

Um modelo para avaliação automática de respostas textuais com uso de regras linguísticas

Evandro Metz Flores¹, Sandro J. Rigo¹, Jorge L. V. Barbosa¹

¹Programa de Pós-Graduação em Computação Aplicada – Universidade do Vale do Rio dos Sinos (UNISINOS)
São Leopoldo – RS – Brazil

evandrometzflores@gmail.com, {rigo, jbarbosa}@unisin.br

Abstract. *This work describes a Model for Recognizing Textual Entailment based on linguistic rules, aiming to overcome some problems known in the literature. We present a new approach that combines syntactic analysis, morphology, voice flexion detection and the treatment of negations and synonyms. The use of these resources was evaluated with a prototype developed. The obtained results are promising, ensuring the textual entailment identification in different textual samples with an average accuracy of 92,12%.*

Resumo. *Este artigo descreve um Modelo de Reconhecimento de Vinculação Textual, cujo objetivo é superar problemas conhecidos em outros trabalhos. Para isso é apresentada uma nova abordagem através do uso combinado da análise sintática, morfológica, de regras linguísticas, detecção da flexão de voz, tratamento de negação e do uso de sinônimos. O uso destes recursos foi avaliado com a construção de diferentes técnicas que permitem validar e comparar seus resultados em um ambiente integrado. Os resultados apresentados até o momento são promissores, permitindo a identificação da vinculação textual de diferentes amostras textuais com precisão média de 92,12%.*

1. Introdução

A ampla adoção dos Ambientes Virtuais de Ensino e Aprendizagem (AVEAs) na permite facilitar o acesso dos alunos aos materiais de aprendizagem. Os grandes volumes de dados gerados pelos alunos nestas interações são, em parte, dados com formato textual, associados com a diversidade de atividades estimuladas nos AVEAs (JOHNSON et al., 2013). No caso específico da EaD, o envolvimento dos professores no acompanhamento detalhado de todo o conteúdo produzido pelos alunos constitui uma dificuldade (BAKER, ISOTAMI e CARVALHO, 2011). Já no caso de cursos na modalidade aberta, tais como os MOOCs (*Massive Online Open Courses*), não existe a preocupação em acompanhamento individualizado dos alunos, mas busca-se estimular a independência e iniciativa dos alunos (CHUONG et al., 2013).

Nos dois casos citados o acompanhamento de dados textuais volumosos pode ser facilitado com o desenvolvimento de ferramentas automatizadas que avaliem qualitativamente a participação dos alunos (RIGO et al., 2013; OLIVEIRA, OLIVEIRA e MARCHESI, 2009). Na modalidade de EaD estas ferramentas proporcionam ao

professor maior quantidade de informações sobre os alunos e seu desempenho. No caso dos MOOCs, estas ferramentas proporcionam aos alunos maior autonomia e estímulo. A verificação de correção de respostas textuais é um dos focos de atuação possíveis para estas ferramentas, apoiando o processo de identificar vinculações entre amostras textuais, que podem ser, por exemplo, diversas respostas textuais a uma pergunta, ou então podem representar diversos comentários sobre um assunto discutido em um fórum de discussões.

A área de reconhecimento de vinculação entre amostras textuais (DAGAN et al., 2013) desenvolve uma série de ferramentas que podem apoiar em atividades de análise de textos em educação (ÁVILA e SOARES, 2013). Dentre as diversas abordagens conhecidas (ANDROUSTSOPOULOS e MALAKASIOTIS, 2009; DAGAN et al., 2013), uma das mais comumente utilizada é a estatística, observada em um número relevante de trabalhos (FABRÍCIO, BEHAR e REATEGUI, 2011; KIM et al., 2007; LIN, HSIEH e CHUANG, 2009). Outra abordagem é a análise sintática de textos (DAGAN et al., 2013), onde a classificação gramatical de cada palavra e sua posição nas sentenças é identificada e utilizada para montar uma árvore dependência, por sua vez utilizada para identificar a existência ou não de vinculação entre amostras textuais (CHEN e LIU 2009; LAMJRI, KOSSEIM e RADHAKRISHNAN, 2007; SHARIATMADARI e MAMAT, 2009). A literatura apresenta ainda o uso de sinônimos, que de acordo com ANDROUSTSOPOULOS e MALAKASIOTIS (2009) apresenta bons resultados e vem sendo utilizado em diversos trabalhos (LAMJRI, KOSSEIM e RADHAKRISHNA, 2007; LIU, ZHAO e YU, 2006). Nesta técnica, o objetivo é avaliar a aproximação entre um corpus e outro considerando, entre outras possibilidades, a consulta em um léxico de sinônimos para cada um dos elementos do corpus.

