

LISTA DE *CORPORA*

NOME DO CORPUS: CORPUSINESC

RESPONSÁVEIS: Diana Santos

INSTITUIÇÃO: INESC

OBJECTIVOS DO CORPUS: Amostra da estrutura sintáctica das frases do português contemporâneo para a sua análise computacional

CONTEÚDO: 1000 frases retiradas de fontes variadas correspondendo a 14873 palavras gráficas identificadas com a categoria sintáctica a que pertencem

DIMENSÃO EM MAIO DE 1996: 1000 frases

ESTADO ACTUAL: Completado em 1992

NÍVEIS DE ANOTAÇÃO: 2 níveis: 1) cada palavra gráfica está classificada como pertencendo às seguintes classes nome/adjectivo, verbo, particípio passado e/ou palavra gramatical, fora do contexto.

2) cada palavra gráfica está classificada como pertencendo a uma das seguintes categorias no contexto em que ocorre: nome, ou adjectivo, ou verbo, ou palavra gramatical, ou particípio passado.

ACESSO:

DOMÍNIO PÚBLICO:

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS: X

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO: Publicado, com a anotação de nível 1, em relatório INESC

OBSERVAÇÕES: Veja-se Santos, D. (ed.) "Processamento de corpora de texto no INESC", Relatório INESC, 1992, para mais informações

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Luzia Helena Wittmann

ENDEREÇO: Grupo de Linguagem Natural do INESC, Rua Alves Redol 9, Apartado 13069, 1000 Lisboa Codex

TELEFONE: 3100303

FAX: 3145843

E-MAIL: luzia@inesc.pt

NOME DO CORPUS: CIPM – Corpus Informatizado de Português Medieval

RESPONSÁVEIS: Equipa do Departamento de Estudos Linguísticos

INSTITUIÇÃO: Faculdade de Ciências Sociais e Humanas-UNL

OBJECTIVOS DO CORPUS: Facilitar a recolha e análise dos dados de fontes textuais medievais a investigadores e docentes de Linguística, de História e de Cultura

CONTEÚDO: Textos não -literários e literários de vários tipos

DIMENSÃO EM MAIO DE 1996: Séc. XIII com 154.122 palavras; Séc. XIV com 244.775 palavras e Séc. XV com 223.095.

ESTADO ACTUAL: em desenvolvimento

NÍVEIS DE ANOTAÇÃO: Em preparação e etiquetagem morfossintáctica

ACESSO: No Departamento de Estudos Linguísticos da FCSM-UNL

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Projectos subsidiados pela JNICT e pelo PRAXIS XXI para tratamento automático e desenvolvimento do CIPM

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Maria Francisca Xavier

ENDEREÇO: FCSM-UNL, Av. de Berna, 26 – 1050 Lisboa

TELEFONE: 7933919 ou 7933519 Ext: 478

FAX: 7977759

E-MAIL: FXavier @FCSH.UNL.Pt

NOME DO CORPUS: Moda 60-90

RESPONSÁVEIS: Anabela Carvalho

INSTITUIÇÃO:

OBJECTIVOS DO CORPUS: Servir de base a uma dissertação de mestrado em Lexicologia e Lexicografia – "Nomes de cor num corpus especializado. Moda e Vestuário nos últimos 30 anos".

CONTEÚDO: Textos retirados de periódicos, de catálogos de moda e de exposições de vestuário e moda, e livros dedicados a esse tema e de manuais escolares da disciplina de têxteis

DIMENSÃO EM MAIO DE 1996: 700 mil palavras

ESTADO ACTUAL: Informatizado

NÍVEIS DE ANOTAÇÃO:

ACESSO: Autorizado para consulta, mediante pedido analisado caso a caso

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Corpus integrado, em parte, no Corpus de Referência do Português Contemporâneo – CLUL.

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Anabela Carvalho
ENDEREÇO: Rua General João de Almeida, 52-2º Dtº, 1300 Lisboa
TELEFONE: 3621798
FAX:
E-MAIL:

NOME DO CORPUS: Corpus Alsácia

RESPONSÁVEIS: Maria Antónia Ramos Coelho da Mota

INSTITUIÇÃO: CLUL/FLUL

OBJECTIVOS DO CORPUS: Geral: Francês falado por imigrantes da primeira geração na Alsácia; específico: estudo de questões de morfologia (associada à fonologia e à fonética) com o objectivo de testar as zonas mais sensíveis do contacto do Português. LM e do Francês LE.

