



FACULDADE VISCONDE DE CAIRU

CEPPEV – Centro de Pesquisa e Pós-graduação da Faculdade Visconde de Cairu

SILVIA MARIA GOMES CALDEIRA

**CARACTERIZAÇÃO DA REDE DE SIGNOS LINGÜÍSTICOS:
UM MODELO BASEADO NO APARELHO PSÍQUICO DE FREUD**

SALVADOR – BAHIA

NOVEMBRO / 2005

SILVIA MARIA GOMES CALDEIRA

**CARACTERIZAÇÃO DA REDE DE SIGNOS LINGÜÍSTICOS:
UM MODELO BASEADO NO APARELHO PSÍQUICO DE FREUD**

SALVADOR - BAHIA

NOVEMBRO/2005

Copyright © 2005 Silvia Maria Gomes Caldeira

É permitida a cópia, distribuição e/ou modificação deste documento sob os termos da Licença de Documentação Livre GNU, Versão 1.2 ou qualquer versão posterior publicada pela Free Software Foundation; sem Seção Invariante, textos da capa e contra-capa. Uma cópia da licença está incluída na seção intitulada “GNU Free Documentation License”.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation,; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

- C146 Caldeira, Silvia Maria Gomes
Caracterização da rede de signos lingüísticos: [manuscrito]
um modelo baseado no aparelho psíquico de Freud/ Silvia Maria
Gomes Caldeira. _ Salvador, 2005.
130 f. : il. ; 21cm x 29,7cm.
- Dissertação (Mestrado Interdisciplinar em Modelagem Computacional) –
Centro de Pós-Graduação e Pesquisa Visconde de Cairu, Fundação Visconde de
Cairu.
- “Orientação: Prof. Dr. José Garcia Vivas Miranda”.
“Co-orientação: Prof. Dr. Hernane Borges de Barros Pereira”.
1. Linguagem. 2. Sistemas Complexos - Linguagem. 3. Redes.
4. Aparelho Psíquico de Freud. I. Miranda, José Garcia Vivas,
orient. II. Pereira, Hernane Borges de Barros, co-orient. III.

Ficha catalográfica elaborada por:
Arthur Manoel Santana de Barros
CRB/ 5-1043

SILVIA MARIA GOMES CALDEIRA

**CARACTERIZAÇÃO DA REDE DE SIGNOS LINGÜÍSTICOS:
UM MODELO BASEADO NO APARELHO PSÍQUICO DE FREUD**

Dissertação apresentada à Coordenação de Mestrado Interdisciplinar em Modelagem Computacional da Fundação Visconde de Cairu para a obtenção do título de Mestre em Modelagem Computacional.

Orientador:
Prof. Dr. José Garcia Vivas Miranda.

Co-orientador:
Prof. Dr. Hernane Borges de Barros Pereira.

Mestrado Interdisciplinar em Modelagem Computacional
Centro de Pós-graduação e Pesquisa Visconde de Cairu
Fundação Visconde de Cairu

Salvador – BA
Novembro/2005

(FOLHA COM ASSINATURAS)

AGRADECIMENTOS

Uma difícil tarefa essa de agradecer. Persiste sempre a sensação de que vamos deixar de lembrar alguém. Vamos ser injustos com quem nos foram sempre caros.

Minha família é a grande responsável por eu ter feito esse trabalho com tanta dedicação. Mesmo incompleta. Mesmo impaciente, e às vezes inconformada com a minha ausência. Vocês são a razão por eu ter feito esse trabalho tão importante para mim. Principalmente o meu filho Ângelo, que tem sido o meu cúmplice nessa busca pelo saber.

Meus amigos, agradeço aos que ficaram e resistiram. Esperando pacientemente que eu acabasse os inúmeros trabalhos, reuniões e leituras. Um agradecimento muito especial aos que me apresentaram com textos, programas e idéias.

Meu orientador que transformou a minha maneira de ver o mundo, entrelaçando conceitos, criando uma verdadeira teia em torno do que poucos poderiam unir: física e Freud. Você é a pessoa mais transdisciplinar que eu já conheci.

Ao meu co-orientador pela grande disponibilidade em tornar possível o meu tumultuado caminho entre as disciplinas.

Um agradecimento especial às psicanalistas Sônia Vicente e Denise Coutinho pela paciência, dedicação e desprendimento em contribuir, com seus conhecimentos, na difícil tarefa de desvendar os escritos de Freud.

Silvia Maria Gomes Caldeira

RESUMO

Freud conjecturou que as nossas percepções de mundo produzem marcas (traços mnêmicos) em nosso cérebro, que se unem formando as representações-objeto, desprovidas de significado consciente, até que passamos a usar a linguagem. Assim, os traços mnêmicos fazem sentido para o indivíduo, quando se juntam em torno de signos que as representem: as palavras. Estas, então, seriam representações de segunda ordem das nossas imagens mnêmicas. Na fala ou na escrita, a escolha das palavras ocorre de acordo com as facilitações existentes entre as diversas representações-objeto. Assim, o fator individual na escolha das palavras ao se escrever um texto resulta, em última análise, das associações existentes nesta rede de representações-objeto.

A hipótese fundamental deste trabalho está em que a estrutura de conexão entre as palavras de um texto escrito representa, de forma simplificada, a topologia da rede de conexões das representações-objeto. Utilizamos, então, a teoria das redes complexas para analisar a dinâmica de escrita de textos individuais, em busca de propriedades emergentes. A rede é construída considerando cada sentença como uma unidade conceitual, e novas sentenças, com palavras em comum com as antigas, são conectadas mediante a palavra compartilhada, formando uma rede complexa.

Utilizando esta metodologia geramos seqüências de redes com quantidades progressivas de sentenças seguindo a seqüência de escrita do texto. A caracterização do crescimento destas redes representa a dinâmica de inclusão de novos conceitos, de forma a construir a semântica geral do escrito. Os valores calculados dos índices das redes, de todos os textos, indicam que estas apresentam características de rede de mundo pequeno (*small world*) e de redes livres de escala (*scale free*). Tais topologias nos levam a crer que a rede de representações-objeto está formada por grupos de associações que funcionam como concentradores, que facilitam as ligações entre elas, o que as tornam, energeticamente, mais econômicas.

ABSTRACT

Freud conjectured that our acuity of world make traces (mnemics marks) in our brain, that when putted together, corresponds to the object-representations, without conscientious mean, until we utilize it in language. According to him, the mnemics marks begin to make individual sense, when they are linked together around signs that represents it: the words. Therefore, the words could be described as second order representations of ours mnemics images. In talking or writing, the choice of words occurs according to pre-existing neural ways between some object-representations. So, the individual factor in choosing words in written texts results, in some way, of the pre-existent links in this complicated network of object-representations.

The fundamental hypothesis of this work lies in the idea that connection structure of the words in a written text represents, in a simple form, the topology of the object-representation network in our mind. Then, we utilized the theory of complex networks to analyze the dynamic of individual written texts, with the aim of founding emergent properties. The network is constructed considering each sentences as the conceptual unit, and new sentences, with words already utilized, are linked with to old sentence using the common word, forming a complex network.

Utilizing this model we generate sequences of networks with progressive quantities of sentences following the natural sequence of written the text. The characterization of this network grow denote the dynamic of including new concepts in text, in order to construct the general semantic of the text. The calculated indices values of the network, for all texts, indicate that they have characteristics of both small world and scale-free networks. This topologies leads us to conclude that the object-representation network are formed by groups of associations that works as concentrators (hubs), that facilitates the links among then, witch make it energetically more economic.

SUMÁRIO

APRESENTAÇÃO	1
CAPÍTULO I – A LINGUAGEM	6
1.1 PENSAMENTO LINGÜÍSTICO	6
1.2 SIGNO, SIGNIFICADO, SIGNIFICANTE, SIGNIFICAÇÃO E PALAVRA	9
1.3 PALAVRAS GRAMATICAIS E LEXICAIS	10
1.4 COMPLEXIDADE DA LINGUAGEM	13
1.5 FUNÇÕES DO CÉREBRO ENVOLVIDAS NA LINGUAGEM	13
1.6 TEORIA LOCALIZACIONISTA	15
1.7 TEORIA FREUDIANA DA LINGUAGEM.....	17
CAPÍTULO II – O APARELHO PSÍQUICO DE FREUD	21
2.1 A MEMÓRIA ψ	24
2.2 PENSAMENTO E LINGUAGEM.....	26
2.3 EXPERIMENTO MENTAL DE CONSTRUÇÃO DE UM TEXTO	28
2.4 CONHECIMENTO ATUAL SOBRE O FUNCIONAMENTO DA LINGUAGEM.....	30
CAPÍTULO III – SISTEMAS COMPLEXOS E LINGUAGEM	32
3.1 REDES COMPLEXAS	36
3.2 CÁLCULO DOS PARÂMETROS DE REDES COMPLEXAS	39
3.3 REDES ALEATÓRIAS	44
3.4 REDES DE MUNDO PEQUENO (<i>SMALL WORLD</i>)	46
3.5 REDES LIVRES DE ESCALA (<i>SCALE FREE</i>).....	49
3.6 LINGUAGEM E SISTEMAS COMPLEXOS.....	51
3.7 ESTUDOS DA LINGUAGEM SOB A ÓTICA DE SISTEMAS COMPLEXOS	54
CAPÍTULO IV – DESENVOLVIMENTO DA PESQUISA	59
4.1 QUESTIONAMENTOS A SEREM RESPONDIDOS	60
4.2 MÉTODO DA PESQUISA	61
4.3 AMOSTRAGEM	63
4.4 IDENTIFICAÇÃO DOS TEXTOS	64
4.5 LIMPEZA DOS TEXTOS	65
4.6 PRÉ-TRATAMENTO DOS TEXTOS	65
4.7 CONSTRUÇÃO DAS REDES.....	75
4.8 VISUALIZAÇÃO DAS REDES	77
4.9 MÉTODOS DE ANÁLISE DAS REDES.....	78
CAPÍTULO V – RESULTADOS	86
5.1 ANÁLISE DOS TEXTOS COMPLETOS	86
5.2 COMPARAÇÃO ENTRE AS TÉCNICAS DE TRATAMENTO DOS TEXTOS.....	86
5.3 RESULTADOS COMPARATIVOS EM RELAÇÃO À AMOSTRAGEM	91
5.4 DISTRIBUIÇÃO DO TAMANHO DAS SENTENÇAS	93
5.5 INFLUÊNCIA DO TAMANHO DAS SENTENÇAS NO COEFICIENTE DE AGLOMERAÇÃO	95
5.6 ANÁLISE DO CRESCIMENTO DOS TEXTOS	96
5.7 RESULTADOS DO CRESCIMENTO DO T-VÉRTICE.....	96
5.8 COMPARAÇÃO DOS RESULTADOS DE TODOS OS TEXTOS COM O CRESCIMENTO DOS MAIORES TEXTOS ANALISADOS.....	98
5.9 CRESCIMENTO INDIVIDUAL DOS TEXTOS.....	99
5.10 RESULTADOS DO CRESCIMENTO DO DIÂMETRO (DAM).....	101
5.11 RESULTADOS DO CRESCIMENTO DO CAMINHO MÍNIMO MÉDIO (CMM)	102
5.12 SIMULAÇÃO DA REDE DE PALAVRAS	103
5.13 ANÁLISE DO CAMINHO MÍNIMO MÉDIO	104
5.14 ANÁLISE DO COEFICIENTE DE AGLOMERAÇÃO MÉDIO	106
5.15 ANÁLISE DA DISTRIBUIÇÃO DE GRAUS.....	107
5.16 COMPARAÇÃO DOS RESULTADOS OBTIDOS COM OUTRAS PESQUISAS	108
CONCLUSÕES	109
REFERÊNCIAS BIBLIOGRÁFICAS	113
APÊNDICE A	117

LISTA DE FIGURAS

<i>Figura 1 – Esquema do circuito da fala (Fonte: SAUSSURE, 2004, p.19).....</i>	<i>7</i>
<i>Figura 2 - Córtex cerebral e suas subdivisões (Fonte: KANDEL et al., 1997, p. 70).....</i>	<i>14</i>
<i>Figura 3 - Áreas de Broca e Wernicke (Fonte: KANDEL et al, 1997, p. 511).....</i>	<i>16</i>
<i>Figura 4 - Ligações entre as representações-palavra e as representações-objeto (Fonte: FREUD, 1977, p. 71).....</i>	<i>20</i>
<i>Figura 5 - Proposta de um esquema ampliado do aparelho de linguagem de Freud.</i>	<i>29</i>
<i>Figura 6 - Imagens PET das regiões cerebrais de maior atividade participantes nos processos de uso da linguagem (Fonte: Carter, 2003, encarte)</i>	<i>31</i>
<i>Figura 7 – Esquema simplificado de uma rede.....</i>	<i>36</i>
<i>Figura 8 – Desenho esquemático das sete pontes da cidade de Königsberg (Fonte: Barabási, 2003).....</i>	<i>38</i>
<i>Figura 9 – Esquema Simplificado de grau de um nó ou vértice.....</i>	<i>40</i>
<i>Figura 10 – Esquema Simplificado de Distribuição de Grau de uma rede.....</i>	<i>40</i>
<i>Figura 11 – Representação dos vizinhos π_1 e π_2 do vértice v.....</i>	<i>41</i>
<i>Figura 12 – Exemplo de cálculo do coeficiente de aglomeração de um vértice ou nó</i>	<i>42</i>
<i>Figura 13 – Esquema simplificado mostrando o caminho entre dois nós.....</i>	<i>43</i>
<i>Figura 14 – Diagrama ilustrativo do modelo de Watts de embaralhamento de redes regulares e o surgimento de redes SW e aleatórias (Fonte: Watts, 1999, p. 68).</i>	<i>49</i>
<i>Figura 15 – A distribuição rank-freqüência de palavras. (A) Análise do livro de James Joyce; (B) Análise de jornais americanos feita por Eldridge; (C) Curva ideal com decaimento negativo (Fonte: Zipf, 1972, p. 25).</i>	<i>53</i>
<i>Figura 16 - Exemplo de uma construção de rede de palavras</i>	<i>62</i>
<i>Figura 17 – Composição da Amostra de Textos.....</i>	<i>63</i>
<i>Figura 18 – Panorama geral do programa de pesquisa lingüístico do LADL.....</i>	<i>65</i>
<i>Figura 19 – Diagrama do pré-tratamento dos textos e Linhas do código do arquivo BAT usado para chamar os programas respectivos.</i>	<i>66</i>
<i>Figura 20 – Exemplo de segmentação do texto em unidades lexicais.....</i>	<i>68</i>
<i>Figura 21 – Exemplo de etiquetagem em uma sentença do texto.....</i>	<i>69</i>
<i>Figura 22 – Exemplo de todas as possibilidades de etiquetas que as palavras do texto analisado anteriormente podem receber, após aplicação do dicionário.</i>	<i>70</i>
<i>Figura 23 – Algoritmo de eliminação das ambigüidades.....</i>	<i>73</i>
<i>Figura 24 – Processo de tratamento da unidade lexical para o formato requerido para o próximo programa de análise, o programa NetIPal.....</i>	<i>75</i>
<i>Figura 25 – Escolha da forma flexionada (a) ou da forma canônica (b) das palavras</i>	<i>75</i>
<i>Figura 26 - Diagrama da construção das redes e linhas de código do arquivo de lote correspondente.</i>	<i>76</i>
<i>Figura 27 - Arquivo texto com a indicação dos pares das palavras</i>	<i>77</i>
<i>Figura 28 - Rede construída pelo programa PAJEK</i>	<i>77</i>
<i>Figura 29 – Exemplo da influência do tamanho e composição da sentença no valor do CAM.</i>	<i>80</i>
<i>Figura 30 - Exemplo da influência do tamanho e composição da sentença no valor do CAM.....</i>	<i>81</i>

<i>Figura 31 - Exemplo da influência do tamanho e composição da sentença no valor do CAM</i>	81
<i>Figura 32 - Exemplo da influência do tamanho e composição da sentença no valor do CAM</i>	82
<i>Figura 33 – Comparação do comportamento da Distribuição de Graus do Livro dos Espíritos</i>	83
<i>Figura 34 – Exemplo de contagem de ligações (a) e frequência de ligações de uma pequena rede (b). Somente para ilustração, pois não é possível o ajuste de curvas com poucos valores</i>	84
<i>Figura 35 – Exemplo do ajuste Linear da Distribuição de Graus do livro MADLiv006m</i>	84
<i>Figura 36 - Exemplo de simulação do crescimento de um texto</i>	85
<i>Figura 37 – Comportamento do CAM das redes dos textos submetidos aos tratamentos 0, 1 e 2</i>	87
<i>Figura 38 – Exemplo do comportamento dos parâmetros nos Tratamentos aplicados aos textos</i>	88
<i>Figura 39 - Comportamento do CMM das redes dos textos submetidos aos tratamentos 0, 1 e 2</i>	89
<i>Figura 40 – Comparação do Grau Médio das redes dos textos tratados nas condições 0, 1 e 2</i>	90
<i>Figura 41 – Distribuição de frequência do tamanho das sentenças dos textos analisados</i>	93
<i>Figura 42 – Distribuição de frequência do valor mais provável de quantidade de palavras nas sentenças</i>	94
<i>Figura 43 – Simulação de possíveis composições e tamanhos de sentenças para alcançar o valor de CAM encontrado nos textos originais analisados</i>	95
<i>Figura 44 – Gráficos do crescimento de textos indicando o comportamento do índice T-vértice</i>	96
<i>Figura 45 – Resultados do CAM para todos os livros analisados, comparando com o crescimento de dois livros: Código de Da Vinci de Dan Brown e Anna Karenina de León Tolstói</i>	98
<i>Figura 46 - Crescimento do CAM em textos em inglês, divididos pelo tamanho</i>	99
<i>Figura 47 - Crescimento do CAM em textos em português, divididos pelo tamanho</i>	99
<i>Figura 48 – Comparação dos textos de dois autores</i>	100
<i>Figura 49 – Gráfico da evolução do diâmetro no crescimento do livro Ulysses, de James Joyce, escrito em português</i>	101
<i>Figura 50 – Gráfico da evolução do Caminho Mínimo Médio na simulação de crescimento do livro Ulysses, de James Joyce, escrito em português</i>	102
<i>Figura 51 – Distribuição de Frequência dos valores de CMM dos textos analisados nas condições de Tratamento 2 e Aleatório Condição 1, 2, 3 e 4</i>	105
<i>Figura 52 – Distribuição da Frequência dos valores de CAM dos textos analisados nas condições de Tratamento 2 e Aleatório Condição 1, 2, 3 e 4</i>	106
<i>Figura 53 – Distribuição de Graus de um mesmo texto nas condições de Tratamento 2 e Aleatório Condição 2, 3 e 4</i>	107

LISTA DE TABELAS

<i>Tabela 1 – Exemplos de Estudos de Redes complexas (Fonte: Artigo de Albert e Barabási, 2001).....</i>	<i>37</i>
<i>Tabela 2 - Recursos utilizados na pesquisa.....</i>	<i>61</i>
<i>Tabela 3 – Significado da primeira letra do nome do arquivo: gênero do autor.....</i>	<i>64</i>
<i>Tabela 4 – Significado da segunda letra do nome do arquivo: tipo geral do texto.....</i>	<i>64</i>
<i>Tabela 5 – Significado da terceira letra do nome do arquivo: tipo específico do texto.....</i>	<i>64</i>
<i>Tabela 6 - Códigos gramaticais usuais.....</i>	<i>71</i>
<i>Tabela 7 – Parâmetro fornecido ao programa ‘AMBISIN’ para determinar o tipo de análise.....</i>	<i>72</i>
<i>Tabela 8 - Símbolos de final de sentença considerados em textos da língua portuguesa.....</i>	<i>79</i>
<i>Tabela 9 – Resumo dos valores encontrados para os tratamentos do texto.....</i>	<i>91</i>
<i>Tabela 10 – Resultados da análise dos textos segmentados nas classes definidas na amostragem.....</i>	<i>92</i>
<i>Tabela 11 – Análise da distribuição de frequência do tamanho das sentenças por idioma.....</i>	<i>94</i>
<i>Tabela 12 – Resumo dos Índices do Embaralhamento / Simulação.....</i>	<i>107</i>
<i>Tabela 13 – Comparação de resultados da análise com pesquisas anteriores.....</i>	<i>108</i>

APRESENTAÇÃO

"E a agulha do real nas mãos da fantasia
Fosse bordando ponto a ponto nosso dia-a-dia"
[...]
Reproduzidos no bordado
A casa, a estrada, a correnteza
O sol, a ave, a árvore, o ninho da beleza"
Gilberto Gil

A linguagem humana, desde que se tornou objeto de estudo da ciência, apresentou-se com um nível de complexidade que poderia ser tratado pela filosofia, psicanálise, lingüística, e inclusive por ramos da ciência que a princípio não a teriam como objeto de estudo principal, como a matemática e a física.

Este fenômeno humano possui características que o leva a ser tratado, pelos cientistas, por uma perspectiva interdisciplinar. Apenas uma visão não seria suficiente para elucidar o que representa a construção da linguagem, seja falada ou escrita. Cada um desses ramos da ciência vê as palavras a partir de um ângulo próprio, e com interesses bem definidos. Faremos uma breve explanação de como a unidade lingüística 'palavra' pode ganhar significados diferentes, a depender de que ramo científico realiza a pesquisa.

Do ponto de vista da filosofia, a linguagem e o seu conjunto de símbolos, as palavras, adquirem a sua importância na capacidade de descrição do mundo, como descreveu o Prof. Waldomiro Filho em artigo publicado no site da Facom (UFBA):

A linguagem não é um aparato, mas um modo de se organizar o mundo e enformar a experiência, de modo que as regras e as palavras que constituem uma linguagem são a maneira de um processo contínuo do espírito humano expressar sua relação com o mundo e tecer os objetos e a mobília deste mundo (SILVA FILHO, 1997).

Segundo Margarida Petter, "o fascínio que a linguagem sempre exerceu sobre o homem vem desse poder que permite não só nomear/criar/transformar o universo real, mas também possibilita trocar experiências, falar sobre o que existiu, poderá vir a

existir, e até mesmo imaginar o que não precisa nem pode existir” (Petter apud FIORIN, 2004, p. 11). Essas considerações deixam evidente a potencialidade da linguagem como construtor de mundos.

Porém, a linguagem não tem como finalidade somente descrever/criar racionalmente o mundo real, mas também a construção do mundo ideal. No âmbito extra-científico, a linguagem é considerada como capaz de explicar e até mesmo criar realidades ideais, constatada na poesia e nas religiões. Do ponto de vista das religiões, podemos lembrar e citar a idéia difundida amplamente pelas religiões ocidentais, que acreditam na palavra proferida por Deus, como precursora para que fatos ocorram ou para que objetos surjam:

No princípio, Deus criou o céu e a terra. A terra, porém, estava informe e vazia, e as trevas cobriam a face do abismo, e o Espírito de Deus movia-se sobre as águas. E Deus disse: Exista a luz. E a luz existiu. E Deus viu que a luz era boa; e separou a luz das trevas. E chamou à luz dia, e às trevas noite. E fez-se tarde e manhã, (e foi) o primeiro dia.

Gênesis, I, 1-5

Já a psicanálise tem interesse específico pela análise do discurso falado onde o contexto em que a palavra aparece é de fundamental importância para definir sua significação. A ocorrência de metáforas e metonímias é o ponto forte da enunciação e esclarecimento do discurso.

Outro ramo de conhecimento que tem como objeto de estudo a linguagem, e que de antemão poderíamos imaginar que fosse o único ou o mais importante, é a lingüística. Segundo Julia Kristeva, lingüista e psicanalista búlgara, em seu livro sobre a História da Linguagem,

a escrita é considerada uma *representação* do falado, o seu duplo fixador, e não uma matéria particular cuja combinatória obrigue a pensar um tipo de funcionamento da linguagem diferente do fonético. Portanto, a ciência da escrita parece estar amarrada a uma concepção de que a linguagem se confunde com linguagem falada, articulada segundo as regras de uma certa gramática (2003, p. 40).

Graças ao desenvolvimento das tecnologias de informação, que permitiram o processamento de grandes quantidades de dados, gerados nas pesquisas, em menor espaço de tempo, podemos descrever um mecanismo que melhor represente as inter-

relações nas quais as palavras participam quando são invocadas para compor uma representação do pensamento humano. Tornando-se, inclusive, objeto de ciências que até então não tinham visto a linguagem como seu objeto de estudo, como a matemática.

A idéia de estudar os textos a partir do ponto de vista matemático, originou-se em meados de 1930, quando na Europa surgiram várias escolas lingüísticas, e uma das mais importantes, que foi a Escola de Praga, se dedicou à fundação da primeira disciplina dentro da nova lingüística, a fonologia (o estudo dos fonemas da língua), oposta à fonética (o estudo dos sons da fala). A Escola de Praga mostrou que se podia tratar o sistema de fonemas de uma língua de modo lógico-matemático, usando ferramentas até então exclusivas das ciências naturais (LEPSCHY, 1975).

O tratamento estatístico de textos escritos, por sua vez, só veio ocorrer em meados do século XX, quando Zipf, lingüista americano, realizou um estudo detalhado de vários textos, incluindo o *Ulysses* de James Joyce, por se tratar, na época do romance mais longo já escrito. O trabalho de Zipf (1972) consistiu na contagem das palavras e num posterior trabalho de correlação entre essas quantidades e o *rank* de sua freqüência no texto. Esse trabalho indicou que há um padrão de comportamento nos textos escritos. Zipf realizou estudos de correlação entre *rank* e freqüência não somente com textos, mas também com outros comportamentos do indivíduo e de grupos de pessoas, pois sua intenção era provar que todas as ações humanas, sejam individuais ou coletivas, eram regidas pelo princípio do mínimo esforço.

O elemento principal deste trabalho é a palavra. A palavra como elemento de construção de um sistema com finalidades bem precisas, os textos escritos. E, por conta de sua natureza humana, a composição de um sistema, cujos elementos sejam as palavras, não poderia ter outro perfil que não o de um sistema complexo. Uma frase de Júlia Kristeva resume a gama de possibilidades do que poderia ser utilizado para estudar as palavras: “O nosso século é (tanto o do átomo e do cosmos, como) o da linguagem” (2003, p. 9). Nesse leque abrangente de possibilidades, do átomo ao cosmos, encontra-se a linguagem. Como agente atuante e modificante da realidade e também como objeto de estudo.

A proposta desse trabalho é transformar as palavras em elementos para manipulação matemática e computacional, trabalhando não somente com a perspectiva de análise das frequências no sistema analisado, textos escritos, como também avaliando o significado das ocorrências conjuntas das palavras e a topologia que é criada quando esses elementos são colocados como uma grande rede, uma rede complexa.

As principais disciplinas envolvidas nesta pesquisa foram a matemática, a computação, a física, a lingüística e a psicanálise. A matemática emprestou os conceitos da teoria dos grafos para a representação simbólica das várias inter-relações existentes entre os elementos do sistema. As estruturas de rede, para serem construídas, necessitaram fundamentalmente dos recursos de formalização algorítmica e de estruturação necessárias para o processamento computacional. Foram desenvolvidos códigos para a manipulação da grande quantidade de dados e dos cálculos dos parâmetros fundamentais para a caracterização de redes.

A física, por sua vez, com a sua abordagem estatística, deu-nos as ferramentas para a análise dos textos como um sistema complexo, usando o modelo de redes complexas, permitindo determinar parâmetros que confirmem a existência de um padrão de comportamento.

Da lingüística tomamos os conceitos, as definições e, o principal, a delimitação do nosso objeto de estudo, a fala e os processos psíquicos envolvidos em sua realização.

O ponto de vista de Freud, sobre o aparelho de linguagem e o aparelho psíquico, serviu como base para a interpretação dos resultados alcançados. O que significa o relacionamento entre as palavras em uma sentença ou texto, e como se dá a representação dos conceitos em nossa mente são pontos que Freud aborda no estudo *A interpretação das Afasias* de 1891 (FREUD, 1979) e no *Projeto para uma psicologia* de 1895 (FREUD, 1996), que utilizamos como modelo teórico para análise dos resultados da pesquisa.

O trabalho consta de cinco partes principais. A primeira, esta apresentação, em que demonstramos a complexidade do objeto de estudo e o caráter interdisciplinar da pesquisa realizada. O fundamento teórico explica o que foi utilizado de cada uma das disciplinas envolvidas. No desenvolvimento da pesquisa explicamos os procedimentos realizados no processamento dos textos, nos critérios de amostragem, no pré-

tratamento empregado, na construção das redes e no cálculo dos parâmetros. Na seção de Resultados, constam as tabelas e gráficos com os índices calculados e suas explicações. Por último, as conclusões, trazendo as nossas reflexões sobre a relação dos resultados encontrados e o modelo de uso da linguagem conforme o aparelho psíquico de Freud.

Pode ser encontrado no Apêndice (Apêndice A.1) a lista de todos os textos analisados, com suas respectivas classificações. Além disso, existem cálculos de outros parâmetros da rede, que a princípio não contribuíram para elucidar o comportamento do fenômeno analisado, mas que entendemos ser de utilidade para trabalhos futuros (Apêndice A.2).

CAPÍTULO I – A LINGUAGEM

Definições sobre conceitos lingüísticos como signo, significado, significante, significação e a palavra foram necessários para compreender o objeto de nosso estudo. Os componentes fisiológicos e os componentes psíquicos envolvidos nessa capacidade foram também analisados, culminando numa descrição minuciosa do aparelho psíquico proposto por Freud, que se tornou o fundamento da abordagem dos textos, produzidos por indivíduos, como um sistema complexo.

Uma seção foi reservada para esclarecimentos sobre o conceito de sistema complexo, sua caracterização através da física estatística, trazendo também um inventário de trabalhos sobre a linguagem realizados sob a ótica dos sistemas complexos e sua modelagem.

1.1 PENSAMENTO LINGÜÍSTICO

Apesar de as línguas contemporâneas terem surgido há tanto tempo, somente no século IV aC ocorreram os primeiros estudos sobre as línguas (FIORIN, 2004). Os hindus iniciaram os estudos de sua língua, motivados por razões religiosas, buscando garantir a integridades dos textos sagrados reunidos no *Veda* durante seu uso nos cultos religiosos.

Já os gregos, segundo Margarida Petter no mesmo livro, buscavam o estudo das palavras e seus significados, acompanhados também pelos latinos, na figura de Varrão que se dedicou ao estudo da gramática, esforçando-se por defini-la como arte e ciência (FIORIN, 2003).

O pensamento sobre a língua como sendo una e universal já havia se manifestado, na Idade Média, entre os *modistas*, para quem as regras gramaticais eram independentes das línguas em que se realizavam.

Nos séculos XVII e XVIII surgiram estudos que serviram para embasar a idéia de que a linguagem se funda na razão, é a imagem do pensamento e que os princípios de

análise estabelecidos não se prendem a uma língua particular, mas servem a toda e qualquer língua.

O século XIX confere um ponto de vista mais científico aos estudos lingüísticos, empregados no estudo da evolução histórica das línguas, que deu origem à gramática comparada indo-européia (LEPSCHY, 1975). Em 1878, Saussure, um jovem de 21 anos, “descobriu alguns princípios fundamentais como a chamada lei das palatais que [...] revolucionava a visão do indo-europeu” (LEPSCHY, 1975, p. 27).

As descobertas de Saussure foram possíveis pelo caráter metodológico de sua análise, que considera a língua como um sistema composto de elementos abstratos, definíveis com base em suas funções estruturais e não em seu aspecto fonético. Apesar de tão precoce produção, somente a sua obra póstuma, *Cours de linguistique générale*, ganhou representatividade no estudo lingüístico, trazendo conceitos largamente aceitos, tais como *langue* (língua) e *parole* (fala), e a noção de entidade lingüística como sistema de signos.

Para entender a diferença entre língua e fala, Saussure esquematiza (Figura 1) e descreve o circuito da fala, simulando a conversa entre duas pessoas *A* e *B*, identificando posteriormente onde ocorre cada um dos aspectos da linguagem.:

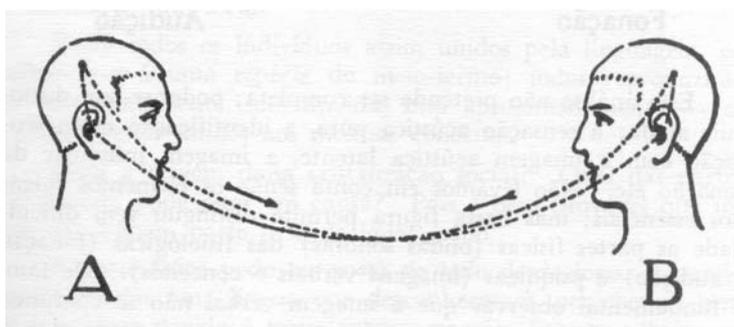


Figura 1 – Esquema do circuito da fala (Fonte: SAUSSURE, 2004, p.19)

O ponto de partida do circuito se situa no cérebro de uma delas, por exemplo *A*, onde os fatos de consciência, a que chamaremos conceitos, se acham associados às representações dos signos lingüísticos ou imagens acústicas que servem para exprimi-los. Suponhamos que um dado conceito suscite no cérebro uma imagem acústica correspondente: é um fenômeno inteiramente *psíquico*, seguido, por sua vez, de um processo *fisiológico*: o cérebro transmite aos órgãos da fonação um

impulso correlativo da imagem; depois, as ondas sonoras se propagam da boca de A até o ouvido de B: processo puramente *físico*. Em seguida, o circuito se prolonga em B numa ordem inversa: do ouvido ao cérebro, transmissão fisiológica da imagem acústica; no cérebro, associação psíquica dessa imagem com o conceito correspondente (SAUSSURE, 2004, p. 19).

Para que a língua funcione como um sistema, torna-se necessário estabelecer um meio-termo entre todos os seus falantes, quando então “todos reproduzirão – não exatamente, sem dúvida, mas aproximadamente – os mesmos signos unidos aos mesmos conceitos” (ibid, p. 21).

As partes do circuito que garantem esse acordo são referentes às faculdades receptiva e coordenativa. A primeira sugere que a imagem acústica, ao ser percebida pelo indivíduo, virá acompanhada de alguma experiência ou sensação que produzirá marcas, que chegam a ser sensivelmente as mesmas em todos. Estas compõem o conceito a ser registrado em associação à imagem acústica em questão.

Já a faculdade coordenativa ocorre de duas maneiras distintas. A primeira refere-se à associação na memória das palavras que possuem algo em comum, formando grupos com relações diversas (relações associativas) e a segunda está relacionada ao caráter temporal da fala que provoca associação entre palavras pelo encadeamento em sintagmas¹ (relações sintagmáticas). Assim, somente uma parte do processo psíquico seria constituinte da língua, aquela referente ao despertar de um conceito a partir de uma imagem acústica.

A parte do processo psíquico que envolve o despertar de uma imagem acústica por um conceito, a partir de fatos da consciência, pelo contrário, não compõe o acordo referente ao caráter social da linguagem. São processos individuais, raramente reproduzidos pelo coletivo. Segundo Saussure, esta parte “é sempre individual e dela o indivíduo é sempre senhor; nós a chamaremos *fala (parole)* (SAUSSURE, 2004, p. 21).

No caso deste trabalho, fica evidente que a busca é com relação a padrões de utilização da fala, já que, além de referir-se à escolha de palavras que irão compor os textos analisados, não há uma preocupação com a estrutura sintática e nem

¹ Sintagma é definido por Saussure como combinações entre termos no discurso em virtude do caráter linear da língua (SAUSSURE, 2004).

propriamente semântica, pois fazemos um tratamento prévio dos textos, eliminando todas as suas características sintáticas, exceto a composição da sentença, e todas as características gramaticais, com a redução das palavras à sua forma canônica, como veremos na metodologia.

A definição, porém, de alguns conceitos lingüísticos se tornam necessários para o entendimento do processo associativo da simulação proposta e para a compreensão do pré-tratamento que aplicaremos aos textos.

1.2 SIGNO, SIGNIFICADO, SIGNIFICANTE, SIGNIFICAÇÃO E PALAVRA

A unidade lingüística é constituída da união de dois termos, e é denominada de signo lingüístico. O signo lingüístico une não uma coisa e uma palavra, mas um conceito (significado) e uma imagem acústica (significante). Segundo Saussure o signo lingüístico apresenta duas características marcantes: é totalmente arbitrário e possui um caráter linear.

A arbitrariedade do signo nos diz que a união entre um significado e um significante não se dá por conta de relação alguma entre a sonoridade do significante que faça lembrar o seu significado. Em relação à segunda característica, nos diz Saussure: “O significante, sendo de natureza auditiva, desenvolve-se no tempo, unicamente, e tem as características que toma do tempo: a) representa uma extensão, e b) essa extensão é mensurável numa só direção: é uma linha” (SAUSSURE, 2004, p. 84).

A significação está relacionada ao *valor lingüístico*, que é obtido a partir da delimitação das entidades constituintes da língua: as idéias e os sons. O pensamento, psicologicamente, não passa de uma massa amorfa e indistinta. Não existem idéias preestabelecidas e nada é distinto sem o recurso dos signos. Por outro lado, a substância fônica também não possui unidades preestabelecidas, às quais os pensamentos devem acomodar-se. Assim esclarece Saussure sobre as delimitações recíprocas entre o pensamento e o som promovida pela língua:

O papel característico da língua frente ao pensamento não é criar um meio fônico material para a expressão das idéias, mas servir de intermediário entre o

pensamento e o som, em condições tais que uma união conduza necessariamente a delimitações recíprocas de unidades (ibid, p. 131).

Delimitações recíprocas promovidas entre pensamento e som produzirão unidades lingüísticas compostas de significado e significante. Porém, identificar essas entidades concretas ou unidades da língua é tarefa difícil, e Saussure, mesmo compreendendo tratar-se de uma aproximação, propõe considerar a palavra como esta unidade.

Pelo caráter concedido à palavra, por Saussure, de ser o resultado de uma delimitação, consideraremos a definição proposta pelo dicionário²: “unidade da língua escrita, situada entre dois espaços em branco, ou entre espaço em branco e sinal de pontuação”.

1.3 PALAVRAS GRAMATICAIS E LEXICAIS

No livro *Introdução à Estilística*, Martins esclarece o conceito de léxico, pontuando a diferença entre palavra gramatical e palavra lexical (MARTINS, 2003).

Baseando-se no estudo de Josette Rey-Debove (apud Martins, 2003), a autora discrimina três definições de léxico:

- a) conjunto de morfemas de uma língua, sendo os morfemas unidades significativas mínimas, presas ou livres, de natureza lexical ou gramatical. Exemplo: cant-aremos (2 morfemas);
- b) conjunto de unidades ou palavras de classe aberta de uma língua, podendo-se considerar essas unidades os morfemas ou palavras lexicais;
- c) conjunto de palavras de uma língua.

Esse último é o conceito tradicional, e o que vamos considerar neste estudo, sendo o termo *palavra* empregado conforme a definição dada anteriormente.

² Dicionário eletrônico disponível em: <http://biblioteca.uol.com.br/>. Acesso em 12 set 2005.

As palavras do léxico de uma língua podem ser classificadas em gramaticais ou lexicais. As palavras gramaticais também podem ser denominadas palavras-formas, palavras vazias e instrumentos gramaticais (ou não-palavras). Segundo Martins,

sua significação só é apreendida no contexto lingüístico, daí dizer-se que é intralingüística ou interna. Diz-se também que são palavras sinsemânticas, por serem significativas quando acompanhadas de outras, em oposição às autosemânticas – as lexicais, que têm significação por si mesmas. [...] Sua função pode estar relacionada com o ato de enunciação, com a organização do discurso ou texto, ou com a estruturação da frase (MARTINS, 2003, p. 72).

As palavras gramaticais são pouco numerosas (em torno de 100), mas de altíssima freqüência nos enunciados. A saber: preposições, pronomes, artigos, numerais, advérbios, conjunções e interjeições. As palavras gramaticais têm a clara função de organizar o texto, seguindo regras mais ou menos fixas. Como quase nunca trazem um conceito, uma idéia associada, por não se referirem a fenômenos e objetos do mundo real, se fizerem parte da rede irão provocar resultados equivocados, principalmente por ocorrerem com freqüência.

Já as palavras chamadas lexicais ou nocionais, reais, plenas ou lexicográficas, mesmo isoladas, fora da frase, despertam em nossa mente uma representação, seja de seres, de ações, de qualidade de seres ou modos de ações. Martins acrescenta,

diz-se que elas têm significação extralingüística ou externa, visto que remetem a algo que está fora da língua e que faz parte do mundo físico, psíquico ou social. São em número muito grande, indeterminável, pois constantemente se formam novas palavras ou se tomam emprestadas palavras de outras línguas (MARTINS, 2003, p. 77).

Essas palavras formam o maior percentual do léxico de uma língua e apresentam crescimento rápido, principalmente os substantivos, pois é necessário sempre criar novos nomes para representar as coisas do mundo real, sempre em constante mutação.

Os substantivos, os adjetivos, os advérbios a eles derivados, os verbos que exprimem ação ou algum processo mental (excluídos, portanto, os auxiliares e os de ligação, que são palavras gramaticais) devem permanecer e, enfim, fazer parte da rede das palavras que iremos construir.

