

# Porquê o Páxico? Razões para uma avaliação conjunta

Diana Santos

Linguatca/FCCN & Universidade de Oslo

d.s.m.santos@ilos.uio.no

## Resumo

Este artigo apresenta a motivação da avaliação conjunta Páxico - Português Mágico, organizada pela Linguatca em 2011-2012 como uma medida para (i) incentivar o desenvolvimento de sistemas de ajuda à procura de informação em português; (ii) avaliar a wikipédia em português; (iii) estudar a interação humana na procura de respostas, e compará-la com as características dos sistemas automáticos. Depois de fazer uma pequena descrição da própria tarefa e de iniciativas relacionadas, mencionando também a organização de anteriores avaliações conjuntas pela Linguatca, cada uma das questões acima mencionadas é descrita e problematizada.

## Palavras chave

Avaliação, extração de informação, recolha de informação, resposta a perguntas, português, lusofonia, avaliação conjunta, wikipédia

## 1 Apresentação

Como descrito no editorial do presente volume, o Páxico foi uma avaliação conjunta que decorreu em 2011-2012 sobre a wikipédia em português, preparada pela Linguatca a partir da sua versão de Abril de 2011, usando apenas o texto (e não as páginas completas).

Muito brevemente, e citando o folheto de divulgação do Páxico (Páxico, 2011) criado na altura da sua disseminação, pretendíamos que sistemas (e pessoas) respondessem a perguntas, ou tópicos, com base na wikipédia em português:

Exemplos de perguntas associadas à cultura e sociedade lusófona em que pretendemos uma resposta agregada, e justificada, com base na informação da wikipédia:

- Que outros resistentes associados a movimentos de libertação privaram com Amílcar Cabral durante a vida deste?
- Que cientistas ou avanços da

ciência podem ser direta ou indiretamente relacionados com os jesuítas da escola de Coimbra?

- Que gramáticos brasileiros se pronunciaram sobre a questão da “língua brasileira”?
- Quais os jogadores de futebol de língua portuguesa que passaram por mais de três países estrangeiros na sua vida profissional?

Cada resposta seria uma página da wikipédia, a que estaria associado um conjunto de páginas adicionais que a justificassem, caso a própria página não contivesse informação suficiente para uma pessoa confirmar que era uma resposta válida.

Os sistemas podiam enviar no máximo cem (100) respostas por pergunta, e podiam enviar 3 corridas diferentes.

Para os resultados, a participação e os recursos criados, veja-se os restantes artigos do presente volume, visto que este artigo tem como único objetivo motivar a própria organização do Páxico.

Assim, tentarei explicar porque é que na Linguatca achámos que seria relevante organizar o Páxico nos moldes em que foi organizado, dividindo a argumentação em três partes: avaliação da wikipédia, desenvolvimento de sistemas realistas aplicados a tarefas comuns, e estudo de utilizadores interessados em cultura lusófona.

Embora a secção de motivação do sítio do Páxico liste cinco diferentes razões que pensámos pudessem motivar possíveis participantes e explicar o interesse numa iniciativa como a nossa, nomeadamente

- As limitações dos sistemas atuais
- A falta de interesse pela lusofonia
- O enviesamento da wikipédia em português
- A necessidade de juntar esforços para o ensino e o ensino da cultura
- O concurso homem-máquina

neste artigo propomos uma categorização diferente, também um pouco com base naquilo que foi acontecendo ao longo do Págico, e que aumentou a nossa compreensão do processo que despoletámos.

Para um balanço, veja-se Santos et al. (2012), mas no que aqui interessa podemos desde já adiantar que os dois últimos pontos não surtiram qualquer efeito: estamos convencidos de que nem o Págico foi – pelo menos até agora – aproveitado em experiências pedagógicas, nem as pessoas se sentiram especialmente interessadas em concorrer com sistemas automáticos.

