

ANAC: um analisador automático de coesão textual em redação

João Carlos Silva Nobre¹, Sérgio Roberto Matiello Pellegrino¹

¹Divisão de Ciência da Computação
Instituto Tecnológico de Aeronáutica (ITA) – São José dos Campos, SP - Brasil

{jcnobre, pell}@ita.br

***Abstract.** This paper presents an approach for automatic analysis of textual cohesion in essays discourse in Portuguese based on a method developed with the aid of the Centering Theory and Theory Focusing. The prototypes take as input a text, and analyze the structure of discourse cohesion in order to identify breaks, terms that undermine cohesion. Preliminary results indicate that 90% of the essays are identified with cohesion problems consistent with the scores assigned by human assessors.*

***Resumo.** Este artigo apresenta o processo de análise automática de coesão textual em redações dissertativas em português com base em um método desenvolvido a partir da Teoria da Centragem e do Foco. O protótipo desenvolvido recebe como entrada um texto e deve analisar a estrutura do discurso visando identificar quebras de coesão local e global, termos que prejudicam a coesão. Resultados preliminares indicam que em 90% das redações identificou-se problemas de coesão compatíveis com as notas atribuídas por avaliadores humanos.*

1. Introdução

Nos anos 80 era impossível analisar sistematicamente a coesão textual, visto faltar elementos teóricos na língua nos múltiplos níveis conjugado às teorias e ferramentas computacionais. Entretanto, os avanços nas áreas de psicolinguística, processamento do discurso, linguística computacional têm propiciado o desenvolvimento de ferramentas computadorizadas que analisam os textos em várias dimensões.

Muitas têm sido as aplicações, com destaque às análises textuais nas suas mais variadas finalidades: análise sintática de termos, localização de sinônimos, sintetizador de textos técnicos.

Diante de sua forma irregular, por não ser um texto proficiente da língua, as dissertações de vestibular apresentam grandes obstáculos a uma análise textual.

Identificar as inconformidades linguísticas e as inadequações na expressão verbal escrita com a finalidade de obter uma análise textual qualificada dessas redações é a proposta deste trabalho.

Com base na fundamentação teórica a ser apresentada, vislumbra-se a possibilidade de aplicação da Teoria da Centragem e Teoria do Foco visando definir um método robusto para avaliar a coesão textual, o qual pode ser aplicado a sistemas que auxiliem a produção de texto por alunos.

O presente artigo encontra-se estruturado da seguinte forma: nas seções 2 e 3 há uma breve fundamentação teórica sobre Coesão e Discurso; na seção 4 explana-se sobre

o processo da análise automática da coesão em texto dissertativo; na seção 5, apresenta-se o experimento realizado; e, na seção 6, o artigo é encerrado com observações decorrentes desta pesquisa.

2. Coesão

Segundo Halliday *and* Hasan (1976), a coesão textual é um conceito semântico o qual se refere às relações de sentido existentes no interior do texto e que o definem como um texto. Quando ocorre a interpretação de algum elemento no discurso que é dependente da de outro se tem a coesão. Assim, a efetiva decodificação só acontece por recurso ao outro e se baseia no sistema léxico-gramatical, de acordo com Townsend *and* Bever (2001). Existe, portanto, formas de coesão realizadas por intermédio da gramática e outras pelo léxico. São mencionados como principais fatores de coesão a referência, a substituição, a elipse, a conjunção e a coesão lexical.

Segundo Louwerse (2002) e McNamara et al. (1996), há coesão local e global. A compreensão desses tipos por parte do leitor permite-lhe identificar as relações coesivas locais entre cláusulas próximas no texto e as relações globais entre grupos de cláusulas e parágrafos, facilitando o processo mental de compreensão do texto.

Os textos localmente coesos marcados pela ausência de coesão global, tendem a inibir a compreensão e a recuperação de informações. Já os textos que possuem coesão global e são desprovidos de coesão local tornam-se, por vezes, difíceis de ler e compreender (McNamara et al. 1996).