Estes conceitos de palavras gramaticais e lexicais são bastante utilizados nas pesquisas, sejam teóricas ou práticas, desenvolvidas com a linguagem. Volnyr Santos, em artigo publicado na Internet sobre a língua portuguesa, refere-se às palavras lexicais e gramaticais, dizendo que

o léxico compreende dois tipos de palavras: as chamadas palavras lexicais, que constituem o sistema aberto de comunicação em face de sua mobilidade (surgimento, desaparecimento, transformações de sentido). São caracterizadas pelos nomes (substantivos, adjetivos, advérbios) e pelos verbos (SANTOS, 2005).

Já as palavras gramaticais, para ele, “são as que exprimem as relações entre as idéias, a contextura da língua, o sistema, a gramática, representadas pelas desinências verbais (tema, vogal temática), artigos, pronomes, conjunções, preposições. Constituem o sistema fechado da comunicação” (SANTOS, 2005).

Ele corrobora a informação do número reduzido das palavras gramaticais e acrescenta a importância sintática que exercem na construção da língua:

As palavras gramaticais são em número reduzido e limitado (em torno de 100 palavras), enquanto que as palavras lexicais, em estado de permanente renovação, incorporam-se e integram-se à língua, através de adaptações gráficas, morfológicas ou semânticas. Ainda que de número reduzidíssimo praticamente imutável, as palavras gramaticais são responsáveis pela estrutura da língua (SANTOS, 2005).

Há ainda uma distinção básica entre os vocábulos da língua em dois grupos: o das palavras nocionais e o das palavras gramaticais (AZEREDO, 2001).

Nem só na lingüística é feita essa distinção entre palavras de conteúdo e palavras gramaticais. Também na Teoria da Informação há essa diferenciação, porém as palavras gramaticais ganham um nome novo, *stopword*, referindo-se à mesma classe de palavras que não possui significado intrínseco (SILVA et al, 2005).

1.4 COMPLEXIDADE DA LINGUAGEM

A linguagem humana, certamente, tem um grau de complexidade elevado, não só pelo volume de informações manipulado, como também pelo intrincado método sintático, que se utiliza de pequenas unidades – palavras – que juntas em sentenças formam significações:

Em uma simples conversação, você anuncia em média 180 palavras a cada minuto. Esse fluxo contínuo da linguagem não apenas ocorre sem esforço, mas é, também, quase perfeito. As palavras são escolhidas em um dicionário mental (ou léxico) contendo cerca de 60.000 a 120.000 palavras. Contudo, só raramente você erra na pronúncia ou no uso de uma palavra; talvez, uma palavra em 1 milhão seja produzida de forma incorreta. Essa performance sem falhas na *geração* da linguagem corresponde a uma notável eficiência na *compreensão* da linguagem. Embora em grande parte não tenhamos consciência disso, esse uso da linguagem exige um extenso conhecimento sobre a estrutura lingüística, incluindo a gramática, o significado e as formas sonoras (KANDEL et al, 1997, P. 505).

Essa capacidade lingüística da raça humana, onde o volume de informação manipulada exige uma estrutura cerebral complexa, necessita de uma estrutura física, ou seja, neurológica, que permita manuseá-lo com eficiência, e de forma otimizada.

1.5 FUNÇÕES DO CÉREBRO ENVOLVIDAS NA LINGUAGEM

A referência mais antiga sobre estudos do cérebro é um papiro do século XVII a.C., o Papiro Cirúrgico de Edwin Smith. Neste documento, encontram-se registrados os sintomas, diagnósticos e prognósticos de dois pacientes que haviam sofrido ferimentos na cabeça (KANDEL et al, 1997).

Desde então, o homem busca compreender como o cérebro produz a notável individualidade da atividade humana. Os avanços tecnológicos, desde o Papiro Cirúrgico até os dias atuais, permitem-nos definir o cérebro como

uma rede precisa, onde mais de 100 bilhões de células neurais individuais, interconectadas em sistemas que produzem nossa percepção do mundo exterior, fixam nossa atenção e controlam a maquinaria da ação (KANDEL et al, 1997, p. 03).

O que se pode afirmar hoje é que a linguagem humana é produzida no córtex cerebral, que se enrola e recobre os hemisférios cerebrais como a casca de uma árvore. Por sobre os dois hemisférios fica a camada, extremamente enrugada, do córtex cerebral, que é dividido em quatro lobos: frontal, parietal, occipital e temporal (Figura 2). Porém, uma questão polêmica ainda persiste, até os dias atuais, a respeito de como a linguagem se processa nessa região.

Hoje, o ramo da ciência que se ocupa em desvendar os mistérios da mente (estudos do comportamento) relacionando com o funcionamento neural (a ciência do cérebro), é a neurociência. Segundo Kandel, um questionamento fundamental é esclarecer se “os processos mentais estão localizados em regiões específicas do cérebro [e.g.: teoria localizacionista da linguagem], ou eles representam uma propriedade coletiva e emergente de todo o cérebro [e.g.: Teoria Freudiana da Linguagem]” (KANDEL et al., 1997, p. 15). Ou ainda, se houver processos mentais que ocorram em diferentes regiões cerebrais, “quais as regras que relacionam a anatomia e a fisiologia de uma região à sua participação específica na percepção, no pensamento ou no movimento”? (KANDEL et al., 1997, p. 15).

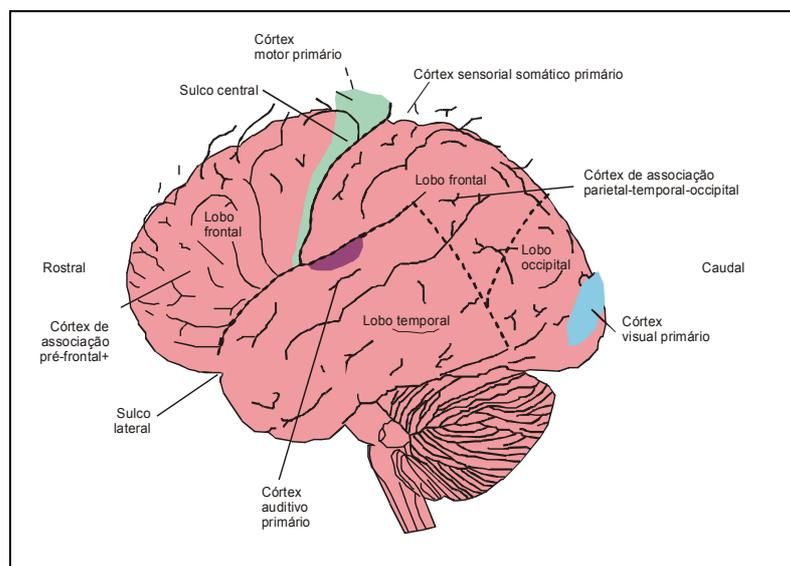


Figura 2 - Córtex cerebral e suas subdivisões (Fonte: KANDEL et al., 1997, p. 70)

1.6 TEORIA LOCALIZACIONISTA

A investigação sobre as regiões do cérebro que participam do processo de linguagem era realizada, no século XIX, através do estudo das afasias, que são uma categoria de distúrbio de linguagem, normalmente resultante de acidente vascular cerebral, ou seja, obstrução ou ruptura de vaso sanguíneo que nutre parte de um hemisfério cerebral. Nesse período, ocorreram as mais importantes dessas descobertas.

A primeira foi feita pelo neurologista francês Pierre Paul Broca, quando estudava um paciente capaz de entender tudo o que era dito, mas que não conseguia falar. Mesmo não apresentado qualquer problema motor em sua língua, boca ou cordas vocais, sendo inclusive capaz de cantar uma melodia completa sem dificuldade. Após sua morte, descobriu-se que ele tinha uma lesão na região posterior do lobo frontal, chamada atualmente de *área de Broca* (Figura 3). Esse neurologista, dando continuidade a seus estudos com mais oito pacientes que apresentavam os mesmos sintomas, descobriu que todos possuíam o mesmo tipo de lesão (FREUD, 1979; KANDEL et al., 1997; STERNBERG, 2000).

Após doze anos, o neurologista alemão, Carl Wernicke, publicou um estudo que identificava um novo tipo de afasia, onde a fala poderia ser produzida, mas a compreensão da fala não poderia ser realizada. Uma nova região do cérebro, a parte posterior do lobo temporal (Figura 3), apresentava lesões detectadas por Wernicke, que foi denominada, por sua vez, de *área de Wernicke* (KANDEL et al., 1997).

Nessa época, havia duas correntes científicas antagônicas que se ocupavam desses estudos. Os frenologistas defendiam que “o córtex seria um mosaico de funções específicas e que até mesmo os atributos mentais abstratos seriam localizados e representados em áreas corticais únicas e funcionalmente específicas” (KANDEL et al., 1997, p. 8). A corrente científica oposta pensava o cérebro como “um campo agregado, onde as funções mentais não estariam localizadas em regiões cerebrais específicas, mas sim, que cada função estaria difusamente representada por todo o córtex” (KANDEL et al., 1997, p. 7).

Wernicke então propõe uma teoria intermediária, na qual as funções mentais mais básicas, as relacionadas com as atividades perceptivas e motoras simples, ocorreriam

em regiões específicas, e que as funções intelectuais mais complexas resultariam nas interconexões entre várias regiões funcionais.

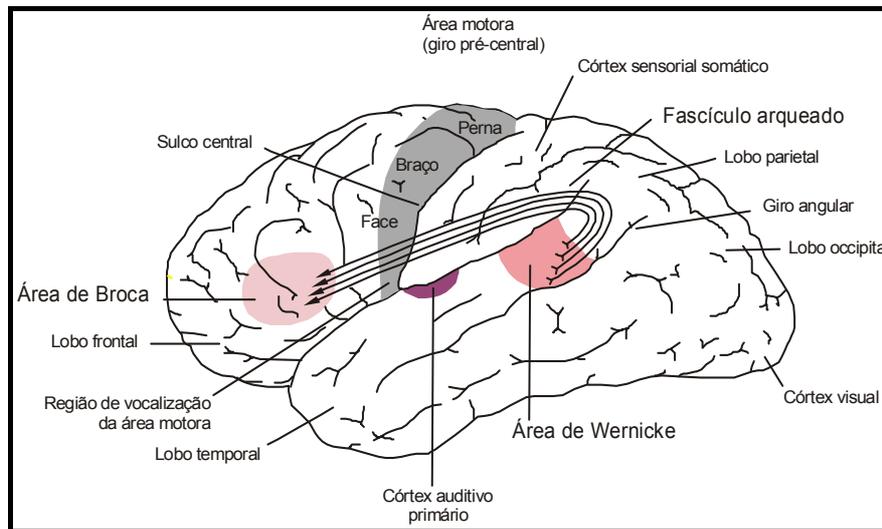


Figura 3 - Áreas de Broca e Wernicke (Fonte: KANDEL et al, 1997, p. 511)

1.7 TEORIA FREUDIANA DA LINGUAGEM

Em 1891, Freud escreve um livro denominado “A interpretação das afasias”, que tem como objetivo se contrapor às idéias localizacionistas da linguagem, defendidas por Wernicke, Lichteim, Grashey e Meynert, dentre outros (FREUD, 1979).

A crítica de Freud desenvolve-se baseada no pensamento de que “a parafasia³ observada em alguns doentes não se distingue em nada daquela troca ou mutilação de palavras que quem é saudável pode encontrar em si próprio em caso de cansaço ou de atenção distraída ou sob a influência de estados afetivos” (FREUD, 1979, p. 35).

A teoria localizacionista restringe a algumas regiões anatômicas limitadas do cérebro a função da linguagem. Para Freud, não podemos procurar o substrato fisiológico da atividade mental na função desta ou daquela parte do cérebro, mas como resultado de processos que abarcam o cérebro em toda sua extensão. Ou seja, tanto a afasia quanto a parafasia seriam distúrbios funcionais e não fisiológicos do cérebro, já que nas pessoas normais não existe nenhuma lesão nas partes indicadas pelos neurologistas que defendem essa teoria (FREUD, 1979).

Segundo Freud, no caso de lesões destrutivas, o aparelho de linguagem responde à lesão de forma solidária, como um todo, apresentando uma perturbação funcional. No caso de uma pequena lesão do aparelho de linguagem, no centro motor, não haverá como efeito uma perda de cinquenta ou cem palavras cuja natureza dependa do local da lesão. Há sim uma redução geral da funcionalidade do centro como um todo (GARCIA-ROZA, 2001).

Para Freud (apud Garcia-Roza, 2001), todas as afasias podem ser pensadas como sendo conseqüência última de interrupções da condução, isto é, sobre rupturas das associações entre os diversos elementos percebidos do mundo externo e que estão marcados no córtex cerebral.

Para se contrapor ao proposto por Wernicke, Freud propõe um aparelho de linguagem, que é descrito em termos puramente neurológicos, já que esse texto é um escrito pré-psicanalítico. A explicação trazida por Freud pretende determinar uma relação entre

³ Por parafasia devemos entender uma perturbação da linguagem em que a palavra apropriada é substituída por uma outra não apropriada que tem no entanto uma certa relação com a palavra exata (FREUD, 1979, p. 35).

uma lesão orgânica e uma perturbação no funcionamento do córtex como um todo, ao invés de supor que uma lesão em um centro específico do cérebro venha a provocar uma perturbação pontuada, como por exemplo a perda do entendimento do significado de uma meia dúzia de palavras (GARCIA-ROZA, 2001).

Garcia-Roza no entanto, faz uma advertência quanto à extensão da crítica de Freud:

Não se deve depreender da abordagem funcionalista de Freud que ele recusa qualquer referência a lugares anatômicos. O que ele nos obriga a fazer é repensar a questão da relação entre funções e localizações, os elementos tópicos sendo submetidos a arranjos e rearranjos que obedecem a exigências funcionais. A antiga teoria da localização afirmava uma relação ponto a ponto entre os *estímulos* provenientes do mundo externo e *representações* localizadas em determinados pontos do córtex cerebral, de tal forma que as representações corresponderiam a uma projeção dos elementos da periferia (2001, p. 29).

No aparelho de linguagem de Freud as atividades fisiológicas se dão em paralelo às atividades psíquicas. Não há o efeito de causalidade entre elas, que poderia levar a dois erros graves de interpretação: (1) que a atividade fisiológica tem que cessar para dar lugar à atividade psíquica e (2) que um estímulo do mundo externo levará a uma única representação relacionada ao que está contido no dado sensorial elementar, e que essa representação única ficará registrada em local específico do córtex cerebral (GARCIA-ROZA, 2001).

O próprio Freud esclarece o funcionamento de seu aparelho de linguagem:

Rejeitamos portanto as hipóteses de que o aparelho da linguagem consista em centros distintos, separados por regiões corticais isentas de funções e além disso que as representações (imagens mnésicas ^[4]) [conceito que será discutido mais adiante] que servem para a linguagem estejam acumuladas em determinadas áreas corticais denomináveis centros, ao passo que à sua associação procederiam exclusivamente as brancas massas subcorticais. Só nos resta pois formular a hipótese de que a *região cortical da linguagem seja um articulado tecido cortical* dentro do qual as associações e as transmissões em que se apoiam as funções da linguagem procederiam com uma complexidade não propriamente compreensível (1979, p. 62).

⁴ Também referida como imagens mnêmicas, pelo autor Garcia-Roza (2001).

A visão das representações como um conjunto distinto de associações exige que o aparelho de linguagem seja visto em termos estruturais, como por exemplo de uma rede, ao invés da soma de áreas corticais distintas. O aparelho é visto como algo unitário e indivisível.

As modificações que se dão no córtex cerebral, ou seja, as imagens mnêmicas gravadas, não possuem uma correspondência de um para um com as representações psíquicas. Antes sim, trata-se de um processo que começa numa região específica do córtex encefálico e que se estende por todo ele, em etapas que podem ser acrescentadas ou modificadas ao longo da vida de uma pessoa, deixando a possibilidade da recordação ou de mudanças de significação, e que envolve as sensações e as associações correspondentes que forem sendo feitas.

Analogamente, podemos explicar esses acréscimos ou alterações como uma teia de percepções registradas que vão se ligando a outras pré-existentes ou provocando a ruptura de ligações anteriormente estabelecidas. A recordação dessas impressões se dá quando as ligações da rede são novamente excitadas. Assim, o psíquico apresenta-se de novo como imagem mnêmica (GARCIA-ROZA, 2001). Subconjuntos de imagens mnêmicas são chamados de representações, e estas podem ser do tipo representação-objeto ou representação-palavra. A relação que se estabelece entre essas duas representações é explicado pelo esquema psíquico proposto por Freud (1979).

- **Representação-objeto**

A representação-objeto é uma série de inscrições feitas no córtex e que estão associadas a uma representação-palavra. Essas inscrições são fruto da experiência de cada indivíduo, assim como as suas associações entre si e com a representação-palavra em questão.

Freud (1979) elabora o conceito de representação como um complexo que reuniria as associações de imagens sensoriais de palavra, por uma parte, e de associações de imagens sensoriais de objeto, por outra. Isso indica que para que haja qualquer tipo de representação é necessário: a) que haja imagens associadas e b) que elas sejam armazenadas. A essas imagens foi dado o nome de imagens mnêmicas.

- **Representação-palavra**

A representação-palavra, segundo Freud, é a reunião de quatro componentes: a imagem acústica, a imagem visual da letra, a imagem motora da linguagem e a imagem motora da escrita. Por conta dessa multiplicidade de formas de representação, é que Freud afirma que qualquer operação da linguagem é um processo de *associação* que envolve várias partes do cérebro, em uma ação funcional simultânea.

- **O esquema psíquico**

A representação-palavra apresenta-se como um complexo representativo fechado e a representação-objeto como um complexo aberto. A ligação entre as duas representações se dá, pelo lado da palavra, por imagem acústica, e pelo lado do objeto por sua imagem visual, como mostra a Figura 4 (FREUD, 1979, p. 71).

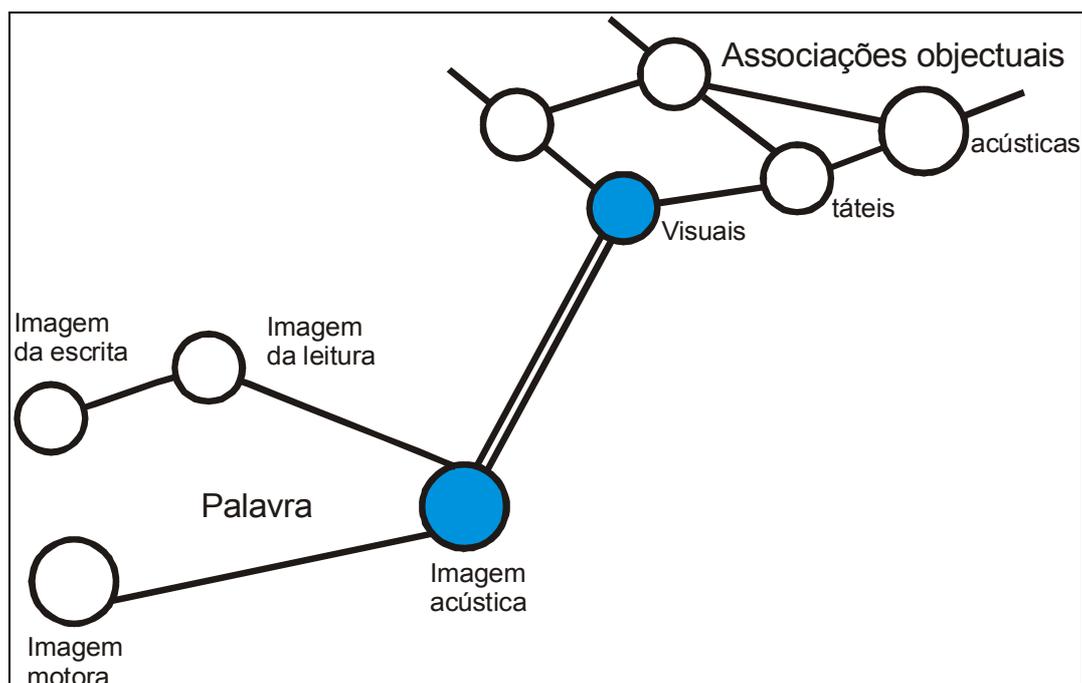


Figura 4 - Ligações entre as representações-palavra e as representações-objeto
(Fonte: FREUD, 1977, p. 71)

CAPÍTULO II — O APARELHO PSÍQUICO DE FREUD

Freud, em sua proposta do aparelho psíquico, deixara explícito que seu esquema era fundamentalmente neurológico, pois, segundo ele, essas associações da palavra e dos objetos não são um registro psíquico, mas sim um registro físico no córtex cerebral (FREUD, 1979). Porém, as teorias desenvolvidas por ele no *Projeto*, quatro anos depois, começavam a definir as bases para a psicanálise, onde o caráter psíquico da atividade mental passa a ter prioridade nas explicações dos atos humanos. Ainda assim, esse livro é considerado como pré-psicanalítico.

Algumas idéias apresentadas no livro *Sobre as Afásias* se configuraram como um primeiro modelo do aparelho psíquico, como descrito por Garcia-Roza: “A verdade é que o aparelho de linguagem produzido por Freud transbordou os limites estritos de um aparelho de linguagem e constituiu-se como o primeiro modelo freudiano de *aparelho psíquico*” (2001, p. 68).

Essas idéias iniciais ganharam novas explicações no *Projeto*, escrito em 1895, onde Freud tinha a intenção de oferecer uma concepção quantitativa do funcionamento dos processos psíquicos. Neste livro, ele propõe um modelo teórico do aparelho psíquico que traz dois conceitos básicos:

- Neurônios = partículas materiais;
- Q = o que distingue a atividade do repouso, e está sujeita às leis gerais do movimento.

Os neurônios são anatomicamente iguais, porém sofrem alterações na maneira como funcionam, dependendo em que processo/atividade estejam envolvidos. Estão ligados numa rede de conexões, formando vias de condução por onde fluirá a energia que circula entre eles. Nos contatos entre os neurônios existe uma resistência variável à passagem da energia (que varia desde a ausência de resistência até a completa resistência). Freud denominou esse mecanismo de *barreiras de contato*.

A energia Q circulante tem duas origens distintas. A primeira (Q), proveniente dos estímulos externos, e a energia ($Q\eta$) é de fonte endógena, e se origina nas células do corpo responsáveis pelas *exigências da vida*, como respiração, sexualidade, etc.

Ao sistema nervoso são atribuídas duas funções: primária e secundária. A função primária refere-se à descarga da energia (Q) nos mecanismos musculares⁵, através das vias correspondentes. Já a função secundária está associada ao acúmulo de $Q\eta$ suficiente para realizar uma ação específica relacionada às *exigências da vida*. A combinação da Q com os neurônios levou à noção de um *neurônio catexizado*⁶, ou seja, cheio de determinada Q , possíveis somente por causa das barreiras de contato entre os neurônios, que fazem oposição à descarga (FREUD, 1996).

As barreiras de contato, que dificultam a passagem da Q , são importantes também para o entendimento de uma função emergente do funcionamento do cérebro: a memória. Mas é preciso especificar de que memória se refere:

não devemos confundir essa ‘memória neuronal’ com a memória tal como é entendida pela psicologia. Não se trata da memória consciente, mas da capacidade do tecido nervoso de ser alterado de forma permanente, contrariamente a uma matéria que permitisse a passagem da energia e retornasse ao seu estado anterior (GARCIA-ROZA, 2001, p. 94).

A idéia da ‘memória neuronal’ admite que, após a passagem da Q , os neurônios deveriam ficar permanentemente modificados, por isso sendo denominados de *células mnêmicas* ou *neurônios impermeáveis* ou *neurônios ψ* . Por outro, entende-se que os neurônios envolvidos na percepção devem sempre retornar à condição anterior, antes da excitação provocada pelo mundo externo, sendo assim denominados de *células perceptuais* ou *neurônios permeáveis* ou *neurônios ϕ* .

A diferença entre os neurônios não estaria em sua anatomia, mas sim se configuraria pela ambiência a que estão submetidos:

Deve-se então supor que pelos neurônios ϕ passam quantidades contra as quais a resistência das barreiras de contacto é praticamente nula, ao passo que aos neurônios ψ só chegam quantidades da mesma ordem de magnitude que essa resistência. [...] A diferença na essência de ambos é substituída por uma diferença na ambiência a que estão destinados (FREUD, 1996, p. 356).

⁵ Toda descarga é em direção à motilidade, ou seja, transformação da Q em movimento.

⁶ Na Edição Standard Brasileira, de 1996, Freud utiliza a palavra *catexia* com o mesmo significado que é dado à palavra *investimento*, no livro de Garcia-Roza. Neste trabalho irá aparecer uma ou outra, porém com o mesmo significado.

Os processos que descrevemos até agora, e que ocorrem nos sistemas ϕ e ψ , são inconscientes. Então, para explicar as questões da consciência e da qualidade⁷, Freud propõe a existência de um terceiro sistema, o sistema ω :

Assim, reunimos ânimo suficiente para presumir que haja um terceiro sistema de neurônios - ω , talvez – que é excitado junto com a percepção, mas não com a reprodução, e cujos estados de excitação produzem diversas qualidades – ou seja, são sensações conscientes (FREUD, 1996, p. 361).

Com a introdução do sistema ω , e de sua função de registro das qualidades, o aparelho psíquico deixa de ser totalmente quantitativo, já que nesse sistema o que é percebido não é a quantidade de $Q\eta$, pois esta não o alcança. O que fica registrado em ω é a ocorrência de uma passagem de Q em ϕ e ψ , através do registro do período, da duração da passagem. Freud esclarece a sua hipótese sobre o caráter qualitativo do sistema ω :

A hipótese, porém, vai mais longe [e presume] que os neurônios ω sejam incapazes de receber $Q\eta$, mas que, em compensação, se apropriem do *período* de excitação, e que nesse estado de serem afetados por um período enquanto são enchidas de um mínimo de $Q\eta$ constitui a base fundamental da consciência (FREUD, 1996, p. 362).

Os neurônios do sistema ϕ são aqueles atingidos pelos estímulos externos (que possuem uma quantidade e uma característica qualitativa), no caso, os neurônios que compõem a massa cinzenta da medula espinhal, que são os únicos a estar em contato com o mundo externo. Mas vale ressaltar que esse contato não se dá diretamente, existindo o que Freud chamou de *aparelhos nervosos terminais* (órgãos do sentido), que atuam como telas protetoras, deixando passar apenas frações da Q para os neurônios ϕ (FREUD, 1996).

Essa quantidade é totalmente descarregada, e sua qualidade é transferida para o sistema ω . Assim, os neurônios ψ não recebem frações da quantidade provenientes do mundo externo (Q), sofrendo, no entanto, uma excitação do sistema ω , indicando as vias a serem tomadas pela energia ψ livre.

⁷ “sensações que são *diferentes* numa ampla gama de variedades e cuja *diferença* se discerne conforme suas relações com o mundo externo” (FREUD, 1996, p. 360).

Os neurônios ψ recebem, porém, a quantidade proveniente dos estímulos endógenos ($Q\eta$), com magnitude suficiente para alterar as barreiras de contato, ou seja, “tornar as barreiras de contato mais capazes de condução, menos impermeáveis, e assim, mais semelhantes às do sistema ϕ ” (FREUD, 1996, p. 352). Esse estado em que se encontram os neurônios ψ após a passagem da $Q\eta$ é denominado de *grau de facilitação*.

O sistema ψ é dividido em ψ *núcleo* e ψ *pallium*. O primeiro receberá as quantidades provenientes das fontes endógenas e o segundo as transferências das qualidades das fontes exógenas.

Resumindo, os neurônios se afetam de três maneiras: (1) transferindo quantidade entre si; (2) transferindo qualidade entre si; (3) exercendo, segundo determinadas regras, um efeito excitante recíproco (*ibid*).

2.1 A MEMÓRIA ψ

O sistema ψ se transforma num aparato de memória, exatamente porque as barreiras de contato entre seus neurônios apresentam graus de facilitação diferenciados emergindo as características da memória: “a memória está representada pelas diferenças nas facilitações entre os neurônios ψ ” (FREUD, 1996, p. 352). Dessa maneira, explica-se porque motivo uma via teria a preferência sobre outra.

A complexidade dessa estrutura está bem representada em um trecho do livro de Garcia-Roza:

As *Bahnungen* [facilitações] formam, na trama dos neurônios, caminhos privilegiados que se entrecruzam formando uma rede complexa, de tal modo que a repetição exata de um mesmo percurso seja praticamente impossível. A memória não é, pois, a reprodução mecânica e idêntica de um traço concebido como algo imutável, mas uma memória constituída pela diferença de caminhos eles mesmos móveis (GARCIA-ROZA, 2001, p. 100).

Para entendermos o funcionamento da memória ψ , no que se refere à impermeabilidade dos neurônios, analisaremos uma atividade psíquica com base na *lei da associação por simultaneidade* e a noção de *investimento (catexia) colateral*.

A catexia é um processo que pode ocorrer em um neurônio, ou grupo de neurônios, ou a uma representação, ou grupo de representações. Para que se mantenham cheios é necessário que algo oponha resistência à descarga total, não só proporcionada pelas *barreiras de contato*. Trata-se de uma estratégia de ocupação que leva em consideração a ocupação recíproca dos neurônios (associação por simultaneidade), a direção da corrente no neurônio, a relação entre o investimento e a facilitação (GARCIA-ROZA, 2001).

Pela lei da associação por simultaneidade, o *investimento colateral* é uma facilitação (via) entre dois neurônios que são investidos simultaneamente, isso porque uma $Q\eta$ passa mais facilmente para um neurônio catexizado (investido) do que para um não catexizado (FREUD, 1996). O conjunto dos investimentos e facilitações constituirá imagens mnêmicas, que Freud denomina de restos ou traços mnêmicos, numa alusão à forma fragmentada e simbólica com que representamos a realidade em nossa mente.

O investimento de neurônios do ψ *núcleo* pela $Q\eta$, quando chega em determinado nível, resulta numa descarga liberada pela via motora. A primeira via a ser seguida é a *alteração interna* (expressão das emoções, gritos, inervação muscular). Porém, como a fonte endógena é contínua, restabelecendo a tensão em ψ novamente, essa descarga não produzirá resultado aliviante. É necessária uma intervenção que impeça provisoriamente essa descarga, ou seja, não permita que o investimento chegue a esse nível.

Essa intervenção se dá, exclusivamente, no mundo externo, como uma *ação específica*, capaz de remover o estímulo endógeno provocador do desprazer. Este evento gerará uma experiência de satisfação, e provocará no sistema ψ as seguintes ocorrências: (1) descarga permanente dos neurônios do ψ *núcleo* eliminando a urgência que causou desprazer em ω ; (2) catexização de um ou de vários neurônios no ψ *pallium*, correspondentes à percepção do objeto associado à *ação específica*; (3) transferência da informação, para o sistema ω , sobre a descarga após a *ação específica* nos neurônios ψ .

2.2 PENSAMENTO E LINGUAGEM

Uma facilitação entre os investimentos ocorridos nas regiões do *núcleo* e *pallium* do sistema ψ é estabelecida quando a ação específica realizada gera uma experiência de satisfação. Então, quando ocorrer novamente um investimento dos neurônios nucleares, a $Q\eta$ passará para os neurônios da região *pallium*, reativando as imagens mnêmicas associadas (representações-lembrança). Em seguida, o sistema ψ passará a analisar as percepções recebidas (representações-percepção), a fim de encontrar alguma semelhante à que foi investida no núcleo ψ .

Se existir uma concordância completa entre a representação-lembrança e a representação-percepção, a indicação de qualidade fornecida pelo sistema ω , promoverá a descarga da $Q\eta$. Porém, se a concordância não existir ou for parcial, a descarga não ocorrerá espontaneamente, necessitando de um mecanismo que aperfeiçoe a semelhança, buscando a *identidade*. Essa busca por identidade foi denominada de *pensamento discernidor*. Assim, o processo de pensamento caracteriza-se pelo investimento dos neurônios ψ em busca da semelhança, através dos *investimentos colaterais* (FREUD, 1996).

Nas duas situações de discordância estabelece-se a busca por vias que unam as duas representações, um estado de expectativa denominado, por Freud, de *atenção psíquica*. A atenção psíquica está voltada para indicações de qualidade (ou realidade), que ocorrem em neurônios previamente investidos em grande quantidade.

Seguindo essas facilitações, a partir das indicações de qualidade em direção às percepções, a atenção psíquica provocará um hiperinvestimento nos neurônios perceptivos, e isso levará, dentre outras coisas, ao pensamento observador, cujo objetivo é o de familiarizar-se ao máximo com as vias que partem da percepção, esgotando o conhecimento do objeto perceptivo (FREUD, 1996).

Nem o pensamento discernidor nem o pensamento observador podem prescindir das indicações de qualidade. Como vimos até agora, as indicações de qualidade originam-se na percepção e são sentidas imediatamente no sistema ψ no caso da consciência, ou quando ocorre a descarga das catexias (investimentos) na ligação entre a representação-lembrança e a representação-percepção, no caso dos processos de pensamento.

Ora, até então os signos de qualidade eram informados ao sistema ψ assim que a $Q\eta$ alcançava a representação-percepção procurada, pois a esta estava associada uma indicação, proveniente de descarga do sistema ω . Em função desse comportamento diferenciado do processo de pensamento observador (ou consciente), que não termina em percepções, Freud propõe uma outra fonte dos signos de qualidade (realidade).

Trata-se de uma nova origem dos signos de qualidade que envolveria a função da linguagem no aparelho psíquico, responsável por fornecer *signos de descarga lingüística*⁸ equivalentes aos signos de qualidade anteriormente explicados. Garcia-Roza expõe a importância dessa descarga para o funcionamento do sistema ψ :

Os signos de descarga lingüística são mais um instrumento, além dos signos de qualidade fornecidos por ω , para que possa se dar a ação específica sem risco de frustração. A *inervação da fala* é um caminho de descarga para ψ , importante na regulação do equilíbrio de $Q\eta$ (GARCIA-ROZA, 2001, p. 173).

As associações da fala consistem na vinculação de neurônios ψ com neurônios participantes das representações sonoras, que estão associados a imagens verbais motoras (ver definição de representações-palavra na Seção 1.5). Essas vinculações sucessivas permitem que a excitação passe da imagem sonora para a imagem verbal e desta para a descarga (FREUD, 1996).

Dessa maneira, a catexia (investimento) entre as imagens mnêmicas dos processos de pensamento é acompanhada por sucessivas descargas através dos *signos de descarga lingüística* que funcionam como signos de qualidade, servindo também como indicação de que a lembrança é consciente. Esses signos de qualidade, assim produzidos, serão o mecanismo que permite conduzir a catexia (investimento) de ψ para as lembranças que emergem durante a passagem da $Q\eta$.

Na Seção 2.3 realizaremos um experimento mental⁹ com o objetivo de descrever a construção de um texto simples de duas sentenças, explicando os mecanismos psíquicos envolvidos nessa tarefa.

⁸ Termo proposto por Garcia-Roza (2001, p. 172).

⁹ En física y en algunos campos de la filosofía un experimento mental (de la expresión Gedankenexperiment) constituye un razonamiento lógico sobre un experimento no realizable en la práctica pero cuyas consecuencias pueden ser exploradas por la imaginación, la física o las matemáticas. Estos experimentos se utilizan para comprender aspectos no experimentables del Universo (Fonte: http://es.wikipedia.org/wiki/Experimento_mental).

2.3 EXPERIMENTO MENTAL DE CONSTRUÇÃO DE UM TEXTO

Na Figura 5, temos um exemplo de construção de uma rede de conceitos baseado na Teoria de Freud. Faremos um experimento mental, a fim de explicar como um texto de duas sentenças, apenas para efeito didático, surgiria na mente de uma pessoa, segundo essa teoria.

Imaginemos que esta pessoa, dirigindo seu carro, esteja com o pensamento vagando por lembranças agradáveis sobre o dia do nascimento de seu filho. Este pensamento, ocorrido ao deixá-lo na porta da escola, antes de ir para o trabalho, talvez tenha surgido pela constatação de que ele cresceu e está quase um homem. São pensamentos inconscientes. A pessoa não se dá conta de que essas lembranças estão sendo revividas. Um estado de alegria, porém, pode ser detectado em seu semblante, mas se alguém perguntasse a ela, subitamente, porque estava com aquele sorriso, decerto não saberia responder.

De repente, vê cair sobre o pára-brisa uma gota de chuva. A imagem percebida, provocaria uma catexização (investimento) naquelas representações-objeto registradas no córtex cerebral, referentes àquele dia chuvoso em que seu filho veio ao mundo (representado pelo círculo amarelo, identificado com o número 1, na Figura 5). Pela teoria de Freud, esse investimento, denominado de *atenção psíquica*, levaria essas lembranças, ou parte delas, à consciência. Para melhor aproveitar o nosso experimento mental, vamos supor que somente uma parte das representações-objeto veio à consciência: o estado de alegria, sem, no entanto, estar associado ao que o motivou.

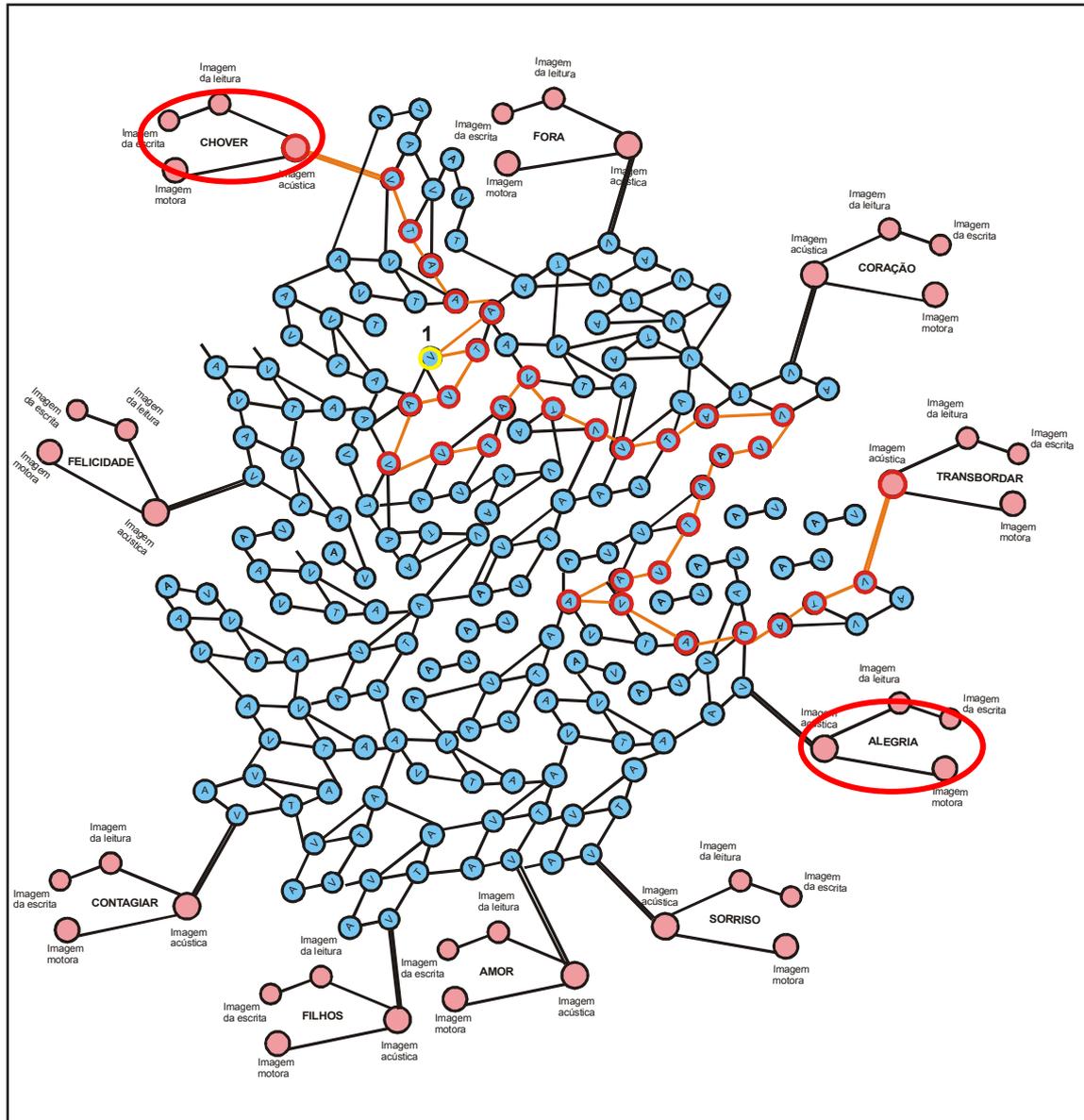


Figura 5 - Proposta de um esquema ampliado do aparelho de linguagem de Freud.

Além de provocar uma descarga na consciência, esse investimento, alcançando um nível suficiente para romper as barreiras de contato, levariam também a uma ou mais descargas, que neste caso do pensamento, se dariam sob a parte motora do sistema ψ , provocando uma série de descargas nos signos lingüísticos.

Ao observar a chuva caindo, estariam surgindo palavras que, juntas, levariam a construir, por exemplo, essas duas sentenças:

“Chove lá fora. Meu coração transborda de alegria”.¹⁰

¹⁰ Todas as frases usadas como exemplo são de autoria de Silvia Caldeira.

Dessa maneira, o que leva alguém a escolher uma determinada palavra, em detrimento de outra, é a forma como as representações-objeto estão encadeadas ou conectadas no cérebro que está produzindo o texto, pois o caminho associativo que será percorrido, de representação em representação, é que definirá, por sua vez, o encadeamento das representações-palavra que serão *despertadas* para a consciência gerando, então, a mensagem.