Pelo contrário, parece que são os desenvolvedores de sistemas automáticos que estão especialmente interessados em publicitar os seus sistemas como capazes de ter um desempenho próximo ou melhor do que o humano (como aconteceu com a participação do Watson na Jeopardy (Thompson, 2010; Ringel, 2011)), e que o espírito de competição das pessoas, pelo contrário, não é especialmente posto em destaque quando os seus competidores são máquinas. Uma das razões para isto pode naturalmente ser que, como o desempenho dos sistemas concorrentes demonstrou (Mota, 2012), ainda estamos muito longe de uma situação em que estes consigam ombrear com os seres humanos.

Quanto à questão pedagógica no ensino de português como língua estrangeira, estamos conscientes de que a causa da não utilização do Págico foi o curtíssimo prazo que houve entre a divulgação da iniciativa e a sua execução, e não deve portanto a observação anterior ser interpretada como uma crítica aos professores que poderiam fazer uso do material do Págico. De facto, esperamos que possam vir mais tarde a fazê-lo, não em “tempo real”, mas em tarefas pensadas e apropriadas aos objetivos do ensino de cada um.<sup>1</sup>

## 2 Apreciação da wikipédia

Não há dúvida de que a wikipédia é uma das fontes mais consultadas mundialmente e, embora não tenhamos dados para os utilizadores da rede de língua portuguesa apenas, pensamos que tal se verifica também ao comparar páginas ou sítios em língua portuguesa. De qualquer forma, o simples facto de ser de acesso aberto e modificável e melhorável por qualquer pessoa que a consulte faz com que seja um dos maiores recursos da Web 2.0 em português, e por isso digna de

<sup>1</sup>Agradeço à Cláudia Freitas por chamar a atenção sobre a possibilidade de o meu comentário sobre este assunto ser entendido como uma crítica.

estudo e de utilização, quer no desenvolvimento de sistemas para o português, quer na criação de formas alternativas de consultar e reutilizar o conhecimento nela contido.

Como é sabido, cada vez mais a informática está dirigida a estudos de utilização para melhorar os seus produtos e para compreender a forma de interagir do público: a massificação da sociedade do conhecimento faz-se também através do teste e escolha de vários procedimentos alternativos, muitas vezes tendo o utilizador contribuído, sem saber, para a melhoria significativa do sistema que usa.

Por outro lado, existem várias controvérsias sobre se deixar que qualquer pessoa edite as fontes de conhecimento permite melhorar a qualidade de uma enciclopédia ou, pelo contrário, inquina a transmissão de conhecimento, e vários estudos – de índoles e objetivos diferentes – têm alegado enviesamento da Wikipédia

- em relação ao peso relativo dos assuntos: Veale (2007) notou a importância desproporcionada de informação sobre livros de ficção científica e personagens de certo tipo de literatura fantástica na wikipédia;
- em relação a questões políticas: Hagen (2008) defende que um grupo de pessoas com uma agenda política extremista é capaz de controlar e modificar as informações na wikipédia sem que a maioria da população se dê conta.

São, aliás, bem conhecidas as variadas discussões que de vez em quando assolam a wikipédia em relação a artigos em que diferentes autores têm opiniões radicalmente diferentes, e que os editores têm de congelar.

Contudo, e pelo menos que eu saiba, não tem havido muitas avaliações da wikipédia em português no que se refere ao seu conteúdo e abrangência, nem à sua capacidade de satisfazer os utilizadores que nela procuram<sup>2</sup>. Existe sim uma prática extremamente positiva e interessante dos wikipedistas brasileiros que têm organizado concursos de melhoria da wikipédia no que se refere ao material em português, como é disso prova o presente I GP Wikimedia Brasil.<sup>3</sup> De qualquer maneira, quando iniciámos

<sup>2</sup>Claro está que a noção de satisfação de um utilizador é complicada e não passa necessariamente pela garantia de qualidade – uma pessoa pode ficar muito satisfeita por ter aprendido muitas coisas que não sabia ... para mais tarde descobrir que não eram verdade – ou, pior ainda, nunca vir a descobri-lo.

