

Procura em grandes quantidades de texto e suas consequências

Diana Santos

d.s.m.santos@ilos.uio.no

9 de outubro de 2014



Inundação de informação

Todos estamos conscientes de que no mundo, agora, há mais informação acessível do que a que podemos ler e digerir.

A pergunta, então, é:

O que se pode procurar sobre grandes quantidades de texto, de forma a

- visualizar?
- filtrar?
- sumarizar?
- encontrar?

Perguntas possíveis

Não necessariamente de resposta fácil, mas das primeiras coisas que vêm à cabeça, são estudos exploratórios, caracterizadores do material:

- termos
- expressões
- perfil lexical
- construções sintáticas
- temas
- comparações
- metáforas
- questões retóricas

E sobre quê: um gênero textual, uma época, um autor, uma categoria profissional, um sexo, uma cor política?

Outra forma de proceder

Na metodologia dos estudos empíricos, faz-se grande diferença entre:

- Os ditos estudos exploratórios, “ateóricos”
- Hipóteses ou conjecturas, ou mesmo só palpites, chamados pomposamente *teste de hipóteses*. Exemplos:
 - os discursos políticos têm mais palavras positivas
 - os discursos políticos têm mais exortações
 - a forma de se referir a mulheres mudou com o tempo
 - as variantes do português aproximaram-se muito

Em teoria, primeiro encontram-se factos interessantes, que nos fazem desenvolver teorias (mais concretamente, hipóteses) que depois (de operacionalizar) testamos.

Na prática, já temos uma teoria/ideia, mas precisamos de a validar, verificar, consolidar, estruturar.

E depois, há as ferramentas...

Dada a grande quantidade de texto, não é possível manuseá-lo sem ferramentas próprias. Ferramentas para

- obter informação fiável
- obter abonações
- obter dados interessantes
- ensinar
- aprender
- visualizar o material

Exemplo

Não resisti a experimentar algo que tivesse a ver com o ensino, mas gostava de afirmar e sublinhar que o aqui apresentarei **não passa de uma brincadeira**. Não é para ser levado a sério, é apenas uma ilustração!

De qualquer maneira, é uma forma de apresentar a ideia do

Palavrómetro: um sistema onde se pode observar uma palavra à lupa.

As palavras associadas ao professor e ao ensino

Olhemos para a vossa profissão nos textos que temos:

- professor/docente
- aluno/estudante/discente
- ensino/pedagogia/didática

Como

- em termos de género
- em termos de ações associadas
- como aposto ou título

E a questão do *colega*?

Números crus

professores	226537
professor	204334
professora	56931
professoras	14166
alunos	315974
aluno	115311
aluna	7939
alunas	6013
estudantes	95303
estudante	39595

Comparando o Brasil com Portugal – só no jornalístico

Procurar: [lema="professora* alun[oa]]estudante" & variante: Corpo: TODOS Distribuir por: lema	Procurar: [lema="professora* alun[oa]]estudante" & variante: Corpo: TODOS Distribuir por: lema
---	---

fundir numa única tabela

Limite mínimo de frequência

procurar

professor	67600	40839
aluno	42150	33282
estudante	37825	28246
professora	11636	4858
aluna	1802	1248

Navigation icons: back, forward, search, etc.

Outras formas de procurar

- o papel do professor: 1399 ocorrências

gênero	casos	tamanho total	tamanho relativo (x100000)
academico	814	563577075	0,14
tecnico	518	89278750	0,58
jorn	51	521009525	0,010
oralmono	10	78543738	0,013
encic	2	38790844	
ensaio	1	2223118	
blog	1	4277183	
cartas	1	64540	1,54
oralent	1	5559594	

Navigation icons: back, forward, search, etc.

- Quais os substantivos mais correntemente associados a *professor*?
- Quais os adjetivos mais correntemente associados a *professor*?
(17871 imediatamente a seguir) *titular, universitário, emérito, catedrático, primário, adjunto, assistente, responsável, leigo, especializado, indígena, orientador...*
(4327 imediatamente antes) *bom, futuro, antigo, primeiro, velho, jovem, grande ilustre, único, novo, eminente, excelente, saudoso, mau...*
(607 predicativos) *fundamental, capaz, responsável, importante, essencial...*
- Quais os verbos mais correntemente associados a *professor*?

