

LÍNGUA FALADA X LÍNGUA ESCRITA
ABORDAGEM DO LÉXICO

INTRODUÇÃO

Além dos aspectos sintáticos da construção da frase que incluem a hesitação, a correcção, a repetição, os anacolutos, as reiteraões e todos aqueles traços característicos da fala, por serem indispensáveis à interação face a face - a língua falada e a língua escrita se diferenciam em relação ao léxico e, não só na chamada densidade lexical, análise quantitativa dos termos, como também no tipo de léxico empregado.

Enquanto a escrita utiliza um léxico especializado, detalhado e variado, a fala tem um léxico generalizante com uma enorme quantidade de palavras-ônibus, e de outras também que só são capazes de nomear os seus referentes na presença dos mesmos, pois são inapropriadas. Partindo deste pressupostos, elaboramos a presente pesquisa, utilizando como mini-corpus a mesma notícia recebida por três canais diferentes.

- a notícia relatada em conversa por um falante comum (mod. falada/falada)
- a notícia transmitida pelo telejornal (modo falado/escrita)
- a notícia veiculada pelo jornal (mod. escrita).

Escolhemos a mesma notícia para que o referente fosse sempre o mesmo. A diferença reside no tipo de canal e

modalidade de língua utilizada. Com isto, observamos algumas particularidades a respeito do uso dos itens lexicais, que não nos permitem contudo ampliar muito sua projeção, dada a exigüidade do corpus.

Para a fundamentação teórica da pesquisa escolhemos como base o capítulo 5 do livro Spoken and Written Language de Halliday - "Written language lexical density" com algumas informações de Sttubs em "Lexical Density - a Technique and some findings" além de outros sobre o léxico, nos quais baseamos os modelos de análise e que constam na bibliografia.

1. Pressupostos Teóricos

Os objectivos da língua escrita e da língua falada são diferentes mas nenhum tem valor superior ao outro. Halliday, ao analisar o léxico da língua escrita chama a atenção para a densidade lexical da mesma.

As palavras de uma língua podem ser classificadas em gramaticais (palavras funcionais) e lexicais (palavras de conteúdo). As gramaticais são aquelas que funcionam no sistema fechado da língua: determinantes, pronomes, preposições, conjunções, algumas classes de advérbios e as terminações verbais. As palavras são aquelas que nomeiam itens do mundo extra-lingüístico: ações, qualidades, objetos, pessoas, processos, sentimentos etc. A principal diferença entre o léxico da língua escrita e o da falada é a maior quantidade de itens lexicais na primeira. A causa não é apenas o assunto: é a densidade com a qual a informação é apresentada. Enquanto a língua escrita é densa, a língua falada é esparsa.

Grande número de fatores contribuem para esta densidade.

É um fenômeno complexo, produto de um número de variáveis. Dizer que a língua escrita é a mais densa é sugerir que ela do ponto de vista lexical é mais complexa, embora a falada seja mais intrincada, isto é, a informação fica emaranhada nas formas do falar - Cada modalidade de língua é complexa à sua maneira. Se a língua falada é mais intrincada, em compensação, a espécie de complexidade típica da escrita é a densidade lexical.

2. Classe de Palavras

Itens lexicais são palavras de conteúdo. A expressão item é mais adequada porque freqüentemente consiste em mais de um termo: pré-escola, vice-presidente, viúva-negra, homem-rã. Estes itens funcionam nos conjuntos lexicais e não em sistemas gramaticais; isto é, entram em contrastes abertos e não fechados. Um item gramatical entra num sistema fechado. Por exemplo ele contrasta como eu, ela, tu, você, nós, vós, vocês - e nada mais. Porém janela, item lexical, pode contrastar com porta, portão, entrada, etc. O conjunto de seus contrastes nunca fica fechado.

Porém, pode-se perceber que há um "continuum" do léxico à gramática. Advérbios modais, por exemplo, ficam no limite entre ambos, como também certos pronomes.

Favorecendo evidenciar o "continuum" entre os itens gramaticais e os lexicais, alguns (bons) dicionários colocam itens gramaticais como eu, mim, ligando-os a itens lexicais como ego, personalidade, espírito.

3. Usos dos itens lexicais

O item lexical caracteriza-se no conjunto do léxico de uma língua, não apenas por um significante e um significado, mas também por uma probabilidade de ocorrência, reflexo de sua freqüência no discurso. O léxico será o resultado da articulação complexa de três estruturas: morfo-sintática, semântica e quantitativa. O conhecimento de todas essas estruturas será determinante para conhecer a natureza do léxico e precisar-lhe a organização e o funcionamento, levando em conta o interrelacionamento.

O reconhecimento da importância da estrutura quantitativa do vocabulário de um texto para a apreensão de unidades estilísticas, por sua condição de mapa das zonas privilegiadas do sistema expressivo como um todo, corresponde a um saber que vem de longe e é compartilhado por muitos.

O conjunto formado pelos cinquenta substantivos, adjetivos e verbos mais freqüentes de um texto, conhecidos em Estatística Léxica como palavra-tema ou palavra-chave, pode conter os itens lexicais que na norma da língua apresentam probabilidade de baixa ocorrência e que podem ser indício de fato estilístico.

Mas, não somente as palavras - chaves são reveladoras: há ausências lexicais que também o são. Os lexemas de baixa freqüência modificam a qualidade do texto nos pontos onde ocorrem.

4. Densidade lexical e freqüência

Se, numa frase como a transcrita "Os médicos do Hospital Central oferecerão um jantar às autoridades presentes no res-

taurante local, às vinte horas do dia 14 de Dezembro de 1990", são observados 12 itens lexicais e 8 gramaticais, isto dá a proposição de 12 para 20, o que nos mostra que a densidade lexical é de 60%. Em geral, quanto mais tipicamente escrita é a linguagem, mais alta é a proporção de palavras lexicais no total de termos do texto.

Além do aspecto da densidade, surge o da probabilidade. Os itens gramaticais são de alta freqüência no texto, enquanto que os lexicais são repetidos muito menos freqüentemente, o que aliás é bastante previsível.

A freqüência e o uso de um item lexical são elementos significantes na situação. Toda língua tem um número de lexemas de alta freqüência que permite descrever um sem número de objetos. São as palavras - ônibus: gente, pessoa, coisa, modo, caminho, fazer, dar, meio, ter, ser, ir, bom, muitos, pouco. Alguns deles ficam no limite entre a classe gramatical e a lexical e contribuem pouco para a densidade lexical.

Ao contrário, um lexema de baixa freqüência na língua, contribui bastante para o efeito de densidade. Comparemos:

Ex. 1. O mecanismo da determinação do sexo varia em diferentes organismos.

Ex. 2. O modo como o sexo é determinado difere com criaturas diferentes.

Ex. 3. Criaturas diferentes tem seu sexo determinado de modos diferentes.

A proporção de itens lexicais é quase a mesma nas três, mas as duas últimas parecem menos densas porque incluem itens muito freqüentes, além de repetir termos com o mesmo radical: diferir/diferente. A repetição reduz o efeito de densidade.

Assim para melhor perceber a diferença entre o léxico da língua escrita e o da língua falada, devemos incluir o critério de frequência, passando a reconhecer três espécies de categorias em vez de duas:

- 1 - itens gramaticais;
- 2 - itens lexicais;
- 3 - itens lexicais de baixa frequência.

Afim de explorar o potencial da linguagem, mapeando a estrutura da transitividade (processos, participantes e circunstâncias) e a estrutura da mensagem (tema/rema - Dado novo e todas as suas possíveis combinações) os falantes têm que estar preparados para expressar-se em formas nominalizadas. Para Halliday é desta forma que a estrutura do mundo moderno e a estrutura da linguagem combinam-se para fazer da linguagem escrita, o que ela é: uma linguagem com alta densidade lexical, medida pelo número e peso informacional de itens lexicais por oração e uma forte tendência a codificar seus conteúdos, em formas nominais: em substantivos - núcleos, em substantivos e adjetivos, em grupos nominais e em frases nominais. Para ele, seriam estas estruturas nominais que dariam as orações enorme elasticidade.

A questão da frequência dos itens lexicais (alta frequência, baixa frequência) ajuda a construir um sistema de análise mais refinada. Contudo, a visão continua sendo unilateral. O léxico da língua falada continuará sendo caracterizado por traços negativos, isto é, o baixo grau de densidade de informação.

Reinterpretando esta noção, traços positivos poderão ser notados na língua falada.

Examinemos como a informação é veiculada: uma palavra de baixa probabilidade carrega muito mais informação. Mas as palavras não vêm só: elas nos chegam através de estruturas que determinam a densidade informacional de um texto. A mais relevante é a oração, unidade gramatical na qual construtos semânticos de diferentes espécies unem-se e integram-se como um todo. Ela é a ligação entre a semântica e a gramática. Assim a melhor forma de observar a densidade lexical, será medi-la também pelo número de orações.

O conceito de orações é fundamental para o estudo da língua escrita e falada, sendo bastante controverso. A oração é o "locus" das escolhas na transividade, modo e tema. Assim a quantidade de informação lexical incorporada numa oração pode ser extremamente variada. Pode não haver nenhuma, como em: Não! É mesmo. Também pode haver uma grande quantidade, com textos sobrecarregados de itens lexicais. A grande flexibilidade da oração possibilita a evolução das formas escritas do discurso. As orações são constituídas de sintagmas nominais e verbais. Quase todo conteúdo lexicalizado está contido nos grupos nominais formados por substantivos e seus modificadores sejam adjetivos ou outros substantivos. Os grupos nominais são assim constituídos: pré-modificadores + núcleo ou substantivo principal + pós-modificadores. O núcleo é a classe de fenômeno a ser referida. Os classificadores ou epítetos também têm informações léxicas: subclasses ou qualidades de várias espécies incluindo as que expressam a atitude do falante.

Os grupos verbais têm apenas um elemento lexical, o próprio verbo. Alguma informação léxica pode ser expressa em grupos adverbiais, porém são muito limitadas.

Muitas coisas só podem ser ditas em construções nominal, especialmente no registro das linguagens da ciência e da tecnologia onde as idéias proliferam rapidamente e as palavras têm de acompanhá-las.

Entre os novos usos da língua, "finalmente" pode ser usado como um substantivo e alerta como um adjetivo. Além de sua organização como representação de um processo - transitividade - e como suporte da função da linguagem - modo - toda oração é também estruturada como uma mensagem. Consiste de duas partes: um tema, ponto de partida da mensagem e outro elemento que constitui o corpo de mensagem conhecido como rema. Na maioria das línguas, o tema vem sempre primeiro. Na língua falada é freqüentemente um pronome: 1ª ou 2ª pessoa. Na escrita é freqüentemente o pronome de 3ª pessoa, porém mais comumente um elemento nominal. Nunca é, contudo, um grupo verbal. Esta é uma das razões pela qual a informação lexical vem sempre veiculada por um substantivo.

Uma outra estrutura oracional é o das frases clivadas. As frases clivadas ou pseudo-clivadas ajudam a chamar a atenção para o tema, exigindo a nominalização da informação.

5. Característica dos corpus lexicais falado/escrito

Só os lexemas carregam o acento de força entonacional. Em português, podemos estabelecer a diferença em pára (verbo) e para (prep.) que pode ser contraído. Alguns ainda recebem um acento chamado subtônico: cafezinho.

Ure (1971-citado por Sttubs) estudou 34 textos falados e 30 escritos ambos com um total de 21.000 termos. Encontrou uma forte tendência entre os textos falados de ter uma densidade

lexical abaixo de 40% variando de 24 a 43%. os textos escritos têm tendência a ter densidade lexical acima de 40%, variando entre 36 a 57%.

As razões destes números são evidentes: em média o texto escrito é mais curto, menos redundante e com raras repetições, se comparado ao texto falado. É permanente, altamente editado, reescrito e cheio de recursos, não sendo espontâneo e não-planejado como a maior parte das conversações. Um texto escrito é relativamente independente de contexto situacional, enquanto que o texto falado pode basear-se em grande parte do contexto físico imediato. O nível de informação é bem mais alto num texto escrito; desde que ele é permanente os leitores podem reler as passagens obscuras.

A maioria dos textos falados, por outro lado, são compreendidos como e quando são produzidos e por isso devem ser mais previsíveis. Textos escritos são menos previsíveis e os falados, mais previsíveis. Também os lexemas são menos previsíveis pois há uma quantidade infinita deles, como vimos.

Enfatizamos também que os itens gramaticais existem em pequeno número e por isso são mais previsíveis. Assim enquanto os textos falados têm uma mais alta proporção de itens gramaticais previsíveis, os textos escritos têm uma mais alta proporção de itens lexicais não-previsíveis.

Há palavras que são mais especificadas do que outras: cão policial é mais específico que cão; mamífero que vertebrado e este mais do que animal. Trabalhador é termo de sentido geral muito amplo; é uma classe. Operário é mais restrito; é o género. Metalúrgico é a espécie e soldador a variedade. Ao descrever uma cena na TV, pode-se dizer simplesmente o trabalha-

dor; mas, na notícia escrita será usado o termo mais preciso pois, não existe o contexto para ligá-la ao referente. Generalização e especificação têm no entanto sentidos relativos. O grau de generalização ou de abstração de um enunciado depende do seu contexto.

Generalizações na escrita são menos usadas pois tornam confusa e vaga a informação, enquanto que podem ser usadas na fala sendo as especificações indicadas pela própria situação de presença. A escrita para conseguir narrar/descrever lança mão de artifícios incluindo lexemas generalizados, mas, indo também ao fim da escala, apresentando termos bastantes específicos.

6. Análise do mini-corpus

O corpus para a verificação dos pressupostos da diferença de usos na vocabulário da fala e escrita tomou como ponto de partida a linguagem dos meios de comunicação de massa.

Analisamos desta forma três modalidades da língua portuguesa:

- a modalidade escrita representada pela notícia de jornal
- a modalidade falada/escrita representada pelo noticiário de TV
- a modalidade falada (espontânea) representada por uma narração vicária.

Um falante comum ouve a notícia e conta o que ouviu. Sendo assim não é uma experiência sua e sim algo que ele ouviu e re-conta.

Gravamos no noite de 24/10 uma notícia escolhida sobre o envolvimento da filha do Governador Leonel Brizola com o

tráfico de drogas no Jornal Nacional da Globo. No dia seguinte (25/10) identificamos a mesma notícia no Jornal do Comércio - Recife.

A seguir, pedimos a um informante, escolhido aleatoriamente, que nos contasse o que tinha lido/ouvido sobre o fato.

Captamos assim 3 modalidades, lembrando que a linguagem de TV é considerada por Blanche Benveniste como língua falada. Para ela o que importa é o canal.

Seguem-se as análises:

Modalidade 1: O escrito

A. Densidade lexical

Os itens lexicais do texto foram em número de 269 enquanto que os itens gramaticais foram número de 179. A proporção foi, pois, de 269 para 448, o que mostra a densidade lexical na ordem de 60%, coincidente com o número observado nas pesquisas de Halliday e Sttubs.

B. Freqüência.

As palavras de alta freqüência surgem obviamente no texto jornalístico: entrar-sair-dizer-pessoas. Porém, o texto analisado, embora dirigido ao grande público, está pleno de itens lexicais de baixa freqüência: indiciada - implicada - apontou - embalar - fixar - constar - aprender.

Observa-se que as expressões são iniciadas por termos denotadores de generalidades sendo concluídas, contudo, por termos denotadores de grande especificidade: Governador eleito Leonel Brizola; Polícia Federal; Tráfico internacional de cocaína; Traficante holandês Erik Jurrian Peter Terlien;

Mandado de busca e apreensão; Sacos plásticos presos ao corpo, na altura da cintura com esparadrapo.

A droga é especificada como cocaína, acrescida de dois litros de éter, um de ácido sulfúrico acompanhada de instrumentos: balança de precisão.

A tendência à nominalização com grande número de grupos nominais é freqüente, como observa Halliday.

A presença de lexemas do campo semântico policia da triologia droga/contravenção/punição é o elemento de coesão.

Modalidade 2: O falado/escrito - Análise do Tele-Jornal

A. Densidade lexical.

A mesma notícia, veiculada pelo Jornal Nacional, constou de três parágrafos, onde as frases curtas na ordem direta simplificam a informação, contextualizada nas imagens.

Os itens lexicais foram em número de 106, enquanto que os gramaticais foram em número de 79 - o que dá a proporção de 269 para 448, sendo a média de 50% abaixo portanto da amostragem anterior.

B. Freqüência.

O texto é composto por palavras de alta freqüência na língua geral, com, igualmente, alto teor de generalização, isto é, freqüentemente ajustadas ao uso em contexto diferente: apresentar, dizer (muito repetido), receber, usar, morar.

As declarações nunca são apresentadas com um verbo mais característico e forte como na escrita: denunciar, indiciar, apontar. Apesar de surgirem grupos nominais estes não apresentam o alto grau de especificidade da escrita. O traficante tem

apenas dois nomes Erik Terlien (a princípio) e depois apenas um: Erik. A droga é nomeada de forma genérica ou apenas como cocaína. A embalagem é feita por "saco plástico e esparadrapo".

A tendência à nominalização é menor, sendo menos freqüentes os grupos nominais.

Como as palavras usadas são de alta freqüência e de uso abrangente, é menos marcante o campo semântico policial (droga/contravenção/repressão) apagado pelo uso de itens lexicais generalizantes.

C. Modalidade 3: O falado/falado

A notícia recebida pelo tele-jornal foi repetida por um informante que acabara de ouvi-la, a nosso pedido. A transcrição foi feita sem processo de edição do texto.

A. Densidade lexical:

Os itens lexicais foram em número de 102, enquanto que os gramaticais foram 149 o que faz com que a densidade lexical seja na proporção de 102 para 351. Neste texto, os itens gramaticais foram em maior número e o percentual da densidade lexical, na ordem de 29% bem abaixo dos casos anteriores.

B. Freqüência

O texto é composto por itens lexicais da alta freqüência e entre eles, surgem alguns considerados da mais alta freqüência na Língua Portuguesa como coisa, pessoa.

Há uma repetição de itens lexicais, receber, preso, como também de parassinônimos ocorrendo como auto-correção: maconha

— droga; por conta — responsável.

Os verbos mais empregados são do tipo palavra - ônibus de sentido geral: dizer, pegar, estar.

O texto apresenta uma alta freqüência de termos generalizantes e uma baixa freqüência de grupos nominais.

7. CONCLUSÃO

Algumas dificuldades foram encontradas na classificação dos itens, como por exemplo o caso dos verbos de ligações e especialmente do verbo estar. Mas, a margem de erro que o levantamento destes dados pode provocar, é insignificante.

Não podemos contudo projetar as conclusões dos dados numéricos como sendo dados universais a serem usados em Língua Portuguesa. Além da exigüidade do corpus não aconselhar essa projeção, funciona como variante não só a modalidade, como também o tipo de discurso veiculado por essa modalidade.

Concluimos, pois, que a densidade lexical é, em si mesmo, uma simples estatística.

Devem ser observados também outros dados que incluem o informante, o tema, a forma de veiculá-lo para que se tenha uma idéia completa da diferença de usos lexicais na língua oral e na língua escrita.

8. BIBLIOGRAFIA

1. BIDERMAN, M. T. Camargo - Teoria Lingüística. (lingüística quantitativa e computacional). Livros Técnicos e Científicos, Ed. São Paulo - 1978.
2. GARCIA - Othon. Comunicação em prosa Moderna - Ed. Fundação Getúlio Vargas - 8ª ed. R. de Janeiro - 1980.

3. HALIDAY, M. A. K. - Spoken an Written Language - Oxford University Press - Oxford. 1989.
- 4 . MARCUSCHI - Luis Antonio. A ação de verbos introdutores de opinião - Conferência pronunciada no 6º encontro de Línguaística da UPC - (Mímeo) - Rio de Janeiro. 12/11/81.
5. PRETI, Dino - A Linguagem da Televisão (mímeo) São Paulo. 1980.
6. STUBBS, Michael - Lexical density: A Technique and some findings - in Talking about Text - University of Birmighan Press - Birmighan - 1986.
7. VANOY, Francis - Usos da Linguagem: Problemas e técnicas na produção oral e escrita. Martins Fontes. São paulo - 1981.