

Linguística de Corpus e Interaccionismo Sociodiscursivo: interface e/ou complementaridade

Rosalice Pinto & Sebastião Silva Filho

Abstract: This article intends to demonstrate the way in which the Corpus Linguistics can be an important methodological tool for the Socio-discursive Interactionism to help describe *text unities* that occur in texts that circulate in society. Our concern is to show that: (1) the lexical statistic may be influenced by the social activity where the text is integrated; (2) some linguistic strategies chosen by the person who answers a *questionnaire* are imposed by the *text genre* and the *social activity*. In order to prove our hypothesis we analyze the answers of some *questionnaires* that were applied in two different classes and institutions, both in the first year of graduation courses, in December 2009.

1 – Introdução

Este trabalho procura descrever de que forma a unidade textual¹ *acção empreendedora* pode vir a ser linguisticamente descrita com o contributo de aspectos metodológicos da Linguística de *Corpus* conjugados com questões teórico-metodológicas do Interaccionismo Sociodiscursivo.

No intuito de atingir tal objectivo, far-se-á um levantamento de suas ocorrências em *inquéritos por questionário* realizados junto a alunos do 1º ano de Licenciatura, em dois cursos distintos (Direito e Ciências da Linguagem) de duas universidades em Lisboa (uma privada e uma pública).

¹ O termo *unidade textual* é definido, neste trabalho, como qualquer elemento verbal ou não-verbal que entra na composição de um texto, com *níveis de complexidade* (cf. Rastier, 2005: 36) distintos em função do interesse de análise do investigador.

De forma a clarificar questões teóricas relevantes para as análises que serão apresentadas, elucidaremos, primeiramente, alguns aspectos relevantes da Linguística de *Corpus* que nortearão alguns estudos preliminares; em seguida, passaremos a abordar conceitos importantes do Interaccionismo Sociodiscursivo, no intuito de complementar algumas análises. Por fim, mostraremos algumas reflexões finais que suscitaram esta contribuição.

2 – Questões teóricas relevantes

O estatuto da linguística de *corpus*, ou como metodologia, ou como abordagem teórica é um ponto de divergência entre os especialistas dessa área. (McEnery e Wilson (1996), Biber, D. et al. (1998).

Independentemente do estatuto adoptado, a Linguística de *Corpus*

comporta-se, fundamentalmente, como *um observatório da língua*, isto é, um local onde podemos observar a evolução dessa entidade, tendo em conta que a modernidade da vida quotidiana está em constante evolução. Como um *observatório da língua*, o *corpus* pode vir a ser usado para diversos estudos, dependendo do objectivo de seu utilizador.

O *corpus*, como um conjunto de dados, é um local onde podemos observar e descrever a estrutura e as situações de uso da língua e nos permite analisar a variação e a mudança que ocorre nessa entidade e ainda permite a observação das análises estatísticas e morfossintáticas.

Falar sobre a linguística de *corpus*, implica ressaltar a importância das ferramentas automáticas e semi-automáticas concebidas em função das análises linguísticas. É a partir da criação de distintas ferramentas automáticas e/ou semi-automáticas que pode haver a extracção de unidades textuais. Vale inclusive salientar que tais ferramentas podem ser consideradas relevantes para a criação de glossários e de dicionários, para a construção de base de dados e a utilização de sistemas de hipertextos.

A análise das informações de um *corpus* feita a partir de ferramentas computacionais permite atingir uma maior eficácia no resultado, tendo em conta a rapidez e a eficiência do computador ao analisar um grande número de informações obtidas através de uma grande quantidade de textos que constitui um determinado corpus.

Lino (2000: 26) observa, inclusive, que "a investigação em linguística transformou-se, beneficiando das contribuições da informática"; por essa razão, os variados pontos de vista evidenciam o desenvolvimento do estudo da língua geral.

A autora refere-se, ainda, ao facto da língua ter deixado de ser apenas um fenómeno cultural e passou a ser observada como um "poderoso agente económico com um lugar importante na actividade económica e industrial à medida que a informática e a telemática se instalam na comunidade internacional" – Lino (2000:27).

Dessa forma, pode-se afirmar que a linguística de *corpus* é fundamentada em distintos processos para analisar os dados encontrados e chegar ao resultado final: a descrição das unidades textuais.

Contudo, lembremos que, nesta contribuição, perspectivada no estudo

de textos empíricos inseridos em géneros textuais e em actividades sociais, como evidencia o Interaccionismo Sociodiscursivo, consideramos a *linguística de corpus* como uma metodologia. E adoptamos a definição de Rastier, para o entendimento da noção de *corpus*, uma vez que este autor compartilha com o ISD a relevância do texto como objecto de estudo da linguística e pontua a relevância de aspectos genéricos nas descrições de linguístico-textuais. Para este autor o *corpus* corresponde a “un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés: (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d’une gamme d’applications.” (Rastier, 2005: 32).

