

Resumo da actividade da Linguateca de 16 de Dezembro de 2006 a 31 de Dezembro de 2007

Coordenado por Luís Costa

Este relatório divide-se em três secções principais que correspondem às 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 a que se refere o relatório. Todo o trabalho aqui descrito é a continuação lógica do trabalho iniciado em 2000, e relatado pormenorizadamente em [105, 106]. O presente relatório foi redigido por vários membros do projecto, cujos nomes aparecem sob as secções respectivas.

1. Avaliação

1.1 HAREM

Diana Santos

Em 2007 iniciou-se a organização da segunda edição do HAREM, uma avaliação conjunta sobre o reconhecimento de entidades mencionadas, organizada pela Linguateca, que recebeu o recorde de 22 inscritos, cf. www.linguateca.pt/HAREM/ e que terá lugar em Março/Abril de 2008.

Além da preparação de um novo sítio, em português e inglês, em 2007 foi desenvolvido um sistema de ajuda à anotação manual da colecção dourada, o Etiquet(H)AREM, e novas directivas e pistas foram aprovadas e definidas, assim como foi disponibilizada uma pequena colecção exemplo.

1.2 CLEF

1.2.1. Resposta automática a perguntas

Paulo Rocha

Realizaram-se as seguintes actividades no âmbito da organização da parte portuguesa da tarefa de resposta automática a perguntas:

- Discussão com os restantes membros da organização sobre os moldes de realização do QA@CLEF em 2007 e respectivas directrizes.
- Preparação dos novos textos a utilizar na parte portuguesa.
- Selecção das 200 perguntas a realizar aos sistemas na pista monolingue portuguesa.
- Tradução para português das 800 perguntas a realizar aos sistemas nas pistas bilingues com o português como fonte.
- Avaliação das respostas da pista monolingue portuguesa e das pistas bilingues com o português como alvo.
- Escrita da secção portuguesa do artigo geral sobre o QA@CLEF para as working notes e colaboração no resto do artigo [41]; reescrita da secção referida para publicação internacional [99].

1.2.2. GeoCLEF

Diana Santos

Em 2007, a Linguateca participou, como em 2006, na organização do GeoCLEF, tendo dado origem a várias publicações e apresentações [39, 40, 49, 101]

Além disso, coube-nos também em Dezembro a edição da parte do GGeoCLEF para o volume resultante da Springer, organizando 11 revisores e as suas recensões para o volume a publicar em 2008.

Foi também efectuada a tradução/localização do DIRECT para português, assim como variadas sugestões de melhoria da usabilidade e das funcionalidades deste sistema.

Finalmente, houve também a participação do pólo da Linguateca no XLDB no GeoCLEF [17, 97]

2. Recursos

2.1 AC/DC

Diana Santos

Neste projecto incorporaram-se três novos corpora: CHAVE (dando acesso à versão anotada pelo PALAVRAS [104] desta colecção), Vercial (contendo textos literários portugueses disponibilizados pelo projecto Vercial) e CDHAREM (dando acesso à colecção dourada do primeiro HAREM, também anotada pelo PALAVRAS).

Além disso, foi reformulada a interface (de HTML para PHP), introduzindo alguma sofisticação adicional e documentação e exemplos específicos para cada corpus, e foram criadas versões novas da anotação (pelo PALAVRAS) de praticamente todos os corpora.

Publicou-se um artigo e houve uma apresentação pública sobre o AC/DC. [62, 60]

2.2 Catálogo de publicações da Linguateca e SUPeRB

Luís Cabral

No ano de 2007 foram implementadas melhorias no catálogo de publicações tais como a criação de novos formulários de inserção de dados e o melhoramento da pesquisa de publicações e da integração do catálogo com o SUPeRB, uma ferramenta desenvolvida no âmbito do mestrado do Luís Miguel Cabral [12]. O objectivo principal foi melhorar o processo de validação e a descoberta de novos candidatos a entradas no catálogo através de métodos automáticos e permitir a conversão entre formatos bibliográficos.

2.3 COMPARA

Ana Frankenberg-Garcia e Diana Santos

Durante este período o corpus sofreu uma série de melhorias: foi concluída a revisão textual de inserção de marcas de parágrafo; iniciou-se uma revisão pormenorizada dos textos digitalizados por terceiros, com o objectivo de garantir a qualidade da digitalização de todo o corpus; e deu-se início à sistematização do registo dos erros detectados nas edições impressas que formam a base do COMPARA. Além disso, algumas novas funcionalidades foram implementadas.

No âmbito da revisão da anotação gramatical da parte portuguesa do corpus, concluiu-se o processo de revisão dos adjectivos, deu-se início à revisão dos verbos e foram revistos os substantivos

dos 17 novos ficheiros acrescentados posteriormente ao corpus.

Iniciou-se a anotação semântica do COMPARA, em ambas as línguas, no campo da cor.

Deu-se continuidade ao processo de documentação de todas as decisões inerentes à anotação gramatical em português. Actualizaram-se as instruções de digitalização dos textos. Foram totalmente revistos e actualizados os tutoriais de utilização do corpus em versão portuguesa e inglesa.

Finalmente, foram efectuadas várias apresentações sobre o COMPARA, assim como escritos, publicados ou enviados para publicação vários artigos sobre o mesmo. [31, 32, 33, 34, 35, 36, 76, 85, 98, 100, 102]

2.4 Corpógrafo

Sérgio Matos

As seguintes funcionalidades foram acrescentadas ao Corpógrafo: a cópia de termos entre bases de dados terminológicas, incluindo as relações já definidas entre os termos copiados; a pesquisa por textos na internet, a partir de termos gravados nas bases de dados terminológicas, permitindo acrescentar novos textos especializados aos corpora já existentes; a criação de bases de dados de expressões lexicais e discursivas, incluindo a definição de relações léxico-semânticas e discursivas para classificar estas expressões; e a integração do sistema de processamento de linguagem natural NooJ (<http://www.nooj4nlp.net/>), que permite pesquisar por expressões lexicais/discursivas utilizando etiquetas gramaticais (em português, inglês e francês).

