

LEXICOMETRIA, GERAÇÃO DE DESCRITORES, CONSTRUÇÃO DE ONTOLOGIAS E ENSINO DE LÍNGUAS: IMPLICAÇÕES E PERSPECTIVAS

Elizabete Aparecida DAMASCENO (UNIFIEO/IFIP)¹

RESUMO: Atualmente tem-se tornado, sobretudo no domínio da educação, extremamente necessário aprofundar aspectos relacionados não só à natureza e propriedades, criação e utilização de base de conhecimento, como também questões relacionadas aos métodos adequados para a geração e aplicação de conhecimento. Neste documento, discute-se a lexicometria como uma metodologia adequada para a identificação de redes de conhecimento para redefinição de pré-requisitos e exigências para desenvolvimento de competências, importantes para o processo de interação, inovação, desenvolvimento e gerenciamento do potencial humano.

Palavras-chave: lexicometria; geração de descritores; ensino de línguas; base de conhecimento

ABSTRACT: Nowadays it has been extremely necessary to delve deeply into the education domain, aspects related not only to nature and attributes, creation and use, but also the suitable methods for the generation and application of knowledge. In this study, lexicometry is analysed as an appropriate methodology to identify knowledge networks to redefine pre-requisites and demands for the development of competences, important for the interaction, innovation, development and management process of the human potential.

Keywords: lexicometry; generation descriptors; education language learning. knowledge-base

1. Introdução

Em sociedade, a dinamicidade dos diferentes e múltiplos agrupamentos sociais confere existência a uma variabilidade de discursos-concretos, produzidos e reproduzidos no quadro geral da comunicação da vida social, expressões de diferentes e múltiplas significações e visões-de-mundo. A relação entre espaço e homem motiva, constantemente, o aparecimento considerável de construções léxico-sintático-semânticas de dois fazeres - individual e coletivo - que possibilitam a existência de uma sintaxe das significações humanas, em suas relações interiores e exteriores, resultando em integração e interação dos sujeitos pela produção de mensagem plurissignificativas e, por isso, plausíveis.

Diversos e diferentes são os sistemas e subsistemas de significação produzidos e em circulação no espaço social, cuja possibilidade de produção e performances emissivas e receptivas dos membros de uma comunidade assinalam a constituição de rede não só das comunicações sociais generalizadas, como das formas e modos de seu exercício. Tanto uma quanto outra apresentam-se como caracteres asseguradores de estatuto sócio-semiótico do indivíduo, que assume papéis e funções sócio-semióticas diversos e participa, por conseguinte, de diferentes grupos e subgrupos, em que trocas são operadas, prestações (com)partilhadas e, de algum modo, determinadas, imperativas em suas formas e inconscientes em sua estrutura.

A comunidade, pelo reconhecimento do lugar de condição de sua verdade, estabelece na palavra social a noção de normalidade: a língua é o lugar de todos os discursos, porque cumpre a presença ideológica dissimulada da realidade, pela condição de falar o objeto no discurso e dissimular a presença da sua ausência no jogo de ocultamento, causa de arbitrariedade. O reconhecimento de uma sociabilidade semiótica permite-nos postular um significado de natureza

1. projeto desenvolvido no UNIFIEO/Instituto FIEO de Pesquisas -IFIP

coletiva; todavia, recoberto por diversas e múltiplas linguagens de manifestação recorrente a diferentes canais e códigos de comunicação, muitas vezes, heteróclitos.

A significação, em processo, remete a uma outra significação e o significante, sozinho, garante a coerência do conjunto como conjunto. A rede significante diz respeito à estrutura sincrônica; e a rede do significado, ao conjunto diacrônico dos discursos concretamente pronunciados. E, na cadeia significante, o sentido insiste sem, contudo, consistir numa significação, de que a única forma possível é o acolchamento,

o ponto de cruzamento entre o percurso do significante (na cadeia) e a elipse deslizante do sujeito, de que o sentido é situação, e o processo é a semiose ilimitada. A força do significante comanda em seus próprios nós o movimento do significado, característica de um sistema de transformações, assumido por um enunciador-enunciatário de discursos e representado por múltiplas e significativas relações intra- e interpessoais.

Desse ponto de vista, a instância da enunciação passa a ser vista como instância actancial e a comunicação de significação não mais se apresenta como simples transmissão do saber, porque pressupõe duas relações actancias: o enunciador e o enunciatário de discursos. Entre destinador e destinatário ocorre não só transmissão de valor e assunção de papéis bem como uma relação entre sujeito e seu objeto causa de desejo: relação de busca e construção de significação, de que participam as dimensões cognitiva, pragmática e afetiva, cujo processo de identificação do sujeito funda-se sobre partilha de um *socius* e múltiplas combinações para a emergência de um fazer transformador.