Salientemos que nem todos os dados recolhidos constituem um *corpus*, uma vez que é necessário que os dados sejam qualificados com os objectivos de serem reconhecidos como dados. "La linguistique de corpus ne pratique pas d’analyse automatique des données, dans la mesure où elles doivent d’abord être qualifiées comme données, puis interprétées avant et

après traitement: les données sont ce qu’on se donne.” (Rastier, 2005: 42).

De acordo com Rastier (2005: 36), é através da *linguística de corpus* que é estabelecida a relação entre o *global* (*corpus*, texto) e o *local* (signo), tanto no plano do conteúdo quanto da expressão, sendo que estes planos podem vir a ser analisados em separado ou em conjunto, constituindo o que o autor denomina *semiosis textual*.

Ainda, esta *semiosis textual* pode vir a ser analisada, a partir do estudo do *corpus*, em *patamares de complexidade* diversos² em função dos objectivos da investigação. Nesta contribuição em especial, centrar-nos-emos em estabelecer correlações entre dois patamares: entre o *léxico* e *actividade social* e entre *elementos linguístico-textuais* que funcionam como *marcadores de género* e o próprio *género textual* em análise.

3 – Algumas opções metodológicas

Para Rastier, um *corpus bom* é aquele constituído por textos pertencentes ao mesmo género. No caso específico desta contribuição foram seleccionados

² Para mais detalhes sobre a questão, ver Rastier (2005: 36).

trechos de *inquéritos por questionário*³, efectuados no 1º ano de Licenciatura de cursos distintos em instituições diferentes, como já salientámos. Estes inquéritos tinham como um dos objectivos principais obter dos inquiridos informações acerca de suas representações relativas a uma *acção empreendedora*⁴. E, a partir dessa selecção prévia, vários níveis de análise podem ser distinguidos, baseando-se também em Rastier (2005: 33).

Num primeiro nível, temos o *corpus* de referência, propriamente dito, constituído por inquéritos efectuados em turmas de horário pós-laboral em licenciaturas e universidades distintas. Esse *corpus* foi transcrito e os dados

³ O instrumento de análise utilizado integra-se numa investigação em andamento financiada pela Fundação para a Ciência e Tecnologia de Portugal que visa a descrição do *agir empreendedor em instituições portuguesas a partir de uma abordagem textual*. Tal pesquisa integra-se no projecto PRETEXTO do grupo Gramática & Texto.

⁴ O termo *acção* é definido diferentemente em função das várias perspectivas teóricas. Para detalhes ver, Filliettaz (2004: 9-23) na apresentação do *Cahiers de linguistique générale* 26. Inclusive, todo este número é dedicado a esta questão. Neste trabalho, consideramos que a *acção* diz respeito a qualquer forma de intervenção humana e também à sua interpretação. Já a acepção de empreendedorismo corresponde a que se encontra socialmente instaurada, isto é, está relacionada à *criação/expansão de ideias inovadoras a partir de oportunidades identificadas em determinada actividade*.

nele contidos foram tratados a partir de uma ferramenta computacional: o CONCAPP⁵.

Num segundo nível, destacamos o *corpus* de estudo. Este corresponde a passagens textuais contendo elementos relacionados com o tópico em estudo: a definição de acção empreendedora. Pontuamos abaixo os conjuntos de questões/ respostas analisadas:

Resposta à questão b do inquérito	Quando pensa numa acção empreendedora, que ideias lhe vêm à mente?
Resposta à questão d do inquérito	Proponha a definição que lhe ocorrer, quando pensa na expressão acção empreendedora.

Este *corpus* de estudo foi estudado em dois níveis, seguindo o *conceito de patamares de complexidade* em que um texto pode vir a ser analisado.

Num primeiro nível, estabelece-se a relação entre lexemas e actividades, analisando-se:

Para o primeiro nível:

⁵ O *Concapp Concordancer* foi a ferramenta automática que nos permitiu utilizar informações. Com o auxílio dessa ferramenta, elaboramos os dicionários de frequência. A partir desses dicionários seleccionamos algumas unidades textuais e construímos dois gráficos, um para cada área de especialidade, com o objectivo de demonstrar a ocorrência dessas unidades nos *corpora*.

(1) Elementos textuais relacionados com o conceito de *acção empreendedora* que circula socialmente a partir do levantamento do campo lexical⁶.

(2) Levantamento das ocorrências mais frequentes de alguns lexemas.

E, num segundo nível, procura-se estabelecer a inter-relação entre secções do texto (questões agentivas/sistema verbal// género textual/actividade), realizando:

(3) Levantamento dos marcadores de retoma /de actualização do conteúdo proposicional/relações predicativas indirectas.