CONTEÚDO: cerca de 10 horas de gravação de cerca de 30 indivíduos dos dois sexos; conversa livre, sobretudo centrada na experiência da imigração (corpus reunido entre 1983 e 85).

DIMENSÃO EM MAIO DE 1996:

ESTADO ACTUAL:

NÍVEIS DE ANOTAÇÃO: Transcrição fonética larga

ACESSO: 26 entrevistas transcritas incluídas no vol. II da minha tese de doutoramento

DOMÍNIO PÚBLICO: transcrições

CONSULTA: gravação

DISPONÍVEL PARA FINS NÃO COMERCIAIS: X

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

Pessoa a contactar para mais informações: Maria Antónia Mota
Endereço: CLUL Av. 5 de Outubro, 85, 6.º – 1050, Lisboa
Telefone:
Fax:
E-mail:

NOME DO CORPUS: Astro

RESPONSÁVEIS: Paula Marquez Neto

INSTITUIÇÃO: CLUL

OBJECTIVOS DO CORPUS: Análise linguística de discurso especializado

CONTEÚDO: Discurso da Astronomia

DIMENSÃO EM MAIO DE 1996: 560.000palavras

ESTADO ACTUAL: on-line, corrigido

NÍVEIS DE ANOTAÇÃO: mínimos

ACESSO: Livre

DOMÍNIO PÚBLICO:

CONSULTA: Sim

DISPONÍVEL PARA FINS NÃO COMERCIAIS: Sim

PRODUTO COMERCIAL: Não

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: O corpus Astro foi cedido ao CRPC e faz parte deste, hoje em dia.

Pessoa a contactar para mais informações: Paula Neto
Endereço: Rua Freitas Reis, 8-1º Esqº, 2750 Cascais
Telefone: 4831192
Fax: 4831192
E-mail: pgc@telepac.pt

*********NOME DO CORPUS:** Leiria**RESPONSÁVEIS:** Isabel Leiria**INSTITUIÇÃO:** FLUL**OBJECTIVOS DO CORPUS:** Investigar a aquisição por falantes de Português europeu, língua não materna, dos aspectos verbais expressos pelo Pretérito Perfeito e Imperfeito**CONTEÚDO:** 218 documentos produzidos por 168 informantes falantes de 16 línguas maternas diferentes, com idades compreendidas entre os 18 e os 55 anos, de ambos os sexos. Cada documento consiste no relato de uma de três narrativas ouvidas pelos informantes aquando da prova de compreensão oral e produção escrita do exame do curso básico de língua Portuguesa.**DIMENSÃO EM MAIO DE 1996:****ESTADO ACTUAL:****NÍVEIS DE ANOTAÇÃO:****ACESSO:** constitui o 2º volume da minha tese de mestrado**DOMÍNIO PÚBLICO:** Sim**CONSULTA:****DISPONÍVEL PARA FINS NÃO COMERCIAIS:** Sim**PRODUTO COMERCIAL:****OUTRAS FORMAS DE ACESSO:****OBSERVAÇÕES:****PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES:** Isabel Leiria**ENDEREÇO:** Rua Alferes Barrilaro Ruas, 50-4º T, 1800 Lisboa**TELEFONE:** 8530880 8536373**FAX:****E-MAIL:**

NOME DO CORPUS: Panorama do Português oral de Maputo

RESPONSÁVEIS: Perpétua Gonçalves e Christopher Stroud

INSTITUIÇÃO: Instituto Nacional de Desenvolvimento da Educação

OBJECTIVOS DO CORPUS: Criação de um banco de dados para apoio em instituições de ensino

CONTEÚDO: Entrevistas individuais a 100 informantes e encontros informais entre cada informante e 2 ou 3 parceiros

DIMENSÃO EM MAIO DE 1996:

ESTADO ACTUAL: Gravação de dados orais: concluída; transcrição ortográfica: 50 entrevistas e 30 encontros; informatização (processador de texto "Word 6" e tratamento através do "Oxford Concordance Program (OCP)": 50 entrevistas e 10 encontros