3. Discurso

Considerando-se sinônimo de enunciado em termos linguísticos, o discurso constitui-se por um conjunto de frases logicamente ordenadas, de forma a comunicar um sentido. Nessa acepção, a frase é considerada hoje uma unidade do discurso e é susceptível de ser analisada na forma como se combina com outras frases para constituir um discurso coeso.

Para compreender melhor o discurso é fundamental que essa estrutura possibilite a organização da informação transmitida, facilitando a sua interpretação por parte do receptor. Segundo Freitas e Lopes (1994), o processo de estruturação do discurso está diretamente relacionado com a comunicação entre o emissor e o receptor, sendo a eficiência do procedimento capaz de ser medida pela rapidez com que se recuperam as interpretações possíveis para um dado trecho do discurso (Blutner 2000). Em termos computacionais, isto equivale a um menor tempo de processamento e, em termos lógicos, a um menor número de inferências sobre o menor número possível de modelos.

3.1. A Teoria do Foco e a Teoria da Centragem

A Teoria do Foco - TC, elaborada por Sidner (1981, 1979), propõe um algoritmo cujas principais funções são: (a) reduzir o universo de possíveis antecedentes introduzidos no universo do receptor durante a interpretação de novas frases proferidas em contexto ou escritas num todo coerente; e (b) propor um caminho mais eficiente para percorrer este universo, já reduzido, em busca de um antecedente. Para tal, a teoria propõe que as entidades mais salientes de uma frase devem ser os antecedentes preferenciais para a resolução de uma anáfora numa frase seguinte.

No processo de resolução de anáforas, Sidner (1981) define dois centros de atenção ou focos: o foco do ator (FA) e o foco do discurso (FD), preferencialmente determinados pelo agente e pelo tema de cada frase, utilizando os seguintes dados: a informação temática (agente e tema) (Gruber 1976), a informação gramatical (sujeito, objeto direto, objeto indireto, etc) e a informação sobre quais são as entidades mais salientes da frase anterior - o foco local (Grosz 1977).

A Teoria da Centragem - TC (Grosz et al. 1983) foi criada para medir como a coerência do discurso é influenciada pela compatibilidade entre os centros de atenção ou foco e a escolha das expressões de referenciamento (Grosz et al. 1995).

É importante destacar que a TC é uma proposta diretamente ligada ao trabalho conjunto de Grosz *and* Sidner (1986) sobre a Teoria da Estrutura do Discurso, com o objetivo de explicar a coerência dentro de um mesmo segmento e permitir a obtenção de ligações entre as entidades das frases, estabelecendo critérios de coesão.

3.2. O Foco do Discurso

Foco do discurso, aqui tomado como sinônimo de centro de atenção, é o termo utilizado para designar as entidades mais salientes do discurso (Sidner 1981, Grosz 1977). Tipicamente, o foco é a entidade sobre a qual o emissor centra sua atenção em determinado ponto do discurso, sendo sua utilização continuada como elo coesivo um forte indício de que esta entidade está em foco (Grosz et al. 1995, Sidner 1981).

O termo foco, no âmbito deste trabalho, será redefinido e sofrerá alteração de nomenclatura como segue:

- Foco Explícito - FE: Lista das entidades explicitamente contidas em cada frase do texto, podendo ser substantivos, pronomes, nomes próprios e elos de coesão.
- Foco Implícito - FI: Lista dos rótulos semânticos das entidades contidas no foco explícito. Quando se tratar de nome próprio, o rótulo semântico será um identificador único para o termo.

4. O processo de Análise

O modelo de análise de coesão proposto é baseado em conhecimentos morfossintáticos e semânticos, os quais foram obtidos por meio do *parser* Palavras, concebido por Bick (2000). Dessa maneira, as redações são primeiramente processadas pelo *parser*, de modo a obter o conteúdo da entrada do processo de análise. A Figura 1 apresenta a estrutura de funcionamento.

Um Corpus de teste utilizando 60 redações de vestibular, previamente analisadas, foi submetido ao sistema para avaliação. Esse conjunto foi chamado de CorpusR, contendo 16870 *tokens*, 731 sentenças e 373 parágrafos.