<sup>3</sup>Em <http://pt.wikipedia.org/wiki/Wikip%C3%A9dia:GP>.

o Páxico não estávamos conscientes desta iniciativa, e pensamos que o Páxico poderia dar um contributo para o trabalho de avaliação. Pelo menos posso dar voz à satisfação de ver que no Brasil a língua portuguesa ainda é o veículo preferido de comunicação e conhecimento, como aliás foi-nos comentado pela Belinda Maia a propósito das suas aulas de terminologia na Universidade do Porto: não há dificuldade em encontrar texto técnico e científico em português do Brasil, enquanto que em Portugal os textos técnicos são geralmente escritos (ou pelo menos publicados) em inglês.<sup>4</sup> Ainda reforçando a convicção de que o futuro está no Brasil, veja-se as experiências pedagógicas universitárias de melhoria da wikipédia em português relatadas por Neto (2012).

Depois de usarmos instantâneos (progressivamente maiores) da wikipédia em português em várias avaliações conjuntas internacionais, como o QA@CLEF (Giampiccolo et al., 2008) para responder a perguntas em português (ou cuja resposta estivesse em português), e o GikiP (Santos et al., 2009) e o GikiCLEF (Santos et al., 2010) para responder a um tipo especial e mais complexo de perguntas cuja resposta também se podia encontrar na wikipédia em português, pensamos ter chegado a altura de dedicar uma avaliação conjunta apenas a responder a perguntas em português cuja resposta estivesse em português.

Isto não só porque a Linguateca se dedica ao processamento computacional do português, mas porque a parcela de atenção e de cobertura do português por oposição às outras línguas resultou muito diminuta nessas avaliações anteriores, como discutimos em Santos e Cabral (2009).

### 3 Desenvolvimento de sistemas realistas

Uma das principais razões para a organização de avaliações conjuntas pela Linguateca (Santos, 2007) é a nossa convicção de que essa organização leva a comunidade a desenvolver sistemas e a resolver problemas práticos que resultam no avanço da área como um todo, e que levam ao eventual surgimento de sistemas com aplicação prática.

Embora seja claro que, pelo menos desta vez,

<sup>4</sup>Realce-se a este propósito, embora não vindo diretamente à baila, que a Linguateca disponibilizou há tempos memórias de tradução português-inglês em áreas especializadas, tal como engenharia industrial, arquitetura, gestão, es estudos literários, e que podem ser úteis para mostrar as diferenças entre as línguas nas áreas respetivas, cf. <http://www.linguateca.pt/Repositorio/>.

tal organização não levou a esse efeito desejado, com apenas dois participantes automáticos que não parecem ter aplicado muito esforço à tarefa oferecida pelo Páxico, usando-o apenas tangencialmente como verificação ou teste de partes do seu trabalho, a motivação desta tarefa era clara nesse sentido.<sup>5</sup>

Ainda existe muito pouco apoio para respostas agregadas, exceto no caso da distribuição geográfica, onde os mapas são populares e usados em muitas interfaces. Além disso, encontram-se também em alguns sistemas os cronogramas ou linhas temporais (Heyer, Holz e Teresniak, 2009; Heyer et al., 2011), e, naturalmente, os grafos, que geralmente ligam objetos semelhantes (Dorow, 2006; Widdows, 2004). Mas os grafos, para que a sua visualização seja humanamente viável, pressupõem tentativas de diminuir o número de dimensões, veja-se por exemplo Speer, Havasi e Lieberman (2008).

Contudo, a obtenção de respostas variadas e múltiplas a uma mesma necessidade de informação (o que tecnicamente se pode chamar respostas abertas a um assunto sobre o qual não conhecemos de antemão as respostas) é uma tarefa muito mais complicada, não só de avaliar: quem pergunta, aprende, e essa aprendizagem depende muito do quanto já sabia, o que torna a utilidade de um tal sistema muito diferente conforme o perguntador (Freitas et al., 2012). O que é, aliás, algo já bem conhecido e problematizado na área de recolha de informação, veja-se por exemplo a discussão de Saracevic (1995). Além disso, não é fácil comparar medidas de desempenho de sistemas com objetivos diferentes, veja-se Su (1998).