2.4 CONHECIMENTO ATUAL SOBRE O FUNCIONAMENTO DA LINGUAGEM

No livro *Sobre as Afasias*, Freud se dedica a uma crítica à teoria localizacionista, vigente na época. E, a partir de uma teoria sobre o aparelho de linguagem propõe uma nova explicação da fala e suas falhas. Após o estudo do aparelho psíquico, ele retoma a sua crítica, no livro *A Interpretação dos sonhos*, onde reitera sua visão sobre onde se processam as associações:

... as representações, os pensamentos e as estruturas psíquicas em geral nunca devem ser encarados como se estivessem localizados em elementos orgânicos do sistema nervoso, mas antes, por assim dizer, *entre* eles, onde as resistências e as vias de acesso fornecem seus correlatos (FREUD, 1987).

As evidências científicas atuais, como o estudo das atividades cerebrais durante processos da linguagem por imagens PET¹¹ (Figura 6), não são suficientes para concluir definitivamente sobre a existência ou não de localizações específicas bem delineadas para cada uma delas. Podemos até interpretar que existem atividades que envolvem a região em torno dos possíveis órgãos da linguagem que apresentam atividades em um grau de energia menor que poderiam ser interpretadas como as associações descritas por Freud em sua teoria.

¹¹ PET é o acrônimo da expressão em inglês Positron Emission Tomography, que significa Tomografia de Emissão de Positrões, e é um exame imagiológico da medicina nuclear que utiliza radionuclídeos que emitem um positrão quando da sua desintegração, o qual é detectado para formar as imagens do exame. (Fonte: Wikipedia - [http://pt.wikipedia.org/wiki/PET\(exame_m%C3%A9dico\)](http://pt.wikipedia.org/wiki/PET(exame_m%C3%A9dico)))

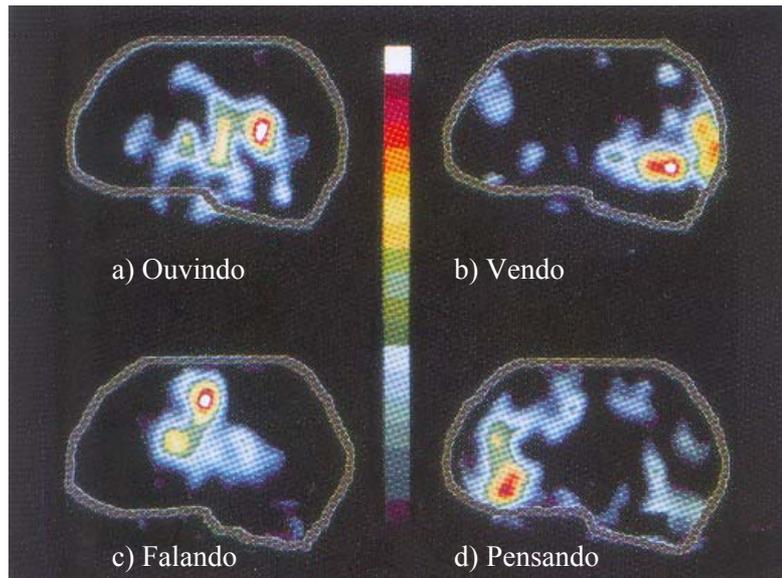


Figura 6 - Imagens PET das regiões cerebrais de maior atividade participantes nos processos de uso da linguagem (Fonte: Carter, 2003, encarte)

CAPÍTULO III – SISTEMAS COMPLEXOS E LINGUAGEM

*'Chaos is a name for any order that produces confusion in our minds.'*¹² - George Santayana

O estudo dos fenômenos naturais teve, no início do século XVII, seu direcionamento definido por René Descartes, que em seu sistema filosófico tinha como um dos princípios a divisão do evento ou objeto de estudo em partes. E, a partir da análise das partes, pretendia chegar ao conhecimento do objeto ou evento como um todo. Ele diz:

O segundo ponto é dividir cada uma das dificuldades em tantas partes quantas forem necessárias para resolvê-las melhor. O terceiro ponto, dirigir ordenadamente meus pensamentos, começando pelos objetos mais simples e fáceis de conhecer para subir, pouco a pouco, como por graus, até o conhecimento do mais complexo e supondo assim mesmo uma ordem entre os que não procederam naturalmente uns dos outros (DESCARTES, 2002, p.9).

É interessante notar que a proposta filosófica dele subentende uma hierarquia de relacionamento, que a ciência facilmente poderia determinar, apenas desenvolvendo métodos capazes de analisar as partes escolhidas retiradas do todo. Tudo muito simples, sem levar em consideração o efeito que as próprias relações entre as partes poderia causar no evento.

Esse pensamento predominou durante duzentos anos, e somente em meados do século XIX, por conta dos avanços científicos na área da termodinâmica, o método de análise dos fenômenos passou a considerar o sistema como um todo, fazendo surgir a mecânica estatística com a determinação da função de probabilidade (PINHO e ANDRADE, 2002).

Nesse ponto da história da física tornou-se, em princípio, fácil, a partir da identificação dos sistemas, construir modelos que representassem o comportamento das partes envolvidas, através de simulações matemáticas (AGUIRRE, 2004). Porém nem todos os sistemas podem ter suas partes devidamente identificadas e

¹² Tradução livre: Caos é um nome para qualquer ordem que produz confusão em nossas mentes.

caracterizadas. A estes foi dado o nome de sistemas complexos, cujo tratamento analítico, dentro da teoria geral de sistemas dinâmicos, ainda é um desafio para os matemáticos (NUSSENZVEIG, 2003).

Um sistema é definido como um conjunto de elementos que mantêm relações entre si. Um elemento é uma unidade primitiva que a cada instante está em um dentre vários estados possíveis. Diferentes elementos possuem relações entre si quando o estado de um condiciona o estado do outro, em um mesmo instante ou em um instante posterior. Na prática, tais relações e estados nem sempre são bem definidos (PESSOA JR, 1996).

É o caso dos textos, um sistema composto de significantes (aqui simplesmente referenciados como palavras) interrelacionados com a finalidade de construção de um pensamento. Não há uma maneira de caracterizar os estados e relacionamentos possíveis entre as palavras em função da construção de um texto com sentido e de maneira a ser melhor compreendido pelo leitor (através das constantes repetições de palavras já citadas).

A classificação de um sistema como complexo refere-se à determinação de medidas sistêmicas, ou seja, propriedades do sistema, e que podem ser quantificadas ou não. Não há, porém, um consenso entre a definição mais adequada para essas propriedades, que são normalmente citadas como “ordem”, “complexidade” ou “organização”. Segundo Pessoa (1996), muitas são as definições para as medidas sistêmicas e que citaremos algumas:

- **Organização enquanto condicionalidade ou restrição** - Essa definição baseia-se na condicionalidade entre os elementos do sistema, focalizando sua estrutura sem levar em consideração a sua finalidade;
- **Organização enquanto adaptabilidade** – A organização de um sistema segundo essa definição indicará o grau de eficiência no cumprimento de metas, sua boa adaptação ao ambiente e a conservação ou otimização das variáveis essenciais que garantem a sobrevivência do sistema enquanto tal;
- **Ordem como regularidade** – relaciona-se à existência de regularidades no espaço ou no tempo, ou de padrões que permitam a descrição de um sistema complicado em termos de poucas variáveis.

- **Ordem enquanto confiabilidade** – um sistema que permanece ou retorna a um mesmo estado macroscópico após alterações aleatórias em seus elementos possui alta confiabilidade, resistência a erros, inércia ante modificações.
- **Complexidade enquanto número de elementos** – empregada na engenharia de sistemas, essa definição leva em conta o número de componentes do sistema, considerar as suas interconexões.
- **Complexidade enquanto beira do caos** – a complexidade estaria associada a um meio-termo entre a ordem do cristal e o caos da fumaça. Admite a existência de um ponto crítico entre eles, onde haveria um regime de máxima evolutibilidade em sistemas vivos, e que surgiria espontaneamente como no regime de criticalidade auto-organizada.

A partir das definições de organização, ordem e complexidade expostas acima, um texto escrito, então, poderia ser caracterizado como um sistema complexo. Somente do ponto de vista em que estamos analisando os textos, ou seja, tomando a sentença como a menor unidade de significado, já seria possível defini-lo como complexo, já que um texto é composto de um grande número de unidades (significantes), que interagem entre si em pequenos grupos (sentenças). Os significantes, que aqui simplesmente chamaremos de palavras, podem apresentar significados diferentes, dependendo de sua relação com os outros significantes a que estão ligados.

O conceito de auto-organização não está associado nem à natureza física dos elementos que compõem o sistema e nem é uma maneira “pré-organizada” por algum tipo de interferência externa. Trata-se de uma característica do sistema como um todo e ocorre sob condições bem definidas.

Existem muitas maneiras de conhecer como as coisas são e como funcionam. Pode-se prever o comportamento de um objeto, fazendo relacionamentos com outros objetos similares, ou com representações deste, quando utilizamos a analogia, uma maneira indireta de aprender que se vale de igualar um fenômeno novo com um fenômeno já conhecido, e prevendo seu comportamento, a partir do modelo escolhido.

Esses modelos podem ser baseados em fenômenos ou objetos escolhidos como referência para o estudo de um outro fenômeno, encontrados na natureza, ou podem ser construídos especialmente para o propósito cognitivo. O modelo escolhido para

representar o mundo dependerá de nossas crenças, pois estas determinarão nossas visões sobre o próprio mundo

Os modelos podem ser criados de duas maneiras. Podemos escolher entre construir mecanismos que analogamente imitem um fenômeno, ou apenas conhecer as suas características sem criar nenhum mecanismo de imitação deste. Aos primeiros é dado o nome de *modelo de simulação*, onde o objetivo é conhecer como os processos ocorrem em sua totalidade ou em partes. Os modelos que possuem como objetivo investigar as características de um sistema, sem no entanto criar mecanismos que reproduzam seu funcionamento são denominados de *modelos de caracterização*.

Freud ao investigar a complexidade do funcionamento dos processos mentais, por exemplo, utilizou o modelo de simulação. A princípio somente avaliando as funções referentes ao uso da linguagem e em seguida estendendo para os processos psíquicos em geral. Trata-se de um modelo teórico de simulação, um *experimento mental*, tão comum à época.

Na pesquisa dos sistemas complexos podemos utilizar os seguintes exemplos nos dois tipos de modelos: autômatos celulares, modelos de crescimento, equações diferenciais (simulação), e mapas logísticos, fractais/multifractais (caracterização). As redes complexas podem ser utilizadas com as duas finalidades, e este foi o modelo escolhido para o nosso estudo.

3.1 REDES COMPLEXAS

Até pouco tempo atrás acreditava-se que nenhum campo científico que estudasse os fenômenos sociais e individuais humanos, desde a história e economia até as ciências políticas e psicologia, poderia criar modelos que se utilizassem de leis simples, como as da física e química, para explicá-los (BUCHANAN, 2002).

Durante seus estudos sobre as redes sociais, em 1998, Duncan Watts e Steve Strogatz (WATTS, 1999) desenharam pontos como pessoas e interligaram esses pontos com linhas, representando suas relações de amizade. Tentavam descobrir como o fenômeno dos ‘seis graus de separação’ ocorre. Como é que seis bilhões de pessoas poderiam estar distantes uma das outras, com não mais do que seis ligações entre elas.

Os autores não são os pioneiros nos estudos sobre redes sociais. Estudos iniciais datam de 1930 (SCOTT, 2000). Porém, Watts e Strogatz foram os primeiros a usar a analogia de uma rede para representar os relacionamentos sociais entre as pessoas, propondo, inclusive, um novo parâmetro de caracterização das mesmas.

A partir daí, nos últimos sete anos, sociólogos, físicos, biólogos e outros cientistas têm descoberto inúmeras conexões inesperadas entre o funcionamento do mundo humano e o funcionamento de outras coisas aparentemente sem relação, desde uma célula e o ecossistema global até a Internet e o cérebro humano, utilizando o modelo de redes complexas.

Os sistemas complexos são estruturas que, de uma maneira simplificada, podem ser representadas como uma rede complexa, formado por elementos que se interligam a partir de regras de relacionamento, como mostrado na Figura 7.

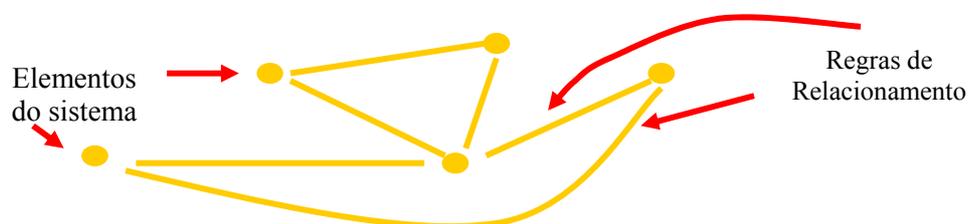


Figura 7 – Esquema simplificado de uma rede

Utilizadas como modelos que representam sistemas reais, as redes complexas geraram estudos em diversas áreas de conhecimento, principalmente pela disponibilidade atual dos bancos de dados com informações apropriadas a esse propósito (ALBERT e BARABÁSI, 2002). A Tabela 1 traz um resumo de estudos nessa direção.

Tabela 1 – Exemplos de Estudos de Redes complexas (Fonte: Artigo de Albert e Barabási, 2001)

Tipo de Rede	Total nós	CAM	CMM	Referência
WWW	153 127	0.11	3.1	Adamic, 1999
Internet	3015	0.18	3.7	Yook et al., 2001a
Atores de hollywood	225 226	0.79	3.65	Watts and Strogatz, 1998
MEDLINE co-autoria	1 520 251	0.07	4.6	Newman 2001a, 2001b, 2001c
SPIRES co-autoria	56 627	0.73	4.0	Newman 2001a, 2001b, 2001c
Co-autoria entre matemáticos	70 975	0.59	9.5	Barabási et al., 2001
Cadeia alimentar	154	0.15	3.4	Montoya e Solé, 2000
Co-ocorrência de palavras	460 902	0.44	2.67	Ferrer i Cancho e Solé, 2001
Rede energia elétrica	4 941	0.08	18.7	Watts e Strogatz, 1998

O modelo de redes complexas gera estruturas com arquitetura não-linear, aberta, descentralizada, dinâmica, e capaz de auto-organização, caracterizadas fundamentalmente pela sua horizontalidade, isto é, pelo modo de inter-relacionar os elementos sem hierarquia. A formulação desses padrões ocorreu originalmente em estudos de matemática pura, com a *Teoria dos Grafos*. Os primeiros fundamentos da Teoria dos Grafos surgiram em meados do século XIX, quando Eüler se dispôs a responder o enigma das sete pontes. O problema consistia em responder se havia um caminho para cruzar as sete pontes de sua cidade natal, Königsberg (Figura 8), passando apenas uma vez por cada uma delas (NEWMAN, 2003).

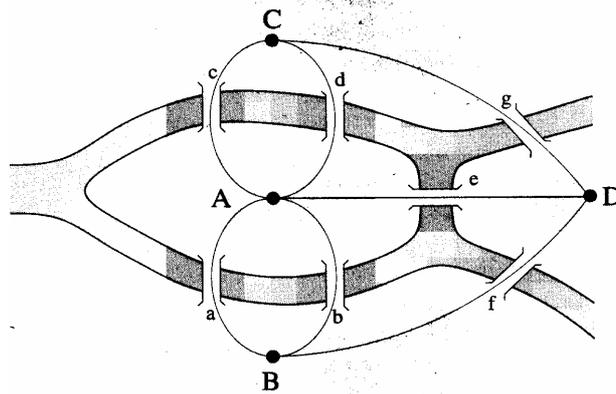


Figura 8 – Desenho esquemático das sete pontes da cidade de Königsberg (Fonte: Barabási, 2003)

Euler, a partir de um estudo matemático, não só provou a inexistência de tal caminho, como fundou a teoria dos grafos, que veio a se tornar a base para os estudos sobre redes. O grande mérito de Euler foi visualizar as pontes da cidade como um grafo, uma coleção de nós conectados por ligações, e concluir que a existência, ou não, do caminho desejado era uma *propriedade* do grafo (BARABÁSI, 2003).

A forma como uma notícia se difunde entre um grupo de pessoas que se encontram pela primeira vez numa festa, é um bom exemplo de problema para a teoria dos grafos. Também podemos citar outros, independentemente da natureza dos nós e ligações, tais como: computadores ligados por linhas telefônicas, moléculas ligadas em nosso corpo por reações bioquímicas, células nervosas conectadas por axônios.

Os grafos são estruturas matemáticas ideais para o estudo e representação das redes complexas, manipulam grandes quantidades de dados, permitindo cálculos de parâmetros que caracterizam suas diversas topologias. Alguns fatores contribuíram significativamente para o avanço da metodologia de estudo de redes:

- Automatização da aquisição de dados em todos os campos de conhecimento, contribuindo para o surgimento de grandes bancos de dados sobre a topologia de várias redes;
- O aumento da capacidade de processamento que permite investigar redes contendo milhões de nós, explorando questões que nem haviam sido previstas antes (ALBERT e BARABÁSI, 2002; KOPPEL et al, 2003);

- A lenta, porém visível, ruptura dos limites entre as disciplinas, que oferece o acesso de pesquisadores a diversos bancos de dados, permitindo descobrir as propriedades genéricas dos sistemas complexos;
- Finalmente, a necessidade cada vez maior de mover-se além das aproximações reducionistas, e tentar entender o comportamento do sistema como um todo.

Segundo Watts (1999), uma rede complexa pode ser descrita a partir de um grafo G , que consiste em um conjunto não vazio de elementos, chamados vértices, e uma lista de pares não ordenados desses elementos, chamado de ligações. Essa estrutura topológica de representação de elementos conectados por algum tipo de regra definidora permite vários cálculos matemáticos capazes de caracterizar as redes em vários tipos de estruturas.

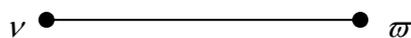
Nesta dissertação, abordamos três tipos de estruturas:

- redes aleatórias;
- redes de mundo pequeno;
- redes livres de escala.

Alguns parâmetros são utilizados para caracterizar essas estruturas, tais como, Grau Médio, Distribuição de Grau, Coeficiente de Aglomeração, Caminho Mínimo Médio, Diâmetro.

3.2 CÁLCULO DOS PARÂMETROS DE REDES COMPLEXAS

O conjunto de vértices do grafo G é chamado de conjunto vertex de G , denotado por $V(G)$, e seus elementos são chamados de vértices ou nós. A lista de ligações é denotada por $E(G)$, e seus elementos são chamados de Arestas. Se ν e ϖ são vértices de G , então uma ligação na forma de $\nu\varpi$ indica que há uma aresta, ou seja, uma relação de adjacência entre ν e ϖ .



O número de vértices em $V(G)$ é denominado ordem do grafo (n), e o número de ligações em $E(G)$ é chamado de tamanho do grafo (M). No exemplo acima, a ordem seria igual a 2, e o grau do vértice v seria igual a 1 (um).

Grau Médio ($\langle k \rangle$)

Sendo o Grau (k) a quantidade de arestas de cada nó, o Grau Médio refere-se à média desse valor. Na Figura 9 temos um exemplo de cálculo do Grau de um único vértice da rede. Essa mesma determinação é feita para todos os nós da rede e em seguida é calculado o valor médio.

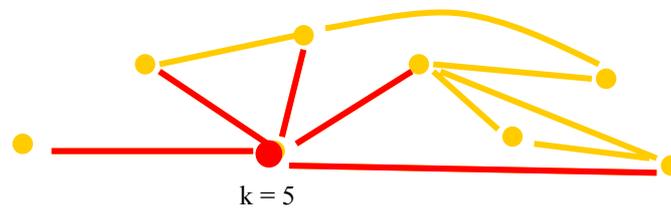


Figura 9 – Esquema Simplificado de grau de um nó ou vértice

O valor do Grau Médio pouco informa sobre as características da rede, porém a distribuição de graus é fundamental para caracterizá-la como aleatória ou não.

Distribuição de Grau

Redes aleatórias apresentam uma distribuição de graus do tipo Normal, já que o número de ligações que os nós terão é definido por uma probabilidade de conexões. Já as redes livres de escala apresentam formato em lei de potência, conforme Figura 10.

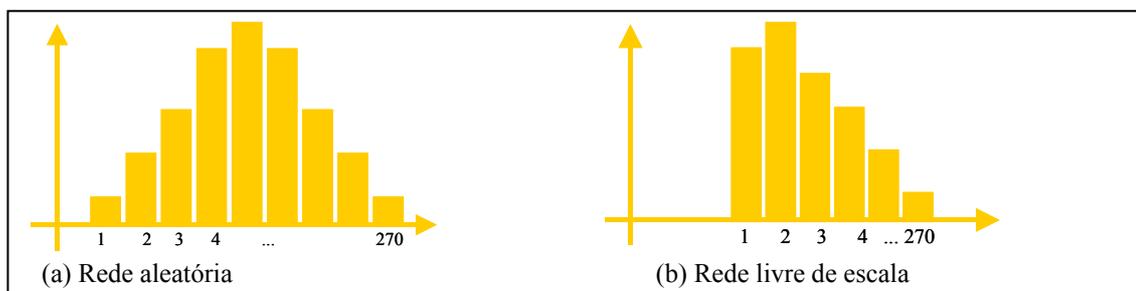


Figura 10 – Esquema Simplificado de Distribuição de Grau de uma rede

Vizinhança de Vértices

Para grafos totalmente conectados, o conceito de vizinhança é importante, pois serve para calcular o Coeficiente de Aglomeração. Segundo Boaventura Netto (2003), vizinho ou vértice adjacente de um vértice v , é todo vértice w que participa de uma ligação com v , e é denotado $\Gamma(v)$, como mostra a Figura 11.

A vizinhança $\Gamma(v)$ de um vértice v é o conjunto dos vértices adjacentes a v (não incluindo o próprio vértice v).

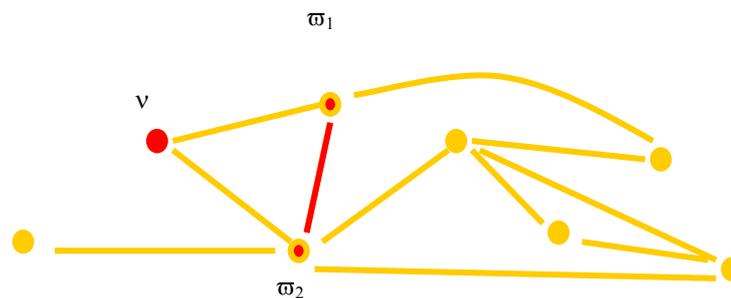


Figura 11 – Representação dos vizinhos w_1 e w_2 do vértice v .

Coeficiente de Aglomeração Médio (CAM)

Um grafo que possui uma ligação associada a cada par de vértices, ou seja, possui todas as arestas possíveis, é chamado grafo completo. Se analisarmos, porém, não o grafo inteiro, mas apenas uma sub-estrutura desse grafo, a esta ocorrência chamamos de clique (BOAVENTURA NETTO, 2003).

Um clique indica que todos os vizinhos do nó analisado estão conectados entre si, e este comportamento constitui-se numa ocorrência comum em redes sociais, representando círculos de amigos, onde cada membro conhece todos os outros. Essa tendência inerente à aglomeração é quantificada pelo Coeficiente de Aglomeração (WATTS e STROGATZ, 1998 apud ALBERT e BARABÁSI, 2002).

O Coeficiente de Aglomeração é definido como a relação entre o número de arestas existentes entre os nós vizinhos ao nó analisado e o número máximo de arestas possíveis entre esses mesmos nós (Equação 1.1).

$$CA_i = \frac{2E_i}{k_i(k_i - 1)}, \text{ onde:} \quad (0.1)$$

CA_i é o Coeficiente de Aglomeração do nó i ;

E_i é o número de arestas existentes entre os nós vizinhos ao nó analisado e

$\frac{k_i(k_i-1)}{2}$ é o número máximo de arestas possíveis entre os nós vizinhos.

No exemplo da Figura 11, o vértice analisado possui dois vizinhos e entre eles existe uma ligação. Para que eles formem um clique deveria existir uma ligação apenas. Logo, o Coeficiente de Aglomeração desse nó é igual a um. Já no exemplo da Figura 12, o nó analisado (azul) possui 3 vizinhos e para que eles formem um clique deveriam existir 3 conexões entre eles, porém só existe uma. Logo, o Coeficiente de Aglomeração desse nó é igual a 1/3.

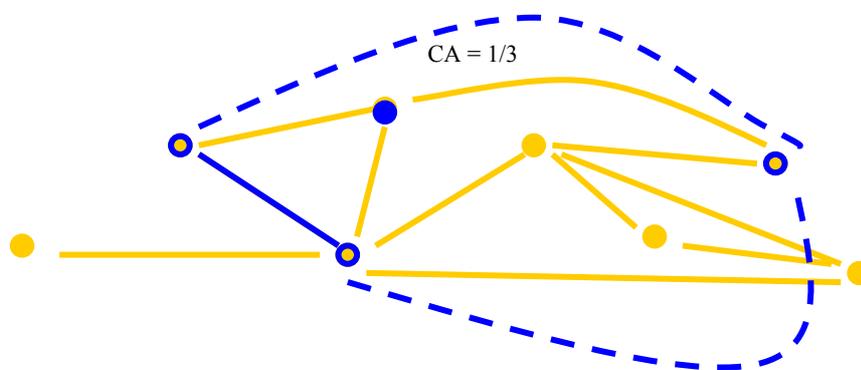


Figura 12 – Exemplo de cálculo do coeficiente de aglomeração de um vértice ou nó

Após a determinação do Coeficiente de Aglomeração de cada um dos vértices da rede é calculado o valor médio que será representativo da topologia desta, aqui denominado de CAM.

Caminho

Um dos mais importantes índices de grafos a ser considerado em redes é o tamanho do caminho característico ($L(G)$), isto é, a distância típica $d(i, j)$ entre cada vértice e todos os outros. “Distância” aqui referindo-se não à distância métrica, e sim ao número mínimo de ligações que separam o vértice j do vértice i (Figura 13). A partir da determinação do caminho podemos calcular o Caminho Mínimo Médio e o Diâmetro.

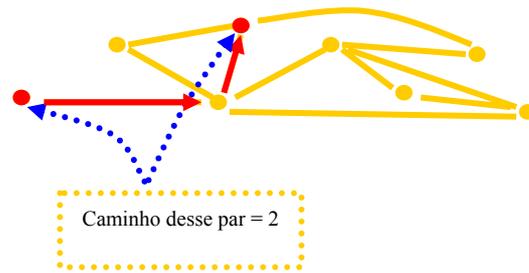


Figura 13 – Esquema simplificado mostrando o caminho entre dois nós

- **Caminho mínimo médio (CMM)**

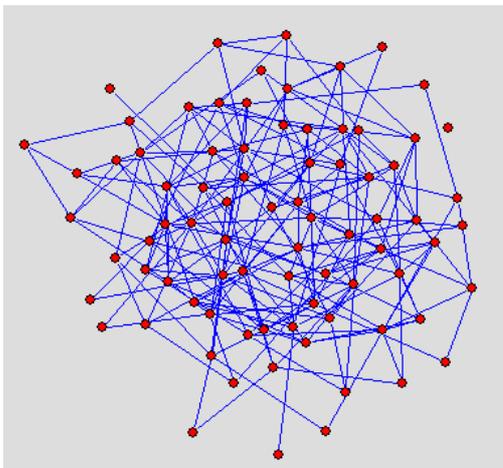
O Caminho Mínimo Médio (*CMM*) de um grafo é a média dos menores caminhos mínimos que conectam cada vértice $i \in V(G)$ a todos os outros vértices. Isto é, calcula-se $d(i, j) \forall j \in V(G)$ e encontra-se \bar{d}_v para cada v . Então *CMM* será a média dos $\{\bar{d}_v\}$.

- **Diâmetro (DAM)**

O maior caminho mínimo médio é denominado de Diâmetro da rede. E representa, em média, a maior distância entre os pontos mais distantes entre si.

Aos estudos de Eüler, que deram início à teoria dos grafos, foram acrescentadas, durante os dois séculos seguintes, contribuições importantes para o entendimento das características das redes, principalmente a partir do trabalho de dois matemáticos húngaros: Paul Erdős e Alfréd Rényi. A fim de responder à questão de como as redes se formam, seus estudos levaram à teoria das redes aleatórias (BARABÁSI, 2003).

3.3 REDES ALEATÓRIAS



Os estudos desenvolvidos por Erdős-Rényi sobre como as redes são formadas culminaram na descoberta de comportamentos universais referentes à sua estrutura, levando-os a propor um modelo para explicar a estrutura das redes na natureza. Segundo esse modelo, bastava apenas uma única ligação por nó de uma rede para que em um determinado momento,

denominado de transição de fase ou percolação, emergisse um gigante aglomerado composto por todos os seus elementos.

Para responder a perguntas fundamentais como essa, os cientistas procuram encontrar a resposta plausível mais simples possível. Segundo Erdős e Rényi (BARABÁSI, 2003), a melhor forma de explicar o mecanismo de construção das redes seria conectar os nós aleatoriamente.

Para eles, uma rede surge a partir de um grupo inicial de elementos, acrescentando conexões aleatoriamente. No início, quando só há poucas conexões, serão formados pares entre os elementos. Continuando a acrescentar ligações, inevitavelmente alguns desses pares serão conectados entre si, formando aglomerados (*clusters*). E com a continuidade de acréscimo de ligações, de tal ordem que cada nó tenha uma média de uma ligação, “milagrosamente” emergirá um único e gigante aglomerado. Dessa forma, a partir de qualquer nó, podemos alcançar qualquer outro nó, navegando através das ligações entre eles (BARABÁSI, 2003).

Os matemáticos chamam esse fenômeno de ‘emergência de um componente gigante’. Os físicos chamam-no de percolação e dirão que se trata de uma transição de fase, similar ao que ocorre quando a água congela. Os sociólogos dizem que os indivíduos formaram uma comunidade. Apesar de usar termos diferentes, todos acreditam que quando aleatoriamente acrescentamos as ligações conectando pares de nós juntos em uma rede, alguma coisa de especial ocorre: a rede, após alcançar um número crítico de ligações, drasticamente se modifica (BARABÁSI, 2003).

Erdős e Rényi disseram que basta apenas uma ligação por nó para que esse aglomerado surja. Porém, esse valor mínimo é ultrapassado freqüentemente nas redes que encontramos na natureza. A constatação desses dois matemáticos não se tornou importante por conta dessa predição do valor mínimo de conexões médio por nó para que surgisse o grande aglomerado, mas sim pelas possibilidades que surgiram para analisar redes bem maiores e com estruturas bem mais complexas do que as redes regulares que, até então, eram os únicos objetos de estudo da teoria dos grafos aleatórios (BARABÁSI, 2003).

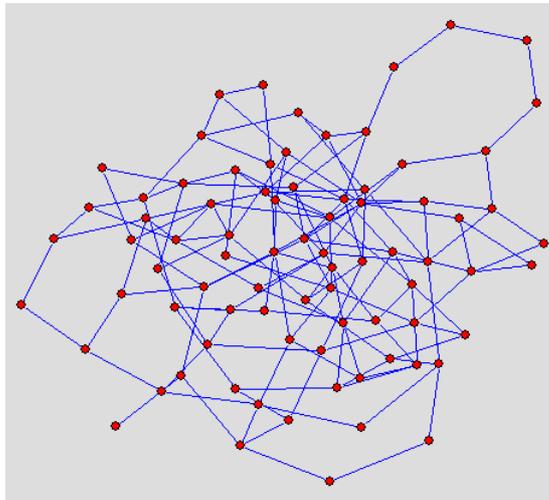
Uma das maneiras de identificar se uma rede é aleatória é fazer uma medida de quantos nós possuem uma, duas, três até k ligações e depois traçar o histograma, que, no caso de uma rede aleatória, seguirá uma distribuição de Poisson. A distribuição de Poisson tem um pico proeminente, indicando que a maioria dos nós possui o mesmo número de ligações próximo de um valor médio. Dos dois lados do pico a distribuição diminui exponencialmente, tornando desvios significativos da média extremamente raros (BARABÁSI, 2003). O autor complementa:

A teoria de redes aleatórias de Erdős e Rényi tem dominado o pensamento científico sobre redes desde sua introdução em 1959. Isso criou vários paradigmas que estão, consciente ou inconscientemente, impressos nas mentes dos que lidam com redes. Isso equiparou a complexidade com a aleatoriedade. Se uma rede era muito complexa para ser capturada em termos simples, isso incitava-nos a descrevê-la como aleatória¹³ (BARABÁSI, 2003, p. 23).

Hoje, observando criticamente alguns modelos de rede, relacionados a fenômenos reais, podemos estranhar essa explicação da aleatoriedade do surgimento das redes, pois vemos claramente que deve haver um princípio organizativo que as distingue das redes aleatórias.

¹³ Tradução livre do texto: The random network theory of Erdős and Rényi has dominated scientific thinking about networks since its introduction in 1959. It created several paradigms that are consciously or unconsciously imprinted on the minds of everyone who deals with networks. It equated complexity with randomness. If a network was too complex to be captured in simple terms, it urged us to describe it a random.

3.4 REDES DE MUNDO PEQUENO (*SMALL WORLD*)



O fenômeno *small world* formaliza a lenda segundo a qual você está afastado a somente ‘seis graus de separação’ de qualquer outra pessoa do planeta (WATTS, 1999). Essa lenda, coincidentemente, teve a sua origem na Hungria, a partir de uma história de autoria do escritor Karinthy, em 1929, em que ele traçava uma rede de relacionamentos em que estavam ligados entre si, por exemplo, ele próprio, o rei Gustavo e Henri Ford, através de conhecidos intermediários, que ele preconizou serem da ordem de 6 (BARABÁSI, 2003).

Essa história foi redescoberta por Stanley Milgram (MILGRAM, 1967), que realizou um experimento com o objetivo de determinar a *distância* entre quaisquer duas pessoas nos Estados Unidos escolhidas aleatoriamente. O experimento consistiu em escolher aleatoriamente duas pessoas, que seriam as destinatárias. A primeira residia em Cambridge e era esposa de um estudante. A segunda era um corretor da bolsa de valores que trabalhava em Boston e morava em Sharon, Massachusetts. Estas pessoas seriam os alvos de cartas enviadas a partir das cidades de Wichita no estado de Kansas e Omaha, no estado de Nebraska.

As cartas enviadas continham uma breve explicação do objetivo da pesquisa, o nome, endereço e uma fotografia da pessoa a ser alcançada, e instruções que esclareciam que se a pessoa conhecesse pessoalmente a pessoa da carta deveria reenviá-la diretamente para ela no endereço indicado. Porém, caso não a conhecesse, deveria reenviar a carta para uma pessoa conhecida sua que considerasse pudesse conhecê-la. Em pouco tempo, 46 das 160 cartas enviadas chegaram a seu destino, exibindo um valor médio de 5.5 intermediários, coincidentemente muito próximo do número encontrado por Karinthy em sua história fictícia (BARABÁSI, 2003).

Quem nunca esteve conversando com alguma pessoa desconhecida, que encontrou em um lugar público, e após alguns minutos de conversa descobriu que possuem um amigo em comum? Nestes momentos, costumamos exclamar: “Que mundo pequeno!”. Esta é a origem do nome para o fenômeno que foi estudado por Duncan Watts¹⁴, em relações sociais, e que, após a formalização feita por ele, tornou-se um dos principais parâmetros de caracterização de redes complexas, associado às descobertas de Albert e Barabási (2002) sobre a identificação de redes de escala livre.

Algumas aplicações de pesquisas que podemos citar são (1) emergência e evolução de comportamento cooperativo em grandes organizações, cuja natureza estrutural permite variações; (2) qualquer tipo de propagação, desde os vírus de computador até doenças sexualmente transmissíveis; (3) o processamento da informação em redes espacialmente estendidas e conectadas irregularmente tais como o cérebro humano (WATTS, 1999; BUCHANAN, 2002; BARABÁSI, 2003).

Um bom exemplo contemporâneo de aplicação de pesquisa sobre redes sociais, usando o conceito de *small world*, é o sistema denominado Orkut, de propriedade da Google, que pretende construir a rede de relacionamentos se não de todas as pessoas do planeta, pelo menos daquelas que têm acesso ao computador e que representavam menos de 1% entre os países latino-americanos e não mais do 20% da população das 10 maiores economias mundiais em 1999 (TAKAHASHI, 2000).

A teoria de Watts diz que uma rede pode ser representada por nada mais do que as conexões existentes entre seus elementos, tratando todas as conexões como iguais e simétricas. A pesquisa de Watts sobre o fenômeno *small world* não se prontifica a responder quais as características que uma rede deve possuir a fim de exibí-lo, e sim determinar se este fenômeno ocorre em algum ponto de transição entre a ordem e a desordem, e em que tipo de sistemas ocorre, determinando se o fenômeno não tem nenhuma correlação com o modelo usado para criar a rede. Ou seja, se essa característica é do sistema e não do modelo utilizado para análise.

Para efeito do estudo sobre redes de mundo pequeno (*small world*), Watts (1999) fez algumas restrições sobre que tipo de grafos poderiam ser analisados, com precisão dos

¹⁴ Físico formado pela University of New South Wales, na Austrália, e atualmente professor do Departamento de Sociologia da Columbia University nos EUA.

resultados referentes ao fenômeno *small world*. Os grafos devem apresentar as seguintes características:

- **Não direcionado** – as ligações entre os elementos do grafo não devem possuir um direcionamento, implicando que qualquer desses relacionamentos será representado como simétrico;
- **Não ponderado** – às ligações não são atribuídas nenhum peso *a priori*. Portanto, qualquer importância que algumas ligações possam ter em relação a outras não será levada em consideração;
- **Simples** – múltiplas ligações entre os mesmos pares de vértices, ou ligações que unam um vértice a ele mesmo serão desconsideradas;
- **Disperso** – para um grafo não direcionado, o tamanho máximo (M) de $E(G) = n(n-1)/2$, corresponde a um grafo totalmente conectado ou completo. A dispersão implica que $M \ll n(n-1)/2$;
- **Conectado** – qualquer vértice pode ser alcançado por qualquer outro vértice, cruzando um caminho composto de um número finito de ligações.

Uma rede de mundo pequeno apresenta um valor de CMM pequeno e um elevado CAM, podendo apresentar porém distribuição de frequência num traçado Normal, que caracteriza redes aleatórias. Isso porque a rede de mundo pequeno é uma rede intermediária entre uma rede aleatória e uma rede regular.

Watts (1999) propôs um modelo de construção de redes de mundo pequeno onde, a partir de uma rede regular¹⁵, fosse escolhido, aleatoriamente, algum nó e também de forma aleatória se definisse a qual outro nó da rede este se conectaria. Se continuamente isso for feito com todos os nós da rede, surgirá uma rede completamente aleatória (Figura 14).

¹⁵ Rede cujos vértices possuem o mesmo número de ligações.

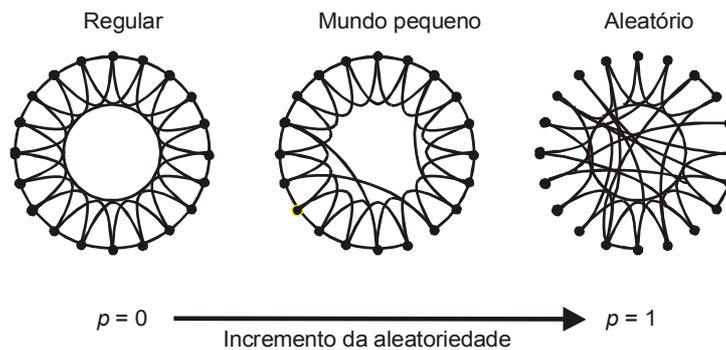
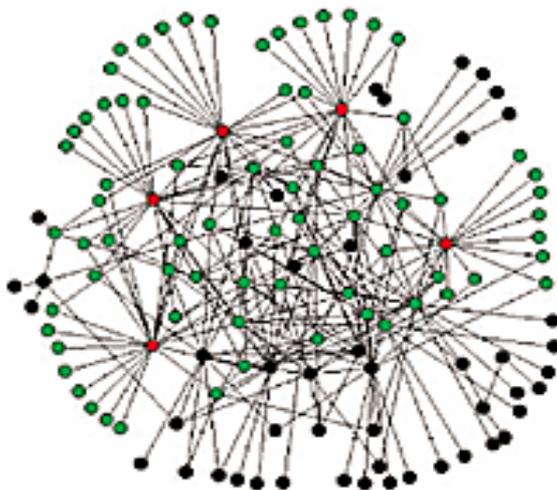


Figura 14 – Diagrama ilustrativo do modelo de Watts de embaralhamento de redes regulares e o surgimento de redes SW e aleatórias (Fonte: Watts, 1999, p. 68).

3.5 REDES LIVRES DE ESCALA (*SCALE FREE*)



Os diversos tipos de redes, sejam naturais, tecnológicas ou sociais, apresentam uma característica em comum, importante para explicar comportamentos como rápida difusão de informação por ela, resistência à falha em caso de quebra de grande parte de sua estrutura, ou até mesmo vulnerabilidade a ataques coordenados.

Trata-se da existência de um número relativamente pequeno de nós que apresentam um grande número de conexões a outros nós. Esses nós altamente conectados são chamados de conectores, e essas redes são denominadas de livres de escala, pois não apresentam um valor padrão médio do número de conexões (k), ou seja, não exibem uma escala característica.