Além destas novas funcionalidades, foram também actualizadas as taxonomias utilizadas na organização e classificação dos ficheiros, corpora e bases de dados (terminológicas e fraseológicas) criados pelos utilizadores. Para tornar estas classificações mais úteis, criou-se também a possibilidade de os utilizadores acrescentarem os seus próprios itens às taxonomias (dentro da sua área de trabalhos pessoal).

2.5 Esfinge

Luís Costa

O Esfinge participou pela quarta vez na tarefa de RAP (resposta automática a perguntas) monolíngue português do CLEF [41, 99]. Uma parte dos esforços foram dirigidos à resposta aos novos desafios propostos pela organização do evento: tópicos e anáforas nas perguntas (perguntas mais complicadas) e o uso da Wikipedia para procurar e justificar as respostas (mais difícil encontrar as respostas). Adicionalmente investigou-se uma abordagem "Várias Perguntas/Várias Respostas" para RAP.

Nesse sentido desenvolveram-se um módulo para resolver a co-referência nas perguntas (que cria igualmente expressões de pesquisa alternativas para seleccionar os textos relevantes para responder às mesmas) usando a análise sintáctica do PALAVRAS [104], um módulo para procurar textos relevantes na Wikipedia e um módulo para escolha da melhor resposta de entre um conjunto limitado de respostas candidatas.

Escreveu-se um artigo [13], apresentou-se um poster e reformulou-se o primeiro artigo para publicação internacional [96] (tudo no âmbito da participação no QA@CLEF 2007).

2.6 Floresta Sintá(c)tica

Cláudia Freitas e Paulo Rocha

O projeto Floresta Sintá(c)tica foi ativamente retomado em junho de 2007. Com o reinício da actividade, foram realizados dois workshops (Oslo, 22-28 Junho e Coimbra, 27-28 Setembro), nos quais foram tomadas decisões relativas principalmente à parte lingüística, o que deu origem a algumas alterações de formato no Bosque (actualmente na versão 7.5) e a uma nova versão da documentação da Floresta, a Bíblia Florestal [38]. Desde o reinício da actividade, as principais tarefas do projeto foram:

- Actualização da documentação, e passagem ao formato HTML.
- Criação de um mini-guia da Floresta e actualização e tradução de um breve tutorial (visita guiada) sobre o sistema de busca da Floresta.
- Mudança do sistema de busca Águia para Coimbra e sua adaptação às mudanças de formato previamente decididas.
- Revisão das árvores do Bosque e da Floresta Virgem.
- Busca de maior homogeneidade no tratamento de determinados fenómenos lingüísticos (flutuação entre adjectivos e substantivos; numerais; locuções adverbiais; expressões multi-vocabulares; datas; prefixos e elipses).
- Criação de um programa de alinhamento de novas etiquetas no Bosque (i.e., nos ficheiros revistos manualmente), o que possibilita a inserção dessas novas etiquetas, relativas principalmente a estruturas sintácticas mais complexas. Criação de programas para validação dessas novas etiquetas

2.7 PAPEL

Hugo Oliveira e Diana Santos

O PAPEL (Palavras Associadas - Porto Editora / Linguateca) é um projecto que pretende construir, de forma automática ou quase automática, uma ontologia lexical para o português, através de informação extraída do Dicionário da Língua Portuguesa da Porto Editora [103].

Neste projecto, iniciado apenas em Setembro de 2007 com a contratação de Hugo Oliveira, mas que pretendia dar continuação ao trabalho iniciado pela bolsa de doutoramento do Nuno Seco interrompida em Setembro de 2006, foram desenvolvidas as seguintes actividades, em colaboração com o Núcleo de Investigação e Desenvolvimento da Porto Editora:

- Actividades exploratórias da base lexical da Porto Editora (PE), incluindo uma primeira ideia dos n-gramas mais frequentes nas definições.
- Foi implementado um chart parser em Java para analisar as entradas no dicionário da PE disponibilizado em (<http://linguateca.dei.uc.pt/index.php?sep=recursos>).
- Foi desenvolvida uma primeira gramática que pretendia extrair relações de causa efeito entre palavras.
- Foi definido e testado o workflow que liga as três localizações do projecto PAPEL, ou seja, o pólo de Coimbra da Linguateca, o pólo de Oslo da Linguateca, e a Porto Editora, no Porto.
- Foi escrito um primeiro relatório PAPEL [55], após extensa revisão da literatura.

2.8 WPT-05

Luís Costa

A colecção WPT-05 será uma actualização da colecção WPT 03 (colecção da Web portuguesa de 2003) disponibilizada pela Linguateca em 2004 [18].

Além da inclusão de mais documentos (prevê-se que a nova colecção venha a ter oito milhões de documentos distintos, quando a colecção WPT 03 não passava dos 1,6 milhões de documentos), da melhoria e enriquecimento da representação dos metadados associados aos documentos e da inclusão das etiquetas HTML originais, a sua especificação levará também em conta os resultados de um inquérito dirigido aos utilizadores da versão actual da colecção.

Dado que o colaborador responsável pela criação deste recurso apenas iniciou o seu contrato a 22 de Outubro, o trabalho neste projecto encontra-se ainda numa fase bastante inicial. Serão de referir apenas a especificação inicial da estrutura XML para a nova colecção e o respectivo ficheiro de validação XSD e algum estudo, adaptação e optimização do código existente para criar a versão anterior da colecção.

3. Instrução e disseminação

3.1 Formação de recursos humanos

Diana Santos

Em 2007, a seguinte tese de mestrado foi concluída e defendida na Faculdade de Engenharia da Universidade do Porto:

- Luís Miguel Cabral. SUPERB - Sistema Uniformizado de PEquisa de Referências Bibliográficas. Tese de Mestrado. Faculdade de Engenharia da Universidade do Porto. 2007. [12].

As duas teses de doutoramento em progresso sofreram avanços significativos, sendo prevista a sua conclusão para 2008:

- Marcirio Silveira Chaves. Uma metodologia para construção de ontologias e integração de conhecimento geográfico. Universidade de Lisboa.
- Alberto Simões. Extracção e Generalização de Recursos de Tradução com base em Dicionários Probabilísticos. Universidade do Minho.

