

Resumo da actividade da Linguateca de 15 de Maio de 2003 a 15 de Dezembro de 2006

Diana Santos

Dividimos este relatório (também em [231]) nas três vertentes do modelo IRA (informação, recursos e avaliação) seguido pela Linguateca, apresentando no final algumas estatísticas tentando dar uma medida do impacto da Linguateca no período de 15 de Maio de 2003 até 15 de Dezembro de 2006. Convém contudo salientar que esta calendarização é artificial e que todo o trabalho aqui descrito é a continuação lógica do trabalho iniciado em 2000, e relatado pormenorizadamente em [22]. Um breve relatório intercalar [85] foi também tornado público em Junho de 2005. O presente relatório foi redigido por vários membros do projecto, cujo nome aparece sob as secções respectivas.

1. Avaliação

Possivelmente o contributo mais valioso do ponto de vista científico dado pela Linguateca foi a organização -- ou participação na organização -- de avaliações conjuntas focando o processamento computacional da língua portuguesa.

1.1. Morfolimpiadas

Os resultados das Morfolimpiadas, a primeira avaliação conjunta para o português cuja organização foi principalmente levada a cabo durante o período anterior deste projecto, foram anunciados no Encontro Avalon, que teve lugar no dia 27 de Junho de 2003 como satélite do PROPOR 2003, e teve como participantes sete diferentes sistemas de Portugal, do Brasil e da Dinamarca, permitindo além disso divulgar este paradigma ao resto da comunidade científica da área. Esta iniciativa levou à publicação de um livro sobre este paradigma pela IST Press [275] e a vários artigos e apresentações [9, 12, 14, 16-18, 25, 37, 240-244, 253, 256-257, 261, 268, 270].

1.2. CLEF

Luís Costa

O CLEF (*Cross-Language Evaluation Forum*, Forum de avaliação entre várias línguas) é uma série de avaliações conjuntas pretendendo promover a pesquisa e desenvolvimento na área de recolha de informação (RI) entre várias línguas (cruzada), veja-se <http://www.clef-campaign.org/> e <http://www.linguateca.pt/CLEF/>. Desde 2004, a Linguateca colabora na organização do CLEF para a parte portuguesa, disponibilizando as colecções, participando na criação de material para ser usado na avaliação e encarregando-se da classificação dos resultados dos participantes. Esta colaboração enquadra-se naturalmente na vertente da Linguateca de promoção e organização de avaliações conjuntas na área do processamento computacional da língua portuguesa, embora aqui num contexto internacional multilingue.

No CLEF 2004, a actividade organizativa centrou-se nas pistas de recolha de informação e de resposta automática a perguntas (RAP), criando-se respectivamente 50 tópicos em português, e uma lista de 700 perguntas e respostas em português, em cooperação com os restantes organizadores do CLEF.

No CLEF 2005, para além da participação na organização nas pistas de RI e de RAP, participámos também pela primeira vez na organização do WebCLEF, ImageCLEF e GeoCLEF, respectivamente criando tópicos, traduzindo legendas e traduzindo tópicos geográficos.

No CLEF 2006, a Linguateca colabora na organização das pistas de RI, RAP e GeoCLEF, nesta última já como organizadora de pleno direito. Além disso, no encontro anual do CLEF, a Linguateca apresentou a proposta de uma pista piloto para o futuro CLEF 2007, o QoIA [206], que foi aceite e estamos presentemente a organizar.

Em 2004 obtivemos a autorização de usar os documentos do Público (notícias completas) de 1994-1995, em 2005 conseguimos autorização para usar os textos de 1994 e 1995 da Folha de S. Paulo, e são essas as duas colecções utilizadas desde então.

O trabalho de organização do CLEF teve dois resultados de grande importância: por um lado, o número de sistemas para o português avaliados tem crescido de ano para ano (veja-se a extensa lista de publicações sobre o assunto); por outro lado, foi criada e publicamente disponibilizada a colecção CHAVE, <http://www.linguateca.pt/CHAVE/>, que permite a todas as pessoas que trabalham em RI ou RAP em português avaliarem os seus sistemas, mesmo que não tenham oportunidade de participar no CLEF. Finalmente, a própria participação na organização deu origem a uma série de publicações e apresentações [34, 42-45, 95-98, 101-104, 152, 195-206, 216, 251-252, 272, 284, 291-292].

1.3 HAREM

Nuno Cardoso

O HAREM constituiu a primeira avaliação (conjunta) para sistemas de reconhecimento de entidades mencionadas (REM) em textos em português [77, 107, 119, 122, 126-127, 130, 144-147, 149, 154, 208, 226],

no âmbito da responsabilidade da Linguateca de organizar avaliações conjuntas para a comunidade científica interessada no processamento computacional do português. O HAREM foca a tarefa de identificação e classificação de entidades mencionadas (EM) no texto, uma tarefa relevante para várias áreas de processamento de linguagem natural (PLN) como a resposta automática a perguntas, a extracção de informação e a tradução automática, entre outras.

No HAREM procurámos desenvolver uma nova metodologia de avaliação em REM que abrangesse as especificidades da tarefa, tendo contemplado questões que não tinham sido ainda abordadas com profundidade suficiente pelos anteriores eventos de avaliação internacionais. Exemplos são a vagueza das EM, a caracterização morfológica das mesmas, e a avaliação de expressões apenas parcialmente identificadas.

Os participantes no HAREM tiveram um papel activo no desenvolvimento da metodologia e na anotação das colecções de texto segundo a metodologia aprovada. Adicionalmente, a organização do HAREM desenvolveu uma plataforma de avaliação de sistemas de REM, para aferir o desempenho dos sistemas participantes, que se encontra publicamente disponível em <http://www.linguateca.pt/HAREM/>, o sítio do HAREM.

O HAREM culminou na organização de dois eventos de avaliação, em Fevereiro de 2005 e em Abril de 2006, e contou com 10 sistemas participantes oriundos de 6 países, que deram origem a um total de 38 saídas que foram avaliadas. Os resultados foram publicados, e os relatórios detalhados do desempenho de cada sistema foram entregues aos respectivos participantes.

A 15 de Julho de 2006, a Linguateca organizou o Encontro do HAREM na Universidade do Porto, logo a seguir à Primeira Escola de Verão da Linguateca, que reuniu os participantes, organizadores e outros interessados no HAREM [179-190]. No encontro, os participantes apresentaram os seus sistemas, a organização apresentou detalhadamente o seu trabalho, e todos debateram várias questões sobre o futuro do HAREM, numa sessão final, que demonstrou bem o interesse da comunidade em prosseguir com mais eventos de avaliação em REM. Está em curso a organização de um livro (electrónico) sobre o HAREM [278] englobando as comunicações nesse encontro [238-239, 241, 245, 248, 250, 254, 260, 262-264, 266-267, 273, 279, 280, 282, 286-287, 290].

2. Recursos

Outra das actividades em que pensamos que o papel da Linguateca tem sido determinante é na construção de recursos públicos para o processamento computacional da língua portuguesa. Damos em seguida uma descrição dos mais importantes. De salientar que a ordem por que os recursos são apresentados neste relatório é arbitrária.

2.1 Corpógrafo

Luís Sarmento

O Corpógrafo, <http://www.linguateca.pt/Corpografo/>, actualmente na versão 3 [156], é um projecto que vem sendo desenvolvido desde 2003 (embora inicialmente denominado “Gestor de Corpora” [4]) seguindo uma estratégia centrada no utilizador, que nos permitiu desenvolver um conjunto de ferramentas para construção e análise de corpora especializados e para extracção de terminologia, que satisfazem as necessidades correntes dos terminólogos e tradutores, assim como do ensino a nível universitário destes. O Corpógrafo foi desenhado de forma a fornecer a partir de um único ponto na web todo este conjunto de funcionalidades, concebidas para utilizadores que não possuam quaisquer conhecimentos de programação. Cada utilizador possui uma área própria num servidor do Corpógrafo instalado no pólo do Porto da Linguateca, na qual pode realizar todo o trabalho sobre os seus corpora privados.

Inicialmente, o Corpógrafo teve como principal objectivo permitir aos utilizadores a construção dos seus próprios corpora de domínio específico para estudos de terminologia, seguindo o conceito “do-it-yourself corpora” proposto por Belinda Maia [298], mas a sua evolução levou-o a tornar-se numa plataforma de pesquisa sobre corpora capaz de auxiliar a construção de recursos lexicais, tais como terminologias ou ontologias, a partir de textos em português (e outras línguas). Note-se que o Corpógrafo oferece uma alternativa ao trabalho com corpora públicos de grandes dimensões que não permitem normalmente a investigação sobre domínios altamente especializados.

O fluxo de trabalho do Corpógrafo permite ao utilizador realizar, nas suas áreas pessoais, aquisição e pré-processamento de texto; combinação de vários textos; criação e gestão de bases de dados terminológicas; extracção de terminologia nos seguintes idiomas: português, ingles, espanhol, francês e italiano; extracção de definições; e exportação de resultados, em formato XML e MySQL.

O Corpógrafo já foi experimentado por cerca de mil utilizadores diferentes, na sua maioria alunos de instituições portuguesas e brasileiras cursando linguística, terminologia e/ou tradução. Além disso, é rotineiramente utilizado nas Universidades do Porto, de Barcelona, de São Paulo, de Aveiro, do Minho, na PUC-RS e na Universidade Federal do Rio Grande do Sul. O Corpógrafo, além da extensa documentação

produzida no âmbito da Linguateca [3-4, 38, 41, 49, 59, 66, 68-69, 90-91, 115, 120, 156-158, 227], foi apresentado em sessões práticas em Braga (Junho de 2003), Barcelona (Janeiro de 2004), Lisboa (Maio de 2004), São Paulo (Agosto de 2004), Lisboa (Setembro de 2004), Leeds (Março de 2005), Montréal (Abril de 2005), Tampere (Abril de 2005), Birmingham (Julho de 2005), Braga (Julho de 2005), Santiago de Compostela (Setembro de 2005), Braga (Março de 2006), Tampere (Abril de 2006), Rio Grande do Sul (Maio de 2006), Rio de Janeiro (Maio de 2006), Génova (Maio de 2006), Tarragona (Junho de 2006), Lisboa (Junho de 2006), Barcelona (Junho de 2006), Istambul (Julho de 2006) e Porto (Julho de 2006) e já foram defendidas sete teses de mestrado na Universidade do Porto fazendo uso deste sistema (também aplicado em três teses de doutoramento e várias outras de mestrado em curso).

Finalmente, desde Outubro de 2005 o código fonte do Corpógrafo está disponível, após o sucesso da instalação de uma versão do mesmo na Universidade Pompeu Fabra de Barcelona, no qual é usado pelo Institut Universitari de Lingüística Aplicada (IULA), com mais de uma centena de utilizadores registados.

2.2 COMPARA

Ana Frankenberg-Garcia

O COMPARA, <http://www.linguateca.pt/COMPARA/>, é um corpus paralelo bi-direccional e extensível, que tem como base uma colecção de textos originais e traduções em português e inglês, postos em correspondência (alinhados) por frase, marcados com diversa informação relevante para o estudo de cada uma das línguas e da tradução e das diferenças ou semelhanças entre elas. A primeira versão foi tornada acessível ao público em Maio de 2000. No período a que se refere este relatório, foram adicionados 46 pares de textos ao corpus (que tem 75 no total) correspondendo a 17 novos autores e foram obtidas autorizações para 10 textos novos, resultando num acréscimo de 1,8 milhões de palavras.

A usabilidade do COMPARA também sofreu várias melhorias, a saber: foi adicionado um texto de ajuda (bilingue) para todos os pontos de preenchimento dos formulários de pesquisa; as páginas de conteúdo do corpus foram reformuladas com textos em ordem alfabética e mais dados quantitativos sobre o corpus; e foram implementadas as opções de adicionar automaticamente diacríticos e não fazer distinção entre letras maiúsculas/minúsculas nos formulários de pesquisa, além de fornecer a distribuição dos resultados por texto original e traduzido, variante do português e variante do inglês.

A marcação do corpus foi enriquecida e consolidada através de: uma revisão completa das etiquetas semânticas <foreign>, <named>, <title>, <voice> e <emph>; um melhor tratamento das aspas; da inserção da marca de parágrafo e de títulos de capítulos; e de um tratamento especial de separação de unidades de alinhamento a algumas obras que faziam uso de pontuação não padrão. Implementou-se também a geração automática das páginas de conteúdo do corpus.

Finalmente, foi incluída no corpus a anotação gramatical automática (apenas por enquanto no lado português) com a consequente possibilidade de fazer procuras com base em elementos gramaticais, além de permitir a distribuição por categoria gramatical e lema, e iniciou-se a revisão humana dessa mesma anotação gramatical, com a consequente documentação detalhada das opções tomadas e dos problemas em aberto [229].

Existem neste momento 49 publicações ou apresentações mencionando e utilizando o COMPARA listadas na página de publicações do mesmo. Apenas no âmbito da Linguateca, foram escritos 33 textos (28 produzidos no período a que se refere este relatório, [2, 27, 29, 30-31, 46-47, 51, 58, 65, 87, 89, 99-100, 109, 124-125, 140-141, 143, 153, 166, 191, 194, 220, 229, 276, 281]), e feitas apresentações públicas sobre o COMPARA em encontros e conferências em Braga (Junho de 2003), Lisboa (Maio de 2004), São Paulo (Agosto de 2004), Fortaleza (Setembro de 2004), Lisboa (Outubro de 2004), Aveiro (Outubro de 2004), Rio de Janeiro (Maio de 2006), Génova (Maio de 2006), Paris (Julho de 2006), Porto (Julho de 2006) e São Paulo (Agosto de 2006).

2.3 O sistema de resposta automática a perguntas Esfinge

Luís Costa

O pólo de Oslo iniciou o desenvolvimento do sistema de resposta automática a perguntas (RAP) Esfinge [114, 139] em finais de 2003. A ideia inicial foi a implementação das técnicas descritas por Eric Brill [293] ao português, investindo na redundância de informação existente na Rede e poupando o uso de recursos linguísticos de construção morosa e cara. O facto de a Rede se encontrar em constante crescimento e o português ser uma das línguas mais utilizadas foi outra motivação para este projecto.