Ao contrário do modelo proposto por Erdős-Renyi, essas redes não podem ser consideradas como aleatórias. São também chamadas de redes democráticas, já que, em média, cada vértice apresenta número de ligações iguais, apresentando uma distribuição de Poisson. As redes livres de escala, que são construídas acrescentando aleatoriamente os novos nós aos já existentes, podem ser chamadas de não democrática já que os nós não terão o mesmo número de links, seguindo uma distribuição em lei de potência.

Em 1999, os pesquisadores Réka Albert, Albert Barabási e Hawonng Jeong (ALBERT et al, 1999), realizaram o mapeamento da rede de ligações das diversas páginas web, a fim de detectar a topologia da mesma. Qual não foi a surpresa desses pesquisadores, quando ao invés de encontrarem uma distribuição de Poisson, como supunha o modelo utilizado para explicar as redes naquela época, encontraram uma distribuição de graus dos nós como uma lei de potência, onde mais de 80% das páginas tinham pouco mais de 4 links, e uma pequena percentagem destas (cerca de 0,01%) tinham mais de 1.000 links.

O modelo de Erdős-Renyi não foi capaz de explicar o comportamento dessa rede analisada porque a forma como pressupõe seu crescimento difere do que ocorre no mundo real em dois pontos. O primeiro é que a construção da rede aleatória inicia já com o total de nós da rede final, pressupondo um crescimento médio gradativo. Na prática, o que temos são redes que iniciam com pouquíssimos nós, e que de um momento para o outro sofrem um crescimento acelerado, alcançando um número extraordinário de nós.

O segundo ponto de divergência do modelo de Erdős-Renyi são as justificativas para os novos nós “escolherem” a que outros nós ele “deseja” se ligar. No modelo da teoria dos grafos todos os nós são basicamente iguais, logo, não ocorrem preferências de escolhas. Na prática, o que vemos é uma ligação preferencial, onde, no processo de crescimento da rede, há sempre uma preferência dos novos nós a se ligarem aos nós mais antigos que tenham mais links. Talvez por serem mais famosos, no caso dos atores de Hollywood ou dos cientistas de uma área de investigação, ou por serem mais fáceis de serem encontrados pelos internautas, como no caso dos sites Google e Yahoo, ou até mesmo para que tenham o acesso facilitado em nossa memória, no caso das palavras lexicais que dão sentido a um texto escrito.

3.6 LINGUAGEM E SISTEMAS COMPLEXOS

Tomando como base o conceito de sistemas complexos anteriormente citado (NUSSENZVEIG, 2003), podemos definir o uso da linguagem humana como tal, já que um texto escrito é formado por um grande número de unidades (palavras) que interagem entre si em pequenos grupos, através das construções sintáticas (frases). O comportamento dessas unidades, em suas inter-relações, faz surgir propriedades emergentes (propriedades de mundo pequeno e livre de escala), como visto nas Seções 3.4 e 3.5.

Um dos mais simples experimentos a serem feitos, para definir se um sistema é complexo ou não, é a determinação da distribuição de frequência dos seus elementos, com o intuito de identificar se essa distribuição é normal ou apresenta uma lei de potência. Um dos primeiros cientistas a detectar esse comportamento em sistemas complexos foi o lingüista George K. Zipf.

- **Lei de Zipf**

Zipf (1972) realizou estudos sobre a linguagem humana, considerando-a como um conjunto de ferramentas, especificamente o vocabulário. Ele fez uma analogia, conjecturando que usamos a linguagem para interagir com o mundo externo, conseguindo obter aquilo de que necessitamos, como o fazemos com qualquer ferramenta que criamos como extensão de nosso corpo. Em seu estudo ele faz uma distinção entre o uso da palavra e a comunicação. Para ele, o estudo sobre as palavras oferece uma chave para o entendimento do processo completo da linguagem, enquanto que o estudo do processo da linguagem abrirá caminhos para o entendimento da personalidade e da dinâmica biosocial.

Para ele, os seres humanos ao utilizarem a fala, o fazem para conseguir algo. Desta maneira, pode-se considerar a linguagem como um conjunto de ferramentas que são utilizadas para alcançar objetivos. A partir dessa constatação, a abordagem sobre o uso das ferramentas e o princípio do menor esforço pode ser aplicado também à fala humana.

Consideramos nesse trabalho apenas as palavras lexicais (as que despertam em nossa mente uma representação, seja de seres, de ações, de qualidade de seres ou modos de ações – ver 1.3). Então podemos supor que existe uma maneira de combinar essas palavras com seus significados de uma forma mais ou menos econômica, tanto do ponto de vista de quem fala quanto do ponto de vista de quem escuta. Do ponto de vista de quem fala, que terá o trabalho de selecionar não somente o significado a ser transmitido mas também a palavra que será usada para representá-lo, sem dúvida alguma seria conveniente se pudesse usar apenas uma palavra que significasse tudo o que desejasse dizer. Logo, se tivesse que transmitir m significados, esta palavra teria m significados ligados a ela. Dessa maneira, o falante pouparia o esforço de adquirir e armazenar um grande vocabulário e selecionar palavras particulares ao desejar a comunicação.

Mas, do ponto de vista do ouvinte, um vocabulário de uma única palavra iria representar o ápice do trabalho verbal, já que teria que determinar o significado da palavra em cada contexto que fosse usada. Dessa maneira, concluímos que do ponto de vista do ouvinte, o ideal seria um vocabulário extenso, com cada palavra tendo apenas um significado, sem ambigüidades. Ele conclui que a fala então seria o resultado dessas duas forças opostas, e chamou-as de força de unificação e força de diversificação respectivamente.

Um dos estudos empíricos de Zipf foi a contagem das palavras do livro *Ulysses*, de James Joyce, num total de 260.430 palavras diferentes. Em seguida, traçou o gráfico correlacionando o rank e a frequência das mesmas, encontrando uma lei de potência (Figura 15).

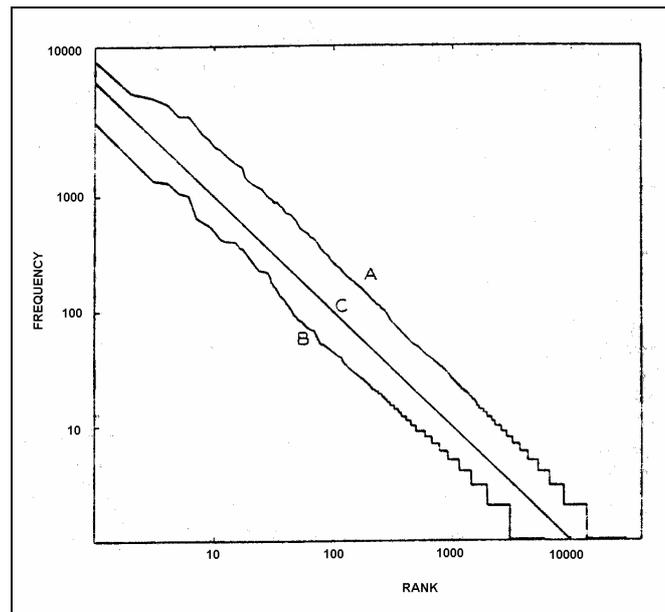


Figura 15 – A distribuição rank-freqüência de palavras. (A) Análise do livro de James Joyce; (B) Análise de jornais americanos feita por Eldridge; (C) Curva ideal com decaimento negativo (Fonte: Zipf, 1972, p. 25).

Nesse trabalho, Zipf (1972) demonstra, a partir das duas curvas oriundas de análises de dados reais, o quão próximas estão da curva ideal com decaimento de 45 graus. Zipf usou um livro denso, em que o autor inventa novas palavras, e recortes de jornais, representando a linguagem mais coloquial. Ambas as produções apresentaram características similares quanto à freqüência das palavras.

Além desse trabalho com palavras, Zipf estendeu suas observações a outros comportamentos humanos, como, por exemplo, o processamento simbólico da mente, a geografia das cidades em relação à densidade populacional, além de conflitos e interações entre nações. Sempre usando como base explicativa a sua lei do menor esforço.

Hoje, sabe-se que em sistemas que apresentam comportamento de freqüência como lei de potência, como por exemplo no caso específico da lei de Zipf aplicada às palavras, essa característica pode ser um prévio indicativo de sua complexidade.

3.7 ESTUDOS DA LINGUAGEM SOB A ÓTICA DE SISTEMAS COMPLEXOS

O estudo da linguagem como uma rede complexa já apresenta um grande volume de experimentos, abordando a questão da linguagem sobre diferentes perspectivas, que levou à construção das redes seguindo várias metodologias. Citaremos alguns trabalhos que tivemos oportunidade de ler durante nossa pesquisa, buscando um relacionamento entre estes e a pesquisa realizada, apontando os coincidentes e os divergentes entre eles.

O trabalho de Motter et al (2002) consistiu em correlacionar as palavras que aparecem em um dicionário com 30.000 entradas, com uma média de 100 palavras sinônimas para cada entrada. As palavras foram conectadas a partir dessa referência cruzada. A rede resultante dessas conexões apresentou propriedades de mundo pequeno e de livre escala, pois apresentou valor de Coeficiente de Aglomeração elevado ($CAM = 0,53$) e valor de Caminho Mínimo Médio baixo ($CMM = 3.16$), e a distribuição de probabilidade $P(k)$ tem um traçado exponencial.

Em seu trabalho, os autores argumentam que o caráter associativo da linguagem, que muito bem representaram construindo a rede a partir de um dicionário de sinônimos, está diretamente associado à forma de representação dos conceitos no cérebro humano. Como se neste existisse uma relação de um para um entre os conceitos, ou seja, pressupõem que nós armazenamos o dicionário em nosso cérebro e as palavras estão diretamente relacionadas entre si.

Essa abordagem difere da fundamentação teórica de nossa pesquisa, pois, segundo a teoria de Freud não há uma relação direta, de um para um, entre os conceitos (palavras), outrossim, passam pelas representações-objeto, que podem ser de diversas naturezas (e.g. imagens, sons, cheiros, enfim, resíduos de memória), até alcançar um outro conceito (palavra). Dessa maneira, entendemos que esse trabalho explica o caráter da linguagem, mas não os processos cognitivos de uso da linguagem. Citando Saussure (2004), esse estudo é sobre a linguagem, e não sobre a fala.

Pesquisas realizadas por Cancho e Solé (2003, 2001) abordaram a linguagem humana pelo ponto de vista da frequência de palavras e levaram em consideração a co-ocorrência delas dentro de sentenças. Na pesquisa da co-ocorrência, porém, somente consideraram as palavras que estavam a uma distância de até duas palavras quando

definiram as ligações entre elas, desprezando todas as co-ocorrências das sentenças com distância maior. Eles chamam essa co-ocorrência até duas palavras de co-ocorrência significativa. A pesquisa foi referente à análise de 10.000.000 de palavras do British National Corpus¹⁶.

Uma outra hipótese que levantaram é a existência de dois tipos de vocabulário: (1) o básico e (2) o especializado. A partir da análise das redes construídas foi identificada a existência de dois conjuntos distintos de palavras chamadas de dicionário básico e dicionário especializado. Os valores encontrados para os parâmetros de análise de rede, para cada uma delas foram: Rede do vocabulário básico ($CAM_1 = 0.687$ e $CMM_1 = 3.03$) e a rede do vocabulário especializado ($CAM_2 = 0.437$ e $CMM_2 = 3,06$), confirmando as características de mundo pequeno desta rede. Já as distribuições de conectividade foram $\gamma_1 = -1.50$ e $\gamma_2 = -2.70$, que levam à afirmação da existência de propriedades de livre escala.

Em 2004, os dois pesquisadores unem-se a Köhler (CANCHO et al., 2004) e publicam o trabalho *Patterns in syntactic dependency networks*, em que realizam pesquisa similar para as línguas checa, alemã e romena.

Há uma divergência metodológica entre o trabalho realizado por Cancho e Solé (2003, 2001) e a nossa pesquisa, pois consideramos que todas as palavras que estão presentes na sentença possuem uma correlação, já que entendemos que ao construir a sentença desejamos expressar um pensamento que está contido nela. Consideramos todas as palavras de uma sentença para correlacionar.

Wentian Li publicou dois trabalhos fazendo uma análise crítica sobre a importância da lei de Zipf como um índice de organização de sistemas complexos. No primeiro (LI, 1992) ele construiu textos aleatórios e provou que também estes apresentam uma lei de potência, com expoente próximo de 1. O segundo trabalho trata de uma prova matemática de que a lei de zipf pode ser considerada mais como uma lei de transformação do que uma lei de sistemas complexos (LI, 1996).

¹⁶ **Corpus lingüístico** é um conjunto de textos escritos ou falados numa língua, disponível para análise. O plural de corpus é corpora. Os corpora nos permitem examinar um vasto material produzido espontaneamente na fala ou na escrita das pessoas, fazendo observações precisas sobre o real comportamento lingüístico de gente real. Portanto os corpora nos proporcionam informações altamente confiáveis e isentas de opiniões e de julgamentos prévios, sobre os fatos de uma língua (Fonte: Wikipedia: http://pt.wikipedia.org/wiki/Corpus_ling%C3%BC%C3%ADstico).

O psicólogo Herbert A Simon (1955) fez um estudo sobre as classes de função de distribuição, e uma de suas pesquisas foi realizada com exemplos de textos, especificamente sobre a distribuição de palavras por sua frequência de ocorrência. Do nosso ponto de vista, esse trabalho é muito interessante, pois traz referências às diferentes maneiras de se tratar textos escritos para análise. O autor cita outras pesquisas como a de Yule (1944) e a de Thorndike (1937) (apud SIMON, 1955), que reduziram as formas flexionadas das palavras às suas formas canônicas antes de realizar a contagem, sendo que o primeiro só trabalhou com os substantivos.

Além dessa referência ao tratamento prévio que os textos recebem, Simon também cita exemplos de pesquisadores que trabalharam com uma grande amostra composta de textos individuais reunidos, como no caso da pesquisa de Dewey, Eldrige e Thorndike (1937) (apud SIMON, 1955), indicando também que outros trabalharam com a análise de textos individuais.

Por conta dessa diversidade de tratamentos, Simon faz uma crítica à interpretação dos resultados encontrados, e logo em seguida levanta uma suspeita de que pode estar enganado:

Eu consideraria essa heterogeneidade como uma evidência maior de que a explicação deve ser procurada em um mecanismo da probabilidade, do que em características mais específicas da língua; mas, ao mesmo tempo, a heterogeneidade complica a tarefa de especificar o mecanismo de probabilidade em detalhes¹⁷ (SIMON, 1955, p. 433).

Simon propõe um modelo estocástico para a explicação da distribuição da frequência. Se esse modelo descreve corretamente a seleção de palavras, então as palavras em um livro não podem ser consideradas como um exemplo aleatório tirado de uma população com uma distribuição anterior.

Analisando um único texto, o *Ulysses* de James Joyce, Simon postula que as suposições anteriores podem explicar o processo de crescimento do mesmo, porém coloca objeções a essa possibilidade:

um autor escreve não somente pelo processo de associação – isto é, fazendo uma amostragem dos segmentos das seqüências de palavras que escreveu

anteriormente – mas também pelo processo de imitação – isto é, fazendo uma amostragem dos segmentos de seqüências de palavras de outros trabalhos que ele tenha escrito, de trabalhos de outros autores, e, é claro, de seqüências que ele tenha ouvido. O modelo anterior aparentemente permite somente as associações, e exclui as imitações (1955, p. 434).

Marcelo Montemurro (2001) propõe uma revisão sobre a utilização da lei de Zipf para diferentes tamanhos de amostra, pois, segundo o autor, a lei de Zipf somente pode descrever o comportamento de amostras relativamente pequenas, e não se aplica à análise de grandes *corpora*. Na análise desenvolvida por Montemurro foram encontrados dois valores de α , sendo 1.05 e 2.3.

Montemurro e Pury (2002) fizeram uma análise da estrutura fractal de grande quantidade de textos, mapeando-os como uma série temporal.

Dorogovtsev e Mendes (2001) propõem um modelo teórico para explicar o desenvolvimento da linguagem humana baseado no conceito de ligações preferenciais, tomando como referência os trabalhos de Herbert Simon e Barabási. Eles encontraram dois expoentes da lei de potência diferentes para distribuição de conectividade: 1.5 e 3.

O artigo de Luciano da F. Costa (COSTA, 2004) sugere que a linguagem pode ser um reflexo do processo associativo que ocorre em nosso cérebro. A investigação consistiu em um procedimento psicofísico, em que uma palavra é apresentada ao indivíduo, e a este é solicitado que associe uma outra palavra àquela que está sendo mostrada. Após o recolhimento dos dados foi montada a rede fazendo um grafo direcionado, com a frequência de cada associação tratada como o peso da respectiva ligação.

Somente a primeira palavra do experimento é sugerida. As outras, que serão usadas para solicitar associação, serão as próprias palavras que o indivíduo irá falando ao longo do processo. A investigação durou uma semana, e foi feita com um único indivíduo, quando foram apresentadas 305 palavras diferentes, e um total de 1930 associações foram registradas. Apesar do experimento ter sido realizado com um único sujeito, o autor indica uma série de características interessantes que foram determinadas com o estudo, por exemplo, a identificação de algumas características

¹⁷ Tradução livre do excerto: I would regard this heterogeneity as further evidence that the explanation is to be sought in a probability mechanism, rather than in more specific properties of language; but at

de dinâmicas caóticas, características de rede de mundo pequeno, além da lei de potência para a distribuição de graus das palavras correlacionadas.

Uma aplicação chamada Automap[©], desenvolvida por Lewis, Diesner e Carley (LEWIS et al, 2001), faz um tratamento lingüístico interessante, pois de forma automática e *online* oferece vários parâmetros estatísticos do texto fornecido para análise, inclusive fazendo um pré-tratamento do mesmo, reduzindo as palavras à sua forma canônica. O objetivo dessa aplicação é construir um mapa de conceitos, que iremos chamar de primeira ordem. Ou seja, é feito um agrupamento das palavras em torno de macro-conceitos.

Outro exemplo de aplicação de lingüística computacional é o trabalho de Koppel, Argamon e Shimoni (KOPPEL et al, 2003), em que um grande *corpus* é analisado e determina-se exatamente o que difere no estilo da escrita de homens e mulheres. Segundo os pesquisadores, foi identificado que os homens utilizam mais substantivos para construir os seus textos, enquanto que as mulheres fazem maior uso de negações, pronomes e algumas preposições. Essa pesquisa provocou a nossa decisão de trabalhar individualmente com os textos, pois entendemos que existem características individuais que podem emergir dessa maneira, ao invés de tratar os textos como um grande conjunto e tentar definir padrões globais.

CAPÍTULO IV – DESENVOLVIMENTO DA PESQUISA

Nessa parte do trabalho será feita uma descrição detalhada de todas as suas etapas, que abrangem uma discussão sobre o modelo escolhido para a pesquisa, a amostragem e seus critérios, os métodos, até o cálculo dos parâmetros analisados.

No intuito de caracterizar o sistema complexo da linguagem, nos fundamentamos em dois mecanismos: a estrutura sintática e a escolha de palavras. Foi analisado também a interferência que esses mecanismos provocam nos sistemas reais dos textos. Para isso, lançamos mão de métodos da física, mais precisamente aqueles referentes à caracterização dos sistemas complexos, para analisar o comportamento de textos escritos, tratando-os como redes complexas e definindo os valores de parâmetros que as caracterizem.

Ao demonstrar, com riqueza de detalhes, o funcionamento do aparelho psíquico, explicando como distintos eventos e parâmetros poderiam exhibir padrões emergentes, como a lembrança e os processos de pensamento, e que esse funcionamento baseava-se em processos associativos, Freud descrevia a caracterização de um *sistema complexo* (Seção 0) como passamos a defini-lo nos anos 1990, na *Teoria dos sistemas complexos*¹⁸.

Os textos existem como uma realidade incontestável, produzidos pela mente humana, capazes de gerar significações, base da transmissão das idéias. Estudar as conexões das palavras nos fornece indícios de como são construídas e como se apresenta um sistema com essas características.

O modelo computacional definido para representar a construção dos textos foi o de redes complexas. À medida que construímos um pensamento a ser transcrito para o papel ou para a tela do computador, escolhemos palavras, que se ligam às anteriormente escolhidas, formando uma rede, como reflexo da construção de significados que queremos transmitir. Como a intenção do trabalho é determinar a rede de significados, foram excluídas as palavras gramaticais (Seção 1.3).

¹⁸ Esta área [teoria] teve seu alicerce na Teoria do Caos e na área dos Fenômenos Não-Lineares, sendo fortemente estimulada pelo uso de métodos computacionais que se propagaram rapidamente nos últimos 20 anos (<http://twiki.im.ufba.br/bin/view/Fractais/Historico>).

Partindo do processo associativo entre as representações, proposto por Freud, identificamos que as representações-palavra emergem à consciência periodicamente, como um mecanismo de descarga da $Q\eta$, permitindo conduzir a catexia (investimento) de ψ para as lembranças que surgem durante a passagem desta $Q\eta$. Dessa maneira, podemos também inferir que a rede formada pelas palavras, quando escritas num texto, pode representar a teia de interconexões entre as imagens mnêmicas gravadas no córtex cerebral, e invocadas durante o processo de pensamento, seja na construção de textos técnicos ou literários.

4.1 QUESTIONAMENTOS A SEREM RESPONDIDOS

Após a análise das redes construídas a partir dos textos escritos, deveremos ser capazes de responder aos seguintes questionamentos:

1. Ao escrevermos um texto, escolhemos aleatoriamente as palavras?

Como se dá a combinação das palavras num texto escrito? Aleatoriamente escolhemos as palavras que desejamos combinar num significado maior (o texto) ou existe uma estrutura semi-ordenada ou ordenada que podemos identificar ao analisar esses textos escritos? Após identificar se há um tipo de rede característica dos textos escritos, carece definir se esse comportamento é comum entre vários autores, comparando os resultados entre textos de estilos diferentes, textos escritos por autor de gêneros diferentes e se o tamanho dos textos influencia no comportamento analisado.

2. A dinâmica de construção também pode ser comum em diferentes textos?

Ao considerarmos o texto como um sistema, podemos inferir que ele passa por um processo de crescimento a cada vez que acrescentamos uma nova sentença a ele. Na resposta ao primeiro questionamento analisamos o texto como um sistema acabado, completo e estático. Na análise da dinâmica de construção dos textos, por sua vez, determinaremos as características do sistema a cada ponto de seu crescimento, ou seja, a cada nova inserção de uma quantidade pré-definida de sentenças. Se houver uma forma característica de crescimento, será definido se este é identificado também em textos de diferentes autores.

3. Qual a característica do sistema formado pelas palavras do texto responsável pelo seu comportamento? Seria a estrutura sintática (sentenças e S-V-O)? Ou seria a escolha das palavras?

A linguagem humana é sintática, combinando os sinais com significados próprios para construir um novo objeto com significado diferenciado das partes (NOWAK, 2000). Esse mecanismo de construção dos textos pode ser o fator que transmite ao sistema do texto a sua característica que iremos detectar nos questionamentos 1 e 2. A fim de realizar a identificação do que contribui para seu comportamento, serão executados processos de simulação baseados tanto na sintaxe quanto no mecanismo de escolha das palavras.

4.2 MÉTODO DA PESQUISA

Na Tabela 2, encontra-se um resumo das etapas desenvolvidas para a caracterização dos textos e os recursos empregados em cada uma delas.

Tabela 2 - Recursos utilizados na pesquisa

Etapas	Recursos
Amostragem	-
Limpeza e nomeação dos textos	Manual
Pré-tratamento	Programa UNITEX (modificado) e programa AMBISIN
Construção da rede estática e dinâmica	Programa NetPal
Cálculo dos Índices	Programa NetAll
Visualização das Redes	Software Pajek
Apresentação e análise dos resultados	Software Origin

Os textos escolhidos foram submetidos a um pré-tratamento lingüístico, e em seguida foi construída a rede de associações entre as palavras, conforme ilustra a Figura 16.

Esse método de construção das redes foi aplicado tanto para as redes de textos completos (rede estática), quanto na simulação da escrita de um texto (rede dinâmica).

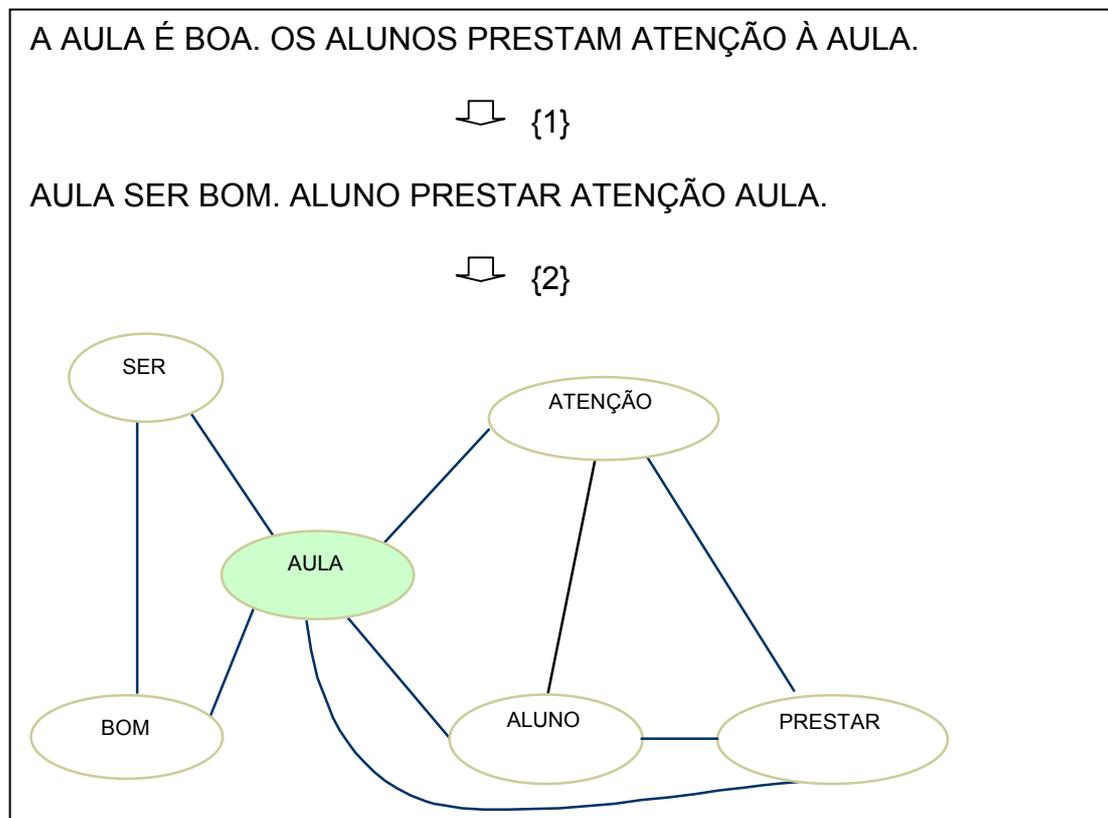


Figura 16 - Exemplo de uma construção de rede de palavras

{1} – Pré-tratamento e análise do texto. Dessa maneira, temos a Representatividade das idéias que o autor está expressando.

{2} Construção da Rede.

As redes, por sua vez, foram criadas no formato PAJEK, um programa para análise e visualização de grandes redes, desenvolvido por Vladimir Batagelj e Andrej Mrvar, da University of Ljubljana, na Slovenia¹⁹. Trata-se também de um formato livre, que nos permitiu tanto a construção quanto a análise posterior das redes pelos programas *NetPal* e *NetAll*.

¹⁹ <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

4.3 AMOSTRAGEM

Os textos foram coletados na Internet em sites oficiais de armazenamento de publicações científicas, no caso de textos técnicos, e em sites que disponibilizam livros eletrônicos gratuitos nos textos literários. No total, foram analisados 312 textos. A preocupação maior era manter quantidades equilibradas entre as diversas opções de classificação, conforme Figura 17. As classes selecionadas foram o estilo dos textos, o gênero do autor, o idioma e o tamanho em sentenças. A quantidade total de textos foi definida de forma a manter o equilíbrio entre o número de textos dentro de cada classe, levando em consideração a disponibilidade dos mesmos na Internet.

Foram escolhidos quatro, dentre os textos analisados, para servirem de referência no estudo do crescimento, pois trata-se dos maiores livros usados do nosso espaço amostral. O Ulysses, de James Joyce, em inglês e português, o livro de León Tolstói, Anna Karenina, e o Código de Da Vinci, de Dan Brown, em sua versão em português.

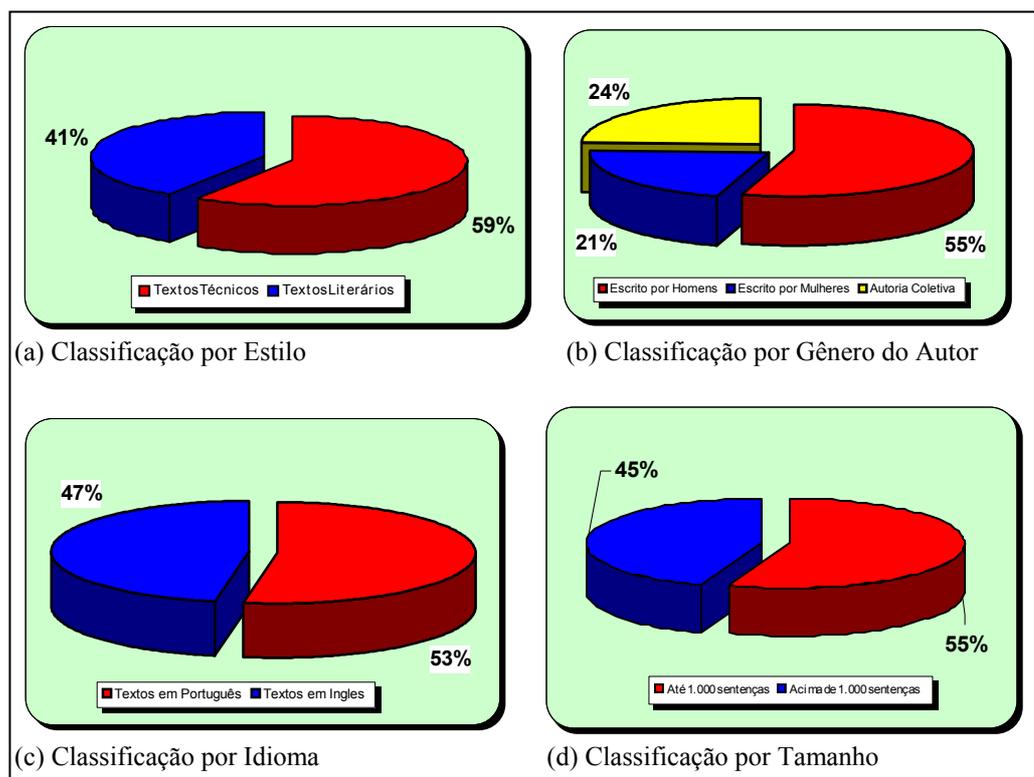


Figura 17 – Composição da Amostra de Textos

No Apêndice A.1 encontra-se uma tabela com a composição dos textos e suas classificações.

4.4 IDENTIFICAÇÃO DOS TEXTOS

Uma forma de identificação rápida que permitisse visualizar algumas características do texto somente pelo nome foi estabelecida. As Tabelas 3, 4 e 5 mostram a simbologia usada para definir as 3 (três) primeiras letras do nome do arquivo:

Tabela 3 – Significado da primeira letra do nome do arquivo: gênero do autor.

Caractere	Descrição
M	Masculino
F	Feminino
Z	Texto coletivo ou de autoria desconhecida

Tabela 4 – Significado da segunda letra do nome do arquivo: tipo geral do texto.

Caractere	Descrição
A	Acadêmico
P	Textos poéticos
C	Contos e Crônicas
L	Livros
J	Texto jornalístico
D	Jurídico

Tabela 5 – Significado da terceira letra do nome do arquivo: tipo específico do texto.

Caracteres		Descrição
<i>2ª letra</i>	<i>3ª letra</i>	Significado
A	A	Monografia
	B	Dissertação
	C	Tese
	D	Artigo
	L	Projetos e relatório de investigação
P	E	Letra de música
	F	Poema
C	Z	Todos os estilos
L	T	Técnico
	K	Romance
J	M	Matéria de revista científica

4.5 LIMPEZA DOS TEXTOS

A hipótese de rede de conceitos tornou necessária a eliminação das partes iniciais do livro ou artigo ou trabalho acadêmico que não estivessem diretamente relacionados com o desenvolvimento do assunto. As partes eliminadas foram: sumário, resumo, índice, lista de figuras e tabelas, prefácio escrito por outros, referência bibliográfica, notas de rodapé e legendas. Desta maneira, garantimos a análise somente da parte do texto construída para desenvolver o raciocínio do autor.

4.6 PRÉ-TRATAMENTO DOS TEXTOS

Para construir a rede de conceitos a analisar, foi feito um tratamento prévio nos textos, a fim de gerar uma amostra das palavras lexicais que o autor(a) foi escolhendo com a finalidade de transmitir a sua idéia. Esse tratamento abrange aspectos puramente técnicos, advindos de teorias lingüísticas. Para tal, utilizamos como base o conjunto de programas vindos no pacote UNITEX, disponibilizado pela Rede Relex Brasil, que faz parte de um programa francês do LADL²⁰, cujas características estão descritas na Figura 18.

O UNITEX (Manual UNITEX, 2002), inicialmente chamado de INTEX, foi desenvolvido originalmente em 1968, quando Maurice Gross fundou o LADL que se transformaria no embrião para o UNITEX e para a Rede Relex. Desde 1970, o LADL desenvolve pesquisas lingüísticas com enfoque nas relações entre léxico e sintaxe. Em 2002, a equipe do LADL foi incorporada ao *Institut d'électronique et d'informatique Gaspard-Monge* (IGM - Paris). O trabalho de Maurice Gross foi inspirado nas pesquisas de seus contemporâneos Noam Chomsky e M. P. Schützenberger, quando ele utiliza-se do princípio segundo o qual a unidade de significado é uma frase elementar, para apoiar a sua teoria léxico-gramatical (1975). Desde então o programa vem sofrendo alterações e acréscimos lexicais a fim de construir um grande léxico, se não de todas as línguas, pelo menos das mais importantes faladas atualmente. E, para isso, tem o apoio da Rede Relex, consórcio de Laboratórios Lingüísticos com estudos referentes a nove idiomas, a saber: inglês, francês, português de Portugal, português do Brasil, grego, italiano, russo, espanhol e tailandês.

Figura 18 – Panorama geral do programa de pesquisa lingüístico do LADL.

²⁰ Laboratoire d'Automatique Documentaire et Linguistique.

O pré-tratamento foi realizado a partir de alterações no programa UNITEX, pois ele segue a filosofia GPL, disponibilizando os códigos do programa para quem desejar utilizá-lo de forma livre. Além disso construímos o programas *Ambisin* para tratar questões pertinentes à nossa metodologia, como a eliminação das ambigüidades, eliminação das palavras gramaticais e separação das formas flexionada ou canônica das palavras do restante dos itens de classificação gramatical gerados pelo UNITEX.

Criamos um programa de lotes, onde cada grupo de linhas de comando corresponde aos programas utilizados em cada etapa do pré-tratamento dos textos, como mostra a Figura 19.

Criação de pastas e subpastas para organizar os diversos arquivos gerados	<code>mkdir Cres_%1 cd Cres_%1 mkdir %1_snt</code>
Conversão do texto ASCII para Unicode, [e transferência do arquivo gerado para a subpasta criada]	<code>..\asc2uni PORTUGUESE ..%1 move ..%1.uni \</code>
Normalização de Separadores [elimina tabulação, retorno de linha (enter) e espaço excedentes]	<code>..\normalize %1.uni</code>
Segmentação em sentenças, [usando arquivo <i>Alphabet</i> correspondente ao idioma do texto]	<code>..\Fst2Txt.exe %1.snt ..\Sentence.fst2 ..\Alphabet.txt -merge</code>
Segmentação em unidades lexicais	<code>..\Tokenize.exe %1.snt ..\Alphabet.txt</code>
Normalização das formas não-ambíguas [e aplicação dos dicionários]	<code>..\dico %1.snt ..\Alphabet.txt ..\Dela\Delaf_pb.bin</code>
Conversão do texto Unicode para ASCII	<code>..\uni2asc PORTUGUESE %1_snt\dlf</code>
Eliminação das ambigüidades, [escolha das palavras que comporão a rede e eliminação das palavras gramaticais]	<code>..\Ambisin %1_snt\dlf.ascii %1_snt\dlf.txt 2</code>

Figura 19 – Diagrama do pré-tratamento dos textos e Linhas do código do arquivo BAT usado para chamar os programas respectivos.

A seguir discutiremos cada etapa do diagrama de pré-tratamento dos textos.

Criação de pastas e subpastas para organizar os diversos arquivos gerados

O processo de pré-tratamento dos textos gera vários arquivos. Para permitir uma maior organização, foram criadas pastas para cada texto analisado.

Conversão de textos ASCII para Unicode²¹

O texto a ser analisado foi gravado no formato TXT para a primeira etapa do pré-tratamento. Essa conversão é uma exigência puramente técnica, necessária para a utilização do programa escolhido para esse tratamento: o UNITEX.

Normalização de Separadores

Eliminação dos separadores usuais excedentes, como a tabulação, retorno de linha (enter) e espaço. Trata-se de um outro requisito técnico do sistema escolhido, e propõe-se a preparar o texto para que seja facilmente tratado pelo algoritmo de identificação das palavras e da segmentação das sentenças.

Segmentação em Sentenças

Permite a definição de unidades de processamento lingüístico. Nesse processo é acrescido o símbolo {S} ao final de cada sentença.

Este é um requisito metodológico do nosso estudo da rede dos textos escritos, pois entendemos que a sentença é a menor unidade para análise dos significados expressos nos textos. Cada palavra isoladamente pode adquirir um significado que somente será identificado a partir do contexto. Esse contexto, para a nossa pesquisa é a sentença em que a palavra participa.

Segmentação em unidades lexicais

Coloca a unidade lexical (em sua maioria as palavras), uma em cada linha do arquivo gerado. Essa nova ordenação do texto é necessária para que seja feita a análise de cada palavra individualmente em relação à classe gramatical a que pertence, além de permitir a geração de informações estatísticas sobre o mesmo. Tomando como

²¹ Unicode é um padrão que descreve uma codificação universal dos caracteres. A cada caractere é atribuído um número único, o que possibilita representar textos independentemente das codificações

exemplo um pequeno texto formado por duas sentenças, podemos ver o resultado na Figura 20.

Ex.: **Chove lá fora e meu coração transborda. A alegria me contagia.**

Chove

lá

fora

e

meu

coração

transborda

.

A

alegria

me

contagia

.

Índice	O	1	2	1	3	1	4	1	5	1	6	1	7	1	8	1	9	1	10	1	11	1	12	1	8
	Chove		lá		fora		e		meu		coração		transborda		.		A		alegria		me		contagia		.

Figura 20 – Exemplo de segmentação do texto em unidades lexicais.

Normalização das formas não ambíguas

Esse processo decompõe as palavras derivadas em suas formas originais. (e.g.: A palavra DESSE será decomposta em DE+ESSE).

O aspecto analítico dessa etapa é que dá condição posterior para que as palavras analisadas sejam etiquetadas com a classe gramatical a que pertencem.

especificas a várias máquinas ou sistemas operacionais. Unitex usa um código com 2 bytes do padrão 3.0, chamado Little-Indian (para mais detalhes, veja <http://www.unicode.org>).

- **Aplicação dos dicionários**

Nesta etapa foi feita a análise de cada palavra, com base no léxico (dicionário) da língua escolhida, armazenado no programa UNITEX. A partir desse dicionário, identifica-se a forma canônica correspondente à unidade lexical analisada, além da(s) classificação(ões) gramatical(is). Observe na Figura 21 um exemplo de etiquetagem típica de palavras constantes em uma sentença. Note o símbolo {S} do final da sentença. Após a palavra transcrita como encontrada no texto, é colocada a sua forma canônica, seguida da classificação gramatical sugerida pelo programa.

```
{S}
modelos,modelo.N:mp:fp
lineares,linear.A:mp:fp
e,.CONJ
modelos,modelo.N:mp:fp
hierárquicos,hierárquico.A:mp
parsing,.NOTFOUND
{S}
```

Figura 21 – Exemplo de etiquetagem em uma sentença do texto.

Uma modificação no programa original do UNITEX foi necessária, pois as palavras não registradas no dicionário eram retiradas do arquivo de saída. A partir dessa modificação, essas palavras passaram a constar na lista geral das palavras (arquivo dlf.txt), etiquetadas com a palavra 'NOTFOUND' (Figura 21).

chove,chover.V:P3s:Y2s
 lá,.ADV
 fora,.ADV
 fora,.N:ms
 fora,.PREP
 fora,ir.V:Q1s:Q3s
 fora,ser.V:Q1s:Q3s
 e,.CONJ
 e,.N:ms
 meu,.N:ms
 meu,eu.PRO+Pos:1ms
 coração,.N:ms:fs
 transborda,transbordar.V:P3s:Y2s
 {S}
 a,.ABREV:ms
 a,.N:ms
 a,.PREP
 a,ele.PRO+Pes:A3fs
 a,o.DET+Art+Def:fs
 a,o.PRO+Dem:fs
 alegria,.N:fs
 me,.ABREV:fs
 me,eu.PRO+Pes:O1ms:A1ms:D1ms:R1ms:O1fs:A1fs:D1fs:R1fs
 contagia,contagiar.V:P3s:Y2s
 {S}
 a,.N:ms
 a,.PREP
 a,ele.PRO+Pes:A3fs
 a,o.DET+Art+Def:fs
 a,o.PRO+Dem:fs

Figura 22 – Exemplo de todas as possibilidades de etiquetas que as palavras do texto analisado anteriormente podem receber, após aplicação do dicionário.