(4) Levantamento de marcadores de agentividade nas passagens seleccionadas.

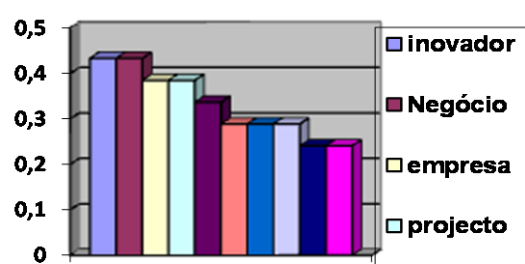
4 – Análises efectuadas

Apresentamos abaixo os resultados apontados para os dois níveis de análise, sendo que foram analisados 30 inquéritos relativos à Licenciatura em Direito e 22, na de Ciências da Linguagem.

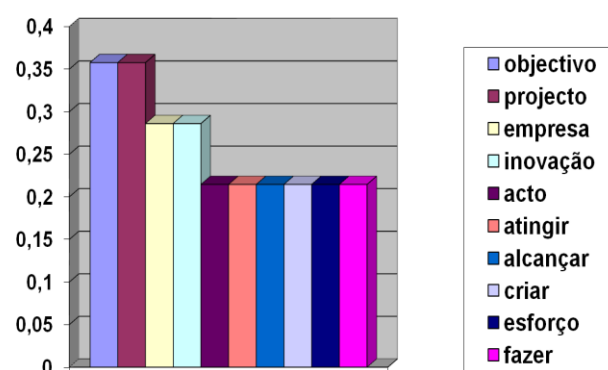
⁶ Campo lexical segundo Koch (1999: 156): “conjunto de palavras empregadas para designar, qualificar, caracterizar, significar uma noção, uma actividade, uma técnica, uma pessoa”.

4.1. 1º nível de análise

Pela estatística lexical realizada, conforme o gráfico 1, observa-se que, na turma de Direito há uma maior incidência dos lexemas *inovador*, *inovação* e *negócio*. Já na de Ciências da Linguagem (como vemos no gráfico 2) há uma maior incidência dos lexemas *objectivo*, *projecto* e *empresa*.



Direito
Gráfico 1



Ciências da Linguagem
Gráfico 2

4.2. 2º nível de análise

Neste segundo nível, procura-se observar de que forma algumas características linguístico-textuais do género textual inquérito em sala de aula são influenciadas por questões de ordem genérica.

Como afirma Bronckart (1999: 103), “a apropriação dos gêneros é um mecanismo fundamental de socialização, de inserção prática nas atividades comunicativas humanas”. E ainda todo gênero só pode vir a ser materializado a partir de um texto que o realiza. Assim, um agente produtor, ao produzir determinado texto, recorre a modelos já disponíveis no arquitexto (processo de adoção), e os adapta em função das representações acerca do contexto de produção e do conhecimento que tem relativo aos gêneros (processo de adaptação). Há de ressaltar que o gênero apresentará, além de aspectos previsíveis para o próprio gênero, propriedades singulares e um estilo individualizado. Além disso, segundo Maingueneau (1998: 51-54), os gêneros devem ser analisados a partir de alguns elementos que os constituem. Por exemplo: lugar e momento de realização, legitimidade e estatuto dos parceiros, finalidade, suporte material e plano de texto.

4.2.1.a. Questões contextuais relevantes

No caso do gênero textual em apreço, o inquérito de investigação aplicado sob forma de questionário, podemos verificar alguns desses elementos

pontuados por Maingueneau que, como veremos, vão intervir na materialização linguístico-textual dos textos analisados.

Lugar de realização – sala de aula
Legitimidade e estatuto dos parceiros – investigador/aluno (no caso da turma de Ciências Sociais) // professor com papel de investigador/aluno (no caso da turma de Direito)

Apontadas algumas características genéricas relevantes, passaremos a identificar alguns elementos linguístico-textuais relevantes observados nos textos em análise.

4.2.1.b. Questões linguístico-textuais relevantes

Nos trechos em análise, pontuamos abaixo algumas recorrências linguísticas importantes que podem nos fornecer pistas para caracterizar este gênero, neste lugar de circulação. De forma a identificar onde tais ocorrências aconteceram, colocamos as abreviaturas (D) para a Licenciatura em Direito e (CL) para a de Ciências da Linguagem.

- Marcadores de retoma e atualização do conteúdo proposicional (1) *Quando penso numa acção empreendedora*, não é só uma ideia

que surge na minha mente mas sim um desejo (...) (D)

(2) *Quando penso numa acção empreendedora*, vem-me à mente a ideia de ganhar dinheiro e enriquecer, mesmo não sendo o meu projecto. (CL)

(3) *Quando penso numa acção empreendedora*, vem-me à mente algo relacionado com a prática (...) (D)

- Marcas de agentividade (uso constante da 1ª pessoa do singular – implicação do actante na própria definição do termo acção empreendedora) + verbo que denota o processo de pensamento/ ou aspectos cognitivos.

