

Linguística e Tecnologias Informatizadas: uma experiência do Grupo Interdisciplinar de Pesquisas em Linguística Informática

Zilda Maria Zapparoli¹, Neide Ferreira Gaspar², Edenis Gois Cavalcanti³

¹ Faculdade de Filosofia, Letras e Ciências Humanas – Universidade de São Paulo (USP) / Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

² Faculdade de Filosofia, Letras e Ciências Humanas – Universidade de São Paulo (USP)

³ Faculdade de Filosofia, Letras e Ciências Humanas – Universidade de São Paulo (USP)

zmq@usp.br, neidegaspar@gmail.com, edenis@usp.br

Abstract. *Certified by the University of São Paulo (USP) and officially registered in the directory of the National Counsel of Technological and Scientific Development (CNPq) since 2002, the Grupo Interdisciplinar de Pesquisas em Linguística Informática (Linguistic Informatics Interdisciplinary Research Group) congregates lecturers, professors, post-graduate and former graduate students who are interested in promoting an interchange between researchers in the fields of mathematical and language sciences. The Group studies are particularly concerned with computer technology applied to lexical and discourse analysis, with the creation of data bases comprising different types of language manifestations (written or oral, of technical, literary, legal, medical, journalistic or advertising character), in different languages and with various aims (to be used in teaching, research, diagnosis and expert verification). The members of the Group are also concerned with the analyses of bases generated with the aid of computer tools and descriptive-statistical methods.*

Key-words: *Linguistic Informatics; Linguistic Informatics Interdisciplinary Research Group; informatics technology in lexical, text and discourse analyses; quanti-qualitative treatment of lexical data*

Resumo. *Certificado pela Universidade de São Paulo e cadastrado no Diretório de Grupos de Pesquisa no Brasil do CNPq em 2002, o Grupo Interdisciplinar de Pesquisas em Linguística Informática reúne docentes, pós-graduandos e ex-alunos de pós-graduação da USP, com o objetivo de promover intercâmbio de experiências entre pesquisadores das ciências exatas e da ciência da linguagem. As investigações do Grupo dão especial atenção a aplicações de tecnologias informatizadas em análises lexicais, textuais e discursivas, através da geração de bases de dados de diferentes tipologias de textos (verbais orais ou escritos, de caráter técnico, literário, jurídico, jornalístico, publicitário, patológico), em diferentes línguas, para*

diferentes finalidades (ensino, pesquisa, perícia, diagnóstico), e da análise das bases geradas mediante a utilização de ferramentas informáticas e de métodos estatístico-descritivos.

Palavras-chave: Lingüística Informática; Grupo Interdisciplinar de Pesquisas em Lingüística Informática; tecnologias informatizadas em análises lexicais, textuais e discursivas; tratamento lexical quantiquantitativo.

1. Grupo Interdisciplinar de Pesquisas em Lingüística Informática

Certificado pela Universidade de São Paulo e cadastrado no *Diretório de Grupos de Pesquisa no Brasil do CNPq* em 2002, o *Grupo Interdisciplinar de Pesquisas em Lingüística Informática*, liderado pela Prof^a Dr^a Zilda Maria Zapparoli, tem o objetivo de promover intercâmbio de experiências entre pesquisadores da área de exatas e das humanidades em geral, em especial, entre as ciências exatas e a ciência da linguagem. Para isso, os seus integrantes desenvolvem e divulgam trabalhos que se situam na interação entre Lingüística e Informática, de forma a que o grupo atue como agente multiplicador de métodos avançados de análise e de tecnologias de ponta na pesquisa lingüística junto a estudiosos de diferentes áreas.

As investigações do grupo dão especial atenção ao exame de usos e aplicações de tecnologias informatizadas nos estudos da linguagem através da geração de bases de dados de diferentes tipologias de textos (verbais orais ou escritos, de caráter técnico, literário, jurídico, jornalístico, publicitário, patológico), em diferentes línguas, para diferentes finalidades (ensino, pesquisa, perícia, diagnóstico), e da análise das bases geradas mediante a utilização de ferramentas informáticas e de métodos estatístico-descritivos.