NÍVEIS DE ANOTAÇÃO:

ACESSO: Consulta disponível para fins não comerciais; eventuais acordos com instituições/projectos

DOMÍNIO PÚBLICO:

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: António Tuzine

ENDEREÇO: INDE Rua Comte. Augusto Cardoso, 81-
Maputo

TELEFONE: 258 1 421701/2

FAX: 258 1 426797

E-MAIL: linguist@inde.uem.mz

*********NOME DO CORPUS:** CPE VAR**RESPONSÁVEIS:** Ernesto d'Andrade / Maria Celeste Rodrigues**INSTITUIÇÃO:** FLUL/CLUL**OBJECTIVOS DO CORPUS:** Variação Linguística no PE**CONTEÚDO:** gravações de 3 entrevistas de aproximadamente 80 minutos, por sexo, por escalão etário e sociocultural com naturais de Lisboa e de Braga.. As gravações incluem leitura e fala .**DIMENSÃO EM MAIO DE 1996:** 30 entrevistas em Lisboa; 30 em Braga.**ESTADO ACTUAL:** Em elaboração.**NÍVEIS DE ANOTAÇÃO:** Actualmente, nenhum.**ACESSO:****DOMÍNIO PÚBLICO:****CONSULTA:** Condicionada a autorização dos responsáveis.**DISPONÍVEL PARA FINS NÃO COMERCIAIS:****PRODUTO COMERCIAL:****OUTRAS FORMAS DE ACESSO:** Disponível apenas depois da apresentação da tese de doutoramento de Celeste Rodrigues**OBSERVAÇÕES:****PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES:** Celeste Rodrigues**ENDEREÇO:** FLUL**TELEFONE:** 7965162**FAX:** 7960063**E-MAIL:***********NOME DO CORPUS:** Corpus de aquisição do Português Europeu (Fonologia)**RESPONSÁVEIS:** Isabel Hub Faria / Maria João Freitas**INSTITUIÇÃO:** Laboratório de Psicolinguística da FLUL

OBJECTIVOS DO CORPUS: Estudo da aquisição do Português Europeu.

CONTEÚDO: Sessões de gravação com 8 crianças entre os 0;11 e os 3;07 anos de idade, efectuadas em casa de cada uma das crianças, em sessões mensais de cerca de 60 minutos.

DIMENSÃO EM MAIO DE 1996: Cerca de 100 horas de gravação.

ESTADO ACTUAL: Em elaboração.

NÍVEIS DE ANOTAÇÃO: transcrição ortográfica e transcrição fonética larga.

ACESSO: Disponível para consulta após a realização da dissertação de doutoramento de M. João Freitas

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: M. João Freitas

ENDEREÇO: Laboratório de Psicolinguística/FLUL

TELEFONE: 7965162

FAX: 7960063

E-MAIL:

NOME DO CORPUS: Subcorpus de interacções mãe/criança do corpus de aquisição do Português Europeu

RESPONSÁVEIS: Dília Ramos / Isabel Hub Faria

INSTITUIÇÃO: Laboratório de Psicolinguística da FLUL

OBJECTIVOS DO CORPUS: Análise da linguagem infantil; análise da linguagem adulta utilizada nas interacções com as crianças (baby talk).

CONTEÚDO: Interações verbais entre a mãe e o filho, gravadas em áudio, em três situações diferentes: banho, refeição e brincadeira

DIMENSÃO EM MAIO DE 1996: 30 díades mãe-filho (cerca de 45 horas de gravação)

ESTADO ACTUAL: Totalmente transcrito e codificado, de acordo com as normas CHILDES.