O desenvolvimento do Esfinge apostou também desde o início no uso de recursos já existentes para o português, permitindo assim uma avaliação indirecta dos mesmos e a sua melhoria. Assim, na sua primeira versão o sistema usava o analisador morfológico jspell [301], ferramentas de atomização e fraseamento desenvolvidas no âmbito da Linguateca (módulo Perl Língua::PT::PLNbase disponibilizado em <http://www.cpan.org>) e o corpus CETEMPúblico [299]. Desde essa altura existe uma interface na Rede para uma versão simplificada do sistema, <http://www.linguateca.pt/Esfinge/> [196].

Em 2005 o Esfinge passou a utilizar também o reconhecedor de entidades mencionadas SIEMÊS e a aceitar perguntas em inglês (respondidas em português) [103], enquanto em 2006 foi testada a integração do BACO [151, 203].

O Esfinge participou desde o início na tarefa de RAP monolíngue português do CLEF [43, 96, 103, 196, 252] e, desde 2005, também na bilingue inglês-português..

Durante 2006, o esforço principal no desenvolvimento do Esfinge focou a modularização do sistema e a disponibilização pública do seu código fonte: <http://linguateca.di.uminho.pt/cgi-bin/viewcvcs.cgi/Esfinge/>. Foi também definido o Esfinge como serviço Web no âmbito do QoLA.

2.4 AC/DC

Embora o serviço mais antigo fornecido pela Linguateca, o AC/DC foi de todos o mais negligenciado durante o período transacto, com apenas três publicações [24, 26, 191]. De facto, apenas podemos mencionar a actualização pontual de alguns dos corpora já existentes e em alargamento constante (o CoNE e o ANCIB) e a criação de novos corpora acessíveis através do AC/DC, como o corpus ConDiv e o corpus Museu da Pessoa, ambos criados pelo pólo de Braga. Também se encontra em progresso a criação de novos corpora AC/DC dando acesso à colecção CHAVE e aos textos do projecto Vercial, ambos também processados pelo pólo de Braga.

Passaram a ser também disponibilizadas as listas de frequências de cada corpus, não só em termos de formas mas também por categoria gramatical.

2.5 WPT 03 e BACO

Nuno Cardoso

A WPT 03 é uma recolha da *web* portuguesa realizada pela equipa do tumba!, um motor de busca para a comunidade portuguesa, em actividade desde Novembro de 2002, desenvolvido pelo grupo XLDB da FCUL [300]. Um dos primeiros trabalhos do pólo da Linguateca no XLDB foi transformar esta recolha num recurso para a comunidade científica interessada no processamento computacional do português, que levou à sua disponibilização em Maio de 2004, <http://www.linguateca.pt/WPT03/>.

Esta recolha é composta por cerca de 3,5 milhões de documentos publicados na web portuguesa entre Março e Junho de 2003. A grande diversidade de autores, de estilos de escrita e de temas abordados nos seus documentos tornam a WPT 03 num recurso valioso para diversas áreas de investigação, nomeadamente a linguística, a computação e a sociologia. Além disso, a WPT 03 inclui um registo de pesquisas feito pelos utilizadores do tumba! sobre um índice com os textos da WPT 03 durante um período de 6 meses [249].

O material textual contido na WPT 03 levou ao desenvolvimento, por Luís Sarmento, de uma base de dados de co-ocorrências de grandes dimensões, denominada BACO [79, 155, 222]. O BACO é uma base de dados relacional que contém informação das co-ocorrências das palavras contidas nos cerca de 1,5 milhões de documentos resultantes do processo de remoção de documentos duplicados do WPT03, e actualmente serve de suporte para o desenvolvimento de diversos trabalhos (por enquanto apenas dentro da Linguateca).

2.6 Floresta Sintá(c)tica

Embora não tenha havido grande investimento neste recurso (por falta de recursos humanos e financeiros), houve um claro aumento da sua projecção a nível internacional: a Floresta, <http://www.linguateca.pt/Floresta/>, foi usada numa avaliação conjunta internacional de aprendizagem automática (ConLL-X) como recurso de avaliação de marcação sintáctica dependencial, http://nextens.uvt.nl/~conll/free_data.html; na Universidade de Austin no Texas foi desenvolvido um analisador sintáctico do português baseado na Floresta [302]; e esta foi, além disso, discutida como exemplo por investigadores alemães no LREC 2006 [294].

Do ponto de vista das funcionalidades, foram criadas pelo pólo de Braga versões em vários outros formatos [64], tais como o Tiger-XML, o Penn Treebank, o SQL, o SimTreeML e Perl, assim como foram disponibilizadas em 2006 versões dos Bosques e das Florestas Virgens num novo formato dependencial desenvolvido por Eckhard Bick, contendo muito mais informação.

Vários artigos de divulgação foram produzidos [8, 11, 56-57, 60, 64, 246, 277], tendo-se esboçado em Julho de 2006 uma proposta de colaboração internacional para revitalizar a Floresta e incluir mais informação, que esperamos que venha a ter efeito já no próximo ano. Algumas experiências encontram-se já em progresso, nomeadamente a marcação semântica, por Eckhard Bick, e a inclusão da informação do HAREM, por Rui Vilela.

Foi também desenvolvido e tornado público um validador dos Bosques pelo pólo de Braga e a documentação [33, 228] extensivamente melhorada.

2.7 Geo-NET-PT

Marcirio Chaves

A Geo-Net-PT01 é a primeira ontologia com dados geográficos administrativos de Portugal publicamente disponível na web (<http://xldb.fc.ul.pt/geonetpt/>). Essa ontologia foi desenvolvida no Pólo XLDB da Linguateca. A Geo-Net-PT01 contém 418.065 dados geográficos administrativos de Portugal, com informação sobre os distritos, concelhos e ruas, entre outros dados. Além da informação geográfica, inclui o âmbito geográfico atribuído a 686 sítios da Internet. Essa ontologia é um recurso para aplicações informáticas que necessitem de informação geográfica de Portugal.

A Geo-Net-PT01 está a ser usada dentro do projecto GREASE por um sistema de reconhecimento automático de locais bem como para suportar consultas em um motor de busca geográfico (Geo-Tumba) desenvolvido no mesmo projecto. Essa ontologia também foi usada para suportar a participação do grupo XLDB em avaliações conjuntas, a saber: GeoCLEF 2005, GeoCLEF 2006, HAREM e Mini-HAREM.

Atualmente, a Geo-Net-PT01 está a ser complementada e expandida com informação de locais obtidas a partir de texto. Maiores detalhes sobre a Geo-Net-PT01 estão disponíveis em [92, 112, 142, 224].

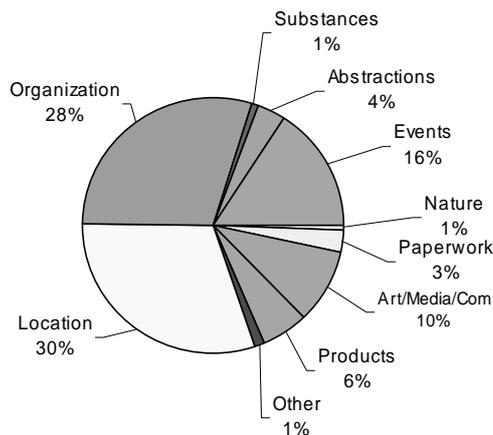
2.8 REPENTINO - REpositório para reconhecimento de ENTidades com NOME

Luís Sarmento

O REPENTINO [77, 146] é um repositório público (<http://www.linguateca.pt/repentino/>) que contém exemplos de entidades com nome, ou seja, de entidades concretas ou abstractas que possuem um nome próprio. O REPENTINO é um almanaque (em inglês, "gazetteer") que serve essencialmente para sistemas de reconhecimento automático de entidades mencionadas para a língua portuguesa.

Os cerca de 450.000 exemplos armazenados no REPENTINO encontram-se divididos por várias categorias conceptuais, cada uma das quais contendo diversas sub-categorias, numa estrutura em árvore, garantindo assim uma razoável organização destes exemplos. As categorias actuais do REPENTINO são: Abstracções, Arte / Média / Comunicação, Natureza, Eventos, Papeladas, Locais, Organizações, Produtos, Seres e Substâncias.

Cerca de 65% das instâncias armazenadas pertencem à classe dos Seres. Das restantes, a distribuição é apresentada no seguinte gráfico (de [146]):



Além de ser possível ir buscar todo o repositório em formato XML a partir do sítio web, encontra-se também disponível um módulo Perl para utilização directa e autónoma dos dados do REPENTINO, que é também consultável e extensível através da interface na web.

2.9 SIEMÊS – Sistema de Identificação de Entidades Mencionadas com Estratégia Siamesa

Luís Sarmento

O SIEMÊS [145, 181, 187, 282] é um sistema que identifica e classifica referências a entidades (pessoas, locais, organizações, expressões numéricas, etc.) e a respectiva forma de menção em texto em português. Neste momento o SIEMÊS usa uma hierarquia de classificação que possui 10 categorias de topo e mais de 100 subcategorias. Estas categorias incluem todas as que estão incluídas no almanaque REPENTINO e outras relacionadas com entidades numéricas (tais como Tempo e Valores).

O princípio de funcionamento do SIEMÊS assenta na convicção de que se for possível, numa primeira fase, através da informação existente em almanaques, gerar um conjunto de hipóteses de classificação para um determinado candidato, torna-se possível numa segunda fase desambiguar semanticamente a classe e a forma de menção do referido candidato, usando regras de análise do contexto relativamente simples. Para a análise do contexto, o SIEMÊS recorre a vários bancos de regras especializados, que estão constantemente a ser ampliados e melhorados e que possuem actualmente mais de 50 regras para entidades alfanuméricas e mais de 100 regras para entidades numéricas.

O SIEMÊS ainda se encontra em desenvolvimento no âmbito da tese de doutoramento de Luís Sarmiento; contudo, já é passível de ser utilizado internamente pela Linguateca como um módulo Perl e foi empregue pelo Esfinge, pelo RAPOSA [200, 284], e nos primeiros estudos da geograficidade de Web portuguesa [142, 193], estando presentemente a ser usado para anotar o BACO.

2.10 NATools

Alberto Simões

O NATools é um pacote de ferramentas que surgiu na sequência do extractor de dicionários probabilísticos de tradução, denominado Twente-Aligner [296], que tem vindo a incorporar um conjunto de ferramentas úteis para o processamento de corpora paralelos [5, 20, 23, 62-63, 84, 136, 159, 207, 220, 289]. Neste sentido, actualmente inclui: 1) um extractor escalável de dicionários probabilísticos de tradução a partir de corpora alinhados à frase; 2) um servidor de corpora (concordâncias, bigramas, trigramas e tetragramas) e de dicionários probabilísticos de tradução, multi-corpora e multilingue; 3) ferramentas para a extracção de terminologia e de exemplos de tradução úteis quer na tradução assistida por computador quer em tradução automática baseada em exemplos; 4) uma biblioteca C, e um conjunto de módulos Perl para facilitar a programação sobre os recursos criados (e um conjunto de programas que os usam para a realização de várias tarefas ligadas a processamento de corpora paralelos); 5) Um alinhador de corpora paralelos ao nível da palavra com base em dicionários probabilísticos de tradução, e 6) um alinhador de corpora paralelos ao nível da frase, baseado na implementação Vanilla Aligner [295].

Foram além disso criados e disponibilizados vários corpora paralelos utilizando o pacote NATools, alinhados ao nível da frase, e com dicionários probabilísticos de tradução extraídos e disponibilizados. Estão disponíveis os EuroParl [297] nos seus pares de línguas português-ínglês, português-francês e português-espanhol, as encíclicas papais nos pares de língua português-francês e português-ínglês, e um corpus francês-português resultante das respectivas edições do *Le Monde Diplomatique*. Para melhorar a qualidade dos corpora foram também criadas ferramentas de pré-processamento e validação de alinhamento capazes de rejeitar unidades de tradução suspeitas de trazer ruído.

Mais informação sobre o NATools, como por exemplo o procedimento de instalação e documentação de uso, pode ser encontrada em <http://linguateca.di.uminho.pt/natools/>.

2.11 Ferramentas de avaliação de tradução (automática)

Belinda Maia

No âmbito da formação de tradutores técnicos e científicos e também aproveitando o entusiasmo à volta da noção de avaliação conjunta provocado pelo EPAV em 2002 e pelo Encontro Avalon em 2003 criaram-se várias ferramentas de avaliação de tradução automática no pólo do Porto [13, 19, 35, 59, 258, 283, 285], denominadas TrAva e CorTA, usadas no âmbito de um seminário de mestrado de sintaxe e semântica por mais de 40 alunos, e resultaram em alguns artigos em conferências por alunos da Universidade do Porto e a uma maior consciencialização dos problemas da avaliação de tradução da Linguateca como um todo.

O sistema METRA, que distribui um pedido de tradução a vários sistemas de tradução automática na Rede e apresenta os resultados em conjunto ao utilizador, é um dos serviços ao público com mais uso na Linguateca (700 pedidos por dia), o que levou a tornar os próprios diários de consulta acessíveis para investigação.

2.12 WebJspell

Rui Vilela

O WebJspell, <http://www.linguateca.pt/WebJspell/>, é um serviço desenvolvido pelo pólo de Braga da Linguateca com o objectivo de tornar o analisador morfológico e corrector ortográfico Jspell [301] (desenvolvido pelo projecto Natura) acessível a um maior número de utilizadores.

Este serviço oferece três funcionalidades distintas para o português e para o inglês (e para latim, parcialmente): acesso ao Jspell como analisador morfológico, como corrector ortográfico e como verificador de páginas Web. Permite além disso recolher sugestões de palavras para adicionar ou remover do dicionários, e adicionar informação morfológica ao Jspell.

O WebJspell facilitou o elo de comunicação com os utilizadores do Jspell, permitindo que estes contribuíssem de uma forma mais activa para o desenvolvimento do(s) seu(s) dicionário(s). Além disso, os logs também são analisados de forma a conhecer o padrão dos erros dos utilizadores, e a descobrir palavras suspeitas de não serem correctamente analisadas pelo Jspell, ou mesmo palavras cuja frequência justificaria a sua entrada no dicionário.

