

**TADEU ROSSATO BISOGNIN**

**DO *INTERNETÊS* AO LÉXICO DA ESCRITA DOS JOVENS  
NO ORKUT**

**PORTO ALEGRE  
2008**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS  
ESPECIALIDADE: TEORIAS DO TEXTO E DO DISCURSO  
LINHA DE PESQUISA: LEXICOGRAFIA E TERMINOLOGIA:  
RELAÇÕES TEXTUAIS

**DO *INTERNETÊS* AO LÉXICO DA ESCRITA DOS JOVENS  
NO ORKUT**

**TADEU ROSSATO BISOGNIN**

**Orientadora: Prof<sup>a</sup> Dr<sup>a</sup> Maria José Bocorny Finatto**

**Dissertação de Mestrado em Teorias do Texto  
e do Discurso, apresentada como requisito  
parcial para a obtenção do título de Mestre  
pelo Programa de Pós-Graduação em Letras  
da Universidade Federal do Rio Grande do Sul.**

**PORTO ALEGRE  
2008**

**BANCA EXAMINADORA**

Prof. Dr. Oto Araújo Vale (UFSCar)

Prof. Dr. Valdir do Nascimento Flores (UFRGS)

Profa. Dra. Cleci Regina Bevilacqua (UFRGS)

---

## **AGRADECIMENTO ESPECIAL**

Sinto-me profundamente agradecido à

Professora Doutora Maria José Bocorny Finatto,

pela maneira segura e competente com que conduziu a orientação deste trabalho, por me fazer gostar de pesquisa científica, por me ensinar que pesquisa pode tocar o coração, por ser exemplo de pesquisadora incansável, pelas aulas de vida, enfim. Menciono, sobretudo, a atenção a mim dedicada, respeitando meu tempo/processo de conhecimento e fazendo com que uma tarefa árdua se transformasse num percurso prazeroso. Mostrou-me como subjugar o Minotauro para que eu conseguisse encontrar a saída do labirinto.

## AGRADECIMENTOS

*Um galo sozinho não tece uma manhã:  
Ele precisará sempre de outros galos.  
(João Cabral de Melo Neto)*

*Deo meo, omnino.*

A meus pais, mentores de mim, pelo exemplo.

Ao Colégio de Aplicação da UFRGS, pela licença concedida para qualificação.

Aos professores da Pós-Graduação, Anna Maria Becker Maciel, Cleci Regina Bevilacqua, Felix Bugueño Miranda, Maity Siqueira e Valdir do Nascimento Flores, pelo crescimento proporcionado, pelas estradas apontadas.

Ao Prof. Aduino Locatelli Taufer, pela amizade, estímulo, sugestões, pela eficiência, rapidez e paciência com que revisou esta dissertação.

Aos professores Aduino Taufer, Andréa Perrot e Juçara Freitas, pela cedência de turmas e aplicação da proposta de produção de textos escolares para elaboração de *corpus*.

Às colegas professoras de Francês, Carolina Fernandes e Maria da Graça Saraiva Marques, pelo estímulo, incentivo e eficiência na transformação do RESUMO em RÉSUMÉ do dia para o dia.

A Glades Dilelio Noble e aos bolsistas do Projeto Termisul, por acolhida e auxílio, em especial a Siane Simioni e Sue Anne Coimbra, que disponibilizaram dados do Projeto TextQuim, enviaram pôsteres, dirimiram dúvidas e incentivaram.

Aos colegas professores Gláucia de Souza, Juçara Benvenuti, Juçara Freitas, Margareth Sudbrack, Marlon de Almeida, pelo estímulo, consideração e confiança.

Às colegas do mestrado, Carolina Fernandes, Karol Ribeiro Silva Ferreira, Renata Beneduzi e Susana de Azeredo, pelo apoio, empatia, sorrisos, estímulo e empréstimo de livros e ouvidos.

Aos meus afetos maiores: Maria Irene, pelo apoio irrestrito, estímulo constante e paciente compreensão de meu trabalho; Alexandre, por me apoiar, fazer e me ensinar a fazer figuras e outros quejandos no computador; Marcos Eduardo e Henrique, pelo auxílio e por compartilharem as dificuldades e alegrias desta tarefa.

A André Luiz Batista Figueredo, por me ouvir pacientemente, por me estimular nos momentos de crise e por partilhar a alegria das conquistas.

A Pollyana Lobos Severo, por digitar *ipsis litteris* todo o *corpus* das redações escolares.

Aos meus alunos, que tanto me ensinam.



Criação de Silveira Neto sobre pintura de Vincenzo Camuccini. Disponível em: [www.eupodiatamatando.com](http://www.eupodiatamatando.com)

“Palavras são tecidas a partir de uma multidão de fios ideológicos e servem de trama a todas as relações sociais em todos os domínios. É, portanto, claro que a palavra será sempre o indicador mais sensível de todas as transformações sociais”.

(Bakhtin, M. *Marxismo e filosofia da linguagem*, 1999)

## RESUMO

Este trabalho descreve características da escrita empregada em textos de recados (*scraps*) e de depoimentos (*testimonials*) do *site* de relacionamentos Orkut. Essa escrita é tomada como uma das variedades do *internetês* e o ponto de partida para a descrição é a frequência dos itens de vocabulário. O suporte teórico para a pesquisa é dado pela Lingüística de *Corpus*, área que analisa com auxílio do computador os padrões de uso da língua em grandes conjuntos de textos reais, para observar de modo empírico as formas gramaticais possíveis e prováveis utilizadas pelos falantes. Os *corpora* analisados foram 1) *corpus* de estudo formado por depoimentos e recados com 553.875 palavras, 2) *corpora* de referência formados por amostras de língua escrita e falada compostos de 1.289.949 palavras e 3) *corpora* de contraste compostos de textos escolares, jornalísticos e didáticos, totalizando 571.090 palavras. Analisaram-se os *corpora* com a ferramenta *WordSmith Tools*, empregada pelos estudos em Lingüística de *Corpus*. Entre as regularidades e especificidades dos dados sob exame, deu-se maior atenção à riqueza vocabular presente nos textos dos internautas, às palavras mais frequentes, às variações ortográficas - inclusive por regiões do Brasil demonstradas em nove *subcorpora* -, aos indícios de oralização sobre a escrita e à variedade de forma e conteúdo. A pesquisa permitiu concluir que o *internetês* é basicamente uma expressão gráfica com alterações ortográficas presentes em torno de 20% de seu léxico, com variações quanto à riqueza vocabular em diferentes regiões do país, variação essa maior nos *scraps* do que nos depoimentos. Percebeu-se que não são significativas as diferenças de riqueza vocabular entre textos do Orkut e os de redações escolares e de textos jornalísticos premiados, não podendo ser o vocabulário dos internautas considerado pobre no quesito riqueza lexical. Quanto à frequência, as palavras da escrita do Orkut mostraram-se bastante similares às do português falado, de modo que o *internetês* se apresenta num *continuum* em que a oralidade e escrita se fundem. O trabalho postula que o *internetês*, tal como representado no Orkut, é um dialeto diastrático, um socioleto dos internautas jovens, com alguns poucos traços repercutindo sobre a escrita escolar. Na sua parte final, a pesquisa traz sugestões sobre como a escrita sob exame pode ser explorada em atividades em sala de aula.

**PALAVRAS-CHAVE:** escrita - *internetês* – Lingüística de *Corpus* – lexicologia –Orkut

## RÉSUMÉ

Ce travail décrit des caractéristiques de l'écrit employées dans des textes de messages et des dépositions au site de rapports Orkut. Cette écriture c'est une des variétés de l'*internetês* et le point de départ pour la description c'est la fréquence du vocabulaire. Le support théorique de la recherche est donné par la Linguistique du *Corpus*, qui analyse avec l'aide de l'ordinateur l'usage de la langue dans des textes réels afin d'observer de façon empirique les formes grammaticales possibles utilisées par les locuteurs. Les *corpora* analysés ont été: 1) *corpus* de l'étude formés par des dépositions et messages avec 553.875 mots, 2) *corpora* de référence formés par échantillons de langue écrite et orale avec 1.289.949 mots, et 3) *corpora* de contraste composés par des textes scolaires, journalistiques et didactiques, au total de 571.090 mots. Les *corpora* ont été analysés l'instrument WordSmith Tools, employé par des études en Linguistique du *Corpus*. Parmi les régularités et les spécificités des données examinées, on a remarqué la richesse du vocabulaire dans des textes des internautes aux mots les plus fréquents, aux variations orthographiques y inclus par des régions du Brésil représentées dans neuf subcorpora, aux indices d'oralisation sur l'écrite et à la variété de forme et de contenu. La recherche a conclu que l'*internetês* est une expression graphique avec des modifications orthographiques de environ 20% de son lexique, avec des variations touchant à la richesse du vocabulaire dans des différentes régions du pays. Cette variation a été plus grande dans des textes de messages. On a remarqué que les différences de richesse vocabulaire entre les textes de Orkut, les rédactions scolaires et les textes journalistiques primés ne sont pas significatives. Ce qui signifie affirmer que le vocabulaire des internautes n'est pas pauvre à propos de la richesse lexicale. À l'égard de la fréquence, les mots de l'écrite d'Orkut sont semblables aux mots d'un *continuum* où l'oralité et l'écrite se fondent. Le travail postule que l'*internetês*, comme il est représenté au Orkut, est un dialecte diastratique, un sociolecte des jeunes internautes avec un peu des traits qui repercutent sur l'écrite scolaire. À la fin, la recherche suggère comme l'écrite examinée peut être explorée en classe.

**MOTS-CLÉS:** écrite – internetês – Linguistique du *Corpus* – lexicologie - Orkut

## LISTA DE TABELAS

Tabela 1: <i>Tokens, types e ratio por corpora</i> (Ordem decrescente da <i>ratio</i> ) .....	133
Tabela 2: Palavras mais freqüentes dos <i>subcorpora</i> Manaus, Belém e Brasília .....	135
Tabela 3: Palavras mais freqüentes dos <i>subcorpora</i> Cuiabá, Porto Alegre e Recife	136
Tabela 4: Palavras mais freqüentes dos <i>subcorpora</i> Rio Branco, Rio de Janeiro e Salvador .....	136
Tabela 5: Freqüência <i>Corpus</i> de <i>scraps</i> e Freqüência <i>Corpus</i> reduzido de <i>scraps</i> ...	137
Tabela 6: Soma de QUE e Q na lista de freqüência dos <i>subcorpora</i> regionais .....	138
Tabela 7: Palavras mais freqüentes nos <i>corpora</i> de produções dos alunos de 7 <sup>a</sup> série e 3 <sup>o</sup> ano do Ensino Médio .....	158
Tabela 8: <i>Tokens, types e ratio por corpora</i> do <i>internetês</i> .....	165
Tabela 9: <i>Tokens, types e ratio</i> do <i>corpus</i> de reportagens sobre biodiversidade .....	168
Tabela 10: <i>Types</i> , hápax e percentagem de hápax dos <i>corpora</i> .....	172
Tabela 11: Percentagem da freqüência de uso na escrita, na fala e no <i>internetês</i> .....	185
Tabela 12: Posição das palavras no <i>ranking</i> das mais freqüentes nos diferentes <i>corpora</i> .....	186
Tabela 13: Verbos e substantivos entre as cem palavras mais freqüentes .....	213

## LISTA DE QUADROS

Quadro 1: Condições de produção de fala e escrita .....	67
Quadro 2: Lista das dez primeiras palavras do <i>Corpus Geral</i> em ordem alfabética .....	118
Quadro 3: Formas de escrever <i>abraço/abraços</i> (da lista de palavras do <i>Corpus Geral</i> )	119
Quadro 4: As dez primeiras palavras mais freqüentes da lista de palavras do <i>Corpus Geral</i> .....	119
Quadro 5: Primeiras dez concordâncias da palavra <i>cmg</i> (=comigo) de um total de 407 ocorrências no <i>Corpus Geral</i> .....	119
Quadro 6: Os dez primeiros <i>clusters</i> (reunião de três palavras) contendo <i>cmg</i> (= comigo) de 64 ocorrências .....	120
Quadro 7: Dados Estatísticos e lista das palavras mais freqüentes do <i>Corpus Geral</i> do Orkut .....	127
Quadro 8: Palavras mais freqüentes e freqüência no entorno do “E” .....	130
Quadro 9: Exemplos de concordâncias para a forma “á” .....	131
Quadro 10: Oito primeiros exemplos da concordância da palavra “Â” .....	132
Quadro 11: Concordâncias do QUE no <i>corpus</i> geral .....	141
Quadro 12: Concordâncias do Q no <i>corpus</i> geral .....	142
Quadro 13: Concordâncias do QUE no <i>corpus</i> reduzido de scraps .....	142
Quadro 14: Concordâncias do Q no <i>corpus</i> reduzido de scraps .....	143
Quadro 15: Estatística do <i>Corpus Falado</i> NURC-RJ .....	147
Quadro 16: As palavras mais empregadas nos recados/depoimentos e as palavras mais utilizadas do Banco de Português escrito .....	148
Quadro 17: Possíveis formas do <i>internetês</i> nos textos escolares .....	152
Quadro 18: Estatística dos <i>corpora</i> de produções dos alunos de 7ª série e 3º ano EM .....	157

Quadro 19: Relação de frequências de palavras na escrita escolar, <i>internetês</i> e norma culta escrita (Banco de Português) .....	159
Quadro 20: Relação fala culta NURC-RJ, escrita escolar e fala culta Banco de Português .....	160
Quadro 21: Dados estatísticos e palavras mais frequentes dos <i>Corpora Superinteressante</i> , Química Geral e entalpia .....	161
Quadro 22: <i>Tokens, types</i> e <i>ratio</i> do <i>corpus</i> do português falado NURC-RJ .....	166
Quadro 23: Dados estatísticos gerais dos <i>corpora</i> de produções dos alunos de 7 <sup>a</sup> série e 3 <sup>o</sup> ano do EM .....	169
Quadro 24: Estatística do <i>corpus</i> reduzido de redações .....	169
Quadro 25: Primeiras e as últimas vinte <i>hapax legomena</i> do <i>corpus</i> do <i>internetês</i> .....	171
Quadro 26: Palavras mais empregadas no Orkut e no Banco de Português .....	175
Quadro 27: Diferentes posições de cinco palavras na ordem de frequência .....	175
Quadro 28: Palavras mais frequentes no Orkut e na fala do Banco do Português ....	177
Quadro 29: Frequência de formas diferentes de mesma palavra e <i>Corpus</i> Geral do Orkut .....	179
Quadro 30: Palavras mais frequentes do <i>corpus</i> reduzido de <i>scraps</i> e do <i>Corpus Geral do Orkut</i> .....	180
Quadro 31: Palavras mais empregadas no português escrito, no português falado e no <i>internetês</i> .....	181
Quadro 32: Palavras mais frequentes nos <i>corpora</i> NURC-RJ e Banco de Português	182
Quadro 33: Palavras mais frequentes nos <i>corpora</i> de língua falada culta, <i>internetês</i> e língua falada .....	184
Quadro 34: Distribuição de diferentes formas lexicais por região do Brasil .....	195
Quadro 35: Características de escrita, fala e <i>internetês</i> , inspirado no quadro das diferenças entre escrita e fala de Shepherd (1984) .....	199

## LISTA DE FIGURAS

<b>Figura 1: Léxicos em volta do núcleo comum .....</b>	<b>42</b>
<b>Figura 2: Exemplificação de núcleo comum e léxicos específicos .....</b>	<b>43</b>
<b>Figura 3: Atos lingüísticos concretos, norma e sistema .....</b>	<b>52</b>
<b>Figura 4: Representação da heterogeneidade da composição do <i>internetês</i> .....</b>	<b>105</b>
<b>Figura 5: Exemplo de página com perfil de usuário do Orkut .....</b>	<b>108</b>
<b>Figura 6: Número de letras das palavras <i>do corpus</i> de estudo .....</b>	<b>129</b>
<b>Figura 7: Influências sofridas pelo <i>internetês</i> .....</b>	<b>130</b>
<b>Figura 8: Equilíbrio da composição dos <i>corpora</i> regionais .....</b>	<b>134</b>
<b>Figura 9: Palavras do <i>internetês</i> alteradas graficamente e sem alteração .....</b>	<b>193</b>
<b>Figura 10: Ocorrências de formas variáveis das palavras .....</b>	<b>206</b>
<b>Figura 11: Uso da língua: equilíbrio entre aceitabilidade e adequabilidade .....</b>	<b>207</b>

## SUMÁRIO

<b>INTRODUÇÃO, POR QUE E PARA QUE .....</b>	<b>15</b>
Justificativa .....	18
Problemas de pesquisa .....	20
Objetivos, questões de pesquisa e hipóteses .....	21
Breve apresentação do <i>corpus</i> sob estudo .....	23
<b>1.REVISÃO DE LITERATURA .....</b>	<b>25</b>
<b>1.1 NOÇÕES FUNDAMENTAIS .....</b>	<b>26</b>
1.1.1 Língua, sistema de signos inter-relacionados .....	26
1.1.2 Palavra, algo difícil de definir .....	33
1.1.3 Léxico, repertório de vocábulos .....	40
1.1.4 Norma, uso e “bom uso” .....	48
1.1.5 Variantes, formas alternativas com mesmo valor e função.....	55
1.1.5.1 Dialeto, algo diatópico e/ou diastrático .....	60
1.1.5.2 Idioleto, forma individual de expressão .....	63
1.1.6 Escrita, uma representação .....	65
<b>1.2 O <i>INTERNETÊS</i> .....</b>	<b>70</b>
1.2.1 O <i>miguxês</i> .....	79
<b>1.3 LINGÜÍSTICA DE <i>CORPUS</i>, empirismo e probabilidade .....</b>	<b>83</b>
<b>2. POSICIONAMENTO DO TRABALHO .....</b>	<b>90</b>
<b>3. MATERIAIS E MÉTODOS, O QUE E COMO .....</b>	<b>106</b>
3.1 O TERRITÓRIO DO <i>CORPUS</i> DE ESTUDO .....	107
3.2 CONSTITUIÇÃO DOS <i>CORPORA</i> .....	108
3.3 AS FERRAMENTAS UTILIZADAS .....	117
<b>4.DESCRICÃO DOS <i>CORPORA</i> .....</b>	<b>121</b>
<b>4.1O TIPO DE TEXTO EM FOCO .....</b>	<b>124</b>
<b>4.2. DIVISÕES DO <i>CORPUS</i> DE PESQUISA .....</b>	<b>126</b>
4.2.1 <i>Corpus</i> geral do Orkut .....	127
4.2.2 <i>Corpora Regionais</i> e <i>corpus</i> de <i>Scraps</i> .....	133

4.3 <i>CORPORA</i> DE REFERÊNCIA .....	145
4.4 <i>CORPORA</i> DE CONTRASTE .....	149
<b>5. REGULARIDADES E ESPECIFICIDADES NOS DADOS SOB EXAME</b> .....	<b>164</b>
5.1 RIQUEZA VOCABULAR .....	164
5.2 PALAVRAS MAIS FREQUÊNTES .....	174
5.3. ALTERAÇÕES DE GRAFIA – POR GRUPOS REGIONAIS DO ORKUT .....	185
5.4. INDÍCIOS DE ORALIZAÇÃO SOBRE A ESCRITA .....	197
5.5 VARIEDADE DE FORMA E CONTEÚDO .....	203
5.6 SÍNTESE DAS OBSERVAÇÕES .....	207
<b>6 RESULTADOS DA DESCRIÇÃO E PERSPECTIVAS</b> .....	<b>211</b>
6.1 RETOMADA DAS QUESTÕES DE PESQUISA .....	211
6.2 AVALIAÇÃO DAS HIPÓTESES .....	217
6.3 CONSIDERAÇÕES E PERSPECTIVAS .....	219
<b>7 SUGESTÕES PARA O PROFESSOR: O <i>INTERNETÊS</i> NA SALA DE AULA</b> .....	<b>230</b>
7.1 . ALGUMAS CONSIDERAÇÕES SOBRE AS ATIVIDADES PROPOSTAS.....	242
<b>REFERÊNCIAS</b> .....	<b>244</b>
<b>ANEXOS</b> .....	<b>250</b>
ANEXO I - Amostra do <i>corpus</i> de <i>scraps</i> .....	250
ANEXO II - Amostra do <i>corpus</i> de depoimentos .....	251
ANEXO III - Proposta de produção de texto para Ensino Médio .....	252
ANEXO IV - Amostra do <i>corpus</i> de redações escolares .....	253
ANEXO V- As cem palavras mais usadas no <i>subcorpus</i> escrito do Banco de Português .....	254
ANEXO VI - As cem palavras mais usadas no <i>subcorpus</i> falado do Banco de Português .....	255
ANEXO VII - As cem palavras mais frequentes do <i>corpus geral do internetês</i> .....	256
ANEXO VIII - As cem palavras mais frequentes do <i>corpus falado NURC- RJ</i> .....	257
ANEXO IX - Palavras típicas do <i>internetês</i> entre as 500 mais frequentes do <i>corpus geral do Orkut</i> .....	258
ANEXO X - Palavras típicas do <i>internetês</i> entre as 500 mais frequentes do <i>corpus apenas scraps</i> .....	259

## INTRODUÇÃO, POR QUE E PARA QUE

**E**ste trabalho dedica-se a descrever o léxico presente na escrita do *internetês*, tomando como referência um *corpus* de 553 mil palavras obtido de textos de depoimentos e de recados do *site* de relacionamentos Orkut. Esse *corpus*, produzido por redatores de diferentes regiões do Brasil, é contrastado com diferentes amostras de uso de escrita em língua portuguesa, incluindo amostras de redação escolar.

A vida na modernidade passou a exigir uma expressão lingüística ajustada à rapidez da comunicação pela Internet. Tal expressão, aparentemente alterada na sua escrita para economizar tempo, utiliza símbolos, gírias, expressões estrangeiras e abreviaturas das mais diferentes formas. À primeira vista, não parece seguir norma alguma. Esse novo jeito de comunicar, de acordo com Bagno (2007, p.20) “do ponto de vista dos usuários da língua, podemos dizer que é bom, porque dá provas de vitalidade e capacidade de adaptação às exigências da vida moderna”.

O objetivo desta pesquisa é descrever um tipo de escrita utilizada principalmente por jovens no cenário da comunicação eletrônica. O foco principal incide sobre o léxico empregado, em um trabalho que parte do quantitativo para o qualitativo. O ponto de vista adotado está guiado pelos princípios e metodologia da Lingüística de *Corpus* (LC). Esse tipo de enfoque permite fazer afirmações fundamentadas sobre o funcionamento da língua em uso real com base na observação em grandes amostras, os *corpora*.

É inegável a relação entre linguagem, língua e sociedade. Saussure (2004, p. 14) a isso já se referiu quando afirmou que “a linguagem é um fato social”. Acrescentou ainda que “os costumes duma nação têm repercussão na língua e, por outro lado, é em grande parte a

língua que constitui a Nação” (SAUSSURE, 2004, p. 29). As línguas refletem sociedades e suas épocas, sendo o léxico a parte que mais nitidamente acompanha as alterações sociais. Uma sociedade com novas descobertas necessita nomeá-las, cria novos vocábulos ou amplia a significação dos já existentes. O vocabulário e também a sua forma de expressão indicam a evolução e o estado de desenvolvimento científico, tecnológico e cultural de uma sociedade. É possível afirmar que, sob alguns aspectos, o século XX foi pródigo em transformações, culminando com o surgimento do computador e da comunicação em rede. Nesse ambiente cultural marcado pelo advento da Internet, destacamos o componente lexical em situação de escrita.

A Internet é um novo ambiente de enunciação<sup>1</sup> cultural, com múltiplas linguagens, possibilidade de interações, velocidade acelerada de informação e estrutura multimidiática. Ela suscita e expressa um ambiente de comunicação diferenciado. Isso pode ser percebido até nas formas de escrever utilizadas pelos internautas, principalmente pelos jovens<sup>2</sup>, na comunicação eletrônica: interferem sobre a escrita culta padrão para interagir. Escrever teclando no computador, especialmente *on-line*, é certamente algo que induziu a inovações, principalmente pela velocidade que se precisa dar àquilo que se transmite por escrito. Como é possível notar, a Internet revolucionou o mundo das comunicações em muito pouco tempo como nenhuma outra invenção antes fez. Nenhuma tecnologia ou instrumento anteriormente empregado conseguiu uma tão grande sinergia.

Com o surgimento da Internet<sup>3</sup>, disseminou-se entre seus usuários, não apenas entre os jovens, uma escrita particular e específica: o chamado *internetês* ou PT-SMS<sup>4</sup>. Em termos

---

<sup>1</sup> O termo *enunciação cultural* entendido como expressão ou manifestação de conhecimentos, informações e saberes do ser humano. Não fazemos aqui relação com a terminologia benvenistiana.

<sup>2</sup> Vale notar que a escrita de internautas mais velhos, em torno dos 45-50 anos de idade, tal como podemos pessoalmente testemunhar, mostra-se em geral diferente justamente por não exibir tais alterações. Desse modo, para um grupo de jovens, seria fácil identificar um redator mais velho.

<sup>3</sup> O termo *Internet*, no *Dicio.net* (2007), consta de dois verbetes: 1) **Internet (com I maiúsculo)**: Um vasto conjunto de redes em todo o mundo, interligadas entre si, que utilizam os protocolos baseados no conjunto TCP/IP. A Internet evoluiu a partir da ARPANET, criada nos finais dos anos 60 e é hoje uma rede em constante evolução em todas as vertentes - técnicas, comerciais e sociais - à qual se ligam milhões de pessoas e empresas em todo o mundo. 2) **internet (com i minúsculo)**: Termos para designar duas ou mais redes interligadas, mas habitualmente usado apenas com "i" maiúsculo, tendo assim um significado muito particular, designando a Internet que todos conhecemos. Já na *Wikipédia* (2007) encontramos: A Internet é um conglomerado de redes em escala mundial de milhões de computadores interligados pelo Protocolo de Internet que permite o acesso a informações e todo tipo de transferência de dados. Ao contrário do que normalmente se pensa, Internet não é sinônimo de World Wide Web. Esta, que utiliza hipermídia na formação básica, é parte daquela. Alguns dos serviços disponíveis na Internet, além da Web, são o acesso remoto a outras máquinas (Telnet e SSH), transferência de arquivos, correio eletrônico, boletins eletrônicos, bate-papo *on-line*, mensagens instantâneas, Skype, Orkut, etc. De acordo com dados de março de 2007, a Internet é usada por 16,9% da população mundial

lingüísticos, em geral se destacam duas modalidades de usos da linguagem: o código escrito e o código oral. Mas, com a difusão da comunicação pela rede mundial de computadores, surgiu uma nova modalidade que parece englobar características das duas. Isto é, há aparentemente, um novo “código escrito oralizado”.

A nova modalidade de expressão, construída em um espaço reduzido para acomodar caracteres, utiliza um grande número de abreviaturas, siglas, neologismos, palavras cifradas, estrangeirismos, desenhos, ícones, símbolos... Também se codificam emoções, como risos, por exemplo, com o auxílio de palavras ou até mesmo apenas com um amontoado de letras. Vejamos alguns exemplos:

ahuahuhahua

ahhhhhhhhhhhhhhh

giuuuuuuuuu xD:\*\*\*\*\*s2 (, =) heuiheuiha, euiehueihueiueiueiueh

Bjooooooooos :\*\*\*\*\* #)

Ainda AmO x ) vc +++ q d+ .:=DDD

"eheheh"

Hehehehheheh

Bjaum t adoruuuuuu

jinhusssssss

pohhhh!!!!!!

!! Bjks gatXinha :@@

“KKKKKKKKKKkk" e "rsrsrs".

Esses sinais justapostos constituem um sentido. Entendê-los, muitas vezes, não é fácil para quem não tem familiaridade com a comunicação via Internet.

---

(em torno de 1,1 bilhão de pessoas). Acreditamos ser pertinente trazer informações da fonte Wikipédia, embora seja considerada controversa como fonte confiável, assim como o próprio *Dicio.net* (op.cit). Por ter estrutura colaborativa e atualização fácil e permanente, a Wikipédia contém assuntos, como o que abordamos na p. 78, publicados bem antes das obras impressas. Trata-se de uma enciclopédia *on-line* livre, colaborativa, isto é, escrita, em conjunto e internacionalmente, por várias pessoas, todas elas voluntárias. *Enciclopédia livre* significa que qualquer artigo publicado nela pode ser transcrito, modificado e ampliado, desde que preservados os direitos de cópia e modificações. É uma obra que possui 1,6 milhões de verbetes e 3,1 milhões de artigos, oferecida em 257 idiomas. Apesar de ser de conteúdo aberto, a Wikipédia conquistou a aprovação — no tocante à confiabilidade de suas informações — da renomada revista científica *Nature*, da Inglaterra. O modelo "wiki" é uma rede de páginas web contendo as mais diversas informações, que podem ser modificadas por qualquer pessoa através de navegadores como Internet Explorer ou Mozilla Firefox. Acesso à versão em português: <http://pt.wikipedia.org>

<sup>4</sup> PT-SMS é a sigla formada por PT (indicação universal de português) e SMS, acrônimo de *Short Message Service* (Serviço de Mensagens Curtas), um serviço disponível em telefones celulares digitais para o envio de mensagens com até 255 caracteres. No Brasil utiliza-se *torpedo* como sinônimo de SMS. O primeiro SMS foi criado em 1992.

O termo *internetês* é aqui utilizado para denominar a linguagem surgida e empregada na Internet. Baseia-se na simplificação informal da escrita para tornar mais ágil a comunicação. Como vimos nos exemplos, vale-se de abreviações, simplificações e símbolos criados por combinação de caracteres. Também pode fazer uso de símbolos gráficos próprios e de uma grande diversidade de recursos de comunicação por imagens disponíveis na Internet. O *internetês* é utilizado principalmente em bate-papos (*chats*), em *sites* de relacionamento como o Orkut, *blogues*<sup>5</sup>, mensagens de celulares, fóruns de Internet e correio eletrônico (*e-mail*)<sup>6</sup>. É um todo de escrita que pode ser subdividido em diversas instâncias.

A nossa opção pelo estudo da escrita do *internetês*, tal como exibida no Orkut, um *site* de relacionamentos que simula um ponto de encontro para trocas assíncronas de vários tipos, distingue-nos de algumas pesquisas já feitas sobre a escrita que partiram dos textos de *e-mails* ou de *blogues*. Ao mesmo tempo em que instaura um recorte para o vasto território da escrita na Internet, oferece uma situação de comunicação que tem características específicas: o Orkut pode ser visto como uma gigantesca “vitrine” de pessoas que se expõem e se comunicam de modo assíncrono. Além disso, os usuários brasileiros são campeões mundiais em sua utilização. Trata-se, assim, de um ambiente massivamente conhecido e utilizado por internautas brasileiros. Isso, sem dúvida, autoriza a sua colocação entre os objetos de estudo desse tipo de comunicação. Uma idéia sobre a constituição do material recolhido para exame pode ser obtida pela observação das amostras de escrita que estão nos Anexos 1 e 2.

### **Justificativa**

Ao fazer um estudo da escrita utilizada na Internet, tomando por base a Lingüística de *Corpus*, acreditamos poder gerar mais instrumentos para análise e compreensão do código lingüístico empregado pelos jovens. Além disso, esperamos contribuir para a inclusão dessa modalidade de expressão também entre os estudos de lexicologia.

O que chama a atenção, já no primeiro contato com o *internetês*, é o seu léxico e as formas pelas quais é registrado graficamente. Pretendemos iniciar a observação e a descrição

---

<sup>5</sup> *Blog*, abreviação de *weblog*, é o registro freqüente de informações *on-line*. A maioria das pessoas o utiliza como diário pessoal no qual publica histórias, idéias ou imagens.

<sup>6</sup> *E-mail*, *blog*, *orkut* e *chat* são ferramentas da Internet que propiciam comunicação entre pessoas do mundo inteiro. As três primeiras são assíncronas, escrevemos mensagem e sua resposta só é recebida em outro momento. Já o *chat* (ou bate-papo) é síncrono: escrevemos mensagem e obtemos a resposta no mesmo instante, *on-line*.

dessa forma de comunicação pela porta do léxico. *Léxico*<sup>7</sup> é um termo entendido aqui no seu sentido mais comum, o de conjunto das palavras de uma língua.

A riqueza vocabular de uma pessoa geralmente é considerada um reflexo de seu nível de educação e, para alguns, também de inteligência. Daí por que haver questões de vocabulário em muitos testes de inteligência. A palavra *vocabulário* abrange vários significados: equivale a léxico (o conjunto dos vocábulos de uma língua), a conjunto de palavras conhecidas por um indivíduo ou empregadas por um autor, e, na terminologia das Ciências da Computação, significa “o conjunto de símbolos com os quais se constrói a linguagem informática e sua decodificação” (HOUAISS, 2001). Para este trabalho, *vocabulário* é visto, inicialmente, como o conjunto de palavras ou expressões empregadas na comunicação, o conjunto de realizações discursivas das unidades lexicais. É justamente o vocabulário do *internetês* que queremos descrever e mensurar neste trabalho

Esse tipo de descrição, acreditamos, ainda é uma carência no cenário dos estudos de Linguística Aplicada. Sem uma descrição sistemática, ficamos reféns de opiniões que são, em geral, impressões bastante subjetivas sobre esse tipo de escrita..

Ao dominar a escrita, o ser humano cresce cognitivamente. A escrita é fator de inserção e de prestígio social. Além disso, nossa consciência lingüística provém muito mais do que lemos ou escrevemos do que daquilo que falamos ou ouvimos. Daí a valorização do escrever de acordo com as convenções em vigor. Kato (1995, p. 31) afirma que a modalidade escrita tem como condições de uso de linguagem: ser menos dependente do contexto situacional, permitir um planejamento verbal mais cuidadoso, ser mais sujeita a convenções prescritivas e ser um produto permanente. Mas, o que valerá disso, dessas considerações, para a escrita da Internet?

Uma vez que a linguagem escrita é um sistema gráfico convencionado em determinada comunidade lingüística, buscar entender quais os meios gráficos utilizados na comunicação dos internautas jovens vale também para auxiliar a entender esses jovens e poder ajudá-los na sua comunicação fora e dentro do ambiente da Internet. Verificar que novas combinações de palavras são empregadas e que novos significados esses internautas dão a velhas palavras pode contribuir para que se compreenda melhor como são os jovens do século XXI. Entendendo sua maneira de ser e de se comunicar no ambiente virtual, a escola poderá aproveitar esse

---

<sup>7</sup> Essa noção encontra-se devidamente discutida no capítulo de Revisão de Literatura, seção 1.1.3.

conhecimento para incentivar o interesse por situações comunicativas diferentes. Confrontar esses redatores com a sua própria escrita pode auxiliar a que reflitam sobre a sua própria comunicação em meio a diferentes cenários de interação.

### **Problemas de pesquisa**

Suportes especiais pedem escritas especiais. A alteração do suporte (tijolo, papiro, papel) foi fator de mudança na escrita e também na leitura. Escrever em letras de forma é o modo mais indicado para pedra e tábuas de argila. Escrever teclando, especialmente *on-line*, também induz a inovações. Mas essas inovações, em geral, são simples e rapidamente julgadas como deturpações indesejáveis.

O primeiro problema de pesquisa a ser enfrentado nesta dissertação é a falta de informação sistematizada e empiricamente embasada sobre as características do vocabulário do *internetês*. Colaborar para diminuir esse vazio nos estudos lingüísticos sobre a linguagem dos jovens na Internet é pretensão desta pesquisa. Queremos, também, contribuir para que tenhamos uma real medida para alguns “julgamentos sumários” sobre esse uso da escrita, muitas vezes bastante apressados.

Há falta de informação sobre como realmente é o fenômeno de comunicação dos jovens na rede mundial de computadores. Não são abundantes os trabalhos científicos sobre o *internetês*<sup>8</sup> no Brasil. Isso deixa professores, principalmente os de Língua Portuguesa, sem um posicionamento seguro sobre como lidar com essa forma de usar a língua. Para o ensino, faltam estudos confiáveis com a descrição do fenômeno “a língua portuguesa da Internet brasileira”.

A Internet proporciona um ambiente para desenvolvimento das capacidades cognitivas e comunicativas do ser humano. A comunicação eletrônica promove a interação social, podendo tornar mais ricas as práticas socioculturais. Aparentemente há uma forma específica de comunicação utilizada em relação com a língua materna culta. Como já mencionamos, alguns tendem a perceber isso como uma deturpação do código culto, capaz inclusive de

---

<sup>8</sup> São dignos de registro, em que pese a falta, os trabalhos de FONSECA (2001), COSTA (2005), ARAÚJO e BIASI-RODRIGUES (2005), POSSENTI (2006), PIMENTEL (2006) e EISENKRAEMER (2006). Esses trabalhos encontram-se indicados na nossa Revisão de Literatura.

“contaminar” a língua portuguesa. Alguns puristas consideram que há um empobrecimento da língua, pelo excesso de abreviaturas, desrespeito às normas ortográficas, utilização de diferentes letras para indicar mesmo fonema, excessiva utilização de onomatopéias, abandono da acentuação e da pontuação, frases curtas e incompletas. Enfim, por tudo fugir à norma-padrão. Mas, seria isso mesmo verdadeiro? Ou, ao contrário, seria um enriquecimento, uma vez que a tela do computador é um novo suporte para registrar a escrita?

O nosso segundo problema de pesquisa surge na necessidade de esclarecer o que é essa modalidade de escrita, com destaque para a feição do seu léxico. Cabe salientar que o cenário específico de observação é fornecido pela escrita no Orkut, um cenário entre tantos da Internet.

### **Objetivos, questões de pesquisa e hipóteses**

Como já mencionado, o objetivo da pesquisa é descrever as características da escrita dos jovens utilizada na comunicação eletrônica, tomando como ponto de partida o vocabulário. A linguagem a ser estudada é a própria língua materna que, aparentemente, pode ter sofrido modificações para economizar tempo e, no caso dos torpedos de celular, também dinheiro. O ponto-chave que inicialmente se destaca é a redução da escrita, a exemplo dos telegramas, que tinham seu preço calculado pela dimensão das palavras e obrigavam o redator a gerar uma mensagem sintética. Mas importante também são as alterações de significação<sup>9</sup>. Queremos saber se “velhas” palavras, colocadas de forma peculiar no contexto da comunicação eletrônica, constituem um novo sentido para os internautas. O foco do trabalho, porém, está muito mais centrado em uma descrição panorâmica do léxico, em suas unidades, de um modo abrangente. As considerações semânticas estão presentes, mas não é objetivo da pesquisa oferecer um estudo aprofundado sobre semântica lexical, tampouco sobre morfologia, sintaxe, fonologia, etimologia ou outro componente gramatical. Nosso trabalho prende-se à observação de formas e de suas variedades, sobretudo em suas frequências e distribuições.

---

<sup>9</sup> Adotaremos neste trabalho os conceitos de Coseriu (1979b) para *significado*, *sentido* e *designação*. Assim, *significado* é o conteúdo de um signo lingüístico em sentido estrito, é a configuração das possibilidades de designação. *Designação* é a aplicação de um significado num ato de fala. Relaciona-se com a realidade extralingüística, com a própria realidade (representação, fato ou acontecimento). *Sentido* é o conteúdo especial dum texto. Só existe sentido no plano do texto, no ato da fala de um falante numa determinada situação, e não no falar em geral ou nas línguas. Empregaremos *significação* como um equivalente de *significado*.

Dado esse objetivo, as questões de pesquisa a serem investigadas são as seguintes:

- 1) Como é a configuração lexical da escrita da Internet no Orkut em relação à variante culta do português brasileiro?
- 2) Qual a relação entre o vocabulário do *internetês* e o vocabulário norma culta de fala?
- 3) Que itens do vocabulário de variações de grafia do *internetês* são verificados em *corpora* que espelhem a escrita escolar?
- 4) O *internetês* pode ser considerado uma variante, um dialeto ou um subsistema em relação aos usos da língua portuguesa espelhados em *corpora* representativos da norma culta escrita?

Levando em conta o objetivo e as questões de pesquisa, pretendemos verificar a validade e a abrangência das seguintes hipóteses de trabalho:

- 1) A variedade do vocabulário do *internetês* é similar à variedade do vocabulário da escrita culta em suas diferentes apresentações.
- 2) As frequências do léxico do *internetês* revelam uma junção entre padrões de fala e de escrita.

Essas duas hipóteses, em conjunto, espelham certa confiança na especificidade e na padronização desse tipo de escrita. Supomos que alie fala e escrita, sendo diferente da escrita padrão, mas também esperamos que se assemelhe a ela por exibir padrões de variação de vocabulário mais ou menos coincidentes com a da escrita culta “normal”.

Será tomado como parâmetro de comparação entre o léxico da escrita do *internetês* e o da variante culta do português o conjunto de dados chamado Banco de Português, composto por dois *subcorpora* distintos: o *subcorpus* escrito e *subcorpus* falado. O primeiro contém palavras de diferentes tipos de textos da linguagem culta, enquanto o segundo é composto de registros falados de aulas e de conversação.

O segundo elemento de comparação será a escrita escolar. Essa escrita será aqui representada por um *corpus* de redações escolares. Essas redações foram produzidas por sujeitos de condições semelhantes aos que produzem a escrita aqui em foco, o *internetês*.

### Breve apresentação do *corpus* sob estudo

Cabe lembrar que o nosso *corpus* sob estudo é constituído de textos do Orkut chamados de depoimentos (*testimonials*) e de recados (*scraps*), mensagens, na maioria das vezes curtas, trocadas entre internautas como se fossem *e-mails*. Tais recados e depoimentos foram coletados de várias regiões do Brasil, totalizando 553.875 palavras (*tokens*). Percebemos, no início da recolha, que em alguns lugares havia algumas peculiaridades. Formamos, então, nove *subcorpora*, cada um deles com mais de trinta mil palavras, nomeados pela capital do estado: Belém, Brasília, Cuiabá, Manaus, Porto Alegre, Recife, Rio Branco, Rio de Janeiro e Salvador. Além desses *subcorpora* regionais, constituímos também um *subcorpus* apenas de *scraps* de todo o Brasil. Esses *subcorpora* visam subsidiar nossas observações sobre possíveis variações diatópicas nessa escrita.

Em sua estrutura, esta dissertação não poderia fugir das formalidades, tendo introdução, objetivos, materiais e métodos, resultados, discussão e conclusões. Assim, após a introdução, no primeiro capítulo tratamos da fundamentação teórica, dividida em três partes: a) noções fundamentais abordando *língua, palavra, léxico, norma, variantes, dialeto* e *escrita*; b) o *internetês* e sua derivação, o *miguxê*; e c) a Lingüística de *Corpus*. Em seguida, temos, no Capítulo 2, o posicionamento teórico em que deixamos claro quais os pressupostos adotados para nos auxiliar a perceber melhor o objeto de estudo. Isso se mostrou necessário principalmente porque nas diferentes opiniões recolhidas percebemos muitas controvérsias sobre determinados conceitos, o que nos obrigou a fazer opções por determinadas abordagens e não por outras. No Capítulo 3 apresentamos o Orkut, como foram constituídos os *corpora* (de estudo, de contraste e de referência) e as ferramentas da Lingüística de *Corpus* utilizadas. Já no Capítulo seguinte, observamos o tipo de texto do *corpus* sob análise e os *corpora* de contraste e de referência. No 5º Capítulo, temos o registro e a análise dos dados, percebendo as regularidades e especificidades do *internetês*. São examinadas as palavras mais freqüentes, a riqueza vocabular dos textos, as variações ortográficas, os indícios de oralização sobre a escrita e a variedade de forma e conteúdo das palavras. No derradeiro capítulo apresentamos as considerações finais, após retomada dos questionamentos e hipóteses mencionadas nesta introdução do trabalho. Finalmente, no último capítulo, propomos sugestões aos professores para aproveitar o *internetês* em sala de aula, utilizando essa escrita como um aliado para levar o aluno ao conhecimento da norma padrão e de outras normas da língua.

Passamos agora ao capítulo da revisão da literatura, no qual trazemos alguns conceitos básicos dos Estudos da Linguagem, úteis para fundamentar e guiar nossa pesquisa.

## 1 REVISÃO DE LITERATURA

**A**presentamos a seguir um panorama, relativamente extenso, de noções teóricas atinentes ao nosso objeto de estudo, o *internetês*. O capítulo está organizado em três partes. Na primeira, abordaremos algumas noções basilares dos Estudos da Linguagem que, acreditamos, poderão orientar este trabalho de pesquisa. Praticamente todos os conceitos revisados foram e ainda são motivo de controvérsia, pois possuem graus de entendimento variados. A discussão sobre cada um deles pode render uma tese, de modo que nos deteremos apenas em alguns tópicos. A revisão, embora longa, não é exaustiva – até por questão de tempo e espaço - de modo que destacamos os posicionamentos que nos pareçam mais úteis para o foco de investigação.

O objetivo da revisão é oferecer um quadro dos conceitos relacionados ao nosso tema de estudo e utilizar alguns posicionamentos teóricos para nos auxiliar a melhor perceber o objeto sob exame: o léxico mais presente na escrita dos jovens interagindo na Internet. Como temos um objetivo bastante imediato e prático, qual seja, o de descrever o léxico do *internetês* e, dessa descrição, apontar sugestões aos professores para seu aproveitamento nas aulas de Língua Portuguesa, trazemos aqui apenas uma seleção de concepções sobre *língua, palavra, léxico, norma, variantes, dialeto, idioleto e escrita*. Esses conceitos, conforme nossa visão, parecem bastante úteis para essa pesquisa, visto que pretendemos refletir sobre a natureza e o estatuto desse tipo de escrita.

Ao identificar diferentes perspectivas sobre um conceito como, por exemplo, *língua*, acreditamos que o destaque a um determinado ponto de vista pode auxiliar a compreender a relação entre a *língua* e o nosso *internetês*. Do mesmo modo, diferentes entendimentos sobre a noção de *palavra*, a partir de suas diversidades, podem nos auxiliar a qualificar as palavras que integram o léxico do *internetês*. Assim, dos diferentes pontos de vista dos itens teóricos abordados nesta revisão esperamos obter referências para observar, explicar e descrever a linguagem utilizada pelos jovens no Orkut.

A segunda parte da Revisão de Literatura trata do *internetês*. A terceira apresenta a Lingüística de *Corpus*. Seleccionamos essa nova corrente dos Estudos Lingüísticos, alavancada no Brasil pelos estudos de Berber Sardinha (2000, 2004, 2004a), porque acreditamos que seus princípios teóricos e suas metodologias harmonizam-se com nossos objetivos de descrição do léxico e dos *corpora* organizados.

A extensão da revisão, acreditamos, é justificada pela amplitude e variedade de tópicos envolvidos. Além disso, é fruto da nossa necessidade de compreensão e de mensuração do estatuto da escrita no Orkut frente à língua portuguesa.

## 1.1 NOÇÕES FUNDAMENTAIS

### 1.1.1 Língua, sistema de signos inter-relacionados

*Uma língua é um sistema de relacionamentos  
entre formas e significados.  
(Mario Alberto Perini, 2006)<sup>10</sup>*

Afinal, o que é língua? Precisamos reconhecer o que se pode considerar como tal para avaliar se o que os usuários do *internetês* estão realizando seria algo concernente à língua, facultado pela própria língua, se está ou não na língua. Língua é apenas um instrumento de comunicação, o que implica existir somente em relação ao seu uso por alguém? Língua falada e língua escrita são a mesma língua? Língua é apenas um sistema ou um conjunto de sistemas interligados? Língua é produção do homem? O que é mesmo uma língua?

Em meio a essas questões, começemos retomando as informações básicas que Dubois (1978, p. 378) nos dá: “No sentido mais corrente, *língua* é um instrumento de comunicação, um sistema de signos vocais específicos aos membros de uma mesma comunidade”. Entretanto, esse instrumento é fonte de ambigüidade quando se trata de defini-lo. O termo *língua* é controverso porque temos de um lado uma infinidade de línguas diversas em suas variações dentro de uma mesma língua e, de outro, todos os membros de uma comunidade

---

<sup>10</sup> PERINI, Mario Alberto. *Princípios de Lingüística Descritiva: uma introdução ao pensamento gramatical*. São Paulo: Parábola Editorial, 2006, p. 45.

lingüística produzindo enunciados individuais, para se comunicar, mesmo com variações individuais, dentro de um mesmo sistema de regras e relações. Sem contar que há, pelo menos, dois meios de comunicação, cada um com um sistema próprio: a língua falada e a língua escrita.

Ao sistema abstrato, subjacente a todo ato de fala, Saussure chamou de *língua*. Ainda de acordo com Dubois, para o lingüista genebrino, para a Escola de Praga e para o estruturalismo americano, a língua é um sistema de relações, ou melhor, um conjunto de sistemas interligados, cujos elementos não têm nenhum valor fora das relações de equivalências ou de oposições que os unem.

A noção de instrumento antes indicada por Dubois encontramos rebatida por Benveniste (1995). Para ele, a comparação da linguagem a um instrumento material deve “encher-nos de desconfiança”. E explica melhor:

Falar de instrumento é pôr em oposição o homem e a natureza. A picareta, a flecha, a roda não estão na natureza. São fabricações. A linguagem está na natureza do homem que não a fabricou. [...] Não atingimos nunca o homem separado da linguagem e não o vemos nunca inventando-a. [...] Todos os caracteres da linguagem, a sua natureza imaterial, o seu funcionamento simbólico, a sua organização articulada, o fato de que tem um *conteúdo*, já são suficientes para tornar suspeita essa assimilação a um instrumento, que tende a dissociar do homem a propriedade da linguagem<sup>11</sup> (BENVENISTE, 1995, p.285) [grifo do autor].

Saussure (2004) concebe língua como um sistema de signos que por si só dão conta da significação. Ele define o signo como a união do sentido e da imagem acústica. O que ele chama de “sentido” é a mesma coisa que conceito ou idéia, a representação mental de um objeto ou da realidade social. Em outras palavras, conceito é sinônimo de significado (plano das idéias), algo como o lado “espiritual” da palavra, a parte inteligível, em oposição ao significante (plano da expressão), que é sua parte corpórea. Por outro lado, a imagem acústica “não é o som material, coisa puramente física, mas a impressão psíquica desse som” (p. 80). Melhor dizendo, a imagem acústica é o significante. Com isso, temos o signo lingüístico como

---

<sup>11</sup> Em Dubois (1978, p. 387) encontramos que “linguagem é a capacidade específica à espécie humana de comunicar por meio de um sistema de signos vocais (ou língua), que coloca em jogo uma técnica corporal complexa e supõe a existência de uma função simbólica e de centros nervosos geneticamente especializados. Esse sistema de signos vocais utilizado por um grupo social (ou comunidade lingüística) determinado constitui uma língua particular”.

uma entidade psíquica de duas faces semelhante a uma moeda. Vejamos as próprias palavras do autor marcando a distinção entre entidades psíquicas e físicas:

[...] os termos implicados no signo lingüístico são ambos psíquicos e estão unidos, em nosso cérebro, por um vínculo de associação. [...] O signo lingüístico une não uma coisa e um palavra, mas um conceito e uma imagem acústica. Esta não é o som material, coisa puramente física, mas a impressão (*empreinte*) psíquica desse som, a representação que dele nos dá o testemunho de nossos sentidos [...] O signo lingüístico é, pois, uma entidade psíquica de duas faces [...] Esses dois elementos estão intimamente unidos e um reclama o outro (SAUSSURE, 2004, p. 79-80).

Pode parecer ao leitor que estamos desviando do tema *língua* ao nos determos na explicitação de *signo*. Nossa intenção é esclarecer o que é signo, uma vez que acreditamos ser a língua um sistema de signos inter-relacionados. Aqueles amontoados de letras, tipo “rsrsrs” e “bjks”, são signos para os jovens que os utilizam no Orkut na Internet. Continuemos, pois, mais um pouco observando esses componentes inter-relacionados, constituintes básicos da língua.

Como vimos na última citação de Saussure, os dois elementos – *significante* e *significado* - são interdependentes e inseparáveis, pois sem significante não há significado e sem significado não existe significante. O laço que une o significante ao significado é arbitrário ou então, visto que entendemos por signo o total resultante da associação de um significante com um significado, podemos dizer mais simplesmente: o signo lingüístico é arbitrário (p.81).

Mas devemos tomar cuidado: ao simplificar que o signo é arbitrário, pode parecer que ele esteja à mercê do falante, que poderia associá-lo livremente a outras significações:

A palavra *arbitrário* requer também uma observação. Não deve dar a idéia de que o significante dependa da livre escolha do que fala [...]; queremos dizer que o significante é imotivado, isto é, arbitrário em relação ao significado, com o qual não tem nenhum laço natural na realidade (SAUSSURE, 2004, p. 83).

O princípio da arbitrariedade do signo, portanto, que é o primeiro princípio enunciado por Saussure e, segundo ele mesmo, o de primordial importância na análise lingüística (p.82), não está relacionado com a conexão do signo com o mundo, com a coisa do mundo real designada pelo signo. Uma observação curiosa em relação ao *internetês* e à arbitrariedade do signo é que alguns signos empregados pelos jovens são motivados. São exemplos as representações para risos, quando há o registro das formas onomatopéicas “rsrsrs”, “eheheeh”, “ahuauhuhahua”, “kkkkkkk” e “uahua”, entre outras.

Voltemos à busca do entendimento do que seja língua. Vejamos o que um dos estudiosos de Saussure tem a nos dizer. Carvalho (2007), da análise da obra saussuriana, depreendeu três concepções para língua: acervo lingüístico, instituição social e realidade sistemática e funcional. Observemos, rapidamente, cada uma delas:

1. A língua, como acervo lingüístico, é “o conjunto dos hábitos lingüísticos que permitem a uma pessoa compreender e fazer-se compreender” (SAUSSURE, 2004, p. 92). A língua é “uma soma de sinais depositados em cada cérebro, mais ou menos como um dicionário cujos exemplares, todos idênticos, fossem repartidos entre os indivíduos” (p. 27)<sup>12</sup>. Saussure nos diz que, além do dicionário, temos também a gramática: “Trata-se de um tesouro depositado pela prática da fala em todos os indivíduos pertencentes à mesma comunidade, um sistema gramatical que existe virtualmente em cada cérebro ou, mais exatamente, nos cérebros dum conjunto de indivíduos [...]” (p.21).

É na condição de acervo que a língua guarda a experiência histórica de uma civilização. Todas as línguas, mesmo as ágrafas, passam adiante a cultura e o conhecimento adquirido em gerações. Exemplos diretamente relacionados a nós são o latim e o grego clássico, que mesmo não sendo mais faladas, estão presentes no nosso cotidiano muito mais do que possamos imaginar.

2. A língua, como instituição social, “não está completa em nenhum (indivíduo), e só na massa ela existe de modo completo” (p. 21), por isso, ela é, simultaneamente, realidade psíquica e instituição social. Para Saussure, a língua “é, ao mesmo tempo, um produto social da faculdade da linguagem e um conjunto de convenções necessárias, adotadas pelo corpo social para permitir o exercício dessa faculdade nos indivíduos” (p. 17); é “a parte social da linguagem, exterior ao indivíduo, que, por si só, não pode nem criá-la nem modificá-la; ela não existe senão em virtude de uma espécie de contrato estabelecido entre os membros da comunidade” (p. 22).

3. A visão da língua como realidade sistemática e funcional é o conteúdo mais importante da concepção saussuriana. Para o autor do *Curso de Lingüística Geral*, a língua é, antes de tudo, “um sistema de signos distintos correspondentes a idéias distintas” (p. 18); é um código, um sistema no qual, “de essencial, só existe a união do sentido e da imagem acústica”

---

<sup>12</sup> Neste capítulo, a indicação da obra *Curso de Lingüística Geral*, de Ferdinand de Saussure, será assinalada apenas com o número da página entre parênteses.

(p. 23). Saussure vê a língua como um objeto de “natureza homogênea” (p. 23) e que, portanto, se enquadra perfeitamente na sua definição basilar: “a língua é um sistema de signos que exprimem idéias” (p. 24). Essa concepção da língua como sistema funcional está inter-relacionada com a noção de valor. Língua como um sistema de valores está intimamente associada à sua célebre frase: “na língua só existem diferenças”, ou seja, ela funciona sincronicamente e com base em relações opositivas (paradigmáticas) no sistema e contrastivas (sintagmáticas) no discurso.

Ao contrário da língua, a fala, ao ser constituída por atos individuais, é múltipla, imprevisível, irreduzível a uma pauta sistemática. Tais atos lingüísticos individuais são ilimitados, não formam um sistema. Já os fatos lingüísticos sociais, bem diferentemente, formam um sistema, pela sua própria natureza homogênea. Devemos ressaltar, no entanto, que tanto o funcionamento quanto a exploração da faculdade da linguagem estão intimamente ligados às implicações mútuas existentes entre os elementos língua (virtualidade) e fala (realidade).

Lembremos ainda que língua e pensamento são indissociáveis, tal uma folha de papel, um sendo o verso e outro o reverso: ao rasgarmos o papel, afetamos ambos os lados da folha. Esta metáfora, utilizada por Saussure, aplica-se aos componentes do signo, o significado e o significante. A língua, para Saussure, é a expressão do pensamento que, sem ela, é uma “massa amorfa e indistinta”. A expressão não se dá diretamente do pensamento aos sons: ela é mediada pela língua, que é um sistema de signos. É na relação que se estabelece no sistema que os signos adquirem seu valor, que significam. E mais ainda: a língua não é um sistema de signos justapostos, mas uma rede de signos que se relacionam e, assim, significam. Entra aqui, na análise de Saussure, a metáfora do jogo de xadrez: cada peça se define, adquire valor, na relação que tem com as outras peças do jogo. Os signos, também, se definem negativamente, pela oposição com outros signos do sistema. Mas há que se distinguir, como acentua Saussure, o valor lingüístico da significação.

A interpretação do signo se dá, então, em duas direções: vertical, entre seus componentes (significante e significado); e horizontal, na relação com outros valores semelhantes. Sem estas relações de diferentes direções não haveria significação.

Mudando um pouco o rumo do raciocínio, passemos a observar as idéias de um professor e economista sobre língua. Dacanal (2006), em sua obra *Linguagem, poder e ensino*

*da língua*, produzida em 1985 para ser um contraponto a *Língua e liberdade*, de Luft (1985), lembra que a língua é uma produção do homem, uma convenção, uma imposição social e histórica e, também, um instrumento de poder.

É interessante analisar o ponto de vista desse autor, observar a sua visão de que a língua deve ser considerada – principalmente no seu ensino nas escolas – como um instrumento de poder (possibilidade de meio de ascensão ou exclusão social pelo domínio (ou não) de sua norma padrão). Enfatiza Dacanal que a língua é uma produção do homem, que cria, entre tantas coisas, símbolos, “*algo físico* (ou material) que carrega consigo, ou em si, um *sentido* não físico (ou imaterial)” (p. 42). Nada de novo percebemos ao ler isso, uma vez que é estabelecida uma relação com aquilo dito por Saussure: a parte material é o *significante* e a parte imaterial, o *significado*. Para Dacanal (2006), a língua é “um conjunto mais ou menos amplo de *símbolos* sonoros convencionados [...], símbolos cujos sons, em princípio, referem-se sempre às mesmas coisas” (p.43). Surge aí, então, o caráter de convenção<sup>13</sup> de qualquer língua. Ela é um acordo entre os membros componentes de um grupo.

Convenção, então, parece ser a palavra que nos faz refletir sobre o *internetês*. Ele se torna algo combatido pelos puristas porque mexe com convenções estabelecidas, principalmente no que se refere à grafia. Mas mesmo com as alterações, alguns acordos entre os usuários da língua da Internet são combinados ou repetidos, tornando-se também uma espécie de convenção, algo que pode ser metodicamente repetido, ou ser transformado em princípios normativos na forma de se comunicar. Exemplo disso são *kkkkkkkkkk* para dizer que está rindo, as trocas de dígrafos por uma letra com em *akela* (aquela) e *axim* (assim) ou ainda uma única letra representar um monossílabo, como em *q* (que) e *t* (te).

A língua, ainda segundo esse autor, é produção de um grupo humano, imposta pela sociedade com todas as suas relações (principalmente de poder) ao longo do tempo. “A língua dominante é – ou tende a ser – sempre a língua daquele que detém o poder econômico, social e político” (DACANAL, 2006, p. 48). Aqui o autor toca num assunto que discutiremos mais adiante – a norma culta, aquela do grupo dominante – relacionado com a língua dos jovens internautas: a norma culta é desrespeitada na comunicação feita entre eles. Como os usuários

---

<sup>13</sup> Porque relacionada ao caráter de convenção da língua, consideramos interessante transcrever - nem que seja apenas por curiosidade - a opinião do autor sobre a capacidade lingüística *inata*, preconizada por Chomsky (1957): “É, portanto, rematada tolice dizer que a criança traz dentro de si a língua e é por isto que ela aprende a falar. O que a criança possui – como qualquer indivíduo que aprendeu uma ou mais de uma língua -, se não apresentar defeitos nos órgãos emissores e receptores de sons e se viver entre humanos que, supostamente, se utilizam de uma língua, é a capacidade de detectar e dominar esta convenção” (p. 45).

utilizam a sua própria língua, afrontam, de certa forma, a língua considerada na sua forma oficial. Rendemo-nos, portanto, ao pensamento do autor cujas idéias há pouco apresentamos, o qual ainda nos diz ser a língua, no sentido diacrônico “produto, parte, instrumento e relicário de uma civilização” (*op. cit.*, p. 55). No sentido sincrônico, “a língua é instrumento de treinamento da mente, de aprendizagem eficiente, de capacitação técnica, de inserção ética, de cidadania efetiva, de ascensão social e de ação política” (*op. cit.*, p. 56).

Dacanal (2006) insiste muito na idéia de a língua ser um instrumento de poder e que quem a domina tem melhores condições de dominar os outros. Relacionando tais idéias ao objeto do nosso estudo, podemos concluir que há um poder manifesto no uso do *internetês*. Utilizar tal forma de interação é ter o poder de inserir-se na comunicação digital. A expressão *analfabeto digital*, por exemplo, é marca e fruto de relação de poder. Tal expressão indica exclusão e mostra também que o *internetês* não constitui uma língua diferente, porque ela tem signos, significantes, é arbitrária. Enfim, é apenas mais uma forma da língua portuguesa. O problema estaria em os jovens dominarem apenas o *internetês*, o que parece ser pouco viável porque ninguém se alfabetiza nessa modalidade lingüística. Nenhum jovem interage com o mundo apenas na língua da Internet, ao contrário, ele precisa dominar duas formas da língua. A primeira é a oficial, com sua norma culta, que lhe dá o poder de expressar-se com competência, de forma clara e organizada, como cidadão. A segunda é a adaptada às necessidades da Internet, que lhe dá o poder de se inserir no mundo digital e no grupo característico de sua faixa etária. Essa forma lingüística é empregada num ambiente em que seus usuários não vêm a obrigatoriedade de utilizar a escrita da norma culta.

Depois do que vimos até agora sobre língua, podemos perguntar, entre outras coisas: O *internetês* pode ser visto como integrante da língua nos termos de Saussure? A resposta aparentemente é sim, mesmo que à primeira vista seja uma representação gráfica diferente daquela preconizada pela norma ortográfica oficial. Mas grafia é apenas uma faceta da língua, e da língua escrita. A modalidade de linguagem utilizada pelos jovens na Internet não contraria Saussure quando ele diz que a “língua é um sistema de signos que exprimem idéias” (p. 24). O problema está em todos entenderem os signos empregados pelos internautas (lembremos, entretanto, que nem todos entendem os signos usados por José de Alencar ou Guimarães Rosa...).

O fundador da Lingüística moderna nos diz ser a língua um produto social da faculdade de linguagem e conjunto de convenções adotadas por um corpo social. Aí temos a chave: o

corpo social de quem emprega o *internetês* é restrito, é língua dos que utilizam a Internet (salientamos ser uma *língua de jovens*, porque nem todo internauta utiliza tal linguagem). Mesmo restrita à comunicação entre jovens, um leitor contumaz consegue decifrar tal escrita oralizada, visto que ela se caracteriza, à primeira vista, por alterar significativamente apenas a grafia das palavras. Saussure diz mais: há “uma espécie de contrato estabelecido entre os membros da comunidade” (p. 22), o que é verdade, jovem usuário de *internetês* entende jovem usuário de *internetês*. Quanto à língua ser “uma soma de sinais depositados em cada cérebro”, assim como haver nele também um tesouro chamado “sistema gramatical virtual”, acreditamos se relacionar com o *internetês*. Antes de utilizar essa nova modalidade de escrita, o internauta já dominava a língua, por isso pode agora fazer as alterações para adaptar-se ao novo meio de comunicação.

Não esqueçamos que a escrita alterada pelos jovens na Internet é uma representação da linguagem, a qual, por sua vez, associa-se ao pensamento. Ela torna presente a fala de alguém que não está falando no local. Por ora, antes de chegar à escrita, vamos revisar algumas concepções sobre palavra, uma unidade também da escrita.

### 1.1.2 Palavra, algo difícil de definir

*Lutar com palavras  
é a luta mais vã,  
entanto lutamos  
mal rompe a manhã.  
São muitas. Eu pouco.*  
(Carlos Drummond de Andrade)<sup>14</sup>

O que provavelmente mais chama a atenção na escrita dos jovens na Internet são as palavras, ou melhor dizendo, a forma que têm as suas “palavras”. O que considerar como palavra quando vemos letras soltas e outras amontoadas, à primeira vista, nada significando para quem não está familiarizado com tal tipo de linguagem? Até onde vai o limite entre símbolo e palavra? No *internetês* há siglas, abreviaturas, ícones, desenhos. Alguns desses códigos podem ser considerados palavras?

<sup>14</sup> ANDRADE, Carlos Drummond de. *Reunião – 10 livros de poesia*. Rio de Janeiro: José Olympio, 1969, p.91.

Uma informação curiosa é que os *softwares* editores de texto e outras ferramentas que cumprem a tarefa de identificar quantas palavras há num texto divergem entre si justamente na contagem. Isso ocorre porque a presença de símbolos e caracteres pouco usuais implicam percepções diferentes sobre o que é palavra. No *corpus* utilizado para nossa pesquisa, ao verificar o número de palavras, o Microsoft Word indicou 541.841, enquanto o Wordsmith Tools, uma ferramenta para observação de palavras em texto, indicou 553.875 palavras. Isso ocorreu porque a ferramenta considerou como palavras os símbolos £ e ¢, entre outros.

Como vimos, mesmo mecânicas simples de identificação de palavras podem ser divergentes entre si. Por outro lado, as teorizações sobre a noção de palavra, guardadas as suas devidas especificidades, também exibem as suas divergências. Que critérios devemos levar em conta para classificar um signo como palavra? Vejamos a seguir.

No VI Congresso Internacional de Lingüística, em 1948, houve lingüista <sup>15</sup> propondo que o termo *palavra* fosse proscrito da nomenclatura lingüística por sua imprecisão e impossibilidade de acordo sobre uma definição ideal (BIDERMAN, 1999, p. 81). Apesar disso, a noção de palavra é um conceito fundamental, que dá sustentação a outras noções básicas construídas sobre e em torno dela. Isso também ocorre “com a noção de frase, assim como a de morfologia e a de sintaxe, pois o sistema da gramática clássica foi montado em torno do eixo *palavra-frase* que, com o abandono de um de seus pólos, desmorona” (*op. cit.*, p. 81).

Diante das polêmicas e da importância que a noção de palavra adquire nos estudos da língua, será necessário optar por um conceito. É o que tentamos fazer neste trabalho, após a revisão de literatura.

Para delimitar o que seja palavra, seguindo o pensamento de Biderman (1999), podemos utilizar os critérios fonológico, morfossintático e semântico. Assim, em cada um dos níveis temos as seguintes concepções:

- **Fonológico:** palavra é uma seqüência fonológica recorrente com o mesmo significado, uma emissão completa seguida de pausa. Mas só o critério fonológico é pouco

---

<sup>15</sup> Biderman registra a informação, citando apenas o nome de Holt dentre os lingüistas que propuseram o banimento do conceito de *palavra* da Lingüística.

para delimitar palavras. No exemplo “Penetra || surdamente || no : reino | das : palavras |||” temos quatro tipos de pausa: longa (representada |||) no final, duas medianas (||), uma ligeira (|) e duas ligeiríssimas (:), estas últimas devido à atonicidade e à dependência gramatical de *no* e *das*. O falante agrupa as sílabas em sua mente utilizando processos gramaticais e semânticos, e não fonéticos.

- **Morfossintático:** leva-se em conta classe e função da unidade, superpondo-se critério formal e funcional. Aqui já aparecem outras dificuldades de delimitação, porque há casos em que uma forma contém duas unidades complexas, como no exemplo de *no* (em + o), e também o contrário, as chamadas locuções, em que duas unidades correspondem a um único significado como em *vai sair* (sairá) ou *em vão* (debalde). A coesão interna da palavra é outro aspecto importante, porque embora possa ter mobilidade no discurso, internamente a palavra é uma seqüência coesa e estável.

- **Semântico:** palavra é unidade semântica indecomponível, com significação autônoma.

Conforme Biderman (1999, p. 87-9), *vocábulo* e *palavra* são termos imprecisos na ciência, tidos como sinônimos para o falante comum. Assim, para a autora, seria melhor convencionar as seguintes definições:

- Lexema: unidade lexical abstrata e virtual que compõe o léxico;
- Lema: representação do lexema no dicionário;
- Léxico: conjunto abstrato das unidades lexicais da língua;
- Vocabulário: conjunto de realizações discursivas das unidades lexicais.
- Nas realizações discursivas <sup>16</sup> pode-se chamar qualquer seqüência significativa de *palavra* ou *vocábulo*.

Biderman (1999) também sugere que possa ser utilizado o termo técnico *lexia*, proposto por Pottier (1975), para designar as manifestações discursivas dos lexemas. As *lexias* podem ser classificadas de *simples* (seqüência gráfica separada por dois brancos, ex.: cesta), *complexas* (várias unidades unidas sem hífen, ex.: cesta básica) e *compostas* (unidades ligadas por hífen, ex.: pai-de-santo). *Lexia* nos parece um termo melhor que *palavra*, uma vez que é uma unidade funcional significativa, podendo ser um vocábulo simples ou composto, uma

<sup>16</sup> Segundo a autora, qualquer manifestação da língua em ação. É a própria *fala* na dicotomia saussuriana *língua/fala*.

locução, uma expressão idiomática ou até uma sentença (provérbio), neste último caso chamando-se *lexia textual*. Vejamos, então, o que o autor evocado entende como *lexia*.

Pottier (1975, p. 26) concebe *lexia* como “unidade lexical memorizada”, e isso é claramente entendido quando se utilizam *lexias* complexas ou textuais, em que não construímos essas combinações no momento da fala, mas as retiramos de nossa “memória lexical”, assim como quando utilizamos qualquer palavra simples. O autor dá como exemplo *pé de cabra*<sup>17</sup>, que “pode ser uma *lexia*, no sentido de ferramenta, ou o resultado de uma construção sintática de discurso se se tratar do pé do animal” (p. 27). A etimologia de *lexia* nos ajuda a entender melhor ainda sua significação. O autor baseou-se em  $\lambda\acute{\epsilon}\xi\iota\varsigma$  (de  $\lambda\acute{\epsilon}\gamma\omega$ , dizer), que no grego clássico indica “ação de falar, maneira de falar, elocução, estilo, palavra, expressão”.

Devemos levar em conta, então, a complexidade do léxico de uma língua, formado por unidades monossilábicas até unidades bastante complexas, como as Unidades Fraseológicas (seqüências de palavras semanticamente coesas e com propriedades morfossintáticas específicas) divididas em colocações e expressões idiomáticas.

Passemos agora a ver como está indicado o termo *palavra* no *Dicionário de Lingüística*, de Dubois, um dos mais citados no meio acadêmico. Dubois (1978, p. 361) considera a palavra como a unidade mínima construída, em oposição a morfema (o menor signo lingüístico) e a *lexia* (a unidade de comportamento léxico). No mesmo dicionário (p. 449) encontramos que, para a lingüística tradicional, “*palavra* é um elemento lingüístico significativo composto de um ou mais fonemas”, que pode ter transcrição escrita entre dois espaços em branco e que conserva sua forma, podendo ser flexionado, nos empregos sintagmáticos. Parece-nos que é essa a concepção de palavra mais usual entre as pessoas comuns, leigas em Lingüística. Continua Dubois (1975, p. 450) informando que a lingüística estrutural evita empregar a noção de palavra devido a sua falta de rigor e que haveria uma oposição entre *termo* (com significação única, de emprego monossêmico, conforme apregoado por Wüster, o fundador da Terminologia) e *palavra* (unidade léxica do vocabulário geral, essencialmente polissêmico). Também informa que a mesma corrente lingüística faz distinção entre *palavra* e *vocábulo*, sendo a primeira, para a estatística léxica, toda unidade gráfica

<sup>17</sup> Quando significar ferramenta, pelas normas ortográficas vigentes, deve ser grafada com hífen: pé-de-cabra.

inscrita entre dois espaços em branco, ao passo que o segundo será apenas cada unidade léxica diferente.

Nesse caso, palavra corresponderia a *token* e vocábulo, a *type*. E segue Dubois explicando como a polêmica *palavra* é encarada por outros lingüistas, ninguém chegando a um consenso. Neste trabalho, então, talvez fosse mais útil empregar palavra com o sentido mais comum, unidade de sentido registrada entre dois brancos gráficos.

Como definiriam *palavra* os dicionários gerais da língua portuguesa? Observemos o que Houaiss e Aurélio têm a nos dizer. Houaiss (2001) nos indica várias acepções para *palavra*, além de nos informar sua etimologia. Observando-a, percebe-se a intuição dos primitivos falantes que associaram palavra (inicialmente *parábola*) como comparação entre o conceito (significado) e a expressão dela (significante). Vejamos: palavra originou-se, então, do latim *parabòla,ae*, um empréstimo do grego pela língua da retórica no sentido de 'comparação'. Vejamos também as primeiras acepções<sup>18</sup> dadas por Houaiss:

1 unidade da língua escrita, situada entre dois espaços em branco, ou entre espaço em branco e sinal de pontuação 2 GRAM unidade pertencente a uma das grandes classes gramaticais, como substantivo, verbo, adjetivo, advérbio, numeral etc., não levando em conta as modificações que nela ocorrem nas línguas flexionais, e sim, somente, o significado; vocábulo 3 GRAM m.q. **vocábulo** ('unidade que se agrega') 4 LING.EST para o estruturalismo norte-americano, unidade mínima com som e significado que pode, sozinha, constituir enunciado; forma livre mínima, vocábulo [...].

Mais adiante, no mesmo verbete, o dicionarista informa as muitas significações que o termo *palavra* adquire em vários contextos, as mais comuns, em sentido figurado ou nas diferentes correntes da lingüística<sup>19</sup>. A noção de palavra que aparece nos vocábulos formados

<sup>18</sup> Ao citarmos, dos dicionários gerais, verbetes ou partes deles, registramos grifos, abreviaturas e símbolos utilizados pelo autor.