E os colegas?

- colega de ... (4288)
equipa, trabalho, partido, bancada, classe, profissão, turma, escola, faculdade, quarto, equipe, curso,...
- colegas de ... (9025)
trabalho, mestrado, curso, turma, escola, profissão, ...
- colega(s) meu(s) ou teu(s) ou seu(s) (670) ou dele/a/es/as?
seus (213), meu (180), meus (157), seu (110), dele (90), dela (41), deles (16), teu (10), delas (3)
- meu(s) ou teu(s) ou seu(s) colegas?
seus (10442), seu (5061), meus (3236), meu (966), teus (42), teu (27)

- Uns e outros no mesmo ambiente? *sempre tive boa relação com professores e colegas*
- um subconjunto dos outros?
 - *os meus colegas professores*
 - *uns professores meus colegas*
- uma mudança de estatuto? *primeiro meu professor, depois meu colega*

Como identificar isto? Usando o VARRA...

Exemplos de concordâncias

- Novos **colegas**, vários **professores**, diferentes disciplinas / conteúdo, novos conhecimentos .
- Os **professores** devem achar difícil mudar o seu comportamento pelas mais diversas razões, inclusive a pressão dos pais, as expectativas dos administradores, a resistência dos **colegas** ameaçados por qualquer sugestão de destruição da tradição educacional, ...
- A leccionar há cinco, seis ou mais anos, estes **professores** são ultrapassados nos concursos pelos **colegas** que se acabaram de formar.
- Aos **professores** e **colegas** de Curso pelo tempo que tivemos juntos buscando vencer mais esta etapa
- em Niterói, onde era **professor**, encontrava outro espaço de intercâmbio cultural com alunos e **colegas** ...

- Relação entre a menção do lema *professor* e uma emoção numa mesma frase (38876 casos):
desejo, amor, coragem, esperança, gratidão, humildade, medo, desejo/esperança, felicidade, surpresa, ...
- quais as palavras emotivas mais frequentes?
querer, pretender, respeito, gostar, expectativa, desejar, incentivar, agradecer, intenção, confiança, ...

Emoções associadas a professores: o respeito

O respeito como lema e o respeito como categoria.

- *a respeito de*: irrelevante, e não está marcado como emoção.
- *dizer respeito a, a esse respeito, a respeito, a meu respeito*: irrelevante, mas está a contar mal
- *respeito por, sem respeito, perder o respeito, ganhar respeito, exigir respeito, por respeito a, impor respeito, despertar respeito, merecer respeito, ...*
- Exemplos únicos: *depende do respeito, escola sobre o respeito, ...*

Categoria? O respeito está marcado como humildade – e não como admiração.

Referências I

- Baayen, R. Harald & Antoinette Renouf. “Chronicling The Times: Productive Lexical Innovations in an English Newspaper”, *Language* 72 (1), 1996, pp. 69-96.
- Bennett, Karen. “Galileo’s Revenge: Ways of Construing Knowledge and Translation Strategies in the Era of Globalization”, *Social Semiotics* 17, 2, June 2007, pp. 171-193.
- Ellis, John M. *Language, Thought and Logic*. Evanston, IL: Northwestern University Press, 1993.
- Sampson, Geoffrey. “Thoughts on two decades of drawing trees”, in Anne Abeillé (ed.), *Treebanks: Building and using parsed corpora*, Kluwer Academic Publishers, 2003, pp. 23-41.
- Santos, Boaventura de Sousa. *Epistemologies of the South. Justice against Epistemicide*. Boulder - London: Paradigm Publishers, 2014. <http://download1.hermes.asb.dk/archive/download/H0314.pdf>

Referências II

- Santos, Diana. “Podemos contar com as contas?”, in Sandra Aluísio & Stella Tagnin (eds.), *New Language Technologies and Linguistic Research: A Two-way Road*, Cambridge Scholars Publishing, 2014, pp. 194-213.
- Wilks, Yorick & Christopher Brewster. “Natural Language Processing as a Foundation of the Semantic Web”, *Foundations and Trends in Web Science* 1, 3, March 2009, pp. 199-327.