Na Figura 21 temos o exemplo de saída das palavras analisadas no item anterior, composta da palavra como se encontra no texto, seguida de sua forma canônica e de sua classificação gramatical.

No exemplo da Figura 22, vale ressaltar a incidência de palavras com várias classificações gramaticais, cujo tratamento será abordado ainda nesta seção.

A seguir iremos demonstrar os tipos de etiquetas que o programa UNITEX possui no dicionário PORTUGUÊS, mostrando as características do dicionário lexical do UNITEX.

- **Dicionários do UNITEX**

Os dicionários fornecidos com o UNITEX contêm as descrições das palavras simples e compostas. Essas descrições indicam a categoria gramatical de cada entrada, eventualmente seus códigos de flexões, assim como as informações semânticas diversas. Nesta pesquisa não iremos trabalhar com o dicionário de palavras compostas, portanto o assunto não fará parte desta explicação.

Os dicionários eletrônicos do UNITEX usam o formalismo DELA – Dicionários Eletrônicos do LADL (REDE RELEX BRASIL, 2002), que possibilita descrever as entradas lexicais de uma língua associando opcionalmente informações gramaticais, semânticas ou flexionais²².

A Tabela 6 mostra os códigos gramaticais usados nos dicionários do UNITEX para identificar a análise gramatical e que foram utilizados para tratar o problema da ambigüidade.

Tabela 6 - Códigos gramaticais usuais

Código	Significado	Exemplos
A	Adjetivo	Fabuloso
ADV	Advérbio	Ontem, de repente
N	Substantivo	Mesa, bolsa de valores
V	Verbo	Cantar, ver, remediar
CONJC	Conjunção de coordenação	Mas, porém, contudo
CONJS	Conjunção de subordinação	Embora, conquanto
DET	Determinante	Uma, seus, vinte
INTJ	Interjeição	Tchau, ei!
PREP	Preposição	Sem, à margem de
PRO	Pronome	Ela, a gente

Conversão do texto Unicode para ASCII

Essa conversão deve ser feita, pois desse ponto em diante será utilizado um programa diferente do UNITEX, desenvolvido para essa pesquisa, que recebe como entrada um arquivo em formato ASCII. Essa conversão é feita pelo programa UNI2ASC que faz parte do pacote UNITEX (REDE RELEX BRASIL, 2002).

²² Para mais detalhes, sugerimos consultar o manual do UNITEX.

Eliminação das ambigüidades

No exemplo da Figura 22, vemos que algumas palavras do texto original, após a etiquetagem, deram origem a várias linhas indicando as possíveis classificações gramaticais que a palavra pode receber. A essas possíveis classificações, dá-se o nome de ambigüidade, um dos problemas mais controversos da análise lingüística de textos, em qualquer idioma.

Em paralelo, nesta etapa, foi feita a escolha das palavras finais levadas em consideração para a construção da rede. Foram criadas 4 opções, a fim de gerar dados comparativos para assegurar a adequação da nossa escolha de trabalhar com as palavras lexicais apenas, e de reduzi-las à sua forma canônica.

- **Escolha das palavras que compuseram a rede**

A metodologia proposta nesta dissertação, para a análise da rede de conceitos presente no texto analisado, é de trabalhar com as palavras em sua forma canônica, eliminando as palavras gramaticais, ou seja, as palavras que servem apenas para criar a estrutura lingüística do texto. Porém, com a finalidade de fundamentar essa escolha fizemos uma série de testes com os textos, conforme a Tabela 7. O número especificado nesta tabela, fornecido como parâmetro do programa AMBISIN, foi escolhido conforme o objetivo da análise a ser feita. Caso não seja passado nenhum parâmetro, o programa assume como escolha a terceira opção (parâmetro 2).

Tabela 7 – Parâmetro fornecido ao programa ‘AMBISIN’ para determinar o tipo de análise.

Parâmetro	Forma de Tratamento	Descrição
0	Texto sem tratamento: Tratamento 0 (original)	Mantém flexões e não elimina palavras gramaticais.
1	Texto Tratado: Tratamento 1	Reduz palavras à forma canônica e não elimina as palavras gramaticais
2	Texto Tratado: Tratamento 2	Reduz palavras à forma canônica e elimina as palavras gramaticais
3	Não foi utilizado nessa pesquisa	Mantém flexões e elimina palavras gramaticais

- **Eliminação das palavras gramaticais**

Caso o tratamento do texto requerido seja referente ao parâmetro 2 ou 3, será necessário eliminar as palavras gramaticais, eliminar as ambigüidades e ao final escolher entre usar a palavra original, flexionada, ou construir a rede com a forma canônica dessas palavras.

```

// Escolhe a ambigüidade correta
char ResolveAmbi(TUniLex *Amb, char fim)
{
    int i=0;
    if(AbLe==3 || AbLe==2)
    {
        for(i=0;i<=fim;i++)
        {
            if(!strcmp(Amb[i].Gra,"PRO",3))return -1;
            if(!strcmp(Amb[i].Gra,"DET",3))return -1;
            if(!strcmp(Amb[i].Gra,"PREP",4))return -1;
            if(!strcmp(Amb[i].Gra,"ABREV",5))return -1;
            if(!strcmp(Amb[i].Gra,"INTJ",4))return -1;
            if(!strcmp(Amb[i].Gra,"CONJ",4))return -1;
        }
    }

    if(fim==0) return 0; // se resta apenas uma retorna esta

    for(i=0;i<=fim;i++) // Caso contrário resolve ambigüidade
    {
        if(Amb[i].Gra[0]=='N')return i; // Escolhe o substantivo
        if(Amb[i].Gra[0]=='V')return i; // Escolhe o Verbo
    }
    return 0;
}

```

Figura 23 – Algoritmo de eliminação das ambigüidades.

Não foi objeto de estudo da pesquisa a resolução desse problema. Optamos por realizar uma eliminação sumária das palavras que, por ventura, não se enquadravam na classificação de ‘palavra lexical’, ou seja, as palavras que constituiriam a nossa representatividade dos conceitos escolhidos pelo autor para transmitir o seu pensamento.

O algoritmo apresentado na Figura 23, na etapa (a) e (b), Caso o parâmetro fornecido ao AMBISIN tenha sido o parâmetro 2 ou 3, a escolha das palavras gramaticais e eliminação das ambigüidades ocorrem da maneira indicada. Todas as palavras que foram classificadas como Pronomes (PRO), Artigos (DET), Preposições (PREP), Abreviações (ABREV), Interjeições (INTJ) e Conjunções (CONJ) sejam

definitivamente eliminadas do arquivo das palavras lexicais do texto, independentemente de terem um comportamento ambíguo ou não.

Após essa etapa, se somente restar uma palavra, esta será assumida como a correta, independente de sua classificação gramatical (*c*). Caso contrário, pelo estudo visual dos textos, notamos que o restante das palavras que continuavam na lista estariam envolvidas em ambigüidades em que prevaleceria a condição de classificação do substantivo (N) sob as palavras etiquetadas como Adjetivos (A) ou Verbos (V). Por isso, a próxima etapa (*d*) manterá as palavras com a etiqueta do Substantivo, descartando as restantes.

Ainda restariam as palavras com ambigüidades entre os Advérbios (A) e Verbos (V). Neste caso, prevalecerá, segundo a observação dos textos que fizemos, as palavras classificadas como Verbos (V), sendo as palavras ambíguas classificadas como Adjetivos (A) eliminadas da listagem, feito também na etapa (*d*). Por último, os Advérbios (ADV) quase nunca estavam envolvidos em situações de ambigüidades, podendo permanecer na listagem.

Porém, se o parâmetro passado ao AMBISIN tiver sido o parâmetro 0 ou 1, significa que a escolha seria a de não eliminar nenhuma palavra, e isso dificulta a eliminação de ambigüidades, já que está foi baseada na possibilidade de eliminação. Neste caso, fizemos uma abstração ainda maior e concluímos que poderíamos escolher da lista qualquer uma das palavras, sem preocupação com a classe gramatical a que pertence.

Na lista restante (dlf.txt) existem, agora, somente as unidades lexicais que nos interessa para compor a rede dos conceitos utilizados para construir o texto. Na próxima etapa do pré-tratamento foi feita a escolha da forma da palavra utilizada para construir a rede, seja forma flexionada, original, ou a forma canônica.

- **Seleção da forma canônica da palavra e de sua classe gramatical**

Nesse ponto, o arquivo das palavras gerado possui todas as palavras ainda flexionadas (Figura 22). Se a escolha de tratamento tiver sido o parâmetro 1 ou 2 (Tabela 7), é preciso extrair, da linha da unidade lexical, a informação da forma canônica dessa palavra. Veja o exemplo da Figura 24. Caso contrário, se a opção passada tiver sido o parâmetro 0 ou 3, a forma flexionada da palavra será a utilizada.

A ocorrência da palavra ‘coração’, no exemplo anterior, irá gerar uma linha de unidade lexical conforme Figura 24 (1), e a informação que precisamos extrair dessa unidade é a sua forma canônica e sua classe gramatical, transformando a linha de unidade lexical para o formato mostrado na Figura 24 (2). A palavra é transformada em maiúscula, pois precisamos uniformizar esse formato para que o tratamento estatístico posterior não sofra distorções.

coração,.N:ms:fs (1)	==> CORAÇÃO N (o N indica que se trata de um substantivo) (2)
-------------------------	--

Figura 24 – Processo de tratamento da unidade lexical para o formato requerido para o próximo programa de análise, o programa NetIPal

Esta etapa do pré-tratamento é realizada pelas linhas de código da Figura 25.

```

// Grava em arquivo a palavra resolvida
void Salva(FILE *Fout,TUniLex *Amb,char Escolha)
{
    if(AbLe==0||AbLe==3)
        fprintf(Fout,"%s\t%s\n",Amb[Escolha].Inf,Amb[Escolha].Gra);           (a)
    else
        fprintf(Fout,"%s\t%s\n",Amb[Escolha].Can,Amb[Escolha].Gra);       (b)
}

```

Figura 25 – Escolha da forma flexionada (a) ou da forma canônica (b) das palavras

4.7 CONSTRUÇÃO DAS REDES

Após a etapa de pré-tratamento, e de posse do conjunto de palavras a ser analisado, a rede foi construída, no formato PAJEK (ver exemplo da Figura 28), e serão calculados os parâmetros que identificam as redes, conforme diagrama da Figura 26.

Constrói a rede a partir do texto tratado	<code>..\NetPal %1 _snt\dlf.txt -%2 %1</code>
Grava a data e hora no arquivo de resumo	<code>date /t >>..\resumo%1 time /t >>..\resumo%1</code>
Coloca o cabeçalho da tabela dos índices	<code>..\NetAll >> ..\resumo%1</code>
Calcula os índices dos textos selecionados	<code>for %%i in (%1*.net) do ..\NetAll %%i.pth -arq %%i -bin >> ..\resumo%1</code>
Grava no arquivo de saída a hora do término	<code>time /t >>..\resumo%1</code>
Retorna à pasta principal	<code>cd..</code>

Figura 26 - Diagrama da construção das redes e linhas de código do arquivo de lote correspondente.

Todas as palavras que co-ocorrem numa sentença estão ligadas, formando cliques²³ individuais. Isso é feito pelo programa *NetPal*. Cliques individuais, de cada sentença são formados, ligando todas as palavras de uma sentença entre si. Porém, caso ocorra uma palavra numa sentença que já tenha sido referenciada em sentença anterior, então não é repetida a palavra e sim, todas as outras da sentença nova serão ligadas à palavra já referenciada no texto. No exemplo da Figura 16, a palavra aula foi repetida, e por isso se tornou o elo de ligação entre as duas sentenças. A essas palavras, dá-se o nome de *hubs* ou conectores, pois, através delas, podemos alcançar as outras palavras do texto, por caminhos mais curtos. Elas aproximam as palavras entre si.

Para a escolha da sentença como unidade de significado do texto, utilizamos como referência o trabalho de Maurice Gross²⁴, onde ele utiliza o princípio segundo o qual a unidade de significado é uma frase elementar, para apoiar a sua teoria léxico-gramatical.

²³ Clique – uma rede ou sub-rede que possui todos os seus vértices interligados entre si. No caso específico de nossa análise, todas as sentenças formam um clique.

²⁴ <http://ladl.univ-mlv.fr/>

Nessa etapa, de construção das redes, é gerado um arquivo com o formato PAJEK, contendo uma listagem de todos os vértices ou nós (palavras) que vão compor a rede, numerados. Logo após essa lista de palavras numeradas, aparecem os relacionamentos representados por pares de números correspondentes às das palavras que estão sendo referenciadas nas sentenças do texto (Figura 27).

```

*Vertices 6
1 "chover"
2 "lá"
3 "coração"
4 "transbordar"
5 "alegria"
6 "contagiar"
*Edges
5 6
6 5
3 4
4 3
2 4
4 2
2 3
3 2
1 4
4 1
1 3
3 1
1 2
2 1

```

Figura 27 - Arquivo texto com a indicação dos pares das palavras

4.8 VISUALIZAÇÃO DAS REDES

Para a visualização da rede foi utilizado o programa PAJEK, que a partir da matriz adjacente representando as co-ocorrências das palavras, gera gráficos como o da Figura 28.

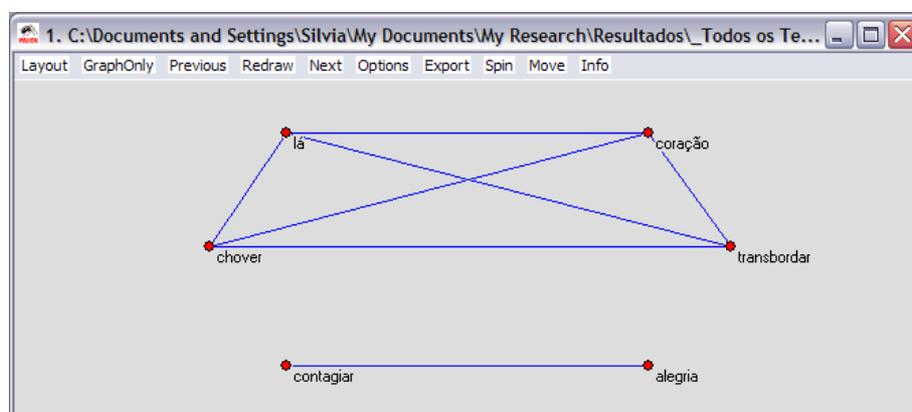


Figura 28 - Rede construída pelo programa PAJEK

4.9 MÉTODOS DE ANÁLISE DAS REDES

O objetivo de caracterização da rede formada a partir dos textos escritos nos levou a determinar os valores dos parâmetros que as classificam. Esses parâmetros foram:

- número total de sentenças (T-sentença);
- número total de vértices (T-vértice);
- número total de arestas (T-aresta);
- diâmetro da rede (DAM);
- coeficiente de aglomeração médio (CAM);
- caminho mínimo médio (CMM);
- grau (k), grau médio ($\langle k \rangle$) e distribuição de graus;
- gama da distribuição de graus (G).

Os índices e seus significados

A seguir são detalhados cada um dos índices de caracterização de redes complexas e seus significados em relação ao objeto de análise, ou seja, os textos escritos.

T-Sentença: quantidade de sentenças, finalizadas por símbolos de final de sentença da língua analisada (Tabela 8), observando que, quando encontrados nas abreviações mais comuns, como nomes de logradouros e pronomes de tratamento, devem ser desconsiderados. Ao final de cada sentença identificada é colocado o símbolo {S} para marcação.

Tabela 8 - Símbolos de final de sentença considerados em textos da língua portuguesa

Símbolo	Nome
.	Ponto
:	dois pontos
?	ponto de interrogação
!	ponto de exclamação
;	ponto e vírgula

T-vértice: quantidade de palavras que representa o vocabulário que o autor utiliza para construir o texto.

Diâmetro (DAM): indica a maior distância mínima entre todas as palavras do texto. Para alcançarmos qualquer palavra, no máximo precisamos de n passos (DAM). Valores médios de diâmetro elevados indicam que esses textos possuem uma maior quantidade de palavras diferentes do que textos que apresentam diâmetro médio menor.

Coefficiente de Aglomeração Médio (CAM): determina o quanto as sentenças possuem dependência entre seus significados. Este índice necessitou de uma análise mais rigorosa de seu significado, pois tem uma influência tanto sobre o modelo estático quanto o dinâmico (crescimento) da rede. Uma sentença única ou sentenças isoladas (i.e. aquelas que não possuem palavras repetidas), e que tenham mais de três palavras em sua composição, apresentam coeficiente de aglomeração igual a 1 (um), como demonstrado no exemplo da Figura 29:

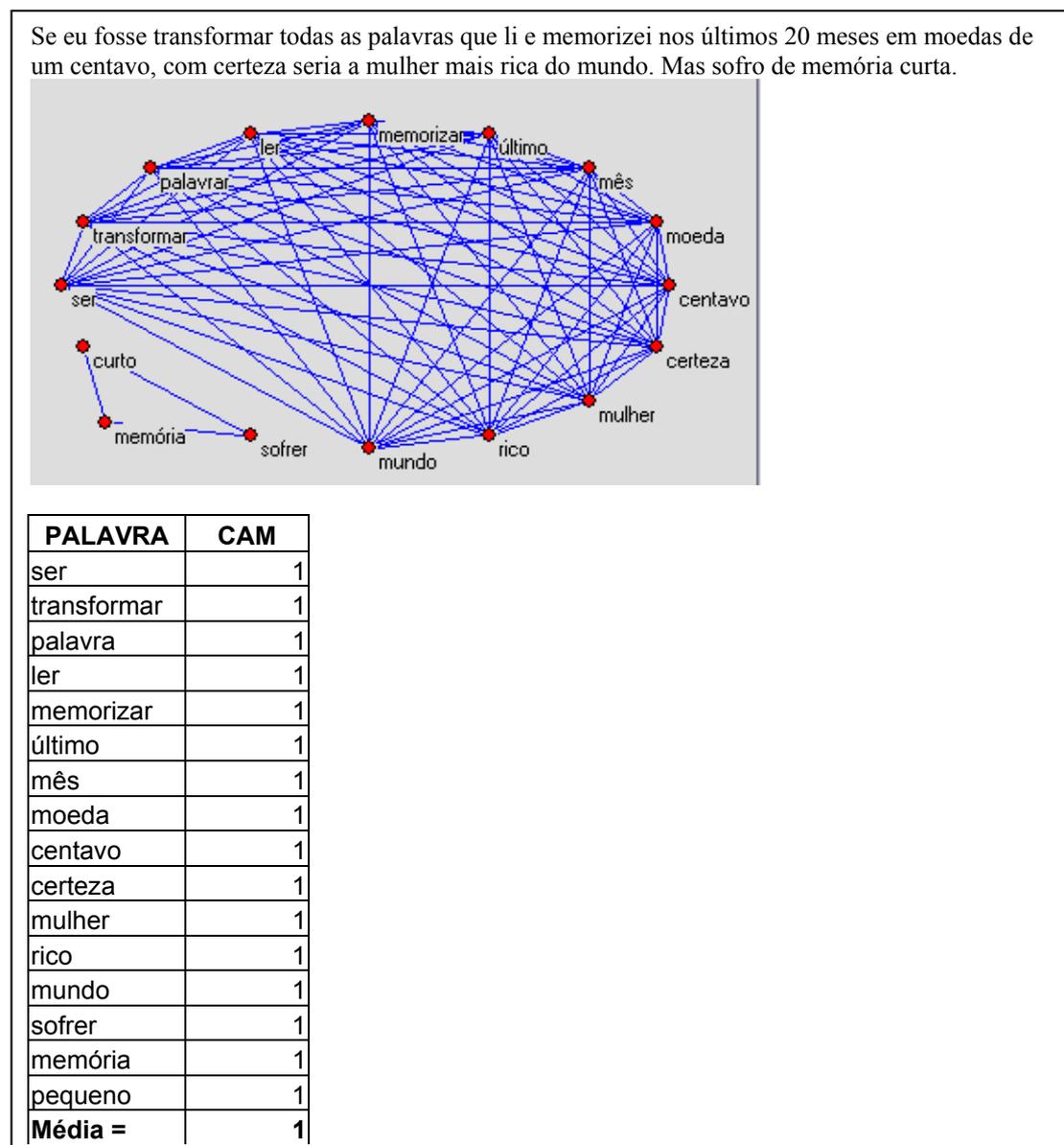


Figura 29 – Exemplo da influência do tamanho e composição da sentença no valor do CAM.

Porém se as sentenças isoladas forem compostas de apenas 2 palavras, o coeficiente de aglomeração será igual a zero, por conta da definição do seu cálculo referir-se ao grau de conexão das palavras vizinhas analisadas. No caso de duas palavras, ao tomarmos uma como referência e determinarmos o coeficiente de aglomeração da segunda, esta não estará conectada a nenhuma outra (Figura 30).

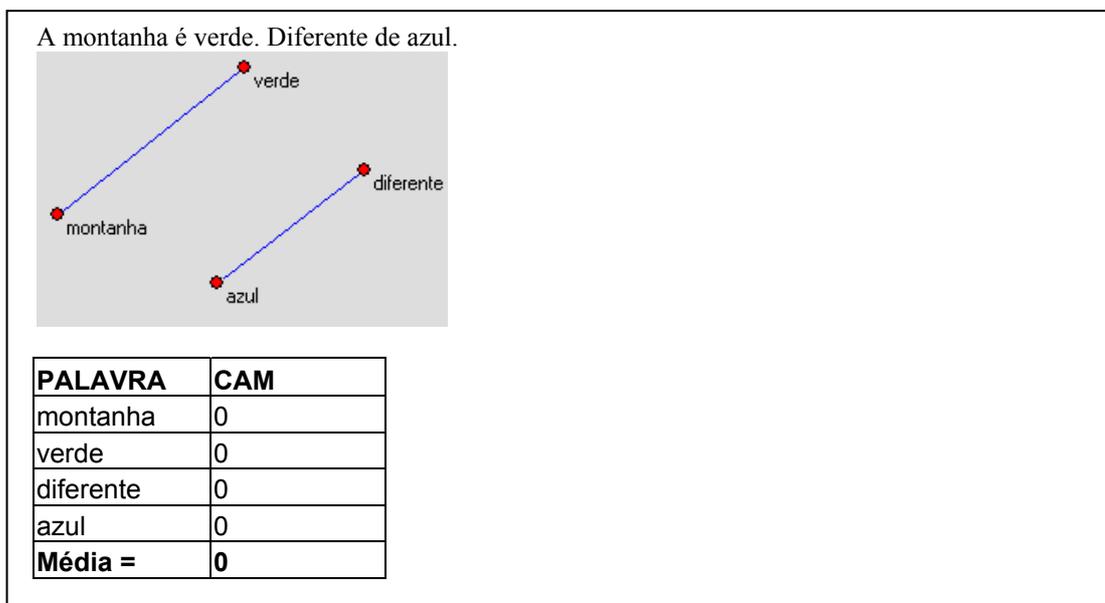


Figura 30 - Exemplo da influência do tamanho e composição da sentença no valor do CAM.

Ao repetirmos palavras em sentenças curtas, essa repetição irá provocar um decréscimo considerável no valor final do coeficiente de aglomeração (Figura 31).

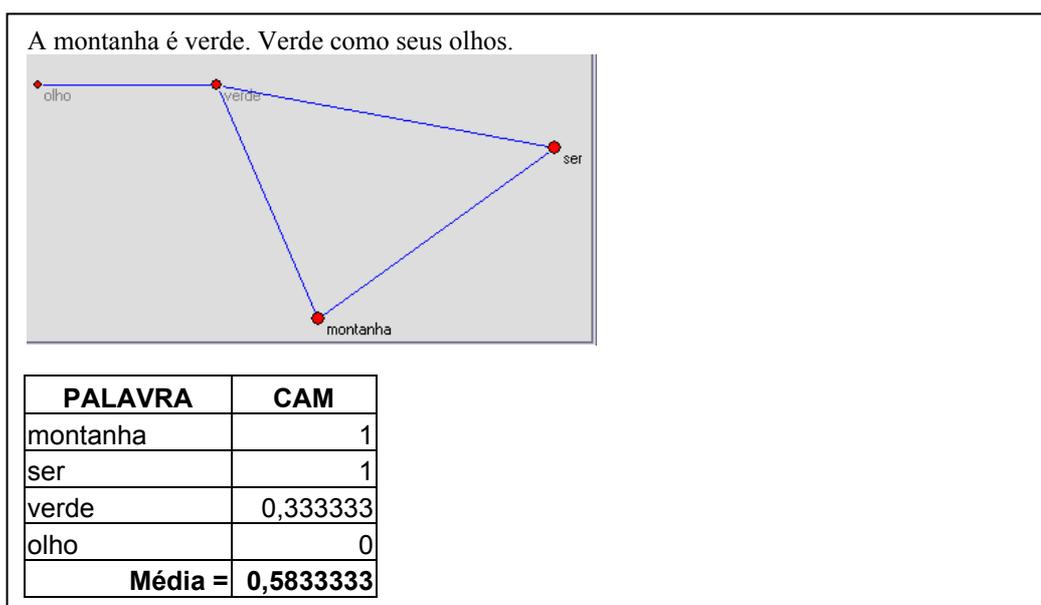


Figura 31 - Exemplo da influência do tamanho e composição da sentença no valor do CAM.

Quando analisamos um texto composto de duas sentenças que possuem palavras comuns, e uma delas é muito maior do que a outra, o coeficiente de aglomeração tende a manter-se próximo de 1 (um), que é o valor do CAM da sentença maior (Figura 32).

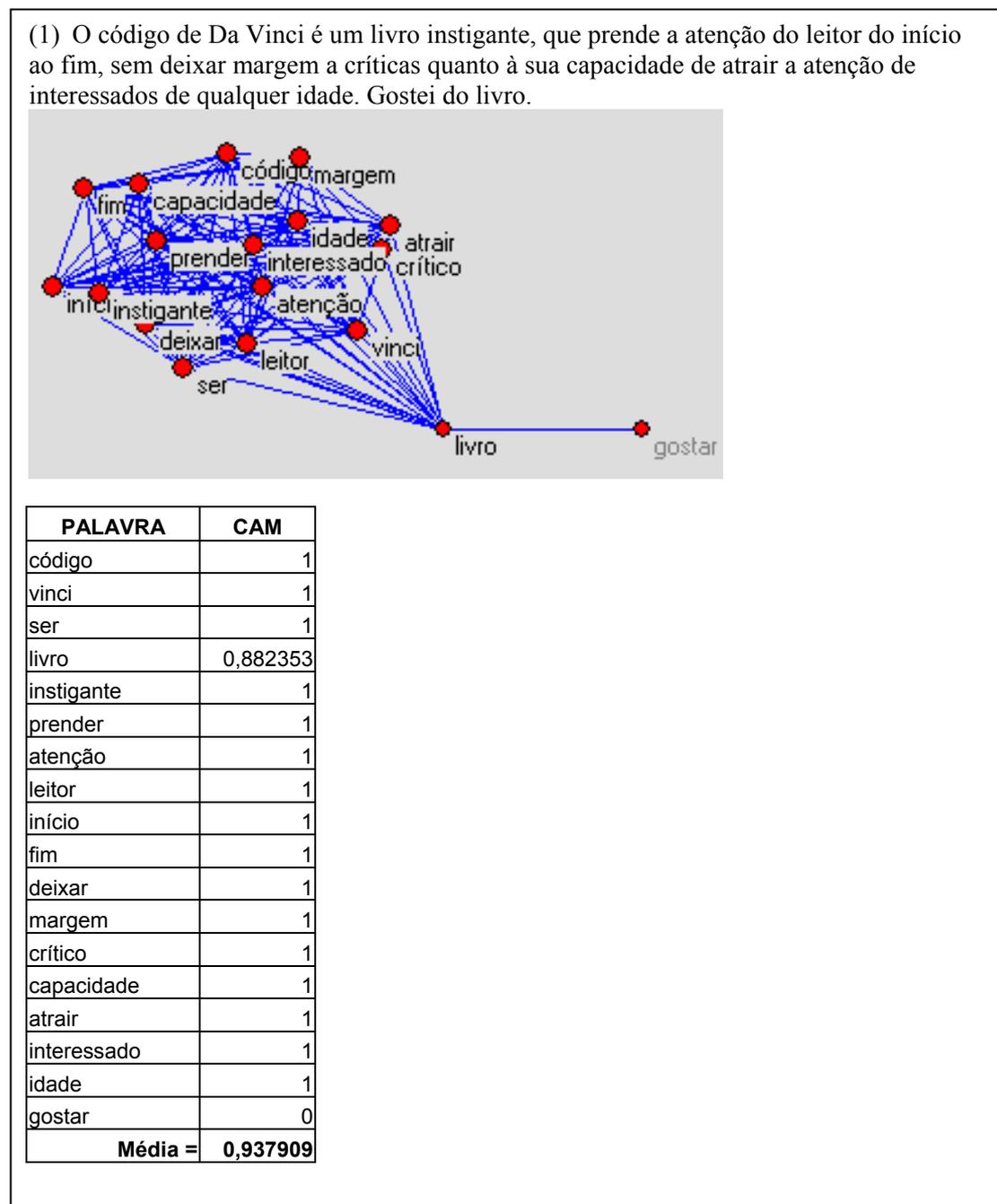


Figura 32 - Exemplo da influência do tamanho e composição da sentença no valor do CAM.

Apesar da palavra repetida ser uma conectora, a influência do tamanho da sentença no valor final do CAM é maior, reduzindo a importância da repetição da palavra no valor final do parâmetro.

Caminho Mínimo Médio (CMM): o caminho mínimo indica a menor distância entre cada uma das palavras do texto e todas as outras. Calculando a média dessas distâncias obtemos o Caminho Mínimo Médio, ou seja, a quantos passos podemos alcançar qualquer palavra do texto a partir de uma qualquer, tomada como referência.

Grau (k), Grau médio ($\langle k \rangle$) e Distribuição de Graus: cada vértice possui um total de arestas, que designa o seu grau. Na Figura 16, por exemplo, a palavra AULA possui ligação com 5 outras palavras do texto, portanto, apresenta grau $k = 5$. No caso de redes aleatórias, a distribuição de graus tem formato de uma distribuição Normal (Figura 33b). Já as redes complexas do tipo livre de escala apresentam distribuição de graus em forma de uma lei de potência (Figura 33a). E as de mundo pequeno podem apresentar tanto uma distribuição em lei de potência quanto uma normal.

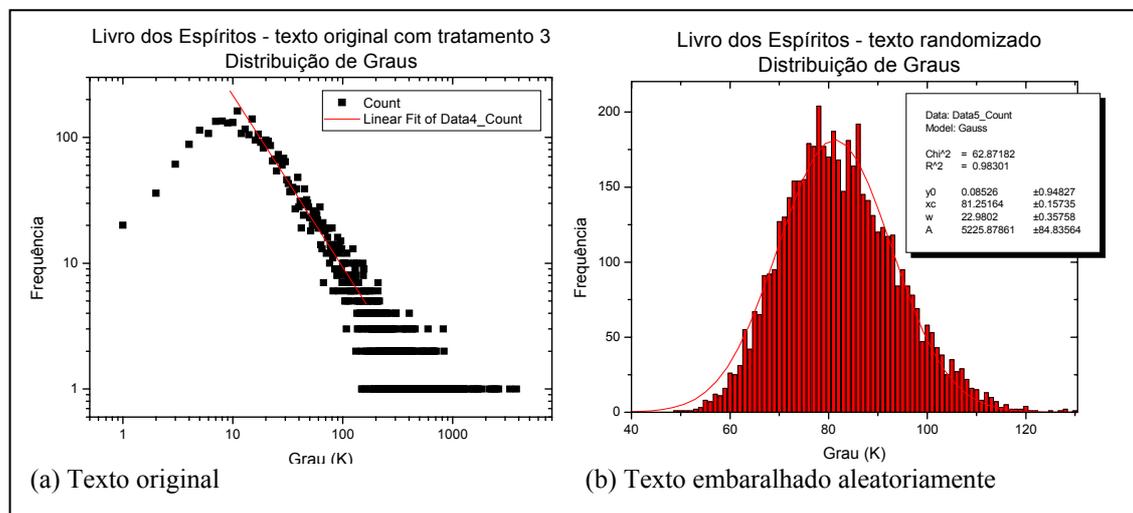


Figura 33 – Comparação do comportamento da Distribuição de Graus do Livro dos Espíritos.

Gama (γ): é determinado, para cada vértice, o número de conexões que este possui com os outros vértices da rede. Dessa contagem da quantidade de ligações calcula-se a frequência com que as ligações presentes na rede ocorrem (Figura 34). Em seguida,

constrói-se o gráfico da frequência, calculando o valor do gama, ou seja, da inclinação da reta ajustada em escala logarítmica (Figura 35).

Vértice	Grau
carro	2
passar	2
veloz	6
percorrer	4
distância	13
dilacera	4
coração	13
teima	10
manter	10
separado	10
deixar	10
alma	10
amante	10
torpor	10
angustiar	10
ansiedade	10

Quantidade de ligações	Frequência
1	0
2	2
3	0
4	2
5	0
6	1
7	0
8	0
9	0
10	9
11	0
12	0
13	2
14	0
15	0

Figura 34 – Exemplo de contagem de ligações (a) e frequência de ligações de uma pequena rede (b). Somente para ilustração, pois não é possível o ajuste de curvas com poucos valores.

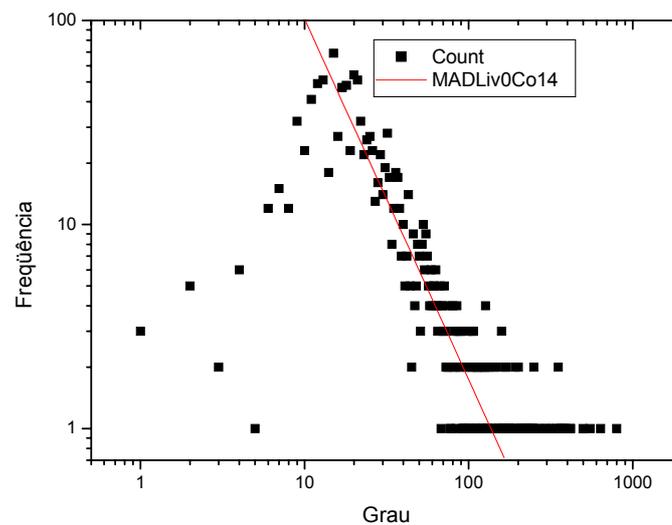


Figura 35 – Exemplo do ajuste Linear da Distribuição de Graus do livro MADLiv006m.

Comparação dos valores absolutos dos parâmetros entre os textos

A primeira análise comparativa entre índices foi feita em relação aos valores dos parâmetros, fazendo segmentação da amostragem (Seção 4.3) com a finalidade de identificar a existência de diferenças significativas por classe de textos, ou se é um comportamento comum às diferentes classes analisadas.

Modelo de Análise do Crescimento de Textos

Com o objetivo de entender a mecânica de crescimento dos textos escritos, fizemos uma simulação onde uma quantidade nova de sentenças era acrescentada ao texto, seguindo um acréscimo de 10% na quantidade de sentenças. Cada acréscimo foi chamado de etapa (Figura 36). A cada etapa foram calculados os parâmetros de caracterização da rede montada, até alcançar o final do texto.

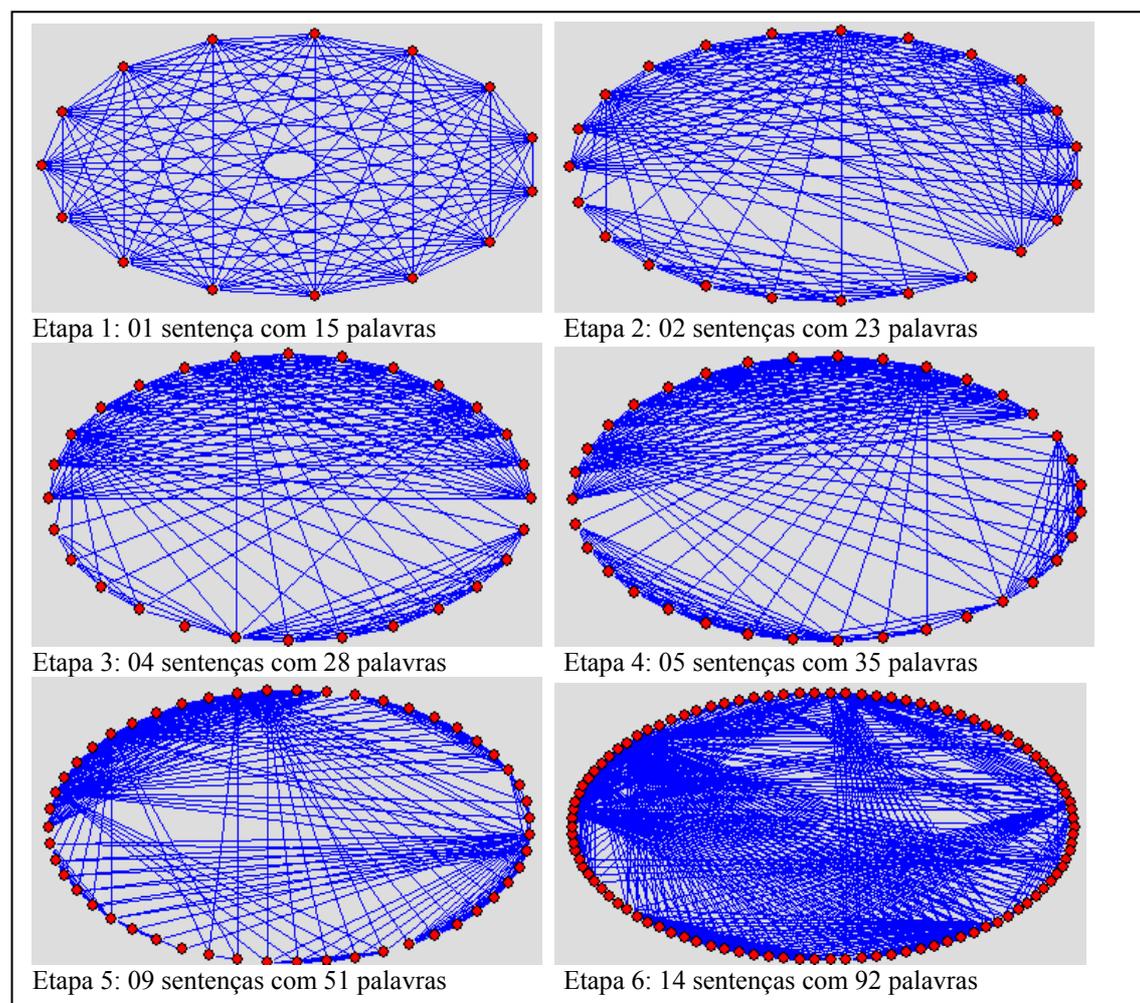


Figura 36 - Exemplo de simulação do crescimento de um texto.

CAPÍTULO V – RESULTADOS

Nesta parte da dissertação, apresentaremos os resultados da análise dos textos escolhidos, que serão apresentados em 3 blocos, que contemplam todos os focos de investigação propostos: Análise dos Textos Completos – denominada de *análise estática* –, análise do Crescimento dos textos e Simulação de construção de textos.

5.1 ANÁLISE DOS TEXTOS COMPLETOS

Serão apresentados os resultados dos parâmetros das redes estáticas, com a finalidade de caracterizá-las, definindo se são redes aleatórias, quase-aleatórias ou determinísticas, ou seja, se são redes aleatórias, redes de mundo pequeno, redes livres de escala ou redes regulares. Antes, porém, analisamos se o tratamento proposto para o texto provoca alterações significativas nos valores analisados.

5.2 COMPARAÇÃO ENTRE AS TÉCNICAS DE TRATAMENTO DOS TEXTOS

A metodologia de tratamento dos textos proposta neste trabalho tem duas finalidades básicas. A primeira é construir uma rede capaz de evidenciar a interligação entre os conceitos que estão sendo evocados pelo autor, a partir das palavras lexicais. A segunda finalidade é construir uma metodologia que otimize o uso dos recursos computacionais, eliminando o máximo possível de palavras processadas, sem no entanto provocar prejuízos significativos na estrutura da rede, no que se refere à rede de significação construída.

Como mostrado na Seção 4.6, foram definidas 4 maneiras de escolha das palavras para compor a rede do texto. Porém, somente realizamos testes com as três primeiras, pois os resultados que foram encontrados já justificavam a não realização dos testes com a quarta maneira proposta.

As escolhas das palavras que aplicamos foram:

- Tratamento 0 - uso das formas flexionadas das palavras, sem eliminar as palavras gramaticais (texto em sua forma original);

- Tratamento 1 - uso das formas canônicas das palavras, sem eliminar as palavras gramaticais;
- Tratamento 2 - uso das formas canônicas das palavras, eliminando as palavras gramaticais.

