumas das palavras esperadas para uma resposta correta e seu estado final é Z, o que determina pontuação 1 no corretor. *“Para o medico Ricardo Veronezi, o empenho dos pesquisadores em dar credibilidade aos cogumelos tem rendido bons frutos. " Hoje a biotecnologia ja viabiliza a identificacao de todas as principis ativos dos cogumelos." " Ainda que esse seja um objetivo a longo prazo, o otimismo ja contagia a*

area de pesquisas.<sup>1</sup> Mas esta resposta é cópia integral do texto original, o que impossibilita uma pontuação maior do que 0, mesmo que a ideia copiada do texto seja relevante para respostas corretas. A identificação de respostas com cópia literal do texto, sem indicador de citação, também faz parte das perspectivas para continuidade do trabalho.

Respostas também podem fazer uso de transcrições, contudo não constituem apenas cópia. Por exemplo, na resposta 45 do experimento, “*Porque a tendência é que sejam ingeridos por indicação médica. Tanto, que estima-se que 12mil espécies de fungos produzam corpos de frutificação para serem considerados cogumelos, sendo que 2mil são de grande valor para o consumo. No Paraná cerca de 1700 espécies já foram catalogadas, mas muitas outras ainda devem ser descobertas.*”. Percebemos que, nesta resposta, não existe relação entre a primeira frase e a segunda frase considerando que a palavra 'tanto' deveria expressar uma relação de causa entre as duas partes da sentença. Além disso, foram utilizados muitos dados de partes diferentes do texto que, apesar de não constituírem totalmente cópia, não são objetivos a ponto de garantir correção da pontuação. O reconhecedor não possui mecanismos automatizados o suficiente para interpretar corretamente este tipo de construção. Como perspectiva de trabalho para tratamento de situações desse tipo, a primeira etapa é capacitar o reconhecedor a detectar ou identificar respostas desse tipo gerando, inicialmente, uma demanda para correção manual (professor).

## 6. CONCLUSÕES

Este trabalho apresentou a especificação e construção de instrumentos computacionais para automatização do processo de correção de respostas discursivas utilizando recursos para tratamento formal de linguagens livres de contexto. A avaliação da proficiência leitora mediante testagem com o uso de Unidades de Leitura, metodologia embasada nos princípios de PISA, gera grande demanda quando da aplicação em grandes grupos. Fato que abre margem para variabilidade de interpretação entre avaliadores distintos e demora nas correções. São justificativas para a construção de instrumentos automatizados.

A correção um conjunto de respostas discursivas para uma UL específica, utilizada como estudo de caso neste trabalho, apresentou comportamento satisfatório. As respostas corrigidas pela ferramenta apresentaram afinidade acima de 90% em comparação com a correção realizada por especialistas humanos para as respostas do grupo de controle do estudo de caso.

Foram encontradas algumas dificuldades, entre elas, meios para criar regras sintáticas capazes de discernir entre citações e cópias de trechos do texto nas respostas. Outro fator limitante no uso de regras no escopo das linguagens livres de contexto, é o tratamento de significados distintos para um mesmo token dependendo do contexto em que é utilizado. Características que requerem tratamento sensível ao contexto.

Os fatores limitantes identificados durante a construção das regras de correção apontam para a continuidade do trabalho em que elementos de probabilidade podem ser incorporados aos instrumentos de correção. Tais elementos podem auxiliar a resolver o problema de ambiguidade citado. Incorpora-se probabilidade a uma linguagem livre de contexto adicionando-se

uma probabilidade de ocorrência a cada uma das produções da gramática (as probabilidades de cada produção de um símbolo não terminal devem ser, somadas, igual a 1). Então, escolhe-se a produção com maior índice probabilístico ao invés de se continuar com todas as derivações possíveis para um mesmo token em uma produção [13].

## 7. REFERÊNCIAS

- [1] Aho, A.V., Sethi, R. and Ullman, J.D. (2007). *Compilers: Principles, Techniques and Tools*. Pearson Education, 2nd edition.
- [2] Bolshakov, Igor A e Gelbukh, Alexander. 2004. *Computational Linguistic: Models, Resources, Applications*. INSTITUTO POLITÉCNICO NACIONAL, México.
- [3] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *Reading for change: performance and engagement across countries*. Paris: Author, 2002.
- [4] Organização para a cooperação e o desenvolvimento econômicos (OCDE), Série: Reviews of National Policies for Education, 2010, ISBN: 978-92-64-09136-8.
- [5] Contier, Ana; Padovani, Djalma; Neto, João José. Linguístico: Uma proposta de reconhecedor gramatical usando tecnologia adaptativa. *Revista de Sistemas e Computação*, Salvador, v.2, n.1, p.70-81, jan/jun 2012.
- [6] Finger-Kratochvil, Claudia. 2010. *Estratégias para o desenvolvimento da competência lexical: relações com a compreensão em leitura*. Tese apresentada ao Programa de Pós-graduação em Letras – Linguística, Universidade Federal de Santa Catarina.
- [7] Silva, Bento Carlos Dias. O Estudo Linguístico Computacional da Linguagem. *Revista Letras de Hoje*, Porto Alegre, v. 41, n.2, p.103-138, jun 2006.
- [8] Santos, T.L.T, Lino, A. P. A. S., Silva, Santos, T.L.T. Harb, M.P.A.H., Favero, E.L. Avaliação automática de questões conceituais discursivas. TIL'2007. Congresso da Sociedade Brasileira de Computação.
- [9] CHOMSKY, N. Three models for the description of language. *IRE Transactions PGIT*, 2. (pp. 113-124), 1956.
- [10] STAHL, S. A. Vocabulary and readability: how knowing word meanings affects comprehension. *Top Language Disorders*, vol. 23, n. 3, 2003, p. 241-247.
- [11] Kari, L. et al. A Formal Language Analysis of DNA Hairpin Structures. *Journal Fundamenta Informaticae*, Amsterdam, Netherlands, Vol 71 Issue 4, September, 2005, pg. 453-475.
- [12] Agirre, E. et al. MLDS: A translator-oriented MultiLingual dictionary system. *Journal Natural Language Engineering*, New York, USA, Volume 5 Issue 4, December 1999 Pages 325 - 353
- [13] Romeiro, Ana Karoline Queiroz. Um protótipo de analisador sintático probabilístico baseado em chart parsing para sentenças em português do Brasil. Cuiabá, MT, 2009.