<sup>19</sup> Por ser um dicionário geral, chamam a atenção os registros especializados: **p. funcional** LING palavra que expressa relações gramaticais, como, p.ex., as preposições e as conjunções; palavra gramatical, palavra estrutural, palavra vazia [A descrição do seu significado ger. é feita usando-se *metalinguagem* gramatical.] □ **p. gramatical** GRAM LING 1 m.q. **palavra funcional** 2 determinada unidade de um paradigma gramatical [...], são todas palavras gramaticais; palavra morfossintática □ **p. lexical** LING aquela que possui um semantema cujo significado remete para o mundo físico e biossocial; palavra plena □ **p. metalingüística** LING palavra da terminologia lingüística ou gramatical, da qual nos valem para falar da própria língua e de seus elementos □ **p. plena** LING m.q. **palavra lexical** □ **p. reservada** INF em linguagem de programação, termo cujo significado é definido pelas regras da própria linguagem e que não pode ser us. com outro sentido ou em diferentes condições em programas escritos nesta linguagem □ **p. vazia** GRAM denominação dada por alguns gramáticos às palavras funcionais

por composição também é indicada e aparece assim no dicionário: “noção de 'palavra', usar *antepos.* lex(e/i)-, log(o)-, palavr- e verb(i/o)-; *pospos.* -fasia, -lalia e -lexia”.

O Dicionário Aurélio, na edição de 1980, por sua vez, definia *palavra* de forma direta, dando como primeira acepção “Fonema ou grupo de fonemas com uma significação; termo, vocábulo, dição” (p.1019). Já na edição de 1999, as informações são mais completas:

1. *E. Ling.* Unidade mínima com som e significado que pode, sozinha, constituir enunciado; forma livre. 2. Unidade pertencente a uma das grandes classes gramaticais, como, p.ex., substantivo, verbo, adjetivo, advérbio, abstraídas as diferentes realizações (marcas flexionais) que ela possa apresentar; lexema. [Quando referimos um verbo como *amar*, temos em mente não apenas o infinitivo, tomado aí como forma de citação, mas todas as demais formas de conjugação]. 3. *E. Ling. V. palavra gramatical.* 4. Nas escritas modernas, unidade constituída por morfemas, delimitada por espaços em branco e/ou sinais de pontuação. 5. Alta expressão do pensamento. 6. Grupo de palavras; frase(s). 7. Faculdade de expressar idéias por meio de sons articulados; fala. 8. Modo de ver; opinião, afirmação, asserto. 9. Alocução, oração, discurso. 10. Doutrina (1). 11. Promessa verbal. 12. Permissão ou direito de falar. 13. Maneira de falar. 14. *Inf.* O maior segmento de dados que um computador é capaz de manipular (processar, transmitir através de barramento (2) interno, armazenar na memória principal ou num registrador, etc.) com uma única operação. [Há diversos tamanhos convencionais, os mais comuns são os de 16, 32 ou 64 bits] *Interj.* 15. Exclamação peremptória.

Para *palavra gramatical*, encontramos:

1. Determinada unidade de um paradigma gramatical; palavra morfossintática [Tb se diz apenas *palavra*]. [No português padrão do Brasil, *amamos* representa duas palavras gramaticais diferentes: a primeira pessoa do plural do presente do indicativo de *amar*, e a primeira pessoa do plural do pretérito perfeito do mesmo lexema] 2. *V. palavra funcional*

Seguindo as remissões indicadas pelo dicionarista, vemos que para *palavra funcional* temos “*E. Ling.* Palavra cujo significado expressa relações gramaticais, como, p.ex., as conjunções; palavra gramatical, palavra estrutural, palavra vazia [Cf. *palavra lexical*]”. Fazendo o confronto, observamos para *palavra lexical* o seguinte registro: “*E. Ling.* Aquela cujo significado refere o mundo biossocial, palavra de conteúdo. [Cf. *palavra funcional*]”.

Já a respeito desse assunto, nosso maior lingüista brasileiro, Mattoso Câmara (1977, p. 59), diz-nos que “a apresentação do vocábulo na escrita se faz pelo critério formal. Deixa-se entre eles, obrigatoriamente, um espaço em branco, porque, mesmo quando sem pausa entre si num único bloco de força cada um é considerado uma unidade mórfica de per si”. O critério

para reconhecer se um vocábulo é formal está em verificar se ele pertence a uma das duas espécies: *forma livre* ou *forma presa*. *Forma livre* é aquela constituída por “uma seqüência que pode funcionar isoladamente como comunicação (ex.: “Que vão fazer?”). Resposta: “Proscreever”. “Proscreever o quê?” Resposta: “Lei”)” (CÂMARA, 1977, p.59). E *forma presa* é aquela que só funciona ligada a outras, sendo o caso dos afixos, por exemplo.

Mattoso Câmara continua explicando que vocábulo formal é a unidade que resta após serem feitas todas as divisões em formas livres, podendo constar de uma forma livre indivisível (ex.: *luz*), de duas ou mais formas presas (ex.: *im + pre + vis + ível*) ou de uma forma livre e forma(s) presa(s) (ex. *in + feliz*). Há também a *forma dependente*, o caso das partículas proclíticas e enclíticas (artigo, preposição e pronomes átonos, por exemplo), que não podem funcionar isoladamente como comunicação suficiente, mas também não estão presas a nenhuma outra.

Para tentar enfeixar o assunto *palavra*, vamos buscar citação de Saussure, que a considera, na falta da delimitação do que seja a unidade da língua, a própria substituta de tal unidade:

Não podendo captar diretamente as entidades concretas ou unidades da língua, trabalharemos com as palavras. Estas, sem recobrir exatamente a definição da unidade lingüística, dão dela uma idéia pelo menos aproximada, que tem a vantagem de ser concreta; tomá-las-emos, pois, como espécimes equivalentes aos termos reais de um sistema sincrônico, e os princípios obtidos a propósito das palavras serão válidos para as entidades em geral (SAUSSURE, 2004, p.132).

Como vimos, confirma-se a heterogeneidade sobre a conceituação de *palavra*. Há pontos de convergência e de divergência entre os diferentes autores revisados aqui. Na impossibilidade de uma conceituação própria, entendemos, provisoriamente, que uma palavra, na escrita, será aquilo que se deixa identificar por dois espaços em branco. Analisando a linguagem empregada pelos jovens no Orkut, conheceremos as palavras utilizadas e modificadas pelos jovens, e com base nelas buscaremos entender o que é o *internetês*. Feito isso, passamos agora ao tema do *léxico*.

### 1.1.3 Léxico, repertório de vocábulos

*O léxico de uma língua constitui uma forma de registrar o conhecimento do universo. Ao dar nomes aos referentes, o homem os classifica simultaneamente.*  
(Maria Tereza Camargo Biderman)<sup>20</sup>

No sentido mais comum, *léxico* é o repertório de palavras de uma língua. Seabra (2006, p.7) a ele assim se refere:

Responsável por nomear e exprimir o universo de uma sociedade, o léxico encontra-se arraigado à história, tradição e costumes de um povo, estando, por isso, em constante processo de expansão, alteração e contração. Devido a essas características, é considerado o subsistema mais dinâmico da língua.

Retornamos a Dubois (1978) para ver como a Lingüística percebe esse importante componente da língua. Afinal, utilizamos signos de forma organizada para constituir um determinado sentido em situações variadas.

Para a Lexicografia, conforme o autor, *léxico* equivale a *dicionário*, mas para a lingüística geral, “a palavra *léxico* designa o conjunto das unidades que formam a língua de uma comunidade, de uma atividade humana, de um locutor, etc.” (DUBOIS, 1978, p. 364). Em assim sendo, *léxico* participa de diversos sistemas de oposição, de acordo com o modo como o conceito é considerado.

A primeira delas é a oposição *léxico* e *vocabulário*, em que *léxico* é reservado à língua e *vocabulário*, ao discurso. O léxico tem como unidades os lexemas, e o discurso, os vocábulos e as palavras. O vocabulário de um texto qualquer é, então, apenas uma amostra do léxico daquele autor ou da comunidade lingüística a que ele pertença. O léxico constitui, portanto, potencialidades ou competência do falante, enquanto o vocabulário é o conjunto de unidades localizadas no *corpus* utilizado pelo mesmo falante. A descrição léxica resulta, então, em um léxico fundamental, a intersecção dos diversos conjuntos de vocabulários ou, ao contrário, a reunião dos conjuntos, um “tesouro” no dizer de Dubois (1978, p. 364). O mesmo autor informa que todo falante possui dupla competência léxica, podendo compreender inúmeros lexemas (participam do modelo de competência do falante), e que nunca são realizados. Por

---

<sup>20</sup> BIDERMAN, Maria Tereza Camargo. Dimensões da palavra. In: *Filologia e Lingüística Portuguesa*, São Paulo, Humanitas, FFLCH/USP, n. 2, 1998. p. 81.

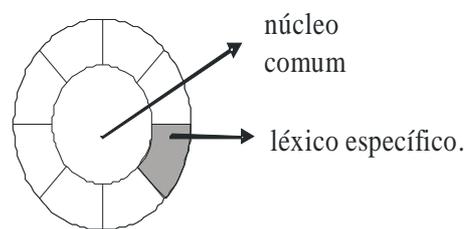
exemplo, certas palavras ouvidas, mesmo sem nenhuma probabilidade de serem utilizadas, são compreendidas pelo falante que as entende por levar em conta a situação em que foram empregadas. Há outros casos em que lexemas são entendidos porque fazem parte de uma regra da gramática, como, por exemplo, as nominalizações. Mesmo que nunca sejam utilizadas, figuram nas potencialidades oferecidas pelas regras e bases léxicas, estando presente na competência léxica do falante-ouvinte.

Sobre a importância do léxico ninguém põe dúvidas, a ponto de Krieger (2006, p. 159) chamá-lo de “componente basilar de todo e qualquer sistema lingüístico”. O que existem são variações de conceitos conforme o enfoque teórico ou escolha de algum ponto descritivo. O léxico é, então, um componente de muitas faces, ocupando um ponto central nas línguas. Mesmo que imprecisa conceitualmente, à palavra estão ligados elementos de ordem fônica, morfológica, sintática e semântica. Além disso, os estudos discursivos dos últimos anos também se valem da palavra, que expressa subjetividade e ideologia. O léxico permite identificar reflexos de diferentes ideologias na produção lingüística dos indivíduos.

Krieger (*op. cit.*, p. 162) nos lembra que “o léxico de uma língua, longe de ser um bloco monolítico, constitui-se como um conjunto heterogêneo em vários ângulos de sua composição”. O tempo, o espaço e o registro determinam sua formação. Esses componentes mostram a heterogeneidade constitutiva do léxico como seu dinamismo. “É também esse dinamismo que torna o léxico o pulmão das línguas, e, simultaneamente, um objeto multifacetado e em constante mobilidade.” (KRIEGER, 2006, p. 163). Mesmo que aparentemente essa mobilidade possa fazer pensar que o léxico seja algo descontínuo, sem regras, ele apresenta alto teor de regularidade constitutiva, aliás, como na língua em geral.

É no nível lexical que podemos observar mais claramente o fenômeno da mudança, o qual pode se manifestar em todos os níveis lingüísticos, do fonológico ao semântico. Renovação do léxico é fenômeno permanente, porque o léxico reflete a dinâmica da língua, que sem se dissociar da sociedade e da cultura, registra a dinâmica da realidade histórica e cultural de uma comunidade lingüística. À medida que ocorrem mudanças na sociedade, a língua a elas se adapta e produz novas unidades léxicas. Criar novas palavras e reutilizar as já existentes com novos significados constituem um processo geral de desenvolvimento da língua.

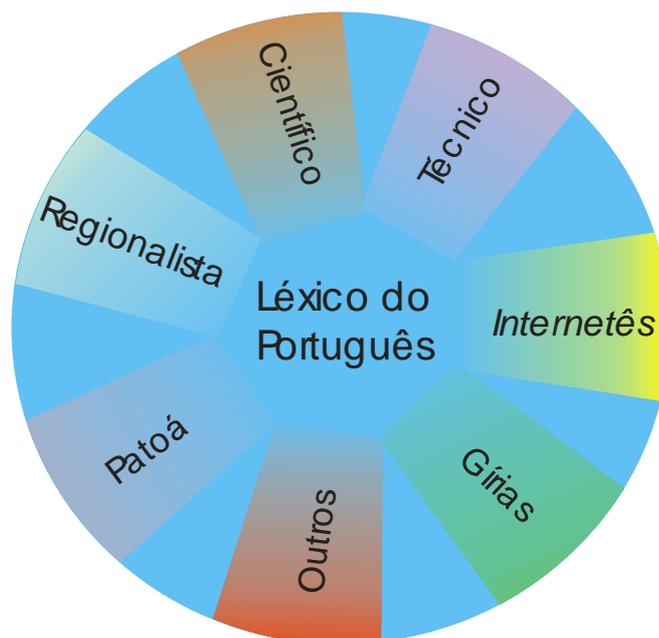
Toda língua se compõe de uma série de regras e de um componente lexical. Em princípio, este constitui um sistema aberto por força de sua função designativa, e aquela forma um sistema fechado. De acordo com Rosengren (1979), o léxico não é um conjunto bem delimitado de lexemas ou morfemas. Consiste em uma série de léxicos dispostos em volta de um núcleo comum, disposição esta que talvez possa ser ilustrada em uma figura como a a seguir:



**Figura 1- Léxicos em volta do núcleo comum (Fonte: Rosengren, 1979, p. 294.)**

Rosengren denominou léxico específico os léxicos técnicos, ou como o nome já diz, específicos de um grupo. Não há falante que domine todos os léxicos específicos de sua língua, mas todos os que dominam os diferentes léxicos também dominam um mesmo núcleo, comum a todos, a língua padrão ou comum, no dizer do autor, língua Standard. O que não varia de falante para falante dentro de seus léxicos específicos é o núcleo da língua. Não são sistemas lingüísticos diferentes, apenas léxicos diferentes, com um núcleo comum. Podemos perceber que esse autor nos aponta para a possibilidade de ver na pesquisa sobre a linguagem empregada pelos jovens na Internet algo de léxico específico.

Baseados na representação de Rosengren (1979, p. 294), procuramos mostrar na Figura 2 nossa interpretação sobre a composição geral do léxico do português, formado pelo núcleo comum e por vários léxicos específicos. O núcleo comum, principalmente formado pelas palavras funcionais, perpassa todos os léxicos, está subjacente a eles.



**Figura 2: Exemplificação de núcleo comum e léxicos específicos (Com base em Rosengren, 1979, p. 294.)**

Busquemos agora a visão saussuriana de léxico. Procurando no *Curso de Lingüística Geral* referências ao léxico, encontramos, na maioria das vezes, citações indiretas ao conjunto de palavras necessárias para que língua e fala possam expressar o pensamento.

Saussure (2004, p. 156) informa-nos que à reunião de morfologia e sintaxe se convencionou chamar Gramática<sup>21</sup>, dela sendo excluída a lexicologia ou ciência das palavras. Com isso o autor não concorda, porque logo depois se refere desta forma à relação lexicologia e Gramática:

À primeira vista, as palavras, tais como estão registradas no dicionário, não parecem dar ocasião ao estudo gramatical, que se limita geralmente às relações existentes entre as unidades. Mas logo em seguida verifica-se que grande número dessas relações pode ser expresso tão bem por palavras como por meios gramaticais. [...] Atribuem-se geralmente as preposições à Gramática; no entanto, a locução preposicional *em consideração a* é essencialmente lexicológica, de vez que a palavra *consideração* nela figura com seu sentido próprio (SAUSSURE, 2004, p. 157).

Na página seguinte, há o reforço da idéia de serem imbricados léxico e gramática:

<sup>21</sup> Saussure (p. 156) inicia o Capítulo VII definindo o que é Gramática (com G maiúsculo): “A Lingüística estática ou descrição de um estado de língua pode ser chamada de *Gramática*, no sentido muito preciso e ademais usual que se encontra em expressões como “gramática do jogo de xadrez”, “gramática da Bolsa” etc., em que se trata de um objeto complexo e sistemático, que põe em jogo valores coexistentes. A Gramática estuda a língua como um sistema de meios de expressão; [...]”.

Vê-se, portanto, que do ponto de vista da função, o fato lexicológico pode confundir-se com o fato sintático. Por outro lado, toda palavra que não seja uma unidade simples e irreduzível não se distingue essencialmente de um membro de frase, de um fato de sintaxe; a disposição das subunidades que a compõem obedece aos mesmos princípios fundamentais que a formação dos grupos de palavras. (*op. cit.*, p. 158)

Depois de definir signo lingüístico como “a combinação do conceito e da imagem acústica” (p.81), Saussure tratou do princípio da arbitrariedade do signo, explicando que é arbitrário por não ter nenhuma ligação com a realidade. O “signo lingüístico une não uma coisa e um nome, mas um conceito e uma imagem acústica” (p. 80), chamando o conceito de Significado e a imagem acústica de Significante. O Significante não tem nenhuma relação natural com a realidade, o signo lingüístico “é imotivado, isto é, é arbitrário com relação ao significado” (p. 83). As línguas com mais imotivação são mais lexicológicas, as com menos, são mais gramaticais: “São como dois pólos entre os quais se move todo o sistema: [...] a tendência a empregar o instrumento lexicológico, o signo imotivado, e a preferência concedida ao instrumento gramatical, isto é, à regra de construção” (p. 154).

Segue o autor tratando da imutabilidade e da mutabilidade do signo. Ensina-nos que em qualquer momento que observarmos uma língua, ela sempre se mostra como uma herança de épocas anteriores. Nunca saberemos em que momento um Significante passou a estar junto a um Significado, transpassando gerações, permanecendo tal signo imutável, resistindo a substituições. Mesmo parecendo paradoxal, Saussure nos lembra que o signo pode ser mutável, tal mutabilidade ocorrendo pelas mudanças culturais ao longo da história. É o caso do desaparecimento de palavras devido ao desaparecimento de costumes, instituições e objetos. Assim, se a coisa designada desaparece, desaparece também o uso da palavra que a nomeava. É o caso de tantas palavras, dentre elas *candeeiro*, *bacamarte* e *palmatória*, só presentes em textos ou referências históricas do passado. Referindo-se a imutabilidade e mutabilidade do signo, os organizadores do *Curso* explicam em nota de rodapé que “a língua se transforma sem que os indivíduos possam transformá-la. Pode-se dizer também que ela é intangível, mas não inalterável” (p. 89).

Parece-nos que essas idéias se aplicam ao estudar o vocabulário dos jovens. Eles recebem uma língua imutável e a usam de um modo particular, podendo parecer que a estão modificando. Outro ponto importante a salientar é o do valor lingüístico de uma palavra. Para Saussure, o valor constitui um dos aspectos do significado e se esse valor fosse excluído, a língua se resumiria a uma simples nomenclatura. “A língua é um sistema onde todos os termos

são solidários” (p. 133). O valor do signo resulta da presença simultânea de outros signos dentro do sistema e aos quais ele se contrapõe, havendo formação de uma complexa rede de significações.

Há palavras que são unidades complexas com subunidades (afixos e radicais) e há unidades maiores que as palavras, os compostos, locuções, formas de flexão. Há um problema de delimitação. O falante não se preocupa com isso: “tudo o que for significativo num grau qualquer aparece-lhes como um elemento concreto, e eles distinguem infalivelmente no discurso” (p. 122-123).

Recorremos também a Lepschy (1984) para ilustrar uma outra concepção de léxico. Inicialmente ele se refere à ambigüidade do termo, usado às vezes para designar um setor, nível ou aspecto de língua ou então para indicar a obra que descreve tal setor. O autor italiano opta, então, por utilizar *léxico* no primeiro sentido e *vocabulário* para o segundo, sendo *léxico* objeto da lexicologia, e *vocabulário*, da lexicografia. Os objetos lingüísticos que compõem o léxico recolhidos do vocabulário são as palavras. Mesmo reconhecendo que a noção de palavra é polivalente, gerando discussões variadas, “todo o funcionamento da língua, nos seus vários níveis, parece constar de sistemas que giram à volta da palavra” (LEPSCHY, 1984, P. 156). Para sabermos uma língua, é necessário conhecer suas palavras, como se pronunciam ou escrevem, o que significam e como se alteram e se ligam nas frases, enfim, conhecer os aspectos lexical e gramatical, que aparecem como capítulos diferentes na descrição das línguas, mas que, na realidade, são indissociáveis, é praticamente impossível estudar a gramática desligada das palavras. “A distinção que se cria entre léxico e gramática é que o primeiro é o reino da irregularidade, e a segunda o reino da sistematicidade” (*op. cit.*, p. 161). Vemos aqui, mais uma vez, a menção ao caráter irregular do léxico.

Referindo-se a fatos de natureza lingüística, é ao léxico (em sentido amplo, às palavras) que as pessoas recorrem para entender um povo. O léxico é considerado um espelho fiel da vida, refletindo a sua organização. A presença ou ausência de certas palavras podem mostrar indício de algum fato historicamente interessante

Vejamos agora como se entende o termo no âmbito da Lingüística de *Corpus*. Trata-se de “um conjunto bastante heterogêneo de unidades lexicais, contendo inclusive nomes próprios, siglas, abreviaturas, entre outras unidades da língua, cada qual associada a uma série

de traços morfossintáticos” (NUNES et al, 2005, p. 37). O léxico é assim, em primeiro lugar, um conjunto.

Berber Sardinha (2004a), respondeu à pergunta *Com o que se preocupa atualmente a Lingüística de Corpus?* destacando o léxico: “A Lingüística de Corpus se ocupa de quase todas as áreas de investigação lingüística. O léxico é a que mais recebe a atenção dos lingüistas de corpus e é a que mais se projeta para o mundo, basta ver os dicionários de inglês atuais, que são produzidos com base em corpus”. O mesmo autor, em sua principal obra (2004), lembra-nos que um atributo inseparável da palavra é a sua freqüência de uso, sua ocorrência, tão inseparável quanto a significação que encerra. O valor da palavra é assumido pela soma de seus empregos ao ser aceita, retomada, repetida. Daí a importância da análise das freqüências para o estudo da linguagem.

Para o estudo do léxico, segundo Berber Sardinha (2004, p.165-6), a Lingüística de *Corpus* considera alguns termos-chave. Tais termos nos ajudam a entender o que a Lingüística de *Corpus* percebe como léxico de uma língua:

- ▶ Palavras, ocorrências ou itens (*tokens*): contagem de palavras corridas; cada palavra conta como uma ocorrência, mesmo que seja repetida.
- ▶ Palavras, formas, vocábulos ou tipos (*types*): número de palavras diferentes.
- ▶ Palavras de conteúdo, lexicais, plenas ou de classe aberta: palavras que são das categorias morfossintáticas substantivo, adjetivo, verbo, advérbio, interjeição e numeral. Algumas classificações incluem aqui apenas os verbos principais.
- ▶ Palavras gramaticais, funcionais, de classe fechada ou instrumentais: palavras de categorias fechadas (pronomes, artigos, preposições, conjunções; algumas classificações incluem aqui verbos auxiliares e modais).
- ▶ Lema ou lexema: palavras que resumem as suas variantes. Por exemplo, o lema *querer* inclui as várias conjugações desse verbo: *quero, queres, queria*, etc.
- ▶ *Hapax legomena*: itens que ocorrem apenas uma vez (isto é, elementos de freqüência 1) no *corpus*.
- ▶ *Hapax dislegomena*: itens que ocorrem mais de uma vez no *corpus*.

O autor alerta para o sentido ambíguo do termo *palavra*, que pode significar tanto *types* como *tokens*. A dúvida será dirimida pelo contexto de uso, como normalmente ocorre na língua em geral.

Das pesquisas já realizadas utilizando os princípios da Lingüística de *Corpus*, alguns resultados (Biderman, 1998) indicam características do léxico do português brasileiro. Um ponto importante a ser destacado é a presença pouco expressiva de estrangeirismos, sejam empréstimos<sup>22</sup> ou palavras de outras línguas com adaptação da grafia ao português. Nos estudos feitos, aparece uma nova classificação para determinados estrangeirismos, as palavras consideradas *sem fronteiras* por aparecerem em vários idiomas praticamente com a mesma grafia, como *airbag, blues, design, fitness, free, Internet, jeans, rally, relax* e *slogan*.

Um melhor entendimento da léxico-gramática do português pode ser feito pela investigação da linguagem baseada em *corpus*. Um trabalho relatado por Berber Sardinha confirma os princípios da descrição léxico-gramática presente na literatura da Lingüística de *Corpus*:

(1) as escolhas léxico-gramaticais não são feitas em isolamento, mas sim em cadeia, segundo o princípio idiomático; (2) as palavras possuem padronização, observável a partir da exploração de um corpus computadorizado; (3) formas derivadas de uma mesma raiz devem ser consideradas itens lexicais distintos; e (4) diferenças entre padrões colocacionais indicam diferenças de sentido (BERBER SARDINHA, 2004, p. 234).

Baayen (1998) refere-se à importância de utilizar *corpora* para estudar diferentes tipos de texto. Diz ele que o uso da gramática e do léxico varia com gênero, autor, assunto, modalidade (língua falada x escrita), região e registro (variantes estilísticas). O estudo de como a gramática e o léxico são afetados pelo tipo de texto é um domínio de pesquisa pelo qual os *corpora* estão se tornando mais importantes na pesquisa. Os *corpora* provam a extensa variedade de textos permitindo ao pesquisador estudar com detalhes como as pesquisas lingüísticas são colocadas em uso através de diferentes tipos textuais. Segundo esse autor, a relação entre gramática e gênero, registro e modalidade tem sido estudada detalhadamente por Biber (1988, 1995). A relação entre léxico e tipo de texto, entretanto, tem recebido menos atenção sistematicamente.

---

<sup>22</sup> Estrangeirismo e empréstimo também são conceitos controversos quanto à definição nos estudos lingüísticos. Referem-se a palavras com grafia adaptada ou ainda grafadas como na língua de origem? Para Houaiss (2001), *estrangeirismo* é a palavra ou expressão estrangeira ainda não incorporada ao léxico da língua receptora. Já para *empréstimo*, informa: “5 LING incorporação ao léxico de uma língua de um termo pertencente a outra língua [Dá-se por diferentes processos, tais como a reprodução do termo sem alteração de pronúncia e/ou grafia (*know-how*), ou com adaptação fonológica e ortográfica (*garçom, futebol*).]”.

Nos estudos literários, aspectos do léxico têm sido considerados por produzirem informações importantes aos autores, indicando diferenciação regional, por exemplo. Estudos literários e estatísticas lexicais podem ser usados para estudar a relação entre tipo de texto e léxico. Além disso, Baayen (1998) diz que a frequência das palavras constitui uma rica fonte de informação na mesma relação entre léxico e tipo de texto. Para um texto grande, a presença de *hapax legomena* também se torna um diagnóstico para distinguir o tipo de texto. É que tanto **a alta frequência** como **a baixa frequência** dão uma rica fonte de informação para o estudo da diferenciação gramatical e morfológica através dos tipos de texto [grifo nosso]. Como vemos, então, pela análise do léxico de um texto poderemos perceber características do tipo ao qual pertence.

A seguir, apresentamos alguns estudos que explicitam o que é norma lingüística.

#### 1.1.4 Norma, uso e “bom uso”

*[...] A língua portuguesa tem avesso e direito:  
o direito é como nós a falamos, e o avesso como a  
falam os naturais [...].  
(Padre Antônio Vieira)<sup>23</sup>*

Formas de comportamento generalizados numa sociedade são considerados padrões culturais. Há os padrões ideais (o que se espera das pessoas) e os reais (o que as pessoas fazem). O padrão ideal seria uma regra para a qual todos tendem, mas que nem todos cumprem. Isso acontece com a língua, que pode ser vista em termos de padrões ideais e padrões reais. O que consideramos língua padrão é um modelo ideal, aceitável em todas as situações. É um modelo tão idealizado que o próprio professor que a ensina não a usa na sua fala. É uma língua prisioneira da gramática normativa, distante da realidade.

A realidade contém variantes, uma questão bastante complexa, que envolve vários fatores, os quais devem ser levados em conta. Assim, temos variantes no tempo (diacrônicas), no espaço (dialetos geográficos), no grupo social (com variantes socioculturais ou dialetos sociais) e por idade, sexo ou assunto, esta última expressa pela terminologia técnica ou científica. Todas essas variantes podem ocorrer de forma imbricada, um mesmo indivíduo utilizando diversas delas. Não esqueçamos que a norma é um fator de coesão social.

<sup>23</sup> Padre Antônio Vieira (*apud* SILVA NETO, Serafim da. *História da Língua Portuguesa no Brasil*. 5ª ed. Rio de Janeiro: Presença, 1988. p. 600).

Exemplifica isso a importância que se dá ao corrigir os desvios da norma em crianças ao lembrá-las de que devem dizer, por exemplo, *eu fiz* e não *eu fazi*. Também corrobora essa valorização a existência de programas de rádio e televisão, colunas em jornais e revistas, além de *sites*, tratando do assunto língua portuguesa, na maioria das vezes abordando apenas a gramática normativa.

Essas informações estão relacionadas à idéia de que norma são os usos e atitudes da classe social de prestígio. Deve ter sido sempre assim, até mesmo antes do auge da cultura grega. Muitas vezes é um dialeto social em nada superior aos demais, mas por ser da classe culta, escolarizada, é apenas por isso prestigiado. Além disso, percebemos que há uma norma escrita, mais conservadora, e uma oral, que aceita inovações. Na norma oral, as diferenças regionais são mais percebidas, notadamente os níveis fonológico e lexical.

Faraco (2002, p. 40) esclarece que

a expressão *norma culta* deve ser entendida como designando a norma lingüística praticada, em determinadas situações (aquelas que envolvem certo grau de formalidade), por aqueles grupos sociais mais diretamente relacionados com a cultura escrita, em especial por aquela legitimada historicamente pelos grupos que controlam o poder social.

Parece-nos que Faraco disse tudo. A camada social mais privilegiada econômica e socialmente determina a língua, isto é, que a língua, fundamentalmente a escrita, usada por aquele grupo seja considerada como a melhor, a que serviria de modelo, porque pelo acesso que tem a todos os meios de cultura e de conhecimento seria o “mais culto”. As outras normas deveriam estar o mais próximo possível daquela para obter algum prestígio.

Um ponto que pode gerar controvérsias é confundir norma culta com *norma-padrão* ou *língua-padrão*, a ponto de vermos a expressão *norma culta-padrão*. Enquanto a norma culta é algo concreto, a norma-padrão é um fenômeno abstrato, como o nome sugere, uma padronização idealizada para abranger todas as regiões geográficas em todos os tempos. São formas contidas e prescritas pela gramática normativa. O padrão almeja ter efeito unificador sobre todas as normas existentes. Compreenderemos isso melhor entendendo como o padrão brasileiro foi construído no século XIX. Não foi levada em conta a norma culta de então, e sim, “a elite letrada conservadora se empenhou em fixar como nosso padrão um certo modo lusitano de escrita, praticado por alguns escritores portugueses do Romantismo” (FARACO, 2002, p. 43).

O que se formou entre o padrão artificial estabelecido e a língua real foi um fosso abissal. O resultado foi a existência até hoje de atitude purista e normativista, que leva a massa de falantes brasileiros achar que não sabe português, que essa é uma língua muito difícil, que todo mundo erra, e assim por diante. Tudo o que não está de acordo com as gramáticas tradicionais tende a não ser considerado. Parece haver uma grande dificuldade de aproximar a norma-padrão da norma culta, afinal, as línguas são um emaranhado de variedades e o padrão é algo abstrato, artificial, idealizado, homogeneizado, até petrificado ousaríamos dizer. Essa pasteurização forçada pelo padrão não está de acordo com a realidade em que as línguas são um conjunto de variedades e sofrem alterações continuamente, principalmente no léxico. Basta observar o que os jovens fazem na sua comunicação via Internet.

Há dois entendimentos para o termo *norma*: 1. modalidade lingüística “normal”, “comum”; 2. “uso regrado como modalidade ‘sabida’ por alguns, mas não por outros” (NEVES, 2003, p. 43). No primeiro caso, essa modalidade é estabelecida pela frequência de uso em camadas sociais, período de tempo e regiões geográficas. Como percebemos, cada camada social, tempo ou região pode ter uma norma, isto é, o normal ou comum no falar de determinados grupos. No segundo caso, há juízo de valor sobre as modalidades em cada zona de variação: umas são mais prestigiadas. Uma modalidade é estabelecida representando o padrão desejável, de forma arbitrária e sustentada por autoridade. Na primeira significação de norma o que está em questão é o apenas o uso, na segunda, o “bom uso”, neste caso apontando para discriminação, havendo estigmas e exclusões. A língua reflete, portanto, os valores da sociedade, a modalidade de prestígio foi eleita como modelo porque socioculturalmente representa o uso de uma elite intelectual do momento e não porque são as “legítimas” e “puras” construções da língua portuguesa (NEVES, 2003, p. 46).

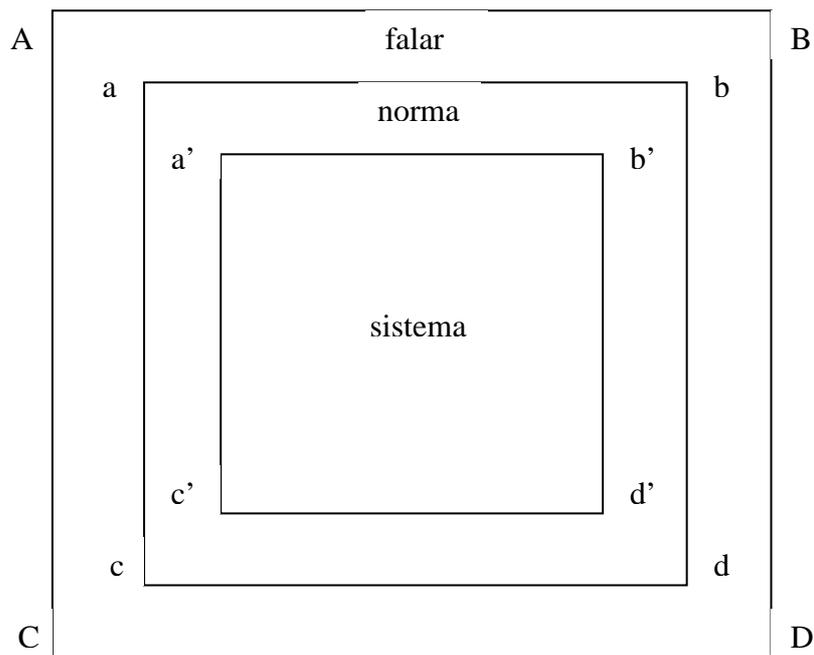
Ao tratarmos de norma, não poderíamos deixar de citar Coseriu e sua contribuição para a lingüística justamente nesse tema. Para a dicotomia saussuriana língua e fala (língua, sistema abstrato, e fala, realização concreta desse sistema), Coseriu (1979) propôs a tríade sistema, norma e fala, considerando importante diferenciar, nas estruturas que constituem a língua, o que é simplesmente normal ou comum (norma), que abrange tudo que é estabelecido e comum nas realizações lingüísticas tradicionais, e o que é oposicional ou funcional (sistema):

O **sistema** é sistema de possibilidades, de coordenadas que indicam os caminhos abertos e os caminhos fechados de um falar compreensível numa comunidade; a **norma**, em troca, é um sistema de realizações obrigatórias, consagradas social e culturalmente: não corresponde ao que se pode dizer, mas ao que já se disse e tradicionalmente se diz na comunidade considerada. O sistema abrange as formas ideais de realização duma língua [...] a **norma**, em troca, corresponde à fixação da língua em moldes tradicionais; e neste sentido, precisamente, a **norma** representa a todo momento o equilíbrio sincrônico (externo e interno) do sistema (COSERIU, p.50).

Sobre *fala*, esse autor entende ser as características concretas, infinitamente variadas e variáveis, dos fatos lingüísticos observados nas infinitas manifestações individuais. Assim sendo, é possível a comprovação de realizações inéditas do sistema. A fala é *como funciona concretamente*. A norma é o modelo de *como funciona*. E sistema é *o que funciona*.

Os conceitos de norma e sistema foram elaborados por Coseriu com base no falar concreto. Os atos lingüísticos são, ao mesmo tempo, criações inéditas e atos de "re-criação" – eles se estruturam na base de modelos já elaborados, e os superam, de alguma forma. Tais modelos aparecem cristalizados numa "língua anterior". Dessa forma, em certo momento, os acontecimentos lingüísticos são normais e tradicionais numa comunidade, compondo, portanto, a norma. Num segundo momento, dessa norma se extraem elementos considerados essenciais, formando oposições funcionais que comporão o sistema.

A concepção de Coseriu é representada por ele mesmo (1979a, p. 72) por meio de quadrados concêntricos: o mais exterior representa os **atos lingüísticos concretos**, o intermediário representa a **norma** (repetição de modelos anteriores no falar, excluindo-se o inédito, o ocasional), e o mais interior corresponde ao **sistema** (com exclusão, na norma, do que não é indispensável, funcional).



**Figura 3: Atos lingüísticos concretos, norma e sistema (Fonte: Coseriu, 1979a)**

A língua deve ser entendida, de acordo com Coseriu, primeiramente, como “função”, depois como “sistema”, uma vez que ela funciona não por ser um sistema, mas, ao contrário, constitui-se num sistema a partir do momento em que cumpre uma função, e essa função liga-se a fatores históricos que, certamente, imprimem marcas na organização sistêmica da própria língua.

Retomando Coseriu, percebemos que ele levou mais longe a distinção que Saussure havia estabelecido entre “langue” e “parole”, ou seja, entre a língua, enquanto sistema abstrato, e a fala, enquanto realização concreta desse sistema. Para Coseriu, a língua pode ser vista a partir de dois níveis de abstração:

- o sistema, conjunto de possibilidades de uma língua, definindo o que pode e não pode ser lingüisticamente realizado, é o que possibilita um falar compreensível numa comunidade;

- a norma, conjunto de imposições sociais e culturais que favorecem o uso de determinadas possibilidades do sistema em detrimento de outras. São realizações obrigatórias, consagradas e compartilhadas numa comunidade. Dentro da norma é possível existirem outras normas (língua familiar, popular, literária, erudita, vulgar, etc.).

A tricotomia do lingüista romeno vai do mais concreto (fala, uso individual da norma) ao mais abstrato (língua, sistema funcional), passando por um grau intermediário: a norma (uso coletivo da língua). Em outras palavras, há realizações consagradas pelo uso e que, portanto, são normais em determinadas circunstâncias lingüísticas, previstas pelo sistema funcional. É à norma que nos prendemos de forma imediata, conforme o grupo social de que fazemos parte e a região onde vivemos. A norma seria assim um primeiro grau de abstração da fala. Considerando-se a língua (o sistema) um conjunto de possibilidades abstratas, a norma seria então um conjunto de realizações concretas e de caráter coletivo da língua.

De acordo com Dacanal (2006, p. 51), “a língua não é apenas sinal e reflexo das estruturas de uma sociedade e da evolução desta ao longo do tempo. Ela é também e diretamente um instrumento de dominação e de exercício do poder” É por isso, informa o autor, que a elite ateniense do século V a.C. e a classe dirigente de Roma davam tanta importância à retórica, ou oratória, a arte de fazer discursos e de convencer. É que a língua é o veículo pelo qual circulam a informação e o conhecimento, sendo também a forma mais simples de mostrar que os possui. O autor está se referindo indiretamente à norma culta, a forma considerada de prestígio pelos que detêm o poder. É por isso que anteriormente ele afirmava que “diante do português de Camões, de D. João VI, do Rio de Janeiro e da TV Globo [...] as variantes *caipira* e *caboclo-sertaneja* do português tendem, se não a desaparecerem definitivamente, como parece mais provável, pelo menos a reduzir-se drasticamente em importância” (DACANAL, 2006, p. 50).

Seria mesmo assim? Há várias correntes contrárias à globalização lingüística, ou pelo menos lutando para salvar as formas peculiares regionais, fazendo, por exemplo, programas de rádio em dialeto do italiano nas regiões colonizadas por imigrantes e publicações de obras regionalistas em que se registram as expressões típicas, além de publicações de dicionários como, entre outros, o de baianês, o de expressões do Ceará e o de Papachibé, a língua paraense. É claro que interesses outros que não lingüísticos (turísticos e artísticos, por exemplo) estão implicados também na manutenção das características regionais. Teremos que concordar, em vista disso, com a idéia que o autor perpassa todo o tempo na sua obra: a língua é um instrumento de poder, adaptável, até certo ponto, aos interesses de quem o detém.

Perguntamo-nos, então, em relação ao *internetês*: há um poder na/da Internet, com suas características de rapidez e grande quantidade de informação, exigindo de seus usuários a adaptação a um ritmo exageradamente acelerado, a ponto de alterar a língua para acompanhar

tal velocidade? Por mais poder que a Internet tenha, parece-nos que a língua, considerada em seu nível culto, continuará com seu status de representação oficial, assim como a maneira de vestir traje e gravata será sempre a forma clássica de um homem bem se apresentar. Pouco importa que os jovens utilizem até formas bizarras de se vestir – são liberdades concedidas à juventude no ambiente dela, em ocasiões especiais estarão os jovens vestindo trajes clássicos e mesmo, toga. Assim, a língua empregada vai ser a “clássica”, por conseguinte, adequada ao momento e à hora determinada pelas relações de poder.

Bechara e Cunha e Cintra são um exemplo do posicionamento dos gramáticos brasileiros sobre norma e língua-padrão. Para Cunha e Cintra (1984, p. 4), a língua-padrão corresponde a “uma entre as muitas variedades de um idioma, é sempre a mais prestigiosa, porque atua como modelo, como norma, como ideal lingüístico de uma comunidade”. Em seguida, os autores se apóiam em Coseriu, concordando que, se uma língua pode abarcar vários sistemas, pode também abranger várias normas “que representam modelos, escolhas que se consagraram dentro das possibilidades de realizações de um sistema lingüístico”.

Já Bechara (2003, p. 37) começa definindo “língua histórica” como um “produto cultural histórico, constituída como unidade ideal, reconhecida pelos falantes nativos ou por falantes de outras línguas, e praticada por todas as comunidades integrantes desse domínio lingüístico”. O termo **língua-padrão** não é usado por esse gramático, que recorre a **norma**, a **língua comum**, língua exemplar. Para Bechara, portanto, **norma** é tudo o “que é tradicional, comum e constante, ou, em outras palavras, tudo o que se diz “assim, e não de outra maneira”. **Língua comum** é uma “unidade lingüística ideal – que nem sempre cala o prestígio de outros dialetos nem afoga localismos lingüísticos”. E também baseado em Coseriu, Bechara (2003, p. 37) refere-se a **língua exemplar** como

um tipo de outra língua comum, mais disciplinada, normatizada idealmente, mediante a eleição de usos fonético-fonológicos, gramaticais e léxicos como padrões exemplares a toda a comunidade e a toda a nação, a serem praticados em determinadas situações sociais, culturais e administrativas do intercâmbio superior.

Esse produto cultural histórico vivo chamado língua varia suas normas do ideal ao real, conforme local, tempo e circunstâncias. Pode até utilizar formas alternativas para estabelecer comunicação. É isso que veremos em seguida.

### 1.1.5 Variantes, formas alternativas com mesmo valor e função

*Variantes lingüísticas são, portanto, diversas maneiras de se dizer a mesma coisa em um mesmo contexto e com o mesmo valor de verdade.*  
(Fernando Tarallo)<sup>24</sup>

Vamos agora percorrer um pouco da Sociolingüística para procurar entender o que acontece com a língua nas mãos dos jovens teclando no computador. Essa subárea da Lingüística

estuda a língua em uso no seio das comunidades de fala, voltando a atenção para um tipo de investigação que correlaciona aspectos lingüísticos e sociais. Esta ciência se faz presente num espaço interdisciplinar, na fronteira entre língua e sociedade, focalizando precipuamente os empregos lingüísticos concretos, em especial os de caráter heterogêneo (MOLLICA, 2003, p. 9).

Se o sistema lingüístico serve a uma comunidade heterogênea e plural, ele deve ser também heterogêneo e plural para conseguir cumprir suas funções. Heterogêneo aqui significa terem as línguas um dinamismo inerente, com variações e mudanças. A variação lingüística ocorre em todas as línguas e pressupõe a existência de formas lingüísticas alternativas, as variantes, um fenômeno variável. Como exemplo, Mollica (2003, p.11) informa que a concordância verbal é uma variável lingüística podendo ser realizada por duas variantes: a marca de concordância do verbo ou a ausência dessa marca. As mesmas formas variantes podem permanecer estáveis nos sistemas durante curto ou longo tempo, podem sofrer mudanças e até desaparecer. As variações, como vimos, tem grau de estabilidade ou mutabilidade. Há fatores que atuam nessas variações. Assim, nas variáveis internas temos os fatores de natureza fono-morfo-sintáticos, os semânticos, os discursivos e os lexicais. Nas variáveis externas encontramos fatores próprios do indivíduo (etnia, sexo), sociais (escolarização, renda, profissão, classe social) e contextuais (grau de formalidade, tensão discursiva). Nos fatores externos, aqueles inerentes ao indivíduo, obviamente, estão centrados traços próprios do falante, enquanto nos demais, por serem características circunstanciais, ora envolvem o falante, ora, o evento da fala.

Como é percebido por todos, as línguas apresentam inovações e modificações e, mesmo assim, se mantêm coesas. É que de um lado há o impulso à variação e de outro, o

---

<sup>24</sup> TARALLO, Fernando. *A pesquisa sociolingüística*. São Paulo: Ática, 2000. p.31.

impulso à convergência, que fundamenta a noção de comunidade lingüística, com seus padrões estruturais e lingüísticos, que poderíamos chamar de normas. Acontece que a heterogeneidade e também a homogeneidade lingüística não são aleatórias, mas reguladas por um conjunto de regras. Temos então a parte fixa e heterogênea em contraponto, conseguindo a façanha de apresentar unidade em meio à heterogeneidade.

As variações lingüísticas podem ocorrer por regiões geográficas (eixo diatópico) ou por estratos sociais (eixo diastrático). Pelo que aparentemente se pode perceber na observação da linguagem dos jovens, há apenas variações típicas da faixa etária ou características de usuários da Internet, sendo o mesmo código utilizado em todo o país. Seria, portanto, variação diastrática a que caracteriza o *internetês*.

São várias as tendências a empregos de formas variáveis motivadas simultaneamente por condicionamentos diversos, como o grau de isolamento geográfico ou social, relações sociais, características das redes sociais, o grau de relação do falante com o meio, os estilos formais e informais na fala e na escrita, o plano da enunciação e o envolvimento com os diferentes gêneros discursivo-textuais. Daí resultarem escolhas que o falante faz para acompanhar o interlocutor, sendo esse o caso dos jovens que necessitam utilizar um mesmo tipo de comunicação para se sentirem inseridos no grupo. Não esqueçamos que as línguas apresentam uma diversidade distribuída em *continuum*, adquirindo o falante primeiramente as variantes informais para depois, paulatinamente, chegar ao domínio da norma culta e, por necessidade e imposição, principalmente da escola, à norma padrão. Toda língua, sabemos, possui variantes mais prestigiadas do que outras. O *internetês*, dentro do contexto da informática e para os jovens, nos parece ser uma variante de prestígio (apenas para aqueles usuários e naquele contexto específico, é claro).

Há formas variantes (normalmente duas) usadas pelos falantes que não alteram a mensagem transmitida. Envolvem principalmente fonologia, morfologia, sintaxe, como nos exemplos *deixe/dexe*, *homem/home*, *pepino/pipino*, *nós/a gente*, *pego/pegado*, *ir a/ir em*, *umas guriinhas inteligentes/umas guriinha inteligente*, *seu casaco/o casaco dele*, *aquele carro/o carro aquele*, *voam as pombas/voa as pombas*. Essas variantes ocorrem em maior ou menor número, dependendo do tipo de falante, sua origem social, acesso a bens materiais e culturais, escolarização, ocupação, grau de inserção em rede social, etc., como já foi anteriormente dito. Outro fator a ser considerado nas variações é o dinamismo das línguas, que se alteram no

tempo, não de um dia para outro, mas de maneira gradual. No caso do *internetês*, em poucos anos surgiu esse escrever que tomou conta dos usuários da rede mundial de computadores. A escrita dos jovens em rede virtual parece ainda não ter impingido modificações à língua. Como a tecnologia modificou a sociedade em grandes dimensões nos últimos 40 anos, essas modificações poderão se refletir também na forma de utilizar a língua. Uma delas é o uso dessa forma chamada *internetês*, por enquanto empregada pelos usuários da Internet, principalmente jovens, mas também pelos usuários em geral devido à praticidade (com ela se ganha tempo e se ocupa menor espaço).

Até agora falamos das mudanças na língua, vamos nos referir também a mudanças que o falante pode realizar ao longo da sua vida, e que não seriam tantas. Naro (2003, p. 44) nos lembra de que no processo de aquisição da linguagem a posição teórica mais aceita é a que

postula que o processo de aquisição da linguagem se encerra mais ou menos no começo da puberdade e que a partir desse momento a língua do indivíduo fica essencialmente estável. Segundo esta posição, a gramática do indivíduo não pode sofrer mudanças significativas porque o acesso aos dispositivos cognitivos que possibilitam a sua manipulação (a chamada faculdade de linguagem) fica bloqueada, uma hipótese que se apóia na psicologia desenvolvimentista. Quaisquer eventuais mudanças seriam apenas esporádicas: troca de uma palavra por outra, troca de pronúncia de uma palavra, etc.

De acordo com essa hipótese, a língua de um adulto reflete o estado de língua adquirido aproximadamente aos 15 anos. Uma pessoa de 80 anos representa a língua de 65 anos atrás e assim todos os falantes revelariam a língua adquirida, *grosso modo*, até os 15 anos. Observemos essa teoria e sua relação com a comunicação dos jovens pela Internet. Se adolescentes são os que mais usam o *internetês*, só daqui a alguns anos iremos perceber se tal forma de comunicação terá repercussões na língua ou não. Também nos parece claro que os jovens não adquirem a língua no computador, sendo ele apenas mais um meio em que eles utilizam a língua adquirida na sociedade em geral para sua interação.

Todas as comunidades têm variações na sua realidade de uso da língua, afinal, as pessoas têm diferenças entre si e é natural que essas diferenças atuem na forma de expressão de cada um (ou de cada grupo). Os primeiros trabalhos variacionistas baseados em Labov<sup>25</sup>

---

<sup>25</sup> William Labov, lingüista americano, no início da década de setenta do século passado, se sobressaiu com seus estudos visando à descrição da heterogeneidade lingüística e à influência de fatores sociais sobre a língua, estudos que fizeram dele o mais renomado representante da teoria da variação lingüística.

abordavam fenômenos morfofonológicos, tendo por premissa básica da variação que duas ou mais formas alternantes ocorrem no mesmo contexto com o mesmo significado. Essa variação de forma escrita percebemos no *internetês*. Na sociolinguística, é consenso que há menos variações na sintaxe do que na fonologia e, agora, na linguagem da Internet, na morfologia, nesse caso, relacionada basicamente à grafia.

Uma das perguntas que ocorre ao tratar das variações é se o significado permanece o mesmo nas formas alternantes. Parece-nos que relacionado à grafia, sim, quando se tratar de uso entre jovens no computador, mas o mesmo não acontece com a sintaxe. Todos sabemos que *uma grande mulher* não é o mesmo que *uma mulher grande*, por exemplo. Outra questão a ser levada em conta é o contexto. Tudo pode variar conforme o contexto, o contexto discursivo e não em sentenças isoladas. Não há significação “estanque no poço dela mesma” como poeticamente se refere João Cabral de Melo Neto à palavra em “situação dicionária”,<sup>26</sup> aqui estendidas às palavras encadeadas em frases na realidade do uso na comunicação. Não podemos esquecer de que a questão do significado e a questão do contexto estão imbricadas.

Ao tentar encerrar, por ora, o assunto variação linguística, vejamos o que Coseriu (1979) nos acrescenta sobre a mudança linguística, relacionado de alguma forma à variação. Diz ele, contrariando a visão saussuriana de que “o sistema é imutável”, que o dinamismo da língua só é possível através da fala: “A língua não existe senão no falar dos indivíduos, e o falar é sempre falar uma língua” (p.33). E nos diz também que “a língua não pode ser isolada dos fatores externos – isto é, de tudo aquilo que constitui a fisicidade, a historicidade e a liberdade expressiva dos falantes” (p. 19). As línguas mudam porque não estão feitas, mas fazem-se continuamente pela atividade linguística, ou seja, mudam porque são faladas. Quem fala é um indivíduo, e por ser este indivíduo histórico, não muda totalmente a sua expressão, utiliza-se de modelos anteriores. As línguas mudam porque é inerente à natureza delas que elas mudem, pois fatores externos exercem influência sobre elas. Se assim não fosse, as línguas estariam fadadas a uma estabilidade eterna, já que somente as línguas abstratas não mudam.

---

<sup>26</sup> Versos 5 a 12 do poema *Rios sem discurso*: Em situação de poço, a água equivale/ a uma palavra em situação dicionária:/ isolada, estanque no poço dela mesma,/e porque assim estanque, estancada; /e mais: porque assim estancada, muda/e muda porque com nenhuma comunica, /porque cortou-se a sintaxe desse rio,/o fio de água por que ele discorria. (MELO NETO, João Cabral de. *A educação pela pedra*. Rio de Janeiro: José Olímpio, 1979).

Variações sistemáticas de uma comunidade de fala, em uma heterogeneidade estruturada, levam à conclusão de que não existe um caos lingüístico, uma impossibilidade de sistematização, e, sim, uma organização heterogênea da língua. As formas em variação são denominadas "variantes lingüísticas", que são as diversas maneiras de se dizer a mesma coisa em um mesmo contexto e com o mesmo valor de verdade. Um conjunto de variantes recebe o nome de "variável lingüística".

Variação é possibilidade de estilos diversos (POSSENTI, 1986), decorre em função de situações formais ou informais, classe social, formação escolar do falante, entre outras. Com o advento da sociolingüística descobriu-se que a gramática é um conjunto de regras variáveis, que as expressões existem para desempenhar um papel e não apenas para ser construídas pelos locutores. "As variantes são tratadas como espécie de sinônimas, distribuídas segundo determinados fatores [...], mas basicamente com a mesma função semântica, porque sua referência é a mesma" (*op. cit.*, p.110). As formas lingüísticas têm uma história, sendo carregadas de marcas com o passar do tempo. Tais marcas selecionam palavras como sinônimas, algumas pronúncias como variantes, determinada construção sintática e não outra, por exemplo. É a história das palavras, sua distribuição social e não só gramatical, que permite efeitos diversos, alguns bem aceitos socialmente e outros não. É o caso das variações de estilo ou das que denotam categoria social (classe, etnia) ou categorias sócio-psicológicas (arrogância, pretensão).

Para ser considerado variação, é necessário um termo de comparação fixo, sendo a FORMA mais usualmente aceita aquela que COINCIDE com a prescrição da norma-padrão. Para variantes de estilo, podemos considerar a noção de escolha e não de desvio (anomalia) da norma prescrita. As variações, em geral, são escolhas entre alternativas possíveis, a mais distante da norma é a que chama mais a atenção, pelo estranhamento causado, que pode ser positivo (artístico) ou negativo (indicar pouco domínio da norma de prestígio e ser estigmatizante).

Observemos um *scrap* retirado do nosso *subcorpus Amazonas*. É um exemplo ilustrativo de variação, na forma de escrita do *internetês* em relação ao padrão de escrita culta "normal". Há aqui um caso de variação diastrática, típica de uma faixa etária e característica de usuários jovens da Internet:



Vejamos o que os Dicionários registram sobre esse assunto. Para Dubois (1978, p. 184), “o dialeto é uma forma de língua que tem seu próprio sistema léxico, sintático e fonético, e que é usado num ambiente mais restrito que a própria língua”. Mattoso Câmara (1984, p. 141) vê o dialeto sob dois pontos de vista: 1º: ponto de vista puramente lingüístico: “os dialetos são falares regionais que apresentam entre si coincidência de traços lingüísticos fundamentais”; 2º: ponto de vista extralingüístico: *a)* existe um sentimento lingüístico comum em que os dialetos são sentidos como variantes de uma língua geral e ideal; *b)* existe uma língua culta, superposta aos dialetos que ficam limitados ao uso cotidiano, sem maior expressão literária; *c)* existe subordinação política das regiões como partes de um estado político nacional. Dubois (1978, p. 184) faz distinção entre dialeto regional e dialeto social. Este último “é sistema de signos e de regras sintáticas usado num dado grupo social ou em referência a esse grupo. [...] Quando esse sistema é criado ou empregado como um conjunto secreto de signos, trata-se de uma gíria (gíria dos malfeitores, dos soldados, de certos trabalhadores migrantes)”. Também podemos chamar de *socioleto* essa variação da língua usada por grupos de indivíduos que, tendo características sociais em comum, se distinguem dos demais falantes na sua comunidade. Parece-nos ser o caso do *internetês*.

Houaiss (2001), por sua vez, dá seis acepções para dialeto, além de informar a etimologia, que vale a pena ser observada<sup>28</sup>. Vejamos, então, a transcrição das duas primeiras informações do Houaiss para o verbete *dialeto*:

**1** LING conjunto de marcas lingüísticas de natureza semântico-lexical, morfossintática e fonético-morfológica, restrito a dada comunidade de fala inserida numa comunidade maior de usuários da mesma língua, que não chegam a impedir a intercomunicação da comunidade maior com a menor [O dialeto pode ser *geográfico* ou *social*.] cf. *registro*, *jargão*, *patoá*, *gíria* **2** LING qualquer variedade lingüística coexistente com outra e que não pode ser considerada uma *língua* (p.ex.: no português do Brasil, o *dialeto caipira*, o *nordestino*, o *gaúcho* etc.) cf. *falar*, *linguajar*.

O Dicionário Aurélio (1980) simplesmente registrava dialeto desta forma: “Variante regional de uma língua”. Mas na edição de 1999 foram acrescentadas mais informações:

<sup>28</sup> ETIM lat. *dialéctós* ou *dialéctus*, *i* 'linguagem particular de um país, maneira de falar', adp. do gr. *diálektos*, *ou* fem. 'conversa, discussão por perguntas e respostas, linguagem corrente, linguagem própria de um país', ligado ao v.gr. *dialégó* 'distinguir, separar' cog. de *légo* 'dizer, falar, pronunciar uma palavra' e de *léksis, eós* 'palavra, uso da fala', por infl. do fr. *dialecte* (1550); cp. ing. *dialect* (1650) e demais línguas.

1. Variedade subpadrão ou não-padrão de uma língua associada a grupos que não contam com prestígio social [Sin.: *linguajar* (2)] 2. Variedade regional de uma língua que conta com forte tradição literária. 3. Variedade lingüística regional que não tem escrita; patoá. 4. Cada uma das subdivisões que se podem aplicar a determinada língua, utilizando como critério básico a região geográfica ou a camada social a que pertence o falante; variedade [V. falar (28)] (FERREIRA, 1999, p. 676).

Como pudemos perceber, dos autores revisados, os dialetos são aspectos regionais de uma língua. Podem ser analisados do ponto de vista puramente lingüístico, como línguas regionais com traços lingüísticos fundamentais entre si, mas a esse conceito lingüístico pode ser acrescentado um conceito extralingüístico de ordem psíquica, social ou política. Foi por motivos extralingüísticos que alguns dialetos passaram a ser a língua-padrão de uma nação, como os conhecidos casos dos dialetos de Castela e de Florença. Aliás, toda língua-padrão não deixa de ser um dialeto, porque lingüisticamente não existe uma língua superior a outra.

Como vimos, *dialeto* pode ser resumido em três acepções bem distintas: a) como sinônimo impróprio de língua e idioma em geral; b) como forma inferior de uma língua, e c) como uma subdivisão de um idioma, geograficamente ou socialmente estabelecido. Tentando fazer distinção entre língua e dialeto que não se apoiasse em fatores políticos ou sociológicos, foram buscados critérios relacionados aos aspectos comunicacionais. Coseriu propôs o chamado critério da intercompreensão, segundo o qual dois falares podem ser considerados dialetos da mesma língua se seus falantes conseguem compreender-se mutuamente, caso contrário, teremos duas línguas diferentes. Seria esse um bom critério, já que se apóia num dado subjetivo? Falantes do português e do espanhol podem entender-se relativamente, portanto seriam dialetos, segundo o lingüista romeno. Seria interessante aplicar esse critério para a escrita da Internet: um adulto pouco familiarizado com computadores teria grande dificuldade de entendimento. Então haveria, nessa ótica, não um dialeto, mas uma língua diferente.

Outra informação sobre o tema o mesmo autor nos traz, ao afirmar que toda língua histórica pode ser constituída por dialetos:

Um dialeto, sem deixar de ser intrinsecamente uma língua, se considera subordinado a outra língua, de ordem superior. Ou, dizendo-se de outra maneira: o termo dialeto, enquanto oposto a língua, designa uma língua menor incluída em uma língua maior, que é, justamente, uma língua histórica (ou idioma). Uma língua histórica - salvo casos especiais - não é um modo de

falar único, mas uma família histórica de modos de falar afins e interdependentes, e os dialetos são membros desta família ou constituem famílias menores dentro da família maior. (COSERIU, 1982, p. 11-12).

É interessante registrar que a discussão sobre variação dialetal vem de longa data. Já os gregos - de quem nos chegou a noção de dialeto - distinguiam quatro variantes regionais de sua língua - o eólico, o jônico, o dórico e o ático. Foi a partir do século IV a.C. que adotaram, com base nesse último, a *koiné dialektos*, ou "língua comum", como meio de intercomunicação. Também entre os romanos - cuja sociedade era estratificada - encontram-se menções à variabilidade de natureza social. A linguagem corrente - como testemunham alguns escritos latinos - recebia subclassificações, como *sermo urbanus*, *sermo plebeius*, *sermo rusticus*.

Variações da língua realizadas pela comunidade de falantes em geral vem sendo produzidas desde há muito tempo. Existiria, também, alguma variação individual a ser considerada como importante produzida pelo próprio falante? Há, como vimos, uma língua abrangente, a norma com seu núcleo comum, que sofre variações por região ou por grupos sociais. Seriam significativas as variações praticadas pelos usuários individualmente? É o que veremos a seguir.

### 1.1.5.2 Idioleto, forma individual de expressão

*A história é sempre a história de uma sociedade,  
mas com toda a certeza a de uma sociedade de indivíduos.*  
(Norbert Elias)<sup>29</sup>

Originado dos radicais gregos ἴδιος, próprio, particular, peculiar e λέγω, dizer, falar, pronunciar, *idioleto* é registrado por Houaiss (2001) da seguinte forma:

/é/ s.m. LING o sistema lingüístico de um único indivíduo num determinado período de sua vida, que reflete suas características pessoais, os estímulos a que foi submetido, sua biografia etc.; idiolecto [Pertence ao campo da *langue*, e não da *parole*, porque trata de particularidades lingüísticas constantes, não fortuitas.]

<sup>29</sup> ELIAS, Norbert. *A sociedade dos indivíduos*. Lisboa: Publicações Dom Quixote, 1993, p. 67.

Essa idéia é confirmada por Mattoso Câmara (1977, p. 9), ao informar que o termo foi criado pelos norte-americanos e que o mesmo se refere à língua de um único indivíduo. Na verdade, é um conjunto de hábitos verbais de um indivíduo, hábito que reflete as influências regionais e sociais. É a variação da língua única de um falante. Manifesta-se pela escolha de palavras, frases ou metáforas únicas desse indivíduo. Cada um de nós tem um idioleto, com arranjos de palavras e frases que nos é único, embora não signifique que usemos palavras específicas, não usadas por outros.

Para Dubois (1978), *idioleto* pode ser considerado o estilo de se comunicar do indivíduo, com seu conjunto de usos de uma língua num determinado momento. “Idioleto é o conjunto dos enunciados produzidos por uma só pessoa, e principalmente as constantes lingüísticas que lhes são subjacentes e que consideramos como idiomas ou sistemas específicos” (DUBOIS, 1978, p. 329).

Coseriu (1979a, p. 23) nos lembra que “a fala é atividade de falar, em geral, mas também cada ato particular de falar. O caráter social ou individual, por si só, não é determinante, pois o ato lingüístico é ao mesmo tempo social e individual”.

Diz mais Mattoso Câmara Jr. (1980, p. 27):

a língua é, de maneira geral, coletiva; mas cada um de nós tem certas peculiaridades lingüísticas, ou pelo menos preferências, e há assim, de certo modo, múltiplas línguas individuais, ou idioletos [...]. O estilo é em princípio, individual, [...] mas os traços estilísticos coincidem, em grande parte, nos indivíduos de uma sociedade lingüística [...].

Retomando a fala de Mattoso Câmara no que se refere às múltiplas formas individuais, perguntamo-nos assim como fez Pimentel (2006, p. 2): No meio eletrônico estaria surgindo uma sociedade lingüística com suas peculiaridades ou o *internetês* seria apenas uma transcrição da linguagem oral para a escrita sem a preocupação da formalidade gramatical?

Encontramos traços do idioleto no *internetês* como em toda língua. Manifesta-se pelo estilo próprio de cada redator-internauta. Esse estilo vê-se pelo uso majoritário de determinadas palavras, repetição de outras, os mesmos *emoticons* empregados muitas vezes e pelas formas para indicar riso de sua preferência, por exemplo.

Como o *internetês* é basicamente um registro escrito, faz-se necessário revisarmos o assunto *escrita*. Vejamos, então, a seguir.

### 1.1.6 Escrita, uma representação.

*A escrita é importante na escola,  
porque é importante fora dela e não o contrário.  
(Emília Ferreiro)<sup>30</sup>.*

Começamos refletindo sobre a escrita desde quando a adquirimos, ainda na infância. Para Ferreiro e Teberosky (1999, p. 26), “sendo a escrita uma maneira de transcrever a linguagem, tudo muda se supomos que o sujeito que vai abordar a escrita já possui um notável conhecimento da língua materna, ou se supomos que não possui”.

As mesmas autoras lembram que a teoria de Piaget lhes permite “introduzir a escrita *enquanto* objeto do conhecimento, e o sujeito da aprendizagem *enquanto* sujeito cognoscente”. A criança, então, ao aprender a escrever não identifica a escrita apenas como cópia de um modelo. Escrever não é apenas uma técnica de reproduzir o traçado gráfico nem regras de transcrição do oral. “Escrever é uma tarefa de ordem conceitual. [...] a escrita não é uma cópia passiva e sim interpretação ativa dos modelos do mundo adulto” (*op. cit.*, p. 37). Ao aprender a escrever, produzindo traços sobre o papel, a criança “põe em jogo suas hipóteses acerca do próprio significado da representação gráfica” (*op. cit.*, p. 37). Ferreiro e Teberosky (1999, p. 70) dizem mais: “A escrita também é um objeto simbólico, é um substituto (significante) que representa algo. Desenho e escrita - substitutos materiais de algo evocado – são manifestações posteriores da função semiótica mais geral”. Com a diferença de que a escrita, ao contrário do desenho, não tem semelhança com o objeto (ou acontecimento) referido e constitui, como a linguagem, um sistema de regras próprias. A escrita constitui um tipo específico de objeto substituto.

Literal e resumidamente, a escrita é a representação gráfica do discurso, aqui entendido segundo as palavras de Mattoso Câmara (1984, p. 99): “em sentido lato, é o termo

---

<sup>30</sup> FERREIRO, Emília. *Cultura escrita e educação*. Porto Alegre: Artes Médicas, 2001, p. 19.

que melhor corresponde em português ao termo francês PAROLE, estabelecido por Saussure”. E é a escrita uma das formas assumidas pela língua. Até boa parte do século XX, de acordo com Sampson (1996, p. 7), a lingüística ignorou completamente a escrita, considerando língua basicamente a expressão oral. Além de poucos, os estudos sobre a escrita ignoraram a teoria e a metodologia lingüísticas. Esse valorizar a língua falada foi uma reação contra uma antiga tradição dos estudos da língua em que se abordava o assunto língua com espírito avaliador, preocupando-se com a identificação do “bom”, do “aprovado”, erradicando-se o que era visto como erros lingüísticos. Quem vê a língua dessa forma se concentra na língua escrita, pois ao escrever procuramos seguir todas as prescrições, a fim de obter um produto final elaborado.

De maneira geral, a escrita é vista como um aspecto de tecnologia, algo que as pessoas usam, não fazendo parte de suas personalidades. Os sistemas de escrita “são claros instrumentos idealizados para a execução de uma tarefa, que podem desempenhar mais ou menos bem [...]” (SAMPSON, 1996, p. 15). O mesmo autor se pergunta o que é a escrita e responde como sendo “um sistema para se representar enunciados da língua falada por meio de marcas permanentes e visíveis” (*op. cit.*, p. 25).

As discussões teóricas entre lingüistas, para quem a língua é basicamente a expressão oral, não chegam ao falante comum. Para ele, a percepção de língua que tem é a da que lhe é cobrada pela escola, nas provas e concursos para ascensão social, aquela que lhe dá prestígio, enfim, há uma norma-padrão a ser seguida. Essa norma se centra muito na escrita, até porque as gramáticas mais difundidas são as normativas. E a maioria das gramáticas trata as relações entre fala e escrita tendo como parâmetro a língua escrita. A impressão que passa para o usuário é que a fala é o lugar do erro, sem regras, sendo a escrita formal, com regras. Cria-se a noção de confundir língua com a gramática codificada. E se esquece que tanto a fala como a escrita são um *continuum* que vai do nível mais informal até o mais formal.

Favero et al. (1994, p. 276) mostram um quadro em que são consideradas as condições de produção entre fala e escrita. Como a escrita na Internet parece ser basicamente um registro gráfico da fala, podemos observar como tais condições distintas em cada modalidade acabam se mesclando na prática comunicativa via computador.

**Quadro 1: Condições de produção de fala e escrita**

Relações entre fala e escrita	
Condições de produção	
- Interação face a face	- Interação à distância (espaço temporal)
- Planejamento simultâneo ou quase simultâneo	- Planejamento anterior à produção
- Impossibilidade de apagamento	- Possibilidade de revisão para operar correções
- Sem condições de consulta a outros textos	- Livre consulta
- Ampla possibilidade de reformulação: é marcada, pública, pode ser promovida tanto pelo falante como pelo ouvinte	- As reformulações não podem ser tão marcadas, é privada e promovida apenas pelo escritor
- Acesso imediato ao <i>feed-back</i> do ouvinte	- Sem possibilidade de <i>feed-back</i> imediato
O falante pode processar o texto, redirecionando-o a partir das reações do ouvinte	- O escritor pode processar o texto a partir das possíveis reações do leitor

(Fonte: Fávero et. al., 1994)

A posição de Mattoso Câmara (1984, p. 108) ilustra a concepção de escrita como representação: “A escrita é representação visível e durável da linguagem, que, de falada e ouvida, passa a ser escrita e lida. [...] Assim se estabelece numa língua dada a escrita ao lado da fala [...]. (A escrita) consiste numa transposição do discurso falado, de que resultam novas condições de funcionamento da linguagem”.

O mesmo Mattoso Câmara (1979) alerta para a diversidade, que classifica de muito sutil e enganadora, existente entre a fala e a escrita. É esta última a focalizada pelas gramáticas normativas escolares. Mesmo que o estudante já vá para a escola sabendo falar satisfatoriamente, é lá que vai aprender a técnica da língua escrita. Por a língua escrita se manifestar em condições muito diversas da língua oral é que estudantes que falam muito bem têm dificuldades para redigir. A fala acontece em situação concreta, estimulada por um ou vários falantes, bem individualizados, o que não ocorre ao produzir a língua escrita. Além disso, a escrita não é reprodução fiel da fala, não sendo verdadeira a metáfora de que “ela é a roupagem da língua oral” (1979, p.9). Há uma diferença fundamental entre esses dois tipos de linguagem.

Mattoso ainda nos lembra que, em sociedades como a nossa, a língua escrita se sobrepõe à língua oral por reger toda a vida geral do país. É por isso que a gramática normativa

dá grande atenção à escrita, e, em assim sendo, é ela que a escola deve ensinar em primeiro lugar.

A escrita é uma face da língua e como sua representação foi útil à ciência. Escrita é expressão do domínio das estruturas lingüísticas, adquiridas através do exercício escrito: a produção de texto (ENDRUWEIT, 2006). É instrumento para comunicar através de textos, o que reforça a concepção de escrita apenas como representação.

Segundo Endruweit (2006, p. 34), “a exclusão da escrita foi certamente um dos dramas de Saussure”. Sobre esse tema, a escrita, o mestre genebrino afirmou:

Língua e escrita são dois sistemas distintos de signos; a única razão de ser do segundo é representar o primeiro; o objeto lingüístico não se define pela combinação da palavra escrita e da palavra falada; esta última, por si só, constitui tal objeto. Mas a palavra escrita se mistura tão intimamente com a palavra falada, da qual é a imagem, que acaba por usurpar-lhe o papel principal; terminamos por dar maior importância à representação do signo vocal do que ao próprio signo. É como se acreditássemos que, para conhecer uma pessoa, melhor fosse contemplar-lhe a fotografia do que o rosto. (SAUSSURE, 2004, p. 34).

Mais adiante, Saussure (p. 40), retoma o assunto, sustentando que “a escrita obscurece a visão da língua; não é um traje, mas um disfarce”. Endruweit (*op. cit.*, p. 36) refere-se a essa afirmação dizendo que “a escrita seria o significante do significante primeiro, capaz de representar a voz, cuja primazia situava-se na relação direta e natural com o sentido”.

Saussure (p. 39) menciona incoerências na escrita, dentre elas, por exemplo, a “multiplicidade de signos para representar um som” e as ortografias flutuantes. Mas Saussure (p. 35) também explica o prestígio da escrita. Para ele:

- 1º . A imagem gráfica impressiona como objeto permanente e sólido;
- 2º. As impressões visuais são mais nítidas e duradouras que as impressões acústicas;
- 3º. “A língua literária aumenta ainda mais a importância imerecida da escrita” (p.35). Há livros e “é conforme o livro e pelo livro que se ensina na escola” (p.35), a língua aparece como um código regulamentado;
- 4º. Havendo desacordo entre língua e ortografia, a escrita apresenta solução.

Endruweit (2006, p. 46), referindo-se à relação fala-escrita, afirma ser a escrita uma presença-ausência, porque é utilizada como “suplemento, como recurso artificial a que se lança

mão quando não se pode usar o meio natural que é a voz. Essa dissimulação pretende tornar presente a fala quando ela está ausente. Uma presença-ausência”.

Podemos entender que a escrita também possui um sistema de relações. Saussure (p.138) nos diz que a escrita é um “outro sistema de signos”, esclarecendo:

- 1º. Os signos da escrita são arbitrários, não há relação entre uma letra e o som que a representa;
- 2º. O valor das letras é puramente negativo e diferencial, uma letra pode conter variantes, mas não pode ser confundida com outra;
- 3]. Os valores da escrita só funcionam pela oposição recíproca dentro de um sistema definido composto por número determinado de letras;
- 4º. O meio de produção do signo é indiferente, não importam as cores das letras, seu tamanho, espessuras, etc.

A essas informações saussurianas, temos Endruweit nos alertando:

Somente em relação à materialidade sonora é possível entender a escrita como representação do som e, por conseguinte, vê-la como sua subordinada. Ao assumir um significante incorpóreo, a escrita perde o contato com o som e pode, a partir de então, ser também considerada um sistema de signos. Pode, principalmente, ser aceita em posição de igualdade em relação à língua, sendo submetida à noção de valor. O que temos agora é a escrita dotada de estrutura, ratificando o conceito saussuriano de que *a língua é forma e não substância*, ao que se pode acrescentar: a língua não é som, ela se manifesta também na substância sonora ou escrita. (ENDRUWEIT, 2006, p. 56)

Nos diz mais a autora, ainda sobre fala-escrita: “Pode-se, portanto, entender a fala como um ato fonatório, e sonoro, tendo como consequência a exclusão da escrita; e/ou como uso da língua, permitindo a inclusão da mesma” (p. 70). E acrescenta que é mito acreditar que só a escrita desenvolve o raciocínio lógico e abstrato. É também errado pensar que a fala é o lugar do pensamento concreto e a escrita do pensamento abstrato. A escrita possibilita acesso a maior número de conhecimentos. A escrita também transcende o espaço e dá a idéia de fidelidade. “De certa forma, é como se a escrita substituísse o homem” (p. 82). Nas sociedades sem escrita, bastava a palavra, um juramento oralmente feito era o suficiente. Com a aquisição da escrita, o juramento deu lugar ao manuscrito, o que passou a valer foi o registro e não a palavra empenhada.

