

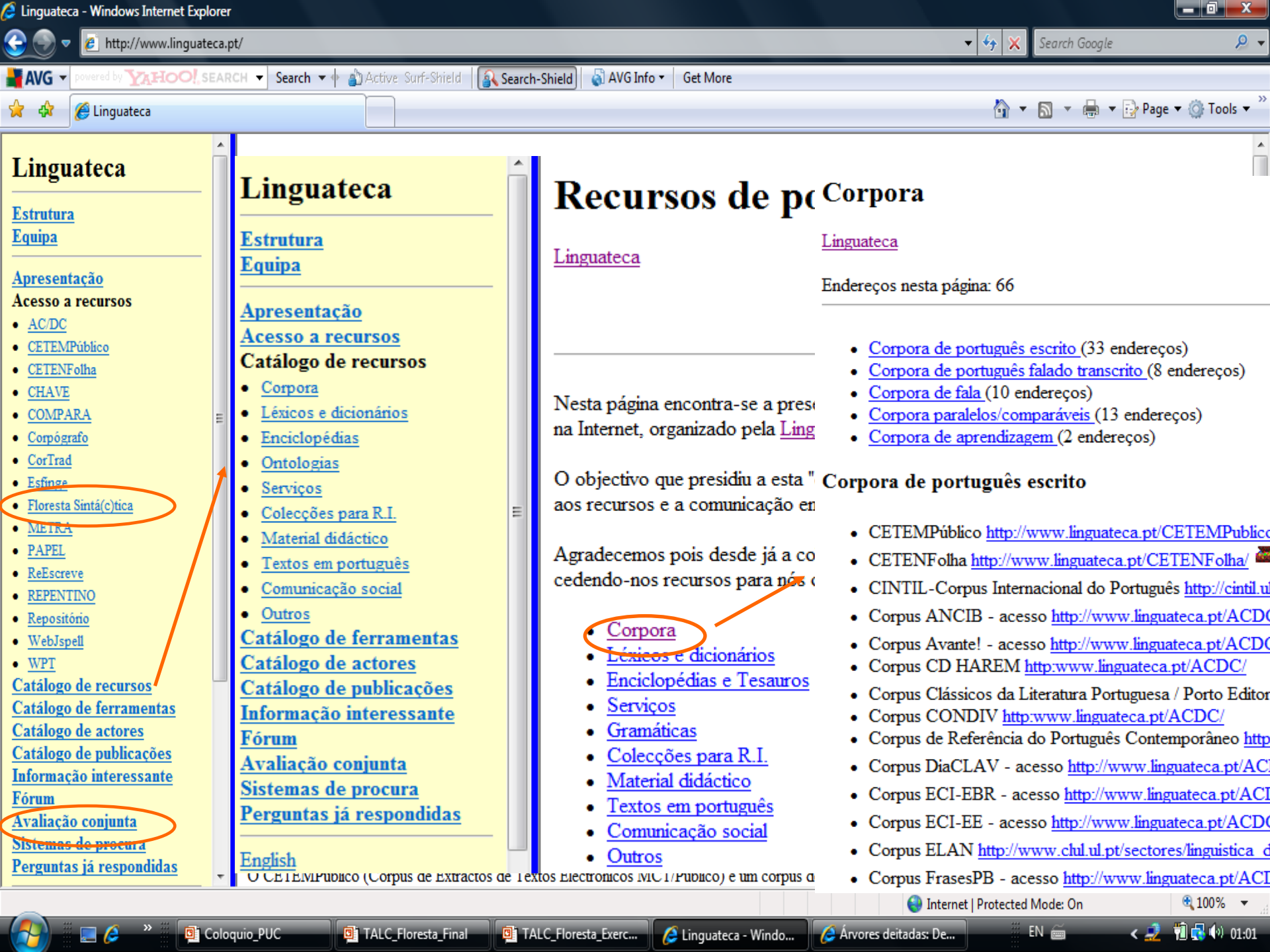
Linguística Computacional, corpus, pesquisa e ensino de língua portuguesa

Cláudia Freitas

PUC-RIO
Abril, 2009

Linguateca (www.linguateca.pt)

- Centro de recursos **distribuído** para o processamento computacional do Português
- Modelo IRA de funcionamento
 - **I**nformação – facilitar acesso a recursos já existentes
 - **R**ecursos – desenvolver os mais necessários
 - **A**valiação – organizar avaliações (comparação de abordagens e métodos aplicados a uma tarefa comum (vs auto-avaliação))



Linguateca

- [Estrutura](#)
- [Equipa](#)
- [Apresentação](#)
- [Acesso a recursos](#)
- [AC/DC](#)
- [CETEMPúblico](#)
- [CETENFolha](#)
- [CHAVE](#)
- [COMPARA](#)
- [Corpógrafo](#)
- [CorTrad](#)
- [Esfinge](#)
- [Floresta Sintá\(c\)tica](#)
- [METRA](#)
- [PAPEL](#)
- [ReEscreve](#)
- [REPENTINO](#)
- [Repositório](#)
- [WebJspell](#)
- [WPT](#)
- [Catálogo de recursos](#)
- [Catálogo de ferramentas](#)
- [Catálogo de actores](#)
- [Catálogo de publicações](#)
- [Informação interessante](#)
- [Fórum](#)
- [Avaliação conjunta](#)
- [Sistemas de procura](#)
- [Perguntas já respondidas](#)

Linguateca

- [Estrutura](#)
- [Equipa](#)
- [Apresentação](#)
- [Acesso a recursos](#)
- [Catálogo de recursos](#)
- [Corpora](#)
- [Léxicos e dicionários](#)
- [Enciclopédias](#)
- [Ontologias](#)
- [Serviços](#)
- [Colecções para R.I.](#)
- [Material didáctico](#)
- [Textos em português](#)
- [Comunicação social](#)
- [Outros](#)
- [Catálogo de ferramentas](#)
- [Catálogo de actores](#)
- [Catálogo de publicações](#)
- [Informação interessante](#)
- [Fórum](#)
- [Avaliação conjunta](#)
- [Sistemas de procura](#)
- [Perguntas já respondidas](#)
- [English](#)

Recursos de pc Corpora

[Linguateca](#)

[Linguateca](#)

Endereços nesta página: 66

Nesta página encontra-se a presença na Internet, organizado pela [Ling](#)

- [Corpora de português escrito](#) (33 endereços)
- [Corpora de português falado transcrito](#) (8 endereços)
- [Corpora de fala](#) (10 endereços)
- [Corpora paralelos/comparáveis](#) (13 endereços)
- [Corpora de aprendizagem](#) (2 endereços)

O objectivo que presidiu a esta " aos recursos e a comunicação em

Corpora de português escrito

Agradecemos pois desde já a cedendo-nos recursos para nós

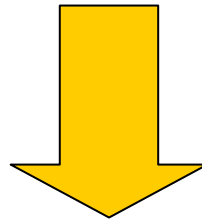
- CETEMPúblico <http://www.linguateca.pt/CETEMPúblico>
- CETENFolha <http://www.linguateca.pt/CETENFolha/>
- CINTIL-Corpus Internacional do Português <http://cintil.ul.pt/>
- Corpus ANCIB - acesso <http://www.linguateca.pt/ACDC/>
- Corpus Avante! - acesso <http://www.linguateca.pt/ACDC/>
- Corpus CD HAREM <http://www.linguateca.pt/ACDC/>
- Corpus Clássicos da Literatura Portuguesa / Porto Editor
- Corpus CONDIV <http://www.linguateca.pt/ACDC/>
- Corpus de Referência do Português Contemporâneo <http://www.linguateca.pt/ACDC/>
- Corpus DiaCLAV - acesso <http://www.linguateca.pt/ACDC/>
- Corpus ECI-EBR - acesso <http://www.linguateca.pt/ACDC/>
- Corpus ECI-EE - acesso <http://www.linguateca.pt/ACDC/>
- Corpus ELAN <http://www.chul.ul.pt/sectores/linguistica/>
- Corpus FrasesPB - acesso <http://www.linguateca.pt/ACDC/>

- [Corpora](#)
- [Léxicos e dicionários](#)
- [Enciclopédias e Tesouros](#)
- [Serviços](#)
- [Gramáticas](#)
- [Colecções para R.I.](#)
- [Material didáctico](#)
- [Textos em português](#)
- [Comunicação social](#)
- [Outros](#)

• CETEMPUBLICO (Corpus de extractos de textos electrónicos MUC1/PUBLICO) e um corpus a

Linguateca

- Processamento de textos em português
- Trabalho interdisciplinar



papel do linguista na criação e na
utilização de recursos

HAREM

Avaliação e Reconhecimento de Entidades Mencionadas

avaliação conjunta:

“modelo de avaliação em que vários grupos comparam, com base num conjunto de tarefas consensuais, o progresso dos seus sistemas numa dada área, usando para isso um **conjunto de recursos** comum e uma métrica consensual.”



HAREM

- EM: “nomes próprios” (“NE: named entities”)
- Por que EM ?

- análise semântica de textos

- Extração de Informação (de textos) , RAP...

- Alto poder informativo

- Instâncias de classes

→ reconhecer (identificar e classificar)
EM segundo categorias pré-definidas

HAREM – escolhas (linguísticas) subjacentes

- Definição / Delimitação de uma EM

- Mal de Alzheimer vs mal de Alzheimer; pastéis de Santa Clara; TV, CD, USB, HD...

- Modelo semântico do HAREM

- EMs classificadas em contexto

- Não há primazia de uma classificação sobre outra
 - Brasil=LOCAL ou BRASIL=EQUIPE

- VAGUEZA

- Mais de uma classificação simultaneamente

HAREM – modelo semântico

□ EM em contexto

Regressou então a **<CATEG="LOCAL">Portugal**, onde iniciou a radiação de origem cósmica, prevista pelo **<CATEG="ABSTRACCAO">meteorológica** **Big Bang **, seria descoberta em 1964...

O acordo político foi obtido durante a Presidência Alemã, tendo cabido a **<CATEG="ORGANIZACAO">Portugal** concluir o processo de revisão.

... pelo qual tem início a expansão das galáxias que os cosmologistas «descrevem como **uma explosão que se dá a** **<CATEG="ABSTRACCAO">Portugal** **Big Bang ** ou dois dentro de si»

<CATEG="PESSOA">Portugal perdeu com a Suíça por 2-0

HAREM – modelo semântico

- Vagueza

Mais de 32 mil pessoas poderiam morrer se uma pandemia de gripe aviária atingisse

`<CATEG="PESSOA|LOCAL" TIPO="POVO|HUMANO">Portugal`

Assim aceitam sacramentos do `<CATEG="ABSTRACCAO|OBRA" TIPO="IDEIA|PLANO">Evangelho `: o Santo Batismo, através do qual....

O carácter diferente da `<CATEG="ABSTRACCAO/ACONTECIMENTO" TIPO="IDEIA/EFEMERIDE">Reforma Inglesa ` deve-se ao facto de ter sido promovida pelas necessidades políticas de Henrique VIII.

HAREM – escolhas (linguísticas) subjacentes

- Classificação semântica

- Escolha das categorias ontológicas

- Granularidade: onde parar?

- Quanto mais categorias, mais vagueza....
- Informações dos textos + interesse dos sistemas participantes

CATEGORIA	TIPO	SUBTIPO
ABSTRACCAO	ESTADO, DISCIPLINA, IDEIA, NOME OUTRO	
ACONTECIMENTO	EFEMERIDE, EVENTO, ORGANIZADO OUTRO	
COISA	CLASSE, MEMBROCLASSE, OBJECTO SUBSTANCIA, OUTRO	
LOCAL	FISICO HUMANO VIRTUAL OUTRO	AGUACURSO, AGUAMASSA, ILHA, PLANETA, REGIAO, RELEVO, OUTRO CONSTRUCAO, DIVISAO, PAIS, REGIAO, RUA, OUTRO COMSOCIAL, OBRA, SITIO, OUTRO
OBRA	ARTE, PLANO, REPRODUZIDA, OUTRO	
ORGANIZACAO	ADMINISTRACAO, EMPRESA, INSTITUICAO, OUTRO	
PESSOA	CARGO, GRUPOCARGO, GRUPOIND, GRUPOMEMBRO, INDIVIDUAL, MEMBRO, POVO, OUTRO	
TEMPO	DURACAO, FREQUENCIA, GENERICO TEMPO_CALEND , OUTRO	DATA, HORA, INTERVALO , OUTRO
VALOR	CLASSIFICACAO, MOEDA, QUANTIDADE, OUTRO	
OUTRO		

HAREM – ideia geral

- Diretivas de anotação (com os participantes (?))
- Criação da CD (e da CH)
 - Anotação
 - Etiqueta(h)rem
 - Revisão da anotação
 - Revisão sequencial dos documentos
 - Revisão por categoria
 - Anotação de casos duvidosos
 - Omissão de discordâncias
 - Metadados
- Criação das métricas de avaliação (com participantes (?))
- Resultados

<DOC DOCID="HAREM-962-05111">

HISTÓRICO Esta seção traz de volta um pouco da longa história do <EM ID="91" CATEG="ORGANIZACAO" TIPO="INSTITUICAO">DCC.

