

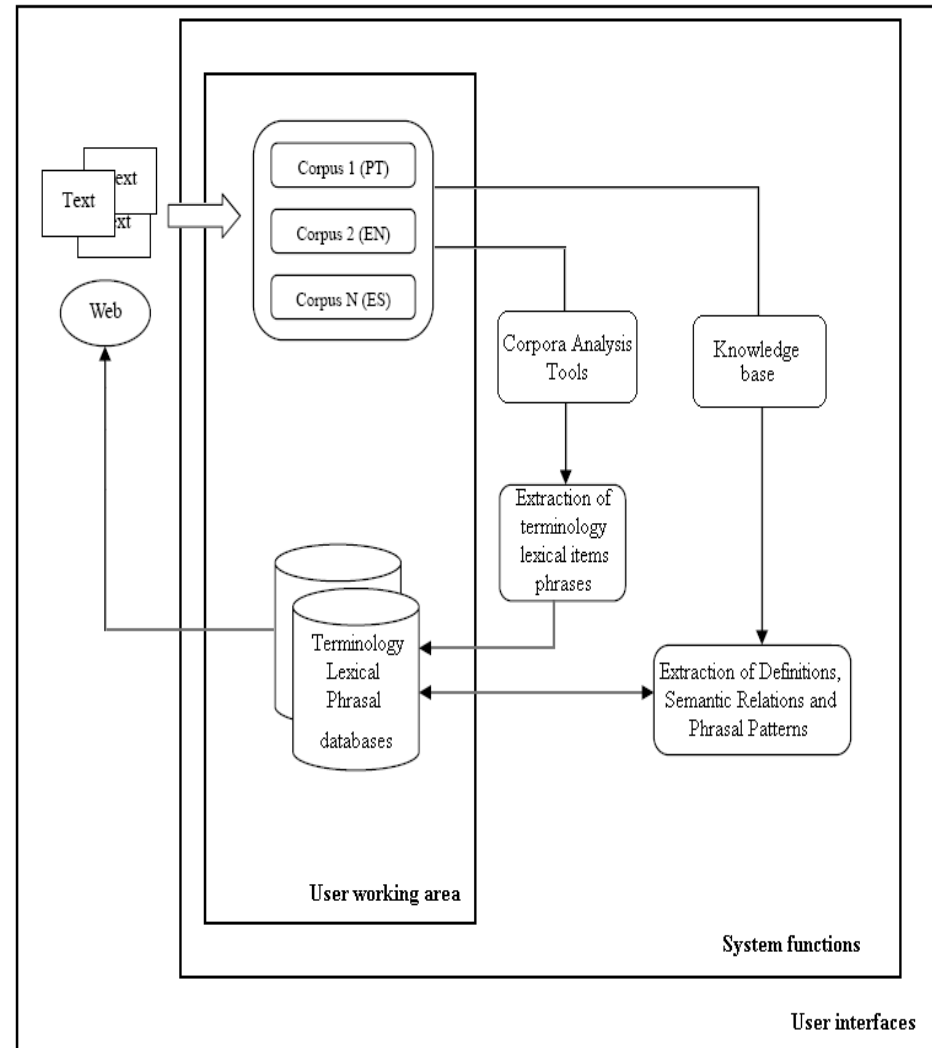
# Corpógrafo and Nooj

## Using Linguistic Resources to Obtain Aligned Concordances from Corpora

Sérgio Matos, Anabela Barreiro, Belinda Maia  
April 2008

# Corpógrafo

- Text manipulation
- Corpora analysis
  - Concordances
  - n-grams
- Terminology
  - Extraction of terms
  - Extraction of relations, definitions



# Corpógrafo

- Concordances
  - Using regular expressions (e.g. “(the|a) \w+ is \w+”)
- N-grams
  - Identification of multi-word expressions / lexical bundles

# Corpógrafo

Applications Places System Sérgio Matos Mon 21 Apr, 15:40

Corpógrafo - smatos21 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

Corpógrafo - Linguateca

- Gestor
- Pesquisa
  - Concordância Frase
  - Estudo de N-Gramas**
  - Concordância Janela
  - Concordância KWIC
  - Concordâncias paralelas
- Centro de Conhecimento
- Centro de Comunicação

**N-Gramas**

- \* Exportar
- \* Voltar ao menu

Imprimir!

**Informações**

Corpus: Natura-PUBLICO (1726679 átomos)  
Lista com 981104 N-Gramas de comprimento 3

||<| <<| >>| ||

#	n-grama	# oc.	freq. %
1	milhões de contos	628	0.036
2	, o que	514	0.029
3	, que se	348	0.020
4	no entanto ,	331	0.019
5	, com a	298	0.017
6	, com o	293	0.016
7	, depois de	292	0.016
8	o presidente da	249	0.014
9	por cento ,	246	0.014
10	milhões de dólares	239	0.013
11	secretário de estado	239	0.013
12	, e a	217	0.012
13	em lisboa ,	216	0.012
14	a partir de	215	0.012
15	por cento ,	215	0.012
16	, e o	213	0.012
17	de acordo com	213	0.012
18	de contos ,	212	0.012
19	, em que	211	0.012
20	, no entanto	192	0.011
21	do ano passado	186	0.010
22	, em lisboa	179	0.010
23	o facto de	179	0.010
24	ano passado ,	177	0.010
25	por cento do	176	0.010
26	, apesar de	174	0.010

Done

Corpógrafo - smatos21 - Mozilla Firefox

# Corpógrafo

- Concordances
  - Using regular expressions (e.g. “(the|a) \w+ is \w+”)
- N-grams
  - Identification of multi-word expressions / lexical bundles
- Term extraction
  - Based on n-grams (using filters / stop-words)

# Corpógrafo

Applications Places System Sérgio Matos Mon 21 Apr, 17:04

Corpógrafo - smatos21 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Corpógrafo - smatos21 Administração do Corpógrafo