Mais ainda nos diz Endruweit (p. 85): a escrita “se presta à análise, à separação de suas partes e ao retorno reparador sobre o que foi escrito, mas talvez sua principal função seja a de armazenar”. As implicações sociais e intelectuais da cultura escrita, na escola e na sociedade como um todo, são compreendidas pela função de arquivar da escrita. Desses pensamentos seria interessante ponderar sobre o papel do arquivamento monumental de escrita que é o *site* do Orkut. Que razões levam os jovens a registrar e a expor tanto de suas vidas nesse veículo?

Para concluir, lembremos sobre o assunto as palavras de Luft (1985), que pode sintetizar os pensamentos vistos e ditos de várias maneiras pelos diferentes autores:

Evidentemente a Lingüística nada tem contra a escrita, onde sem dúvida a língua atinge os níveis mais elevados de expressão, ao menos mais rebuscados, sofisticados, sutis, até pela possibilidade de revisar, corrigir, aperfeiçoar o texto. Mas talvez seja bom que a ciência da linguagem nos lembre que não haveria língua escrita se primeiro não houvesse a língua falada. E que a escrita é um sinal secundário: representa (imperfeitamente, claro) a fala, que por sua vez representa o pensamento. É, pois, a escrita, um “sinal de sinal” (LUFT, 1985, p. 72).

## 1.2 O INTERNETÊS

*[..] uma horda de zulus eletrônicos, que pularam o livro e foram direto para o teclado [...] está mais do que na hora de nós fazermos uma defesa da norma culta da língua portuguesa. Eu já disse isso, eu considero língua portuguesa assassinada a tecladas.*

(Deonísio da Silva)<sup>31</sup>

Embora sejam encontrados vários trabalhos e publicações relacionados à linguagem da Internet, apenas alguns artigos de curta extensão tratam de como é essa linguagem utilizada na comunicação eletrônica, também conhecida por *internetês*. Relacionado a esse termo, é oportuno esclarecer que ele também se refere ao glossário dos termos utilizados na Internet, a maioria em inglês, a ponto de existir desde 1995 o *Dicionário de Internetês*, de António Miguel Caetano Ferreira, editado em livro no Brasil e em Portugal em 1997. A versão eletrônica mais atualizada dessa obra se chama *Dicio.net*. Conforme já foi dito, o termo

---

<sup>31</sup> Resumo do programa. Linguagem de Internet. Disponível em:

<[http://www.tvebrasil.com.br/observatorio/arquivo/principal\\_050329.asp#programacao](http://www.tvebrasil.com.br/observatorio/arquivo/principal_050329.asp#programacao)> Acesso em: 2/4/2007.

*internetês* neste trabalho tem o sentido de linguagem surgida no ambiente da Internet, baseado na simplificação informal da escrita, em uso principalmente nos bate-papos (*chats*), Orkut, blogues, troca de arquivos, mensagens de celulares, fóruns de Internet e correio eletrônico. Trata-se de um todo de escrita que pode ser subdividido em diferentes instâncias.

Consultando o Banco de Teses da CAPES em janeiro de 2007, obtivemos a informação de haver 75 teses/dissertações tratando do assunto “língua da Internet”. Dessas, muito poucas estão relacionadas à área da Lingüística, e dentro dela, tratam sobre ensino e aprendizagem de línguas fazendo uso da Internet, ou de como utilizar a rede como recurso pedagógico ou fonte de pesquisas. Os demais trabalhos foram realizados nas mais diversas áreas, como Educação, Informática, Artes, Administração, Comunicação e Engenharia.

A autora que mais se aproxima do proposto nesta pesquisa é Fonseca (2001), com trabalho inserido nos campos da educação tecnológica e da linguagem, procurando observar em sua dissertação de mestrado as características dessa linguagem, se tal linguagem eletrônica estaria se constituindo numa re-oralização da língua e, ainda, se a Internet estaria modificando o português da norma culta. A autora verificou que a linguagem da Internet é considerada um registro híbrido, com características da fala e da escrita, constituindo um *contínuum* topológico, que vai do mais informal ao mais formal. Quanto à re-oralização da língua, constatou marcas da oralidade na linguagem da Internet, mais fortes no bate-papo do que no correio eletrônico. Infelizmente, a referida autora nada registrou sobre a escrita no Orkut porque na época de sua pesquisa (2001) ainda não existia essa comunidade virtual ou rede social (criada em janeiro de 2004). Fonseca analisou especificamente a linguagem de *chats* e *e-mails*. Nossa pesquisa, ao contrário do que ofereceu, detém-se apenas na observação das formas mais presentes na escrita dos *scraps* e dos depoimentos do Orkut.

Observando o *internetês* no nosso *corpus* do Orkut, constatamos ser ele um amálgama de fala/escrita, uma vez que é uma “escrita oralizada”. Utiliza pontuação e alguns acentos, mas faz transcrição dos fonemas tais como usados ao falar, por exemplo, *brigadu* e *naum*. Sobre isso, Eisenkraemer (2006, p.2), referindo-se ao *internetês*, lembra que ele apresenta características do código escrito e do código oral, “ou seja, um ‘código escrito oralizado’, ou uma ‘fala/escrita criptografada’”.

Ainda de acordo com Eisenkraemer, à primeira impressão, o *internetês* aparece como uma simplificação da língua, em que são abundantes as abreviações e siglas, no entanto, há algumas ocorrências contrárias, em que aparece complexificação. É o caso do acréscimo de letras à palavra original, intencionalmente, a fim de expressar, por exemplo, a força do significado de uma palavra, ou então, o uso das letras “n”/“m” e “h” para indicar nasalização e acento agudo das oxítonas, respectivamente, como em *bejaum* e *ateh*. Percebemos que, em princípio, nem as abreviações e nem os acréscimos são aleatórios, sendo necessário maiores observações para fazermos uma afirmação categórica a respeito disso. Parece-nos que há normatização, uma certa lógica nessa escrita simplificada ou alterada graficamente.

No *internetês* há muito de simbologia. Cada sinal, um asterisco, um ponto-final ou de exclamação, por exemplo, tem sua significação. Não podemos analisar esses sinais isoladamente, mas somente em seu contexto. A pontuação é um grande indicador de expressividade e é muito empregada<sup>32</sup>. Como ferramenta de representação do gestual e da entonação, o uso de exclamações, interrogações e reticências é bastante enfático. Alguns símbolos tomam o lugar de palavras, como o sinal “+” para indicar a palavra “mais” e, algumas vezes, até a conjunção “mas”, o “-” para substituir “menos” e “T+” significando “até mais”. Outro recurso também muito aproveitado é a letra maiúscula, como também a repetição de letras, que representam uma alteração de voz, um grito ou uma ênfase na sílaba ou palavra. São mais regras estabelecidas pelo emprego entre os usuários de uma linguagem que aparentemente parece não se prender a regra nenhuma.

Outra característica atual é a utilização dos chamados *emoticons* ou *charactereta*, desenhos que representam caretinhas e personagens, usados para expressar as emoções de quem está teclando ou para sinalizar despedidas. Antes desses desenhos, colocados prontos em menus para escolha imediata, havia o recurso de utilizar combinação de caracteres disponíveis no teclado. Vejamos alguns exemplos:

<sup>32</sup> Conforme se pode verificar no trecho abaixo extraído do nosso *corpus*, chegamos a ter 44 sinais de pontuação em um segmento de 83 palavras. Um segmento “normal”, como o parágrafo em que se insere a nota 29, tem 146 palavras e 19 sinais de pontuação (11 vírgulas e 8 pontos finais).

*ola igor cm vc esta um abraço!!!!!!!!!!*

*amor da minha vidaa... como ce tá?!!!*

*biruta foi massa... foi eu e a pri.. e uma galerrraa lá da bio.. e um povo que era migo do povo.. foi massa viu..*

*muita putaria.. a bruna foi tb...*

*e ontem ainda fui pro dragão com a sil.. assiti o otto.. e a gente depois encontrou a mon com o felipe...*

*show ontem tb foi irado..*

*mas dragao do mar tá decadente...*

*saudade de tu*

*te amo*

: )	Feliz	: ((	Chorando	: o)	Palhaço
: (	Triste	: ))	Gargalhando	: #	Guardar segredo
: x	Apaixonado	:	Cara de tacho	: -@	Gritando.

Acreditamos que o *internetês*, por ser utilizado basicamente no teclar *on-line*, similar à fala, é um registro efêmero. Essa efemeridade, porém, não se revela em todas as ocorrências. Ela existe, sim, nas comunicações síncronas via computador, como nos *chats*. Já na escrita do Orkut, por exemplo, fica armazenada no *site*, podendo permanecer por tempo indeterminado, o que ocorre nos casos em que o usuário tenha morrido. Essa sensação de ser efêmera talvez ocorra pela similaridade com a língua falada, a qual raramente se corrige. A língua escrita para comunicação na *web*, então, também não implicaria correções por conter essa impressão de ser fugaz (mesmo que depois permaneça registrada).

A maioria dos autores que escrevem sobre a língua da Internet se refere a ela como um fenômeno dos tempos pós-modernos. Aceitam a legitimidade de sua existência e a vêem, em situações específicas de comunicação, adaptada ao teclar, à tela do computador, às mudanças da sociedade, enfim. Um autor, porém, esbraveja desde 2005 contra o *internetês*. Silva (2007), ao utilizar como título de um texto “Português assassinado a tecladas”, já mostra que seu objetivo é menosprezar essa forma de comunicação. Para ele, o *internetês* é “mais um besteiro, a que os apressados de sempre já deram até nome: idioma cibernético”. É interessante observar a sua dificuldade explícita em nomear o *internetês*. Para esse autor, nem sequer deve ser nomeado.

Também afirma não entender como respeitadas intelectuais consideram normal o que está ocorrendo com a língua utilizada pelos jovens na Internet. Do seu ponto de vista, não há como defender essa forma de comunicação. Diz-nos que só emburrece os jovens, levando-os a não conseguirem se comunicar:

Pois na dita linguagem cibernética a língua portuguesa está sofrendo de diarreia e tenesmo ao mesmo tempo. Ora o jovem diz demais e confusamente, economizando em letras, mas se perdendo em prolixias, ora está preso ao reduzido universo vocabular que o vitima principalmente na escola. Como aprender um texto sofisticado, se professores e livros, por melhores que sejam, não conseguem contato com repolhos e alfaces ali matriculados? (SILVA, 2007).

Diz também que com o vocabulário utilizado pelos jovens nesse tipo de comunicação não é possível pensar, e que eles nem conseguiriam se comunicar. Resumindo: “[...] se se restringirem à linguagem cibernética perderão via de acesso indispensável ao êxito no trabalho, no amor, na vida: a capacidade de entender e de serem entendidos”.

Parece-nos um exagero tais afirmações. Os jovens não utilizam o computador o tempo todo, nem estão todo o tempo em contato com tal tipo de linguagem. Há situações e situações. Qualquer ser pensante sabe distinguir, em tese, quais são ocasiões em que se pode ou não fazer determinadas coisas. Assim, os jovens saberão que na comunicação entre eles na rede mundial de computadores é possível usar um código simplificado, típico de determinada faixa etária na comunicação informal. Se tal linguagem for empregada em situações que requeiram o nível padrão, mais errados do que os jovens serão aqueles que não os orientaram sobre isso ou que aceitam tal forma inadequada de escrita para aquele contexto.

Por outro lado, Silva (2007) acaba reconhecendo algo positivo: que nunca se escreveu tanto como nesses tempos de correspondências eletrônicas. Mas, alerta, estão “botando os carros na frente dos bois”. Quer dizer, esses adolescentes têm acesso à internet e ao celular, mas não à norma culta da língua escrita. Nas palavras do autor, “os pequenos burgueses tinham internet e celular, mas não dominavam a língua escrita. E por isso criaram a deles. Nada espantoso”. Ao procurar as causas que levaram ao, na sua opinião, desastre da língua cibernética, indica que a escola e a sociedade não se preocupam com educação e cultura tanto quanto deveriam:

A norma culta da língua portuguesa não tem mais quem a defenda nem em legendas de filmes na televisão! A confusão é geral. E a escola deu, por atos, palavras e omissões, grande contribuição ao atual descalabro de que o idioma cibernético é um dos mais óbvios sintomas.

Demos telefones celulares também aos pobres, que podem comprá-los bem baratinhos e em suaves prestações no crediário. Não lhes demos o direito de comprar livros com tamanhas facilidades. Para exemplificar: se os livros fossem alardeados e promovidos como são celulares e computadores, o idioma cibernético não teria lugar.

Concordamos em parte com o autor. O *internetês* não é apenas um sintoma da grave falência educacional, que gera a exclusão dos jovens ao mundo letrado ao qual só poucos têm acesso. Há jovens bem letrados, universitários, alunos de cursos de prestígio (leia-se: rigorosa seleção pelo vestibular) que em ambiente de comunicação pela Internet utilizam o código típico dos jovens. É o que disse Silva há pouco: “criaram a língua deles”, o *internetês*, para ser

usado em ambiente informal, assim como também vai ser empregado nos blogues, diários íntimos, recados no celular, bilhetes entre amigos, etc. Mais ainda: na Internet, os *sites*, em sua grande maioria, utilizam a língua padrão. Portanto, a língua da Internet é apenas utilizada para a interação *on-line*, e os jovens não estão em constante interação, há momentos em que contatam a língua oficial nos *sites* que visitam. Eles precisam dominar duas modalidades da língua, a oficial e a “deles”.

Temos certeza de que os jovens sabem que, num exame ou prova de concurso, certas formas simplificadas que escrevem *on-line* não podem ser empregadas. Eles até tentam utilizá-las em sala de aula (porque lhes parecem mais fácil, mais simples ou estão mais habituados a elas), cabe à Escola esclarecer que “uma coisa é uma coisa, outra coisa é outra coisa”.

Voltemos ao posicionamento radical de Silva, para quem o *internetês* é um besteiro que nem deve ser nomeado. Ele mesmo, no decorrer de sua argumentação, se contradiz ao reconhecer ser algo bem mais do que um besteiro. Chama-o de *idioma* e *linguagem* cibernéticos e de a *língua escrita deles*. Confirma ser algo que estabelece comunicação, faz interações utilizando signos. É principalmente sobre a forma desses signos que ele esbraveja.

Passemos agora à visão de um outro autor. Costa afirma que as novas tecnologias digitais produzem formas de leituras e escrita com características próprias e específicas: “Leitor e autor confundem-se nos (hiper)textos, produzidos/construídos sem fronteiras nítidas, misturando formas, processos e funções da oralidade, da leitura e da escrita” (COSTA, 2005, p. 103). Para esse autor, leitor e autor/escritor utilizam uma linguagem multissemiótica, eles se cruzam, participando da edição do texto que lêem e escrevem *on-line*.

Diz-nos mais: Há mutações do/no ler/escrever no novo espaço em que a escrita é acelerada, fazendo emergir novos gêneros discursivos e textuais. Tais mutações fogem, às vezes, da sucessividade canônica das ferramentas ou dos suportes da escrita tradicionais.

Costa (2005) mostra que esse produto histórico-social que circula num novo suporte eletrônico se constitui num novo gênero também por ser produzido num novo suporte. Ele esclarece que a linguagem eletrônica utilizada pelos jovens é construída, em relação à escrita tradicional, com *ferramentas* (computador, teclado e programas de interface), novo *suporte* (tela) e novos *dispositivos* (meios eletrônicos), um somatório de recursos técnicos que permitem comunicação, organização e construção textual *sui generis*. Em razão disso, o autor afirma haver

surgimento de novos gêneros discursivos e textuais e de novas variedades de linguagem, expressas por uma escrita de plasticidade e heterogeneidade diferentes das tradicionais, e provocam mudanças no ler/escrever. A *leitura/escrita* de (*hiper*)*textos virtuais* de estrutura reticular que circulam na internet pode exigir estratégias (meta)cognitivas diferentes das da leitura/escrita do texto-papel linear (COSTA, 2005, p.104) [grifo do autor].

Com todas as ferramentas e dispositivos, a escrita torna-se fluída, imaterial. O espaço virtual “permite ao usuário modificar o texto a seu bel-prazer: corrigir, cortar e colar, limpar, inserir, editar, formatar, movimentar elementos. Podemos ler e escrever ao mesmo tempo. Somos escreventes e escritores ao mesmo tempo. Escrever é ler. Ler é escrever” (COSTA, 2005, p.106).

O autor ainda diz que muitos pesquisadores<sup>33</sup> de textos da Internet apontam o surgimento de novos gêneros, semelhantes aos já existentes, mas não os mesmos. E que essas diferenças se refletem na linguagem, existindo um consenso de que se trata de uma hibridação entre o oral e o escrito, de uma “escrita oralizada”, de um “falar-escrito” ou de uma “escrita espontânea rápida”. Mais: estamos diante de “falar textos” com uma linguagem-código que produz um tipo de escrita que se equilibra entre o espaço da tela e o tempo usado.

Uma informação importante e que não apareceu em outros textos tão claramente quanto no de Costa (2005) é o da relevância da materialidade do código usado pelos internautas. É exatamente essa *materialidade do código* que caracteriza os textos telemáticos como exemplos de um novo gênero. Só poderá ser classificado como gênero específico – e com isso concordamos totalmente – porque em nenhum outro aparecem ícones, logogramas<sup>34</sup>, *emoticons*, sinais de pontuação, abreviações, alongamentos gráficos, combinações de sinais, uso de maiúsculas, etc. Segundo ele, é a maior reprodução da conversa na escrita que se possa ter. Vejamos, nas cores originais, alguns exemplos retirados de nosso *corpus* de pesquisa que ilustram as afirmações feitas pelo autor:

q vx me Traz Sorte  
 shashahsh, bjuxxxxxx, vixxi  
 escolhe tbm 😊  
 findi =D

<sup>33</sup> O autor não enumera ou identifica esses pesquisadores.

<sup>34</sup> Desenho que corresponde a uma noção ou a uma seqüência fônica, nas escritas ideográficas.



Amizade são como as flores...

```
.....@@@@
.....o00o.....@ @()@ @
.....vVVv..00()00....._...@@@@
.....(____).`0000`..._(*)_.....L.....wWWW
.....~Y~.....|.....( )@( ).....|.....(____)
.....\|.....\|...../.....( ).....\|.....~Y~
.....\\|/.....\\|/.....\|/.....\\|/.....\|/
```

Outro dado trazido por Costa (2005, p.111) é que os “textos escapam aos usuários porque não possuem ancoragem concreta e são abertos a múltiplas interpretações. Nesse universo ciberespacial, virtual, novas formas e novos tipos de espaços são reinventados e adicionados”. Com isso, o usuário necessita desenvolver outras habilidades/competências para ler e escrever, porque se trata de uma escritura interativa em rede. Estamos diante de algo revolucionário para as cabeças acostumadas à lógica linear que, agora, na Internet, passa a ser uma lógica multidimensionada. Ainda nas palavras do autor, “a internet é virtual, múltipla, multimidiática, heterogênea, multifacetada, não-linear, autônoma, desterritorializada, desmaterializada: um ciberespaço, um hiperespaço, uma entidade enunciativa nova” (COSTA, 2005, p. 110).

Cabe aqui fazer uma relação com as idéias de Benveniste sobre enunciação. Diz-nos ele que “é na linguagem e pela linguagem que o homem se constitui como *sujeito*; porque só a linguagem fundamenta na realidade, na *sua* realidade que é a do ser, o conceito de ‘ego’” (BENVENISTE, 1995, p. 286). E a consciência de si mesmo só vai ser experimentada em contraste com um outro, que passa a ser um *tu*. A linguagem, então, só é possível porque cada locutor se apresenta como *sujeito*, sendo ele mesmo um *eu* no seu discurso. Mas esse *eu* propõe outra pessoa, “aquela que, sendo embora exterior a ‘mim’, torna-se o meu eco – ao qual digo *tu* e que me diz *tu*” (p. 286). É isso que acontece na Internet: os jovens na sua apropriação da linguagem se fazem sujeitos, um sendo o *eu* e o *tu* nessa interação recíproca, em que nenhum deles se concebe sem o outro.

Observando tantos adjetivos anteriormente utilizados por Costa (principalmente os iniciados por *multi-* e *des-*: *multidimensionada*, *múltipla*, *multimidiática*, *multifacetada*, *desterritorializada*, *desmaterializada*), caracterizando a Internet como algo complexo, complexa, portanto, é a comunicação que nela se produz. Como resultado, só poderiam surgir novos gêneros textuais. Os textos produzidos *on-line*, ou não, em que podemos “saltitar” de um

a outro acessando *links* diversos são muito diferentes do discurso tradicional linear, pela Internet são construídas superposições de discursos<sup>35</sup>. As ações realizadas em rede são complexas, podemos até dizer labirínticas, resultando em textos bastante diferentes daqueles produzidos em papel.

### 1.2.1 O Miguxês

*Pode parecer cômico (e por vezes é mesmo) para alguns ou trágico (e na maioria das vezes é também) para outros, mas há pelo menos três idiomas usados na Web brasileira de hoje. O miguxês, o geek chic<sup>36</sup> e o português coloquial. Todos apresentam variações que vão do tempero leve ao radicalismo total.*

(Gilberto Pavoni Junior)<sup>37</sup>

A linguagem está sempre mudando e a escrita da Internet é uma adaptação ao mundo moderno. Como as alterações na língua são inevitáveis, até o *internetês* já tem variações. É o *miguxês*, termo proveniente de *miguxa*, variante de "amiga", forma carinhosa usada no tratamento entre duas adolescentes. Como é uma criação recente, não encontramos informações seguras sobre esse assunto na literatura. A explicação fornecida pela Wikipédia nem sequer informa fontes ou referências, mas esclarece que “*miguxês* é o termo utilizado em alguns lugares do Brasil para se referir à forma de escrita comum utilizada por alguns emos<sup>38</sup> em mensagens trocadas pela Internet ou em *blogs*”.

O *miguxês* identifica os participantes de um determinado grupo, adolescentes que criam uma nova linguagem para que os outros não os entendam. Eles substituem letras,

<sup>35</sup> Mais uma vez esclarecemos que *discurso* aqui é empregado como sinônimo de *fala* na dicotomia saussuriana *língua/fala*. Refere-se a qualquer manifestação da língua em ação.

<sup>36</sup> O *geek chic* mescla variados caracteres, letras e números. Funciona melhor no inglês, com o uso de números para formar fonemas: 4=for 2=tw, tu 8=eit, etc (*gr8 minds = great minds*). Exemplo de frases em *geek chic*: Qu3 d v3RDd3 53jd d!td, !55o é o /\d!5 p3rto qu3 3u já chĩ39u3! d3 1 c3r!/\dn!d d3 for/\dturd. (Que a verdade seja dita. Isso é o mais perto que eu já cheguei de uma cerimônia de formatura). Outro nome para *geek chic* é *leet*. Originou-se da pronúncia da palavra *elite* em inglês, no sentido de que o usuário que escreve em *leet* faz parte de uma elite dentro da comunidade. A expressão surgiu nos anos 80 como uma forma de taquigrafia, devido à baixa velocidade dos *modems* da época.

<sup>37</sup> Miguxês x geek chic. Disponível em < <http://www.techboogie.blogspot.com/2007/09/miguxs-x-geek-chic.html-37k> > Acesso em: 20 dez. 2007.

<sup>38</sup> Adolescentes que seguem uma moda caracterizada não somente pela música punk, mas também pelo comportamento geralmente emotivo e tolerante, e pelo visual (em geral, uso de trajes pretos e franjas do cabelo caídas sobre os olhos). São contra a agressividade e a favor da livre expressão de sentimentos, especialmente a tristeza.



## INTRODUÇÃO AO MIGUXÊS 🤪🤪🤪

Aprenda os segredos desse idioma tão culto, a coqueluche do futuro. Seja um poliglota você também. Impressiona as pessoas com sua capacidade e renomada destreza com a língua. Acrescente aquele algo mais no seu currículo. Você também pode .

Ps : LEIA EM VOZ ALTA!

Para quem não sabe "miguxês" é um dialeto internético de gente com meiguice aguda. Criada pelas miguxas (amiguinhas que querem tudo fofo). Enfim : "MiGuXaS Ke AdOrAm sCrEvEr aXxIm! Pq fIcA mTo FOFu!!!"

Miguxês para Iniciantes - Exemplo: "Oiiii, Gentiiii, LinduXo, BonituXo, Xau! BJUXXX..." (cuidado começa assim...).

Miguxês Intermediário - Exemplo: "poix eh exa gentii ignorantii ki nah intendii qq noix fala i xohh reklamaa neh!!"

Miguxês Avançado usa maiúsculas e minúsculas - Exemplo: "Mi IndeNtiFikEi CuM ExXa CoMuNitye!!VuXxeiX XabiaM Ki KiaLaM Uma CoMuNidadEnHa KoNta NoIx"

Miguxês Numérico - Exemplo : "M1gUx3s NuM3r1c4 3 p4r4 m3d10cR3s."

<< PROTESTO >>

xi xe jah tah di xako xeiu di gnti xeim noxaum ki fik implikndu kum xeu jeitu di ixkleve exa komunidadadi eh peifeita pa voxe..ou xeilah xi xe num teim nd kontla keim ixklevi aXim ou ateh axa fofinhu =} xe tbm eh beim vindu aki..puixke euxi iXkLeVu aXim XiM i dAiH?

[\*hahahahhahah....se conseguir entender algo ....]

Falando sério, porra gente tá foda, vamos parar com esse miguxês...isso é irritante,dá vontade de socar a pessoa que está conversando com você, **a língua portuguesa é tão legal escrita corretamente**<sup>41</sup>, e escrevendo assim no miguxês gasta-se mil vezes mais tempo ... PAREM COM O MIGUXÊS... \*hahahaha

[\*hahahahhahah....se conseguir entender algo ....]

Por outro lado, há *sites* valorizando-o, a ponto de ser criado um tradutor para o *miguxês*, indicado da seguinte forma no endereço <http://aurelio.net/web/miguxeitor.html>:

# MiGuXeIToR © ® TM

Tradutor Online de Português para Miguxês/Fofolês

<sup>41</sup> Grifo nosso.

Esse tradutor solicita a digitação de um texto em português e possibilita sua obtenção de três versões. Vejamos o exemplo de um diálogo que o próprio *site* sugere em português, intitulado-o “Paquera moderna”:

<p><b>Português</b></p> <ul style="list-style-type: none"> <li>- Você tem ICQ?</li> <li>- Não, isso é coisa do passado.</li> <li>- Então vou te adicionar no MSN!</li> <li>- Tudo bem.</li> <li>- Daí você me passa o teu Orkut?</li> <li>- O endereço está no meu blog.</li> <li>- Tem muitas fotos no álbum?</li> <li>- Deve ter umas nove mais ou menos.</li> <li>- Beleza, depois teclamos, tchau!</li> <li>- Falou, até mais.</li> </ul>	<p><b>Miguxês arcaico – dialeto ICQ</b></p> <ul style="list-style-type: none"> <li>- vc tem icq??</li> <li>- nao, isso eh coisa do passado.</li> <li>- entao vou te adicionar no msn!!</li> <li>- tudo bem.</li> <li>- daih vc me passa o teu orkut??</li> <li>- o endereco tah no meu blog.</li> <li>- tem mtas fotos no album??</li> <li>- deve ter umas 9 +-.</li> <li>- blz, depois tc, tchau!!</li> <li>- falou, t+</li> </ul>
<p><b>Miguxês moderno Dialeto MSN</b></p> <ul style="list-style-type: none"> <li>- vc tem icq??!?!?</li> <li>- naum...isso eh koiza du passadu.....</li> <li>- entaum vo t addeah nu msn!!!!</li> <li>- tudu bem.....</li> <li>- daih vc me passa u teu orkut??!?!?</li> <li>- u enderessu tah nu meu blog.....</li> <li>- tem mtas foteenhas nu album??!?!?</li> <li>- devi te umas 9 +-.....</li> <li>- blz...dpois tc...xau!!!!</li> <li>- flw...t+.....</li> </ul>	<p><b>Neo-Miguxês - Dialeto Orkut, Fotolog, ...</b></p> <ul style="list-style-type: none"> <li>- VuxXxE TeM icQ??!?!?</li> <li>- NAUM...IXXu EH KOIzaH Du PAXXAdU.....</li> <li>- eNTaUM vow tI adD NU MSn!!!!</li> <li>- TuDu BEM.....</li> <li>- DaIH VuxXxE Me PaXXAH u TEu ORkuT??!?!?</li> <li>- U enDEREXXu TAH Nu mEu blOG.....</li> <li>- tEm mTaxXx FoteEnhAxXx nu album??!?!?</li> <li>- DEvI TE UMaxXx 9 +-.....</li> <li>- bLz...dpoIxXx TC...xXxAu!!!!</li> <li>- Flw...t+.....</li> </ul>

Como podemos ver, a intenção desses adolescentes é escrever num formato que apenas outros adolescentes consigam entender. Mas ao ser criado um tradutor automático, percebemos que normatização e convenções estão presentes, basta seguir a norma de alterar e alternar com certa regularidade as maiúsculas e minúsculas, por exemplo, trocar algumas letras, abreviar palavras e usar símbolos como no *internetês* e teremos mais uma forma de comunicação.

Mesmo com variações do *internetês*, o qual já é uma variação, o *miguxês* mostra-se como “língua” porque proporciona comunicação entre seus usuários. Essa maneira de querer estabelecer diferenças no próprio grupo nos remete ao pensamento de Benveniste, para quem o ato de apropriação da linguagem é utilizado para se fazer sujeito. É o *miguxês*, também chamado *fofolês*, muito individual e coletivo, é a busca de ser único no grupo, por isso, por

exemplo, em vez de pôr duas letras “x” (para representar o fonema /s/) numa palavra, o adolescente põe três “x” para dar seu toque pessoal.

Mesmo sendo forma única e própria de um grupo, o *miguxês* também tem suas regularidades. Passemos agora ao tema da Lingüística de *Corpus*.

### 1.3 LINGÜÍSTICA DE *CORPUS*, empirismo e probabilidade

A Lingüística de *Corpus*, que vai nos auxiliar a descrever o *internetês* tal como utilizado no Orkut, talvez seja, dentro dos Estudos Lingüísticos, a parte ainda menos conhecida ou divulgada. Muitos professores não a conhecem e nem sequer têm idéia das possibilidades que ela oferece para um entendimento mais abrangente do funcionamento da língua ou de como utilizá-la no ensino. Nesta síntese em que a apresentamos, podemos dizer, de maneira bem ampla, que a Lingüística de *Corpus* analisa os padrões de uso real da língua em grandes conjuntos de textos da realidade, observando de modo empírico as formas gramaticais possíveis e prováveis utilizadas pelos falantes reais e não por falantes idealizados. Pode ser definida como a área da Lingüística que se

ocupa da coleta e exploração de *corpora*, ou conjunto de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador (BERBER SARDINHA, 2004, p. 3).

Estudos de estrutura e estudo de uso são as duas grandes áreas em que se dividem os estudos de linguagem, de acordo com Biber (1998). Segundo a perspectiva dos estudos de estrutura, focaliza-se uma característica lingüística, vêem-se as formas em que estruturas similares ocorrem nos vários contextos e como elas servem para diferentes funções. Da perspectiva do uso, em vez de julgar a gramaticalidade, dá-se atenção aos padrões típicos da linguagem, investigando como os falantes usam os recursos da linguagem, indo além da descrição gramatical. São, portanto, duas perspectivas para observar um mesmo objeto, a linguagem. Perspectivas que não se excluem, ao contrário, se complementam, “uma vez que para observar o uso da linguagem é preciso saber em que estrutura o uso está inserido, e para observar a estrutura é preciso ver em que contextos de uso tal estrutura é usada” (AZEREDO,

2007, p. 71). Temos então, claramente, que é nessa perspectiva de estudos de uso da língua que se encontra a Lingüística de *Corpus*.

É grande a importância do empirismo para a Lingüística de *Corpus*. Isso se deve ao fato de muitos teóricos, entre eles Firth, Halliday e Sinclair, defenderem que a observação da linguagem real é a única maneira segura de se descrever uma língua. Sem dados empíricos não há como garantir que um dado enunciado seja realmente utilizado por falantes de uma determinada língua. Essa é a mesma opinião de Biber (1998), para quem a visão da Lingüística de *Corpus* difere da visão tradicional, que punha sua ênfase na estrutura-identificação de unidades estruturais. A nova perspectiva tem seu foco no uso da linguagem, na análise de textos naturais, com os especialistas dando atenção a padrões típicos, em vez de tentar julgar a gramaticalidade. Outro fato destacado pelo autor é a importância dada aos relevantes padrões de associação das palavras, sendo necessário fazer abordagem quantitativa e qualitativa. Com essas associações, o pesquisador focaliza características lingüísticas e também as características de textos ou de variedades de uso.

É interessante destacar, então, que a Lingüística de *Corpus* trabalha dentro de um quadro conceitual formado por uma abordagem empirista e uma visão da linguagem como sistema probabilístico. Sabemos que o empirismo considera ser o conhecimento originado da experiência. Em Lingüística, o empirismo prioriza os dados provenientes da observação da linguagem, geralmente reunidos na forma de um *corpus*. Em oposição ao empirismo está o racionalismo, de acordo com o qual o conhecimento advém de princípios, estabelecidos *a priori*. De um lado temos Halliday, seguidor da tradição empirista, e de outro, Chomsky, o maior expoente do racionalismo na Lingüística.

Como já foi dito, o outro elemento central da conceituação em que a Lingüística de *Corpus* se baseia é a visão probabilística da linguagem. Mais uma vez se evidencia a diferença entre Halliday e Chomsky. O primeiro vê a linguagem como *probabilidade* enquanto o segundo a percebe como *possibilidade*. A visão da linguagem como sistema probalístico pressupõe ser a linguagem padronizada e “embora muitos traços lingüísticos sejam possíveis teoricamente, eles não ocorrem com a mesma frequência” (BERBER SARDINHA, 2000, p. 350). Um exemplo é a frequência de substantivos ser a maior das categorias no nível morfossintático. Assim sendo, a probabilidade de um substantivo existir no texto é maior do

que outra classe gramatical. E constatar que uma linguagem é padronizada só é possível por meio de análises de grandes quantidades de dados empíricos.

Uma metáfora difundida para mostrar a crítica entre os lingüistas de *corpus* e os demais é a do lingüista de poltrona, criada por Charles Fillmore. É a caricatura dos dois tipos de lingüistas. Na visão fillmoriana, o lingüista de *corpus* possui os fatos primários de que necessita na forma de *corpus* imenso de palavras e se dedica a obter fatos secundários de fatos primários. O outro tipo, o “lingüista de poltrona”, é assim descrito por Fillmore apud Berber Sardinha (2000, p. 355):

Ele se senta numa poltrona bem confortável com os olhos fechados e com a cabeça apoiada nas mãos por trás. De vez em quando ele abre os olhos, se mexe todo, berra ‘Nossa, que fato interessante’, pega o lápis e toma algumas notas [...] ficando entusiasmado por ter chegado mais perto de entender como a linguagem funciona.

Embora o estudo e o registro de *corpus* pareçam algo surgido com o computador, na verdade já existiam há muitos séculos, com o mesmo sentido que o conhecemos hoje: *corpus*<sup>42</sup>, *conjunto de documentos sobre determinado tema*, conforme o dicionário Houaiss (2001). Berber Sardinha (2004) informa que já na Grécia Antiga, Alexandre, o Grande, definiu o *Corpus* Helenístico, existindo na Antiguidade e na Idade Média a produção de *corpora* de citações bíblicas. Durante o século XX houve vários pesquisadores descrevendo a linguagem por meio de *corpora*, como Thorndike<sup>43</sup> e lingüistas como Boas<sup>44</sup> e Fries<sup>45</sup>. As diferenças fundamentais entre aquela época e a atual é que os *corpora* não eram obviamente, eletrônicos (eram coletados, mantidos e analisados manualmente) e tinham a ênfase no ensino de línguas. Hoje predomina nos trabalhos a descrição de linguagem e não a pedagogia, embora recentemente haja ressurgido um interesse do emprego de *corpora* em sala de aula.

<sup>42</sup> /korpus/ [lat.] s.m. (1873 cf. DV) **1** coletânea ou conjunto de documentos sobre determinado tema <ele estuda o c. juris canonici (coletânea de direito canônico)> **2** p.ana. repertório ou conjunto da obra científica, técnica e/ou artística de uma pessoa ou a ela atribuída <o c. da poética camoniana> **3** FISL estrutura com características ou funções especiais no corpo de um homem ou de um animal **4** LING conjunto de enunciados numa determinada língua, ger. colhidos de atos reais da fala, que servem como material para análise lingüística **5** LING SEMIO conjunto de enunciados (que são indefinidamente possíveis, i.é., inesgotáveis), constituído por amostras significativas da gramática de determinada língua <o c. sintagmático da língua portuguesa> □ GRAM pl.: *corpòra* □ ETIM emprt. lat. *còrpus* nom. sing. de *corpùs,òris* 'corpo'

<sup>43</sup> Edward Lee Thorndike (1874-1949), psicólogo norte-americano - um dos cientistas que estabeleceu a base do condicionamento instrumental ou operante.

<sup>44</sup> Franz Boas (1858-1942) - antropólogo cultural norte-americano que enfatizou nitidamente a análise lingüística. As modernas técnicas de trabalho de campo em lingüística derivam de seu trabalho.

<sup>45</sup> Charles C. Fries (1887-1967), estruturalista, um dos pioneiros da Lingüística Descritiva do Inglês e precursor da Lingüística de *Corpus*.

Mas o que caracteriza ou constitui um *corpus*? Das definições de vários autores, Berber Sardinha (2000, p. 18) considera a mais completa a de Sanches (1995)<sup>46</sup>:

Um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.

São, portanto, itens importantes a considerar para que um conjunto de textos seja considerado um *corpus*:

- a) **A origem:** deve ser composto por textos autênticos, em linguagem natural. Não devem ter sido produzidos “de encomenda” para serem alvo de pesquisa, não ter sido criados em linguagem artificial, como a de programação de computador ou notação matemática. Devem ser também textos de falantes nativos (quando não for, deve ser qualificado de “*corpora* de aprendizes”);
- b) **O propósito:** deve ter a finalidade de ser objeto de estudo lingüístico;
- c) **A composição:** o conteúdo deve ser criteriosamente escolhido, valorizando-se sobremaneira as condições de naturalidade e autenticidade;
- d) **A formatação:** os dados devem ser legíveis por computador;
- e) **A representatividade:** deve ser representativo de uma língua ou variante, atendendo às perguntas “é representativo de quê?” e “é representativo para quem?”;
- f) **A extensão:** deve ser vasto para ser representativo. São três as dimensões a serem consideradas ao se tratar da extensão do *corpus*. A primeira é o número de palavras, porque quanto maior esse número, maior a chance de o *corpus* conter palavras de baixa frequência. A segunda é o número de textos. Quanto mais textos, maior a garantia de que tal tipo textual, gênero ou registro esteja bem representado. A terceira é o número de gêneros, registros ou tipos textuais que representam uma língua como um todo. Maior número de textos de tipologias diferentes permite maior abrangência do espectro genérico da língua.

As bases teórico-metodológicas da Lingüística de *Corpus* são encontradas em J. R. Firth, considerado um pioneiro na área. Ele não teve acesso aos modernos recursos da

---

<sup>46</sup> SANCHEZ, A. et al. (Orgs.) *CUMBRE: Corpus Lingüístico dl Español Contemporâneo – Fundamentos, Metodología y Aplicacione*. Madrid: SGEL, 1995.

Informática, mas mesmo assim percebeu seu valor para a Lingüística. Observando que há regularidade nos tipos de associações a que se submetem as palavras de uma língua, ele percebeu que o significado de uma palavra se configura pelo contexto de uso, isto é, pelas combinações que ela estabelece no evento comunicativo. Recorrência, freqüência e estabilidade dessas combinações mostram que elas não ocorrem ao acaso. Esse autor enfatizou a importância das relações sintagmáticas e paradigmáticas do léxico, salientando o aspecto sociolingüístico da comunicação.

Assim sendo, os princípios da Lingüística de *Corpus* podem ser sintetizados desta forma:

- 1) A Lingüística é essencialmente uma ciência social e uma ciência aplicada;
- 2) A linguagem deve ser estudada em instâncias de uso atestadas, autênticas e reais, não como sentenças isoladas, inventadas e intuitivas;
- 3) A Lingüística estuda o significado: forma e significado são inseparáveis;
- 4) Não há fronteira entre léxico e gramática: léxico e gramática são interdependentes;
- 5) A maior parte da língua em uso é rotina.

Como vemos, essa área da Lingüística é essencialmente prática, baseando-se no que existe de verdade, recolhendo do uso lingüístico o que é a língua. Deve ter sido por isso que Habert et al. (1997, p. 14) disseram que a Lingüística de *Corpus* se situa na fronteira da Lingüística, nos domínios de especialidade (análise sintática automática, semântica, representação de conhecimento e estatística textual). Segundo esses autores, falta uma cultura lingüística para avaliar e apreender a multiplicidade de trabalhos que podem ser feitos a partir de *corpus*. São trabalhos interdisciplinares para perceber e sentir o aspecto fundamental de certas pesquisas.

Habert et al. lembram que para encontrar rapidamente informações sobre determinada massa de dados, por exemplo, examinam-se documentos que tratam de uma determinada noção. Além de informações, descobrem-se a terminologia, seus empregos e sinônimos (equivalentes). A evolução extremamente rápida de certos setores faz com que só se possa obter um “instantâneo” seguro dos termos em uso pela Lingüística de *Corpus*. Por exemplo, com ela são prontamente detectadas as alterações do vocabulário da navegação, que sofre modificações incessantes.

Sobre a variação lexical, os autores nos dizem que pelas combinações também se pode desambigüizar palavras. Como exemplo, citam *circulação* e *artérias*, que podem ser empregadas para *sangue* e *automóveis*. As palavras em questão serão entendidas levando-se em conta as relações de sinonímia (*circulação/tráfego*), (*artéria/via*) e polissemia (*circulação sanguínea/ circulação de automóveis*). Salientam os autores que tais vocábulos, para indexar um documento, não devem ser utilizados como palavras-chave isoladamente (*circulação*), mas pelo seu sentido naquele texto (*circulação de automóveis*).

Ainda de acordo com esses autores, a exploração semântica do *corpus* é largamente empírica. A significação de uma palavra varia de texto pra texto, sabemos, e para a Lingüística de *Corpus*, a semântica passou de uma concepção lógica a uma concepção distribucional, conforme uma unidade textual possa ser descrita pelo contexto em que figure. O sentido se constrói em contexto mas também pelo contexto, dando um papel central ao *corpus*. Assim, no contexto dos jovens teclando na Internet várias palavras alteraram fortemente seu significado. Temos como exemplos *sinistra*, *irado e bruxo*, que na linguagem jovem pouco ou nada lembram do significado usual registrado pelos dicionários gerais da língua portuguesa. Os conhecimentos projetados pela análise de um *corpus* são, então, altamente reveladores.

Houve a discussão do estatuto da Lingüística de *Corpus*, se metodologia, disciplina, área ou abordagem, com diferentes opiniões. Segundo Berber Sardinha (2000, p. 355-7), a Lingüística de *Corpus* não é uma simples *metodologia*, pois produz conhecimentos novos, e nem uma *disciplina*, pois não possui objeto próprio, servindo aos mais variados objetivos. Para o autor, ela é uma *abordagem*, pois pressupõe princípios filosóficos. O fato de não ser uma *disciplina* faz com que possa ser usada em pesquisa com diversos propósitos.

A proposta metodológica da Lingüística de *Corpus* fundamenta-se num *corpus*, num computador e na análise dos dados pelo pesquisador. Segundo Araujo (2006, p. 107),

o *corpus* é essencial, o computador torna-se imprescindível sob o ponto de vista da coleta, armazenamento e análise de dados, e o pesquisador constitui-se como elemento fundamental por determinar quais são os dados que devem ser coletados e como devem ser organizados para que seja viável o processo de pesquisa.

Por que o computador se torna imprescindível? Porque os *corpora*, sendo tão extensos, jamais poderiam ser analisados manualmente, mesmo que contássemos com uma equipe enorme. Segundo Maciel (2002), “tal equipe gigantesca e poderosa, ainda que existisse, estaria sujeita ao cansaço, à imperfeição e a subjetividade inerentes ao ser humano. Ao contrário, a

máquina, uma vez alimentada inteligente e tecnicamente pelo homem, é sempre a mesma, incansável e consistente em seus achados”.

Assim, depois do que vimos, retornemos a Biber (1998), que nos caracteriza a Lingüística de *Corpus*:

- É empírica, observa os padrões de uso em textos naturais;
- Utiliza uma grande coleção de textos naturais;
- Faz uso extensivo de computador;
- Depende de técnicas de análise quantitativas e qualitativas;
- Tem como objetivo explorar a importância dos resultados quantitativos para aprender sobre os padrões de uso.

Essa, portanto, a Lingüística de *Corpus*, que utilizaremos para entender e descrever o *internetês*. Com ela é possível descrever com maior precisão um uso da língua em extensão, quantitativa e qualitativa, e também a distribuição das formas. Afinal, analisar um *corpus* de 553 mil palavras de *internetês*, sem a ajuda dos princípios da Lingüística de *Corpus* e do ferramental utilizado para o levantamento dos dados, seria uma tarefa impossível de ser feita manualmente, pelo menos dentro dos prazos estabelecidos para a realização desta pesquisa. Acreditamos que a Lingüística de *Corpus* nos possibilita ver de forma clara a configuração do léxico no *corpus* que constituímos. Essa abordagem foi a eleita tendo em vista a descrição de um uso da linguagem de forma extensiva, para indicar resultados quantitativos sobre padrões de escrita em determinado contexto e utilizando um suporte específico, no caso, jovens em comunicação pela Internet escrevendo diante da tela do computador.

## 2 POSICIONAMENTO DO TRABALHO

**N**o longo capítulo anterior, fizemos um extenso panorama de conceitos básicos que nos auxiliarão a analisar e descrever nosso objeto de estudo: o *internetês* representado pela escrita utilizada pelos jovens no Orkut. Muitos autores foram convocados para a tarefa de nos auxiliar a melhor perceber o *internetês*, assim como são vários os assuntos implicados. Nas diferentes opiniões recolhidas, percebemos controvérsias sobre determinados conceitos (na verdade, quase todos são polêmicos). Somos obrigados, agora, a realizar alguns recortes a fim de definir um ângulo para observação de nosso objeto de estudo.

Passemos agora à caracterização dos principais pressupostos adotados em relação ao nosso objeto de estudo, o *internetês*. Logo em seguida, fazemos uma breve discussão sobre cada um. Ao final do capítulo, sintetizamos nossas opções.

O *internetês*, tal como depreendemos da relação com a revisão da literatura:

- 1) é um tipo de escrita;
- 2) essa escrita espelha características de um gênero textual e de comunicação;
- 3) suas “palavras” se identificam formalmente por um espaço em banco antes e depois;
- 4) seus signos incluem uma dimensão de significado e outra de significante, tal qual o signo lingüístico caracterizado por Saussure. Sua característica marcante é ter esse significante peculiarmente grafado;
- 5) tem um léxico;
- 6) possui atipicidade e padronização;
- 7) mostra-se como um dialeto, é um conjunto de marcas lingüísticas restrito a uma comunidade de fala inserida numa comunidade maior de usuários da mesma língua;
- 8) espelha uma relação de poder;
- 9) é uma forma de uso da língua portuguesa;

- 10) usa novas significações e valores em relação ao cânone lingüístico porque a coletividade os consagrou;
- 11) segue a norma entendida como modalidade lingüística normal, comum, em uso entre os internautas;
- 12) é uma variante que ocorre pela não realização da prescrição normativa oficial;
- 13) é uma espécie de sinônimo da própria língua com mesma função semântica.

Buscamos fundamentar nosso trabalho revisando mais de duas dezenas de autores, desde lingüistas do porte de Saussure, Benveniste, Coseriu e Mattoso Câmara até dicionaristas como Aurélio e Houaiss. Até mesmo uma obra de autoria coletiva não identificada, a *Wikipédia*, necessitamos consultar. Isso ocorreu devido a um desdobramento do nosso objeto de estudo, o *miguxês*, algo muito recente sem referência em fontes usuais.

Como o *internetês* é uma forma de os jovens utilizarem a língua, iniciamos refletindo qual concepção de língua é que utilizaremos nesta pesquisa. Assim sendo, Saussure (2004, p. 24) em suas considerações, nos lembra que língua é “um sistema de signos que exprimem idéias”, um sistema de relações ou conjunto de sistemas interligados, com valor nas relações de equivalências ou de oposições que os unem. Também é afirmado pelo fundador da Lingüística que a língua, como acervo lingüístico, é “o conjunto dos hábitos lingüísticos que permitem a uma pessoa compreender e fazer-se compreender” (p. 92). Cabem todas essas afirmações saussurianas quando se considera o *internetês*. É por meio dele, mesmo que alterando graficamente a forma dos signos, que os jovens exprimem suas idéias. Não deixa ele de ser um sistema interligado, com equivalências ou oposições unindo valores existentes nesse sistema. A linguagem dos jovens prova que a língua possui sistema de valores, ela só existe na coletividade. E ela, segundo Saussure (2004, p. 132), “é necessária para estabelecer os valores cuja única razão de ser está no uso e no consenso geral: o indivíduo, por si só, é incapaz de fixar um que seja”. Assim é com a comunicação jovem, novas significações e novos valores em relação ao cânone lingüístico passam a existir porque a coletividade os consagra. E o valor de qualquer unidade está determinado por aquilo que o rodeia, por relações e diferenças com outros elementos. Cremos que nos bastam essas observações de Saussure, não sendo necessário aprofundar suas afirmações relacionadas à língua como sistema abstrato subjacente ao ato da fala ou instituição social que só existe de modo completo na massa (embora essa última também se relacione ao *internetês*, que só existe na massa dos jovens, mas já referida anteriormente ao tratarmos da coletividade).

De Dubois (1978) validamos a idéia de que a língua é um sistema de regras e de relações, com dois meios de comunicação também com sistemas próprios, a língua falada e a língua escrita. Não discutiremos a noção de “instrumento de comunicação” a que o mesmo autor se refere como sendo a língua, uma vez que Benveniste (1995) refuta esse pensamento. Benveniste lembra que falar em instrumento é opor homem e natureza, sendo qualquer instrumento uma fabricação. E a linguagem está na natureza do homem que não a inventou, tendo ela natureza imaterial, funcionamento simbólico, organização articulada, não sendo, portanto, um instrumento que dissocia do homem a propriedade da linguagem. A esse pensamento converge o de Saussure, quando afirmou ser a língua um produto social da faculdade de linguagem e conjunto de convenções, adotados pelo corpo social para o exercício dessa faculdade.

Do mesmo Benveniste, temos a confirmação, mais uma vez, de que a língua é um sistema conforme podemos observar: “Princípio fundamental é de que a língua constitui um sistema do qual todas as partes são unidas por uma relação de solidariedade e de dependência. Esse sistema organiza unidades, que são os signos articulados, que se diferenciam e se delimitam mutuamente” (BENVENISTE, 1995, p. 104).

Podemos estender essa reflexão para o *internetês*. Ele pode ser visto como uma parte da língua, uma forma alterada por um grupo de usuários que a utiliza, em princípio, para comunicação na rede mundial de computadores. O *internetês* é solidário com a língua portuguesa e depende dela, da sua estrutura básica para existir. Ele é a língua modificada em alguns aspectos, mantendo as unidades organizadas, os signos articulados, mesmo que sejam signos registrados de forma, às vezes, bastante diferente daquela consagrada oficialmente.

Essas alterações refletidas na escrita estão relacionadas com a noção de língua em uso da Lingüística de *Corpus* e confirmam Saussure e Benveniste: língua é produto social, conjunto de convenções adotado por determinado corpo social e sistema com partes relacionadas por solidariedade e dependência. A Lingüística de *Corpus* se preocupa com a descoberta de padrões de associação no estudo do uso, entendendo padrão como um conjunto de traços típicos que co-ocorrem. De acordo com Biber (1998), o ser humano não é dotado da capacidade de perceber o que é típico, pelo contrário, ele é equipado para notar aquilo que se destaca, o atípico. A abordagem baseada em *corpus* permite buscar respostas à questão da tipicidade porque faz uso do computador, algo naturalmente programado para detectar

ocorrências e co-ocorrências. O *internetês* chama a atenção justamente por sua atipicidade, mas também a tipicidade ou padronizações interessam à sua descrição.

A polêmica entre defensores e críticos do *internetês* provavelmente ocorra devido à condição de a língua ser utilizada como instrumento de poder. A norma culta da língua reflete o poder de quem domina essa norma, porque bem dominá-la é ter a possibilidade de ascensão social, ocorrendo o contrário, exclusão social, na falta de seu domínio. A crítica ocorre porque o *internetês* prejudicaria a aquisição ou o emprego do português oficial.

Além do poder existente no uso da língua oficial, há o poder que o próprio *internetês* adquire entre os jovens na Internet. Cada vez mais - e até abandonando a faixa etária jovem - formas do *internetês* estão sendo usadas, para acompanhar a rapidez necessária no uso do computador *on-line* ou por outros motivos práticos. Utilizar tal forma de interação é uma maneira de mostrar o poder de inserir-se nesse tipo de comunicação, fugindo de ser um analfabeto digital. Essa expressão (“analfabeto digital”) é marca e fruto de relação de poder advindo com o uso cada vez maior do computador. Como vemos, então, a língua é um instrumento de poder, muito mais na sua forma oficial, mas também nas expressões específicas para determinados ambientes, como é no caso da comunicação via Internet, principalmente entre os jovens. Para eles, estar inserido no grupo, dominar o mundo da comunicação virtual e saber utilizar a sua linguagem lhes concede superioridade, força, possibilidade de fazer coisas, supremacia, enfim, lhes é mais uma forma de exercer e obter poder.

O próximo conceito a ser estabelecido é o de *palavra*, talvez o mais difícil de todos. Todo mundo sabe identificar uma palavra, mas dizer o que é não está claro nem para os mais renomados lingüistas. Não há consenso, apenas convergência para defini-la pelo critério formal, que a estabelece na escrita como algo colocado entre dois espaços em branco. Uma definição insuficiente e incompleta, com certeza. Como entender, então, o que sejam palavras no *internetês*, em que várias letras aparentemente sem significado se juntam ou uma simples letra contém significado? Uma simples letra convencionalizada também contém significado na língua escrita oficial, basta lembrar de “e”, “a” e “o”, o problema, parece-nos, é considerar as novas convenções válidas apenas no contexto da Internet.

Não podemos esquecer que palavra é um conceito fundamental para a comunicação, ela dá sustentação a outras noções construídas sobre e em torno dela, a ponto de a gramática tradicional ter sido construída em torno do eixo *palavra-frase*. Diante da dificuldade de definir palavra, Biderman (1999) sugere a substituição de *palavra* por *lexia*. Seria uma solução, mas também não totalmente, uma vez que as *lexias* podem ser simples (uma única palavra) ou complexas, compostas ou textuais, envolvendo palavras unidas ou não por hífen e até em frases inteiras, como no caso dos provérbios. Usar *lexia* não resolveria, então, a definição de palavra, uma vez que precisaríamos definir aquelas unidades que comporiam, por exemplo, as próprias *lexias* representadas por frases feitas ou provérbios.

Os dicionários gerais da língua, notadamente Houaiss e Aurélio, utilizam como definição o critério formal, estar delimitada por espaços em branco na escrita, mas também acrescentam que palavra é “unidade mínima de som e significado”. Reconhecemos como válida e satisfatória essa idéia, também defendida por Mattoso Câmara (1977). Já as outras concepções mattosianas não servem para a análise do *internetês*, têm seu valor noutras situações. Para o lingüista brasileiro, há ainda formas livres, formas presas e formas dependentes. Tais pormenores, porém, pouco auxiliam para o entendimento do que seja palavra, que tentamos buscar pesquisando em vários autores. O próprio Saussure pouco esclarece, dizendo que, por não poder captar as entidades concretas ou unidades da língua, trabalha com palavras, o que dão delas idéias aproximadas e são concretas.

Diante da falta de uma conceituação definitiva sobre o que seja palavra, para fins deste trabalho, consideraremos como *palavra* toda unidade lingüística mínima que pode constituir significado, delimitada na escrita por dois espaços em branco e/ou sinal de pontuação. Essa definição simples dá conta de classificar como palavras as entidades registradas pelos internautas na sua comunicação. Será considerada como palavra qualquer caracter que contenha significado no *internetês*. É um conceito elástico, porque a interação comunicativa entre internautas se faz por meio de um código escrito muitas vezes cifrado, com símbolos e junção de caracteres aparentemente sem conexão para transmitir idéias. Assim *rsrsrsrsrs* (risos) passa a ser considerada palavra como *snif* (choro) e tantas outras que passam a ser convencionadas e estabelecem comunicação como *Ctai* (Você está aí?), *D+* (demais), *FALOW* (Adeus, até mais), *KD* (Cadê?), *Pru6* (Para vocês) e *T+* (Até mais, tchau). Tais formas têm estatuto de palavra nesse cenário, sendo, em muitos casos, *lexias* complexas ou até uma frase, um enunciado.

Passemos agora a refletir sobre o que é comumente visto como o conjunto de palavras de uma língua, o léxico. É esse também o pensamento de Seabra (2006), para quem o léxico é o responsável por nomear e exprimir o universo de uma sociedade, estando ele em constante expansão, alteração ou contração. Isso podemos observar na comunicação entre jovens internautas, o léxico em suas mãos é alterado, modificado não apenas na forma mas também na significação. Dentre muitos exemplos, lembremos que para eles *irado* não tem nada a ver com *ira*, *bruxo* não se relaciona a bruxaria nem *sinistro* com coisas ruins.

Para Dubois (1978), o léxico possibilita várias oposições conforme o conceito for considerado. Para fins de nossa pesquisa, porém, é irrelevante opor léxico a vocabulário, quando o termo *léxico* é reservado à língua, e *vocabulário*, ao discurso. Também não contribui para nosso objetivo nos prendermos à concepção de que léxico tem como unidades os lexemas, e o discurso tem os vocábulos e as palavras. Para analisar o *internetês* vale isto: entender que o conjunto de palavras (unidades significativas) utilizadas numa língua ou numa de suas variações se chama léxico.

Ainda segundo Dubois, vocabulário é produção realizada e léxico é visto como competência que não precisa ser totalmente concretizada. Ora, não nos interessam as potencialidades que os internautas possuem, interessam as realizações exibidas nos depoimentos e *scraps*, objeto de nossa pesquisa. Sobre léxico é interessante lembrar, ainda de acordo com Dubois (1978), que há potencialidades oferecidas pelas regras e bases léxicas, as quais permitem ao falante compreender inúmeros lexemas, mesmo que nunca os utilizem. Essa compreensão ocorre pela situação em que são empregadas tais bases léxicas. A leitura de textos em *internetês* é um exemplo disso, pessoas pouco familiarizadas a ele conseguem entendê-lo.

É pelo léxico, visto como repertório de palavras de uma língua, que se identificam subjetividade e ideologia de uma comunidade lingüística. Ele não se dissocia da sociedade e da cultura, registra a dinâmica da realidade histórica e cultural de um povo. À medida que surgem mudanças na sociedade, a língua a elas se adapta e produz novas unidades léxicas. É o caso do objeto de estudo desta pesquisa, uma forma peculiar de registrar a língua, adaptada à exigência de novos costumes e tecnologias. Qual componente lingüístico mais se alterou com as mudanças ocorridas nos últimos cinquenta anos? O léxico, claro, com a incorporação (criação ou importação) de novas palavras, com a modificação da sua escrita, embora essa última apenas no ambiente restrito da Internet e inicialmente relacionada à comunicação dos jovens.

Concordamos com Rosengren (1979), que afirma não haver um único léxico na língua, mas um conjunto de léxicos dispostos em torno de um núcleo comum. Essa afirmação faz sentido. Há o léxico comum, dominado por todos os falantes e outros com especificidades de cada grupo. Nesse último temos os vocabulários técnicos e todos os utilizados por distintos grupos, inclusive aquele dos internautas. Cada conjunto de falantes de léxico específico necessita dominar o núcleo comum que, como o nome já diz, tem base abrangente a todos os léxicos. Isso fica claro nos usuários do *internetês*, é necessário dominar antes o português para utilizá-lo modificado na Internet.

Um outro autor, Lepschy (1984) trouxe a idéia de que os aspectos lexical e gramatical são indissociáveis, pensamento corroborado por Saussure (2004). Não é possível utilizar a gramática desligada das palavras, porque para saber uma língua é necessário conhecer bem as palavras, como se pronunciam e escrevem, o que significam, como se alteram e se ligam nas frases, etc. Segundo Lepschy (1984, p. 161), “a distinção que se cria entre léxico e gramática é que o primeiro é o reino da irregularidade, e a segunda o reino da sistematicidade”. Essa afirmação nos leva a uma maior reflexão. De que a gramática tenha sistematicidade ninguém duvida, mas seria o léxico realmente o reino da irregularidade? Não nos parece ser exatamente assim. Mesmo nas observações preliminares do *internetês* percebemos certa sistematicidade. Até na desorganização que aparenta ser, o léxico da linguagem dos internautas jovens não foge ao sistema, há um conjunto de regras repetidas mesmo sendo diferentes das seguidas na forma oficial, especialmente as que se referem à grafia. A que irregularidades Lepschy estaria se referindo? Possivelmente às da liberdade de criação, alteração e importação de palavras. Para contrapor o pensamento desse autor, trazemos Basílio (2004, p. 7), que afirma o contrário: “... o léxico apresenta um alto teor de regularidade e é um componente fundamental da organização lingüística, tanto do ponto de vista semântico e gramatical quanto do ponto de vista textual e estilístico”.

Também concordamos com Baayen (1998), para quem o uso da gramática e do léxico varia conforme gênero, autor, assunto, modalidade (língua falada x escrita), região e registro (variantes estilísticas). Mais ainda, para o autor, gramática e léxico são afetados pelo tipo de texto que é produzido. Confirmamos essas afirmações pela observação do *internetês*, sendo muito mais marcantes tais variações no léxico do que na gramática.

Resta-nos, ainda, registrar como a Lingüística de Corpus percebe o léxico, área a que ela mais dá atenção. Considera como tal o conjunto heterogêneo de unidades, incluindo nomes próprios, siglas, abreviaturas. E separa itens lexicais em formas (types) e ocorrências (tokens), em palavras *lexicais*, plenas, de conteúdo e *gramaticais*, funcionais. Também valoriza muito a frequência de uso de uma palavra, sendo visto como um atributo dela tão inseparável quanto a significação que encerra.

E assim como para Lepschy e Saussure, para a Lingüística de *Corpus* não há separação entre o léxico e a gramática, a ponto de utilizar a expressão léxico-gramática para se referir a essa união, dizendo que as escolhas léxico-gramaticais são feitas em cadeia e não isoladamente, e que as diferenças entre padrões colocacionais indicam diferenças de sentido.

Um outro assunto que em tudo se relaciona com nossa pesquisa sobre o *internetês* é a norma lingüística, sobre a qual nos posicionaremos agora. Na sociedade há padrões culturais advindos de formas comportamentais generalizados, uns são padrões ideais, esperados das pessoas, e outros são padrões reais, efetivamente realizados pelas pessoas. Com a língua isso também ocorre, havendo a língua padrão, modelo idealizado e aceitável em todas as situações, e as outras todas, inclusive a da Internet. Não esqueçamos, porém, que a norma é fator de coesão social, como anteriormente dito, aceitável em todas as ocasiões. As formas que fogem à norma podem ser causa de várias exclusões, inclusive de exclusão social. E assim como em todos os demais padrões impostos na sociedade, o padrão lingüístico considerado é o da classe social de prestígio, a que detém cultura e poderes (Faraco, 2002). Aqui podemos lembrar do anteriormente afirmado por Dacanal (2006): a língua é um instrumento de poder. A norma que seguimos não foi estabelecida pela maioria dos falantes, mas por aqueles que detêm mais poder, mesmo que sejam minoria. E a língua também serve para exercer o poder, tanto que o mesmo Dacanal (2006) salientou não ser por acaso que a elite ateniense do século V a.C. e a classe dirigente de Roma davam tanta importância à retórica, ou oratória, a arte de fazer discursos e de convencer. É a língua o veículo pelo qual circulam a informação e o conhecimento, sendo também a forma mais simples de mostrar que os possui.

Há dois entendimentos para o que seja norma. Primeiramente vista como modalidade lingüística “normal”, “comum”, uso, e depois, como uso regrado, modalidade “sabida” por uns mas não por outros, mais prestigiada, “bom uso”. De acordo com Coseriu (1979), nessa segunda concepção, a norma são imposições obrigatórias e consagradas social e culturalmente para o uso de determinadas possibilidades em detrimento de outras. É essa última significação

que nos interessa mais, porque poderemos observar que transgressões a ela os internautas realizam na sua comunicação.

No dia-a-dia, percebemos uma norma escrita, mais conservadora, e uma oral, que aceita inovações. Além dessas variações de modalidade, há outros tipos que abordaremos a seguir. O sistema lingüístico está a serviço de uma comunidade heterogênea e plural, resultando, então, por refletir em si tais características, a de ser heterogêneo e plural. As línguas possuem dinamismo inerente, com variações e mudanças, apresentando formas lingüísticas alternativas com mesmo valor e função. Mesmo havendo inovações e modificações, as línguas se mantêm coesas. Percebemos aqui um movimento curioso: ao mesmo tempo em que há o impulso à variação, há um outro, voltado à convergência, fundamentando a noção de comunidade lingüística, com seus padrões estruturais e lingüísticos (as normas). A heterogeneidade e a homogeneidade lingüística não são aleatórias, mas reguladas por um conjunto de regras.

As variações, pelo que vimos, têm grau de estabilidade e de mutabilidade. Mollica (2003) faz considerações sobre essas variações da língua. Com ele concordamos, quando nos lembra que:

- Variantes ocorrem pela realização ou não da prescrição normativa;
- Formas variantes podem permanecer por tempo estável, mudar ou desaparecer;
- Variações são originadas por variáveis internas (fatores de natureza fono-morfo-sintática, semântico, discursivo e lexicais) e por variáveis externas (etnia, sexo, escolarização, classe social, tensão discursiva);
- Variações ocorrem por regiões (eixo diatópico) ou por estratos sociais (eixo diastrático).

Pelo que aparentemente se pode perceber na observação da linguagem dos jovens, é a variação diastrática a que mais caracteriza o *internetês*. O que mais aparece são variações típicas da faixa etária ou características de usuários da Internet, sendo o mesmo código utilizado praticamente em todo o país. Essas e outras variações acontecem porque, segundo Coseriu (1979), as línguas mudam porque se fazem continuamente, não estão prontas. Elas existem no falar, e não podem ser isolada dos fatores externos, daí decorrendo adaptações e formas alternativas de uma mesma língua. É como Possenti (1986) se refere às variantes, como espécie de sinônimas da própria língua, basicamente com a mesma função semântica. E nos

lembra que é necessário um termo de comparação fixo (afinal, uma variante é uma variante de quê?), sendo o mais usual comparar a variante com a prescrição da norma-padrão.

Uma espécie de variação especial da língua, mais complexa, é o chamado *dialeto*. Também esse conceito não tem uma definição unânime entre lingüistas e dicionaristas. O que aparece como traço comum em praticamente todos os autores é a afirmação de ser algo diatópico, prática lingüística restrita a algumas regiões. Dubois (1978), por exemplo, afirma ser a forma de língua com sistema léxico, sintático e fonético usado num ambiente restrito, podendo ser classificado como gíria quando sistema criado ou empregado como conjunto secreto de signos. O mesmo autor faz distinção entre dialeto regional e dialeto social. O que nos interessa neste trabalho é ver se o *internetês* pode ser visto como um dialeto social que “é sistema de signos e de regras sintáticas usado num dado grupo social ou em referência a esse grupo” (DUBOIS, 1978, p. 184). Para Houaiss, a definição abrange as duas modalidades, tanto que faz a indicação disso entre colchetes no final da primeira aceção indicada:

conjunto de marcas lingüísticas de natureza semântico-lexical, morfossintática e fonético-morfológica, restrito a dada comunidade de fala inserida numa comunidade maior de usuários da mesma língua, que não chegam a impedir a intercomunicação da comunidade maior com a menor [O dialeto pode ser *geográfico* ou *social*]. cf. *registro, jargão, patoá, gíria* (HOUAISS, 2001).

Opinião semelhante tem Ferreira (2001, p. 676), acrescentando também a idéia de que pode ser uma “variedade subpadrão ou não-padrão de uma língua associada a grupos que não contam com prestígio social [Sin.: *linguajar* (2)]”. Seria, então, o *internetês*, um dialeto? Seria ele uma variedade subpadrão? E seus usuários não contariam com prestígio social? Respondamos, então. O *internetês* se coloca como um dialeto porque apresenta os elementos/condições registrados por Dubois, Houaiss, Ferreira e, como veremos a seguir, também por Mattoso Câmara: forma de língua com sistema léxico, sintático e fonético usado em ambiente restrito (geográfico e social), inserido numa comunidade maior de usuários dessa língua. Sob o ponto de vista relacionado a preconceito, seria sim o *internetês* uma variedade subpadrão, principalmente considerada pelos puristas que o vêem como uma deturpação da língua, um linguajar menor, como o usado por grupos sem prestígio social. Não seria esse, porém, o caso do internautas. O dialeto empregado na Internet é uma forma de comunicação daquele grupo de usuários, que a utilizaria basicamente na comunicação via rede mundial de computadores e não nas ocasiões em que a norma culta deve ser empregada, nos registros de qualquer espécie de escrita a ser feita conforme prescrição da gramática normativa.

Mattoso Câmara (1984, p. 141) vê o dialeto sob ponto de vista puramente lingüístico: “os dialetos são falares regionais que apresentam entre si coincidência de traços lingüísticos fundamentais”, mas também sob ponto de vista extralingüístico, em que: a) existe um sentimento lingüístico comum em que os dialetos são sentidos como variantes de uma língua geral e ideal; b) existe uma língua culta, superposta aos dialetos que ficam limitados ao uso cotidiano, sem maior expressão literária; c) existe subordinação política das regiões como partes de um estado político nacional. Das afirmações de Mattoso Câmara sob o ponto de vista extralingüístico, inicialmente pensamos não aplicar a última delas, todas as demais estão claramente relacionadas com nosso objeto de pesquisa. Faremos uma verificação por região analisando subcorpora de várias partes do Brasil para constatar se não haveria subdialetos do *internetês*.

É discutível se devemos levar em consideração para definir de dialeto o chamado critério da intercompreensão, proposto por Coseriu (1982), segundo o qual dois falares podem ser considerados dialetos da mesma língua se seus falantes conseguirem compreender-se mutuamente. Todos os falantes do português conseguem entender *internetês*? Não. Por essa definição o *internetês* não é um dialeto, porque um falante na faixa dos 50 anos de idade não familiarizado nada compreende da escrita modificada dos jovens no computador. Quanto às outras opiniões do mesmo autor, podemos concordar plenamente, principalmente quando diz que o dialeto é uma língua menor incluída em uma língua maior. De acordo com ele, uma língua, em princípio, “é uma família histórica de modos de falar afins e interdependentes, e os dialetos são membros desta família ou constituem famílias menores dentro da família maior” (COSERIU, 1982, p. 11-2).

Como vimos, os dialetos são variações da língua realizadas de forma comunitária. Veremos agora as variações produzidas pelo falante, de forma individual. É o chamado *idioleto*, no dizer de Houaiss (2001), o sistema lingüístico, num determinado período da vida de um indivíduo, que reflete características pessoais, estímulos recebidos, sua biografia, enfim. A mesma idéia é confirmada por Mattoso Câmara, que explica manifestar-se pela escolha de palavras, frases ou metáforas únicas de cada indivíduo, num arranjo muito particular de cada um. Dubois (1978), por sua vez, acrescenta que é o estilo de se comunicar indicado pelo conjunto de enunciados produzidos por um indivíduo. E partindo da idéia de que o ato lingüístico é ao mesmo tempo social e individual, Coseriu (1979a) lembra que a fala é o falar em geral, mas também cada ato particular de falar. Deduzimos que esteja se referindo ao *idioleto* de cada um.

Como podemos perceber, os autores convergem para o mesmo ponto: é variação da língua única de um falante. Até onde podemos dizer que no *internetês* há muito do idioleto de cada internauta? Quanto daquela forma de comunicação é coletiva e quanto é peculiar, exclusiva do jovem que a utiliza? Um internauta que comete muitos erros de digitação reflete dessa forma seu idioleto, registra, mesmo sem querer, seu estilo característico de se comunicar. E isso acontece muitas vezes no *internetês*, na pressa de enviar a mensagem, principalmente na comunicação *on-line*, muitos “atropelos” à grafia são cometidos. Como é usual esse fato, e com a ajuda do contexto, o interlocutor consegue decodificar a mensagem enviada. A observação do *miguxês* leva-nos a pensar que ele tenha surgido com fortes características de idioleto, registrando as formas particulares dos componentes de um determinado grupo de usuários, cada um querendo marcar-se como sujeito pelos seus hábitos verbais gráficos únicos.

Como o *internetês* é basicamente um registro escrito, faz-se necessária uma posição sobre a escrita. De acordo com Ferreiro e Teberosky (1999), ela é a maneira de transcrever a linguagem e também um objeto simbólico, um substituto (significante) que representa algo, assim como o desenho. Só que a escrita não tem semelhança com o objeto (ou acontecimento) referido e constitui, como a linguagem, um sistema de regras próprias. A escrita constitui um tipo específico de objeto substituto. São as regras desse objeto substituto que os adolescentes quebram ou simplificam ao escrever *on-line*.

Um aspecto interessante a observar é que de maneira geral, e esse pensamento é corroborado por Sampson (1996), a escrita é vista como um aspecto de tecnologia, algo que as pessoas usam, não fazendo parte de suas personalidades. Para ele, a escrita é “um sistema para se representar enunciados da língua falada por meio de marcas permanentes e visíveis” (SAMPSON, 1996, p. 25). É isso, provavelmente, que vem ocorrendo desde a invenção da escrita que, segundo os estudiosos, ocorreu por volta de 4000 a.C, quando os sumérios desenvolveram a forma cuneiforme de escrever. Temos confirmada a afirmação de Sampson por haver a presença até hoje de placas de argila com as marcas permanentes e visíveis dos sumérios. Claro que as “placas” mudaram ao longo do tempo, sendo bem diferentes as utilizadas atualmente pelos internautas. Por ser algo externo ao ser humano, a escrita pode ser modificada, afinal, ela é apenas uma maneira de representar graficamente o significado. Claro que essas modificações devem fazer parte de uma convenção, senão, por ser simbólico, modificando aleatoriamente a forma de representação perdemos a relação com o objeto representado, uma vez que essa relação é imotivada. É a isso que Ferreiro e Teberosky (1999) se referiam como um sistema de regras próprias. A relação língua e norma (regras) é mais

considerada na língua escrita, até pelo uso generalizado das gramáticas normativas, as quais tratam as relações entre fala e escrita tendo como parâmetro a língua escrita. Isso passa para o usuário a impressão de que fala é o lugar do erro, sem regras, enquanto a escrita é formal, com regras.