A existência de estruturas narrativas permite a recorrência de um texto a outro(s), bem como a articulação de conteúdos, definidos como forma de organização do imaginário humano. Estruturas se submetem não só a um sistema de organização determinado por fatores históricos, lingüísticos e culturais, como também se sujeitam a uma gramática narrativa geral com a qual se encontram interligadas. Nesse sentido, sintaxe passa a ser entendida como “bombardeio” dos objetos e sujeitos para definir o conjunto de inter-relações e de interações entre sujeitos e objetos, de que o fazer humano - processo e produto - aparece significativo, mas estágio disforizante, porque lugar de dessemantização e de alienação.

Eis uma questão epistemológica: a passagem do actante-indivíduo ao actante-coletivo, dotado e investido de modalidades axiológicas. Ora, um sujeito coletivo é constituído pela integração de um poder-fazer coletivo. A assunção comum das modalidades que compõem e constituem a competência do sujeito, inscrito num programa de enunciados regulamentadores da instauração de um sujeito-actante-coletivo passa a ser uma problemática de ordem taxionômica e pertencer à estrutura e tipologia sociais, já que o que se tem é um sujeito hiponímico representante de um actante-coletivo.

Desta feita, compreender um discurso, inscrito em Universo de Discurso delimitado, não é uma questão de simples construção passiva de uma representação do objeto verbal, porque parte de um processo interacional, em que pares encontram-se envolvidos como sujeito do saber e do fazer manipulatório-discursivo. Componentes lingüísticos, retóricos e pragmáticos participam da produção e da interpretação de textos, unidade de manifestação discursiva concreta, de gêneros e tipos diversos e diferentes, além de níveis e registros lingüísticos singulares.

Um estudo científico da utilização da linguagem parte de um ato metafórico, que tem como objeto direto enunciados teóricos, e como objeto indireto suas aplicações, de acordo com setores específicos da realidade. Desta feita, processar um discurso envolve a interação de dimensões afetivas, cognitivas e pragmáticas, dado que produzir discursos é atividade relacionada a processos funcional e contextual, pois usuários de uma língua natural constroem uma representação mental tanto do texto e do contexto social, quanto da interação das dimensões envolvidas. E na própria comunidade em que o sujeito-falante se insere, a interpretação é já um pressuposto internalizado, assumido sob a condição de “formação discursiva”.

Por um lado, em relação a um modelo de interpretação de discurso(s), várias são as limitações, tais como: 1.processos de informação semântica; 2.representação do conhecimento e do uso lingüístico; 3.informações contextuais. Por outro, devem ser considerados os seguintes pressupostos da operação interpretativa: 1.representação mental e acontecimentos e compreensão deles por (re)construções seqüentes e subseqüentes de eventos e armazenagem de informações visuais e lingüísticas e a conseqüente disponibilidade para atualizações e reatualizações em novos atos discursivos; 2.representação simultânea do processamento da informação, compreensão e armazenagem simultânea de informações em que crenças, opiniões ou atitudes específicas de cada comunidade ou agrupamento sócio-lingüístico-cultural aparecem como elementos singulares e motores; 3.motivações, objetivos e/ou tarefas singularizadoras e contribuintes do processamento de acontecimentos, de sua compreensão e também do processamento da interpretação de informações exteriores, ativação e uso, cuja resultante é a construção de derivações, segundo a capacidade de uso, flexível, de informações.

O processo da descoberta de conhecimento é um processo de análise complexo. Desse aspecto, interroga-se sobre as expressões, por exemplo, no tocante à construção de ontologias que, sem alguma relação significando substância do ser, são categorizadas em quantidade; qualidade; relação; lugar; tempo; posição; possessão; fazer, cuja organização retórica tem base semântico-gramatical, constituindo quadro de descrição (ou exposição) prontos para posterior definição funcional, como nas gramáticas escolares. E, além disso, um outro fator é a disposição de metaforização interna (transferência interior), a partir da visada de uma língua, em que a explicação de universais lingüísticos se dá por correspondência a um jogo de estratégias contextuais, para além do exercício da categorização.

Ainda, do ponto de vista de uma arqueologia do saber lingüístico (COSERIU:1992), um modelo teórico, definido em razão de teorias e modelos, insere-se num contato social, global e específico sob três componentes: 1.formalização ou conjunto de técnicas de descrição como um inventário constituído de regras descritivas ou eventualmente normativas e de constituintes ou meta-regras; 2.conceptualização ou nível de formação teórica do próprio modelo em que são alvo as macroestruturas do objeto descrito e a articulação das unidades da descrição e seu processo de funcionamento lingüístico-discursivo; 3.modelização ou auto-imagem do objeto e dos procedimentos da descrição e da relação ente objeto/sujeito, para o que se desdobra o sujeito lingüístico e o sujeito da descrição.

