# Avaliador Automático de Coesão Textual em Redação Dissertativa - AVAC

**João Carlos Silva Nobre[1], Sérgio Roberto Matiello Pellegrino[1]**

[1] Divisão 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 assessment of textual cohesion in essays discourse in Portuguese based on a method developed with the aid of the Centering Theory and Theory Focusing, and application of Fuzzy Logic. The prototype receives a text as input, and it assesses the structure of discourse cohesion in order to identify breaks, terms that undermine cohesion. The experimental results indicate that 85% of the essays are scored in the same range of scores assigned by human assessors.*

***Resumo.*** *Este artigo apresenta o processo de avaliação automática de coesão textual em redações dissertativas em português com base em um método desenvolvido a partir da Teoria da Centragem e do Foco, e aplicação da Lógica Difusa. O protótipo desenvolvido recebe como entrada um texto e deve avaliar a estrutura do discurso visando identificar quebras de coesão local e global, termos que prejudicam a coesão. Resultados do experimento indicam que em 85% das redações avaliadas são pontuadas na mesma faixa das notas atribuídas por avaliadores humanos.*

## 1. Introdução

Segundo Leffa (1996), atualmente, há uma preocupação com a macroestrutura do texto além da microestrutura, em que fatores como a organização do texto, a coesão, a coerência, o conceito do texto sensível ao leitor são visto como facilitadores da compreensão. Um texto coeso apresenta características que podem facilitar a compreensão por parte do leitor por meio do uso de elementos anafóricos, de marcadores discursivos entre as orações, por apresentação de informações completas e definições explícitas.

De acordo com DuBay (2004), até 1980 já existiam por volta de 200 fórmulas superficiais de inteligibilidade para a língua inglesa, as quais não conseguem capturar a coesão e dificuldade de compreensão de um texto (McNamara et al. 2002), nem avaliar mais profundamente as razões e correlações de fatores que tornam um texto difícil de ser entendido.

Desenvolver métodos e modelos para auxiliar os alunos na produção de textos escritos com um mínimo de inteligibilidade parece ser um desafio grande, pois representar conhecimento, estabelecer relações entre conceitos, entender e compreender uma passagem de texto, e atribuir um valor que represente o grau de coesão de um texto são tarefas computacionalmente complexas.

O processo de avaliação de coesão em redação dissertativa constitui uma habilidade específica, pois exige tanto conhecimentos de mundo quanto linguístico e,

ainda, é caracterizado por um alto grau de subjetividade. Ao considerar esses motivos, vislumbra-se a aplicação da Lógica *Fuzzy* para estabelecer o grau de coesão de um texto, tendo em vista ela propiciar a utilização de conhecimento e procedimentos inferenciais para resolver problemas que normalmente requerem muita perícia humana.

O presente artigo encontra-se estruturado da seguinte forma: na seção 2 há uma breve fundamentação teórica sobre Coesão e Lógica *Fuzzy*; na seção 3 explana-se sobre os trabalhos correlatos, na seção 4, é conduzida a caracterização à avaliação automática da coesão em texto dissertativo e, na seção 5, o artigo é encerrado com observações decorrentes desta pesquisa.

## 2. Fundamentação Teórica

### 2.1. Coesão

Segundo Halliday *and* Hasan (1976), a coesão textual refere-se à compreensão de um escrito, formada pelas relações entre seus termos, às quais lhe denotam sentido e lhe definem como texto. A efetiva decodificação acontece devido a essas relações e se baseia no sistema léxico-gramatical Assim, há formas de coesão realizadas por intermédio da gramática e do léxico. São fatores de coesão a referência, a substituição, a elipse, a conjunção e a coesão lexical.

Para Beaugrande *and* Dressler (1981) apud Koch (2009), a coesão está relacionada à maneira como os componentes da superfície textual, quais sejam palavras e frases, encontram-se conectadas entre si numa superfície linear, por meio de dependências de ordem gramatical.

### 2.2. Lógica Difusa – Lógica *Fuzzy*

Os Conjuntos *Fuzzys* - CF (Zadeh 1965) e a Lógica *Fuzzy* - LF provêem a base para geração de técnicas que visam solucionar problemas nas áreas de controle e de tomada de decisão. Mais recentemente, tem-se considerado a possibilidade de representar informação semântica na forma de relações *fuzzys*.

A Lógica *Fuzzy* fundamenta-se na teoria dos Conjuntos *Fuzzys*, sendo esta uma generalização da teoria dos Conjuntos Tradicionais visando resolver os paradoxos gerados a partir da classificação “verdadeira ou falsa” da Lógica Clássica. Na Lógica *Fuzzy* uma premissa varia em grau de verdade de 0 a 1, inclusive, ou seja, falsa, parcialmente falsa, parcialmente verdadeira ou verdadeira, segundo Klir (2005).

#### 2.2.1. Conjuntos Difusos – Conjuntos *Fuzzys*

Um Conjunto *Fuzzy* é definido em um universo de discurso (conjunto base) X, e caracterizado pela sua função de pertinência:

$$A(x) : X \rightarrow [0,1] \quad (1)$$

onde A(x) representa o grau com que x pertence a A e expressa a extensão com que x se enquadra na categoria representada por A.

Uma função de pertinência particular pode ser visualizada por meio da Equação 2. Como se constata, esta função é triangular e as variáveis a, b e c são os parâmetros.

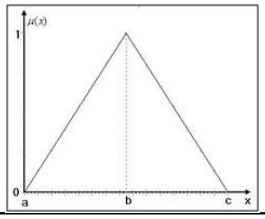
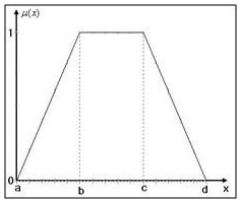
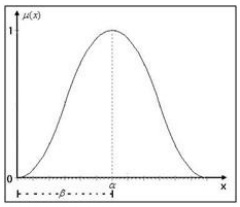
$$\mu(x) = \begin{cases} \dfrac{x-a}{b-a} & se\ x\ \in [a,b] \\ \dfrac{c-x}{c-b} & se\ x\ \in [b,c] \\ 0, caso\ contrário \end{cases} \quad (2)$$

As principais operações e relações entre Conjuntos *Fuzzys* são definidas como extensão das operações e relações da teoria dos Conjuntos Tradicionais

### 2.2.2. Funções de Pertinência

Na Tabela 1 são apresentadas as principais funções de pertinência e suas representações gráficas. As funções de pertinência podem ser lineares ou não-lineares. As lineares, aqui, são a triangular, a trapezoidal; e a não linear é a Gaussiana.

**Tabela 1. Principais Funções de Pertinência**

| Função | Equação | Representação Gráfica |
|---|---|---|
| Triangular | Equação 2 |  |
| Trapezoidal | $\mu(x) = \begin{cases} 0, & se \quad x \le a \\ \dfrac{x-a}{b-a}, & se\ a < x \le b \\ 1, & se\ b < x \le c \\ \dfrac{d-x}{d-c}, & se\ c < x \le d \\ 0, & se \quad x > d \end{cases}$ (3) |  |
| Gaussiana | $\mu(x) = \begin{cases} 0, & se \quad x \le \alpha - \beta\ e\ x \ge \alpha + \beta \\ \frac{2}{\beta^2}(x-\alpha+\beta)^2, & se\ \alpha-\beta \le x \le \alpha - \frac{\beta}{2} \\ 1-\frac{2}{\beta^2}(x-\alpha)^2, & se\ \alpha-\frac{\beta}{2} \le x \le \alpha+\frac{\beta}{2} \\ \frac{2}{\beta^2}(x-\alpha-\beta)^2, & se\ \alpha+\frac{\beta}{2} \le x \le \alpha+\beta \end{cases}$ (4) |  |

### 2.2.3. Regras Difusas e Variáveis Linguísticas

Conjuntos *Fuzzys* e operadores *fuzzys* são os sujeitos e os verbos da Lógica *Fuzzy*. As declarações de regras “if-then” são usados para formular as instruções condicionais que compreendem a LF.

A simples regra *fuzzy IF-Then* assume a forma,

Se x é A Então y é B,

onde A e B são valores linguísticos definidos por conjuntos difusos nas escalas (universos de discurso) X e Y, respectivamente. A parte “Se” da regra "x é A" chama-se

antecedente ou premissa, enquanto que a parte "Então" da regra "y é B" nomina-se consequente ou conclusão.

Uma variável linguística difusa é expressa: (a) qualitativamente - por um termo linguístico; e (b) quantitativamente - por uma função de pertinência, sendo caracterizada pela tupla {n,T,X,m(t)}, onde n é o nome da variável, como, por exemplo, índice de Coesão; T é o conjunto de termos linguísticos de n, como, por exemplo, baixo, médio e alto; X é o domínio de valores de n sobre o qual o significado do termo linguístico é determinado; e m(t) é uma função semântica que assinala para cada termo linguístico $t \in T$ o seu significado, que é um conjunto difuso em X, ou seja, $m : T \rightarrow X$, onde X é o espaço dos conjuntos difusos.

## 3. Trabalhos Relacionados

Os trabalhos relacionados à inteligibilidade de texto se utilizam de índices e/ou fórmulas para avaliar a dificuldade ou a facilidade de compreensão de textos escritos. Para textos em língua inglesa existem diversas fórmulas. Já para o idioma português do Brasil são escassas as pesquisas nesta área. O trabalho de Martins et al. (1996) adaptou o índice Flesch e, recentemente, Scarton (2009) adequou 30 métricas do Coh-Metrix para o português, as quais são utilizadas no projeto Simplificação Textual do Português para Inclusão e Acessibilidade Digital - PorSimples (Aluisio et al. 2008).

As fórmulas de inteligibilidade superficiais *Flesch Reading Ease* e *Flesch-Kincaid Grade Level* são as mais conhecidas no Brasil. Elas são limitadas e se baseiam, somente, na média de palavras por sentenças e na média de sílabas por palavra na avaliação do texto.

A ferramenta computacional Coh-Metrix produz índices de representações linguísticas e discursivas de um texto. Esses indicadores podem ser usados de muitas maneiras para investigar a coesão em si e a coerência da representação mental do texto. A versão livre da ferramenta possui 60 índices, os quais são divididos em 6 classes: Identificação Geral e Informação de Referência; Índices de Inteligibilidade; Palavras Gerais e Informação do Texto; Índices Sintáticos; Índices Referenciais e Semânticos; e Dimensões do Modelo de Situações.

O AVAC, assim como o Coh-Metrix, se utiliza de conhecimentos linguísticos na determinação dos índices, além de identificar e localizar os problemas de coesão no texto. Diante do que, é possível afirmar, em uma comparação simples, não obstante este trabalho não ter a mesma amplitude de índices do Coh-Metrix, a sua abordagem quanto à coesão é mais profunda, visto que identifica as partes de texto que comprometem a coesão textual, permitindo a avaliação da coesão de forma subjetiva por intermédio da lógica *fuzzy*.

## 4. O Processo de Avaliação

O modelo de avaliação 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 Palavras*, de modo a obter o conteúdo de entrada para o processo de análise e, em seguida, processado pelo avaliador automático de coesão. A Figura 1 apresenta a estrutura de funcionamento.

Um Corpus de teste utilizando 60 redações de vestibular, previamente analisado, foi submetido ao Sistema para treinamento. Esse conjunto foi chamado de CorpusR, contendo 16870 tokens, 731 sentenças e 373 parágrafos.

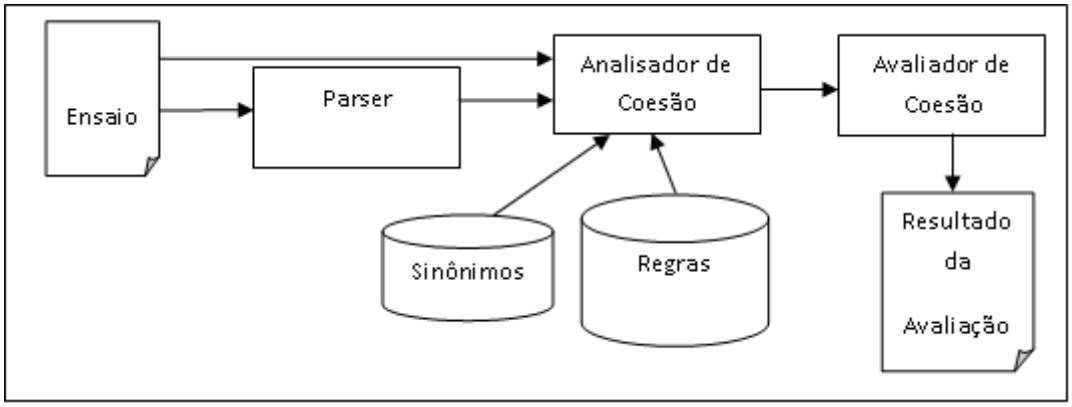



**Figura 1. Estrutura do processo de avaliação de coesão.**

O analisador de coesão apoia-se na Teoria da Centragem - TC (Grosz et al. 1983) e na Teoria do Foco - TF (Sidner 1981, 1979), 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 algoritmos para analisar a coesão local, descrito na Figura 2, e a global que segue os mesmos princípios da local (Nobre e Pellegrino 2010).

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 $FE_1$ que possuem representante na lista de $FE_2$ 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 FIi e FIi+1 que são representados em RFFEi, i+1
  Calcular a intersecção entre os elementos do FIi e FIi+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.**

A Tabela 3, adaptada da TC e TF, apresenta os possíveis relacionamentos existentes entre o foco explícito e implícito na determinação da coesão e a pontuação atribuída a cada relação. Algumas leituras do relacionamento entre o FE e FI são:

- Se existe um elemento de $FE_i$ em $FE_{i+1}$ ($FE_i$ C $FE_{i+1}$) e, também, existe um elemento de $FI_i$ em $FI_{i+1}$ ($FI_i$ C $FI_{i+1}$), então as frases $F_i$ e $F_{i+1}$ estão num processo de elaboração, visto que compartilham as mesmas entidades.

- Se não existe elemento de $FE_i$ em $FE_{i+1}$ ($FE_i$ NC $FE_{i+1}$), nem de $FI_i$ em $FI_{i+1}$ ( $FI_i$ 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 2. Relação entre FE e FI para estabelecimento da coesão e pontuação.**

| | $FE_i$ C $FE_{i+1}$ | Pontuação | $FE_i$ NC $FE_{i+1}$ | Pontuação |
|---|---|---|---|---|
| $FI_i$ C $FI_{i+1}$ | Elaboração | 1 | Mudança de tópico | 0,5 |
| $FI_i$ NC $FI_{i+1}$ | Manutenção do tópico | 0,75 | Mudança de assunto | 0,0 |

Para avaliar qualitativa e quantitativamente a coesão textual utilizar-se-á a abordagem da Lógica *Fuzzy* - LF, empregando-se o *Fuzzy Logic Toolbox* do MatLab 7, devido a habilidade em propiciar a inferência de conclusões e a geração de respostas baseadas em informações imprecisas.

Trabalhar-se-á com 3 variáveis de entrada: (a) Índice de Coesão - IC, exibido na Equação 5; (b) Índice de Ambiguidade e Contradição – ICA, representado na Equação 6; e (c) Índice de Forma – IF, representado na Equação 7, e uma variável de saída, denominada Coesão *Fuzzy* – CF, na construção de um Sistema de Inferência *Fuzzy* – SIF.

O IC estabelece o peso das relações constituídas entre frases e os parágrafos do ensaio textual visando determinar a força coesiva do texto:

$$IC = \left( \frac{\sum_{i=1}^{s-1} PSf_{i,i+1}}{s-1} + \frac{\sum_{j=1}^{p-1} PSP_{j,j+1}}{\sum_{j=1}^{p-1} j} \right) / 2 \quad (5),$$

onde “s” e “p” representam, respectivamente, o total de sentenças e de parágrafos do texto, $PSf_{i,\ i+1}$ representa a pontuação da relação entre frases adjacentes, $PPf_{i,\ i+1}$ representa a pontuação da relação entre parágrafos adjacentes.

O IAC determina o índice imprecisão do texto sob avaliação, ou seja, o peso das estruturas que comprometem a compreensão do escrito:

$$IAC = (TA + TC)/s \quad (6),$$

onde “TA” e TC representam, nessa ordem, o Total de Ambiguidades e o Total de Contradições encontradas no texto.

Já o IF relaciona o peso das estruturas que comprometem a coesão, mas não interferem na coerência do texto:

$$IF = TPr/s \quad (7),$$

onde TPr indica a quantidade de termos com problemas, tais como: períodos incompletos, introdução de pronomes ou palavras que não estabelecem ligação com o que já foi dito.

A estrutura básica do Processo de Inferência *Fuzzy* - PIF da CF é exibida na Figura 3. O fluxo de informação segue da esquerda para direita, a partir de 3 entradas para a saída. A natureza paralela das regras é um dos aspectos mais importantes dos sistemas de LF. Em vez de alternar entre os modos baseados em ponto de quebra, a

lógica flui sem problemas de regiões onde o comportamento do sistema é dominado por uma regra e/ou por outra.

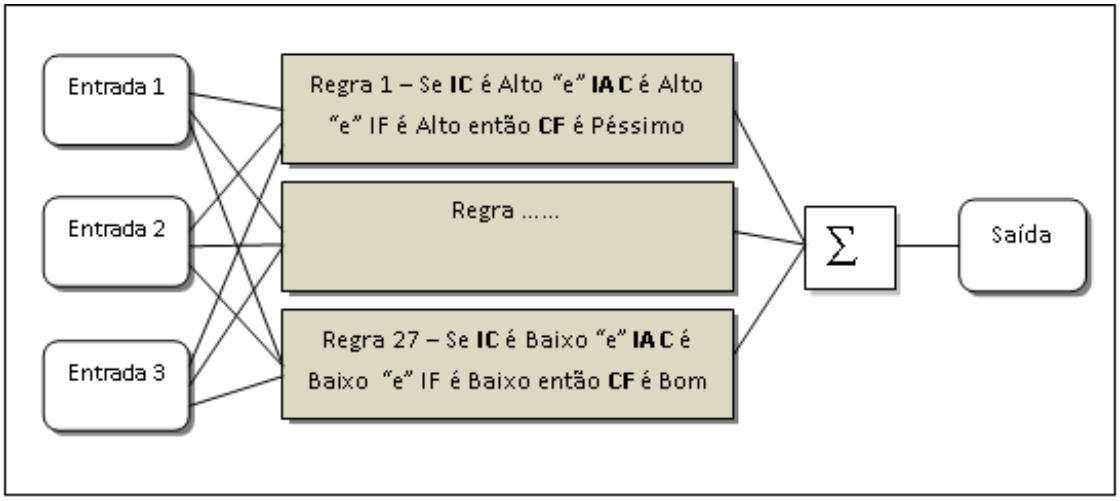



**Figura 3. Estrutura básica do processo de inferência *fuzzy*.**

O PIF é composto por 5 partes: fuzificação das variáveis de entrada; aplicação do operador *fuzzy*; aplicação do método de implicação; agregação de todas as saídas; e desfuzificação.

A Tabela 3 exibe os dados referentes às variáveis de entrada e saída utilizadas, e a Tabela 4 mostra as 9 primeiras regras de um total de 27 regras necessárias à determinação do valor de coesão do texto (CF). Adotou-se o método Mamdani com os operadores *fuzzy* "and" – min(A,B) e "or" – max(A,B), método de implicação – min(A,B) e agregação – max(A,B), método de desfuzificação – centróide na configuração do SIF.

**Tabela 3. Parâmetros de configuração do SIF.**

| Variável | Tipo | Faixa de valores | Função de Pertinência-FP | Parâmetros da FP |
|---|---|---|---|---|
| IC | Entrada | [0 1] | Gaussiana | Baixo – [0.1 0]<br>Médio – [0.1 0.4]<br>Alto – [1.5 1] |
| IAC | Entrada | [0 1] | Gaussiana | Baixo – [0.1 0]<br>Médio – [0.1 0.4]<br>Alto – [1.5 1] |
| IF | Entrada | [0 1] | Gaussiana | Baixo – [0.1 0]<br>Médio – [0.1 0.4]<br>Alto – [1.5 1] |
| CF | Saída | [0 2] | Triangular | Péssimo – [0 0 0.4]<br>Ruim – [0.4 0.6 0.8]<br>Boa – [0.8 1.0 1.2]<br>Ótima – [1.2 1.4 1.6]<br>Excelente – [1.6 1.8 2] |

**Tabela 4. Conjunto parcial de regras.**

| Regra | IC | Operador. | ICA | Operador | IF | Implicação | CF |
|---|---|---|---|---|---|---|---|
| 1. | Alto | And | Alto | and | Alto | ➔ | Péssima |
| 2. | Alto | And | Alto | and | Médio | ➔ | Péssima |
| 3. | Alto | And | Alto | and | Baixo | ➔ | Péssima |
| 4. | Alto | And | Médio | and | Alto | ➔ | Péssima |
| 5. | Alto | And | Médio | and | Médio | ➔ | Péssima |
| 6. | Alto | And | Médio | and | Baixo | ➔ | Ruim |
| 7. | Alto | And | Baixo | and | Alto | ➔ | Boa |
| 8. | Alto | And | Baixo | and | Médio | ➔ | Boa |
| 9. | Alto | And | Baixo | and | Baixo | ➔ | Excelente |

### 4.1. Experimentos

Com o propósito de validar o método proposto, as redações foram analisadas e avaliadas qualitativa e quantitativamente seguindo a dinâmica apresentada e os resultados comparados com a pontuação estabelecida pela comissão de avaliação de redações.

A pontuação da coesão nas redações varia de 0 a 2 pontos e é subdividida conforme exposto em Parâmetros da FP da variável CF na Tabela 3.

A comissão avaliadora é composta por três membros. As redações são corrigidas por dois membros e, caso haja discrepância de quatro décimos nas notas atribuídas pelos dois julgadores, torna-se necessária a avaliação do terceiro membro da comissão para dirimir os pontos controversos.

Cada redação teve seus Índices estabelecidos de forma automática, conforme determinado nas Equações 5-7 e, em seguida, aplicaram-se tais números no SIF, o qual foi configurado seguindo os parâmetros das Tabelas 3 e 4, obtendo-se como resultante a Coesão *Fuzzy*.

O resultado numérico da CF de cada redação foi comparado com média da pontuação atribuída pelos avaliadores visando determinar se o valor da CF está classificado no mesmo intervalo da pontuação atribuída pelos avaliadores humanos, estabelecendo a diferença entre o valor da CF e da média das notas dos julgadores humanos.

Ao analisar os dados comparativos, verificou-se que 70% das redações avaliadas automaticamente foram classificadas na mesma faixa de pontuação que as valoradas pelos humanos. Dos 30% fora da faixa, 15% apresentaram uma diferença inferior a quatro décimos e 15% divergiram em mais de quatro décimos.

A Figura 4 apresenta um roteiro de todo o processo de inferência *fuzzy*. As três primeiras colunas tratam, respectivamente, das variáveis IC, IAC e IF, ou seja, os elementos que compõem a parte “IF” ou antecedente das regras *fuzzy*. A quarta coluna refere-se à parte “THEN” ou consequente das regras. As linhas representam as regras *fuzzy*. Os gráficos em amarelo exibem as regras e as funções de pertinência que estão

sendo referenciadas como antecedentes. Os gráficos em azul exibem as funções de pertinência que são referenciadas pelos consequentes. O gráfico em azul, após a última regra, representa a decisão ponderada agregada do SIF. Para coesão do texto com IC = 1 – Alto, IAC = 0,63 – Médio/Alto e IF = 0 – Baixo, a Coesão *Fuzzy* resultante é 0,3 – Péssima, coincidindo com a pontuação atribuída pelo avaliador humano.

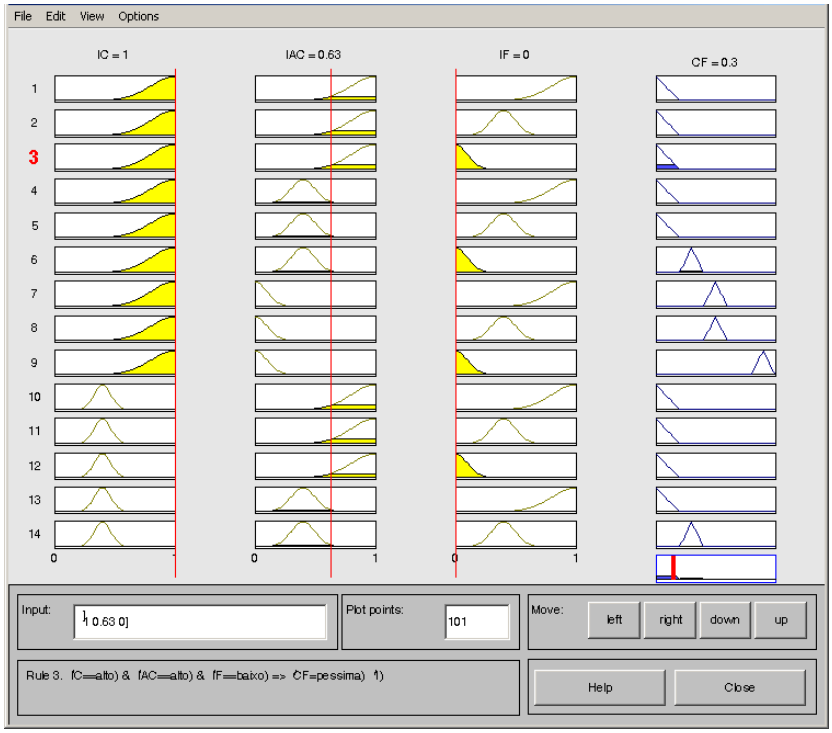



**Figura 4. Resultado da aplicação do SIF com IAC Alto.**

## 5. Conclusões

Os resultados experimentais indicam que o uso da abordagem *fuzzy* proposta é válida e promissora dentro das condições experimentais, pois se conseguiu com a aplicação da Lógica *Fuzzy*, resultados muito próximos aos padrões humanos, para este tipo de atividade.

O método de determinação dos Índices e os intervalos das funções de pertinência são importantes no estabelecimento do valor da CF, visto que concentram as principais características referentes à coesão textual e à natureza subjetiva do processo de avaliação.

O sucesso do método proposto reflete-se no resultado, considerando o alcance de 85% de concordância com as notas atribuídas pelos avaliadores humanos, bem como os 15% de discordância são aceitáveis. Para pesquisas futuras, pretende-se testar o método em diferentes temas, tipos de redação, níveis de estudante e permitir ao aluno escrever um texto dissertativo e submetê-lo à avaliação automática a fim de melhorar a habilidade escrita de forma interativa.

## Referências Bibliográficas

Aluísio, Sandra Maria et al. (2008) Towards Brazilian Portuguese Automatic Text Simplification Systems. Em Proceedings of The Eight ACM Symposium on Document Engineering (DocEng 2008), páginas 240-248, São Paulo, Brasil.

Beaugrande, R. and Dressler, W. U. (1981) Introduction to Text linguisticsLinguistics. London: Longman.

Bick, E. (2000) The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. PhD Thesis, Arhus University, Arhus.

DuBay, Willian H. (2004) The Principles of Readability. <http://www.eric.ed.gov/PDFS/ED490073.pdf> Acesso em janeiro de 2010.

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.

Halliday, M.A.K. and Hasan, R. (1976) Cohesion in English. London:Longman.

Klir, J.G. and Yuan, B. (1995) "Fuzzy Sets and Fuzzy Logic – Theory and Applications". P.Hall.

Koch, I. G. V. (2009) A Coesão Textual. São Paulo: Contexto.

Leffa, V. J. (1996) Fatores da compreensão na leitura. Em Cadernos no IL, v.15, n.15, páginas 143-159, Porto Alegre <http://www.leffa.pro.br/textos/trabalhos/fatores.pdf)>. Acesso em janeiro de 2010.

Martins, T. B. F. et al. (1996) Readability formulas applied to textbooks in brazilian portuguese. Notas do ICMC, N. 28, 11p.

McNamara, D. S. et al. (2002) Coh-MetrixCoh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Institute for Intelligent Systems, University of Memphis, Memphis, TN.

Nobre, J. C. S. e Pellegrino, S. R. M. (2010) **ANAC: um analisador automático de coesão textual em redação.** Trabalho a ser apresentado no 21º Simpósio Brasileiro de Informática na Educação – XXI SBIE, João Pessoa. No prelo.

Scarton, C. E. et al. (2009) Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. In: 7o Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana. São Carlos, Brasil: Universidade de São Paulo. ISSN 2175-6201.

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.

Zadeh, L. A. (1965) Fuzzy Sets, Inf. Control 8, 338-353.