Mattoso Câmara também se posiciona favorável à concepção de escrita como representação visível e durável da linguagem, tornando-a, de falada e ouvida, a escrita e lida. Ele alerta para a diversidade existente entre fala e escrita, afirmando que a escrita não é reprodução fiel da fala, não sendo verdadeira a metáfora de que “ela é a roupagem da língua oral” (MATTOSO CÂMARA, 1979, p.9). Saussure (2004, p. 40) se refere à mesma imagem, sustentando que “a escrita obscurece a visão da língua; não é um traje, mas um disfarce”. De acordo com os lingüistas, há uma diferença fundamental entre as duas modalidades da linguagem, a falada e a escrita. No *internetês*, porém, parece-nos não mais existir tal diferença ou, pelo menos, ter sido bastante minimizada.

Sobre a escrita, Saussure (2004) diz mais, referindo-se a incoerências dela, como as várias letras para representar um som e a existência de ortografias flutuantes. Mas ele também reconhece o prestígio da escrita afirmando que a imagem gráfica impressiona como objeto permanente e sólido; as impressões visuais são mais nítidas e duradouras que as impressões acústicas; a língua literária valoriza a escrita, a qual aparece como um código regulamentado, e a escrita apresenta solução se houver desacordo entre língua e ortografia. E o que tem isso a ver com o *internetês*? Aparentemente nada, uma vez que a língua da Internet pouco segue as regras ortográficas e não respeita muito as regras de pontuação, marcas típicas da escrita. Mas é pela escrita que se dá a comunicação no computador, mesmo quando usada para substituir a fala. O caso das incoerências ortográficas referidas por Saussure a escrita na Internet procura solucionar simplificando ao máximo, utilizando equivalências entre letras e fonemas, abreviaturas, códigos e símbolos. Nesse aspecto, a escrita parece contrariar o anteriormente afirmado pelos lingüistas, sendo o *internetês*, sim, uma roupagem ou representação muito próxima da língua oral. Nos depoimentos do Orkut, a imagem gráfica tende a ser permanente como Saussure afirma, uma vez que é um registro para ficar exposto por muito tempo. Quanto à produção literária, por ser o *internetês* um fenômeno recente na língua, ainda não temos conhecimento de obras produzidas nessa modalidade de escrita. Certamente, se existirem, serão registradas nesse código, provavelmente, narrativas semelhantes às histórias contadas oralmente, como os velhos causos gauchescos ou outras histórias orais.

Fávero (1994) informa que escrita realiza interação à distância, tem planejamento anterior à produção, tem possibilidade de ser revisada, tem reformulações promovidas apenas pelo escritor e não tem possibilidade de *feed-back* imediato. Com exceção da interação à distância, nenhuma das afirmações da autora se aplica à escrita na Internet. Escrever digitalmente *on-line* pode alterar até as concepções sobre escrita, porque suas condições de produção foram alteradas pela tecnologia computacional. Há, sim, possibilidade de revisão, mas numa condição diferente da escrita tradicional, em que se usava rascunho, o qual poderia ser riscado e rabiscado, substituindo-se o que não deveria constar no texto final. No *internetês*, após dar o “enter”, após enviar a mensagem, a revisão só poderá ser feita por acréscimo, por um novo texto, uma explicação tentando “consertar” o anteriormente escrito.

Independentemente de onde se escreva, concordamos com o que Endruweit (2006) afirma sobre a escrita, tendo aplicação no *internetês*:

- É instrumento para comunicar através de textos, o que reforça a concepção de escrita apenas como representação.

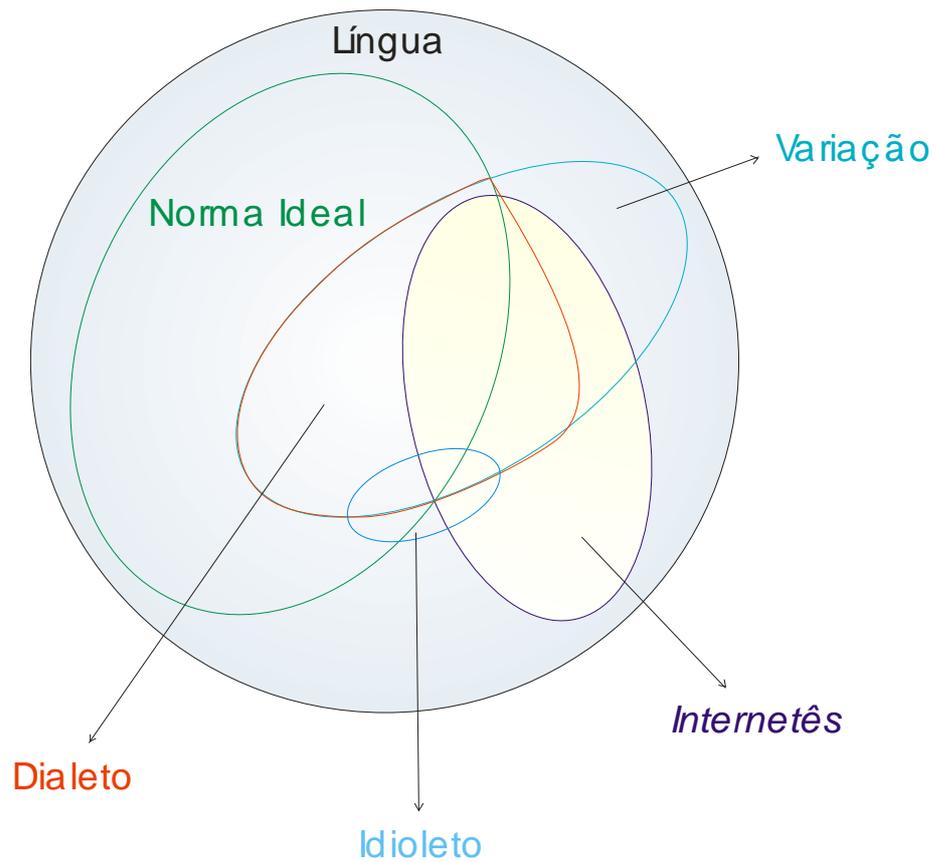
- É uma face da língua.
- É significante de significante primeiro. O primeiro significante é a fala.
- É presença-ausência, recurso artificial, torna presente a fala sem a voz.
- É um sistema de signos.
- A escrita dá idéia de fidelidade, substitui a presença do homem.
- Tem por função armazenar (a cultura). É reflexo de uma cultura.

O que escrevemos neste capítulo mostra os conceitos nos quais nos apoiaremos para entender esse fenômeno da escrita dos jovens no Orkut. Em resumo, os entendimentos que guiam nossa metodologia de observação do *corpus* são os seguintes:

- 1) Escrita é uma representação da linguagem, um sistema de signos convencionados com regras próprias, um substituto que representa algo. Realiza interação à distância, numa presença-ausência ao tornar presente a fala sem a voz.
- 2) Língua é um sistema interligado de signos, com equivalências ou oposições, unindo valores existentes nesse sistema. Além disso, é também um sistema probalístico de combinatórias, que inclui aspecto quantitativo e estatístico. É um instrumento de poder.

- 3) Palavra, por não haver consenso sobre sua definição, é entendida formalmente como toda unidade mínima com significado entre espaços em branco e/ou sinal de pontuação. Muitas vezes tais unidades correspondem a lexias complexas e até a uma frase.
- 4) Léxico é visto como o conjunto de palavras (unidades significativas) de uma língua ou de uma de suas variações. Está em constante expansão e alteração, sofrendo modificações tanto na forma como na sua significação. Há junção entre léxico e gramática. Na comunicação as escolhas léxico-gramaticais são feitas em cadeia, havendo diferenças de significação conforme os padrões colocacionais das palavras.
- 5) Norma é vista de duas maneiras: como modalidade de uso comum, normal e como uso regrado, imposição de forma consagrada de prestígio, “bom uso”.
- 6) Variações lingüísticas obedecem ou não à prescrição da norma (“bom uso”), originam-se por variáveis internas (gramaticais) e externas (sociais) e ocorrem por regiões ou estratos sociais. É uma outra forma de utilizar a própria língua com mesma função semântica.
- 7) Dialeto é uma variação mais complexa da língua, usada em ambiente restrito, inserido numa comunidade maior de usuários da mesma língua. Pode ser diatópico (por regiões) ou diastrático (por camadas sociais), nesse caso também chamado de socioleto.
- 8) Idioleto é a variação produzida pelo falante, expressa na escolha de palavras e frases únicas de cada um. É o estilo de se comunicar de um indivíduo.

A figura a seguir visa sintetizar esses posicionamentos.



**Figura 4: Representação da heterogeneidade da composição do *internetês***

Após o exame dos nossos *corpora*, voltaremos a discutir esses posicionamentos. Passemos, agora, à etapa seguinte, em que vamos detalhar o modo como a descrição será conduzida.

### 3 MATERIAIS E MÉTODOS, O QUE E COMO

**N**osso *corpus* de estudo, como já dissemos, é composto por textos de depoimentos (*testimonials*) e de recados (*scraps*) coletados no *site* de relacionamentos Orkut. Nossa metodologia básica de estudo é a comparação entre esse objeto e um conjunto de outros *corpora* (escrita culta, fala, textos jornalísticos sobre temas de ciências, textos didáticos universitários e textos de redações escolares). Como se verá mais adiante, prestamos maior atenção às palavras mais freqüentes e também à riqueza vocabular, entendida como: a) relação entre o número de palavras (*tokens*) e a variedade de formas (*types*) em um dado texto ou corpus, e b) a percentagem de palavras de único emprego (*hapax legomena*) num texto.

A observação da freqüência das palavras mais empregadas em um dado tipo de uso da língua é algo que, no nosso trabalho, teve muita importância. Afinal, a percepção das diferentes freqüências das palavras empregadas nos diferentes usos da língua portuguesa possibilita um entendimento da língua como um todo multifacetado. Desenha-se, assim, uma integração de subsistemas que estão em um *continuum* de usos. Pela análise da repetição de emprego das palavras por vários usuários, podemos avaliar a probabilidade de ocorrência de determinados traços ou de determinadas estruturas da língua nessa nova situação de uso. Essa análise, ainda que limitada às formas mais e menos presentes, pode revelar toda uma amplitude de usos, indicar o compartilhamento maior ou menor dos itens do léxico da chamada “norma culta”. Além disso, permite revelar movimentos de adesão ou de não aceitação para determinadas novas grafias.

Como o *internetês* é basicamente um tipo de escrita, optamos por um levantamento das principais modificações relacionadas à grafia oficial e a outros usos decorrentes. Para isso, serão utilizadas as ferramentas *WordList* e *Concord* do *WordSmith Tools* para examinar detidamente a grafia das duas mil palavras mais freqüentes do nosso *corpus* de estudo. Também indicaremos se há indícios de oralização sobre a escrita no *Corpus Geral do Orkut*.

Partindo da lista das palavras mais freqüentes nessa escrita, também faremos o levantamento das variedades de forma e conteúdo, quer sejam diferenças morfofonêmicas de conteúdo distinto, embora de mesmo lema (como *pude/pode*), quer sejam apenas variações de forma para um mesmo significado (*vc/você*). Podemos dizer que uma das características do *internetês* é a utilização de mais de uma forma para registrar uma mesma palavra.

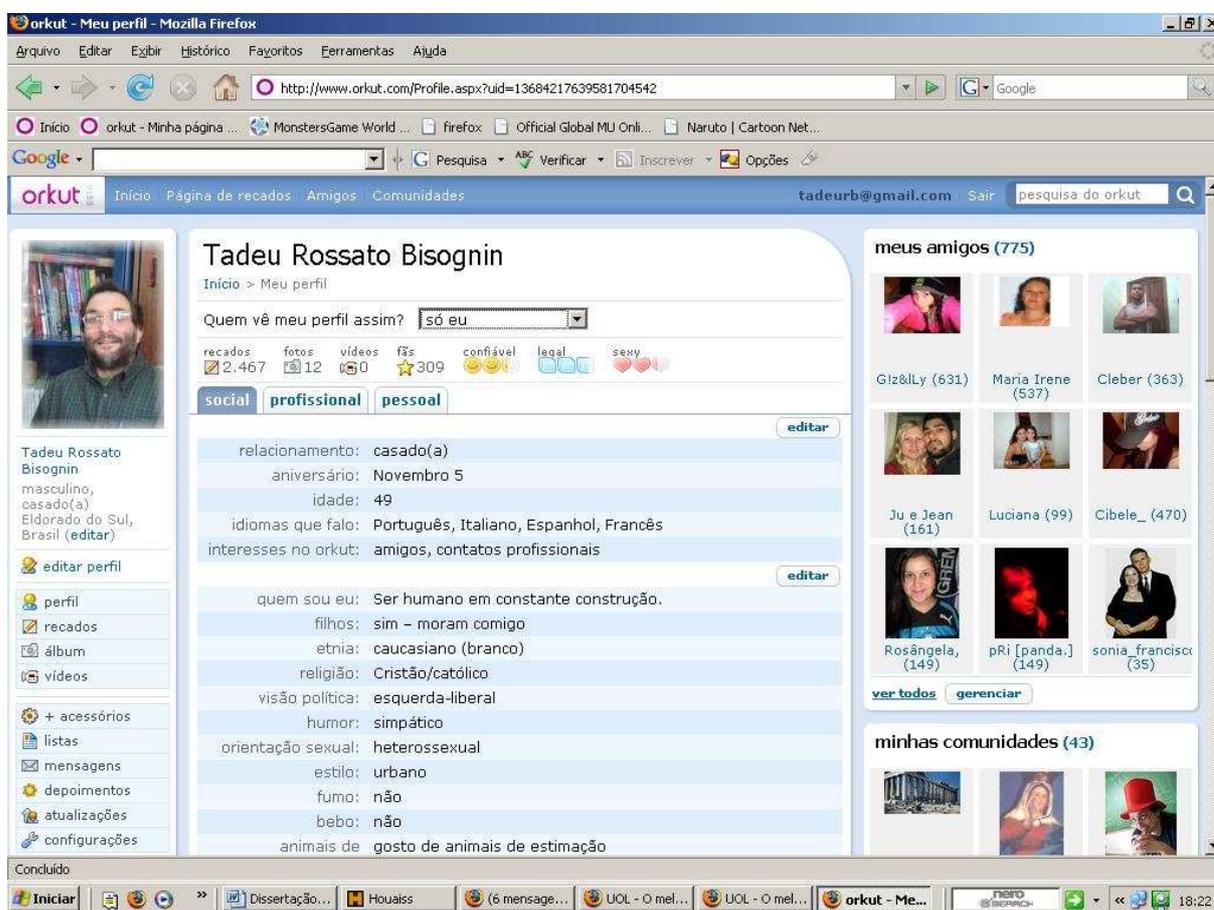
Com a comparação entre as palavras mais freqüentes do *corpus* de estudo retirado do Orkut com as dos *corpora* de referência (português escrito, português falado e VOLP – Vocabulário Ortográfico da Língua Portuguesa), planejamos poder perceber as coincidências e divergências, principalmente com vistas a obter as respostas para nossas questões de pesquisa e confirmar nossas hipóteses. O contraponto com os *corpora* para contraste (redações escolares, textos didáticos de Química e textos jornalísticos) tem o mesmo objetivo. Além disso, utilizaremos dados de estudos sobre riqueza vocabular e repetição de palavras em um conjunto reportagens premiadas sobre biodiversidade, contrastando a riqueza lexical do *internetês* e dos textos dos demais *corpora*, para buscar descrever o vocabulário da escrita no Orkut.

### 3.1 O TERRITÓRIO DO *CORPUS* DE ESTUDO

O campo de pesquisa deste trabalho sobre o emprego de palavras em situação de escrita na Internet centra-se nos depoimentos escritos sobre amigos nas páginas do Orkut e nos recados deixados e/ou trocados entre eles. O Orkut é uma comunidade virtual afiliada ao Google, criada em 22 de janeiro de 2004 com o objetivo de ajudar seus membros a criar novas amizades, relacionar amigos (não necessariamente no sentido literal da palavra, mas no aspecto de se relacionar com pessoas com quem se tenha alguma(s) afinidade(s)) e participar de comunidades sobre os mais diversos assuntos. Seu nome é originado no projetista chefe, Orkut Büyükkökten, engenheiro do Google. Tais sistemas, como esse adotado pelo projetista, também são chamados de rede social.

O Orkut se compõe de três partes principais: *perfil*, *meus amigos* e *minhas comunidades*. É junto ao *perfil* que se localizam os depoimentos, havendo nele um *link* para os *scraps*. São os textos retirados dos depoimentos e dos recados do Orkut que servirão de objeto de estudo neste trabalho. Essa opção se justifica pela importância dos depoimentos de amigos sobre uma pessoa. É por meio deles que podemos conhecê-la melhor, completar o perfil que ela quer mostrar aos que a visitarem virtualmente no *site*. Lembramos também que os

depoimentos só são visíveis com a autorização de quem os recebe, assim como os *scraps*, que podem ser deletados pelo receptor (e também pelo emissor) deles.



**Figura 5:** Exemplo de página com perfil de usuário do Orkut, junto ao qual estão os tópicos *meus amigos* e *minhas comunidades*.

### 3.2 CONSTITUIÇÃO DOS CORPORA

Como dissemos há pouco, as unidades-objeto de nossa investigação foram extraídas de um *corpus* de *depoimentos* (*testimonials*) e de *recados* (*scraps*). Os recados são mensagens, na maioria das vezes curtas, trocadas entre internautas como se fossem e-mails. Os depoimentos coletados são em maior número devido à facilidade da coleta, por conterem mais palavras, caracterizando o que classicamente se tem como texto (tecido, com coesão). Os recados, em menor número, são formados pelas mais diferentes mensagens, sendo empregado todo tipo de palavra e até figuras obtidas pela organização de caracteres, como nos exemplos abaixo.



Os recados do *subcorpus* foram coletados de várias cidades do Brasil, em maior número nas capitais onde também se coletaram outros *subcorpora* compostos por recados e depoimentos. Esclarecendo melhor: O *corpus* para esta pesquisa é constituído de um total de 553.875 palavras (*tokens*), entre elas sendo encontradas as palavras de dez *subcorpora*, cada um deles com mais de trinta mil palavras. A formação do *corpus* e dos *subcorpora* ocorreu da seguinte maneira:

- O início da recolha da linguagem utilizada no Orkut pelos jovens se deu aleatoriamente: procuramos no Orkut jovens de 15 a 23 anos de variados lugares do Brasil.
- Observamos alguns recados recebidos e os depoimentos sobre eles e com os recursos do computador, copiamos tais textos para um arquivo em formato “.doc”.
- Percebemos que em alguns lugares havia algumas palavras escritas de forma peculiar, que pouco se repetiam em outras partes do país. Então, passamos a formar outros arquivos por regiões, para investigar quantos dados seriam particularidades regionais. Formamos, então, nove *subcorpora*: Belém, Brasília, Cuiabá, Manaus, Porto Alegre, Recife, Rio Branco, Rio de Janeiro e Salvador.
- Para facilitar a escolha de usuários jovens da rede mundial de amigos nas diferentes partes do território brasileiro, procuramos comunidades ligadas a escolas ou a organizações tradicionalmente relacionadas à educação.
- Procuramos observar não mais apenas a idade, mas também se era habitante da região de nosso interesse para ver posteriormente se sua comunicação continha alguma peculiaridade diatópica. Para ilustrar, podemos citar como local de busca as comunidades do Orkut *Maristas Salvador*, *Salesianos Cuiabá*, *La Salle Belém*, *Jesuítas Rio de Janeiro*, *Escolas Brasília*, *Rio Branco*, *ACM Amazonas* e *Colégio Recife*. Normalmente os participantes dessas comunidades estavam na faixa etária de interesse para observação de sua linguagem, objeto desta pesquisa.

Após copiar os depoimentos e *scraps*, foi necessário excluir nomes, fotos e datas, o que foi feito. Para que os *corpora* pudessem ser lidos pelo software *WordSmithTool*, foi necessário que cada *corpus* fosse convertido para o formato “.txt”. Depois de prontos, os *corpora* foram salvos com novos nomes. Assim, os *corpora* foram nomeados em ordem crescente, a eles sendo acrescentados também os *corpora* de contraste e de referência. Ficaram estabelecidos, então, os seguintes *corpora*:

**CORPORA DE ESTUDO**

<b>Nome1</b>	<b>Nome2</b>	<b>Conteúdo</b>
corp1.txt	Geral do Orkut	Recados e <i>scraps</i> de todo o Brasil
corp2.txt	Manaus	Recados e <i>scraps</i> de Manaus e região
corp3.txt	Belém	Recados e <i>scraps</i> de Belém e região
corp4.txt	Brasília	Recados e <i>scraps</i> de Brasília e região
corp5.txt	Cuiabá	Recados e <i>scraps</i> de Cuiabá e região
corp6.txt	Porto Alegre	Recados e <i>scraps</i> de Porto Alegre e região
corp7.txt	Recife	Recados e <i>scraps</i> de Recife e região
corp8.txt	Rio Branco	Recados e <i>scraps</i> de Rio Branco e região
corp9.txt	Rio de Janeiro	Recados e <i>scraps</i> do Rio de Janeiro e região
corp10.txt	Salvador	Recados e <i>scraps</i> de Salvador e região
corp11.txt	<i>Scraps</i>	<i>Scraps</i> de todo o Brasil
corp12.txt	<i>Scraps reduzido</i>	<i>Scraps</i> de todo o Brasil (apenas 2.371 tokens)

**CORPORA DE CONTRASTE**

<b>Nome1</b>	<b>Nome2</b>	<b>Conteúdo</b>
corp13.txt	Redações <sup>7<sup>a</sup></sup>	Textos produzidos por alunos de 7 <sup>a</sup> série
corp14.txt	Redações <sup>3<sup>o</sup></sup>	Textos produzidos por alunos do 3 <sup>o</sup> ano do EM
corp15.txt	Redações	Reunião dos textos de 7 <sup>a</sup> série e do 3 <sup>o</sup> ano do EM
corp16.txt	Entalpia	Texto técnico sobre entalpia em livro de Química
corp17.txt	Superinteressante	Reportagens sobre assuntos da Química
corp18.txt	Química Geral	Textos e exercícios de livros didáticos de Química
corp22.txt	Redações reduzido	Textos de 7 <sup>a</sup> série e 3 <sup>o</sup> ano (apenas 2.466 <i>tokens</i> )

Para compor o *corpus* de redações, solicitamos a colaboração de professores de Língua Portuguesa de 7<sup>a</sup> série e do 3<sup>o</sup> ano do Ensino Médio do Colégio de Aplicação da UFRGS. Eles gentilmente cederam as turmas para realização da produção textual a ser utilizada na pesquisa em busca da possível influência do *internetês* nos textos escolares.

Tanto para as turmas do Ensino Fundamental como para as do Ensino Médio foi feita a seguinte proposta de redação:

Você participou de uma festa na qual vários fatos inusitados aconteceram. No outro dia, um amigo que não esteve na festa, muito curioso, queria saber o que tinha acontecido. RELATE POR ESCRITO O ENCONTRO COM O AMIGO E A CONVERSA QUE TIVERAM.

O dicionário Houaiss indica INUSITADO assim: *adj.* (1572 cf. IAVL) **1** não usual **1.1** que não é corrente, que não se usa ou emprega com frequência, que foge a padrões costumeiros; incomum, insólito, estranho; inabitual <procedimento i.> <recurso i.> <escolha i.> <os tons berrantes são i. na paleta desse pintor> **1.2** que causa surpresa, estranhamento; que é diferente do que se espera ou se imagina <festa i.>

Para os alunos do Ensino Médio, a proposta foi inserida num texto maior, semelhante ao das provas de redação da UFRGS (Anexo III).

O tipo de solicitação proposto aos alunos para sua produção foi pensado tendo em vista um texto que se assemelhasse ao que os jovens escrevem na Internet. Uma redação com diálogos ou a simples narração de um fato para um interlocutor, de alguma forma, lembra o que é escrito nos *scraps*. A reprodução da fala de diálogos num texto escolar deve ser registrada de acordo com a forma oficial de grafia, mesmo no registro das falas consideradas “erradas”. O que pretendíamos observar é se a presença de algumas formas (“erradas” ou não) seria escrita repetindo as formas peculiares do *internetês*.

Os professores solicitaram a produção dos textos sem mencionar sua utilização em pesquisa. Como no Colégio de Aplicação há a prática de produção de texto semanal, para os alunos essa foi apenas mais uma, realizada no dia 6 de julho na 7ª série e no dia 9 de julho no 3º ano do Ensino Médio.

Recebemos 63 textos da 7ª série e 90 do 3º ano. Numa primeira observação praticamente nada relacionado ao *internetês* chamou a atenção. Selecionamos 15 textos de uma mesma turma para cada série, constituindo, então, um *corpus* de 30 textos escolares. Nomeamos esse primeiro *corpus* de *Redações*. Como as redações das duas séries se mostraram bastante distintas, o que poderia evidenciar dados peculiares de escrita referentes à faixa etária dos alunos, resolvemos constituir mais dois *subcorpora*, *Redações7<sup>a</sup>* e *Redações3<sup>o</sup>*, com as redações de 7ª série e do 3º ano, respectivamente.

Os textos integrantes dos *corpora* foram digitados tal qual o registro feito pelos alunos, sendo cópia fiel do manuscrito. Cada produção escolar foi numerada e identificada pelo

primeiro nome do aluno, para conferir o original e a cópia, caso necessário posteriormente. Essas informações foram colocadas entre os caracteres < e > para não serem computadas como palavras do *corpus*, quando o *corpus* fosse analisado pelo *WordSmith Tools*. Para fins de citação dos autores na pesquisa, utilizamos os termos Aluno 1, Aluna 2, por exemplo, para garantir privacidade. A indicação de gênero também pode ser um dado relevante, porque sabemos, por exemplo, que certas formas são mais típicas de emprego feminino.

Os outros textos componentes dos *corpora* de contraste fazem parte do acervo do Projeto TextQuim<sup>47</sup>, da UFRGS. Esse projeto, desde 2003, estuda diferentes perfis da linguagem, da produção textual e das terminologias da Química em língua portuguesa. O primeiro desses *corpora* é composto de um pequeno texto, que trata do tema entalpia. O segundo tem como conteúdo reportagens publicadas na revista *Superintessante* sobre assuntos de Química, em linguagem simplificada. O terceiro contém textos de livros didáticos de Química utilizados no ensino superior.

## REFERÊNCIA

Nome1	Nome2	Conteúdo
corp19.txt	BcoPort	Textos escritos de variados tipos
corp20.txt	BcoFalado	Textos falados de aulas e de conversação
corp21.txt	VOLP	Vocabulário Ortográfico da Língua Portuguesa
corp.23.txt	FaladoNURC	Textos falados de entrevistas orais

Utilizaremos como primeiro *corpus* de referência o Banco de Português, um *corpus* monitor do português do Brasil, atualizado constantemente. Em 2003, já registrava 240 milhões de palavras (*tokens*) recolhidas. Uma parte é franqueada ao público, que pode utilizar uma amostra do *corpus*, com cerca de 1,1 milhão de palavras. O Banco de Português foi criado e é mantido no âmbito do projeto DIRECT e faz parte dos Bancos de Dados do CEPRIL, LAEL, PUC/SP. Maiores informações sobre esse *corpus* são obtidas no *site* <http://lael.pucsp.br/corpora/>. Na verdade, quanto ao modo, esse grande *corpus* do português brasileiro se compõe de dois *subcorpora* distintos, o *subcorpus* escrito e *subcorpus* falado, ambos utilizados como *corpus* de referência neste trabalho. O primeiro disponibiliza para pesquisa 985.092 palavras, enquanto o segundo *subcorpus* possibilita acesso à fala de aulas

<sup>47</sup>**TextQuim:** sigla para pesquisa que se ocupa de TEXTOS DE QUÍMICA. Projeto desenvolvido pelo Instituto de Letras e Área de Educação Química da UFRGS. Disponível em [www.ufrgs.br/textquim](http://www.ufrgs.br/textquim).

(84.910 palavras) e de conversação (112.991 palavras) totalizando 197.901 palavras de registros falados.<sup>48</sup>

O Banco de Português é um *corpus* de língua geral do português do Brasil e tem como recursos disponíveis:

- Concordanciador de uma amostra de 1,1 milhão de palavras do Banco de Português;
- Lista de palavras do *subcorpus* escrito, formato .txt (zip, 1,67 MB);
- Lista de palavras do *subcorpus* falado, formato .txt (zip, 288 KB);
- Listas de palavras integrais do *corpus* (acesso restrito);
- Acesso ao *corpus* integral / Browse the corpus (acesso restrito);
- Índice do *corpus* / Corpus contents (acesso restrito).

Observando atentamente, o Banco do Português pode ser visto como um *corpus de corpora*, por possuir diversos *corpora* independentes, coletados inicialmente sem intenção de comporem o Banco, como os variados anos do jornal *Folha de S. Paulo*, publicados em CD-ROM. Na seção da fala, o Banco de Português é composto de transcrições de variados projetos, como o Porcufort<sup>49</sup>, disponível na Internet. Na parte referente a negócios, por exemplo, a fonte de dados é o *corpus* de linguagem de negócios do Projeto Direct<sup>50</sup>, que abrange o próprio Banco de Português.

Como segundo *corpus* de referência, para o segmento fala, utilizamos os textos disponibilizados *on-line* do Projeto NURC (Projeto de Estudo da Norma Lingüística Urbana Culta)<sup>51</sup> do Rio de Janeiro. O Projeto NURC teve por objetivo documentar e descrever o uso urbano do português falado no Brasil, em seus aspectos fonético-fonológicos, morfológicos, sintáticos e vocabulares. Desenvolvido em cinco capitais brasileiras (Recife, Salvador, Rio de Janeiro, São Paulo e Porto Alegre), visou ao estudo da fala culta média, habitual. Foi constituído um *corpus* no país com base em critérios rigorosos na seleção dos informantes e no controle de variáveis, perfazendo cerca de 1500 horas de registros magnetofônicos. O *corpus* que utilizamos neste trabalho foi formado pela recolha das entrevistas realizadas no Rio de Janeiro, cujo Arquivo Sonoro da fala culta daquela cidade reúne 330 horas de elocuições de 493 locutores, em 394 inquéritos. Esse material representa o desempenho lingüístico de falantes de

<sup>48</sup> Informação obtida no site <http://www2.lael.pucsp.br/corpora/bp/conc/index.html> acessado em 20/07/2007.

<sup>49</sup> O projeto PORCUFORT (Português Oral Culto de Fortaleza) é uma aplicação, na cidade de Fortaleza-CE, do projeto NURC (Norma Urbana Culta), cujo objetivo é a descrição do português falado culto no Brasil. Organizado pelo professor José Lemos Monteiro, é disponibilizado na internet, no site <http://www.geocities.com/Paris/Cathedral/1036>.

<sup>50</sup> Projeto desenvolvido pela PUCSP, que estuda a linguagem de todas as profissões, em português, inglês e espanhol.

<sup>51</sup> Disponível em: <http://www.lettras.ufrj.br/nurc-rj/>. Acesso em: 30 nov. 2007.

ambos os sexos, nascidos na cidade do Rio de Janeiro, com escolaridade universitária, distribuídos por três faixas etárias.

Copiados os textos, foram salvos em formato .txt para poderem ser processados pelo *software WordSmith Tools*. O cabeçalho com as informações sobre cada entrevista foram colocados entre os caracteres < e > para que suas palavras não fossem lidas como integrantes do *corpus*. Para termos uma idéia do que contém o cabeçalho, transcrevemos um deles:

<NURC RJ

**AMOSTRA COMPLEMENTAR: Inquérito 23 (masculino / 33 anos)**

**TEMA: Vida Social e Diversões**

**LOCAL/DATA: Rio de Janeiro, 2 de julho de 1996.**

**TIPO DE INQUÉRITO: Diálogo entre informante e documentador**

**DOCUMENTADOR: C L>**

A seguir, uma pequena amostra, contendo o início e o fim da reprodução da entrevista que se segue ao cabeçalho:

**Doc.** É sobre futebol... descrever... como é que é uma partida de futebol... quais são os elementos... os componentes... estrutura...

**Loc.** Bom... descrever como é que é uma partida de futebol ((risos)) eu prefiro jogar futebol... é melhor do que descrever como é que é uma partida...

**Doc.** Você joga futebol há muito tempo?

**Loc.** Olha... eu jogo futebol desde que eu me entendo por gente... [

**Loc.** [sempre foi uma grande paixão na minha vida

**Doc.** [ como é que você começou?

**Doc.** Como é que você começou a jogar futebol? Como é que você... começou a jogar futebol?

**Loc.** Olha... eu nasci e fui criado em Ramos... subúrbio da Leopoldina... e acho que praticamente todo filho homem que nasce em Ramos ou em um outro subúrbio da Leopoldina ganha logo de presente do pai uma bola de futebol... a camisa do time do coração do pai... que acaba sendo a extensão do time do coração do filho... e... desde que eu me lembro... desde garoto... eu me lembro que... frequentemente... nos aniversários eu ganhava bola de presente... eu jogava com meu pai... jogava com o meu avô... ele fazia... visitas semanais... a minha casa... e praticamente com o meu avô que eu aprendi a... começar a chutar a bola... e a ter interesse pelo esporte

[...]

**Doc.** E você acha que... se diz que as crianças criadas em condomínios só se divertem em *shoppings*... eh... sei lá... podem... viver num mundo... numa redoma de vidro... não ter... não saber... o que se passa na realidade... o que você acha disso?

**Loc.** Eu acho que essa... essa realidade ela é... naturalmente multiforme... quer dizer... eu acho que... na perspectiva humanista... de... o que serão essas pessoas... quando crescerem... que postura terão... que visão de mundo... acho que... a princípio... tudo leva a crer... que não sejam lá muito positivo... a visão de mundo deles deve ser certamente muito restrita... muito vinculada àquela realidade mais... direta e... quase ( )... que eles conhecem... né... agora... a perspectiva geral... eh... a relação das pessoas... com a... a realidade... a realidade é bem diferenciada mesmo... quer dizer... eu acho que... se pode também... estar alheio a uma série de problemas... a uma série de... de conflitos... e de misérias sem... estar isolado dentro de um bairro daqueles... não acho que... isso necessariamente... pode ser sim uma experiência traumática... quando lá na adolescência se começa a ter a noção que o mundo não se reduz àquilo né...

Apresentaremos mais detalhes sobre esse material na **Seção 4.3 Corpus de Referência**.

Depois de ver os *corpora* de contraste e de referência, voltemos à tipologia do *corpus* (e de seus *subcorpora*) de estudo que formamos. Com base em Berber Sardinha (2000, p. 340), é classificado quanto a modo, tempo, seleção, conteúdo, finalidade, autoria e integralidade, respectivamente. Assim, nosso *corpus* de estudo é:

- escrito, mesmo que não impresso, foi registrado na tela do computador por meio eletrônico;
- contemporâneo, por ser do período de tempo corrente;
- de amostragem, composto de textos, porções de textos e variedades textuais. É uma amostra finita da linguagem dos jovens na Internet;
- dialetal, inicialmente considerado, por serem textos provenientes de uma variedade sociolinguística específica de falantes de determinada faixa etária em ambiente da rede mundial de computadores;
- de língua nativa, todos os autores são falantes do português;
- de estudo, porque é o que pretendemos descrever;
- com pluralidade de autoria, por serem produzidos por centenas de autores;
- de textos integrais, depoimentos completos e, mesmo que os recados pareçam fragmentos, essa é uma característica de sua tipologia.

Nesta subseção ficamos conhecendo como foram constituídos e quais são os *corpora* (de estudo, de contraste e de referência) utilizados, totalizando duas dezenas se considerarmos todos os *subcorpora* que foram sendo elaborados ou localizados para o desenvolvimento da pesquisa. Para realizá-la, foram necessárias também ferramentas para extrair o que se pretendia

verificar no *corpus*, a fim de atingir o objetivo proposto: descrever o léxico da escrita no Orkut. A seguir, então, apresentamos as ferramentas e sua utilização na pesquisa com Linguística de *Corpus*.

### 3.3 AS FERRAMENTAS UTILIZADAS

Para operacionalização da descrição lingüística com *corpus* existem alguns programas de computador como o *Simple Concordance Program* e o *WordSmith Tools*, em várias versões, sendo este último o mais empregado pelas facilidades e recursos de que dispõe. Berber Sardinha (1994, p. 86) assim a ele se refere:

O programa coloca à disposição do analista uma série de recursos que, bem usados, são extremamente úteis e poderosos na análise de vários aspectos da linguagem, como a composição lexical, a temática de textos selecionados e a organização retórica e composicional de gêneros discursivos.

O programa *WordSmith Tools* funciona com base em três princípios abstratos básicos, quais sejam:

- 1) Ocorrência: só são considerados os itens<sup>52</sup> presentes, caso contrário, não são incorporados porque não observáveis;
- 2) Recorrência: os itens devem ocorrer mais de uma vez, sendo considerados também os de frequência 1, os *hapax legomena*, formadores da maioria dos itens de linguagem;
- 3) Coocorrência: os itens devem estar com outros, porque isolados são pouco informativos. É na relação com outros (parte de um conjunto) que são interpretados.

É uma suíte de aplicativos composta de ferramentas, utilitários e funções. Para a pesquisa foram utilizadas apenas duas ferramentas: a *WordList* e a *Concord*. A primeira possibilita a criação de listas de palavras, disponibilizadas em ordem alfabética ou pela ordem de frequência. Também fornece dados estatísticos como o número total de palavras (*tokens*), número de palavras diferentes (*types*), a razão vocábulo/ocorrência (*Type-Token Ratio*) e outras informações. É pela razão vocabulário/ocorrência que podemos perceber a riqueza vocabular de um texto. Quanto maior o número de palavras diferentes (*types*) num conjunto de palavras (*tokens*), maior será a riqueza desse texto quanto ao seu vocabulário. Além disso, a *WordList*

<sup>52</sup> *Itens* corresponde ao que comumente consideramos palavras, tanto podem ser *types* quanto *tokens*.

nos possibilita definir também o número de palavras que um agrupamento tenha, os chamados *clusters*.

Já a *Concord* é assim explicada por Berber Sardinha (2004, p. 105): “Essa ferramenta produz concordâncias de um item específico (chamado palavra de busca ou nóculo, que pode ser formado por uma ou mais palavras) acompanhado do texto ao seu redor (o co-texto)”. O tipo de concordância empregada neste trabalho para a compreensão de vocábulo foi o mais convencional deles, chamado de KWIC (‘Key Word In Context’). Mostra a palavra de busca no centro da listagem ladeada pelas palavras que ocorreram no texto junto a ela (normalmente 5, à direita e à esquerda).

A seguir, apresentamos alguns exemplo de como utilizamos essas ferramentas, com o objetivo de tornar mais acessível a compreensão de seu uso.

**Quadro 2: Lista das dez primeiras palavras do *Corpus Geral* em ordem alfabética**

N	Palavra	Freq.	%
1	£	2	
2	¢	10	
3	çÊ	4	
4	çS	9	
5	çSLA	1	
6	A	10.122	1,84
7	À	90	0,02
8	Á	43	
9	Â	134	0,02
10	Ã	3	

**Quadro 3: Formas de escrever *abraço/abraços* (da lista de palavras do *Corpus Geral*)**

N	Palavra	Freq.
277	ABRASO	2
278	ABRASS	2
279	ABRASSAO	1
280	ABRASSAUMM	1
281	ABRASSO	2

282	ABRASSOOO	1
283	ABRAXO	2
284	ABRAXOS	1
285	ABRAXUSSSSSSSS	1
286	ABRAXXXXXXXXXXX+	1
287	ABRÇO	1
288	ABRÇS	3
289	ABRÇSS	2

**Quadro 4: As dez primeiras palavras mais frequentes da lista de palavras do *Corpus Geral***

N	Palavra	Freq.	%
1	E	13.877	2,52
2	QUE	11.500	2,09
3	EU	10.571	1,92
4	A	10.122	1,84
5	DE	9.899	1,80
6	Q	9.016	1,64
7	TE	8.877	1,61
8	O	8.305	1,51
9	É	7.827	1,42
10	VC	6.342	1,15

**Quadro 5: Primeiras dez concordâncias da palavra *cmg* (=comigo) de um total de 407 ocorrências no *Corpus Geral***

N	Concordance	Set	Tag	Word No.	File	%
1	ENDO ANIVER JUNTO	<b>CMG</b>	NÉÉÉÉH? TE h	34.317	d:\corp1.txt	7
2	nunca iria acabar.....	<b>Cmg</b>	está tudo ótimo!! T	9.903	d:\corp1.txt	2
3	E MUNDOO.. CONTA	<b>CMG</b>	PÁ TUDO VIU?!	381.309	d:\corp1.txt	70
4	ry vamu compra lanche	<b>cmg</b> !?	a piri vamu fica	123.974	d:\corp1.txt	23
5	mia cmg, sai pra come	<b>cmg</b> ,	sai pra compra co	391.524	d:\corp1.txt	72
6	lhe, me da colo, brinca	<b>cmg</b> ,	tira fotos cmg, que	262.478	d:\corp1.txt	49
7	tece vo te leva sempre	<b>cmg</b> !	TE AMO E NA	496.825	d:\corp1.txt	91
8	oro mto e conta sempre	<b>cmg</b>	viu? bjo ooooooo	463.862	d:\corp1.txt	85

9	o muito e sempre conta <b>cmg</b> . Bjinhos Pocahonta	430.210	d:\corp1.txt	79
10	gnt jah viveu.. thi cont <b>cmg</b> qnd precisarr i a	143.13	d:\corp1.txt	26

**Quadro 6: Os dez primeiros *clusters* (reunião de três palavras) contendo *cmg* (= comigo) de 64 ocorrências**

N	cluster	Freq.
1	pode contar <b>cmg</b>	38
2	conta <b>cmg</b> pra	17
3	conta sempre <b>cmg</b>	16
4	pode conta <b>cmg</b>	16
5	conta <b>cmg</b> sempre	13
6	sempre contar <b>cmg</b>	12
7	contar <b>cmg</b> sempre	11
8	pode sempre contar	9
9	pod conta <b>cmg</b>	8
10	contar <b>cmg</b> pra	7

No próximo capítulo, empreendemos, finalmente, a descrição dos *corpora* que utilizaremos para investigar nosso objeto de estudo: a escrita do léxico dos jovens no Orkut.

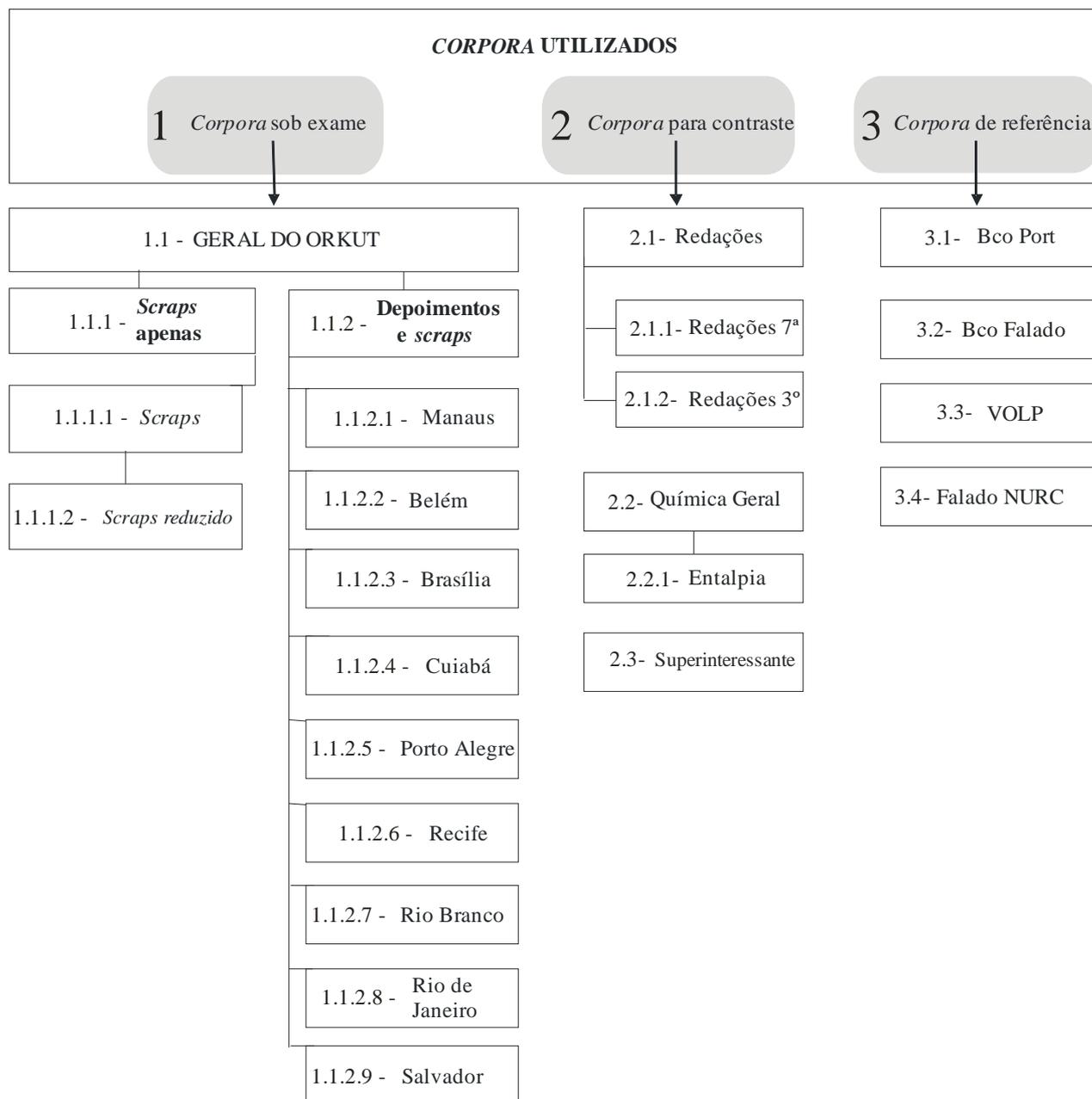
## 4 DESCRIÇÃO DOS *CORPORA*

Por realizarmos pesquisa do uso da linguagem utilizando os princípios da Linguística de *Corpus*, necessitamos, é óbvio, de vários *corpora*, que serão examinados e descritos neste capítulo. Os textos para os *corpora* foram selecionados de diferentes fontes, conforme já explicitado no Capítulo 3. Não esqueçamos que estudos de uso, no dizer de Biber (1998), requerem análise empírica de textos autênticos. É o caso desses coletados para a composição dos *corpora*.

Neste capítulo, primeiramente veremos o tipo de texto pesquisado formador do *corpus geral* sob exame, com exemplos, as divisões desse *corpus geral*. Depois temos os *corpora* de referência e os *corpora* de contraste. Em seguida, trazemos comparações entre os diferentes *corpora*, entre o *internetês* geral do *corpus* de todo o Brasil e o dos *corpora* das regiões. Por último, vem uma síntese contendo, entre outros, dados sobre a riqueza vocabular e algumas ponderações.

Em função dos diferentes *corpora* sob exame, com 19 recortes entre o principal, os de referência e os de contraste, o Esquema 1 a seguir visa ser uma representação didática e um auxílio para nosso leitor.

### Esquema 1 – *Corpora* utilizados no trabalho



Como vemos, esse esquema reúne todos os *corpora* utilizados na pesquisa:

- Para **exame**, criamos o *corpus* geral do Orkut, formado por dez *subcorpora*, nove com textos de depoimentos e *scraps* de diferentes regiões do Brasil e um apenas composto por *scraps*.
- Foi necessário organizar mais um *subcorpus*, ao qual chamamos de *Scraps Reduzido*, extraído do próprio *corpus* de *scraps*, para comparação entre a riqueza vocabular de reportagens premiadas sobre diversidade ambiental e a dos *scraps*.

- Para **contraste**, utilizamos um *corpus* de redações formado por dois *subcorpora*, um de redações de alunos de 7ª série do Ensino Fundamental e outro de alunos do último ano do Ensino Médio.
- Utilizamos também mais três *corpora* para contraste: um com textos de manual de Química, outro contendo um segmento desse material e um terceiro com reportagens sobre assuntos científicos publicados na revista *Superinteressante*.
- Como *corpus* de **referência** ou de **controle**, com a função de fornecer norma para comparação com o que aparece no *internetês*, utilizamos o Banco do Português Escrito, o Banco do Português Falado, o VOLP (Vocabulário Ortográfico da Língua Portuguesa) e o Falado NURC-RJ.

Além das informações extraídas dos *corpora* de estudo, de contraste e de referência, nos valem também dos dados levantados por Belmonte (2007) sobre riqueza vocabular em textos jornalísticos considerados de excelência por terem sido premiados em concurso. Faremos um contraponto entre a riqueza lexical do *internetês* empregado nos *scraps* e a dos textos respeitados no meio jornalístico

Cabe ainda lembrar aqui Biber (1998) nos informando que a Linguística de *Corpus* permite ao pesquisador identificar “padrões de associação”, por focalizar o uso de uma característica lingüística e por focalizar as características de textos ou variedades. Nos alerta que é preciso considerar como um item lexical ou uma construção gramatical são distribuídos entre: 1) variedades definidas por situação; 2) variedades definidas por grupo social, e 3) período de tempo. Diz-nos mais o autor: identificar padrões não significa que aquilo encontrado é “lei” ou é o que sempre acontece ou o que nunca acontece. É preciso pensar em certos padrões como “raros” e “comuns” e, então, estabelecer o que é raro e o que é comum na língua. Outra informação importante é sobre os contextos em que as palavras são usadas. Eles mostram o significado das palavras, que podem ser encontrados nas listas de concordâncias. Muitas palavras podem ter mais significados do que os indicados nos dicionários. Essas e outras possibilidades poderemos encontrar analisando os *corpora* que estamos descrevendo nesta seção.

#### 4.1 O TIPO DE TEXTO EM FOCO

Assim como a estenografia, ainda utilizada em situações especiais, a linguagem dos jovens na Internet, à primeira vista, aparece registrada como uma escrita simplificada do original para aproveitar melhor o tempo e o espaço. É uma linguagem que prima pela informalidade, aparentemente sem compromisso nenhum com normas lingüísticas oficiais, assim tornando o contato entre os usuários mais natural. Se alguém entrar numa sala de bate-papo entre jovens, por exemplo, e colocar todas as letras corretamente está se denunciando como não sendo do meio ou que não está acostumado com a Internet. Já no Orkut, um ambiente não apenas para jovens e para comunicação informal, em que fica registrado o que se escreve, há textos escritos de acordo com a norma culta. Observamos, entretanto, que mesmo nos depoimentos registrados por jovens para pessoas mais velhas é utilizado o *internetês*.

O tipo de texto dos *corpora* recolhidos do Orkut para análise se caracteriza por usar abreviações sem seguir regras de abreviatura (*bjim, cervj*), várias formas para um mesmo vocábulo (*bj, bjo, bjus, bju, bjuxx*), grande número de sinais de pontuação e repetição de letras para dar ênfase (*foi boooooooooom!!!!*), escrita praticamente como transcrição da fala (*akeli, fikei*), desconsideração com as normas ortográficas básicas como emprego de maiúsculas e acentuação (*iai katiuuuuuscia beleuza contigo!!!!!!!!!!*).

Além disso, há alguns acréscimos como indicação de sílaba tônica das oxítonas pelo emprego do “h” (*ateh*), da nasalização pelo “m” ou “n” (*naum*) e uso de onomatopéias, principalmente para indicar riso e choro (*hehehehe, shuashuashua* e *tsc tsc*).

Vejamos dois exemplos desse tipo de texto, extraídos do *corpus geral*, o primeiro é um depoimento e o segundo, um *scrap*.

*hhhhhm...boum mew õ.O u qui fla da tinuxa*

*bah...essa guria eh pah laaaaa di ixpecial...eu amu muuuuuito ela pqqqqq*

*ela eh fofaaaaa # ela eh liiinda # ela eh minha migaaaaaa # ela eh loook # ela eh especial na minha vida # ela eh tuuuuuuudu di boum*

*boum mew...sei lah xD apesar di ela as vezis quasi seimpri mi cortah dus bagulhu*

*eu amu muuuuuuuuuito muito muito elaaaaa...ela sb qui eu voh tah seimpri aew pah ela #D ela*



Como pudemos perceber, os depoimentos, conforme o nome já indica, são textos declarativos, com o objetivo de registrar a opinião sobre um amigo na página pessoal dele no Orkut. Tais textos recebem o nome de depoimento ou testemunho e ficam expostos na página principal, junto ao perfil do usuário. São textos pequenos, de cunho pessoal, produzidos com carga emotiva para o destinatário, expressando a opinião do autor sobre ele.

Já *scrap* é uma palavra inglesa que significa recado. No Brasil se tornou sinônimo de recado no Orkut, porque quando só existia a versão em inglês as mensagens enviadas apareciam escritas com essa denominação. Portanto, *scrap* é um recado, bilhete ou anotação enviados para pessoas do Orkut, pode ser falando oi, dando aviso, xingando ou abordando qualquer assunto. Normalmente é um pequeno texto, porque há um espaço limitado para escrevê-lo.

## 4.2 DIVISÕES DO CORPUS DE PESQUISA

Conforme o esquema 1 já apresentado, o *corpus* de estudo é composto de depoimentos e *scraps* coletados do Orkut no período de abril a junho de 2007. Foi recolhido um número maior de depoimentos porque são textos maiores, enquanto a maioria dos *scraps* tem poucas linhas, o que tornaria a recolha muito trabalhosa.

Inicialmente pensamos realizar a pesquisa em apenas um *corpus*, mas por algumas peculiaridades no léxico vistas em depoimentos de usuários de diferentes regiões do Brasil, foram criados nove *subcorpora*. São indicados pelo nome das capitais onde nelas e em cidades do seu entorno foram recolhidos os textos. Também foi criado um *subcorpus* apenas de *scraps*, totalizando onze *subcorpora* com seus conteúdos reunidos no *Corpus Geral do Orkut*. Do *corpus* de *scraps* originou-se o décimo segundo, com a redução para 2.371 *tokens* a fim de comparar a riqueza vocabular dos recados com aquela presente em reportagens premiadas sobre biodiversidade. O contraste entre os dois tipos de texto está descrito no próximo capítulo.

### *Corpus Geral do Orkut*

Apresentamos, a seguir, as principais informações sobre o *Corpus*. O primeiro quadro contém todos os dados indicado pelo *WordSmith Tools*. Para entender melhor o que esse programa de computador para análise lingüística nos indicou, lembremos que ele nos informa por meio de três instrumentos fundamentais: *WordList* (informa listas de palavras, uma ordenada alfabeticamente e outra por ordem de freqüência), *KeyWords* (compara freqüência de palavras entre *corpus* de estudo e *corpus* de referência) e *Concord* (produz lista de ocorrências de palavra acompanhada de texto ao seu redor, o cotexto). Os dados dos quadros desta seção são informados pelo *WordList*.

Destacamos como números fundamentais neste estudo, informados na janela que indica a estatística, aqueles que indicam a quantidade de *tokens* (itens ou palavras), *types* (ocorrências, formas ou palavras distintas) e a relação ou razão *type/token*. São os dados mais importantes para entendermos o léxico de uma língua e a riqueza vocabular de seus falantes.

Ao lado, no mesmo quadro, também informado pelo *WordList*, temos a lista das palavras do *corpus*, indicadas pela ordem crescente do número de vezes que elas apareceram. Assim, encabeça a lista a mais freqüente, no caso, a conjunção “e”, com 13.930 ocorrências, ocupando 2,52% do *corpus*. Observemos que na mesma lista das mais empregadas aparecem as formas “que” e “q”, ambas com mesma significação no *internetês*.

#### **Quadro 7: Dados Estatísticos e lista das palavras mais freqüentes do *Corpus Geral do Orkut***

<i>CORPUS GERAL</i> Estatística	<i>CORPUS GERAL</i>			
	Nº	Palavra	Freq.	%
Bytes 3.114.941	1	E	13.930	2,52
Tokens 553.875	2	QUE	11.537	2,08
Types 38.803	3	EU	10.619	1,92
Relação Type/Token 7,01	4	A	10.173	1,84
Padronização Type/Token 46,11	5	DE	9.950	1,80
Ave. Tamanho da palavra 4,01	6	Q	9.028	1,63
Sentenças 16.349	7	TE	8.937	1,61
Tamanho da Sentença 21,28	8	O	8.396	1,52

sd. Tamanho da Sentença 36,01	9	É	7.855	1,42
Parágrafos 13.630	10	VC	6.347	1,15
Tamanho dos parágrafos 40,64	11	PRA	5.640	1,02
sd. Tamanho dos parágrafos 102,37	12	UM	5.252	0,95
Cabeçalho 0	13	AMO	4.960	0,90
Tamanho do cabeçalho	14	EH	4.525	0,82
sd. Tamanho do cabeçalho	15	UMA	4.386	0,79
Palavras de 1 letra 62.538	16	MAIS	4.249	0,77
Palavras de 2 letras 104.328	17	ELA	4.051	0,73
Palavras de 3 letras 113.600	18	SEMPRE	3.905	0,71
Palavras de 4 letras 66.226	19	SE	3.851	0,70
Palavras de 5 letras 76.645	20	MAS	3.683	0,66
Palavras de 6 letras 46.750	21	NÃO	3.564	0,64
Palavras de 7 letras 30.558	22	ME	3.524	0,64
Palavras de 8 letras 21.425	23	MUITO	3.399	0,61
Palavras de 9 letras 11.866	24	COM	3.369	0,61
Palavras de 10 letras 7.583	25	POR	3.215	0,58
Palavras de 11 letras 4.779	26	DO	3.175	0,57
Palavras de 12 letras 2.528	27	PQ	3.158	0,57
Palavras de 13 letras 1.270	28	MEU	3.148	0,57
Palavras de 14 letras ou mais 830	29	DA	2.965	0,54

Observando os dados estatísticos, percebemos que a maior concentração é formada por palavras de duas e três letras, havendo um decréscimo significativo de ocorrências à medida que aumenta o número de letras. A palavra curta reflete a lógica do *internetês*, a ponto de encurtar as longas, abreviando-as de diversas maneiras, desde que possam ser legíveis ou decifráveis pelo interlocutor. A Figura 6, a seguir, busca ilustrar a distribuição da extensão das palavras no nosso *corpus*. As palavras de três letras são as mais utilizadas.

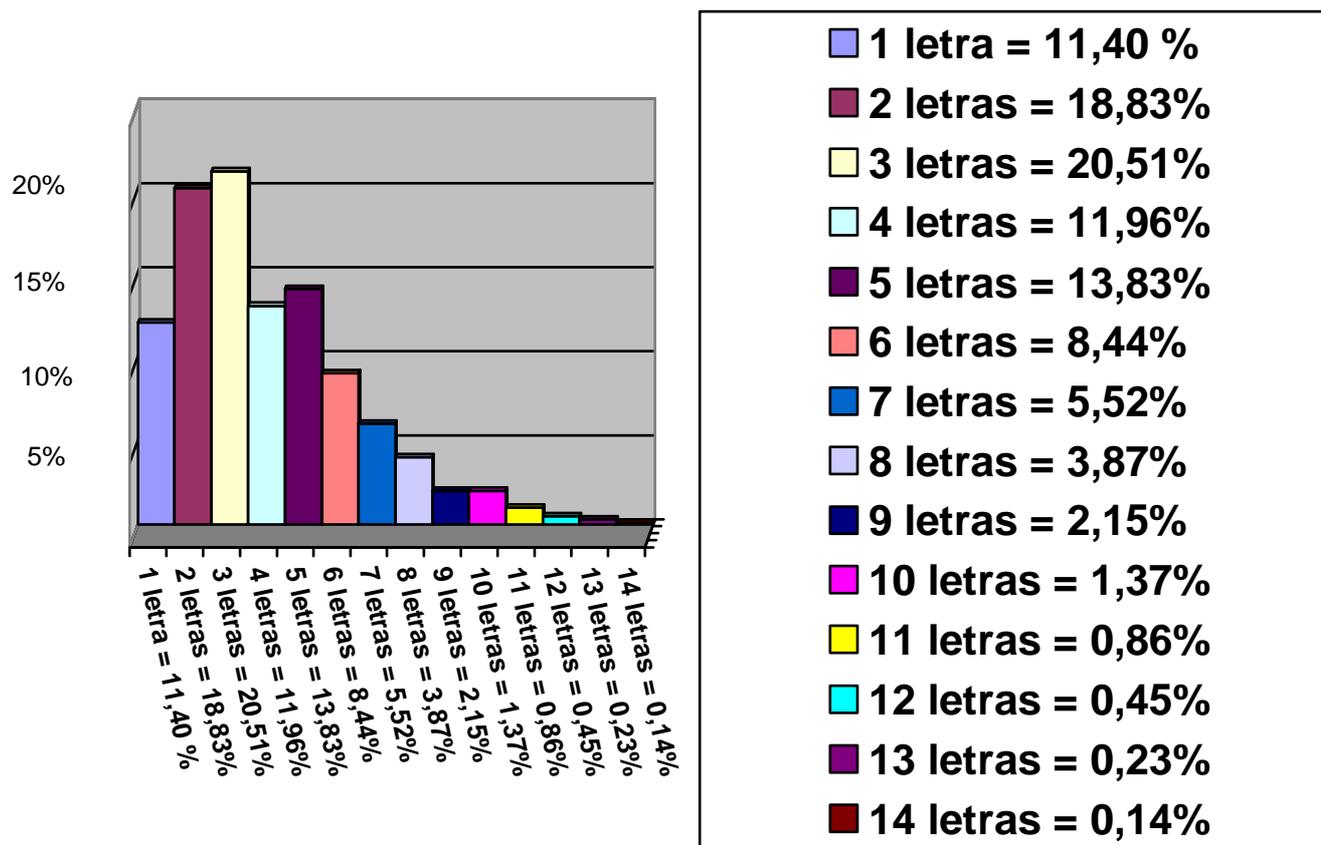


Figura 6: Número de letras das palavras *do corpus* de estudo

O que veremos, em seguida, no Quadro 8 são duas amostras de como o *WordSmith Tools* fornece a lista de freqüência em ordem alfabética. O programa informa o que considera palavra conforme foi planejado para desenvolver tal ação. Em assim sendo, indica Algarismos e símbolos antes de iniciar a listagem das palavras em ordem alfabética. Seriam esses caracteres ou símbolos nos textos escritos intencionalmente, tendo, portanto, significado, ou poderiam ser erros de digitação, muito plausíveis quando se tecla rapidamente *on-line*? Essa questão será respondida mais adiante utilizando-se a ferramenta *Concord*, que verifica em que frases aparecem tais signos, aparentemente, sem significação. Certamente sem significação fora do contexto. O que pode significar um £ numa frase em português? Dependendo do caso poderia indicar uma inovação convencionalizada entre os jovens que, para sua afirmação como sujeitos, alterariam de alguma forma a comunicação já estabelecida. Por ser produto de interferência, é complicado afirmar com certeza qual o tipo de interferência que receberam: se influência da fala, desconhecimento da norma culta ou outra. Lembremos que o *internetês* é o resultado de várias influências, tal como vemos representado na Figura 7 a seguir.

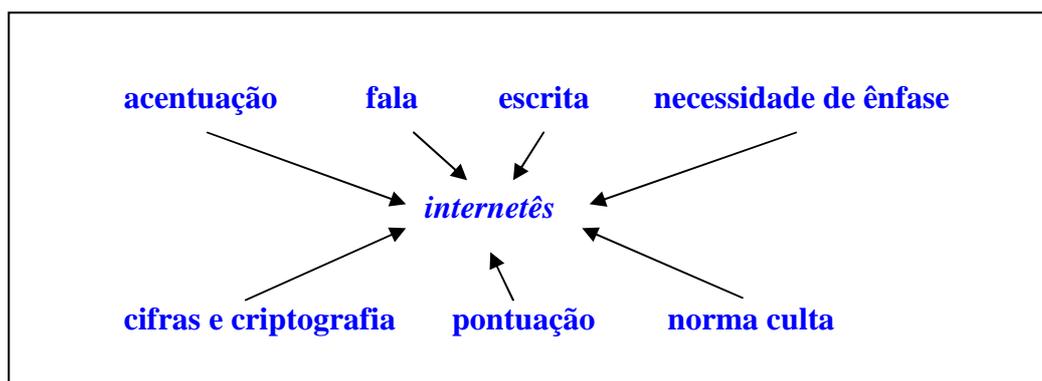


Figura 7: Influências sofridas pelo *internetês*

Quadro 8: Palavras mais frequentes em ordem alfabética e frequência no entorno do “E”

CORPUS GERAL				CORPUS GERAL			
N	Palavra	Freq.	%	N	Palavra	Freq.	%
1	£	2		12363	DZ	5	
2	ç	10		12364	DZÉÉ	1	
3	çÊ	4		12365	DZENDO	1	
4	çS	9		12366	DZER	7	
5	çSLA	1		12367	DZR	6	
6	A	10.173	1,84	12368	E	13.930	2,52
7	À	87	0,02	12369	È	34	
8	Á	42		12370	É	7.855	1,42
9	Â	134	0,02	12371	Ê	6	
10	Ã	3		12372	EA	3	

Como observamos, na lista das palavras mais frequentes em ordem alfabética, aparecem várias formas usualmente pouco empregadas na escrita. Podem ter sido erro de digitação, emprego de uma forma por outra, forma apenas para registrar presença ou casos especiais, que talvez somente quem escreveu saiba qual intenção teve ao registrar daquele jeito. Há aqui um registro importante a ser feito referente a uma limitação do instrumental *WordSmith Tools*. Ao elaborar a lista de palavras (*wordlist*), ele exclui pontos, vírgulas, ponto-e-vírgula e travessões, o que afetará a indicação de *emoticons* formados por sinais de pontuação e letras ou outro caracter.

Vejam os alguns exemplos, extraídos do *corpus* com o auxílio da ferramenta *Concord*, ao buscarmos a concordância para a forma “á”:

**Quadro 9: Exemplos de concordâncias para a forma “á”**

1	??sa q çs? ç????za ?á ?? ??s ?????ç?al ??a
2	S A! qi vc é! agradeço á papai do céu por ter c
3	?s ls?g?... ?? ??? q ?á ?s??s çs??a çs? ?çê
5	dela sabem disso! Eu á conheci recentemente,
6	va ak?la? cs?b? :x ?á ?á ?á :? a g???? gs??a
7	porque ela me aguenta á uns 3 anos :O porqu
8	MmMm!! pRimerona á fLaR deSSE meninoo
9	bu.. deixa eu pensar.. á, foi muito massa o jei
10	.. Veja em meus olhos á paixão... Descubra, si

Os autores dos segmentos expostos nas concordâncias 1, 3 e 6 provavelmente escreveram “á” como algo a preencher uma frase cifrada, típico de brincadeiras na comunicação entre jovens. Nos outros casos, provável uso de crase por ser uma preposição “a” (exemplos 2 e 8), substituição do pronome pessoal (5), emprego como “há” indicando tempo decorrido (7), equivalente à interjeição “ah” (9) e indicação do artigo “a” (10). Pelo que percebemos, então, tal configuração gráfica não passou de falta de atenção ou desconhecimento da gramática normativa. O mesmo não pode ter ocorrido nos três casos estranhos ao falante comum, talvez nem quem escreveu saiba depois de um tempo o que quis dizer com aquela “revoada” de letras e sinais de interrogação. A mesma situação deve ter acontecido com o sinal “£”, cujas duas ocorrências nada indicam para o leitor comum. Ou será um código secreto da comunicação entre adolescentes, um mistério do *internetês*? Vejamos como ocorreram:

ri , a nem lembro mais ;£ só sei que depois a g  
du apelido deli (piripu) ;£ que que eu conte??

Observando as ocorrências, notamos que nos dois exemplos o sinal £ está posposto a ponto-e-vírgula, formando uma figura, que nos lembra um *emoticon*. Como ao construir a *wordlist* a pontuação não é considerada – uma limitação da ferramenta<sup>53</sup> – o sinal foi lido como

<sup>53</sup> A limitação da ferramenta, nesse caso, acreditamos, não invalida os demais dados que ela nos fornece.

uma palavra, tal qual o *q* (que) ou o *t* (te). Consultando três jovens internautas, obtivemos a confirmação de que ;£ é um emoticon, é algo do *internetês* mesmo, uma criação para uso nesse ambiente de comunicação.

Já a forma “Â”, considerada palavra pelo *WordSmith Tools*, apareceu 134 vezes. É um dos componentes de figuras formadas por letras, espaços e diacríticos. Observando as concordâncias, vemos que muitas foram indicadas pelo mesmo número de palavra, o que leva a deduzir que pertençam a um mesmo desenho. É o caso das oito abaixo exemplificadas, transcritas com todas as informações expedidas pelo *WordSmith Tools*.

#### Quadro 10: Oito primeiros exemplos da concordância da palavra “Â”

N	Concordance Set	Tag	Word	No.	File	%
1	_____Â!Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46
2	_____Â!Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46
3	Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46
4	Â!Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46
5	Â!Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46
6	Â!_Â!Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46
7	Â!Â!_Â!Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46
8	_Â!Â!Â!_Â!Â!Â!_____	Â!Â!Â!	Â!Â!Â!	250.258	c:\corp1.txt	46

Podemos ter o mesmo pensamento a respeito do carácter “ç” que aparece em frases ininteligíveis para os simples falantes. É apenas um sinal num código utilizado pelos jovens para expressar sua confusão, seus sentimentos pouco claros em relação a algo? Vejam os exemplos:

N	Concordance Set	Tag	Word	No.	File	%
1	??a ???a ?????sa q çs? ç????za ?á ?? ??s ???			275.644	c:\corp1.txt	51
2	. =) a g???? ?? çs???ç?? ?s ?a?.. (pur curio			275.655	c:\corp1.txt	51
3	a a??a? çsla?a? ?? ??? ç??ç????íí ?}~ a? ???			275.684	c:\corp1.txt	51
4	?a ??? é + ?s q?? ??s?ç?al... çs??a ç?g ??			380.164	c:\corp1.txt	70
5	? ??s?ç?al... çs??a ç?g ??????... BjOoO			380.166	c:\corp1.txt	70

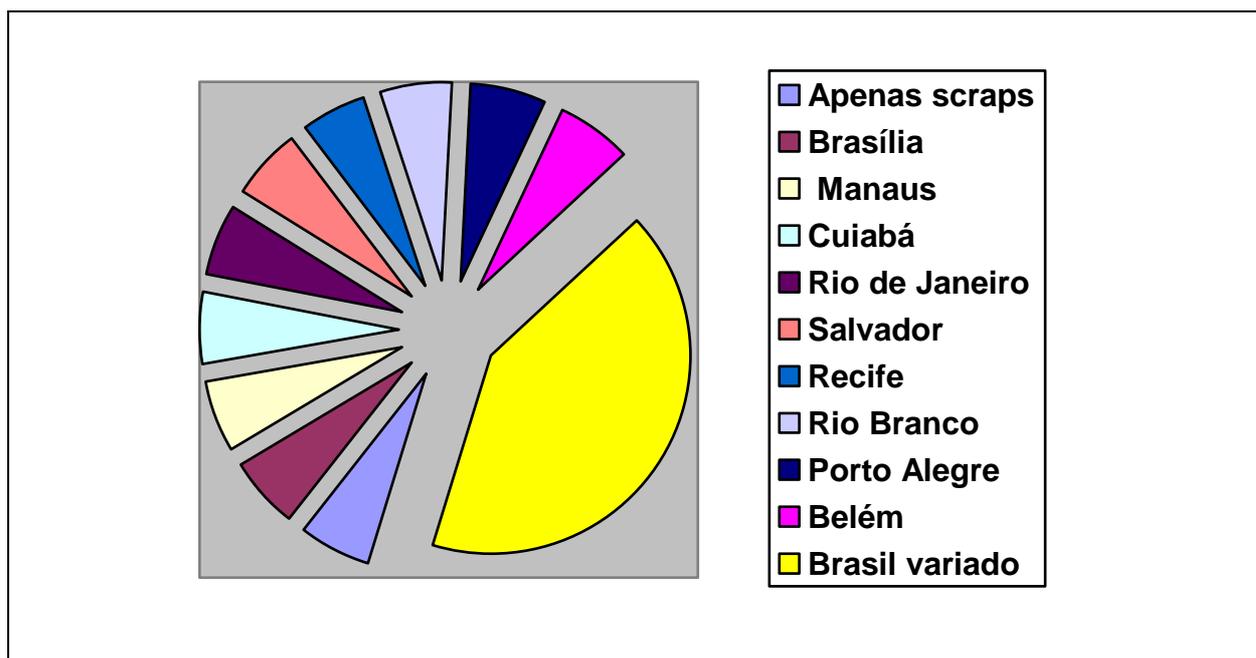
#### 4.2.2 *Corpora Regionais e corpus de Scraps*

Os principais dados levantados na Tabela 1, a seguir, mostram o número total de *tokens* (itens), o número de *types* (formas) e *type/token ratio* (razão forma/item), esta última indicadora da riqueza vocabular de um texto, que abordaremos no próximo capítulo.

**Tabela 1: *Tokens, types e ratio por corpora (Ordem decrescente da ratio)***

<i>CORPUS</i>	<i>TOKENS</i> Itens repetidos	<i>TYPES</i> Formas	<i>TYPE /TOKEN</i> <i>RATIO</i> Riqueza vocabular	RELAÇÃO MÁXIMA
<i>SCRAPS APENAS</i>	32.895	6.653	20,22	100,00
BRASÍLIA	32.761	5.956	18,18	100,00
MANAUS	32.119	5.785	18,01	100,00
CUIABÁ	32.832	5.852	17,82	100,00
RIO DE JANEIRO	32.251	5.744	17,81	100,00
SALVADOR	31.643	5.601	17,10	100,00
RECIFE	30.385	5.006	16,48	100,00
RIO BRANCO	32.282	5.282	16,36	100,00
PORTO ALEGRE	33.140	5.389	16,26	100,00
BELÉM	34.220	5.392	15,76	100,00

Vejamos, na Figura 8, o equilíbrio dos diferentes *corpora* regionais compostos para estudo. A “fatia” intitulada *Brasil variado* é composta aleatoriamente por *scraps* e depoimentos de diferentes partes do Brasil.



**Figura 8: Equilíbrio da composição dos corpora regionais**

A palavra de maior frequência de uso no português escrito no Brasil é a preposição “de”, devido à sua natureza de funcionar como elemento associativo entre diferentes tipos de palavra. Villavicencio, Finatto e Possamai (2006) afirmam que “a alta frequência dessa preposição é confirmada em diversos corpora de linguagem escrita, seja de registros de linguagem cotidiana ou de linguagem técnica/especializada”. E comprovam com os resultados pesquisados nos corpora Banco de Português (com percentual de ocorrência de 4,42 % em relação ao total de palavras), Folha de S. Paulo (4,7 %), NILC<sup>54</sup> (4,2%) e Lacio-Ref<sup>55</sup> (4,5%), esses de linguagem cotidiana. Para os de linguagens especializadas, a pesquisa utilizou os corpora do TextQuim, Manuais didáticos acadêmicos (com percentual de ocorrência de 5,12), revista Química Nova (5,45%) além do CORTEC<sup>56</sup> segmento Ecoturismo (5,8%). Dos dados levantados em nossos subcorpora - e conseqüentemente no corpus geral – sobre a frequência das palavras, vemos que a preposição “de” não aparece em nenhum deles como a palavra mais

<sup>54</sup> Núcleo Interinstitucional de Linguística Computacional da Universidade de São Paulo, com sede em São Carlos, SP.