Como a linguagem é determinada, em essência e manifestação, por dimensões universais, como criatividade; materialidade; semanticidade; alteridade; historicidade, e definida como atividade histórica e social, três são os níveis relativamente autônomos a serem considerados: 1.nível do falar em geral ; 2.nível histórico; 3.nível do falar concreto individual e particular (o texto, definido como unidade de significação), níveis aos quais correspondem o saber elocucional, idiomático e expressivo, respectivamente. E a esses saberes correspondem, ainda, a designação, o significado e o sentido, três tipos específicos de conteúdo.

Desta feita, decorre a definição, teórica, de língua histórica como o conjunto complexo de dialetos, níveis e estilos, marcada por criatividade, manifestada pelas variedades diatópica ou espacial; diastrática ou sócio-cultural e diafásica ou estilística. Por outro lado, marcada pela homogeneidade ou unidades sinstrática (dialeto); sintópica (níveis de linguagem); e sinfásica (estilos de linguagem), já que o falante é o ponto de convergência e conflito de n-normas de discurso(s) como também é seu (re)produtor social.

Desse ponto de vista, o campo metodológico coloca em foco questões epistemológicas que demandam estudos, pois importam as condições de possibilidade, uma vez que no espaço do saber aparecem configurações que dão lugar a formas diversas do conhecimento empírico. Descrever, então, uma estrutura é dar conta das dependências e independências, sem ignorar a problemática semântica, porque não se pode reduzi-la a apenas um repertório de significações diferentes, pois uma variante é sempre função de uma forma; e a significação particular, da significação fundamental, deduzida, logicamente, de significações particulares. Significa isto dizer que a função constitui dependência; e inerência, a estrutura.

2. A análise lexicométrica e a produção de sentido(s)

A lexicometria é um procedimento metodológico e tecnológico - objetivo, descritivo, indutivo e científico - para tratar estatisticamente dados qualitativos sob fundo quantitativo para a caracterização topológica e combinatória de elementos lexicais de um corpus dado e delimitado, a fim de que a trajetória do discurso, por operações conhecidas e controladas, seja balizada, e a topologia e a combinatória dos elementos lexicais do corpus caracterizadas.

Permite descobrir co-relacionamentos e dados implícitos nos registros de um corpus, pelo estudo e desenvolvimento de um processo de extração por: a) detecção de dependências entre os dados, para identificar atributos e relações de interdependência; b) detecção de desvios para identificar elementos fora dos padrões estabelecidos/esperados; c) identificação e análise de estruturas e de objetos com características comuns; d) descoberta de fórmulas ou modelos para descrever conceitos envolvidos.

A análise lexicométrica, auxiliada por ferramenta tecnológica, no caso deste estudo pela ferramenta *STABLEX* (CAMLONG & BERTRAND - *Stablex* - Software PC 2004), permite descobrir co-relacionamentos e dados implícitos nos registros de um corpus, pelo estudo e desenvolvimento de um processo de extração, pela:

- a) detecção de dependências entre os dados: identificação dos atributos e suas relações de interdependência;
- b) detecção de desvios: identificação de elementos que se encontram fora dos padrões estabelecidos ou esperados;
- c) identificação e análise de estruturas e de objetos com características comuns;
- d) descoberta de fórmulas ou modelos capazes de descrever os objetos ou conceitos envolvidos.

Uma metodologia adequada é construída e compreende etapas de:

- a) pré-processamento ou mineração de dados: processos de análise, integração, reinformações e limpeza de dados;
- b) processamento: processamento de dados, predição, regressão, classificação, constituição de agrupamentos, associações, visualizações;

- c) pós-processamento de objetos: filtragem, seleção, ordenação das descobertas, mapeamentos de representação e apresentação de base de conhecimento.

A base de dados, considerada neste estudo, constituiu-se de redações de alunos de 8ª série, última série do ciclo fundamental de ensino, de seis diferentes escolas da rede estadual paulista, selecionados aleatoriamente dez alunos de cada uma das seis escolas, cujos textos foram produzidos em três momentos diferentes do processo de ensino-aprendizagem, totalizando, então, três diferentes momentos da produção textual dos alunos - períodos inicial, medial e final do processo de ensino-aprendizagem de um ano letivo.

Os textos coletados receberam tratamento especial para a classificação dos objetos (palavras), em que técnicas de análise de informação de dados lexemáticos foram aplicadas para a identificação de padrões e relacionamento entre os objetos (vetores), de sorte que os textos, depois desse tratamento, foram convertidos em formato TXT ou ASC II, que é semi-estruturado ou desestruturado, por apresentarem características complexas.

