

# ANÁLISE MANUAL E AUTOMÁTICA DE PISTAS LEXICAIS DE PRESENÇA SOCIAL EM CHAT EDUCACIONAL

**Resumo:** O artigo apresenta um estudo de interações discursivas feitas em um chat em contexto educacional de forma a verificar a presença social dos sujeitos em ambiente de aprendizagem a distância. Os procedimentos usados foram a análise textual convencional e automática com base nos critérios de afetividade, interatividade e coesão. Verificou-se que a mineração da linguagem diferenciada usada no gênero textual chat demanda o aprimoramento do software usado no experimento.

**Abstract:** This paper presents a study of discursive interactions in an educational chat aiming at detecting social presence in on line learning environments. Methods used in the study include conventional and automatic text analysis based on the following criteria: affect, interactivity and cohesion. Results show that processing the distinct language of chats demands improvement of further resources in the software used in the experiment.

## 1.Introdução

Este trabalho se apóia na premissa de que as interações baseadas em texto em contextos virtuais de aprendizagem revelam diferentes aspectos socioafetivos do aluno, entre eles, seu grau de satisfação, envolvimento e sentimento de pertencimento no grupo. Nesse sentido, a análise de postagens feitas em fóruns e chats educacionais pode fornecer indícios do grau de envolvimento dos sujeitos em cursos a distância, porquanto são indicadoras da *presença social* dos alunos nesses ambientes (Rourke et al., 2001).

A noção de *presença social* (PS) é definida por Gunawardena (1995) como o grau de como os sujeitos se percebem e são percebidos como “reais” em interações mediadas por computador. Para Tu e McIsaac (2002), PS refere-se ao sentimento ou percepção dos indivíduos de estarem conectados com outros por recursos de comunicação mediada por computador (CMC). Segundo Garrison (2007), as variáveis e a complexidade do discurso em meio eletrônico oferecem, ainda, vasto campo de investigação qualitativa ou quantitativa. Para Giordan e Dotta (2008), os desafios que se apresentam nessa área se devem, em parte, ao fato de que as novas formas de expressão lingüística usadas na CMC não constituem, ainda, um sistema consolidado.

O estudo teve como objetivos (i) mapear pistas lexicais denotadoras de *presença social* por meio de análise textual convencional em um chat e, (ii) verificar o potencial e as limitações de um programa de mineração de texto no processamento das marcas lingüísticas diferenciadas usadas nesse gênero comunicacional. O chat selecionado para estudo foi extraído de um curso de extensão sobre gestão de redes sociais para profissionais na área de Meio-Ambiente. A análise textual foi realizada conforme os critérios de Rourke et al. (2001) para avaliação do grau de PS em ambientes virtuais de ensino e aprendizagem (AVEAs). O mesmo *corpus* de análise foi utilizado no teste de processamento automático feito no software Eureka (Wives, 1999).

## 2. Presença Social em comunicações mediadas por computador

Segundo Palloff e Pratt (2002, p. 53), “[...] na educação a distância, deve-se prestar atenção ao desenvolvimento da sensação de comunidade entre os participantes do grupo a fim de que o processo seja bem-sucedido [...]”. Essa percepção de *comunidade* é viabilizada, entre outras estratégias, pelas ferramentas de CMC – recursos que potencializam os sentimentos de *pertencimento* e de *presença* do aluno em AVEAs e minimizam a freqüente sensação de isolamento observada em experiências de educação a distância (EaD). Aykol, Garrison e Ozden (2009) explicam que a noção de PS envolve a *identificação* dos sujeitos *com* a comunidade de aprendizagem e o desenvolvimento de relações interpessoais que projetem as personalidades individuais.

Conforme a perspectiva de Short, Williams e Christie (1976), as interações baseadas em texto apresentam *baixo grau de presença* devido à ausência de pistas não-verbais (expressões faciais, gestos, articulações sonoras, distância entre os falantes), sendo, portanto, menos pessoais e “inferiores” do que as face a face (FaF). Em oposição, Walther (1992) argumenta que os dois contextos de interlocução são propícios ao desenvolvimento de relacionamentos e que os sujeitos conseguem formar impressões sobre os outros com base somente no conteúdo lingüístico das mensagens.