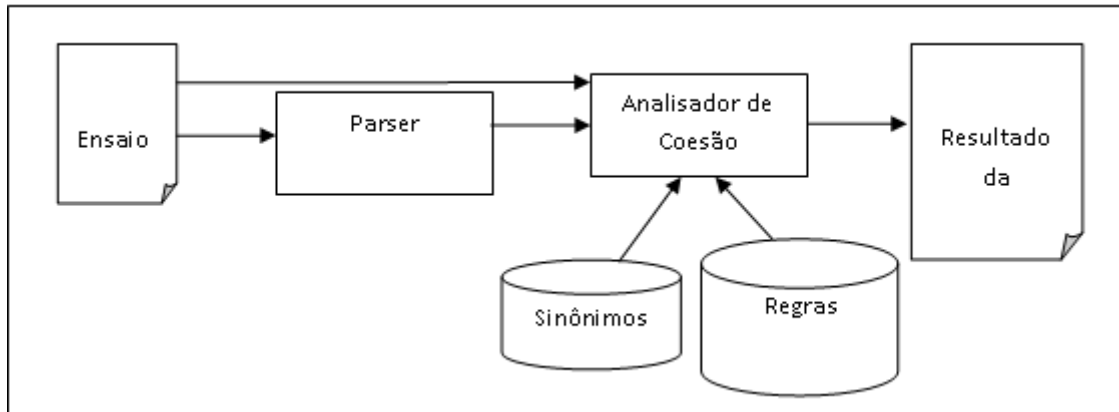


Figura 1 – Estrutura do processo de análise de coesão

4.1. Algoritmo de Análise de Coesão

A análise da coesão apoia-se na teoria da centragem e na teoria do foco, utilizadas para resolução de anáforas e, no caso da TC, para medir como a coerência do discurso é influenciada pela compatibilidade entre os centros de atenção ou foco.

A partir do conhecimento obtido com as teorias, e as modificações visando possibilitar a análise de coesão, elaboraram-se os algoritmos para analisar a coesão local e global, descrito, respectivamente, nas Figuras 2 e 3.

Identificar de forma única cada frase - F

Para cada F

Identificar os substantivos, os nomes próprios, pronomes e/ou outros elementos anafóricos e a que termo refere-se, gerando a lista de Foco Explícito – FE.

Criar a lista de Foco Implícito - FI com o rótulo semântico, quando possível, de cada elemento da lista de FE.

Incluir os sinônimos na lista de FE₁ que possuem representante na lista de FE₂ cujos rótulos semânticos são iguais nas listas dos Focos Implícitos

Localizar elementos que geram ambiguidade e contradição locais

Para cada par de F adjacente (F_i e F_{i+1})

Calcular a intersecção entre os elementos do FE_i e FE_{i+1} por meio da fórmula

$$RFFE_{i, i+1} = FE_i \cap FE_{i+1}$$

Excluir os elementos em FI_i e FI_{i+1} que são representados em RFFE_{i, i+1}

Calcular a intersecção entre os elementos do FI_i e FI_{i+1} por meio da fórmula RFFI_{i,}

$$_{i+1} = FI_i \cap FI_{i+1}$$

Analisar RFFE_{i, i+1} e RFFI_{i, i+1} com base na tabela de coesão definida

Figura 2 - Algoritmo para avaliar Coesão Local – CL.

Identificar de forma única cada parágrafo

Para cada Parágrafo - P

Criar lista de Foco Explícito por intermédio do cálculo da união das intersecções dos elementos entre suas frases, com adjacência máxima: $RFEP_j = (FEP_j F_1 \cap FEP_j F_2) \cup (FEP_j F_1 \cap FEP_j F_n) \cup \dots \cup (FEP_j F_{n-1} \cap FEP_j F_n)$

Criar lista de Foco Implícito por intermédio do cálculo da união das intersecções dos elementos entre suas frases com adjacência máxima: $RFIP_j = (FIP_j F_1 \cap FIP_j F_2) \cup (FIP_j F_1 \cap FIP_j F_n) \cup \dots \cup (FIP_j F_{n-1} \cap FIP_j F_n)$

Para cada par de P, com adjacente máxima (P1 e P2, ..., P2 e Pn, ..., Pn-1 e Pn)