O WebJspell foi tornado público pela primeira vez em Fevereiro de 2006 e conta de momento com cerca de 300 perguntas mensais. Quase diariamente também, são enviadas por voluntários sugestões anónimas ou identificadas de palavras ou listas de palavras para apreciação.

2.13 O catálogo de publicações da Linguateca e o SUPERB

Para melhorar a interacção da comunidade com o catálogo de publicações sobre o processamento computacional da língua portuguesa, foi desenvolvido pelo pólo de Oslo (Paulo Rocha) um serviço de adição de publicações e outro de procura, além de se ter reformulado e aumentado a informação associada a cada publicação, nomeadamente através da atribuição de marcações internas (tags) para objectivos distintos.

Mais tarde foi efectuada uma melhoria significativa do catálogo de publicações no âmbito do trabalho de mestrado do Luís Miguel Cabral [86, 210, 221, 247]. O objectivo principal era melhorar o processo de validação e a descoberta de novos candidatos a entradas no catálogo através de métodos automáticos. Ao mesmo tempo, contudo, foram desenvolvidos vários programas para converter entre formatos bibliográficos, para corrigir erros de digitação em referências, e para verificar a consistência do catálogo actual. Foram também produzidos alguns programas que facilitam a criação de listas especializadas, por exemplo empregues nas páginas de publicações dos vários projectos da Linguateca, assim como foram incluídas várias novas opções de visualização do material, que serão empregues na secção 4 abaixo.

2.14 T2O

Alberto Simões

Thesaurus-to-Ontologies [150] é uma plataforma para a criação de uma mega-ontologia, multi-lingue e multi-domínio com base na exploração de recursos estruturados existentes livremente na Internet, como thesaurus, ontologias, taxonomias e simples tabelas relacionais. Foram construídas ferramentas para a manipulação destes vários recursos e a sua conciliação numa única ontologia. Foram também desenvolvidas ferramentas de disponibilização destes recursos [160].

2.15 Outros

Como actividades dignas de menção, embora não suficientemente abrangentes para merecerem secção separada, indicamos:

- Disponibilização de ferramentas computacionais como módulos Perl;
- Tentativa de melhoria do sistema Busca usando tecnologias da área da terminologia [91];
- Estudo preliminar dos logs do tumba incluídos no WPT 03 [128, 233];
- Interface na Rede para os recursos do LABEL [7, 54, 106, 108];
- Participação no HAREM dos sistemas ELLE [107] e SIEMÊS;
- Primeiros estudos na construção de ontologias lexicais para o português [81, 111, 123, 192, 234-235];
- Primeiros estudos no dimensionamento da informação geográfica na Web [142, 193];
- Continuação da manutenção do fórum e dos catálogos de recursos e de actores.

3. Instrução e disseminação

Com um conjunto de actividades cuja principal razão de ser é o serviço à comunidade científica da área do processamento computacional da língua portuguesa e todas as que lhe estão ou podem estar de certa forma associadas, é natural que parte do investimento pelo menos dos séniores associados à Linguateca passe por actividades de formação e disseminação. Produzimos, assim, alguns trabalhos com o fito de apresentar genericamente a Linguateca, tais como [55, 116, 138, 225].

3.1 Escola de Verão

A Primeira Escola de Verão da Linguateca teve lugar na Universidade do Porto de 10 a 14 de Julho de 2006 [161-178] e ultrapassou em muito as expectativas, tanto pelo número elevado de inscrições e pela assiduidade dos alunos, como pela apreciação global extremamente positiva que estes fizeram do curso, por escrito e oralmente durante a sessão de encerramento. Igualmente de louvar foi a participação de nove alunos provenientes de empresas, assim como de três alunas brasileiras. Todos os materiais de ensino, juntamente com extensas listas de referências, estão acessíveis do sítio da Linguateca.

3.2 Simpósios doutorais

Devido à actividade científica e de investigação da Linguateca ter crescido significativamente nestes anos, pareceu útil juntar regularmente todos os membros que tinham em mãos um trabalho académico (mestrado ou doutoramento) e fazer uma reunião informal de orientação e troca de ideias, fomentando a colaboração. O primeiro simpósio teve lugar em Maio de 2005, <http://www.linguateca.pt/documentos/SimpósioDoutoral2005.html> o segundo em Abril de 2006, <http://www.linguateca.pt/documentos/SimpósioDoutoral2006.html> e o terceiro em Outubro do mesmo ano, <http://www.linguateca.pt/documentos/SimpósioDoutoralOut2006.html>.

3.3 Livro sobre avaliação conjunta

Pensando na documentação das actividades da Linguateca a mais longo prazo, e considerando que a avaliação conjunta vinha marcar um novo paradigma na nossa área, impunha-se um livro que apresentasse as ideias, documentasse a primeira (as Morfolimpíadas) e reunisse para o futuro todos os autores, brasileiros e portugueses, que participaram neste marco histórico. Este livro, denominado Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*, IST Press (ISBN: 972-8469-60-8) encontra-se no prelo [275], prevendo-se a sua publicação para Março de 2007.

3.4 Formação de recursos humanos (bolsas e teses)

A Linguateca teve, no período a que se refere o presente relatório, cinco bolsheiros de investigação (por ordem alfabética: Ana Sofia Pinto, Débora Oliveira, Isabel Marcelino, Luís Miguel Cabral e Susana Inácio), e quatro de doutoramento (por ordem alfabética: Alberto Simões, Marcirio Chaves, Nuno Seco e Rachel Aires), tendo sido finalizadas as seguintes três teses no âmbito da Linguateca no período a que se refere o relatório, estando a quarta em fase de conclusão:

Rachel Virgínia Xavier Aires. Uso de marcadores estilísticos para a busca na Web em português. Tese de Doutoramento. ICMC - USP - São Carlos. Agosto de 2005 [94]

Alberto Manuel Simões. Parallel Corpora word alignment and applications. Tese de Mestrado. Escola de Engenharia - Universidade do Minho. Setembro de 2004. [62]

Nuno Cardoso. Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas. Tese de Mestrado. Faculdade de Engenharia da Universidade do Porto. Outubro de 2006. [208]

Luís M. Cabral. SUPERB - Sistema Uniformizado de **P**esquisa de **R**eferências **B**ibliográficas. Tese de Mestrado. Faculdade de Engenharia da Universidade do Porto. 2007. [247]

3.5 Preparação de novos projectos

Fez também parte da actividade da Linguateca a procura de fontes de financiamento e projectos que pudessem complementar e enriquecer o trabalho feito. Faz portanto sentido que a participação em projectos associados, ou a preparação de propostas para a criação dos mesmos seja mencionada no presente relatório.

3.5.1 GREASE e GREASE II

Mário J. Silva

Alguns do trabalho executado no âmbito do pólo XLDB da Linguateca teve ampla relação com o projecto GREASE (POSI/SRI/47071/2002), como se vê pela extensa documentação conjunta [39, 42, 67, 92, 95, 101-102, 110, 112, 199, 201-202, 215, 224, 236, 251, 259, 262-263].

Como continuação do GREASE, que deu contributos importantes na nova área de recolha de informação geográfica (GIR), em particular na atribuição automática de âmbitos geográficos a páginas Web, propusemos um novo projecto GREASE II para a investigação de métodos de extracção de informação de grandes quantidades de texto e metadados, através da construção de um resumo constituído por um conjunto de etiquetas geográficas, correspondendo quer a conceitos geo-referenciados quer a etiquetas não geo-referenciáveis (i.e., sem uma definição geográfica precisa).

3.5.2 POP

Paulo Gomes

Existe neste momento um défice para o português por não existir, publicamente acessível, nenhuma ontologia de grande porte que modele e explicita as relações entre as palavras e os conceitos da nossa língua, ao contrário do que se passa para o inglês e outras línguas. Um dos objectivos do presente projecto, proposto à FCT com outros parceiros, é pois colmatar esta falha, quer partindo de um grande dicionário da língua geral para construir uma tal ontologia, como usando a grande quantidade de recursos textuais já publicamente acessíveis para extrair ontologias a partir de texto.

3.5.3 CLARA, PONTE e iBOOKAS

Também no âmbito do SINTEF foram desenvolvidas várias propostas de projectos com o objectivo de aumentar o conjunto de actividades associadas ao PLN no pólo de Oslo. Em particular foram enviadas as propostas PoNTE ao Instituto Camões (através da embaixada portuguesa em Oslo), a proposta CLARA à chamada de projectos europeus FP6 FET Open (como principal parceiro), e a proposta iBookAS (como parceiro). Embora tenham sido recusadas, produziram ideias e parcerias interessantes para o futuro, e ajudaram à disseminação internacional do trabalho da Linguateca.

3.5.4 Queryonomy

Encontra-se neste momento em desenvolvimento uma nova proposta de projecto europeu, desta feita na área de resposta automática a perguntas, a ser enviada à chamada FP7.

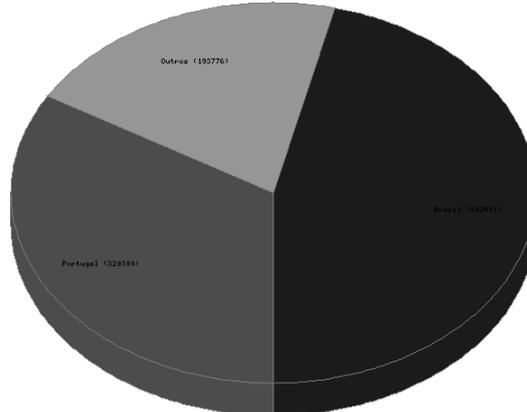
4. Medidas de impacto (catálogo e estatísticas)

Luís Costa e Luís Miguel Cabral

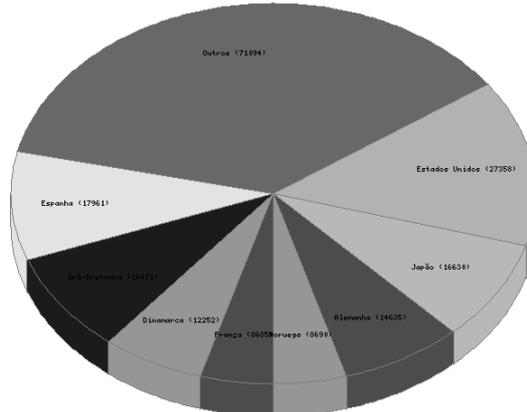
Os valores aqui apresentados referem-se, se mais nada for acrescentado, à data de 9 de Novembro de 2006. O número de visitas às páginas da Linguateca, de 1 de Maio de 2003 a 1 de Dezembro de 2006 foi de 2.094.236 (no total, desde o início do projecto, atinge os 2.812.510).

Para dar uma ideia do impacto internacional do nosso portal, apresentamos nas figuras seguintes a distribuição do perfil geográfico dos visitantes das nossas páginas (os valores referem-se apenas aos acessos para os quais foi possível determinar a sua origem, 35% do total de acessos):

Repartição geográfica acumulada dos acessos até 1 de Novembro de 2006

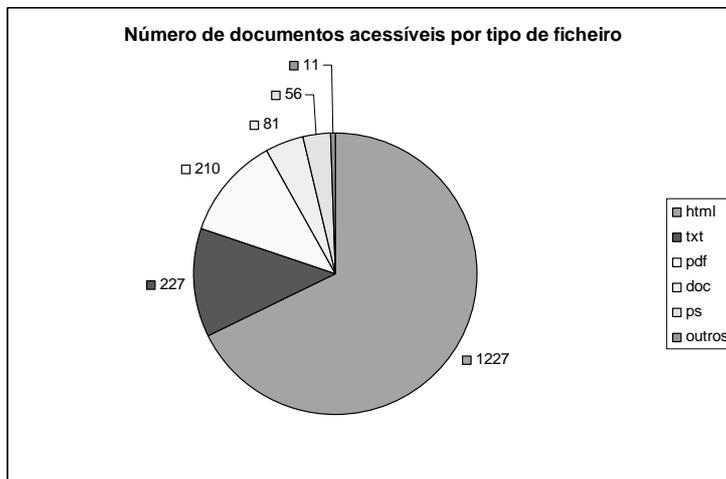


Repartição geográfica acumulada dos acessos excluindo Portugal e Brasil

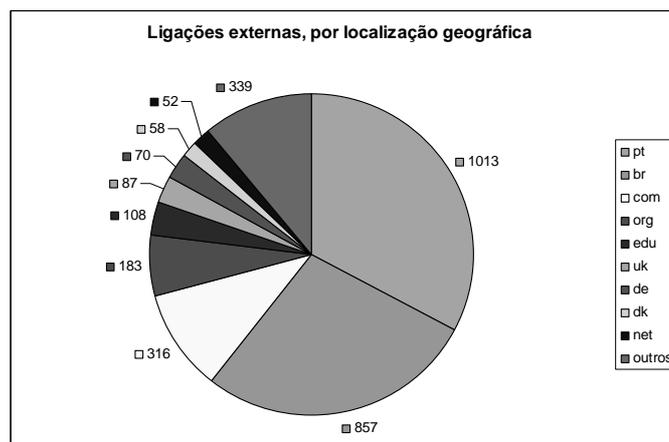


O número de documentos acessíveis do sítio da Linguateca é de 1812, constituindo 1375 documentação própria da Linguateca. A figura seguinte pormenoriza o tipo de documentos disponibilizados.