Corpógrafo - Linguateca Base de dados activa: teste copia

**Candidatos a Termo**

Corpus: Natura-PUBLICO  
Tamanho: 1726679 Tokens  
BD Terminológica: teste copia

Candidatos encontrados: 74478  
Apresentados candidatos de 1 a 200

[ << ] [ < ] [ > ] [ >> ]

Candidatos encontrados: 74478  
Apresentados candidatos de 1 a 200

#	candidato a termo	#	OPM	Inserir!
1	<a href="#">ano passado</a>	240	138.99	<input type="checkbox"/>
2	<a href="#">próximo dia</a>	182	105.40	<input type="checkbox"/>
3	<a href="#">fc porto</a>	157	90.92	<input type="checkbox"/>
4	<a href="#">semana passada</a>	123	71.23	<input type="checkbox"/>
5	<a href="#">guerra do golfo</a>	111	64.28	<input type="checkbox"/>
6	<a href="#">Médio Oriente</a>	84	48.64	<input type="checkbox"/>
7	<a href="#">passado dia</a>	80	46.33	<input type="checkbox"/>
8	<a href="#">conferência de imprensa</a>	75	43.43	<input type="checkbox"/>
9	<a href="#">presidente da república</a>	71	41.11	<input type="checkbox"/>
10	<a href="#">Assembleia da República</a>	59	34.16	<input type="checkbox"/>
11	<a href="#">Arábia Saudita</a>	59	34.16	<input type="checkbox"/>
12	<a href="#">união soviética</a>	57	33.01	<input type="checkbox"/>

Done

Corpógrafo - smatos21 - Mozilla Firefox

# Corpógrafo

- Concordances
  - Using regular expressions (e.g. “(the|a) \w+ is \w+”)
- N-grams
  - Identification of multi-word expressions / lexical bundles
- Term extraction
  - Based on n-grams (using filters / stop-words)
- Extraction of definitions / semantic relations
  - Pattern-based

# Corpógrafo

- Extraction of Definitions

Example patterns:

```
(^|the |a |an | the | a | an )__TERMO__ is (a|an|the) \w+ \w+ of the
(^|the |a |an | the | a | an )__TERMO__ is (a|an|the) \w+ of
(^|the |a |an | the | a | an )__TERMO__ is \w+ with
(^|the |a |an | the | a | an )__TERMO__ is made of
(^|the |a |an | the | a | an )__TERMO__ is made up of
(^|the |a |an | the | a | an )__TERMO__ is the main
(^|the |a |an | the | a | an )__TERMO__ is (that|the) part of
(^|the |a |an | the | a | an )__TERMO__ is (part|composed) of
```



# Corpógrafo

Corpógrafo - neurodemo - Windows Internet Explorer

http://polodup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

File Edit View Favorites Tools Help

Google G Go Bookmarks PageRank 79 blocked Check AutoLink AutoFill Send to Settings

Corpógrafo - neurodemo

### Corpógrafo - Linguateca

- Gestor
- Pesquisa
- Centro de Conhecimento
  - BD Terminológicas**
  - Gestor de Relações
  - BD Fraseológicas
- Centro de Comunicação

#### Definições do Termo

- \* [Listar Definições](#)
- \* [Inserir ou Pesquisar](#)
- \* [Ficha do Termo](#)

#### Informações

## neuron

Pesquisa efectuada sobre o corpus **Neurons**.  
Definições propostas [19]:

The neuron is the functional unit of the nervous system .	Guardar
<b>Padrão de suporte:</b> (^ the  ^a  ^an   the   a   an )neuron is (a an the) \w+ \w+ of the	
<b>Ficheiro de origem:</b> EN_neouvssystem.htm (The Nervous System)	

A neuron is the functioning unit of the nervous system ; specialized to receive , integrate , and transmit information .	Guardar
<b>Padrão de suporte:</b> (^ the  ^a  ^an   the   a   an )neuron is (a an the) \w+ \w+ of the	
<b>Ficheiro de origem:</b> EN_StructureoftheNeuron-Introduction.htm (Structure of the Neuron - Introduction)	

A neuron has many of the same " organelles " , such as mitochondria , cytoplasm and a nucleus , as other cells in the body .	Guardar
<b>Padrão de suporte:</b> (^ the  ^a  ^an   the   a   an )neuron has (many  several  a  an  all  one  two  various  the )	
<b>Ficheiro de origem:</b> EN_Neurons.htm (Neurons)	

The Myelin Sheath of a neuron consists of fat-containing cells that insulate the axon from electrical activity .	Guardar
<b>Padrão de suporte:</b> (^ the  ^a  ^an  ^each   the   a   an   each )neuron (results  releases  prevents  holds  dominates  causes  originates  represents  controls  contains  conducts  defines  spreads  influences  determines  produces  becomes  presents  occurs  describes  regulates  acts  enables  equals  brings  consists  corresponds  leads  depends  takes  receives  derives  plays )	
<b>Ficheiro de origem:</b> EN_StructureoftheNeuron-MyelinSheath.htm (Structure of the Neuron - Myelin Sheath)	

The Soma ( Latin , som a : body ) or cell body of a neuron contains the nucleus and other structures common to living cells .	
<b>Padrão de suporte:</b> (^ the  ^a  ^an  ^each   the   a   an   each )neuron (results  releases  prevents  holds  dominates	

Done Internet 100%

start Corpógrafo - neurod... EN 21:48

# Corpógrafo

- Extraction of Semantic Relations

Example patterns for holonymy / meronymy:

```
(^|a |an |the | a | an | the )__TERMO2__ is located in the __TERMO1__
(^|a |an |the | a | an | the )__TERMO1__ is (composed|made up) of .* : the .* ,
(the|a|an) __TERMO2__
(^|a |an |the | a | an | the )__TERMO1__ is (composed|made up) of .* :
(the|a|an) __TERMO2__
(^|a |an |the | a | an | the )__TERMO1__ which is composed of \w+ __TERMO2__
(^|a |an |the | a | an | the )__TERMO2__ is the \w+ of the __TERMO1__
(^|a |an |the | a | an | the )__TERMO2__ , the \w+ part of the __TERMO1__
(^|a |an |the | a | an | the )__TERMO1__ contains the \w+ , __TERMO2__
(^|a |an |the | a | an | the )__TERMO1__ consists of (the|a|an) __TERMO2__
```