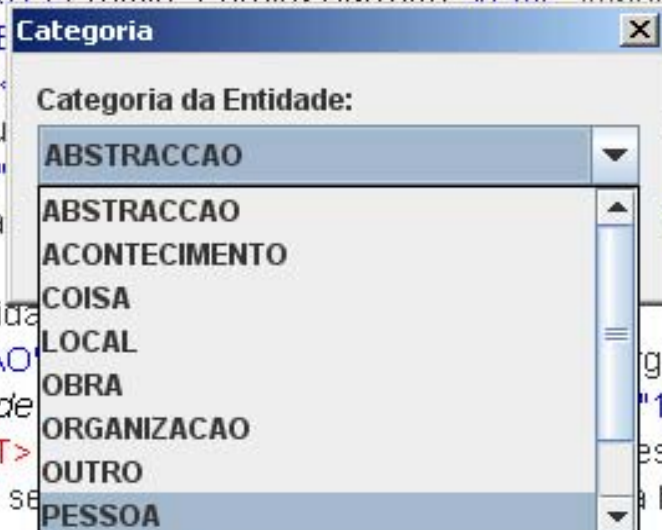
O <EM ID="92" CATEG="ORGANIZACAO" TIPO="INSTITUICAO">DCC- <EM ID="93" CATEG="ORGANIZACAO" TIPO="INSTITUICAO">Departamento de Cultura Científica do Centro Acadêmico Pereira Barretto (<EM ID="94" CATEG="ORGANIZACAO" TIPO="INSTITUICAO">DCC/CAPB), órgão responsável pela representação e encaminhamento científico dos alunos da <EM ID="95" CATEG="ORGANIZACAO" TIPO="INSTITUICAO">UNIFESP/EPM, fundado em <EM ID="96" CATEG="TEMPO" TIPO="DATA">1937, atua junto aos alunos promovendo vários cursos extracurriculares, palestras, conferências e discussões de interesse à área médica.

Organiza o <EM ID="97" CATEG="OUTRO">Prêmio "Pereira Barretto" importante e conceituada premiação anual de trabalhos científicos acadêmicos, o <EM ID="98" CATEG="OUTRO">Congresso Acadêmico Paulista de Medicina (<EM ID="99" CATEG="OUTRO" TIPO="ORGANIZADO">CAPM), cu <EM ID="100" CATEG="ORGANIZACAO" TIPO="INSTITUICAO">Escola Paulista de Medicina estiveram envolvidos na estruturação da <EM ID="101" CATEG="OUTRO" TIPO="ORGANIZADO">Revista científica.

Poucos anos após sua fundação as atividades <EM ID="102" CATEG="ORGANIZACAO" TIPO="INSTITUICAO">Escola Paulista de Medicina começaram. O <EM ID="103" CATEG="TEMPO" TIPO="DATA">cinco de maio de 1937 <EM ID="104" CATEG="TEMPO" TIPO="DATA">1937 marcou o início da história de trabalho, dedicação e abnegação.

Seus primeiros diretores foram os então acadêmicos **Wladimir da Prússia Gomes Ferraz** (Presidente), Arulemo Santos Novaes e Jair Xavier Guimarães (Secretários).

O DCC/CAPB iniciou seus trabalhos coordenando atividades extracurriculares como Cursos, Congresso Acadêmico e instituindo Prêmios Científicos, como o Prêmio Pereira Barreto, e organizando revista científica voltada para a divulgação de trabalhos produzidos por alunos da escola e por integrantes do Hospital São Paulo.



Revista científica.

começou Os alunos sempre

<EM ID="105" CATEG="OUTRO" TIPO="INSTITUICAO">Escola Paulista de Medicina

ge em <ALT><EM ID="103">

"104" CATEG="TEMPO"

espírito virtuoso de acadêmicos que, no

a história de trabalho, dedicação e

HAREM



- CD

- Material para a análise

- 129 DOCS
- 7836 EM
- 633 EM vagas

- Estudo aprofundado das categorias ...

- Categorias omitidas: COMSOC..

- Relação entre categorias/vagueza e tipo de texto...

- Metonímia...

- Outros modelos semânticos (?)

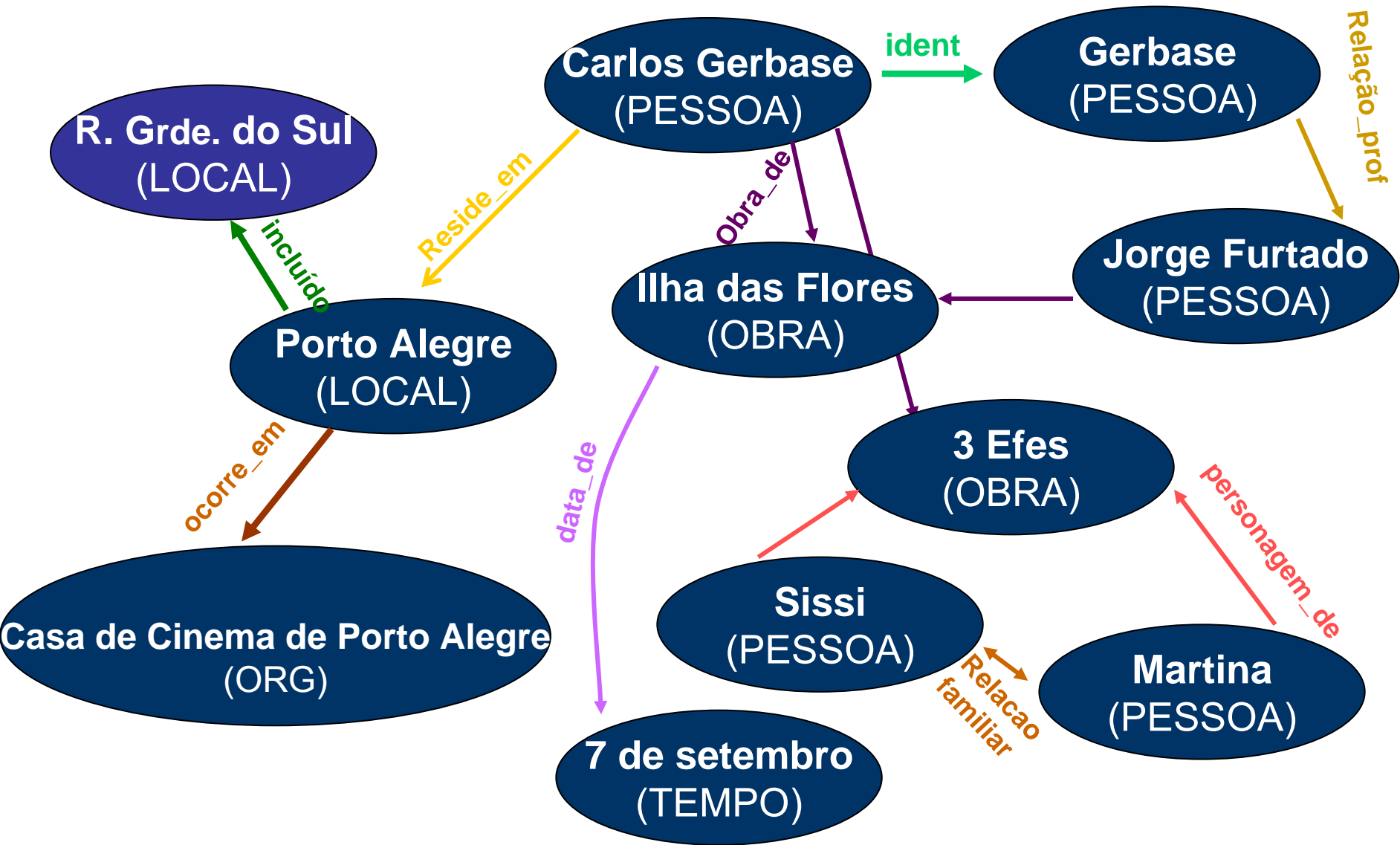


ReReEM

Reconhecimento de Relações entre EM

- Tarefa piloto (2008)
- + análise semântica de textos
- Dependente do HAREM
 - EM
 - Relações estabelecidas em contexto
 - Vagueza

ReReIEM





ReReIEM

HAREM

Portugal perdeu para a **Alemanha** nas quartas de final da **Eurocopa**. Vi o jogo na **Praça da República**, e mesmo com a derrota os bares de **Coimbra** continuaram cheios.

HAREM → ReReIEM

Portugal

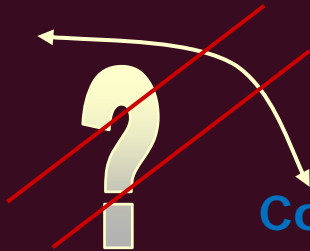


Coimbra

<EM ID="h-37" CATEG="PESSOA" TIPO="GRUPOMEMBRO">Portugal perdeu para a <EM ID="h-38" CATEG="PESSOA" TIPO="GRUPOMEMBRO">Alemanha na <EM ID="h-39" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Eurocopa. Vi o jogo na <EM ID="h-40" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="RUA">Praça da República, e mesmo com a derrota os bares de <EM ID="h-41" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="DIVISAO">Coimbra estavam cheios.

HAREM → ReReIEM

Portugal



Coimbra

<EM ID="h-37" CATEG="PESSOA" TIPO="GRUPOMEMBRO">Portugal

<EM ID="h-41" CATEG="LOCAL" TIPO="HUMANO"

SUBTIPO="DIVISAO">Coimbra

HAREM → ReReIEM

Portugal



Coimbra

Portugal	→	Eurocopa
Alemanha	→	Eurocopa
Pr. da República	→	Coimbra

~~<EM ID="h-37" CATEG="PESSOA" TIPO="GRUPOMEMBRO">Portugal perdeu para a <EM ID="h-38" CATEG="PESSOA" TIPO="GRUPOMEMBRO">Alemanha na <EM ID="h-39" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Eurocopa. Vi o jogo na <EM ID="h-40" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="RUA">Praça da República, e mesmo com a derrota os bares de <EM ID="h-41" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="DIVISAO">Coimbra estavam cheios.~~

ReReIEM – o que anotar?

Portugal

Alemanha

Pr. da República **INCLUSÃO**

Portugal

Alemanha

Pr. da República

Eurocopa

Eurocopa

Coimbra

INCLUSÃO

Portugal → Eurocopa

Alemanha → Eurocopa

Pr. da República → Coimbra

ReReEM – escolhas (linguísticas) subjacentes

Portugal

Alemanha

Pr. da República



Eurocopa

Eurocopa

Coimbra

INCLUSÃO

■ Consulta à literatura ☹️

■ Análise de corpus

■ Sintonia com HAREM

■ Especificidade vs Generalidade



relações informativas

relações aplicáveis a diferentes domínios

ReReIEM – o que anotar?

Identidade

...poderia ser resolvida com um simples teste de **ADN (DNA)**.

Inclusão (inclui / incluído)

Lobos recebidos em apoteose

(...), o capitão **Vasco Uva** explicou por que houve uma empatia tão grande entre ...

Localização (ocorre_em / sede_de)

GP Brasil

Não faltou emoção (...) no **Circuito José Carlos Pace** desde a primeira volta...

Outra (outras relações relevantes)

ReRelEM – o que é “outra” relação ?

- Co-ocorrência no mesmo texto/parágrafo?