Um modelo teórico que destaca a “presença” em atividades de aprendizagem em espaços físicos e/ou virtuais é o *Community of Inquiry – CoI* (Comunidade de Investigação), desenvolvido por Garrison (1991). Esse modelo se fundamenta na premissa de que a “experiência de aprendizagem” (*learning experience*) envolve o desenvolvimento e integração de três tipos de presença – *cognitiva, docente e social*.

A *presença cognitiva* refere-se ao quanto o aprendiz consegue construir significados por meio de reflexão e expressão discursiva de teor crítico. A *docente* (ou *de ensino*) relaciona-se ao desenho e gestão das sequências pedagógicas e promoção de aprendizagem significativa. A *presença social* refere-se à habilidade do aprendiz em se projetar social e emocionalmente em AVEAs. Para Rourke et al. (2001), a presença social é um aspecto relevante na criação de uma atmosfera propícia à aprendizagem *on line*, porque contribui para a permanência do aluno, desenvolvimento da responsabilidade individual, apreensão e domínio dos conteúdos. Esse tipo de presença é caracterizado pela expressividade discursiva, podendo ser verificada pela análise das interações textuais feitas nas ferramentas de comunicação disponíveis em AVEAs.

Rourke et al. (2001) propõe que a verificação de presença social em trocas discursivas via CMC seja feita segundo os indicadores listados no quadro abaixo.

**Quadro 1. Modelo de Rourke et al. (2001) para detecção de PS**

CATEGORIA	INDICADORES
<i>Aspectos afetivos</i>	Expressão da emoção, utilização de humor, auto-revelação
<i>Aspectos interativos</i>	Participação contínua numa discussão, referência explícitas a outras postagens, encorajamento, apreciação e concordância



### 3.1.Verificação de indicadores de presença social no chat (CH1)

Para avaliar o grau de PS nessa experiência, foram utilizados os critérios indicados por Rourke et al (2001): aspectos *afetivos*, *interativos* e *coesivos*.

#### a) Exemplos de marcas linguísticas de afetividade:

- Expressões convencionais ou não-convencionais de afetividade: *tudo blz!! / kogueeia, vo ti add nao<sup>2</sup>/ hum...vlw! Toninho amor!*
- Uso enfático de maiúsculas: *CHEGUEIII! / MEIO AMBIENTE? / BOA TARDE /só a gente foi APROVADO / VCS JÁ ENVIARAM O TRABALHO ...*
- Pontuação repetida: *só a gente???* *OI SUZANAAA!!!!* / *ME atrasei mas cheguei!!!*
- Implicâncias, brincadeiras, ironia, sarcasmo: *To brincando/ kd todo mundo???* / *até agora nadaaa / n vms falar sobre a bolsa de valres? / to zuando / vc não é mole,... // Integração ué? Não é isso que estamos estudando? rrrs...*
- Emoticons: / =P / (n.a.:apenas uma ocorrência neste chat)
- Detalhes da vida extraclasse: *Vc ta on no MSN?/ sai e deixei o pc ligado / estou no trabalho...rrrs / to em casa tb / so q moro numa cidade....*
- Vulnerabilidade: *eu nm sei!/ NÃO SEI!! / vou entrar + não consigo falar ...*

#### b) Exemplos de marcas linguísticas de interatividade

- Perguntas endereçadas a colegas ou tutores: *pod mandar vídeo tb como tarefa? / isso é pra mim Sam? / é pra levar uma webcam amanhã? / e aonde está essa reportagem .../ aglm já fez a tarefa d casa? / tem qe fazer algum comentario, resumo sei laa?*
- Referências e comentários positivos acerca das mensagens de outros: *Boa essa! / Verdade.. / Com certeza / HÁ QUE ÓTIMO... / caramba q bacana!! / Tb pensei nessa hipótese /parece ser mto interessante...*
- Concordância com comentários de outros: *Concordo / SIM / sim sim / concerteza\*\*\*/ verdade!! Eu tenho! / SERIA ÓTIMO/ HÁ TÁ... / tb já pensei nessa hipótese...*