NÍVEIS DE ANOTAÇÃO: Linha principal; linhas dependentes variáveis

ACESSO: restrito até conclusão da dissertação de doutoramento

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS: X

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Consulta mediante prévia autorização

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Dília Ramos

ENDEREÇO: FLUL

TELEFONE: 7965162

FAX: 7960063

E-MAIL: diliarp@mail.telepac.pt

NOME DO CORPUS: Batoréo 94

RESPONSÁVEIS: Hanna Batoréo (sob orientação da Prof^a Isabel Hub Faria)

INSTITUIÇÃO: Laboratório de Psicolinguística da FLUL

OBJECTIVOS DO CORPUS: Reunir narrativas de crianças e adultos, falantes nativos do PE

CONTEÚDO: 60 narrativas adultas e 60 narrativas infantis (faixa etária de 5, 7 e 10 anos) transcritas no sistema CHILDES

DIMENSÃO EM MAIO DE 1996: 120 narrativas (cerca de duas horas de gravação)

ESTADO ACTUAL:

NÍVEIS DE ANOTAÇÃO: Transcrição e codificação segundo o sistema CHILDES

ACESSO: Público (a partir da defesa da dissertação de doutoramento)

DOMÍNIO PÚBLICO:

CONSULTA: No sistema CHILDES (Laboratório de Psicolinguística)

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Sobre o funcionamento do corpus CHILDES cf. Faria & Batoréo (1994) na RILP nº 10

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Hanna Batoréo

ENDEREÇO:

TELEFONE: 7145584

FAX:

E-MAIL: Batoreo@ip.pt

NOME DO CORPUS: 100 Falantes Adultos do Português Europeu

RESPONSÁVEIS: Isabel Hub Faria

INSTITUIÇÃO: Laboratório de Psicolinguística da FLUL

OBJECTIVOS DO CORPUS: Amostra da população falante do PE, organizada por estrato socioprofissional, sexo e local de produção.

CONTEÚDO: Entrevistas estruturadas.

DIMENSÃO EM MAIO DE 1996:

ESTADO ACTUAL: Em fase de codificação no sistema CHILDES

NÍVEIS DE ANOTAÇÃO: CHILDES

ACESSO: Final de 1997

DOMÍNIO PÚBLICO:
CONSULTA:
DISPONÍVEL PARA FINS NÃO COMERCIAIS:
PRODUTO COMERCIAL:
OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Isabel Hub Faria
ENDEREÇO: Lab. de Psicolinguística, DLGR, FLUL
TELEFONE: 796 51 62
FAX:
E-MAIL: ihf@di.fc.ul.pt

NOME DO CORPUS: CPE FACES – '*Corpus* de Português Europeu Falado por Adolescentes em Contexto Escolar'

RESPONSÁVEIS: Ana Isabel Mata

INSTITUIÇÃO: Departamento de Linguística Geral e Românica, Faculdade de Letras da Universidade de Lisboa

OBJECTIVOS DO CORPUS: Reunir um conjunto de textos falados exemplificativos de actividades de produção oral familiares para os alunos no contexto escolar. Tem como principal finalidade o estudo de aspectos prosódicos do Português Europeu e a caracterização de estilos de fala.

CONTEÚDO: Integra gravações audio e video de diálogos e exposições (não lidas) de adolescentes (dos dois sexos) do último ano da escolaridade obrigatória, em Escolas Secundárias de Lisboa.

DIMENSÃO EM MAIO DE 1996: Actividades de produção de fala diferentes por informante: 4. Total de informantes: – Adolescentes (com idades entre os 14 e os 17 anos): 25; – Adultas: 2 professoras de Língua Portuguesa. Equivale a um total de aproximadamente 15 h de gravação.

ESTADO ACTUAL: A transcrição ortográfica e "miscelânea" dos dados de cerca de 60% dos informantes está completa. A transcrição prosódica de uma parte dos dados de cerca de 20% dos informantes e a constituição de uma base de dados com a informação dos múltiplos níveis de transcrição do *corpus* encontra-se em curso.

NÍVEIS DE ANOTAÇÃO: Transcrição ortográfica, prosódica e comentários "miscelâneos" (riso, tosse, dsfluências, etc.). O formato de transcrição adaptado para o CPE Faces segue, como referência de base, o sistema de transcrição prosódica ToBI (*Tones and Break Indices*).

ACESSO: O *corpus* só estará disponível depois da conclusão do projecto de doutoramento para o qual foi propositadamente recolhido.

DOMÍNIO PÚBLICO:

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Nas *Actas* do XI Encontro Nacional da APL (1995) encontra-se uma apresentação preliminar do CPE FACES.