Além disso, Marcirio Chaves efectuou a prova de qualificação para o doutoramento, exigida pela Universidade, a 22 de Fevereiro [23].

Também em 2007, foi iniciado um doutoramento no âmbito da Linguateca, de Nuno Cardoso (Universidade de Lisboa), sobre reformulação de pesquisas na área de recolha de informação geográfica, e um mestrado em integração de sistemas de RAP e serviços Web semânticos, de Yuzhu Yang (Universidade de Oslo), também no domínio geográfico.

3.2 Actividades de orientação

Diana Santos

Em 2007 tiveram lugar o quarto e o quinto simpósios doutorais da Linguateca, ambos no pólo de Lisboa do XLDB, respectivamente em Março e em Outubro, cujo desenrolar pode ser visto em <http://www.linguateca.pt/documentos/SimposioDoutoralMarco2007.html> e <http://www.linguateca.pt/documentos/SimposioDoutoralOutubro2007.html>.

Além disso, foi dado um curso sobre Avaliação na *European Summer School on Language, Logic and Information* (ESSLI 2007) (Dublin, 6-17 de Agosto de 2007), suportado em conjunto por esta escola, pelo SINTEF e pela Linguateca [67].

3.3 Livro sobre o HAREM

Diana Santos

Em 2007 a Linguateca publicou electronicamente o seguinte livro sobre o Primeiro HAREM [78]:

- Diana Santos & Nuno Cardoso (eds.). *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área. Linguateca. 2007. ISBN: 978-989-20-0731-1. 413 págs, 20 capítulos, 6 anexos.* <http://www.linguateca.pt/LivroHAREM/>,

tendo além disso divulgado o mesmo amplamente em várias listas electrónicas em português e em inglês e noutras publicações.

3.4 “Um passeio pela Floresta Sintá(c)tica”

Cláudia Freitas e Paulo Rocha

Foi realizada uma acção de divulgação da Floresta Sintá(c)tica em Coimbra no dia 28 de Setembro. Destinada principalmente a linguistas e informáticos, o evento, chamado "Um Passeio pela Floresta Sintá(c)tica", contou com 7 palestras que, além da Floresta, apresentaram também a Linguateca, o projecto AC/DC e o Pólo de Coimbra [9, 37, 42, 60, 68, 69, 71].

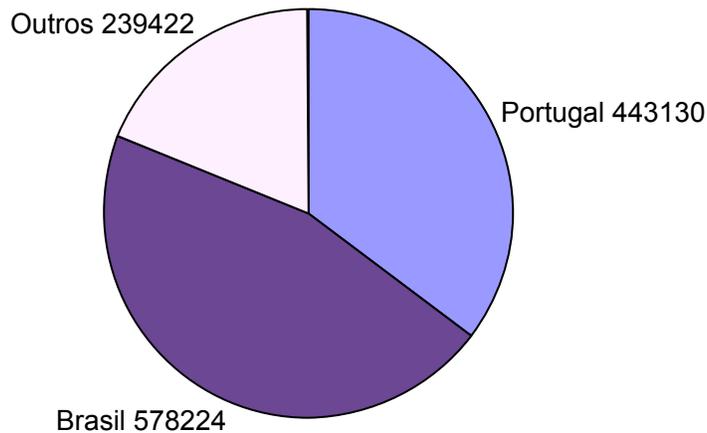
4. Medidas de Impacto (Catálogo e estatísticas)

Luis Cabral

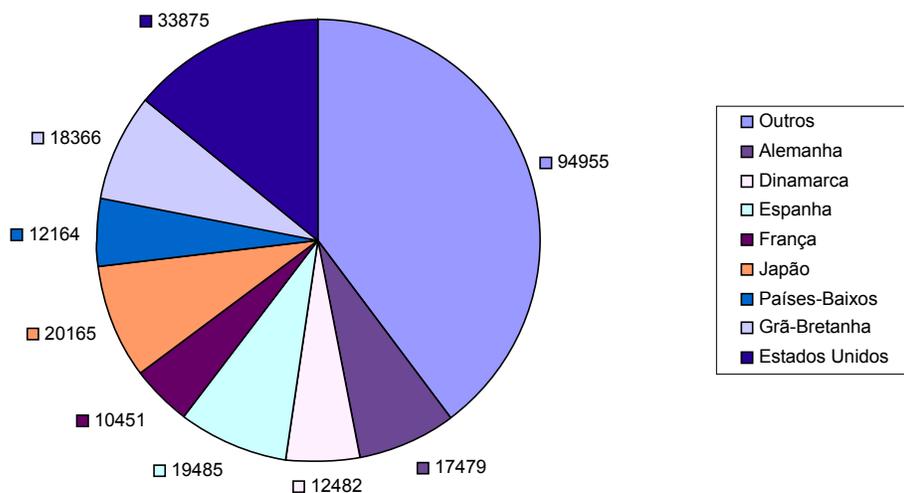
Os valores aqui apresentados referem-se, se mais nada for acrescentado, à data de 1 de Janeiro de 2008. O número de visitas às páginas da Linguateca, de 1 de Janeiro de 2007 a 1 de Janeiro de 2008 foi de 2.576.953 (no total, desde o início do projecto, atinge os 5.523.097).