Calcular a intersecção entre os elementos do FEP_i e FEP_{i+1} por meio da fórmula $RFEP_{j,j+1} = FEP_j \cap FEP_{j+1}$

Excluir os elementos em FIP_j e FIP_{j+1} que são representados em $RFEP_{j,j+1}$

Calcular a intersecção entre os elementos do FIP_j e FIP_{j+1} por meio da fórmula $RFIP_{j,j+1} = FIP_j \cap FIP_{j+1}$

Analisar $RFEP_{j,j+1}$ e $RFIP_{j,j+1}$ com base na tabela de coesão definida.

Localizar elementos que geram ambiguidade e contradição em todo um parágrafo

Figura 3 - Algoritmo para avaliar Coesão Global – CG.

A Tabela 1, adaptada da TC, apresenta os possíveis relacionamentos existentes entre o foco explícito e implícito na determinação da coesão. As possíveis leituras do relacionamento entre o FE e FI são:

- Se existe um elemento de FE_i em FE_{i+1} ($FE_i \subset FE_{i+1}$) e, também, existe um elemento de FI_i em FI_{i+1} ($FI_i \subset FI_{i+1}$), então as frases F_i e F_{i+1} estão num processo de elaboração, visto que compartilham as mesmas entidades.
- Se existe um elemento de FE_i em FE_{i+1} ($FE_i \subset FE_{i+1}$), mas não existe um elemento de FI_i em FI_{i+1} , então as frases F_i e F_{i+1} ($FI_i \text{ NC } FI_{i+1}$) estão num processo de Manutenção de tópico, visto que compartilham alguns elementos explícitos.
- Se não existe um elemento de FE_i em FE_{i+1} ($FE_i \text{ NC } FE_{i+1}$), mas existe um elemento de FI_i em FI_{i+1} ($FI_i \subset FI_{i+1}$), então as frases F_i e F_{i+1} estão num processo de Mudança de tópico, haja vista compartilhar as mesmas entidades semânticas.
- Se não existe elemento de FE_i em FE_{i+1} ($FE_i \text{ NC } FE_{i+1}$), nem de FI_i em FI_{i+1} ($FI_i \text{ NC } FI_{i+1}$), então as frases F_i e F_{i+1} estão num processo de Mudança de assunto, pois não compartilham as mesmas entidades explícitas e nem semânticas.

Tabela 1 - Relação entre FE e FI para estabelecimento da coesão.

	$FE_i \subset FE_{i+1}$	$FE_i \text{ NC } FE_{i+1}$
$FI_i \subset FI_{i+1}$	Elaboração	Mudança de tópico
$FI_i \text{ NC } FI_{i+1}$	Manutenção do tópico	Mudança de assunto

Como exemplo considera-se as frases (1) e (2) para apresentar o processo de análise. Os rótulos semânticos aplicados ao FI significam: H → humano, HH → grupo humano, f-psyq → característica psicológica humana, Lh → construção usada por humano, inst → instituição, act → ação

O ser humano constrói seu caráter em sociedade. (1)

Primeiro no lar, na família e depois na Igreja ou qualquer outra instituição. (2)

A Frase 1 possui como FE_1 as seguintes entidades “ser humano”, “caráter”, “sociedade” que são representadas no FI_1 pelas entidades “H”, “f-psyc”, “HH”, respectivamente.

Na Frase 2, as entidades do FE_2 são “lar”, “família”, “igreja”, “instituição” e as respectivas entidades do FI_2 são “Lh”, “HH”, “inst”, “inst/act”.

Durante o processo de definição das entidades relevantes, busca-se os sinônimos, por intermédio da base de sinônimos Tep2 (Maziero et al. 2008), para compor a lista de entidades, gerando uma lista intermediária ($LI = \{\text{ser humano} = \{\text{homem, humanidade, gente}\}, \text{caráter} = \{\text{feição, índole, expressão, propriedade, coração, cunho, marca, estilo, feito, qualidade, veia, constituição, formação}\}, \text{sociedade} = \{\text{entidade, instituição, agremiação, associação, clube, grêmio, companhia, empresa, firma, organização, parceria, comunidade, clã, facção, grei, união}\}\}$). Após a composição da LI, compara-se cada elemento dos subconjuntos da LI com os elementos da lista de entidades da Frase 2, ou seja, os elementos de FE_2 . Os elementos coincidentes terão comparados também seus rótulos semânticos, de forma a garantir que estão sendo empregados com o mesmo sentido.