Durante a organização do QA@CLEF, esse foi um assunto muito discutido, mas devido à inércia provocada por um grande número de organizadores, foi apenas atacado em avaliações conjuntas mais pequenas e, infelizmente, com muito menos participação, nomeadamente o Wika (Jijkoun e de Rijke, 2007) e o GikiP (Santos et al., 2009).

Outra linha de discussão, embora ainda não muito seguida, tem a ver com a variedade e diversidade na apresentação de resultados — veja-se as tentativas nesse campo, principalmente em avaliações de imagens (Karlgrén, Clough e Gonzalo, 2006) ou de cariz geográfico (Bucher et al., 2005).

Pensámos, de qualquer maneira, que a tarefa do Páxico poderia levar à mentalização da

<sup>5</sup>Mais uma vez, isto não é para ser lido como crítica, dado o pouco tempo entre a divulgação e a participação: é apenas uma constatação.

comunidade informática para a necessidade de desenvolver sistemas que dessem alguma resposta a este tipo de problemas.

#### 4 Estudo de utilizadores interessados em cultura lusófona

Outra área em que precisamos absolutamente de compensar a falta do peso da cultura lusófona na investigação em recolha de informação a nível mundial é a da investigação de assuntos relacionados com a nossa cultura e língua.

De facto, o português, sendo uma das línguas mais faladas no mundo, tem um peso comercial e cultural comparativamente muito reduzido, ou pelo menos anda arredado da atenção de muitos atores no campo da recolha de informação e/ou do processamento computacional da língua. Senão atente-se aos seguintes indicadores

- Muito raramente o português é uma língua de investigação internacional – contam-se pelos dedos as avaliações envolvendo o português, numa altura em que não é exagero dizer que existe uma avaliação (ou “shared task”) semana sim semana não;
- A maior parte das grandes iniciativas do Google não contemplam o português (por exemplo, o Google books ngram viewer<sup>6</sup>, além de inglês e chinês, só contempla, o francês, o alemão, o espanhol, o russo – e o hebraico!)

A própria União Latina, cujo objetivo é dinamizar e estudar as línguas românicas, não dá muito peso ao português, e o barómetro que propõe<sup>7</sup> confere apenas a ordem dezasseis ao português. Mas, considerando com mais atenção a forma como os valores foram calculados, apercebemo-nos de que o único elemento original é a fórmula: Todos os outros indicadores provêm de organizações diferentes.<sup>8</sup>

Embora o Observatório da Língua Portuguesa<sup>9</sup> nos últimos tempos tenha ganho um dinamismo apreciável (ao contrário do relatado

<sup>6</sup><http://books.google.com/ngrams>

<sup>7</sup>Acessível de <http://www.portalingua.info/fr/poids-des-langues/>

<sup>8</sup>Por exemplo, os indicadores relativos à Internet têm origem numa companhia, provavelmente americana, que à data da escrita (e consequente consulta do seu sítio, no dia 2 de janeiro de 2012), tem em lugar proeminente o anúncio “Date sexy African women”, o que não abona em favor da seriedade da mesma – embora não sejam necessariamente responsáveis pelo conteúdo da publicidade, o facto de terem publicidade, na minha opinião, reduz a impressão geral de confiabilidade.

<sup>9</sup><http://www.observatorio-lp.sapo.pt/>

em Santos (2009)), pode constatar-se que os dados que apresenta provêm... exatamente da mesma fonte.

Voltando ao barómetro da União Latina, é muito estranho que seja o francês a língua mais cotada (entre as línguas românicas). Olhando com mais atenção, outra das bases para calcular o peso das línguas é algo chamado “Index translatorum”<sup>10</sup>, que – mais uma vez na mesma data – continha a seguinte observação relativa a Portugal: “1989, 1990, 1991, 1992, 1997, 1998, 1999, 2000, 2001, 2002, 2005, 2006, 2007, 2008 and 2009 currently being processed by the INDEX team”. Ou, por outras palavras: os dados ainda não se encontram lá.