Comparando, então, os valores do Coeficiente de Aglomeração (CAM), Caminho Mínimo Médio (CMM) e Grau Médio (k), considerados os principais parâmetros de caracterização de redes complexas (ALBERT e BARABÁSI, 2002), verificamos que o tratamento proposto provocava alguma alteração em seus valores, em relação ao texto original, porém mantendo as características globais da rede.

Nos gráficos das figuras 37, 39 e 40 apresentamos as distribuições de frequência para os índices. Nestes gráficos, o eixo x representa o valor do índice e no eixo y a quantidade de textos que assumiram o valor x .

- **Comparação do Coeficiente de Aglomeração Médio**

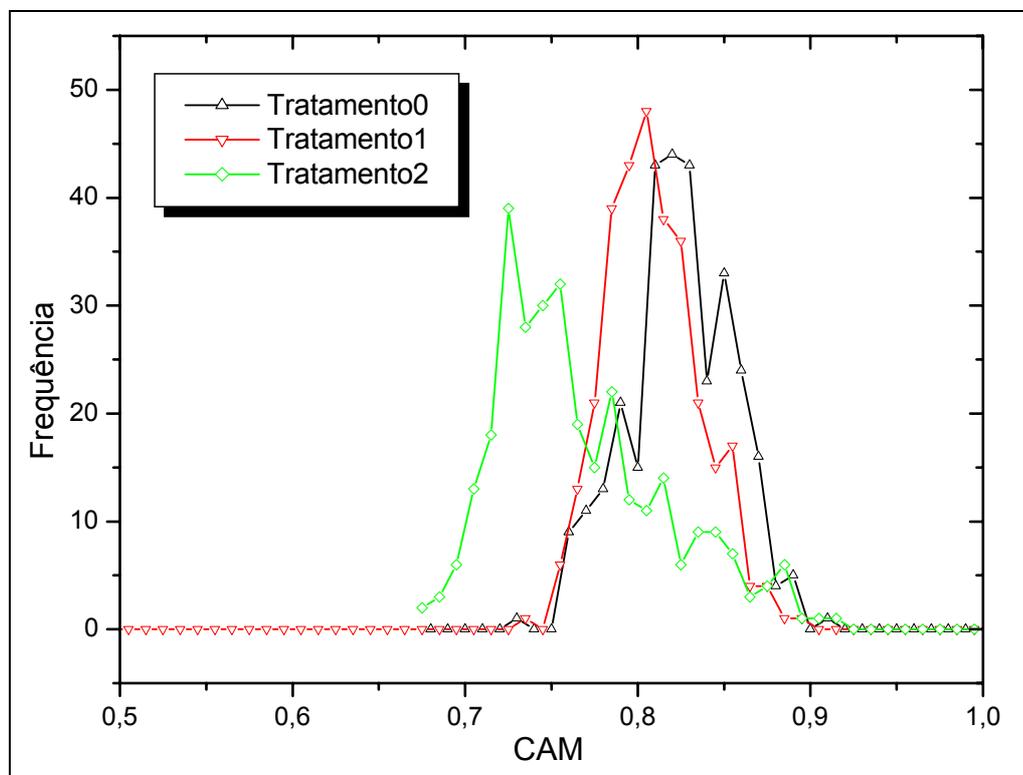


Figura 37 – Comportamento do CAM das redes dos textos submetidos aos tratamentos 0, 1 e 2.

O Coeficiente de Aglomeração Médio da rede de textos sofreu redução ao realizar tanto o Tratamento 1 quanto o Tratamento 2 (Figura 37). Um dos fatores responsáveis por essa diminuição é a redução das palavras à forma canônica, nos dois Tratamentos. Isso porque o número de cliques isolados (que possuem $CAM = 1$) da rede é reduzido com esse tratamento e passa a ter mais palavras conectoras, formando um grande aglomerado, que apresentam valores de CAM menores que 1, reduzindo o valor médio. A Figura 38 traz um exemplo que explica esse comportamento.

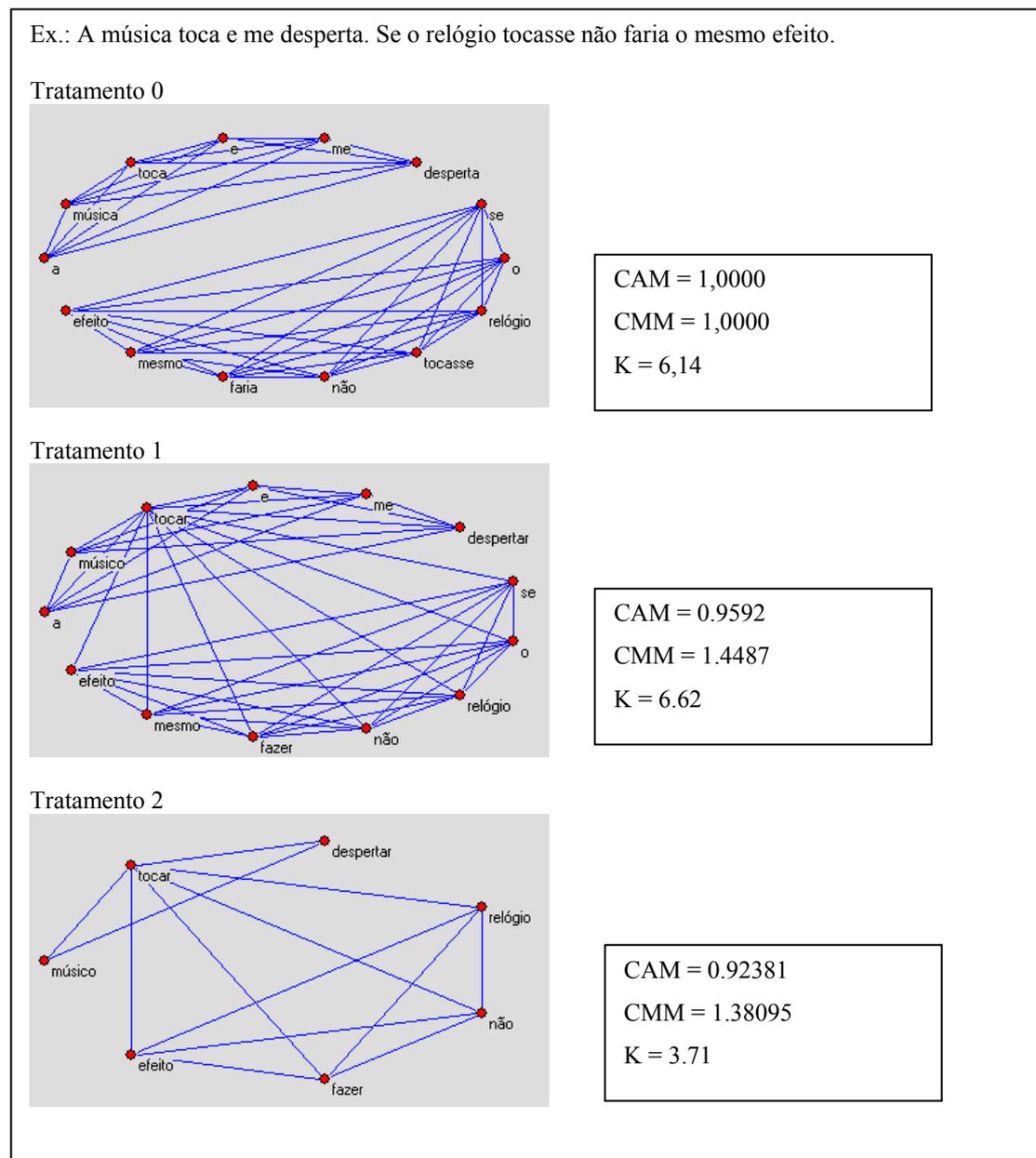


Figura 38 – Exemplo do comportamento dos parâmetros nos Tratamentos aplicados aos textos.

- **Comparação do Caminho Mínimo Médio**

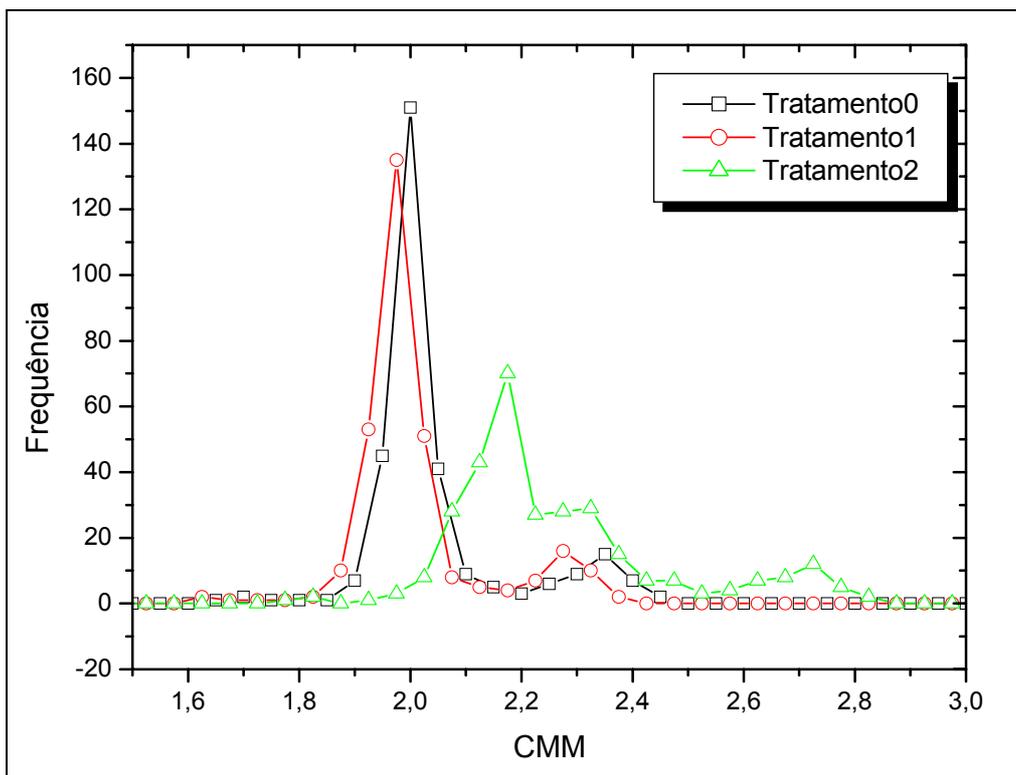


Figura 39 - Comportamento do CMM das redes dos textos submetidos aos tratamentos 0, 1 e 2.

A eliminação das palavras gramaticais do Tratamento 2 é que causa maior interferência nesse parâmetro, pois são estas palavras que criam ‘pontes’ em todo o texto, aproximando-as, e conseqüentemente reduzindo o CMM (Figura 39).

- **Comparação do Grau Médio**

A redução das palavras à sua forma canônica quase não interfere no comportamento do Grau Médio, onde observamos um pequeno acréscimo das palavras com graus (ligações) acima de 100. O efeito pode ser notado na Figura 40, com a suave modificação do formato dos valores de Grau Médio encontrados no texto com o Tratamento 1.

A eliminação das palavras gramaticais provoca uma alteração maior no comportamento desse parâmetro, pois como se trata de uma análise de valores

médios, ao eliminarmos as palavras com os maiores valores de grau (palavras gramaticais) a média tenderá a diminuir (Figura 40).

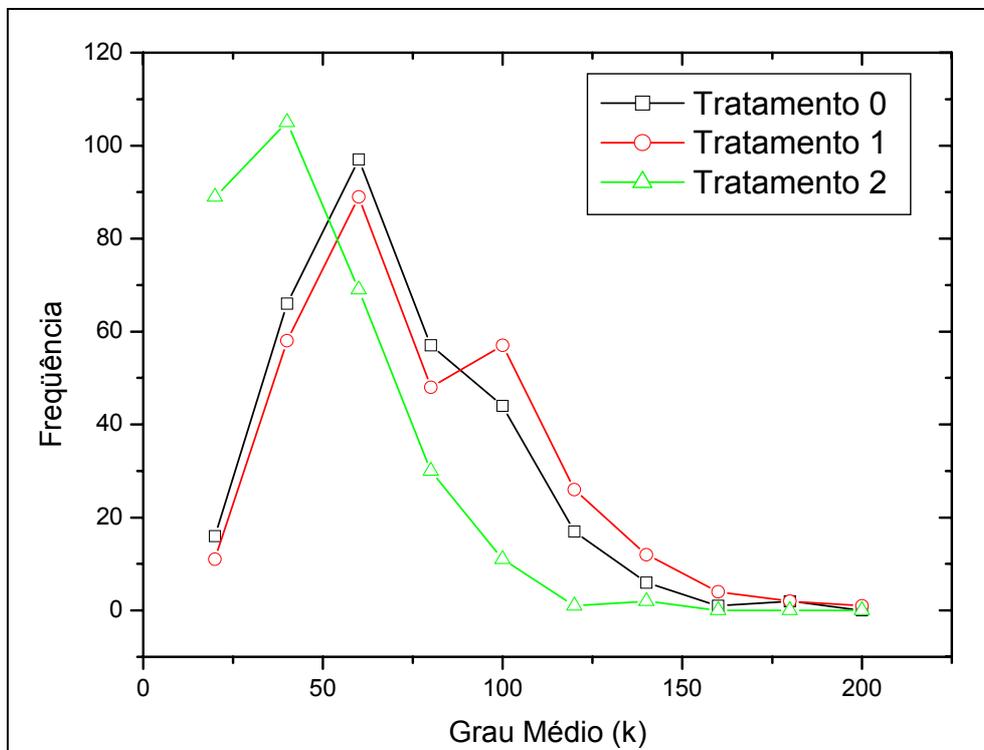


Figura 40 – Comparação do Grau Médio das redes dos textos tratados nas condições 0, 1 e 2

Podemos então concluir que os resultados dos parâmetros da rede sofrem uma certa alteração por conta do tratamento realizado nos textos analisados. Porém, para efeito do nosso estudo será tomado o Tratamento 2 para análise e caracterização do comportamento dos textos, pois a abstração que fazemos refere-se à rede de conceitos trazidos pelo autor ao escrever um texto, e analisando os resultados vemos que se trata de uma rede com topologia distinta das anteriores.

Podemos então concluir que o tratamento escolhido neste trabalho (Tratamento 2) está relacionado com uma estrutura de associações diferente dos outros tratamentos (Tratamentos 0 e 1). Dessa maneira, as avaliações seguintes referem-se apenas à condição do Tratamento 2, ou seja, o uso das palavras em sua forma canônica e

eliminação das palavras gramaticais para a formação da rede. Na Tabela 9 encontra-se o resumo dos valores do CAM e do CMM para os três tratamentos. Nela se confirmam os resultados comentados anteriormente.

Tabela 9 – Resumo dos valores encontrados para os tratamentos do texto

Tipo do tratamento	CAM	CMM	<k>
Tratamento 0	0,82 ± 0,03	2,0 ± 0,1	69 ± 29
Tratamento 1	0,81 ± 0,03	2,0 ± 0,1	76 ± 33
Tratamento 2	0,77 ± 0,05	2,3 ± 0,2	45 ± 23

5.3 RESULTADOS COMPARATIVOS EM RELAÇÃO À AMOSTRAGEM

A amostragem dos textos analisados foi feita de tal maneira que não apresentasse tendências de resultados diferenciados em função de sua composição. Calculando os parâmetros característicos da rede para cada grupo identificado, podemos notar que não há diferenças significativas entre seus valores que possam ser utilizadas para caracterizar o grupo a que pertence, seja em relação ao idioma do texto, ao estilo do texto, ao gênero do autor ou até mesmo ao seu tamanho, medido em quantidade de sentenças (Tabela 10).

Tabela 10 – Resultados da análise dos textos segmentados nas classes definidas na amostragem

		CAM	CMM	<K>	GAMA
Idioma	Inglês	0,76 ± 0,04	2,3 ± 0,2	44 ± 28	1,6 ± 0,2
	Português	0,78 ± 0,05	2,2 ± 0,2	46 ± 19	1,7 ± 0,2
Estilos do Texto	Técnicos	0,77 ± 0,05	2,2 ± 0,1	49 ± 20	1,6 ± 0,2
	Literários	0,77 ± 0,05	2,4 ± 0,3	40 ± 26	1,6 ± 0,2
Gênero do autor	Feminino	0,77 ± 0,05	2,2 ± 0,1	49 ± 20	1,6 ± 0,2
	Masculino	0,76 ± 0,05	2,3 ± 0,2	41 ± 25	1,6 ± 0,2
	Coletivo	0,77 ± 0,04	2,2 ± 0,3	51 ± 21	1,5 ± 0,2
Tamanhos do Texto	Até 1000	0,83 ± 0,04	2,2 ± 0,2	29 ± 11	1,5 ± 0,2
	Acima de 1000	0,75 ± 0,03	2,3 ± 0,2	50 ± 24	1,6 ± 0,2

Há um padrão comum na forma como os indivíduos escolhem as palavras para construir os seus textos, pois quase todos os valores são sobrepostos pela barra de erro. Os desvios encontrados não justificam uma separação. Uma ressalva porém deve ser feita em relação à análise dos textos com diferentes tamanhos, visto que os valores do CAM e do $\langle k \rangle$ apresentaram valores distintos. Esse fato será melhor explicitado na análise do crescimento dos textos, na próxima seção.

As redes dos textos apresentaram características de redes de mundo pequeno e de redes livres de escala. As redes de mundo pequeno são caracterizadas por apresentarem Coeficiente de Aglomeração alto (CAM = 0,77 ± 0,05) e Caminho Mínimo Médio pequeno (CMM = 2,3 ± 0,20). Já as redes livres de escala são identificadas por apresentarem Distribuição de Graus em forma de lei de potência. As redes de texto apresentaram $\gamma = 1,60 \pm 0,19$.

5.4 DISTRIBUIÇÃO DO TAMANHO DAS SENTENÇAS

Outro fator importante na definição do comportamento do sistema como um todo é a forma como construímos as sentenças, já que estas foram consideradas como a menor unidade de análise.

Foi observado para o conjunto dos textos analisados que a sentença menor (ou seja, a que apresenta quantidade mínima de palavras) mais freqüente possui apenas uma palavra, e a sentença maior (a que apresenta quantidade máxima de palavras) mais freqüente apresenta 41 palavras em sua composição, conforme podemos ver na Figura 41, em que o gráfico da esquerda representa a distribuição de freqüência das sentenças menores em todos os textos analisados, e no gráfico da direita podemos ver a distribuição de freqüência das sentenças maiores.

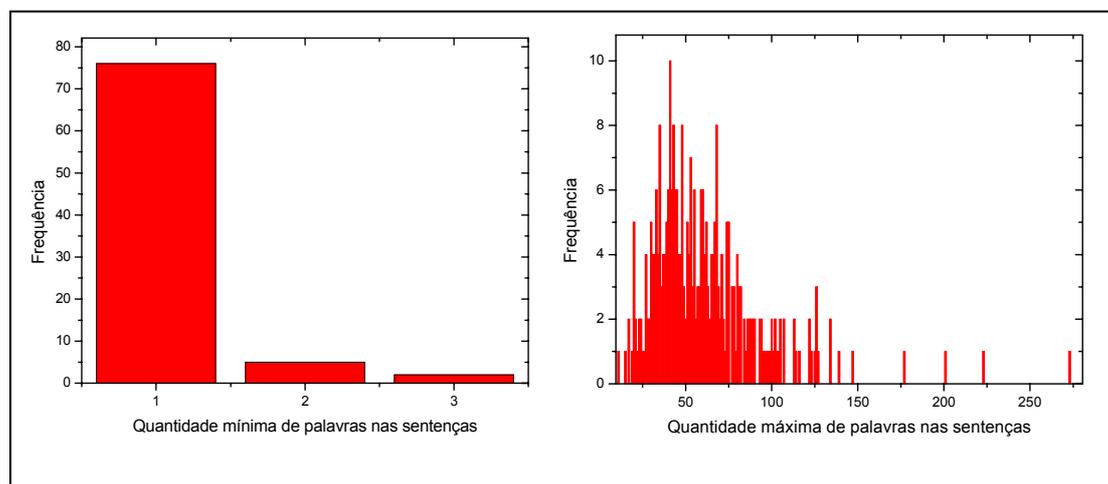


Figura 41 – Distribuição de freqüência do tamanho das sentenças dos textos analisados.

Apesar de encontrarmos textos com tamanhos de sentenças muito grandes, chegando a 273 o valor máximo, podemos notar que a quantidade de palavras mais provável de se encontrar nas sentenças é da ordem de 4, conforme Figura 42. E, a quantidade média de palavras por sentença nos textos analisados é 11.

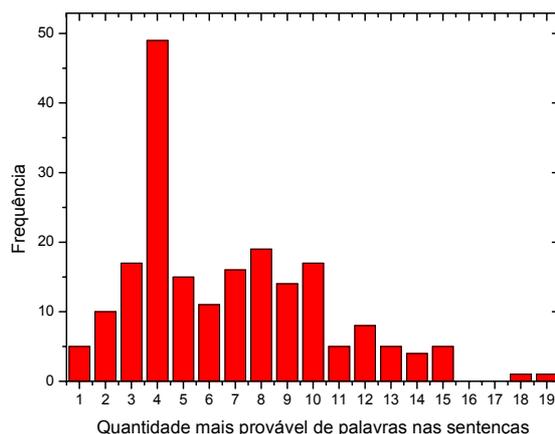


Figura 42 – Distribuição de frequência do valor mais provável de quantidade de palavras nas sentenças.

Fazendo uma análise separada dos textos nos idiomas analisados, ficou evidenciado que as sentenças em português tendem a ser construídas com menor quantidade de palavras lexicais, conforme Tabela 11. Além disso, as sentenças maiores mais prováveis, ou seja, com maior frequência de serem encontradas nos textos analisados, são de tamanho menor nos textos em português, comparados aos textos em inglês.

Tabela 11 – Análise da distribuição de frequência do tamanho das sentenças por idioma

Idioma	Mínimo	Máximo	Máximo mais provável	Valor mais provável
Textos em Inglês	1	273	53 e 68	4
Textos em Português	1	223	41	3

Uma observação interessante, é que, pelo senso comum, a língua portuguesa sempre foi considerada muito mais prolixa, mais redundante do que a língua inglesa. Aqui, pelos resultados experimentais, conclui-se que isso é incorreto, pelo menos no que se refere às palavras escolhidas para compor a rede, isto é, as palavras lexicais. Seria interessante que um estudo mais detalhado pudesse ser realizado para a constatação dessa evidência preliminarmente caracterizada nesse estudo.

5.5 INFLUÊNCIA DO TAMANHO DAS SENTENÇAS NO COEFICIENTE DE AGLOMERAÇÃO

Dos índices analisados o que mais informações trouxe sobre a topologia da rede foi o Coeficiente de Aglomeração (CAM), por isso, um maior detalhamento foi feito sobre o seu significado tanto nas redes dos textos completos quanto na avaliação das etapas de crescimento do texto. Na Figura 43 verifica-se uma simulação de construção de um pequeno texto e como a sua composição influencia no valor encontrado para todos os textos analisados, demonstrando que o caráter comum deste valor é uma relação direta de como o texto vai sendo escrito.

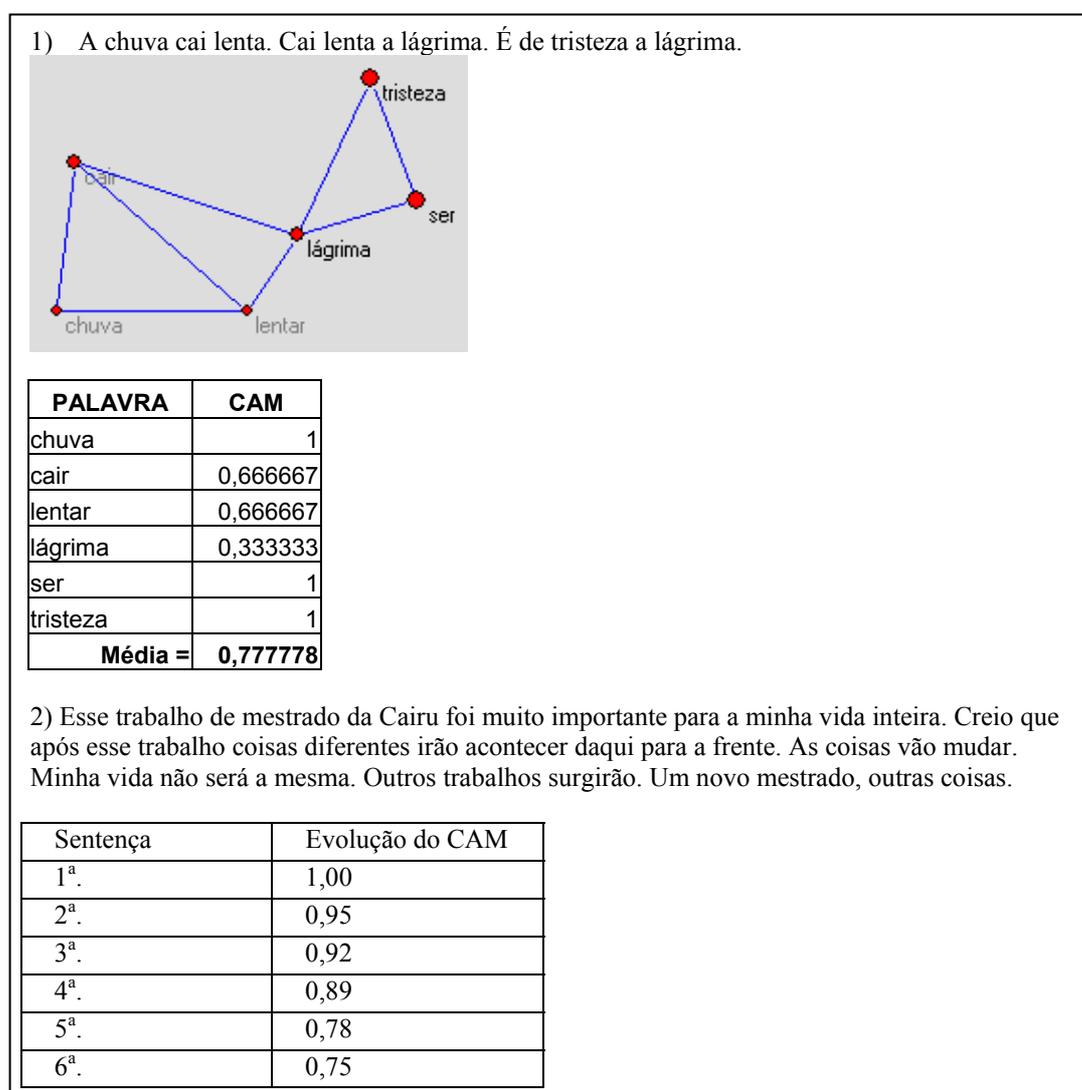


Figura 43 – Simulação de possíveis composições e tamanhos de sentenças para alcançar o valor de CAM encontrado nos textos originais analisados.

Várias combinações entre tamanhos de sentenças e palavras repetidas podem ser simuladas a fim de exemplificar como o valor médio do coeficiente de aglomeração (0,7664) poderia ser gerado. No exemplo da Figura 43(1) temos 3 sentenças que formam uma cadeia de significados, com repetição de palavras seqüenciadas. Na Figura 43, exemplo 2, temos uma simulação em que o acréscimo de novas sentenças foi promovendo o decréscimo gradativo do valor do coeficiente de aglomeração.

Os resultados obtidos nessa análise referem-se a um processo de simulação do comportamento do texto relacionado ao tamanho das sentenças. Sem no entanto averiguar com uma quantidade de textos maior. Sugerimos que outros estudos sejam desenvolvidos, gerando uma quantidade de dados mais significante.

5.6 ANÁLISE DO CRESCIMENTO DOS TEXTOS

A caracterização das redes formadas pelos textos completos, ou seja, a análise estática, apresentou resultados indicativos de um padrão comum a todas as classes de amostragem. O interesse nessa parte do trabalho está na caracterização em diferentes etapas de crescimento das mesmas, a fim de identificar se na construção dos textos também há ocorrência de padrões comuns.

5.7 RESULTADOS DO CRESCIMENTO DO T-VÉRTICE

Na Figura 44, apresentamos um gráfico com a dinâmica do acréscimo de novas palavras ao texto. Notamos que à medida em que o texto vai sendo escrito, o número de vértices cresce numa relação direta com a quantidade de sentenças acrescentadas.

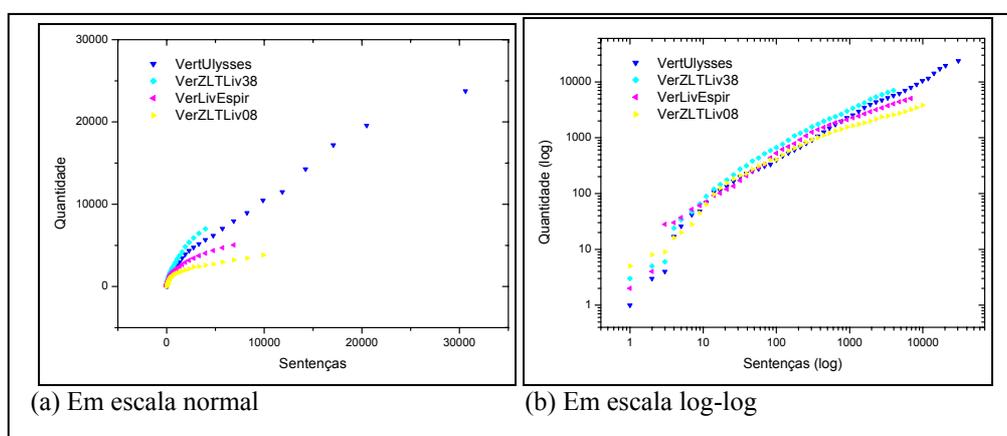


Figura 44 – Gráficos do crescimento de textos indicando o comportamento do índice T-vértice.

Os gráficos da Figura 44 apresentam a dinâmica de acréscimo de novas palavras ao texto. No primeiro (a), em escala normal, pode-se verificar que cada texto possui uma taxa de acréscimo diferente e esta não é a mesma à medida em que o texto vai sendo escrito. No segundo (b), em escala logarítmica, observamos que a equação que descreve as curvas é a mesma, um crescimento em lei de potência (com expoente aproximado de 0,74), ou seja, isso significa que a forma com que as taxas de crescimento se modificam no decorrer da escrita, é a mesma para todos os textos.

Analisando, como exemplo, o livro *Ulysses* na Figura 44(a), temos que ele apresentou uma taxa média de acréscimo de novas palavras a cada sentença de 84%, ou seja, a cada 100 sentenças escritas, 84 trazem uma nova palavra que ainda não tinha sido referenciada. Porém esse acréscimo não é constante. No início do texto (até 1594 sentenças) surgem, em média, 2,24 novas palavras, estabilizando em 0,83 até o final, quando alcançamos 30.000 palavras. Embora o autor apresente um comportamento de retornar a falar palavras já citadas, de uma forma redundante, e esse comportamento seja o responsável pela topologia da rede, ele não deixa de trazer conceitos novos ao seu texto até o final.

O livro *Ulysses* de James Joyce configurou-se num exemplo ideal para esse estudo, pois o autor é conhecido pela sua técnica de “inventar” novas palavras em seus livros, inclusive palavras com até 40 caracteres.

Segundo KANDEL et al (1997), o ser humano possui um vocabulário em torno de 60.000 a 120.000 palavras, porém num livro de 256.000 palavras, como neste caso, somente foram utilizadas cerca de 25.000 palavras diferentes, embora o autor claramente quisesse “esgotar” todo o seu vocabulário, inclusive acrescentando outras “inventadas” especificamente para aquele livro.

Mesmo James Joyce, que tinha a intenção de usar a linguagem de um modo diferenciado do escritor comum, apresentou um comportamento similar de acréscimo de novas palavras em seu texto, não alcançando, finalmente o seu objetivo de esgotar o seu vocabulário.

Outras similaridades entre o texto desse autor e os outros analisados nesta pesquisa serão mostradas em seguida.

5.8 COMPARAÇÃO DOS RESULTADOS DE TODOS OS TEXTOS COM O CRESCIMENTO DOS MAIORES TEXTOS ANALISADOS

A análise do Coeficiente de Aglomeração Mínimo (CAM) de todos os textos, relacionado à quantidade de vértices da rede, demonstra uma tendência de variação, que também é percebida quando simulamos o crescimento de textos, no caso o Código de Da Vinci de Dan Brown e Anna Karenina de León Tolstói (Figura 45).

Pode-se notar um decaimento do valor do CAM no início do texto, em função, tanto das repetições das palavras (que passam a ocorrer com uma acentuada frequência no início do texto), quanto do tamanho das sentenças, que em média estão em torno de 3 a 4 palavras, depois se estabilizando em valores próximos a 0,7 – 0,8.

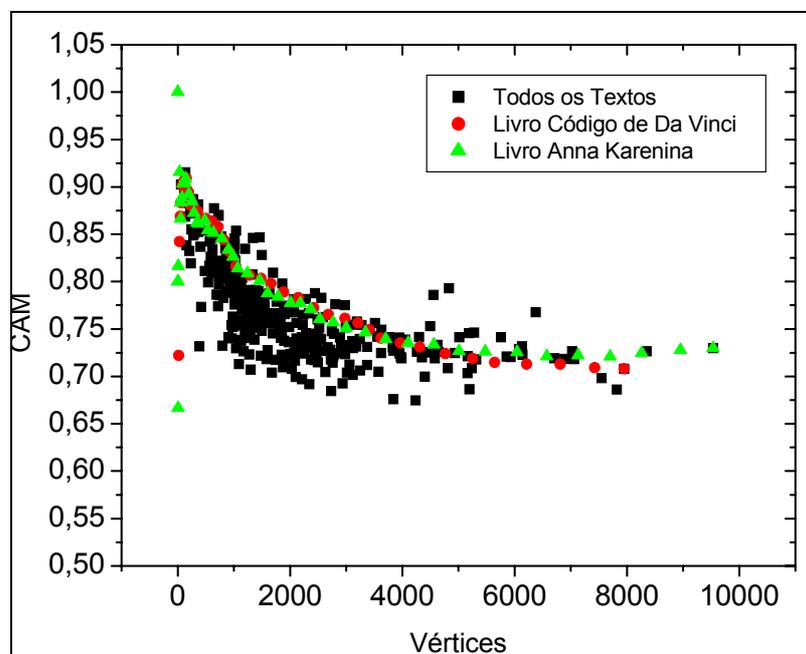


Figura 45 – Resultados do CAM para todos os livros analisados, comparando com o crescimento de dois livros: Código de Da Vinci de Dan Brown e Anna Karenina de León Tolstói.

Os valores de CAM encontrados para os 310 textos apresentam similaridade com a dinâmica do crescimento dos dois textos, Código de Da Vinci de Dan Brown e Anna Karenina de León Tolstói. Isso sugere que, caso as pessoas que escreveram os textos menores continuassem a construção dos mesmos, apresentariam valores finais de CAM próximos dos textos maiores. Isso significa que existe um padrão comum na forma como os textos são construídos para todos os autores analisados.

5.9 CRESCIMENTO INDIVIDUAL DOS TEXTOS

Não somente os valores de Coeficiente de Aglomeração dos textos completos demonstrou um padrão comum. Também a própria dinâmica do crescimento destes apresentou similaridade em seu comportamento, sempre decrescendo com a mesma forma, mesma inclinação, em lei de potência, mesmo para textos em idiomas e tamanhos diferentes, tendendo a um valor assintótico em torno de 0,77, confirmando os resultados obtidos para textos de diferentes tamanhos (Figura 46 e Figura 47).

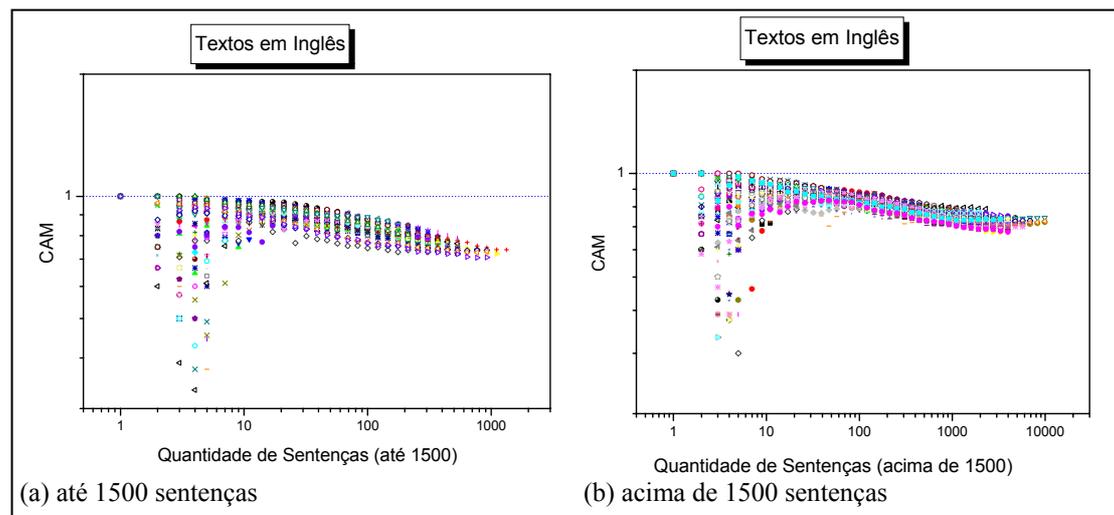


Figura 46 - Crescimento do CAM em textos em inglês, divididos pelo tamanho

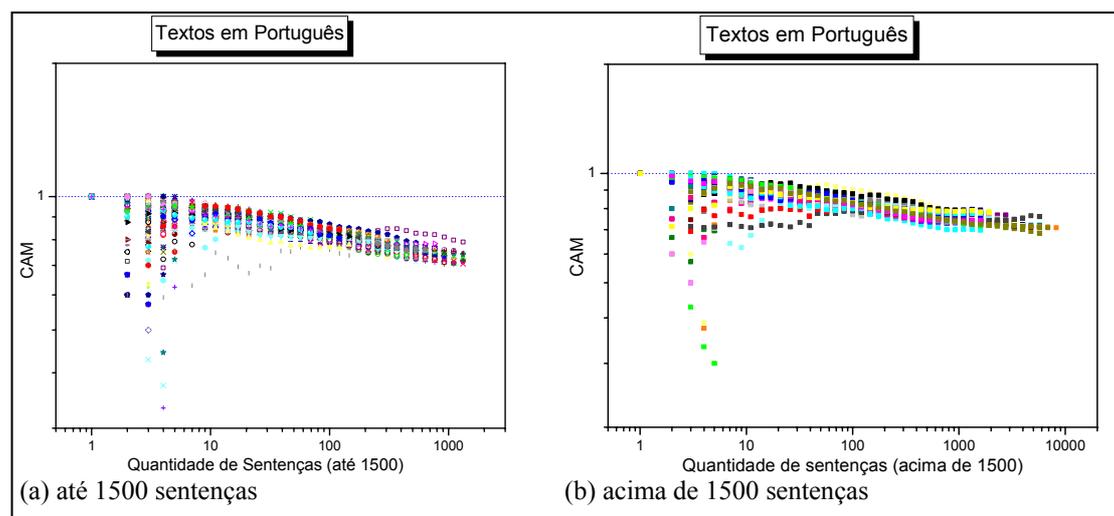


Figura 47 - Crescimento do CAM em textos em português, divididos pelo tamanho

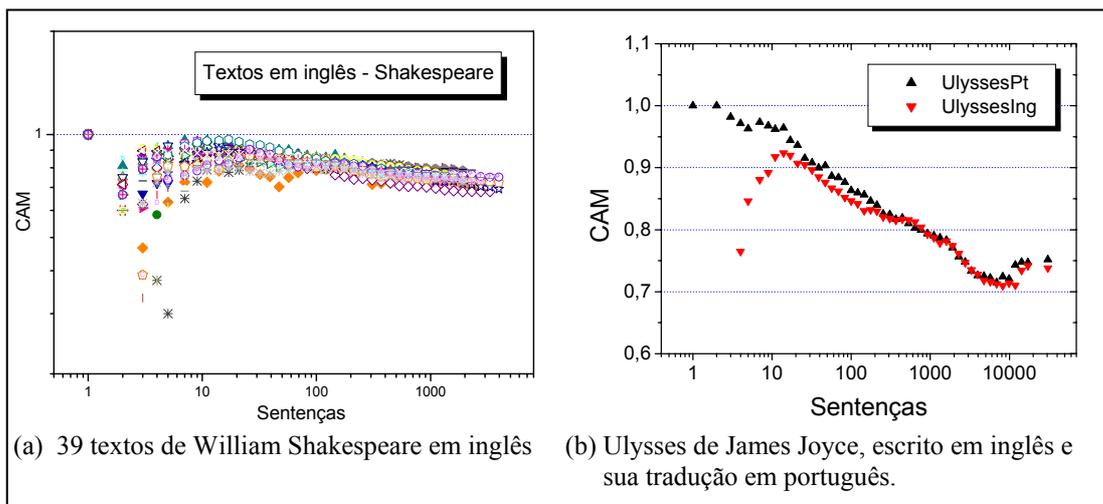


Figura 48 – Comparação dos textos de dois autores

Quando analisados textos de mesmo autor, como na

Figura 48(a), as curvas apresentam um comportamento melhor definido, indicando seu modo de construção do texto. Isso ocorre mesmo em textos que sofreram tradução, no caso de Ulysses, conforme

Figura 48(b).