Número de documentos acessíveis por tipo de ficheiro



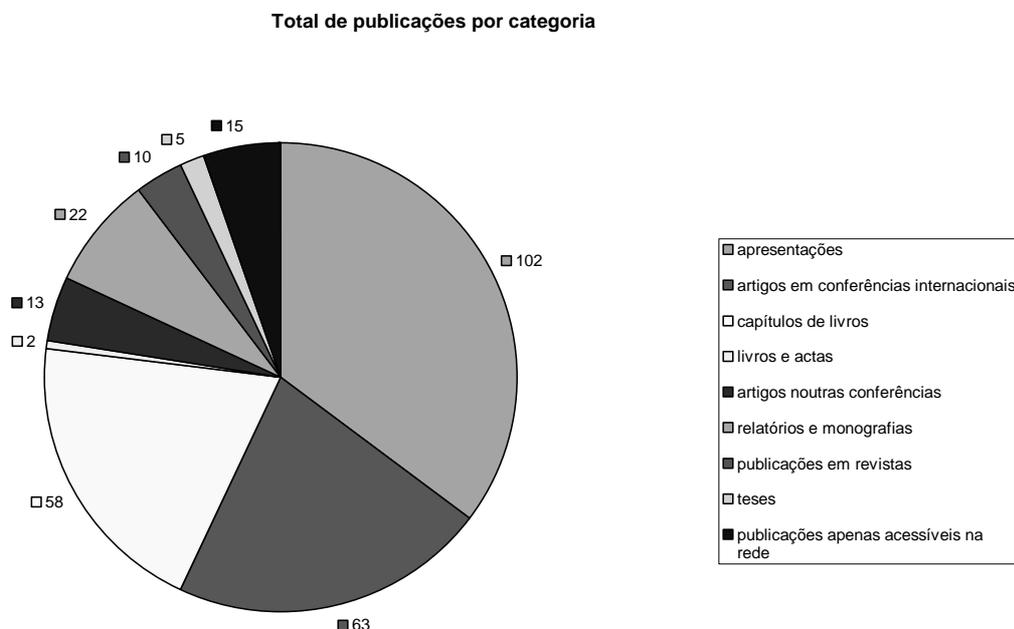
O número de ligações externas incluídas nos nossos catálogos é de 3083, e a distribuição geográfica das localizações para que apontam encontra-se na figura seguinte:



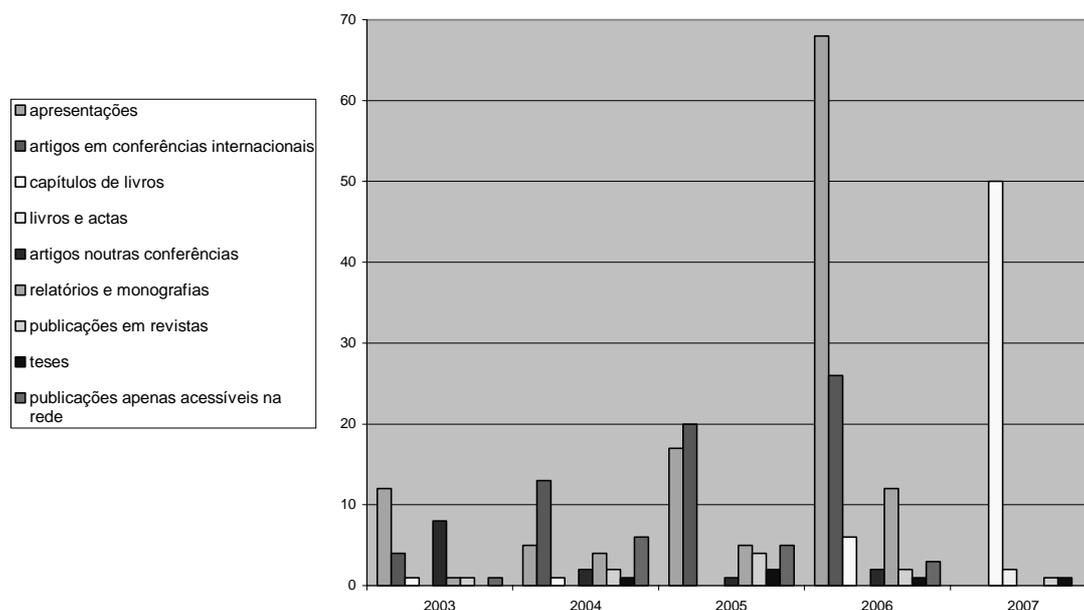
Apresentamos também o número de pedidos dos recursos que exigem registo, desde 15 de Maio de 2003 (entre parênteses, desde o início da sua disponibilização) até 1 de Dezembro de 2006:

Corpógrafo: 1036 (1036)
 CETENFolha: 196 (247)
 CETEMPúblico: 195 (418)
 CHAVE: 36 (36)
 WBR 99: 24 (24)
 WPT 03: 11 (11)
 GeoNet-PT-01: 15 (15)

Fianlmente, a distribuição das publicações (listadas na secção seguinte) por categoria, seguindo as normas do catálogo de publicações da Linguateca, e por categoria e por ano, encontra-se nas duas figuras seguintes:



Publicações por categoria por ano



5. Lista de publicações actualizada

Luís Miguel Cabral

As publicações da Linguateca no período a que se refere o presente relatório foram listadas com o auxílio do SUPeRB, e são apresentadas por ordem cronológica, e por ordem alfabética em cada ano, a partir de 15 de Maio de 2003.

Listamos também as publicações criadas no âmbito da Linguateca (por exemplo incluídas em livros ou material de ensino editado ou criado pela Linguateca) mas não por pessoas com vínculo formal à Linguateca, com o marcador **ÂMBITO**, para as distinguir claramente.

As publicações que, por lapso, ou por terem sido anteriores à informação recebida não mencionam correctamente o número do projecto POSI estão marcadas com o marcador **SemRec**.

2003 (28)

[1] J.J. Almeida, Alberto Manuel Simões, José Castro, Bruno Martins & Paulo Silva. "Projecto TerminUM". In José João Almeida (ed.), *Corpora Paralelos, Aplicações e Algoritmos Associados (CP3A)* (Braga, 3 de Junho), Braga: Universidade do Minho, pp. 7-14. **SemRec**

[2] Ana Frankenberg-Garcia. "A Construção (e alguns usos) do corpus Compara". In José João Almeida (ed.), *Corpora Paralelos, Aplicações e Algoritmos Associados (CP3A)* (Braga, 3 de Junho), Braga: Universidade do Minho, pp. 15-23. **SemRec**

[3] Belinda Maia. "The pedagogical and linguistic research implications of the GC to on-line parallel and comparable corpora". In José João Almeida (ed.), *Corpora Paralelos, Aplicações e Algoritmos Associados (CP3A)* (Braga, 3 de Junho), Braga: Universidade do Minho, pp. 31-32. **SemRec**

[4] Luís Sarmiento & Belinda Maia. "Gestor de corpora - Um ambiente Web integrado para Linguística baseada em Corpora". In José João Almeida (ed.), *Corpora Paralelos, Aplicações e Algoritmos Associados (CP3A)* (Braga, 3 de Junho), Braga: Universidade do Minho, pp. 25-30. **SemRec**

[5] Alberto Manuel Simões. "Alinhamento de corpora paralelos". In José João Almeida (ed.), *Corpora Paralelos, Aplicações e Algoritmos Associados (CP3A)* (Braga, 3 de Junho), Braga: Universidade do Minho, pp. 71-77. **SemRec**

[6] Rachel Aires, Sandra Aluísio, Paulo Quaresma, Diana Santos & Mário J. Silva. "An initial proposal for cooperative evaluation on information retrieval in Portuguese". In Nuno J. Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *Computational Processing of the Portuguese Language, 6th International Workshop (PROPOR 2003)* (Faro, 26-27 June 2003), Springer Verlag, pp. 227-234. **SemRec**

[7] Cristina Mota & Pedro Moura. "ANELL: A Web System for Portuguese Corpora Annotation". In Nuno J. Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *Computational Processing of the Portuguese Language, 6th International Workshop (PROPOR 2003)* (Faro, 26-27 June 2003), Springer Verlag, pp. 184-188. **SemRec**

- [8] Diana Santos. "Timber! Issues in treebank building and use". In Nuno J. Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *Computational Processing of the Portuguese Language, 6th International Workshop (PROPOR 2003)* (Faro, 26-27 June 2003), Springer Verlag, pp. 151-158. **SemRec**
- [9] Diana Santos, Luís Costa & Paulo Rocha. "Cooperatively evaluating Portuguese morphology". In Nuno J. Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *Computational Processing of the Portuguese Language, 6th International Workshop (PROPOR 2003)* (Faro, 26-27 June 2003), Springer Verlag, pp. 259-266. **SemRec**
- [10] Rachel Aires. "Avaliação de RI". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [11] Eckhard Bick. "Treebank Trouble". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [12] Eckhard Bick. "PALMORF's morpholimpical headaches". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003). **ÂMBITO**
- [13] Belinda Maia, Anabela Barreiro & Luís Sarmento. "EVAL - Evaluation of Machine Translation at FLUP". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [14] Ronaldo Martins. "ReGra's Lexical Database". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003). **ÂMBITO**
- [15] Cristina Mota. "Avaliação Conjunta de Sistemas de Reconhecimento de Entidades Mencionadas". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [16] Elisabete Ranchhod. "Vantagens da integração de dicionários e palavras compostas em sistemas PLN". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003). **ÂMBITO**
- [17] Diana Santos. "AvalON'2003: Abertura". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [18] Diana Santos & Luís Costa. "Morfolimpíadas - Apresentação detalhada da metodologia e dos problemas identificados". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [19] Luís Sarmento & Belinda Maia. "Avaliação de Tradução Automática - Alguns Esforços de Abordagem". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [20] Alberto Simões & José João Almeida. "Avaliação de Alinhadores à Frase". Apresentação no *AvalON'2003* (Universidade do Algarve, Faro, 28 de Junho de 2003).
- [21] Belinda Maia & Luís Sarmento. "Linguatca@Porto". Apresentação no *SINTEF* (SINTEF, 10 September 2003).
- [22] Diana Santos. "Relatório Linguatca 2000-2003". Setembro de 2003.
- [23] Alberto Manuel Simões & J. João Almeida. "NATools — A Statistical Word Aligner Workbench". *Revista da SEPLN - Sociedade Española para el Procesamiento del Lenguaje Natural* **31** (2003), pp. 217-226. **SemRec**
- [24] Susana Afonso. "Clara e sucintamente: um estudo em corpus sobre a coordenação de advérbios em -mente". In Amália Mendes & Tiago Freitas (orgs.), *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)* (Porto, 2-4 de Outubro de 2002), Lisboa: APL, pp. 27-36. **SemRec**
- [25] Diana Santos & Paulo Rocha. "AvalON: uma iniciativa de avaliação conjunta para o português". In Amália Mendes & Tiago Freitas (orgs.), *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)* (Porto, 2-4 de Outubro de 2002), Lisboa: APL, pp. 693-704. **SemRec**
- [26] Diana Santos & Luís Sarmento. "O projecto AC/DC: acesso a corpora/disponibilização de corpora". In Amália Mendes & Tiago Freitas (orgs.), *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)* (Porto, 2-4 de Outubro de 2002), Lisboa: APL, pp. 705-717. **SemRec**
- [27] Ana Frankenberg-Garcia & Diana Santos. "Introducing COMPARA, the Portuguese-English parallel translation corpus". In Federico Zanettin, Silvia Bernardini & Dominic Stewart (eds.), *Corpora in Translation Education*. Manchester: St. Jerome Publishing, 2003, pp. 71-87. **SemRec**
- [28] Eugénio Oliveira, Maria Francisca Xavier, Jorge Baptista, Isabel Trancoso, Luis Oliveira, Nuno Mamede, Violeta Quental, Graça Nunes, Graciete Teixeira, Lucia Rino, Renata Vieira, Tony Berber Sardinha, Isabel Hub Faria & Osvaldo Oliveira Jr. "Pareceres sobre Linguatca: Relatório relativo ao período 2000-2003". 2003. **ÂMBITO**
- 2004 (34)**
- [29] Ana Frankenberg-Garcia. "Are translations longer than source texts? A corpus-based study of explicitation". In *Third International CULT (Corpus Use and Learning to Translate) Conference* (Barcelona, 22-24 January 2004). Reprinted in CIC-ISLA Working Papers. Lisboa: ISLA, 1-2004, pp. 2-9 **SemRec**
- [30] Ana Frankenberg-Garcia. "COMPARA - Checking automatic alignment". 18 February 2004.
- [31] Ana Frankenberg-Garcia. "COMPARA - Verificação da etiquetagem automática". 18 de Fevereiro de 2004.