Não são observados trabalhos que busquem implementar recursos de análise qualitativa entre trechos textuais através do uso combinado da análise sintática, morfológica, de regras linguísticas e do uso de sinônimos. Por outro lado, a crescente utilização da modalidade de Educação a Distância, dos cursos abertos (MOOCs) e a ampla adoção de AVEAs como apoio em atividades de ensino e aprendizagem configuram um cenário que pode se beneficiar da utilização de ferramentas que permitam o tratamento automático de trechos textuais, o que pode ser um elemento de apoio para os professores e para os alunos.

Um exemplo de avaliação da vinculação textual entre frases curtas e suas diversas implicações pode ser visto na figura 1, onde quatro frases curtas indicam a resposta para uma pergunta hipotética. Considera-se que a primeira frase (no item “a”) é a frase contendo a resposta correta. As demais frases supostamente foram providenciadas pelos alunos para avaliação. O processo de identificação da vinculação textual é realizado, em alguns trabalhos, com a comparação léxica entre os elementos das frases. Neste caso, a frase do item “b”, que está correta, seria penalizada por empregar um sinônimo para o verbo (a palavra “representa” em vez de “descreve”). A frase do item “c” seria penalizada neste contexto por estar estruturada em outra forma de escrita. Já a última frase candidata, do item “d”, seria penalizada por estes mesmos itens e também poderia incorporar um erro de interpretação caso a implicação da palavra “não” seja incorretamente tratada. Com este breve exemplo, pretende-se indicar as possibilidades de melhorias neste processo, que são obtidas com a incorporação de

maior quantidade de recursos processo de vinculação textual, bem como demonstrar as fragilidades de métodos baseados em poucos recursos de comparação.

-
- a) A equação **descreve** uma **parábola**.
- b) A equação **representa** uma **parábola**.
- c) Existe uma **parábola** implicada nesta equação.
- d) Não existe uma **parábola** implicada nesta equação.

Figura 1. Exemplos de abordagens para reconhecimento de vinculação textual

O objetivo deste artigo é apresentar um modelo que possibilita identificar a vinculação entre uma mensagem curta textual contendo a resposta correta para uma pergunta e outras mensagens textuais contendo as respostas de alunos para esta mesma pergunta. Sabe-se que a linguagem natural possui uma natureza complexa e ampla, sendo bastante variadas as possibilidades de descrição de respostas de perguntas. Apesar deste contexto, parte-se do pressuposto de que em ambientes controlados e com domínios de conhecimento específicos, torna-se viável o tratamento de mensagens textuais em linguagem natural, para a finalidade de avaliação automática da vinculação entre frases curtas. O modelo foi implementado em um protótipo e o principal diferencial é o reconhecimento de vinculação textual com base no conjunto amplo de informações que envolve informações sintáticas, informações morfológicas, regras linguísticas e a descrição de sinônimos. Este modelo possibilita a identificação correta de respostas breves para perguntas sobre autoria e composição. O protótipo foi validado com estudos de caso que demonstraram resultados com boa precisão e desempenho.

2. Trabalhos Relacionados

Um dos aspectos importantes em trabalhos nesta área é a abordagem utilizada para representação das informações, sendo que muitos trabalhos adotam a forma de representação léxica como base. Esta forma utiliza as palavras dos textos analisados e a sua grafia como a informação principal a ser tratada. O sistema de Herrera, Penas e Verdejo (2006) por exemplo, utiliza esta abordagem e também amplia a representação possível com as relações descritas no WordNet (GEORGE et al., 1993), a partir do qual são incluídas as relações descrevendo sinônimos, hipônimos ou antônimos, entre outros, na descrição das informações. Já no trabalho de Montalvo-Huhn (2008) utiliza-se também a representação léxica, mas neste caso são aplicadas várias métricas de comparação envolvendo as palavras originais, os sinônimos e antônimos.