# Corpógrafo

Corpógrafo - neurodemo - Windows Internet Explorer

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

File Edit View Favorites Tools Help

Google G Go Bookmarks PageRank 79 blocked Check AutoLink AutoFill Send to Settings

Corpógrafo - neurodemo

### Corpógrafo - Linguateca

- Gestor
- Pesquisa
- Centro de Conhecimento
  - BD Terminológicas
  - Gestor de Relações
  - BD Fraseológicas
- Centro de Comunicação

#### Pesquisar relações

- [\\* Usando restrições lexicais](#)
- [\\* Definindo manualmente](#)

#### Informações

**neuron é holónimo de (composto por) ...**

<b>dendrites</b> ["neuron" já está definido como <i>holónimo de (composto por)</i> "dendrites"]	Guardar!
Thus , the <b>dendrites</b> of a neuron provide a surface for receiving synaptic inputs from other neurons .	
Padrão de suporte: (^ ^the   the )dendrites of (the a an) neuron	
The arbor formed by the <b>dendrites</b> of a neuron often has a characteristic shape as determined by the spatial domains into which the <b>dendrites</b> ramify .	
Padrão de suporte: (^ ^the   the )dendrites of (the a an) neuron	
The axon is considerably thicker and longer than the <b>dendrites</b> of a neuron .	
Padrão de suporte: (^ ^the   the )dendrites of (the a an) neuron	
Thus , the <b>dendrites</b> of a neuron provide a surface for receiving synaptic inputs from other neurons .	
Padrão de suporte: (^ ^the   the )dendrites of (the a an) neuron	
The arbor formed by the <b>dendrites</b> of a neuron often has a characteristic shape as determined by the spatial domains into which the <b>dendrites</b> ramify .	
Padrão de suporte: (^ ^the   the )dendrites of (the a an) neuron	
The axon is considerably thicker and longer than the <b>dendrites</b> of a neuron .	
Padrão de suporte: (^ ^the   the )dendrites of (the a an) neuron	
One neuron usually has several <b>dendrites</b> , and these branch out in a tree-like fashion .	
Padrão de suporte: (^ ^a  ^an  ^the  ^one   a   an   the   one )neuron (typically usually normally) has (several many various) dendrites	
A neuron typically has many <b>dendrites</b> and one axon .	
Padrão de suporte: (^ ^a  ^an  ^the  ^one   a   an   the   one )neuron (typically usually normally) has (several many various) dendrites	

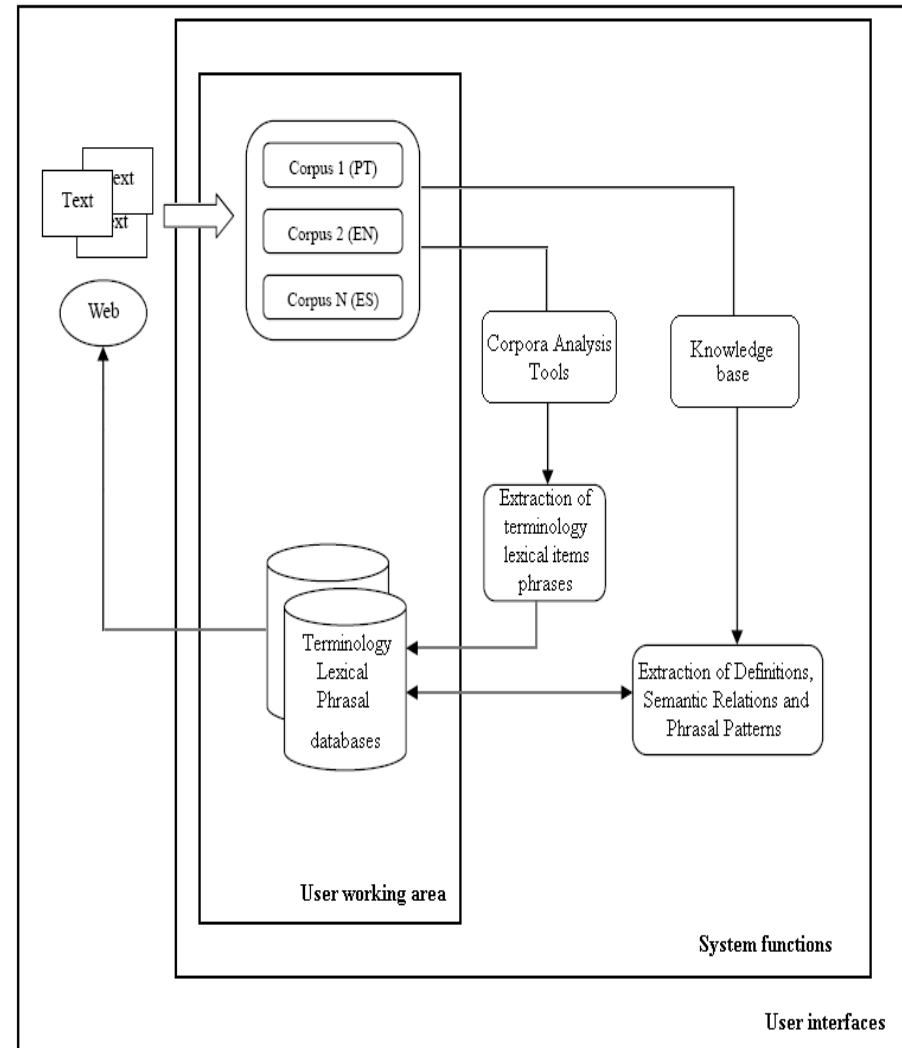
<b>axon</b> ["neuron" já está definido como <i>holónimo de (composto por)</i> "axon"]	Guardar!
However , each neuron has only one <b>axon</b> ( examples : spinal motor neurons , pyramidal neurons , Purkinje cells ) .	