Na Frase 2, a palavra “sociedade” tem como um dos sinônimos a palavra “instituição”, mas ao se comparar o rótulo semântico de “sociedade” em FI_1 com seu pretense sinônimo “instituição” em FI_2 percebe-se que são diferentes, o que impede seu uso como sinônimo.

Para se determinar a coesão entre as Frases 1 e 2, avalia-se $FE_1 \cap FE_2$ e $FI_1 \cap FI_2$. Caso $RFI_{1,2} = \{\}$, busca-se estabelecer um relacionamento mínimo entre as entidades do FI_1 e FI_2 , por meio da comparação da primeira letra maiúscula do rótulo semântico de cada entidade em FI_1 e FI_2 , no exemplo: “H” e “HH” = H, sendo que H representa o termo humano. Finalmente, tem-se $RFE_{1,2} = \{\}$ e $RFI_{1,2} = \{\text{“H”}\}$ o qual, após comparado com a Tabela 1, significa que há uma Mudança no tópico tratado pelas frases.

Para se determinar a coesão entre dois parágrafos, avalia-se da maneira exemplificada acima. Ressalte-se que a avaliação dos pares de parágrafos tem adjacência máxima. A exploração dos detalhes de avaliação desse componente é apresentada em Nobre e Pellegrino (2010). A Figura 4 apresenta o resultado da avaliação de um texto.

Avaliação das frases 1 - 2 --> MD Tópico	
Avaliação das frases 2 - 3 --> MN Tópico	
Avaliação das frases 3 - 4 --> MD Tópico	
Avaliação das frases 4 - 5 --> Elaboração	Avaliação dos parágrafos 1 - 2 --> Elaboração
Avaliação das frases 5 - 6 --> MD Tópico	Avaliação dos parágrafos 1 - 3 --> MD Tópico
Avaliação das frases 6 - 7 --> MN Tópico	Avaliação dos parágrafos 1 - 4 --> MD Tópico
Avaliação das frases 7 - 8 --> MD Tópico	Avaliação dos parágrafos 2 - 3 --> Elaboração
Avaliação das frases 8 - 9 --> MD Tópico	Avaliação dos parágrafos 2 - 4 --> MD Tópico
Avaliação das frases 9 -10 --> Elaboração	Avaliação dos parágrafos 3 - 4 --> MD Tópico
Avaliação das frases 10 -11 --> MN Tópico	
Avaliação das frases 11 -12 --> Elaboração	
Avaliação das frases 12 -13 --> MD Tópico	
Avaliação das frases 13 -14 --> MD Tópico	
Avaliação das frases 14 -15 --> MD Tópico	
Avaliação das frases 15 -16 --> MD Tópico	
Avaliação das frases 16 -17 --> MD Tópico	

Figura 4 - Apresenta a avaliação da coesão local e global.

Com intuito de complementar o processo de análise de coesão, foram elaborados módulos específicos para detectar: o uso de recursos coesivos responsáveis por problemas de coesão, como ambiguidade e contradições locais que não afetam o sentido global do texto; mudança da forma impessoal para a pessoal ou vice-versa; períodos incompletos; uso inadequado do léxico.

A detecção de ambiguidade leva em consideração a ordem dos vocábulos na oração, pois o deslocamento desses pode levar a diferentes interpretações de um mesmo enunciado, observe-se o exemplo:

O menino admirava a cantora que dançava com um olhar lânguido. (3)

Na Frase 3, quem estava “com um olhar lânguido”: o menino ou a bailarina? A posição do adjunto adverbial na oração não permite estabelecer com exatidão a quem se refere o termo.

Tanto a identificação de termos que causam ambiguidade como de períodos incompletos se faz por meio do uso de expressões regulares sobre a estrutura da frase, a qual é concebida especificamente para cada tipo de problema a detectar.