A abordagem léxica pode ser ampliada, por exemplo, como no trabalho de Litkowski (2009), que consiste em privilegiar a utilização de comparações entre entidades nomeadas, dentro das etapas de avaliações dos textos comparados. Já em uma abordagem relacionada, Majumdar e Bhattacharyya (2010) descrevem um sistema baseado em representação léxica, porém que faz uso de correferência entre os elementos dos textos tratados. A ideia principal do trabalho de Varma et al. (2009) é identificar estruturas linguísticas, gerando padrões que permitem relacionar elementos dos dois textos sendo comparados. Já o trabalho de Tsuchida e Ishikawa (2011) combina a

utilização de uma pontuação gerada com base em nível de representação lexical e ampliada com mecanismos baseados em aprendizagem de máquina, que empregam diversas características originadas no nível léxico, tais como estruturas de sentenças e predicados. O trabalho de Zhang et. al (2012) descreve um sistema chamado SNRTE, que combina informações léxicas, sintáticas, e três níveis de análises semânticas, com o auxílio de ferramentas de PLN que incluem analisadores, rotuladores de parte no discurso e o WordNet. O trabalho de Rios et. al (2012) utiliza semântica em combinação com uma abordagem de edição de distância semântica, onde pontos são concedidos ao texto candidato de acordo com similaridade léxica e semântica entre os termos de ambas as amostras textuais.

A análise dos trabalhos estudados indica uma diversidade de abordagens para a representação dos elementos e seu tratamento. Os elementos utilizados para representação são predominantemente léxicos, mas também existem trabalhos relevantes com uso de relações adicionais, tais como sinônimos e anotação semântica. O trabalho apresentado neste artigo tem como diferencial a combinação de informações provenientes de análises distintas e de regras linguísticas.

3. Modelo Proposto

A abordagem adotada no modelo proposto permite realizar o reconhecimento de vinculação textual de respostas candidatas quando comparadas a uma resposta ótima. Neste caso, se considera que uma resposta candidata seja aquela resposta fornecida por um aluno e a resposta ótima seja a resposta fornecida pelo professor. Devido à grande complexidade para o tratamento de linguagem natural, somente são tratados neste modelo perguntas do tipo “Qual” ou “Quais”, sendo considerado apenas o tratamento de frases curtas. Frases que exemplificam estas perguntas poderiam ser “Quais são os estados que compõem a região sul?” ou “Quais são as cores do padrão CMYK?”. A expectativa de respostas curtas para estas duas perguntas poderia ser exemplificada como “É composta por três estados: Rio Grande do Sul, Santa Catarina e Paraná” ou “As cores do padrão CMYK são: ciano, magenta, amarelo e preto”, respectivamente.

A visão geral do modelo proposto é apresentada na Figura 2. A etapa 1 considera o tratamento de amostras textuais presentes em um AVEA. Nesta etapa inicial o usuário do sistema deve indicar qual é a pergunta e qual é o texto que contém as informações importantes e que devem estar contidas nas respostas dos alunos. O usuário também deve indicar quais são as respostas dos alunos que serão avaliadas, gerando assim as mensagens selecionadas presentes na etapa 2. Estas operações estão representadas na Figura 2 através dos componentes “Texto ótimo”, “Texto Candidato” e “Filtro”. Após selecionadas, as mensagens contendo tanto o texto ótimo como os textos candidatos são anotadas pelo analisador morfossintático. No caso da implementação deste modelo foi empregado o *parser* morfossintático Palavras (BICK, 2000), devido a sua precisão. Esta atividade de anotação dos textos está indicada na etapa 3, que deve gerar como resultado as respectivas mensagens anotadas em um formato adequado para manipulação, na etapa 4.

As mensagens textuais anotadas são manipuladas na etapa 5. Inicialmente é realizado o tratamento da negação nas frases, com o objetivo de armazenar qualquer informação sobre negação que possa estar presente no texto. Na segunda operação é verificada a flexão de voz, para garantir que diferenças entre o uso de voz ativa ou voz

passiva nas frases não gerem interferências na avaliação final das respostas. Para tal, se faz necessária a alteração da estrutura da frase e a troca de tempo verbal do verbo principal quando existe esta diferença no uso de tipos de voz no texto. Após a alteração da voz, o modelo prevê uma operação onde os textos candidatos têm seus verbos comparados aos do texto ótimo. Estes verbos são comparados entre si e nesta comparação se utiliza o auxílio de uma base de informações extraída do repositório WorldNet Br, que proporciona o acesso aos sinônimos para os verbos encontrados nas frases. A última operação da etapa 5 trata da execução de regras linguísticas sobre os textos ótimo e candidato. O objetivo desta operação é identificar qual regra linguística abrange maior parte do texto sendo analisado. Estas regras linguísticas são provenientes de um repositório de regras linguísticas (etapa 6) que está associado a um editor de regras linguísticas (etapa 7), onde estas podem ser adicionadas ou removidas por especialistas em linguística.