O *Grupo Interdisciplinar de Pesquisas em Lingüística Informática* substituiu o *Núcleo Interdisciplinar de Pesquisa em Novas Perspectivas de Análises do Discurso*, criado em 1997 e integrado por docentes e pós-graduandos da Universidade de São Paulo. Atualmente, reúne docentes, pós-graduandos e ex-alunos de pós-graduação da Universidade de São Paulo, dentre eles docentes da Universidade Estadual de São Paulo (campus de Bauru), Universidade Federal de Pernambuco, Pontifícia Universidade Católica de São Paulo e Universidade Paulista.

Este artigo prioriza trabalhos do Grupo que usam tecnologias informatizadas para investigações do comportamento do léxico, do texto e do discurso, tecendo considerações sobre um programa de computador para análise lingüística – *Stablex* –, sobre o método de análise de textos para o qual serve de ferramenta e sobre o uso que pesquisadores brasileiros vêm deles fazendo.

2. Lingüística e Tecnologias Informatizadas: o método quantiquantitativo de análise de textos

O presente artigo dá destaque ao emprego do programa computacional *Stablex* (STA – de statistique, TAB – de tableaux, LEX – de lexique e T...EX – de texte), desenvolvido por André Camlong e Thierry Beltran, Universidade de Toulouse II. O programa possibilita a geração de léxicos, indexação, extração de seqüências e concordâncias, lematização e tratamento estatístico. Construído inicialmente para a plataforma

Macintosh (1991), o programa *Stablex* conta, a partir de 2004, com a sua versão PC (São Paulo, Pirus Tecnologia, 2004), que inclui novas funções estatísticas para a análise de textos. Por contemplar uma confluência de áreas – Linguística, Matemática, Estatística, Computação –, o *Stablex* facilita e otimiza não somente a busca, organização e quantificação, mas também a análise de dados lingüísticos. Destaca-se o fato de o programa ter sido desenvolvido em função de um modelo de análise lexical, textual e discursiva – o *método matemático-estatístico-computacional de análise de textos* de André Camlong. Trata-se, por conseguinte, não apenas da aplicação de um programa computacional, mas, de forma mais ampla, de um programa que serve de ferramenta para um método de análise de textos.

Fundado na matemática e na estatística paramétrica, o método permite a análise quantiquantitativa do léxico, que indica apontamentos para a análise textual e discursiva. Da aplicação do método resultam: (a) levantamento lexical, com constituição de vocabulários de frequência e de tabelas de distribuição de frequências – cálculo aritmético – tratamento quantitativo; (b) constituição de vocabulários preferenciais, normais e diferenciais, a partir de tabelas de desvios reduzidos – cálculo algébrico – tratamento quantiquantitativo; (c) aplicação de testes estatísticos – normalidade de distribuição lexical, correlação, entre outros. Esses resultados dão a conhecer o léxico através de uma descrição lexical quantiquantitativa.¹

3. Aplicações do Grupo

Esta seção relaciona trabalhos do Grupo que vêm utilizando o modelo de análise de textos de Camlong para a descrição e análise da tessitura lexical, textual e discursiva de *corpora* compostos por diferentes tipologias de textos, os quais têm ressaltado as inúmeras vantagens de pesquisas lingüísticas baseadas em *corpora*, por sua vez explorados por programas de análise lingüística, que respondem, de forma satisfatória, às necessidades do pesquisador cujo objeto de trabalho é o texto, porque possibilitam não somente a busca, organização e quantificação, mas também a análise de dados lingüísticos: (a) Corpus de Discursos Orais - Português Falado de São Paulo (Zilda Maria Zapparoli); (b) Corpus de Discursos Literários - Fernando Pessoa (João Martins Ferreira); Machado de Assis (Daniela Fregonese Bragazza); Realismo Fantástico (Neide Ferreira Gaspar); Guimarães Rosa (Márcia Angélica dos Santos); (c) Corpus de Discursos Bíblicos - Epístolas de São Paulo (Edenis Gois Cavalcanti); (d) Corpus de Discursos Públicos – Greve da Educação em Pernambuco, 1987-1990 (Maria Cristina Hennes Sampaio); (e) Corpus de Discursos Escolares (Luís Rogério da Silva).

4. Alguns resultados de aplicações da abordagem de análise quantiquantitativa do léxico

A fim de dar mostras das aplicações da abordagem de análise quantiquantitativa do léxico, expõem-se, a seguir, alguns resultados de duas investigações.

¹ Sobre os procedimentos utilizados para a geração dos léxicos, vocabulários e tabelas, consultar Camlong (1996) e Zapparoli (2002).