Operou-se com a descoberta de padrões e de associações entre os padrões para se efetuar, a partir dessa etapa, a definição de classes ou de categorias, uma vez que não se conhecia de antemão a classificação a que os objetos seriam rotulados. Identificou-se a melhor distribuição ou configuração não pela comparação de dados a objetos já conhecidos, mas com base na identificação de similaridade entre os objetos, segundo atributos comuns, cujas variáveis analisadas geraram grupos homogêneos, conforme análise combinatória e de alta complexidade.

Do tratamento do corpus, feitas as primeiras extrações ou ordenação dos dados foram obtidas a confecção do Léxico - Léxico Alfa e Léxico Delta - disposição dos dados segundo o critério da frequência e indexação das palavras, a que se seguiu a apresentação da tabela de frequência, e desta as operações subseqüentes.

3. Inventário e segmentação léxica: dependências e interdependências

Segmentar o enunciado para reconhecer as unidades léxicas interpõe dificuldades. A língua é inacessível diretamente, pois constitui um universo de abstrações, cuja manifestação concreta se dá por meio da fala pelo exercício físico-psíquico da linguagem. A língua é produto social da faculdade da linguagem e um conjunto de convenções necessárias adotadas pelo corpo social para permitir o exercício dessa faculdade nos indivíduos (SAUSSURE *apud* BIDERMAN:2001), de que a fala é uma prática individual, concreta e particular.

Concretamente, porém, como atividade lingüística só existem os atos de fala ou de escrita. Importante assinalar, de acordo com COSERIU, a tríplice série de conceitos sobre língua - 1.acervo funcional; 2.instituição social; 3.sistema funcional - que, se de um lado confirma a própria heterogeneidade do fenômeno lingüístico, destaca, de outro, a dificuldade encontrada para a análise de nosso objeto: a unidade lingüística.

Salientamos o caráter de forma e produto da língua, ao mesmo tempo que assinalamos sua natureza inter-subjetiva, extra-individual e social. A fala, todavia, nota-se pela ação verbal, porque é concreta, individual e subjetiva, e também caracterizada como produto social, a posteriori. Assim, cada ato de fala oral ou escrito de um falante e registrado eletronicamente pode, então, ser considerado um produto, como um documento da atividade lingüística desse falante (*energéia*) e amostra de sua língua (*érgon*), um encontro de duas realidades: língua e fala (a virtual e a realizada). Os falantes têm, então, um acervo das infinitas realizações de linguagem armazenado em memória, um sistema de possibilidades ou um conjunto de possibilidades, a que poderemos considerar como língua, embora o falante não a possua toda.

No entanto, essa dicotomia necessita de reformulação. Entre língua e fala há uma instância intermediária, que é a norma, definida como o conjunto dos hábitos lingüísticos dos falantes de uma dada e definida comunidade lingüístico-cultural; instância reguladora dos atos de fala e escrita da comunidade e de seus membros; ou instância coercitiva de realizações obrigatórias e de imposições sociais e culturais, que é variável segundo fatores diversos, como comunidade, situação e contextos históricos, culturais, econômicos, regionais e outros. A norma, então, “contém o sistema e mais os elementos normais no falar de uma comunidade: as oposições não-pertinentes mais usuais”. (BIDERMAN:2001).

Entende-se o porquê de as maiores dificuldades estarem justamente no Campo Léxico, onde funções representativas e associativas estão atuando, dificuldades advindas da vasta e variada complexidade das oposições, e mais se considerarmos as unidades léxicas não como unidades ou entidades isoladas, mas elementos organicamente opostos e associados constitutivos de um sistema. É nesse sentido que se depara o pesquisador com uma dificuldade: como considerar a unidade léxica? Trata-se de lexia simples, composta ou

complexa? Como definir a unidade lingüística para dar conta das criações, das associações vocabulares e valores de significação inéditos na norma léxica da língua, donde resulta a riqueza do léxico de uma língua?

4. Geração de descritivos: ambigüidades e (re)categorizações

O esquema representativo levou em conta a representação por vetores de palavras - seleção e extração - pelo critério de seleção de características, conforme representabilidade, efetividade e relevância, em que a frequência absoluta é medida importante, mas insuficiente, porque não demarca a possibilidade combinatória, característica dos signos lingüísticas.

Dá-se que enquanto a língua sente necessidade da classificação das unidades como formas gramaticais de distribuição organizada em compartimentos específicos, a fala, dada sua natureza, furta-se a esse modelo de ordenação, uma vez que o discurso enxerta variedades e multiplicidades, razão de riqueza e não de fragilidade do sistema lingüístico, dada a sua natureza não-exclusivamente lógica.