Quanto ao estudo relacionado com a cultura (outro estudo que podemos consultar do sítio da União Latina, denominado “Línguas e culturas na Web: Estudo 2007”), e que data (pelo título) de há cinco anos (União Latina – Direção de Terminologia e Indústrias da Língua (DTIL), s.d.), as imprecisões e erros são de tal forma gritantes que nos convencem da pouca fiabilidade em relação ao português: desde “amalia rodriguez” a “antónio cavaco silva”, passando por “Otel de Carvalho”, “sofia de mello bryner” e “Luis de Camões”, até ao facto de que nas “Letras” o valor de Maio de 2008, 3.752.201 é vinte vezes menor do que o de julho de 2005, 65.323.792<sup>11</sup>, ao contrário de todas as outras categorias, tudo nos leva a duvidar de que possamos confiar nos resultados.

Seja como for, em Prado, Pimenta e Álvaro Blanco (2009), os autores discutem os variados problemas que tiveram com as mudanças no funcionamento e existência dos motores de procura, e sugerem que uma das grandes vantagens do seu trabalho é ter usado o mesmo método ao longo de mais de uma década, podendo portanto os números serem usados para estudar a evolução da presença das línguas na rede.

Não querendo criticar estas duas instituições – o observatório da Língua Portuguesa e a União Latina – que, pelo menos, lutam contra a corrente, achamos contudo que estes números demonstram bem o quanto ainda é preciso fazer para levantar o português ao nível que merece.

De qualquer maneira, gostávamos de salientar o trabalho de Calvet (Calvet, 2006; Calvet, 2008), que está na origem do dito barómetro, como merecedor de reflexão e do maior respeito,

<sup>10</sup><http://www.unesco.org/xtrans/bscontrib.aspx>

<sup>11</sup>Os valores referidos foram lidos em 3 de janeiro de 2012 dos seguinte endereço: [http://dtil.unilat.org/LI/2007/pt/cultura\\_letras\\_pt.htm](http://dtil.unilat.org/LI/2007/pt/cultura_letras_pt.htm).

e encorajar os leitores a tomá-lo em conta.

Já em Aires e Santos (2002) tentámos estimar o tamanho da rede em português, mas nessa altura não estávamos interessadas em comparações com outras línguas, e por isso os nossos resultados não podem ser invocados como dados relativos.

De qualquer maneira, não só os estudos lusófonos não são muito difundidos, como as principais fontes quantitativas sobre os mesmos não nos inspiram suficiente confiança, o que demonstra a urgência de neles insistirmos e a eles dedicarmos os nossos esforços.

## 5 Estudos de utilizadores

---

É bem conhecida a dificuldade de avaliar, em termos de utilizadores, as vantagens e desvantagens de um sistema, sobretudo quando se trata de um sistema interativo.

O nosso caso, de tentar avaliar um recurso com o qual se interage, ainda é mais complexo, por duas razões:

- Para garantir a comparabilidade, desenvolvemos um método diferente (e pior) de interagir com a Wikipédia, que não só impedia o utilizador de ter acesso a imagens e a tabelas, como limitava a sua navegação para escolher e marcar justificações, o que torna o estudo da interação fundamentalmente diferente da verdadeira interação com a wikipédia.
- Por outro lado, as condições de participação também não eram naturais: é pouco provável, pelo menos, encontrar utilizadores realmente interessados em todos os tópicos que oferecemos, o que é diferente de estudos feitos com utilizadores naturais e interessados numa tarefa comum.

Assim sendo, a tarefa do Páxico torna-se artificial por duas razões...

De qualquer maneira, o trabalho de Su (Su, 1998) mostrou que uma avaliação que entre em conta com a impressão geral de uma tarefa (e de um conjunto de resultados) é mais indicada do que outra que apenas classifique e avalie um a um. Por isso, o facto de procurarmos informação sobre um tópico e não sobre respostas individuais pode ser meritório, sobretudo se conseguirmos quantificar medidas com base no todo e não apenas em resultados individuais.