- [32] Alberto Manuel Simões, José João Almeida & Xavier Gomez Guinovart. "Memórias de Tradução Distribuídas". In José Carlos Ramalho & Alberto Simões (eds.), *XML, Aplicações e Tecnologias Associadas (XATA 2004)* (Porto, Fevereiro de 2004), pp. 59—68. **SemRec**
- [33] Susana Afonso. "Avaliação do grau de concordância entre anotadores: análise e discussão dos resultados do processo de re-revisão". 25 de Março de 2004. Relatório Floresta Sintá(c)tica
- [34] Paulo Alexandre Rocha & Diana Santos. "Portuguese at CLEF". Apresentação no *SINTEF* (Oslo, 27 Abril de 2004).
- [35] Diana Santos, Belinda Maia & Luís Sarmento. "Gathering empirical data to evaluate MT from English to Portuguese". In Lambros Kranias, Nicoletta Calzolari, Gregor Thurmair, Yorick Wilks, Eduard Hovy, Guðrún Magnúsdóttir, Anna Samiotou & Khalid Choukri (eds.), *Proceedings of LREC 2004 Workshop on the Amazing Utility of Parallel and Comparable Corpora* (Lisboa, Portugal, 25 May 2004), pp. 14-17.
- [36] Rachel Aires, Aline Manfrin, Sandra Maria Aluísio & Diana Santos. "What Is My Style? Stylistic features in Portuguese web pages according to IR users' needs". In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of LREC 2004* (Lisboa, Portugal, 26-28 May 2004), pp. 1943-1946.
- [37] Diana Santos & Anabela Barreiro. "On the problems of creating a consensual golden standard of inflected forms in". In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of LREC 2004* (Lisboa, Portugal, 26-28 May 2004), pp. 483-486.
- [38] Luís Sarmento, Belinda Maia & Diana Santos. "The Corpógrafo - a Web-based environment for corpora research". In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of LREC 2004* (Lisboa, Portugal, 26-28 May 2004), pp. 449-452.
- [39] Mário J. Silva, Bruno Martins, Marcirio Silveira Chaves, Nuno Cardoso & Ana Paula Afonso. "Adding Geographic Scopes to Web Resources". In *ACM SIGIR Workshop on Geographic Information Retrieval* (Sheffield - UK, June 2004), s/pp. **SemRec**
- [40] Alberto Manuel Simões, Xavier Gomez Guinovart & José João Almeida. "Distributed Translation Memories implementation using WebServices". In *Sociedade Espanola para el Procesamiento del Lenguaje Natural (SEPLN)* 33 (Barcelona, Julho de 2004), pp. 89-94.
- [41] Belinda Maia. "Trabalhando com o CORPÓGRAFO: no ensino e na pesquisa". Apresentação no *IV Encontro de Corpora - Da compilação à anotação lingüística, padronização e análise de corpora* (Universidade de São Paulo, 23 de Agosto de 2004).
- [42] Nuno Cardoso, Mário J. Silva & Miguel Costa. "The XLDB Group at CLEF 2004". In Carol Peters & Francesca Borri (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004)* (Bath, UK, 15-17 September 2004), Pisa, Italy: IST-CNR, pp. 183-191. Versão revista republicada como Cardoso et al. (2005)
- [43] Luís Costa. "First evaluation of Esfinge - a question-answering system for Portuguese". In Carol Peters & Francesca Borri (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004)* (Bath, UK, 15-17 September 2004), Pisa, Italy: IST-CNR, pp. 393-402. Versão revista republicada como Costa (2005)
- [44] Diana Santos & Paulo Rocha. "CHAVE: topics and questions on the Portuguese participation in CLEF". In Carol Peters & Francesca Borri (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004)* (Bath, UK, 15-17 September 2004), Pisa, Italy: IST-CNR, pp. 639-648. Versão revista republicada como Santos & Rocha (2005)
- [45] Alessandro Vallin, Bernardo Magnini, Christelle Ayache, Gregor Erbach, Anselmo Peñas, Maarten de Rijke, Paulo Rocha, Kiril Simov & Richard Sutcliffe. "Overview of the CLEF 2004 Multilingual Question answering track". In Carol Peters & Francesca Borri (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004)* (Bath, UK, 15-17 September 2004), Pisa, Italy: IST-CNR, pp. 281-294. Versão revista republicada como Vallin et al. (2005)
- [46] Ana Frankenberg-Garcia. "COMPARA English Tutorial". 28 September 2004.
- [47] Ana Frankenberg-Garcia. "O corpus COMPARA". Apresentação no *IX Encontro Nacional / III Encontro Internacional de Tradutores* (Fortaleza (CE), Setembro 2004).
- [48] Rachel Aires, Aline Manfrin, Sandra Aluísio & Diana Santos. "Which classification algorithm works best with stylistic features of Portuguese in order to classify web texts according to users. needs?". Relatório Técnico n° 241 : ICMC/USP. Outubro de 2004.
- [49] Ana Sofia Pinto & Débora Oliveira. "Extração de definições no Corpógrafo". Outubro de 2004.
- [50] Alberto Manuel Simões, Tiago Bezerra & Pedro Henriques. "A importância das Ontologias num Museu Virtual". In *II Congresso Internacional de Investigação e Desenvolvimento Sócio-cultural* (Paredes de Coura, Outubro de 2004), s/pp. **SemRec**
- [51] Ana Frankenberg-Garcia. "Working with COMPARA, an online parallel corpus of English and Portuguese fiction". Apresentação no *ISLA* (Lisboa, 1 de Outubro de 2004).

- [52] Diana Santos. "Working with Portuguese corpora". Apresentação na *Universidade de Oslo* (Oslo, 22 Outubro de 2004).
- [53] Rachel Aires. "Uso de características lingüísticas para a apresentação dos resultados de busca na Web de acordo com a intenção da busca do usuário". In *IX Simpósio de teses e dissertações do ICMC-USP São Carlos* (São Carlos (SP), 19 e 20 de Novembro de 2004).
- [54] Isabel Marcelino. "Descrição do trabalho para o sistema AnELL". 26 de Novembro de 2004.
- [55] Diana Santos, Alberto Simões, Ana Frankenberg-Garcia, Ana Pinto, Anabela Barreiro, Belinda Maia, Cristina Mota, Débora Oliveira, Eckhard Bick, Elisabete Ranchhod, José João Dias de Almeida, Luís Cabral, Luís Costa, Luís Sarmento, Marcirio Chaves, Nuno Cardoso, Paulo Rocha, Rachel Aires, Rosário Silva, Rui Vilela & Susana Afonso. "Linguateca: um centro de recursos distribuído para o processamento computacional da língua portuguesa". In Guillermo De Ita Luna, Olac Fuentes Chávez & Mauricio Osorio Galindo (eds.), *Proceedings of the international workshop "Taller de Herramientas y Recursos Lingüísticos para el Español y el Portugués"*, IX Iberoamerican Conference on Artificial Intelligence (IBERAMIA 2004) (Puebla, México, Novembro de 2004), pp. 147-154. **SemRec**
- [56] Susana Afonso. "A Floresta Sintá(c)tica como recurso". 2004.
- [57] Eckhard Bick. "Looking at the Floresta Sintá(c)tica with a CorpusEye: A user-friendly cross-language search interface". 2004.
- [58] Ana Frankenberg-Garcia. "Lost in parallel concordances". In Guy Aston, Silvia Bernardini & Dominic Stewart (eds.), *Corpora and language learners*. Amsterdam/Philadelphia: John Benjamins Publishing, 2004, pp. 213-229. **SemRec**
- [59] Belinda Maia, Luís Sarmento, Stella E. O. Tagnin & Sandra Maria Aluísio. "Idéias que cruzam o Oceano". *CROP - Revista da Área de Língua e Literatura Inglesa e Norte-Americana* **10** (2004), pp. 43-64. Brasil: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. **SemRec**
- [60] Raquel Marchi. "Revisão humana da Floresta Sintá(c)tica: exemplos e método". 2004.
- [61] Diana Santos. "Aonde vamos em relação a aonde". *The ESpecialist* **25.1** (2004), pp. 85-103. São Paulo. **SemRec**
- [62] Alberto Manuel Simões. Parallel Corpora word alignment and applications. Tese de Mestrado. Escola de Engenharia - Universidade do Minho. 2004. **SemRec**

2005 (54)

- [63] Alberto Simões. "EBMT: Example Based Machine Translation". Apresentação no *Simpósio Doutoral do Departamento de Informática da Universidade do Minho* (Braga, 5-7 de Janeiro de 2005).
- [64] Rui Vilela, Alberto Manuel Simões, Eckhard Bick & José João Almeida. "Representação em XML da Floresta Sintáctica". In José Carlos Ramalho, Alberto Simões & João Correia Lopes (eds.), *3ª Conferência Nacional em XML, Aplicações e Tecnologias Aplicadas (XATA 2005)* (Braga, Fevereiro de 2005), Departamento de Informática, Universidade do Minho, pp. 351-361. **SemRec**
- [65] Ana Frankenberg-Garcia. "COMPARA - Aula Prática em Português". 14 de Março de 2005.
- [66] Belinda Maia & Luís Sarmento. "The Corpógrafo - an Experiment in Designing a Research and Study Environment for Comparable Corpora Compilation and Terminology Extraction". In *Proceedings of eCoLoRe / MeLLANGE Workshop, Resources and Tools for e-Learning in Translation and Localisation* (Centre for Translation Studies, University of Leeds, UK, 21-23 March 2005), pp. 45-48.
- [67] Bruno Martins, Marcirio Silveira Chaves & Mário J. Silva. "Assigning Geographical Scopes To Web Pages". In *27th European Conference on Information Retrieval* (Santiago de Compostela, Spain, 21-23 March 2005). **SemRec**
- [68] Belinda Maia. "Terminology and Translation -Bringing Research and Professional Training Together Through Technology". In *META Symposium - For a Proactive Translatology* (Université de Montréal, Québec, Canadá, 7-9 de Abril de 2005). **SemRec**
- [69] Luís Sarmento. "A Simple and Robust Algorithm for Extracting Terminology". In *META Symposium - For a Proactive Translatology* (Université de Montréal, Québec, Canadá, 7-9 de Abril de 2005).
- [70] Luís Cabral. "Bolsa SAGI". Apresentação no *1º Simpósio Doutoral da Linguateca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [71] Marcirio Chaves. "Ontologias: Terminologia e exemplos de uma ontologia geográfica". Apresentação no *1º Simpósio Doutoral da Linguateca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [72] Marcirio Silveira Chaves. "Plano de Doutoramento". Apresentação no *1º Simpósio Doutoral da Linguateca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [73] Marcirio Silveira Chaves. "Escarafunchando* o sistema KnowItAll". Apresentação no *1º Simpósio Doutoral da Linguateca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [74] Cristina Mota. "A Statistical Profile of the Named Entity Task (Palmer & Day)". Apresentação no *1º Simpósio Doutoral da Linguateca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).

ÂMBITO

- [75] Cristina Mota. "Aprendizagem Automática: Uma introdução". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005). **ÂMBITO**
- [76] Cristina Mota. "Plano de tese: How to Keep up with Language Dynamics? - A case study on Named Entity Recognition". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005). **ÂMBITO**
- [77] Luís Sarmento. "Descrição técnica do REPENTINO". Maio 2005.
- [78] Luís Sarmento. "Extracção Automática de Terminologia". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [79] Luís Sarmento. "BACO: BAsE de Co-Ocorrências". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [80] Luís Sarmento. "Revisão de "Automatic Acquisition and Expansion of Hypernym Links" Emmanuel Morin e Christian Jacquemin". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [81] Nuno Seco. "Determining Similarity and Inferring Relations in a Lexical Knowledge Base (Richardson)". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [82] Nuno Seco. "WordNet: Relações Semânticas e Métricas de Associação/Semelhança". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [83] Nuno Seco. "Plano de Doutoramento: Construção de uma Base de Reconhecimento Lexical para o Português". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [84] Alberto Simões. "EBMT: Example Based Machine Translation". Apresentação no *1º Simpósio Doutoral da Linguatca 2005* (Faculdade de Ciências - Universidade de Lisboa, Maio de 2005).
- [85] Diana Santos. "Relatório da Linguatca de 15 de Maio de 2004 a 14 de Maio de 2005". 2 de Junho de 2005.
- [86] Luís Cabral. "SUPeRB - Sistema Uniformizado de Pesquisa de Referências Bibliográficas". Junho de 2005. Proposta de Tese de Mestrado. **SemRec**
- [87] Ana Frankenberg-Garcia. "Pedagogical uses of Monolingual and Parallel Concordances". *English Language Teaching Journal* **59.3** (2005), pp. 189-198.
- [88] Rachel Aires, Sandra Aluísio & Diana Santos. "'Yes, user!': compiling a corpus according to what the user wants". In Pernilla Danielsson & Martijn Wagenmakers (eds.), *Proceedings from the Corpus Linguistics 2005 Conference Series* (Birmingham, UK, 14-17 July 2005), s/pp.
- [89] Ana Frankenberg-Garcia. "A corpus-based study of loan words in original and translated texts". In Pernilla Danielsson & Martijn Wagenmakers (eds.), *Proceedings from the Corpus Linguistics 2005 Conference Series* (Birmingham, UK, 14-17 July 2005), s/pp. **SemRec**
- [90] Belinda Maia, Luís Sarmento, Diana Santos, Luís Cabral & Ana Sofia Pinto. "The Corpógrafo - a Web-based environment for corpora research". Poster na *Corpus Linguistics 2005 Conference* (Birmingham, UK, 14-17 July 2005).
- [91] Débora Oliveira, Luís Sarmento, Belinda Maia & Diana Santos. "Corpus analysis for indexing: when corpus-based terminology makes a difference". In Pernilla Danielsson & Martijn Wagenmakers (eds.), *Proceedings from the Corpus Linguistics 2005 Conference Series* (Birmingham, UK, 14-17 July 2005), s/pp.
- [92] Marcário Chaves, Bruno Martins & Mário J. Silva. "GKB - Geographic Knowledge Base". DI/FCUL TR 05—12 : Department of Informatics, University of Lisbon. July 2005.
- [93] Rachel Aires, Sandra Aluísio & Diana Santos. "User-aware page classification in a search engine". In *Proceedings of Stylistic Analysis Of Text For Information Access, SIGIR 2005 Workshop* (Salvador, Bahia, Brasil, 19 de Agosto de 2005), pp. 61-68.
- [94] Rachel Virgínia Xavier Aires. Uso de marcadores estilísticos para a busca na Web em português. Tese de Doutoramento. ICMC - USP - São Carlos. Agosto de 2005.
- [95] Nuno Cardoso, Mário J. Silva & Miguel Costa. "The XLDB Group at CLEF 2004". In Carol Peters, Paul Clough, Julio Gonzalo, Gareth Jones, Michael Kluck & Bernardo Magnini (eds.), *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum (CLEF 2004)* (Bath, UK, 15-17 September 2004), Heidelberg, Alemanha: Springer. Lecture Notes in Computer Science, pp. 245-252. Versão revista de Cardoso et al. (2004)
- [96] Luís Costa. "First Evaluation of Esfinge - a Question Answering System for Portuguese". In Carol Peters, Paul Clough, Julio Gonzalo, Gareth Jones, Michael Kluck & Bernardo Magnini (eds.), *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum (CLEF 2004)* (Bath, UK, 15-17 September 2004), Heidelberg, Alemanha: Springer. Lecture Notes in Computer Science, pp. 522-533. Versão revista de Costa (2004)
- [97] Diana Santos & Paulo Rocha. "The key to the first CLEF in Portuguese: Topics, questions and answers in CHAVE". In Carol Peters, Paul Clough, Julio Gonzalo, Gareth Jones, Michael Kluck & Bernardo Magnini