Analisamos também outros índices no crescimento dos textos (Diâmetro e Caminho Mínimo Médio), porém não nos pareceu revelar nenhuma nova informação a respeito da dinâmica do crescimento. Os resultados seguem abaixo, com a indicação prévia de que não são conclusivos, exigindo outras pesquisas melhor compreensão desse comportamento observado.

5.10 RESULTADOS DO CRESCIMENTO DO DIÂMETRO (DAM)

Como não foram encontradas referências bibliográficas que tratassem da dinâmica de crescimento de textos, foram também analisados os outros parâmetros de caracterização de redes complexas. Na Figura 49, apresentamos um típico comportamento para a evolução do livro *Ulysses*, de James Joyce. Nele podemos notar que a rede do texto escrito analisado apresenta, logo no início, um pequeno acréscimo no valor do diâmetro (Seção 3.2), pois entende-se que no princípio da escrita de um texto estamos delimitando quais os assuntos que iremos abordar ao longo do texto. O escopo do que será dito. Depois, vamos acrescentando as palavras que irão explicar os tópicos, contextualizando-os. Isso não significa que ao longo do texto não teremos novos assuntos. Apenas serão diluídos no todo, já que o agrupamento (*cluster*) principal já está formado (Figura 49).

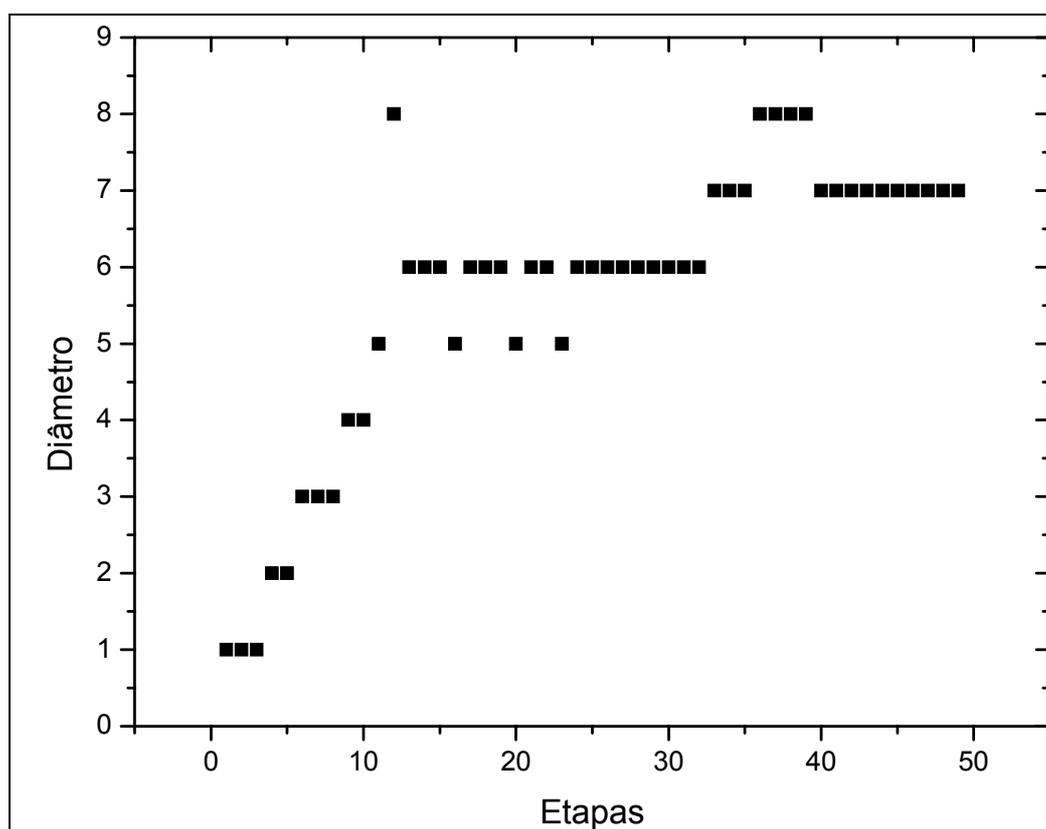


Figura 49 – Gráfico da evolução do diâmetro no crescimento do livro *Ulysses*, de James Joyce, escrito em português.

5.11 RESULTADOS DO CRESCIMENTO DO CAMINHO MÍNIMO MÉDIO (CMM)

No gráfico da Figura 50, apresentamos um exemplo típico da evolução do CMM, onde vemos que no início, quando o texto possui poucas sentenças, o CMM apresenta valores baixos, que logo vão crescendo, indicando que a rede em crescimento está fazendo com que as palavras novas, acrescentadas ao texto, para serem alcançadas a partir de alguma das primeiras, necessite passar por mais palavras. À medida em que vamos repetindo palavras, vamos criando o que poderia se chamar de “pontes” entre elas, provocando um decréscimo com posterior estabilização em valores de CMM que variaram entre 1,77 e 3,24 nos textos analisados.

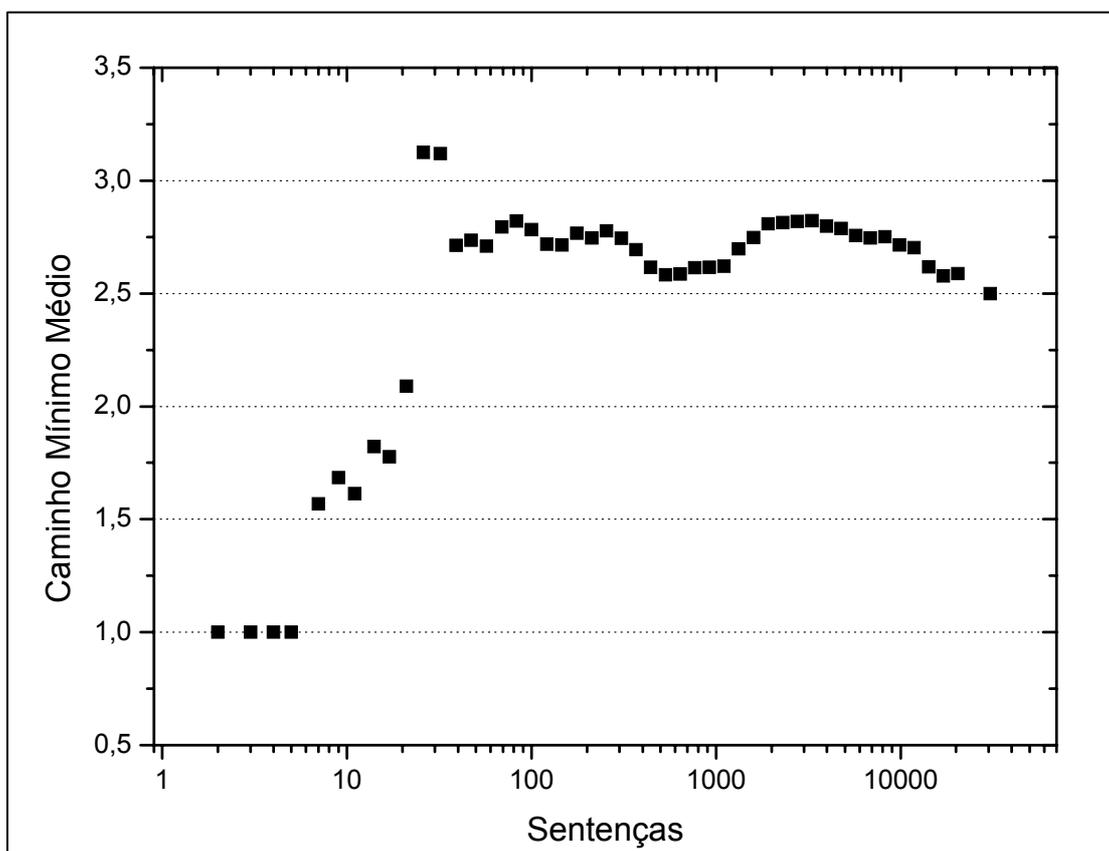


Figura 50 – Gráfico da evolução do Caminho Mínimo Médio na simulação de crescimento do livro *Ulysses*, de James Joyce, escrito em português.

5.12 SIMULAÇÃO DA REDE DE PALAVRAS

Outra perspectiva de análise dos textos refere-se ao entendimento do modo como eles são gerados, ou seja, se existe um padrão de comportamento determinando qual o principal elemento dos textos responsável pela topologia de suas redes apresentarem características de redes de mundo pequeno e de redes livres de escala.

O modelo de crescimento dos textos foi simulado, realizando vários testes de embaralhamento de seus elementos, com o propósito de identificar qual deles se aproximava mais das condições reais de construção de um texto. Além disso, conclusões importantes foram retiradas com relação ao modo como o dicionário de palavras é utilizado quando da construção de um texto. Todos os testes de simulação realizados trataram o texto como um sistema composto de elementos (palavras e marcadores de sentença) e o embaralhamento realizado não teve qualquer preocupação com o sentido semântico do texto ao final do processo.

Para avaliar essa perspectiva foram feitas quatro simulações com o texto, com a finalidade de detectar os pontos propostos. As simulações seguiram a seguinte estratégia:

- **Aleatório Condição 1**

Nesse embaralhamento mantivemos os elementos (i.e. palavras, frequência de palavras e quantidade de marcadores de sentença) e fizemos o embaralhamento da matriz adjacente, ou seja, envolvendo tudo. Nesta simulação não analisamos todos os textos, compomos uma amostragem pois com os resultados encontrados já sabíamos que as redes não apresentavam comportamento de redes aleatórias. Escolhemos entre textos de diferentes idiomas, com gênero do autor diferentes e tamanhos de textos de até 1500 sentenças e acima de 1500 sentenças.

- **Aleatório Condição 2**

A condição 2 de embaralhamento manteve os tamanhos das sentenças e a ordem em que aparecem no texto, ou seja, a localização dos pontos, porém as palavras foram embaralhadas, também sem qualquer preocupação com o sentido semântico final do texto. A frequência das palavras continuou a mesma do texto original tratado.

- **Aleatório Condição 3**

A condição 3 de embaralhamento calcula o tamanho médio da sentença (quantidade de palavras que a compõe), construindo um novo texto com sentenças de tamanho igual à média, apenas deslocando o marcador para sua nova posição, mantendo a seqüência de aparecimento das palavras iguais aos do texto original tratado. Neste teste a freqüência das palavras também continua a mesma.

- **Aleatório Condição 4**

Por último, mantivemos as sentenças em seus tamanhos e ordem do texto original tratado, fazendo uma simulação da escolha das palavras para compor cada sentença, a partir do próprio vocabulário trazido pelo autor do texto, conseqüentemente alterando suas freqüências.

Os resultados dos principais parâmetros de caracterização das redes em crescimento foram então comparados com os dos textos nas condições de Tratamento 2, ou seja, submetido ao pré-tratamento proposto neste trabalho, de uso da forma canônica das palavras e eliminação das palavras gramaticais.

5.13 ANÁLISE DO CAMINHO MÍNIMO MÉDIO

Na Figura 51 estão traçados os gráficos de distribuição de freqüência dos valores de CMM encontrados para os diversos textos analisados. Em média, nenhuma das condições de embaralhamento provocou alterações neste parâmetro. Isso porque redes aleatórias também apresentam CMM baixo, não sendo este parâmetro isoladamente adequado para diferenciar entre redes aleatórias, redes de mundo pequeno e redes livres de escala.

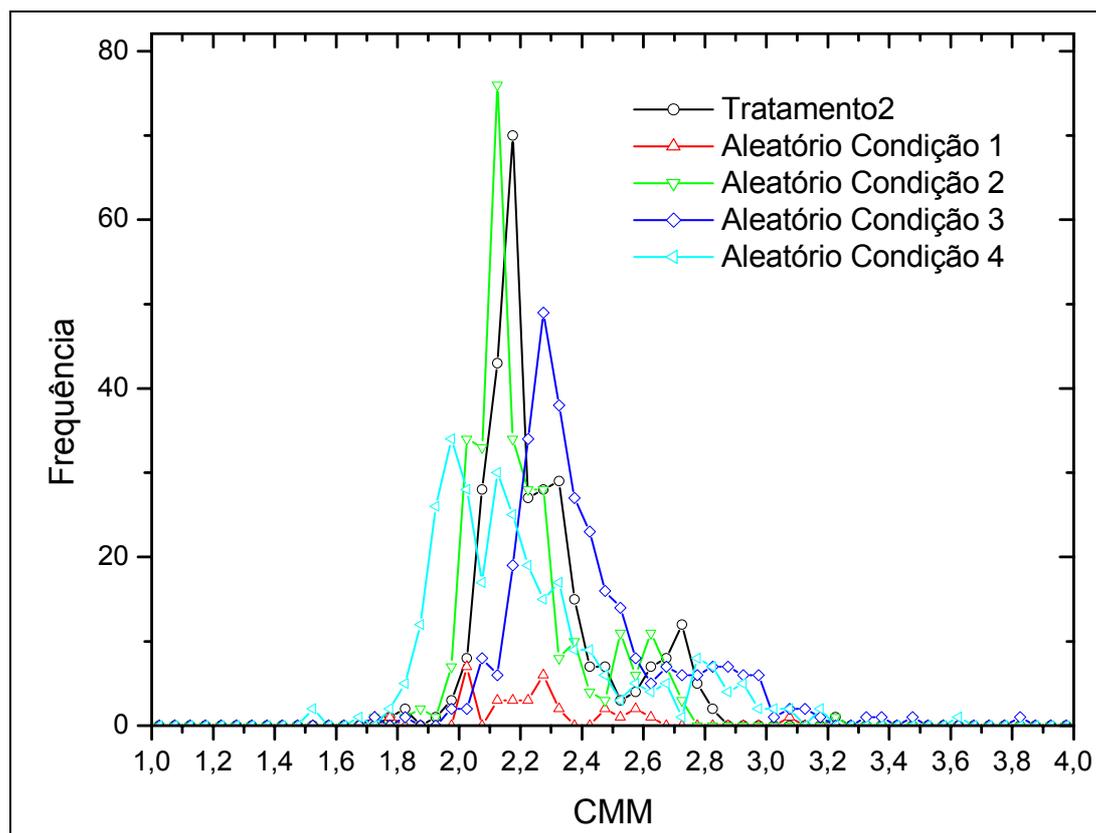


Figura 51 – Distribuição de Frequência dos valores de CMM dos textos analisados nas condições de Tratamento 2 e Aleatório Condição 1, 2, 3 e 4.

5.14 ANÁLISE DO COEFICIENTE DE AGLOMERAÇÃO MÉDIO

Na

Figura 52 temos o comportamento do parâmetro CAM para os diversos textos, indicando a diferença entre textos totalmente aleatórios (Aleatório Condição 1) e os textos com Tratamento 2. As simulações Aleatório Condição 2 e Aleatório Condição 3 apresentam resultados similares ao Tratamento 2, numa demonstração de que tanto o tamanho das sentenças quanto a ordem de aparecimento das palavras no texto são fatores que não influenciam nos valores do CAM. Já a frequência com que as palavras são utilizadas (Aleatório Condição 4) é o fator principal que interfere no comportamento do parâmetro CAM.

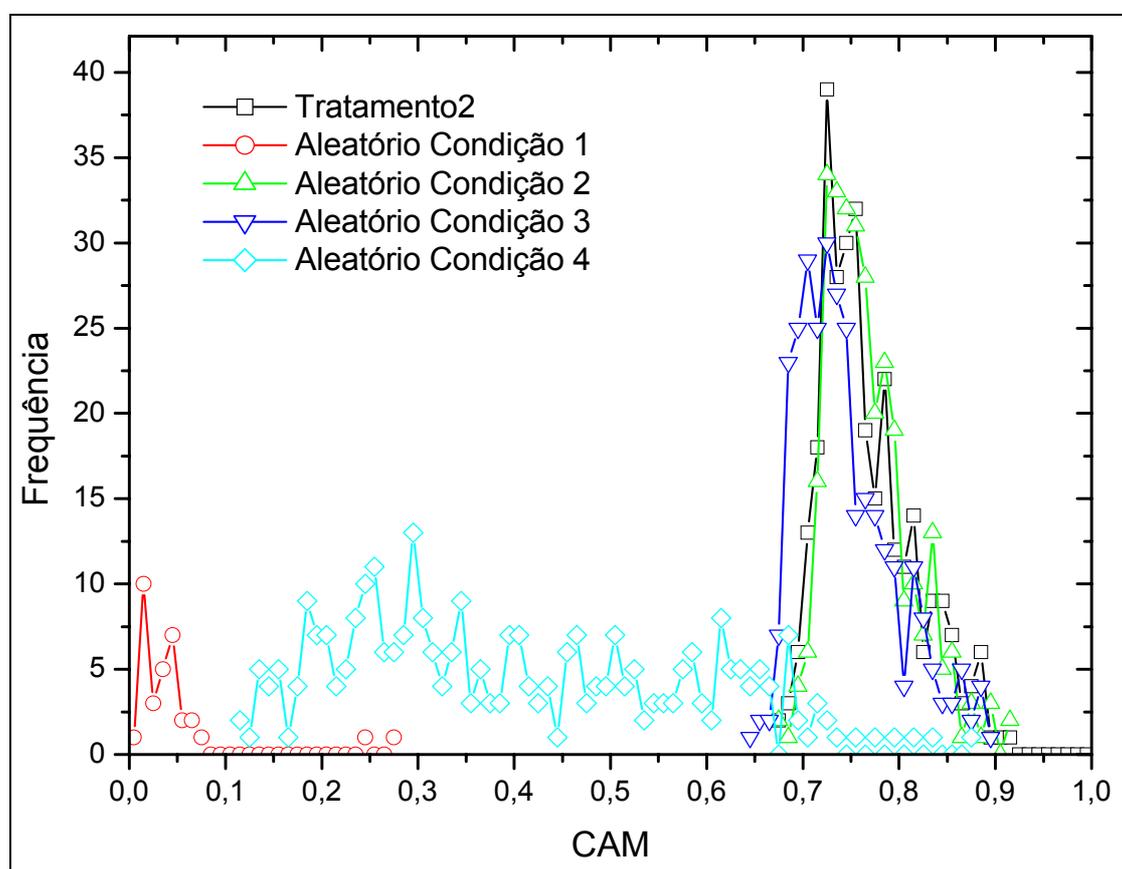


Figura 52 – Distribuição da Frequência dos valores de CAM dos textos analisados nas condições de Tratamento 2 e Aleatório Condição 1, 2, 3 e 4.

5.15 ANÁLISE DA DISTRIBUIÇÃO DE GRAUS

Esse parâmetro é capaz de indicar com clareza que quebramos a estrutura de uma rede do tipo livre de escala, quando alteramos a frequência de palavras no texto, no caso da simulação Aleatório Condição 4 (Figura 53).

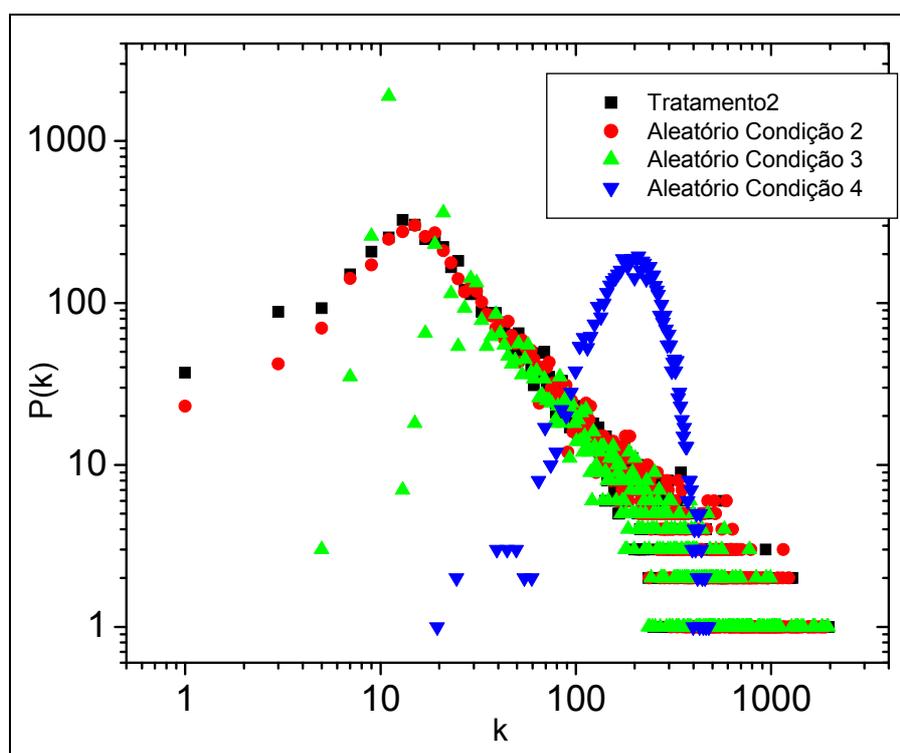


Figura 53 – Distribuição de Graus de um mesmo texto nas condições de Tratamento 2 e Aleatório Condição 2, 3 e 4.

Na Tabela 12 encontra-se o resumo dos índices obtidos para as diferentes formas de embaralhamento/simulação.

Tabela 12 – Resumo dos Índices do Embaralhamento / Simulação

	CAM	Desvio padrão	CMM	Desvio padrão
Tratamento 2	0,77	0,05	2,27	0,20
Aleatório Condição 2	0,77	0,04	2,21	0,19
Aleatório Condição 3	0,74	0,05	2,42	0,27
Aleatório Condição 4	0,40	0,18	2,22	0,32

5.16 COMPARAÇÃO DOS RESULTADOS OBTIDOS COM OUTRAS PESQUISAS

Na Tabela 13 encontram-se os valores dos índices CAM, CMM, e γ das principais pesquisas utilizando redes complexas para a análise da linguagem, comparando com os valores obtidos em nosso trabalho.

Tabela 13 – Comparação de resultados da análise com pesquisas anteriores

Pesquisas	CAM	CMM	γ
Costa (2004)	0.075	3.32	-
Motter et al (2002)	0.53	3.16	3.5
Cancho e Solé (2001)	0.687	2.63	1.5
Resultado desta pesquisa	0.77	2.3	1.6

O método de construção das redes dos trabalhos de COSTA (2004) e MOTTER et al (2002) sugere que estes tratam da língua, pois fazem associações entre palavras sem o objetivo de construir um discurso. O primeiro associando termos aleatoriamente que levou a índices característicos de uma rede aleatória. O segundo tratou de uma associação com uma regra definida: somente as palavras sinônimas estariam conectadas. E esta regra caracteriza um relacionamento lingüístico.

Por outro lado, o método do nosso trabalho e o de Cancho e Sólé (2001) apresentaram similaridades, pois ambos consideraram a associação das palavras em textos discursivos. Cabe, porém, ressaltar que o último trata do inventário de uma língua, enquanto que analisamos os textos individuais, ou seja, a fala, segundo Saussure (2004), constatando porém que os resultados obtidos estiveram mais próximos que os resultados dos trabalhos anteriores.

CONCLUSÕES

A complexidade da linguagem pode ser constatada pelo intrincado mecanismo que utiliza e pela excessiva quantidade de elementos que mobiliza. O desenvolvimento dos recursos de processamento computacional, em paralelo com as novas perspectivas da física moderna, os sistemas complexos, permitiram a análise da linguagem pela ótica da física. Um sistema composto de muitos elementos (palavras) que interagem entre si em pequenos grupos (sentenças), dando origem a um organismo auto-organizado maior (o texto), criando uma ordem emergente, o seu significado.

Outros estudos anteriores foram realizados, provando a complexidade do sistema de uma língua. Porém, nesta pesquisa foram acrescentadas sutis contribuições que conduziram a uma interpretação diferente dos experimentos até então realizados.

A primeira contribuição foi a exclusão das palavras gramaticais ou funcionais, e a constatação de que as características de ordem encontradas no sistema, e que foram atribuídas a esses elementos²⁵, mantiveram-se, mesmo eliminando-as. Houve uma leve alteração nos valores, porém as redes ainda conservaram as características de mundo pequeno e livre de escala. Podemos dizer então que as características topológicas das redes de textos não estão diretamente relacionadas à existência das palavras gramaticais. Existem palavras de outra natureza, as lexicais, que são utilizadas com frequência suficiente para lhes dar essas características. A estrutura da rede não é resultado do mecanismo sintático de construir os textos em sentenças e nem tampouco de utilizar palavras que possuem a função de estruturar o texto gramaticamente. A rede montada com as palavras lexicais, ou seja, aquelas que despertam em nossa mente representações de seres e ações, e suas respectivas qualidades, teve o mesmo comportamento.

A segunda contribuição refere-se à natureza dos sistemas analisados. Constatamos a existência de estudos que apesar de utilizarem os mesmos elementos (as palavras), geram sistemas com semelhante topologia de redes, porém com valores de parâmetros distintos, podendo ser um indicativo de tratarem-se de sistemas de natureza diferentes. Essa constatação leva-nos a propor uma pesquisa mais detalhada a fim de comprovar

²⁵ “Hubs in lexical networks are function words (e.g., prepositions, articles, and determiners). Hubs, and thus function words are crucial for the lexical network small-worldness”.

esse comportamento, porém sugerimos que essas diferenças podem ser atribuídas, às regras empregadas para ligar os elementos do sistema, podendo estar diretamente relacionadas à sua finalidade, que classificaremos em dois tipos: *estudos da linguagem* (experimentos psicofísicos com associação livre de palavras e associação de sinônimos de palavras constantes de um dicionário) e *estudos da língua* (análise de corpora com janelas de co-ocorrência e análises com textos individuais levando em conta as sentenças). Porém, ressalto que para uma maior clareza e certeza desta afirmação será necessário um aprofundamento das pesquisas.

O estudo da dinâmica de construção dos textos escritos, baseado no acréscimo de novas sentenças, com a análise dos parâmetros em cada etapa foi, tecnicamente, a contribuição mais inovadora desse trabalho, já que não havíamos encontrado até então esse método de análise, e se constitui como a nossa terceira contribuição à análise de textos como um sistema complexo.

Essa análise comparativa da dinâmica de crescimento dos textos individuais, mesmo que escritos por autores diferentes, que abordem assuntos diferentes e até mesmo escrito em línguas diferentes demonstrou a similaridade de comportamento da evolução dos sistemas dos textos em todos os casos analisados. Sendo a taxa de crescimento de novas palavras nos textos, sem dúvida, a contribuição mais eficiente, já que identificamos que a taxa para alguns textos escolhidos aleatoriamente demonstrou ser diferente, porém a lei que descreve o crescimento parece ser a mesma. Aqui caberia um aprofundamento nessa avaliação, estendendo esse estudo para um número maior de textos.

Os parâmetros analisados nas redes construídas foram suficientes para comprovar que o sistema composto pelas palavras de um texto é um sistema complexo e, portanto, apresenta características de não-aleatoriedade, além de apresentar uma dinâmica de crescimento bem definida.

Essas são características de sistemas que apresentam propriedades emergentes, que, no caso estudado, acreditamos ser a significação do texto, otimizada pela maneira como ocorrem as ligações entre as palavras, destacando aqui a forma como repetidamente são evocadas as palavras citadas anteriormente no texto, de tal forma que as redes apresentaram características de mundo pequeno e livres de escala,

características que estão intrinsecamente ligadas à frequência dos elementos no sistema analisado. Sistemas que apresentam propriedades emergentes possuem um princípio auto-organizativo, que, neste trabalho consideramos o processo associativo proposto por Freud em seu aparelho psíquico, que nos sugere uma topologia de rede.

Uma rede. Esse foi o modelo escolhido para representar as ligações que as palavras possuem entre si, quando consideradas as sentenças como menor unidade de análise, e quando supomos que cada repetição de uma palavra fará com que as novas palavras que a acompanham se unam firmemente à estrutura antiga, formando um complexo intrincado chamado rede complexa.

Essa estrutura complexa e auto-organizada ganha maior importância e beleza quando lembrarmos que essas palavras, que por ora repousam inertes no papel ou na tela do computador, foram objeto de uma acirrada disputa na mente do autor do texto em análise. Qual a palavra mais adequada a usar? Com que outra palavra esta, agora escolhida, deverá fazer par, a fim de construir para o leitor o real sentido da mensagem que se pretende difundir? E a pergunta mais elucidativa: quantas vezes devo repetir a(s) palavra(s) que mais esclarecem a mensagem para o leitor?

Sem dúvida que esse processo de escolhas é totalmente inconsciente e, segundo Freud, na descrição do seu aparelho psíquico, as palavras vão se tornando conscientes como consequência de um mecanismo complexo de ligações, similar ao descrito para os textos escritos, que por sua vez são o resultado desse processo. Dessa maneira, analisando os textos podemos ter uma idéia de como o mecanismo que os gerou funciona. No caso, a mente humana, descrita pelo aparelho psíquico de Freud. Não de uma forma organizada, lógica e determinística, como nos fazem crer algumas abordagens da cibernética. E sim, de uma maneira caótica, porém auto-organizada, a ponto de emergir um texto, que chamamos de mensagem, de tal modo estruturado que facilita o seu entendimento pelo receptor e que apresenta características de otimização semelhantes a outros fenômenos naturais como a cadeia alimentar e a propagação de doenças, dentre outros.

Ao considerarmos o mecanismo de construção dos textos como uma consequência das facilitações entre as imagens mnêmicas que levam às descargas nos signos de descarga lingüística, tornando-os conscientes, podemos inferir que analisando um,

estaremos determinando o mecanismo do outro. Ou seja, a análise da topologia dos textos escritos é uma caracterização também da topologia do uso dos neurônios na formação do pensamento. Sua complexidade, sua característica auto-organizativa e a emergência de propriedades como a consciência e a cognição.

O significado de uma representação-palavra, para Freud, está relacionado à forma como os traços mnêmicos se associam no córtex cerebral, formando representações-objeto, a partir das facilitações (trilhamentos) produzidas pela passagem da energia Q por entre os neurônios. Esses traços mnêmicos estão associados às percepções do mundo exterior e se constituem em função da experiência vivenciada pelo indivíduo. Embora as significações sejam individualizadas, segundo Freud, o mecanismo, o processo como as associações ocorrem são similares, seja na criação dos traços mnêmicos ou na memorização dos mesmos quando realizamos o processo de pensamento. Essa similaridade na topologia da rede de conexões entre os signos de descarga lingüística, resultantes desse processo, pode ser observada na caracterização das redes complexas formadas pelos textos escritos.

E aqui apresentasse a quarta e mais significativa das contribuições sutis propostas. Considerarmos que a estrutura complexa do texto é um retrato do caminho percorrido em nossa “memória neuronal”, e que reflete um caminho possível, dentre os inúmeros disponíveis, na rede de ligações presentes em nossa mente. E, a escolha do caminho a ser percorrido, trata-se de um processo eficiente, já que apresenta características de funcionamento otimizado (auto-organização).

Apesar da inferência sobre a representatividade da análise dos textos relacionada à atividade cerebral, cremos que uma análise física dessa atividade nos daria mais segurança à afirmação. Estudos recentes (SPORNS, 2004; EGUÍLUZ, 2005) sobre o mapeamento da atividade cerebral em ações como ouvir música e estalar os dedos, usando o modelo de redes complexas, encontraram padrões de redes do tipo mundo pequeno e redes livres de escala, nas redes construídas pelas áreas do córtex cerebral envolvidas nessas ações. Para uma determinação mais efetiva, futuros trabalhos são sugeridos relacionados à caracterização da atividade cerebral em ações envolvendo a linguagem, mais precisamente à criação de sentenças e não associações livres de significados.

REFERÊNCIAS

AGUIRRE, Luis Antonio. Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais. 2. ed. rev. e ampl. Belo Horizonte: Editora UFMG, 2004.

ALBERT, R., BARABÁSI, A-L. **Statistical mechanics of complex networks**. *Reviews of Modern Physics* 74, 47, 2002. Disponível em: <<http://cond-mat/0106096>>. Acesso em 15 jun. 2003.

ALBERT, R., JEONG, H., BARABÁSI, A-L. **The diameter of the world wide web**. *Nature* 401, p. 130-131, 1999. Disponível em: <<http://xxx.lanl.gov/abs/cond-mat/9907038>>. Acesso em: 20 jul. 2003.

AZEREDO, José Carlos de. A palavra e suas classes. **Revista eletrônica IDIOMA 21**. Rio de Janeiro: Centro Filológico Clóvis Monteiro – UERJ, 2001. Disponível em: <<http://www2.uerj.br/~letras/idioma-n21-azer.htm>>. Acesso em: 17 jan. 2005.

BARABÁSI, Albert-László. **Linked: How everything is connected to everything else and what it means for Business, Science, and everyday life**. Cambridge: Plume, 2003.

BÍBLIA. Português. **Bíblia sagrada**. Tradução por João Ferreira de Almeida. Brasília: Sociedade Bíblica do Brasil, 1969.

BOAVENTURA NETO, Paulo Osvaldo. **Grafos: teoria, modelos, algoritmos**. 3. ed. rev. e ampl. São Paulo: Editora Edgard Blücher, 2003.

BUCHANAN, Mark. **Nexus: small world and the groundbreaking science of networks**. 1. ed., New York: W. W. Norton & Company, Inc., 2002.

CANCHO, Ramon Ferrer i., SOLÉ, Ricard V. **The National Academy of Sciences**. *Proc Natl Acad Sci U S A.*, 2003 February 4; v. 100, n.3: p. 788–791. doi: 10.1073/pnas.0335980100. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=298679>>. Acesso em: 12 jun. 2004.

CANCHO, Ramon Ferrer i., SOLÉ, Ricard V. **The small world of human language**. *Proc. R. Soc. London, Ser. B* 268, 2261, 2001.

CANCHO, Ramon Ferrer i., SOLÉ, Ricard V., KÖHLER, R. Patterns in syntactic dependency networks. **Physical Review E**, 69, 051915, 2004.

CARTER, Rita. **O livro de ouro da mente: o funcionamento e os mistérios do cérebro humano**. Tradução por Vera de Paula Assis. 3. ed. Rio de Janeiro: Ediouro, 2003.

CAVALLI-SFORZA, Luigi Luca. **Genes, povos e línguas**. Tradução por Carlos Afonso Malferrari. São Paulo: Companhia das Letras, 2003.

COSTA, Luciano da F. **What's in a Name?** Intl. J. Mod. Phys. C, v. 15, n. 1, p. 371-379, 2004. Disponível em: <<http://arxiv.org/abs/cond-mat/0309266>>. Acesso em: 20 fev. 2005.

DESCARTES, René. **Discurso do Método**. 2002. Disponível em: <<http://www.phoenix-library.org/index.php?page=search&ref=7002>>. Acesso em: 21 jan. 2005.

DOROGOVTSSEV, S. N. e MENDES, J. F. F. Language as an evolving word web. [Proceedings of The Royal Society of London. Series B, Biological Sciences](#), v. 268, n. 1485, p. 2603-2606, 2001.

EGUÍLUZ, M., CHIALVO Dante R., CECCHI, Guillermo A., BALIKI, Marwan, APKARIAN, A. Vania. **Scale-Free Brain Functional Networks**. PRL 94, 018102, 2005.

FIORIN, José Luiz (org). **Introdução à Lingüística: objetos teóricos**. 3. ed. São Paulo: Editora Contexto, 2004. v. 1

FIORIN, José Luiz (org). **Introdução à Lingüística: princípios de análise**. 2. ed. São Paulo: Editora Contexto, 2003. v. 2

FREUD, Sigmund. **A Interpretação das Afasias**. Tradução por Antonio Pinto Ribeiro. Lisboa: Edições 70, 1979.

_____. **A Interpretação dos Sonhos**. Tradução por Walderedo Ismael de Oliveira. Licença editorial da Imago. São Paulo: Editora Círculo do Livro, 1987. vol. 2

_____. **Edição Standard das obras psicológicas completas de Sigmund Freud**. Projeto para uma psicologia. Rio de Janeiro: Imago, 1996. vol. 1

GARCIA-ROZA, Luiz Alfredo. **Introdução à Metapsicologia Freudiana**. 5. ed. Rio de Janeiro: Jorge Zahar Editor, 2001. (Vol. 1- Sobre as Afasias e O Projeto de 1895)

KANDEL, Eric R., SCHWARTZ, James H., JESSEL, Thomas M. **Fundamentos da Neurociência e do Comportamento**. Tradução por Charles Alfred Esbérard e Mira Casrilevitz Engelhardt. Rio de Janeiro: Editora Prentice-Hall do Brasil, 1997.

KOPPEL, M., ARGAMON, S., SHIMONI, R. **Automatically categorizing written texts by author gender**. Lit. and Ling. Comp., v. 17, n. 4, 2003. Disponível em: <<http://www.cs.biu.ac.il/~koppel/papers/male-female-llc-final.pdf>>. Acesso em: 31 mar. 2003.

KRISTEVA, Julia. **História da Linguagem**. Tradução por Maria Margarida Barahona. Portugal: Edições 70, 2003.

LEPSCHY, Giulio C. **A Lingüística Estrutural**. 2. ed. São Paulo: Editora Perspectiva, 1975.

LEWIS, E.T., DIESNER, J., CARLEY, K.M. **Using Automated Text Analysis to Study Self-Presentation Strategies.** Computational Analysis of Social and Organizational Systems (CASOS) Conference, Pittsburgh PA, 2001. Disponível em: <http://www.casos.ece.cmu.edu/casos_working_paper/ESP-all.pdf>. Acesso em: 23 nov. 2004.

LI, Wentian. **Comments on bell curves and monkey languages.** *Complexity*. 1(6): pp. 6, 1996. Disponível em: <http://www.nslj-genetics.org/wli/pub/complexity96_lett_pre.pdf>. Acesso em: 14 jun 2004.

LI, Wentian. **Random texts exhibit Zipf's-law-like word frequency distribution.** IEEE transactions on information theory. v. 38, n. 6, p. 1842-1845, 1992.

MARTINS, Nilce Sant'anna. **Introdução à estilística: a expressividade na língua portuguesa.** 3. ed. rev. e aum. São Paulo: T. A. Queiroz, 2003. (Biblioteca universitária de língua e lingüística; v. 8).

MILGRAM, Stanley. **The Small-World Problem.** Psychology Today, v. 2, 1967, p. 60-67.

MONTEMURRO, Marcelo A. **Beyond the Zipf-Mandelbrot law in quantitative linguistics,** 2001. Disponível em: <http://arxiv.org/PS_cache/cond-mat/pdf/0104/0104066.pdf>. Acesso em: 10 jul. 2004.

MONTEMURRO, Marcelo A. e PURY, Pedro A. **Long-range fractal correlations in literary corpora.** ArXiv:cond-mat/0201139, v. 1, 09 jan. 2002.

MOTTER, Adilson E., MOURA, Alessandro P S. de, LAI, Yiang-Cheng, DASGUPTA, Partha. **Topology of the conceptual network of language.** Physical Review E, 65, 065102, 2002.

NEWMAN, M. E. J. **The structure and function of complex networks.** *arXiv.org e-Print archive*, 2003. Disponível em: <http://arxiv.org/PS_cache/cond-mat/pdf/0303/0303516.pdf>. Acesso em: 19 fev. 2005.

NOWAK, M. A., PLOTKIN, J. B., and JANSEN, V. A. A. **The evolution of syntactic communication.** *Nature Magazine*, vol. 404, 2000.

NUSSENZVEIG, H Moysés (org). **Complexidade e Caos.** 2. ed. Rio de Janeiro: Editora UFRJ/COPEA, 2003.

PESSOA JR, O. Medidas Sistêmicas e Organização, in Debrun, M.; Gonzales, M.E.Q. & Pessoa Jr., O. (orgs.), **Auto-Organização**, Coleção CLE 18, CLE-Unicamp, 1996, p. 129-61.

PINHO, Suani T. Rubim de, ANDRADE, Roberto F. Silva. Evolução das idéias da termodinâmica e da mecânica estatística. In: ROCHA, José Fernando M. (org). **Origens e evolução das idéias da física.** Salvador: EDUFBA, 2002. p. 170-180.

REDE RELEX BRASIL. **Manual Unitex 0-1-2-3-4**. Tradução de Aléxis Neme. Bahia: Rede Relex Brasil, 2002.

SANTOS, Volnyr. **Idioma e Soberania**: nossa língua, nossa pátria - empréstimos lingüísticos: tradição e atualidade (CÂMARA NOS 500 ANOS). Disponível em: <http://www.camara.gov.br/internet/Camara500/Seminarios/Id_Volnyr_P3.htm>. Acesso em: 17 jan. 2005.

SAUSSURE, Ferdinand de. **Curso de Lingüística Geral**. 26 ed. São Paulo: Editora Cultrix, 2004.

SCOTT, John. *Social Network Analysis: a handbook*. 2. ed. London: SAGE Publications Ltd., 2000.

SILVA FILHO, Waldomiro José da. Discurso da Ação. In: O Sentido e a Época: **Ensaios sobre cultura na era da comunicação**. 1997. Disponível em: <<http://www.facom.ufba.br/sentido/wal.html>>. Acesso em: 08 ago. 2004.

SILVA, Altigran S. da, CAVALCANTI, João Marcos B., MOURA, Edleno S. de. **Tecnologia Web: Busca na Web**. Disponível em: <<http://www.dcc.fua.br/~alti/tw/slides/TW-Aulas%205%206%207%20-%20Busca%20-%20Alunos.pdf>>. Acesso em: 17 jan. 2005.