Não podemos contudo deixar de afirmar que o Páxico é uma gota de água no oceano, e que veio demonstrar, se precisássemos de demonstração, que a usabilidade e o desenho de sistemas que

de facto ajudam as pessoas em tarefas reais é algo que requer uma atenção específica e um trabalho aturado, e que estamos muito longe de ter conseguido definir um ambiente em que, fora de uma competição específica e experimental, pudesse atrair de facto pessoas para a utilizar.

Contudo, uma das motivações do Páxico – sobre a qual falaremos mais no artigo que faz o seu balanço, Santos et al. (2012) – era sair para a realidade dos falantes de português e não nos confinarmos sempre e apenas à comunidade estreita de desenvolvedores de sistemas de PLN ou RI.

Como motivação, é importante reafirmá-la e refletir em como a implementar, quiçá, de maneira diferente.

## 6 Construção de recursos e de competência

---

Ao embarcarmos no Páxico, alimentava-nos pelo menos uma certeza: o trabalho que realizaríamos iria ser benéfico à comunidade porque criaríamos recursos públicos e fomentariamos a discussão e a reflexão numa área ou conjunto de áreas que nos pareciam – e parecem – importantes.

Esse é um dos traços distintivos da Linguateca (Santos, 2009), e que aproveito para repetir aqui: a construção de recursos públicos e a dinamização de áreas de investigação e de desenvolvimento no processamento e uso da língua portuguesa.

## 7 Comentários finais

---

O Páxico distinguiu-se das avaliações conjuntas anteriores da Linguateca principalmente por duas características: quase não teve colaboração dos participantes, ou seja, pese embora o adjetivo “conjunto”, foi totalmente decidido pela equipa, com base no raciocínio descrito no presente artigo, e mais globalmente, na presente edição da Linguamática; e dedicou-se a uma tarefa que tentámos que fosse muito mais próxima do dia-a-dia de uma população falante do português, em vez de uma tarefa técnica, automática, de apoio. Com esta decisão tentámos afastar-nos do efeito protótipo e também abordar os problemas da interação pessoa-máquina.

Continuamos a achar que um sistema que permitisse encontrar uma série de respostas a uma dada necessidade de informação na wikipédia podia servir como um auxiliar poderoso em várias profissões relacionadas com o conhecimento, tal como jornalista, escritor de temas de divulgação,

e mesmo estudante de uma dada área ou assunto. Parece-nos também que um tal sistema podia ser uma peça fulcral numa tarefa maior, como a que outros investigadores se têm atrevido, nomeadamente a criação automática de novas páginas da wikipédia (Sauper e Barzilay, 2009; Balasubramanian e Cucerzan, 2009), ou de outras formas de visualização da mesma.

Por outro lado, esperamos que, como efeitos laterais desta iniciativa, possamos obter mais conhecimento sobre o problema e sobre a cultura em português que possam servir para aumentar os recursos que a Linguateca põe à disposição da comunidade em geral, tanto a do processamento computacional da língua portuguesa, como a mais geral dos interessados na cultura lusófona.

## Agradecimentos

O trabalho aqui descrito enquadra-se no âmbito da Linguateca, co-financiada desde o seu início pelo Governo Português, pela União Europeia (FEDER e FSE), sob o contrato POSC/339/1.3/C/NAC, pela UMIC e pela FCCN, e em 2011, pela Fundação da Ciência e da Tecnologia (FCT) e pela Fundação para a Computação Científica Nacional (FCCN).

O Págico, e em particular o meu trabalho nesta avaliação conjunta, foi também financiado pela Universidade de Oslo.

Agradeço aos organizadores do PROPOR o convite feito para integrar o Págico neste contexto, ao Fernando Perdigão especialmente, pelo empenho e entusiasmo em relação ao mesmo, e a toda a organização do Págico. Estou também especialmente grata à Belinda Maia e à Cláudia Freitas pelos comentários pertinentes sobre versões preliminares deste artigo.