- (eds.), *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum (CLEF 2004)* (Bath, UK, 15-17 September 2004), Heidelberg, Alemanha: Springer. Lecture Notes in Computer Science, pp. 821-832. Versão revista de Santos & Rocha (2004)
- [98] Alessandro Vallin, Bernardo Magnini, Christelle Ayache, Gregor Erbach, Anselmo Peñas, Maarten de Rijke, Paulo Rocha, Kiril Simov & Richard Sutcliffe. "Overview of the CLEF 2004 Multilingual Question answering track". In Carol Peters, Paul Clough, Julio Gonzalo, Gareth Jones, Michael Kluck & Bernardo Magnini (eds.), *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum (CLEF 2004)* (Bath, UK, 15-17 September 2004), Heidelberg, Alemanha: Springer. Lecture Notes in Computer Science, pp. 371-391. Versão revista de Vallin et al. (2004)
- [99] Ana Frankenberg-Garcia. "COMPARA - Instruções para a revisão do Reconhecimento Óptico de Caracteres (ROC) e etiquetagem preliminar". 19 de Setembro de 2005.
- [100] Ana Frankenberg-Garcia. "COMPARA - Optical Character Recognition (OCR) editing and preliminary markup instructions". 19 September 2005.
- [101] Nuno Cardoso, Leonardo Andrade, Alberto Simões & Mário J. Silva. "The XLDB Group participation at CLEF 2005 ad hoc task". In Carol Peters (ed.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (CLEF 2005)* (Vienna, Austria, 21-23 September 2005), s/pp.
- [102] Nuno Cardoso, Bruno Martins, Leonardo Andrade, Marcirio Silveira Chaves & Mário J. Silva. "The XLDB Group at GeoCLEF 2005". In Carol Peters (ed.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (CLEF 2005)* (Vienna, Austria, 21-23 September 2005), s/pp.
- [103] Luís Costa. "20th Century Esfinge (Sphinx) solving the riddles at CLEF 2005". In Carol Peters (ed.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (CLEF 2005)* (Vienna, Austria, 21-23 September 2005), s/pp.
- [104] Diana Santos & Nuno Cardoso. "Portuguese at CLEF 2005: Reflections and Challenges". In Carol Peters (ed.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (CLEF 2005)* (Vienna, Austria, 21-23 September 2005), s/pp.
- [105] Alessandro Vallin, Bernardo Magnini, Danilo Giampiccolo, Lili Aunimo, Christelle Ayache, Petya Osenova, Anselmo Peñas, Maarten de Rijke, Bogdan Sacaleanu, Diana Santos & Richard Sutcliffe. "Overview of the CLEF 2005 Multilingual Question Answering Track". In Carol Peters (ed.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2005 Workshop (CLEF 2005)* (Vienna, Austria, 21-23 September 2005), s/pp. **SemRec**
- [106] Isabel Marcelino. "Documentação dos recursos linguísticos subjacentes ao AnELL". 27 de Setembro de 2005.
- [107] Isabel Marcelino. "Documentação do ELLE". 29 de Setembro de 2005.
- [108] Isabel Marcelino. "Ajuda à Revisão do AnELL". 30 de Setembro de 2005.
- [109] Ana Frankenberg-Garcia. "A Peek into What Today's Language Learners as Researchers Actually Do". *International Journal of Lexicography* (2005), pp. 335-355. **SemRec**
- [110] Bruno Martins, Mário J. Silva & Marcirio Silveira Chaves. "Challenges and Resources for Evaluating Geographical IR". In *Proceedings of the Workshop on Geographic Information Retrieval at CIKM* (Bremen, Germany, October 2005), ACM Press, pp. 65-69. **SemRec**
- [111] Nuno Seco. Geração de uma ontologia lexical para o português: métodos e avaliação. Outubro de 2005. Plano de Estudos para a investigação conducente à obtenção do grau de doutor pela Faculdade de Ciências e Tecnologia da Universidade de Coimbra, especialidade de Engenharia Informática.
- [112] Marcirio Silveira Chaves, Mário J. Silva & Bruno Martins. "A Geographic Knowledge Base for Semantic Web Applications". In C. A. Heuser (ed.), *Proc. do 20º Simpósio Brasileiro de Banco de Dados (SBBDD)* (Uberlândia, MG, Brasil, 3-7 de Outubro de 2005), pp. 40-54.
- [113] Diana Santos. "Recursos da Linguateca". Apresentação no *Centro de Linguística da Universidade do Porto* (Porto, 4 de Outubro de 2005).
- [114] Luís Costa. "Esfinge - Resposta a perguntas usando a Rede". In José María Gutiérrez, Flavia Maria Santoro & Pedro Isaías (eds.), *Proceedings da conferência IADIS Ibero-Americana WWW/Internet 2005* (Lisboa, Portugal, 18-19 de Outubro de 2005), IADIS Press, pp. 616-619.
- [115] Belinda Maia, Luís Sarmiento & Diana Santos. "Introduzindo o Corpógrafo - um conjunto de ferramentas para criar corpora especializados e comparáveis e bases de dados terminológicas". *Terminómetro - Número especial nº 7 - A terminologia em Portugal e nos países de língua portuguesa em África* (2005), pp. 61-62. **SemRec**
- [116] Diana Santos & Luís Costa. "A Linguateca e o projecto 'Processamento Computacional do português'". *Terminómetro - Número especial nº 7 - A terminologia em Portugal e nos países de língua portuguesa em África* (2005), pp. 63-69.

- [117] Luís Sarmiento. "A expansão de conjuntos de co-hipónimos a partir de colecções de grandes dimensões de texto em Português ". In *Actas da 1ª Conferência em Metodologias de Investigação Científica (CoMIC'06)* (Porto, 9 de Janeiro 2006), pp. 18-30. **ÂMBITO**
- [118] Diana Santos & Marcirio Chaves. "Linguateca: presente e futuro". Apresentação no *Encontro de Processamento de Linguagem Natural* (São Leopoldo, 11 de Janeiro de 2006).
- [119] Nuno Cardoso. "Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas". Janeiro de 2006. Proposta de Tese de Mestrado.
- [120] Ana Sofia Pinto. "Neurodemo: um exemplo de extracção semi-automática de definições e relações semânticas usando o Corpógrafo". 1 de Fevereiro de 2006.
- [121] Marcirio Silveira Chaves. "Expansão de ontologia geográfica com textos em Português". Apresentação no *Seminário Doutoral do DI - FCUL* (Lisboa, 23 de Fevereiro de 2006).
- [122] Diana Santos. "HAREM: the first evaluation contest for Named Entity Recognition in Portuguese". Apresentação no *Departamentp de Matemática do IST* (Lisboa, 24 de Fevereiro de 2006).
- [123] Nuno Alexandre Lopes Seco. "Geração de uma Ontologia Lexical para o Português: Métodos e Avaliação: estado da arte". 21 de Março de 2006.
- [124] Ana Frankenberg-Garcia. "COMPARA - Instruções de alinhamento por parágrafo". 27 de Março de 2006.
- [125] Ana Frankenberg-Garcia. "COMPARA - Paragraph alignment". 27 March 2006.
- [126] Nuno Cardoso & Diana Santos. "Directivas e categorias para identificação e classificação semântica na colecção dourada do HAREM". Relatório técnico DI-FCUL TR-06-18. 29 de Março de 2006.
- [127] Nuno Cardoso, Diana Santos & Rui Vilela. "Directivas para a identificação e classificação morfológica na colecção dourada do HAREM". Relatório técnico DI-FCUL TR-06-19. 29 de Março de 2006.
- [128] Nuno Seco & Nuno Cardoso. "Detecting user sessions in the *tumba!* Web Log". March 2006. Technical report
- [129] Anabela Barreiro. "Extraction and Formalization of Support Verb Paraphrases from Corpora - Applications in Machine Translation ". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006). **ÂMBITO**
- [130] Nuno Cardoso. "Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006).
- [131] Marcirio Silveira Chaves. "Em Direção à Construção de uma Ontologia Geográfica a partir de Textos em Português na Web Portuguesa". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006).
- [132] Marcirio Silveira Chaves. "Web 2.0 e Folksonomy". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006).
- [133] Diana Santos. "Pontes no trabalho de investigação". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006).
- [134] Diana Santos. "Não deitar fora o bebé com a água". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006).
- [135] Luís Sarmiento. "A expansão de conjuntos de co-hipónimos a partir de colecções de grandes dimensões de texto em Português ". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006). **ÂMBITO**
- [136] Alberto Simões. "Tradução Automática baseada em Exemplos". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006).
- [137] Alberto Simões. "T2O". Apresentação no *2º Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 10-11 de Abril de 2006).
- [138] Diana Santos, Luís Costa & Luís Cabral. "Linguateca, a distributed resource center for language technology for Portuguese". Apresentação no *Hafjell, Norway* (Hafjell, Norway, 20-21 de Abril de 2006).
- [139] Luís Costa. "Esfinje - A Question Answering System in the Web using the Web". In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)* (Trento, Italy, April 3-7, 2006), pp. 127-130.
- [140] Ana Frankenberg-Garcia, Diana Santos & Rosário Silva. "COMPARA - Revisão do alinhamento por frase". Abril de 2006.
- [141] Ana Frankenberg-Garcia, Diana Santos & Rosário Silva. "COMPARA: Sentence alignment revision and markup". Abril de 2006.
- [142] Marcirio Silveira Chaves & Diana Santos. "What Kinds of Geographical Information Are There in the Portuguese Web?". In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *7th Workshop on Computational Processing of Written and Spoken Language (PROPOR'2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 264-267.

- [143] Susana Inácio & Diana Santos. "Syntactical Annotation of COMPARA: Workflow and First Results". In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *7th Workshop on Computational Processing of Written and Spoken Language (PROPOR'2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 256-259.
- [144] Diana Santos & Nuno Cardoso. "A Golden Resource for Named Entity Recognition in Portuguese". In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *7th Workshop on Computational Processing of Written and Spoken Language (PROPOR'2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 69-79.
- [145] Luís Sarmento. "SIEMÊS - A Named Entity Recognizer for Portuguese Relying on Similarity Rules". In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *7th Workshop on Computational Processing of Written and Spoken Language (PROPOR'2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 90-99.
- [146] Luís Sarmento, Ana Sofia Pinto & Luís Cabral. "REPENTINO - A Wide-Scope Gazetteer for Entity Recognition in Portuguese". In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *7th Workshop on Computational Processing of Written and Spoken Language (PROPOR'2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 31-40.
- [147] Nuno Seco, Diana Santos, Rui Vilela & Nuno Cardoso. "A Complex Evaluation Architecture for HAREM". In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *7th Workshop on Computational Processing of Written and Spoken Language (PROPOR'2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 260-263.
- [148] Diana Santos. "What is natural language? Differences compared to artificial languages, and consequences for natural language processing". Apresentação no PROPOR 2006 e SBLP2006 (Itatiaia, RJ, Brasil, 15 de Maio de 2006). Invited lecture.
- [149] Diana Santos. "Reconhecimento de entidades mencionadas". Apresentação na PUC-RJ (Rio de Janeiro, Brasil, 18 de Maio de 2006). Palestra convidada.
- [150] José João Almeida & Alberto Manuel Simões. "T2O - Recycling Thesauri into a Multilingual Ontology". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May 2006), pp. 1466-1471. **SemRec**
- [151] Luís Costa & Luís Sarmento. "Component Evaluation in a Question Answering System". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May 2006), pp. 1520-1523.
- [152] Bernardo Magnini, Danilo Giampiccolo, Lili Aunimo, Christelle Ayache, Petya Osenova, Anselmo Peñas, Maarten de Rijke, Bogdan Sacaleanu, Diana Santos & Richard Sutcliffe. "The Multilingual Question Answering Track at CLEF". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May 2006), pp. 1156-1163. **SemRec**
- [153] Diana Santos & Susana Inácio. "Annotating COMPARA, a grammar-aware parallel corpus". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May 2006), pp. 1216-1221.
- [154] Diana Santos, Nuno Seco, Nuno Cardoso & Rui Vilela. "HAREM: An Advanced NER Evaluation Contest for Portuguese". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May 2006), pp. 1986-1991.
- [155] Luís Sarmento. "BACO - A large database of text and co-occurrences". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May 2006), pp. 1787-1790. **SemRec**
- [156] Luís Sarmento, Belinda Maia, Diana Santos, Ana Pinto & Luís Cabral. "Corpógrafo V3: From Terminological Aid to Semi-automatic Knowledge Engine". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May 2006), pp. 1502-1505.
- [157] Belinda Maia & Luís Sarmento. "Corpógrafo - Applications". In *Third International Workshop on Language Resources for Translation Work Research & Training, Satellite event of LREC 2006 (LR4Trans-III)* (28 May 2006), pp. 55-58. **SemRec**

- [158] Luís Sarmiento. "Corpógrafo - Um ambiente livre para o ensino e desenvolvimento de Terminologia". Apresentação na *PUCRS e PUCRJ* (PUCRS (Porto Alegre, Brasil) e PUCRJ (Rio de Janeiro, Brasil), Maio de 2006).
- [159] Alberto Manuel Simões & José João Almeida. "Combinatory Examples Extraction for Machine Translation". In *EAMT 11th Annual Conference* (Oslo, Norway, 19-20 June, 2006), pp. 27-32.
- [160] José João Almeida & Alberto Manuel Simões. "Publishing multilingual ontologies: a quick way of obtaining feedback". In *Electronic Publishing 2006 - Digital Spectrum: Integration Technology and Culture (ELPub 2006)* (Bansko, Bulgaria, June 2006). **SemRec**
- [161] J. João Dias de Almeida. "Tradução Automática.". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). **ÂMBITO**
- [162] J. João Dias de Almeida. "Corpora". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). **ÂMBITO**
- [163] Eckhard Bick. "Automatic Syntactical Annotation". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). **ÂMBITO**
- [164] Ana Frankenberg-Garcia. "Corpora e Tradução". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). **ÂMBITO**
- [165] Ana Frankenberg-Garcia. "Corpora e Ensino de línguas". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). **ÂMBITO**
- [166] Ana Frankenberg-Garcia. "Entrando em linha de conta com os utilizadores: Exemplo concreto do COMPARA". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). ,
- [167] Belinda Maia. "Terminologia e Ontologias". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [168] Belinda Maia. "Tradução Automática". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [169] Belinda Maia. "Corpora Comparáveis". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [170] Thiago Pardo. "Sumarização Automática: Métodos e Avaliação". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). **ÂMBITO**
- [171] Elisabete Marques Ranchhod & Paula Carvalho. "Expressões Multipalavra - Questões lexicais e sintáticas". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006). **ÂMBITO**
- [172] Diana Santos. "Abertura científica: processamento computacional da língua portuguesa". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [173] Diana Santos. "Desenho, construção e utilização de corpora". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [174] Diana Santos. "Tradução automática: a tecnologia e sua aplicação". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [175] Mário J. Silva. "Ontologias & Terminologias: Perspectivas da engenharia". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [176] Diana Santos. "Avaliação: fundamentos e sua aplicação ao PLN". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [177] Diana Santos. "Corpora". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [178] Diana Santos. "Entrando em conta com os utilizadores". *Material de ensino na Primeira Escola de Verão da Linguateca* (UP, Porto, Portugal, 10-14 Julho de 2006).
- [179] José João Dias de Almeida & Edgar Alves. "RENA - Reconhecedor de Entidades". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006). **ÂMBITO**,
- [180] Eckhard Bick. "Functional Aspects on Portuguese NER". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006). **ÂMBITO**
- [181] Luís M. Cabral. "SIEMES e uma breve introdução ao REPENTINO". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006).
- [182] Nuno Cardoso. "HAREM e MiniHAREM: Uma análise comparativa". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006).
- [183] Bruno Martins, Marcirio Chaves & Mario J. Silva. "O sistema CaGE para Reconhecimento de referências geográficas em textos na Língua Portuguesa". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006).
- [184] Cristina Mota & Max Silberztein. "Stencil-NooJ no HAREM". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006). **ÂMBITO**