Internet 100%

start Corpógrafo - neurod... EN 21:53

# Corpógrafo

- Aligned concordances
  - Parallel Corpora
  - Comparable Corpora
- Phraseology / Lexicography
  - Identification of phrases / lexi
  - Discourse analysis
  - Genre analysis



# Corpógrafo

- Parallel Corpora

The screenshot displays the Corpógrafo web application interface, which is used for managing and aligning parallel corpora. The interface is divided into several sections:

- Top Navigation:** Includes a menu with options like 'Gestor', 'Ficheiros', 'Corpora', 'Autores', 'Organizações', 'Repositório multimédia', 'Pesquisador', 'Corpora paralelos', 'Pesquisa', 'Centro de Conhecimento', and 'Centro de Comunicação'.
- Left Sidebar:** Contains a tree view of the application structure and a 'Ficheiros alinhados' section with a 'Voltar' button.
- Main Content Area:**
  - Alinhar ficheiros:** A table showing the process of aligning files from two corpora. The files are identified by their IDs (e.g., FXAC07306ENC\_002.pdf) and are grouped under the heading 'FICHEIROS \ (Tratado de Lisboa)'. A yellow 'Alinhar' button is visible below the table.
  - Resultado do alinhamento:** A section showing the results of the alignment process. It indicates that 3300 phrases were aligned and the first 50 are being displayed. Below this, a table shows the aligned text for three specific file pairs:
- Right Panel:** Contains a 'Ficheiros alinhados' section with a 'Novo alinhamento' button and an 'Informações' section.

#	Ficheiro origem	Ficheiro destino	Ver / Editar	Eliminar
1	FXAC07306PTC_002.pdf	FXAC07306ENC_002.pdf	<a href="#">Ver</a>	<a href="#">Eliminar</a>

#	FXAC07306PTC_002.pdf	FXAC07306ENC_002.pdf
1	ISSN 1725-2482 C 306Jornal Oficial da União Europeia Número de informação Índice Página 2007/C 306/01 Tratado de Lisboa que altera o Tratado da União Europeia e o Tratado que institui a Comunidade Europeia , assinado em Lisboa em 13 de Dezembro de 2007...	ISSN 1725-2423 C 306Official Journal of the European Union Notice No Contents Page 2007/C 306/01 Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community , signed at Lisbon , 13 December 2007...
2	1 ALTERAÇÕES INTRODUZIDAS NO TRATADO DA UNIÃO EUROPEIA E NO TRATADO QUE INSTITUI A COMUNIDADE EUROPEIA...	1 AMENDMENTS TO THE TREATY ON EUROPEAN UNION AND TO THE TREATY ESTABLISHING THE EUROPEAN COMMUNITY... 10 Article 1...
3		10 Article 2...

# Integrating NooJ

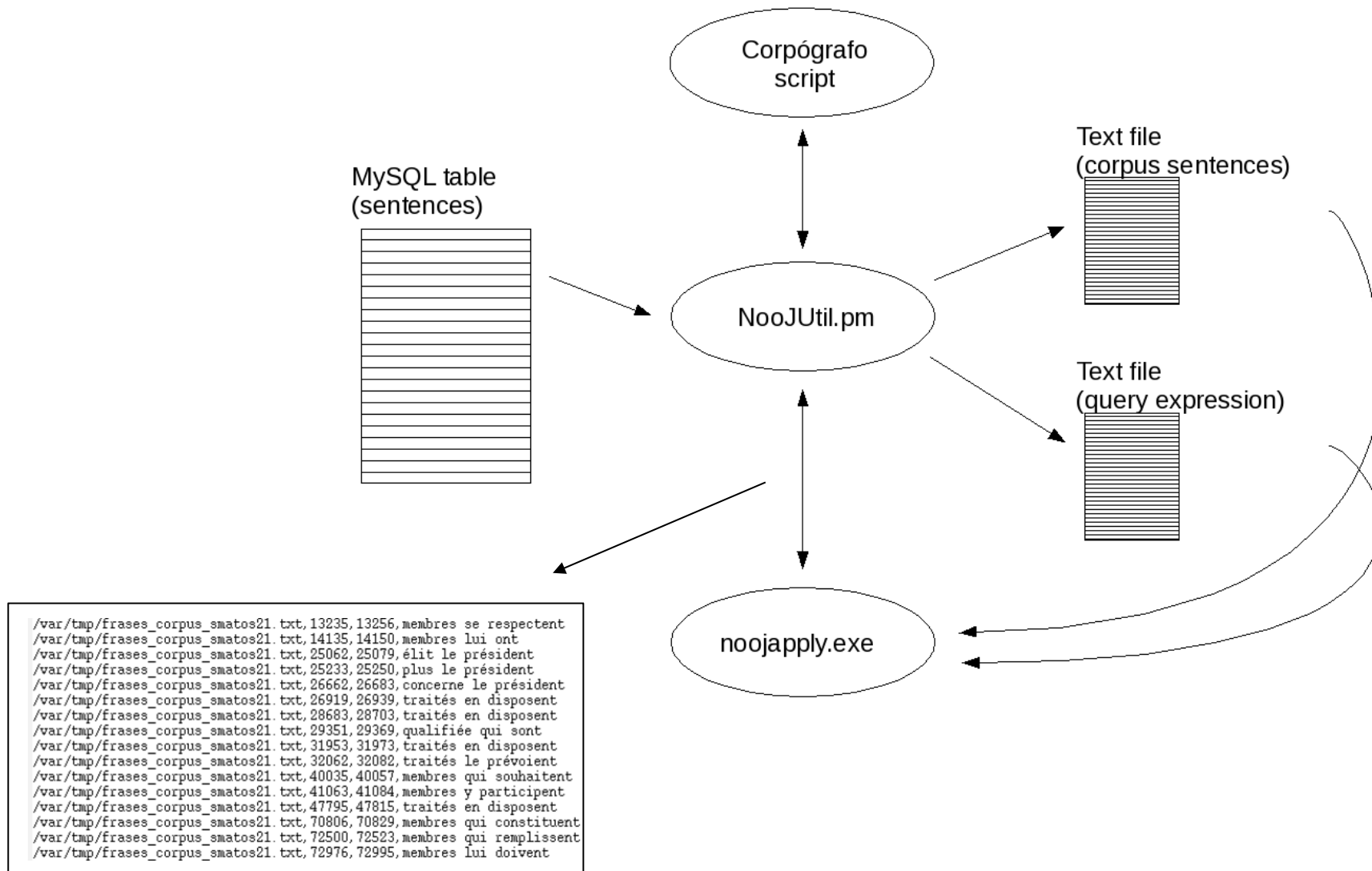
- Motivation
  - Integrate POS information for obtaining concordances and extracting lexical items / phrases
  - Syntactic and semantic analysis
- Why NooJ?
  - Comprehensive corpus processing engine
  - Extensible through the use of lexical and syntactic grammars
  - Linguistic resources available for various languages