#### c) Exemplos de marcas linguísticas de coesão

- Fala direcionada com nomes próprios ou vocativos: *Toni, comentei sobre.../ Ric, você poderia / gente, tenho que ir / é verdade, Toni... / vamos interagir [...] também, gente!*
- Fala direcionada ou referência a pessoas e fatos usando pronomes pessoais inclusivos: *nós fizemos nossa parte / aqui estamos nós! // se vc quiser / eles sao tranquilos! / e vc?*

- Termos de sociabilidade, saudações e despedidas: *Oii / oiiiiii / Olá, gente / Boa Tarde galera! / Tudo bem??? / bj / bjão / até!!! / Saudades já!!! / see you...*

Observou-se que, nesse chat, houve poucas menções direcionadas à discussão proposta, mas alto teor de interação social. Como é comum nesse gênero, os enunciados são bastante curtos e interpolados pela escrita simultânea feita por vários participantes. Devido à informalidade da conversa, encontram-se inúmeras instâncias de presença social, particularmente pelo uso intenso de marcas de afetividade e coesão. Destaca-se, no conjunto, o uso intensivo de recursos paralinguísticos.

#### 4. Experimento de mineração do CH1

A mineração de textos (MT) é uma área da Ciência da Computação com o objetivo de extrair informações a partir de coleções de documentos. Segundo Feldman e Sanger (2007), a MT pode ser definida como um processo intensivo de conhecimento no qual um usuário interage com uma grande quantidade de documentos utilizando ferramentas para análise dos mesmos.

A MT explora técnicas e métodos provenientes das áreas de recuperação de informação, extração de informação, e linguística computacional. Para extrair informações úteis de documentos textuais, são descobertas características relevantes nos mesmos; sendo as mais utilizadas: caracteres, palavras, termos e conceitos. As tecnologias empregadas na MT incluem extração de informação, rastreamento de tópicos, produção de sumários, categorização de textos, agrupamento de textos, links conceituais, visualização de informação, análise de perguntas e respostas (Fan et al, 2006).

##### 4.1. Mineração do CH1 no Eureka

O Eureka é um programa de mineração de texto desenvolvida por Wives (1999) que permite o processamento de textos não formatados de gêneros variados. O Eureka (EKH) possui recursos que permitem a realização de análise lexical com índices de frequência e relevância, e geração de gráficos das palavras mais relevantes. O EKH também faz agrupamento de textos (*clustering*) – técnica que analisa o conteúdo de textos, identifica os que têm conteúdo semelhante e organiza-os num único grupo (*cluster*).

A interface do software apresenta três tabelas (“orelhas”), nomeadas em língua inglesa, e correspondentes, cada uma, às seguintes etapas (Wives, 1999, p. 43):

#### Quadro 2. Interface do Eureka: tabelas e funções

TABELAS	FUNÇÃO
<i>Stopwords</i>	Definição e manipulação de palavras negativas
<i>Collections</i>	Definição e manipulação de coleções de documentos, definição dos atributos para o processo de agrupamento e início do processo de agrupamento

	(cálculo de similaridades e geração de matriz de similaridades)
<i>Clusters</i>	Agrupamento das coleções cujas matrizes de similaridades já tenham sido calculadas na etapa anterior.