Para dar uma ideia do impacto internacional do nosso portal, apresentamos nas figuras seguintes: a distribuição do perfil geográfico dos visitantes das nossas páginas desde o início do projecto (os valores referem-se apenas aos acessos para os quais foi possível determinar a sua origem, 29,6% do total de acessos):

Repartição geográfica acumulada dos acessos até 1 de Janeiro de 2008



Repartição geográfica acumulada dos acessos excluindo Portugal e Brasil

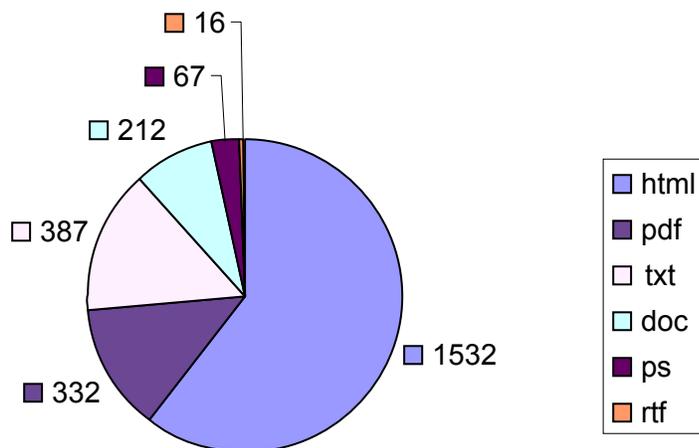


Apresentamos também o número de pedidos dos recursos que exigem registo durante o ano de 2007 (o valor entre parênteses refere-se ao número de pedidos desde o início da disponibilização de cada um dos recursos):

Corpógrafo: 352 (1388)
CETENFolha: 46 (293)
CETEMPúblico: 53 (471)
CHAVE: 21 (57)
WBR 99: 2 (26)
WPT 03: 0 (11)
GEONET-PT-01: 13 (28)

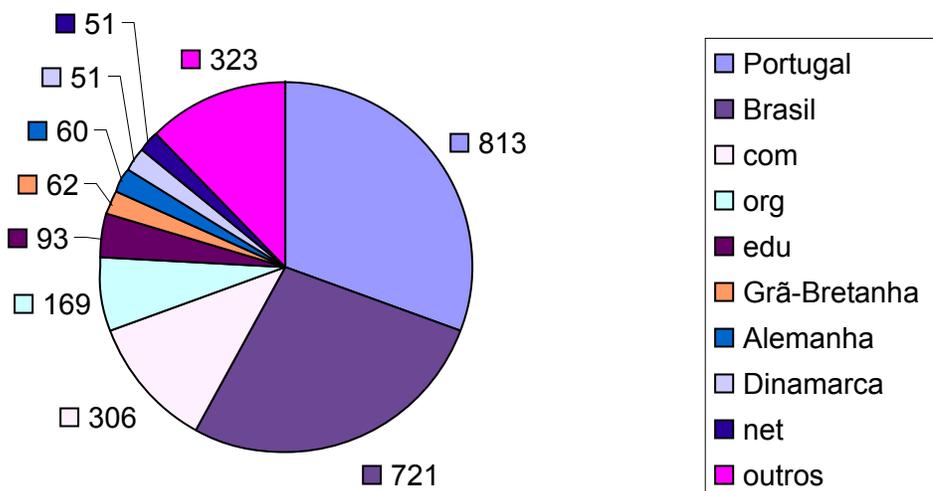
O número actual de documentos acessíveis do sítio da Linguateca é de 2546, constituindo 1480 documentação da própria da Linguateca. A figura seguinte pormenoriza o tipo de documentos disponibilizados.

Número de documentos acessíveis por tipo de ficheiro

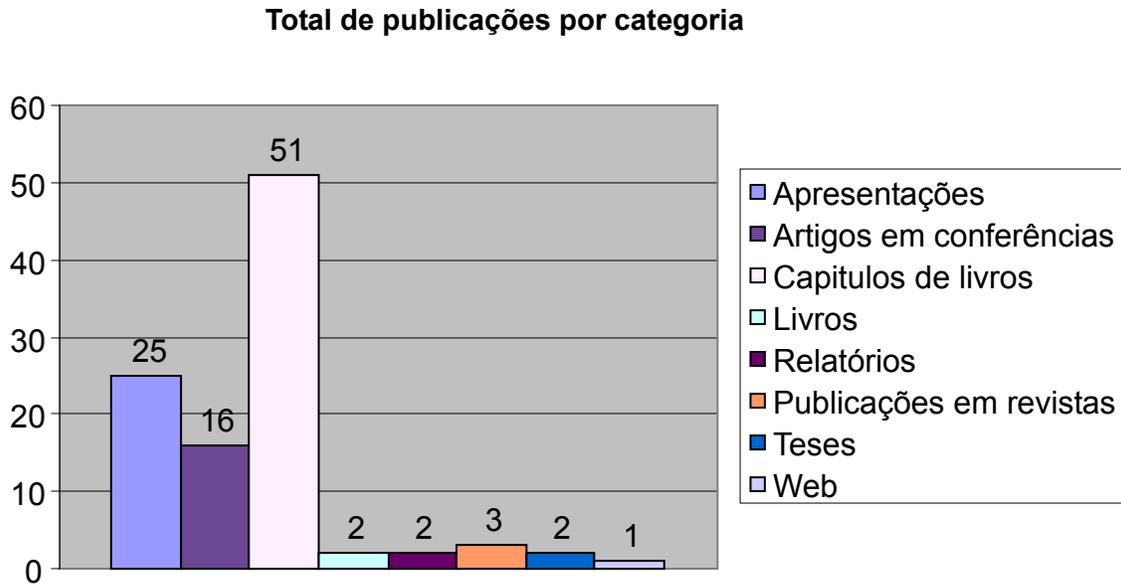


O número de ligações externas incluídas nos nossos catálogos é de 2649, e a distribuição geográfica das localizações para que apontam encontra-se na figura seguinte:

Ligações externas, por localização geográfica



Finalmente, a distribuição das publicações (listadas na secção seguinte) por categoria, seguindo as normas do catálogo de publicações da Linguateca, e por categoria, encontra-se na figura seguinte:



De notar que algumas das publicações, embora apenas tenham sido publicadas em 2007, foram escritas e referem-se a trabalho produzido em anos anteriores, como por exemplo as publicações no livro *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa* [77].

5. Lista de publicações

As publicações da Linguateca no período a que se refere o presente relatório foram listadas com o auxílio do SUPeRB [12], e são apresentadas por ordem alfabética.