# Integrating NooJ

- Practical aspects
  - Executable in Linux under the Mono platform (<http://www.mono-project.com/>)
  - Command-line programme (nooapply.exe)
  - Communication through text files
- Implementation
  - Perl module (NooJUtil.pm)
    - **init**: write corpus sentences to file; select dictionary
    - **query**: obtain NooJ concordances with nooapply.exe
    - **concordance**: obtain results and convert to Corpógrafo's concordance format

# Integrating NooJ

- NooJUtil module





# Integrating NooJ

- Converting results

Select corresponding sentence

```
/var/tmp/frases_corpus_smatos21.txt,13235,13256,membres se respectent
/var/tmp/frases_corpus_smatos21.txt,14135,14150,membres lui ont
/var/tmp/frases_corpus_smatos21.txt,25062,25079,élit le président
/var/tmp/frases_corpus_smatos21.txt,25233,25250,plus le président
/var/tmp/frases_corpus_smatos21.txt,26662,26683,concerne le président
/var/tmp/frases_corpus_smatos21.txt,26919,26939,traités en disposent
/var/tmp/frases_corpus_smatos21.txt,28683,28703,traités en disposent
/var/tmp/frases_corpus_smatos21.txt,29351,29369,qualifiée qui sont
/var/tmp/frases_corpus_smatos21.txt,31953,31973,traités en disposent
/var/tmp/frases_corpus_smatos21.txt,32062,32082,traités le prévoient
/var/tmp/frases_corpus_smatos21.txt,40035,40057,membres qui souhaitent
/var/tmp/frases_corpus_smatos21.txt,41063,41084,membres y participent
/var/tmp/frases_corpus_smatos21.txt,47795,47815,traités en disposent
/var/tmp/frases_corpus_smatos21.txt,70806,70829,membres qui constituent
/var/tmp/frases_corpus_smatos21.txt,72500,72523,membres qui remplissent
/var/tmp/frases_corpus_smatos21.txt,72976,72995,membres lui doivent
```

Concordancia.pm

- left and right context
- origin file
- occurrences

# Using NooJ Expressions in Corpógrafo

Applications Places System Sérgio Matos Fri 18 Apr, 11:52

Corpógrafo - smatos21 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Corpógrafo - Linguateca

- Gestor
- Pesquisa
  - Concordância Frase
  - Estudo de N-Gramas
  - Concordância Janela
  - Concordância KWIC**
  - Concordâncias paralelas
- Centro de Conhecimento
- Centro de Comunicação

Informações

### Concordância KWIC

- Selecione corpus a pesquisar:
- Introduza expressão de pesquisa:
- Contexto de visualização
  - palavras
  - caracteres

Done

Corpógrafo - smatos21 - Mozilla Firefox

# Using NooJ Expressions in Corpógrafo

Applications Places System Sérgio Matos Wed 16 Apr, 14:43

Corpógrafo - smatos21 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

**Corpógrafo - Linguateca**

- Gestor
- Pesquisa
  - Concordância Frase
  - Estudo de N-Gramas
  - Concordância Janela
  - Concordância KWIC**
  - Concordâncias paralelas
- Centro de Conhecimento
- Centro de Comunicação

**Pesquisa em janela**

- \* [Voltar ao menu](#)
- \* [Contexto Frase](#)
- \* [Contexto Janela 1-1](#)
- \* [Contexto Janela 2-2](#)

**Imprimir!**

**Informações**

Expressão de pesquisa: <V><PRO><V+PR+3+p>  
 Corpus pesquisado: Traité de Lisbonne (103899 átomos : 1 ficheiros)

Nº de concordâncias obtidas: 48 distribuídas por 1 ficheiros  
 Frequência da totalidade das instâncias da concordância: 0.04 %

[||<](#) [<<](#) [>>](#) [>>||](#)

#	POS (C/P)	esq.	conc.	dir.	nº frase	info
1	67 / 14	et les États	<b>membres se respectent</b>	et s' assistent	61	
2	106 / 21	que les États	<b>membres lui ont</b>	attribuées dans les	69	
3	4 / 2	Il	<b>élit le président</b>	de la Commission	142	
4	49 / 10	cent cinquante ,	<b>plus le président</b>	.	145	
5	139 / 28	en ce qui	<b>concerne le président</b>	de la Commission	162	
6	73 / 15	cas où les	<b>traités en disposent</b>	autrement .	165	
7	69 / 15	cas où les	<b>traités en disposent</b>	autrement .	180	
8	71 / 11	de la majorité	<b>qualifiée qui sont</b>	applicables jusqu' au	186	
9	115 / 24	cas où les	<b>traités en disposent</b>	autrement .	210	
10	75 / 13	Commission lorsque les	<b>traités le prévoient</b>	.	211	
11	10 / 3	Les États	<b>membres qui souhaitent</b>	instaurer entre eux	274	
12	290 / 49	moins neuf États	<b>membres y participent</b>	.	278	
13	130 / 28	cas où les	<b>traités en disposent</b>	autrement .	313	
14	10 / 3	Les États	<b>membres qui constituent</b>	entre eux des	432	
15	10 / 3	Les États	<b>membres qui remplissent</b>	des critères plus	441	
16	102 / 21	les autres États	<b>membres lui doivent</b>	aide et assistance	443	
17	10 / 3	Les États	<b>membres qui participent</b>	à la rédaction	457	