## Referências

- Aires, Rachel e Diana Santos. 2002. Measuring the web in portuguese. Em Brian Matthews, Bob Hopgood, e Michael Wilson, editores, *Euroweb 2002 conference*, pp. 198–9, 17-18 de Dezembro, 2002.
- Balasubramanian, Niranjan e Silviu Cucerzan. 2009. Automatic generation of topic pages using query-based aspect models. Em *CIKM'09*, pp. 2049–52, 2-6 de Novembro, 2009.
- Bucher, Bénédicte, Paul Clough, Hideo Joho, Ross Purves, e A. K. Syed. 2005. Geographic IR Systems: Requirements and Evaluation. Em *Proceedings of the 22nd International Cartographic Conference ICC 2005*, 11-16 de Julho, 2005.
- Calvet, Louis-Jean. 2006. *Towards an ecology of world languages*. Polity, Londres.
- Calvet, Louis-Jean. 2008. Le 'poids' des langues: une présentation de la situation linguistique du monde à l'heure de la mondialisation. Colloque de l'ACFAS 2008, Association des universités de la francophonie canadienne, <http://www.francophoniecanadienne.ca/DATA/ANNONCE/159.pdf>.
- Dorow, Beate. 2006. *A Graph Model for Words and their Meanings*. Tese de doutoramento, Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart.
- Freitas, Cláudia, Paulo Rocha, Cristina Mota, Luís Costa, e Diana Santos. 2012. O que é uma resposta? Notas de uns avaliadores estafados. *Linguamática*, 4(1), Abril, 2012. Neste volume.
- Giampiccolo, Danilo, Pamela Forner, Anselmo Peñas, Christelle Ayache, Dan Cristea, Valentin Jijkoun, Petya Osenova, Paulo Rocha, Bogdan Sacaleanu, e Richard Sutcliffe. 2008. Overview of the CLEF 2007 Multilingual Question Answering Track. Em Carol Peters, Valentin Jijkoun, Thomas Mandl, Henning Müller, Doug W. Oard, Anselmo Peñas, Vivien Petras, e Diana Santos, editores, *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Revised Selected Papers*, volume 5152, pp. 200–236, Berlim, Setembro, 2008. Springer.
- Hagen, Arnulf. 2008. Battlefield wikipedia. Em *Dagbladet.no*. 21 de Outubro, <http://www.dagbladet.no/2011/10/21/kultur/debatt/kronikk/wikipedia/manipulasjon/18716526/>.
- Heyer, Gerhard, Florian Holz, e Sven Teresniak. 2009. Change of topics over time and tracking topics by their change of meaning. Em Ana L. N. Fred, editor, *KDIR 2009: Proc. of Int. Conf. on Knowledge Discovery and Information Retrieval*. INSTICC Press, Outubro, 2009.
- Heyer, Gerhard, Daniel Keim, Sven Teresniak, e Daniela Oelke. 2011. Interaktive explorative Suche in großen Dokumentbeständen. *Datenbank-Spektrum*, 11(3):195–206, Outubro, 2011. 10.1007/s13222-011-0072-4.