- [185] Christian Nunes. "Cortex Intelligence". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006). **ÂMBITO**
- [186] Diana Santos. "Visão do futuro do HAREM". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006).
- [187] Luís Sarmiento. "Apresentação do SIEMES2". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006).
- [188] Luís Sarmiento & Cristina Mota. "HAREM 2.0 Proposta". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006). **ÂMBITO**
- [189] Nuno Seco. "MUC e HAREM: Uma visão contrastiva ". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006).
- [190] Rui Vilela. "HAREM - Avaliação conjunta de sistemas de reconhecimento de entidades mencionadas". Apresentação no *Encontro HAREM* (Porto, Portugal, 15 de Julho de 2006).
- [191] Diana Santos. "Breves explorações num mar de língua". *Ilha do Desterro* **50** (2006). Aceite para publicação: Outubro de 2004.
- [192] Nuno Seco. "Building a Large Scale lexical Ontology for Portuguese". Apresentação no *SINTEF* (Oslo, 16 de Agosto de 2006).
- [193] Diana Santos & Marcirio Chaves. "The place of place in geographical IR". In *Proceedings of GIR06, the 3rd Workshop on Geographic Information Retrieval (SIGIR 2006)* (Seattle, 10th August 2006), pp. 5-8.
- [194] Ana Frankenberg-Garcia. "Using a parallel corpus in translation practice and research". In *Actas da Contrapor 2006, 1ª Conferência de Tradução Portuguesa, Faculdade de Ciências e Tecnologia* (Universidade Nova de Lisboa, 11-12 de Setembro de 2006), pp. 142-148. **SemRec**
- [195] Nuno Cardoso, Mario J. Silva & Bruno Martins. "The University of Lisbon at CLEF 2006 Ad-Hoc Task". In Alessandro Nardi, Carol Peters & José Luís Vicedo (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop (CLEF 2006)* (Alicante, Espanha, 20-22 de Setembro de 2006).
- [196] Luís Costa. "Esfinge - a modular question answering system for Portuguese". In Alessandro Nardi, Carol Peters & José Luís Vicedo (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop (CLEF 2006)* (Alicante, Espanha, 20-22 de Setembro de 2006).
- [197] Fredric Gey, Ray Larson, Mark Sanderson, Kerstin Bischoff, Thomas Mandl, Christa, Womser-Hacker, Diana Santos, Paulo Rocha, Giorgio M. Di Nunzio & Nicola Ferro. "GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview". In Alessandro Nardi, Carol Peters & José Luís Vicedo (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop (CLEF 2006)* (Alicante, Espanha, 20-22 de Setembro de 2006).
- [198] Bernardo Magnini, Danilo Giampiccolo, Pamela Forner, Christelle Ayache, Valentin Jijkoun, Petya Osenova, Anselmo Peñas, Paulo Rocha, Bogdan Sacaleanu & Richard Sutcliffe. "Overview of the CLEF 2006 Multilingual Question Answering Track". In Alessandro Nardi, Carol Peters & José Luís Vicedo (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop (CLEF 2006)* (Alicante, Espanha, 20-22 de Setembro de 2006).
- [199] Bruno Martins, Nuno Cardoso, Marcirio Silveira Chaves, Leonardo Andrade & Mário J. Silva. "The university of Lisbon at GeoCLEF 2006". In Alessandro Nardi, Carol Peters & José Luís Vicedo (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop (CLEF 2006)* (Alicante, Espanha, 20-22 de Setembro de 2006). **SemRec**
- [200] Luís Sarmiento. "Hunting Answers with RAPOSA (FOX)". In Alessandro Nardi, Carol Peters & José Luís Vicedo (eds.), *Cross Language Evaluation Forum: Working Notes for the CLEF 2006 Workshop (CLEF 2006)* (Alicante, Espanha, 20-22 de Setembro de 2006). **SemRec**
- [201] Nuno Cardoso, Leonardo Andrade, Alberto Simões & Mário J. Silva. "The XLDB Group participation at CLEF 2005 ad hoc task". In Carol Peters, Frederic Gey, Julio Gonzalo, Henning Müeller, Gareth J.F. Jones, Michael Kluck, Bernardo Magnini & Maarten de Rijke (eds.), *Accessing Multilingual information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005. Vienna, Austria, September 2005. Revised Selected papers*. Springer Berlin / Heidelberg, Lecture Notes in Computer Science, 2006, pp. 54-60.
- [202] Nuno Cardoso, Bruno Martins, Leonardo Andrade, Marcirio Silveira Chaves & Mário J. Silva. "The XLDB Group at GeoCLEF 2005". In Carol Peters, Frederic Gey, Julio Gonzalo, Henning Müeller, Gareth J.F. Jones, Michael Kluck, Bernardo Magnini & Maarten de Rijke (eds.), *Accessing Multilingual information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005. Vienna, Austria, September 2005. Revised Selected papers*. Springer Berlin / Heidelberg, Lecture Notes in Computer Science, 2006, pp. 997-1006. Short version of Cardoso et al. (2005)
- [203] Luís Costa. "20th Century Esfinge (Sphinx) solving the riddles at CLEF 2005". In Carol Peters, Frederic Gey, Julio Gonzalo, Henning Müeller, Gareth J.F. Jones, Michael Kluck, Bernardo Magnini & Maarten de Rijke (eds.), *Accessing Multilingual information Repositories: 6th Workshop of the Cross-Language Evaluation*

- Forum, CLEF 2005. Vienna, Austria, September 2005. Revised Selected papers.* Springer Berlin / Heidelberg, Lecture Notes in Computer Science, 2006, pp. 467-476.
- [204] Diana Santos & Nuno Cardoso. "Portuguese at CLEF". In Carol Peters, Frederic Gey, Julio Gonzalo, Henning Müeller, Gareth J.F. Jones, Michael Kluck, Bernardo Magnini & Maarten de Rijke (eds.), *Accessing Multilingual information Repositories:6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005. Vienna, Austria, September 2005. Revised Selected papers.* Springer Berlin / Heidelberg, Lecture Notes in Computer Science, 2006, pp. 1007-1010.
- [205] Alessandro Vallin, Bernardo Magnini, Danilo Giampiccolo, Lili Aunimo, Christelle Ayache, Petya Osenova, Anselmo Peñas, Maarten de Rijke, Bogdan Sacaleanu, Diana Santos & Richard Sutcliffe. "Overview of the CLEF 2005 Multilingual Question Answering Track". In Carol Peters, Frederic Gey, Julio Gonzalo, Henning Müeller, Gareth J.F. Jones, Michael Kluck, Bernardo Magnini & Maarten de Rijke (eds.), *Accessing Multilingual information Repositories:6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005. Vienna, Austria, September 2005. Revised Selected papers.* Springer Berlin / Heidelberg, Lecture Notes in Computer Science, 2006, pp. 307-331. Short version of Vallin et al. (2005)
- [206] Diana Santos & Luís Costa. "Where is future QA?: One pilot and two main proposals". Apresentação no *CLEF 2006 Workshop* (Alicante, 22 de Setembro de 2006).
- [207] Alberto Simões & José João Almeida. "NatServer: A Client-Server Architecture for building Parallel Corpora applications". *Sociedade Española para el Procesamiento del Lenguaje Natural* (2006). **SemRec**
- [208] Nuno Cardoso. Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas. Tese de Mestrado. Faculdade de Engenharia da Universidade do Porto. Outubro de 2006.
- [209] Anabela Barreiro. "Tradução Automática com Paráfrases Bilingues Inglês-Português". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006). **ÂMBITO**
- [210] Luís Miguel Cabral. "SUPeRB". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006).
- [211] Nuno Cardoso. "Automatic Query Reformulation using Contextual Information". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006).
- [212] Paula Carvalho. "Análise e Representação de Construções Adjectivais para Processamento Automático de Texto". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006). **ÂMBITO**
- [213] Marcirio Chaves. "Estado da arte: Extracção de informação (geográfica)". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006).
- [214] Marcirio Chaves. "Uma metodologia para construção de ontologias e integração de conhecimento geográfico". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006).
- [215] Bruno Martins. "Geographically Aware Web Text Mining". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006). **ÂMBITO**
- [216] Paulo Rocha. "CLEF e CHAVE, alguns comentários". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006).
- [217] Diana Santos. "Objectivos do III simpósio doutoral da Linguatca, comentários e pontes". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006).
- [218] Luís Sarmiento. "Estado da arte de aprendizagem automática em EI". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006). **ÂMBITO**
- [219] Luís Sarmiento. "Medidas de associação". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006). Tutorial **ÂMBITO**
- [220] Alberto Simões. "Extracção e Generalização de Recursos de Tradução com base em Dicionários Probabilísticos". Apresentação no *3º Simpósio Doutoral da Linguatca 2006* (Faculdade de Ciências - Universidade de Lisboa, 3-4 de Outubro de 2006). ,
- [221] Luís Miguel Cabral. "Diagramas e Exemplos do SUPeRB ". 15 de Dezembro de 2006.
- [222] Luís Sarmiento. "BACO e a sua API". Apresentação nas *Mini-Jornadas de PLN* (Departamento de Informática - Universidade do Minho, 8 de Novembro de 2006). **ÂMBITO**
- [223] Diana Santos. "A tradução na sociedade do conhecimento". Apresentação no *IX Seminário de Tradução e Técnica em Língua Portuguesa* (Instituto Franco-Português, Lisboa, 13 de Novembro de 2006). Palestra convidada.
- [224] Catarina Rodrigues, Marcirio Chaves & Mário J. Silva. "Uma representação ontológica da geografia física de Portugal". In *IX Encontro de utilizadores de Informação Geográfica* (Tagus Park, Oeiras, 15-17 de Novembro de 2006), s/pp. **SemRec**

- [225] Diana Santos, Luís Miguel Cabral & Luís Costa. "Linguatca: seven years working for the computational processing of Portuguese". Apresentação no *SINTEF* (Oslo, 23rd November 2006).
- [226] Diana Santos, Nuno Cardoso & Nuno Seco. "Avaliação no HAREM: Métodos e medidas". Relatório técnico DI-FCUL TR-06-17. Novembro de 2006.
- [227] Ana Sofia Pinto. "Introdução à Utilização do Corpógrafo: Um pequeno tutorial". 13 de Dezembro de 2006. Primeira versão: 21 de Outubro de 2004 (Em constante redacção)
- [228] Susana Afonso. "Árvores deitadas: Descrição do formato e das opções de análise na Floresta Sintáctica". 2006. Primeira versão: 2004 (Em constante redacção)
- [229] Susana Inácio & Diana Santos. "Documentação da anotação da parte portuguesa do COMPARA". 2006. Primeira versão: 9 de Dezembro de 2005 (Em constante redacção)
- [230] Diana Santos. "Tradução: uma tecnologia humana de ponta OU A tradução na sociedade do conhecimento OU Ciência E Tradução". In *Actas do IX Seminário de Tradução e Técnica em Língua Portuguesa* (2006), Paris: União Latina.
- [231] Diana Santos. "Resumo da actividade da Linguatca de 15 de Maio de 2003 a 15 de Dezembro de 2006". 2006. Com a colaboração (por ordem alfabética) de Alberto Simões, Ana Frankenberg-Garcia, Belinda Maia, Luís Costa, Luís Miguel Cabral, Luís Sarmento, Marcirio Chaves, Mário J. Silva, Nuno Cardoso, Paulo Gomes e Rui Vilela
- [232] Luís Sarmento. "Carros, barcos, mulheres e outras 'coisas semelhantes' (ou 'Porque é que nem tudo é o que parece!') ". Apresentação na *PUCPR, UFRGS, PUCRS e PUCRJ*. (PUCPR (Curitiba), UFRGS (Porto Alegre), PUCRS (Porto Alegre) e PUCRJ (Rio Janeiro), Brasil, Maio de 2006). **ÂMBITO**
- [233] Nuno Seco. "Detecção de Sessões de Utilização nos diários do Tumba!". Apresentação na *Universidade de Coimbra* (Coimbra, 2006).
- [234] Nuno Seco. "Extracção Automática de Ontologias a partir de Texto: Parte I: Introdução". Apresentação na *Universidade de Coimbra* (Coimbra, 2006).
- [235] Nuno Seco. "Extracção Automática de Ontologias: Métodos e Avaliação". Apresentação na *Universidade de Coimbra* (Coimbra, 2006).
- [236] Mário J. Silva, Bruno Martins, Marcirio Silveira Chaves, Ana Paula Afonso & Nuno Cardoso. "Adding Geographic Scopes to Web Resources". In *CEUS - Computers Environment and Urban Science*. Elsevier Science, 2006, pp. 378-399. **SemRec**

2007 (54)