Done

Corpógrafo - smatos21 - Mozilla Firefox

# Using NooJ Expressions in Corpógrafo

Applications Places Desktop

Corpógrafo - RTraducao - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Corpógrafo - Linguateca

Gestor

Ficheiros

Corpora

Autores

Organizações

Repositório multimédia

Pesquisador

Corpora paralelos

Pesquisa

Centro de Conhecimento

Centro de Comunicação

Pesquisar corpora alinhados

\* Voltar

Informações

Done

Corpógrafo - RTraducao - Mozilla Firefox

## Pesquisar corpora paralelos

Resultado da pesquisa por: <ADV+MP="mente\$"> <WF>

Resultados encontrados: 180  
Apresentando resultados de 1 a 50

[ << ] [ < ] [ > ] [ >> ]

#	Texto original	Texto alinhado
1	Em virtude do princípio da cooperação leal , a União e os Estados-Membros respeitam-se e assistem-se <b>mutuamente</b> no cumprimento das missões decorrentes dos Tratados .	Pursuant to the principle of sincere cooperation , the Union and the Member States shall , in full mutual respect , assist each other in carrying out tasks which flow from the Treaties .
2	Em virtude do princípio da subsidiariedade , nos domínios que não sejam da sua competência exclusiva , a União intervém apenas se e na medida em que os objectivos da acção considerada não possam ser <b>suficientemente alcançados</b> pelos Estados-Membros , tanto ao nível central como ao nível regional e local , podendo contudo , devido às dimensões ou aos efeitos da acção considerada , ser mais bem alcançados ao nível da União .	Under the principle of subsidiarity , in areas which do not fall within its exclusive competence , the Union shall act only if and insofar as the objectives of the proposed action cannot be sufficiently achieved by the Member States , either at central level or at regional and local level , but can rather , by reason of the scale or effects of the proposed action , be better achieved at Union level .
3	Os cidadãos estão <b>directamente representados</b> , ao nível da União , no Parlamento Europeu .	Citizens are directly represented at Union level in the European Parliament .
4	Os Estados-Membros estão representados no Conselho Europeu pelo respectivo Chefe de Estado ou de Governo e no Conselho pelos respectivos Governos , eles próprios <b>democraticamente responsáveis</b> , quer perante os respectivos Paramentos nacionais , quer perante os seus cidadãos .	Member States are represented in the European Council by their Heads of State or Government and in the Council by their governments , themselves democratically accountable either to their national Parliaments , or to their citizens .
5	As instituições , recorrendo aos meios adequados , dão aos cidadãos e às associações representativas a possibilidade de expressarem e partilharem <b>publicamente os</b> seus pontos de vista sobre todos os domínios de acção da União .	The institutions shall , by appropriate means , give citizens and representative associations the opportunity to make known and publicly exchange their views in all areas of Union action . 2 .
6	Artigo 8.º -C Os Paramentos nacionais contribuem <b>activamente para</b> o bom funcionamento da União : a ) Sendo informados pelas instituições da União e notificados dos projectos de actos legislativos da União , de acordo com o Protocolo relativo ao papel dos Paramentos nacionais na União Europeia; b ) Garantindo o respeito pelo princípio da subsidiariedade de acordo com os procedimentos previstos	Article 8 C National Parliaments contribute actively to the good functioning of the Union : ( a ) through being informed by the institutions of the Union and having draft legislative acts of the Union forwarded to them in accordance with the Protocol on the role of national Parliaments in the European Union; ( b ) by seeing to it that the principle of subsidiarity is respected in accordance with the procedures provided for in the Protocol

# Using NooJ Expressions in Corpógrafo

Applications Places Desktop Fri Mar 7, 5:27 PM

Corpógrafo - RTraducao - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Corpógrafo - Linguateca Base de dados activa: BD Fraseológica

Gestor  
Pesquisa  
Centro de Conhecimento  
BD Terminológicas  
Gestor de Relações  
BD Fraseológicas  
Centro de Comunicação

**Resultado da Pesquisa**

- \* Voltar...
- \* Menu principal da BD
- \* Listar e editar Itens Lexicais

**Informações**

Base de dados activa: BD Fraseológica

**Candidatos a Item Lexical**

Corpus: Seguros  
Tamanho: 20734 Tokens  
BD Fraseológica: **BD Fraseológica**

Expressão de pesquisa: <ADV> <WF>

Candidatos encontrados: 452  
Apresentados candidatos de 1 a 200

[ << ] [ < ] [ > ]

Candidatos encontrados: 452  
Apresentados candidatos de 1 a 200

#	candidato a item lexical	#	OPM	Inserir
1	<a href="#">Atualmente a</a>	1	48.22	<input type="checkbox"/>
2	<a href="#">agora confrontadas</a>	1	48.22	<input type="checkbox"/>
3	<a href="#">ainda as</a>	1	48.22	<input type="checkbox"/>
4	<a href="#">ainda definir</a>	1	48.22	<input type="checkbox"/>
5	<a href="#">ainda mais</a>	1	48.22	<input type="checkbox"/>
6	<a href="#">amplamente melhorada</a>	1	48.22	<input type="checkbox"/>
7	<a href="#">apenas destinado</a>	1	48.22	<input type="checkbox"/>
8	<a href="#">Apenas deve</a>	1	48.22	<input type="checkbox"/>
9	<a href="#">apenas dos</a>	1	48.22	<input type="checkbox"/>
10	<a href="#">apenas necessitam</a>	1	48.22	<input type="checkbox"/>
11	<a href="#">apenas passa</a>	1	48.22	<input type="checkbox"/>
12	<a href="#">Apenas sábados</a>	2	96.45	<input type="checkbox"/>
13	<a href="#">apenas se</a>	2	96.45	<input type="checkbox"/>
14	<a href="#">automaticamente ajusta</a>	1	48.22	<input type="checkbox"/>
15	<a href="#">automaticamente em</a>	1	48.22	<input type="checkbox"/>
16	<a href="#">automaticamente na</a>	1	48.22	<input type="checkbox"/>
17	<a href="#">automaticamente para</a>	1	48.22	<input type="checkbox"/>
18	<a href="#">automaticamente pelo</a>	1	48.22	<input type="checkbox"/>
19	<a href="#">automaticamente todas</a>	1	48.22	<input type="checkbox"/>
20	<a href="#">Basicamente são</a>	1	48.22	<input type="checkbox"/>
21	<a href="#">bem como</a>	4	144.68	<input type="checkbox"/>
22	<a href="#">cada ambiente</a>	1	48.22	<input type="checkbox"/>
23	<a href="#">cada apólice</a>	2	96.45	<input type="checkbox"/>

http://localhost - Concordâncias de "bem como" - Mozilla Firefox

Pesquisa de contexto para o candidato a termo: **bem como**  
Corpus pesquisado: Seguros (20734 átomos)  
Ocorrências encontradas: 4  
Ocorrências por milhão no corpus (OCM): 192.910

Listando resultados...

#	concordância no contexto de uma frase
1	Criação de novos registos no protocolo Está disponível o tratamento via protocolo de valores de balanço por apólice <b>bem como</b> as cargas da apólice .
2	Assim , a solução GIS é constituída por um conjunto de módulos aplicativos que suportam todas as actividades fundamentais de qualquer interveniente da actividade seguradora , <b>bem como</b> permite facilmente interligar e compatibilizar os sistema de informação dos diferentes intervenientes .
3	Através do módulo de configuração é possível definir as múltiplas informações requeridas nos questionários dos processos de sinistro de cada ramo/tipo de seguro <b>bem como</b> as regras de tratamento associadas ( provisões de sinistros , controlo de parcelas por natureza , agravamento/bónus do prémio , regras de aplicação de franquias ou participações , etc. )
4	Todo o funcionamento do sistema é baseado numa configuração de produtos , que depois é interpretada pelos componentes de negócio , de forma a disponibilizar o conjunto de funcionalidades e serviços associados ao produto , <b>bem como</b> a sua total integração com o sistema de gestão de back office , implementando assim uma arquitectura em três camadas .

[ Fechar ]

Done

Corpógrafo - RTraducao - Mozi... http://localhost - Concordância...



# Using NooJ Expressions in Corpógrafo

Applications Places System Sérgio Matos Fri 18 Apr, 11:37

Corpógrafo - neurodemo - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Corpógrafo - Linguateca

Expressão de pesquisa: <DET> (neuron + neurons) <V+PR>  
Corpus pesquisado: Neurons (29181 átomos : 37 ficheiros)

Nº de concordâncias obtidas: 39 distribuídas por 16 ficheiros  
Frequência da totalidade das instâncias da concordância: 0.13 %

Navigation icons: << >> >>> >>>>

#	f	Frase onde ocorre a concordância	Info
1	1	The Axon of <b>a neuron</b> is a singular fiber that carries information away from the soma to the synaptic sites of other neurons ( dendrites and somas ) , muscles , or glands .	Info
2	1	The Dendrites ( Greek , dendro : tree ) of <b>a neuron are</b> its many short , branching fibers extending from the cell body or soma .	Info
3	1	<b>A neuron is</b> the functioning unit of the nervous system ; specialized to receive , integrate , and transmit information .	Info
4	1	The Myelin Sheath of <b>a neuron consists</b> of fat-containing cells that insulate the axon from electrical activity .	Info
5	1	The Soma ( Latin , som a : body ) or cell body of <b>a neuron contains</b> the nucleus and other structures common to living cells .	Info
6	1	The Terminal Buttons of <b>a neuron are</b> the small knobs at the end of an axon that release chemicals called neurotransmitters .	Info
7	1	<b>A neuron is</b> a cell specialized to conduct electrochemical impulses called nerve impulses or action potentials .	Info
8	1	This page will describe the way <b>that neurons work</b> .	Info
9	3	It is at these junctions <b>that neurons are</b> excited , inhibited , or modulated .	Info
10	4	<b>Many neurons do</b> not have axons .	Info
11	5	Information from <b>one neuron flows</b> to another neuron across a synapse .	Info
12	6	<b>The neuron is</b> the functional unit of the nervous system .	Info
13	6	The dendritic branching of <b>a neuron is</b> changeable in the mature nervous system , either growing or retracting .	Info
14	7	The axons of <b>many neurons are</b> wrapped in a myelin sheat , which is composed of the membranes of interstitial cells and is wrapped around the axons to form several concentric layers .	Info
15	8	While variable in size and shape , <b>all neurons have</b> three parts .	Info
16	9	Thus , the dendrites of <b>a neuron provide</b> a surface for receiving synaptic inputs from other neurons .	Info
17	12	Dendrites of <b>some neurons are</b> smooth , tapered processes , such as in motor neurons of the spinal cord .	Info

Done

Corpógrafo - neurodemo - Mozilla Firefox

# Using NooJ Expressions in Corpógrafo

The screenshot shows the Corpógrafo web application in a Mozilla Firefox browser. The browser's address bar displays the URL: `http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl`. The application's title bar reads "Corpógrafo - neurodemo - Mozilla Firefox".

The main content area displays the search results for the expression: `<DET> <neurone> <V+PR>`. The corpus searched is "Le Neurone" (22748 atoms, 12 files). It reports 19 concordances found across 6 files, with a frequency of 0.08%.

Navigation controls include: `[<] [ << ] [ >> ] [ > ]`.