Além do estudo da frequência absoluta (TDF), também, e, principalmente, fez-se a análise da frequência relativa, pela geração da(s) Tabela(s) de Desvios Reduzidos (TDR), que pode ser dada pela relação do vetor a todos os outros, para identificar o peso ou a força de representabilidade de cada palavra, e se operar um normalização do número de ocorrências em relação ao número total de ocorrências de todas as palavras presentes nos documentos, a exemplo das tabelas 01 e 02.

TDF -Tabela de Frequência

	TDF	47		1550	421	509	620
		3		p	0,2716	0,3284	0,4
				q	0,7284	0,6716	0,6
LÉXICO	linha	freq	n°	freq	P1	P2	P3
que	1	208	1	208	51	73	84
a	2	128	1	128	42	30	56
de	3	121	1	121	29	35	57
...	...	...	...	...	...	...	...
eu	12	55	1	55	11	44	0
felicidade	13	53	1	53	53	0	0
...	...	...	...	...	...	...	...
...	46	344	172	2	86	115	143
...	47	849	849	1	200	285	364

Tabela 01 - TDF/ escola 01

TDR -Tabela de Desvios Reduzidos

0,784	Total	9,679	-5,207	-3,688
0,017	Média	0,206	-0,111	-0,078
0,061	Khi2	0,042	0,012	0,006
	0,996			
Desvio	média	Maximo	Mínimo	
	0,017	12,048	-5,670	
Linha	Fréq	P1	P2	P3
1	-0,061	-0,726	-0,379	1,044
2	0,119	1,551	-3,040	1,608
3	-0,062	-0,691	-1,700	2,329
...	...	...	...	...
12	-0,073	-1,130	6,728	-5,670
13	0,979	12,048	-5,503	-5,566
...	...	...	...	...
47	-0,179	-2,099	-1,692	3,612

Tabela 02 -TDR /escola 01

Foram trabalhadas medidas de frequência absoluta, relativa e também inversa: vetores foram selecionados em algoritmos de classificação ou de recuperação de informação com perspectivas de identificação de centróides ou matrizes. O corte se fez a partir do próprio peso pela estipulação de valores mínimos e máximos, descritos segundo graus de liberdade (03 ddl.) e escala de repartição (*lei de Laplace-Gauss*) e também pelo cálculo das probabilidades.

A análise paramétrica, ferramenta de análise e síntese - objetiva, indutiva e científica - permitiu observar, decompor e manipular dados, auxiliada pelo software *STABLEX*, e chegar, pela própria arquitetura dos textos, à composição do discurso.

Essa análise da representação da situação textual se deu de forma controlada pela especificidade, por planos de produção e compatibilizações sêmicas e semêmicas (BARBOSA:1996), em que elementos lingüísticos foram reduzidos a dados mensurados e proporcionalmente dimensionados, de tal modo que, pelo exercício do processo de lematização, foi possível examinar o índice nocional, quantitativo e qualitativo dos elementos do *corpus*.

Para a solução de ambigüidades classificatórias e polivalência de algumas formas, e para a produção de novas características mais salientes, ou mais discriminantes ou mais descritivas, criaram-se novos descritores, quer pelo mapeamento de conceitos, pela transformação de palavras em estruturas mais complexas, quer pela análise fatorial em que as palavras foram reduzidas à dimensionalidade, pelo agrupamento de variáveis correlacionadas, operações possíveis pelo exercício do processo de lematização.

A operação também contou com a medição da proximidade, associação ou semelhança entre padrão, o que tornou possível descrever os documentos por valores quali-quantitativos. Esses valores de proximidade - proximidade, associação e semelhança - correlacionam-se à distância (física, no espaço euclidiano) entre os vetores (objeto-palavra), em que, na operação, por se relacionar mais ao valor quantitativo, o conceito de distribuição foi substituído pelo conceito de similaridade, dado por medidas de distância, coeficientes de correlação, coeficientes de associação e medidas probabilísticas de similaridade, como se pode observar pelo **gráfico 01**, correspondente ao exame do conjunto de textos/períodos (período inicial-P1; período medial-P2 e período final-P3, no gráfico T1, T2 e T3 respectivamente) de apenas uma das seis escolas estudadas, a escola 01.