<sup>55</sup> Lacio-Ref é um segmento do Lácio Web, um site brasileiro com um corpus contemporâneo de língua geral, subdividido em subcorpora que representam vários gêneros e tipos textuais - textos disponíveis para download e para serem usados com as próprias ferramentas do site. Disponível em <http://www.nilc.icmc.usp.br/lacioweb/index.htm>

<sup>56</sup> O CorTec - Corpus Técnico-Científico - é um corpus comparável de textos técnicos e/ou científicos originalmente escritos em português brasileiro e em inglês. Inclui as seguintes áreas: Culinária, Hipertensão, Informática e Instrumentos Contratuais. Disponível em <http://www.fflch.usp.br/dlm/comet>.

usada no *internetês*. Isso é um grande indicativo sobre que tipo de escrita é essa utilizada pelos jovens no Orkut.

Sabendo-se que na escrita o “de” e o “a” são os vocábulos mais empregados, seriam eles também os de maior frequência na fala? Para uma resposta segura, nada melhor do que buscar a informação extraída com ferramenta utilizada pela Lingüística de *Corpus*, a qual examina com isenção *corpora* extensos, com milhões de palavras. Pesquisamos, então, o *subcorpus* falado do Banco do Português e obtivemos como resposta serem “e” (com percentual de ocorrência de 3,73 %) e “que” (3,59%) as palavras mais frequentes na língua falada no Brasil. Já o *corpus* do NURC-RJ nos informa serem “que” (3,71%) e “de” (2,91%) as mais utilizadas na fala culta do Rio de Janeiro.

Partindo da hipótese de que as frequências do léxico do *internetês* revelam uma junção entre fala e escrita, passamos a observar quais palavras estariam presentes nos registros escritos do *internetês* no Orkut. Com a intenção de chamar visualmente a posição das duas palavras mais usadas na escrita e na fala nas listas de frequência dos *subcorpora*, assinalamos em azul a palavra mais frequente na escrita e em vermelho, a mais frequente na fala (do Banco de Português).

**Tabela 2: Palavras mais frequentes dos *subcorpora* Manaus, Belém e Brasília**

(Em azul a palavra mais frequente na escrita, em vermelho a mais frequente na fala)

MANAUS Corp2.txt				BELÉM Corp3.txt				BRASÍLIA Corp4.txt			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
MANAUS				BELÉM				BRASÍLIA			
1	E	892	2,78	1	E	867	2,53	1	E	847	2,59
2	QUE	617	1,92	2	EU	866	2,53	2	QUE	650	1,98
3	Q	574	1,79	3	Q	716	2,09	3	A	642	1,96
4	DE	559	1,74	4	TE	688	2,01	4	DE	604	1,84
5	VC	536	1,67	5	QUE	616	1,80	5	EU	576	1,76
6	EU	526	1,64	6	DE	591	1,73	6	VC	502	1,53
7	O	512	1,59	7	A	590	1,72	7	Q	477	1,46
8	A	502	1,56	8	O	463	1,35	8	O	472	1,44
9	É	470	1,46	9	VC	428	1,25	9	TE	471	1,44
10	TE	402	1,25	10	PRA	357	1,04	10	É	435	1,33

**Tabela 3: Palavras mais freqüentes dos *subcorpora* Cuiabá, Porto Alegre e Recife**

CUIABÁ Corp5.txt				PORTO ALEGRE Corp6.				RECIFE Corp7.txt			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	E	763	2,32	1	E	863	2,60	1	E	787	2,59
2	QUE	705	2,15	2	EU	798	2,41	2	TE	767	2,52
3	DE	612	1,86	3	QUE	784	2,37	3	QUE	753	2,48
4	EU	570	1,74	4	A	671	2,02	4	DE	646	2,13
5	VC	565	1,72	5	DE	553	1,67	5	É	597	1,96
6	É	541	1,65	6	Q	545	1,64	6	EU	517	1,70
7	TE	535	1,63	7	O	475	1,43	7	O	502	1,65
8	Q	508	1,55	8	TU	414	1,25	8	AMO	449	1,48
9	A	492	1,50	9	TE	400	1,21	9	Q	449	1,48
10	O	484	1,47	10	É	382	1,15	10	A	427	1,41

**Tabela 4: Palavras mais freqüentes dos *subcorpora* Rio Branco, Rio de Janeiro e Salvador**

RIO BRANCO Corp8.txt				RIO DE JANEIRO Corp9.txt				SALVADOR Corp10.txt			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	QUE	791	2,45	1	E	887	2,75	1	TE	1.172	3,70
2	E	789	2,44	2	QUE	696	2,16	2	AMO	771	2,44
3	TE	789	2,44	3	EU	663	2,06	3	QUE	703	2,22
4	EU	604	1,87	4	VC	623	1,93	4	E	701	2,22
5	O	586	1,82	5	DE	618	1,92	5	DE	684	2,16
6	DE	570	1,77	6	Q	593	1,84	6	EU	634	2,00
7	A	529	1,64	7	A	590	1,83	7	VC	625	1,98
8	É	523	1,62	8	O	463	1,44	8	A	516	1,63
9	Q	501	1,55	9	TE	446	1,38	9	Q	469	1,48
10	AMO	474	1,47	10	PRA	386	1,20	10	É	423	1,34

**Tabela 5: Frequência *Corpus* de scraps e Frequência *Corpus* reduzido de scraps<sup>57</sup>**

SCRAPS				SCRAPS REDUZIDO			
Corp11.txt				Corp12.txt			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	E	814	2,47	1	E	59	2,49
2	DE	640	1,95	2	DE	56	2,36
3	O	572	1,74	3	VC	46	1,94
4	A	562	1,71	4	O	37	1,56
5	QUE	551	1,68	5	Q	36	1,52
6	VC	458	1,39	6	É	33	1,39
7	EU	424	1,29	7	A	32	1,35
8	É	403	1,23	8	TE	31	1,31
9	Q	327	0,99	9	EU	29	1,22
10	COM	308	0,94	10	QUE	27	1,14

O que podemos observar é que, com exceção dos *subcorpora* Rio Branco e Salvador, em todos eles a palavra mais freqüente foi a conjunção “e”, repetindo na escrita uma característica da fala, a de “emendar” um assunto no outro. Essa situação de agregar idéias, quando por escrito, graficamente é resolvida pela utilização de vírgula ou ponto para evitar o uso repetitivo da conjunção aditiva. Enquanto nos *corpora* escritos o “e” aparece pela quarta ou quinta colocação no *ranking* das palavras mais freqüentes, no do português falado ele aparece em primeiro lugar.

Um outro dado que chama a atenção na observação das palavras mais freqüentes do *internetês* é que a mais usada não seria o conetivo “e”, a mais empregada na língua falada, e sim, a que aparece em segundo lugar na listas de freqüência. Senão vejamos: nas listas aparecem sempre duas formas para indicar uma mesma palavra, o “que” e o “q”. Em todas as listas dos 12 *subcorpora* as duas formas de grafia se fazem presentes, inclusive estando uma subsequente a outra nos *corpora* Manaus, Belém e Brasília.

<sup>57</sup> O *Corpus* reduzido de *scraps* é composto por 2371 *tokens* e 1030 *types*, criado para comparar a riqueza vocabular presente nas reportagens premiadas, examinadas por Belmonte (2007), com a dos *scraps*, na **Seção 5.1 Riqueza vocabular**.

O que devemos levar em consideração nesta pesquisa, cada forma gráfica (representada pelo *type*) ou o sentido (podendo ser representado por vários *tokens*)? Se valorizarmos o sentido da palavra e não a sua forma, o léxico do *internetês* tem como palavra mais freqüente o “que”. Para comprovar isso, montamos mais um quadro estatístico:

**Tabela 6: Soma de QUE e Q na lista de freqüência dos *subcorpora* regionais**

<i>Subcorpus</i>	Freq. de QUE	Freq. de Q	Que + Q TOTAL	Freq. de E
MANAUS	617	574	1191	892
BELÉM	616	716	1332	867
BRASÍLIA	650	477	1127	847
CUIABÁ	705	508	1213	763
PORTO ALEGRE	784	545	1329	863
RECIFE	753	449	1202	787
RIO BRANCO	791	501	1292	789
RIO DE JANEIRO	696	593	1289	887
SALVADOR	703	469	1172	701
SCRAPS APENAS	551	327	878	814

Como observamos na Tabela 6, em todos os *subcorpora* e, por conseqüência, também no *corpus* geral, a freqüência do “que”, quando somadas as formas “que” e “q”, é superior à da palavra “e”. A freqüência dessas duas palavras nos leva a refletir sobre características gramaticais do *internetês*. Pelos estudos de Villavicencio, Finatto e Possamai (2006) sobre a preposição “de”, a palavra de maior freqüência na escrita, o português mostra-se como uma língua de associações. Segundo as autoras, a presença de palavras que operam como “ponte” ou como “cola” entre outras unidades mostra um traço intrínseco ao padrão gramatical do português e pode inclusive caracterizar nossa língua frente a outras. Citam como exemplo o inglês, em que a palavra mais freqüente é o “the”, um determinante, o que as fez concluir, à primeira vista, ser o português uma língua de associações enquanto o inglês é uma língua de determinações.

Sabemos que o “e” também faz associação, é “cola” ou “ponte”, pois une vocábulos ou orações. Realiza conexão ou adição (mais raramente também pode apresentar idéia contrária à

que foi expressa, sendo sinônimo de ‘mas’). Resta-nos, então, para confirmar que nossa língua é de associações, analisar a presença tão frequente de “que” no *internetês*. Afinal, defendemos que esse *internetês* integra, a seu modo, o português brasileiro.

De acordo com os preceitos da gramática normativa, o “que” pode desempenhar até 21 funções na língua. Pode ser desde pronome relativo (um nome significativo numa língua de associações) até interjeição ou partícula expletiva. Na maioria de suas funções, o “que” funciona como conector, estabelecendo nexos, ligação, vínculo, união, ligação entre acontecimentos ou idéias. Na fala as idéias são rapidamente atadas, entrelaçadas ou presas, numa ciranda rápida para acompanhar o pensamento. Isso justifica a presença de tantos “e”, como também deve explicar o porquê de tantos “que”. Isso nos leva a pensar que o *internetês* seja realmente uma escrita oralizada. Vejamos, então, que papéis o “que” pode desempenhar nas frases da escrita culta padrão, recorrendo a exemplos retirados principalmente do dicionário Houaiss (HOUAISS, 2001).

São suas funções:

1. substituir um antecedente, *nome* ou *pronome*, assumindo-lhe as funções sintáticas próprias (p.ex. a função *objeto direto* de 'a melodia' na frase: *a melodia que compôs especialmente para a ocasião foi um sucesso.*)

2. conferir à oração que inicia a função de *adjetivo*: 'que acabaram de construir' equivale a 'recém-construído' na frase: *o bangalô que acabaram de construir receberia novos hóspedes.*

3. ser usado com sentido indefinido em frases que interrogam: *que significa esse rabisco?*

4. indicar indefinição, sinônimo de “quanto”: Que tristeza, meu Deus!

5. expressar pasmo, surpresa, perplexidade, impaciência, contrariedade, raiva, admiração ou pedido de explicação do que se acaba de ouvir: *Quê!?! Infringiste um regulamento secular?*

6. conferir à oração subordinada valores circunstanciais, em numerosas locuções (antecedido de preposições, advérbios ou de participios). 6.1 causa: já que, visto que, pois

que, por que: *Já que as pernas lhe tremiam, sentou-se*; 6.2 finalidade: para que, a fim de que: *Afastaram-se para que outros não os ouvissem*. 6.3 concessão: posto que, ainda que, mesmo que, se bem que: *Ainda que lhe pagassem, jamais comeria carne de cobra*; 6.4 condição: dado que, sem que, desde que, uma vez que: *Desde que preferia a noite ao dia, ofereceram-lhe um jantar*; 6.5 tempo: antes que, assim que, depois que, sempre que: *Sempre que liga a televisão, adormece*; 6.6 proporção: à proporção que, à medida que: *À proporção que as autoridades iam se retirando, os trabalhadores ficavam mais descontraídos*; 6.7 comparação: que nem, mais ... (do) que, menos.... (do) que; *Ficou menos alto (do) que o pai*; 6.8 consequência: tal, tanto, tão, tamanho ... que: *Era tal (tanto, tamanho) o seu entusiasmo que acabou contagiando todos*.

7. conferir à oração subordinada as funções próprias do substantivo, ser conjunção integrante (p.ex., 'fique bem claro' que é *sujeito* na frase: *é necessário que fique bem claro*).

8. ligar vocábulos ou orações do mesmo nível sintático, atribuindo ao termo que inicia diversos sentidos. 8.1 explicação: *Espere um pouco que a chuva já vai parar*; 8.2 adição: *Luta que luta com tenacidade admirável*.

9. exprimir intensidade; correspondendo a “quão”: *Que bela estava aquela noite!*

10. indicar algo indeterminado, indefinido, sinônimo de “alguma coisa”: *Há um quê difícil de definir nessa questão*.

11. Substituir a preposição “de”: *A gente tem que explicar com franqueza certas coisas*.

12. Servir apenas para completar ou realçar: *Olha que esperto que é teu irmão*.

Além dessas doze situações, temos ainda expressões ou locuções nas quais o “que” aparece, como em “que de” (expressa indefinição e/ou partição; quanto: *Que de inexperiência aquelas palavras denunciavam*) e “que tal” (do mesmo gênero; semelhante, similar, parecido: *Amava quadros, livros e coisas que tais*). E temos também o “quê”, acentuado, dificilmente utilizado com acento circunflexo no *internetês*, significando: a) algo indeterminado, indefinido (*Há um quê difícil de definir nessa questão*); b) aquilo que é difícil ou torna uma coisa custosa; complexidade, complicação, dificuldade (*Informática não é difícil, mas tem seus quês*).

Afora esses empregos, há ainda as locuções em que é usado, como em “como quê” (de modo incomparável ou em grande quantidade: Ela declama como quê, comeu-se como quê), “não há de ou por quê” (não existe motivo para agradecimentos, fórmula de cortesia em resposta a agradecimento manifestado por outrem), “sem quê nem pra quê (sem motivo aparente, sem razão, de repente), um quê (alguma coisa, algo: *Ele tem um quê que não se consegue explicar*) e “um não sei quê” (algo indefinido, incerto: *Ela tem um não sei quê que atrai*).

Na rubrica GRAM/USO, o mesmo dicionário antes citado traz mais estas observações sobre o uso de “que”:

- a) em frases exclamativas, é freq. us. como intensificador expletivo: *que fantástico!*; *que maravilha!* b) ainda como expletivo, é corrente no Brasil em frases interrogativas informais, tanto diretas (*o que que você pretende fazer?*) quanto indiretas (*não entendiam o que que ele tinha na cabeça*) (Grifo do autor).

Tendo em vista todas as utilizações nas frases, não é de estranhar que seja uma das palavras mais empregadas no dia-a-dia, principalmente por seu papel de conectivo subordinativo, principalmente em orações adjetivas. No *internetês*, especialmente no nosso *corpus* de depoimentos do Orkut, a presença de orações adjetivas deve ser significativa, devido à especificidade do texto, um registro - normalmente positivo - de aspectos sobre quem se faz o depoimento. Deixando a ambiência do dicionário, vejamos agora alguns empregos do “que” e do “q” no *corpus* geral.

#### Quadro 11: Concordâncias do QUE no *corpus* geral

N	Concordance	Set	Tag	Word No.	File	%
1	___000	aconteça o	<b>que</b> acontecer... passe	499.255	c:\corp1.txt	91
2	004)	xD~: dae garota	<b>que</b> eu AMOOOO sim e	417.611	c:\corp1.txt	76
3	/2005)	: Vc é a pessoa	<b>que</b> mais me odeia com	84.657	c:\corp1.txt	16
4	Juh #):	o único	<b>que</b> me entendi nos meu	422.149	c:\corp1.txt	77
5	rtida, gulosa;	2.Menina	<b>que</b> tem o prazer de rir,	387.259	c:\corp1.txt	71
6	anos, nem a distância		<b>que</b> existe hoje entre a	64.314	c:\corp1.txt	12
7	diiii!	Minha companhia,	<b>que</b> eu adoro tá junto o	456.325	c:\corp1.txt	83
8	noção mais 100	noção	<b>que</b> eu amo nesse mund	227.265	c:\corp1.txt	42
9	fronteiras, e é	*CLARO*	<b>que</b> não estou mentindo	242.951	c:\corp1.txt	45
10	a de pessoas como tu..		<b>que</b> nos faz muito feliz!!!!	353.455	c:\corp1.txt	69

**Quadro12: Concordâncias do Q no *corpus* geral**

N	Concordance	Set	Tag	Word No.	File	%
1	..vc eh...d+	sem contar	<b>q</b> eh 1000d	201.497	c:\corp1.txt	37
2	deixar 1 bjaõ e p/ dizer		<b>q</b> é terminantemente pro	356.281	c:\corp1.txt	65
3	caso assim, naum digo		<b>q</b> a perdi, soh q tomam	387.635	c:\corp1.txt	71
4	1 ano só..quero q saiba		<b>q</b> mesmo não nos vendo	324.743	c:\corp1.txt	59
5	/ 1 pergunta: -Claro + o		<b>q</b> ele tem? O filho com	260.832	c:\corp1.txt	48
6	0 msm meu... Um cara		<b>q</b> gosto muito msm...ge	231.757	c:\corp1.txt	43
7	mto 10.... po nem sei o		<b>q</b> fala dele gosto mtoooo	130.521	c:\corp1.txt	24
8	00%. EM VC JA SABE		<b>Q</b> TE ADORO!QUE E U	228.474	c:\corp1.txt	42
9	+ 10.000 + 1.000.000		<b>q</b> eu conheçu na minha	55.690	c:\corp1.txt	10
10	000000... , pq toda hora		<b>q</b> a gente precisa ela e	197.757	c:\corp1.txt	36

O que percebemos nessas amostras é que no primeiro quadro realmente apareceram várias orações subordinadas adjetivas, o mesmo não ocorrendo no segundo, o das concordâncias com o “q”. Neste último, a incidência maior é de orações subordinadas substantivas, o que nos indica serem diversos os variados empregos desse vocábulo. Vejamos a seguir todos os exemplos do emprego de “que” e “q” no *corpus* reduzido de *scrap*s, composto de recados, sem nenhum depoimento. Nos recados a tendência é reproduzir a fala mais diretamente. Que funções as palavras observadas teriam nesses bilhetes postados no *site* do Orkut?

**Quadro 13: Concordâncias do QUE no *corpus* reduzido de *scrap*s**

N	Concordance	Set	Tag	Word No.	File	%
1	o restrito a ele... espero		<b>que</b> compreenda. abs	412	c:\corp12.txt	17
2	dindo aki so pra te dizer		<b>que</b> te amooooooooooooo	1.507	c:\corp12.txt	64
3	m intão! agora eu quero		<b>que</b> tu vá neh!! brigado	2.313	c:\corp12.txt	98
4	ho de semana... acho		<b>que</b> vai ser dificil levá-lo	545	c:\corp12.txt	22
5	ja coloca o video do edi		<b>que</b> é bom nadaaaaaaaa	764	c:\corp12.txt	31
6	como c ta? Só agora		<b>que</b> vc lembrou...rs Obri	186	c:\corp12.txt	8
7	ano (IDH) mais baixos,		<b>que</b> normalmente não co	1.075	c:\corp12.txt	45
8	é justa ou injusta, mas		<b>que</b> simplesmente é...	715	c:\corp12.txt	29
9	o encontramos alguém		<b>que</b> nos transforme no	832	c:\corp12.txt	34
10	nômicas (Fipe)apontam		<b>que</b> cerca de 12 mil pe	1.146	c:\corp12.txt	48

11	isto ai da maneira com <b>que</b> vc coloca as coisas	319	c:\corp12.txt	13
12	).. ihuul te amo mais <b>que</b> tudo nessa vidah :	1.827	c:\corp12.txt	78
13	mande uma mensagem <b>que</b> te passo meu celula	334	c:\corp12.txt	14
14	erra Negra... nd melhor <b>que</b> uma bela noite de	1.903	c:\corp12.txt	81
15	ce o que há de melhor.. <b>que</b> tu esteja sempre ro	446	c:\corp12.txt	18
16	orkut e te achei... Será <b>que</b> tem problema?	1.233	c:\corp12.txt	52
17	io hehehe bjos hola? <b>que</b> tal? nos falamos ao	639	c:\corp12.txt	26
18	s transforme no melhor <b>que</b> podemos ser." Bo	837	c:\corp12.txt	34
19	bar!!! Me diz uma coisa <b>que</b> gostarias de roubar	1.279	c:\corp12.txt	53
20	a saudade e o saber de <b>que</b> a morte nao é justa	706	c:\corp12.txt	29
21	preee.. tu só merece o <b>que</b> há de melhor.. que	442	c:\corp12.txt	18
22	u celular pro meu email <b>que</b> eu te ligo. Abços!	305	c:\corp12.txt	13
23	<sup>3</sup> Muitaum <sup>123</sup> .....e saiba <b>que</b> pode contar sempre	1.478	c:\corp12.txt	62
24	a faiz sua caminha da <b>que</b> eu fasso a minha	1.840	c:\corp12.txt	78
25	xo de chuva. Desejo <b>que</b> tenhas um ótimo fin	1.192	c:\corp12.txt	50
26	.. pela msg! Abraço! <b>que</b> bonito concerteza m	1.349	c:\corp12.txt	56

#### Quadro 14: Concordâncias do Q no corpus reduzido de scraps

N	Concordance Set	Tag	Word No.	File	%
1	até terça	aew! pow <b>q</b> saudades!! hehe só n	2.187	c:\corp12.txt	93
2	... eu acho ki foi a gabi	<b>q</b> apagou os dela.. num	1.960	c:\corp12.txt	83
3	uitooooo... bom pe achu	<b>q</b> ja invadi geral por hoj	1.668	c:\corp12.txt	71
4	ado. Aqui está td ok!! O	<b>q</b> vc tem feito de bom?	2.146	c:\corp12.txt	91
5	as horas ,e por incrível	<b>q</b> pareça a maioria dos	1.418	c:\corp12.txt	59
6	semana xau beijo O	<b>q</b> vc roubaria de mim?	1.254	c:\corp12.txt	52
7	r meu sleep na portaria	<b>q</b> a Chris queria emprest	207	c:\corp12.txt	9
8	kkkkkkk aneinn sabia	<b>q</b> eu amo de paixão ele	1.869	c:\corp12.txt	80
9	eu melhor amigo, msm	<b>q</b> vc mora longe vc é	1.652	c:\corp12.txt	70
10	mmmmmmmm querida	<b>q</b> q eu fiz dessa vez nene	737	c:\corp12.txt	30
11	o sua. entao nao sabia	<b>q</b> vc era vc. fiquei na d	164	c:\corp12.txt	7
12	ho vc pode ficar esperto	<b>q</b> eu volto de novo viu...	1.693	c:\corp12.txt	72
13	e só no whisk! como	<b>q</b> tao as coisas ai? flw!	2.194	c:\corp12.txt	93
14	o fazer um dia!!! Nussa	<b>q</b> mundo pekeno!! numa	37	c:\corp12.txt	2

15	peee brigada por tudo <b>q</b> vc tem feito por mim	1.528	c:\corp12.txt	65
16	temente... Sem contar <b>q</b> vx me Traz Sorte uah	1.435	c:\corp12.txt	60
17	u@ não sumi não... vc <b>q</b> sumiu q nem ficou gr	1.204	c:\corp12.txt	50
18	as mulheres..... Bom o <b>q</b> posso falar sobre mim	56	c:\corp12.txt	3
19	ovo viu... a unik coisa <b>q</b> eu te peço é q nunk s	1.702	c:\corp12.txt	72
20	... o q importa mesmo é <b>q</b> eu te amo muuuuuuito!	949	c:\corp12.txt	39
21	maioria dos momentos <b>q</b> mais precisei vx tava l	1.424	c:\corp12.txt	59
22	. .. pe so quero te pedir <b>q</b> nunk me deixe viu!?!	1.616	c:\corp12.txt	69
23	ois é... eu quero... será <b>q</b> rola? te amo taaaaa	992	c:\corp12.txt	41
24	nik coisa q eu te peço é <b>q</b> nunk se esqueça de	1.707	c:\corp12.txt	72
25	q nunk me deixe viu!?! <b>q</b> eu tbm nunk vo te de	1.621	c:\corp12.txt	69
26	oraçãozinho.. aproveita <b>q</b> tô boazinha pq é teu	475	c:\corp12.txt	20
27	rs Mas tudo bem... o <b>q</b> importa mesmo é q e	945	c:\corp12.txt	39
28	srs... Mas como tudo o <b>q</b> é bom acaba... precis	886	c:\corp12.txt	36
29	m!! bjim vc vai ter <b>q</b> deixar outro 12.000 rs	1.944	c:\corp12.txt	83
30	sumi não... vc q sumiu <b>q</b> nem ficou grudadinho	1.206	c:\corp12.txt	50
31	sto... :\$ Mas vc sabe <b>q</b> eu te amo muuuuuuito,	853	c:\corp12.txt	35
32	Quanto aos livros, sinto <b>q</b> sou uma pessoa mais	875	c:\corp12.txt	36
33	i tudo bom? ai meu, eh <b>q</b> vc nao tem foto sua.	155	c:\corp12.txt	7
34	c nao "x apenas disse <b>q</b> seus recados sumiram	1.757	c:\corp12.txt	74
35	isitar.... vc é um amigo <b>q</b> todo mundo queria, vc	1.593	c:\corp12.txt	68

O que se percebe, numa análise superficial, é a predominância de orações subordinadas substantivas e adjetivas. Apareceram outras funções do “que”, como conjunções subordinativas adverbiais, pronomes interrogativos e até partículas expletivas como é o caso das frases: *mmmmmm querida q q eu fiz dessa vez nen e só no whisk! como q tao as coisas ai? flw!*

Observando mais atentamente, das 61 ocorrências somadas de QUE e Q no *corpus* reduzido de *scraps*, 38 são de emprego em orações subordinadas substantivas e adjetivas (21 e 17 respectivamente). Há alguns exemplos que parecem ser apenas para reforço de expressão, como se vê nas concordâncias retiradas dos quadros anteriores:

- 6        *como c ta? Só agora **que** vc lembrou...rs Obri*  
8        *é justa ou injusta, mas **que** simplesmente é...*  
26      *.. pela msg! Abraço! **que** bonito concerteza m*

- 14 o fazer um dia!!! Nussa **q** mundo pekeno!! Numa  
 17 u® não sumi não... vc **q** sumiu q nem ficou gr

Os dados obtidos do *corpus* do *internetês* sobre tão elevada freqüência de QUE (e sua variante Q) ratificam, em primeiro lugar, o caráter associativo da língua portuguesa (não esqueçamos que o *internetês* também faz parte da língua), conforme estudos de Villavicencio, Finatto e Possamai (2006). Esse caráter também se confirma ao verificarmos ser a segunda palavra mais freqüente o E, um “associador” ou “adicionador” por excelência. Em segundo lugar, começa a confirmar-se nossa hipótese de haver grande similaridade entre o *internetês* e a língua falada. Como veremos mais adiante, o QUE é a primeira palavra mais usada na fala culta do *corpus* NURC e a segunda mais empregada do *corpus* Banco do Português Falado.

Depois de vermos as principais informações do *Corpus* Geral do Orkut e dos *corpora* regionais e de *scraps*, principalmente as freqüências de palavras, revelando como as mais empregadas QUE e o E, passemos agora à comparação desses dados com as dos *corpora* de referência.

#### 4.3. CORPORA DE REFERÊNCIA

Nosso *corpus* de referência – também conhecido como *corpus* de controle -, nas palavras de Berber Sardinha (2004, p. 97), “funciona como termo de comparação para análise. A sua função é fornecer uma norma com a qual se fará a comparação das freqüências do *corpus* de estudo”.

Também sobre *corpus de referência*, temos a informação dada por Habert et al (1997) de que é aquele concebido para fornecer uma informação em profundidade sobre uma língua. Visa ser suficientemente extenso para representar todas as variedades pertinentes da língua e de seu vocabulário característico, de modo a poder servir de base para gramáticas, dicionários ou para outros interesses confiáveis de usuários em geral. O *corpus* de referência serve para que se realmente individualize um *corpus* de estudo, de modo que nos mostrará suas coincidências e divergências em relação a um todo mais amplo.

Como já explicitado na seção **3. 2 Constituição dos corpora**, utilizaremos como *corpora* de referência uma parte do Banco de Português franqueada ao público e, também, parte das transcrições orais disponibilizadas na Internet pelo Projeto NURC-RJ (Projeto da Norma Urbana Oral Culta do Rio de Janeiro). Lembremos que o Banco de Português é um *corpus* de linguagem geral, sendo a análise de suas frequências de interesse para o entendimento da língua como um todo. Utilizaremos como dados para comparação com o léxico do *internetês* aqueles encontrados nos dois *subcorpora* distintos do Banco do Português, o *subcorpus* escrito e *subcorpus* falado. O primeiro dispõe para consulta 985.092 palavras de diferentes tipos de textos na linguagem culta, enquanto o segundo possibilita acesso a 197.901 palavras de registros falados de aulas (84.910 palavras) e de conversação (112.991 palavras).

A seguir, apresentamos as informações<sup>58</sup> detalhadas da constituição do nosso principal *corpus* de referência, os tipos de texto e o número de palavras que contêm. São dados apenas da parte franqueada ao público do Banco de Português:

- |   |
|---|
| <p>Todo o corpus (1.182.993 palavras)</p> <ul style="list-style-type: none"> <li>• Registro acadêmico (198.652 palavras) <ul style="list-style-type: none"> <li><input type="checkbox"/> Artigos e teses acadêmicas (198.652 palavras)</li> </ul> </li> <li>• Registros de negócios, comerciais e técnicos (386.138 palavras) <ul style="list-style-type: none"> <li><input type="checkbox"/> Cartas comerciais (19.736 palavras)</li> <li><input type="checkbox"/> Cartas de pedido de emprego (14.306 palavras)</li> <li><input type="checkbox"/> Editais (30.829 palavras)</li> <li><input type="checkbox"/> Fax comerciais (16.131 palavras)</li> <li><input type="checkbox"/> Relatórios anuais de negócio (104.255 palavras)</li> <li><input type="checkbox"/> Manuais de informática (200.881 palavras)</li> </ul> </li> <li>• Registros falados (197.901 palavras) <ul style="list-style-type: none"> <li><input type="checkbox"/> Aulas (84.910 palavras)</li> <li><input type="checkbox"/> Conversação (112.991 palavras)</li> </ul> </li> <li>• Registro da imprensa (199.285 palavras) <ul style="list-style-type: none"> <li><input type="checkbox"/> Jornal diário, impresso (199.285 palavras)</li> </ul> </li> <li>• Registro da literatura (201.018 palavras) <ul style="list-style-type: none"> <li><input type="checkbox"/> Literatura de ficção (201.018 palavras)</li> </ul> </li> </ul> |
|---|

<sup>58</sup> Dados disponíveis no site [www2.lael.pucsp.br/corpora/bp/conc/index.html](http://www2.lael.pucsp.br/corpora/bp/conc/index.html), página inicial do Banco de Português. Acesso em 12 maio 2007.

Os tipos de textos que compõem esse *corpus* estão agrupados em cinco categorias. A proporção de cada uma na composição é a seguinte:

- 1) Escrita acadêmica: 16,79%
- 2) Documentos escritos de negócios, comércio e técnica: 32,64%
- 3) Fala: 16,72%
- 4) Imprensa: 16,84%
- 5) Ficção literária: 16,99%

O segundo *corpus* de referência foi formado recolhendo parte das transcrições do acervo oral do Projeto NURC-RJ (Projeto da Norma Urbana Oral Culta do Rio de Janeiro). Trata-se de entrevistas gravadas na década de 90 com informantes com nível superior completo, nascidos no Rio de Janeiro e filhos de pais cariocas. As entrevistas foram gravadas em fitas de áudio, transcritas de acordo com normas previamente definidas. São diálogos entre informante e documentador abordando diferentes temas.

Com base no *corpus* NURC, construímos um conjunto para exame. Para formar nosso *corpus* de 106.956 palavras (*tokens*), utilizamos duas amostras de textos transcritos, a *Amostra Recontato Década de 90*, contendo 11 gravações com os mesmos entrevistados da Amostra dos anos 1970 e a *Amostra Complementar década de 90*, com 16 gravações com novos informantes. As entrevistas foram realizadas de 1992 a 1998 com informantes distribuídos em quatro faixas etárias, a saber: 25 a 35 anos (8 pessoas), 36 a 55 anos (8 pessoas), 56 a 73 anos (8 pessoas) e de 74 a 80 anos (3 pessoas). Vejamos os dados estatísticos do *corpus* que chamamos de *Falado NURC*:

#### Quadro 15: Estatística do *Corpus* Falado NURC-RJ

<i>CORPUS 23. TXT</i>
<i>Bytes</i> 618.802
<i>Tokens</i> 106.956
<i>Types</i> 8.549
<i>Type/Token Ratio</i> 7,99

Antes de passarmos para a seção referente a *Corpus* de Contraste, apresentamos uma breve amostra de comparação entre a frequência escrita do nosso *internetês* do Orkut e a da escrita do Banco do Português. Maior aprofundamento será feito adiante, na **Seção 5.2 Palavras mais freqüentes**.

**Quadro 16: As palavras mais empregadas nos recados/depoimentos e as palavras mais utilizadas do Banco de Português escrito**

FREQUÊNCIA ORKUT				FREQUÊNCIA BANCO DE PORTUGUÊS ESCRITO			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	E	13.930	2,52	1	DE	1.537.460	4,42
2	QUE	11.537	2,08	2	A	1.082.233	3,11
3	EU	10.619	1,92	3	O	1.026.380	2,95
4	A	10.173	1,84	4	E	726.548	2,09
5	DE	9.950	1,80	5	QUE	667.850	1,92
6	Q	9.028	1,63	6	DO	609.521	1,75
7	TE	8.937	1,61	7	DA	545.271	1,57
8	O	8.396	1,52	8	EM	443.567	1,28
9	É	7.855	1,42	9	PARA	353.847	1,02
10	VC	6.347	1,15	10	NO	308.932	0,89

Nesse primeiro contraponto, já percebemos diferenças nas posições das frequências, estando as mais empregadas da escrita “normal” na 4ª e 5ª posição na lista do *internetês*. Curiosamente, as mesmas posições ocorrem com as mais usadas do *internetês* na lista da escrita oficial. Há ainda a registrar a presença do “q” como uma palavra do *internetês* entre as mais freqüentes, e também a existência de palavras que não são comuns às duas listas de palavras. A seguir, as comparações entre a escrita do *internetês* e outros *corpora*, os de contraste.

#### 4.4. CORPORA DE CONTRASTE

Nossos *corpora* de contraste são menores que os *corpora* de estudo e, assim como os *corpora* de referência<sup>59</sup>, são utilizados como recurso de comparação, fornecendo uma norma para comparação das frequências do *corpus* de estudo.

Como *corpus* de contraste se relaciona a *corpus* de referência, vejamos mais uma opinião sobre este último. Para Sinclair e Ball (1996), denomina-se *corpus de referência* um *corpus* suficientemente amplo para representar uma variedade relevante da língua. Deve apresentar-se suficientemente consistente para pesquisas lingüísticas e, dada a sua dimensão e diversidade, pode servir de base para a construção de *corpora* específicos.

Nesta parte do trabalho, rompemos com a idéia de referência e faremos um contraste entre o nosso *corpus* do *internetês* do Orkut e os *corpora* de produções textuais escolares e de textos de Química (texto técnico, reportagens e excerto de manual didático). O objetivo desse contraste é verificar a frequência do vocabulário, o que há em comum entre os *corpora*. No que se refere à escrita escolar, buscaremos também ver se nela haveria presença de elementos do *internetês*, fato recorrentemente citado como um dos maiores “perigos” desse novo tipo de escrita.

Costa (2006, p. 25) afirma que “se formos observar, hoje, manchetes, textos de revistas, propagandas, publicações diversas, etc., certamente encontraremos o estilo *on-line* influenciando a escrita *off-line*”. Como seria, então, tal influência na produção de textos escolares? Conseguiria a escola (leia-se professores) sobreviver a tal poder e ainda ser ouvida no seu trabalho de divulgação/ensino da língua padrão? Essas são questões intrigantes, uma vez que na educação básica encontram-se em maior número os usuários do *internetês*.

Se o estilo *on-line* estiver mesmo influenciando a escrita fora da Internet, acreditamos que essa influência se faria notar na produção de alunos desse nível escolar. Estaria clara para o adolescente a existência de linguagens adequadas somente a determinados ambientes, devendo utilizá-las apenas onde e quando fossem apropriadas? Para observar a presença ou não dessa influência organizamos um pequeno *corpus* com produções textuais de alunos de 7ª série do Ensino Fundamental e do 3º ano do Ensino Médio. Nele temos um total de trinta redações escolares, com amostras na seção de anexos (Anexo IV).

---

<sup>59</sup> O *corpus* de referência, diferente do *corpus* de contraste, deve ser bem maior do que o *corpus* de estudo. O recurso de *corpora* de contraste não é usual na metodologia da Lingüística de *Corpus*. Foi, entretanto, um recurso já utilizado por Azeredo (2007).

Apresentamos a seguir o que observamos no *corpus*<sup>60</sup> de redações de 7ª série, produções de alunos na faixa etária de 12-13 anos. Podemos adiantar que muitas observações, as quais inicialmente poderiam ser creditadas à influência do *internetês*, já apareciam em textos de adolescentes bem antes do surgimento da comunicação *on-line*. Portanto, fazem parte do processo de aquisição da escrita, não sendo necessariamente reflexo do *internetês*. Um exemplo típico é a falta de pontuação, havendo muitas frases siamesas, as idéias são escritas em seqüência sem uso de vírgulas para separá-las. Observemos isso no segundo parágrafo do texto do Aluno 2:

Ninguém saiu “BV” (boca virgem, alguém que nunca beijou na boca). Nem o Jhow e o Tucano, eles são os caras mais envergonhados do estado, também eles beberam mais de 2 barris de cachassa cadaum, sem contar a Gabi F., era amenina mais patricinha de todo o colégio, ela tomou uns 50 litros de cerveja e ficou com o Isaque, ele era (e continua) apaixonado por ela, e com Pedro também apaixonado por ela e o Dudu, eu só sei que foi a festa mais “bala” que eu já fui.

Observemos também a falta de pontuação no parágrafo inicial do texto do Aluno 8:

- Bá meu tu tinha que ter visto a festa, foi muito irado pois no meio da festa faltou luz e ocorreu um assalto a mão armada onde participavam 4 homens mascarados que pareciam aqueles terroristas dos filmes e na vida real também, mas não roubaram nada de mim pois eu estava escondido alias eu tirei umas fotos que posso te mostrar lá em casa ba falando nisso me surgiu uma ideia que tal tu dormir la em casa e daí eu te conto aos poucos e com detalhes a festa mas não tão detalhada porque eu tive de me esconder bom te vejo la em casa mais tarde falou.

Não empregar vírgulas seria apenas um recurso para reproduzir a rapidez da fala ou seria falta de conhecimento do emprego da pontuação, uma vez que adolescentes ainda não teriam completado todo o processo de aquisição da escrita? O mesmo valeria para a acentuação, pouco ou nada empregada na comunicação *on-line*?

Em mais um exemplo, desta vez no texto do Aluno 4, observemos a falta de vírgulas, de ponto final e de acentuação:

- Foi bem assim:  
Tava rolando a festa dele tudo normal tinha varios guris ficando com as gurias a musica rolando D.J. a o normal mesmo la em um salão de festas quando ve!  
- O que? O que?  
- Calma Wagner vo conta  
- a Rodolpho to morto de curiosidade

<sup>60</sup> A formação do pequeno *corpus* de redações de 7ª série não foi aleatória. Agregamos a ele todos os textos com possíveis marcas do *internetês*, encontradas em apenas 12 de um conjunto de 63 produções. Para completar o número estabelecido para o *corpus* (15 produções escolares), acrescentamos os primeiros textos (1 aluna e 2 alunos) redigidos “tradicionalmente”.

- mas tu sabe que eu vo ti conta

- ok, vou ficar quieto.

Quando ve! Entra um cara tri estranho no meio da festa. Ele pegou sacou a arma e deu uns tiros para sima para todos ficarem quietos e ele disse:

- Vamo logo, passando as coisas de valor!

ta ok todos foram assautados mas continuou a festa. Depois por azar nosso deu uma queda de luz.

- Bá Wagner foi aquele azar.

- É Rodolpho deu para perceber.

- E a vida Wagner.

- Ta brigado por me conta. Xau Rodolpho

- Até Wagner

Enfim, queremos deixar claro que nem tudo o que parece escrita do *internetês* tem a ver com *internetês* nos textos escolares. É preciso considerar que há fenômenos advindos do processo de aquisição da escrita, sem relação com tal influência. Vejamos, recolhidas do *corpus* de textos da 7ª série, algumas formas em desacordo com as normas da grafia que já deviam acontecer bem antes do teclar *on-line*: *sô eu*, *Puts*, *achegar* (a chegar), *agente* (a gente), *derrepente*, *mais* (mas), *caio* (caiu), *vio* (viu), *vai enche* (vai encher) e *lucura*.

Mas, claro, há também ocorrências que oferecem dúvidas se são fruto do *internetês* ou se ainda fazem parte da aquisição da escrita na 7ª série. São exemplos: registrar todo o texto em maiúscula, escrever nome de pessoa com letra minúscula, não usar travessão no diálogo, pôr letra minúscula após ponto final ou início de frase e escrever *demanhã!?*, *Nem ti conto*, *qui eu ja fui*, *Ata* (ah, tá), *de pois* e *sobre tudo Preto* (sobretudo preto).

Temos um terceiro grupo de registros com alterações gráficas e de pontuação. Parece-nos que tais alterações, essas sim, têm relação com a maneira de escrever *on-line*, seriam, portanto, influências do *internetês* na escrita formal. Para comprovar isso, extraímos do *corpus* as seguintes ocorrências: presença de emoticons, textos sem pontuação ou acentuação, muitos pontos de exclamação e reticências, pontos de exclamação e de interrogação juntos, falta de travessão no registro de diálogo, letra maiúscula sem razão no meio da frase (*festa muito "Bala"*), justaposição ou aglutinação de palavras (*cadaum* e *colé meu* (qual é, meu)) e grafias típicas do *internetês*, como *mew*, *xau* (tchau), *umonte* (um monte), *calmaaa*, *"UAUU! Adoro esta, vamo dança?"*<sup>61</sup>, *bah*, *Báh*, *"dai ja vio ne"*, *"Eai dudu."* (E aí, Dudu?), *hahaha*, *"Báá!"*, *"Ai! Queodio!"* (Que ódio!), *portoens* (portões) e *"Ahf!"*.

<sup>61</sup> Indicamos frases ou expressões com pontuação entre aspas, para não confundir a pontuação original do aluno com a que empregamos para separar os exemplos.



Observemos agora mais detalhadamente o texto do Aluno 2, que chama a atenção por tentar reproduzir graficamente recursos do *internetês*. Inicialmente chama a atenção as palavras desnecessariamente destacadas para a escrita, um típico recurso para reproduzir a ênfase dada na fala. Talvez pretendendo deixar sua marca como sujeito no texto (lembramos a enunciação de Benveniste), o aluno recorre também aos símbolos das histórias em quadrinhos para substituir um palavrão.

Fiesta  
 Após uma SUPER FESTA (BOCA-LIVRE) que meu amigo Adauto não vai, me encontrei com ele no bar do Julio para lhe falar sobre a festa.  
 - Bã meu tu tinha que tá lá, tinha tudo que tu pode imaginar gente rica, pobre, travessa, balbum, traficante e policial. O sor boquiaberto falou:  
 - e tinha muita comida? - Falou Adauto  
 - tinha tudo que possa se imaginar e em quantia, muito grande.  
 - A que #000000. Por que eu não fui ia me divertir bastante! - Adauto reclamou  
 - Mas táva muito parada na iniciação, mas depois teve muita dança e diversão. - Tu balei.

Vemos aqui também a falta do emprego de vírgula em *A que* (Ah, que..) e também ausência de vírgula e ponto de interrogação (*Por que eu não fui ia me divertir bastante.*). Fica a dúvida: esse aluno estaria reproduzindo uma influência do *internetês* ou teria esquecido ou ainda não teria internalizado o emprego da pontuação?

Um parágrafo do mesmo texto mostra claramente a influência do teclar *on-line* ao destacar em maiúscula e sublinha uma palavra que quer enfatizar:

- Da próxima vez tu me liga e nós vamos na festa no meu carro, aliás, como tu vai pra festa?  
 - Eu fui de BUSÃO pra lá até a metade do caminho, depois os ônibus pifou e nós todos (passageiros) desceram e fomos apé até a parada mas como estava perto da festa fui a pé.

Possibilidade de influência do *internetês* é o registro de *apé* e falta de ponto ou vírgula após *caminho*. Percebemos que toda a fala da personagem é típica da linguagem oral. O autor “jogou” as idéias em seqüência, num único período composto pretendendo transmitir rapidamente as idéias e o sentimento da personagem ao reviver uma experiência. Há presença de “e”, típico da fala, substituindo vírgula na escrita. Ainda sobre pontuação, podemos observar a despedida, em que numa frase as palavras são simplesmente registradas como na fala, já encerrando o diálogo. Curiosa foi a forma encontrada para o último ponto final, escrito de forma ampliada, como a fechar definitivamente o texto.

- Interessante - Falou Adauto  
 - Bom Tchau Adauto  
 - Tchau.

Relacionado também à pontuação como um indicador de ênfase, a Aluna 15 pretendeu registrar na pontuação do título toda a intensidade que queria transmitir sobre o assunto do texto. Observemos o ponto de exclamação criado pela autora, algo admitido no *internetês* ou no *miguxês*, talvez mais ainda neste último, em que o adolescente utiliza formas para estabelecer diferenças no próprio grupo, busca ser único no grupo, servindo-se da linguagem para se fazer sujeito. Vejamos um ponto de exclamação “estrelado” criado pela aluna:

Que Festa!

Do que observamos até aqui sobre a influência do *internetês* na escrita escolar, podemos chegar a, no mínimo, duas conclusões:

- Há mitificação e exagero ao afirmar que alunos utilizam *internetês* na escola. Nossos dados indicam que há alguns traços, sim, mas não em número significativo, nem empregados por todos os alunos. Às vezes é um “tb” (= também) ou “p/” (= para) utilizados no texto, elementos que provavelmente quem reclama também já escreveu numa situação apressada. O problema está em o aluno escrever ainda tais formas num texto formal. Aí entra o papel do professor em chamar a atenção para tal inadequação.

- Muito do que aparece como possível influência do *internetês* nada mais é do que o processo de aquisição da escrita. São ocorrências que já se manifestavam nas escritas escolares bem antes do teclar *on-line*. Quantos alunos do Ensino Fundamental (e até do Ensino Médio) tiveram dúvidas se escreveriam, por exemplo, *oque, derrepente, emcima, de baixo, sempri e porexemplo?* Professores mais experientes lembram de casos curiosos de pré-adolescentes escrevendo, entre outras, *fachetária* e *jornau nacionau* para indicar *faixa etária* e *Jornal Nacional*. Há, então, que relativizar essa atribuição à Internet pela responsabilidade dos erros ortográficos. Seria interessante fazer estudo sobre a ortografia empregada por alunos sem acesso à Internet para compará-la com a dos alunos de mesmo nível de escolaridade usuários da rede. Teríamos, então, dados concretos do quanto certas ocorrências são inerentes ao processo de aquisição da escrita oficial.

Já nas observações dos textos produzidos por alunos do 3º ano do Ensino Médio, poucos registros possíveis referentes ao *internetês* chamaram a atenção. Apareceram apenas uma vez as formas *aqela, Bah, Ahhhh, “HaHaHaHa!!!”* e *Aham* (expressão correspondente a *sim*). Quanto à acentuação, percebemos falta de acento em muitas palavras que deveriam ser acentuadas. Relacionado à pontuação percebemos no *corpus* pesquisado algumas ocorrências do uso de dois pontos de exclamação juntos (!!), vírgula depois de reticências, vocativo sem vírgula, uso de mais de uma exclamação ou interrogação e reticências desnecessárias, como em “*Ele gosto de mim...!!*”, “*Mentira?!?*”, “*Capaz?!?*”, “*Teve até polícia no final!!!*” e “*E o resto???*”. Estariam os pontos de exclamação e de interrogação enfraquecendo o seu valor, a ponto de não bastar um deles no final da frase? Por que é necessário escrever dois, três ou mais, a não ser para dar mais ênfase, como nas conversas *on-line*?

Um que outro parágrafo, mesmo em textos produzidos ao término do Ensino Médio, são escritos sem usar a pontuação adequada, tentando reproduzir a rapidez da fala. O curioso no

exemplo que transcrevemos a seguir é que se trata da introdução do texto, não é parte do diálogo entre personagens:

Era 13:30 da tarde, eu estava acordando, me arrumei e fui comer algo. De repente a campanha começa a tocar “afobadamente”, fui ver quem era, abri a porta, era obvio que era a Laurita – Se enrolando toda de tão rápido e ansiosa que estava falando, ela me perguntou:  
 - Oi Tetê, tudo bem?  
 - Oi Laurita, é... estou um pouco triste. E você como está?

Assim, **pelos dados observados, percebemos não ser muito significativa a influência do *internetês* nos textos escolares.** É bastante compreensível que na 7ª série, em relação ao 3º ano do Ensino Médio, haja maior interferência da escrita *on-line*, por uma série de razões. Poderíamos citar algumas, tais como menor tempo de contato com as normas da língua padrão, alunos com faixa etária ainda em processo de aquisição da escrita e necessidade de utilizar a linguagem diferenciada como participante de determinado grupo. De um lado, é preciso ter cautela ao reconhecer a influência do *internetês* em meio à aquisição da escrita entre os mais jovens.

As produções textuais analisadas foram realizadas em julho, período do ano letivo em que os professores já teriam detectado possíveis reproduções da linguagem *on-line*. Perguntamos, via *e-mail*, ao professor<sup>62</sup> responsável pelas turmas sobre a presença do *internetês* nos textos ao longo do semestre e se os professores chamavam a atenção sobre isso, orientando os alunos para redigirem seguindo as normas da gramática normativa. Pela resposta obtida, a seguir transcrita, podemos entender por que há baixa presença do estilo *on-line* nos textos escolares observados:

Com relação às redações, eu tenho sempre informado os alunos sobre a diferença entre escrever textos no ambiente de internet e na sala de aula. Uso um exemplo para ilustrar isso. Digo a eles que não vamos à praia de "terno e gravata", assim como não pedimos "por obséquio" no MSN. Dessa forma, fica claro para eles que o *internetês* pode e deve ser usado no ambiente virtual; nas demais produções textuais, entretanto, tal linguagem é inadequada. Na sétima série ainda há, embora pouca, ocorrência do *internetês*; no terceiro ano, acredito ser rara tal ocorrência. Ao longo do ano, na medida em que foram aparecendo tais ocorrências, eu (e acredito que os demais professores) chamamos a atenção para esse aspecto.

<sup>62</sup> Adauto Locatelli Taufer, professor de Língua Portuguesa do Colégio de Aplicação da UFRGS, responsável por duas turmas de 7ª série e pelas aulas de redação em uma turma de 3º ano do Ensino Médio. Identificamos o professor e publicamos a resposta à questão formulada com sua autorização por escrito.

Cabe salientar que, no Colégio de Aplicação da UFRGS, no qual foram recolhidas as redações para constituição do *corpus*, há o hábito, já instituído há alguns anos, de solicitar semanalmente produção escrita de texto aos alunos a partir de 7ª série. Para isso, destinava-se um período da carga horária de Língua Portuguesa. No ano de 2007 foi implantado no Ensino Médio o chamado *Projeto de Redação*, em que todas as séries, num mesmo horário, trabalham produção de texto durante dois períodos, realizando as mais diferentes propostas. Assim, todos os alunos de uma série realizam o mesmo tipo de atividade proposto para ela. É uma alternativa encontrada pelos professores, esperando obter resultados melhores nas produções escritas de todos os alunos da escola.

Consideramos o papel da escola (com seus professores) de fundamental importância para que o adolescente adquira os mecanismos para a elaboração adequada de textos segundo a norma padrão, aquela exigida pela sociedade, porque para escrever em *internetês* não é necessário, aparentemente, grandes exercícios de aprendizagem. Lembremos aqui, oportunamente, Dacanal (2006, p. 24), para quem a função do professor “é, precisamente, a de ensinar seus alunos a dominarem da melhor forma possível a convenção, para poderem assim utilizá-la no futuro como instrumento de trabalho e como ganha-pão”. Porque “na prática, o que ocorre é que esta convenção que se chama *língua* se torna uma imposição. A língua é uma imposição histórica e social, apesar de funcionar tecnicamente como uma convenção.” (DACANAL, 2006, p. 78). Essencial, portanto, proporcionar aos alunos o desenvolvimento de habilidades para utilizarem em seu benefício e da melhor forma essa convenção imposta, conseguindo discernir onde e quando podem empregar suas variantes, como nos parece ser o *internetês*.

Passemos agora a observar outros dados sobre o léxico, obtidos nos *corpora* das produções escolares. Seria o vocabulário empregado nas redações similar ao utilizado na comunicação *on-line* ou estaria mais próximo ao da norma padrão? Ao que parece, é semelhante. Vejamos os dados.

**Quadro 18: Estatística dos *corpora* de produções dos alunos de 7ª série e 3º ano do EM**

REDAÇÕES 7ª SÉRIE	REDAÇÕES 3º ANO EM	REDAÇÕES SOMADAS
<i>Tokens</i> 3.357	<i>Tokens</i> 3.514	<i>Tokens</i> 6.871
<i>Types</i> 1.016	<i>Types</i> 1.025	<i>Types</i> 1.675
<i>Type/Token Ratio</i> 30,27	<i>Type/Token Ratio</i> 29,17	<i>Type/Token Ratio</i> 24,38

Inicialmente, vale comentar que os dois *corpora* de redações, no seu todo, parecem bastante semelhantes em termos de número de palavras e até em termos de riqueza lexical.

Vejamos agora quais são as palavras mais freqüentes nesse universo de textos:

**Tabela 7: Palavras mais freqüentes nos *corpora* de produções dos alunos de 7ª série e 3º ano do Ensino Médio**

REDAÇÕES 7ª SÉRIE				REDAÇÕES 3º ANO EM				TOTAL DE REDAÇÕES			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	E	130	3,87	1	QUE	145	4,13	1	QUE	256	3,73
2	A	117	3,49	2	A	121	3,44	2	A	238	3,46
3	QUE	111	3,31	3	O	100	2,85	3	E	227	3,30
4	EU	90	2,68	4	E	97	2,76	4	O	186	2,71
5	O	86	2,56	5	EU	72	2,05	5	EU	162	2,36
6	FESTA	84	2,50	6	NÃO	70	1,99	6	FESTA	149	2,17
7	DE	61	1,82	7	DE	67	1,91	7	DE	128	1,86
8	COM	45	1,34	8	FESTA	65	1,85	8	NÃO	115	1,67
9	FOI	45	1,34	9	PARA	50	1,42	9	FOI	88	1,28
10	NÃO	45	1,34	10	ESTAVA	43	1,22	10	COM	84	1,22

Curiosa é a posição ocupada pela palavra DE (7ª), uma vez que, já referimos, na escrita usualmente aparece em primeiro lugar. Seis palavras da escrita escolar também aparecem como as primeiras do internetês (E, QUE, EU, DE, A, O) e cinco também no português escrito (QUE, A, E, O, DE). Estão entre as dez mais freqüentes das redações também FESTA, NÃO, FOI e COM, relacionadas, certamente, à proposta solicitada para produção de texto, uma narrativa. Podemos ver no Quadro 19 essas freqüências relacionadas:

**Quadro 19: Relação de freqüências de palavras na escrita escolar, *internetês* e norma culta escrita (Banco de Português)**

<i>INTERNETÊS</i> GERAL DO ORKUT				ESCRITA ESCOLAR TOTAL DE REDAÇÕES				ESCRITA BCO DE PORTUGUÊS			
Nº	Palavra	Freq.	%	N	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	E	13.930	2,52	1	QUE	256	3,73	1	DE	1.537.460	4,42
2	QUE	11.537	2,08	2	A	238	3,46	2	A	1.082.233	3,11
3	EU	10.619	1,92	3	E	227	3,30	3	O	1.026.380	2,95
4	A	10.173	1,84	4	O	186	2,71	4	E	726.548	2,09
5	DE	9.950	1,80	5	EU	162	2,36	5	QUE	667.850	1,92
6	Q	9.028	1,63	6	FESTA	149	2,17	6	DO	609.521	1,75
7	TE	8.937	1,61	7	DE	128	1,86	7	DA	545.271	1,57
8	O	8.396	1,52	8	NÃO	115	1,67	8	EM	443.567	1,28
9	É	7.855	1,42	9	FOI	88	1,28	9	PARA	353.847	1,02
10	VC	6.347	1,15	10	COM	84		10	NO	308.932	0,89

Quanto à freqüência das palavras, observamos uma tendência de a escrita escolar estar mais próxima da freqüência do *internetês*, uma vez que o QUE aparece como a palavra mais usada em ambos os *corpora*, o E também estar bem colocado no *ranking* e o DE, em ambos, aparecer distanciado da posição de destaque no *corpus* Banco do Português escrito.

Passamos agora a um contraponto entre as palavras mais freqüentes da fala culta nos nossos dois *corpora* de referência, o Português Falado Projeto NURC-RJ e o Banco de Português Falado.

**Quadro 20: Relação fala culta NURC-RJ, escrita escolar e fala culta Banco de Português**

FREQUÊNCIA PORTUGUÊS FALADO PROJETO NURC				ESCRITA ESCOLAR TOTAL DE REDAÇÕES				FREQUÊNCIA BANCO DE PORTUGUÊS FALADO			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	QUE	3.973	3,71	1	QUE	256	3,73	1	E	113.061	3,73
2	DE	3.109	2,91	2	A	238	3,46	2	QUE	108.883	3,59
3	EU	2.885	2,70	3	E	227	3,30	3	A	77.882	2,57
4	NÃO	2.739	2,56	4	O	186	2,71	4	É	75.609	2,49
5	É	2.566	2,40	5	EU	162	2,36	5	O	71.329	2,35
6	A	2.534	2,37	6	FESTA	149	2,17	6	DE	66.922	2,27
7	E	2.422	2,26	7	DE	128	1,86	7	N É	64.870	2,14
8	O	2.300	2,15	8	NÃO	115	1,67	8	NÃO	62.445	2,06
9	NÉ	1.439	1,35	9	FOI	88	1,28	9	EU	55.733	1,84
10	UM	1.349	1,26	10	COM	84		10	F	45.235	1,49

A presença do “não” em todos os *corpora* do Quadro 20 até a 8ª posição é significativa. Lembremos que ele não apareceu entre as mais frequentes do *internetês* e do Banco do Português escrito e, como podemos ver adiante, em nenhum dos outros *corpora* de contraste relacionados à Química. Podemos deduzir que, nesse tipo de escrita escolar, há mais similaridade com a fala do que com a escrita no que tange à frequência das palavras. Como o *internetês* também possui traços da fala, estaria ele, então, também próximo da escrita escolar, ou vice-versa em termos das palavras mais frequentes.

O Quadro 21, a seguir, traz estatísticas e as palavras mais frequentes dos nossos outros três *corpora* de contraste, todos relacionados a temas de Química. Esses *corpora* pertencem ao Projeto TextQuim, conforme já explicitado na Seção 3.2 Constituição dos Corpora. O primeiro é composto por reportagens da revista *Superinteressante* sobre assuntos de Química, o segundo, por textos de livros didáticos de Química para nível universitário e o terceiro, por texto que aborda o tema entalpia. O que veremos é um contraponto entre as palavras mais frequentes em *corpora* de textos de ciências e as palavras do *internetês*. Haveria pontos em comum entre a frequência lexical de textos voltados a temas específicos com a do *internetês*? Vejamos, então.

**Quadro 21: Dados estatísticos e palavras mais freqüentes dos *Corpora* Superinteressante, Química Geral e Entalpia**

Textos da revista <i>Superinteressante</i>				Textos sobre Química Geral – texto didático universitário				Texto Entalpia - Recorte de texto sobre entalpia			
Tokens: 37.395 Types: 7.317 Type/Token Ratio: 19,54				Tokens: 526.354 Types: 10.553 Type/Token Ratio: 20,04				Tokens: 470 Types: 168 Type/Token Ratio: 35,74			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	DE	1.879	5,02	1	DE	29.772	5,66	1	A	27	5,74
2	A	1.306	3,49	2	A	21.142	4,02	2	DE	26	5,53
3	O	1.142	3,05	3	O	12.073	2,29	3	É	12	2,55
4	QUE	874	2,34	4	É	10.492	1,99	4	E	11	2,34
5	E	675	1,81	5	QUE	10.140	1,93	5	H	11	2,34
6	DO	574	1,53	6	E	10.137	1,93	6	QUE	11	2,34
7	EM	548	1,47	7	UM	8.186	1,56	7	ENTALPIA	10	2,13
8	DA	503	1,35	8	PARA	7.590	1,44	8	O	10	2,13
9	SE	496	1,33	9	UMA	7.089	1,35	9	SISTEMA	10	2,13
10	OS	440	1,18	10	DO	6.421	1,22	10	U	8	1,70

Da comparação entre esses três *corpora*, há pouco explicitados, e o nosso *corpus* do *internetês* do Orkut obtivemos as constatações a seguir. Vale lembrar que os *corpora* contrastados com o do Orkut são esses que abordam temas de Química e o *corpus* das redações escolares.

- 1) A palavra DE aparece como a mais freqüente (em 2º lugar no *corpus* Entalpia, muito próxima da primeira colocada, com 26 e 27 ocorrências, respectivamente), confirmando essa preposição como a palavra mais usada na escrita, enquanto no *internetês* o DE aparece como a 5ª mais freqüente.
- 2) O E, diferentemente da sua 2ª colocação no *internetês*, está em 5º lugar nos textos escritos sobre Química (no *corpus* Entalpia aparece com a mesma freqüência de H e QUE, empatados em 4º lugar, com 11 ocorrências cada).
- 3) O QUE (com sua variante Q no *internetês*) está na posição 4 ou 5 nos textos escritos de acordo com a norma oficial.

- 4) Nos textos denotativos tratando sobre Química, as palavras EU, TE e VOCÊ (VC), ao contrário da lista do *internetês*, não constam no *ranking* das mais freqüentes.
- 5) Os artigos A e O aparecem como 2º e 3º colocados nos *corpora* de reportagens da revista *Superinteressante* e de Química Geral, sendo A o 1º no *corpus* Entalpia (embora com diferença mínima para o DE, como anteriormente vimos).
- 6) O verbo É só não aparece entre as dez palavras mais freqüentes do léxico de reportagens sobre assuntos químicos.
- 7) Mesmo no peculiar *corpus* sobre entalpia observamos 6 palavras freqüentes em praticamente todos os *corpora* estudados neste trabalho (A, DE, É, E, QUE, O). As especificidades técnicas apareceram nas outras 4 palavras mais freqüentes (H, ENTALPIA, SISTEMA, U). Resta saber se o O que aparece não representa também o símbolo do oxigênio, como o H é o do hidrogênio e de entalpia, entre outros<sup>63</sup>.

Das observações feitas, em termos gerais, podemos afirmar que, quanto à freqüência, há diferença entre o léxico do *internetês* e o das linguagens escritas, sejam elas mais “comuns”, como as registradas no Banco de Português, sejam mais “especializadas”, como as dos textos abordando Química em manuais didáticos ou reportagens. A aproximação da freqüência, mais uma vez, não a percebemos tão fortemente com a da língua escrita, mas sim, com a da língua falada. Assim, como podemos ver mais adiante nos Quadros 27 e 33, relacionado às dez palavras mais freqüentes, percebemos que:

- 1) O DE aparece em 5º lugar no *corpus* do *internetês* e em 6º no do Banco de Português falado (nos *corpora* de escrita está em 1º lugar);
- 2) O E é a palavra mais freqüente no Banco Falado de Português e a 2ª no Orkut (na escrita está em 4º, 5º e 6º lugares);
- 3) O QUE é a palavra mais freqüente no *internetês* e no NURC, e é a 2ª no Banco de Português (nos *corpora* de escrita está em 4º, 5º e 6º lugares, alternando com o E);
- 4) A palavra EU aparece nos três *corpora* (fala e *internetês*) e em nenhum dos de escrita;
- 5) A palavra A na escrita é a 1ª ou 2ª mais freqüente, enquanto na fala e no *internetês* é a 3ª, 4ª e 6ª colocada;

<sup>63</sup> De acordo com o Dicionário Houaiss (2001), H:

*s.m.* MÚS a nota si, na notação alfabética alemã □ 1 ELETR FÍS *símb.* de **henry** 2 FÍS *símb.* de **entalpia** 3 FÍS.PART *símb.* de **higgs** 4 QUÍM *símb.* de **hidrogênio**

6) O verbo *É* aparece nos *corpora* de fala e *internetês*, e apenas no *corpus* de Química (manual didático e seu recorte Entalpia);

7) Nos *corpora* de fala e *internetês* há sete palavras comuns (E, QUE, DE, EU, É, O, A) enquanto nos de escrita e *internetês* há cinco (E, QUE, DE, O, A).

Essas frequências são mais um indício que nos leva à confirmação de que o *internetês* tem traços marcantes de fala.

No capítulo seguinte, faremos uma análise dos resultados descritos com algumas considerações qualitativas. Trataremos de riqueza vocabular, das palavras mais frequentes, das variações ortográficas, dos elementos de oralização no *internetês* e das variedades de forma e conteúdo.

## 5 REGULARIDADES E ESPECIFICIDADES NOS DADOS SOB EXAME

### 5.1. RIQUEZA VOCABULAR

A primeira constatação que nos chamou a atenção ao ver os números que o *Wordsmith Tools* nos forneceu é que há variação da riqueza vocabular conforme as diferentes regiões do Brasil, no tocante aos *corpora* que envolvem depoimentos e recados no Orkut. Já no *corpus* composto apenas de recados (*scraps*), o emprego de diferentes vocábulos é bem maior do que em qualquer outro *corpus*. Há, portanto, maior riqueza vocabular quando se considera somente o léxico na escrita de recados.

Os depoimentos são um tipo de texto de cunho pessoal, em que se destacam as características da pessoa sobre quem se escreve. Talvez por isso haja certa uniformidade de palavras ou um vocabulário mais restrito. Na situação dos recados, os jovens internautas abordam os mais diversos assuntos, o que provavelmente acaba se refletindo na diversidade das palavras.

Sobre a riqueza vocabular nas diferentes regiões, os dados indicam que em Manaus e em Brasília há maior emprego de palavras diferentes nos textos. Belém, Porto Alegre, Recife e Rio Branco, entretanto, são os locais onde a interação pelo Orkut tem menor variedade de palavras. Podemos fazer essas afirmações observando a proporção (*ratio*) entre as palavras diferentes (*types*) e o total de palavras (*tokens*) que compõe o *corpus*. Quanto maior for o número de *types* maior a variedade. Observemos na Tabela 7, as informações estatísticas dos onze *corpora* examinados. Os dados sobre a riqueza (*type/token/ratio*) estão em ordem decrescente:

**Tabela 8: Tokens, types e ratio por corpora do internetês**

<i>CORPUS</i>	<i>TOKENS</i> (itens)	<i>TYPES</i> (palavras distintas)	<i>TYPE /TOKEN RATIO</i> (Riqueza vocabular)	RELAÇÃO MÁXIMA
<i>SCRAPS APENAS</i>	32.895	6.653	<b><u>20,22</u></b>	100,00
BRASÍLIA*	32.761	5.956	<b><u>18,18</u></b>	100,00
MANAUS*	32.119	5.785	<b>18,01</b>	100,00
CUIABÁ*	32.832	5.852	<b>17,82</b>	100,00
RIO DE JANEIRO*	32.251	5.744	<b>17,81</b>	100,00
SALVADOR*	31.643	5.601	<b>17,10</b>	100,00
RECIFE*	30.385	5.006	<b><u>16,48</u></b>	100,00
RIO BRANCO*	32.282	5.282	<b><u>16,36</u></b>	100,00
PORTO ALEGRE*	33.140	5.389	<b><u>16,26</u></b>	100,00
BELÉM*	34.220	5.392	<b><u>15,76</u></b>	100,00
<i>CORPUS</i> GERAL DO ORKUT	553.875	38.614	<b><u>7,01</u></b>	100,00

\* Inclui *scraps* e depoimentos

Como o leitor deve ter percebido na Tabela 8, os dados sobre riqueza vocabular, quando tomados num pequeno segmento, demonstram um grau determinado de variação do vocabulário. Entretanto, a observação do *Corpus Geral do Orkut*, na última linha da Tabela 7, ao exibir menor proporção *type/token*, mostra que, vendo-se no todo, a riqueza vocabular é muito baixa. Talvez apenas essa visão de todo, associada a uma pobreza vocabular seja o que tem movido os críticos contra o uso da língua da Internet. Falta-lhes, provavelmente, a percepção mais detalhada. Tal perspectiva, do detalhamento e da extensão, só se consegue com a utilização das ferramentas e dos princípios da Linguística de *Corpus*. Assim, vemos que a observação da língua, em amplitude, pode não coincidir com a observação em seus segmentos.

Na prática, segundo Berber Sardinha (2004, p. 94), a proporção entre *types* e *tokens*, que o autor chama de “a razão forma/item”, indica a riqueza lexical de um texto. Quanto maior a razão, mais palavras distintas há no texto. Essa razão (*type/token ratio*) é a percentagem de

palavras diferentes (formas, vocábulos) presentes no total de palavras (itens, ocorrências). Ela é obtida dividindo-se o total de *types* pelo total de *tokens* dividido por cem. Assim, por exemplo, para obtermos a razão *type/token* no corpus *Scraps*, dividimos 6.653 por 32.895 anteriormente dividido por 100, resultando em 20,22 a percentagem de palavras não repetidas em tal *corpus*. Do ponto de vista do vocabulário, um texto será menos rico ou variado quanto menor for esse número, porque maiores serão as repetições de palavras. Claro está que, quanto maior for esse texto, maior será a possibilidade de muitas palavras serem usadas repetidamente. É o que acabamos de comprovar pela observação dos *corpora*, quanto maior seu número de *tokens*, menor a proporção *type/token*. É impossível se comunicar utilizando infinito número de palavras sem que sejam, em algum grau, repetidas.

Podemos agora, com o auxílio do Quadro 22, comparar a riqueza lexical do *internetês* com a riqueza exibida em um *corpus* de língua falada, no caso, o do Projeto NURC (Projeto de Estudo da Norma Lingüística Urbana Culta).

**Quadro 22: Tokens, types e ratio do corpus do português falado NURC-RJ**

<i>CORPUS</i>	<i>TOKENS</i> (itens)	<i>TYPES</i> (palavras distintas)	<i>TYPE /TOKEN RATIO</i> (Riqueza vocabular)	RELAÇÃO MÁXIMA
FALADO NURC	106.956	8.549	7,99	100,00

O que percebemos é uma riqueza lexical da língua falada próxima à do *internetês* (7,99 NURC-RJ/ 7,01 *Internetês*). Mesmo levando em consideração a grande diferença do número total de *tokens* dos dois *corpora*, a proporção final é equivalente. Senão vejamos,

se 553.875 *tokens* (do *internetês*) correspondem a 106.956 (do Falado NURC), então 38.614 *types* (do *internetês*) correspondem a x (no Falado NURC).

O resultado obtido é 7.456 *types* (próximo aos 8.549 indicados no NURC), o que dá uma riqueza lexical de 6,97, praticamente o mesmo número de 7,01 presente no *internetês*. Temos, assim, uma observação que ajusta em proporção os dois escores, dado que os *corpora* têm tamanhos diferentes. Isso revela, mais uma vez, uma aproximação do *internetês* com a língua falada, desta vez relacionada à riqueza lexical. Vale destacar que essa similaridade se dá em relação à riqueza extraída de um *corpus somente de fala culta*. Tais dados contradizem, portanto, a opinião dos “defensores” da língua sobre a pobreza da linguagem dos jovens. Há que relevar, porém, que os *types* do *internetês* podem conter formas diferentes para um mesmo vocábulo, como *muito/mto/mtu*. Mesmo assim, como veremos adiante, tais variações gráficas

de forma não são em número tão elevado a ponto de dominar o *internetês*. A maioria expressiva das palavras utilizadas no Orkut pelos jovens são as registradas no VOLP (Vocabulário Ortográfico da Língua Portuguesa), só que acompanhadas por repetidas variantes típicas, o que causa estranhamento e leva à generalização de ser uma linguagem totalmente modificada.

Sobre a repetição de palavras, os manuais de redação (principalmente para jornalistas) recomendam evitá-la.. O Manual de Estilo da Editora Abril (1990), por exemplo, chega a recomendar que em um parágrafo não seja usado o mesmo substantivo e nem o mesmo verbo. Já Belmonte (2007, p. 4), afirma que “a repetição de palavras tem uma função coesiva quando ajuda a dar unidade e a fazer fluir um texto. Mas também é verdade que a repetição de palavras em demasia ‘suja’ uma reportagem ao empobrecer o vocabulário”.

Para observar a riqueza lexical (razão vocábulo/ocorrência), em textos jornalísticos considerados de excelência, o mesmo autor examinou quatro reportagens sobre biodiversidade. Todas foram premiadas em primeiro lugar em um concurso. Trata-se, assim, de textos consideradas referência de qualidade no meio jornalístico. Os dados de Belmonte (2007) mostraram que a palavra indicadora do tema teve repetição em redor de 1,4 % e que a riqueza lexical ficou ao redor de 40%. Esse número é praticamente o dobro da riqueza vocabular que observamos no *corpus* de *scraps*. Há, porém, uma curiosidade estatística que precisamos considerar. Cada reportagem compôs um *corpus* de aproximadamente 2.500 *tokens*, enquanto o nosso *corpus* de *scraps* tem 30.000 *tokens*. Se considerarmos a mesma proporção, realizando uma simples regra de três, então a riqueza lexical dos *scraps* seria de 48%. Para esclarecer essa questão, resolvemos reduzir nosso *corpus* para haver paridade de palavras entre os *corpora* de *scraps* e de reportagens. Para isso, fomos aleatoriamente deletando páginas de *scraps* até chegar ao número de palavras (*tokens*) que desejávamos. Tendo composto um *corpus* reduzido com 2.371 palavras a partir do *corpus* de *scraps* (de 32.895 palavras), buscamos um contraste com os dados do autor em novas condições. Assim, a comparação com o *corpus* de Belmonte (2007) poderia ser mais equilibrada. Eis dados do nosso novo *corpus*, agora reduzido:

*Tokens* 2.371

*Types* 1.030

*Type/Token Ratio* 43,44

Como podemos perceber pelas informações fornecidas pelo *software*, a riqueza vocabular indicada pela proporção entre *types* e *tokens* encontrada no *corpus* reduzido de *scraps* não é significativamente diferente da riqueza vocabular presente nas quatro reportagens premiadas, conforme mostrado pela transcrição da tabela do artigo de Belmonte (2007), a nossa Tabela 8. Enfim, o vocabulário exibido pelos internautas não pode ser considerado pobre no quesito variedade, à medida que se aproxima do vocabulário exibido em textos jornalísticos de excelência. Temos a riqueza dos *scraps* em 43,44 e a dos textos jornalísticos em torno de 40%. Isso é o que vemos na tabela a seguir.

