

## PARA O ESTUDO LÉXICO-SEMÂNTICO DE UMA LÍNGUA DE ESPECIALIDADE

Cilene da Cunha Pereira  
Maria Emília Barcellos da Silva  
UFRJ

I. Pouco se precisa argumentar em favor da organização de um banco de dados lexicais que seja realmente representativo do vocabulário português contemporâneo, variante brasileira.

Até onde se sabe, realizam-se atualmente recolhas restritas à língua escrita nos gêneros romance, conto, livros de ciência e tecnologia, textos dos meios de comunicação de massa. A heterogeneidade desse acervo justifica-se pela amplitude do leque de atividades presentificadas em tal amostragem, reiterando a lição aprendida com outras línguas de cultura: quanto maior for o volume dos dados recolhidos, tanto maior será a possibilidade de os resultados espelharem a estrutura e repartição do vocabulário idiomático em estudo.

Da tarefa – como se expôs – compor-se-ia um **corpus** avaliado em cinco milhões de palavras, o que repetiria o **quantum** do *American Heritage World Frequency Book*, por exemplo. Trabalhos dessa natureza soem gerar números que ultrapassam a casa dos seis zeros, haja vista o **corpus** em que se estrutura o *Trésor de la Langue Française* (séculos XIX e XX), que coletou nada menos de cem milhões de palavras.

Para se ter idéia do universo da linguagem humana – em constante expansão –, calculou-se que o curso universitário dardeja seu alunado com cerca de setecentos milhões de palavras/ano.

Em que pese ao reconhecido esforço que representa a organização de um banco de dados lexicais da língua escrita no Brasil – tal como se relatou –, sabe-se que esse trabalho não será bastante e suficiente para dar conta do manancial vocabular ativo de que se serve o falante da lusofonia americana: ficaria fora da amostragem a constrangedora fatia de analfabetos e semialfabetizados que, a seu modo e com o seu sacrifício, metaforizam os braços da construção nacional. Deixá-los mais uma vez à margem do que seria a fotografia da expressão lingüística brasileira é reeditar o hábito de maquiar os dados de brasilidade com aviamentos tais que o retratado, ao nele se mirar, mal se reconheceria ou reconhecer-se-ia mal.

É óbvio que se entende ter de se começar a tarefa por uma das pontas possíveis, e a língua escrita aí está comprovada, encadernada, pronta, à espera de um tratamento adequado e – de mais a mais – os estudos sobre a fala de comunidades periféricas são raros e, quando existem, não carregam por si sós prestígio a quem os realiza.

A equipe de professores da Faculdade de Letras/UFRJ que responde pelo Projeto do Atlas Etnolingüístico dos Pescadores do Estado do Rio de Janeiro – Projeto APERJ –, tendo dedicado ao estudo das variantes populares da língua portuguesa no Brasil boa parte do seu fazer acadêmico, propõe engajar-se no afã de erigir o **corpus** do idioma brasileiro, juntando à listagem eliciada da língua escrita, a contribuição dos iletrados.

II. A vertente LÉXICO-SEMÂNTICA do APERJ elabora, por localidade e faixa etária:

- a) **arquivo das unidades léxico-textuais e lexemáticas** eliciadas do discurso do pescador, dispostas em ordem alfabética, em ordem decrescente de frequência e distribuídas em classes de palavras — substantivos, adjetivos, verbos, advérbios **-mente**;
- b) **arquivo do vocabulário da pesca distribuído em áreas conceptuais, em campos semânticos e campos lexicais** acrescido do significado atribuído a cada item e das abonações do mesmo.

Com a pesquisa lexicológica pretende-se oferecer o vocabulário do falar dos pescadores norte-fluminenses desdobrado em índices e a norma lexicológica a eles subjacente, acompanhada de alguns quadros de distribuição de frequências, primeiro esboço do perfil numérico do vocabulário desse grupo.

A existência de índices de vocábulos de profissionais constitui uma fonte de renovação para glossários e dicionários tanto no que diz respeito à terminologia quanto às definições semânticas.

Pierre Guiraud, em *Problèmes et méthodes de statistique linguistique*,<sup>1</sup> considera na língua três níveis – fonético, lexical e sintático – e admite que essa tríade de signos se define **qualitativamente** por suas funções diacríticas e semânticas e **quantitativamente** pela frequência que possui valor funcional. Daí poder-se afirmar que um vocábulo se caracteriza no léxico de uma língua não apenas por um significante e um significado mas também por uma probabilidade de ocorrência, de que sua frequência no discurso é um reflexo.

Pela frequência com que emprega determinado signo, o indivíduo caracteriza não só seu discurso, mas também a linguagem do seu grupo e da localidade onde vive, num determinado momento.

Para o estudo **semântico** e **quantitativo** do vocabulário de uma língua de especialidade, é indispensável a elaboração de índices de vocábulos, que constituirão o "conjunto arquivo".

Neste trabalho, aplicar-se-á o método estatístico no campo do léxico em cada inquérito, no conjunto de inquéritos por localidade e no conjunto das localidades que constituirá o universo da pesquisa.

Indicam-se também, neste estudo, os critérios adotados para a segmentação do texto e para o levantamento e classificação dos vocábulos visando à elaboração dos seguintes índices:

Índice 1 – **o das unidades léxico-textuais**, com os vocábulos como aparecem no texto, ordenados alfabeticamente, seguido do código da classe léxica e da indicação da linha em que se localiza o vocábulo.

Índice 2 – **o das unidades lexemáticas com as formas básica ordenadas alfabeticamente**, acompanhadas do código da classe lexical e do número de ocorrências em cada inquérito seguidas das formas

flexionadas e das derivacionais de grau, como aparecem no texto com as frequências parciais e a distinção entre formas homônimas.

**Índice 3 – o das unidades lexicáticas básicas distribuídas em ordem de frequência decrescente.**

**Índice 4 – o das unidades lexicáticas organizadas em classes:** a dos substantivos, a dos adjetivos, a dos verbos e a dos advérbios em **-mente**, distribuídas em ordem decrescente de frequência.

**Índice 5 – o do vocabulário específico distribuído em áreas conceituais, campos semântico e lexical.**

**Índice 6 – o de vocábulos específicos da atividade, com as indicações da classe gramatical, do significado e com sua abonação.**

A análise de tais índices poderá fornecer informações numéricas como a extensão do texto, a extensão do vocabulário e a distribuição das unidades léxicas básicas por valores de frequência.

Esta pesquisa pretende estabelecer uma sistemática de análise que permita, através do estudo da frequência de utilização das palavras, configurar os principais universos de significação dos pescadores consideradas as variáveis faixa etária e localidade.

O levantamento e a classificação dos vocábulos impuseram o trabalho prévio de estabelecimento de uma norma lexicológica ou **norme de dépouillement** na terminologia de Charles Muller, que a define como

o conjunto das regras ou de convenções que, no levantamento quantitativo de um texto, garantissem a constância do tratamento e de seus resultados.<sup>2</sup>

O estudo lexical que ora se propõe realizar opera com levantamentos que tomam a palavra (lexema ou vocábulo) como sua unidade elementar, sem considerações relativas à fonologia ou à sintaxe.

É impossível prever-se o número de unidades que formam o léxico dos pescadores norte-fluminenses, pois este é um inventário que varia num só informante, de um para outro indivíduo, segundo as faixas etárias e consoante as localidades que circunstanciam tais e quais falantes.

Dessa forma, sendo o vocábulo uma concretização do léxico no discurso, os levantamentos que se possam fazer para a análise dos vocabulários devem seguir critérios pré-estabelecidos que garantam aquele mínimo de confiança nos resultados a que se chegar.

O presente estudo pressupõe a adoção de alguns conceitos básicos, expressos a seguir, nos quais se fundamentará este trabalho. Para tanto, utilizaram-se as definições de Charles Müller<sup>3</sup> no que respeita ao

**léxico** – conjunto de unidades virtuais que pertencem à língua;

**vocabulário** – conjunto de unidades virtuais da língua, atualizadas no discurso, daí poder-se falar de vocabulário de um inquérito ou do conjunto de inquéritos;

**lexema** – unidade do léxico, um conjunto unitário de formas que se atualizarão no discurso; entrada nos dicionários;

**vocábulo** – unidade do texto, unidade fonético-gráfica que em textos escritos costuma figurar entre dois espaços em branco.

O termo **palavra** serve para designar toda a ocorrência de um vocábulo qualquer. Serão considerados como **vocábulo** os termos simples, primitivos, compostos e derivados.

Esse levantamento levou à decomposição do texto em unidades léxico-textuais, reagrupando cada uma delas sob um vocábulo, separando as formas homógrafas pertencentes a vocábulos distintos e juntando as formas heterógrafas ou flexionadas pertencentes a um mesmo vocábulo. Assim os nomes (substantivo e adjetivo) – passíveis de flexão de gênero e número – e os verbos – passíveis de flexão de tempo-modó, número-pessoa – vêm, respectivamente, fixados no masculino singular ou no infinitivo com as formas flexionadas arroladas sob a forma básica; isso porque o padrão lexical, a forma não marcada, é o singular para o substantivo, o masculino para o adjetivo e o infinitivo para o verbo.

Dos inquéritos levantar-se-ão os morfemas lexicais – substantivo, adjetivo, verbo – por serem unidades cujo sentido relaciona o homem a seu universo biossocial. Quanto aos advérbios foram levantados apenas os em **-mente** por serem derivados de adjetivos e, dessa forma, estabelecer-se, na análise, o equilíbrio entre "os adjetivos de substantivos" e os "adjetivos de verbos". Deixou-se de lado, por ora, o levantamento dos morfemas gramaticais em função dos numerosos e delicados problemas que o envolvem e que, para um estudo de natureza léxico-semântica, não se fazem fundamentais.

Para a estruturação dos arquivos lexicológicos e léxico-semânticos, foi necessário definir os conceitos de lexia simples, composta e complexa, terreno movediço e controvertido.

Como lexia simples, enquadraram-se as unidades léxico-gráficas ocorrentes entre dois espaços em branco.

Ex.:           #peixe#  
                  #marola#  
                  #dourado#

Como lexia composta, operaram-se aquelas em que houvesse ligação mórfica entre os elementos componentes, com certo esvaziamento do significado dos constituintes em prol do novo todo arquitetado; apresentassem possibilidade de uso de artigo

diante do elemento determinante, impossibilidade de expansão e, por fim, dela fosse depreensível o caráter polissêmico do significado.

Ex.: #barco a motor#  
#cabeça d'água#  
#peixe-rei#

Os nomes próprios constituídos de vários elementos foram tratados com uma **lexia composta**. A razão disso se deve ao fato de que, sendo o nome próprio um rótulo, haverá tantas entidades dessa espécie, numa instância de discurso, quantos forem os objetos a listar.

Ex.: #São João da Barra#  
#São Fidélis#

Os numerais compostos foram tratados como uma lexia composta, pois separar os elementos de um número seria eliminar um referente.

Ex.: #quarenta oito#

Como **lexias complexas – sinapsias** –, definiram-se os compósitos que apresentassem ligação de natureza sintática entre os membros (determinante + determinado), plenitude tanto das formas léxicas quanto dos significados dos componentes, possibilidade de expansão e, por fim, fosse depreensível o caráter monossêmico do significado em que pese à polissemia dos seus constituintes.

Ex.: #andar à rola#  
#fio de náilon#  
#pescaria de tarrafa#

A definição desses critérios enfatiza o mecanismo de criação léxica marcadamente suscetível a novas formações; daí esse estudo ser extremamente sensível a processos de **derivação** e de **aglutinação** de elementos mórficos numa única forma-radical.

Nos inquéritos, marcam-se as unidades do discurso que participam das seguintes classes e subclasses gramaticais, definidas a partir do critério morfossemântico:

- 1 — substantivo comum
- 2 — substantivo próprio
- 3 — adjetivo
- 4 — verbo principal
- 5 — verbo auxiliar
- 6 — advérbio em **-mente** e suas formas apocopadas.

Por **substantivo**, entenderam-se os itens designativos de pessoa, animal e coisa que se submetessem à flexão de gênero e número e à derivação de grau aumentativo e diminutivo. Palavras inseridas no texto por derivação imprópria foram assumidas na sua situação textual.

o [vento] **Nordeste**  
subst.

o **normal**  
subst.

o **certo**  
subst.

**outrozinho**  
subst.

Por **adjetivo** entenderam-se os designativos de estado, qualidade e matéria e os participios de sonância adjetiva que se submetessem a flexões de gênero e número à derivação de grau superlativo:

molinete **improvisado** na lata  
adj.

a frente **fundada** na pedra  
adj.

faz-se uma esteira de bambu **unido**  
adj.

eles estão **perdidos**  
adj.

lavar bem **lavada**  
adj.

dois metros **quadrados**  
adj.

a rede é **preservada**  
adj.

vela **quadrada**  
adj.

o mestre é **respeitado** por todos.  
adj.

Dáí participarem dessa classe os adjetivos participiais de voz passiva.

Os numerais ordinais foram classificados como **adjetivos**, por cumprirem o estatuto morfossemântico dessa classe.

Foram considerados **verbos auxiliares ter e haver** acompanhados de participio passado ou de infinitivo e **estar** seguido de gerúndio ou de infinitivo na mesma oração.

Ex.: **tem** que preservar  
aux.  
**está** cuidando  
aux.



país se todas as suas formas de expressão forem havidas como igualmente importantes e merecedoras de atenção. Repetindo as palavras do emérito Professor Celso Cunha, "deve-se estar atento à voz do povo, porque os ventos das mudanças sopram de lá".<sup>4</sup>

### NOTAS

1. GUIRAUD, Pierre. *Problèmes et méthodes de la statistique linguistique*. Dordrecht, D. Reidel Publishing Company, 1959. p. 16-7.
2. MULLER, Charles. *Initiation à la statistique linguistique*. Paris, Larousse, 1968. p. 142.
3. opus cit p. 136
4. CUNHA, Celso. *Discurso de emergência proferido na Faculdade de Letras da Universidade Federal do Rio de Janeiro*, 1988.

\*\*\*