



Corpora: o que são e para quê servem

Corpora para a Terminologia

Stella E. O. Tagnin

Abril 2004

Esquema

1. O que é um corpus
 2. Como construir um corpus
 3. Como explorar um corpus
 4. Como usar o WordSmith Tools
-

O que é um corpus? (cf. Aurélio Eletrônico)

(1) Conjunto de documentos, dados e informações sobre determinada matéria.

(3) Conjunto de materiais significantes (enunciados lingüísticos, capas de revistas) constituído com vistas à análise semiológica.

O que é um corpus hoje?

Uma coletânea de textos
em formato eletrônico,
compilada segundo critérios específicos,
considerada representativa de uma língua
(ou da parte que se pretende estudar),
destinada à pesquisa

Critérios de compilação

- Origem: textos autênticos
 - Objetivo: pesquisa
 - População: seleção
 - Formato: eletrônico
 - Representatividade: de que, para quem?
 - Extensão: de acordo com objetivos
-

Qual a grande vantagem dos corpora?

**Observação da
linguagem em uso
- natural -**

Para que servem os corpora?

- Descrição da língua
 - léxico
 - gramática
 - Tradução
 - Terminologia
 - Ensino: DDL (data-driven learning)
-

Para que servem os corpora?

- Detectar padrões lexicais
 - Fraseológicos (colocações): palavras que co-ocorrem em frequência maior do que esperada: sweeping victory, high and low > search high and low
 - Terminológicos: lucros e perdas, controle de qualidade, plantas geneticamente modificadas (alteradas??)
- Compilação de dicionários e glossários

Para que servem os corpora?

- Detectar padrões gramaticais
 - “coligações”: verbos seguidos de infinitivo ou gerúndio:
 - begin + to vs. start + -ing
 - Detectar a prosódia semântica:
cause vs causar
-

O que é uma concordância?

- Lista de todas as ocorrências da **palavra de busca** em contexto
 - **KWIC**: “key word in context” –
 - em geral centralizada
 - **Leitura vertical**: permite identificar padrões gramaticais e lexicais
 - **Leitura horizontal**: permite identificar colocados e diferentes sentidos
-



now is a time to [look](#) at the other side of
the style I want?" Well, [look](#) no further. I, Adam Polselli
on in and take a " [look](#) ," literally! Be on the lookout
update is a brand new [look](#) that I call "Chic Simplicity"
call "Chic Simplicity." This new [look](#) was inspired by the outdated
Centre is opposite Coral Reef - [look](#) out for the brown signs
homepage trn: How not to [look](#) like an idiot on Usenet
you want to do is [look](#) like an idiot or be
group.name or take a [look](#) at news.answers. If you
careful of advertising. Take a [look](#) at `trn: How to advertise
Service Pack 2 Beta first [look](#) by Kurt Hutchinson A

harder [look](#) at after-school help By Lisa
in the take a closer [look](#) page for several weeks. My
for several weeks. My students [look](#) forward to the picture each
Friday March 5, 2004 Hey [look](#) , it's a new comic. That's
seen it yet, take a [look](#) . The new aqua appearance is

Concordância

family thinks Vicki Rock's former job at the plant **caused** her son's cancer.

but he denies that the plant **caused** leukemia in families that have brought suits against the company

One of two things could have **caused** this error: The Pobox.com user whose web site you

Ok: BM **caused** TONS of problems over last few weeks...31: Holes that **caused** fuel leak on shuttle located in nozzle July 28: Columbia

reduce deaths, injuries, property damage, economic losses and human suffering **caused** by natural disasters.

ventos mais fortes do oeste e aumento de tempestades violentas, **causou** danos sem precedentes nas propriedades situadas na costa

de 1992 na região central dos Estados Unidos. Por isso, **causou** grande **impacto** a descoberta do oceanógrafo americano Dudley Chelton

membro do conselho executivo deve fazer é "A nossa empresa **causou** uma primeira **boa impressão** em consumidores potenciais suficientes para garantir de 1995, quando já assumira a presidência do Banco Central, **causou** **repercussão** na imprensa o fato de ter passado o carnaval fato teve não teve a repercussão equivalente ao **prejuízo** que **causou** a centenas de empresas que hoje dependem prioritariamente da rede

Áreas de Pesquisa

Ensino e aprendizagem de língua

- o Uso de corpora no aprendizado
- o CALL: ênfase no aprendiz
- o Preparação de materiais

Linguagem do aprendiz

- o Análise de erros
- o Preparação de materiais

Língua de especialidade

- o Intuitivo vs. Natural

Áreas de Pesquisa

- Lexicologia e lexicografia
 - Estudos lexicais
 - Compilação de dicionários

 - Estudos contrastivos
 - Tradução
 - Naturalidade da língua
-

Áreas de Pesquisa

- Tradução
- Normalização/ Simplificação/ Explicitação
- Estratégias de tradução
- Normas de tradução
- Ensino e treinamento
- Fontes de referência

O diferencial da LC

■ **Intuição**

■ **VS**

■ **Observação**

O diferencial da LC

■ Probabilidade

■ VS

■ Possibilidade

Vantagens de um Corpus

- Linguagem natural
 - Atualidade
 - Confiabilidade
 - Riqueza de informações
 - lexicais
 - sintáticas
 - contextuais
 - discursivas/textuais
 - Facilidade de consulta
-

Tipos de Corpora

Quanto à língua

- monolíngües
 - bilíngües
 - multilíngües
-

Corpora monolíngües

- língua geral: jornalísticos e literários
- língua de especialidade: técnicos
- mistos
- traduções
 - TEC - Translation English Corpus
- corpus de referência
 - língua geral
 - variedade de gêneros
 - variedade de registros

TEC (Translational English Corpus)

- Traduções completas: biografias, ficção, jornais, revistas de bordo
 - ficção:
 - Saramago - diversas
 - Mario de Sá Carneiro: uma
 - Clarice Lispector: Discovering the World, The Hour of the Star
 - Chico Buarque: Turbulence
-

Corpora Online Inglês

- **BNC:** <http://sara.natcorp.ox.ac.uk/lookup.html>
 - **COBUILD:**
<http://titania.cobuild.collins.co.uk/form.html>
 - **WebCorp:**
<http://www.webcorp.org.ul/webcorp.html>
-

Corpora Online com Português

- **COMPARA:**

<http://www.linguateca.pt>

- **Lácio-Web:**

<http://www.nilc.icmc.usp.br/lacioweb>

Corpora multilíngüe

- comparáveis:

- originais em duas ou mais línguas

- paralelos:

- originais + tradução/traduições

- **COMPARA:** <http://www.linguateca.pt>
-

Then – Então – Aí – Daí - Depois

EBJB1(1018):

Then there are the three live parrots, two at Trouville and one at Venice; plus the sick parakeet at Antibes.