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

5.1 Publicações e apresentações com data de 2007

- [1] 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*. Lisboa, Portugal: IST Press, 2007, pp. 247-254. **ÂMBITO**
- [2] José João Dias de Almeida "RENA - Reconhecedor de Entidades". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e*

actas do HAREM, a primeira avaliação conjunta na área. Linguateca:, 2007, pp. 157-172.

- [3] José João Almeida & Alberto Simões. "XML::TMX - Processamento de Memórias de Tradução de Grandes Dimensões". In José Carlos Ramalho, João Correia Lopes & Luís Carriço (eds.), *XML: Aplicações e Tecnologias Associadas (XATA2007)* (FCUL, LISBOA, 15-16 de Fevereiro), Universidade do Minho, pp. 83-93.
- [4] José João Almeida & Alberto Simões "Jspellando nas morfolimpiadas: Sobre a participação do Jspell nas Morfolimpiadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 83-89.
- [5] Christian Nunes Aranha "O Cortex e a sua participação no HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 113-122.
- ÂMBITO**
- [6] Jorge Baptista "Alguns comentários sobre a lista dourada". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 121-131. **ÂMBITO**
- [7] Anabela Barreiro & Susana Afonso "Construção da lista dourada para as primeiras Morfolimpiadas do português". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 107-118.
- [8] Eckhard Bick "Morfologia com sintaxe debaixo de olho: Sobre a participação do PALMORF nas Morfolimpiadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 53-67. **ÂMBITO**
- [9] Eckhard Bick. "Computational Syntax". *Encontro Um passeio pela Floresta Sintá(c)tica* (Coimbra, Portugal, 28 de Setembro de 2007). **ÂMBITO**
- [10] Eckhard Bick "Functional Aspects on Portuguese NER". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 145-155. Este artigo foi publicado pela Springer, na série LNAI, vol. 3960, ISBN-10: 3-540-34045-9
- ÂMBITO**
- [11] 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*. Lisboa, Portugal: IST Press, 2007, pp. 291-300.
- [12] Luís Miguel Cabral. SUPeRB - Sistema Uniformizado de Pesquisa de Referências Bibliográficas. Tese de Mestrado. Faculdade de Engenharia da Universidade do Porto. Março 2007.
- [13] Luís Miguel Cabral, Luís Fernando Costa & Diana Santos. "Esfinge at CLEF 2007: First steps in a multiple question and multiple answer approach". In Alessandro Nardi & Carol Peters (eds.), *Working Notes for the CLEF 2007 Workshop (CLEF 2007)* (Budapest, Hungary, 19-21 September 2007), s/pp.
- [14] Nuno Cardoso. "Diz que é uma espécie de survey sobre Query Expansion". *IV Simpósio Doutoral da Linguateca 2007* (Faculdade de Ciências - Universidade de Lisboa, 20 de Março de 2007).
- [15] Nuno Cardoso. "Novos rumos na expansão de pesquisas para RIG". *V Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 4 de Outubro de 2007).

- [16] Nuno Cardoso. "Geographic IR Challenges". (Oslo, Noruega, 4 de Dezembro de 2007). Apresentação realizada no SINTEF **ÂMBITO**
- [17] Nuno Cardoso, David Cruz, Marcirio Chaves & Mário J. Silva "The University of Lisbon at GeoCLEF 2007". In Alessandro Nardi & Carol Peters (eds.), *Working Notes for the CLEF 2007 Workshop (CLEF 2007)* (Budapest, Hungary, 19-21 September 2007), s/pp.
- [18] 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*. Lisboa, Portugal: IST Press, 2007, pp. 279-288.
- [19] Nuno Cardoso & Diana Santos "Directivas para a identificação e classificação semântica na colecção dourada do HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 211-238. Documento original publicado no sítio do HAREM a 29 de Março de 2006. Republicado como Relatório técnico DI-FCUL TR-06-18 : Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa, Novembro de 2006
- [20] Nuno Cardoso, Diana Santos & Rui Vilela "Directivas para a identificação e classificação morfológica na colecção dourada do HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 239-244. Republicação do Relatório técnico DI-FCUL TR-06-19 : Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa. Novembro de 2006
- [21] Nuno Cardoso, Mário J. Silva & Marília Antunes "Validação Estatística dos resultados do Primeiro HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 59-77.
- [22] 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, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de_Rijke & Maximilian Stempfhuber (eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval - 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, pp. 51-56.
- [23] Marcirio Silveira Chaves. Uma metodologia para construção de ontologias e integração de conhecimento geográfico. Prova de qualificação. Faculdade de Ciências, Universidade de Lisboa. 22 de Fevereiro de 2007.
- [24] Marcirio Chaves. "Sistema de Extração e Integração de Conhecimento Geográfico - SEI-Geo". *IV Simpósio Doutoral da Linguateca 2007* (Faculdade de Ciências - Universidade de Lisboa, 20 de Março de 2007).
- [25] Marcirio Chaves. "SEI-Geo: Primeiros Resultados". *V Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 4 de Outubro de 2007).
- [26] Marcirio Chaves, Catarina Rodrigues & Mário J. Silva. "Data Model for Geographic Ontologies Generation". In José Carlos Ramalho, João Correia Lopes & Luís Carriço (eds.), *XML: Aplicações e Tecnologias Associadas (XATA2007)* (FCUL, LISBOA, 15-16 de Fevereiro), Universidade do Minho, pp. 47-58.
- [27] Luís Costa "Question answering beyond CLEF document collections". In Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de_Rijke