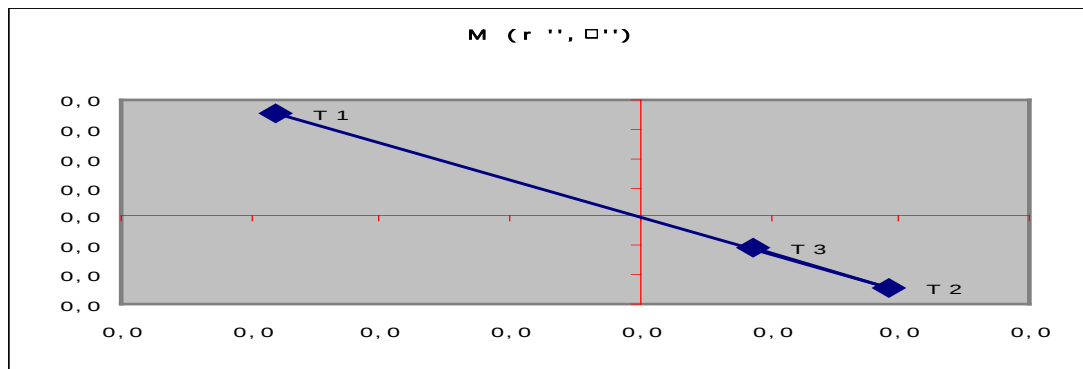


Gráfico 01 - exame da proximidade/similaridade do conjunto de textos/período da escola 01

Foi também considerada a métrica, uma medida estabelecida de forma objetiva, reaplicável e quantificada, segundo a natureza das características do próprio objeto. Foram obtidas a simetria, a densidade triangular e a densidade de não-identidades e de equivalências. A distância nos deu a distribuição dos objetos em espaço euclidiano, em que a cada coordenada correspondeu uma característica; e a similaridade foi dada pela proximidade, associação e semelhança das palavras nesse espaço, como se pode analisar pela **Tabela 03**, correspondente à correlação entre os conjuntos de textos/períodos correspondentes apenas à escola 01.

Correlação	0,923	0,974	0,973
	T1	T2	T3
T1	1,000	0,926	0,920
T2	0,926	1,000	0,946
T3	0,920	0,949	1,000

Tabela 03 - Tabela da correlação referente à análise do conjunto de textos/período da escola 01

Deste estudo obteve-se a análise do agrupamento de entidades similares ou a análise de regiões contíguas, conforme o **Gráfico 02, 03 e 04**, correspondentes aos períodos inicial (P1), medial (P2) e final (P3) do processo de ensino-aprendizagem de um ano letivo, referentes ao estudo do conjunto dos textos da escola 01, apresentando alta densidade relativa de pontos, e baixa densidade de pontos, operação caracterizada por densidade, variância ou grau de dispersão.

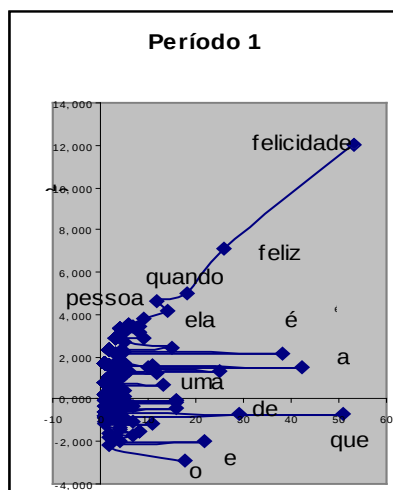


Gráfico 02 - período 1/escola 01

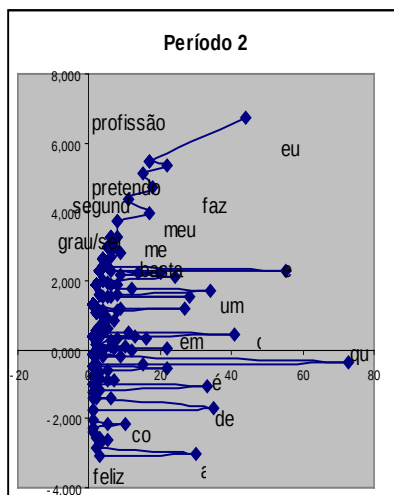


Gráfico 03 - período 2/escola 01

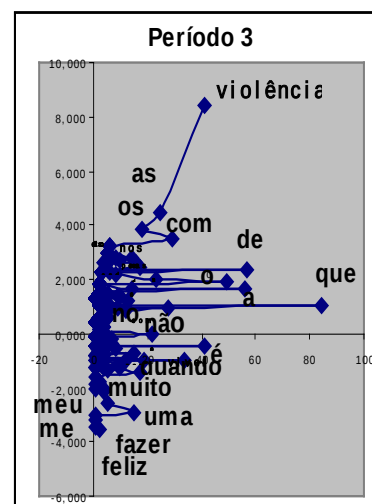


Gráfico 04 - período 3/escola 01

5. Identificação e análise de agrupamentos por conceitos

É fundamental a adequação de uma metodologia para a identificação e análise de agrupamentos, decorrendo de que a análise apenas pelo critério de frequência é ainda insuficiente.