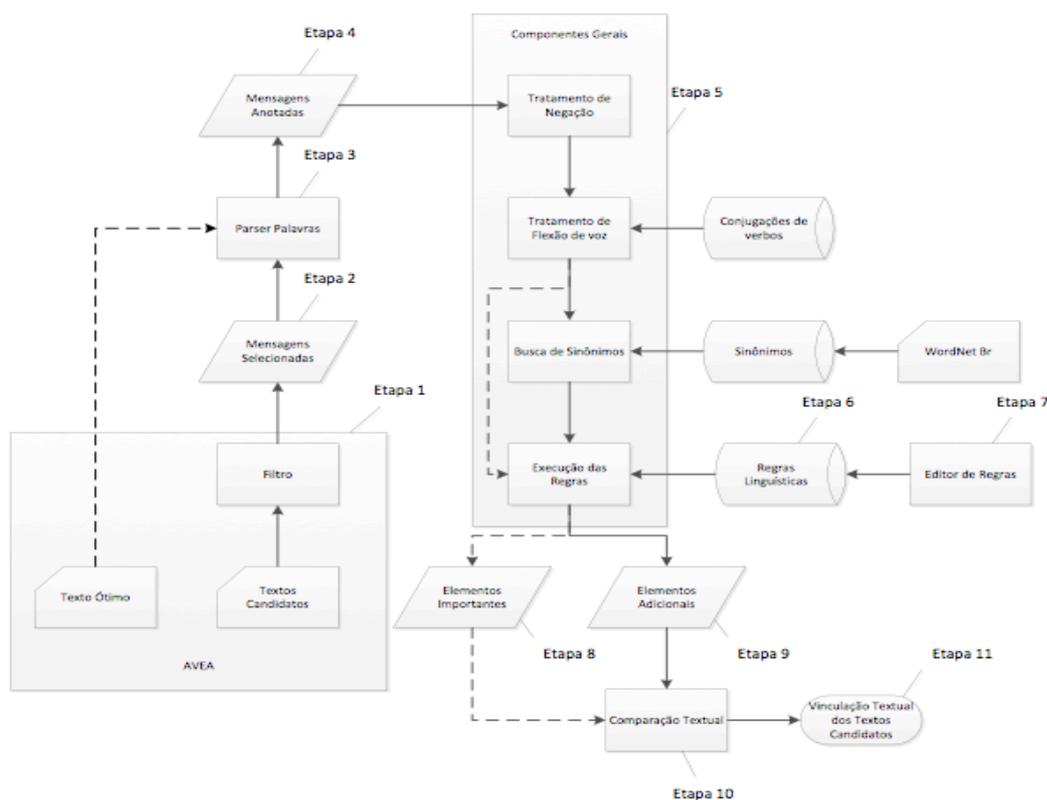


Figura 2. Visão geral do modelo proposto

Na etapa 8 está descrita uma operação que utiliza a regra linguística que foi selecionada como a de melhor aderência para com o texto ótimo com a finalidade de destacar no texto a localização das palavras chaves importantes para a resposta. Ao final desta operação estas palavras são armazenadas e identificadas como elementos importantes para análise da resposta. A etapa 9 tem como objetivo aplicar a regra linguística selecionada para cada um dos textos candidatos com o objetivo de destacar no texto a localização das palavras utilizadas pelo aluno que são candidatas a estarem vinculadas com as do professor. Estas palavras são armazenadas e identificadas como elementos candidatos. A identificação dos termos considerados como elementos importantes (no texto ótimo) e como elementos candidatos (no texto candidato) é realizada a partir da descrição das regras linguísticas, detalhadas adiante. Na etapa 10 e

na etapa 11 são realizadas as comparações entre os elementos importantes e elementos candidatos de cada conjunto de textos. A este processo é adicionada a execução de uma métrica definida no modelo para possibilitar a identificação de uma categorização entre estas frases. Desta forma, a quantidade de palavras encontradas de forma coincidente nos dois conjuntos indica o percentual de vinculação textual entre texto ótimo e candidato.