4.1. *Corpus* de Discursos Literários - Realismo Fantástico (Neide Ferreira Gaspar)

As questões propostas eram: como mensurar a proximidade entre seis contos de duas autoras diferentes, Lygia Fagundes Telles e Marguerite Yourcenar? Que peso teriam componentes com carga simbólica elevada, como cores e elementos da natureza, na criação da atmosfera peculiar desses contos?

O uso do *Stablex* foi válido para confirmar, com dados mais objetivos, as premissas de elementos comuns nos contos de duas autoras. Gráficos e tabelas apontam um *corpus* homogêneo, centrado em torno da média reduzida a zero, o que não teria acontecido se houvesse grandes diferenças nos planos narrativo e discursivo.

Apenas alguns, entre os inúmeros recursos do programa *Stablex*, foram selecionados para esta pequena demonstração das possibilidades de utilização do programa em análise de texto.

Para analisar a Tabela de Desvios Reduzidos, da qual se destacam apenas as linhas do total e das médias, deve-se ter em mente que as distâncias são medidas em relação a um centro de gravidade, que é a média reduzida a zero. Quanto mais próximos do zero, mais equilibrados são os valores. No *corpus* estudado, observa-se a grande proximidade dos valores do khi², tanto na comparação das variáveis entre si, como nos valores de cada variável em relação ao zero.

A título de comparação, são apresentados os valores obtidos no Estudo das *Elocuções Formais*, feito por Zapparoli, que se mostram mais discrepantes. No caso, o *corpus* era constituído por transcrições de aulas, conferências e entrevistas, e o método aponta a maior heterogeneidade entre os textos. Já no estudo de Bragazza, a homogeneidade é maior, com valores mais próximos entre si e do zero. Note-se que, neste caso, são textos do mesmo gênero (contos), do mesmo autor (Machado de Assis). No estudo sobre contos fantásticos, os valores estão ainda mais próximos entre si, apesar de se tratar de duas autoras diferentes, o que aponta para a existência de aspectos em comum (variáveis 1 a 3, respectivamente: *A chave na porta*, *Emanuel* e *Natal na barca*, de Lygia Fagundes Telles; variáveis 4 a 6: *O homem que amou as Nereidas*, *Nossa Senhora das Andorinhas* e como *Wang-Fô se salvou*, de Marguerite Yourcenar).

Valores	T1	T2	T3	T4	T5	T6
Total	1,742	4,381	8,598	-14,841	-4,307	5,176
Média	0,025	0,063	0,123	-0,212	-0,062	0,074
Khi ²	0,001	0,004	0,015	0,045	0,004	0,005

Tabela 1- Desvios Reduzidos (Realismo Fantástico)

Valores	T1	T2	T3	T4	T5	T6	T7
Total	9,401	-17,830	-19,453	-30,692	9,434	52,509	15,433
Média ¹	0,091	-0,171	-0,187	-0,295	0,091	0,505	0,148
Khi ²	0,437	0,029	0,035	0,087	0,008	0,255	0,022

Tabela 2- Valores Apurados Por Zapparoli (Estudo das *Elocuções Formais*)

Valores	T1	T2	T3	T4	T5	T6	T7	T8
Total	5,365	-20,449	-7,068	7,954	10,466	6,725	4,243	-9,871
Média	0,054	-0,204	-0,071	0,080	0,105	0,067	0,042	-0,099
Khi ²	0,003	0,042	0,005	0,006	0,011	0,005	0,002	0,010

Tabela 3 - Valores Apurados Por Bragazza (Contos de Machado de Assis)

A questão sobre a importância de itens lexicais com valor simbólico foi respondida pela análise dos dados nas Tabelas de Valores Lexicais (TVL) de cada variável. Nelas, os itens lexicais são elencados em ordem decrescente de preferência de emprego no texto, e as análises foram direcionadas aos itens do léxico preferencial ($Z \geq 2,0$) e do vocabulário de base com tendência positiva ($1,0 \leq Z < 2,0$). Os elementos com carga simbólica elevada (cores e elementos da natureza, como *água*, *cavernas* e *árvores*) apareceram, em sua maioria, nessas faixas, em todos os contos, indicando emprego preferencial.

4.2. *Corpus* de Discursos Bíblicos (Edenis Gois Cavalcanti)

Este trabalho² apresenta o olhar matemático-estatístico-computacional de análise de textos sobre um *corpus* (do Novo Testamento) constituído de cinco epístolas – quatro do apóstolo Paulo e uma "sem" autoria –, com a intenção de investigar a paternidade da epístola aos Hebreus a partir dos dados fornecidos pela estatística paramétrica, especificamente, a partir do Teste de Correlação.