**Tabela 9: Tokens, types e ratio do corpus de reportagens sobre biodiversidade**

Frequência da palavra-tema e riqueza lexical do texto					
REPORTAGEM	TEMA	%	PALAVRAS DISTINTAS	TOTAL DE ITENS	RIQUEZA LEXICAL
Jardim de luxo	Bromélia(s)	1,4	812	2.355	34,48%
Siga a anta	Anta(s)	1,39	731	1.790	40,84%
Projeto Muriqui	Muriqui(s)	1,46	965	2.461	39,21%
A floresta renasce	Vegetação	0,89	921	2.357	39,08%

*Riqueza lexical: é a razão vocábulo/ocorrência calculada pelo WordSmith Tools. Quanto maior a porcentagem, mais palavras diferentes há no texto (SARDINHA, 2004, p.94).*

Fonte: Belmonte, 2007

Belmonte (2007) examinou também a riqueza vocabular de um conjunto de artigos científicos que tratavam sobre o mesmo tema das reportagens. Constatou que nesses textos a riqueza lexical era de 29%.

Um contraste entre o léxico de recados e de depoimentos do Orkut e o das redações é algo importante de ser visto. Afinal, os produtores desses textos escolares são da mesma faixa etária dos que escrevem no Orkut, quando não são os próprios autores de *scraps* e depoimentos. Além disso, é interessante observar se haveria influências do *internetês* na escrita fora da rede mundial de computadores. Analisemos, então, um pouco mais detalhadamente os dados dos nossos *corpora* de redações:

**Quadro 23: Dados estatísticos gerais dos *corpora* de produções dos alunos de 7ª série e 3º ano do EM**

REDAÇÕES 7ª SÉRIE	REDAÇÕES 3º ANO EM	REDAÇÕES SOMADAS
<i>Tokens</i> 3.357	<i>Tokens</i> 3.514	<i>Tokens</i> 6.871
<i>Types</i> 1.016	<i>Types</i> 1.025	<i>Types</i> 1.675
<i>Type/Token Ratio</i> 30,27	<i>Type/Token Ratio</i> 29,17	<i>Type/Token Ratio</i> 24,38

O que vemos é uma riqueza vocabular bem acima de 20%, um número considerável comparando-se com o do *internetês*. Claro que por ser um *corpus* pequeno, quanto menor o número de palavras no total (*tokens*) maior a probabilidade de haver mais palavras diferentes (*types*).

Assim como reduzimos o *corpus* de *scraps* para ter um número similar aos *corpora* das reportagens premiadas de Belmonte (2007), fizemos o mesmo com o *corpus* total de redações. Dessa maneira, com semelhante número de palavras no total, é possível obter uma visão realmente comparativa, partindo do mesmo ponto em comum – o mesmo número total de palavras – para obter o número de palavras diferentes. Para chegar ao número desejado de palavras no novo *corpus*, fomos eliminando da lista de textos as primeiras e as últimas redações componentes do *corpus de redações*. O número esperado de palavras (*tokens*) foi atingido quando restaram as seis últimas produções escolares do Ensino Fundamental e as cinco primeiras do Ensino Médio.

A seguir, então, os números e as frequências do novo *corpus*, composto das palavras das redações agora diminuídas para apenas 2.466 *tokens*.

**Quadro 24: Estatística do *corpus* reduzido de redações**

Text File <i>CORPUS22.TXT</i>
<i>Bytes</i> 13.933
<i>Tokens</i> 2.466
<i>Types</i> 781
<i>Type/Token Ratio</i> 31,67

Mais uma vez, podemos perceber que a riqueza vocabular encontrada no *corpus* reduzido de redações (31,67%) é um pouco diferente da riqueza vocabular presente nas quatro reportagens premiadas (34,48%, 40,84%, 39,21% e 39,08%). A maior variação presente no léxico dos internautas no Orkut (43,44%), conforme indicado pelo *corpus* reduzido de *scraps*, se deve, provavelmente, pelo uso mais intenso de variantes ortográficas, conforme veremos na **Seção 5.5 Variedade de forma e conteúdo.**

Que indicam as repetições mais ou menos elevadas de formas na escrita dos jovens no Orkut? Essa riqueza lexical inclui realmente palavras diferentes ou abrange apenas as mesmas palavras escritas de forma diferente tal como em *mtu*, *mtu*, *mt*, *mtooo*, *mtoooo* e *bjo*, *bjus*, *bju*, *bjuxx*, *bj*? Esses elementos, justamente, “fazem” a riqueza do *internetês* em relação a outros *corpora*. Vale comentar, já antecipando, que o percentual de alteração de grafia no *internetês* fica em torno de 20%, tal como será explicitado adiante na página 192.

Há repetições, mas também palavras empregadas apenas uma vez no *corpus*. Podemos considerar que há uma relação bastante intensa entre o número de palavras de ocorrência única em um texto e o seu grau de variedade vocabular. Dizemos isso porque, em geral, as palavras de único uso, as *hapax legomena*<sup>64</sup>, constituem a grande “massa” de palavras de um texto. Vale aqui, então, um pequeno contraponto entre o nosso *corpus* de *internetês* e alguns outros que se referem à incidência desses hápax.

Observemos as primeiras e as últimas vinte *hapax legomena* do *corpus* do *internetês*:

<sup>64</sup> Expressão grega (*hapax*, “uma só vez”, *legomenon*, “dito”, “o que se diz”) que se utiliza para referir uma palavra da qual apenas se conhece uma única referência. De acordo com o Aurélio (2001), já está aportuguesada: “hápx é uma abreviação do grego hapax legomenon, ‘o que foi dito apenas uma vez’[...] Palavra, termo locução, etc., que ocorre apenas uma vez em documento, obra literária ou científica, etc.[...]”.

**Quadro 25: Primeiras e últimas vinte *hapax legomena* do corpus do *internetês***

..... PRIMEIRAS VINTE.....				.....ÚLTIMAS VINTE.....			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
15863	çSLA	1		38784	ZUAÇÃO	1	
15864	Æ	1		38785	ZUAÇÕES	1	
15865	AÁ	1		38786	ZUADINHA	1	
15866	ÃÃÃ	1		38787	ZUADORA	1	
15867	ÃÃÃÃ	1		38788	ZUAMU	1	
15868	ÃÃÃÃÃ	1		38789	ZUANDU	1	
15869	ÁÁÁÁÁÁ	1		38790	ZUANON	1	
15870	ÁÁÁÁÁÁÁÁ	1		38791	ZUANUU	1	
15871	AAAAAAAAA	1		38792	ZUBIOLO	1	
15872	ÁAAAAAAAAA	1		38793	ZUEERAS	1	
15873	AAAAAAAAAAAAA	1		38794	ZUEIRAAAA	1	
15874	ÁÁÁÁÁÁÁÁÁÁÁÁ	1		38795	ZUEIRAAAAA	1	1
15875	AAAAAAAAAAAAAA	1		38796	ZUERAA	1	
15876	AAAAAAAAAAAAAH	1		38797	ZUERAAA	1	
15877	AAAAAAAAAAAAADOR+	1		38798	ZUERAAAAAAAAA	1	
15878	AAAAAAAAAAAAAMMM+	1		38799	ZUERRRRRRRRRRRA	1	
15879	AAAAAAAAAAAAAMO	1		38800	ZUMBIZERA	1	
15880	AAAAAAAAAAAAAMO	1		38801	ZUNA	1	
15881	AAAAAAAAAAAAAMO	1		38802	ZUOU	1	
15882	AAAAAAAAAHHHHHH+	1		38803	ZUVA	1	

São palavras características do *internetês* as primeiras vinte que aparecem, com repetição de letras, provavelmente muitas delas formadoras de figura para transmitir recado em lugar de palavras. Também há repetição para dar ênfase à interjeição AH ou ao verbo AMO, cada internauta utilizando o número de letras que lhe aprouver. Vemos, entre as últimas, outras palavras que são variações de uma mesma forma, como ZUEIRA, a qual aparece de sete maneiras diferentes. Esse tipo de escrita típico do *internetês* deve estar, certamente, em outros casos de hápax na grafia orkutiana.

Vejam os seguintes quadros contrastivos que ilustram os diferentes índices de hápax em variados *corpora*, com a indicação de types (palavras diferentes):

**Tabela 10: Types, hápax e percentagem de hápax dos corpora**

	1	2	3	4	5	6
	<i>INTERNETÊS</i>	REDAÇÕES	ENTALPIA TEXTO TÉCNICO	BCO PORT. ESCRITO	BCO PORT FALADO	NURC- RJ FALADO
<i>TYPES</i>	38.803	1.675	168	255.035	46.128	8.549
HÁPAX	22.940	1.048	99	93.094	18.340	4.143
% DE HÁPAX	59 %	62%	59%	36%	39%	51%

Em linhas gerais, o quadro nos mostra que há maior riqueza vocabular<sup>65</sup> nas redações escolares, enquanto os textos escritos do Banco de Português exibem o menor índice de palavras de única ocorrência. Uma explicação possível para essa baixa incidência de hápax no Banco do Português pode estar no tipo de texto formador do *corpus*, que além de textos jornalísticos e de ficção (normalmente com grande número de palavras pouco repetidas) há também textos acadêmicos e documentos comerciais, técnicos e de negócios (cartas e fax comerciais, manuais, editais, relatórios).

Observemos os dados de cada *corpus*, em um plano mais detalhado.

1. **INTERNETÊS no Orkut:** Da lista de 38.803 palavras distintas, temos 22.940 que só foram empregadas uma única vez. O primeiro hápax começa na posição 15.863 e termina na 38.803. Isso significa que 59% de toda a lista são de palavras que ocorrem apenas uma vez.

2. **REDAÇÕES:** Já o vocabulário das redações de 7ª série e do 3º ano do E.M., composto de 1.675 palavras, tem seu primeiro hápax no ranking 627. Contém, então, 1048 palavras de única ocorrência, o que representa 62% do total de types.

Os dez primeiros e os dez últimos hápax que apareceram: ABAIXACE, ABDUZIU, ABOBADA, ABORDADO, ABORRECEU, ABRI, ABRIR, ABRIU, ACALMA, ACALMÁ, VOLTA, VOLTANDO, VOLTE, VOLTOU, VONTADE, WILLIAM, XAU, XD, XINGARAM, ZÍPER.

<sup>65</sup> Podemos perceber a riqueza vocabular de um texto pela proporção entre types e tokens e pelo número de hápax. O uso de hápax oferece uma outra face da riqueza, estando dentro dela. As palavras de único emprego num texto são uma maneira de ver o estilo de um indivíduo, a forma de ele mostrar também autoria pelo arranjo dessas palavras variadas.

3. **ENTALPIA – Texto didático:** O *corpus* do texto de Química (Entalpia), de apenas 168 types, inicia a lista de *hapax legomena* no número 69, totalizando 99 itens só empregados uma vez, o que resulta num percentual de 59%.

Os dez primeiros e os dez últimos hápax: ABERTO, ACORDO, AFIRMA, ALÉM, AQUECEDOR, AQUECIMENTO, ARBITRÁRIA, ATRAVÉS, AUMENTA, BÉQUER, TEMOS, TERMO, TERMODINÂMICA, TERMOQUÍMICA, ÚLTIMO, ÚNICO, VDQ, VEMOS, VEREMOS, VISTA.

4. **BANCO DE PORTUGUÊS ESCRITO:** O *corpus* do português escrito do Banco de Português mostrou hápax da posição 161.941 até a 255.035, num total de 93.094 palavras, o que revela 36% de palavras de única ocorrência.

Os dez primeiros e os dez últimos hápax: ¥ANDUTI, ¥UBLE, A'SABER, AAAAAHHHHH, AAAADOOOREEEI, AAAAI, AAIIII, AAAAMPLO, AAADT, AAAH, AAII, AAIII, ZYBRAINICS, ZYCIE, ZYGOTE, ZYL, ZYLBERMANN, ZYLLBERSZTAJN, ZYMAN, ZYSMAN, ZYUNDAL, ZYXEL.

5. **BANCO DE PORTUGUÊS FALADO:** Na fala do Banco de Português, dos 46.128 itens lexicais não repetidos, iniciam os hápax na palavra de número 27.788 da lista, num total de 18.340 hápax, o que significa 39% de vocábulos de único emprego.

Os dez primeiros e os dez últimos hápax: AÄ, AAA, AAÍ, AAS, AAULA, ABACA, ABACATADA, ABADONA, ABAFA, ABAFADAS, ZORA, ZORBA, ZORÓ, ZOTET, ZUADA, ZULEIDE, ZULEIKA, ZUNINO, ZURICH, ZURIQUE.

6. **NURC-RJ FALADO:** Verificando o *corpus* do português falado do NURC-RJ, na lista de 8.549 palavras diferentes, o primeiro hápax aparece no número 4.406, totalizando 4.243 types só usados uma vez. O percentual de *hapax legomena* é, então, de 51 %.

Os dez primeiros e os dez últimos hápax: ABACATEIRO, ABACATES, ABAFADA, ABAFAMENTO, ABAIXAVA, ABANDONADO, ABANDONADOS, ABANDONO, ABC, ABÓBORA, XVIII, ZA, ZARITA, ZECA, ZIGUEZAGUEAR, ZINCO, ZOEIRA, ZONAS, ZOMERIDE, ZOOLÓGICO.

Chama-nos a atenção a elevada percentagem de palavras usadas uma única vez nos textos escritos. O *internetês*, nesse quesito, acompanha os indicativos da escrita, tendo o mesmo número de hápax do texto técnico.

Pela análise da pequena amostra de exemplos dos primeiros e dos últimos hápax em ordem alfabética, podemos fazer algumas afirmações (mais como curiosidade):

1. Há grande incidência de verbos (15 em 20 palavras) entre os hápax de textos escolares (talvez por ser *corpus* composto por narrativas);

2. Substantivos são as palavras mais empregadas uma única vez na linguagem técnica e nos dois *corpora* de fala;

3. No *corpus* do português escrito do Banco de Português aparecem palavras semelhantes às típicas do *internetês*, inclusive com repetição de letras, estrangeirismos e sobrenomes.

Num texto, os *hapax legomena* contribuem para a qualidade estilística, mostrada pelos recursos retóricos utilizados pelo autor, mas também pelo vocabulário rico e variado, além do uso de nexos diferenciados (normalmente de única ocorrência). Pelo uso de léxico diferenciado, podemos perceber características do falante, o quanto domina o assunto que aborda e também o modo pelo qual busca se mostrar. Enfim, mais uma vez lembramos Benveniste (1991, p. 228): “É na linguagem e pela linguagem que o homem se constitui como sujeito; porque só a linguagem fundamenta na realidade, na sua realidade que é a do ser, o conceito de ego”. Esse sujeito busca, com suas palavra escolhidas e arranjadas de forma particular, significar e constituir sentidos.

## **5.2. PALAVRAS MAIS FREQUENTES (FREQUÊNCIAS ISOLADAS E POR CLASSES DE PALAVRAS)**

Iniciemos observando quais são as dez palavras mais frequentes do Banco do Português no seu todo, com mais de 120 milhões de elementos lexicais, e do nosso *Corpus Geral do Orkut*.

**Quadro 26: Palavras mais empregadas em *scraps*/depoimentos e palavras mais utilizadas no Brasil, segundo o Banco de Português**

FREQUÊNCIA <i>CORPUS</i> GERAL DO ORKUT				FREQUÊNCIA BANCO DE PORTUGUÊS NO TODO		
Nº	Palavra	Freq.	%	Nº.	Palavra	Ocorrências em 120 milhões
1	<b>E</b>	13.930	2,52	1	<b>DE</b>	(6.022.939)
2	<b>QUE</b>	11.537	2,08	2	A	(4.289.463)
3	EU	10.619	1,92	3	O	(4.135.372)
4	A	10.173	1,84	4	<b>E</b>	(2.906.593)
5	<b>DE</b>	9.950	1,80	5	<b>QUE</b>	(2.763.756)
6	Q	9.028	1,63	6	DO	(2.433.919)
7	TE	8.937	1,61	7	DA	(2.169.947)
8	O	8.396	1,52	8	EM	(1.760.984)
9	É	7.855	1,42	9	PARA	(1.403.295)
10	VC	6.347	1,15	10	NO	(1.233.337)

O que vemos, então, é a ocorrência de cinco palavras em ambos os *corpora*: DE, QUE, E, A e O, em posições diferentes na ordem de frequência. Vejamos:

**Quadro 27: Diferentes posições de cinco palavras na ordem de frequência**

	ORKUT ( <i>internetês</i> )	BANCO DE PORTUGUÊS (norma padrão)
PALAVRA	POSIÇÃO	POSIÇÃO
E	1º lugar	4º lugar
QUE	2º lugar	5º lugar
DE	5º lugar	1º lugar
A	4º lugar	2º lugar
O	8º lugar	3º lugar

Interessante observar que, das dez palavras do Banco de Português, sete são preposições (A, PARA, DE e EM e suas combinações), dois artigos (A e O), uma conjunção (E) e o QUE, empregado como conjunção, pronome relativo e tendo outras funções também para unir palavras ou orações. Por essas informações confirma-se, mais uma vez, o caráter de língua associativa do português. Já no *internetês*, não aparecem DO, DA, EM, PARA e NO.

Aparecem, por outro lado, pronomes pessoais (EU, TE e VC<sup>66</sup>) e verbo (É). A presença desses pronomes e do verbo, um caracterizador, se deve, provavelmente, ao tipo de texto formador do *corpus*, depoimento do Orkut, o qual expressa opinião e sentimentos de quem escreve sobre um amigo. Vemos a marca da presença dos autores no *internetês* do orkut, pelo uso – diríamos até excessivo – dessas palavras entre as de maior ocorrência (TE, EU, VC e É). Elas são indicadoras de subjetividade na comunicação dos jovens, em que expõem suas idéias, na maioria das vezes, em textos pessoais. No *corpus* total (mais de 120 milhões de palavras) do Banco de Português essas quatro palavras aparecem na seguinte posição: TE (1.855<sup>a</sup>) (surpreendamo-nos todos!), EU (87<sup>a</sup>), VOCÊ (119<sup>a</sup>) e É (13<sup>a</sup>).

Salientamos também que para o *internetês* Q e VC são consideradas palavras diferentes de QUE e VOCÊ, sendo, portanto, dois *types* na grafia, mesmo que tenham idêntico significado e mesma pronúncia quando lidos (ninguém lê algo que não seja “que” para Q ou “você” para VC, no contexto da Internet relativo à comunicação dos jovens). A palavra VOCÊ, como vimos, aparece no *corpus* do Banco do Português em 119<sup>o</sup> lugar.

A análise das freqüências das palavras é importante para o entendimento da língua como um todo, nos alerta Berber Sardinha (2004). Saberemos a probabilidade de ocorrência de um traço ou estrutura pela observação empírica da freqüência de emprego por diversos usuários em determinado contexto. É isso que fizemos observando a freqüência do léxico do *internetês* no Orkut para entender melhor suas características. A freqüência é um atributo inseparável da palavra, tendo sua ocorrência em grau alto, baixo ou intermediário de uso “um papel definidor da palavra, fornecendo um traço inseparável quanto o sentido” (BERBER SARDINHA, 2004, p. 162). O autor também se refere a Guiraud<sup>67</sup>, para quem a palavra é originalmente uma criação individual, mas “também e sobretudo uma criação coletiva: a palavra, criada pelo indivíduo não assume seu valor senão na medida em que é aceita, retomada, repetida; por isso ela se define, afinal, pela soma de seus empregos”.

<sup>66</sup> Na há dúvida de que na prática VOCÊ equivale a um pronome pessoal.

<sup>67</sup> GUIRAUD, P. Les caracteres statistiques du vocabulaire, essai de methodologie. Paris: PUF, 1954. p. 306.

Biderman (1998a) também destaca a importância da análise da frequência ligando-a à instituição da norma lingüística, que se baseia na frequência dos usos lingüísticos, e às mudanças lingüísticas que ocorrem com o passar do tempo. Diz-nos a autora:

Assim a norma lingüística nada mais é do que a média dos usos frequentes das palavras que são aceitas pelas comunidades dos falantes. E não é só isso. Também as mudanças lingüísticas que, no decorrer da história, levam de um estado de língua a outro, advêm das frequências de certos usos em detrimento de outros (BIDERMAN, 1998a, p.162).

Contrastando os dados disponíveis, verificamos a lista das palavras mais frequentes do *subcorpus* falado do Banco de Português. Para nossa surpresa, as duas palavras mais frequentes são as que aparecem também na lista da frequência do *Corpus* do Orkut. Comparemos:

#### Quadro 28: Palavras mais frequentes no Orkut e na fala do Banco do Português

FREQUÊNCIA <i>CORPUS</i> GERAL DO ORKUT				FREQUÊNCIA BANCO DE PORTUGUÊS <i>SUBCORPUS</i> FALADO				
Nº	Palavra	Freq.	%		Nº	Palavra	Freq.	%
1	E	13.930	2,52	→	1	E	113.061	3,73
2	QUE	11.537	2,08	→	2	QUE	108.883	3,59
3	EU	10.619	1,92		3	A	77.882	2,57
4	A	10.173	1,84		4	É	75.609	2,49
5	DE	9.950	1,80		5	O	71.329	2,35
6	Q	9.028	1,63		6	DE	66.922	2,21
7	TE	8.937	1,61		7	NÉ	64.876	2,14
8	O	8.396	1,52		8	NÃO	62.445	2,06
9	É	7.855	1,42		9	EU	55.733	1,84
10	VC	6.347	1,15		10	F	45.235	1,49

Conforme já detalhado, o *subcorpus* falado do Banco de Português, disponível para livre consulta na Internet, é composto por 197.901 palavras, 84.910 delas recolhidas de aulas e 112.991, de conversação.

A décima palavra mais freqüente do *corpus* falado chama a atenção como algo intrigante, porque poucas vezes falamos “efe” ou utilizamos o som “fff” isoladamente. Qual seria a razão de tão alta incidência do uso de “F”? Para descobrir, recorreremos a concordâncias, e a curiosidade foi desfeita. F ocorre principalmente em três situações: nas listagens metodicamente indicadas por letras do alfabeto, em siglas com tal letra e para indicar um dos falantes na transcrição dos turnos de fala das aulas e conversações registradas no *corpus*. Entenderemos melhor a situação observando algumas concordâncias de F retiradas do Banco de Português:

I As notas (E,F,G,H e I) , variando de 0 a 10, são determinadas (vozes de alunos) de uma PROva de N.E.F. então vocês teriam aí encheriam h

2.7.- Certidão de Regularidade com o F.G.T.S., bem como cópia autenticada s

1. F: Você chegou a usar o livro nessa aula?

3. F: Como? Para quê?

1. F: E, além do livro ... você falou que usou o livr

3. F: Como você usou?

É interessante registrar que, mesmo sendo a letra F (isolada ou seguida de sinal de pontuação) reconhecida como uma palavra pelo *WordSmith Tools*, isso não invalida a pesquisa nem desvaloriza os outros dados fornecidos. É apenas o reconhecimento de que o ferramental possui limitações, sendo programado para reconhecer como palavra um sinal gráfico sozinho ou acompanhado de sinal de pontuação. Lembremos que isso se faz necessário porque existem algumas palavras de uma só letra, como E, O, A ou É. Já nos referimos a essa forma de registro em *Wordlist* na **Seção 4.2.1 Corpus Geral do Orkut**, quando ficamos intrigados com a informação de haver no *internetês* a ocorrência de duas “palavras” grafadas como £.

Agora é preciso explicitar/comentar os pontos de coincidência entre ambos, o *Corpus Geral do Orkut* e o *subcorpus* falado do Banco de Português. O primeiro ponto a chamar a atenção é a coincidência das palavras que aparecem em primeiro lugar: E e QUE, bem diferente do *subcorpus* escrito, em que DE aparecia em primeiro lugar e o E em quarto. Outra coincidência é a proximidade de freqüência de DE, em 5º lugar no *corpus* do Orkut e em 6º no do português falado. A presença do verbo É também chama a atenção, embora aparecendo em 4º lugar na fala e em 10º na escrita orkutiana. Por sua condição de determinante, os artigos não poderiam deixar de figurar entre as palavras mais empregadas.

Das palavras da fala que não se repetem como as mais freqüentes no *internetês* destaca-se o operador conversacional NÉ (em 7º lugar no falado), que está em 124º no *corpus* do

Orkut, com apenas 644 ocorrências (há, porém outras formas dele, como NÉÉÉ, NEEH e NEHHH, mas não em número significativo). Sobre as diferentes formas de grafar uma mesma palavra, característica do *internetês*, verificando as palavras que aparecem no Quadro, teríamos alteração de posições na ordem das frequências. Isso porque VC/VOCÊ (sem contar o C, às vezes equivalendo a VC) somados resultam 8.148 ocorrências, É/EH (sem contar ÉH) tem 12.340, o E/I apresenta 15.109, TE/TI/T mostra 11.681, o O/U com 9.089 e o DE/DI/D exibe 12.321.

Assim, se fôssemos computar cuidadosamente todos os possíveis significantes para um mesmo significado, teríamos um novo quadro das palavras mais frequentes. Superficialmente, sem contar *hapax legomena* e variações de baixo uso, teríamos como as mais frequentes, numa nova ordem, as que colocamos à esquerda no Quadro 27.

**Quadro 29: Frequência de formas diferentes de mesma palavra e *Corpus* Geral do Orkut**

Soma de diferentes formas das maiores frequências no <i>Corpus</i> Geral do Orkut			Palavras mais frequentes do <i>Corpus</i> Geral do Orkut			
Nº	Palavra	Frequência	Nº	Palavra	Frequência	%
1	QUE/Q	20.565	1	E	13.930	2,52
2	E/I	15.100	2	QUE	11.537	2,08
3	É/EH	12.340	3	EU	10.619	1,92
4	DE/DI/D	12.321	4	A	10.173	1,84
5	TE/TI/T	11.681	5	DE	9.950	1,80
6	EU	10.619	6	Q	9.028	1,63
7	A	10.173	7	TE	8.937	1,61
8	O/U	9.089	8	O	8.396	1,52
9	VC/VOCÊ	8.148	9	É	7.855	1,42
			10	VC	6.347	1,15

Comparemos agora as frequências do *corpus* reduzido de *scraps*, de apenas 2.371 palavras, com o *Corpus* Geral do Orkut, com 553.875 itens lexicais.

**Quadro 30: Palavras mais frequentes do *corpus* reduzido de *scraps* e do *Corpus Geral do Orkut*.**

CORPUS REDUZIDO DE SCRAPS				CORPUS GERAL DO ORKUT			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	E	59	2,49	1	E	13.930	2,52
2	DE	56	2,36	2	QUE	11.537	2,08
3	VC	46	1,94	3	EU	10.619	1,92
4	O	37	1,56	4	A	10.173	1,84
5	Q	36	1,52	5	DE	9.950	1,80
6	É	33	1,39	6	Q	9.028	1,63
7	A	32	1,35	7	TE	8.937	1,61
8	TE	31	1,31	8	O	8.396	1,52
9	EU	29	1,22	9	É	7.855	1,42
10	QUE	27	1,14	10	VC	6.347	1,15

Observando o Quadro 28, percebemos:

- Novamente o E aparece em primeiro lugar, proporcionalmente mantendo a mesma percentagem de ocorrência (2,49 e 2,52, respectivamente) nos dois *corpora*.
- Mesmo numa diferença tão grande do número componentes dos *corpora*, **as dez palavras mais frequentes continuam sendo as mesmas**, apenas havendo alteração de posição na ordem delas.
- Uma diferença gritante surge na posição do Q e do QUE, que passam para o 5º e 10º lugares no *corpus* reduzido de *scraps*, com maior frequência da forma Q, enquanto no *Corpus Geral do Orkut* o Q é a segunda mais usada.
- No *corpus* de *scraps* o DE aparece como a segunda mais frequente, seguida de VC, que é a décima colocada no *Geral do Orkut*.
- As palavras mais usadas no *internetês* permanecem as mesmas, seja em *corpora* grandes ou reduzidos. A ordem em que aparecem pode variar, mas na análise dos variados *subcorpora* regionais, vistos anteriormente na **Seção 4.2.2 *Corpora Regionais e Corpus de Scraps***, essa tendência se mostrou estável.
- Com base nas asserções anteriores, podemos afirmar ser o léxico do *internetês*, quanto à frequência, bastante recorrente.

Passemos agora a observar as palavras mais usadas no português escrito, no português falado e no *internetês*, para estabelecermos comparações.

**Quadro 31: Palavras mais empregadas no português escrito, no português falado e no *internetês***

ESCRITA				INTERNETÊS				FALA			
FREQUÊNCIA BANCO DE PORTUGUÊS ESCRITO				FREQUÊNCIA CORPUS GERAL DO ORKUT				FREQUÊNCIA BANCO DE PORTUGUÊS FALADO			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	DE	1.537.460	4,42	1	E	13.930	2,52	1	E	113.061	3,73
2	A	1.082.233	3,11	2	QUE	11.537	2,08	2	QUE	108.883	3,59
3	O	1.026.380	2,95	3	EU	10.619	1,92	3	A	77.882	2,57
4	E	726.548	2,09	4	A	10.173	1,84	4	É	75.609	2,49
5	QUE	667.850	1,92	5	DE	9.950	1,80	5	O	71.329	2,35
6	DO	609.521	1,75	6	Q	9.028	1,63	6	DE	66.922	2,21
7	DA	545.271	1,57	7	TE	8.937	1,61	7	NÉ	64.870	2,14
8	EM	443.567	1,28	8	O	8.396	1,52	8	NÃO	62.445	2,06
9	PARA	353.847	1,02	9	É	7.855	1,42	9	EU	55.733	1,84
10	NO	308.932	0,89	10	VC	6.347	1,15	10	F	45.235	1,49

O que temos no quadro acima é uma visualização geral das palavras mais empregadas na escrita, na fala e no *internetês*. Está explícita a relação existente entre o *internetês* e a língua falada, senão vejamos: o *DE* é a palavra mais usada na escrita, mas na fala está em sexto lugar e no *internetês*, em quinto. O mesmo ocorre com as palavras *A* e *O*, em segundo e terceiro lugar, respectivamente, no português escrito e que nas outras duas formas de emprego não permanecem na mesma posição. Há uma aproximação muito clara entre a frequência das duas palavras mais usadas na fala e no *internetês*, diríamos até que seriam as mesmas se não houvesse duas maneiras de representar o *QUE* pelos jovens na sua comunicação no Orkut, tornando esse conector o item lexical mais empregado. O português é uma língua associativa, o *QUE* (representado pelas formas *QUE* e *Q*) confirma isso assim como o *E* na fala e o *DE* na escrita indicados pelos dados do Banco do Português. Uma é associação ADITIVA e a outra, associação VINCULADORA. Esta última é claramente demonstrada pela palavra mais

empregada pelos jovens na Internet. O QUE vincula fortemente idéias, com suas muitas e variadas funções, como o uso em orações subordinadas substituindo antecedentes, transmitindo noção de adjetivo e conferindo valores circunstanciais. Em situação de menor associação entre palavras ou orações, pode indicar intensidade, surpresa ou indefinição, por exemplo. Assim, podemos considerar que as vinculações seriam uma característica mais marcante da gramática da escrita no Orkut, enquanto a escrita “tradicional” aparece caracterizada por associações, que estabelecem significações entre um antecedente e seu conseqüente. São, assim, dois “territórios” de escrita: o do E e o do QUE.

Para comparação com os dados do Banco de Português, *subcorpus* falado, e confirmação do que afirmamos, vamos agora verificar as dez palavras mais frequentes do *corpus* do NURC-RJ, este composto exclusivamente por entrevistas orais feitas com habitantes de nível universitário do Rio de Janeiro. Observe o leitor a correspondência entre as palavras dos dois *corpora*:

**Quadro 32: Palavras mais frequentes nos corpora NURC-RJ e Banco de Português**

FREQUÊNCIA PORTUGUÊS FALADO – PROJETO NURC-RJ				FREQUÊNCIA BANCO DE PORTUGUÊS FALADO			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	QUE	3.973	3,71	1	E	113.061	3,73
2	DE	3.109	2,91	2	QUE	108.883	3,59
3	EU	2.885	2,70	3	A	77.882	2,57
4	NÃO	2.739	2,56	4	É	75.609	2,49
5	É	2.566	2,40	5	O	71.329	2,35
6	A	2.534	2,37	6	DE	66.922	2,21
7	E	2.422	2,26	7	NÉ	64.870	2,14
8	O	2.300	2,15	8	NÃO	62.445	2,06
9	NÉ	1.439	1,35	9	EU	55.733	1,84
10	UM	1.349	1,26	10	F	45.235	1,49

O que vemos no quadro? As mesmas palavras, com exceção das colocadas na 10ª posição. O intrigante F do Banco de Português e o UM (artigo e numeral). Verificando a lista das 100 palavras mais frequentes do Banco de Português falado (Anexo III), nela o UM aparece em 12º lugar, algo bastante próximo ao indicado pelo *corpus* NURC-RJ. Os diferentes

tipos de texto (conversação, aulas e entrevistas) que compõem os *corpora* não alteraram as palavras na lista, apenas sua frequência. Chama a atenção o QUE como o vocábulo mais empregado nas entrevistas-conversa sobre variados assuntos que compõem o *corpus NURC-RJ*, o mesmo mais freqüente do *corpus* do internetês, neste expresso nas formas QUE e Q. Sobre essa palavra ainda há que se registrar, no Banco de Português falado, que a diferença entre ela e a mais usada é de apenas 0,14%, uma vez que percentagem de E é de 3,73 e a de QUE é de 3,59.

Outro dado a chamar a atenção é a presença do DE como a segunda palavra mais usada na fala culta do Rio de Janeiro. Estaria a fala culta mais próxima da língua escrita como se poderia acreditar? A resposta seria sim, porque enquanto na escrita o DE é a palavra mais empregada, no internetês aparece em 5º lugar e no *corpus* de conversações e de aula aparece em 6º. Já na norma da fala culta, em 2º lugar.

O verbo É aparece quase na mesma posição (5º e 4º lugares). A maior diferença no *ranking* das mais freqüentes dos dois quadros é dada pela palavra E, que de 1ª mais usada no Banco de Português aparece como a 7ª no NURC-RJ. Mesmo assim ainda está entre as mais usadas. Outra observação que fazemos é que em ambos os *corpora* falados aparecem as palavras NÃO e NÉ, que não estavam entre as mais freqüentes do *internetês*: na escrita da Internet o NÃO está em 21º com 0,64% e a forma NEH em 89º, com participação de 0,17% na formação do léxico utilizado. A presença do EU em 4º lugar na fala recolhida pelo NURC-RJ aproxima-se da posição no *internetês*, em que esse pronome pessoal aparece em 3º lugar. Sobre os artigos A e O (haveria pronome pessoal e a preposição A computados junto?), nos dois *corpora* aparecem bem colocados, com maior freqüência no Banco de Português (3º e 5º lugares) do que no NURC-RJ (6º e 8º), aparecendo a forma feminina mais freqüente do que a masculina em ambos.

Seria possível, ainda, recorrer a outros *corpora* de fala. Entretanto, acreditamos que o *corpus* NURC-RJ e o Banco de Português sejam suficientemente representativos para observação das palavras mais freqüentes.

Vejamos a seguir as palavras mais freqüentes na fala culta, no *internetês* e na fala de conversação e aulas:

**Quadro 33: Palavras mais frequentes nos *corpora* de língua falada culta, *internetês* e língua falada.**

FALA CULTA				INTERNETÊS				FALA			
FREQUÊNCIA PORT. FALADO PROJETO NURC				FREQUÊNCIA CORPUS GERAL DO ORKUT				FREQUÊNCIA BANCO DE PORTUGUÊS FALADO			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	QUE	3.973	3,71	1	E	13.930	2,52	1	E	113.061	3,73
2	DE	3.109	2,91	2	QUE	11.537	2,08	2	QUE	108.883	3,59
3	EU	2.885	2,70	3	EU	10.619	1,92	3	A	77.882	2,57
4	NÃO	2.739	2,56	4	A	10.173	1,84	4	É	75.609	2,49
5	É	2.566	2,40	5	DE	9.950	1,80	5	O	71.329	2,35
6	A	2.534	2,37	6	Q	9.028	1,63	6	DE	66.922	2,21
7	E	2.422	2,26	7	TE	8.937	1,61	7	NÉ	64.870	2,14
8	O	2.300	2,15	8	O	8.396	1,52	8	NÃO	62.445	2,06
9	NÉ	1.439	1,35	9	É	7.855	1,42	9	EU	55.733	1,84
10	UM	1.349	1,26	10	VC	6.347	1,15	10	F	45.235	1,49

Observamos no Quadro 33 que sete palavras são comuns aos três *corpora*, havendo em todos a presença destacada do QUE. As palavras do *internetês* que não aparecem entre as dez mais usadas são VC (você) e TE, além da forma Q, “embutida” no QUE dos outros *corpora*. A ausência do UM entre as dez primeiras não é tão significativa, uma vez que, como já dissemos, sua posição na lista está em 12ª posição, a mesma que essa palavra ocupa no *corpus* do Banco de Português. Como vemos há muitas semelhanças entre as posições das palavras na lista das palavras mais frequentes do *internetês* e dos *corpora* falados.

Mais uma vez, vemos que os dados dos *corpora*, tomados em amplitude, revelam similaridade da riqueza lexical entre o *corpus* NURC-RJ (7,99) e o *internetês* (7,01). Mas, ao observarmos esses mesmos dados em detalhe, considerando frequências específicas de QUE, E e DE, vemos que a aproximação do léxico do *internetês* é maior em relação ao Banco de Português do que em relação ao NURC-RJ.

Se voltarmos a observar o Quadro 31, na página 180, em que comparamos as palavras mais empregadas no *internetês*, no português escrito e no português falado, veremos que o *internetês* revela-se como uma junção de fala e escrita, ora os dados aproximando-se de uma, ora de outra. Já ao analisarmos as percentagens das frequências na Tabela 10, a seguir,

notamos a predominância da oralidade. Isso pode ser percebido visualmente observando-se para que coluna a seta indicadora de maior aproximação se dirige. A seta parte da coluna do *internetês* voltando-se para a da escrita ou a da fala, dependendo da similaridade do percentual de uso das cinco palavras mais usadas na escrita.

**Tabela 11: Percentagem da frequência de uso na escrita, na fala e no *internetês***

Palavra	Português escrito %	<i>Internetês</i> %	Port. falado NURC %	Port. falado Bco Port. %
DE	4,42	1,90 →	2,91	2,21
E	2,09	2,50 →	2,26	3,73
QUE	1,92 ←	2,26 (3,83) <sup>68</sup>	→ 3,71	3,59
A	3,11	1,72 →	2,37	2,57
O	2,95	1,50 →	2,15	2,35

O uso de artigos parece ser a característica mais peculiar do *internetês* frente à fala ou à escrita tal como vemos no Banco de Português. Os pontos de maior proximidade com a fala são os usos de DE e de QUE/Q. O uso isolado de QUE no *internetês* é próximo ao uso da escrita.

Na Tabela 13, as mesmas palavras mais empregadas na escrita aparecem com o número da posição em que aparecem no *ranking* das mais usadas. Em cores diferentes, indicamos cada posição, assim o leitor poderá relacionar a posição em que cada palavra se encontra nos vários *corpora* e estabelecer a semelhança entre elas.

**Tabela 12: Posição das palavras no *ranking* das mais frequentes nos diferentes *corpora*.**

Palavra	Português escrito	<i>Internetês</i>	Port. falado NURC	Port. falado Bco Port.
DE	1 <sup>a</sup>	5 <sup>a</sup>	2 <sup>a</sup>	6 <sup>a</sup>
E	4 <sup>a</sup>	1 <sup>a</sup>	7 <sup>a</sup>	1 <sup>a</sup>
QUE	5 <sup>a</sup>	2 <sup>a</sup> <sup>69</sup>	1 <sup>a</sup>	2 <sup>a</sup>
A	2 <sup>a</sup>	4 <sup>a</sup>	6 <sup>a</sup>	3 <sup>a</sup>
O	3 <sup>a</sup>	8 <sup>a</sup>	8 <sup>a</sup>	5 <sup>a</sup>

<sup>68</sup> Somando-se a percentagem de uso de QUE e Q obtemos 3,83%.

<sup>69</sup> QUE passa a ser a palavra mais usada no *internetês* quando somada a sua variante Q.

Já os dados retirados do *corpus Scraps Reduzido*, de apenas 2.371 *tokens*, mostram como mais usadas as palavras *E* (a mais freqüente na fala) e *DE* (a mais freqüente na escrita). Se somarmos, porém, as formas *QUE* e *Q*, temos o número 63, passando a ser essa a palavra mais empregada nos recados do Orkut, a mesma que apareceu em primeiro lugar no *Corpus Geral do Orkut*. Detendo-nos na percentagem das ocorrências, notamos não haver muita diferença entre os números do *Corpus Geral* e os do *subcorpus reduzido*.

Podemos dizer, então, após observar os últimos quadros comparativos, que o *internetês* se apresenta formalmente como uma escrita, mas a quantidade de características da oralidade é muito marcante no que se refere às freqüências.

### 5.3. ALTERAÇÕES DE GRAFIA – POR GRUPOS REGIONAIS DO ORKUT

Iniciemos observando dois textos<sup>70</sup> de autoria desconhecida que circulam pela Internet. Estão transcritos tal como os recebemos por *e-mail*:

**De aorcdo com uma peqsiusa de uma uinrvesriddae ignlsea, não ipomtra em qaul odrem as Lteras de uma plravaa etãso, a úncia csioa iprotmatne é que a piremria e útmlia Lteras etejasm no lgaur crteo. O rseto pdoe ser uma bçguana ttaol, que vcoê anida pdoe ler sem pobrlmea. Itso é poqrue nós não lmeos cdaa Ltera isladoa, mas a plravaa cmoo um tdo.**

**Sohw de bloa.**

**Fixe seus olhos no texto abaixo e deixe que a sua mente leia corretamente o que está escrito.**

35T3 P3QU3N0 T3XTO 53RV3 4P3N45 P4R4 M05TR4R COMO NO554 C4B3Ç4  
 CONS3GU3 F4Z3R CO1545 1MPR3551ON4ANT35! R3P4R3 N155O! NO COM3ÇO  
 35T4V4 M310 COMPL1C4DO, M45 N3ST4 L1NH4 SU4 M3NT3 V41 D3C1FR4NDO O  
 CÓD1GO QU453 4UTOM4T1C4M3NT3, S3M PR3C1S4R P3N54R MU1TO, C3RTO?  
 POD3 F1C4R B3M ORGULHO5O D155O! SU4 C4P4C1D4D3 M3R3C3! P4R4BÉN5!

<sup>70</sup> A escrita do segundo texto, que mescla caracteres, letras e números, é chamada de *geek chic* ou *leet*. Abordamos esse assunto na nota 33.

Pelo que podemos ver, esses textos se relacionam à percepção das formas escritas, de um modo que somos quase obrigados a entender a grafia de uma palavra, mesmo que tenha sido a sua forma alterada.

No *internetês*, encontramos “uma escrita na qual não há uma preocupação com a correção, mas sim com a comunicação com o outro” (FREITAS, 2004, p. 8). É uma escrita teclada, abreviada, com características próprias como qualquer leitor percebe no primeiro contato com ela. A mesma autora salienta que os internautas investem muita criatividade para conseguir condições ideais de interação social:

Essa criatividade se manifesta na criação de códigos discursivos complexos, pois usam, ao mesmo tempo, o alfabeto tradicional, as caracteretas, os scripts e outros, que marcam a natureza processual e dinâmico-discursiva dessa “conversação”, aproximando-a da conversação face a face cotidiana, mas materializada na escrita “teclada” (FREITAS, 2004, p. 8).

Essa nova e diferenciada forma de escrita se apresenta num *continuum* em que oralidade e escrita se fundem, contestando a perspectiva tradicional da natureza fragmentada e dicotimizada da fala em relação à escrita (PALMIERE, 2006). Sobre a oralidade na escrita do *internetês* dedicaremos a próxima seção deste trabalho. Ao que parece, para dar seu recado como se estivessem falando numa “conversa escrita-teclada”, os jovens deixam de lado a escrita convencional, exatamente para dar um caráter “falado” ao que escrevem: muitas palavras grafadas num código próximo à língua fonética, como em *kero* (quero), *keto* (quieto), *c tah loko?* (você está louco?), *soh tava pensanu* (só estava pensando) e *c q c vire* (você que se vire).

Antes de passar à enumeração das características da escrita com a qual nos ocupamos nesta pesquisa, vamos ainda retomar o que nos tem a dizer Palmiere (2006) sobre isso. Lembra-nos que essa nova escrita tem-se constituído em um contexto específico – o ciberespaço – com finalidades específicas. O que em geral a caracteriza e justifica são suas condições de produção/recepção: interação em que a agilização da escrita é um dos fatores fundamentais para imprimir certa dinamicidade que mais a aproxime da oralidade. “E tal aproximação faz com que a escrita convencional seja re-elaborada/re-significada, num movimento que não se submete às convenções da escrita tradicional” (PALMIERE, 2006, p. 576). O que então podemos dizer que foi re-elaborado nessa escrita? O que tem mesmo de diferente essa escrita? Em seguida, veremos um conjunto de características da escrita do

*internetês* no Orkut. Antes, porém, precisamos explicitar a forma como foi observado o *corpus* para que pudéssemos chegar a essas características.

A nossa preocupação inicial foi verificar quantas e quais de suas palavras obedeciam à norma culta ou não, se seriam modificadas por meio de abreviaturas ou por outro tipo de simplificação ou redução. Tendo em vista a extensa lista de palavras gerada a partir do nosso *corpus*, utilizamos como dados principais para exame apenas as primeiras duas mil palavras mais freqüentes. Fizemos o levantamento daquelas que se apresentavam diferentes da grafia registrada no Vocabulário Ortográfico da Língua Portuguesa (VOLP)<sup>71</sup>, indicando entre parênteses o número de ocorrências. Assim, de um *corpus* de 553.875 palavras, as primeiras 25<sup>72</sup> com modificações foram:

q (9.016), vc (6.342), eh (4.513), pq (3.129), d (2.960), mto (1.570), t (1.324), p (1.232), naum (1.210), i (1.149), neh (970), mt (853), c (813), aki (786), tah (786), ki (772), u (770) mtu (219), c (202), tah (185), u (177), soh (666), tbm (565), td (565),mtu (543).

Em seguida, passamos a agrupá-las conforme as semelhanças de alterações encontradas ou por algum outro critério, como o de haver mais de uma mesma forma para o mesmo vocábulo. Considerando as transformações da forma gráfica oficial, verificamos ser possível distribuir o léxico do *corpus* de depoimentos e *scraps* em dezessete categorias.

Para o entendimento de muitas palavras encontradas que pudessem oferecer dúvidas de compreensão fora do contexto, afinal, era uma lista de palavras, algumas esquisitas, muitas vezes tivemos que recorrer à geração de *concordâncias*. Concordâncias são um instrumento típico da Linguística de *Corpus*, que consiste na listagem de co-textos (palavras ao redor) nos quais um dado item (palavra isolada, composta, estrutura, etc.) ocorre. Esse recurso foi muito útil para dirimir dúvidas sobre palavras utilizadas nesse tipo de escrita. Com ele pudemos ver seu significado esclarecido pelo contexto, muitas vezes na leitura da primeira frase da listagem de concordâncias. Vejamos exemplos de concordâncias para *ctg*, *seim*, *tm*, *qnt*, *flr* e *ngm*:

<sup>71</sup> O Vocabulário Ortográfico da Língua Portuguesa registra a forma oficial de escrever as palavras conforme o Formulário Ortográfico aprovado pela Academia Brasileira de Letras em 1943, com as alterações de 1971. Atualmente a obra impressa está na quarta edição, lançada em 2004, com cerca de 350 mil verbetes. O VOLP é administrado no Rio de Janeiro por um Conselho de Lexicografia, constituído pelos especialistas Antônio José Chediak, Sílvio Elia, Evanildo Bechara e Diógenes de Almeida, este último representando a Academia Brasileira de Ciências. É possível consultar a versão on-line do VOLP acessando:

<http://www.academia.org.br/abl/cgi/cgilua.exe/sys/start.htm?sid=23>

<sup>72</sup> Entre parênteses indicamos o número de ocorrências no *corpus*.

6909 eu nunk vo dexe d tah sempri **ctg** em tudu q eh lugar, d ti dah  
 800 maxuca!!! mas mesmu fikandu **seim** fala kum eli kuandu eli mi  
 100 e mt arreganho, mas o cara **tm** o dom d ser gente fina!!! heheeh  
 4799 /goria qro q tu saiba u **qnt** t considero e q pra qualquer  
 4338 pode fikr um sééclo sem c **flr**, que não influi na nossa intimidade  
 3730 apaxonadus um pelu otro... i **ngm** m ingana.....hehehe /hm...

Apresentando algumas variantes de significação em relação à norma culta ou em relação a um sentido literal, pois há muitas gírias, que sempre existiram em todos os tempos, o *internetês* mostra-se basicamente como um conjunto de alterações de grafia. Com base no levantamento feito, exibimos a seguir, com exemplos do *corpus* pesquisado, algumas modificações constatadas em relação à grafia oficial e outros usos recorrentes.

### 1. Indicação de monossílabos por uma simples letra:

q = que, d = de, t = te, c = se, p = pra, m = me.

### 2. Substituição do acento agudo pela letra h em final de palavra:

Eh, neh, tah, lah, bah, jah, ateh, poh, voh, feh, toh, keh, quiseh<sup>73</sup>.

### 3. Reprodução da fala:

u, ki, aki, du, di, so = sou, agente, mi, amu, nu, dexe, issu, cum, qi, meo, gent, loka, tenhu, comu, ke, genti, possu, qui, axu, muito, us, poku, sabi, comigu, dxa, kirida, tamu, ovi, pelu<sup>74</sup>.

### 4. Nasalização indicada por UM ou UN em final de palavra:

naum, naun, bjaum, taum, intaum, noçaum, paixaum.

### 5. Sequência de consoantes representando palavra, sem uso de vogais:

pq = porque, cmg = comigo, gnt = gente, tbm = também, tb = também, td = tudo, qm = quem, tm = tem, bm = bem, qnd = quando, cm = como, qlqr = qualquer, qnt = quanto, flw = falou, bjs = beijos, vcs = vocês, ctg = contigo, dps = depois, qd = quando, flr = falar, ngm = ninguém, sb = sabe, vz = vezes, qq = qualquer.

### 6. Várias formas para um mesmo vocábulo:

mtu, mtu, mt, mtooo, mtoooo, muito

bjo, bjus, bju, bjuxx, bj, beijo, beeeejo, beijos

td, tdu, tudu, todo, tudo

qndo, qnd, qdo, qdu, quando

msm, msmu, msmo, mesmo

tb, tbm, tbem, também.

<sup>73</sup> O *h* já foi utilizado em português para marcar a tonicidade final em lugar do acento agudo, como ainda se pode ver em palavras como *Dinorah* e em algumas transcritas do hebraico, como *Javeh*, *torah*, *chanucah*, *hanucah* e *menorah*.

<sup>74</sup> Na fala, as vogais na posição final, de atonicidade máxima, tendem a sofrer expressiva redução. Assim, o som vocálico de “e” e “o” que normalmente ocorre é um débil [i] e de um débil [u]. O mesmo tende a ocorrer na posição pretônica, que aparece exemplificado no *internetês* em “kirida”.



o meu pegou essa gripe frte demais n sei mais o q fazer...to desesperada ele n come de jeito nenhum amiga...bjsssssss manda noticia...me add no teu msn

### 15. Substituição de palavras e expressões por símbolos ou algarismos:

T+, t+ = até mais, D+ = demais, 9dade = novidade, v6 = vocês, 6 = vocês

### 16. Transformação de expressão ou fraseologia em sigla:

TDB ou tdb = tudo de bom, FDS ou fds = fim de semana, FDP ou fdp = filho da puta  
RDTR = rolando de dar risada, MDDR = morrendo de dar risada

### 17. Uso de caracteretas ou emoticons:

Bjão !!! 😊

bjim

\$(".)\$

diculpa (>.<)bjos

Pra mim nem é muito esforço!ashuashuashasuhsa :\*

xeru =\*\*\*\*\*

antes estarei por aí 🙏 fica com Deus... beijos ;\*\*

**é q ando p lah d ocupada!** 😊

eu tow beem \o/

a parada pode dizerrr 😊 ;\*\*

Conforme vimos nos dados, as modificações realizadas nas palavras se referem basicamente ao registro mais simplificado possível da fala, sem preocupação com grafia oficial, interessando aos jovens apenas estabelecer contato, passar adiante suas idéias. A regra básica é uma simplificação da língua, com abundantes abreviações, siglas ou reduções. Elas vão do uso de uma simples consoante para representar uma palavra (q, p = que, para) passando pela ausência de uma vogal no vocábulo (fla, tmpo, pod, kda = fala, tempo, pode, cada) até a ausência total de vogais (cmg, gnt, tb, qnd, ngm, sb = comigo, gente, também, quando, ninguém, sabe).

Essas reduções, entretanto, não são aleatórias, pois há uma lógica. Observamos que são privilegiadas as consoantes, cujos nomes suprem as vogais não escritas. Por exemplo, “dxa”, em que “d” corresponde à sílaba “dei → de” (na fala o ditongo “ei” não é fortemente pronunciado) e o “xa” é a segunda sílaba de “deixa”. Mais claro fica em “kbça” (k = ca + b = be + ça), estendendo-se a regra para “blz” em que o “z”, pela seqüência de sons só poderá ser a sílaba “za” e não o nome da letra (zê), porque não existe “beleze” em português. Dessa forma, percebemos claramente que ler esses textos não é ler letras e que a nova grafia não é aleatória, mesmo na aparente transgressão gráfica na comunicação dos jovens.

Sobre quantas variantes de forma possíveis, percebemos que não são muitas, até quatro, como em “td, tdu, tudu, tdo”. O que pode acontecer são variações com o acréscimo de vogais

ou “s” para indicar a intensidade da idéia a ser passada. Assim, uma coisa é escrever “T amu, bjs, bjo, mto”, outra bem mais forte é registrar “T amuuuuuuuu, bjsssssssssss, bjoooooooooooooooo, mtooooooooooooooooo”. Essas são variações livres, aliás, todas as regras no internetês não são seguidas com muito rigor. Aparentemente não há leis de espécie alguma cobrando coerência, apenas que sejam grafias inteligíveis para os “iniciados” nessa linguagem até certo ponto cifrada. Se houvesse leis, os jovens, por serem jovens e contestadores, talvez as abandonassem e criassem outras formas de se comunicar.

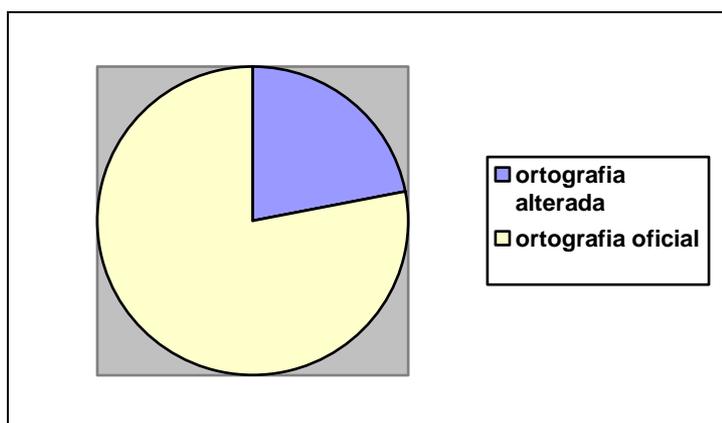
O *internetês* utiliza um sistema gráfico que procura reproduzir as falas da forma bastante fiel. Pouco interessam os grafemas utilizados, a prioridade é transmitir rapidamente uma mensagem ao interlocutor. As transformações existentes visam a simplificar e facilitar a escrita. Dessa forma, por exemplo, o dígrafo “qu” (e às vezes o “c”) é trocado pela consoante que sequer faz parte do alfabeto oficial do português, o “k” (aki, ke, pokko), o “ch” transforma-se em “x” (axo, maxuca) e até o “ss” pode virar um “x” como pronunciado em “máximo” (axim). Mas a maior transformação refere-se à supressão de letras, sejam de todas as vogais (vc = você, tc = teclar, tb = também, ctg = contigo), seja de somente uma delas (nunk = nunca, dpois = depois, kra = cara, pod = pode). Uma transformação que chama a atenção foi a de representar a nasalidade sem utilizar o til, escrevendo “un” ou “um” após a vogal nasal (naun, noçaum). Outra forma de mudança gráfica está na maneira de indicar a sílaba acentuada das oxítonas, em que o acento agudo foi alterado para “h” (neh, ateh, tah, bah). Um bom exemplo de registro variado é a expressão de risos ou gargalhadas, visto nos diálogos literários como “ahahahah”, no internetês é comum aparecer “rsrsrsrsrsr”, “rssssssss”, “KKKKKKKK”, “hauhauhauhauaha”, “hahaahaha”, “hehhehehe”, ou até mesmo um misto de letras maiúsculas e minúsculas, como “aUAheUheAHuaha”.

Sobre que símbolos são equivalentes a palavras, não foram muitos os encontrados por ser o *corpus* pesquisado composto de depoimentos e não de conversa de *chat*, por exemplo, em que a linguagem é mais espontânea. Nos depoimentos, algumas palavras são transformadas em símbolos ou desenhos, como acontece com o sinal “+” para significar “mais” ou, mais raramente, para indicar a conjunção “mas”, a forma gráfica “-” para “menos” e o “T +” para “até mais”. Na verdade, o internetês é uma recriação gráfica da língua, com representações e simbologias. Os sinais de pontuação, por exemplo, podem ser dispostas na frase de tal modo que signifiquem uma interjeição ou frase inteira. Todos os sinais não podem ser analisados isoladamente, mas em seu contexto, como representação das emoções humanas. E para dizer

muito com poucos meios foram criados os *emoticons* (*emotional icons*). São uma fusão de caracteres, que formam as chamadas “carinhas” ou “smilies”, expressão de sentimentos e situações com o uso criativo do teclado, que substituem palavras.

Como pudemos ver e ler nos dois pequenos textos iniciais desta seção, não lemos letra por letra, vemos o conjunto de caracteres e entendemos o seu significado como um todo. É assim com as “palavras” e palavras utilizadas pelos jovens na Internet. Normalmente lemos a abreviatura como palavra inteira, a forma resumida é um substituto apenas da palavra. Quando vemos “cap.”, por exemplo, não falamos “cap” mas “capítulo”, assim como na forma “p/”, vista como “para” e não “pê barra”. Poucos casos de abreviaturas são falados como siglas, é o caso de “tv” (tevé) ou “wc” (vecê). Já “ap.” (apartamento) nos últimos anos passou também a ser dito como “apê”, mas aqui cabe a pergunta: é a leitura da abreviatura ou é uma redução de palavra longa, como as utilizadas pelos jovens (sempre eles!) para “refri” e “bici” (refrigerante e bicicleta)? Quando no *internetês* aparecem “tb” ou “vc” ninguém que domina esse código vai ler “tebê” ou “vecê”, mas “também” e “você”.

Há que se levar em conta que o *internetês* não é escrito só com palavras grafadas que transgridem as normas ortográficas vigentes hoje. Das 2000 palavras mais frequentes do *Corpus Geral do Orkut*, apenas 439 possuíam alguma alteração (o que nos dá 21,9%) em relação à grafia “cultura”. As quase 80% demais estavam grafadas de acordo com a norma oficial. Isso é o que buscamos ilustrar na Figura 9, a seguir.



**Figura 9: Proporção de palavras do internetês alteradas graficamente e sem alteração (das 2000 palavras mais frequentes do *Corpus Geral do Orkut*).**

Por que então a grafia do internetês provoca tantos incômodos, “mexe” tanto com as pessoas, se apenas 20% de sua grafia é alterada? Porque tradicionalmente há maior

valorização com a forma do que com o conteúdo da escrita. Aliás, Bagno se refere a isso como *paranóia ortográfica*. E acrescenta: “Ora, saber ortografia nada tem a ver com saber a língua. São dois tipos diferentes de conhecimento. A ortografia não faz parte da *gramática* da língua, isto é, das regras de funcionamento da língua (grifo do autor)” (BAGNO, 2002, p. 131).

Outra informação importante é salientar que não nos comunicamos por uma lista de palavras que se lêem isoladamente, mas em contexto. É nesse contexto que descobrimos o que significam. Assim, uma palavra do *internetês* vista isoladamente que, a princípio, não apresenta significado algum, ao ser observada na frase onde ocorre desfaz o “mistério” e passa a ter significação. Isso ocorreu conosco ao obtermos a *wordlist*. Como não são todos os itens lexicais grafados de modo incomum numa frase, então pela relação com as palavras grafadas de modo “normal” entendemos a mensagem escrita.

Passemos agora à observação de alguns itens lexicais do *internetês*, analisando sua frequência de uso nas diferentes regiões do Brasil. Pretendemos averiguar se essa escrita eletrônica tem emprego uniforme em todo o Brasil ou se, assim como o léxico em geral de uma língua, tem variações diatópicas. Para a coleta dos dados apresentados no Quadro 34 utilizamos o recurso do *WordSmith Tools* chamado Concordância, em que a palavra solicitada aparece contextualizada por 4 palavras antes e 4 depois. As concordâncias de *gnt* (= gente), por exemplo, no *subcorpus* Rio Branco, mostraram:

N	Concordance	Set	Tag	Word No.	File	%
1	! Cante novamente ki a <b>gnt</b> pede bis!		É pique,	410	c:\corp8.txt	1
2	teu ladoo, por mais q a <b>gnt</b> naum se fale muuuu			464	c:\corp8.txt	1
3	ra Saidinhu q uma vez a <b>gnt</b> brigo e td mais....fic			13.209	c:\corp8.txt	41
4	inhu tem qnt tempo q a <b>gnt</b> se conhece heim???			13.187	c:\corp8.txt	41

Os itens lexicais pesquisados foram inicialmente buscados dentre as 500 primeiras palavras típicas do *internetês* (Anexo VII). Depois, algumas outras formas diferentes de grafar foram acrescentadas à lista, mesmo que pouco frequentes, para efeitos de comparação, como *vlw/valeu* e *dolu* (= adoro).

Quadro 34: Distribuição de diferentes formas lexicais por região do Brasil

TYPE	Ma- naus	Be- lém	Bra- sília	Cuia- bá	Porto Alegre	Reci- fe	Rio .Branco	RJ	Sal- vador	Scraps apenas	GERAL
<i>abs</i>	0	<b>0</b>	0	<b>0</b>	0	<b>0</b>	0	<b>0</b>	0	<b>2</b>	2
<i>bah</i>	0	<b>1</b>	0	<b>0</b>	31	<b>0</b>	0	<b>1</b>	0	<b>3</b>	284
<i>bjim</i>	1	<b>0</b>	1	<b>7</b>	0	<b>0</b>	2	<b>0</b>	2	<b>9</b>	25
<i>bjv</i>	7	<b>3</b>	3	<b>9</b>	3	<b>8</b>	4	<b>4</b>	6	<b>10</b>	96
<i>bjo</i>	8	<b>4</b>	11	<b>7</b>	4	<b>3</b>	5	<b>4</b>	4	<b>21</b>	245
beijo	5	<b>5</b>	5	<b>5</b>	5	<b>8</b>	4	<b>10</b>	11	<b>26</b>	157
<i>bjs</i>	14	<b>6</b>	7	<b>3</b>	5	<b>8</b>	17	<b>5</b>	4	<b>67</b>	211
<i>bjaum</i>	4	<b>2</b>	10	<b>10</b>	5	<b>4</b>	2	<b>3</b>	6	<b>9</b>	121
<i>dolu</i>	0	<b>0</b>	1	<b>11</b>	1	<b>0</b>	3	<b>1</b>	2	<b>0</b>	192
<i>fallow</i>	5	<b>4</b>	1	<b>0</b>	0	<b>6</b>	5	<b>2</b>	0	<b>4</b>	36
<i>flw</i>	5	<b>3</b>	3	<b>5</b>	2	<b>13</b>	2	<b>3</b>	0	<b>20</b>	89
<i>gnt</i>	9	<b>31</b>	14	<b>11</b>	40	<b>10</b>	4	<b>26</b>	9	<b>8</b>	363
gente	86	<b>47</b>	119	<b>88</b>	113	<b>89</b>	105	<b>118</b>	94	<b>52</b>	604
<i>mt</i>	93	<b>95</b>	75	<b>11</b>	19	<b>46</b>	39	<b>148</b>	33	<b>10</b>	855
<i>mtu</i>	18	<b>5</b>	30	<b>20</b>	37	<b>7</b>	13	<b>9</b>	14	<b>2</b>	530
<i>mto</i>	79	<b>40</b>	165	<b>144</b>	95	<b>37</b>	74	<b>100</b>	57	<b>26</b>	1587
muito	246	<b>210</b>	185	<b>170</b>	167	<b>197</b>	224	<b>165</b>	156	<b>113</b>	3401
<i>naum</i>	105	<b>110</b>	64	<b>71</b>	42	<b>59</b>	40	<b>80</b>	38	<b>50</b>	1192
<i>nao</i>	89	<b>54</b>	106	<b>127</b>	151	<b>53</b>	73	<b>91</b>	66	<b>88</b>	1418
não	144	<b>176</b>	199	<b>195</b>	207	<b>243</b>	204	<b>172</b>	186	<b>281</b>	3567
<i>namo</i>	0	<b>0</b>	0	<b>0</b>	1	<b>3</b>	0	<b>0</b>	0	<b>1</b>	11
<i>ngm</i>	1	<b>1</b>	8	<b>5</b>	1	<b>3</b>	5	<b>11</b>	2	<b>9</b>	74
<i>ninguem</i>	7	<b>19</b>	9	<b>7</b>	9	<b>9</b>	8	<b>5</b>	9	<b>6</b>	143
ninguém	11	<b>8</b>	3	<b>5</b>	10	<b>14</b>	8	<b>4</b>	10	<b>15</b>	147
<i>pq</i>	135	<b>226</b>	157	<b>157</b>	151	<b>177</b>	148	<b>185</b>	190	<b>67</b>	3158
porque	22	<b>19</b>	24	<b>19</b>	28	<b>34</b>	34	<b>21</b>	23	<b>17</b>	485
<i>q</i>	574	<b>722</b>	477	<b>508</b>	<b>545</b>	449	<b>501</b>	593	<b>469</b>	<b>327</b>	9028
que	617	<b>593</b>	649	<b>705</b>	784	<b>753</b>	791	<b>695</b>	703	<b>552</b>	11537
<i>tah</i>	35	<b>33</b>	38	<b>56</b>	48	<b>30</b>	25	<b>52</b>	19	<b>43</b>	789
<i>tb</i>	16	<b>12</b>	24	<b>25</b>	19	<b>28</b>	10	<b>28</b>	38	<b>31</b>	382
<i>tbm</i>	43	<b>44</b>	30	<b>34</b>	24	<b>18</b>	28	<b>19</b>	92	<b>56</b>	568
também	3	<b>15</b>	8	<b>6</b>	17	<b>11</b>	23	<b>13</b>	12	<b>23</b>	239
<i>t</i>	40	<b>167</b>	55	<b>72</b>	88	<b>73</b>	89	<b>59</b>	33	<b>26</b>	1369
<i>ti</i>	46	<b>117</b>	41	<b>61</b>	108	<b>17</b>	169	<b>43</b>	23	<b>26</b>	1407
te	402	<b>679</b>	471	<b>536</b>	400	<b>767</b>	789	<b>446</b>	1172	<b>239</b>	8922
<i>tu</i>	103	<b>161</b>	31	<b>9</b>	414	<b>192</b>	139	<b>27</b>	8	<b>110</b>	2699
<i>vlw</i>	2	<b>1</b>	4	<b>2</b>	6	<b>4</b>	4	<b>14</b>	5	<b>19</b>	88
valeu	8	<b>4</b>	9	<b>4</b>	11	<b>4</b>	5	<b>8</b>	4	<b>32</b>	139
<i>vc</i>	537	<b>425</b>	502	<b>564</b>	195	<b>372</b>	469	<b>623</b>	626	<b>459</b>	6347
<i>voce</i>	21	<b>8</b>	32	<b>36</b>	17	<b>13</b>	7	<b>36</b>	25	<b>8</b>	239
você	94	<b>88</b>	163	<b>143</b>	58	<b>167</b>	155	<b>127</b>	175	<b>128</b>	1816
<i>vcs</i>	8	<b>17</b>	6	<b>11</b>	8	<b>6</b>	7	<b>12</b>	7	<b>36</b>	188

Os dados do Quadro 34 permitem várias constatações sobre a variação ortográfica presente no *internetês* do Orkut. Eles indicam, pela grande frequência de uso, que formas

foram adotadas, quais estão sendo substituídas e se há emprego de formas alteradas e oficiais com igual intensidade:

1) No geral, não há significativo predomínio de determinadas alterações gráficas por região.

2) As variações existentes por região são relativas a vocábulos, como o expressivo emprego de *bah* em Porto Alegre, *bjim* e *dolu* em Cuiabá, *tu* em Porto Alegre, Recife e Belém.

3) Poucas formas típicas do internetês têm maior incidência por regiões. Exemplos são *flw* em Recife, *gnt* em Porto Alegre, *mt*, *vlw* e *ngm* no Rio de Janeiro e *tbm* em Salvador.

4) Algumas formas têm mesmo padrão de uso em todo Brasil, como *pq*, *q*, *tah* e *vc*.

5) Tendência a certos usos parece indicar preferência por região. É o caso de no Rio de Janeiro aparecer com destaque a maneira mais reduzida de escrever (apenas consoantes): *mt*, *ngm*, *tb* e *vlw*.

6) Algumas formas reduzidas estão substituindo, em todo país, as formas ortográficas oficiais: *pq* x porque, *vc* x você, *tb* e *tbm* x também.

7) Mesmo com variações, certas formas oficiais ainda são as mais empregadas em todo o Brasil. Ex.:

Manaus: muito (246 ocorrências) x *mtu* (79), *mtu* (18), *mt* (93)

Salvador: muito (156) x *mtu* (57), *mtu* (14) *mt* (33)

Recife: não (243) x *nao* (53), *naum* (59)

Rio de Janeiro: não (172) x *nao* (91), *naum* (80)

Brasília: te (471) x *t* (55), *ti* (41)

Rio Branco: te (789) x *t* (89), *ti* (69)

8) Algumas formas coexistem na escrita com certa similaridade de frequência em todas as regiões:

Porto Alegre: que (784) e *q* (545)

Belém: que (593) e *q* (722)

9) Não se percebe significativo uso das formas típicas do internetês no subcorpora de *scraps*, conforme esperávamos. Em algumas regiões, chega a ser menor o emprego de formas características nos *scraps* do que nos depoimentos. É o caso de *mtu*, *mt*, *pq*, *t* e *ti*.

Percebemos que o *internetês* não tem um número tão elevado de formas modificadas a ponto de ameaçar a escrita da Língua Portuguesa. Quando comparadas as formas em uso na Internet com a forma oficial, quantitativamente, são poucos os exemplos em que novas formas

substituem a norma padrão de escrever. Alguns exemplos têm uso pouco expressivo, conforme pudemos verificar com os levantamentos feitos em nossos *corpora*.

O *internetês* revela-se, nos nossos dados, como uma fala digitalizada e cifrada, surgida na Internet e manifestada por caracteres alfanuméricos. Para entender essa sua característica, também devemos levar em conta o aspecto relacionado ao suporte em que tal escrita é registrada. Afinal, os historiadores da escrita são unânimes em dizer que a alteração do suporte (tijolo, papiro, papel, monitor) é fator de mudanças na escrita e na leitura. Por exemplo, escrever de forma não cursiva é a mais indicada para pedra ou tijolo. Escrever, portanto, teclando no computador, especialmente *on-line*, é certamente algo que induz a transformações, principalmente pela velocidade que se quer dar àquilo que se transmite por escrito. Suportes especiais explicam escritas especiais.

#### 5.4. INDÍCIOS DE ORALIZAÇÃO SOBRE A ESCRITA

Na **Seção 5.2**, ao observarmos as palavras mais frequentes em *corpus* de fala, escrita e *internetês* (E, QUE, DE, A, O, É e EU) afirmamos haver uma junção entre fala e escrita nesse último, com características da oralidade muito marcantes. Aprofundaremos agora o exame desse indicativo buscando reconhecer e categorizar traços/especificidades e exemplos de oralização na escrita do Orkut.

As características que depreendemos da observação do *Corpus Geral do Orkut* são as seguintes:

##### 1) Presença de marcadores conversacionais

*Olá.....*

*Putz ciencias sociais..eh dmais*

*Ainda kero fazer um dia!!!*

*Nossa q mundo pekeno!! numa comunidade e ambos conehecemos a Tica!!*

*Conheço ela da marcha mundial das mulheres.....*

*Bom o q posso falar sobre mim*

*Sou de Poá...*

*faço geografia*

*E rs*

*Luto p uma eskerda..kk isso nem preciso falar*

*entre outras coisas*

*ops meu e-mail correto é karlosrabelo@hotmail.com nao rabalo...*

## 2) Presença de muitos períodos curtos e simples

*To achando q fds vou viajar!!!! mais te falo durante a semana ta????? Ah, detalhe perdi meu celular agora estou incomunicável. só tenho o tel da clinica mesmo.*

*mais de 10 anos devendo....eu?!?!?!rsrsrsr...ta d+...veja bem, minha vida agora e só trabalho...estudo..estudo...filho...trabalho....trabalho...nem encontro mais direito c as pessoas...tô na maior correria, aff!!!saudades, viu:) beijinhos\*\*\*\**

## 3) Emprego de léxico coloquial

ow **falcatrua....**

aonde tu anda??????

o que **anda aprontando??????**

manda noticias ai **cavaloooooooo**

e surplus **eh bom pra caralho** neh!

e ai tudo bom?

**ai meu**, eh q vc nao tem foto sua.

entao nao sabia q vc era vc.

## 4) Uso de frases truncadas

ola igor cm vc esta um abraço!!!!!!!!!!!!

amor da minha vidaa... como ce tá!!!!

biruta foi massa... foi eu e a pri.. e uma galerraa lá da bio.. e um povo que era migo do povo..

foi massa viu.. muita putaria.. a bruna foi tb...

e ontem ainda fui pro dragão com a sil.. assiti o otto.. e a gente depois encontrou a mon com o felipe...

show ontem tb foi irado..

mas dragao do mar tá decadente...

## 5) Pouca densidade informacional

fosse assistir o filme?!

e tu quer que eu fale o que?! Oo

foi legalzim! só isso a dizer, quem pode ter nova é alguém que seja solteira e pá! :P ;\*

eeeeeei... nova nenhuma! :T

As características acima se referem ao ponto de vista das escolhas léxico-sintáticas dos gêneros orais e, como vimos, são visíveis no *internetês*. Tal forma de escrever, mais uma vez, exhibe traços da linguagem empregada na fala. Vejamos agora alguns outros contrapontos entre nossos dados e indicativos da literatura que trata sobre escrita/fala.