Segundo Wives (1999, p.52), o programa requer a realização da etapa de “limpeza de dados” para evitar erros nos resultados finais. Esse processo de limpeza das chamadas *stopwords* (palavras negativas ou pouco relevantes) inclui “correção ortográfica, normalização de termos e substituição de pronomes por seus respectivos substantivos”. O processo de remoção de *stopwords* é facilitado pelo banco de categorias gramaticais e listas de palavras já existentes no programa. Um aspecto a ser destacado é o fato de o EKH permitir o acréscimo de novas categorias para remoção de *stopwords*, isto é, não listadas no programa, mas desejáveis num dado experimento. A Figura 1 apresenta uma tela do EKH após a seleção dos termos considerados como negativos para o teste.

Como indicado acima, o *corpus* usado na testagem das funcionalidades do EKH foram as postagens feitas no CH1 (cf. seção 3). No processo de remoção de *stopwords*, foram selecionadas as seguintes classes de palavras: artigos definidos e indefinidos, pronomes indefinidos, pronomes relativos e preposições. As conjunções não foram removidas porque, como recursos coesivos, constituem estratégias usadas pelo falante para estabelecer *contraste* ou *oposição* (conjunções concessivas), *propósito* (conjunções finais), *razão* (conjunções causais) etc.

Os pronomes pessoais, possessivos e demonstrativos não foram excluídos porque são recursos para se fazer referências no texto. Conforme Vieira et al. (2008), a “cadeia referencial” estabelecida no texto salienta as implicações dos recursos usados na construção do sentido, assim como as dificuldades que essa estratégia discursiva apresenta para o processamento da linguagem natural. A manutenção dos pronomes no processo de mineração justifica-se, também, porque são termos bastante expressivos de presença social, pois o falante se posiciona ou se dirige explicitamente a seu interlocutor ou acerca de outros sujeitos e fatos sobre o qual estão dialogando.

#### 4.2. Discussão sobre os resultados do processamento no Eureka

Uma característica da escrita em fóruns e chats relevante para este trabalho é seu caráter híbrido, isto é, uma escrita com fortes traços da fala oral. O uso de maiúsculas é um dos recursos do escritor para enfatizar seu discurso e “marcar sua presença” numa discussão. O EKH, entretanto, não faz distinção entre letras maiúsculas e minúsculas, o que descaracteriza o texto de análise segundo os critérios apontados por Rourke et al. (2001) para detecção de presença social. Por outro lado, o programa “aceita” a inserção de traços típicos do “internetês” como palavras abreviadas (*vc*, *tb*) e onomatopéias como “rs” (*risos*), além de interjeições.

Observou-se, também, que o Eureka possui um banco com palavras em inglês e francês. Esse dado é importante para a mineração das interações em ferramentas de CMC devido ao uso frequente de estrangeirismos, convencionais ou não, nesses contextos discursivos. Palavras estrangeiras são utilizadas com frequência, também,

quando a discussão envolve termos da área de Informática ou em referência a produções acadêmicas ou culturais.

Um aspecto a ser considerado é a impossibilidade do EKH em processar palavras compostas e sintagmas. Por essa razão, o desmembramento de certas expressões (*ao longo de, de forma que*), topônimos compostos (*Porto Alegre, Belo Horizonte*), sintagmas nominais (*problema ambiental, decreto-lei*), entre tantas, compromete o sentido real dos significados. A possibilidade de minerar palavras compostas é um encaminhamento sugerido para melhoria do programa.

Outra limitação do EKH é a não localização da palavra no parágrafo ou na oração, assim como os termos que a antecedem e / ou sucedem. Essa lacuna pode comprometer a correta identificação do sentido de um termo.