A Figura 5 apresenta um fragmento da expressão regular, representada pela palavra “gramatica”, utilizada para localizar expressões que designam finalidade e uma ação a ser executada dentro de uma oração. A palavra “sentença” representa os constituintes da frase, bem como a sua classificação gramatical. O termo “oração” designa a sentença na forma original.

A mudança da forma pessoal para impessoal dentro de um mesmo texto dissertativo representa problemas no foco do emissor do discurso, o qual interfere diretamente na coesão textual, motivo pelo qual é importante verificar a uniformidade de tratamento. A verificação se dá por intermédio da mudança da flexão verbal da 1ª pessoa do plural para a 3ª pessoa do singular ou vice-versa.

```

gramatica = r""""
FIN: {<PRP><artd>*<N>*<PRP>*<VINF>} #
ACAO: {<VFIN><VINF>} #
""""

sentenca = [('com','PRP'),('o','artd'),('intuito','N'),('de','PRP'),('acabarmos','VINF'),
('com','PRP'),('o','artd'),('bullying','N'),('','PU'),('esse','dem'),('comportamento','N'),
('ridículo','ADJ'),('','PU'),('que','rel'),('vem','VFIN'),('atrapalhando','VGER'),('a','artd'),
('vida','N'),('e','KC'),('o','artd'),('aprendizado','N'),('de','PRP'),('crianças','N'),('','PU'),
('adolescentes','ADJ'),('','PU'),('jovens','ADJ'),('e','KC'),('adultos','ADJ'),('','PU'),
('devemos','VFIN'),('repreender','VINF'),('qualquer','DET'),('abuso','N'),('físico','ADJ'),
('e','KC'),('psicológico','ADJ'),('de','PRP'),('um','NUM'),('indivíduo','N'),('sobre','PRP'),
('outro','DET'),('','PU')]

oracao = "Para acabarmos com o "bullying", esse comportamento ridículo, que vem
atrapalhando a vida e o aprendizado de crianças, adolescentes, jovens e adultos,
devemos repreender qualquer abuso físico e psicológico de um individuo sobre outro,
levando em conta, que independentemente de nossa classe social, somos todos
humanos."

```

Figura 5 – Fragmento da heurística de localização de período incompleto.

5. Experimentos

Com o propósito de validar o método proposto, a avaliação automática de coesão textual foi aplicada ao CorpusR, e os resultados comparados com a pontuação estabelecida por dois avaliadores humanos, membros da banca avaliação de redação de vestibular.

Para efetivar a comparação com as notas atribuídas pelos avaliadores humanos tornou-se necessária a criação de uma classificação que pudesse relacionar a pontuação atribuída com os erros identificados no processo de análise de coesão, conforme exibido na Tabela 2.

A classificação é estabelecida pela severidade do erro, ou seja, o erro mais crítico determina a classe à qual pertence à redação sob análise.

Tabela 2 – Relacionamento entre classe e nota atribuída por banca.

Classes	Formas de utilização dos elos coesivos	Nota
1	Utiliza recursos coesivos da língua que afetam a coerência global do texto.	0,0 – 0,4
2	Utiliza recursos coesivos da língua que não afetam a coerência global do texto.	0,5 – 0,8
3	Utiliza, ainda que com alguns problemas, recursos coesivos da língua, adequadamente.	0,9 – 1,2
4	Utiliza, sem demonstrar grande refinamento, os recursos da língua para o desenvolvimento do tipo de texto	1,3 – 1,6
5	Utiliza com proficiência os recursos coesivos da língua para o desenvolvimento do tipo de texto.	1,7 – 2,0

Importante observar que as redações analisadas não apresentavam problemas de elos coesivos capazes de afetar a coerência global, nem de utilização proficiente dos recursos coesivos. Assim, os extremos da classificação ficaram fora da avaliação do método proposto.

Após análise e classificação das redações bem como de comparação com a classificação das notas originais, verificou-se que 90% das redações analisadas automaticamente foram corretamente classificadas. As redações classificadas em classe distinta da original representaram 10%, sendo que duas redações tiveram uma pior classificação e quatro foram mais bem classificadas.