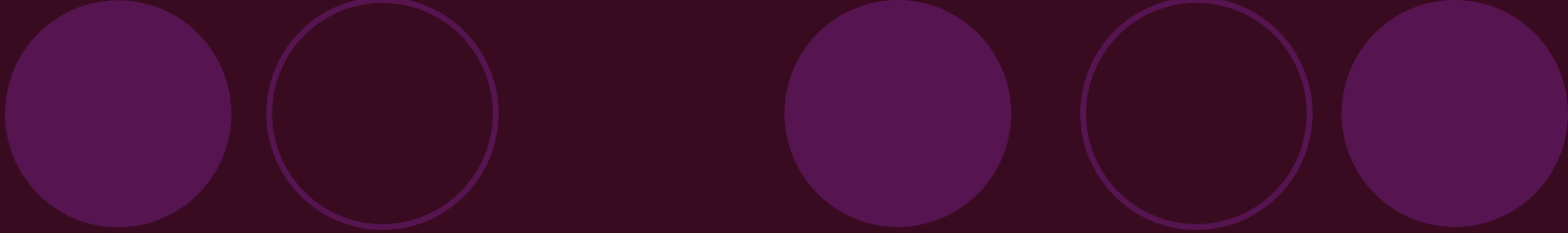
Conhecimento linguístico

X

Conhecimento enciclopédico

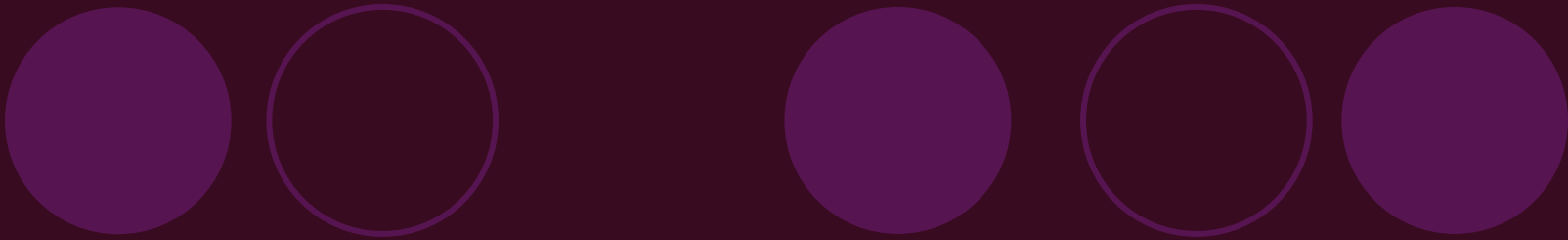
X

Conhecimento de mundo



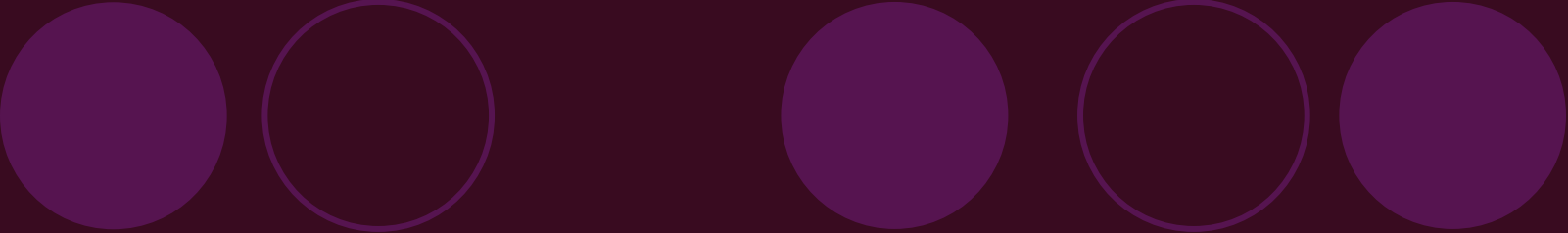
Depois de ser exibida no Rio, chega a São Paulo a mostra Carmen Miranda Para Sempre, que será inaugurada hoje para convidados no Memorial da América Latina. Fotos, roupas, objetos, são mais de 700 peças reunidas para contar a história da "Pequena Notável " ou a Brazilian Bombshell- não há no mundo quem não conheça essa genial estrela que conquistou o Brasil, a Broadway e Hollywood.

A mostra tem percurso cronológico e está dividida em núcleos. Inicia com o nascimento em Portugal e inclui imagens de sua família. Depois, vem a fase brasileira (...).Era uma "mulher art déco dos anos 30", que usava calças, ternos e vestidos belos - em particular, há uma sala especial com retratos da artista feitos em 1931, em Buenos Aires, pela alemã Annemarie Heinrich



Depois de ser exibida no Rio, chega a São Paulo a mostra Carmen Miranda Para Sempre, que será inaugurada hoje para convidados no **Memorial da América Latina**. Fotos, roupas, objetos, são mais de 700 peças reunidas para contar a história da "**Pequena Notável**" ou a **Brazilian Bombshell**- não há no mundo quem não conheça essa genial estrela que conquistou o **Brasil**, a **Broadway** e **Hollywood**.

A mostra tem percurso cronológico e está dividida em núcleos. Inicia com o nascimento em Portugal e inclui imagens de sua família. Depois, vem a fase brasileira (...). Era uma "mulher art déco dos anos 30", que usava calças, ternos e vestidos belos - em particular, há uma sala especial com retratos da artista feitos em 1931, em **Buenos Aires**, pela alemã Annemarie Heinrich



Depois de ser exibida no Rio, chega a São Paulo a mostra Carmen Miranda Para Sempre, que será inaugurada hoje para convidados no **Memorial da América Latina**. Fotos, roupas, objetos, são mais de 700 peças reunidas para contar a história da "**Pequena Notável**" ou a **Brazilian Bombshell** ~~que não há no mundo quem não conheça essa~~ genial estrela que conquistou o **Brasil**, a **Broadway** e **Hollywood**. **?**

A mostra tem percurso cronológico e está dividida em núcleos. Inicia com o nascimento em Portugal e inclui imagens de sua família. Depois, vem a fase brasileira (...). Era uma "mulher art déco dos anos 30", que usava calças, ~~as pernas e os pés descalços~~ **INCLUSÃO?** particular, há uma sala especial com retratos da artista feitos em 1931, em **Buenos Aires**, pela alemã Annemario Heinrich **MAS É RELAÇÃO MARCADA NO ReReIEM!**

ReReEM: o que anotar?

- Compatibilizar anotação linguisticamente motivada e interesses (e capacidades) dos sistemas

*Visitei uma exposição de cavalos, no Peru, e vi raças que só conhecia de fotografia: ↑ ?
Falabella, Hunter, Berbere, Andaluz e Paso*

- Necessidades de informação imprevisíveis

ReRelEM – o que anotar?

- TODAS as relações

- relações não inferíveis a partir do texto: INDEP
 - América Latina inclui Buenos Aires (INDEP)

ReReEM – “outras” relações

NATURAL_DE / LOCAL_DE_NASCIMENTO_DE	CAUSA_DE / CONSEQUENCIA_DE
RESIDENTE_DE/ RESIDENCIA_DE	OUTRA_EDICAO
POVO_DE / LOCAL_DE	REPRESENTANTE_DE / REPRESENTADO_POR
VINCULO_INST	PRATICADO_EM / PRATICA-SE
RELACAO_PROFSSIONAL	PARTICIPANTE_EM / TER_PARTICIPACAO_DE
RELACAO_FAMILIAR	NOME_DE / NOMEADO_POR
AUTOR_DE/OBRA_DE	NASCIDO_EM
PRODUTOR_DE/ PRODUZIDO_POR	MORTO_EM
PROPRIETARIO_DE/ PROPRIEDADE_DE	PERIODO_VIDA
DATADO_DE/DATA_DE	LOCALIZADA_EM / LOCALIZACAO_DE
PERSONAGEM_DE	OUTRAREL

ReReEM — escolhas linguísticas

□ Vagueza

- Relações entre as facetas

conta <EM ID="h-11" CATEG="PESSOA" TIPO="INDIVIDUAL">Diogo Barbosa
Machado na <EM ID="h-64" CATEG="LOCAL|OBRA"
TIPO="VIRTUAL|REPRODUZIDA" COREL="h-11" TIPOREL="OBRA**outra**hub-49343-
11**PESSOA">Bibliotheca Lusitana

ReReIEM

- Relações múltiplas
- Simetria, inversao e transitividade

A ident B e B ident C	→	A ident C
A inclui B e B inclui C	→	A inclui C
A inclui B e B sede_de C	→	A sede_de C
A ident B e B qualquer_relação C	→	A qualquer_relação C

ReReIEM



- Directivas
- CD
 - Anotação
 - Revisão
 - ...
- Métricas de avaliação
- Resultados

ReReIEM

- CD do ReReIEM

- Material para análise

- 12 DOCS
 - 573 EM
 - 582 relações manualmente anotadas
 - 6790 relacoes após expansão

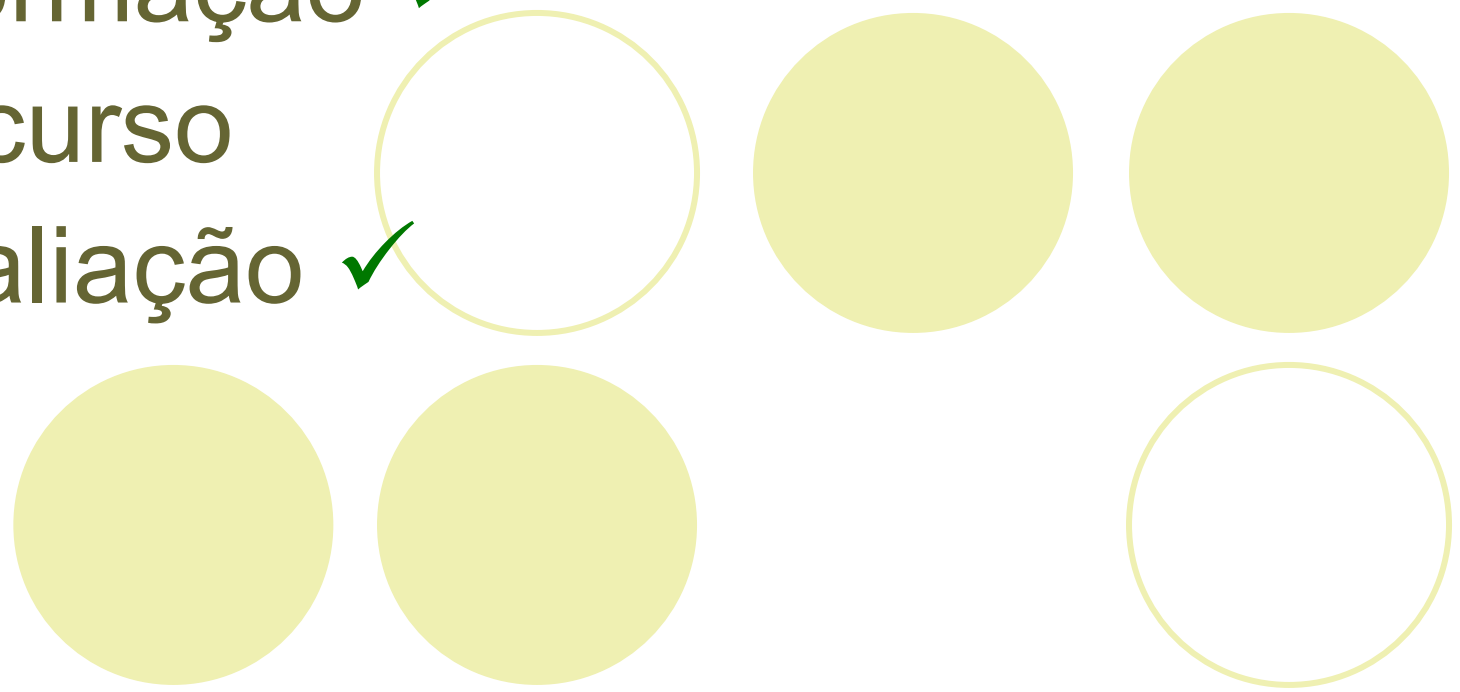
- Muito por fazer!

- anotação de relações em outros textos (validação)
 - Como tratar relações inversas...
 - Comparação entre as relações do ReReIEM e relações de outras avaliações (ACE, MUC), criadas para outras línguas
 - comparação das relações entre SN “simples” e EMs...
 - influências do gênero textual nos tipos de relações...
 - anáforas...
 - coesão textual...

Informação ✓

Recurso

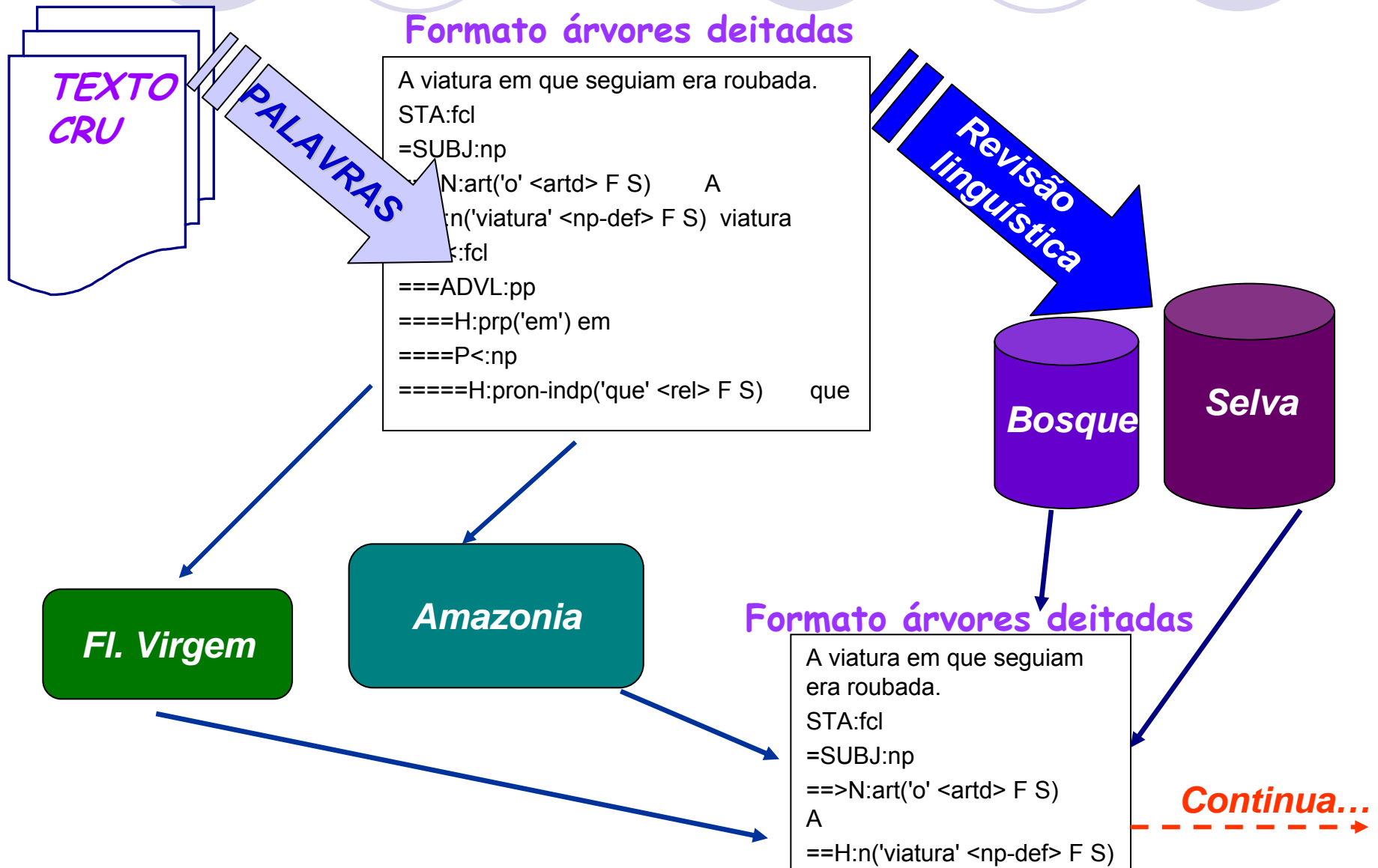
Avaliação ✓



Floresta Sintá(c)tica

- Treebank público do português
 - TREEBANK
 - Corpus em que cada frase foi anotada com informação sintática
 - Informação sintática e HIERÁRQUICA
 - Anotação automática (PALAVRAS) + REVISÃO HUMANA
 - PÚBLICO
 - Corpus publicamente disponível
 - www.linguateca.pt/Floresta
 - do PORTUGUÊS
 - europeu e brasileiro
- Colaboração entre a Linguateca e o projeto VISL (**V**isual **I**nteractive **S**yntax **L**earning)