- Jijkoun, Valentin e Maarten de Rijke. 2007. WiQA: Evaluating Multi-lingual Focused Access to Wikipedia. Em *The First International Workshop on Evaluating Information Access (EVIA)*, 15 de Maio, 2007.
- Karlgren, Jussi, Paul Clough, e Julio Gonzalo. 2006. Multilingual Interactive Experiments with Flickr. *ERCIM News*.
- Mota, Cristina. 2012. Resultados páxicos: participação, medidas e pontuação. *Linguamática*, 4(1), Abril, 2012. Neste volume.
- Neto, Lauro. 2012. Wikipédia pede ajuda aos universitários. Em *O Globo Educação*, 4 de março, 2012. <http://oglobo.globo.com/educacao/wikipedia-pede-ajuda-aos-universitarios-4201782>.
- Prado, Daniel, Daniel Pimienta, e Álvaro Blanco. 2009. *Douze années de mesure de la diversité linguistique sur Internet : bilan et perspectives*. Unesco.
2011. Páxico: português mágico. Folheto de divulgação do Páxico, <http://www.linguateca.pt/Pagico/Pagico.pdf>.
- Ringel, Martin. 2011. IBM Watson and Jeopardy! Apresentação na conferência NoTur 2011, Oslo. [http://www.notur.no/notur2011/material/Martin\\_Watson\\_Notur\\_handout.pdf](http://www.notur.no/notur2011/material/Martin_Watson_Notur_handout.pdf).
- Santos, Diana. 2007. Avaliação conjunta. Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, Lisboa, Portugal, capítulo 1, pp. 1–12.
- Santos, Diana. 2009. Caminhos percorridos no mapa da portuguesificação: A Linguateca em perspectiva. *Linguamática*, 1(1):25–59, Maio, 2009.
- Santos, Diana e Luís Miguel Cabral. 2009. Summing GikiCLEF up: expectations and lessons learned. Em Carol Peters, Giorgio Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peñas, Giovanna Roda, Francesca Borri, Alessandro Nardi, e Carol Peters, editores, *Multilingual Information Access Evaluation, Vol. I: Text Retrieval Experiments*, volume Vol. I: Text Retrieval Experiments, pp. 212–222, Berlim / Heidelberg. Springer.
- Santos, Diana, Luís Miguel Cabral, Corina Forascu, Pamela Forner, Fredric Gey, Katrin Lamm, Thomas Mandl, Petya Osenova, Anselmo Peñas, Álvaro Rodrigo, Julia Schulz, Yvonne Skalban, e Erik Tjong Kim Sang. 2010. GikiCLEF: Crosscultural Issues in Multilingual Information Access. Em Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, e Daniel Tapias, editores, *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, Maio, 2010. European Language Resources Association (ELRA).
- Santos, Diana, Nuno Cardoso, Paula Carvalho, Iustin Dornescu, Sven Hartrumpf, Johannes Leveling, e Yvonne Skalban. 2009. GikiP at GeoCLEF 2008: Joining GIR and QA forces for querying Wikipedia. Em Carol Peters, Tomas Deselaers, Nicola Ferro, Julio Gonzalo, Gareth J.F.Jones, Mikko Kurimo, Thomas Mandl, Anselmo Peñas, e Viviane Petras, editores, *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers*. Springer, pp. 894–905.
- Santos, Diana, Cristina Mota, Alberto Simões, Luís Costa, e Cláudia Freitas. 2012. Balanço do Páxico e perspetivas de futuro. *Linguamática*, 4(1), Abril, 2012. Neste volume.
- Saracevic, T. 1995. Evaluation of evaluation in information retrieval. Em *Proceedings of the 17th Annual International ACM SIGIR'95, Conference on Research and Development in Information Retrieval*, pp. 138–146, Seattle, WA, EUA.
- Sauper, Christina e Regina Barzilay. 2009. Automatically Generating Wikipedia Articles: A Structure-Aware Approach. Em *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pp. 208–216, 2-7 de Agosto, 2009.
- Speer, Robert, Catherine Havasi, e Henry Lieberman. 2008. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge. Em *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008*, pp. 548–553, 13-17 de Julho, 2008.
- Su, Louise T. 1998. Value of search results as a whole as the best measure of information retrieval performance. *Information Processing and Management*, 34:557–579.
- Thompson, Clive. 2010. Smarter Than You Think: What Is I.B.M.'s Watson? Em

New York Times, 20 de Junho, 2010.  
[http://www.nytimes.com/2010/06/20/magazine/20Computer-t.html?\\_r=1&ref=homepage&src=me&pagewanted=all](http://www.nytimes.com/2010/06/20/magazine/20Computer-t.html?_r=1&ref=homepage&src=me&pagewanted=all).

União Latina – Direção de Terminologia e Indústrias da Língua (DTIL). s.d. Línguas e culturas na web: Estudo 2007. Relatório técnico, União Latina. [http://dtil.unilat.org/LI/2007/index\\_pt.htm](http://dtil.unilat.org/LI/2007/index_pt.htm).

Veale, Tony. 2007. Enriched Lexical Ontologies: Adding new knowledge and new scope to old linguistic resources. Curso na ESSLI 2007, Dublin, Irlanda.

Widdows, Dominic. 2004. *Geometry and meaning*. CSLI Publications.