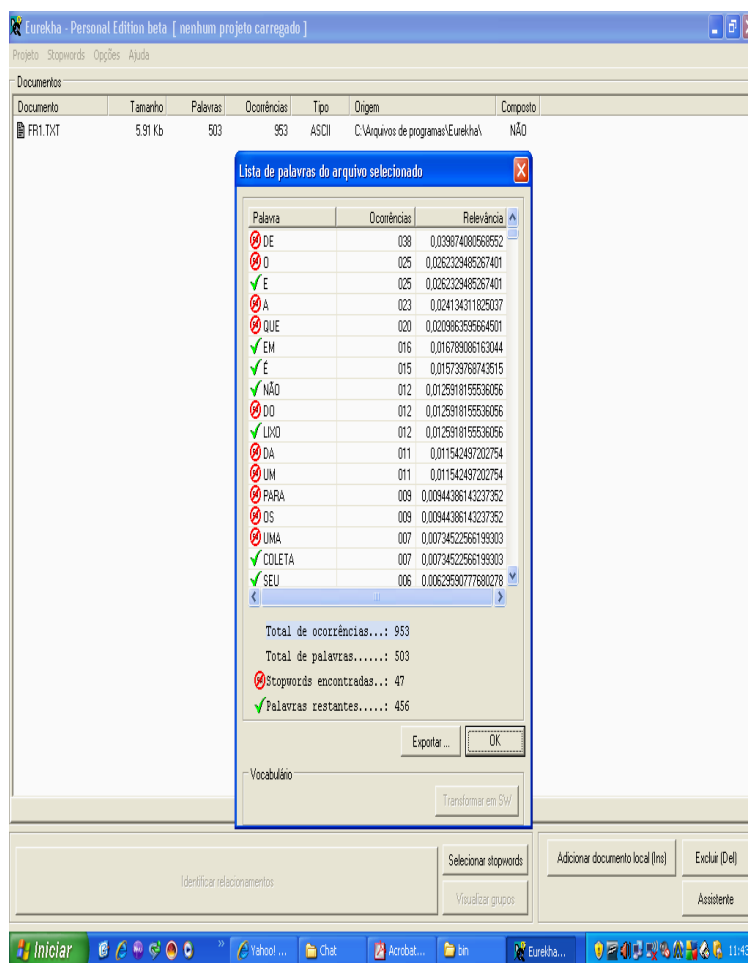


Figura 1. Eureka: janela com *stopwords* e número de ocorrências no CH1

## 5. Discussão

O teste realizado no Eureka mostrou que o software permite o processamento de parte das pistas indicadoras de presença social, e que o levantamento do número de ocorrências dos termos pode fornecer dados quantitativos auxiliares na interpretação dos dados.

Da mesma forma, verificou-se que, o processamento automático de textos não contempla as diversas estratégias e recursos usados na produção textual, tais como o uso de elementos paralingüísticos, entre eles: imagens, *emoticons*, uso expressivo de sinais de pontuação, aspas, e recursos de edição, como negrito, itálico e sublinhado, para citar alguns. Da mesma forma, ficam excluídas outras características do discurso eletrônico informal, como, por exemplo, erros de digitação.

Destaca-se, ainda, que pela análise textual convencional é possível perceber as implicações do contexto tanto na apreciação das escolhas linguísticas feitas na produção do texto quanto no entendimento das referências intra e extra textual. Essas marcas estilísticas são recursos disponibilizados aos usuários da língua para expressar uma gama de intenções comprometidas pelo distanciamento físico entre os interlocutores. Devido às restrições resultantes da ausência de gestos, expressões faciais e entonação, os sujeitos utilizam os recursos gráficos e hipermidiáticos dados pela máquina não só para se mostrarem presentes em conversações via CMC, mas também para que seus interlocutores possam perceber suas intenções afetivas ou referentes ao próprio tema em discussão.

## 6. Considerações finais

Como indicado neste estudo, a análise de texto convencional pode detectar marcas lexicais indicativas de afetividade, interatividade e coesão – aspectos indicadores de presença social. O levantamento de pistas lexicais denotadoras de presença social no CH1 mostrou que os sujeitos usam diferentes recursos linguísticos para marcar seu envolvimento e participação no evento comunicativo, particularmente, pelo uso intenso de traços da fala oral. Neste estudo, os critérios de Rourke et al (2001) para verificação de presença social em interações discursivas feitas em ambientes de aprendizagem on line mostraram-se úteis e pertinentes.