4.2.1 Objeto de estudo, justificativa e objetivos

Aplicamos o método de análise estatística destinado ao tratamento informático dos dados lexicais, textuais e discursivos, desenvolvido por André Camlong³, a um *corpus* constituído de cinco variáveis (textos), com a finalidade de verificar, através da análise objetiva, qual a probabilidade de a epístola canônica aos Hebreus (Novo Testamento) ser do apóstolo Paulo, pois ainda, após séculos de Cristianismo, não há uma conclusão definitiva sobre a sua autoria. O que há é apenas a *suposição* de que seja de Paulo. Seria, pois, muito esclarecedor estabelecer a paternidade, não através de critérios subjetivos, mas científicos, objetivos, por meio de um relacionamento intertextual.

Não obstante existam outros supostos autores, optamos em confrontar textos de Paulo com a epístola aos Hebreus⁴, pois quase a metade dos textos do Novo Testamento são de autoria paulina. Ou seja, de um total de 27, 13 são do apóstolo. Além do aspecto da produção textual de Paulo, há o da sua importância para as Igrejas Cristãs Primitivas. É, portanto, por sua produção literária no contexto do Novo Testamento e por sua importância na fundação das Igrejas Primitivas que se justifica o desejo de descobrir a autoria – ou a *probabilidade* da autoria – de Hebreus *a partir das epístolas de Paulo*.

² Este estudo pertence à Parte I – *O Olhar Estatístico* – da dissertação de mestrado *A Estatística e a Semiótica: imbricação de olhares sobre textos neotestamentários* de Edenis Gois Cavalcanti, defendida em dezembro de 2005 (USP).

³ Sobre os princípios da abordagem teórico-metodológica, conferir Cavalcanti (2005) e Camlong (1996).

⁴ Segundo Pearlman (1977, p. 310), "não há outro livro do Novo Testamento cuja autoria seja mais disputada, nem cuja inspiração seja mais incontestável. O próprio livro é anônimo. Por causa da diferença de estilo, comparado com os outros escritos de Paulo, muitos eruditos ortodoxos negam que foi ele quem o escreveu. Tertuliano, no terceiro século, declarou que Barnabé foi o autor. Lutero sugeriu que fosse Apolo". Ou, nas palavras de Tenney (1972, p. 378), "a Igreja Oriental desde os primeiros tempos que considera esta epístola como produto de Paulo, se bem que talvez indireto. Eusébio declarou que Clemente de Alexandria afirmava que Paulo a escrevera em hebraico e que Lucas a traduzira para o grego [...]. Orígenes freqüentemente a citava como havendo sido escrita por Paulo [...]"

4.2.2 Corpora de estudo

Os textos objetos de análise são cinco: quatro de autoria do apóstolo Paulo e um de autoria desconhecida, que fazem parte do conjunto de textos canônicos – sagrados – do Novo Testamento: (a) 1ª epístola aos Coríntios – T1; (b) Gálatas – T2; (c) Romanos – T3; (d) Efésios – T4; (e) Hebreus (autoria desconhecida) – T5.

Como os textos em questão têm várias versões, em várias línguas, decidimos selecionar dois *corpora*: um *corpus* em português e um em grego.

4.2.3. Descrição e análise dos resultados - teste de correlação simples e múltipla

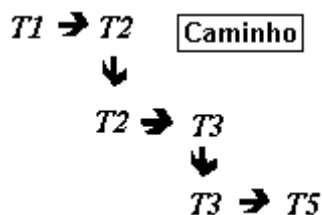
Do grau de dependência de uma variável em função de uma outra variável, extraído a partir do coeficiente de correlação, podemos estabelecer a probabilidade da *autoria-paternidade de um texto, a probabilidade de o texto analisado ser deste ou daquele autor*.

Como dissemos na seção 4.2.1, o objetivo é avaliar a probabilidade de T5 (epístola aos Hebreus) ser ou não do apóstolo Paulo. A autoria de T1, T2, T3 e T4 é indiscutível, pois são epístolas assinadas por Paulo e aceitas como sendo de Paulo por toda a Igreja Cristã, tanto ortodoxos, como não-ortodoxos, tanto católicos como protestantes. Portanto, o *alvo* aqui é a variável T5.