SIMON, H. A. **On a Class of Skew Distribution Functions**. *Biometrika*, vol. 42, partes 3 y 4, diciembre, pp. 425-440, 1955.

SOLÉ, Ricard V. FERRER-CANCHO, Ramon. MONTOYA, Jose M. VALVERDE, Sergi. **Selection, tinkering, and emergence in complex networks**. *Wiley periodicals v. 8 n. 1, 2003, p. 20- 33*.

SPORNS, Olaf, CHIALVO, Dante R., KAISER, Marcus, HILGETAG, Claus C. **Organization, Development and function of complex brain networks**. *Trends in Cognitive Sciences*, v. 8, n. 09 , 2004, p. 418-425.

STERNBERG, Robert J. **Psicologia Cognitiva**. Porto Alegre: Artes Médicas Sul Editora, 2000.

TAKAHASHI, Tadao (org.). **Livro Verde**: sociedade da informação no Brasil. Brasília: Ministério da Ciência e Tecnologia, 2000.

WATTS, Duncan J. Small Worlds. **The Dynamics of Networks between Order and Randomness**. USA : Princenton University Press, 1999.

ZIPF, George Kingsley. **The Principle of Least Effort**: an introduction to human ecology. New York: Hafner Publishing Company, 1972.

APÊNDICE A

A.1. Composição dos textos analisados e suas classificações

Código	Quant Palavras Originais	Quant Sentenças	Estilo			Área	Idioma	Nome original
			T ou L	Disse	Tipo			
faaheed8	12972	429	Técnico	Monog		Humanas	Português	A lingüística e a ciência da informação: estudos de uma interseção
fabegep3	49750	1649	Técnico	Disse		Exatas	Inglês	Development Policies for the Biotechnology Industry in Portugal
fabehb7	33754	991	Técnico	Disse		Exatas	Português	A aplicação de ambientes educacionais informatizados de aprendizagem na produção do texto escrito: um estudo de caso
fabehb9	30021	1264	Técnico	Disse		Exatas	Português	Ambiente computacional para o ensino de amplificadores para pequenos sinais
fabehed1	31604	1012	Técnico	Disse		Exatas	Português	Sistemas Inteligentes de Arquitetura Híbrida Neuro-Difusa para Reconhecimento de Padrões
fabehed5	30579	539	Técnico	Disse		Exatas	Português	Normalização de locutor em Sistema de Reconhecimento de Fala
fabehed3	31408	1096	Técnico	Disse		Exatas	Português	Autoria Adaptativa de Hipermídia Educacional
fabhaed2	39262	1511	Técnico	Disse		Humanas	Português	Religião e Razão Comunicativa: as comunidades eclesiais de base no contexto da redemocratização
fabhaed3	74967	3263	Técnico	Disse		Humanas	Português	Homicídios na periferia de Santo Amaro: Um estudo sobre a sociabilidade e os arranjos de vida num cenário de exclusão
fabhaed6	69876	2007	Técnico	Disse		Humanas	Português	TrAbAlHo e QuImErAS: dilema vivido pelo jovem operário
fabhceb5	23858	712	Técnico	Disse		Humanas	Português	Mal-estar, progresso e felicidade em uma vila pesqueira: a construção subjetiva de um grupo de mulheres da Praia Nova
fabhcee5	27285	1060	Técnico	Disse		Humanas	Português	ATLETAS DE ALTO RENDIMENTO: identidades em construção
fabheeb4	25161	809	Técnico	Disse		Humanas	Português	ALGUNS ASPECTOS SOCIAIS EM A PATA DA GAZELA
fabheed0	37205	1270	Técnico	Disse		Humanas	Português	A marcação de caso acusativo na interlíngua de brasileiros que estudam o espanhol
fabheed8	22713	381	Técnico	Disse		Humanas	Português	Análise da fluência verbal de surdos oralizados em português brasileiro e usuários de língua brasileira de sinais

fabhfeb0	26585	839	Técnico	Disse		Humanas	Português	O Perfil das Empresas Familiares de Pequeno e Médio Portes de Fortaleza: um estudo da gestão, da propriedade e da família
fabhfeb3	19733	355	Técnico	Disse		Humanas	Português	Os impactos da reestruturação produtiva no processo de mudança organizacional do setor bancário cearense: um estudo de caso do Banco do Brasil
fabhfeb6	23716	827	Técnico	Disse		Humanas	Português	Sistema de Informação de Marketing: Estudo de caso em uma pequena indústria do setor de estruturas metálicas de Fortaleza
facheed0	46878	1465	Técnico	Tese		Humanas	Português	A vivência e a invenção no cotidiano em Rosa, minha irmã Rosa (Alice Vieira) e O sofá estampado (Lygia Bojunga)
facheed4	75707	2991	Técnico	Tese		Humanas	Português	FERNANDO PESSOA: POETA CÉTICO?
facheep2	59651	1693	Técnico	Tese		Humanas	Português	Análise do comportamento sintático-semântico dos participios passados em português europeu
fadhced1	4852	170	Técnico	Artigo		Humanas	Português	Considerações sobre o lugar do corpo na organização perversa
fadhced3	3359	116	Técnico	Artigo		Humanas	Português	A construção do dispositivo analítico
fadhced4	4438	192	Técnico	Artigo		Humanas	Português	Dois Leonardos
fadhced7	3066	110	Técnico	Artigo		Humanas	Português	Autismo e defesa primária: questões sobre o sujeito e a transferência
fadheab4	10243	347	Técnico	Artigo		Humanas	Português	O MODELO DAS MOTIVAÇÕES COMPETIDORAS NO DOMÍNIO FUNCIONAL DA NEGAÇÃO
fadhead0	12913	554	Técnico	Artigo		Humanas	Português	A LINGÜÍSTICA INDÍGENA NO BRASIL
fadhead1	12329	373	Técnico	Artigo		Humanas	Português	ESTUDOS SOBRE EDUCAÇÃO BILÍNGUE E ESCOLARIZAÇÃO EM CONTEXTOS DE MINORIAS LINGUISTICAS NO BRASIL
fadhead2	5873	117	Técnico	Artigo		Humanas	Português	A etimologia das palavras escondidas na linguagem do jornal
fadheap2	10645	205	Técnico	Artigo		Humanas	Português	Introdução ao Processamento de Linguagem Natural através das Aplicações
fadhebd1	8414	108	Técnico	Artigo		Humanas	Português	A PALAVRA DÍVIDA -Relações semânticas: Um estudo das relações semânticas das palavras em uma nova tipologia de análise do discurso.
fadhebd2	8513	292	Técnico	Artigo		Humanas	Português	A Interação Sincronia/Diacronia no Estudo da Sintaxe
fadhebd4	7984	234	Técnico	Artigo		Humanas	Português	Formas de Funcionalismo na Sintaxe
fadhedd0	5259	170	Técnico	Artigo		Humanas	Português	O DESENVOLVIMENTO DA LINGÜÍSTICA TEXTUAL NO BRASIL
fadheed1	5864	166	Técnico	Artigo		Humanas	Português	Vigotski e a perspectiva enunciativa da relação entre linguagem,

								cognição e mundo social
fadheed2	3961	102	Técnico	Artigo		Humanas	Português	LINGÜÍSTICA TEXTUAL: QUO VADIS?
fadheed7	16955	436	Técnico	Artigo		Humanas	Português	Aquisição da Linguagem: Uma Retrospectiva dos Últimos Trinta Anos
fadheed8	9200	301	Técnico	Artigo		Humanas	Português	MANIFESTO REGIONALISTA DE 1926: proclamação e sagração da "auctoritas" gilbertiana
FADLivro003d	7690	338	Técnico	Artigo		Exatas	Inglês	Technologies of the Self in the Age of Genetics
FADLivro003f	7015	293	Técnico	Artigo		Exatas	Inglês	Animal Species/Genetic Resources
FADLivro003g	9137	464	Técnico	Artigo		Natural	Inglês	For the Love of a Good Dog Webs of Action in the World of Dog Genetics
FADLivro003k	5648	296	Técnico	Artigo		Natural	Inglês	Reflections and Prospects for Anthropological Genetics in South Africa
FADLivro003o	5596	306	Técnico	Artigo		Natural	Inglês	The Good, the Bad, and the Ugly Promise and Problems of Ancient DNA for Anthropology
FADLivro006g	7299	431	Técnico	Artigo		Natural	Inglês	Mouse and Rat Genome Informatics
falhaeb0	6576	164	Técnico	Artigo		Humanas	Português	História, política e erotismo no cinema brasileiro (1960-1985): uma visão histórica
falhaeb9	20098	444	Técnico	Projeto		Humanas	Português	A HISTÓRIA NAS TELAS: um estudo teórico das especificidades, alcances e limites dos discursos históricos cinematográficos, através da análise de bibliografia especializada e dos filmes de Glauber Rocha
falhdeb2	10091	227	Técnico	Projeto		Humanas	Português	Jogos eletrônicos e violência: Desvendando o imaginário dos screenagers
falhdeb8	8801	241	Técnico	Projeto		Humanas	Português	O audiovisual no ensino da História: em busca de uma proposta teórico-metodológica
fczzzeb7	998	72	Literário		Croni		Português	Verdades e mentiras
fczzzeb8	775	72	Literário		Croni		Português	Passado I
fczzzeb9	476	60	Literário		Croni		Português	Passado II
fjmhced2	5641	149	Técnico	Matéria		Humanas	Português	A escuta e o corpo do analista
fjmhced5	2153	77	Técnico	Matéria		Humanas	Português	Os Difíceis Caminhos da Angústia
fjzhced5	718	39	Técnico	Artigo		Humanas	Português	A psicanálise e a memória corporal
fjzzzed0	2106	81	Técnico	Artigo			Português	De corpo inteiro

fjzzzed4	796	36	Técnico	Report			Português	Cia. Matrix de Dança
FLKLivro049	79580	5350	Literário		Livro		Inglês	THE PICTURE OF DORIAN GRAY
FLKLivro052	75727	4057	Literário		Livro		Inglês	FRANKENSTEIN OR THE MODERN PROMETHEUS
FLKLivro053	78083	4424	Literário		Livro		Inglês	NORTHANGER ABBEY
FLKLivro054	84102	4767	Literário		Livro		Inglês	PERSUASION
FLKLivro055	122990	7537	Literário		Livro		Inglês	PRIDE AND PREJUDICE
FLKLivro056	120695	6209	Literário		Livro		Inglês	SENSE AND SENSIBILITY
FLKLivro058	72988	3460	Literário		Livro		Inglês	THE MYSTERIES OF UDOLPHO
FLTLivro005	130942	11479	Técnico	Livro			Inglês	C++ by Dissection
FLTLivro046	60726	3744	Técnico	Livro			Inglês	VALUE SWEEP: Mapping Corporate Growth Opportunities
maahbeb4	18075	803	Técnico	Monog		Humanas	Português	ESTRATÉGIAS PARA FIDELIZAÇÃO DE CLIENTES – MARKETING DE RELACIONAMENTO – O POVO UM “CASE” DE MARKETING
mabegeed5	31624	879	Técnico	Disse		Exatas	Português	PROJETO ESTRUTURAL DE VIGAS DE CONCRETO DE ALTO DESEMPENHO
mabegeed9	44106	1575	Técnico	Disse		Exatas	Português	Implementação de um Sistema de Conversão Texto-Fala para o Português do Brasil
mabegee0	25832	813	Técnico	Disse		Exatas	Português	DESENVOLVIMENTO DE UM AMBIENTE VIRTUAL TRIDIMENSIONAL MULTIUSUÁRIO PARA INTERNET.
mabegep4	42597	1220	Técnico	Disse		Exatas	Inglês	Perspectives for technological innovation for the Portuguese Hand-made glass sector
mabeheb1	27773	643	Técnico	Disse		Exatas	Português	IRME – Inspeção Radiográfica de Materiais e Equipamentos - um software educativo.
mabeheb2	36997	1339	Técnico	Disse		Exatas	Português	Cl@ssAuto - Uma ferramenta WEB de ensino em classificação automática de dados
mabeheb6	27685	951	Técnico	Disse		Exatas	Português	ORGANIZADOR PRÉVIO EM HIPERMÍDIA PARA O ESTUDO DE MÁQUINAS ELÉTRICAS
mabehed2	36110	1041	Técnico	Disse		Exatas	Português	Segmentação e extração de características para reconhecimento automático de caracteres - Estudo e propostas
mabehed3	48320	1138	Técnico	Disse		Exatas	Português	Modelos da Língua para o Português do Brasil Aplicados ao Reconhecimento de Fala Contínua: Modelos Lineares e Modelos

								Hierárquicos (Parsing)
mabehed4	54211	1793	Técnico	Disse		Exatas	Português	Sistemas Inteligentes de Navegação Autônoma: Uma Abordagem Modular e Hierárquica Com Novos Mecanismos de Memória e Aprendizagem
mabehep8	48157	2418	Técnico	Disse		Exatas	Português	Processamento Morfológico e Correção Ortográfica do Português
mabehzb4	13957	948	Técnico	Disse		Exatas	Português	WEBFLAUTA - UMA APLICAÇÃO EAD PARA O ENSINO DA FLAUTA DOCE
mabejed1	15483	309	Técnico	Disse		Exatas	Português	A DISTRIBUIÇÃO DE LEI DE POTÊNCIA GRADUALMENTE TRUNCADA NA EDUCAÇÃO: EXAME DE VESTIBULAR DA UNESP
mabhadd5	93414	3369	Técnico	Disse		Humanas	Português	SEVERINOS NA METRÓPOLE: a negação do trabalho na cidade de São Paulo
mabheed1	44177	1535	Técnico	Disse		Humanas	Português	A POETA AO ESPELHO (Cecília Meireles e o Mito de Narciso)
mabheed9	30488	1537	Técnico	Disse		Humanas	Português	A Mulher na Literatura Portuguesa: sua Imagem e seus Questionamentos através do Gênero Epistolar
maceged6	31885	1103	Técnico	Tese		Exatas	Português	Jogos Computacionais Fonoarticulatórios para Crianças com Deficiência Auditiva
machade0	93720	6349	Técnico	Tese		Humanas	Português	O ATLÂNTICO AÇORIANO: uma antropologia dos contextos globais e locais da açorianidade
macheed2	47861	1671	Técnico	Tese		Humanas	Português	A CONSTRUÇÃO DA METALINGÜÍSTICA (FRAGMENTOS DE UMA CIÊNCIA DA LINGUAGEM NA OBRA DE BAKHTIN E SEU CÍRCULO)
macheed3	78880	2826	Técnico	Tese		Humanas	Português	A FICÇÃO DO REAL EM GILBERTO FREYRE
madhadd6	5197	170	Técnico	Artigo		Humanas	Português	O sentir das estruturas e as estruturas do sentir: a poesia que Lévi nos trouxe
madhaez9	3850	175	Técnico	Artigo		Humanas	Português	A virtualização Lévyana
madhcad0	11976	614	Técnico	Artigo		Humanas	Português	Os jogos de verdade da Psicanálise
madhcad2	11546	495	Técnico	Artigo		Humanas	Português	PSICANÁLISE E SURTO PSICÓTICO: CONSIDERAÇÕES SOBRE ASPECTOS TÉCNICOS
madhcad3	35284	338	Técnico	Artigo		Humanas	Português	SEXUALIDADE E FUNÇÃO ALFA: UM NOVO MODELO PARA AS PERVERSÕES
madhcdz4	5476	266	Técnico	Artigo		Humanas	Português	Agorafobia. Uma tópica evanescente

madhced6	3942	144	Técnico	Artigo		Humanas	Português	Aspectos metapsicológicos das psicopatologias contemporâneas *
madhded6	3310	88	Técnico	Artigo		Humanas	Português	Educação
madhded8	2332	103	Técnico	Artigo		Humanas	Português	Redes Técnicas
madheab5	21531	206	Técnico	Artigo		Humanas	Português	AS DUAS GRANDES VERTENTES DA HISTÓRIA SOCIOLINGÜÍSTICA DO BRASIL (1500-2000)
madhead9	13478	575	Técnico	Artigo		Humanas	Português	LINGÜÍSTICA DE CORPUS: HISTÓRICO E PROBLEMÁTICA
madheee1	1744	70	Técnico	Artigo		Humanas	Português	Os Estudos de Sintaxe Generativa em Portugal nos Últimos Trinta Anos.
madheep7	2558	85	Técnico	Artigo		Humanas	Português	Estudo léxico-informático de 10 canções de Camões
madhlbd4OK	8855	304	Técnico	Artigo		Humanas	Português	COMPILAÇÃO E ANOTAÇÃO DE UM CORPUS DE PORTUGUÊS DE LINGUAGEM PROFISSIONAL.
MADLivro003b	7490	340	Técnico	Artigo		Natural	Inglês	Human Populations/Genetic Resources
MADLivro003c	6400	295	Técnico	Artigo		Natural	Inglês	Victor McKusick's Fieldwork with the Old Order Amish
MADLivro003h	7743	422	Técnico	Artigo		Natural	Inglês	98% Chimpanzee and 35% Daffodil The Human Genome in Evolutionary and Cultural Context
MADLivro003j	9560	523	Técnico	Artigo		Natural	Inglês	Future Imaginaries Genome Scientists as Sociocultural Entrepreneurs
MADLivro003m	7774	380	Técnico	Artigo		Natural	Inglês	Human Races in the Context of Recent Human Evolution A Molecular Genetic Perspective
MADLivro003n	7347	342	Técnico	Artigo		Natural	Inglês	Buried Alive The Concept of Race in Science
MADLivro006d	11719	595	Técnico	Artigo		Natural	Inglês	Human Genetic Variation: Databases and Concepts
MADLivro006e	8107	439	Técnico	Artigo		Natural	Inglês	Finding, Delineating and Analysing Genes
MADLivro006f	8390	421	Técnico	Artigo		Natural	Inglês	Assembling a View of the Human Genome
MADLivro006h	7632	380	Técnico	Artigo		Natural	Inglês	Genetic and Physical Map Resources-an Integrated View
MADLivro006i	5261	253	Técnico	Artigo		Natural	Inglês	From Linkage Peak to Culprit Gene: Following Up Linkage Analysis of Complex Phenotypes with Population-based Association Studies
MADLivro006j	8850	480	Técnico	Artigo		Natural	Inglês	Genetic Studies from Genomic Sequence
MADLivro006m	7759	359	Técnico	Artigo		Natural	Inglês	Predictive Functional Analysis of Polymorphisms: An Overview
MADLivro006n	5781	287	Técnico	Artigo		Natural	Inglês	Functional In Silico Analysis of Non-coding SNPs

MADLivro006o	7882	502	Técnico	Artigo		Natural	Inglês	Amino Acid Properties and Consequences of Substitutions
MADLivro006r	1968	98	Técnico	Artigo		Natural	Inglês	Concluding Remarks: Final Thoughts and Future Trends
MAEC	1469	75	Técnico	Artigo		Humanas	Português	Email sobre religião, ciencia, fé e razão.
malehee00K	12289	553	Técnico	Projeto		Exatas	Português	Uma Análise de Linguagens de Especificação para Sistemas Distribuídos
malhddb00K	1232	40	Técnico	Projeto		Humanas	Português	Educação para pescadores
MAnietszche	84526	6653	Técnico	Livro		Humanas	Português	Assim falou Zaratustra
MASonhos	58139	3091	Técnico	Livro		Humanas	Português	Análise de Sonhos
mcedgaralanpoe01	2794	417	Literário		Poema		Inglês	AL AARAAF
mcedgaralanpoe02	3790	163	Literário		Cronica		Inglês	THE ANGEL OF THE ODDIAN EXTRAVAGANZA
mcedgaralanpoe03	4530	215	Literário		Cronica		Inglês	THE ASSIGNATION
mcedgaralanpoe04	5180	319	Literário		Cronica		Inglês	THE BALLOON-HOAX
mcedgaralanpoe05	3999	206	Literário		Cronica		Inglês	THE BLACK CAT
mcedgaralanpoe06	5598	268	Literário		Cronica		Inglês	BON-BON
mcedgaralanpoe07	2387	179	Literário		Cronica		Inglês	THE CASK OF AMONTILLADO
mcedgaralanpoe08	3428	203	Literário		Cronica		Inglês	THE COLLOQUY OF MONOS AND UNA
mcedgaralanpoe10	7111	240	Literário		Cronica		Inglês	A DESCENT INTO THE MAELSTROM
mcedgaralanpoe11	3186	158	Literário		Cronica		Inglês	THE DEVIL IN THE BELFRY
mcedgaralanpoe12	3962	256	Literário		Cronica		Inglês	DIDDLING
mcedgaralanpoe13	5894	264	Literário		Cronica		Inglês	THE DOMAIN OF ARNHEIM
MCedgaralanpoe14	2397	109	Literário		Cronica		Inglês	ELEONORA
MCedgaralanpoe16	3588	211	Literário		Cronica		Inglês	THE FACTS IN THE CASE OF M. VALDEMAR
MCedgaralanpoe17	7283	347	Literário		Cronica		Inglês	THE FALL OF THE HOUSE OF USHER
MCedgaralanpoe18	2732	210	Literário		Cronica		Inglês	FOUR BEASTS IN ONE THE HOMO-CAMELEOPARD
MCedgaralanpoe19	13934	585	Literário		Livro		Inglês	THE GOLD-BUG

MCedgaralanpoe20	15217	617	Literário		Livro		Inglês	HANS PHAALL
MCedgaralanpoe21	3588	166	Literário		Cronica		Inglês	HOP-FROG OR THE EIGHT CHAINED OURANG-OUTANGS
MCedgaralanpoe22	2409	140	Literário		Cronica		Inglês	THE IMP OF THE PERVERSE
MCedgaralanpoe23	1995	79	Literário		Cronica		Inglês	THE ISLAND OF THE FAY
MCedgaralanpoe24	4749	149	Literário		Cronica		Inglês	KING PEST
MCedgaralanpoe25	4804	215	Literário		Cronica		Inglês	LANDOR'S COTTAGE
MCedgaralanpoe26	2474	158	Literário		Cronica		Inglês	X-ING A PARAGRAB
MCesquizof1	323	11	Literário		Comum		Português	Texto dito de um esquizofrênico
MCesquizof2	169	7	Literário		Comum		Português	Texto dito de um esquizofrênico II
MCesquizof3	677	25	Literário		Comum		Português	Texto dito de um esquizofrênico III
mjmhced6	4305	146	Técnico	Matéria		Humanas	Português	Por uma teoria da depressão generalizada
mjmzzed4	3266			Matéria		Interdisciplinar	Português	CONGRESSO SOBRE INTERDISCIPLINARIDADE REALIZADO EM AGOSTO NA UNIVERSIDADE STANFORD (EUA) DEU ORIGEM AO DOCUMENTO, QUE VÊ NO CONCEITO DE EMERGÊNCIA A "MÃO INVISÍVEL" QUE ARTICULA A COLABORAÇÃO ENTRE ÁREAS COMO FILOSOFIA, QUÍMICA, BIOLOGIA E TEORIA LITERÁRIA
		120	Técnico					
mjzzzed1	2269	86	Técnico	Artigo		Humanas	Português	Da mobilidade inventiva da escuta à inteligibilidade dos conceitos.
mjzzzed2	963	46	Técnico	Artigo		Natural	Português	O Fenômeno Chamado NBA
mjzzzed8	1081	48	Técnico	Artigo		Natural	Português	FCT-Presidente Prudente participará do Campeonato Paulista de Basquetebol sobre Rodas 2002
mjzzzed9	1107	23	Técnico	Artigo		Natural	Português	O Básico no Futsal
mL1984	87794	5950	Literário		Livro		Português	1984
mLAdmMundoNovo	72748	4045	Literário		Livro		Português	ADMIRÁVEL MUNDO NOVO
mLalegria	37727	1550	Literário		Livro		Português	PROJETO PERIFERIA - A Alegria da Revolução Ken Knab
mLAlmaHomem	13372	850	Literário		Livro		Português	A ALMA DO HOMEM - (SOB O SOCIALISMO)
mLAmoreAnarquia	4063	336	Literário		Poema		Português	Amor e Anarquia

MLcodvinci	151065	9818	Literário		Livro		Português	O Código Da Vinci
MLContosInac	16744	719	Literário		Contos		Português	Contos Inacabados
MLdidatica.txt	67389	3340	Técnico	Livro		Humanas	Português	DIDÁTICA
mljhed5OK	28873	1498	Técnico	Livro		Exatas	Português	Curso de Linguagem C - CPDEE / UFMG
mljhaez9OK	20152	489	Técnico	Livro		Humanas	Português	DISCURSO DO MÉTODO - DESCARTES
mljhdez5OK	68051	2781	Técnico	Livro		Humanas	Português	DIDÁTICA
MLKDubliners	67636	4773	Literário		Livro		Inglês	Dubliners - James Joyce
MLKedgaralanpoe09	78487	4929	Literário		Livro		Inglês	CRITICISM
MLKedgaralanpoe15	38638	1505	Literário		Livro		Inglês	EUREKA A PROSE POEM
MLKfinnegansWake	206548	14751	Literário		Livro		Inglês	Finnegans Wake - James Joyce
MLKLivro012	350577	16791	Literário		Livro		Inglês	Anna Karenina
MLKLivro050	253100	17243	Literário		Livro		Inglês	EARTHBOUND
MLKPortrait	86033	5067	Literário		Livro		Inglês	A Portrait of the Artist as a Young Man
MLKshakespeare01	24261	3940	Literário		Teatro		Inglês	All's Well That Ends Well
MLKshakespeare02	11708	1801	Literário		Teatro		Inglês	As You Like It
MLKshakespeare03	14707	2461	Literário		Teatro		Inglês	The Comedy of Errors
MLKshakespeare04	17388	3178	Literário		Teatro		Inglês	Cymbeline
MLKshakespeare05	11079	1886	Literário		Teatro		Inglês	Loves Labours Lost
MLKshakespeare06	4525	802	Literário		Teatro		Inglês	Measure for Measure
MLKshakespeare07	12150	1882	Literário		Teatro		Inglês	The Merry Wives of Windsor
MLKshakespeare08	22082	2804	Literário		Teatro		Inglês	The Merchant of Venice
MLKshakespeare09	17062	2262	Literário		Teatro		Inglês	A Midsummer Night's Dream
MLKshakespeare10	22439	2285	Literário		Teatro		Inglês	Much Ado About Nothing
MLKshakespeare11	14728	1961	Literário		Teatro		Inglês	Pericles, Prince of Tyre
MLKshakespeare12	16272	2008	Literário		Teatro		Inglês	The Taming of the Shrew

MLKshakespeare13	12235	1819	Literário		Teatro		Inglês	The Tempest
MLKshakespeare14	27413	3670	Literário		Teatro		Inglês	Troilus and Cressida
MLKshakespeare15	21328	2321	Literário		Teatro		Inglês	Twelfth Night
MLKshakespeare16	13067	1651	Literário		Teatro		Inglês	Two Gentlemen of Verona
MLKshakespeare17	25808	3543	Literário		Teatro		Inglês	Winter's Tale
MLKshakespeare18	25908	2957	Literário		Teatro		Inglês	The First part of King Henry the Fourth
MLKshakespeare19	27702	3126	Literário		Teatro		Inglês	The Second part of King Henry the Fourth
MLKshakespeare20	27366	3150	Literário		Teatro		Inglês	The Life of King Henry the Fifth
MLKshakespeare21	16017	2122	Literário		Teatro		Inglês	The First part of King Henry the Sixth
MLKshakespeare22	26667	3212	Literário		Teatro		Inglês	The Second part of King Henry the Sixth
MLKshakespeare23	25808	3172	Literário		Teatro		Inglês	The Third part of King Henry the Sixth
MLKshakespeare24	25870	3870	Literário		Teatro		Inglês	The Life of King Henry the Eighth
MLKshakespeare25	21648	2905	Literário		Teatro		Inglês	The Life and Death of King John
MLKshakespeare26	23791	3078	Literário		Teatro		Inglês	The Life and Death of Richard the Second
MLKshakespeare27	31245	3980	Literário		Teatro		Inglês	The Life and Death of Richard the Third
MLKshakespeare28	26837	4273	Literário		Teatro		Inglês	Antony and Cleopatra
MLKshakespeare29	29125	4124	Literário		Teatro		Inglês	The Tragedy of Coriolanus
MLKshakespeare30	32065	4164	Literário		Teatro		Inglês	The Tragedy of Hamlet, Prince of Denmark
MLKshakespeare31	19293	2751	Literário		Teatro		Inglês	The Life and Death of Julies Caesar
MLKshakespeare32	27619	3904	Literário		Teatro		Inglês	King Lear
MLKshakespeare33	13083	1996	Literário		Teatro		Inglês	The Tragedy of Macbeth
MLKshakespeare34	27827	3866	Literário		Teatro		Inglês	Othello, the Moore of Venice
MLKshakespeare35	17950	2319	Literário		Teatro		Inglês	Romeo and Juliet
MLKshakespeare36	19562	2664	Literário		Teatro		Inglês	Timon of Athens
MLKshakespeare37	21620	2780	Literário		Teatro		Inglês	Titus Andronicus

mlkzzed6OK	47717	2223	Literário		Livro		Português	Pedro Mundim
mlkzzez3OK	25104	2832	Literário		Livro		Português	Romeu e Julieta - William Shakespeare
mlkzzez4OK	65591	4007	Literário		Livro		Português	A Volta ao Mundo em 80 Dias
mlkzzez6OK	61720	4310	Literário		Livro		Português	A ALMA ENCANTADORA DAS RUAS
mlkzzez7OK	22638	2210	Literário		Livro		Português	Muito Barulho por nada - William Shakespeare
mlkzzez8OK	58151	1985	Técnico	Livro		Humanas	Português	Crítica da Razão Pura - Emmanuel Kant
MLTLivro001	70771	3659	Técnico	Livro		Exatas	Inglês	UML Pattern Language, A
MLTLivro003	79615	4942	Literário		Livro		Inglês	THE ALTIAN PLAGUE
MLTLivro011	45574	3732	Técnico	Livro		Exatas	Inglês	THE DESIGN PATTERNS JAVA COMPANION
MLTLivro013	29795	2656	Técnico	Livro		Exatas	Inglês	Extreme Programming Explored
MLTLivro020	76123	3593	Técnico	Livro		Exatas	Inglês	Effective Java: Programming Language Guide
MLTLivro022	73024	4599	Técnico	Livro		Exatas	Inglês	Beginning Perl for Bioinformatics
MLTLivro025	79225	2633	Técnico	Livro		Exatas	Inglês	Computationalism
MLTLivro030	82373	3959	Técnico	Livro		Humanas	Inglês	The Business Start-Up Kit
MLTLivro033	96077	5715	Técnico	Livro		Exatas	Inglês	Enterprise Application Integration
MLTLivro034	99244	7963	Técnico	Livro		Exatas	Inglês	Building a Geodatabase
MLTLivro035	65086	3273	Técnico	Livro		Exatas	Inglês	THE ORIGINS OF FIBER OPTIC COMMUNICATIONS
MLTLivro042	166233	8314	Técnico	Livro		Exatas	Inglês	NETWORK HACK PROOFING YOUR INTERNET TRADECRAFT
MLTLivro044	73737	3249	Técnico	Livro		Humanas	Inglês	Geeks & Geezers: How Era, Values, and Defining Moments Shape Leaders
MLTNielsBohr1	31300	928	Técnico	Livro		Exatas	Inglês	Atomic theory and the description of nature
MLTNielsBohr2	38213	964	Técnico	Livro		Exatas	Inglês	Essays 1933-1957 on Atomic Physics and Human knowledge
MLTNielsBohr3	40202	1085	Técnico	Livro		Exatas	Inglês	Essays 1958-1962 on Atomic Physics and Human knowledge
MLTNielsBohr4	78633	2143	Técnico	Livro		Exatas	Inglês	Causality and complementarity
MLTomBombadil	9054	620	Literário		Livro		Português	AS AVENTURAS DE TOM BOMBADIL E OUTRAS HISTÓRIAS

mpfz3bz3OK	1394	168	Literário		Livro		Português	Poemas
MLKulysses-ing	265534	23753	Literário		Livro		Inglês	Ulysses - James Joyce (inglês)
MLKulysses-port	276425	30617	Literário		Livro		Português	Ulysses - James Joyce (português)
zabeied6	28813	798	Técnico	Disse		Exatas	Português	Problemas de otimização de natureza numérica
zadehaz1	19019	587	Técnico	Disse		Exatas	Português	utilização da computação evolutiva, em particular de sistemas classificadores
zadehdd5	5243	183	Técnico	Artigo		Exatas	Português	A caminho de um ambiente de avaliação e acompanhamento contínuo da aprendizagem em Programação de Computadores
zadhadd7	5514	151	Técnico	Artigo		Humanas	Português	De onde surgem os modelos? As origens e expansões Tupi na Amazônia Central
zadhcd1	4743	170	Técnico	Artigo		Humanas	Português	DEBATE (29): Psicanálise e políticas públicas
zadhde1	4798	167	Técnico	Artigo		Humanas	Português	Modelagem e Implementação da Interface para Apresentação de Comportamentos Animados e Emotivos de um Agente Pedagógico Animado
ZADLivro003a	11186	542	Técnico	Artigo		Natural	Inglês	Anthropology and Science beyond the Two-Culture Divide
ZADLivro003e	7247	368	Técnico	Artigo		Natural	Inglês	The Commodification of Virtual Reality
ZADLivro003i	8048	371	Técnico	Artigo		Natural	Inglês	Political and Cultural Identity
ZADLivro003l	4729	277	Técnico	Artigo		Natural	Inglês	The Genetics of African Americans Implications for Disease Gene Mapping and Identity
ZADLivro006a	999	51	Técnico	Artigo		Natural	Inglês	BIOINFORMATICS FOR GENETICISTS
ZADLivro006b	6945	308	Técnico	Artigo		Natural	Inglês	Introduction: The Role of Genetic Bioinformatics
ZADLivro006c	6907	374	Técnico	Artigo		Natural	Inglês	Internet Resources for the Geneticist
ZADLivro006k	4083	247	Técnico	Artigo		Natural	Inglês	SNP Discovery and PCR-based Assay Design: From In Silico Data to the Laboratory Experiment
ZADLivro006l	8420	548	Técnico	Artigo		Natural	Inglês	Tools for Statistical Analysis of Genetic Data
ZADLivro006p	9427	559	Técnico	Artigo		Natural	Inglês	Gene Expression Informatics and Analysis
ZADLivro006q	7785	380	Técnico	Artigo		Natural	Inglês	Proteomic Informatics
zalhdbd5	7907	444	Técnico	Relat		Humanas	Português	Framework para Aprendizagem à Distância Através da Internet

zBíbliaNT	165937	6895	Literário		Livro		Português	A Bíblia Sagrada - Novo Testamento
zBíbliaVT_cantico	2727	103	Literário		Livro		Português	A Bíblia Sagrada - Cântico dos Cânticos
zBíbliaVT_Cronicas	43496	1718	Literário		Livro		Português	A Bíblia Sagrada - I Crônicas
zBíbliaVT_Daniel	11674	438	Literário		Livro		Português	A Bíblia Sagrada - Daniel
zBíbliaVT_Deuteronomio	27027	858	Literário		Livro		Português	A Bíblia Sagrada - Deuteronômio
zBíbliaVT_Eclesiastes	5514	221	Literário		Livro		Português	A Bíblia Sagrada - Eclesiastes
zBíbliaVT_Esdras	7689	212	Literário		Livro		Português	A Bíblia Sagrada - Esdras
zBíbliaVT_Ester	5412	157	Literário		Livro		Português	A Bíblia Sagrada - Ester
zBíbliaVT_Exodo	32484	1290	Literário		Livro		Português	A Bíblia Sagrada - Êxodo
zBíbliaVT_ezequiel	36473	1490	Literário		Livro		Português	A Bíblia Sagrada - Ezequiel
zBíbliaVT_final	23170	1086	Literário		Livro		Português	A Bíblia Sagrada - Joel
zBíbliaVT_Genesis	35854	1890	Literário		Livro		Português	A Bíblia Sagrada - Antigo Testamento - Gênesis
zBíbliaVT_Isaias	34776	1631	Literário		Livro		Português	A Bíblia Sagrada - Isaías
zBíbliaVT_jeremias	42637	1898	Literário		Livro		Português	A Bíblia Sagrada - Jeremias
zBíbliaVT_Jo	17864	654	Literário		Livro		Português	A Bíblia Sagrada - Jó
zBíbliaVT_Josue	18919	675	Literário		Livro		Português	A Bíblia Sagrada - Josué
zBíbliaVT_juizes	16869	786	Literário		Livro		Português	A Bíblia Sagrada - Juizes
zBíbliaVT_Levitico	22040	736	Literário		Livro		Português	A Bíblia Sagrada - Levítico
zBíbliaVT_Neemias	9852	382	Literário		Livro		Português	A Bíblia Sagrada - Neemias
zBíbliaVT_Numeros	30007	1087	Literário		Livro		Português	A Bíblia Sagrada - Números
zBíbliaVT_Oseias	4872	273	Literário		Livro		Português	A Bíblia Sagrada - Oséias
zBíbliaVT_Proverbios	15206	655	Literário		Livro		Português	A Bíblia Sagrada - Provérbios
zBíbliaVT_Reis	44073	2002	Literário		Livro		Português	A Bíblia Sagrada - I Reis
zBíbliaVT_Rute	2987	167	Literário		Livro		Português	A Bíblia Sagrada - Rute

zBíbliaVT_Salmos	41420	1724	Literário		Livro		Português	A Bíblia Sagrada - Salmos
zBíbliaVT_Samuel	42783	2197	Literário		Livro		Português	A Bíblia Sagrada - I Samuel
zdzbbz7	7574	250	Técnico	Artigo		Humanas	Português	Direitos das Crianças - Documentos das Nações Unidas
zdzbbz8	896	47	Técnico	Artigo		Humanas	Português	Declaração dos Direitos da Criança
zjzhcdd7	4782	203	Técnico	Artigo		Humanas	Português	A tradição freudiana de Donald Winnicott - A situação edípica. E sobre o pai?
zjzhced3	3457	145	Técnico	Artigo		Humanas	Português	Artigos
zjzhced4	3325	142	Técnico	Artigo		Humanas	Português	Artigos II
zjzhced5	3177	118	Técnico	Artigo		Humanas	Português	Artigos III
zjzhced6	3525	141	Técnico	Artigo		Humanas	Português	Artigos IV
zjzhced8	2206	79	Técnico	Artigo		Humanas	Português	Artigos V
zjzzzed0	479	13	Técnico	Artigo		Humanas	Português	APREENDENDO CIDADANIA NA EDUCAÇÃO FÍSICA ESCOLAR
zjzzzed6	381	10	Técnico	Artigo		Humanas	Português	Quem foi James Naismith?
zjzzzed7	885	35	Técnico	Artigo		Humanas	Português	O Valor do Esporte no Desenvolvimento Social das Crianças
zjzzzed9OK	2581	97	Técnico	Artigo		Natural	Português	Carta Brasileira de Educação Física:
zjzzzee1OK	796	33	Técnico	Artigo		Natural	Português	Saúde e Coração: Especialista Destaca Importância dos Exercícios Físicos Para a
zjzzzee7OK	565	34	Técnico	Artigo		Natural	Português	Paraná Esporte - Apresentação
Zlivespir	140848	7457	Literário		Livro		Português	Livro dos Espíritos
ZLKLivro051	165806	6448	Literário		Livro		Inglês	EMMA
ZLKLivro057	11854	7319	Literário		Livro		Inglês	THE LEGEND OF SLEEPY HOLLOW
ZLTLivro006	170294	10067	Técnico	Livro		Biológicas	Inglês	BIOINFORMATICS FOR GENETICISTS
ZLTLivro007	132423	3340	Técnico	Livro		Exatas	Inglês	Agile Software Development Ecosystems
ZLTLivro008	153935	3803	Técnico	Livro		Exatas	Inglês	Formal Syntax and Semantics of Programming Languages A Laboratory Based Approach
ZLTLivro009	50249	4679	Técnico	Livro		Exatas	Inglês	Extreme Programming Explained

ZLTLivro010	56228	4752	Técnico	Livro		Exatas	Inglês	Installing Locks in the Global Village
ZLTLivro015	105117	2331	Técnico	Livro		Exatas	Inglês	Programming/Software Engineering
ZLTLivro016	97961	4381	Técnico	Livro		Exatas	Inglês	J2EE Connector Architecture and Enterprise Application Integration
ZLTLivro017	68981	2162	Técnico	Livro		Exatas	Inglês	The Java™ Native Interface - Programmer's Guide and Specification
ZLTLivro018	72760	5882	Técnico	Livro		Exatas	Inglês	LDAP PROGRAMMING WITH JAVA
ZLTLivro019	35194	3428	Técnico	Livro		Exatas	Inglês	Planning Extreme Programming
ZLTLivro024	80622	4006	Técnico	Livro		Exatas	Inglês	Process and performance Isolated problem Process
ZLTLivro028	89887	4725	Técnico	Livro		Exatas	Inglês	Data Munging with Perl
ZLTLivro037	102564	2049	Técnico	Livro		Exatas	Inglês	Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications
ZLTLivro038	98065	10302	Técnico	Livro		Exatas	Inglês	Anthropology and Science beyond the Two-Culture Divide
ZLTLivro045	44731	539	Técnico	Livro		Humanas	Inglês	Hiring and Keeping the Best People

"Enquanto eu tiver perguntas e não houver respostas, continuarei a escrever."

Clarice Lispector