- & Maximilian Stempfhuber (eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval - 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, pp. 405-414.
- [28] 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*. Lisboa, Portugal: IST Press, 2007, pp. 15-33.
- [29] Luís Fernando Costa & Diana Santos. "Question answering systems: a partial answer". (SINTEF, Oslo, 22 de Janeiro de 2007).
- [30] 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.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 137-144. **ÂMBITO**
- [31] Ana Frankenberg-Garcia. "Que referências consultam os estudantes de tradução?". *IV Ibero-American Conference on Translation and Interpretation ((IV CIATI))* (São Paulo, Brazil, 14 - 17 May, 2007).
- [32] Ana Frankenberg-Garcia. "COMPARA". *IV Ibero-American Conference on Translation and Interpretation ((IV CIATI))* (São Paulo, Brazil, 14 - 17 May, 2007). Mini-curso sobre o COMPARA
- [33] Ana Frankenberg-Garcia. "O que é o COMPARA?". (Lisboa, Portugal, Julho de 2007). Apresentação do COMPARA na FCCN
- [34] Ana Frankenberg-Garcia. "COMPARA". (Cascais, Portugal, Outubro de 2007). Apresentação do COMPARA para professores da Escola Secundária Ibn Mucana
- [35] Ana Frankenberg-Garcia. "Building a Parallel Corpus for Translation Research and much more". (Universitat Jaume I, Castellón, Novembro de 2007).
- [36] Ana Frankenberg-Garcia "Lost in Parallel Concordances". In W. Teubert (ed.), *Corpus Linguistics: critical concepts in linguistics*. Londres: Routledge, 2007, pp. 176-190 . Previously published in G. Aston, S. Bernardini and D. Stewart (eds.) *Corpora and language learners*. Amsterdam: John Benjamins, 2004. **SemRec**
- [37] Cláudia Freitas. "Alguns caminhos da Floresta". *Encontro Um passeio pela Floresta Sintá(c)tica* (Coimbra, Portugal, 28 de Setembro de 2007).
- [38] Cláudia Freitas & Susana Afonso. "Bíblia Florestal: Um manual lingüístico da Floresta Sintá(c)tica". Primeira edição: 25 de Outubro de 2007.
- [39] Fredric Gey, Ray Larson, Mark Sanderson, Kerstin Bischoff, Thomas Mandl, Christa Womser-Hacker, Diana Santos, Paulo Rocha, Andres Montoyo, Giorgio M. Di Nunzio & Nicola Ferro. "Challenges to Evaluation of Multilingual Geographic Information Retrieval in GeoCLEF". In *Workshop on Evaluation of Information Access (EVIA) May 15* (Tokyo, Japan, May 15 2007), s/pp.
- [40] 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, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de_Rijke & Maximilian Stempfhuber (eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval - 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, pp. 852-876.

- [41] Danilo Giampiccolo, Pamela Forner, Anselmo Peñas, Christelle Ayache, Dan Cristea, Valentin Jijkoun, Petya Osenova, Paulo Rocha, Bogdan Sacaleanu & Richard Sutcliffe. "Overview of the CLEF 2007 Multilingual Question Answering Track". In Alessandro Nardi & Carol Peters (eds.), *Working Notes for the CLEF 2007 Workshop (CLEF 2007)* (Budapest, Hungary, 19-21 September 2007), s/pp.
- [42] Paulo Gomes. "Apresentação do Pólo de Coimbra". *Encontro Um passeio pela Floresta Sintá(c)tica* (Coimbra, Portugal, 28 de Setembro de 2007).
- [43] 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*. Lisboa, Portugal: IST Press, 2007, pp. 231-245. **ÂMBITO**
- [44] Caroline Hagège "Segmentação e análise morfológica do português com recursos léxicos limitados: Sobre a participação do Smorph nas Morfolimpiadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 35-40. **ÂMBITO**
- [45] Ricardo Ueda Karpischeck "As Morfolimpiadas e a avaliação da verificação ortográfica: Sobre a participação do br.ispell nas Morfolimpiadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 69-73. **ÂMBITO**
- [46] 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, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de_Rijke & Maximilian Stempfhuber (eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval - 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, pp. 223-256.
- [47] 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*. Lisboa, Portugal: IST Press, 2007, pp. 205-216.
- [48] Belinda Maia, Ricardo Silva, Anabela Barreiro & Cecília Fróis. "N-grams in search of theories". In *Proceedings of the 6th International Conference of Practical Applications in Language and Computers (PALC 2007)* (Lodz University, Poland, April 19-22, 2007).
- [49] Thomas Mandl, Fredric Gey, Giorgio Di Nunzio, Nicola Ferro, Ray Larson, Mark Sanderson, Diana Santos, Christa Womser-Hacker & Xing Xie. "GeoCLEF 2007: the CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview". In Alessandro Nardi & Carol Peters (eds.), *Working Notes for the CLEF 2007 Workshop (CLEF 2007)* (Budapest, Hungary, 19-21 September 2007), s/pp.
- [50] 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, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de_Rijke & Maximilian Stempfhuber (eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval - 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, pp. 986-994.

- [51] Ronaldo Teixeira Martins & Maria das Graças Volpe Nunes "Dos processos de individuação e de categorização lexical: sobre a participação do ReGra nas Morfolimpiadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 43-50. **ÂMBITO**
- [52] Bruno Martins, Mário Silva & Marcirio Chaves "O sistema CaGE no HAREM - reconhecimento de entidades geográficas em textos em língua portuguesa". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 97-112.
- [53] 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*. Lisboa, Portugal: IST Press, 2007, pp. 161-176.
- [54] Cristina Mota & Max Silberztein "Em busca da máxima precisão sem almanaques. O Stencil/NooJ no HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca:, 2007, pp. 191-208. **ÂMBITO**
- [55] Hugo Oliveira, Paulo Gomes & Diana Santos. "PAPEL - Trabalho relacionado e relações semânticas em recursos semelhantes". Departamento de Engenharia Informática, FCTUC, CISUC. Dezembro de 2007. Primeiro Relatório PAPEL.
- [56] Viviane Moreira Orengo & Diana Santos "Radicalizadores versus analisadores morfológicos: Sobre a participação do Removedor de Sufixos da Língua Portuguesa nas Morfolimpiadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 91-104.
- [57] Paulo Quesada & 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*. Lisboa, Portugal: IST Press, 2007, pp. 257-265. **ÂMBITO**
- [58] Elisabete Ranchhod & Cristina Mota "Unidades lexicais multipalavra, um osso duro de roer: Sobre a participação do LABEL nas Morfolimpiadas". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 75-80.
- [59] 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*. Lisboa, Portugal: IST Press, 2007, pp. 267-176. **ÂMBITO**
- [60] Paulo Rocha. "Arte dos corpora em português: o projecto AC/DC". *Encontro Um passeio pela Floresta Sintá(c)tica* (Coimbra, Portugal, 28 de Setembro de 2007).
- [61] 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*. Lisboa, Portugal: IST Press, 2007, pp. 143-158.
- [62] Paulo Rocha & Diana Santos "Disponibilizando a <OBRA>Colecção Dourada</OBRA> do <ACONTECIMENTO>HAREM</ACONTECIMENTO> através do projecto <LOCAL|ORGANIZACAO|ABSTRACCAO>AC/DC</LOCAL|ORGANIZACAO|ABSTRACCAO>". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em*

português: Documentação e actas do HAREM, a primeira avaliação conjunta na área.
Linguatca., 2007, pp. 307-326.