A ligação de uma variável só tem sentido a partir do momento em que o coeficiente de correlação atinge ou ultrapassa o valor de 0,866 ($r \geq 0,866$). Nas análises efetuadas, constata-se que todos os coeficientes estão muito acima de 0,866. Isso demonstra uma forte ligação entre as variáveis.

Observações:

- (a) a correlação de T5 com T1, T2, T3 é forte, tornando-se um pouco mais fraca com T4, mas com valores ainda altos em relação à tabela de correlação, que exige um coeficiente acima de 0,866;
- (b) na tradução em pauta (de língua sintética para língua analítica), os coeficientes "sintéticos" são maiores que os "analíticos", demonstrando maior ligação nos textos originais – nas duas correlações;
- (c) *em grego, observando-se T1/T2, T2/T3, T3/T5, temos um coeficiente de correlação acima de 98% na simples e acima de 96% na múltipla, provando que T5, a partir de T2 e T3 (epístolas de Paulo), é uma epístola de autoria muito provável de Paulo. O caminho da ligação – T1 → T2 → T3 → T5:*



- (d) a tradução enfraquece a ligação intertextual. Isso pode mostrar uma tendência natural nas traduções.

(e) os coeficientes de correlação na tradução não são maiores que os coeficientes de correlação nos originais – há perda de ligação na tradução.

Vejamos, graficamente, como fica a ligação entre as variáveis na projeção ortogonal da correlação simples (CS) e da correlação múltipla (CM).