Como é feita a Floresta



Floresta Sintá(c)tica: anotação

- Tipo de enunciado
 - QUE (?); STA (.); EXC (!); CMD (.)
- Tipo de oração
 - Finita (fcl); não-finita (icl); averbal (acl)
- Estrutura de constituintes (F:f)
 - Função: SUBJ, ACC, DAT, SC, OC, Complementos adverbiais, complementos nominais, PASS ...
 - forma: SNs, SAdj, SVs, SPreps, pronome, adverbio, conjuncao ...

Floresta Sintá(c)tica: anotação

- Etiquetas secundárias < >
 - lema
 - inf. morfossintáctica (artd; arti; refl; gênero, número, tempo, modo, pessoa verbal...)
 - elipse...
- “Procuráveis” ** < >
 - Verbos de orações *concessivas, causais, conformativas, condicionais, consecutivas, temporais, finais, relativas, substantivas, passivas, passivas com –se, sem sujeito*
 - Estruturas com quantificadores
 - Estruturas com partitivos

O **portal** da Quinta do Castro, o **único** que resta dos dois que já existiram, está acentuadamente inclinado e pode cair a qualquer momento.

STA:fcl

SUBJ:np

=>N:art('o' <artd> M S) O

=H:n('portal' M S) portal

=N<:pp

==H:prp('de' <sam->) de

==P<:np

===>N:art('o' <-sam> <artd> F S) a

===H:prop('Quinta_do_Castro' F S) Quinta_do_Castro

=,

=APP:np (<Eg>)

==>N:art('o' <artd> M S)o

==H:adj('único' <Eg> M S) único

==N<:fcl

===SUBJ:np

====H:pron-indp('que' <rel> M S) que

===P:vp

====MV:v-fin('restar' <fs-rel> PR 3S IND) resta

CP362-1 A **criação** da comissão de inquérito da AMS foi motivada por uma denúncia de moradores de São Marcos,...

STA:fcl

SUBJ:np

=>N:art('o' <artd> F S) A

=H:n('criação' F S) criação

=N<ARGO:pp

==H:prp('de' <sam->) de

==P<:np

====>N:art('o' <-sam> <artd> F S) a

====H:n('comissão' F S) comissão

====N<:pp

====H:prp('de') de

====P<:np

====H:n('inquérito' M S) inquérito

====N<:pp

====H:prp('de' <sam->) de

====P<:np

=====>N:art('o' <-sam> <artd> F S) a

====H:prop('AMS' F S) AMS

P:vp

=AUX:v-fin ('ser' PS 3S IND) foi

=MV:v-pcp ('motivar' <passive> F S) motivada

PASS:pp

=H:prp('por') por

A Floresta e a hierarquia

- Adjetivos que co-ocorrem com “dinheiro” (AC/DC)
 - “volume de *dinheiro compatível*” **OK!**
 - “manter o *dinheiro disponível*” **OK!**
- Adjetivos que modificam “dinheiro” (Floresta)

Floresta Sintáctica			AC/DC	
Fl. Virgem	Selva	Bosque	Chave (90 milhões de palavras)	
Resultados 28	Géneros: LIT FAL CIE	Resultados 2	245 casos.	2227 casos.
<u>público</u> 12	Resultados 4	<u>bom</u> 1	Distribuição	
<u>fresco</u> 4	<u>fresquinho</u> 1	<u>público</u> 1	Houve 45 valores diferentes de lema .	
<u>alemão</u> 2	<u>grosso</u> 1		bom 55	público 557
<u>comum</u> 1	<u>intacto</u> 1		próprio 33	disponível 144
<u>devolvido</u> 1	<u>necessário</u> 1		novo 28	fresco 98
<u>egípcio</u> 1			preciso 21	vivo 93
<u>estrangeiro</u> 1			dado 14	sujo 84
<u>extra</u> 1				
<u>fácil</u> 1				

Conteúdo da Floresta

Floresta Virgem

(Público e Folha S. Paulo)

não revista

1.700.000 palavras

Bosque

revisto

185.000 p.

Amazônia

(Overmundo)

Apenas PB

não revista

3.800.000 palavras

Selva

Parcialmente revisto

300.000 p.

oral, lit, cient.

Conteúdo da Floresta

corpus	Floresta Virgem	Amazônia	Bosque	Selva Lit.	Selva Fal.	Selva Cie.
palavras	c. 1.640.000	c. 4.580.000	c. 186.000	c. 105.000	c. 170.000	c. 125.000
frases	c. 96.000	c. 275.000	9.368	c. 7.900	c. 14.000	c. 6.200
revisão	não	não	integral	parcial	parcial	parcial
variantes	PT BR	BR	PT BR	PT BR	PT BR	PT BR
gênero	jornalístico	opinião	jornalístico	literário	entrevistas / debates	acadêmico / informativo
domínio	genérico	cultura brasileira	genérico	genérico	biografia / política	educação / psicolinguística / computação / economia / ciências [*]
registro	formal	formal e informal	formal	formal	formal e informal	formal
modo	escrito	escrito	escrito	escrito	falado	escrito
origem	jornais Folha de São Paulo e Público	blog Overmundo	jornais Folha de São Paulo e Público	livros [*]	Museu da Pessoa (PT, BR) debates parlamentares	bibliotecas universitárias banco centrais Wikipedia

Para que(m) serve uma floresta sintática?

- treino e avaliação de analisadores morfosintáticos

Refletir consenso entre possibilidades de análise / permitir escolha informada

- estudos baseados em dados
 - investigação / descrição da língua

- auxiliar no ensino

ESPAÇO DE PESQUISA!

Como trabalhar com a Floresta ?

O assessor de imprensa do Ministério da Fazenda disse ontem que..

=SUBJ:np
==>N:art('o' M S) o
==H:n('assessor' M S) assessor
==N<:pp
===H:prp('de') de
===P<:np
====H:n('imprensa' F S) imprensa
==N<:pp
===H:prp('de' <sam->) de
===P<:np
====>N:art('o' <-sam> M S) o
====H:prop('Ministério_da_Fazenda' M S) Ministério_da_Fazenda
=P:vp
==MV:v-fin('dizer' PS 3S IND) disse
=ADVL:advp
==H:adv('ontem') ontem
=ACC:fcl
SUB:coni o('que') que

A INFORMAÇÃO É ÚTIL SE
DISPONÍVEL 😊
FACILMENTE ENCONTRADA 😞



(tentativa de)Compatibilização

- Usuário Linguista

- Diferentes modelos teóricos
- Diferentes graus de conhecimento linguístico
- Pouca familiaridade com formalizações mais utilizadas em informática (expressões regulares etc)

- Projeto Floresta

- Único modelo de anotação subjacente ao parser (CG)
- Modelo pouco conhecido do lado “linguístico não-PLN”
- Interface capaz de lidar com um objeto altamente complexo como a língua

Mozilla Firefox

Ficheiro Editar Ver Histórico Marcadores Ferramentas Ajuda

http://linguateca.dei.uc.pt/Floresta/BibliaFlorestal/

Customize Links Free Hotmail Windows Marketplace Windows Media Windows lista_participantes.txt

Bíblia Florestal: Um manual lingüístico da Floresta Sintá(c)tica

Bosque 7.4

[Versão para impressão](#)

- [1. Introdução](#)
- [2. O formato ad \(árvores deitadas\)](#)
- [3. Orações](#)
- [4. Sintagmas](#)
- [5. Unidades/palavras](#)
- [6. Funções sintáticas](#)
- [7. Ambiguidades](#)
- [8. Não especificação de formas](#)
- [9. Constituintes descontínuos](#)
- [10. Frases com um constituinte](#)
- [11. Coordenação](#)
- [12. Fenómenos sintácticos](#)
- [13. Partitivos](#)
- [14. Tipos de orações](#)
- [15. Informação detalhada sobre icl e acl](#)
- [16. Outras marcações](#)
- [17. Revisão futura](#)
- [18. A documentar](#)

[Anexo 1: Glossário de etiquetas florestais](#)
[Anexo 2: Membros das classes fechadas](#)
[Anexo 3: Expressões multi-palavras](#)
[Bibliografia](#)

Bíblia Florestal: Um manual lingüístico da Floresta Sintá(c)tica.

Bosque 7.4

Cláudia Freitas / Susana Afonso

25 de Outubro de 2007

[Versão para impressão](#)

[Próxima página](#)

1. Introdução

A Floresta Sintá(c)tica é um projecto de construção de um "treebank": um conjunto de árvores sintacticamente anotadas e revistas manualmente/intelectualmente.

O "treebank" (a que se chama Bosque) cobre duas variantes do português: português europeu e português do Brasil. Dois corpora de texto jornalístico são utilizados: o [CETEMPúblico](#) (mais concretamente, o primeiro milhão de palavras) (Rocha e Santos, 2000) do jornal Público em português europeu, e o ([CETENFolha](#)) do jornal Folha de S. Paulo, em português do Brasil. Os corpora foram anotados pelo analisador automático PALAVRAS (Bick 2000) e estão a ser manualmente revistos em formato de árvores. A presente documentação é o resultado das discussões conjuntas com os restantes membros do projecto durante a construção da Floresta Sintá(c)tica.

Apenas o formato de árvores presente no Bosque (a parte revista da Floresta) está aqui documentado: não só as etiquetas utilizadas e o formato, mas também as escolhas que foram sendo feitas ao longo do projecto. A sua descrição será acompanhada de exemplos.

Uma vez que o processo de construção do treebank continua, este documento é aberto, isto é, em constante modificação em face das novas decisões que possam surgir no decurso do trabalho de revisão.

Nota 1: O projecto Floresta Sintá(c)tica foi/é (parcialmente) desenvolvido no âmbito da Linguateca, financiada através dos projectos POSI/PLP/43931/2001 e POSC 339/1.3/C/NAC, e co-financiada pelo POSI.

Nota 2: O projecto Floresta Sintá(c)tica existe desde 2000, e a sua documentação foi produzida majoritariamente entre 2000 e 2005. Ligeiramente "adormecido" desde então, o projecto foi retomado em 2007, com alguns novos integrantes, e com isso houve também a revisão de alguns pontos da anotação, que deu origem ao presente documento - que é, portanto, uma revisão de Susana Afonso. "Árvores deitadas: Descrição do formato e descrição das opções de análise na Floresta Sintá(c)tica". Última versão: 12 de Fevereiro de 2006. Primeira versão: 2004. Esse documento pode ser encontrado em <http://www.linguateca.pt/documentos/Afonso2006ArvoresDeitadas.pdf>

Concluído

start Microsoft PowerPoint ... Mozilla Firefox 12:08

Nossa solução

- Modelo teórico

- Opção pela granularidade

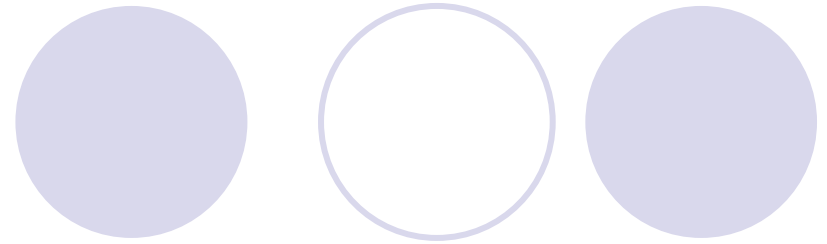
- Complementos vs modificadores de nomes

- N<ARG / N<ARGS / N<ARGO
- Encontrei-o na entrada do prédio... -> AA (>N ; N<)
- ..a entrada de insectos (= os insectos entram) N<ARGS
- ..perdi com a venda da casa (=vender a casa) N<ARGO
- As leituras de J.L.Borges (= Borges leu -> N<ARGS)
(= ler Borges -> N<ARGO)

- n-adj

- um integrante da campanha;
- em favor dos evangélicos;
- os responsáveis pelos menores;
- abuso sexual em bulímicas...
- as multinacionais surgiram no final do século XIX...
- a indicação de um tranquilizante só é feita quando...
- a direção do novo semanal será assinada

Nossa solução



- Modelo teórico
 - Opção pela granularidade
 - Partitivos

partitivos com não contáveis	bebeu um copo de água ; ...pesou metade da farinha	<np-partitive1>
de quantif. /pron. com art. def. e subst. contável	maioria dos feridos , boa parte dos problemas ; metade dos leitores ; alguns dos meninos , três dos meninos	<np-partitive2>
"outros" quantificadores com PP	uma série de questões , milhões de usuários	<np-dequant>

Nossa solução

A decorative graphic at the top of the slide consists of two groups of three circles. The left group has a solid light purple circle on the left, a white circle with a light purple outline in the middle, and a solid light purple circle on the right. The right group has a solid light purple circle on the left, a white circle with a light purple outline in the middle, and a solid light purple circle on the right.