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Ana Isabel Mata

ENDEREÇO: Departamento de Linguística Geral e Românica

Faculdade de Letras da Universidade de Lisboa

Cidade Universitária

1699 Lisboa Codex

TELEFONE: 7965162

FAX: 7960063

E-MAIL:

NOME DO CORPUS: CORPUS DE REFERÊNCIA DO PORTUGUÊS CONTEMPORÂNEO

RESPONSÁVEIS: J. Malaca Casteleiro e M^a Fernanda Bacelar do Nascimento

INSTITUIÇÃO: Centro de Linguística da Universidade de Lisboa

OBJECTIVOS DO CORPUS: Realização de estudos linguísticos sobre dados empíricos

CONTEÚDO: oral: (+espontâneo; + formal): Portugal – Brasil – PALOPS – Macau

escrito: *Lit:* Portugal, Brasil, PALOPS, Macau. *Jornal:* Portugal, Brasil, C. Verde; *Político:* Portugal, C. Verde; *Técnico:* / *Científico,* *Didáctico, Jurídico, Folheto e outros:* Portugal.

DIMENSÃO EM MAIO DE 1996: 45 milhões de palavras.

ESTADO ACTUAL: Acessível com software adequado à pesquisa lexical quantitativa, estatística (em implementação), concordâncias e combinatórias lexicais; análises morfossintáticas.

NÍVEIS DE ANOTAÇÃO: Representação do texto: cerca de 13 milhões de palavras; amostragens com anotação morfossintáctica automática; pequenas amostragens com anotação sintáctica.

ACESSO:

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO: pedidos a analisar caso a caso.

OBSERVAÇÕES:

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: M^a Fernanda Bacelar do Nascimento

ENDEREÇO: Av. 5 de Outubro, 85, 6^o, 1050 Lisboa

TELEFONE: (351 1) 796 71 10

FAX: (351 1) 7939299

E-MAIL: clul@dial-up.rccn.net

NOME DO CORPUS: TEXTOS METALINGUÍSTICOS PORTUGUESES DO SÉCULO XVI

RESPONSÁVEIS: M^a Helena Pinto Novais Paiva

INSTITUIÇÃO: Faculdade de Letras da U. P.

OBJECTIVOS DO CORPUS: Relações entre Variação e Mudança; Evolução da Língua Quinhentista.

CONTEÚDO:

DIMENSÃO EM MAIO DE 1996: 65730 ocorrências.

ESTADO ACTUAL: Concluído o tratamento informático

NÍVEIS DE ANOTAÇÃO: Análise Linguística e Metalinguística

ACESSO: Temporariamente privado.

DOMÍNIO PÚBLICO:

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Produzido um Índice de Vocábulo Seguido de Listas Alfabéticas de Formas Não Pertencentes ao Português

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: M^a Helena Paiva

ENDEREÇO: R. de Vilarinho, 913, 4100 Porto

TELEFONE: (01)618 05 18

FAX:

E-MAIL:

NOME DO CORPUS: EUROM.1 (Português)

RESPONSÁVEIS: Isabel Trancoso (INESC)

INSTITUIÇÃO: INESC, CLUL e Portugal Telecom

OBJECTIVOS DO CORPUS: Investigação Básica em Reconhecimento e Síntese de Fala

CONTEÚDO: passagens, frases e números ditos por 60 falantes; CVCs por falantes

DIMENSÃO EM MAIO DE 1996: 6500 números, 2200 frases, 1260 palavras ou logátomos

ESTADO ACTUAL: Concluído

NÍVEIS DE ANOTAÇÃO: 2 níveis: ortográfico e fonético (transcrição fonética larga)

ACESSO: Ainda em fase de neociação

DOMÍNIO PÚBLICO:
CONSULTA:
DISPONÍVEL PARA FINS NÃO COMERCIAIS:
PRODUTO COMERCIAL:
OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Corpus de fala recolhido no âmbito do projecto europeu SAM A

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Isabel Trancoso
ENDEREÇO: INESC; R. Alves Redol, 9, 1000 Lisboa
TELEFONE: 310 02 68
FAX: 314 52 43
E-MAIL: Isabel.Trancoso@inesc.pt

NOME DO CORPUS: BDFALA

RESPONSÁVEIS: Isabel Trancoso (INESC) e M. Céu Viana (CLUL)

INSTITUIÇÃO: INESC e CLUL

OBJECTIVOS DO CORPUS: Investigação básica sobre o Português falado e treino e teste de sistemas de reconhecimento e síntese de fala