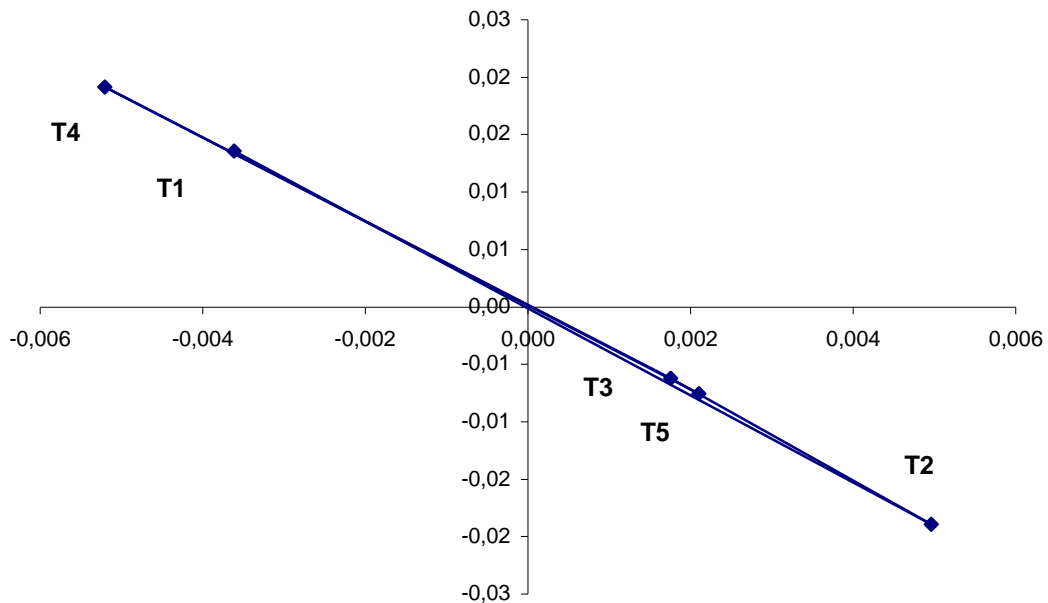


Figura 1 - Projeção Ortogonal CS Português

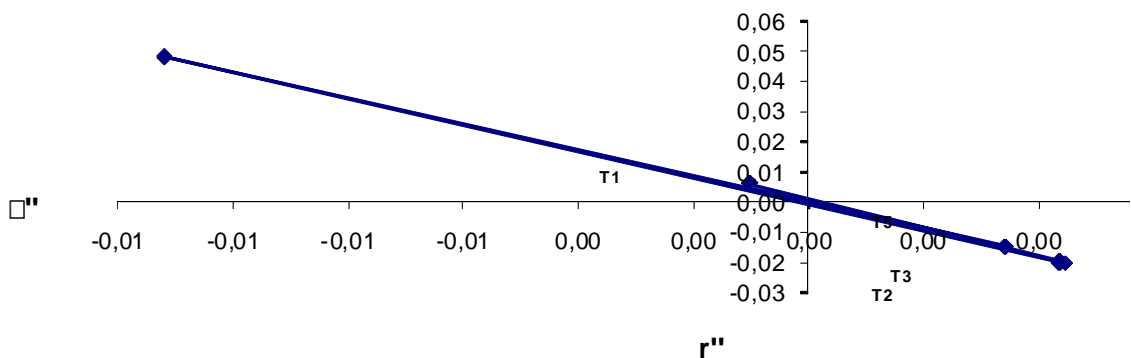


Figura 2 - Projeção Ortogonal CS Grego

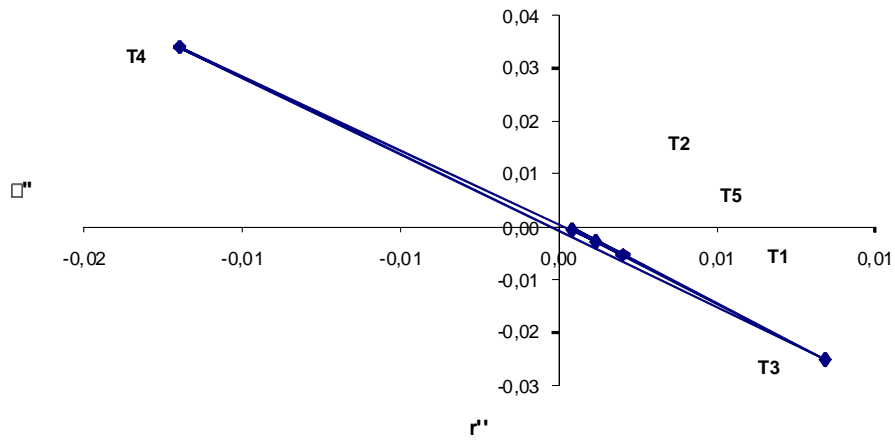


Figura 3 - Projeção Ortogonal CM Português

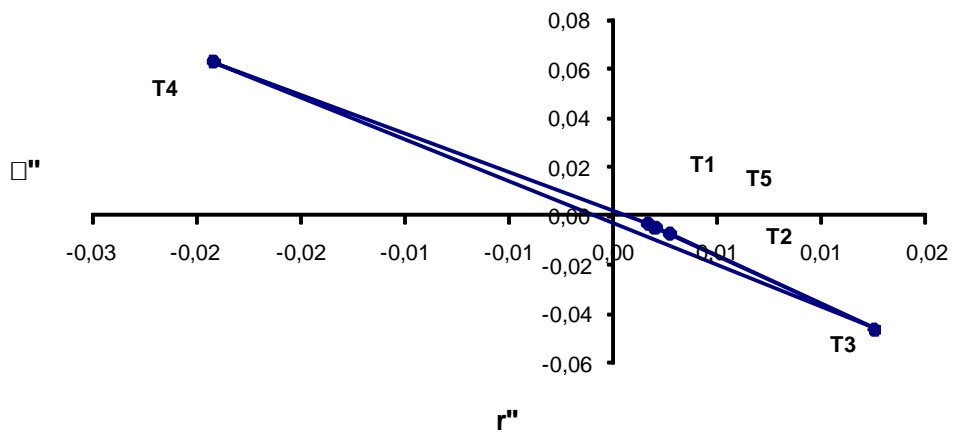


Figura 4 - Projeção Ortogonal CM Grego

Observamos os seguintes agrupamentos na projeção ortogonal:

- a) CS Português: T3, T5 e, um pouco distante, T2;
- b) CS Grego: T5, T3 e T2;

- c) CM Português: T2, T5 e T1;
- d) CM Grego: T1, T5 e T2.

Como se vê, os gráficos confirmam o caminho da ligação das variáveis com T5.

5 – Considerações Finais

A investigação realizada é mostra da possibilidade do diálogo entre áreas do conhecimento aparentemente opostas: de um lado, as Ciências Exatas e, de outro, as Ciências Humanas. A síntese dessa oposição sempre será rica, porque sempre será o resultado de um trabalho interdisciplinar. Acreditamos que a ferramenta informática já é parte constitutiva das pesquisas que envolvem o tratamento de *corpora* lingüísticos, e a matemática, pelo viés da estatística paramétrica, a fornecedora dos elementos analíticos necessários à compreensão do discurso.