As regras linguísticas usadas no modelo proposto têm como objetivo principal extrair e rotular cada informação existente tanto na resposta ótima quanto nas respostas candidato. Estas regras também possibilitam a identificação de voz passiva e ativa presentes tanto no texto ótimo quanto no candidato. As regras são geradas manualmente por especialistas em linguística, com uso de um editor desenvolvido junto do escopo do modelo e no protótipo. O processo de execução das regras trata de identificar qual das regras presentes no banco de regras é capaz de representar o texto analisado. Cada regra é definida com base na análise do corpus tratado, gerando padrões compostos por informações morfossintáticas intercaladas com a indicação dos elementos importante para a resposta.

O modelo proposto adota uma pontuação numérica para a indicação da regra que melhor representa a frase analisada. O processo de identificação de elementos utilizados na pontuação ocorre depois da execução das regras e a indexação da resposta ao texto ótimo. Para realizar esta pontuação são efetuados dois passos onde são comparadas as informações morfológicas das duas frases e os elementos e possíveis sinônimos. Este cálculo é usado para se obter o percentual de correspondência entre o texto candidato e o ótimo.

4. Aspectos de implementação e avaliação

O protótipo desenvolvido, em linguagem C#, permite a identificação de conjuntos de frases para testes incluindo e também a edição de regras linguísticas. Um arquivo de saída de resultados foi definido para a avaliação do funcionamento do protótipo, contendo as informações detalhadas sobre a execução das regras linguísticas e a aplicação das métricas. Neste arquivo estão contidos o texto ótimo e seu arquivo original, o número de palavras armazenadas em elementos importantes, a frase que compõe originalmente o texto ótimo, a regra eleita dentre os presentes no repositório que melhor representa esta amostra, a frase em sua forma alterada, as palavras armazenadas como elementos importantes e, por fim, as informações morfológicas destas palavras de acordo com a anotação do *parser* morfossintático. Para cada frase candidata analisada, o conjunto equivalente de informações é descrito neste arquivo.

A utilização das regras linguísticas foi implementada com a premissa de buscar tanto uma representação simples como eficiente computacionalmente. Após executadas as alterações nas amostras das frases sendo avaliadas, a primeira regra linguística associada com o arquivo de texto em análise é armazenada em uma lista. Esta é então lida e repassada para uma outra estrutura contendo diversas listas, onde cada uma apresenta um grupo das sequências de informações morfológicas que estão presentes nas regras. Os elementos que definem as separações dos grupos de elementos das regras linguísticas são justamente os elementos indicados como “Elementos importantes” nestas regras, ou seja, os elementos que, segundo a definição da regra, identificam os componentes principais das respostas. Estas sequências de elementos são

utilizadas em um processo de busca sobre a lista geral de dados das frases sendo avaliadas. A cada identificação de elementos da sequência ocorre uma operação de pontuação desta regra. Um processo similar ocorre com os elementos importantes descritos nas regras. A cada identificação positiva destes elementos na frase sendo analisada a regra em uso tem sua pontuação incrementada.

Ao final do processo de ponderação das regras, apenas uma é escolhida e representa todo o conteúdo da frase. A etapa seguinte utiliza esta regra selecionada para gerar a pontuação final da frase, compondo assim o reconhecimento de vinculação textual que será atribuído à mesma. A pontuação consiste na identificação dos elementos importantes constantes nas regras e na verificação de sua existência junto dos dados textuais da frase. Desta forma, foi possível tratar a questão das pesquisas de elementos de forma flexível e com um custo computacional baixo, devido ao formato de estruturas de dados utilizados e à escolha do elenco de operações para sua manipulação.