○ Interface

- terminologia baseada em “gramática tradicional”
(quando possível)
- operações em linguagem natural

- [AC/DC](#)
- [CETEMPúblico](#)
- [CETENFolha](#)
- [CHAVE](#)
- [COMPARA](#)
- [Corpógrafo](#)
- [CorTrad](#)
- [Esfinge](#)
- [Floresta Sintá\(c\)tica](#)
- [- Acesso](#)
- [- Documentação](#)
- [METRA](#)
- [PAPEL](#)
- [ReEscreve](#)
- [REPENTINO](#)
- [Repositório](#)
- [WebJspell](#)
- [WPT](#)

procurar expressão 1º elemento
limpar expressão apagar 1º elemento

visualizar pedido

Bosque
 Selva
 Floresta Virgem
 Amazônia
 variante europeia
 variante brasileira

morfologia

Apresentação dos resultados:
 frases completas apenas extracto

Agrupar resultados por
 função forma lema morfologia texto

contém imediatamente (<)
contém (<<)
começa com (<<.)
termina com (<<`)
está contido imediatamente (>)
está contido (>>)
contém exactamente (<:)
é o unico nó de (>:)
é seguido por (.)
é imediatamente seguido por (.)
é precedida por (.)
é imediatamente precedido por (.)
está no mesmo nó (\$)
», imediatamente seguido por (\$.)
», imediatamente precedido por (\$.)
», seguido por (\$..)
», precedido por (\$..)
é o mesmo nó que (\$=)

relações:
1º elemento
 não contém imediatamente
2º elemento OK

forma **função**

Formas oracionais
 finita (fc1)
 não-finita (ic1)
 averbal (ac1)

Sintagmas
 nominal (np)
 adjectival (adjp)
 adverbial (advp)
 verbal (vp)
 preposicional (pp)
 evidenciador coord. (cu)
 seq. discursivas (sq)

Item lexical
Palavra:
Lema:
Procurável:
 substantivo (n)
 subst/adjectivo (n-adj)

Enunciado
 frase (UTT)
 declarativa (STA)
 interrogativa (QUE)
 imperativa (CMD)
 exclamativa (EXC)

Oração
 sujeito (SUBJ)
 obj. directo (ACC)
 part. apassivante (ACC-PASS)
 obj. ind. pronominal (DAT)
 obj. ind. preposicional (PIV)
 agente da passiva (PASS)
 adjunto adverbial ⊕
 predicativos (SC/OC/PRED) ⊕
 vocativo (VOC)
 apostos ⊕
 compl. nominais (N<ARG) ⊕
 predicador (P)

Floresta para Linguistas

● Pesquisa ↔ Ensino

● fenômenos da língua em ambiente real

○ Seleção de estruturas

● Contraste / comparação com gramáticas e dicionários

● Buscas por generalização

elaboração de material didático
gramática/linguística “experimental”
aluno.....→ pesquisador

Incentivo à reflexão
Incentivo à crítica
Sujeito de investigação

Seleção / Reflexão / Generalização

SN

● Sintagmas

- *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
-
- *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.
 - *CF2-7*: Sob o comando de Ronaldo Rosas, o programa mostrará reportagens especiais de Sônia Pompeu.

<i>CF2-7</i>	o comando de Ronaldo Rosas
<i>CF2-7</i>	Ronaldo Rosas
<i>CF2-7</i>	o programa
<i>CF2-7</i>	reportagens especiais de Sônia Pompeu
<i>CF2-7</i>	Sônia Pompeu

Apenas extracto
(SN)

Núcleo do SN

Que classes de palavras podem ser núcleo do SN?

Apresentação dos resultados:

frases completas apenas extracto

Agrupar resultados por

função forma lema morfologia texto

elementos e relações inseridos

Elementos:

clique sobre um elemento para editá-lo ou apagá-lo
selecione os elementos a sublinhar na resposta

- forma*: np
- função*: H
-
-
-
-
-
-
-
-

Relações:

1º elemento ▾

não contém imediatamente (<) ▾

2º elemento ▾

◆ 2º elemento está contido imediatamente no 1º elemento

Variantes: BR PT

Resultados 59000

n 1000 de 39945

prop 10808

pron-indp 3209

pron-pers 2496

pron-det 928

n-adj 564

cu 375

num 279

adj 86

icl 83

np 82

v-pcp 69

prp 18

v-inf 14

art 11

adjp 10

cu- 9

adv 4

np- 3

pp 3

? 1

fcl 1

Quais as funções exercidas pelos SN?

Apresentação dos resultados:

frases completas apenas extracto

Agrupar resultados por

função forma lema morfologia texto

elementos e relações inseridos

Elementos:

clique sobre um elemento para editá-lo ou apagá-lo
selecione os elementos a sublinhar na resposta

- forma: np**
- função: H
-
-
-
-
-
-
-
-

Relações:

2º elemento ▾

não está contido imediatamente (>) ▾

1º elemento ▾

◆ 1º elemento contém imediatamente o 2º elemento

Resultados 58994

P< 1000 de 27896

SUBJ 11827

ACC 8584

CJT 5355

SC 1473

N<PRED 1290

UTT 919

APP 672

ADVL 263

N< 252

H 91

>N 77

OC 62

VOC 29

PRED 28

SA 28

>A 23

S< 17

-ACC 16

A< 12

ACC-PASS 12

-SUBJ 9

QUE 9

-P< 8

TOP 6

- *CF1-1*: PT no governo
- *CF1-3*: Brasília Pesquisa Datafolha publicada hoje revela um dado surpreendente; recusando uma postura radical, a esmagadora nr
- *CF1-3*: Brasília Pesquisa Datafolha publicada hoje revela um dado surpreendente; recusando uma postura radical, a esmagadora nr
- *CF1-5*: Muito mais do que nos tempos na ditadura, a solidez do PT está, agora, ameaçada.
- *CF1-5*: Muito mais do que nos tempos na ditadura, a solidez do PT está, agora, ameaçada.
- *CF1-5*: Muito mais do que nos tempos na ditadura, a solidez do PT está, agora, ameaçada.
- *CF1-8*: Muitas das prioridades do novo governo coincidem com as prioridades do PT.
- *CF1-8*: Muitas das prioridades do novo governo coincidem com as prioridades do PT.
- *CF1-8*: Muitas das prioridades do novo governo coincidem com as prioridades do PT.
- *CF1-8*: Muitas das prioridades do novo governo coincidem com as prioridades do PT.
- *CF2-1*: « Confissões » chega a Portugal
- *CF2-2*: Desde o último dia 13, « Confissões de Adolescente » pode ser vista pelos teens portugueses.
- *CF2-2*: Desde o último dia 13, « Confissões de Adolescente » pode ser vista pelos teens portugueses.
- *CF2-3*: A série exibida aqui pela Cultura estreou na TVI de Portugal.
- *CF2-3*: A série exibida aqui pela Cultura estreou na TVI de Portugal.
- *CF2-3*: A série exibida aqui pela Cultura estreou na TVI de Portugal.
- *CF2-4*: Além disso, a co-produção com o canal francês TCF1 para a realização de mais 30 episódios continua sendo negociada.
- *CF2-4*: Além disso, a co-produção com o canal francês TCF1 para a realização de mais 30 episódios continua sendo negociada.
- *CF2-4*: Além disso, a co-produção com o canal francês TCF1 para a realização de mais 30 episódios continua sendo negociada.
- *CF2-4*: Além disso, a co-produção com o canal francês TCF1 para a realização de mais 30 episódios continua sendo negociada.

Reflexão sobre a língua

- Correlação entre forma e função

- Mesma função, formas diferentes

- Adjunto adnominal

anterior ao núcleo

Resultados 39150

[art](#) 1000 de 29330

pron-det 3929

adj 2443

num 2172

adv 373

adjp 260

ec 162

v-pcp 128

cu 87

np 78

n 65

pp 65

pron-indp 11

advp 10

posterior ao núcleo

Resultados 24919

[pp](#) 1000 de 13383

adjp 6000

fcl 1735

prop 910

icl 895

cu 703

v-pcp 518

np 255

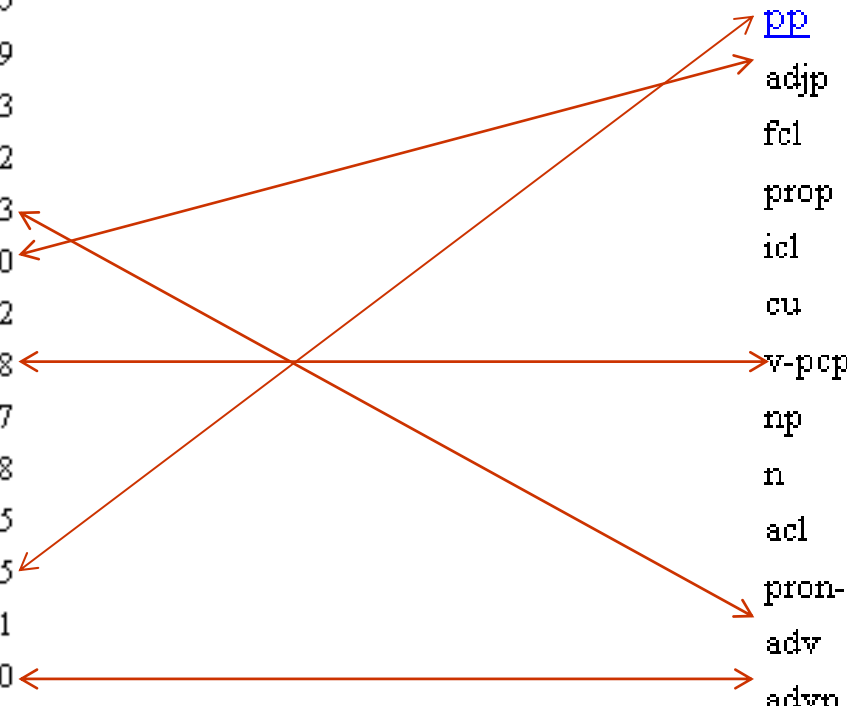
n 182

acl 90

pron-det 72

adv 5

advp 50



(seleção e) Reflexão sobre a língua

● A hierarquia

○ Spreps no mesmo nível de um núcleo N

CF922-7: A Mercedes-Benz deve instalar a linha de montagem do Swatch, carro compacto com design do fabricante dos relógios da marca, em

CF938-3: Para Spinelli, o espaço representa uma nova tendência para as áreas de conveniência dos shoppings.

CF964-3: Segundo a assessoria de imprensa do governador, Fleury já pediu um levantamento de preço das ações, junto a consultores especializa

CF980-1: Segundo ele, o tubo era independente do sistema de propulsão do submarino e o vapor que escapou não é radiativo.

CF994-3: VERA LÚCIA DA COSTA ANTUNES é coordenadora de geografia do curso e colégio Objetivo.

CP3-4: « Estamos a dotar os computadores de um novo sentido »" disse Steve d' Averio, director de marketing para a Europa da Logitech.

CP28-9: Durante o primeiro semestre do corrente ano, o mercado de computadores pessoais da Coreia do Sul registou um aumento de 42 por c

- CP348-2: O dirigente do PSD da Madeira aproveitou a ocasião para dizer que, « até 1985, não houve qualquer c
- CP350-5: Cunha Rodrigues diz que tudo está nas mãos do procurador do Tribunal de Cascais, de quem aguarda
- CP351-3: Esta legislação, altamente contestada pelos serviços de protecção de dados, não foi alterada a tempo p
- CP356-2: A resolução aprovada pelos deputados europeus foi apresentada pelo grupo da coligação de esquerda

(seleção e) Reflexão sobre a língua

- Ausência de concordância gramatical entre sujeito e predicado
 - Consulta a gramáticas
 - Hipóteses sobre processos de prod & compr. da lngem

Resultados 20

CF38-2: Segundo o BC, mais da metade dos recursos da caderneta são de poupadores de médio e grande porte em tese, menos sujeitos a ach

CF156-1: Segundo Milton Julio Zanluqui, da Paraklin Pára-raios, entre os tipos mais comuns (veja exemplos ao lado) estão o sistema de gaiola aéreo, cabo de cobre, suportes isoladores, tubo de proteção, malha de aterramento e conector de medição.