Ao inspecionar as redações que divergiram na classificação, observou-se que as melhores classificadas foram favorecidas pela falta de regras para identificar seus problemas de coesão. Quanto às duas que tiveram pior classificação, não se percebeu qualquer falha nas regras.

6. Conclusões

A abordagem apresentada neste artigo é consistente e passível de adoção em processos de avaliação de coesão textual em larga escala. Contudo, requer a anotação morfosintática e semântica dos textos, sendo que a anotação semântica pode ser crítica à automatização do processo, visto depender da identificação correta do sentido empregado em cada porção de texto. Tal identificação pode ser efetuada manualmente por especialista, como exemplos a WordNet.PT (Marrafa 2002) e a WordNet.BR (Dias da Silva et al. 2002), ou por intermédio de modelos estatísticos.

Os resultados obtidos a partir dos experimentos reforçam a possibilidade de uso do método proposto a fim de avaliar qualitativa e quantitativamente a coesão em texto de redação dissertativa e argumentativa. Além disso, uma análise automatizada baseada em método que explora as potencialidades do computador, e não sofre interferências externas como fadiga, alteração de humor, permite avaliar e analisar sempre de forma equânime. Entretanto, percebe-se a necessidade de revisão das expressões regulares visando detectar problemas não identificados até o presente.

Por fim, a adoção do método proposto acelera o processo de análise e diminui a carga de trabalho do avaliador humano, contudo, e não menos necessária é a avaliação humana no processo, para que seja sempre justa, e servindo de ferramenta para apoiar o desenvolvimento correto da escrita, melhorando a produção de textos.

References

- Bick, E. (2000) The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. PhD Thesis, Arhus University, Arhus.
- Blutner, R. (2000) Some aspects of optimality in natural language interpretation. *Journal of Semantics*, v. 17, p. 189-217.
- Dias da Silva, B. C. et al. (2002) Groundwork for the Development of the Brazilian Portuguese Wordnet. In: Proc. Advances in Natural Language Processing: 3rd International Conference, LNAI, pp. 189-196. Springer Verlag.
- Freitas, S. A. A. de e Lopes, J. G. P. (1994) Discourse segmentation: Extending the centering theory. In: XI Simpósio Brasileiro de Inteligência Artificial. UFCE - Fortaleza - CE: [s.n.].
- Grosz, B. and Sidner, C. L. (1986) Attention, intentions and the structure of the discourse. *Computational Linguistics*, v. 12, n. 3, p. 175-204.
- Grosz, B. J. (1977) The Representation and Use of Focus in a System for Understanding Dialogs. SRI International, Menlo Park, California.

- Grosz, B. J. et al. (1983) Providing a unified account of definite noun phrases in discourse. In: Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics. MIT, USA. p. 44-50.
- Grosz, B. J. et al. (1995) Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, v. 21, n. 2, p. 203-225.
- Gruber, J. (1976) *Lexical Structure in Syntax and Semantics*. [S.I.]: North Holland Publishers.
- Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. London:Longman.
- Louwerse, M. M. (2002). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, 12, 291-315.
- Marrafa, Palmira (2002) Portuguese wordnet: general architecture and internal semantic relations. *DELTA*, 18:131-146.
- Maziero, E.G. et al. (2008) A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. VI Workshop em Tecnologia da Informação e da Linguagem Humana (TIL), pp. 390-392.
- McNamara, D. S. et al. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction*, 14, 1-43.
- Nobre, J. C. S. e Pellegrino, S. R. M. (2010) **Avaliador Automático de Coesão Textual em Redação Dissertativa - AVAC**. Trabalho a ser apresentado no 21º Simpósio Brasileiro de Informática na Educação – XXI SBIE, João Pessoa. No prelo.
- Sidner, C. L. (1979) *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse*. Tese (Doutorado) - MIT, Cambridge, MA, USA.
- Sidner, C. L. (1981) Focusing for interpretation of pronouns. *American Journal for Computational Linguistics*, v. 7, n. 4, p. 217-231.
- Townsend, D. J. and Bever, T. C. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.