Por outro lado, a detecção de presença social pela análise automática no Eureka mostrou-se insatisfatória devido à impossibilidade do programa em processar as diferentes estratégias discursivas indicadas no modelo de Rourke et al. (2001). Além disso, os resultados indicam que a verificação de presença social pela análise automática de mensagens veiculadas em chats demanda o uso de ferramentas não encontradas no programa.

Considerando o exposto, o experimento apresentado neste trabalho pretende se constituir uma contribuição para o aprimoramento de programas de mineração de forma a superar a complexidade da linguagem usada nas comunicações baseadas em texto.

## Referências

- Aykol, Z.; Garrison, D.R. e Ozden, M.Y. (2009) “Online and blended communities of inquiry: exploring the developmental and perceptual differences”. In *International Review of Research in Open and Distance Learning*. v.10, n.6.
- Fan, W.; Wallace, L.; RICH, S.e ZHANG, Z. (2006) “Tapping the power of text mining”. In *Communications of ACM*, Nova York, v. 9, n. 49, p.76-82.



Feldman, R. e Sanger, J. (2007) *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge, MA: Cambridge University Press.

Garrison, D. R. (2007) "Online community of inquiry review: Social, cognitive, and teaching presence issues". In *Journal of Asynchronous Learning Networks*, v.11, n.1, p. 61-72.

\_\_\_\_\_. (1991) "Critical thinking and adult education: a conceptual model for developing critical thinking in adult learners". In *International Journal of Lifelong Education*, v.10, n.4.

Giordan, M.; Dotta, S. (2008) "Estudo das interações mediadas por um serviço de tutoria pela Internet". In *Linguagem & Ensino*, Pelotas, v.11, n.1, p.127-143.

Gunawardena, C. (1995) "Social Presence Theory and Implications for Interactive and Collaborative Learning in Computer Conferences". In *International Journal of Educational Telecommunications*, v. 1, n. 2/3, p. 147-166.

Kerbrat-Orecchioni, C. (2006) *Análise da conversação: princípios e métodos*. São Paulo: Parábola.

Koch, I. V. (2009) *Desvendando os segredos do texto*. 6.ed. São Paulo: Cortez.

Marcuschi, L.A.. (2004) "Gêneros textuais emergentes no contexto da tecnologia digital". In: \_\_\_\_\_; XAVIER, A. C. (orgs.). *Hipertexto e gêneros digitais: novas formas de construção de sentido*. Rio de Janeiro: Lucerna.

Palloff, R. e Pratt, K. (2002) *Construindo comunidades de aprendizagem no ciberespaço: estratégias eficientes para salas de aula on-line*. Porto Alegre: Artmed.

Rourke, L; Anderson, T.; Garrison, D.A. e Archer, Walter. (2001) "Assessing social presence in asynchronous text-based, computer conference". In *Journal of Distance Education*, v.14, n.2.

Short, J.; Williams, E.e Christie, B. (1976) *The social psychology of telecommunications*. London: John Wiley & Sons.

Sschuelter, W. e Reis, M.S. (2008) "O internetês em comunidades virtuais: a interação pela linguagem cifrada". In *Interletras*, v.6, n. 6-7.

Tu, C.H.e McIsaac, M. (2002) "The relationship of social presence and interaction in online classes". In *The American Journal of Distance Education*, v.16, n.3, p. 131-150.

Vieira, R.; Gonçalves, P.N.e Souza, J.G. (2005) "Processamento computacional de anáfora e correferência". In *Revista Estudos da Linguagem*. Belo Horizonte, n. 16, n.1, p. 263-84.

Walther, J.B. (1992) "Interpersonal effects in computer-mediated interaction: relational perspective". In *Communication Research*, v.19, n.1, p. 52-90.

Wives, L. K. (1999) *Um estudo sobre Agrupamento de Documentos Textuais em Processamento de Informações não Estruturadas Usando Técnicas de Clustering*. Dissertação de Mestrado. Instituto de Informática. PPGC-UFRGS. Porto Alegre.