Shepherd (1984) estabeleceu um quadro mostrando as principais diferenças entre fala e escrita. Apresentamos tal quadro ampliado em mais uma coluna referente ao *internetês*, na qual percebemos uma mescla de várias características da escrita e da fala. Vejamos:

**Quadro 35: Características de escrita, fala e *internetês*, inspirado no quadro das diferenças entre escrita e fala de Shepherd (1984)**

Escrita	Fala	<i>Internetês</i>
<p>1. A qualidade abstrata é intensificada pelo deslocamento do tempo.</p> <p><b>2. Se torna um registro permanente do acontecimento, um artefato documentário da história.</b></p> <p>3. O formato visual é de convenção, etiqueta, de acordo com estilo e função.</p> <p><b>4. Tende ao formal e conservador; menos inclinado a mudar.</b></p> <p>5. O que recebe a mensagem está ausente.</p> <p><b>6. Não recíproca, nenhuma resposta imediata.</b></p> <p>7. O escritor tem duplo papel; o leitor é uma presença psicológica.</p> <p><b>8. O receptor é um leitor; ler requer esforço.</b></p> <p>9. Conhecimento pressuposto. Tem que se fazer explícito.</p> <p><b>10. Redundância de natureza sintática – semântica deliberadamente adicionada com finalidade de clarificação.</b></p> <p>11. É possível a monitoração parar, reler, riscar, reescrever.</p> <p><b>12. Ritmo vagaroso.</b></p> <p>13. Convenções de sintaxe, ortografia, coesão e coerência.</p>	<p>1. É tempo real-“agora”.</p> <p><b>2. Vem e vai; é efêmera e transitória.</b></p> <p>3. É não visual (exceto com espectógrafos).</p> <p><b>4. Inclui modas, coloquialismos, gíria.</b></p> <p>5. O que recebe a mensagem está presente, usa uma variedade de características paralingüísticas como “feedback”.</p> <p><b>6. Consciência constante de uma “audiência”.</b></p> <p>7. Papel “simples” em interação face a face.</p> <p><b>8. O receptor é um ouvinte, menor esforço necessário.</b></p> <p>9. De um certo modo. Não necessário – pode ser verificado concomitantemente.</p> <p><b>10. Repetição, rephraseamento, pausas, marcadores de atenção.</b></p> <p>11. Monitoração através de “feedback” da audiência.</p> <p><b>12. Ritmo variado.</b></p> <p>13. Menos controlada, produção oral e desenvolvimento simultâneos.</p>	<p>1. O produtor do texto deixa a impressão de estar escrevendo como se estivesse em tempo real-“agora”.</p> <p><b>2. Torna-se um registro permanente. É efêmero quando <i>on-line</i>.</b></p> <p>3. Tem formato visual.</p> <p><b>4. Inclui mudanças, modas, coloquialismos, gírias.</b></p> <p>5. O receptor da mensagem não está presente, sem resposta imediata (menos quando <i>on-line</i>).</p> <p><b>6. Possibilidade de resposta imediata ou a curto prazo.</b></p> <p>7. Leitor é presença psicológica ou “visual à distância” com <i>webcam</i>.</p> <p><b>8. Requer esforço do receptor para lê-la.</b></p> <p>9. Pode ser verificado quando <i>on-line</i>.</p> <p><b>10. Repetições, marcadores extralingüísticos gráficos (maiúsculas indicando gritos, riso, choro e uso de <i>emoticons</i>).</b></p> <p>11. Não há monitoração para reelaborar após a mensagem enviada.</p> <p><b>12. Ritmo acelerado.</b></p> <p>13. Pouco controladas, pensamento e escrita desenvolvidos simultaneamente.</p>





*que eu tenho por ti ninguém apaga e ninguém tira  
com tudo posso dizer de verdade que TE AMO*

A questão do quanto a oralidade está presente na escrita dos *scraps* e depoimentos pode ser polêmica e instigante, não resolvida. Isso porque, além das mudanças gráficas, a Internet proporciona mudanças sociais, provocando novos modos de operação cognitiva. No dizer de Marcuschi (2005, p. 10), na Internet, “pela natureza da escrita que nela se desenvolve, aflora uma certa primazia de estratégias orais sobrepondo-se às estratégias da escrita”. Isso tem muito a ver com a escrita *on-line*, em que a recepção da escrita pode ser simultânea à sua produção. Não é o caso do texto do Orkut, mas nele se refletem algumas estratégias de textualização da oralidade que passaram para a escrita, como no *internetês* como um todo. Marcuschi ainda nos lembra que esse formato de escrita é mais complexo do que um simples falar por escrito. “Pois aí ocorrem interferências que não vêm da oralidade e sim de outros modos discursivos. Atente-se para o fato de que na fala usual existem gestos, mas não existem as carinhas, os *emoticons* e outros recursos da Internet” (MARCUSCHI, 2005, p. 11).

Costa (2006) acrescenta que a interface oralidade/escrita parece se dissolver de maneira muito forte no uso da Internet, que se torna responsável pelo surgimento de novos gêneros (hiper)textuais (*chat, e-mail, fórum, site, home-page, Orkut*) ligados à interatividade verbal. Para Costa (2006, p. 37):

Os (hiper)textos, produzidos/construídos sem fronteiras nítidas, misturam formas, processos e funções da oralidade, da leitura e da escrita. Leitor e autor/escritor se cruzam, *on-line*, no esforço de releitura, correção e recriação de um texto, participando da edição do texto que lêem e escrevem. Nesse ciberespaço, leitor e escritor deparam-se com novos conceitos, novo léxico, novos gêneros discursivos, novas formas de linguagem, novo código, novo estilo de ler, escrever e conversar.

Há uma sensação de interatividade no Orkut, semelhante à interação face a face, como a conversa em tempo real. Isso é facilitado pela foto presente na página e pela resposta normalmente esperada ao que deixamos escrito. Se escrevemos com a sensação de estarmos próximo do interlocutor na comunicação, é compreensível que haja muitas características da linguagem falada nessa escrita. São, portanto, de acordo com Hilgert (2000, p. 17) enunciados produzidos num “texto falado” por escrito.

Marcuschi (1991) lembra que a atividade comunicacional é uma realização que não pode ser separada do contexto em que se efetiva. O mesmo já fora observado por Bakhtin

(1999), ao afirmar que a linguagem é produzida no e pelo contexto sociocultural. Ora, a comunicação no Orkut ou na Internet como um todo – e ainda mais feita pelo grupo específico de internautas jovens – deve ter seu contexto altamente considerado antes de ser criticada. Em tal contexto, as características da linguagem são modificadas devido a necessidades práticas, como teclar rapidamente para não perder tempo e poder acompanhar o raciocínio, o que, por exemplo, leva a escrever *naum* por ser mais fácil e rápido, sem precisar utilizar a tecla *shift* (em alguns teclados). Nas palavras de Bakhtin (1999, p. 41): “É, portanto, claro que a palavra será sempre o indicador mais sensível de todas (o grifo é nosso) as transformações sociais”.

A transformação social ainda em implantação pelo rápido disseminar do uso da Internet traz consigo esse fenômeno ainda não totalmente esclarecido: a língua usada para interação na grande rede. Ela é usada em textos escritos que também recorrem à modalidade da fala. Apresenta-se com uma nova articulação das linguagens oral e escrita.

Encerramos esta seção com uma citação que consideramos válida para a escrita no Orkut, principalmente nos textos dos *scraps*. São observações de Bernardes e Vieira (2006, p. 52) para a linguagem dos *chats*, as quais também estendemos ao texto orkutiano:

A recorrência de períodos curtos e simples, o aparecimento de marcas de envolvimento entre os interlocutores, o alto tom de informalidade e descontração que predomina no diálogo, assim como a presença de marcadores conversacionais, são apenas algumas das características que parecem indicar uma possível aproximação deste texto com aspectos da fala cotidiana; entretanto, como pudemos observar os enunciados que o compõem apresentam uma nova “reconfiguração das formatações tradicionais da escrita” (XAVIER; SANTOS<sup>75</sup>, 2000, p. 53).

## 5.5. VARIEDADE DE FORMA E CONTEÚDO

A variedade de formas em uma lista de palavras retirada de um texto qualquer pode ser vista sob pelo menos duas perspectivas diferentes. A primeira perspectiva diz respeito às formas diferentes que as palavras assumem, independente de seu conteúdo. Esse é o caso, por exemplo, da variação morfofonêmica tal como se vê em *pude/pode*. De outro lado, na segunda

<sup>75</sup> XAVIER, Antônio Carlos; SANTOS, Carmi Ferraz. O texto eletrônico e os gêneros do discurso. In: *Veredas – Revista de Estudos Lingüísticos da UFJF*, v.4, n.1, jan./jun. 2000. Juiz de Fora: Ed. UFJF, 2000, p. 51-57.

perspectiva, é possível que *pude/pode*, embora sejam formas diferentes, correspondam a um mesmo conteúdo, lematizável na forma *poder*.

No caso do *internetês*, vemos a alternância *vc/você*, *mt/mto/muito/muitoo* e tantas outras. Aqui, diferentemente das formas lematizáveis (diferentes porque indicam uma diferença de significado), temos variedades apenas de forma, jeitos diferentes de expressar a mesma palavra e o mesmo conteúdo. Assim, salientamos, a noção de variedade lexical precisa ser relativizada.

Fizemos uma lista das cem palavras mais frequentes dos *corpora* do português escrito, do português falado e do *internetês* (Anexos V, VI, VII e VIII) para verificar e comparar essas duas formas de registro escrito, a segunda delas basicamente um uso na Internet<sup>76</sup>. Foram consideradas como variações de mesma palavra sua apresentação no plural e no feminino, as combinações com preposições e as formas dos verbos conjugados. Vejamos *corpus* por *corpus*.

No **português escrito**, salientam-se, pela ordem de maior ocorrência:

- |   |                                 |
|---|---------------------------------|
| 1. <b>DE/DO/DA/DOS/DAS</b> (8,53%) <sup>77 78</sup> | 2. A/O/AS/OS/AO/AOS/À/ÀS (8,25) |
| 3. <b>EM/NO/NA/NOS</b> (3,07)                       | 4. <b>É/SER/ERA/SÃO</b> (1,44)  |
| 5. FOI/FORAM (ser ou ir?) (0,40)                    | 6. PELO/PELA (0,39)             |
| 7. SUA/SEU/SEUS                                     | 8. <b>TEM/TER</b>               |
| 9. DISSE/DIZ  |                                 |

No **português falado**:

- |                                |                                      |
|--------------------------------|--------------------------------------|
| 1. A/O/OS/AS (5,68)            | 2. <b>DE/DO/DA/DOS</b> (3,71)        |
| 3. <b>É/ERA/SÃO/SER</b> (3,49) | 4. UM/UMA (1,83)                     |
| 5. PRA/PARA (1,07)             | 6. <b>TEM/TINHA/TENHO/TER</b> (0,68) |
| 7. ELE/ELA/ELES                | 8. <b>EM/NA/NO/NUM</b>               |
| 9. ESTÁ/ESTAVA                 | 10. MEU/MINHA                        |
| 11. ESSE/ESSA                  | 12. SENHORA/SENHOR                   |

<sup>76</sup> Há alguns casos de variantes de mesma palavra sem alteração de significado, como em *aluguel/aluguer*, *assoalho/soalho*, *assobiar/assoviar*, *bílis/bile*, *bravo/brabo*, *cãibra/câimbra*, *espuma/escuma*, *loira/loura*, *percentagem/porcentagem*, *quota/cota*, *quatorze/catorze*, *taberna/taverna*, *voleibol/volibol*, etc.

<sup>77</sup> Das primeiras seis ocorrências, indicamos entre parênteses a percentagem somada das formas variantes no *corpus*.

<sup>78</sup> Indicamos em negrito as variações que apareceram nos três *corpora*.

No *internetês*:

a) Formas com conteúdo diferente:

- |                               |                           |
|-------------------------------|---------------------------|
| 1. A/O/AS/OS (3,97)           | 2. <b>DE/DO/DA</b> (2,91) |
| 3. PRA/PARA (2,31)            | 4. UM/UMA (1,74)          |
| 5. <b>EM/NA/NO/NOS</b> (1,31) | 6. MEU/MINHA (1,04)       |
| 7. <b>TEM/TENHO/TER</b>       | 8. AMIGA/AMIGO            |
| 9. <b>É/SER</b>               | 10. SEI/SABE              |
| 11. SUA/SEU                   | 12. ESSE/ESSA             |

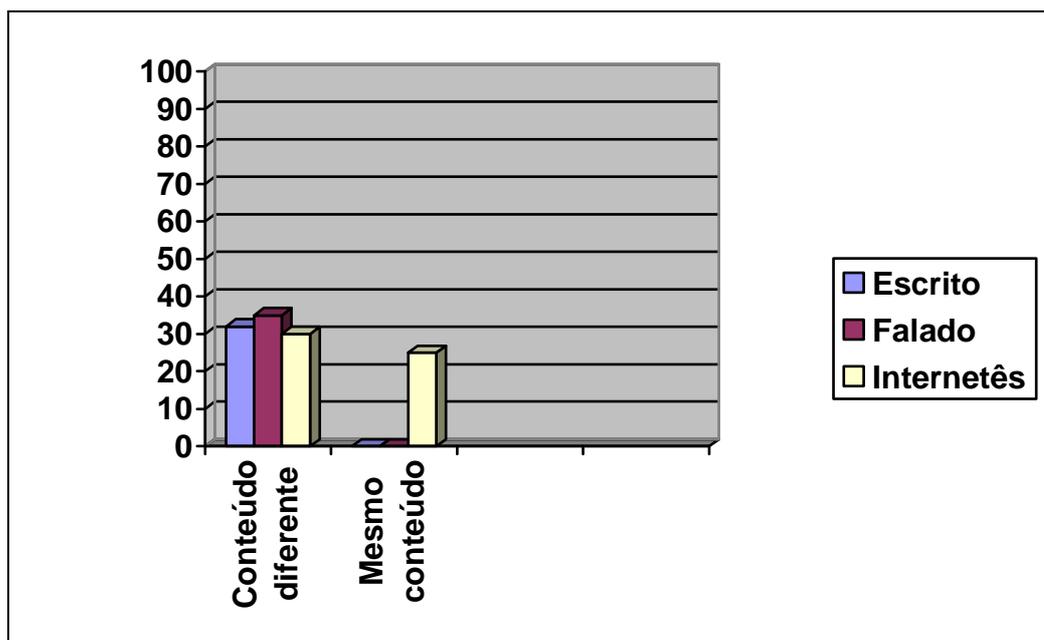
b) Formas com mesmo conteúdo:

- |                                 |                   |
|---------------------------------|-------------------|
| 1. Q/QUE (3,71)                 | 2. D/DE (2,33)    |
| 3. EH/É (2,24)                  | 4. T/TI/TE (1,86) |
| 5. P/PRA (1,53)                 | 6. VC/VOCÊ (1,48) |
| 7. MAIS/MAS (mais, mas ou más?) | 8. NAUM/NÃO       |
| 9. MT/MTO/MUITO                 | 10. I/E           |
| 11. C/VC/VOCÊ (se ou você?)     |                   |

Observando os dados, são evidentes as semelhanças quanto à variedade de formas para conteúdos diferentes extraídas dos três *corpora*. Repetem-se, com suas variações, as grandes incidências das preposições DE e EM, os artigos definidos O<sup>79</sup>, os verbos SER e TER e os pronomes possessivos MEU e SEU.

Vejamos os mesmos dados antes citados colocados sob uma outra apresentação. No gráfico temos colunas indicando as porcentagens das ocorrências de formas com variação de conteúdo e colunas para formas diferentes para um mesmo conteúdo presentes nos *corpora* escrito, falado e do *internetês*.

<sup>79</sup> Causa estranheza não estar entre as cem mais usadas do *corpus* escrito o artigo indefinido UM. Também a contração PRA não apareceu no *corpus* escrito.



**Figura 10:** Ocorrências de formas variáveis das palavras em amostra de cem palavras nos corpora.

Tal como vemos na Figura 9, são praticamente os mesmos números para variações morfofonêmicas, lematizáveis numa mesma forma, em todos os corpora (colunas da esquerda). Foram 35 itens diferentes no português escrito, 32 no falado e 30 no *internetês*. A novidade, como já foi percebida pelo leitor, é o conjunto de 25 itens lexicais com formas variantes para mesma significação, uma característica do *internetês* que enfurece os guardiães da língua portuguesa. São criações dos internautas, duas ou mais formas de dizer a mesma coisa, só que de maneira abreviada, truncada, como vimos na **Seção 5.3 Variação Ortográfica**, sem seguir nenhuma regra da gramática normativa. Os usos na Internet são, portanto, bem diferentes das variações da norma culta, de tipos como, por exemplo, *loura/loira* ou *aluguel/aluguer*.

Pelas observações feitas, podemos concluir que o *internetês* se mostra como língua, em nada diferindo das modalidades escrita e falada no que se refere a variações de forma para alterar a significação de uma palavra. Já no referente às diferentes formas de escrever uma mesma palavra, o *internetês* é pródigo em exibir as criações dos internautas. Ousaríamos afirmar que não há texto produzido por jovens na grande rede, quando na comunicação informal principalmente *on-line*, sem a presença de palavras grafadas de maneira truncada, abreviada, alguns praticamente códigos cifrados na visão de pessoas não familiarizadas com a comunicação internetiana.

## 5.6. SÍNTESE DAS OBSERVAÇÕES

Neste capítulo registramos as descobertas de nossa pesquisa sobre o léxico da escrita no *internetês* do Orkut: palavras mais freqüentes, riqueza vocabular, alterações de grafia dos itens lexicais, elementos de oralidade na escrita da Internet e variedade de forma e conteúdo das palavras empregadas.

Retomamos agora, sinteticamente, o que pudemos perceber acerca do léxico do *internetês*. Como no Brasil, no nível da escrita, as pessoas confundem *português* com *ortografia oficial* (Bagno, 2002, p. 125), e sendo o *internetês* basicamente uma expressão gráfica com alterações ortográficas (algumas estranhas), é esperável que passe a ser atacado, criticado, enfim, visto como algo prejudicial à língua (aqui confundida com ortografia).

O que deve ser levado em conta é o grau de adequabilidade à situação de uso da língua: em situação formal, linguagem formal, em situação descontraída, linguagem descontraída. A adequação se baseia no grau de aceitabilidade por parte dos interlocutores. Ora, a situação de comunicação entre jovens não é formal e sua aceitação pelos próprios jovens é total. Logo, para aquele contexto, não há nada de errado. O problema é levar tal forma de expressão escrita para outros contextos. Aí entra o papel da Escola e da sociedade como um todo, ensinando que, como sempre, tudo depende de *quem diz o que, a quem, como, quando, onde, por que e visando que efeito*. Qualquer tipo de expressão, então, poderia ser adequado e aceito. Graficamente, poderíamos representar essa situação da seguinte forma:

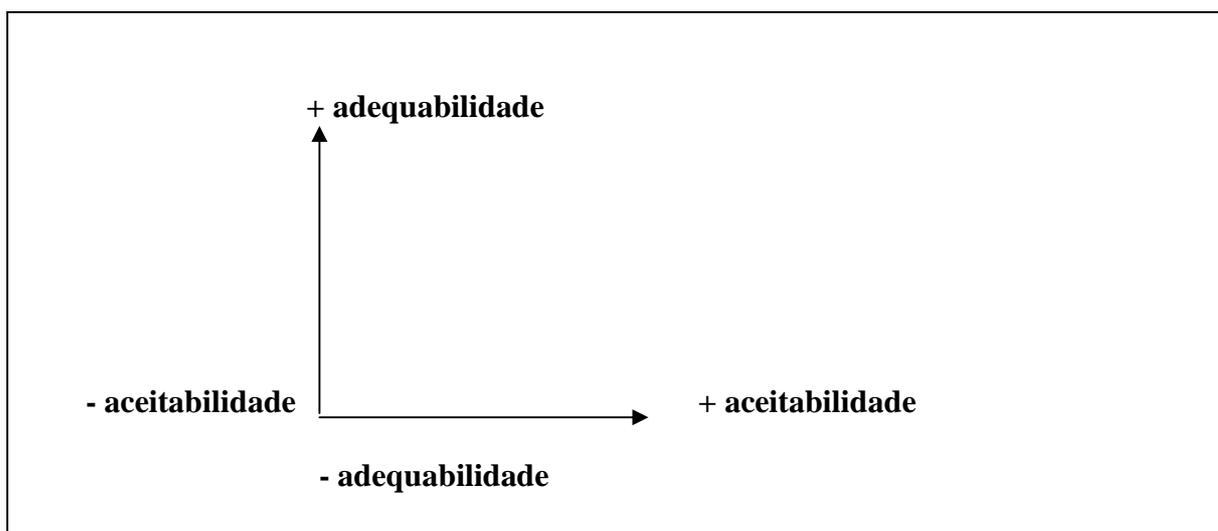


Figura 11: Uso da língua: equilíbrio entre aceitabilidade e adequabilidade (Fonte: BAGNO, 2002, p. 131)

O assunto “uso da língua na Internet” é polêmico, mas, conforme pretendemos demonstrar, é importante que tal uso seja alvo de descrições sistematizadas e criteriosas. Não se pode apenas julgar o fenômeno sem uma abordagem racional de suas características, motivações e efeitos.

Quanto ao *internetês* tal como visto nos textos de *scraps* e de depoimentos do Orkut de nove regiões do Brasil, obtivemos os seguintes dados:

1. Quanto à riqueza vocabular, há variações nas diferentes regiões brasileiras. Manaus e Brasília foram as cidades-pólo das regiões que apresentaram maior variedade de palavras na sua comunicação. Já Porto Alegre e Belém foram as de menor variedades.

2. Os jovens utilizam maior variação vocabular ao escreverem *scraps* do que ao compor seus depoimentos.

3. Quanto menor o número de palavras de um *corpus*, maior a possibilidade de haver palavras diferentes, resultando em maior riqueza vocabular.

4. Não são percebidas grandes diferenças de riqueza vocabular no contraste entre redações escolares, reportagens premiadas e textos de depoimentos e *scraps*.

5. Não há diferença de riqueza vocabular entre textos de *scraps* e textos de reportagens premiadas sobre biodiversidade. Portanto, o vocabulário dos internautas não pode ser considerado pobre no quesito variedade.

6. Comparando-se a frequência das dez primeiras palavras mais utilizadas no *Corpus Geral do Orkut* (553.735 tokens) e no *Corpus do Banco do Português* (120 milhões), aparecem em ambos os *corpora* as palavras QUE, E, DE, A e O. QUE é a mais freqüente do *internetês* e DE a do Banco de Português.

7. As duas palavras mais freqüentes do Banco de Português falado e do *internetês* são as mesmas (E e QUE), havendo duas formas gráficas para a mais usada no *internetês*: QUE/Q.

8. São muito similares as palavras presentes nos *corpora* do *internetês* e do português falado. Das dez mais freqüentes, sete aparecem em ambos os *corpora*, sendo as duas mais freqüentes as mesmas (E, QUE).

9. As dez palavras mais freqüentes de um *subcorpus* reduzido de *scraps* (2.731) se repetem no *corpus* de *scraps* e recados (553.735), apenas se alterando de posição.

10. As cinco primeiras palavras mais freqüentes do *corpus* do português escrito são as que aparecem também no *corpus* do *internetês* (DE, A, O, E, QUE).

11. O uso de artigos é menor no *internetês* do que na fala e na escrita. A maior proximidade do *internetês* com a fala se dá pelo uso de DE e de QUE/Q.

12. O *internetês* se apresenta formalmente como uma escrita, mas com muitas características da oralidade no que se refere às frequências.

13. A escrita dos internautas jovens se apresenta num *continuum* em que a oralidade e a escrita se fundem.

14. São ocorrências no *internetês*: indicação de monossílabos por uma simples letra, substituição do acento agudo pela letra h em final de palavra, reprodução da fala, nasalização indicada por UM ou UN em final de palavra, seqüência de consoantes sem vogais representando palavra, várias formas para um mesmo vocábulo, registro de palavras sem acentuação, palavra com ausência de uma letra, onomatopéias para indicar riso e choro, repetição de letra para indicar intensidade, redução do nome de pessoas, criações especiais só entendidas no contexto, repetição de sinais de pontuação para enfatizar sentimento, supressão de sinais de pontuação que marcam fronteiras oracionais, substituição de palavras e expressões por símbolos ou algarismos, transformação de expressão ou fraseologia em sigla, uso de caracteretas ou *emoticons*.

15. A percentagem de palavras com alteração gráfica é de 21,19 % entre as 2000 mais frequentes do *Corpus Geral do Orkut*. Portanto, 78,81% das palavras empregadas pelos jovens são as mesmas encontradas no VOLP – Vocabulário Ortográfico da Língua Portuguesa.

16. Evidências da presença de elementos da oralidade na escrita da Internet são: existência de marcadores conversacionais, emprego de muitos períodos curtos e simples, emprego de léxico coloquial, uso de frases truncadas, pouca densidade informacional e muitas frases soltas.

17. Existem duas formas de variação de palavras, uma delas modificando o conteúdo e outra, não. Apenas no *internetês* é característico haver duas ou mais formas de escrever um mesmo item lexical, como MT/MTO/MTU.

Pela observação dos dados, percebemos que o *internetês* é uma escrita com características próprias. Apresenta-se como um registro híbrido, misto de fala e escrita, constituindo um *continuum* que vai da total informalidade e transgressão de normas ortográficas até a linguagem formal. É marcado por sinais de oralidade na escrita, mais fortes nos *scraps* do que nos depoimentos. Sua riqueza vocabular não é muito diferente da riqueza lexical presente em redações escolares e reportagens premiadas. Temos como ocorrências recorrentes nessa escrita típica da Internet, basicamente, a forma de grafar as palavras: redução dos itens lexicais (39,34% do léxico é constituído por palavras de 2 e 3 letras), supressão de vogais, ausência e alteração de acentuação e pontuação, reprodução da fala, transformação de

dígrafos e uso de formas diferentes para indicar um mesmo vocábulo. Além dessas modificações, destacam-se o emprego de *emoticons*, onomatopéias e repetição de letras para registrar intensidade. Essas alterações gráficas chamam a atenção e causam diversas reações (principalmente de rejeição) e generalizações do tipo “o *internetês* é todo errado”. Constatamos, porém, em nossa pesquisa que o percentual de palavras com alteração gráfica é de 21,19%, estando, portanto, a maioria das palavras grafadas de acordo com a norma oficial. O estranhamento causado pelas formas inusitadas de grafia aliadas ao emprego de léxico coloquial, frases truncadas e pouca densidade informacional na comunicação gera preconceitos e faz com que seja visto, principalmente pelas pessoas comuns, como uma deturpação da língua.

## 6 RESULTADOS DA DESCRIÇÃO E PERSPECTIVAS

Neste capítulo retomamos nossas questões de pesquisa e apresentamos as respostas que depreendemos da observação do *corpus* e de suas variadas relações com os *corpora* de referência e os de contraste. Também fazemos a retomada das hipóteses. Por último, estão as conclusões a que chegamos, as perspectivas futuras e algumas considerações finais.

### 6.1. RETOMADA DAS QUESTÕES DE PESQUISA

**1) Como é a configuração lexical da escrita da Internet no Orkut em relação à variante culta da escrita do português brasileiro?**

O léxico do *internetês* no Orkut tem suas peculiaridades bem definidas, muito centradas, porém, na grafia das palavras. A forma de escrita é o que mais se destaca, embora o número de palavras escritas fora da norma oficial não ultrapasse muito a 20% do total do léxico. Esses 20%, entretanto, são grafados de maneira bastante diferente da escrita usual e chamam tanto a atenção a ponto de o leitor generalizar o *internetês* como algo totalmente fora do padrão gráfico em uso. Convenhamos, em dez palavras duas estarem “diferentes” não é caso de declarar o fim da língua. Acontece que às vezes há uma seqüência de palavras registradas de forma truncada e abreviada, sem seguir norma alguma, e isso dá a impressão de ser a norma no *internetês*. Como já vimos anteriormente, a tendência é de nos fixarmos mais no atípico. Outro dado é que são repetidas as palavras grafadas de forma pouco comum, o que causa a impressão de serem em número muito maior as formas diferentes (*types*).

Como vimos na **Seção 5.3 Alterações de grafia - por grupos regionais do Orkut**, conseguimos agrupar em 17 categorias as modificações constatadas em relação à grafia oficial e a outros usos recorrentes na expressão internetiana. Outra marca do *internetês*, poderíamos dizer também retratada pelo léxico, é a presença de marcadores conversacionais e o uso de vocabulário coloquial. Isso ocorre devido à grande influência da fala na escrita, ou seja, ser

quase uma “conversa por escrito”, mesmo no Orkut, em que o interlocutor não está presente *on-line*. É interessante destacar que, afora a grafia das palavras, um recado no Orkut pouco difere de um recado manuscrito no papel. São utilizadas palavras de uso corrente entre jovens, gírias, modismos, etc. Outros problemas de texto, como coesão e clareza de idéias, ocorrem independentemente de estar escrevendo em *internetês* ou não.

## **2) Qual a relação entre o vocabulário do *internetês* e o vocabulário norma culta de fala?**

É grande a relação do *internetês* com a fala no que tange ao vocabulário. Nossos dados indicaram que entre as trinta primeiras palavras mais freqüentes dos *corpora* (*internetês*, fala NURC-RJ, fala Banco de Português e escrita) há certa coincidência das palavras: 20 são as mesmas do *corpus* NURC-RJ e 17 são idênticas às dos *corpora* de fala e escrita do Banco de Português. Já entre as cem mais empregadas, as palavras que se repetem são concomitantes à fala e ao *internetês*, com número pouco significativo de itens do *corpus* da escrita. Por que as primeiras são as mesmas? Porque são palavras funcionais. Substantivos e outras palavras com significado que remete ao mundo extralingüístico aparecem bem depois na lista. Por exemplo, o primeiro substantivo (página) está na 25ª posição na escrita, 53ª (coisa) na fala Banco de Português, 40ª (coisa) na fala NURC-RJ e 43ª (vida) no *internetês*. Pela análise dos *corpora* que espelham a norma culta tanto na fala como na escrita, percebemos como as mais freqüentes palavras monossilábicas, dentre elas palavras de caráter associativo (QUE, DE, E), artigos e pronomes. Em praticamente todas as listas de palavras neste trabalho observamos a ocorrência das mesmas palavras entre as cinco mais empregadas: QUE, E, DE, A e O.

Como pôde ser visto na página 176, as sete primeiras palavras mais freqüentes da fala e do *internetês* no Orkut são as mesmas. À medida, porém, que começam a aparecer as palavras lexicais, as ocorrências não se repetem nas mesmas freqüências, havendo, como já dissemos, certa similaridade entre o vocabulário do *internetês* e o da fala. Outro ponto comum entre *internetês* e fala é o número de verbos e substantivos presentes na lista das cem palavras mais freqüentes, como podemos ver na Tabela 14, a seguir:

**Tabela 13: Verbos e substantivos entre as cem palavras mais freqüentes**

	ESCRITA	FALA- NURC-RJ	FALA Bco. Port.	<i>INTERNETÊS</i>
Verbos	14	18	17	18
Substantivos	22	10	8	10

Emerge, então, a diferença da freqüência do emprego dessas palavras, mostrando que na fala há menor uso de substantivos e maior de verbos, enquanto na escrita é o contrário. Um outro levantamento que fizemos verificando quais desses verbos e substantivos eram comuns ao *internetês*, vimos que sete deles estavam na lista da escrita, enquanto dez estavam nas listas do português falado. Como os números do *internetês* estão próximos aos da fala, podemos, mais uma vez, afirmar que ele é uma escrita com fortes traços de oralidade, o que se confirma também quanto ao vocabulário.

### **3) Que itens do vocabulário de variações de grafia do *internetês* são verificados em corpora que espelhem a escrita escolar?**

Surpreendentemente foram poucos os itens do vocabulário típico do *internetês* presentes nas redações escolares do nosso *corpus*. Até mesmo nas de sétima série, produzidas por alunos em faixa etária que usa freqüentemente computador e Internet para se comunicar e que tiveram menos tempo do que os do Ensino Médio para contato com o sistema gráfico do português. Além disso, esses alunos ainda têm necessidade de utilizar uma linguagem diferenciada para se marcar como participantes de determinado grupo. Em mais de 50% dos textos do pequeno *corpus* de redações, não houve nenhum indício da presença do *internetês*. E o que apareceu foi uma ou outra ocorrência de emoticons (XD~), reprodução da fala, marcas de intensidade pela repetição de letras, textos sem pontuação ou acentuação, alguns registros de minúsculas ou maiúsculas quando deveria ser o contrário e justaposição de palavras. Essas três últimas nem mesmo podem estar relacionadas ao *internetês*, pode ser esquecimento ou não internalização do uso de acentuação e pontuação, por exemplo.

Outras marcas típicas do *internetês* como excesso de reticências e exclamações apareceram em alguns textos, mas em número inexpressivo de ocorrências. Algumas formas que aparecem na fala e nos recados do Orkut estavam em alguns textos, escrever, por exemplo, como última palavra ao se despedir: “Falou!”, “Tchau”, “Até mais” e “Beijo”. Novamente fica

a dúvida: é influência do *internetês* ou é algo inerente ao tipo de texto solicitado, um relato (que o aluno narrou com diálogo)? Outra presença marcante no *internetês* - mas não só nele - é a seqüência de idéias “jogadas”, num único período composto para transmitir rapidamente idéias e sentimentos unidos pelo conector “e”, típico da fala. Estaria o aluno sendo influenciado pela fala ou pelo *internetês*, ou por ambos, uma vez que a fala também está no *internetês*? Uma ou outra marca de pontuação modificada ao escrever com letra cursiva apareceram, como um ponto final reforçado (uma bolinha) no final do texto e uma exclamação (no título) bem maior que a usual, tendo o ponto embaixo da barra vertical sido transformada numa estrelinha (Veja na p. 154).

Conforme anteriormente afirmamos, não foi significativa a presença de itens do vocabulário mais típico do *internetês* nas produções escolares sob exame. Assim, por exemplo, pouco vemos ocorrências como *muuito*. Mas, se um estudante raramente é solicitado a produzir textos em aula, escrevendo muito mais na Internet, *on-line*, do que na escola, seus hábitos de escrita fatalmente aparecerão no que escrever para seus professores, seja em que matéria for. Portanto, frente à escrita facilitada e sem compromisso na Internet cabe o trabalho (árduo mais do que nunca) do professor para insistir na norma culta, com sua grafia oficial. O aluno entenderá que no mundo atual ele, de alguma forma, é “bilíngüe”, conectado em rede usa uma língua e, no mundo fora dela, outra. Aqui podemos lembrar Bechara (1993, p. 12) quando afirma que “nenhum falante conhece toda uma língua histórica, mas sim usa uma variedade sintópica (um dialeto regional), sinstrática (um nível social) e sinfásica (um estilo de língua)”. O autor acrescenta que esse falante é capaz de compreender mais de uma dessas variedades lingüísticas de sua língua, que, para Coseriu (1980), são línguas funcionais, porque constituem o instrumento que lhe serve de intercomunicação na comunidade em que está inserido. Assim, à medida que reconhece a existência de outros usos da língua, chegando a perceber o que os caracteriza e faz uso dessas variedades de forma conveniente, cada falante se torna um “poliglota” de sua própria língua. Torna-se, pois, capaz de reconhecer, ainda que inconsciente, as gramáticas de cada sistema funcional do macrossistema, que é a própria língua, e amplia a sua gramática internalizada. Dessa forma, um aluno, ao escrever para a escola, deverá ter presente a forma adequada desse sistema funcional, o mesmo ocorrendo quando escrever na Internet. O que vimos é que a influência do *internetês* é quase zero no nosso *corpus* de redações.

Acreditamos que, também por serem escritas por alunos que semanalmente produzem textos, as redações do nosso *corpus* não exibem itens do vocabulário típico do *internetês*. Tais alunos são capazes de compreender as diferentes variedades lingüísticas e de se expressar adequadamente em cada uma. Mas, reiteramos: se um aluno apenas praticar sua escrita no *internetês*, dificilmente deixará de mostrar num texto fora da Internet a influência do que nela escreve. É indispensável reconhecer que a redação é, enfim, produzida em um ambiente bastante diferente em relação ao ambiente do Orkut. Talvez fosse interessante para uma expansão desta pesquisa colher algum outro tipo de registro escrito mais extenso, tal como o dos *blogs*,<sup>80</sup> por exemplo.

#### **4) O *internetês* pode ser considerado uma variante, um dialeto ou um subsistema em relação aos usos da língua portuguesa espelhados em *corpora* representativos da norma culta escrita?**

Como, aparentemente, não há um *internetês falado*, teríamos um caso peculiar de forma de uso da língua apenas por escrito. Essa forma de escrita é uma variação diastrática, ocorre por estratos sociais - usuários na rede mundial de computadores - e não por regiões. Podemos também dizer que é um dialeto (todo dialeto é uma variação), por ser usado num ambiente mais restrito que a própria língua. É que superposta ao dialeto existe sempre uma língua culta, no nosso caso, o português ao qual se subordina ou nele “varia” o *internetês*. Como vimos, a língua da Internet é um dialeto social, o qual, segundo Dubois (1978), é um sistema de signos usado por um grupo social e em referência a esse grupo. É o caso típico do *internetês*. Está claro que essa escrita de uma comunidade está inserida numa comunidade maior de usuários da mesma língua. Até pela etimologia de dialeto, podemos considerar o *internetês* um deles, porque *dialectus* era em latim “a linguagem particular de um país, maneira de falar”. Aqui, tanto tempo depois dos romanos do latim, é a linguagem particular dos jovens na Internet, a maneira de eles se comunicarem de forma própria na camada social em que estão inseridos e ligados por meio do computador. **Temos, assim, um dialeto de escrita que reproduz elementos da fala.**

---

<sup>80</sup> Há vários trabalhos abordando blogues, entre eles Komesu (2005), Di Luccio (2005) e Prange (2003). Acreditamos ser suficiente o contraste entre o léxico do *internetês* e o das redações escolares para obtenção das respostas que procuramos com nossa pesquisa. A falta de comparação com a linguagem dos blogues, acreditamos, não invalida nosso trabalho. Isso porque o texto do orkut parece ser mais “urgente” do que o do *blog*, além de ter a figura do interlocutor bastante mais marcada.

A língua da Internet e a língua culta são aspectos da variação sociolingüística relacionados à classe social, ao nível de escolaridade dos usuários como também ao registro utilizado em cada situação de comunicação. São, portanto, socioletos ou dialetos sociais. A Sociolingüística analisa essa variação vendo, em cada tipo, as estruturas utilizadas e, nelas, a relação entre a linguagem e o grupo social que a utiliza naquele momento específico. A base para distinção entre linguagem culta e popular (na Internet também) é o *status* social e o grau de escolaridade dos falantes. A esse respeito, Preti diz que o dialeto culto, ou linguagem erudita, é o de maior prestígio, reflete o nível de cultura do falante, é usado na literatura e se constitui na língua padrão. Enquanto que para a linguagem popular autor lista o seguinte:

Dialeto Popular: Subpadrão lingüístico; menor prestígio; situações menos formais; falantes do povo menos culto; linguagem e escrita popular; simplificação sintática; vocabulário mais restrito; gíria; linguagem obscena; fora dos padrões da gramática tradicional (PRETI, 1982, p. 32).

Como podemos ver nas indicações do autor, algumas características por ele lembradas como pertencentes ao dialeto popular também se aplicam ao *internetês*. Com isso, tivemos mais um indicativo para o classificarmos como um dialeto social. Vale destacar, porém, que o quesito “vocabulário mais restrito” parece não se aplicar ao *internetês*.

Não podemos esquecer que, se o sistema lingüístico serve a uma comunidade heterogênea e plural, ele também deve ser heterogêneo e plural. Assim, por exemplo, para servir à comunidade jovem na Internet, a língua se transformou também em *internetês* do Orkut. Isso não quer dizer que deixe de ser a língua portuguesa, coesa, mesmo com todas as modificações e inovações na grafia. Lembremos de Coseriu (1979, p. 19) ao nos dizer que “a língua não pode ser isolada de fatores externos – isto é, de tudo aquilo que constitui a fisicidade, a historicidade e a liberdade expressiva dos falantes”. Pelas novas necessidades advindas do uso do computador, a língua a ele se adaptou para acompanhá-lo na rapidez da comunicação que a máquina proporciona. A língua, dessa forma e num de seus usos, se adaptou porque ela é uma realidade essencialmente dinâmica, à imagem e semelhança da sociedade, daqueles que a usam, criam e recriam sem parar. Além do mais, as línguas mudam porque não estão feitas, mas fazem-se continuamente pela atividade lingüística. Coseriu (1982) lembra-nos ainda de que uma língua não é um único modo de falar, mas uma família de modos de falar afins e interdependentes, e que os dialetos são membros dessa família. Não nos resta dúvida, o *internetês* é um dos membros dessa família chamada língua portuguesa.

## 6.2. AVALIAÇÃO DAS HIPÓTESES

Passamos agora à avaliação hipóteses, lançadas no início desta pesquisa.

### 1) A variedade do vocabulário do *internetês* é similar à variedade do vocabulário da escrita culta em suas diferentes apresentações.

Confirmada. A variedade do vocabulário do *internetês* é semelhante à do vocabulário da escrita culta. Isso pôde ser constatado verificando a riqueza lexical dos *corpora*. O que num primeiro momento poderia parecer um baixo número de vocábulos diferentes (*types*) no *corpus* do *internetês*, logo foi esclarecido porque para estabelecermos uma comparação devemos observar o número de palavras diferentes em *corpora* de equivalente número total de palavras (*tokens*). Quando em *corpora* de mesmo tamanho, a variedade do vocabulário chegou a ser maior no *internetês* do que em reportagens, artigos científicos e redações escolares. Devemos levar em conta que aqueles 20% de vocábulos escritos de outra maneira que não a da norma oficial podem fazer a diferença, porque há algumas palavras que são escritas de formas variadas (MUITO, MT, MTO, MTU, MUTOO). Mesmo assim, em *corpora* similares no tamanho, o número de *types* do *corpus* formado apenas de *scraps* foi superior a todos os outros, resultando uma riqueza lexical também maior. Lembremos que uma característica do *internetês* é ter algumas palavras escritas de diferentes formas “abreviadas” (bj, bjo, bju, bjus, bjuxx), o que contribui para a variedade (riqueza) do vocabulário dos jovens.

Outro destaque que devemos fazer são as palavras de ocorrência única, os hápax, elementos lexicais importantes na constituição do vocabulário, porque aparecem em grande número e contribuem significativamente para a riqueza do texto. Verificando a percentagem de hápax nos *corpora* *Internetês*, Redações, Entalpia, Escrita Banco de Português, Fala Banco de Português e Fala NURC-RJ, o percentual de hápax do *corpus* *Internetês* (59%) foi menor do que o de Redações (62%), igual ao do Entalpia (59%) e menor do que o do Escrito Banco de Português (36%), Falado Banco de Português (39%) e Falado NURC-RJ (51%). Pela observação dos verbos e substantivos presentes nas 100 palavras mais frequentes do *Internetês* e dos *corpora* de fala e escrita (Anexos V, VI, VII e VIII), da mesma forma podemos perceber mais uma face da diversidade lexical da língua dos internautas no Orkut: o *internetês* apresenta tantos verbos quanto a fala e mais verbos (23%) do que a escrita. Em relação aos substantivos, porém, essa riqueza não se mantém quanto à escrita, em que esta apresenta 55% a mais de

substantivos do que o *internetês* e a fala. Afora esse detalhe, podemos afirmar que, quanto ao quesito variedade vocabular, o *internetês*, então, pode ser considerado tão rico quanto o da língua portuguesa na sua forma culta falada e escrita.

## **2) As frequências do léxico do *internetês* revelam uma junção entre padrões de fala e de escrita.**

Confirmada. No léxico do *internetês* percebemos haver coincidência de palavras entre as mais frequentes tanto da língua escrita quanto da falada. Há uma aproximação entre as palavras mais usadas da fala e as do *internetês*, embora muitas das palavras mais frequentes da escrita também estejam presentes entre as mais usadas pelos jovens na Internet. Das nove palavras mais frequentes nos dois corpora de língua falada, sete eram as mesmas encontradas no *internetês* (E, QUE, EU, A, DE, O, É). E todas as cinco mais frequentes do *corpus* escrito estavam entre as oito mais utilizadas no *internetês* (DE, A, O, E, QUE). Ampliando um pouco mais o levantamento, buscamos quais palavras se repetiam entre as trinta mais frequentes dos quatro *corpora* e obtivemos o seguinte resultado: 17 palavras dos *corpora* da fala estavam no *internetês* (E, QUE, EU, A, DE, O, É, PRA, UM, VOCÊ, UMA, MAIS, SE, NÃO, COM, DO, DA), e do *corpus* escrito, 15 delas (E, QUE, A, DE, O, É, UM, UMA, MAIS, SE, NÃO, COM, POR, DO, DA).

Como podemos observar pela lista das palavras mais frequentes tanto na fala quanto na escrita, a maioria são palavras funcionais, nenhuma delas é substantivo ou adjetivo, havendo apenas um verbo, o É. Procurando então ver se há uma correlação equivalente na lista em relação às palavras lexicais, estendemos nossa análise para as cem mais frequentes. O resultado mostra grande diferença entre as frequências, não havendo mais correspondência de palavras nas listas do português escrito e do *internetês* no que se refere às palavras lexicais. Isso é tão marcante que somente o substantivo DIA consta nas duas listas e apenas os verbos É, TEM, SER, FOI, PODE e TER. À medida que apareciam substantivos, advérbios e verbos entre as palavras mais usadas na escrita, desapareciam as “coincidências” de estarem tais palavras na lista do *internetês*. Das cem palavras mais frequentes apenas 35 eram as mesmas do *internetês*. Já das cem da língua falada, 74 estavam presentes na língua da Internet, aqui computando-se também as variantes Q, VC, EH, D, PQ, MTO, MT, I, T e NEH. O que se pode deduzir dessa constatação? Que há no *internetês* uma junção de fala e escrita quanto à frequência e variedade das palavras, com maior aproximação da língua falada.

Com a confirmação de nossas hipóteses, percebemos que ninguém precisa ter medo do *internetês*. Essa forma de escrever para se comunicar no computador não vai “assassinar” a língua. É apenas mais uma faceta da versatilidade da língua, se ajustando ao meio eletrônico para exercer sua função de estabelecer rapidamente comunicação entre os usuários da rede. Como na maioria das vezes o computador medeia a interação como se fosse carta e telefone ao mesmo tempo, temos, então, junção de fala e escrita. Afora o uso “diferente” de grafar 20% do seu léxico, o *internetês* é expressão da língua como qualquer outra, inclusive, com riqueza lexical igual ou superior à da fala e da escrita, segundo os dados de nossos *corpora*. Assim como todo dialeto tem suas características próprias, juntar fala e escrita é uma das particularidades do *internetês*, um dialeto diastrático espalhado por todo o país.

### 6.3. CONSIDERAÇÕES E PERSPECTIVAS

Os resultados desta pesquisa confirmam o quanto pode ser facilitada uma investigação sobre a descrição da língua com suporte informatizado, tendo-se adotado uma metodologia estatístico-lexical.

Após a revisão da literatura, formação de *corpora*, levantamento e análise de dados para descrever o *internetês* e verificar suas influências na escrita formal - principalmente na de estudantes - chegamos às conclusões. Essa forma de escrita na Internet:

1. é um dialeto diastrático, um socioleto, utilizado pelo estrato social (formado principalmente de jovens) que utiliza a Internet para se comunicar;
2. é uma escrita com certa similaridade com a língua falada, mostrando-se uma junção de fala e escrita;
3. possui algo apenas em torno de 20% de formas divergentes em seu léxico em relação à grafia oficial;
4. tem influência na escrita escolar, embora tenha revelado apenas alguns poucos traços nas redações do nosso *corpus*. Não podemos afirmar com certeza se alterações (“erros”) de grafia se devem ao *internetês* ou à não internalização das normas ortográficas. Características típicas da escrita na Internet como uso de *emoticons*, marcas de intensidade e excesso de exclamações pouco apareceram nas produções escolares.

5. possui uma conformação vocabular que permite ser explorada em sala de aula para que o aluno tenha consciência de que é uma expressão legítima da língua, mas que é adequada apenas no contexto da Internet.

Analisemos, agora, os indicativos anteriormente enumerados.

**1. O internetês é um dialeto diastrático** como já visto na resposta à questão de pesquisa número 4. É uma variação diastrática ou sociocultural, um socioleto, usado por grupo de falantes em diferentes locais, não se atendo a um limite físico-geográfico. Isso foi possível conferir ao formar nosso *corpus* com textos de distintas regiões do Brasil.

É um tipo de linguagem especial criada e utilizada por um grupo de pessoas unidas por algo em comum (faixa etária, estilo de vida ou comportamento social), apresentando uma variedade lingüística distinta da linguagem corrente. Dessa forma, partimos do pressuposto de que o desempenho lingüístico e o comportamento social estão perenemente relacionados de forma dialética. Pelo que foi exposto até aqui, podemos afirmar que na formação e no estudo de uma língua é necessário levar em consideração a influência que os aspectos sociais – relacionados ao usuário ou ao ambiente – exercem sobre ela, uma vez que a língua é produto e, ao mesmo tempo, veículo das experiências sociais. Palavras ou formas lingüísticas são aceitas como elementos da língua quando passam a exprimir os valores de um determinado grupo social e, sobretudo, a satisfazer suas necessidades de comunicação. Assim, a atuação social dos falantes favorece o surgimento de alterações e acréscimos à língua resultando na ampliação lexical por meio de vocábulos e expressões na linguagem comum, ou ainda no aparecimento de linguagens especiais utilizadas restritamente, como no caso do *internetês*.

**2. É escrita com similaridade com a fala, junção de fala e escrita.** Mesmo no Orkut, numa comunicação assíncrona, os internautas sentem-se numa interação falada por escrito. Isso aparece nos recados deixados como se fossem frases escritas nos *chats on-line*, ou ditas ao telefone. Contêm muitas marcas dos gêneros orais. Notamos em alguns textos certa tensão que leva quem escreve à digitação apressada, como se estivesse em presença ou sob o olhar do interlocutor. Isso provoca grafias estranhas, sem razões explícitas ou qualquer preocupação de revisar antes de enviar pela rede. É como se estivesse falando, escreve como se dispusesse do mesmo tempo da fala para produção de seu fluxo verbal.

Recursos não-verbais utilizados na fala face a face (gestos, mímica, entonação) aparecem “compensados” no ciberespaço. Assim surgiram, entre outros, os *emoticons*, abreviações, redução de palavras, neologismos, palavras estrangeiras, letras maiúsculas para gritar, repetição de letras para indicar intensidade, uso excessivo de sinais de pontuação, etc.

No *internetês*, há uma quebra das concepções tradicionais que polarizam oralidade e escrita. Onde termina o limite de uma e começa o da outra? Pelo que vimos nos *corpora*, é uma mescla de ambas, uma conversa escrita, com características próprias, porque pelo teclado a comunicação tem que ser por escrito. Mas uma escrita que procura reproduzir a fala. Os internautas, uma vez garantida a interação, tendem a se livrar da coerção da codificação da língua escrita, recodificando-a num xadrez de fala e escrita. Marcas que confundem as duas modalidades aparecem, entre outras, nas frases curtas e truncadas, sem abertura ou fechamento, falta de pontuação, operadores conversacionais, alto tom de informalidade e descontração. Não há dúvidas de que, embora escritos, os textos do *internetês* também põem em uso a modalidade da fala. É uma nova articulação das linguagens oral e escrita, porque mediados pelo computador, os jovens querem conversar mas se vêm compelidos a escrever. Sobre isso, Pereira e Moura (2006, p. 82) afirmam:

Procuramos situar essa escrita num *continuum* em que linguagem falada e linguagem escrita se fundem, dando origem a uma forma mais farta e potencialmente mais completa de escrita. Não se trata de suprimir uma ou outra modalidade de linguagem, mas de condensá-las e redirecioná-las, produzindo um novo estilo de linguagem que vai além da escrita alfabética, englobando os sistemas logográfico e semiótico.

**3. O *internetês* no Orkut exhibe apenas em torno de 20% de formas variantes não oficiais no seu léxico.** Num primeiro levantamento com as 2000 palavras mais freqüentes do *corpus* geral (depoimentos e *scraps*) percebemos a presença de 21,19% de palavras escritas em desacordo com a grafia oficial. É principalmente essa grafia que chama a atenção enfurecendo puristas e desavisados sobre as lentas e constantes transformações da língua. Biber (1998) nos lembra que nossa mente tem sua atenção chamada pelo atípico, não se apegando ao que é típico. E Bagno (2002, p. 133) nos alerta haver uma “tradição que atribui ao domínio da escrita um elemento de *distinção social*, que é na verdade um *elemento de dominação* por parte dos letrados sobre os iletrados” (grifos do autor).

Chamou-nos a atenção o número da percentagem de variantes de grafia não oficiais da escrita na Internet. Pensávamos que fosse bem maior, uma vez que a queixa difusa, mas generalizada, sobre essa grafia é de que “os jovens escrevem *tudo* diferente”. Esse *tudo* está longe de ser verdadeiro, porque há 80% de palavras registradas de acordo com a escrita convencional. Ainda intrigados com tal resultado, pesquisamos, então, num recorte menor de nosso *corpus*, observando na lista as 500 palavras mais usadas do *corpus* geral (Anexo IX) para ver se havia maior incidência de variantes. Constatamos a presença de 93 palavras em desacordo com a norma oficial. Esse número indica percentual de 18,6% de variantes entre as 500 palavras mais usadas em depoimentos e *scraps* no Orkut. Pensamos, então, que ao escrever *scraps* os jovens ignorariam mais a grafia oficial, cometendo mais “erros”. Fizemos o levantamento (Anexo X) e constatamos 89 palavras “diferentes”, o que dá 17,8% do total de 500. O que concluir disso? Que há liberdade usada nas variações, sim, mas que não chegam a comprometer toda a grafia. Há, enfim, a língua portuguesa identificada e preservada na nova grafia. Como vimos nas características do *internetês* na **Seção 5.3 Alterações de grafia** (p. 186), há algumas regras possíveis de serem estabelecidas, e que muitos vocábulos não podem ser modificados sob pena de se tornarem ilegíveis. A lista de frequência tem mostrado que nas variantes do *internetês* há monossílabos e dissílabos, na maioria sendo reduções de palavras maiores para esse limite de sílabas. Tudo, provavelmente, em decorrência da pressa para teclar.

Outros elementos do *internetês*, como uso excessivo de sinais de pontuação, *emoticons*, repetições de letras para indicar intensidade, salientam-se muito, mas parece que não tanto quanto a grafia. Na verdade, todas as características dessa forma de escrita a tornam algo estranho, que deve ser combatido segundo alguns. Pelos dados observados, os “defensores do idioma” não precisam temer tanto: não serão apenas esses 20% de modificações gráficas que irão mudar o português. O que é preciso lembrar sempre é que há modos e modos de dizer (escrever) adequados para o que se quer dizer, para quem se quer dizer, onde se diz e que efeito se quer produzir.

**4. A influência do *internetês* é pequena na escrita escolar.** Na escola os alunos têm contato com a escrita padrão em todas as matérias de estudo. Logo, esperamos que esse modelo de escrita seja seguido e realizado nas suas produções. Se o jovem pouco produz nas aulas e muito escreve na Internet, qual a representação gráfica internalizada que lhe fluirá espontaneamente papel afora? É preciso, então, um trabalho de conscientização e reflexão

como anteriormente dissemos: saber o contexto em que se pode escrever ou não certas coisas e com que tipo de escrita. É preciso que, assim como o professor, cada aluno se torne um investigador, abandonando as atitudes repetidoras e reprodutoras de pensar que estuda (a língua) quando na verdade apenas tem contato com a gramática normativa, muitas vezes perdendo tempo com conteúdos dos quais não se lembrará um dia após realizar uma prova sobre o assunto estudado. Que conteúdos são esses que os alunos estudam somente para a prova, não vendo sentido em estudar para aplicar na vida?

É preciso ter presente que o ensino tem que evoluir, o próprio Ministério da Educação mostra pelos *Parâmetros Curriculares Nacionais* que se deve fazer uma renovação do ensino da língua portuguesa. A língua é dinâmica, modifica-se, **contém** variedades lingüísticas e é um instrumento de poder, um nível ideal de padrão a ser alcançado, por isso, também e principalmente, deve ser estudada. Então não é brigando com o *internetês*, nem o negando – afinal, ele veio e vai ficar na cibercultura – que estaremos ensinando, estudando ou aprendendo português.

Nos PCN- Ensino Médio (1999, p. 134), lemos que “a mais nova das linguagens, a Informática, faz parte do cotidiano e do mundo do trabalho. [...] Conviver com todas as possibilidades que a tecnologia oferece é mais do que uma necessidade, é um direito social”. Na página 129, o mesmo documento registra como algo importante que “o aluno saiba analisar as especificidades, sem perder a visão do todo em que elas estão inseridas, e perceba que as particularidades têm um sentido socialmente construído”. Essa afirmação refere-se às diferentes Áreas do Conhecimento, mas podemos estendê-la à língua com o seu *internetês*, até porque o primeiro objetivo da Área *Linguagens, Códigos e suas Tecnologias* é este: “Analisar, interpretar e aplicar recursos expressivos das linguagens, **relacionando textos com seus contextos**, mediante a natureza, função, organização das manifestações, **de acordo com as condições de produção e recepção**” (Grifo nosso) (PCN-EM, 1999, p. 126).

Marcuschi supõe que o maior problema da comunicação internetiana não está no tipo de escrita, mas “nos desafios cognitivos, no acúmulo de informações e na necessidade de maior formação para enfrentar problemas de compreensão. Percebe-se com nitidez que a Internet não é perigosa nem ameaça a Língua Portuguesa” (MARCUSCHI, 2005, p. 10). O autor ainda lembra que fatores sociais, culturais, históricos e tecnológicos exercem papel importante sobre os usos da língua. Com certeza é o caso da Internet influenciando a maneira

de pensar, de escrever e de agir. Ela “incentiva ações novas que permitem profundas mudanças sociais, de um lado, e o surgimento de novos modos de operação cognitiva, de outro lado” (MARCUSCHI, 2005, p. 10). É muita pobreza as pessoas se preocuparem apenas com a escrita do *internetês*, hipervalorizando a forma gráfica, sem se importar tanto com o conteúdo e outras influências que o mundo dos computadores exerce, principalmente, sobre os jovens em formação.

Nossa pesquisa com textos produzidos em escola que tem a prática semanal de aula de redação mostrou pouquíssima influência da escrita internetiana. Houve algumas palavras ou formas num ou noutro texto, assim como apareceram outras formas de grafias não oficiais, os tradicionais “erros de grafia” ou, mais que isso, os erroneamente chamados “erros de português”. Como se ortografia fosse toda a língua ou a coisa mais importante dela!

Apesar da polêmica exagerada com o *internetês* – melhor dizendo, com sua grafia – o aluno escreve com gosto, vendo sentido e tendo motivação para redigir. Nunca os jovens escreveram tanto como no tempo das correspondências eletrônicas. Ao se utilizar o *internetês*, o indivíduo, além de usufruir de um diálogo e interação ricos, se torna um “poliglota”, desenvolvendo uma capacidade de se comunicar de forma diferente. Basta apenas se conscientizar de que a forma de escrita de tal comunicação só é adequada no computador.

**5. Trabalhar o *internetês* em aula.** Desmitificar o *internetês* também na escola. Acreditamos ser essa mais uma tarefa do professor, porque mesmo os jovens confundem português com ortografia, uma idéia corrente entre a população em geral. Em geral, chamam “erro de português” a meras incorreções ortográficas. Nem erro de gramática é, gramática entendida como regras de funcionamento da língua. Até um analfabeto sabe e usa bem a gramática, sem saber nada de ortografia. Entender que a língua é um organismo vivo, dinâmico, com alterações, criações e recriações. Que a língua não é algo estanque como pode nos fazer crer a gramática normativa que a trata como uma coisa morta, não considerando que é falada pelos vivos numa sociedade em constante transformação.

Levar a vida para a sala de aula. Refletir a língua voltada para seu uso real. Isso se realiza quando são utilizados principalmente textos reais, partindo do “pressuposto de que a língua se realiza no uso, nas práticas sociais e de que os indivíduos apropriam-se dos

conteúdos, transformando-os em conhecimento próprio, por meio da ação sobre eles” (PCNs, 2000, p. 43).

Não esquecer que todas as modalidades lingüísticas são legítimas desde que adequadas ao contexto, aceitando variações sociais, usos regionais, variações determinadas ou condicionadas pelos contextos. Concordamos com Scarton (2007, p. 17) quando se refere ao papel da escola: “Tornar o falante plurilíngüe no interior da própria língua: respeitar/apoiar todas as modalidades lingüísticas dos alunos e, sob a forma de acréscimo, ampliar a competência comunicativa com o favorecimento da linguagem culta formal”. Ao se referir ao papel do professor, o mesmo autor é incisivo:

Ver a língua como uma realidade essencialmente dinâmica, à imagem e semelhança da sociedade, daqueles que a usam, criam e recriam sem parar; ver na língua um conjunto heterogêneo de variedades, cuja igualdade e legitimidade reconhece; finalmente, ver na língua, pelo domínio de recursos de expressão, a possibilidade de levar os alunos ao acesso aos bens culturais, à articulação desses bens e desse saber aos interesses de sua classe são atitudes que devem caracterizar o professor de Português, são atitudes que o professor de Português deve manifestar na prática da sala de aula.

A realidade marcada pela diversidade lingüística já é reconhecida pelas instituições oficiais que planejam a educação no Brasil. Vemos nos *Parâmetros Curriculares Nacionais* (1998, p. 29) que

a variação é constitutiva das línguas humanas, ocorrendo em todos os níveis. Ela sempre existiu e sempre existirá, independentemente de qualquer ação normativa. Assim, quando se fala em “Língua Portuguesa” está se falando de uma unidade que se constitui de muitas variedades. [...] A imagem de uma língua única, mais próxima da modalidade escrita da linguagem, subjacente às prescrições normativas da gramática escolar, dos manuais e mesmo dos programas de difusão da mídia sobre o “o que se deve e o que não se deve falar e escrever”, não se sustenta na análise empírica dos usos da língua.

Há que se pensar a língua como um conjunto heterogêneo e dinâmico. Ela, com o tempo e a contínua variação, sofre mudanças ainda não presentes nas prescrições apresentadas nas Gramáticas Normativas utilizadas pelo professor com seus livros didáticos. Esse tipo de aula torna o aluno um mero receptor, como se ele fosse uma tábula rasa, o discente nada construindo, apenas reproduzindo seu próprio conhecimento com base na repetição.

Então a escola não deve ensinar gramática normativa, nem se preocupar com a norma culta? Não! Ao contrário, devemos ensinar a norma-padrão, até porque só se ensina algo que não se conhece (e a língua materna o aluno já conhece). De acordo com Bagno (1999, p. 179),

ensinar o padrão se justificaria pelo fato dele ter valores que não podem ser negados – em sua estreita associação com a escrita, ele é o repositório dos conhecimentos acumulados ao longo da história. Esses conhecimentos, assim armazenados, constituiriam a cultura mais valorizada e prestigiada, de que todos os falantes devem se apoderar para se integrar de pleno direito na produção/condução/transformação da sociedade que fazem parte.

Reforçando: a gramática normativa, que se ocupa da norma-padrão, entendida como um conjunto de normas que nos orientam no uso padrão-culto quando este nos for exigido. Como tal deve mesmo ser priorizado nas escolas, mas não deve ser o único elemento de estudo. Não acreditamos que o simples ensinar de regras gramaticais ajude o aluno a falar e escrever bem. Há outros padrões privilegiados como práticas lingüísticas, que também precisam ser valorizadas nas escolas. Não podemos esquecer, no entanto, de lembrar o poder dado ao falante pelo domínio da norma culta. O autor que defende de forma enfática a língua como poder é Dacanal, reconhecendo ser toda língua “um conjunto mais ou menos amplo de símbolos sonoros convencionados” (2006, p.43). Destaca-se em Dacanal a importância que dá ao convencionado na língua, ao caráter de convenção, em outras palavras, à arbitrariedade do signo. Essa convenção oficial, no que diz respeito ao *internetês*, está sendo desrespeitada por seus usuários. Mexer com o convencionado gera polêmica porque “a língua dominante é – ou tende a ser – sempre a língua daquele que detêm o poder econômico, social e político” (DACANAL, 2006, p.48).

Acreditamos que este trabalho cumpriu os objetivos a que se propôs: descrever a escrita em uso na Internet, principalmente pelos jovens. Analisamos o léxico empregado, partindo do quantitativo para o qualitativo a fim de elencar características e peculiaridades desse emprego da língua. Sobre descrição, Coseriu (1982, p.138) nos diz que ela “não é alheia à história, como tão frequentemente se pretende, mas está incluída nela, pois a descrição de um objeto, no momento da sua história, pertence, naturalmente, a essa história”. Diz-nos mais, em seguida, o autor: “no caso da língua, se o funcionamento implica a possibilidade da mudança, a descrição do funcionamento e das suas condições deve justificar tal possibilidade” (ibidem).

Pelos dados obtidos – só possíveis em tal número pela utilização do instrumental metodológico da Lingüística de *Corpus* – vários aspectos pouco ou nada estudados/divulgados foram aclarados. Um deles é a qualificação do *internetês* como um dialeto diastrático ou socioleto, outro é que seu léxico possui apenas algo próximo a 20% de palavras com formas divergentes da grafia oficial, o que não o torna algo tão pernicioso à língua escrita culta.

Durante a realização desta pesquisa – para embasá-la – foi de fundamental importância aprofundar os conteúdos vistos na Revisão de Literatura. Tais noções básicas deveriam ser seguidamente lembradas por todos os professores de Português. Muitas vezes as críticas ao *internetês* não têm embasamento científico algum, realizadas mais pelo estranhamento das formas em desacordo com a “boa norma” do que por outros motivos (sérios). Lembremos que são as diferenças o que sempre choca as pessoas. Como falar de um novo uso (que choca/causa estranhamento) da língua sem entender o que seja *língua, palavra, léxico, norma, variante, dialeto e escrita*? Equacionar essas noções básicas foi essencial para entender a estrutura do funcionamento do uso da língua na Internet, uma faceta da língua adaptada às circunstâncias exigidas para a comunicação utilizando o computador. Sobre adaptações e alterações, é oportuno ainda lembrar Coseriu (1979, p.31), para quem toda língua muda “para continuar funcionando como tal. O latim de Cícero deixou de funcionar como língua histórica justamente porque deixou de mudar; e nesse sentido é uma ‘língua morta’. [...] Em compensação a língua viva não permanece nunca em repouso, está em contínua transformação”.

Coseriu (1979, p.106) lembra-nos também que “a língua *está em fazimento* a todo instante. Um sistema lingüístico desde que realizado em formas tradicionais, longe de ser ‘por definição equilibrado’, é, pela sua própria natureza, um sistema ‘imperfeito’ (no sentido de ‘não terminado’)” (grifos do autor).

Em vista disso tudo e da nossa observação empírica, resta-nos aceitar o *internetês* do Orkut como um legítimo uso da língua que deve ser mais estudado, inclusive em sala de aula, e não temido ou abominado. Pelo que pudemos constatar em nossa pesquisa, a escola que cumprir sua missão de trabalhar bem a norma padrão, centrando atividades em leitura e produção de textos, terá menos problemas com a interferência da escrita internetiana. Pelo menos por enquanto. Nada nos assegura que alguns casos, que algumas palavras venham a ter oficializada a sua forma de escrita simplificada para atender à praticidade da comunicação. Não esqueçamos de que *você* era *Vossa Mercê* há alguns séculos, quando a escrita estabelecia comunicação via barco ou lombo de burro. Hoje, via satélite, é aceitável, então, pela rapidez dos contatos, que haja várias modificações, interferências e alterações no constante *fazimento* da língua.

É claro que esta pesquisa não pôde atender a muitas das várias questões que se apresentam quando se aborda um assunto tão instigante quanto o *internetês*, seja no Orkut ou em outros “recantos”. São muitos os caminhos, as indagações e as possibilidades para estudar o

texto eletrônico. Há muito por fazer ainda. Optamos pelo posicionamento teórico da Lingüística de *Corpus*, que cobre alguns aspectos, não atingindo todos. Muito mais deveria ter sido feito, mas esta pesquisa foi o melhor que pudemos fazer, dentro das condições possíveis, do tempo disponível e com os *corpora* que reunimos. Eles poderão ter novos aproveitamentos, afinal, são 553.875 palavras de *scraps* e depoimentos do Orkut divididos por regiões do Brasil. Devem servir para indicar mais elementos de como os jovens se comunicam no suporte tela de computador, um espaço aberto, novo lugar da escrita que vai além da folha de papel e que utiliza formas verbais e não verbais.

Se tivéssemos escolhido outra direção, laboviana ou funcionalista, teríamos, provavelmente, outras constatações, mas tivemos que fazer uma escolha. E a Lingüística de *Corpus*, acreditamos, forneceu-nos bons caminhos para a busca de respostas para o que procurávamos, confirmando nossas hipóteses.

Para o futuro do trabalho, para continuação do que aqui pesquisamos, muitas possibilidades se apresentam. Por exemplo, ver o *internetês* sob a perspectiva do texto e do discurso, o vocabulário relacionado, uma vez que a frase, a sintaxe, a concordância, a estrutura morfológica, a neologia e o texto dos *scraps* e depoimentos não foram examinados. Ficamos apenas no âmbito da palavra e, mesmo assim, nem analisamos o interior delas, suas formas de composição ou derivação, por exemplo. Além dessas, há muitas outras possibilidades de estudo da linguagem da Internet, uma vez que utilizamos aquela presente no Orkut. Podem ser feitas comparações ou estudos com a escrita dos *blogs*, dos e-mails e dos *chats*, a fim de estabelecer, entre outros, o *continuum* tipológico, do informal ao formal, entre fala e escrita. Há muito o que explorar nos diferentes gêneros textuais que ocorrem na Internet.

Ainda com os pressupostos e o ferramental da Lingüística de *Corpus* poderíamos verificar outros aspectos da língua utilizada pelos jovens na Internet. Um deles seriam as metáforas utilizadas pelos internautas, uma vez que já há sistemas informatizados que reconhecem candidatos a metáforas em um *corpus*<sup>81</sup>. Outros seriam as fraseologias empregadas e as colocações recorrentes. Isso utilizando-se apenas a Lingüística de *Corpus*. Pensemos, então, nas outras perspectivas, como a Análise Crítica do Discurso e a da Enunciação, no quanto sob elas pode render esse objeto de comunicação escrita chamado *internetês*.

---

<sup>81</sup> Vide <http://www2.lael.pucsp.br/corpora/metaphor/index.html>.

O estudo desta dissertação foi apenas um passo para entender melhor como se apresenta um componente da linguagem dos jovens na Internet. Como já dissemos, há muitos outros fatores a serem observados. Há a necessidade de dar continuidade ao estudo sobre essa nova forma de expressão escrita, cada vez mais usada porque cada vez mais os computadores se disseminam na população. E no computador, pelo menos na comunicação entre jovens, se utiliza um texto especial, objeto que estudamos e que precisa ser ainda muito mais estudado, o *internetês*.

Ainda não existe um "internetês" (ai!) semelhante ao economês, mas não é impossível que cheguemos lá, graças, por exemplo, ao uso sistemático de abreviaturas e de interjeições. O importante, porém, é que a Internet está de algum modo recuperando a palavra escrita entre os jovens - e ajudando a difundir textos literários. Mas esse é ainda um veículo muito novo e de alcance restrito. Precisamos esperar para ver o que vai acontecer. E precisamos estudar melhor as possibilidades de seu aproveitamento.

(SCLIAR, 2000, p. 6.)

Vejamos agora, no próximo capítulo, algumas sugestões para o professor de Língua Portuguesa aproveitar os dados desta pesquisa.

## 7 SUGESTÕES PARA O PROFESSOR: O *INTERNETÊS* NA SALA DE AULA

Tendo em vista a preocupação de como a Escola agir diante do *internetês*, sugerimos algumas atividades neste capítulo. Elas poderão despertar outras idéias de como utilizar em aula a forma de comunicação dos jovens na Internet.

Os Parâmetros Curriculares Nacionais, tanto de Ensino Fundamental (1998) quanto do Ensino Médio (1999) se referem à importância dos recursos tecnológicos na sociedade contemporânea. Como a escola é e faz a sociedade, deve levar para a sala de aula tais recursos, notadamente os relacionados a tecnologias da comunicação e da informação. “Elas fazem parte da vida das pessoas, não invadem a vida das pessoas” (PCN-Ensino Médio, 1999, p. 133). Essas tecnologias, além de informações, “possibilitam novas formas de ordenação da experiência humana, com múltiplos reflexos, particularmente na cognição e na atuação humana sobre o meio e sobre si mesmo” (PCN, 1998a, p. 135). O documento ainda registra que há pouco tempo não existia possibilidade de comunicação *on-line*,

nem de compartilhar imagens instantaneamente em vários lugares do mundo, assim como não era possível conceber que uma pessoa pudesse aprender tendo como interlocutor uma máquina, como é o caso da aprendizagem intermediada pelo computador. Essas mudanças [...] geram transformações na consciência individual, na percepção do mundo, nos valores e nas formas de atuação social. (PCN, 1998a, p. 136)

Os PCN-Língua Portuguesa (1998) se referem superficialmente à relação das Tecnologias da Informação e a Língua Portuguesa, limitando-se a sugerir propostas de trabalho com computador, CD-ROM, multimídia, hipertexto, rádio, televisão e vídeo. Nos PCN-Língua

Estrangeira (1998b), porém, há análise da leitura e escrita no meio digital. O documento lembra que, por exemplo, o conceito de leitura “passa a ser primordialmente exercício de uma opção de trajetória pela página<sup>82</sup> e a subsequente aquisição seletiva de informações parciais presentes em diversos locais na mesma página” (PCN-LE, p. 105). O leitor, de certa forma, assume o papel de autor do texto lido, uma vez que vai clicando em *links* e construindo seu texto, “a leitura passa a ser algo seletivo, parcial, dependendo do interesse ou do objetivo do leitor” (PCN-LE, 1998b, p. 105). Não é de estranhar, portanto, que a escrita na Internet também reflita, de alguma forma, essa fragmentação, e tenha características peculiares. Como leitura e escrita são duas faces da comunicação no computador, esses exercícios de liberdade para escolher o que ler, utilizando hipertextos, podem se refletir na liberdade que o internauta julga ter para escrever também como quiser. Ele faz registros gráficos em desacordo com a ortografia, desde inusitados usos de letras para representar fonemas (axim = assim), diferentes abreviaturas e truncamentos de palavras (cmg = comigo, namo = namorado, dolu = adoro), palavras estrangeiras (brother), até junção de letras e caracteres para formar *emoticons*.

As propostas que apresentamos a seguir abordam diferentes atividades e habilidades, sendo recomendadas para alunos do 1º ano do Ensino Médio. Nada impede, porém, que sejam aplicadas em outras séries. Das atividades, as mais simples envolvem observação de grafia empregada no *internetês*, pesquisa e transposição para a grafia oficial. Sugerimos, também, a reescritura de frases e de textos conforme a norma culta, pesquisa, análise e interpretação de palavras e textos, análise do funcionamento do *internetês*, leitura e produção de texto, participação em fórum de debates com posterior produção de texto. O foco, enfim, é o do aproveitamento dessa modalidade de escrita nas aulas de língua portuguesa.

- Como a interação entre internautas se faz por uma escrita que mostra traços de fala, na maioria das vezes há muitas palavras repetidas, falta de conjunções ou de outros recursos para fazer a ligação entre as idéias, frases incompletas, enfim, é uma comunicação que se faz de uma forma mais espontânea, descuidada até. Isso ocorre porque na fala se pode responder em seguida e esclarecer o que não ficou claro, e na Internet também quando se estiver *on-line*. Na escrita tradicional, o leitor vai ler bem depois da produção feita, o que exige de quem escreve maior cuidado, maior clareza para transmitir com precisão e fidelidade o pensamento. Um trabalho a ser feito com os alunos é a reescritura de depoimentos e *scraps*, que podem ser

---

<sup>82</sup> Página digital.

fornecidos pelo professor e coletados pelos próprios estudantes. Como exemplo, podemos sugerir o seguinte:

### **ATIVIDADE 1 Reescritura e produção de texto, observação de grafia**

*O exercício a seguir consiste em refazer os scraps abaixo conforme a escrita tradicional. Escreva aquilo que você entendeu, porque certos recados são entendidos totalmente apenas no contexto do Orkut, como resposta a outro recado. Utilize as normas da gramática, empregando a grafia oficial, a pontuação adequada e fazendo as concordâncias (plural, sujeito + verbo, etc.). Evite o emprego de muitos que e a repetição de palavras, utilizando no lugar delas sinônimos ou pronomes (ele, esse, aquele, o mesmo, tal, algum...). Procure observar a adequação dos tempos verbais. Não esqueça de relacionar as idéias principais, jamais as deixando “soltas”, isto é, apenas agrupadas numa seqüência. Elas devem ser ligadas por expressões como desde que, por isso, por outro lado, assim sendo, desta forma, ainda que, a menos que, visto que ou então utilizando orações reduzidas (Ex.: Acreditando estar certa, tomou a atitude drástica).*

eh...  
 eu acho ki foi a gabi q apagou os scrap dela..  
 num sei pq..  
 mais..  
 cade vc??  
 nuss cheguei agora...  
 hihi..  
 to com sdd..  
 nao entendi..  
 ontem tava brigando cmg..hj deixou um scrap..  
 pq ?  
 =[[[[[[[[[[  
 bjux  
 te amo

bixu !  
 eh ki vaum ter muitos concurços aki por agora.. i u povo la du meu curço ta otimista ! pq ano retrasado concurço do tj aki xamou ate mais ki u 900 expedente ! aew eu toh aki na luta num eh neim pra passar entre us 400.. eh pra fica pelo menos em 1000 ! hehe  
 mais eh dificiu dimais... soh nu meu curço lah tem mais de 800 pessoas estudanu.. assim logico ki em turmas diferentes neh ¬¬  
 mais u kra naum podi fica pensanu como eh dura a realidade .. tem ki dream !  
 sempre realçando aki u meu ingles quasi fluente  
 a prova aki .. u povo ta axando ki vai ser dia 24 de junho.. soh sei ki voh fazer a prova i  
 Amando!!!!  
 E vc tah aonde?? Fazendo oq??? tah gostando???  
 Bjão

*david pediu pra tu ir direto pro jogo pra encontrar com o pai dele lah, pq guardar 3 lugares na parte superior da social ateh 9:00 eh pau pow ateh pq soh vai o pai de david e a irma dele.  
caiu nada...td mundo cai  
Cientistas norte americano estão dizendo q pessoas retardadas da kbça, lêem os scrap com a mão no mouse !!!  
não adianta tirar agora...kkkkkk  
eu tbm cai nessa, passa para os amigos.  
huahuahuahua hehehehe rrsrsrs hahahaha*

*aoooooo renan....puta tempo eim...pois é, finalmente sai da amer. Entrei numa multinacional aki em pira...da hora pra caraio.  
Dai vim morar aki...mais sucegado eim!!!!  
e vc ta em campinas ainda????  
eu mudei de casa tbem, to morando perto do centro de lazer agora...meu pai vai fazer um churrasquinho logo logo la, dai eu dou uma ligada pra vc falow...sua presença eh essencial!!!  
falow veio  
abração*

• Um outro trabalho a ser realizado com alunos, que pode atingir diferentes faixas etárias, é o de refletir sobre a adequação do tipo de texto ao contexto social. Assim, podemos apresentar-lhes uma tarefa como esta:

## **ATIVIDADE 2 – Transposição do *internetês* para língua culta, análise de grafia**

*Algumas formas de comunicação só são adequadas a determinados ambientes. Ninguém solicita algo a uma autoridade numa carta cheia de gírias, por exemplo. Saber se comunicar é usar a linguagem adequada a cada contexto (ambiente). Assim, entre jovens é aceitável usar gírias e outras expressões informais, que não devem ser empregadas em documentos ou qualquer tipo de provas e textos da escola, por exemplo.*

*O exercício a seguir consiste em transpor para a linguagem culta (com pontuação, acentuação, concordâncias, etc.) o que você acha que o autor da frase quis dizer ao se comunicar no Orkut:*

*Um grande xeirão pra vc.  
Vc eh Mtuuu Gatu, Mas Tenhu Namó!  
boi na linha aki tb amiguinhoóóóó!!!  
ae pateta, fds muito show heim.  
para de ficar xeretando no meu orkut  
puta tempo eim!!!!  
qse q eu vim aki pirar o kbçote hen...  
td bm?  
Mó correria*

*qlq coisa da o tok*  
*xD...Ow meeu...fuui mimi taava cansadonaaa...\**  
*bjin migowwwwww*  
*voltai um dia pra gente zuar porra*  
*Pow vc ta sumido hem!! To cum xodad!!*  
*fala ae troxa blz maluco*  
*comenta eu aee... pliiix*  
*...eh mto bom irritar ela ao cubo*  
*e a iiii loko so na boa rrsrsrs*

• Uma característica do *internetês* é apresentar muitas abreviaturas ou reduções de palavras para melhor aproveitamento do tempo e do espaço. Muitas vezes somente pelo contexto em que aparecem podemos identificar seu significado. O próximo exercício exige que o estudante observe, descreva e procure explicar a lógica para a modificação de tais palavras.

### ATIVIDADE 3 - Análise e interpretação, observação de grafia

Observe as palavras “abreviadas” segundo critérios criados pelos usuários da Internet. O significado delas está entre parênteses. Depois, escreva respostas para o que é solicitado.

<i>blz (beleza)</i>	<i>msm (mesmo)</i>	<i>rpz (rapaz)</i>	<i>ngm (ninguém)</i>
<i>cm (como)</i>	<i>flw (falou)</i>	<i>pq (porque)</i>	<i>gnt (gente)</i>
<i>td (tudo)</i>	<i>bj (beijo)</i>	<i>cmg (comigo)</i>	<i>dps (depois)</i>
<i>hj (hoje)</i>	<i>tbm (também)</i>	<i>qlq (qualquer)</i>	<i>fds (fim de semana)</i>

1) A maioria das palavras é composta por quantas sílabas?

2) O que foi retirado das palavras para abreviá-las?

3) Por que você acha que é possível entender o significado delas, mesmo sendo abreviadas de forma diferente daquela oficial?

4) Você acha que determinadas formas diferentes de escrita acabam se consagrando como “normais” com o uso repetido? Que outros exemplos de abreviaturas da Internet estão de acordo com a regra (vista neste exercício) criada pelos internautas para reduzir palavras?

• A ortografia brasileira é o conjunto de regras oficiais que determina a escrita da língua. Nossa ortografia tem um vínculo histórico e social com a variante culta do idioma.

Tradicionalmente, as regras ortográficas consideram apenas essa variante e não existe nenhuma ortografia oficial para as demais variantes da nossa língua.

No Brasil, as regras ortográficas oficiais têm sido definidas pela ABL (Academia Brasileira de Letras) e por iniciativas do Ministério de Educação<sup>83</sup>. As lacunas deixadas pela documentação ortográfica oficial são preenchidas por outros agentes da língua, como gramáticos, lexicógrafos, editores e estudiosos variados. Mais recentemente, ganharam importância os manuais de estilo das grandes editoras e dos jornais de grande circulação, que servem como referência em questões ortográficas, principalmente nas formações recentes. Há, enfim, hoje, diferentes “discursos” normativos que vão de instituições, publicações até sites de professores de língua portuguesa ou colunas de jornais de professores “famosos”.

Sobre o *internetês*, é necessário fazer a apreensão desse objeto cientificamente. Quem se pronuncia sobre o *internetês*? A ABL, que nada fez ainda? Professores que têm colunas “oráculo” na imprensa em jornais e revistas de amenidades (como Deonísio da Silva em *Caras*) dão suas opiniões, às vezes sem embasamento teórico, e passam a ser repetidas como verdades. No Brasil, parece-nos, não há manifestação muito clara do que seja essa nova forma de se comunicar por escrito na Internet.

A língua escrita é uma forma de comunicação que deve ser acessível a todos, por isso foram criadas as várias regras, inclusive de ortografia, para dar certa uniformização. O exercício a seguir propõe reflexão sobre a grafia, observando regularidades e registrando-as pela criação de regras.

#### **ATIVIDADE 4 – Análise e interpretação, observação de grafia**

*Na língua as regras foram convencionadas, isto é, em determinado tempo passou-se a usar a escrita e outros elementos de determinada forma e não de outra. No internetês começam a ser convencionados determinados usos, que passam a ser repetidos pelos outros usuários da mesma maneira.*

<sup>83</sup> Por exemplo, em 1958, o MEC aprovou a NGB (Nomenclatura Gramatical Brasileira), elaborada pelos professores Antenor Nascentes, Clóvis do Rêgo Monteiro, Cândido Jucá (filho), Rocha Lima, Celso Cunha, Antônio Chediak, Serafim Silva Neto e Sílvio Elia. Concebida com base nos conhecimentos de 1958, em alguns aspectos está desatualizada, fazendo com que os livros mais sérios coloquem notas de rodapé, contestando-a. A NGB precisa ser reformulada para estar adequada aos avanços dos estudos lingüísticos.

Com base na observação de certos grupos de palavras retiradas da Internet, você pode criar regras. Observe, então, a escrita das palavras e crie uma regra para cada grupo (para algumas palavras do internetês foram indicadas as formas da língua padrão entre parênteses).

Grupo 1: *tah, neh, eh, jah, bah, soh, pah, deixah*

Regra:

Grupo 2: *naum, taum, consideraçaum, feijaum, eraum (eram), vaum (vão), anaum (anão)*

Regra:

Grupo 3: *sab, pod, nunk, unik*

Regra:

Grupo 4: *kim pokim, dakele, loka*

Regra:

Grupo 5: *mi (me), nu(no), dus (dos), ti (te), cum (com), tudu (tudo), pur (por), issu (isso)*

Regra:

Grupo 6: *d (de), t(te), c(ce, você), c (se)*

Regra:

Grupo 7: *mtooo, feituuuuuu, carinhuuuuuu, amoooo*

Regra:

Grupo 8: *prcbr (perceber), esqc (esquece), gnt (gente), rpz (rapaz), tbm (também)*

Regra:

● Não é fácil definir o conceito de palavra. Na língua escrita é fácil demonstrar o que é uma delas, porque em muitos sistemas de escrita, uma palavra em um texto é delimitada por espaços em ambos os lados. Mas é apenas um aspecto formal. O próximo exercício leva o aluno a refletir sobre o que é palavra, observando exemplos do *internetês* num texto.

### ATIVIDADE 5: Leitura, análise e interpretação

O trecho a seguir é de um recado do Orkut, texto formado por palavras. Vemos que algumas estão escritas de forma bastante diferente daquela encontrada nos dicionários e, mesmo assim, transmitem idéias.

tudo blza...

rpz, na verdade eu naum mandei pra um monte de gente, pq deixei o link no meu album, e tbm na minha lista de videos...q bom q gostou

abraço

Responda:

1. Quantas palavras há no texto?

2. Todas elas são necessárias? Pode-se tirar alguma sem alterar o significado do texto?

2. Quantas dessas palavras você encontra num dicionário de português assim como apareceram?
3. Quais as principais modificações que ocorreram nas palavras não encontradas em dicionário?
3. Para você, no internetês, o que é uma palavra?

• O *internetês* chama a atenção por utilizar a língua, em alguns casos, de forma bastante diferente daquela preconizada pela norma culta. Além de modificar as palavras com abreviaturas em número considerável e fazer algumas criações que fogem aos padrões costumeiros, também é uma junção de fala e escrita. É esse o desafio proposto ao aluno: observar as freqüências de palavras da fala e da escrita e suas presenças no *internetês*.

### ATIVIDADE 6 – Análise e interpretação

Observe o quadro abaixo, que indica em três colunas as palavras mais usadas no português escrito, no internetês e no português falado.

ESCRITA				INTERNETÊS				FALA			
FREQUÊNCIA BANCO DE PORTUGUÊS ESCRITO				FREQUÊNCIA CORPUS GERAL DO ORKUT				FREQUÊNCIA BANCO DE PORTUGUÊS FALADO			
Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%	Nº	Palavra	Freq.	%
1	DE	1.537.460	4,42	1	E	13.930	2,52	1	E	113.061	3,73
2	A	1.082.233	3,11	2	QUE	11.537	2,08	2	QUE	108.883	3,59
3	O	1.026.380	2,95	3	EU	10.619	1,92	3	A	77.882	2,57
4	E	726.548	2,09	4	A	10.173	1,84	4	É	75.609	2,49
5	QUE	667.850	1,92	5	DE	9.950	1,80	5	O	71.329	2,35
6	DO	609.521	1,75	6	Q	9.028	1,63	6	DE	66.922	2,21
7	DA	545.271	1,57	7	TE	8.937	1,61	7	NÉ	64.870	2,14
8	EM	443.567	1,28	8	O	8.396	1,52	8	NÃO	62.445	2,06
9	PARA	353.847	1,02	9	É	7.855	1,42	9	EU	55.733	1,84
10	NO	308.932	0,89	10	VC	6.347	1,15	10	F	45.235	1,49

Da observação feita, tendo em vista a língua falada e a língua escrita, o que você conclui sobre o internetês?

- Assim como a atividade anterior, que leva o aluno a perceber a presença de fala e escrita no *internetês* pelos dados apresentados, a proposta a seguir também está relacionada à verificação de elementos da fala. Desta vez exigirá do aluno pesquisa em textos da Internet para levantamento de marcas da fala presentes nesses textos para sua conclusão e produção de texto.

### ATIVIDADE 7: Pesquisa, análise e produção de texto

Numa conversa, principalmente num diálogo, existem vocábulos e expressões que têm por objetivo *indicar o início e o fim de cada intervenção, a vontade de cada um deixar o outro falar (ou de lhe retirar a palavra) e realçar a dimensão emotiva dos conteúdos tratados*. Tais palavras são chamadas de **marcadores conversacionais**. Muitas vezes, para destacar o envolvimento dos falantes com as idéias tratadas são empregadas *interjeições*. Tanto os marcadores conversacionais como as interjeições são expressões próprias do registro oral informal, não aparecendo em textos escritos.

Exemplos de **marcadores conversacionais**: *então, está bem, pois, pois é, deixa lá, vá lá, diz lá, pronto, assim assim, e tal, e tudo, não sei quê, nem por isso, não dá para querer, não pode ser, não me digas*.

Exemplos de **marcadores conversacionais com valor interrogativo**: *que tal?, não é?, não é verdade?, não é assim?, não achas?, como assim?, achas bem?, que te parece?, e tu?, como assim?, diz quem?*

Exemplos de **interjeições**: *ah!, oh!, vamos!, viva!, oxalá!, ai!, ó!, hum!, psiu!, olá!, silêncio!, alto!, valha-me Deus!, raios te partam!*

Há discussões sobre a língua empregada pelos internautas, principalmente se é um misto de fala e escrita.

1. Procure textos do *internetês* (Orkut, chats, blogs, MSN).
2. Anote todos os marcadores conversacionais, as interjeições e outras expressões típicas da fala que aparecerem.
3. Escreva pequeno texto sobre a escrita usada na Internet, verificando se ela tem ou não muitas características da língua falada. Dê exemplos para sustentar sua opinião.

- Esta atividade resulta na produção textual após a participação dos alunos num Fórum eletrônico. Ao participar de algo vivo e dinâmico, diferente da produção escrita tradicional, estarão descobrindo ao natural as características e potencialidades do desenvolvimento retórico-argumentativo. Há vários desses fóruns nos portais de acesso à Internet. Como

exemplo, no portal Terra havia um fórum com a seguinte pergunta: *Há vida inteligente na TV?*<sup>84</sup> Uma pergunta instigante para jovens, que além de darem sua opinião, ainda lêem várias outras escritas anteriormente. Claro que o aluno deverá selecionar o que de melhor houver para compor seu texto final, de acordo com suas próprias idéias. Vendo várias argumentações, terá maior oportunidade para desenvolver habilidades de construir seus próprios pontos de vista e defendê-los. Não consideramos prejudicial a produção inicial no *internetês*, porque depois o texto será reescrito e passado a limpo de acordo com a norma culta. Xavier e Santos (2005, p. 38) defendem que a participação freqüente de crianças e adolescentes em fóruns da Internet desenvolve a capacidade de argumentar sobre variados temas, “levando-os a aprender a refletir dialeticamente sobre as diversas opiniões e construir sua própria síntese sobre as questões em discussão”.

### ATIVIDADE 8 – Leitura, produção de texto

1. *Você deverá localizar um Fórum de discussão num dos portais da Internet.*
2. *Procure um Fórum que aborde um assunto de seu interesse.*
3. *Leia algumas opiniões e discussões já escritas pelos outros participantes.*
4. *Participe do Fórum escrevendo o seu ponto de vista, defendendo-o. Pode ser escrito em internetês.*
5. *Copie sua opinião e depois a reescreva em linguagem culta, redigindo um texto coerente. Pode aproveitar as idéias de outros participantes para enriquecer seu texto.*

● Para a próxima atividade proposta aos alunos, tenhamos presentes as palavras de Bechara (1993, p. 14-5):

No fundo a grande missão do professor de língua materna — no ensino da língua estrangeira o problema é outro — é transformar seu aluno num poliglota dentro de sua própria língua, possibilitando-lhe escolher a língua funcional adequada a cada momento de criação e até, no texto em que isso se exigir ou for possível, entremear várias línguas funcionais para distinguir, por exemplo, a modalidade lingüística do narrador ou as modalidades praticadas por seus personagens.

Assim sendo, haverá *opressão* em “impor”, indistintamente, tanto a língua funcional da modalidade culta a todas as situações de uso da linguagem, como a língua funcional da modalidade familiar ou coloquial, nas mesmas

<sup>84</sup> A pergunta era acompanhada pelo texto: “*Você está satisfeito com o que assiste nas emissoras de TV aberta ou prefere migrar para canais de assinatura? Acha que a programação atual apenas reflete as expectativas do público, oferecendo o que este deseja ver? O governo deve ter voz ativa no debate sobre a qualidade da programação ou isso seria a volta da censura?*”.

circunstâncias, a todas as situações de uso da linguagem, pois que ambas as atitudes não recobrem a complexa e rica visão da língua como fator de manifestação da liberdade de expressão do homem.

Por outro lado, haverá “liberdade” quando se entender que uma língua histórica não é um sistema homogêneo e unitário, mas um diassistema, que abarca diversas realidades diatópicas (isto é, a diversidade de dialetos regionais), diastráticas (isto é, a diversidade de nível social) e diafásicas (isto é, a diversidade de estilos de língua), e que cada porção da comunidade lingüística realmente possui de direito sua língua funcional, que resulta de uma técnica histórica específica.

### **ATIVIDADE 9 – Produção de texto**

*Podemos fazer uma comparação entre o uso da roupa e o da língua. Devemos nos adequar à situação em que nos encontramos: se em situação formal, oficial ou solene usamos roupa “formal”, se em situação descontraída, roupa “descontraída”. Ninguém vai de maiô ao shopping center nem entra na praia em dia quente com blusão ou de terno e gravata. Essas não são roupas adequadas para aqueles ambientes.*

*Escreva pequeno texto refletindo sobre os diferentes usos da língua e a adequação desses usos conforme o ambiente (contexto) em que é empregada. Não esqueça de citar o uso do internetês. Dê outros exemplos comparando língua e vestimenta.*

- A Internet traz inegáveis benefícios às pessoas, mas também alguns problemas, muitas vezes decorrentes da falta de precaução de seus usuários. Nesta proposta de trabalho com os alunos, sugerimos que o professor faça um debate em aula sobre possíveis “perigos” do Orkut, o que é adequado expor na rede, como se prevenir, que exemplos de problemas conhecem, etc. É oportunidade de refletir sobre os riscos que podem advir da capacidade pouco crítica dos jovens ao exporem suas vidas em público de um modo muito intenso. Como finalização da atividade, solicitar que os alunos organizem por escrito as idéias, produzindo um texto autocrítico, que poderia ter como título “Problemas de segurança e privacidade no Orkut”. O encaminhamento poderia iniciar com os alunos indicando pontos positivos e negativos da Internet e do Orkut. Um aluno ou o próprio professor anotaria no quadro as idéias levantadas pela turma. Solicitar que indiquem exemplos de problemas que conhecem, que digam o que sabem sobre a legislação relacionada ao uso da Internet e onde buscariam apoio caso necessitassem (o professor pode falar da ONG SaferNet, indicada na nota de rodapé 86). Finalmente distribuir o texto da Atividade 10 e solicitar que exponham por escrito suas idéias após o que foi discutido e lido.

## ATIVIDADE 10 – Debate oral e produção de texto

### *Problemas de segurança e privacidade no Orkut*

*Os brasileiros aderiram em massa ao Orkut e são os principais usuários dessa rede, representando, em outubro de 2005, 74,54% dos participantes, contra apenas 7,93% do país segundo colocado, os EUA. Isso significa também que os brasileiros podem ser os mais atingidos pelos problemas de segurança e privacidade que essa rede pode oferecer.*

*Os problemas de privacidade são os maiores. Empolgadas, muitas pessoas expõem aspectos da vida pessoal que podem ser acessados por qualquer outro participante, provocando uma verdadeira "evasão de privacidade"<sup>85</sup>. Há registro de vários casos de abuso do Orkut no Brasil, com pessoas tendo seus perfis clonados e utilizados de má-fé. Até para usuários de Internet Banking, o Orkut registrou problemas que comprometem sua segurança nas transações bancárias on-line. O Orkut virou pretexto para o envio de mensagens falsas que induziam os usuários a baixar arquivos mascarados (cavalos-de-troia) para roubar senhas bancárias, dispensando o uso do sistema de e-mail. Quem utiliza essas redes deve tomar os mesmos cuidados recomendados ao acessar sua conta de e-mail. Deve desconfiar de mensagens com conteúdo inusitado, ameaçador ou muito vantajoso, e não deve clicar em links desconhecidos.*

*A ONG Safernet<sup>86</sup>, que luta contra crimes na web, apresenta documento de 150 páginas no qual relata pedofilia, racismo, xenofobia, homofobia, venda ilegal de remédios e violência contra animais em comunidades e páginas pessoais do Orkut.*

*Para evitar problemas, especialistas recomendam não traçar o perfil de sua pessoa de um modo muito fácil e transparente. Nas suas "comunidades" e "perfil" alguns deixam expostas suas vidas e suas relações (número de telefones, RG, CPF, e-mail, nomes de familiares, etc.). Deve-se fazer uma análise das informações contidas no seu Orkut e ver a necessidade delas. São descartáveis ou*

<sup>85</sup> Ao contrário da "invasão de privacidade", em que alguém acessa e utiliza sem autorização informações da vida particular de outra pessoa, "evasão de privacidade" é a exposição detalhada da vida pessoal na rede mundial de computadores pelo próprio usuário. Segundo Komesu (2005), podemos chamar essa exposição voluntária de "publicização de si".

<sup>86</sup> A SaferNet Brasil é uma organização não governamental, que reúne cientistas da computação, professores, pesquisadores e bacharéis em Direito para defender e promover os Direitos Humanos na Sociedade da Informação no Brasil. Através da Central Nacional de Denúncias de Crimes Cibernéticos, operada em parceria com o Ministério Público Federal, oferece um serviço anônimo de recebimento, processamento, encaminhamento e acompanhamento *on-line* de denúncias sobre qualquer crime ou violação aos Direitos Humanos praticados através da Internet. Acesso: <http://www.safernet.org.br>.

*não? Recomenda-se deixar apenas o e-mail de contato como fonte, e participar de uma forma descontraída com comunidades e contatos divertidos, sem expor muitos detalhes da vida pessoal.*

*Tendo em vista vantagens e desvantagens do relacionamento no Orkut, escreva um texto expondo sua opinião sobre esse assunto. Procure fundamentar seu ponto de vista utilizando exemplos.*

## 7.1. ALGUMAS CONSIDERAÇÕES SOBRE AS ATIVIDADES PROPOSTAS

Muitas pessoas se preocupam com o uso do *internetês*, consideram-no uma “transgressão” à norma-padrão e temem que a língua seja influenciada e prejudicada pelos recursos lingüísticos criados pelos usuários da grande rede. Há o temor de que essa linguagem possa estar prestes a invadir o mundo real, passando a fazer parte do dia-a-dia fora da Internet. Não há como ignorar sua existência e seu poder, é uma maneira legítima de se comunicar, surgida, principalmente, para atender a necessidade de rapidez que os tempos modernos exigem. Não podemos esquecer o quanto os jovens se comunicam “conversando” por escrito no MSN e nos *chats*, por exemplo. Esse mesmo ritmo, de alguma forma, e a mesma linguagem empregada *on-line* são transferidos para o Orkut. Pelo menos é isso que mostram as ocorrências de palavras exibidas pelo *corpus*, principalmente o de *scraps*: quem conhece as utilizadas no MSN logo percebe que ambas têm as mesmas características.

Professores, especialmente os de Língua Portuguesa, não devem ignorar esse novo jeito de usar a língua, não mitificá-lo, ao contrário, utilizá-lo como um aliado para o objetivo de levar os alunos ao conhecimento da norma culta e de outras normas do português. O próprio Ministério da Educação favorece a abordagem atual da comunicação. Os PCNs encaram e tratam a linguagem como algo vivo, em constante evolução e inserida nas práticas sociais. Destacam também o fato de que o domínio da língua tem estreita relação com a possibilidade de plena participação social, já que é por meio da língua que o homem se comunica, tem acesso à informação, mostra e defende seus pontos de vista, partilha ou constrói visões de mundo, isto é, produz conhecimento.

Ora, as práticas sociais cada vez mais se utilizam da grande rede. Claro que não devemos fazer apologia do *internetês*, apenas colocá-lo no seu devido lugar: uma variante

dialetal<sup>87</sup> utilizada principalmente pelos jovens para se comunicar na Internet. Somente adequado para esse ambiente e não para outros. Para os demais contextos deve ficar claro que a língua escrita oficial é a que impera, deve ser estudada e empregada. Inclusive devem os alunos saber que a grafia é o único componente da língua regulamentado por lei (sua última alteração ocorreu pela Lei 5765, de 18/12/1971). Então devem os professores esclarecer que “uma coisa é uma coisa, outra coisa é outra coisa”, como popularmente se fala. A língua na sua norma culta é a que deve ser estudada por ser aquela aceitável em todas as situações, ser fator de coesão, integração e ascensão social. Quem não se expressa de acordo com essa norma não é bem visto, é discriminado. Já o *internetês* deve ser entendido como um dialeto social, um socioleto, a forma mais adequada para o uso pelos jovens na apressada comunicação da Internet.

O que a escola pode fazer é aproveitar a presença da língua da grande rede, já conhecida e empregada pela maioria dos alunos, e utilizá-la como um instrumento para se refletir sobre a heterogeneidade da língua em suas diferentes modalidades e situações de comunicação. Não se espera ver o *miguxês* e o *internetês* nas redações e outras produções. Cada coisa tem o seu lugar e a sua hora. A função do professor é ajudar o aluno a dar-se conta de que ele é um poliglota dentro da própria língua. Os jovens devem saber da existência das várias formas de expressão e entender que a língua em sua norma culta é, em princípio, a forma reconhecida, consagrada e compartilhada por todos. Aqui podemos repetir Dacanal (2006), para quem a língua é instrumento de poder e quem a domina bem tem melhores condições de dominar os outros. Qual o poder adquirido por quem apenas bem domina o *internetês*?

---

<sup>87</sup> Aqui lembramos o que seja dialeto para nós anteriormente explicitado no posicionamento teórico. No caso do *internetês*, corresponde a socioleto, assim definido em Houaiss (2001): SOCIOLETO /é/ *s.m.* (sXX) SLING cada uma das variedades de uma língua us. pelos grupos de indivíduos que, tendo características sociais em comum (p.ex., a profissão, os passatempos, a geração etc.), usam termos técnicos, ou gírias, ou fraseados que os distinguem dos demais falantes na sua comunidade; dialeto social, variante diastrática □ ETIM soci(o)- + el. - *lecto* (ou *-lecto*), depreendido de *dialeto*; ver *dialect-* ou *dialet-*

## REFERÊNCIAS

ARAÚJO, Vera Maria Araujo Pigozzi de. *Documentação, terminologia e lingüística: uma interface produtiva*. Porto Alegre, 2006. 163 f. Dissertação (Mestrado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, 2006.

ARAÚJO, Júlio César; BIASI-RODRIGUES, Bernadete (Org.). *Interação na Internet: novas formas de usar a linguagem*. Rio de Janeiro: Lucerna, 2005.

AZEREDO, Susana de. *Expressões anunciadoras de paráfrase em manuais acadêmicos de Química: um estudo baseado em corpus*. Porto Alegre, 2007. 222 f. Dissertação (Mestrado em Teorias do texto e do discurso) – Instituto de Letras, Universidade Federal do Rio Grande do Sul.

BAAAYEN, R. Harald. Lexis, word frequencies and text types. In: JORNADES DE CORPUS LINGUISTICS, 4.-5., 1996-1997, Barcelona. [Actes...] Barcelona: Institut Universitari de Lingüística Aplicada, 1998. p. 16-17.

BAGNO, Marcos. In: VOLPATO, Cadão. Q língua eh essa? *Revista da Cultura*, São Paulo, Edição 158, p. 20, ago. 2007.

BAGNO, Marcos. *Preconceito lingüístico*. São Paulo: Loyola, 2002.

BAKHTIN, Mikhail. *Marxismo e filosofia da linguagem*. 9 ed. São Paulo: HUCITEC, 1999.

BASILIO, Margarida. *Formação e classe de palavras no português do Brasil*. São Paulo: Contexto, 2004.

BECHARA, Evanildo. *Ensino da Gramática. Opressão? Liberdade?* 7. ed. São Paulo: Ática, 1993.

BECHARA, Evanildo. *Moderna Gramática Portuguesa*. Rio de Janeiro: Lucerna, 2003.

BELMONTE, Roberto Villar. *A coesão textual frente à regra jornalística da não-repetição de palavras*. Disponível em: <[www.rvb.jor.br/linguagem.htm](http://www.rvb.jor.br/linguagem.htm)>. Acesso em: 13 maio 2007.

BENVENISTE, Emile. Da subjetividade na linguagem. In: \_\_\_\_\_. *Problemas de Lingüística Geral I*. Campinas, SP: Pontes, 1995. p. 284-293.

BERNARDES, Alessandra Sexto; VIEIRA, Paula M. Teixeira. O *chat* como produção de linguagem. In: FREITAS, Maria Teresa de Assunção; COSTA, Sérgio Roberto (Org.). *Leitura e escrita de adolescentes na internet e na escola*. Belo Horizonte: Autêntica, 2006. p. 45-64.

BIBER, Douglas. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998.

BIDERMAN, Maria Teresa Camargo. Conceito lingüístico de palavra. In: BASILIO, Margarida (Org.). *Palavra*. Rio de Janeiro: Grypho, 1999. v. 1, p. 81 -97.

BIDERMAN, Maria Teresa Camargo. Dimensões da palavra. In: *Filologia e Lingüística Portuguesa*, São Paulo, Humanitas, FFLCH/USP, n. 2, p. 81 -118, 1998.

- BIDERMAN, Maria Teresa Camargo. A face quantitativa da linguagem: um dicionário de frequências do português. *Alfa*, Araraquara, Unesp, v. 42, p. 157-78, 1998a.
- BRASIL. Ministério da Educação. Secretaria de Educação Fundamental. *Parâmetros Curriculares Nacionais. Língua Estrangeira*. Brasília: MEC/SEF, 1998b.
- BRASIL. Ministério da Educação. Secretaria de Educação Fundamental. *Parâmetros Curriculares Nacionais. Língua Portuguesa*. Brasília: MEC/SEF, 1998.
- BRASIL. Ministério da Educação. Secretaria de Educação Fundamental. *Parâmetros Curriculares Nacionais. Terceiro e quarto ciclos do ensino fundamental: introdução aos parâmetros curriculares nacionais*. Brasília: MEC/SEF, 1998a.
- BRASIL. Ministério da Educação. Secretaria de Educação Fundamental. *Parâmetros Curriculares Nacionais. Introdução: Língua Portuguesa, Matemática, Geografia, História e Ciências*. Brasília: MEC/SEF, 2000.
- BRASIL. Ministério da Educação. Secretaria de Educação Média e Tecnológica. *Parâmetros Curriculares Nacionais: ensino médio*. Brasília: Ministério da Educação, 1999.
- CÂMARA JUNIOR, Joaquim Mattoso. Joaquim. *Dicionário de Lingüística e Gramática: Referente à Língua Portuguesa*. Petrópolis: Vozes, 1984.
- CÂMARA JUNIOR, Joaquim Mattoso. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes, 1977.
- CÂMARA JUNIOR, Joaquim Mattoso. *Princípios de lingüística geral*. Rio de Janeiro: Padrão, 1980.
- CARVALHO, Castelar de. Saussure e a Língua Portuguesa. In: SEMANA NACIONAL DE ESTUDOS FILOLÓGICOS E LINGÜÍSTICOS (SENEFIL), 7. Rio de Janeiro: s.n, 2004. Disponível em: <[www.filologia.org.br/viisenefil/09.htm](http://www.filologia.org.br/viisenefil/09.htm)>. Acesso em: 10 fev. 2007.
- COSERIU, Eugenio. *O homem e a sua linguagem*. Rio de Janeiro: Presença; São Paulo: Universidade de São Paulo, 1982.
- COSERIU, Eugenio. *Lições de lingüística geral*. Rio de Janeiro: Ao Livro Técnico, 1980.
- COSERIU, Eugenio. A Perspectiva Funcional do Léxico. In: VILELA, M. (Org.). *Problemas de lexicologia e lexicografia*. Porto: Civilização, 1979b. p. 15-33.
- COSERIU, Eugenio. *Sentido y tareas de la dialectologia*. México: Instituto de Investigaciones Filológicas, 1982.
- COSERIU, Eugenio. *Sincronia, diacronia e história*. Rio de Janeiro: Presença; São Paulo: USP, 1979.
- COSERIU, Eugenio. *Teoria da Linguagem e Lingüística Geral*. Rio de Janeiro: Presença; São Paulo, EDUSP, 1979a.
- COSTA, Sérgio Roberto. (Hiper)textos ciberespaciais: mutações do/no ler/escrever. *Caderno Cedes*. Campinas, v. 25, n. 65, p. 102-116, jan./abr. 2005.

COSTA, Sérgio Roberto. Oralidade, escrita e novos gêneros (hiper)textuais na Internet. In: ASSUNÇÃO, Maria Teresa Freitas de; COSTA, Sérgio Roberto. (Org.) *Leitura e escrita de adolescentes na internet e na escola*. Belo Horizonte: Autêntica, 2006. p.19-28.

CUNHA, Celso; CINTRA, L. F. Lindley. *Nova Gramática do Português Contemporâneo*. Lisboa: Sá da Costa, 1984.

DACANAL, José Hildebrando. *Linguagem, poder e ensino da língua*. Porto Alegre: WS Editor, 2006.

DI LUCCIO, Flavia. *As múltiplas faces dos blogs: um estudo sobre as relações entre escritores, leitores e textos*. Rio de Janeiro, 2005. 113 f. Dissertação (Mestrado em Psicologia Clínica) – Departamento de Psicologia, Pontifícia Universidade Católica, 2005.

DICIO.NET. Disponível em: < <http://dicio.net> >. Acesso em: 01 nov. 2007.

DUBOIS, Jean. *Dicionário de Lingüística*. São Paulo: Cultrix, 1978.

EISENKRAEMER, Raquel Eloísa. Leitura digital e linguagem cifrada dos internautas. *Revista Texto Digital*, ano 2, n. 3, 2006. Disponível em: <<http://www.textodigital.ufsc.br>>. Acesso em: 10 abril 2007.

ENDRUWEIT, Magali Lopes. *A escrita enunciativa e os rastros da singularidade*. Porto Alegre, 2006. 205 f. Tese (Doutorado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, 2006.

FARACO, Carlos Alberto. Norma-padrão brasileira. Desembaraçando alguns nós. In: BAGNO, Marcos (Org.). *Lingüística da Norma*. São Paulo: Loyola, 2002. p. 37-62.

FAVERO, Leonor Lopes; ANDRADE, Maria Lúcia C. V.; AQUINO, Zilda G. Oliveira de. Fala e escrita: diferença e integração. In: SEMINÁRIO DO GEL (41: 1993: Ribeirão Preto). *Anais*. São Paulo: Ed. da USP, 1994. v.1, p. 273-288 .

FERREIRA, Aurélio Buarque de Holanda. *Novo Aurélio Século XXI: o dicionário da língua portuguesa*. Rio de Janeiro: Nova Fronteira, 1999.

FERREIRO, Emilia; TEBEROSKY, Ana. *Psicogênese da língua escrita*. Porto Alegre: Artes Médicas Sul, 1999.

FONSECA, Lucia de Carvalho. *A linguagem da Internet: entre o dito e o escrito*. Belo Horizonte, 2001.150 f. Dissertação (Mestrado em Tecnologia) – Centro Federal de Educação Tecnológica de Minas Gerais, CEFET/MG, 2001.

FREITAS, Maria Thereza. O universo criativo da escrita “teclada”. *Nós da escola*, Rio de Janeiro, ano 2 , n. 23, 2004 . Disponível em: < [www.multirio.rj.gov.br/nosdaescola](http://www.multirio.rj.gov.br/nosdaescola)>. Acesso em: 20 nov. 2007.

HABERT, Benoit; NAZARENKO, Adeline; SALEM, André. *Les linguistiques de corpus*. Paris: A. Colin, 1997.

HILGERT, José Gaston. A construção do texto “falado” por escrito: a conversação na Internet. In: PRETI, Dino (Org.). *Fala e escrita em questão*. São Paulo: Humanitas, 2000. p. 70-84.

HOUAISS, Antônio. *Dicionário Eletrônico Houaiss*. Rio de Janeiro: Objetiva, 2001

INTERNET. In: *Wikipedia*, a enciclopédia livre. Disponível em: <<http://pt.wikipedia.org/wiki/Internet>>. Acesso em: 20 maio 2007.

KATO, Mary Aizawa. *No mundo da escrita: uma perspectiva psicolinguística*. 2. ed. São Paulo: Ática, 1987.

KOMESU, Fabiana Cristina. *Entre o público e o privado: um jogo enunciativo na constituição do escrevente de blogs da internet*. Campinas, 2005. 270 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 2005.

KRIEGER, Maria da Graça. Lexicografia: o léxico no dicionário. In: SEABRA, Maria Cândida Trindade Costa de (Org.). *O léxico em estudo*. Belo Horizonte: Faculdade de Letras da UFMG, 2006. p. 157-172.

LEPSCHY, Giulio. Léxico. In: *Enciclopédia EUNAUDI. Linguagem – Enunciação*. Lisboa: Imprensa Nacional - Casa da Moeda, 1984. v.2, p. 156 – 178.

LUFT, Celso Pedro. *Língua e Liberdade*. Porto Alegre: LPM, 1985.

MACIEL, Anna Maria Becker. Linguística de *Corpus*: bases teórico-metodológicas. In: COLÓQUIO NACIONAL LETRAS EM DIÁLOGO E EM CONTEXTO: rumos e desafios, 2002, Porto Alegre. *Anais...* Porto Alegre: Programa de Pós-Graduação em Letras, 2003. 1 CD-ROM.

MARCUSCHI, Luis Antônio. *Análise da Conversação*. São Paulo: Ática, 1991.

MARCUSCHI, Luis Antônio. Apresentação. In: ARAÚJO, Júlio César; BIASI-RODRIGUES, Bernadete (Org.). *Interação na Internet: novas formas de usar a linguagem*. Rio de Janeiro: Lucerna, 2005. p. 9-12.

MOLLICA, Maria Cecília. Fundamentação teórica: conceituação e delimitação. In: MOLLICA, Maria Conceição, BRAGA, Maria Luiza (Org.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003. p. 13-27

NARO, Anthony Julius. O dinamismo das línguas. In: MOLLICA, Maria Conceição, BRAGA, Maria Luiza (Org.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003. p.56-80.

NEVES, Maria Helena de Moura. *Que gramática estudar na escola?* São Paulo: Cultrix, 2003.

NUNES et al. Desafios na construção de recursos linguísticos para o processamento do português do Brasil. In: SARDINHA, Tony Berber (Org.). *A língua portuguesa no computador*. Campinas: Mercado das Letras; São Paulo: Fapesp, 2005. p.33-70.

PALMIERE, Denise Telles Leme. A escrita da *Internet*: da tela do computador para a tela da tevê. *Estudos Linguísticos*, Araraquara, SP, v. 35, p.572-581, 2006.

PEREIRA, Ana Paula M. S.; MOURA, Mirtes Zoe da Silva. A produção discursiva nas salas de bate-papo: formas e características processuais.. In: FREITAS, Maria Teresa de Assunção;

COSTA, Sérgio Roberto (Org.). *Leitura e escrita de adolescentes na internet e na escola*. 2. ed. Belo Horizonte: Autêntica, 2006. p. 65-84.

PERINI, Mario Alberto. *Princípios de Lingüística Descritiva: uma introdução ao pensamento gramatical*. São Paulo: Parábola Editorial, 2006.

PIMENTEL, Carmen. Vc tb escreve axim? Uma análise do léxico nos blogs de adolescentes. *Revista Querubim*, Rio de Janeiro, ano 02, n. 02, vol. 01, jan/jun 2006.

POSSENTI, Sírio. Uma variante é uma variante de quê? *ABRALIN: boletim da Associação Brasileira de Lingüística*, Salvador, n. 8, p. 107-117, set. 1986.

POSSENTI, Sírio. Você entende Internetês? *Revista Discutindo a Língua Portuguesa*, São Paulo, ano 1, n. 2., 2006. p. 28-34.

POTTIER, Bernard; AUDUBERT, Albert; PAIS, Cidmar Teodoro. *Estruturas lingüísticas do português*. São Paulo: DIFEL, 1975.

PRANGE, Ana Paula Lobão. *Da Literatura aos blogs: um passeio pelo território da escrita de si*. Rio de Janeiro, 2003. 129 f. Dissertação (Mestrado em Psicologia Clínica) – Departamento de Psicologia, Pontifícia Universidade Católica, 2003.

PRETI, Dino. *Sociolingüística: os níveis de fala*. 4. ed. São Paulo: Nacional, 1982.

ROSENGREN, Inger. O léxico fundamental como problema teórico e prático. In: VILELA, M. (Org.). *Problemas de lexicologia e lexicografia*. Porto: Civilização, 1979. p. 281-300.

SAMPSON, Geoffrey. *Sistemas de escrita. Tipologia, história e psicologia*. São Paulo: Ática, 1996.

SARDINHA, Tony Berber. *Lingüística de Corpus*. São Paulo: Manole, 2004.

SARDINHA, Tony Berber. Lingüística de Corpus: histórico e problemática. *D.E.L.T.A.*, v. 16, nº 2, p. 323-367, 2000.

SARDINHA, Tony Berber. Lingüística de *Corpus*: uma entrevista com Tony Berber Sardinha. *Revista Virtual de Estudos da Linguagem – ReVEL*, v. 2, n. 3, ago. de 2004a. Disponível em: <<http://paginas.terra.com.br/educacao/revel/index.htm>>. Acesso em: 20 abril 2007.

SCARTON, Gilberto. Concepções de Língua e Reflexos na Prática do Professor. In: *Manual de Redação*. Disponível em em: <<http://www.pucrs.br/manualred/textos/texto7.php>> . Acesso em: 12 dez. 2007.

SCLIAR, Moacyr. + três questões sobre língua portuguesa. *Folha de São Paulo*, São Paulo, p. 6, 02 jul. 2000. Caderno MAIS!.

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

SEABRA, Maria Cândida Trindade Costa de (Org.). *O léxico em estudo*. Belo Horizonte: Faculdade de Letras da UFMG, 2006.

SILVA, Deonísio da. *Português assassinado a tecladas*. Disponível em: <<http://observatorio.ultimosegundo.ig.com.br/artigos.asp?cod=320JDB001#>>. Acesso em: 10 abril 2007.

SINCLAIR, Julie; BALL, J. *Preliminary Recommendations on Corpus Typology – EAG-TCWG-CTYP/P, 1996*. Disponível em: <[www.ilc.pi.cnr.it/EAGLES96/corpusstyp/corpusstyp.html](http://www.ilc.pi.cnr.it/EAGLES96/corpusstyp/corpusstyp.html)> Acesso: 12 ago 2007.

VILLAVICENCIO, Aline; FINATTO, Maria José; POSSAMAI, Viviane. Padrões da Preposição “de” entre Sintagmas Nominais em Linguagem Cotidiana e Linguagens Técnico-Científicas. In: ENCONTRO DE CORPORA, 5., 2005, São Carlos. Disponível em: <http://www.nilc.icmc.usp.br/vencontro/vencontro.htm>. Acesso em: 10. jun. 2007.

WIKIPEDIA. *Miguxês*. Disponível em: <[pt.wikipedia.org/wiki/Miguxês](http://pt.wikipedia.org/wiki/Miguxês)>. Acesso em: 10 jun. 2007.

XAVIER, Antônio Carlos; SANTOS, Carmi Ferraz. E-forum na Internet: um Gênero Digital: In: ARAÚJO, Júlio César; BIASI-RODRIGUES, Bernadete (Org.). *Interação na internet: novas formas de usar a linguagem*. Rio de Janeiro: Lucerna, 2005. p. 30-38.



**ANEXO II – Amostra do *corpus* de depoimentos**  
**<Início do Corpus Geral do Orkut>**

EU TE AMO!!!!!!!!!!!!

CONTE COMIGOO SEMPRE!!!!!!!!!!

Tava vendo e eu num fiz nenhum depoimento pra vc...=//

eh dificil fla algma coisa de vc...

te conheço des d q agnte era uma sementinha...hauahuahauhauha...agnte tinha 4

anos...sempre nos entendemos mtuu bem...sempre fomos da msma classe...vc sempre foi esse palhaço..esse menino maravilhoso, humilde, carinhoso que td mundo gosta e mtuuu!!!

Agnte cresceu e eu comecei a gosta d vc...uahuahuahuahauhauha...tadinha d mim...vc nem me dava bola...=//...bem plo contrario...fikava flando dah mina q vc gostava...eu fikava

taum triste...hauhauhauhauhaha

mais essa fase pasooo meu lindo..

nao sei pq eu quis faze um depoimento pra vc pq eu nao ligo pra essas merdas de orkut e bla bla bla e sei que merda nenhuma disso pode mostra o valor da nossa amizade de se eu nao me engano 12 anos...

sao 12 anos fazendo merda, rindo que nem uns bobo, chorando, indo ir dormir na casa um do outro (no bom sentindo sempre) auhIAHiuAuAGuya e sei la irmao eh nois sempre nessa merda de vida mesmo vc sendo palmeirense e sempre tendo que me aguenta quando o meu tricolor ganha (o que nao eh nada incomum)

auhUIAHiuAHuiAHuiHAuia

eh nois meu irmao

abrassao com dois S como diz o trovo

nm sei por ondi começa...soh vo faze pq hj eu to insipirada uahauahuah.

meu...vc eh um garoto q consegue me tira do serio...me enxe o saco, mi irritah, mi dexa nervosa...mais ao msm tempo tah me fazendo da risada, me diverti i esquece de todos os problemas, me faze ri enquanto eu divia tah chorando.

se eu fosse conta desde qndo vc me conheceu ateh agora qntas vzs vc jah me zuo...naum tem nm como

uahauahuahauhauh

todo dia uma gracinha, uma piadinha q dexa meu dia fik mais alegre !

soh tenhooo q te diz que vc eh mtoooo especial um dos melhores amigos q eu tenho e que eu te amooo mtoooooo0 chatoo ! x)

naum faz nm tanto tempo q a gente viro amigo mais desde intaum vc conquisto a minha agora aguenta pq eu naum desgrudo mais

auahuahuahauhauh

naum vai equecendo das sextas feiras que eu vo lah na sua ksa enxe o saco...naum queru sabe de desculpas ! xp

te amoooo marcelo !

Cara

Por mais q vc sempre perca no winning eleven , or um gol perfeito de um macaco corintiano ( AHuaHauhauhau )eu vou continuar sendo seu amigo

Mesmo tu me dando odio<sup>3</sup> e NINGUEM sendo mais odioso q vc , eu vou continuar sendo seu amigo.

Mesmo tu sendo um jogador frenetico de tibia , eu vou continuar sendo seu aigo.

Mesmo tu sendo branco (eh foda dizer isso) eu vou continuar sendo seu amigo.

## ANEXO III - Proposta de produção de texto para o terceiro ano do Ensino Médio.



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
COLÉGIO DE APLICAÇÃO



DEPARTAMENTO DE COMUNICAÇÃO – ÁREA DE LÍNGUA PORTUGUESA E LITERATURA

**Projeto de Redação 110** – PROFS. ADAUTO TAUFER, ANDRÉA PERROT E JUÇARA FREITAS

**H**ouve um tempo em que a proposta de redação do Vestibular da UFRGS exigia que o candidato narrasse – em pelo menos um parágrafo do desenvolvimento – uma experiência pessoal que tivesse vivido ou presenciado. Esse aspecto predominou na década de 90 do século XX. Naquela época, os vestibulandos eram convidados a redigirem textos que abarcassem duas modalidades discursivas: **a dissertação e a narração**. Esta, com efeito, não deveria predominar sobre aquela, mas servir como exemplo para a conduta argumentativa do texto dissertativo.

**Pois bem**, sua produção textual – **exclusivamente de caráter narrativo** – versará sobre a seguinte questão: *Você participou de uma festa na qual vários fatos inusitados aconteceram. No outro dia, um amigo que não esteve na festa, muito curioso, queria saber o que tinha acontecido. RELATE POR ESCRITO O ENCONTRO COM O AMIGO E A CONVERSA QUE TIVERAM.*

O dicionário Houaiss indica **INUSITADO** assim: adj. (1572 cf. IAVL) **1** não usual **1.1** que não é corrente, que não se usa ou emprega com frequência, que foge a padrões costumeiros; incomum, insólito, estranho; inabitual <procedimento i.> <recurso i.> <escolha i.> <os tons berrantes são i. na paleta desse pintor> **1.2** que causa surpresa, estranhamento; que é diferente do que se espera ou se imagina <festa i.>

## ANEXO IV AMOSTRA DO CORPUS DE REDAÇÕES ESCOLARES

### <Aluna 1 – 7ª. série do Ensino Fundamental>

Vocês não vão acreditar no que eu vou contar. LEIAM PARA VER!

\_Alô! Mary? Perguntou Lumah.

\_Oi Lumah sô eu. Fala!

\_Mari tu pode dá uma passadinha aqui em casa hoje, não estou muito bem. Mari saiu rapidamente direto para casa da Lumah ver o que estava acontecendo.

Mari chegou na casa de Lumah, e logo foi perguntar o que tinha acontecido. Lumah começa a falar:

\_ Mari sabe a festa que eu fui ontem?

\_ Sei. O que aconteceu?

\_ Eu cheguei na festa tri contente porque eu ia vê o meu namorado, né? Daí eu não achava ele, daí eu tava nervosa e fui no barzinho que tinha lá na festa e comprei um refri, fui botar a latinha do refri na lata do lixo, quando eu olho pro lado duas pessoas se beijando, daí eu olhei melhor e era o meu namorado.

\_ Bah. E o que tu fez?

\_ A Mari não fiz nada né, eu tenho que me valoriza não vou correr atrás dele.

Lumah ficou bem melhor quando desabafou tudo para sua melhor amiga. Lumah chorou muito mas foi bom descobrir o cafageste que ele era, antes de iludi-la mais.

### <Aluna 1 - 3º ano do Ensino Médio>

NÃO GOSTO DE FOFOCA

HOUVE UMA FESTA EM QUE MINHA AMIGA ANA NÃO PODE IR, POIS ESTAVA COM DENGUE, A FESTA FOI NO SÁBADO A NOITE E EU FUI VISITÁ-LA NO DOMINGO À TARDE ANCIOSA PARA CONTAR A ELA OS FATOS QUENTES QUE ACONTECERAM NA BALADA. CHEGANDO EM SUA CASA, ANA PARECIA ESQUECER QUE ESTAVA DOENTE E QUERIA LOGO SABER COMO FOI A FESTA E PRINCIPALMENTE SE O SEU GATINHO HAVIA SE COMPORTADO, EU NÃO FIZ UMA EXPRESSÃO MUITO BOA E COMECEI A FALAR:

- ANA, EU JÁ TE FALEI PRA TI ESQUECER ESSE GURI!. ELA NEM LIGANDO PARA O QUE EU DISSE INSISTIU NA PERGUNTA:

- AI, ME FALA DE UMA VEZ!

- SEI LÁ, NEM REPREI NELE, MAS ACHO QUE ELE ANDAVA COM A TAMARA. A ANA INDIGNADA DISSE:

- AQELA...

- PSST! TU SABE QUE ELE NÃO VALE NADA, QUANDO VOCÊS NAMORAVAM ELE TE TRAIA E JÁ ATÉ BATEU EM TI.

COM UMA EXPRESSÃO CANSADA ELA DISSE:

- MAS EU AMO ELE!

- TU NÃO AMA ELE, É QUE TU PRECISA DE OUTRA PESSOA PARA ESQUECER ELE!

E A DISCUSSÃO COMO SEMPRE FOI LONGE, ATÉ QUE EU MUDEI DE ASSUNTO PARA TRATAR DOS OUTROS ACONTECIMENTOS DA FESTA. E PARA VARIAR ERAM SÓ FATOS RUINS QUE SE TORNAM FOFOCAS PARA QUEM ESTA DE FORA, MAS ESSES A ANA NÃO QUIS NEM SABER POIS SE ABORRECEU COM O FATO DO MALANDRO DELA ESTAR ANDANDO COM A TAMARA, ENTÃO EU QUIS IR EMBORA PARA DEIXAR A ANA REFLETIR SOBRE A SITUAÇÃO. NOS DESPEDIMOS, E EU FUI TRATAR DE CONTAR AS FOFOCAS PARA OUTRA PESSOA POIS EU NÃO PODIA MAIS SEGURAR “AQUILO” DENTRO DE MIM.

**ANEXO V - As cem palavras mais usadas do *subcorpus* escrito do Banco de Português**

Nº	Pal.	Freq.	%	Nº	Pal.	Freq.	%
1	DE	1.537.460	4,42	51	SOBRE	55.459	0,16
2	A	1.082.233	3,11	52	JÁ	54.383	0,16
3	O	1.026.380	2,95	53	ATÉ	53.416	0,15
4	E	726.548	2,09	54	DISSE	53.108	0,15
5	QUE	667.850	1,92	55	TAMBÉM	51.106	0,15
6	DO	609.521	1,75	56	OBSERVAÇÕES	50.426	0,15
7	DA	545.271	1,57	57	ONTEM	48.037	0,14
8	EM	443.567	1,28	58	ANOS	47.899	0,14
9	PARA	353.847	1,02	59	DIA	46.784	0,13
10	NO	308.932	0,89	60	PRINCIPAIS	45.495	0,13
11	COM	299.306	0,86	61	GOVERNO	45.367	0,13
12	OS	282.086	0,81	62	CRÉDITO	42.822	0,12
13	É	267.713	0,77	63	PRESIDENTE	42.744	0,12
14	UM	267.556	0,77	64	VAI	41.816	0,12
15	NA	251.255	0,72	65	HOJE	41.006	0,12
16	NÃO	243.630	0,70	66	US\$	40.975	0,12
17	SE	222.518	0,64	67	ASSUNTOS	40.522	0,12
18	UMA	207.063	0,60	68	HÁ	40.437	0,12
19	AS	172.066	0,50	69	QUANDO	40.104	0,12
20	DOS	170.109	0,49	70	ANO	39.818	0,11
21	POR	170.041	0,49	71	DIZ	39.725	0,11
22	MAIS	138.812	0,40	72	ÀS	39.588	0,11
23	AO	133.615	0,38	73	SÓ	38.871	0,11
24	SÃO	133.381	0,38	74	LEGENDA	38.516	0,11
25	PÁGINA	121.922	0,35	75	DOIS	38.417	0,11
26	NACIONAL	118.428	0,34	76	SEM	37.678	0,11
27	EDIÇÃO	118.159	0,34	77	RIO	37.584	0,11
28	COMO	117.136	0,34	78	AINDA	37.435	0,11
29	EDITORIA	114.457	0,33	79	CONTRA	37.373	0,11
30	FOI	106.965	0,31	80	NAS	37.008	0,11
31	DAS	103.585	0,30	81	PODE	36.851	0,11
32	À	101.925	0,29	82	MESMO	34.256	0,10
33	PAULO	85.082	0,24	83	MUITO	33.708	0,10
34	MAS	80.797	0,23	84	AOS	33.583	0,10
35	FOTO	78.264	0,23	85	ERA	32.391	0,09
36	FOLHA	73.000	0,21	86	LOCAL	32.247	0,09
37	ELE	71.290	0,21	87	EU	31.881	0,09
38	PELO	70.093	0,20	88	FORAM	31.753	0,09
39	SER	69.639	0,20	89	TER	31.642	0,09
40	SUA	69.059	0,20	90	MIL	31.606	0,09
41	BRASIL	66.925	0,19	91	SEUS	31.232	0,09
42	PELA	64.346	0,19	92	MUNDO	30.893	0,09
43	AUTOR	63.671	0,18	93	IMAGEM	30.225	0,09
44	SEU	63.488	0,18	94	DEPOIS	30.089	0,09
45	OU	61.669	0,18	95	ESPECIAL	29.870	0,09
46	NOS	61.159	0,18	96	TRÊS	29.800	0,09
47	TEM	59.668	0,17	97	SUB	29.771	0,09
48	SEGUNDO	58.266	0,17	98	ISSO	29.179	0,08
49	ESTÁ	57.091	0,16	99	SEÇÃO	29.045	0,08
50	ENTRE	55.609	0,16	100	R\$	28.054	0,08

**ANEXO VI - As cem palavras mais usadas do *subcorpus* falado do Banco de Português**

Nº	Pal.	Freq.	%	Nº	Pal.	Freq.	%
1	E	113.061	3,73	51	ESTÁ	9.523	0,31
2	QUE	108.883	3,59	52	VAI	9.482	0,31
3	A	77.882	2,57	53	COISA	9.273	0,31
4	É	75.609	2,49	54	QUANDO	8.586	0,28
5	O	71.329	2,35	55	AGORA	8.145	0,27
6	DE	66.922	2,21	56	ATÉ	8.050	0,27
7	NÉ	64.875	2,14	57	ACHO	7.954	0,26
8	NÃO	62.445	2,06	58	ME	7.926	0,26
9	EU	55.733	1,84	59	OU	7.834	0,26
10	F	45.235	1,49	60	BEM	7.670	0,25
11	EST	38.952	1,29	61	AH	7.509	0,25
12	UM	29.923	0,99	62	SABE	7.267	0,24
13	PRA	28.731	0,95	63	HOJE	6.977	0,23
14	TEM	26.316	0,87	64	ALI	6.931	0,23
15	UMA	25.499	0,84	65	DAÍ	6.931	0,23
16	ELE	23.104	0,76	66	SÃO	6.869	0,23
17	ASSIM	22.951	0,76	67	NUM	6.702	0,22
18	DO	22.261	0,73	68	MEU	6.674	0,22
19	ENTÃO	21.466	0,71	69	SEI	6.606	0,22
20	SE	21.455	0,71	70	NÓS	6.583	0,22
21	NA	21.016	0,69	71	DEPOIS	6.526	0,22
22	EM	20.400	0,67	72	MESMO	6.492	0,21
23	DA	19.638	0,65	73	CASA	6.400	0,21
24	GENTE	19.330	0,64	74	RISOS	6.231	0,21
25	NO	18.928	0,62	75	FAZER	6.214	0,21
26	ERA	18.845	0,62	76	ESSE	6.209	0,20
27	COM	18.755	0,62	77	MINHA	5.871	0,19
28	MAIS	18.238	0,60	78	SEMPRE	5.690	0,19
29	AÍ	17.316	0,57	79	ANOS	5.591	0,18
30	AQUI	16.582	0,55	80	BOM	5.358	0,18
31	LÁ	16.271	0,54	81	TU	5.342	0,18
32	MUITO	15.926	0,53	82	ESSA	5.236	0,17
33	PORQUE	14.652	0,48	83	DIA	5.130	0,17
34	ELES	14.145	0,47	84	SENHORA	5.035	0,17
35	MAS	13.711	0,45	85	NEM	4.760	0,16
36	VOCÊ	13.672	0,45	86	RUÍDO	4.641	0,15
37	OS	13.373	0,44	87	SENHOR	4.616	0,15
38	TINHA	13.281	0,44	88	SER	4.512	0,15
39	ELA	13.126	0,43	89	DIZER	4.503	0,15
40	JÁ	13.121	0,43	90	TEMPO	4.500	0,15
41	COMO	12.704	0,42	91	AINDA	4.280	0,14
42	HES	12.704	0,42	92	SIM	4.071	0,13
43	FOI	12.249	0,40	93	TODO	4.004	0,13
44	TAMBÉM	11.910	0,39	94	ESTAVA	3.962	0,13
45	ISSO	11.718	0,39	95	ÉPOCA	3.850	0,13
46	POR	11.115	0,37	96	TENHO	3.768	0,12
47	ININT	10.374	0,34	97	TER	3.744	0,12
48	TUDO	9.899	0,33	98	IA	3.646	0,12
49	SÓ	9.871	0,33	99	PARA	3.611	0,12
50	AS	9.831	0,32	100	DOS	3.594	0,12

**ANEXO VII - As cem palavras mais frequentes do *corpus* geral do *internetês*.**

Nº	Pal.	Freq.	%	Lema	Nº	Pal.	Freq.	%	Lema
1	E	13.930	2,52		51	ISSO	1.578	0,28	
2	QUE	11.537	2,08		52	ADORO		1.572	0,28
3	EU	10.619	1,92		53	SER	1.438	0,26	
4	A	10.173	1,84		54	NAO	1.430	0,26	
5	DE	9.950	1,80		55	TI	1.408	0,25	
6	Q	9.028	1,63		56	OS	1.404	0,25	
7	TE	8.937	1,61		57	NOS	1.386	0,25	
8	O	8.396	1,52		58	SEI	1.382	0,25	
9	É	7.855	1,42		59	BEM	1.365	0,25	
10	VC	6.347	1,15		60	SABE	1.353	0,24	
11	PRA	5.640	1,02		61	AMIGO		1.324	0,24
12	UM	5.252	0,95		62	T	1.324	0,24	
13	AMO	4.960	0,90		63	FALAR		1.307	0,24
14	EH	4.525	0,82		64	MESMO		1.291	0,23
15	UMA	4.386	0,79		65	DIA	1.290	0,23	
16	MAIS	4.249	0,77		66	AMIZADE		1.275	0,23
17	ELA	4.051	0,73		67	SUA	1.255	0,23	
18	SEMPRE	3.905	0,71	0,71	68	SÓ	1.251	0,23	
19	SE	3.851	0,70		69	P	1.230	0,22	
20	MAS	3.683	0,66		70	TEMPO		1.214	0,22
21	NÃO	3.564	0,64		71	ESSE	1.210	0,22	
22	ME	3.524	0,64		72	ESSA	1.194	0,22	
23	MUITO		3.399	0,61	73	NAUM	1.192	0,22	
24	COM	3.369	0,61		74	NEM	1.186	0,21	
25	POR	3.215	0,58		75	FOI	1.147	0,21	
26	DO	3.175	0,57		76	I	1.145	0,21	
27	PQ	3.158	0,57		77	SEU	1.143	0,21	
28	MEU	3.148	0,57		78	QUANDO		1.141	0,21
29	DA	2.965	0,54		79	SEM	1.129	0,20	
30	D	2.960	0,53		80	COISA	1.061	0,19	
31	ELE	2.722	0,49		81	PODE	1.059	0,19	
32	TU	2.699	0,49		82	TER	1.046	0,19	
33	MINHA		2.624	0,47	83	VOU	1.035	0,19	
34	NA	2.492	0,45		84	TA	1.033	0,19	
35	EM	2.442	0,44		85	DIZER	1.026	0,19	
36	NO	2.340	0,42		86	TO	1.009	0,18	
37	AS	1.989	0,36		87	NUNCA		1.005	0,18
38	TUDO	1.939	0,35		88	ESPECIAL		982	0,18
39	TEM	1.837	0,33		89	NEH	966	0,17	
40	VOCÊ	1.822	0,33		90	ASSIM	945	0,17	
41	BOM	1.792	0,32		91	AI	937	0,17	
42	COMO	1.735	0,31		92	AMOR	908	0,16	
43	VIDA	1.715	0,31		93	CARA	907	0,16	
44	MIM	1.649	0,30		94	FAZ	901	0,16	
45	VAI	1.640	0,30		95	NOSSA		877	0,16
46	AMIGA		1.617	0,29	96	X	874	0,16	
47	PARA	1.617	0,29		97	MT	855	0,15	
48	PESSOA		1.617	0,29	98	FALA	829	0,15	
49	GENTE		1.607	0,29	99	C	810	0,15	
50	MTO	1.587	0,29		100	TENHO		797	0,14

**ANEXO VIII - As cem palavras mais frequentes do *corpus* do português falado NURC –****Norma culta**

N	Palavra	Freq.	%	Lema	Nº	Pal.	Freq.	%	Lema
1	QUE	3.973	3,71		52	TUDO	317	0,30	
2	DE	3.109	2,91		53	CASA	314	0,29	
3	EU	2.885	2,70		54	ME	309	0,29	
4	NÃO	2.739	2,56		55	BEM	298	0,28	
5	É	2.566	2,40		56	AGORA	295	0,28	0,28
6	A	2.534	2,37		57	SÃO	295	0,28	
7	E	2.422	2,26		58	QUER	288	0,27	
8	O	2.300	2,15		59	ELES	286	0,27	
9	NÉ	1.439	1,35		60	SEI	281	0,26	
10	UM	1.349	1,26		61	OU	273	0,26	
11	UMA	1.259	1,18		62	RIO	270	0,25	
12	DOC	1.031	0,96		63	MESMO	259	0,24	0,24
13	VOCÊ	1.010	0,94		64	AQUI	255	0,24	
14	MUITO	977	0,91	0,91	65	SUP	253	0,24	
15	MAS	962	0,90		66	MINHA	250	0,23	0,23
16	TEM	953	0,89		67	FAZER	235	0,22	
17	EM	939	0,88		68	HOJE	235	0,22	
18	PRA	876	0,82		69	ESTÁ	232	0,22	
19	SE	863	0,81		70	MEU	232	0,22	
20	DO	862	0,81		71	DIA	221	0,21	
21	LOC	841	0,79		72	TÁ	221	0,21	
22	NA	789	0,74		73	ESSE	214	0,20	
23	DA	761	0,71		74	TENHO	211	0,20	0,20
24	MAIS	730	0,68		75	PESSOAS	207	0,19	0,19
25	ERA	729	0,68		76	ANOS	204	0,19	
26	NO	724	0,68		77	TODO	199	0,19	
27	COM	716	0,67		78	NÓS	198	0,19	
28	ENTÃO	708	0,66	0,66	79	ESSA	192	0,18	
29	PORQUE	620	0,58	0,58	80	SABE	188	0,18	
30	ISSO	589	0,55		81	SEMPRE	188	0,18	0,18
31	LÁ	563	0,53		82	COISAS	185	0,17	0,17
32	POR	528	0,49		83	DEPOIS	184	0,17	0,17
33	GENTE	519	0,49	0,49	84	EXEMPLO	182	0,17	0,17
34	FOI	504	0,47		85	AH	176	0,16	
35	ACHO	502	0,47		86	CIDADE	176	0,16	0,16
36	ASSIM	502	0,47		87	INF	167	0,16	
37	AS	472	0,44		88	BOM	166	0,16	
38	OS	448	0,42		89	ÉH	165	0,15	
39	AÍ	441	0,41		90	FUI	162	0,15	
40	COISA	435	0,41		91	REALMENTE	162	0,15	0,15
41	TINHA	434	0,41		92	TER	160	0,15	
42	TAMBÉM	425	0,40	0,40	93	SER	159	0,15	
43	ELE	421	0,39		94	HÁ	157	0,15	
44	JÁ	417	0,39		95	TEMPO	155	0,14	0,14
45	COMO	390	0,36		96	MUITA	152	0,14	0,14
46	ELA	367	0,34		97	AINDA	151	0,14	
47	QUANDO	363	0,34	0,34	98	VEZES	151	0,14	
48	ATÉ	355	0,33		99	AO	150	0,14	
49	VAI	333	0,31		100	EH	148	0,14	
50	DIZER	325	0,30		101	GOSTO	147	0,14	0,14
51	SÓ	324	0,30		102	DOS	146	0,14	

**ANEXO IX - Palavras típicas do *internetês* entre as 500 mais frequentes do *corpus geral do Orkut***

N	Palavra	Freq.	%	N	Palavra	Freq.	%
6	Q	9.028	1,63	241	HEHEHE	301	0,05
10	VC	6.347	1,15	248	BAH	284	0,05
14	EH	4.525	0,82	253	AE	275	0,05
27	PQ	3.158	0,57	257	AMOOO	273	0,05
30	D	2.960	0,53	258	Ñ	273	0,05
50	MTO	1.587	0,29	265	CUM	261	0,05
55	TI	1.408	0,25	270	HEHE	257	0,05
62	T	1.324	0,24	278	M	250	0,05
69	P	1.230	0,22	285	NUNK	246	0,04
73	NAUM	1.192	0,22	286	BJO	245	0,04
76	I	1.145	0,21	288	NE	242	0,04
84	TA	1.033	0,19	303	LAH	227	0,04
86	TO	1.009	0,18	305	MSN	223	0,04
89	NEH	966	0,17	307	QNDO	221	0,04
96	X	874	0,16	316	BJS	211	0,04
97	MT	855	0,15	319	POW	210	0,04
99	C	810	0,15	332	ATEH	199	0,04
102	TAH	789	0,14	335	SI	198	0,04
103	AKI	787	0,14	337	QDO	196	0,04
107	KI	768	0,14	343	AMOO	192	0,03
112	SO	721	0,13	344	DOLU	192	0,03
116	U	693	0,13	345	PAH	192	0,03
117	SOH	684	0,12	351	VCS	188	0,03
126	N	629	0,11	358	BJUS	184	0,03
129	NUM	603	0,11	361	DEPO	183	0,03
142	TBM	568	0,10	398	FAZE	167	0,03
145	TD	563	0,10	404	NOIS	164	0,03
150	MTU	530	0,10	415	KRA	158	0,03
158	AMU	502	0,09	417	TÔ	158	0,03
167	JAH	478	0,09	418	MTA	157	0,03
179	XD	422	0,08	422	V	156	0,03
182	ELI	414	0,07	423	FLA	155	0,03
184	DI	408	0,07	428	COMU	151	0,03
186	CMG	405	0,07	429	DEXA	150	0,03
189	MSM	402	0,07	431	QM	150	0,03
191	TB	387	0,07	436	MIGA	147	0,03
199	AGENTE	371	0,07	439	DIZE	145	0,03
203	DU	363	0,07	446	POD	143	0,03
204	GNT	363	0,07	450	AXO	140	0,03
209	HJ	350	0,06	459	TDS	137	0,02
218	MI	337	0,06	463	NAH	136	0,02
223	OQ	320	0,06	471	Â	134	0,02
227	S	319	0,06	474	OH	132	0,02
228	NU	317	0,06	481	ISSU	128	0,02
233	TAVA	311	0,06	484	FIK	127	0,02
237	L	307	0,06	487	AMOOOOOOOOOOOOO+	1260	0,02
239	PA	306	0,06	497	MTAS	123	0,02

**ANEXO X - Palavras típicas do *internetês* entre as 500 mais frequentes do *corpus* apenas *scraps***

Nº	Palavra	Freq.	%		Nº	Palavra	Freq.	%
6	VC	458	1,39		245	VLW	19	0,06
9	Q	327	0,99		246	XD	19	0,06
43	EH	105	0,32		249	I	18	0,05
44	TO	105	0,32		250	LAH	18	0,05
46	P	102	0,31		252	MANDE	18	0,05
49	D	95	0,29		258	X	18	0,05
59	ADD	75	0,23		266	Ñ	17	0,05
63	BJS	67	0,20		272	WWW	17	0,05
64	PQ	67	0,20		284	R	16	0,05
82	N	56	0,17		302	PA	15	0,05
83	TBM	56	0,17		308	BEIJINHOS	14	0,04
85	NAUM	55	0,17		309	BJ	14	0,04
91	TD	51	0,16		310	BR	14	0,04
106	AKI	45	0,14		321	MSM	14	0,04
111	TAH	43	0,13		322	OO	14	0,04
116	NEH	42	0,13		323	PO	14	0,04
118	SO	42	0,13		326	U	14	0,04
121	MSN	40	0,12		327	ATEH	13	0,04
126	FDS	39	0,12		339	RSRS	13	0,04
129	C	38	0,12		341	VISS	13	0,04
138	VCS	36	0,11		343	AEW	12	0,04
141	TB	35	0,11		354	HEHE	12	0,04
144	BLZ	34	0,10		357	OQ	12	0,04
145	KI	34	0,10		359	OW	12	0,04
146	KKKKKKKKKKKKKKK+			34	387	RSRSRS	11	0,03
	0,10				395	XAU	11	0,03
149	VO	34	0,10		396	AII	10	0,03
152	BJOS	32	0,10		399	BJOO	10	0,03
155	HJ	32	0,10		400	BJU	10	0,03
157	S	31	0,09		419	MI	10	0,03
166	T	30	0,09		420	MT	10	0,03
172	AE	28	0,09		422	NU	10	0,03
173	AH	28	0,09		423	OBS	10	0,03
177	TI	28	0,09		424	OIEE	10	0,03
188	MTO	26	0,08		442	BJAUM	9	0,03
192	SOH	25	0,08		443	BJIM	9	0,03
193	TÔ	25	0,08		444	BJINHOS	9	0,03
197	POW	24	0,07		458	HEHEHEHE	9	0,03
201	JAH	23	0,07		462	KKKKK	9	0,03
208	BJUS	22	0,07		465	NE	9	0,03
213	OK	22	0,07		466	NGM	9	0,03
214	RS	22	0,07		469	OIII	9	0,03
216	BJÃO	21	0,06		481	TUH	9	0,03
217	BJO	21	0,06		485	VX	9	0,03
228	ESPECIAL	20	0,06	0,06	489	AXO	8	0,02
229	FLW	20	0,06		492	BJAO	8	0,02
231	L	20	0,06		495	CEL	8	0,02
239	HEHEHE	19	0,06	0,06	498	CMG	8	0,02
240	OIE	19	0,06					

**DADOS INTERNACIONAIS DE CATALOGAÇÃO-NA-PUBLICAÇÃO(CIP)**  
**BIBLIOTECÁRIO RESPONSÁVEL: Rafael Bertoglio**  
**CRB-10/1608**

B622d            Bisognin, Tadeu Rossato  
                    Do *internetês* ao léxico da escrita dos jovens no Orkut /  
Tadeu Rossato Bisognin. – Porto Alegre, 2008.  
                    259 f.

                    Dissertação (Mestrado em Teorias do Texto e do Discurso)  
- Universidade Federal do Rio Grande do Sul. Instituto de  
Letras, Programa de Pós-Graduação em Letras. Porto Alegre,  
BR-RS, 2008. Orientadora: Prof. Dra. Maria José Bocorny  
Finatto.

                    1. Lingüística de *corpus*. 2. Lexicologia. 3. Lexicografia.  
4. Oralidade e escrita. 5. Comunicação eletrônica. 6.  
Internetês. 7. Orkut. 8. *Miguxês*. 9. Atividades em sala de aula.  
I. Título.

CDD 410