O método matemático-estatístico-computacional de André Camlong apresentou-se adequado para a análise dos *corpora* em estudo, fornecendo elementos objetivos para o *desvelamento* do discurso.

No estudo dos contos do *Realismo Fantástico*, a proximidade entre os textos e o emprego preferencial dos itens lexicais com carga simbólica foram atestados pela aplicação do cálculo algébrico do desvio reduzido.

A partir da análise dos coeficientes de correlação simples e múltipla (ligação intertextual), obtiveram-se os elementos objetivos para apontar, do ponto de vista da probabilidade, *a paternidade da Epístola aos Hebreus*. Com mais de 96% de probabilidade no *corpus* português e mais de 98% de probabilidade no *corpus* grego, a epístola aos Hebreus é, provavelmente, de autoria paulina. Veja-se bem: *é muito provável que seja do apóstolo Paulo*.

Referências

BRAGAZZA, Daniela Fregonese. *SUBLIMES PORMENORES: escolhas lexicais privilegiadas na composição das personagens e das temáticas em contos de Machado de Assis*. São Paulo, 2005. Tese (Doutorado em Letras - Lingüística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. /Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/.

CAMLONG, André. Lexicométrie. In: *Stablex version PC*. São Paulo: Pirus Tecnologia, 2004. CD-ROM.

_____. *Méthode d'analyse lexicale textuelle et discursive*. Paris: C.R.I.C. & OPHRYS, 1996.

_____. *Stablex pratique*. Toulouse: Teknea, 1991.

_____; BELTRAN, Thierry. *A propósito do Stablex – versão PC*. In: *Stablex version PC*. São Paulo: Pirus Tecnologia, 2004.

_____; BELTRAN, Thierry. *Stable version PC*. São Paulo: Pirus Tecnologia, 2004.

CAVALCANTI, Edenis Gois. *A Estatística e a Semiótica: imbricação de olhares sobre textos neotestamentários*. São Paulo, 2005. Dissertação (Mestrado em Letras -

Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. /Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/.

GASPAR, Neide Ferreira. *Mágico, Fantástico, Maravilhoso: em busca dos elementos constituintes de um realismo onírico*. São Paulo, 2005. Dissertação (Mestrado em Letras - Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. /Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/.

GRUPO INTERDISCIPLINAR DE PESQUISAS EM LINGÜÍSTICA INFORMÁTICA. Traz apresentação, membros, pesquisas, eventos e cursos relacionados ao Grupo. São Paulo, 2002. Disponível em <<http://www.fflch.usp.br/dl/li>>.

MARTINS FERREIRA, João. *O Discurso de Fernando Pessoa em Mensagem*. São Paulo, 2000. Tese (Doutorado em Estudos Comparados de Literaturas de Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. Disponível em <<http://www.artemusical.com/martinsferreira>>.

SAMPAIO, Maria Cristina Hennes. *Democracia, cidadania e linguagem em tempos de globalização*. Recife: Programa de Pós-Graduação em Letras da UFPE, 2005.

_____. *Democracia, cidadania e produção de um espaço público democrático em tempos de globalização: práticas discursivas entre Estado-Sociedade no movimento grevista da educação em Pernambuco (1987-1990)*. São Paulo, 2002. Tese (Doutorado em Letras - Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. /Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/.

SANTOS, Márcia Angélica dos. *Quando as veredas se encontram: diálogos entre a Estatística e a Semiótica em uma abordagem de seis contos de Guimarães Rosa*. São Paulo, 2008. Tese (Doutorado em Letras - Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. / Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/.

SILVA, Luís Rogério da. *Operadores na Constituição Textual*. São Paulo, 2006. Dissertação (Mestrado em Letras - Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. /Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/.

ZAPPAROLI, Zilda Maria. *Análise do comportamento fonológico da junção intervocabular no português do Brasil (variante paulista). Uma pesquisa linguística com tratamento computacional*. São Paulo, 1980. Tese (Doutorado em Letras - Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. / Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/.

_____. Um pouco da história da análise informatizada do léxico no Brasil. In: NUNES, José Horta; PETTER, Margarida (Orgs.). *História do Saber Lexical e Constituição de um Léxico Brasileiro*. São Paulo / Campinas: Humanitas / Pontes, 2002, p. 223-253.

_____; CAMLONG, André. *Do Léxico ao Discurso pela Informática*. São Paulo: EDUSP/FAPESP, 2002, 256 p. + CD-ROM.