CF221-3: Segundo Calliari, o barulho na agência causado pela reforma do prédio abafaram o ruído da queda.

CF613-4: Ficando novamente vago o cargo, o presidente do Senado e, sucessivamente, do Supremo Tribunal Federal (STF) assumem a incumb

CF657-4: Todos os fax/modem saem de fábrica com dois programas um para gerenciamento das transmissões dfa e outro para gerenciamento

CF918-3: A maioria dos voluntários são jovens.

CF956-3: Esse sinal de fraqueza associado a crises internas levam à queda do sistema em 1867.

CF988-2: O público alvo são os três milhões de pedestres que circulam diariamente na região.

CP29-1: O anúncio divulgado na sexta-feira pela OMS dizia que uma série de testes iniciais, realizados pelo Instituto Pasteur de Paris, em anos diagnóstico de febre de Ebola.

CP320-5: Um par de presenças enlutadas fazem aqui as vezes de personagens das trevas, no meio de uma vitrina de bordados barrocos, azuis .

Reflexão sobre dados do corpus...

Verbos que antecedem sujeito

○ Por lema

Resultados	1066
dizer	136 de 136
ser	86 de 86
estar	42 de 42
existir	39 de 39
afirmar	30 de 30
explicar	22 de 22
fazer	14 de 14
surgir	14 de 14
contar	10 de 10
observar	10 de 10
bastar	9 de 9
caber	9 de 9
chegar	9 de 9
dar	9 de 9
garantir	9 de 9
morrer	9 de 9
parecer	9 de 9
passar	9 de 9
restar	9 de 9

○ Por função da frase

Resultados	1052
STA	583 de 583
ADVL	107 de 107
CJT	91 de 91
ACC	82 de 82
P<	54 de 54
N<	41 de 41
N<PRED	35 de 35
QUE	7 de 24
SC	8
SUBJ	8
UTT	7
-ACC	2
KOMP<	2
PRED	2
-P<	1
APP	1
EXC	1
OC	1
PIV	1
S<	1

Reflexão sobre dados do corpus...

- ...e gramática

“ *inversão verbo + sujeito verifica-se em geral:*

- a) nas orações interrogativas*
- b) nas orações que contêm imperativo*
- c) em que o verbo está na passiva pronominal*
- d) ...*
- e) nas orações construídas com verbos do tipo dizer, sugerir (...) que arrematam enunciados em discurso direto...*

(Cunha e Cintra, 1984, 5ª ed., p. 163-165)

- ...e discurso

- Por que os verbos *dicendi* são freqüentes no corpus?

(seleção e) Compreensão e produção de texto

- Orações passivas sem agente

CF4-3: O modelo Lx 810, da Epson, é vendido em Miami por US\$ 178.

CF36-2: Caio entra no lugar de Catê, caso ele seja suspenso hoje pela manhã

CP587-6: Agora os dois títulos quebraram um jejum de sete e oito anos e foram conquistados, para mais, quando o Benfica era dado como favorito ».

CF5-2: Parte dos recursos para a formação do FSE foi deslocado do orçamento da saúde e educação.

CP594-3: Ontem foram assassinados cinco militantes do partido Bharatiya Janata, da direita hindú.

(seleção e) Compreensão e produção de texto

- Orações reduzidas de gerúndio
 - (or. finita que contém uma or. não finita, que contém gerúndio)

CF71-1: Os empresários abriram mão de posições históricas, eventualmente visando sua proteção, p

CF95-2: A distribuição acontecerá até o final do mês, partindo das capitais para o interior do país.

CF101-3: Fácil subir nos palanques, defendendo que é necessário « mudar tudo o que está aí ».

CF103-3: Além do contrato com a Servaz, há outros do mesmo tipo, envolvendo mais empreiteiras.

CF104-4: Só assim se evitará a estratificação da legislação, permitindo que o sistema acompanhe de

- Tentativa de reescritura com conectivos
- Ambiguidade vs concisão

(seleção e) Compreensão e produção de texto

- Paralelismo sintático
 - (coordenação de elementos sintaticamente diferentes)

CF346-3: No PPR, a discussão será sobre a política de alianças e se ela inclui o PSDB, além dos novos candidatos do partido a presidente.

CF584-2: Por dois motivos; primeiro, a indisponibilidade do escritor, que simplesmente não estava interessado em percorrer os Estados Unidos ou debater suas idéias; segundo, seus comentários sobre a « decadência moral » do Ocidente logo o tornaram inconveniente

CP274-4: Na altura, Manuel António alegou sede e que queria beber água, justificando desta forma a sua entrada na esquadra através do muro.

CP812-6: Não lhe diz nada Ernest Hemingway, Prémio Nobel da Literatura, um homem cativado pela festa e que escolheu a terra de Espanha para sua última morada?

CP866-3: Entre os primeiros, conta-se o advogado Bernard Nussbaum, amigo de Clinton e que se demitiu na semana

(seleção e) Produção de texto

● Regência

CP220-4: Mais tarde ou mais cedo, quando nos encontrarmos, usufruirei, eu também, das experiências de que ele agora desfruta.

CP224-4: Segundo o comandante da GNR do Cacém, entidade a quem primeiro foi comunicada a ocorrência, não foi consumado...

CP470-7: Levaram-me ao Ministério da Defesa para encontrar alguém a quem me entregar, mas o edifício estava vazio.

CP521-2: Neste último caso, o destaque vai para um filme de guerra, « Jornada Trágica », que terá menos a ver com o conflito a que se referia do que ...

CP549-3: Segundo o estudo publicado ontem pelo Expresso, a CDU perderia em Setúbal quatro dos sete deputados de que dispõe, passando o PS...

CP613-2: « A pobreza e a solidariedade » é assim o tema de um colóquio a que Soares assiste em Setúbal, no dia 6, onde intrevêm o presidente...

CP775-7: Outro Orçamento Relativo que contém um SPBE de que todos falavam é

(seleção e) Fluência – uso de tempos verbais

● Pres. do Subjuntivo vs Fut. do Subjuntivo em or. relativas

Resultados 70

- *CF44-3*: Países produtores e consumidores de café entraram em um tipo de pré- pacto internacional que não intervenha nos preços do produto, afirmou o po
- *CF98-2*: No início do mês, o presidente Carlos Menem criou um fundo especial de US\$ 1 milhão para recompensar quem fornecesse informações que ajudem
- *CF137-5*: Na sua defesa encaminhada ontem ao TSE, ela admitiu ter cometido um « erro de interpretação » na produção da reportagem na qual Franco afirmou calote na dívida interna ».
- *CF348-4*: « Não podemos aceitar medidas que contrariem o combate à inflação », disse FHC, ao responder sobre o assunto.
- *CF367-5*: O governo não está errado em formular regras que possibilitem o alongamento.
- *CF417-2*: « Há uma busca frenética por eletrônicos que facilitem o dia-a-dia do cidadão em casa e no escritório », afirma Quintana.
- *CF449-1*: Para ele, os clubes precisam encontrar uma fórmula que concilie os dois aspectos.
- *CF473-3*: « Nós não vamos dar corda ao debate rasteiro », acrescentou o presidente do PT, que acha improvável o STF permitir a candidatura de José Sarney.
- *CF567-3*: Ora, se o presidente chegar à posse sem que estejam no mínimo delineados, divulgados e negociados com os setores organizados da sociedade os difícil fazê-los aprovar no Congresso no prazo de um semestre definido pelo próprio FHC.
- *CF639-1*: O ex- jogador defendeu também a adoção de penas para menores de 18 anos que cometam crimes.
- *CF811-1*: Eduardo Lourenço -- Não se deve esperar uma intervenção que tenha efeitos imediatos como a dos políticos.

Resultados 70

- *CF192-3*: Itamar disse que até agora poucas demissões foram efetivadas porque « há sempre um processo legal e burocrático que precisar ser cumprido ».
- *CF275-6*: Juan Manuel Fangio, sobrinho do argentino pentacampeão de Fórmula 1, está testando na Califórnia um novo protótipo que vai dar o que falar na
- *CF311-5*: Os que preferirem carne, podem optar por um nhoque preparado à base de vitela.
- *CF324-8*: Os contribuintes que pagarem seus tributos (IPTU, ISS etc.) hoje sairão ganhando, pois o cálculo será feito pela UFM de R\$ 26,54.
- *CF329-2*: « Não sei se ele vai querer meu apoio, mas farei tudo que puder por ele. »
- *CF452-2*: O humor de Altman vai na direção das convenções de uma sociedade que, tendo que aceitar tudo sob as aparências, acaba perpetuando uma vast
- *CF481-1*: Convencido de que Munhoz tem uma boa retórica, o publicitário pretende deixar o candidato « falar o que quiser » para, antes da gravação final, «
- *CF583-3*: O motivo é que ele obriga a empresa que contratar jovens a manter um tutor para ensinar o trabalho aos novos empregados.
- *CF663-4*: Os senadores que faltarem às sessões do Congresso revisor, a partir de hoje, terão cortes no valor de um dia de trabalho por falta.

(seleção e) Fluência – uso de tempos verbais

● Pres. do Subjuntivo vs Fut. do Subjuntivo em or. relativas

R Resultados 70

Resultados 25

Frase	/^MV:v-fin.:+PR_.*_SUBJ.:.*\$/	/^.*:vp.*\$/
CF44-3	intervenha	entraram
CF98-2	ajudem	fornecesse
CF137-5	ganhe	ter cometido
CF348-4	contrariem	podemos aceitar
CF367-5	possibilitem	formular
CF417-2	facilitem	Há
CF449-1	concilie	encontrar
CF473-3	seja	permitir
CF567-3	consubstanciem	chegar
CF639-1	cometam	defendeu
CF811-1	tenha	deve esperar
CF833-1	reflita	haja
CF919-5	renove	tem
CF927-3	invista	estive
CF950-6	estimule	é
CP17-3	acabe	deixa

Frase	/^MV:v-fin.:+FUT_.*_SUBJ.:.*\$/	/^.*:vp.*\$/
CF192-3	precisar	há
CF275-6	falar	está testando
CF311-5	preferirem	podem optar
CF324-8	pagarem	sairão
CF329-2	puder	farei
CF452-2	aceitar	vai
CF481-1	quiser	falar
CF583-3	contratar	obriga
CF663-4	faltarem	terão
CF676-5	corresponderem	procederem
CF922-4	descumprirem	pode chegar
CP80-5	inventar	abrirá
CP145-3	quisermos	tornou-
CP156-3	for	está
CP189-6	chegarem	comece
CP241-6	escolherem	podem começar

♦ CF663-4: Os senadores que faltarem às sessões do Congresso revisor, a partir de hoje, terão cortes no valor de um dia de trabalho por falta.

(seleção e) Fluência – uso de tempos verbais

- Pres. do Subjuntivo vs Fut. do Subjuntivo em or. relativas

CF919-5: Pepe ainda tem esperança que o centroavante Guga o seu contrato, vencido em 31 de dezembro.

CF927-3: Durante esta semana, o agora dirigente esteve no Japão, em busca de uma empresa que, no futuro, no Santos.

CF950-6: Nossa preocupação é um projeto que no jovem uma vida sem drogas, promovendo-se um comportamento construtivo.

(listas e) Transitividade dos nomes

Apresentação dos resultados:
 frases completas apenas extracto

Agrupar resultados por
 função forma lema morfologia texto

elementos e relações inseridos

Elementos:
clique sobre um elemento para editá-lo ou apagá-lo
selecione os elementos a sublinhar na resposta

- função: N<ARG, N<ARGS, N<ARGO
- função: H, forma: n
-
-
-
-
-
-
-
-

Relações:

1º elemento ▼

não está no mesmo nó (\$) ▼

2º elemento ▼

◆ 2º elemento está no mesmo nó do 1º elemento



Resultados	1887
milhão	129 de 129
relação	33 de 33
parte	28 de 28
criação	26 de 26
metade	26 de 26
aumento	21 de 21
nome	21 de 21
pagamento	21 de 21
venda	18 de 18
entrada	17 de 17
produção	17 de 17
realização	17 de 17
construção	16 de 16
uso	16 de 16
apoio	15 de 15
milhar	15 de 15
tentativa	15 de 15
abertura	14 de 14
redução	14 de 14
aplicação	13 de 13
centena	13 de 13
desenvolvimento	13 de 13
tonelada	13 de 13
acordo	11 de 11
série	11 de 11

(listas e) Transitividade dos nomes

Apresentação dos resultados:
 frases completas apenas extracto

Agrupar resultados por
 função forma lema morfologia texto

elementos e relações inseridos

Elementos:
clique sobre um elemento para editá-lo ou apagá-lo
selecione os elementos a sublinhar na resposta

- função: N<ARG, N<ARGS, N<ARGO
- função: H, forma: n
-
-

Resultados	1887
milhão	129 de 129
relação	33 de 33
parte	28 de 28
criação	26 de 26
metade	26 de 26
aumento	21 de 21
nome	21 de 21
pagamento	21 de 21
venda	18 de 18
entrada	17 de 17
produção	17 de 17
realização	17 de 17

criação

- *CF188-3*: O programa propõe a monetização da dívida interna (pagamento com dinheiro impresso) acompanhada da criação de uma nova moeda (de rígido controle monetário).
- *CF219-5*: Isso significa que, tudo indo bem, a criação da URV fica para março.
- *CF220-4*: Elas se algemaram entre si e nos móveis do escritório em protesto contra a utilização de peles animais em criações do estilista.
- *CF423-4*: E anuncia, no momento mesmo da criação do indexador, que a partir de determinada data a conversão será obrigatória, conforme regras
- *CF674-2*: O atentado teria sido planejado pelo grupo radical islâmico Gama'a al-Islamiya, que lançou campanha contra o governo egípcio em 1992
- *CF714-6*: Pode se reorientar para uma maior produção industrial e contribuir para a criação de mais emprego e riqueza, consequência natural do pr
- *CF715-6*: O governo promete à sociedade que, de agora em diante, vai mudar de vida, vai garantir o equilíbrio das contas públicas e vai parar de a
- *CF752-1*: O documento também sugere a criação de uma sistemática mensal de acompanhamento.

(listas e) Transitividade dos nomes / verbos

- Complementos de CRIAÇÃO

		N<ARGS		
			Resultados	1
			estilista	1
Resultados	22			
região	4			
moeda	3			
grupo	2			
posto de trabalho	2			
URV	1			
ambiente	1			
depósito	1			
emprego	1			
estado	1			
estilista	1			
estrutura	1			
indexador	1			
negócio	1			
sistema	1			
sistemática	1			
		N<ARGO		
			Resultados	21
			região	4
			moeda	3
			grupo	2
			posto de trabalho	2
			URV	1
			ambiente	1
			depósito	1
			emprego	1
			estado	1
			estrutura	1
			indexador	1
			negócio	1
			sistema	1
			sistemática	1

- Complementos de CRIAR

		OBJECTO		
			Resultados	43
			condição	4
			dificuldade	2
			URV	1
			ambiente	1
			cardápio	1
			chance	1
			comissão	1
			desajustamento	1
			embaraço	1
			emprego	1
			escola	1
			espécie	1
			estilo	1
			família	1
			forma	1
			fundo	1
			fórum	1
			gabinete	1
		SUJEITO		
			Resultados	12
			anomalia	1
			desactualização	1
			efeito	1
			expulsão	1
			fim	1
			governo	1
			marca	1
			ministério	1
			ocorrência	1
			presidente	1
			situação	1
			sócio	1

(listas e) Reflexão sobre a transitividade em contexto

● Verbos com OD

Resultados	11159
ter	652 de 652
dizer	348 de 446
fazer	376
haver	273
dar	212
querer	141
afirmar	137
conseguir	130
ver	126
saber	120
encontrar	104
receber	92
apresentar	88
considerar	84
mostrar	83
permitir	82
deixar	74
manter	74
pedir	70
tentar	70
levar	67
tornar	65
tratar	58
defender	57

● Verbos com OI

Resultados	2925
dar	116 de 116
falar	76 de 76
tratar	65 de 65
levar	61 de 61
dizer	60 de 60
passar	49 de 49
acusar	43 de 43
ligar	42 de 42
fazer	37 de 37
pedir	36 de 36
dispor	34 de 34
chegar	32 de 32
gostar	30 de 30
participar	30 de 30
destinar	27 de 27
contar	26 de 26
servir	26 de 26
chamar	25 de 25
depende	22 de 22
obrigar	22 de 22
atribuir	21 de 21
permitir	21 de 21
contribuir	20 de 20
dedicar	20 de 20

● Verbos só com OD

Resultados	9963
ter	646 de 646
dizer	354 de 391
fazer	353
haver	270
querer	139
afirmar	132
conseguir	130
ver	125
saber	120
dar	114
encontrar	97
receber	86
considerar	83
apresentar	80
mostrar	76
manter	74
deixar	72
tentar	70
tornar	64
permitir	61
defender	55
decidir	51
ganhar	51
explicar	50

(listas e) transitividade e fluência

● Ol com prep. “de”

Resultados	43
<u>tratar</u>	6
<u>falar</u>	5
<u>passar</u>	5
<u>acusar</u>	4
<u>gostar</u>	2
<u>sair</u>	2
<u>variar</u>	2
<u>abster</u>	1
<u>acompanhar</u>	1
<u>convencer</u>	1
<u>decorrer</u>	1
<u>depende</u>	1
<u>morrer</u>	1
<u>participar</u>	1
<u>prescindir</u>	1
<u>proteger</u>	1
<u>pular</u>	1
<u>receber</u>	1
<u>recuperar</u>	1
<u>reduzir</u>	1
<u>ressarcir</u>	1
<u>resultar</u>	1
<u>vir</u>	1
<u>viver</u>	1

● Ol com prep. “a”

Resultados	37
<u>dizer</u>	4
<u>caber</u>	2
<u>dever</u>	2
<u>adequar</u>	1
<u>aplicar</u>	1
<u>arriscar</u>	1
<u>associar</u>	1
<u>conduzir</u>	1
<u>continuar</u>	1
<u>convencer</u>	1
<u>dedicar</u>	1
<u>destinar</u>	1
<u>ensinar</u>	1
<u>entregar</u>	1
<u>expressar</u>	1
<u>forçar</u>	1
<u>ganhar</u>	1
<u>habituar</u>	1
<u>interessar</u>	1
<u>ligar</u>	1
<u>obrigar</u>	1
<u>passar</u>	1
<u>pedir</u>	1
<u>pertencer</u>	1

(listas e) Fluência – Reflexão

Variantes: BR PT

Resultados 180

[conta](#) 11

[origem](#) 6

[nome](#) 5

[lugar](#) 4

[resposta](#) 4

[volta](#) 4

[apoio](#) 3

[consulta](#) 3

[exemplo](#) 3

[impressão](#) 3

[início](#) 3

[passo](#) 3

[razão](#) 3

[resultado](#) 3

[ar](#) 2

[credibilidade](#) 2

[entrevista](#) 2

[oportunidade](#) 2

[procuração](#) 2

[sentido](#) 2

[tom](#) 2

[vantagem](#) 2

Variantes: BR PT

Resultados 878

[entrevista](#) 23

[conta](#) 19

[lugar](#) 18

[passo](#) 16

[volta](#) 16

[apoio](#) 14

[exemplo](#) 14

[origem](#) 14

[aula](#) 13

[resposta](#) 13

[espaço](#) 12

[sinal](#) 12

[razão](#) 10

[entrada](#) 9

[informação](#) 9

[prazo](#) 9

[declaração](#) 8

[vitória](#) 8

[início](#) 7

[nome](#) 7

[oportunidade](#) 7

[prioridade](#) 7

[atenção](#) 6

[continuidade](#) 6

[explicação](#) 6

Variantes: BR PT

Resultados 11

[questão](#) 2

[automotora](#) 1

[cabeça](#) 1

[discussão](#) 1

[dúvida](#) 1

[espantalho](#) 1

[joelho](#) 1

[necessidade](#) 1

[número](#) 1

[parte](#) 1

Variantes: BR PT

Resultados 51

[dúvida](#) 5

[problema](#) 5

[braço](#) 4

[questão](#) 3

[suspeita](#) 3

[hipótese](#) 2

[arma](#) 1

[bola](#) 1

[cartão](#) 1

[celeuma](#) 1

[cinema](#) 1

[dado](#) 1

[dinheiro](#) 1

[embargo](#) 1

[idéia](#) 1

[imunidade](#) 1

[incompatibilidade](#) 1

[informação](#) 1

[inquiétude](#) 1

[licença](#) 1

[mão](#) 1

[número](#) 1

[obstáculo](#) 1

[olhar](#) 1

[polêmica](#) 1

DAR: Bosque

DAR: FI. Virgem

LEVANTAR: Bosque

LEVANTAR: FI. Virgem

(listas e) Reflexão

Variante: BR PT

Resultados	180
conta	11
origem	6
nome	5
lugar	4
resposta	4
volta	4
apoio	3
consulta	3
exemplo	3
impressão	3
passo	3
razão	3
resultado	3
ar	2
credibilidade	2
entrevista	2
oportunidade	2
procuração	2
sentido	2
tom	2
vantagem	2

Variante: BR PT

Resultados	878
entrevista	23
conta	19
lugar	18
passo	16
volta	16
resposta	14
exemplo	14
origem	14
aula	13
resposta	13
espaço	12
razão	10
entrada	9
informação	9
prazo	9
declaração	8
vitória	8
início	7
nome	7
oportunidade	7
prioridade	7
atenção	6
continuidade	6
explicação	6

Variante: BR PT

Resultados	11
questão	2
automotora	1
cabeça	1
discussão	1
dúvida	1
espantalho	1
joelho	1
necessidade	1
número	1
parte	1

Variante: BR PT

Resultados	51
dúvida	5
problema	5
braço	4
questão	3
suspeita	3
hipótese	2
arma	1
bola	1
cartão	1
celeuma	1
cinema	1
dado	1
dinheiro	1
embargo	1
idéia	1
imunidade	1
incompatibilidade	1
informação	1
inquietação	1
licença	1
mão	1
número	1
obstáculo	1
olhar	1
polêmica	1

Quantos “tipos” de “LEVANTAR”?

Quantas acepções no dicionário?

DAR: Bosque

DAR: Fl. Virgem

LEVANTAR: Bosque

LEVANTAR: Fl. Virgem

(listas e) Reflexão

Variantes: BR PT

Resultados 180

conta	11
origem	6
nome	5
lugar	4
resposta	4
volta	4
apoio	3
consulta	3
exemplo	3
impressão	3
início	3
passo	3
razão	3
resultado	3
ar	2
credibilidade	2
entrevista	2
oportunidade	2
procuração	2
sentido	2
tom	2
vantagem	2

Variantes: BR PT

Resultados 878

entrevista	23
conta	19
lugar	18
passo	16
volta	16
apoio	14
exemplo	14
origem	14
aula	13
resposta	13
espaço	12
sinal	12
razão	10
entrada	9
informação	9
prazo	9
declaração	8
vitória	8
início	7
nome	7
oportunidade	7
prioridade	7
atenção	6
continuidade	6
explicação	6

Variantes: BR PT

Resultados 11

questão	2
automotora	1
cabeça	1
discussão	1
dúvida	1
espantalho	1
leito	1
necessidade	1
número	1
parte	1

Variantes: BR PT

Resultados 51

dúvida	5
problema	5
braço	4
questão	3
suspeita	3
hipótese	2
arma	1
cartão	1
celeuma	1
cinema	1
dado	1
dinheiro	1
idéia	1
imunidade	1
incompatibilidade	1
informação	1
inquietude	1
licença	1
mão	1
número	1
obstáculo	1
olhar	1
polêmica	1

O que é o item lexical ?

Onde está o significado do verbo?

DAR: Bosque

DAR: Fl. Virgem

LEVAR: Bosque

LEVAR: Fl. Virgem

(listas e) Semântica

“INFORMAÇÃO” é objecto de:

Variantes: BR PT

Resultados 17

[haver](#) 3

[ter](#) 2

[armazenar](#) 1

[considerar](#) 1

[contestar](#) 1

[divulgar](#) 1

[fornecer](#) 1

[ganhar](#) 1

[oferecer](#) 1

[pedir](#) 1

[publicar](#) 1

[recolher](#) 1

[trazer](#) 1

[trocar](#) 1

Resultados 90

[dar](#) 9

[ter](#) 9

[obter](#) 8

[receber](#) 5

[desmentir](#) 4

[pedir](#) 4

[fornecer](#) 3

[guardar](#) 3

[haver](#) 3

[prestar](#) 3

[recolher](#) 3

[comentar](#) 2

[confirmar](#) 2

[controlar](#) 2

[encontrar](#) 2

[tornar](#) 2

[trazer](#) 2

[achar](#) 1

[analisar](#) 1

[checar](#) 1

[circular](#) 1

[conseguir](#) 1



[considerar](#) 1

[converter](#) 1

[difundir](#) 1

[entregar](#) 1

“INFORMAÇÃO” é modificada por

Frase	
CF5-4	suficientes
CF35-4	pelo telefone M P)  (034) 60
CF64-3	de mercado
CF64-3	seguras e atuais
CF98-2	que ajudem nas investigações
CF246-1	sobre o abrigo
CF288-7	dando conta de que o ex- presid
CF334-4	sobre as questões da mulher
CF505-1	sobre 2.100 eventos
CF505-2	sobre 1.600 missões
CF518-2	da entidade
CF562-8	pelo tel. M S)  (011) 887-87
CF751-5	pelo tel. 212-2051
CF858-2	que todos ouvem, no sentido de :
CF881-1	comerciais
CF881-1	sobre 1,6 milhão de empresas
CF966-1	de agricultores
CP24-3	de que o dia 23 será um dia « no
CF72-1	sobre a evolução verificada no cu
CP98-3	disponíveis
CP254-3	sobre alternativas de formação e

Bosque

Floresta Virgem

<u>dar</u>	2
<u>ser</u>	2
<u>chegar</u>	1
<u>confirmar</u>	1
<u>existir</u>	1
<u>fornecer</u>	1

“INFORMAÇÃO” é sujeito de...

dar

- ♦ CP98-3: As informações disponíveis dão conta da presença de quatro fragatas, um porta-helicópteros e aviões canadianos a vigiar os barcos ibéricos.
- ♦ CP757-3: Informações do Exército, em termos globais, dão conta de 23 palestinianos mortos e 211 feridos por soldados, desde o dia 1 de Janeiro, na Cisjordânia.

ser

- ♦ CP905-4: Mas o orçamento de US\$ 5 milhões insinua que a informação não é séria.
- ♦ CP803-2: Mas são escassas as informações sobre as várias frentes de combate, desconhecendo-se até que ponto as forças governamentais progredem no terreno.

chegar

- ♦ CP24-3: Ontem, começou a chegar às direcções-gerais (através de uma circular) a informação de que o dia 23 será um dia « normal » para os funcionários públicos.

confirmar

- ♦ CP274-6: Esta informação não foi confirmada pelos agentes da PSP que o acompanharam a estaleiros sitos na Foz, Boavista e no Carvalho.

existir

- ♦ CP837-4: E muitas associações de consumidores, incluindo a Associação Internacional para a Educação das Vítimas do Dalkon Shield, continuam a fazer campanha contra o dispositivo, afirmando que não existe informação sobre os seus riscos.

fornecer

- ♦ CP497-2: Não foram fornecidas informações sobre o balanço destes confrontos.