- [63] Diana Santos. "A tradução na sociedade do conhecimento OU Tradução: uma tecnologia humana de ponta OU Ciência E Tradução". In *Actas do IX Seminário de Tradução Científica e Técnica em Língua Portuguesa* (Lisboa, Fevereiro de 2007), Lisboa: União Latina.
- [64] Diana Santos "Avaliação conjunta". In Diana Santos (ed.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press, 2007, pp. 1-12.
- [65] Diana Santos. "Avaliação: ensaio de mais algum material para o ESSLLI'07". *IV Simpósio Doutoral da Linguatca 2007* (Faculdade de Ciências - Universidade de Lisboa, 20 de Março de 2007).
- [66] Diana Santos. "A Linguatca, e em especial a Floresta, o PAPEL e o HAREM". (Universidade de Coimbra, 4 de Abril de 2007).
- [67] Diana Santos. "Evaluation in natural language processing". *European Summer School on Language, Logic and Information (ESSLLI 2007)* (Dublin, 6-17 August 2007).
- [68] Diana Santos. "Aplicações de uma floresta sintáctica". *Encontro Um passeio pela Floresta Sintá(c)tica* (Coimbra, Portugal, 28 de Setembro de 2007).
- [69] Diana Santos. "Linguatca: sete anos devotados ao processamento computacional do português". *Encontro Um passeio pela Floresta Sintá(c)tica* (Coimbra, Portugal, 28 de Setembro de 2007).
- [70] Diana Santos "O modelo semântico usado no Primeiro HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguatca., 2007, pp. 43-57.
- [71] Diana Santos, Eckhard Bick & Susana Afonso. "Floresta Sintá(c)tica: apresentação e história do projecto". *Encontro Um passeio pela Floresta Sintá(c)tica* (Coimbra, Portugal, 28 de Setembro de 2007).
- [72] Diana Santos & Nuno Cardoso "Balanço do primeiro HAREM e perspectivas de trabalho futuro". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguatca., 2007, pp. 87-94.
- [73] Diana Santos & Nuno Cardoso "Breve Introdução ao HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguatca., 2007, pp. 1-16.
- [74] Diana Santos, Nuno Cardoso & Nuno Seco "Avaliação no HAREM: Métodos e medidas". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguatca., 2007, pp. 245-282. Republicado como Relatório técnico DI-FCUL TR-06-17 : Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa. Novembro de 2006
- [75] Diana Santos & Luis Costa. "QoIA: fostering collaboration within QA". In Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de_Rijke & Maximilian Stempfhuber (eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval - 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Alicante, Spain, September, 2006. Revised Selected papers* (2007), Berlin / Heidelberg: Springer. Lecture Notes in Computer Science, pp. 569-578.
- [76] 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), pp. 335-374.

- [77] Diana Santos (ed.). *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. Lisboa, Portugal: IST Press. 2007. ISBN: 972-8469-60-8.
- [78] Diana Santos & Nuno Cardoso (eds.). *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca. 2007.
- [79] Luís Sarmento "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*. Lisboa, Portugal: IST Press, 2007, pp. 193-203.
- [80] Luís Sarmento "O SIEMÊS e a sua participação no HAREM e no Mini-HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca., 2007, pp. 173-189.
- [81] Luís Sarmento "A first step to address biography generation as an iterative QA task". In Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de_Rijke & Maximilian Stempfhuber (eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval - 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, pp. 473-482.
- [82] Luís Sarmento, 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*. Lisboa, Portugal: IST Press, 2007, pp. 181-190.
- [83] Nuno Seco "MUC vs HAREM: a contrastive perspective". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca., 2007, pp. 35-41.
- [84] Nuno Seco, Nuno Cardoso, Rui Vilela & Diana Santos "A arquitectura dos programas de avaliação do HAREM". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca., 2007, pp. 283-306.
- [85] Rosário Silva, Susana Inácio & Diana Santos. "Documentação da anotação relativa à cor no COMPARA". Primeira edição: 27 de Novembro de 2007.
- [86] 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*. Lisboa, Portugal: IST Press, 2007, pp. 133-140. **ÂMBITO**
- [87] Alberto Manuel Brandão Simões. "Extracção de Recursos de Tradução com base em Dicionários Probabilísticos". *IV Simpósio Doutoral da Linguateca 2007* (Faculdade de Ciências - Universidade de Lisboa, 20 de Março de 2007).
- [88] Alberto Simões. "Makefile::Parallel - Uma ferramenta para paralelização de processos". *V Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 4 de Outubro de 2007).
- [89] Alberto Simões. "Extracção de Recursos de Tradução com base na Hipótese das Palavras-Marca". *V Simpósio Doutoral da Linguateca 2006* (Faculdade de Ciências - Universidade de Lisboa, 4 de Outubro de 2007).
- [90] Alberto Manuel Simões. "Automatic Extraction of Translation Resources from Parallel Corpora". *Engenharia'07. e as cidades* (Universidade do Minho, Portugal, 22 a 30 de Outubro de 2007).
- [91] Alberto Simões & José João Almeida "Avaliação de alinhadores". In Diana Santos (ed.),

Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa. Lisboa, Portugal: IST Press, 2007, pp. 219-228.