CONTEÚDO: Logátomos, palavras isoladas, frases, textos recolhidos em situação laboratorial (10 falantes) excertos de entrevistas, debates e noticiários

DIMENSÃO EM MAIO DE 1996:

ESTADO ACTUAL: Recolha e anotação em curso

NÍVEIS DE ANOTAÇÃO: Variáveis em função do subcorpus: mínimo 2, máximo 6

ACESSO: Ainda por definir

DOMÍNIO PÚBLICO:
CONSULTA:
DISPONÍVEL PARA FINS NÃO COMERCIAIS:
PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Corpus de fala recolhido no âmbito do Projecto BDFALA (JNICT-PLUS/C/LIN/(801/93)

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Isabel Trancoso ou M. Céu Viana

ENDEREÇO: INESC, R. Alves Redol, 9 – 1000 Lisboa

CLUL, Av 5 de Outubro, 85-6º – 1000 Lisboa

TELEFONE: 3100268 7967110

FAX: 3145243 793753

E-MAIL: Isabel.Trancoso@inesc.pt – Ceu.Viana@inesc.pt

NOME DO CORPUS: ONOMASTICA

RESPONSÁVEIS: Isabel Trancoso (INESC)

INSTITUIÇÃO: INESC e CLUL

OBJECTIVOS DO CORPUS: Léxico de pronúncia multilingue para nomes próprios, siglas e acrónimos

CONTEÚDO: Nomes próprios, siglas e acrónimos

DIMENSÃO EM MAIO DE 1996: 100 000 entradas

ESTADO ACTUAL: *concluído*

NÍVEIS DE ANOTAÇÃO: Ortográfico e fonético (transcrição fonética larga com marcas de silabificação)

ACESSO: Em fase de negociação

DOMÍNIO PÚBLICO:

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Corpus de pronúncia recolhido no âmbito do projecto europeu ONOMASTICA (Proj.LRE-61-004) para o Português Europeu.

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Isabel Trancoso
ENDEREÇO: INESC, R. Alves Redol, 9 – 1000 Lisboa
TELEFONE: 3100268
FAX: 3145243
E-MAIL: Isabel.Trancoso@inesc.pt

NOME DO CORPUS: SPEECHDAT (Português)

RESPONSÁVEIS: Isabel Trancoso (INESC)

INSTITUIÇÃO: Portugal Telecom

OBJECTIVOS DO CORPUS: Recolha de chamadas telefónicas para treino e teste de reconhecedores de fala

CONTEÚDO: 5000 chamadas telefónicas, com 40 itens cada

DIMENSÃO EM MAIO DE 1996: 1000 chamadas

ESTADO ACTUAL: 1000 chamadas recolhidas e anotadas em 3 CD-ROM

NÍVEIS DE ANOTAÇÃO: 1 (ortográfico)

ACESSO: Ainda em fase de negociação

DOMÍNIO PÚBLICO:

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Corpus fala recolhido no âmbito do projecto europeu SPEECHDAT

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Rui Chaves
ENDEREÇO: PORTUGAL TELECOM
RUA DE ENTRECAMPOS, Nº 28 – 1700 LISBOA
TELEFONE: 500301
FAX:
E-MAIL: RChaves@telepac.pt

NOME DO CORPUS: PF-FONE

RESPONSÁVEIS: Maria do Céu Viana (CLUL) e Ernesto d'Andrade (CLUL/FLUL)

INSTITUIÇÃO: CLUL

OBJECTIVOS DO CORPUS: Estudos estatísticos de ocorrência de grafemas, fones, padrões silábicos e padrões de palavras; treino e teste de sistemas de conversão grafia-som

CONTEÚDO: 26000 palavras correspondentes a cerca de 750000 formas de citação e formas flexionadas

DIMENSÃO EM MAIO DE 1996: 26000 entradas correspondentes a cerca de 750000 formas de citação e formas flexionadas

ESTADO ACTUAL: Concluído

NÍVEIS DE ANOTAÇÃO: Ortográfico e fonético (transcrição fonética larga com silabificação e transcrição estreita com máximo de redução)

ACESSO:

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Corpus de pronúncia construído a partir do conjunto de entrevistas realizadas para o Português Fundamental

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: M^a Céu Viana

ENDEREÇO: CLUL, Av. 5 de Outubro, 85-6^o – 1000

Lisboa

TELEFONE: 7967110

FAX: 7939299

E-MAIL: mcv@inesc.pt

*********NOME DO CORPUS:** ATLAS LINGUARUM EUROPAE**RESPONSÁVEIS:** Manuela Barros**INSTITUIÇÃO:** Centro de Linguística da Universidade de Lisboa**OBJECTIVOS DO CORPUS:** Realização de um Atlas multilingue e plurinacional do Continente Europeu**CONTEÚDO:** Respostas a um questionário lexical de 546 perguntas**DIMENSÃO EM MAIO DE 1996:** 53 Inquéritos (rede total do território continental português para o referido projecto)**ESTADO ACTUAL:** Suporte magnético, transcrição nos cadernos de inquérito e tratamento etimológico das formas lexicais obtidas**NÍVEIS DE ANOTAÇÃO:** Transcrição fonética**ACESSO:****DOMÍNIO PÚBLICO:****CONSULTA:** X**DISPONÍVEL PARA FINS NÃO COMERCIAIS:****PRODUTO COMERCIAL:****OUTRAS FORMAS DE ACESSO:****OBSERVAÇÕES:****PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES:** Manuela Barros**ENDEREÇO:** CLUL, Av. 5 de Outubro, 85, 6º, 1050 Lisboa**TELEFONE:** (01) 7967110**FAX:** (01) 799299**E-MAIL:***********NOME DO CORPUS:** ATLAS LINGUÍSTICO DE PORTUGAL E DA GALIZA**RESPONSÁVEIS:** Equipa do grupo de estudos de Dialectologia**INSTITUIÇÃO:** Centro de Linguística da Universidade de Lisboa

OBJECTIVOS DO CORPUS: realização de um Atlas Linguístico Nacional

CONTEÚDO: respostas a um questionário linguístico, cujos conceitos estão agrupados por campos semânticos

DIMENSÃO EM MAIO DE 1996: 71 inquéritos com respostas a cerca de 3500 conceitos e 115 inquéritos com respostas a cerca de 2500 conceitos

ESTADO ACTUAL: Suporte magnético e transcrição parcial nos cadernos de inquérito.

NÍVEIS DE ANOTAÇÃO: Transcrição fonética

ACESSO:

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

PESSOAS A CONTACTAR PARA MAIS INFORMAÇÕES: Luísa Segura da Cruz, Gabriela Vitorino, Manuela Barros e João Saramago

ENDEREÇO: CLUL, Av. 5 de Outubro, 85, 6º, 1050 Lisboa

TELEFONE: (01) 7967110

FAX: (01) 7939299

E-MAIL:

NOME DO CORPUS: ATLAS LINGUÍSTICO DO LITORAL PORTUGUÊS

RESPONSÁVEIS: Gabriela Vitorino

INSTITUIÇÃO: Centro de Linguística da Universidade de Lisboa

OBJECTIVOS DO CORPUS: Estudo da Linguagem Específica dos Pescadores do Litoral Português

CONTEÚDO: Respostas a um questionário linguístico que contempla a fauna e a flora marítimas, a pesca e processos de pesca, embarcações e navegação, a campanha e comercialização do pescado, a costa e o mar, fenómenos atmosféricos .

DIMENSÃO EM MAIO DE 1996: 32 inquéritos consagrados à fauna e flora (cerca de 200 perguntas); 8 inquéritos completos (cerca de 1400 perguntas).

ESTADO ACTUAL: Suporte magnético e transcrição nos cadernos de inquérito

NÍVEIS DE ANOTAÇÃO: Transcrição fonética

ACESSO:

DOMÍNIO PÚBLICO:

CONSULTA: X

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Gabriela Vitorino

ENDEREÇO: CLUL, Av. 5 de Outubro, 85, 6º, 1050 Lisboa

TELEFONE: (01) 7967110

FAX: (01) 7939299

E-MAIL:

NOME DO CORPUS: NATURA – PÚBLICO

RESPONSÁVEIS: José João Dias de Almeida

INSTITUIÇÃO: Universidade do Minho – Dptº de Informática – Projº NATURA

OBJECTIVOS DO CORPUS: Em aberto

CONTEÚDO: Corpus jornalístico – constituído pelos dois primeiros parágrafos de muitos milhares de artigos que apareceram no "Público"

DIMENSÃO EM MAIO DE 1996: cerca de 6 milhões de palavras

ESTADO ACTUAL: Arranjado de modo a eliminar ruído associado a formação e fotocomposição.