Há **depois** os três papagaios vivos, dois em Trouville e um em Veneza; mais o periquito doente em Antibes.

EBJB1(1033):

Then the parrot, returning its master's gaze with an unflinching eye, would murmur the cabbalistic word, and Henri's soul would be filled with the memory of his lost happiness.

Então o papagaio, retribuindo o olhar ao dono com um olho inflexível, murmurava a palavra cabalística e a alma de Henri enchia-se com a memória da sua felicidade perdida.

ESNG1(496):

Then there was the matter of her job.

Daí vinha a questão do seu emprego

Alguns sites úteis

- Michael Barlow:
<http://www.ruf.rice.edu/~barlow/corpus.html>
- ICAME (International Computer Archive of Modern and Medieval English):
 - Brown Corpus, LOB, London-Lund, Helsinki, Kolhapur<ftp://nora.hd.uib.no/pub>
- Corpus Linguistics:
<http://www-clg.bham.ac.uk/>
- David Lee's Bookmarks for Corpus-based Linguists: devoted.to/corpora

Mais sites

- University Centre for Computer Corpus Research on Language
<http://www.comp.lancs.ac.uk/computing/research/ucrel/>

Site da Universidade de Lancaster com informações sobre diversos projetos baseados em corpora, inclusive: ATR, BNC, CRATER, MARSEC e COLT.

Corpora Paralelos

- Novamente o site de Michael Barlow:

<http://www.ruf.rice.edu/~barlow/para.html>

Lista projetos de pesquisa que estudam textos traduzidos para diferentes línguas, com descrição e links.

Como construir um corpus

Construção de um Corpus

1. Objetivo do corpus

perguntas de pesquisa

2. Projeto do corpus

- a. estático ou dinâmico
- b. falado ou escrito
- c. monolíngüe ou multilíngüe (comparável ou paralelo)
- d. gêneros e tipos de textos a serem incluídos

Construção de um Corpus

2. Projeto do corpus

- e. domínios a serem incluídos
- f. proporção dos textos
- g. quantidade de textos
- h. completos ou trechos
- i. extensão dos textos
- j. fonte dos textos
- k. diversificados: autoria, gênero, tipologia
- l. originais
- m. tamanho do corpus

Construção do Corpus

- 3. Formatação: .txt
 - 4. Organização em pastas
-

4. Estruturação do Corpus

□ Paralelos

□ Inglês Português

- FoodEnginPIP1IO.doc
- EngAlimPIP1PT.doc

□ Português Inglês

- TransgenicosPPI1PO.doc
- TransgenicPPI1IT.doc

➤ PIP = Paralelo Inglês > Português

➤ PPI = Paralelo Português > Inglês

Construção de um Corpus

5a. Código de nomeação

(Manual Lácio-Web)

JO-I F-FSP-mu-05fev99_01

- Texto divulgado em jornal
- Gênero textual: informativo
- Nome do periódico: Jornal "Folha de São Paulo"
- Caderno: "Mundo"
- Data: 05 de fevereiro de 1999
- Primeiro texto (deste caderno, nesta data de publicação)

Construção de um Corpus

5b. Código de nomeação

RE-I F-NE-cea-mar01_05

- Texto divulgado em revista
- Gênero textual: informativo
- Nome do periódico: Revista "Nova Escola"
- Caderno: "Cresça e Aconteça"
- Data: mês de março de 2001
- Quinto texto (deste caderno, nesta data de publicação)

Construção de um Corpus

5c. Código de nomeação

RE-I F-CI -# -nov00_03

- Texto divulgado em revista
- Gênero textual: informativo
- Nome do periódico: Revista "Cerâmica Industrial"
- Caderno: não há subdivisões no periódico
- Data: mês de novembro de 2000
- Terceiro texto (deste caderno, nesta data de publicação)

Construção de um Corpus

5d. Código de nomeação

RE-I F-CI -# -agodez01_02

- Texto divulgado em revista
- Gênero textual: informativo
- Nome do periódico: Revista "Cerâmica Industrial"
- Caderno: não há subdivisões no periódico
- Data: período compreendendo os meses de agosto a dezembro de 2001
- Segundo texto (deste caderno, nesta data de publicação)

Construção de um Corpus

6. Cabeçalho

a. quais informações são relevantes para o projeto?

b. informação para recorte

- título
- autor
- gênero: informativo -> jornalístico
- tipo: notícia
- domínio: Ciências Biológicas -> Ecologia
- data
- fonte

Cabeçalho

```
<Header>
  <title>
    <filename> </filename>
  </title>
  <author>
    <name></name>
  </author>
  <sourceText>
    <language></language>
    <mode>[mode of delivery of textual
      content]</mode>
    <publisher></publisher>
    <pubPlace>[place of publication]</pubPlace>
    <date></date>
    <copyright>[copyrights holder]</copyright>
  </sourceText>
</Header>
```

Construção de um Corpus

7. Etiquetagem

- a. morfossintática (POS-tagging)
 - b. sintática (parsing)
 - c. semântica
 - d. discursiva
 - e. terminológica
-

Etiquetagem morfosintática

- <s>
- Foi_VAUX
- cercada_PCP
- de_PREP|+
- o_ART
- maior_ADJ
- sigilo_N
- a_ART
- chegada_N
- de_PREP|+
- a_ART
- agência=de=publicidade_N
- Saatchi_NPROP
- \$&_NPROP
- Saatchi_NPROP
- a_PREP|+
- o_ART
- Brasil_NPROP
- ._.
- </s>

Como explorar um corpus
- a compilação de um
glossário

Ferramentas de Exploração

- alinhadores
 - concordanciadores
 - contadores de frequência
 - etiquetadores
-

Como é feito o processamento?

Ferramentas de busca que produzem:

- **Concordâncias:**
 - KWIC (key word in context)
 - **Listas de colocados**
 - **Listas de palavras**
-

Corpora online com ferramentas embutidas

Results of your search - BNC

Your query was **salt**

Here is a random selection of 50 solutions from the 2943 found...

[ABB](#) 131 Add the pimentos, salame and boiling water or stock to the pan with most, but not all of, the parsley and a little **salt and pepper**.

[ABB](#) 105 A **pinch of salt** is taken for granted in many cake recipes and is added simply to bring out the flavour of the other ingredients.

[ABB](#) 1332 Return the veal to the pan, add the fresh and dried tomatoes, rosemary, wine, stock, **salt and pepper**.

[AMU](#) 1667 The sea churned the banalities of his life into flotsam: sheets, shirts, sandals, books, charts, **salt cellar**...

[B77](#) 931 But it is possible to reduce salt consumption further by `;placing the **salt shaker** at some distance from the table';.

[BPG](#) 1548 freshly ground black pepper and **salt**

[C97](#) 618 SALT is the spice of life .

[CFS](#) 1681 Substitute LoSalt for common salt, at the table and in cooking to reduce your family's **salt intake**.

[G36](#) 1259 Sieve flour into a bowl with **pinch salt**.

Query Results - Cobuild

NOTE: no more than 40 lines will be displayed here, since a threshold has been implemented. If there were more than 40 instances found, a random selection will have been applied.

are more effective than a pinch of salt. [p] Fold in with a metal spoon, chutney [h] 2 oz walnuts [p] 1/4 tsp salt [p] 1/4 tsp cayenne pepper [p] 2 sea `vegetables" of all types. Sea salt also provides some as does sw3 (tel: 071-276 5599). 4 Topiary salt and pepper pots by Swid Powell, add the saffron and stir. Season with salt and pepper. markets and collected cartoon animal salt and pepper shakers, plastic cuckoo entertaining even Hollywood moguls to salt-beef sandwiches in mainly with boiled water, sugar and salt, can save most diarrhea victims' salt. [p] Herb, vegetable and spice salt: compounds of salt with other few leaves crisp iceberg lettuce [p] salt, freshly ground black pepper [p] a great lover of liberally sprinkling salt on her food at the table, thereby served fried egg and crisp slices of salt pancetta. [p] Caesar salad was tray at my head. I A large pinch of salt should be applied to this story. mousse-like. Sift over the flour and salt, then fold in to the eggs and Lanzarote round potatoes and rock salt. Tomatoes, sweet potatoes, I took her words with a grain of salt, went home, put the sample on a

T- Score - salt

<u>Collocate</u>	<u>Corpus Freq</u>	<u>Joint Freq</u>	<u>Significance</u>
And	1369241	813	17.533492
Pepper	903	285	16.869713
With	364279	237	9.984592
Lake	2689	93	9.579897
Sugar	2472	90	9.427256
Water	15678	86	8.887077
Black	16881	84	8.744025
City	20496	85	8.711252
Pinch	405	73	8.533166
Ground	8804	64	7.748380
Sea	5756	59	7.509810
Tsp	290	53	7.271002
Flour	676	51	7.119786
Add	5006	52	7.052378
Freshly	404	45	6.694434
Season	16627	51	6.609096

Mut ual I nf or mat ion - salt

<u>Collocate</u>	<u>Corpus Freq</u>	<u>Joint Freq</u>	<u>Significance</u>
Pepper	903	285	10.431903
Monosodium	12	3	10.095633
Glutamate	13	3	9.980145
Dampier	18	4	9.925691
Tsp	290	53	9.643600
Pinch	405	73	9.623633
Teaspoon	151	27	9.612068
Paprika	51	9	9.593083
Crinkle	17	3	9.593083
Vinegar	272	40	9.330022
Nutmeg	82	12	9.322967
Sodium	150	18	9.036634
Oregano	43	5	8.991187
Freshly	404	45	8.929159

WebCorp output for search term "sal"

Additional filter: "culinaria - Domain: ".br"

Click on the red text to view the originating web page

e a noz-moscada. Tempere com [sal](#) a gosto. 2. Mexa bem de caldo de galinha quente [sal](#) e [pimenta preta moída na hora](#) tempere com um pouco de [sal](#) e prepare-se para ficar perto dendê e [uma pitada de \[sal\]\(#\)](#). Coloque sobre a folha de colher (sopa) de salsa picada [sal](#) e pimenta-do-reino a gosto Preparo frango cozido em água e [sal](#) 100 g de cogumelos cortados uma lata de milho verde [sal](#) e pimenta-do-reino branca a gosto dois litros de água com [sal](#), adicione os camarões e a mais usada em comidas de [sal](#) CAFOFA - farofa de torresmo ou de açúcar, 1 colherinha de [sal](#), 2 ovos inteiros 100 a de açúcar, 1 colher de [sal](#) 3 ovos inteiros 1 cebola 1 colher de sobremesa de [sal](#) 1 colherinha de açúcar 2 fermento no açúcar, juntar o [sal](#), os líquidos e bater um

Word	Total	L4	L3	L2	L1		R1	R2	R3	R4	Left Total	Right Total
com	189	22	16	19	116		1	2	6	7	173	16
pimenta	157	4	1	3	3		46	78	15	7	11	146
gosto	96			4	2			62	2	26	6	90
colher	83	29	34	3				12	3	2	66	17
alho	78	6	9	14	15		15	8	4	7	44	34
chá	70	1	4	52	1				5	7	58	12
água	63	5	16	28	6		1	3	1	3	55	8
pimenta-do-reino	50	2	1		1		10	20	10	6	4	46
pitada	49		1	39				9			40	9

WebCorp (sem “stop words”)

Key Phrases:

com sal alho sal sem sal ovos sal água sal
leite sal limão sal trigo sal óleo sal
ralado sal ossobuco sal sal pimenta sal grosso
sal alho sal Modo sal pimenta-do-reino
sal limão sal ½ sal Quando sal 200g
sal Misture sal 250g

Corpora FVM

- Personalizados: tema dos textos
 - Adequados às suas necessidades: critérios de seleção
 - Sempre à mão
 - Sempre atualizados - ou atualizáveis
 - Necessitam de ferramenta de busca: Wordsmith Tools
-

Como usar o Wordsmith Tools para compilar um glossário

Primeiramente

baixar a versão demo do
Wordsmith Tools do site do Mike
Scott:

<http://www.liv.ac.uk/~ms2928>

<http://www.oup.co.uk/elt/software/wsmith>

Wordsmith Tools

- **Concord**
 - KWIC = Key Word In Context
 - Collocates
 - Clusters
- **WordList**
 - S = General Statistics
 - F = Frequency
 - A = Alphabetical
- **KeyWords**
 - Corpus de estudo vs Corpus de referência

Compilação do Glossário

- Identificação das palavras-chave:
 - Lista de Palavras do corpus
-

Wordlist

N	Word	Freq.	%	N	Word	Freq.	%
1	DE	18.528	7,70	17	SAL	1.463	0,61
2	E	9.725	4,04	18	NO	1.364	0,57
3	O	7.291	3,03	19	ATÉ	1.331	0,55
4	A	6.569	2,73	20	AO	1.307	0,54
5	COM	4.513	1,87	21	BEM	1.302	0,54
6	EM	4.254	1,77	22	COLOQUE	1.293	0,54
7	PARA	2.578	1,07	23	DA	1.277	0,53
8	OS	2.210	0,92	24	NA	1.235	0,51
9	UMA	2.177	0,90	25	ÁGUA	1.206	0,50
10	AS	2.065	0,86	26	LEITE	1.203	0,50
11	DO	2.027	0,84	27	COLHER	1.186	0,49
12	UM	2.006	0,83	28	MINUTOS	1.151	0,48
13	POR	1.924	0,80	29	SE	1.125	0,47
14	QUE	1.648	0,68	30	DEIXE	1.087	0,45
15	OU	1.524	0,63	31	CHÁ	1.067	0,44
16	SOPA	1.507	0,63	32	XÍCARA	1.045	0,43

Wordlist

N	Word	Freq.	%	N	Word	Freq.	%
1	DE	18.528	7,70	17	SAL	1.463	0,61
2	E	9.725	4,04	18	NO	1.364	0,57
3	O	7.291	3,03	19	ATÉ	1.331	0,55
4	A	6.569	2,73	20	AO	1.307	0,54
5	COM	4.513	1,87	21	BEM	1.302	0,54
6	EM	4.254	1,77	22	COLOQUE	1.293	0,54
7	PARA	2.578	1,07	23	DA	1.277	0,53
8	OS	2.210	0,92	24	NA	1.235	0,51
9	UMA	2.177	0,90	25	ÁGUA	1.206	0,50
10	AS	2.065	0,86	26	LEITE	1.203	0,50
11	DO	2.027	0,84	27	COLHER	1.186	0,49
12	UM	2.006	0,83	28	MINUTOS	1.151	0,48
13	POR	1.924	0,80	29	SE	1.125	0,47
14	QUE	1.648	0,68	30	DEIXE	1.087	0,45
15	OU	1.524	0,63	31	CHÁ	1.067	0,44
16	SOPA	1.507	0,63	32	XÍCARA	1.045	0,43

Compilação do Glossário

- Identificação das palavras-chave:
 - Lista de Palavras-Chave
 - corpus de estudo vs corpus de referência
-

Wordsmith Tools - Key Words

N	WORD	FREQ.	CULINPTG.LST %	FREQ TEK.	KEYNESS
1	SOPA	1.507	0,63	0	4.651,8
2	SAL	1.463	0,61	10	4.400,7
3	COLOQUE	1.293	0,54	4	3.938,0
4	COLHER	1.186	0,49	4	3.608,0
5	DE	18.528	7,70	40.115	3.478,0
6	LEITE	1.203	0,50	28	3.458,4
7	CHÁ	1.067	0,44	0	3.292,0
8	DEIXE	1.087	0,45	11	3.235,9
9	XÍCARA	1.045	0,43	0	3.224,1
10	ÁGUA	1.206	0,50	92	3.101,3
11	COLHERES	989	0,41	0	3.051,1
12	MINUTOS	1.151	0,48	71	3.043,7
13	AÇÚCAR	976	0,41	1	2.995,7
14	MANTEIGA	955	0,40	0	2.946,1
15	FARINHA	924	0,38	0	2.850,4
16	FOGO	948	0,39	7	2.845,1

Português Europeu - positivas

N	WORD	FREQ.	TOTPOP.LST %	FREQ.	TOTPOB.LST %	KEYNESS	P
1	SE	385	2,18	90	0,36	314,8	0,000000
2	LUME	82	0,46	0		144,1	0,000000
3	Q	77	0,44	0		135,3	0,000000
4	B	77	0,44	0		135,3	0,000000
5	PREPARAÇÃO	77	0,44	3	0,01	113,0	0,000000
6	COZER	59	0,33	0		103,7	0,000000
7	DL	57	0,32	0		100,2	0,000000
8	GR	89	0,50	13	0,05	92,6	0,000000
9	NATAS	49	0,28	0		86,1	0,000000
10	SUMO	39	0,22	0		68,5	0,000000
11	GRS	43	0,24	1		67,1	0,000000
12	CONFECÇÃO	37	0,21	0		65,0	0,000000
13	DURANTE	61	0,35	9	0,04	63,1	0,000000
14	S	29	0,16	0		50,9	0,000000
15	DEITE	27	0,15	0		47,4	0,000000
16	JUNTA	23	0,13	0		40,4	0,000000
17	TACHO	21	0,12	0		36,9	0,000000
18	MARGARINA	57	0,32	19	0,08	35,1	0,000000
19	ARREFECER	19	0,11	0		33,4	0,000000
20	PURÉ	19	0,11	0		33,4	0,000000
21	AS	246	1,39	202	0,81	32,5	0,000000
22	LHE	17	0,10	0		29,8	0,000000
23	DEITA	17	0,10	0		29,8	0,000000
24	BOCADOS	17	0,10	0		29,8	0,000000
25	COM	445	2,52	434	1,75	29,8	0,000000
26	COZA	16	0,09	0		28,1	0,000000
27	DEIXA	25	0,14	3	0,01	28,1	0,000000
28	C	40	0,23	12	0,05	27,0	0,000000
29	TAPE	15	0,08	0		26,3	0,000000
30	ALOURAR	14	0,08	0		24,6	0,000001

Português Brasileiro - negativas

N	WORD	FREQ.	TOTPOP.LST %	FREQ.	TOTPOB.LST %	KEYNESS	P
31	XÍCARAS	2	0,01	34	0,14	24,6	0,000001
32	GELADEIRA	0		24	0,10	25,8	0,000000
33	REFOGUE	2	0,01	36	0,15	26,6	0,000000
34	RESERVE	9	0,05	60	0,24	26,9	0,000000
35	OLIVA	0		30	0,12	32,3	0,000000
36	SALSINHA	0		30	0,12	32,3	0,000000
37	COZINHE	6	0,03	58	0,23	33,1	0,000000
38	PORÇÕES	1		38	0,15	33,3	0,000000
39	GOSTO	12	0,07	78	0,31	34,3	0,000000
40	PARA	89	0,50	252	1,02	35,8	0,000000
41	RENDIMENTO	0		35	0,14	37,6	0,000000
42	BAIXO	1		43	0,17	38,5	0,000000
43	ACRESCENTE	9	0,05	81	0,33	44,4	0,000000
44	DO	70	0,40	233	0,94	46,1	0,000000
45	PREPARO	9	0,05	87	0,35	49,7	0,000000
46	POR	77	0,44	262	1,06	53,9	0,000000
47	PANELA	15	0,08	115	0,46	57,1	0,000000
48	SUCO	0		59	0,24	63,5	0,000000
49	CHÁ	21	0,12	156	0,63	75,9	0,000000
50	REINO	0		73	0,29	78,6	0,000000
51	COLOQUE	23	0,13	183	0,74	93,3	0,000000
52	FOGO	6	0,03	135	0,54	106,3	0,000000
53	XÍCARA	0		119	0,48	128,2	0,000000

Compilação do Glossário

Análise do Contexto

- Concordâncias
 - informações lexicais - colocações
 - informações sintáticas - coligações
-

Concord - [KIDNEY: 532 entries (sort: 1R,1L)]

File View Settings Window Help

Concordance

N	Concordance	Set	Tag	Word No.	File
1	ium and potassium retention and how the failing kidney adapts to loss of nephrons by increased			2.932	ni039.txt
2	this admission, the transplant surgeon places a kidney allograft into the recipient's iliac fossa. A			6.389	ni040.txt
3	th T and B lymphocytes are important players in kidney allograft rejection. B lymphocytes make			4.154	ni040.txt
4	Small lymphocytes are central to the problem of kidney allograft rejection. Both T and B lymphoc			4.142	ni040.txt
5	s sole immunosuppressive agent in recipients of kidney allografts from cadaver donors: preliminar			19.921	ni043.txt
6	splantation. In animals and human recipients of kidney allografts, matching for both class I and			4.526	ni040.txt
7	ered at the glomerulus, is not metabolized in the kidney, and does not undergo tubular reabsorpti			2.226	ni044.txt
8	lobulin is largely metabolized and cleared by the kidney, and the plasma concentration of J2-micr			5.822	ni043.txt
9	rugs, which are in whole or part excreted by the kidney, and the resulting endogenous insulin pr			4.034	ni039.txt
10	er an MI or stroke has deleterious effects on the kidney and estimating the risk for ESRD among			3.444	ni033.txt
11	ble and inexorable loss of renal function. Once a kidney begins to fail, most patients whose glom			145	ni041.txt
12	al problems remained, however, that the artificial kidney (by itself) did not correct. Anemia, bone			1.587	ni040.txt
13	INCIPLES OF HEMODIALYSIS When a failing kidney can no longer adequately excrete waste			128	ni043.txt
14	oses or unilateral disease in a single functioning kidney can sometimes be prevented from progre			1.209	ni039.txt
15	nal bleeding disease, medullary sponge kidney Coagulation disorders: anticoagulation t			7.164	ni044.txt
16	gnosis Chronic renal failure is defined as either kidney damage or glomerular filtration rate less t			499	ni021.txt
17	nsplant Initiation factors (Factors that initiate kidney damage) * Diabetes* * Hypertension*			806	ni021.txt
18	tors have shown acute, "reversible," and chronic kidney damage. Cyclosporine slows recovery fro			7.061	ni040.txt
19	stion related to anemia correction in the chronic kidney disease period and the likelihood of goin			5.034	ni030.txt
20	SRD), myocardial infarction (MI), stroke, chronic kidney disease (CKD), risk factors			412	ni033.txt
21	sed in all groups as they transition from chronic kidney disease to either ESRD or death. These			4.825	ni030.txt
22	ith the exception of those with medullary cystic kidney disease (see Chapter 115) have fixed pro			1.359	ni039.txt
23	ase, medullary sponge kidney, medullary cystic kidney disease Interstitial Disease Acute: Inter			1.098	ni044.txt
24	the causative disease into question. Poly-cystic kidney disease can be characterized by recurre			1.915	ni039.txt
25	ant polycystic kidney disease Medullary cystic kidney disease *Systemic disease involving the			1.636	ni039.txt

14:22

C Concord - [collocates [total]]																
C File View Settings Window Help																
     																
      																
N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5	
1	KIDNEY	726	17	26	2	8	4	2	1	683	2	3	4	10	7	
2	THE	396	299	97	40	39	22	39	159	0	0	35	20	19	23	
3	DISEASE	154	62	92	8	13	21	9	11	0	73	0	5	3	11	
4	RENAL	119	94	25	32	29	14	19	0	0	0	2	9	12	2	
5	PATIENTS	53	40	13	9	3	16	1	11	0	0	6	2	2	3	
6	TRANSPLANTATION	51	12	39	1	4	4	1	2	0	32	2	3	1	1	
7	CHRONIC	50	41	9	12	3	9	2	15	0	0	2	6	0	1	
8	WITH	48	33	15	8	10	3	11	1	0	3	1	7	2	2	
9	FOR	46	34	12	3	12	5	5	9	0	0	2	2	2	6	
10	FUNCTION	43	15	28	4	5	5	1	0	0	24	0	0	1	3	
11	DISEASES	34	21	13	1	2	16	2	0	0	11	0	0	1	1	
12	FAILURE	30	20	10	5	4	5	2	4	0	8	0	0	1	1	
13	POLYCYSTIC	29	28	1	2	1	0	0	25	0	0	0	1	0	0	
14	TRANSPLANT	27	4	23	0	2	1	1	0	0	20	1	1	1	0	
15	SUPPL	26	0	26	0	0	0	0	0	0	12	1	1	12		
16	ARE	24	13	11	4	7	1	1	0	0	3	0	4	2	2	
17	DIALYSIS	23	22	1	4	4	5	7	2	0	0	0	0	0	1	
18	HEMODIALYSIS	21	20	1	2	8	2	4	4	0	0	0	1	0	0	
19	THAT	21	16	5	4	3	4	4	1	0	0	0	1	2	2	
20	FROM	19	8	11	0	2	1	5	0	0	3	6	0	0	2	
21	NOT	17	6	11	1	4	0	1	0	0	1	5	2	3	0	
22	ERYTHROPOIETIN	16	11	5	2	5	1	0	3	0	1	1	2	0	1	
23	THAN	15	7	8	2	1	2	2	0	0	1	1	2	1	3	
24	BRENNER	14	2	12	0	2	0	0	0	0	0	0	1	9	2	
25	MEDULLARY	14	8	6	1	0	0	7	0	0	0	3	0	1	2	
26	NORMAL	14	9	5	0	1	0	0	8	0	0	0	4	0	1	

Iniciar



14:29

Compilação do Glossário

Seleção dos Termos

- “Clusters” - unidades multipalavras
-

N	cluster	Freq.
1	tempere com sal e pimenta	38
2	o sal e a pimenta	34
3	tempere com sal e pimenta-do-reino	32
4	sal e pimenta a gosto	24
5	sal e pimenta-do-reino a gosto	21
6	sal e pimenta-do-reino modo de	20
7	com sal e pimenta e	16
8	sal e pimenta do reino	15
9	de oliva sal e pimenta	14
10	e pimenta-do-reino modo de preparo	14
11	e pimenta moída na hora	11
12	sal e pimenta moída na	11
13	al e pimenta-do-reino a gosto	9
14	com sal e pimenta-do-reino a	8
15	oliva sal e pimenta a	8
16	com sal e pimenta-do-reino branca	7
17	e pimenta-do-reino a gosto modo	7
18	pimenta-do-reino a gosto modo de	7
19	sal e a pimenta e	7
20	e pimenta a gosto azeite	6
21	e pimenta-do-reino branca modo de	6
22	e pimenta-do-reino modo de fazer	6
23	sal e pimenta-do-reino branca modo	6
24	salsinha picada sal e pimenta	6
25	colher de chá de pimenta	5
26	de manteiga sal e pimenta-do-reino	5
27	e pimenta do reino a	5
28	e pimenta-do-reino a gosto rendimento	5
29	pimenta do reino a gosto	5
30	tempere-os com sal e pimenta	5

Compilação do Glossário

- Identificação dos equivalentes
 - L1: padrão lexical, sintático: **black pepper - freshly ground black pepper – season with salt**
 - L2: candidatos a equivalentes: **pimenta preta, pimenta-do-reino**
 - L2: padrão lexical, sintático: **pimenta-do-reino moída na hora – tempere com sal**
 - corpus paralelo
 - obras de referência
 - **Termo provável - não apenas possível**
- Contextualização

C Concord - [RIM: 123 entries (sort: 1R,1L)]

File View Settings Window Help

Concordance

N	Set	Tag	Word No.	File	%
98			1.183	ni103.txt	41
99			1.775	ni141.txt	59
100			2.128	ni149.txt	54
101			2.097	ni149.txt	53
102			2.732	ni141.txt	90
103			1.186	ni149.txt	31
104			2.447	ni104.txt	75
105			2.789	ni142.txt	41
106			626	ni118.txt	25
107			153	ni123.txt	2
108			1.432	ni125.txt	13
109			907	ni103.txt	32
110			917	ni103.txt	32
111			4.899	ni123.txt	46
112			2.712	ni124.txt	17
113			2.634	ni124.txt	16
114			3.902	ni124.txt	24
115			2.783	ni124.txt	17
116			2.808	ni124.txt	17
117			2.890	ni125.txt	27
118			494	ni118.txt	21
119			2.028	ni139.txt	51
120			119	ni113.txt	17
121			2.336	ni125.txt	22
122			2.644	ni142.txt	39
123			1.175	ni103.txt	41

Renal Neoplasm

Introdução

O rim, apesar de sua incapacidade funcional na

e ocorrer precocemente, numa fase em que o rim ainda está desnervado, sugerindo que out

a produção de eritropoietina foi relacionada ao rim. A anemia na IRC é tipicamente normocíti

berem transfusões próprias de seu doador de rim. A isso se denomina transfusão específi

bém afeta o fluxo plasmático renal (FPR), no rim a resistência ao fluxo por meio das arterio

deles, que mostrava cistos em apenas um rim; a impossibilidade de se detectar cistos n

15:24

Concord - [DOENÇA: 48 entries (sort: 1R,2R)]

File View Settings Window Help

Concordance

N	Set	Tag	Word No.	File	9
1			1.545	ni118.txt	5
2			414	ni103.txt	1
3			796	ni113.txt	9
4			1.904	ni103.txt	6
5			7.889	ni123.txt	7
6			798	ni147.txt	6
7			1.041	ni142.txt	1
8			1.017	ni147.txt	8
9			27	ni147.txt	
10			569	ni126.txt	2
11			1.283	ni135.txt	3
12			1.186	ni142.txt	1
13			1.403	ni123.txt	1
14			1.194	ni123.txt	1
15			667	ni118.txt	2
16			449	ni103.txt	1
17			6.786	ni125.txt	6
18			147	ni113.txt	2
19			664	ni142.txt	
20			765	ni142.txt	1
21			1.005	ni142.txt	1
22			951	ni126.txt	3
23			2.363	ni135.txt	6
24			1.456	ni118.txt	5
25			2.633	ni118.txt	9

19 anos, P<0,05) e tinham um maior tempo de doença renal ($78,2 \pm 46,6 \times 26,3 \pm 26,3$ mese
lise. INTRODUÇÃO A doença renal cística adquirida (DRCA) é defini
bservada em hemodialisados que desenvolvem doença renal cística da diálise constitui um su
ritropoetina, e que quanto maior a gravidade da doença renal cística, ou seja, quanto maior o t
o chegam à uremia terminal. As crianças com doença renal congênita e aquelas com disfunç
e o período de sono noturno de pacientes com doença renal crônica terminal (n = 20); pacient
1. HIPERTENSÃO ARTERIAL SISTÊMICA A doença renal crônica (DRC) e a hipertensão ar
terial com o sono é causa ou consequência da doença renal crônica é ainda um ponto a ser e
TERIAL EM PACIENTES PORTADORES DE DOENÇA RENAL CRÔNICA Subtítulo: - Lín
se tipo de proteinúria é quase sempre sinal de doença renal estrutural, mesmo quando a funç
nados à raça possam exercer papel causai na doença renal hipertensiva (Fig. 32.1). Mecani
rceira associação entre HAS e DRC ocorre na doença renal isquêmica, em que a lesão renal
ão a nefropatia da gota e a tuberculose renal. Doença renal policística É uma doença heredi
Nefropatia diabética Causas menos comuns Doença renal policística Síndrome de Alport L
tidade distinta dos cistos renais simples ou da doença renal policística do adulto,4 desenvolv
a tratamento dialítico por longo tempo, cuja doença renal primária não era doença cística (
e da artéria renal, rejeição crônica, recidiva da doença renal primária, obesidade, uso de corti
s particularidades existem na dependência da doença renal primária. A anemia pode ser pre
uidados contínuos e intensivos à medida que a doença renal progredisse. Porém, a realidade
de cuidados ideais para o paciente portador de doença renal progressiva é a participação do n
e no tratamento proposto para o paciente com doença renal progressiva, já que cuidados ade
a em diabéticos refletiria a associação de uma doença renal séria e específica52. Um século
16 (67%) apresentavam evidência sugestiva de doença renal significativa no início do seguime
efrite crônica (GNC) foi a principal etiologia da doença renal terminal (35,2%). No grupo em tr
relação direta com idade, raça e etiologia da doença renal terminal. Além disto, ressalta a i

15:34

WordList - [key words (keyness)]							
File Settings Comparison Index Window Help							
N	WORD	FREQ.	ING.LST %	FREQ.	BNCS.LST %	KEYNESS	P
1	RENAL	2.203	1,10	0		10.839,6	0,000000
2	PATIENTS	1.845	0,92	57		8.573,0	0,000000
3	DISEASE	1.046	0,52	35		4.838,4	0,000000
4	ESRD	836	0,42	0		4.108,2	0,000000
5	DIALYSIS	693	0,35	1		3.390,1	0,000000
6	KIDNEY	683	0,34	11		3.244,8	0,000000
7	FAILURE	637	0,32	67		2.699,1	0,000000
8	PROTEIN	532	0,27	8		2.531,7	0,000000
9	CHRONIC	514	0,26	12		2.412,8	0,000000
10	J	566	0,28	153		2.063,7	0,000000
11	CREATININE	368	0,18	0		1.807,6	0,000000
12	TRANSPLANTATIO+	352	0,18	2		1.704,6	0,000000
13	GLOMERULAR	334	0,17	0		1.640,5	0,000000
14	STUDY	507	0,25	250	0,01	1.575,0	0,000000
15	BLOOD	424	0,21	130		1.502,4	0,000000
16	TREATMENT	421	0,21	129		1.492,0	0,000000
17	HEMODIALYSIS	296	0,15	0		1.453,8	0,000000
18	SERUM	307	0,15	10		1.420,9	0,000000
19	OF	7.702	3,85	50.741	2,38	1.410,2	0,000000
20	LUPUS	276	0,14	0		1.355,6	0,000000
21	ET	331	0,17	56		1.315,9	0,000000
22	PLASMA	260	0,13	2		1.253,8	0,000000
23	FUNCTION	337	0,17	89		1.234,6	0,000000
24	HYPERTENSION	241	0,12	0		1.183,6	0,000000
25	NEPHRITIS	243	0,12	1		1.180,7	0,000000
26	AL	321	0,16	93		1.152,2	0,000000

Concord - [RENAL: 1273 entries (sort: 1L, Centre)]

File View Settings Window Help

Concordance

N	Set	Tag	Word No.	File	%
1	oxalate/creatinine ratio was 0.42	Renal	660	ni104.txt	21
2	ulin (Tamm-Horsfall glycoprotein): a renal	ligant for lymphokines. Scienc	2.346	ni126.txt	96
3	L, Grassi C. Hepatitis C: reality of a renal	unit. Nephron 1992; 61:369-	3.134	ni102.txt	100
4	de eliminação da pré-albumina é a renal,	podendo estar falsamente el	1.422	ni144.txt	49
5	ratos Munich-Wistar com ablação renal	de 5/6, que a correção da ane	750	ni143.txt	45
6	ratos Munich-Wistar com ablação renal	de 5/6 que a correção da ane	198	ni110.txt	31
7	ratos Sprague-Dawley com ablação renal	de 3/4, mais semelhante à re	296	ni110.txt	45
8	mg/dL). The prevalence of acquired renal	cystic disease was 80%. No	2.173	ni103.txt	75
9	Summary Acquired renal	cystic disease (ARCD) attack	331	ni118.txt	14
10	ostic considerations in acquired renal	cystic disease. Kidney Int 19	2.309	ni103.txt	80
11	tzet LH. Natural history of acquired renal	cystic disease in dialysis p	2.580	ni103.txt	88
12	Keywords: Acquired Renal	Cystic Disease, Hemodialysi	486	ni118.txt	20
13	) de fósforo cai, há uma adaptação renal	caracterizada por um declínio	598	ni100.txt	13
14	ão dos mecanismos de adaptação renal	na insuficiência renal crônica:	1.919	ni120.txt	31
15	min. Assim, diferente da adaptação renal	quase total vista no balanço d	4.310	ni123.txt	41
16	ios outros processos de adaptação renal,	como veremos adiante. Os	1.990	ni120.txt	33
17	nvolvimento de adenocarcinoma renal	após período prolongado de t	1.616	ni103.txt	56
18	(estágio V), ou de adenocarcinoma renal	em rim único ou bilateral, tam	2.085	ni138.txt	78
19	esenvolvimento do adenocarcinoma renal.	Embora esses cistos carregie	1.845	ni103.txt	63
20	istinose Causas raras Amiloidose renal	Mieloma múltiplo Poliarterite n	1.209	ni123.txt	12
21	Cistinose Amiloidose renal	Doen	1.892	ni123.txt	19
22	oi realizada em 40% e a angiografia renal	em 25% dos pacientes, send	2.589	ni135.txt	71
23	DINGS 3/4 Hepatitis C antibodies. Renal	dialysis. Immunoblotting. </	3.373	ni109.txt	78
24	mitivos, deve-se fazer a arteriografia renal	para afastar a estenose da art	6.372	ni125.txt	59
25	mente com urografia e arteriografia renal.	Esta última permite ao cirurg	1.402	ni125.txt	13
26	stico é confirmado pela arteriografia renal.	A conduta terapêutica que s	6.675	ni125.txt	62

14:45

C Concord - [RENAL: 2203 entries (sort: 1R,2R)]

File View Settings Window Help

Concordance

N	Set	Tag	Word No.	File	%
33			2.970	d:\ing les~5\ca pit~13\m ni039.txt	38
34			4.974	d:\ing les~5\ca pit~13\m ni039.txt	65
35			6.362	d:\ing les~5\ca pit~13\m ni037.txt	89
36			12.753	d:\ing les~5\ca pit~13\m ni043.txt	60
37			108	d:\in gles~5\la rtig~11\m ni029.txt	2
38			4.562	d:\ing les~5\ca pit~13\m	22

33 ssed. Progressive metabolic acidosis develops with CRF. The major cause of the failure to excrete enough acid is diminished renal ammonia production and excretion. Although the metabolic acidosis of CRF is commonly referred to as an anion gap acidosis, this gap does not develop until the serum creatinine concentration approach

34 cause of systemic amyloidosis, which often involves the kidneys. Hepatosplenomegaly and macroglossia also suggest renal amyloidosis. Laboratory studies should include measurement of serum electrolytes, calcium, phosphorus, alkaline phosphatase, and albumin. Careful urinalysis and urinary microscopy should be perfor

35 UCLIDE IMAGING. Radiopharmaceuticals may be used to obtain general images of the kidney. These are only rough estimates of renal anatomy but have some advantage in that allergic reactions or induced nephropathy is essentially non-existent to radioactive compounds as compared with radiocontrast materials. In this technique, radiopharmac

36 porine The administration of cyclosporine, a unique cyclic en-decapeptide derived from a fungal metabolite, has revolutionized renal and all other types of solid-organ transplantation.156 Numerous clinical trials have shown a striking improvement in renal allograft survival in patients treated with cyclosporine and prednisone as compared wit

37 rônica Meio: Internet TRADUÇÃO Nome do tradutor: - Sexo: - Direção: - Abstract Access to living related renal and cadaveric renal transplantation was compared between 5,863 adults with incident end-stage renal disease (ESRD) caused by lupus nephritis and 463,141 adults with other causes of ESRD who were in

38 o some patients to increase erythropoiesis and appear to be moderately effective.56-57 Androgens are thought to stimulate both renal and extrarenal production of erythropoietin but the response rate is variable and unpredictable Nandrolone

15:44

Disponibilização dos Glossários

- www.fflch.usp.br/citrat
 - --> Glossários
 - --> Tradução
-

WordList - [new wordlist (S)]							
File Settings Comparison Index Window Help							
							
							
	N	1	2	3	4	5	6
Text File		OVERALL	MNI100.TXT	MNI101.TXT	MNI102.TXT	MNI103.TXT	MNI104.TXT
Bytes		1.251.646	35.368	31.163	21.239	20.446	23.589
Tokens		188.319	5.457	4.685	3.256	2.995	3.443
Types		14.908	1.464	1.172	981	956	1.137
Type/Token Ratio		7,92	26,83	25,02	30,13	31,92	33,02
Standardised Type/Token		43,15	43,40	42,03	43,45	44,20	48,20
Ave. Word Length		5,19	5,04	5,11	4,89	5,20	5,25
Sentences		5.986	302	218	177	213	169
Sent. length		31,05	18,07	21,46	18,39	14,06	20,33
sd. Sent. Length		28,98	17,34	21,64	18,58	12,76	26,61
Paragraphs		25	0	0	0	0	0
Para. length		959,52					
sd. Para. length		1.221,14					
Headings		0	0	0	0	0	0
Heading length							
sd. Heading length							
1-letter words		21.702	692	535	432	311	414
2-letter words		34.845	1.164	1.010	743	571	695
3-letter words		20.208	560	527	363	336	349
4-letter words		13.584	403	341	247	237	251
5-letter words		16.323	421	373	290	255	261
6-letter words		13.742	403	323	186	306	237
7-letter words		14.594	506	358	217	248	320
8-letter words		13.254	316	267	190	200	237
9-letter words		11.287	400	314	192	148	176

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.