#	f	Frase onde ocorre a concordância	Info
1	1	<b>Les neurones reçoivent</b> simultanément des milliers d'informations activant des récepteurs-canaux et/ou des récepteurs liés aux protéines G .	
2	6	Ceci est basé sur l'éloignement entre la cellule endocrinienne et sa cible et le fait que le message emprunte la voie systémique , alors que <b>les neurones sont</b> en contact quasi direct par l'intermédiaire des synapses .	
3	9	La sécrétion , très focalisée et dirigée uniquement vers les cellules avec lesquelles <b>le neurone est</b> connecté , se fait au niveau des synapses .	
4	10	Les dendrites de <b>certains neurones contiennent</b> un nombre important de canaux sodiques et/ou calciques sensibles au potentiel .	
5	10	<b>Le neurone diffère</b> ainsi ( à quelques exceptions près ) des autres cellules sécrétrices de l'organisme comme les cellules hormonales , qui libèrent leur produit de sécrétion dans la circulation sanguine ( sécrétion endocrine ) .	
6	11	<b>Les neurones ont</b> des formes très diverses et leur morphologie ( dimension et orientation des champs dendritiques , taille du soma ) joue un rôle fondamental dans leurs capacités d'intégration et de traitement de l'information .	
7	12	<b>Les neurones sont</b> amitotiques ( incapables de se reproduire ) et ont donc une longévité extrême .	
8	15	<b>LES NEURONES ONT UN CORPS CELLULAIRE D' OU PARTENT DEUX TYPES DE PROLONGEMENTS : LES DENDRITES ET L'AXONE</b> L'ARBRE SOMATO-DENDRITIQUE REPRÉSENTE LE PÔLE RÉCEPTEUR DU NEURONE	
9	19	<b>Le neurone comprend :</b>	
10	20	<b>Les neurones effectuent</b> donc des opérations complexes nécessitant la sommation de tous les PPSE pour produire une dépolarisation postsynaptique significative : c'est l'intégration synaptique .	
11	21	2 . <b>Le neurone est</b> une cellule dite polarisée ( asymétrique ) avec d'un côté les dendrites , qui constituent le pôle normalement récepteur .	
12	24	<b>Le neurone est</b> la cellule qui élabore , modifie et transmet l'influx nerveux .	
13	25	<b>Les neurones sont</b> ainsi dits " épineux " ou " lisses " .	
14	28	Lorsqu'ils ont atteint leur position définitive , <b>les neurones acquièrent</b> un cône de croissance , qui s'allonge progressivement pour former l'axone .	
15	39	Entre <b>les neurones sont</b> entremêlés de nombreux neurites .	

The interface also includes a sidebar with navigation options like "Gestor", "Pesquisa", and "Concordâncias", and a footer with "Done" and "Corpógrafo - neurodemo - Mozilla Firefox".

# Using NooJ Expressions in Corpógrafo

Applications Places System Sérgio Matos Fri 18 Apr, 12:09

Corpógrafo - neurodemo - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://poloclup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

Corpógrafo - Linguateca

Expressão de pesquisa: <WF> <V> <WF> (of + for + in) <WF>  
Corpus pesquisado: Neurons (29181 átomos : 37 ficheiros)

Nº de concordâncias obtidas: 323 distribuídas por 31 ficheiros  
Frequência da totalidade das instâncias da concordância: 1.10 %

Concordâncias

- \* Voltar ao menu
- \* Analisar Frequências
- \* Contexto Janela 1-1
- \* Contexto Janela 2-2

Imprimir!

#	f	Frase onde ocorre a concordância	Info
1	0	Activation of the transcription factor NF-[kappa]B in Schwann <b>cells is required for peripheral</b> myelin formation	
2	1	The human body <b>is made up of trillions</b> of cells .	
3	1	Nerve cells - or neurons - are the cells in our body <b>that are responsible for bringing</b> information from the surrounding , analyze this input , and finally create a perception .	
4	1	<b>Dentrites branch out in treelike</b> fashion and serve as the main apparatus for receiving signals from other nerve cells .	
5	1	This site will help you learn about the <b>five major sections of a</b> Nerve cell :	
6	1	A neuron is <b>the functioning unit of the</b> nervous system ; specialized to receive , integrate , and transmit information .	
7	1	The gaps ( approximately 1 micrometer wide ) formed between myelin sheath cells long the axons <b>are called Nodes of Ranvier</b> .	
8	1	Plant cells are eukaryotic <b>and have many of the</b> structures found in animal cells .	
9	1	The brain <b>is made up of more</b> than just nerve cells ( neurons ) .	
10	1	Activation of the transcription factor NF-B in Schwann <b>cells is required for peripheral</b> myelin formation .	
11	2	Nervous <b>tissue is composed of two</b> main cell types : neurons and glial cells .	
12	2	The axon is the <b>main conducting unit of the</b> neuron , capable of conveying electrical signals along distances that range from as short as 0.1 mm to as long as 2 m .	
13	2	In addition , all eukaryotic cells <b>contain elaborate systems of internal</b> membranes which set up various membrane-enclosed compartments within the cell .	
14	2	The energy <b>released is trapped in the</b> form of ATP for use by all the energy-consuming activities of the cell.	
15	2	The membrane with its mosaic of <b>proteins is responsible for many</b> important functions : keeping certain ions and small molecules out of the cell and letting others in , accumulating nutrients , rejecting harmful substances , catalyzing enzymatic reactions , and conducting an impulse being sensitive to particular neurotransmitters and modulators .	
16	2	Cilia and Flagella Cells <b>contain elaborate arrays of protein</b> fibers that serve such functions as : establishing cell shape providing	



# What next?

- Annotations
- Grammars
  - Syntactic / semantic analysis
  - Disambiguation
- Lexical / phrasal items
  - Rules and grammars
- Discourse analysis
  - Discourse markers

# Acknowledgment

- This work was done in the scope of the Linguateca, contract nº339/1.3/C/NAC, project jointly funded by the Portuguese Government and the European Union.