- [237] Rachel Virgínia Xavier Aires & Sandra Maria Aluísio. "As avaliações atuais de sistemas de busca na Web e a importância do usuário". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [238] J. João Dias de Almeida. "RENA - Reconhecedor de Entidades". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguatca, 2007. **ÂMBITO**,
- [239] J. João Dias de Almeida. "HAREM: subsídio para a discussão sobre futuras edições". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguatca, 2007.
- [240] José João Almeida & Alberto Simões. "Jspellando nas morfolimpíadas: Sobre a participação do Jspell nas Morfolimpíadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [241] Christian Nunes Aranha. "Cortex Entity-Extractor". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguatca, 2007. **ÂMBITO**
- [242] Jorge Baptista. "Alguns comentários sobre a lista dourada". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [243] Anabela Barreiro & Susana Afonso. "Construção da lista dourada para as primeiras Morfolimpíadas do português". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [244] Eckhard Bick. "Morfologia com sintaxe debaixo de olho: Sobre a participação do PALMORF nas Morfolimpíadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [245] Eckhard Bick. "Functional Aspects on Portuguese NER". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguatca, 2007. Este artigo foi publicado pela Springer, na série LNAI, vol. 3960, ISBN-10: 3-540-34045-9. **ÂMBITO**

- [246] Eckhard Bick, Diana Santos, Susana Afonso & Rachel Marchi. "Floresta Sintá(c)tica: Ficção ou realidade?". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [247] Luís M. Cabral. SUPeRB - Sistema Uniformizado de Pesquisa de Referências Bibliográficas. Tese de Mestrado. Faculdade de Engenharia da Universidade do Porto. 2007.
- [248] Nuno Cardoso. "Panorama de REM em português". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [249] Nuno Cardoso, Bruno Martins, Daniel Gomes & Mário J. Silva. "WPT 03: a primeira colecção pública proveniente de uma recolha da web portuguesa". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [250] Nuno Cardoso, Mário J. Silva & Marília Antunes. "Validação Estatística dos resultados do HAREM". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [251] Nuno Cardoso, Mario J. Silva & Bruno Martins. "The University of Lisbon at CLEF 2006 Ad-Hoc Task". In Carol Peters, Paul Clough, Fredric C. Gey, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini & Maarten de Rijke Julio Gonzalo (eds.), *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September 2006. Revised Selected papers*. Berlin / Heidelberg: Springer, Lecture Notes in Computer Science, 2007.
- [252] Luís Costa. "Question answering beyond CLEF document collections". In Carol Peters, Paul Clough, Fredric C. Gey, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini & Maarten de Rijke Julio Gonzalo (eds.), *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September 2006. Revised Selected papers*. Berlin / Heidelberg: Springer, Lecture Notes in Computer Science, 2007.
- [253] Luís Costa, Paulo Rocha & Diana Santos. "Organização e resultados morfolímpicos". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [254] Óscar Ferrández, Zornitsa Kozareva, Antonio Toral, Rafael Muñoz & Andrés Montoyo. "Tackling HAREM's Portuguese Named Entity Recognition task with Spanish resources". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**
- [255] Marco Gonzalez, Leonardo Cavalheiro Langie & Vera Lúcia Strube de Lima. "Avaliação de Sistemas de recuperação e categorização de textos: métodos e aplicações". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [256] Caroline Hagège. "Segmentação e análise morfológica do português com recursos léxicos limitados: Sobre a participação do Smorph nas Morfolimpíadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [257] Ricardo Ueda Karpischeck. "A Morfolimpíada e a avaliação da verificação ortográfica: Sobre a participação do br.ispell nas Morfolimpíadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [258] Belinda Maia & Anabela Barreiro. "Uma experiência de recolha de exemplos classificados de tradução automática de inglês para português". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [259] Bruno Martins, Nuno Cardoso, Marcirio Silveira Chaves, Leonardo Andrade & Mário J. Silva. "The university of Lisbon at GeoCLEF 2006". In Carol Peters, Paul Clough, Fredric C. Gey, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini & Maarten de Rijke Julio Gonzalo (eds.), *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September 2006. Revised Selected papers*. Berlin / Heidelberg: Springer, Lecture Notes in Computer Science, 2007.
- [260] Ronaldo Martins, Ricardo Hasegawa & Maria das Graças Volpe Nunes. "Sobre a participação do ReGra no HAREM 2005". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**
- [261] Ronaldo Teixeira Martins & Maria das Graças Volpe Nunes. "Dos processos de individuação e de categorização lexical: sobre a participação do ReGra nas Morfolimpíadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [262] Bruno Martins & Mário Silva. "O HAREM e a avaliação de sistemas para o reconhecimento de entidades geográficas em textos da língua portuguesa". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**

- [263] Bruno Martins, Mário Silva & Marcirio Chaves. "O sistema CaGE no HAREM : Reconhecimento de Entidades Geográficas em Textos na Língua Portuguesa". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [264] Cristina Mota. "Documentação pormenorizada do primeiro ensaio de SREM para português". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**
- [265] Cristina Mota, Diana Santos & Elisabete Ranchhod. "Avaliação de reconhecimento de entidades mencionadas: princípio de AREM". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [266] Cristina Mota, Luís Sarmento & Paula Carvalho. "O Bom, o Mau e o Vilão: 3 perspectivas ao REM". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**
- [267] Cristina Mota & Max Silberztein. "Stencil/NooJ no HAREM". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**
- [268] Viviane Moreira Orengo & Diana Santos. "Radicalizadores versus analisadores morfológicos: Sobre a participação do Removedor de Sufixos da Língua Portuguesa nas Morfolimpíadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [269] Paulo Quaresma & Irene Rodrigues. "Avaliação de sistemas interactivos de recuperação de informação em bases de texto jurídicas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [270] Elisabete Ranchhod & Cristina Mota. "Unidades lexicais multipalavra, um osso duro de roer: Sobre a participação do LabEL nas Morfolimpíadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [271] Lucia Helena Machado Rino & Thiago Alexandre Salgueiro Pardo. "A Coleção TeMário e a avaliação de sumarização automática". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [272] Paulo Rocha & Diana Santos. "CLEF: Abrindo a porta à participação internacional em avaliação de RI do português". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [273] Diana Santos. "O modelo semântico usado no primeiro HAREM: Defesa e discussão". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [274] Diana Santos. *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. Aceite para publicação: Maio de 2006.
- [275] Diana Santos. "Avaliação conjunta". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [276] Diana Santos. "Perfect mismatches: Result in English and Portuguese". In Margaret Rogers & Gunilla Anderman (eds.), *Incorporating Corpora - The Linguist and the Translator*. Clevedon: Multilingual matters, 2007.
- [277] Diana Santos, Eckhard Bick, Raquel Marchi & Susana Afonso. "A Floresta Sintá(c)tica como recurso para estudar a língua portuguesa". In Regina Gerber & Vera Vasilévski (eds.), *Um percurso para pesquisas com base em corpus*. Florianópolis - SC: Editora da UFSC, 2007.
- [278] Diana Santos & Nuno Cardoso. *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **SemRec**
- [279] Diana Santos & Nuno Cardoso. "Balanço do primeiro HAREM e perspectivas de trabalho futuro". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [280] Diana Santos, Nuno Cardoso, Nuno Seco & Rui Vilela. "Breve Introdução ao HAREM". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [281] Diana Santos & Ana Frankenberg-Garcia. "The corpus, its users and their needs: a user-oriented evaluation of COMPARA". *International Journal of Corpus Linguistics* **12** (2007). Aceite para publicação: Agosto de 2006.
- [282] Luís Sarmento. "O SIEMÊS e a sua participação no HAREM e no Mini-HAREM". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**

- [283] Luís Sarmiento. "Ferramentas para experimentação, recolha e avaliação de exemplos de tradução automática". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [284] Luís Sarmiento. "A first step to address biography generation as an iterative QA task". In Carol Peters, Paul Clough, Fredric C. Gey, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini & Maarten de Rijke Julio Gonzalo (eds.), *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September 2006. Revised Selected papers*. Berlin / Heidelberg: Springer, Lecture Notes in Computer Science, 2007.
- [285] Luís Sarmiento, Anabela Barreiro, Belinda Maia & Diana Santos. "Avaliação de Tradução Automática: alguns conceitos e reflexões". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [286] Nuno Seco. "MUC vs HAREM: A Contrastive Perspective". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [287] Nuno Seco, Diana Santos, Nuno Cardoso & Rui Vilela. "A arquitectura dos programas de avaliação do HAREM". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007.
- [288] Mário Silva, Bruno Martins & Miguel Costa. "Avaliação conjunta de recuperação de informação da Web portuguesa". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007. **ÂMBITO**
- [289] Alberto Simões & José João Almeida. "Avaliação de alinhadores". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, 2007.
- [290] Tamar Solorio. "MALINCHE: A NER system for Portuguese that reuses knowledge from Spanish". In Diana Santos & Nuno Cardoso (eds.), *HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro*. Linguateca, 2007. **ÂMBITO**
- [291] Fredric Gey, Ray Larson, Mark Sanderson, Kerstin Bischoff, Thomas Mandl, Christa, Womser-Hacker, Diana Santos, Paulo Rocha, Giorgio M. Di Nunzio & Nicola Ferro. "GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview". In Carol Peters, Paul Clough, Fredric C. Gey, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini & Maarten de Rijke Julio Gonzalo (eds.), *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September 2006. Revised Selected papers*. Berlin / Heidelberg: Springer, Lecture Notes in Computer Science, 2007.
- [292] Bernardo Magnini, Danilo Giampiccolo, Pamela Forner, Christelle Ayache, Valentin Jijkoun, Petya Osenova, Anselmo Peñas, Paulo Rocha, Bogdan Sacaleanu & Richard Sutcliffe. "Overview of the CLEF 2006 Multilingual Question Answering Track". In Carol Peters, Paul Clough, Fredric C. Gey, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini & Maarten de Rijke Julio Gonzalo (eds.), *7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September 2006. Revised Selected papers*. Berlin / Heidelberg: Springer, Lecture Notes in Computer Science, 2007.

6. Outras fontes citadas no presente relatório

- [293] Brill, Eric. "Processing Natural Language without Natural Language Processing", in A. Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing: 4th International Conference, CICLing 2003 Mexico City, Mexico, February 16-22, 2003. Proceedings*, LNCS Volume 2588, Springer-Verlag Heidelberg, 2003, pp.360-369.
- [294] Buchholz, Sabine & Darren Green. "Quality control of treebanks: documenting, converting, patching". In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odjik & Daniel Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)* (Genoa, Italy, 22-28 May de 2006), pp. 26-31.
- [295] Danielsson, Pernilla & Daniel Ridings. In Ann Lawson (ed.), "Practical Presentation of a "Vanilla" Aligner". In *TELRI Workshop in alignment and exploitation of texts* (February 1997).
- [296] Hiemstra, Djoerd. Using statistical methods to create a bilingual dictionary. Master's thesis, Department of Computer Science, University of Twente, August 1996.
- [297] Koehn, Philipp. "Europarl: A Multilingual Corpus for Evaluation of Machine Translation" Unpublished Draft, <http://people.csail.mit.edu/~koehn/publications/europarl.ps>.
- [298] Maia, Belinda. "Do-it-yourself corpora... with a little bit of help from your friends". In Lewandowska-Tomaszczyk, B. & P.J. Melia (eds.), *PALC'97: practical applications in language corpora* Lodz. Lodz University Press, 10-14 April 1997, pp. 403-410.
- [299] Rocha, Paulo Alexandre & Diana Santos. "CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa", in Maria das Graças Volpe Nunes (ed.), *Actas do V Encontro para o*

processamento computacional da língua portuguesa escrita e falada (PROPOR'2000), Atibaia, São Paulo, Brasil (19 a 22 de Novembro de 2000), pp.131-140.

[300] Silva, Mário J. "The Case for a Portuguese Web Search Engine". Technical Report TR-03-3, Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa, March 2003.

[301] Simões, Alberto Manuel & José João Almeida. "jspell.pm -- um módulo de análise morfológica para uso em Processamento de Linguagem Natural", In Anabela Gonçalves & Clara Nunes Correia (eds.), *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística (APL 2001)* (Lisboa, 2-4 de Outubro de 2001), Lisboa: APL, pp. 485-495.

[302] Wing, Benjamin & Jason Baldrige. "Adaptation of Data and Models for Probabilistic Parsing of Portuguese". In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *VII Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 140-149.

Indicadores de realização física

		Unidade: em número
A- Publicações		
Livros	2	<input type="checkbox"/> <input type="checkbox"/>
Capítulos de livros	42	<input type="checkbox"/> <input type="checkbox"/>
Artigos em revistas internacionais	9	<input type="checkbox"/> <input type="checkbox"/>
Artigos em revistas nacionais	0	<input type="checkbox"/> <input type="checkbox"/>
B- Comunicações		
Em congressos científicos internacionais	64	<input type="checkbox"/> <input type="checkbox"/>
Em congressos científicos nacionais	13	<input type="checkbox"/> <input type="checkbox"/>
C- Relatórios	38	<input type="checkbox"/> <input type="checkbox"/>
D- Organização de seminários e conferências	6	<input type="checkbox"/> <input type="checkbox"/>
E- Formação Avançada		<input type="checkbox"/> <input type="checkbox"/>
Teses de Doutoramento	1 (3)	<input type="checkbox"/> <input type="checkbox"/>
Teses de Mestrado	3 (4)	<input type="checkbox"/> <input type="checkbox"/>
Outra		<input type="checkbox"/> <input type="checkbox"/>
F- Modelos		<input type="checkbox"/> <input type="checkbox"/>
G- Aplicações computacionais	18	<input type="checkbox"/> <input type="checkbox"/>
H- Instalações Piloto		<input type="checkbox"/> <input type="checkbox"/>
I- Protótipos laboratoriais		<input type="checkbox"/> <input type="checkbox"/>
J- Patentes		<input type="checkbox"/> <input type="checkbox"/>
L- Outros (discriminar)	18	<input type="checkbox"/> <input type="checkbox"/>
Recursos	10	
Serviços na rede	9	

Recursos: CETEMPúblico, CETENFolha, CHAVE, WPT 03, GeoNet-PT-01, Floresta, REPENTINO, BACO, COMPARA, "Yes, User!"

Aplicações: Corpógrafo, NATools, Esfinge, GKB, PLN::Base, SIEMÊS, Esfinge, Águia, Ambiente de avaliação do HAREM, Validador da Floresta, DISPARA, SUPeRB, T2O, RAPOSA, SEI-Geo, Leva-e-Traz, METRA, TrAva

Serviços: AC/DC, COMPARA, Floresta, WebJSPELL, NATools, Esfinge, AnELL, Busca no Catálogo, SUPeRB