As avaliações realizadas contaram com a participação de um especialista em linguística. Este especialista desenvolveu trinta questões criadas especificamente para esta avaliação, sendo vinte e duas no formato “Qual/Quais”, cinco no formato “Quem”, duas no formato “Quando” e duas no formato “De que”. Um exemplo destas questões e respostas pode ser visto na Tabela 1. Junto a estas questões também foram desenvolvidas uma resposta ótima e quatro respostas candidatas para cada pergunta, totalizando cento e vinte grupos de respostas a serem avaliadas pelo protótipo. Estas respostas também foram observadas pelo especialista em linguística para a construção das regras linguísticas a serem cadastradas no protótipo, para ficarem disponíveis junto ao repositório de regras linguísticas. As regras presentes no repositório durante o estudo de caso inicial totalizam sessenta e nove regras.

O especialista em linguística também indicou para cada resposta candidata uma nota em percentual, para possibilitar uma posterior comparação com o resultado gerado com o uso do protótipo. O critério usado pelo especialista para indicar a nota de cada resposta foi definido como sendo a medida dos elementos contidos na resposta para responder corretamente a pergunta. Como o grupo de perguntas tratado é voltado para situações objetivas e para respostas curtas, este critério é traduzido de forma bastante precisa e representa a quantidade de elementos corretos existentes em cada resposta.

Tabela 1. Exemplo de um grupo de perguntas e respostas

| Perguntas e respostas | | Avaliação do especialista (%) |
|-----------------------|--|-------------------------------|
| Pergunta: | Quais cores apresenta o padrão CMYK? | |
| Resposta ótima: | São: ciano, magenta, amarelo e preto. | 100 |
| Respostas candidatas: | As cores padrão CMYK são: ciano, magenta, amarelo e preto. | 100 |
| | São as cores amarelo e preto. | 50 |
| | As cores são verde, vermelho e azul. | 0 |
| | São ciano e vermelho. | 25 |

A métrica de avaliação utilizada neste caso, para comparar as respostas anotadas pelo especialista manualmente com as respostas anotadas pelo protótipo foi a métrica de precisão (P), que é calculada dividindo-se o número de avaliações geradas pelo protótipo que são corretas e idênticas às do especialista (AC) pela soma deste número AC com o número de avaliações geradas pelo protótipo que são incorretas e diferentes das avaliações do especialista (AI). Para as oitenta e oito respostas candidatas sobre a

forma de pergunta “Qual/Quais”, o protótipo não apresentou exatamente o mesmo percentual que a avaliação do especialista em apenas três casos, resultando em uma precisão de 96,6%. Das perguntas do tipo “Quem”, duas das vinte respostas candidatas não foram avaliadas da forma como indicada pelo especialista, resultando em uma precisão de 90%. Depois de avaliado o sistema nos moldes das perguntas “Qual/Quais” e “Quem”, foram inseridos dois tipos de perguntas extras, para avaliar o comportamento do protótipo. A pergunta do tipo “De que” teve suas respostas avaliadas corretamente em 100% dos casos. Já as respostas do tipo “Quando” não puderam ser avaliadas pelo protótipo, isso porque o *parser* morfossintático utilizado não identificou as datas presentes nas amostras textuais, impossibilitando qualquer tipo de análise. Sendo assim, do total de pares de repostas possíveis de serem analisadas pelo protótipo (cento e dezesseis), cento e onze puderam ser avaliadas pelo protótipo, e sendo que cinco análises não apresentaram exatamente o mesmo percentual de avaliação definido pelo especialista em linguística. O protótipo apresentou uma precisão de 95,7% de acerto no estudo de caso inicial. A tabela 2 sumariza estas informações.

Tabela 2. Sumarização dos resultados do primeiro teste

| Tipo de questão | Total de respostas | Corretas | Incorretas | Precisão % |
|-----------------|--------------------|----------|------------|------------|
| Qual/Quais | 88 | 85 | 3 | 96,6% |
| Quem | 20 | 18 | 2 | 90% |
| De que | 8 | 8 | 0 | 100% |
| Total | 116 | 111 | 5 | 95,7% |

Para o segundo estudo de caso foram desenvolvidas pelo especialista em linguística onze perguntas de estilos diferentes, sendo quatro relacionadas ao padrão “Qual/Quais”, quatro ao “Quem”, duas ao “Que” e a última ao “Como”, assim como as respostas ótimas para cada pergunta. Dezoito pessoas responderam a estas perguntas gerando assim 198 respostas candidatas, analisadas pelo especialista em linguística e também pelo protótipo. É importante salientar que nenhuma regra linguística nova foi adicionada ao protótipo entre o primeiro e o segundo estudo de caso. Os resultados provenientes da análise das demais perguntas podem ser acompanhados na tabela 3.

Tabela 3. Sumarização dos resultados do segundo teste

| Tipo de questão | Total de respostas | Corretas | Incorretas | Precisão % |
|-----------------|--------------------|----------|------------|------------|
| Qual/Quais | 72 | 72 | 0 | 100% |
| Quem | 72 | 57 | 15 | 79,16% |
| De que | 36 | 35 | 1 | 97,22% |
| Total | 180 | 164 | 16 | 92,12% |

As quatro perguntas da categoria “Qual/Quais” apresentaram setenta e duas respostas candidatas, e que quando comparadas pelo protótipo a resposta ótima, apresentaram o mesmo percentual que o especialista indicou em seu estudo, sendo assim, nenhuma análise foi indicada como incorreta, e a precisão do algoritmo alcançou 100%. As quatro perguntas da categoria “Quem” apresentaram também setenta e duas respostas candidatas, mas quinze delas obtiveram um percentual de vinculação textual diferente do indicado pelo especialista. Sendo assim, o protótipo atingiu neste tipo de pergunta uma precisão de 79,16%. Por último foram analisadas as duas perguntas do tipo “De que”, que resultaram em trinta e seis respostas candidatas. Estas respostas, após avaliadas pelo protótipo, não apresentaram o mesmo percentual de vinculação

textual que o especialista em linguística em apenas um dos casos, resultando em uma precisão de 97,22%.

5. Considerações finais

Este artigo apresentou um modelo para identificação de vinculação textual, que apresenta como principais componentes a utilização de regras linguísticas em combinação com o *parser* morfossintático Palavras, identificação de voz das frases, negação e busca de sinônimos no WordNet. Para a exploração destas técnicas, foi desenvolvido e avaliado um protótipo que permitiu demonstrar os resultados possíveis de serem alcançados com a combinação de recursos utilizados. O protótipo possui uma funcionalidade para que os especialistas em linguística possam acrescentar padrões morfológicos ao banco de regras linguísticas.

A área de identificação de vinculação textual está sendo amplamente explorada ao longo dos anos, mas existem dificuldades na adaptação de outros trabalhos conhecidos em seu uso na língua portuguesa. Para contornar esta característica, foi apresentado nesta proposta um enfoque com uso das regras linguísticas, pouco exploradas pelos pesquisadores na área, e que pode viabilizar a construção de um sistema de vinculação textual voltado a língua portuguesa. O uso da detecção de voz, como proposto, também é um diferencial na área, pois poucos trabalhos pesquisados apresentam tal preocupação.

O primeiro teste realizado sobre uma base desenvolvida propositalmente para ser avaliada pelo protótipo, resultou em um nível de precisão alto e também em algumas informações específicas inesperadas, como a anotação inconsistente do *parser* morfossintático utilizado e a necessidade de se ter uma espécie de repositório de significância para nomes próprios ou termos específicos. Este último evitaria que na avaliação de respostas contendo nomes próprios, por exemplo, o algoritmo possa compreender quais são as formas possíveis de se referir a aquela pessoa em específico, seja pelo primeiro, último ou todos os nomes de uma pessoa, e ainda assim ser avaliado como correta a sua resposta.

O segundo teste confirmou os resultados observados no primeiro teste, apresentando uma boa precisão geral e também a mesma flexibilidade para o seu uso em diversos contextos, além de indicar que o uso das regras linguísticas pode ser considerado em diversos contextos, de forma efetiva. Como trabalhos futuros está sendo realizada uma terceira etapa de avaliações com um conjunto mais extenso de dados e provenientes de questionários do AVEA MOODLE, e também a implementação dos componentes para permitir a automatização desta análise de vinculação textual.

Referências

- ANDROUTSOPOULOS, I.; MALAKASIOTIS, P. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, v. 38, p. 135–187, 2009.
- ÁVILA, R. L. F.; SOARES, J. M. Uso de técnicas de pré-processamento textual e algoritmos de comparação como suporte à correção de questões dissertativas: experimentos, análises e contribuições. In *Simpósio Brasileiro de Informática na Educação*, 2013.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de Dados Educacionais: oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(02). P. 15-35, 2011.

- BICK, E. The parsing system Palavras. Tese de Doutorado, Universidade de Aarhus, 2000.
- CHEN, L.; LIU, Y.; Automated Scoring System Using Dependency-Based Weighted Semantic Similarity Model. 2009 Second International Symposium on Knowledge Acquisition and Modeling, v. 1, p. 241–244, 2009.
- CHUONG, B. Do et al. Daphne; Self-Driven Mastery in Massive Open Online Courses. MOOCs FORUM. September 2013, 1(P): 14-16. doi:10.1089/mooc.2013.0003, 2013.
- DAGAN, I. et al. Recognizing Textual Entailment – Models and Applications. Synthesis Lectures on Human Language Technologies, Toronto. Morgan & Claypool Publishers. 2013
- FABRÍCIO, B. T. A.; BEHAR, P. A.; REATEGUI, E. B. Análise das mensagens de fóruns de discussão através de um software para mineração de textos SBIE. n. 2004, p. 20–29, 2011.
- GEORGE A. M. et al. Introduction to WordNet: An On-line Lexical Database. MIT Press, 1993.
- HERRERA, J.; PENAS, A.; VERDEJO, F. Textual Entailment Recognition Based on Dependency Analysis and WordNet. Machine Learning Challenges. Lecture Notes in Computer Science Volume 3944, pp 231-239, 2006.
- JOHNSON, L. et al. NMC Horizon Report: 2013 Higher Education Edition. Austin, Texas: The New Media Consortium, 2013.
- KIM, J. et al. An Intelligent Discussion-Bot for Guiding Student Interactions in Threaded Discussions. Proceedings of the AAAI 2007.
- LAMJIRI, A. K.; KOSSEIM, L.; RADHAKRISHNAN, T. Comparing the Contribution of Syntactic and Semantic Features in Closed versus Open Domain Question Answering. International Conference on Semantic Computing ICSC 2007, p. 679–685, 2007.
- LIN, F.; HSIEH, L.; CHUANG, F. Discovering genres of online discussion threads via text mining. Computers & Education, v. 52, n. 2, p. 481–495, 2009.
- LITKOWSKI, K. Overlap analysis in textual entailment. recognition. In Proc. of TAC, 2009.
- LIU, P.; ZHAO, T.; YU, X. Application-Oriented Comparison and Evaluation of Six Semantic Similarity Measures Based on Wordnet. 2006 International Conference on Machine Learning and Cybernetics, p. 2605–2610, 2006.
- MAJUMDAR, D.; BHATTACHARYYA, P. Lexical Based Text Entailment System for Sum-marization Settings of RTE6. Proceedings of the Text Analysis Conference (TAC 2010) November 15, 2010.
- MONTALVO-HUHN, O.; TAYLOR, S. Textual Entailment – Fitchburg State College. In Proceedings of TAC08, Fourth PASCAL Challenges Workshop on Recognising Textual Entailment, 2008.
- OLIVEIRA, M. G. de; OLIVEIRA, E. e MARCHESI, R. Z. Um QAsystem para Interação de Alunos em Avaliações Somativas a Distância. In Simpósio Brasileiro de Informática na Educação, 2009.
- RIGO, S. J. et al. O papel do Processamento de Língua Natural e da Representação de Conhecimento na extração de informações em mensagens textuais na Educação a Distância. In Simpósio Brasileiro de Informática na Educação, 2013.
- RIOS, M.; GELBUKH, A. Recognizing Textual Entailment with a Semantic Edit Distance Metric. 11th Mexican International Conference on Artificial Intelligence. 2012.
- SHARIATMADARI, S.; MAMAT, A. Discovering semantic similarity association in semantic search system. Proceedings of the 11th WAS, p. 331, 2009.
- TSUCHIDA, M.; ISHIKAWA, K. IKOMA at TAC2011: A Method for Recognizing Textual Entailment using Lexical-level and Sentence Structure-level features. In TAC 2011.
- VARMA, V. et al. IIIT Hyderabad at TAC 2009. Maryland USA, 2009.
- ZHANG, S.; et al. "Recognizing Textual Entailment with synthetic analysis based on SVM and feature value control." Intelligent Computer Communication and Processing (ICCP), 2012 IEEE vol., no., pp.91,94, Aug. 30 2012-Sept. 1, 2012.