Colocam-se, aqui, dificuldades de se estabelecer, em língua, categorizações, dado que as classes de palavras não são “compartimentos estanques”, não possuem, pois, etiquetagem definitiva, dada a potencialidade combinatória dos signos lingüísticos, considerados como unidades ou de língua, ou de norma ou de fala, segundo os diferentes referentes contextuais e situacionais.

Para dar conta da questão e minimizar a questão, a lexicometria oferece técnicas, como as sumariamente descritas, que tornam possível um estudo controlado e padronizado, de tal sorte que foi possível para cada um dos conjuntos de textos de cada um dos períodos de cada uma das seis escolas analisadas, estudar os valores lexicais e obter o Vocabulário Preferencial, o Vocabulário Comum de base positiva e de base negativa e o Vocabulário Diferencial.

Para uma amostragem do estudo realizado, o **gráfico 05** apresenta os dados relacionados somente ao conjunto de textos dos alunos da escola 01, de acordo com os períodos inicial, medial e final do processo de ensino-aprendizagem de um ano letivo, de cujos valores - desvios reduzidos/peso - o leitor poderá tecer suas próprias considerações.

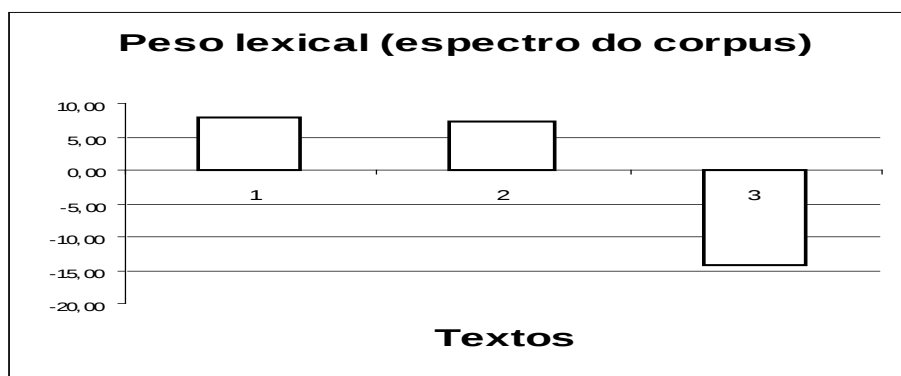


Gráfico 05- peso do corpus do conjunto dos textos de alunos da escola 01 / período inicial, medial e final do processo de ensino-aprendizagem de um ano letivo

A análise paramétrica também registrou a presença de descritores específicos: presença de *hápax* ou ocorrências altamente discriminantes, caracterizadas pela especificidade, pois ocorrem uma única vez no conglomerado e especificamente num único e determinado aglomerado, o que dá um forte índice de significância de aspectos identitários do próprio aglomerado em que ocorrem. Como exemplo, o **quadro 01** registra alguns dos vetores do conjunto de textos da escola 01, relacionados aos três períodos (P1 início, P2 - meio e P3 -final) do processo de ensino-aprendizagem de um ano letivo:

P1/início	P2/meio	P3/fim	P1/início	P2/meio	P3/fim
abstrata	adiado	aborrescente	contada	comentários	causador
abstrato	adio	acabaria	converte	completa	causas
acaba	administração-de-empresas	acharmos	coração	complicada	causem
achem	administrar	acontecem	cujo	compromissos	cérebro
agradar	adolescência	acontece-se	cumprir	concluindo	chamar
ah	advocacia	aconteciam	dançando	conclusão	chega
á-há	a-há	acordo	danceteria	conhecer	cidade
alcançam	alcançar	adolescentes	dará	conseguem	classe-social
alegrias	alexandre-de-gusmão	afirma	daria	conservatório	clássico
ama	amanhã	aflige	data	continuação	cobradas
amem	amo	agimos	delicada	cooperar	comando
amigo	a-nível-de	agir	demais	copa-do-mundo	combatem
anda	ano-que-vem	agravando	demonstra	craques	começam
ano-novo	antes	agride	demonstre	cursando	comecem
aparecer	apenas	à-há	dentes	cursinho	comparado
assunto	aperfeiçoar	ajudasse	descontrações	curso	compõe
astral	apresentações	alcance	desejado	curso-profissionalizante	comumente
atitude	apresentam	alcansar	dessa	custa	concentradas
através	aquisitivo	algumas	desta	daí	condições*
bairro	área	aliás	de-v	dão	confraternização
bar	arrepender	amplo	dizem	decidido	confusão
basta	aspiro	andam	dizendo	decidir	conscientizar
beleza	ator	animal	emoções	decisiva	conscientize
bens	basear	antiga-capital-do-brasil	encarar	dedicar	conseguiria
brasileiras	boa	ao-invés-de	erguida	definida	contrabandiar
brincando	cabeças	aprendem	espaço	definiram-ão	contrabando
brincar	cálculos	aprendendo	espaços	de-novo	contribuiremos
busca	cargo	assaltados	esperando	der	controlada
buscar	carrefour	assaltos	esquecemos	derem	corações
calquer*	carreira	associar	esquecer	descoberta	cordão
caminhos	casar	atenção	estamos	descobrir	cordão-umbilical
campeonato	cedo	atinge	estavam	desempenho	corinthians
capacidade	chance	atrai	esteve	desenhar	correremos
carros	chances	atualidade	eternidade	dessas	corrupção
chamamos	chegando	auge	exatamente	destacando	costumes
chegamos	cia-suzano	aventura	existir	diante	cotidiano
cheia	ciência	bailes	famílias	diferente	cresce
colorida	ciências	barbaridades	fazemos	difíceis	crescimento
comigo	claro	bilhões	felisidade*	dificuldade	cria
competição	classe	bomba	fica	digamos	crianças
confiar	clube	brigar	fiz	digno	culpa
conhecê-la	clube-do-mundo	brutal	fortes	distância	cura
conjunto	clubes	cá	futebou	d-três	danos
conquistar	colegas	campo-de-batalha	fúteis	dúvidas	decladem*
consegui	colegial-técnico	cara	ganhar	eletrônica	demorar