- [92] Alberto Simões & José João Almeida. "Parallel Corpora based Translation Resources Extraction". *Procesamiento del Lenguaje Natural* **39** (2007), pp. 265-272. SEPLN.
- [93] Alberto Simões, Rúben Fonseca & José João Almeida. "Makefile::Parallel Dependency Specification Language". In Anne-Marie Kermarrec, Luc Bougé & Thierry Priol (eds.), *Parallel Processing, 13th International Euro-Par Conference (Euro-Par 2007)* (Berlin/Heidelberg, August 28-31, 2007), Springer. Lecture Notes in Computer Science, pp. 33-41.
- [94] Tamar Solorio "MALINCHE: A NER system for Portuguese that reuses knowledge from Spanish". In Diana Santos & Nuno Cardoso (eds.), *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguatca., 2007, pp. 123-136. **ÂMBITO**
- [95] Rui Vilela. "WebJspell, an Online Morphological Analyser and Spell Checker". *Procesamiento del Lenguaje Natural* **39** (2007), pp. 291-292. SEPLN.

5.2 Publicações no prelo (escritas ou iniciadas em 2007)

Em alguns casos o nome da obra ou dos editores ainda é provisório.

- [96] Luís Miguel Cabral, Luís Fernando Costa & Diana Santos "What happened to Esfinge in 2007?". In Carol Peters, Valentin Jijkoun, Thomas Mandl, Henning Müller, Doug Oard, Anselmo Peñas, Vivien Petras & Diana Santos (eds.), *8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, Revised Selected Papers*. Berlin: Springer, Lecture Notes in Computer Science, 2008.
- [97] Nuno Cardoso, David Cruz, Marcirio Chaves & Mário J. Silva "Using Geographic Signatures as Query and Document Scopes in Geographic IR". In Carol Peters, Valentin Jijkoun, Thomas Mandl, Henning Müller, Doug Oard, Anselmo Peñas, Vivien Petras & Diana Santos (eds.), *8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, Revised Selected Papers*. Berlin: Springer, Lecture Notes in Computer Science, 2008.
- [98] Ana Frankenberg-Garcia. "Compilação e uso de corpora paralelos". In *Linguística Computacional - o Estado da Arte* (2008). trabalhos apresentados na 1^a. Escola Brasileira de Linguística Computacional
- [99] Danilo Giampiccolo, Pamela Forner, Anselmo Peñas, Christelle Ayache, Dan Cristea, Valentin Jijkoun, Petya Osenova, Paulo Rocha, Bogdan Sacaleanu & Richard Sutcliffe. "Overview of the CLEF 2007 Multilingual Question Answering Track". In Carol Peters, Valentin Jijkoun, Thomas Mandl, Henning Müller, Doug Oard, Anselmo Peñas, Vivien Petras & Diana Santos (eds.), *8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, Revised Selected Papers*. Berlin: Springer, Lecture Notes in Computer Science, 2008. (Título do artigo e lista de autores provisórios)
- [100] Susana Inácio, Diana Santos & Rosário Silva. "COMPARando cores em português e inglês". In *Actas do XXIII Encontro nacional da Associação Portuguesa de Linguística (APL 2007)* (Universidade de Évora, 1-3 de Outubro de 2007), APL, 2008.
- [101] Thomas Mandl, Fredric Gey, Giorgio Di Nunzio, Nicola Ferro, Ray Larson, Mark Sanderson, Diana Santos, Christa Womser-Hacker & Xie Xing "GeoCLEF 2007: the CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview". In Carol Peters, Valentin

Jijkoun, Thomas Mandl, Henning Müller, Doug Oard, Anselmo Peñas, Vivien Petras & Diana Santos (eds.), *8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, Revised Selected Papers*. Berlin: Springer, Lecture Notes in Computer Science, 2008.

- [102] Rosário Silva, Susana Inácio & Diana Santos. "COMPARA: contrastive and monolingual colour studies in English and Portuguese". In *Proceedings of the conference of the American Association for Corpus Linguistics (AACL)* (Provo, UT, USA, March 2008).

5.3 Outras fontes citadas no presente relatório

- [103] Dicionário PRO da Língua Portuguesa. Porto Editora, Porto, 2005.
- [104] Eckhard Bick. The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. PhD thesis, Aarhus University, 2000.
- [105] Diana Santos. "Relatório Linguateca 2000-2003". Setembro de 2003.
- [106] Diana Santos. "Resumo da actividade da Linguateca de 15 de Maio de 2003 a 15 de Dezembro de 2006". Linguateca. Dezembro de 2006. Com a colaboração (por ordem alfabética) de Alberto Simões, Ana Frankenberg-Garcia, Belinda Maia, Luís Costa, Luís Miguel Cabral, Luís Sarmiento, Marcirio Chaves, Mário J. Silva, Nuno Cardoso, Paulo Gomes e Rui Vilela

6. Indicadores de realização física

Unidade: em número

A- Publicações

Livros	2
Capítulos de livros	51
Artigos em revistas internacionais	3
Artigos em revistas nacionais	0

B- Comunicações

Em congressos científicos internacionais	12
Em congressos científicos nacionais	4

C- Relatórios 2

D- Organização de seminários e conferências 4

E- Formação Avançada

Teses de Doutorado	3 em curso
Teses de Mestrado	1 terminada, 1 em curso
Outra	
F- Modelos	
G- Aplicações computacionais	12
H- Instalações Piloto	
I- Protótipos laboratoriais	
J- Patentes	
L- Outros (discriminar)	
Recursos	7
Serviços na rede	5

Recursos: CETEMPúblico, CETENFolha, CHAVE, WBR 99, GeoNet-PT-01, Floresta, COMPARA

Aplicações: Corpógrafo, Esfinge, GKB, PLN::Base, Águia, Ambiente de avaliação do HAREM, Alinhador/validador de novas etiquetas do Bosque, DISPARA, SUPeRB, SEI-Geo, Chart parser para analisar entradas no dicionário da PE, DIRECT para português (tradução/localização)

Serviços: AC/DC, COMPARA, Floresta, Busca no Catálogo, SUPeRB