(4) *Penso* em algumas acções conducentes à concretização de algum projecto, construir algo de útil à comunidade (CL)

(5) *Considero* que a minha frequência universitária é a minha acção empreendedora, esperando dela o fruto da capacidade de poder exercer a advocacia (D).

- Relações predicativas directas (verbos denotando actos/gestos)

(6) *Investir* em algo que me proporcione sustentabilidade financeira como *abrir* uma empresa de advogados (D)

(7) *Contribuir* para o desenvolvimento social e humano (D.)

(8) As ideias que me ocorrem é *abrir e criar* uma empresa minha (CL)

5 – Algumas reflexões

Ao compararmos as estatísticas lexicais, verificamos que na turma de Direito há uma incidência maior de lexemas relacionados ao campo lexical “inovação”. Por outro lado, noutra turma predomina o campo lexical “projecto” e “empresa”. Podemos supor que essas representações colectivas são constringidas pela própria actividade a que estes alunos estão insertos e também ao fato de pertencerem a instituições com orientações distintas (a primeira, privada; a segunda, pública).

De uma forma consensual, os lexemas relacionados à acção empreendedora não sofrem muita variação de uma turma para a outra. Contudo, observam-se ainda uma pequena diferença entre as frequências de lexemas relacionados com a acção empreendedora nas duas turmas (maior homogeneização do colectivo em CL). Como afirma Bronckart (2006), “Le psy est d’abord dans le social (representations collectives). L’individu a des représentations qui

l'instaurent comme personn.” E esta coloração individual só poderá ser atestada a partir do levantamento de alguns aspectos linguístico-textuais.

Como constatámos, a estatística lexical foi um instrumento indicador importante para o início da descrição da unidade textual *acção empreendedora*, mas não se mostrou suficiente para tal objectivo. Ao considerarmos, num outro patamar de análise, a relevância da concepção do inquérito em sala de aula como um género textual, complexificamos essa descrição do ponto de vista linguístico-textual.

Dessa forma, atestámos que o género inquérito em sala de aula condiciona determinadas expressões fixas que retomam conteúdos proposicionais das perguntas. Dessa forma, este género, produzido em sala de aula, é constrangido pelo lugar de circulação (sala de aula) e talvez pelo próprio papel dos interlocutores.

Vale salientar que o nosso grande desafio foi de olhar para o mesmo objecto a partir de focalizações diferentes. Na verdade, os dois investigadores que desenvolveram esta contribuição integram linhas de pesquisa diferentes e tinham como objectivo final mostrar que realmente

as perspectivas teóricas com as quais trabalhavam poderiam *dialogar*.

Como etapa futura, pretende-se continuar a inventariar aspectos linguístico-textuais que possam vir a caracterizar esta acção empreendedora do ponto de vista mais individual. E ainda resta-nos lançar um desafio: por que não desenvolver uma ferramenta computacional que seja capaz de atender à complexidade da descrição da língua?

Referências Bibliográficas

Biber, Douglas; Conrad, Susan; Reppen Randi (1998). *Corpus Linguistics Investing Language Structure and Use*. Cambridge: Cambridge University Press.

Bronckart, Jean-Paul (1999). *Atividade de linguagem, textos e discursos*. São Paulo: Educ.

Bronckart, Jean Paul (2006) *Notas da conferência proferida no INPLA*. São Paulo: PUC/SP.

Costa, Rute (2005) *Corpus de spécialité: une question de types ou de genres de textes ou de discours*. In Béjoint, Henri / Manie François (eds) *Actes du Colloque en Hommage à Philippe Thoiron*. Lyon: PUL, 313-323.

- Coutinho, Antónia (2005) Para uma linguística dos géneros de texto. *Diacrítica* 19 (1), pp. 73-88.
- Koch, Ingedore. (1999) *Argumentação e linguagem*. São Paulo: Cortez.
- Lino, Teresa (2000). Terminologia e industrias das línguas. *Actas do VII Simpósio Ibero-Americano de Terminologia e Indústrias da Língua*. Lisboa: Fundação Calouste Gulbenkian, Lisboa, pp. 25-39.
- Maingueneau, Dominique (1998). *Analyser les textes de communication*. Paris: Dunod.
- McEnery, Tony & Wilson, Andrew (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press, pp. 209.
- Rastier, François (2004). Poétique et textualité. *Langages* 153, pp. 121-126.
- Rastier, François (2005). *Enjeux épistémologiques de la linguistique de corpus*. In. Geoffray, Williams (org.). *La linguistique de corpus*. Rennes: Presses Universitaires de Rennes.