Quadro 01 - apresentação de hápax do conjunto de textos/períodos da escola 01

Além dos *hápax*, foi estudado o vocabulário específico, pela técnica da estimação e expressão das densidades fatoriais. Da análise de frequência ao cálculo algébrico, foi possível observar e estudar tendências e convergências que enriqueceram a compreensão tanto da natureza dos elementos codificados quanto do cruzamento das variáveis contextuais. Da análise lexical e textual determinaram-se valores, hábitos e força de certas cargas emotivo-afetivas, e evidenciaram-se operações lógicas da superfície discursiva e redes produtoras de efeitos de sentidos. A quantificação do vocabulário tornou possível atingir três níveis fundamentais da linguagem para cada uma das seis escolas em seus respectivos períodos de ensino-aprendizagem:

- ▶ **vocabulário maciço ou preferencial** ($z \geq +1,96$);
- ▶ **vocabulário comum, base média positiva e negativa** ($-1,96 \leq z \leq +1,96$);
- ▶ **vocabulário único ou diferencial** ($z \leq -1,96$).

E pela técnica da discriminação, foram examinadas as variedades do vocabulário e seu valor, explicitada a configuração temática e demarcada a linha melódica discursiva e construção predicativa na constitutiva da arquitetura do tecido discursivo.

Desse estudo é possível, então, do reconhecimento e constituição de uma rede léxico-sintático-semântica estabelecer o estudo das relações e inter-relações das unidades lingüísticas, de modo que fenômenos lexicais, textuais e discursivos possam ser reconhecidos e descritos por uma análise suficientemente balizada.

6. Conclusão

Vive-se hoje num ambiente sistêmico, complexo e dinâmico. A identificação, a análise e a manipulação de informações e a recuperação de informações relevantes, bem como a descoberta de novas informações, não se constituem tarefa das mais simples ou fáceis. A identificação de redes de conhecimento, dimensão de análise transversal, torna possível a revelação de perfis importantes para redefinição de pré-requisitos e exigências para desenvolvimento de competências e, conseqüentemente, esforços num trabalho conjunto, importantes para o processo de interação, inovação, desenvolvimento e gerenciamento do potencial humano, em que novas formas e abordagens de socialização do conhecimento são indispensáveis às novas arquiteturas de aprendizagem, que situações de sala de aula podem dinamizar.

7. Referências bibliográficas

BARBOSA, M. A. - *Língua e discurso: contribuições aos estudos semântico-sintáticos*. SP: Ed.Plêiade, 1996.

BIDERMAN - M. T.- *Teoria Lingüística - (Teoria lexical e lingüística computacional)*, 2ªed., SP:Martins Fontes, 2001,Coleção leitura e crítica.

CAMLONG, A. & BERTRANSD, T. *STABLEX – Version PC 2004- Université de Toulouse II- The Art Editora Ltda, Pirus Tecnologia,2004*

CAMLONG, A. - *Méthode d'analyse lexicale, textuelle et discursive*. Paris: OPHRYS, 1996

COSERIU, E. - *Communication & Cognition*. V.25, nº 213, 1992.

GECKLER, H. *Semântica Estructural y Teoria del Campo Léxico*, Madrid, 1994.

HJELMESLEV, L. - *Essais Linguistiques*. Les Editions Minuit, nº 47 - Langue et Parole.

LYONS, J. - *Introdução àLingüística Teórica*, EDUSP/Cultrix Ed., 1979.