“INFORMAÇÃO” é sujeito de...

Agrupar resultados por
 função forma lema morfologia texto

elementos e relações inseridos

Elementos:
 clique sobre um elemento para editá-lo ou apagá-lo
 seleccione os elementos a sublinhar na resposta

- função: MV
- forma: vp
- função: SUBJ
- função: H, lema: informação
- proc.: passive
- ...
-
-
-
-

Relações:

1º elemento

não é o mesmo nó que (\$=)

5º elemento

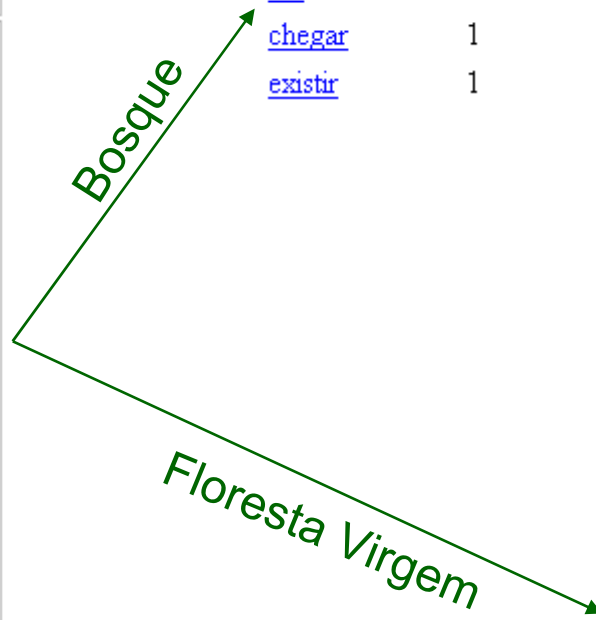
- ♦ 1º elemento está contido imediatamente no 2º elemento
- ♦ 2º elemento está no mesmo nó do 3º elemento
- ♦ 3º elemento contém imediatamente o 4º elemento
- ♦ 1º elemento não é o mesmo nó que o 5º elemento

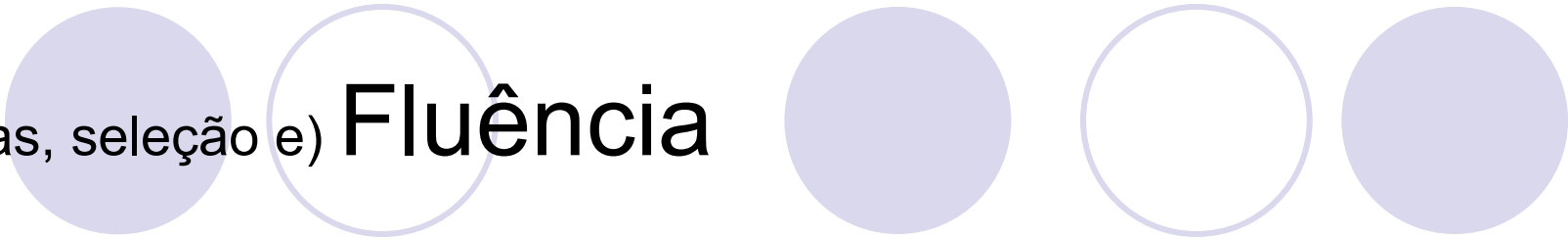
Resultados	
6	
2	dar
2	ser
1	chegar
1	existir

Resultados	
47	
12	ser
5	constar
3	chegar
3	estar
3	surgir
2	ficar
2	indicar
1	aparecer
1	apontar
1	assegurar
1	compensar
1	complicar
1	confirmar
1	existir
1	ir
1	levar
1	mobilizar
1	mostrar
1	nascer
1	ocorrer
1	proceder
1	sugerir
1	tardar
1	vir

Bosque

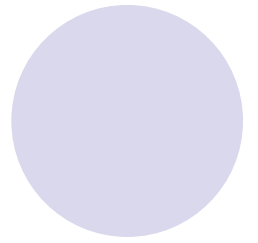
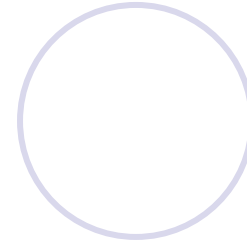
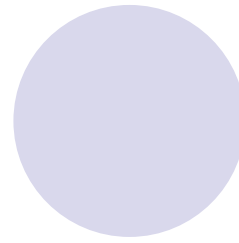
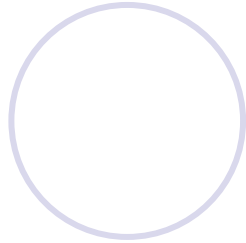
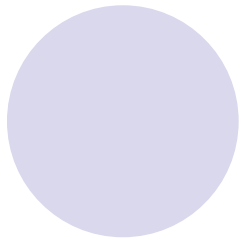
Floresta Virgem



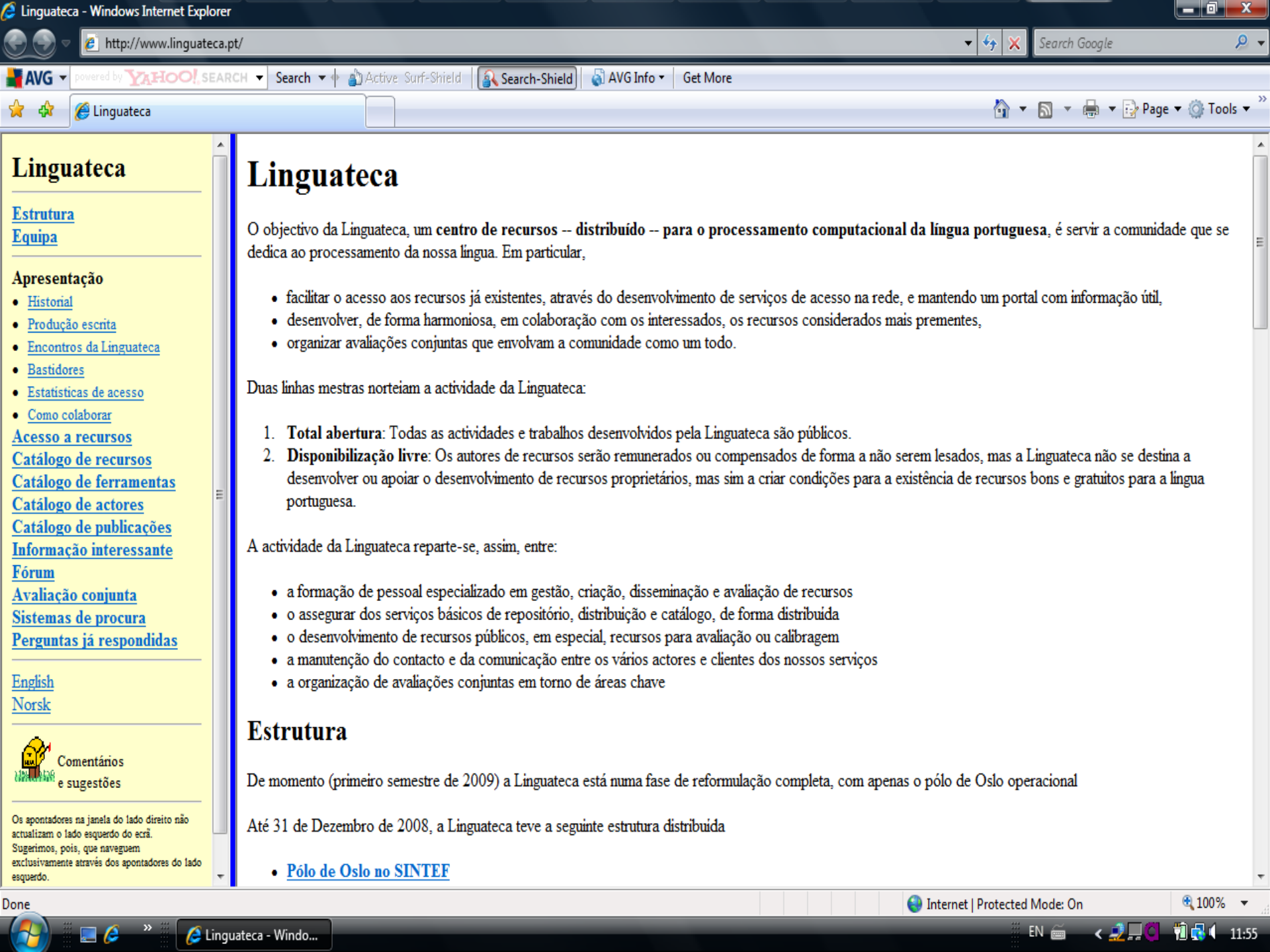


(listas, seleção e) **Fluência**

- *CF9970-2*: As operadoras não _____ a informação, já que o assunto é para elas sigilo comercial.
- *CF6059-5*: O PT, que _____ informações sobre os adversários de Lula, tem dificuldades com Flávio Rocha (PL).
- *CF2812-3*: Os agentes de saúde _____ para a população informações sobre aleitamento materno, soro caseiro e vacinação.
- *CF8647-3*: Os bombeiros _____ a



www.linguateca.pt



Linguateca


[Estrutura](#)
[Equipa](#)

Apresentação

- [História](#)
- [Produção escrita](#)
- [Encontros da Linguateca](#)
- [Bastidores](#)
- [Estatísticas de acesso](#)
- [Como colaborar](#)

[Acesso a recursos](#)
[Catálogo de recursos](#)
[Catálogo de ferramentas](#)
[Catálogo de actores](#)
[Catálogo de publicações](#)
[Informação interessante](#)
[Fórum](#)
[Avaliação conjunta](#)
[Sistemas de procura](#)
[Perguntas já respondidas](#)

[English](#)
[Norsk](#)

 Comentários e sugestões

Os apontadores na janela do lado direito não actualizam o lado esquerdo do ecrã. Sugerimos, pois, que naveguem exclusivamente através dos apontadores do lado esquerdo.

Linguateca

O objectivo da Linguateca, um **centro de recursos -- distribuído -- para o processamento computacional da língua portuguesa**, é servir a comunidade que se dedica ao processamento da nossa língua. Em particular,

- facilitar o acesso aos recursos já existentes, através do desenvolvimento de serviços de acesso na rede, e mantendo um portal com informação útil,
- desenvolver, de forma harmoniosa, em colaboração com os interessados, os recursos considerados mais prementes,
- organizar avaliações conjuntas que envolvam a comunidade como um todo.

Duas linhas mestras norteiam a actividade da Linguateca:

1. **Total abertura:** Todas as actividades e trabalhos desenvolvidos pela Linguateca são públicos.
2. **Disponibilização livre:** Os autores de recursos serão remunerados ou compensados de forma a não serem lesados, mas a Linguateca não se destina a desenvolver ou apoiar o desenvolvimento de recursos proprietários, mas sim a criar condições para a existência de recursos bons e gratuitos para a língua portuguesa.

A actividade da Linguateca reparte-se, assim, entre:

- a formação de pessoal especializado em gestão, criação, disseminação e avaliação de recursos
- o assegurar dos serviços básicos de repositório, distribuição e catálogo, de forma distribuída
- o desenvolvimento de recursos públicos, em especial, recursos para avaliação ou calibragem
- a manutenção do contacto e da comunicação entre os vários actores e clientes dos nossos serviços
- a organização de avaliações conjuntas em torno de áreas chave

Estrutura

De momento (primeiro semestre de 2009) a Linguateca está numa fase de reformulação completa, com apenas o pólo de Oslo operacional

Até 31 de Dezembro de 2008, a Linguateca teve a seguinte estrutura distribuída

- [Pólo de Oslo no SINTEF](#)