NÍVEIS DE ANOTAÇÃO: Nenhum

ACESSO:

DOMÍNIO PÚBLICO: X

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS:

PRODUTO COMERCIAL:

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Dispomos ainda de um outro *corpus*, não disponível, de cerca de 17.000.000 de palavras

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: José João Dias de Almeida

ENDEREÇO:

TELEFONE:

FAX:

E-MAIL: jj@di.uminho.pt

NOME DO CORPUS: NATURA – PÚBLICO – Etiquetado

RESPONSÁVEIS: José João Dias de Almeida

INSTITUIÇÃO: Universidade do Minho – Dptº de Informática – Projº NATURA

OBJECTIVOS DO CORPUS: Treino de um etiquetador

CONTEÚDO:

DIMENSÃO EM MAIO DE 1996: cerca de 4 mil palavras

ESTADO ACTUAL: Anotado

NÍVEIS DE ANOTAÇÃO: Anotação morfossintáctica

ACESSO:

DOMÍNIO PÚBLICO: X

CONSULTA:

DISPONÍVEL PARA FINS NÃO COMERCIAIS:
PRODUTO COMERCIAL:
OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES: Existem ainda uma lista para corrector ortográfico (ispell) e outra para analisador morfológico (jspell) de cerca de 40.000 radicais cada

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: José João Dias de Almeida

ENDEREÇO:

TELEFONE:

FAX:

E-MAIL: jj@di.uminho.pt

miguel@ci.uminho.pt

NOME DO CORPUS: PROPER – (PE)

RESPONSÁVEIS: Amália Andrade

INSTITUIÇÃO: Centro de Linguística da Universidade de Lisboa

OBJECTIVOS DO CORPUS: Estudo da relação entre a produção e a percepção da fala.

CONTEÚDO: Material de fala (produção) e dados perceptivos de uma amostragem de informantes nativos do PE.

DIMENSÃO EM MAIO DE 1996: 30%

ESTADO ACTUAL: 30%

NÍVEIS DE ANOTAÇÃO:

ACESSO: N

DOMÍNIO PÚBLICO: N

CONSULTA: N

DISPONÍVEL PARA FINS NÃO COMERCIAIS: N

PRODUTO COMERCIAL: N

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Amália Andrade
ENDEREÇO: CLUL, Av. 5 de Outubro, 85, 6º, 1050 Lisboa
TELEFONE: (01) 7967110
FAX: (01) 7939299
E-MAIL:

NOME DO CORPUS: RED – I

RESPONSÁVEIS: Amália Andrade

INSTITUIÇÃO: Centro de Linguística da Universidade de Lisboa

OBJECTIVOS DO CORPUS: Estudo Acústico de fenómenos fonéticos do PE como lenição vocálica, desvozeamento, elisão, efeitos coarticulatórios intra e trans-silábicos.

CONTEÚDO: Palavras isoladas e um conjunto de 250 frases (presentemente) produzidas por 10 adultos (M e F); material espontâneo.

DIMENSÃO EM MAIO DE 1996: 60%

ESTADO ACTUAL: O material recolhido está digitalizado quase na totalidade

NÍVEIS DE ANOTAÇÃO: Parcial: 2 níveis, presentemente.

ACESSO:

DOMÍNIO PÚBLICO: N

CONSULTA: N

DISPONÍVEL PARA FINS NÃO COMERCIAIS: N

PRODUTO COMERCIAL: N

OUTRAS FORMAS DE ACESSO:

OBSERVAÇÕES:

PESSOA A CONTACTAR PARA MAIS INFORMAÇÕES: Amália Andrade
ENDEREÇO: CLUL, Av. 5 de Outubro, 85, 6º, 1050 Lisboa
TELEFONE: (01) 7967110
FAX: (01) 7939299
E-MAIL: