

Balanço do Págico e perspetivas de futuro

Diana Santos
Linguatca/FCCN &
Universidade de Oslo
d.s.m.santos@ilos.uio.no

Cristina Mota
Linguatca/FCCN
cmota@ist.utl.pt

Alberto Simões
Instituto de Letras e Ciências Humanas
Universidade do Minho
ambs@ilch.uminho.pt

Luís Costa
Linguatca/FCCN
luis.f.kosta@gmail.com

Cláudia Freitas
Linguatca/FCCN & PUC-Rio
maclaudia.freitas@gmail.com

Resumo

Uma avaliação só está concluída quando se faz um balanço e se tiram ilações para o futuro. Neste artigo discutimos o que foi obtido com o Págico, o que podia ter sido melhor, e propomos o que nos parece interessante de realizar numa próxima edição. Tentámos estruturar o balanço segundo as vertentes descritas na motivação do Págico, resumindo e opinando sobre outros artigos nesta edição, e depois fazendo uma apreciação crítica da participação e dos resultados. Descrevemos ainda formas de utilizar, numa fase pós-Págico, o extenso trabalho desenvolvido, quer na avaliação de novos sistemas, quer no ensino e/ou divulgação de temas de cultura lusófona por exemplo no estrangeiro. Em relação à área de usabilidade, deixamos em aberto o desafio de como melhorar significativamente a interface do SIGA de forma a poder ser usada em casos reais.

Palavras chave

Recolha de informação, Resposta a perguntas, Avaliação, Cooperação pessoa-máquina, Wikipédia, Usabilidade

1 Apresentação

Como primeiro acontecimento dedicado a avaliar e estudar a procura de informação na wikipédia em português, o Págico terá merecido o seu lugar na história, mas a participação ficou muito aquém das expetativas, sobretudo se tivermos em conta que se dedicava a todo o mundo lusófono: participaram três grupos de investigação que fazem investigação em recolha de informação (RI) e resposta a perguntas (RAP), mas um deles sem sistema e como equipa humana, e ainda outras quatro equipas humanas, das quais apenas uma respondeu a todos os tópicos e as restantes não responderam a mais de um terço dos tópicos

cada uma.

De facto, a participação é comparável ao que por exemplo é relatado na tese de Rachel Aires Aires (2005), em que seis utilizadores criaram corpora personalizados para o estudo em questão, ou, em geral, estudos feitos em contextos de doutoramento, e muito inferior, a nível de participação de grupos de investigação, ao que foi o caso nas anteriores avaliações conjuntas organizadas pela Linguatca, como as Morfolimpíadas (Costa, Rocha e Santos, 2007), com sete sistemas participantes, ou o HAREM (Santos e Cardoso, 2007; Mota e Santos, 2008), em que participaram nove sistemas.

Podemos tentar entender esse decréscimo de duas maneiras, não necessariamente excludentes. Por um lado, temos consciência de que a tarefa proposta no Págico é muito mais complexa que a das avaliações conjuntas anteriores organizadas pela Linguatca, o que pode ter assustado alguns. A nossa intenção foi realmente propor um desafio que fosse o mais próximo possível dos “desafios do mundo real,” em oposição às tarefas da Morfolimpíadas ou do HAREM (veja-se a secção 8 sobre a comparação com avaliações conjuntas anteriores). Por outro lado, podemos também supor que, aparentemente, as áreas de recolha de informação e de resposta a perguntas em Portugal e no Brasil não consideraram o desafio válido ou interessante, nem as áreas de estudos de usabilidade ou de estudos de utilizadores se interessaram pela nossa iniciativa. A área de PLN, por outro lado, achou provavelmente esta uma tarefa difícil demais, aliás como aconteceu no GikiCLEF e no GikiP, onde a participação de língua portuguesa foi mínima. Isto pode significar ainda que estas disciplinas não estão focadas no português, ou que talvez devêssemos ter uma base de organizadores muito maior para que cada grupo desafiasse os seus membros, e para que dessem

o aval científico nas áreas respetivas (RI, RAP e usabilidade).

Entre outras coisas, este artigo pretende de certa forma contrariar essa segunda visão, e tentar que, à posteriori, o trabalho feito possa de facto ser usado nessas áreas e por esses investigadores (veja-se a secção 7). Mas antes disso queremos olhar para o que foi feito e ver o que aprendemos e o que podemos apresentar para que outros possam aprender.

2 A wikipédia melhorou?

Certamente que não! Ou seja, por muito que tenhamos estudado e interagido com esta enciclopédia comum, de forma alguma isso teve impacto na realidade.¹ Mas é nossa esperança que, ao distribuímos e iniciarmos um estudo quantitativo de várias questões, veja-se (Simões, Costa e Mota, 2012), possamos entusiasmar outros a fazerem algo semelhante ou muito mais interessante.

Outra das coisas que será possível fazer, se houver interesse, é medir, de acordo com critérios semelhantes ou pelo menos comparáveis, versões futuras da Wikipédia, lançando, quem sabe, um wikiavaliómetro, à semelhança do barómetro das línguas românicas da União Latina, referido por Santos (2012), mas referindo-se ao português e à wikipédia em português mais especificamente.

3 A comparação pessoa-máquina

O Págico acabou por não dar um contributo especial nesta matéria—embora tenhamos progredido no desenvolvimento de ferramentas que podem levar a esse objetivo, ao melhorar o SIGA e equipá-lo com capacidades de reflexão ou investigação do comportamento dos utilizadores, como ilustrado já em (Costa, Mota e Santos, 2012).

A principal razão foi a já mencionada falta de participantes, sobretudo automáticos, mas também humanos, que não nos permite generalizar com um mínimo de confiança, e, por outro lado, a confirmação de que os seres humanos ainda não têm par na resposta ao tipo de perguntas do Págico. Contudo, houve algumas respostas corretas encontradas pelos sistemas e não propostas pela participação humana, o que leva a esperar que de facto o concurso, no sentido de ajuda, já é e será cada

vez mais benéfico na resposta a perguntas a uma base de grande informação.

Foi de qualquer forma importante que, mais uma vez, todos fôssemos obrigados a refletir na questão extremamente complexa do que é uma resposta (Freitas et al., 2012) e na dificuldade de delimitação rigorosa do que é útil ou apropriado na procura ou descoberta de informação sobre um dado tema.

4 O realismo da tarefa

Uma das questões metodologicamente mais complexas na organização de uma avaliação conjunta é a obtenção de uma tarefa finita e bem delimitada que seja por outro lado passível de repetição e extensão. Que não seja simplesmente uma demo ou uma curiosidade, mas que seja de certa forma representativa de problemas práticos e autênticos na vida de utilizadores da wikipédia (como fonte de informação sobre a cultura lusófona).

A primeira coisa que tivemos de decidir e que, de certa forma, contraria o realismo da tarefa, mas que era essencial no nosso caso, foi a limitação à versão em português. Todos nós sabemos da possibilidade de navegação entre várias línguas na wikipédia, portanto não é realista obrigar a procurar só em português. Mas tínhamos razões de sobra para fazer esta escolha: por um lado, criar uma coleção da wikipédia com todas as línguas potenciais dos participantes incluindo o inglês seria uma tarefa demasiado grande para os nossos meios; por outro—e esta talvez seja a razão mais importante—, fizemo-lo no GikiCLEF e acabámos por não conseguir medir o impacto ou interesse da parte portuguesa.

Já referimos em diversos outros artigos (por exemplo, em (Santos, 2012) e (Mota et al., 2012)), mas importa aqui novamente realçar, que também o desenvolvimento de um sistema de navegação na wikipédia para indicar respostas e justificações, que concorresse com a forma humana e habitual com que os participantes contactam e interagem com esse recurso, não foi fácil, e possivelmente nem mesmo bem sucedido.

Este é um dado que precisamos de levar em conta mais tarde, se viermos a organizar mais avaliações conjuntas com participação humana. Se, para sistemas automáticos, é só definir uma sintaxe rígorosa de entrada e saída, e escrever validadores que a verifiquem ou corrijam, a situação é totalmente diferente quando queremos que participantes humanos não sejam impedidos ou contrariados, em vez de ajudados, numa dada

¹No entanto, foram pontualmente feitas, e marcadas para fazer no futuro, correções ortográficas, gramaticais, de conteúdo e a hiper-ligações com problemas diversos.

tarefa. Deveriam ter sido feitos estudos de usabilidade e ter sido dada muita mais atenção à forma de desenvolver sistemas realmente apropriados à tarefa em mente, e o facto de termos (todos, organizadores e participantes) tido pouco tempo e termos dado pouco retorno acabou por espantar muitos participantes, ou levar a que a maior parte deles usasse, não o nosso sistema, mas a interface normal da wikipédia. É interessante ver que isto aponta para duas linhas de desenvolvimento que já têm sido mencionadas:

- tentar diminuir a novidade ou diferença nas interfaces: idealmente, apenas adicionar algo àquilo que os utilizadores já conhecem e de que gostam, e não obrigá-los a criar novos hábitos ou raciocínios;
- não pedir para fazer mais do que é preciso. . . veja-se a interface do Webpaper descrita em (Simões, Rocha e Fonseca, 2009), que aliás reuniu muito mais participantes que o Páxico, mostrando que atividades lúdicas, ou didacto-lúdicas, têm um grande potencial para recolher informação sobre participantes humanos.

Uma outra questão que surgiu, em particular através da chamada de atenção da Belinda Maia, foi a de os temas escolhidos serem maioritariamente de letras, e não de ciência, economia ou técnica/tecnologia.

A questão aqui é até que ponto existe ciência em português, ou seja, até que ponto a informação que tentaríamos obter era escrita de raiz em português ou especialmente relacionada à cultura lusófona.

Mas esta é uma área e atividade que, a nível pedagógico nos próprios países lusófonos, poderá ter um impacto fundamental e mais importante do que aquele relacionado com a cultura lusófona para estrangeiros, e que fica pois agendada como ideia para o futuro.

5 Os problemas do reuso

É evidente que sempre que não se começa do princípio, mas se usa algo já desenvolvido, isso tem vantagens. Mas é preciso também referir que nem tudo são rosas numa tal abordagem, sobretudo se os novos desenvolvedores não são os mesmos do sistema anterior, como foi o caso do SIGA.

Assim, a escolha da forma de desenvolvimento de um dado sistema passa a obedecer a dois princípios que por vezes são contraditórios:

1. minimizar as alterações ao que já está feito, procedendo de forma incremental;
2. adicionar novas funcionalidades de acordo com o mais adequado ao utilizador.

Esta questão, que não é de fácil resolução, teve impacto nas três mudanças principais realizadas ao SIGA: a adição de utilizadores humanos, já comentada acima, a melhoria da interface de avaliação, e a apresentação dos resultados com novas medidas.

Nos três casos poderíamos ter desenvolvido soluções mais inovadoras e capazes. Todavia, talvez não tivéssemos ainda realizado a própria avaliação, que foi aquela que foi conseguida num prazo tão curto.

6 Autocrítica

Existe uma série de pontos em que fizemos as opções erradas ou não conseguimos dar conta do recado, e que parece mais natural indicar aqui por atacado numa lista, sem tentar justificar ou desculpar. Obviamente, esses casos são automaticamente casos a melhorar, se houver um próximo Páxico:

- Não houve qualquer sugestão de perguntas ou tópicos por parte dos participantes. Se tivéssemos conseguido que os tópicos fossem/fizessem parte de uma partilha de vários investigadores sobre questões que lhes interessavam e sobre as quais queriam saber mais, e ao mesmo tempo na área ou em questões sobre as quais os seus sistemas brilhariam ou estavam especialmente interessados em ser avaliados, o processo teria sido muito melhor e tido muito mais participação.
- Não publicámos as medidas de avaliação a tempo de serem discutidas, internalizadas ou sequer tomadas em conta no desenvolvimento dos sistemas. De facto, foram publicadas só depois de os participantes terem enviado as suas respostas.
- A coleção do Páxico foi disponibilizada em várias versões, e nem a última estava imune a problemas, conforme descrito em (Simões, Costa e Mota, 2012).
- Devido a vários problemas detetados demasiado tarde, a interface de participação humana foi alterada várias vezes durante o próprio mês em que o Páxico esteve aberto, o que pode ter levado a confundir os utilizadores, e de dificultar os nossos estudos da sua interação.

- Ao contrário das nossas intenções, não conseguimos publicar um manual de utilização do sistema para os participantes humanos, o que, estamos convencidos, afastou alguns inscritos e muitos que poderiam ter tentado se fossem mais ajudados.
- Devido a um problema apenas descoberto tarde, algumas respostas não foram sequer avaliadas, resultando em erros nos resultados finais.
- Muitas das modificações e melhorias feitas ao sistema de avaliação foram-no à posteriori, não tendo os avaliadores a possibilidade de delas beneficiar.
- Devido à enorme quantidade de respostas, não foi possível usar uma das funcionalidades mais interessantes do SIGA, nomeadamente a avaliação sobreposta e a análise subsequente de possíveis conflitos, a não ser num número muito diminuto de casos, como referido em (Freitas et al., 2012)
- Devido a um engano, parte da avaliação sobreposta foi feita conhecendo a anterior avaliação: ou seja, as respostas a avaliar foram atribuídas referindo que “estas são as respostas duvidosas do avaliador Y”, o que impediu observar se também levantariam dúvidas a outros avaliadores não precavidos.
- Por causa dos prazos, foi necessário publicar os resultados sem fazer uma revisão completa às respostas, o que implicou que estamos conscientes de ainda haver erros no material disponibilizado.
- Não foi possível traduzir toda a interface do SIGA, que estava em inglês, para português, nem vice-versa no que se refere às funcionalidades novas, que estão apenas em português, nem documentar exaustivamente as ditas.
- Também não foi possível incluir ou processar convenientemente as corridas não oficiais dos sistemas automáticos, o que claramente melhoraria o Cartola (Mota, 2012; Simões, Costa e Mota, 2012).

Por todas estas imprecisões ou faltas, estamos convencidos de que seria muito interessante ter um período de consolidação pós-Págico em que tanto os recursos, como o sistema, como a coleção pudessem ser polidos, melhorados e investigados em mais detalhe—não só por nós, mas por todos quantos acham o assunto interessante. Consideramos que os problemas e imperfeições existentes no recurso são em muito superados

pelo facto de não guardarmos o nosso trabalho só para nós ou esperarmos que esteja perfeito para disponibilizar... pelo contrário, disponibilizamo-lo a todos assim que o consideramos útil (o que não nos impede de continuar a melhorar e criar novas versões), para que ajudem a melhorá-lo e possam aprender com os nossos erros também.

7 Contributos para o futuro

Como já dissemos, o Págico foi organizado e concluído num tempo recorde, o que faz com que muita prospeção sobre os dados recolhidos, que nós teríamos gostado de fazer, ficou para o futuro.

Mas além disso gostávamos de mencionar aqui algumas ideias de aproveitar criativamente o material, em ocasiões posteriores:

- refazer a coleção noutras datas, e confirmar / reclassificar as respostas encontradas nessa altura, muito provavelmente pedindo a novas equipas/pessoas para encontrar as respostas;
- com base em temas ou super-temas, tentar criar perguntas automaticamente, à semelhança da avaliação conjunta QG (“question generation”) (Rus et al., 2012), veja-se <http://www.questiongeneration.org/>;
- usar o SIGA para adicionar mais tópicos e respostas de forma a ir criando uma base maior de perguntas respondidas pela wikipédia;
- usando as mesmas perguntas e respostas, mas com novas pessoas, verificar como é que elas reagiriam para justificar ou negar uma dada resposta: ou seja, criar algo parecido com a AVE (Rodrigo, Peñas e Verdejo, 2009), para identificar processos típicos ou comuns de raciocínio humano, e também clarificar a dificuldade ou não de diferentes pares de perguntas e respostas;
- desenvolver sistemas interativos que apresentassem, para cada tópico, o conjunto de respostas de maneira agregada e satisfatória.

Todas estas iniciativas poderiam aumentar o valor do Cartola e capitalizar o trabalho feito no Págico, e oferecemo-las a quem as quiser desenvolver.

8 Comparação com avaliações conjuntas anteriores

Finalmente, parece-nos apropriado fazer alguns comentários baseados na longa experiência de

avaliações conjuntas que a Linguateca iniciou há precisamente dez anos.

Não há dúvida que essa organização é trabalhosa e tem de ser independente dos participantes; por outro lado, desde o início que quanto mais associados ou ligados à Linguateca (ou à organização), maior a probabilidade de um grupo ou investigador participar.

Se por um lado é natural que grupos próximos partilhem opiniões científicas sobre o progresso na área, o perigo da participação próxima é que isso pode levar a que as avaliações conjuntas sejam vistas como apenas valendo a pena para a própria Linguateca e para fomentar os nossos objetivos, desvirtuando pois a vertente de serviço à comunidade, que é o que nos move.

Em 2008, fizemos uma consulta à comunidade, e a resposta que tivemos foi a de que alguns grupos (dois) estariam interessados em avaliar sistemas de deteção de terminologia. Contudo, nós não considerámos apropriado fazer uma avaliação conjunta nesse domínio visto que não havia nenhuma forma independente de obter recursos dourados, e pareceu-nos que uma avaliação à posteriori iria ser demasiado subjetiva. Além disso, e ao contrário do Págico, não conseguiríamos juntar resultados que fossem relevantes para futuras edições, ou seja, seria algo que apenas serviria para um dado conjunto de textos e domínios, fixo na altura.

Mas por outro lado estamos conscientes de que, e dado o decréscimo significativo do número de participantes das Morfolimpíadas para o HAREM para o CLEF/GikiCLEF e agora para o Págico, sem garantir uma base real de participantes suficientemente alargada não faz sentido a Linguateca organizar mais avaliações conjuntas.

Uma outra vertente que faz sentido comentar é o tamanho das coleções disponibilizadas e que se esperava que os sistemas processassem. De um conjunto de pequenos textos nas Morfolimpíadas para coleções de textos pequenas no HAREM passámos a coleções jornalísticas e depois para versões da wikipédia cada vez maiores. Se por um lado isso espelha o progresso no PLN, por outro pode também consistir um problema para arranjar participantes—como mencionado em (Cardoso, 2012), o sistema não conseguiu processar a coleção toda em tempo útil.

Finalmente, repare-se que as tarefas em que tentamos avaliar um sistema são cada vez mais próximas das tarefas de um utilizador humano: analisar corretamente uma dada palavra fora do contexto em todas as interpretações morfológicas é algo que só se faz (?) na escola, estando

claramente na província de sub-sistemas transparentes ao utilizador, para usar a terminologia de Gaizauskas (1998), classificar/identificar um nome próprio como mencionando uma pessoa, uma instituição, uma abstração ou um local é já algo que um ser humano faz “automaticamente”, sem pensar, enquanto procurar informação sobre um tema para responder a perguntas concretas usando a wikipédia é algo que se faz conscientemente no mundo atual.

Dito isto podemos portanto identificar uma diminuição no número de participantes e de movimento à volta da avaliação – que, pensamos, congregou praticamente toda a comunidade no caso das Morfolimpíadas, pelo menos se virmos a audiência do encontro satélite do PROPOR 2003 na altura—e que no Págico foi mínima, e por outro lado um aumento na dificuldade da tarefa oferecida, tanto a nível de tamanho de recurso como a nível de comparação com desempenho humano.

9 Comentários finais

Pensamos que mesmo com as restrições temporais e de financiamento que a Linguateca sofreu, e que podemos adjectivar de drásticas sem exagero², o Págico conseguiu alguns resultados interessantes, e congregar mais alguns interessados na área da inter-relação entre a recolha de informação e o processamento computacional das línguas.

Pensamos que o grande contributo, além do início de uma avaliação científica da wikipédia em português, foi a criação do Cartola, que permitirá que muitos outros investigadores, no futuro, possam treinar os seus sistemas e/ou avaliá-los com base no material por nós coligido, além de fazer prospeção de outros assuntos que nós nem sequer tenhamos (ainda) abordado.

Também apelamos à criação de novas iniciativas, por exemplo com valor pedagógico, ou noutras áreas mais relacionadas com interesses específicos e em que a wikipédia pode ser ou vir a ser uma fonte importante, que possam (re)usar o nosso trabalho, e os sistemas desenvolvidos.

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

²Em relação a 2011, o panorama está descrito em Santos (2011), em relação a 2012, não obtivemos qualquer financiamento.

POSC/339/1.3/C/NAC, pela UMIC e pela FCCN, e durante 2011, pela Fundação da Ciência e da Tecnologia (FCT) e pela Fundação para a Computação Científica Nacional (FCCN).

A Universidade de Oslo também contribuiu para a organização do Páxico, e especificamente financiou substancialmente a organização do encontro do Páxico. Um dos membros da organização foi parcialmente suportado pela bolsa da Fundação da Ciência e da Tecnologia SFRH/BPD/73011/2010. Também agradecemos à PUC-Rio e à universidade de Coimbra pelo apoio prestado.

Agradecemos sinceramente a todos os participantes, sem os quais o Páxico teria sido inútil, ao Fernando Ribeiro e ao Hernâni Costa, da equipa da Linguateca, pelo retorno sobre o SIGA no que se refere à participação humana, e ao Paulo Rocha pelo esforço titânico na avaliação e pela cultura geral e imaginação empregada na escolha de tópicos.

Agradecemos também à comissão científica do presente volume, em particular ao Xavier Guinovart e ao António Teixeira, que, mais uma vez submetida a prazos sobre-humanos, conseguiu mesmo assim contribuir significativamente para a sua qualidade e para a variedade de ideias interessantes que podemos apresentar.

Referências

- Aires, Rachel Virgínia Xavier. 2005. *Uso de marcadores estilísticos para a busca na Web em português*. Tese de doutoramento, ICMC - USP - São Carlos, Agosto, 2005. <http://www.linguateca.pt/documentos/TeseDoutRachelAires.pdf>.
- Cardoso, Nuno. 2012. Medindo o precipício semântico. *Linguamática*, 4(1), Abril, 2012. Neste volume.
- Costa, Luís, Cristina Mota, e Diana Santos. 2012. SIGA, a Management System to Support the Organization of Information Retrieval Evaluations. Em Helena Caseli, Aline Villavicêncio, António Teixeira, e Fernando Perdigão, editores, *Computational Processing of the Portuguese Language, PROPOR'2012*, pp. 284–290, Berlim/Heidelberg. Springer.
- Costa, Luís, Paulo Rocha, e Diana Santos. 2007. Organização e resultados morfolímpicos. Em Diana Santos, editor, *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*. IST Press, Lisboa, Portugal, capítulo 2, pp. 15–33.
- 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.
- Gaizauskas, Robert. 1998. Evaluation in language and speech technology. *Computer Speech and Language*, 12(4):249–62.
- Mota, Cristina. 2012. Resultados páxicos: participação, medidas e pontuação. *Linguamática*, 4(1), Abril, 2012. Neste volume.
- Mota, Cristina e Diana Santos, editores. 2008. *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*. Linguateca.
- Mota, Cristina, Alberto Simões, Cláudia Freitas, Luís Costa, e Diana Santos. 2012. Páxico: Evaluating Wikipedia-based information retrieval in Portuguese. Em *Language Resources and Evaluation Conference*, Maio, 2012.
- Rodrigo, Álvaro, Anselmo Peñas, e Felisa Verdejo. 2009. Overview of the answer validation exercise 2008. 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*, pp. 296–313, Berlim/Heidelberg. Springer.
- Rus, Vasile, Brendan Wyse, Paul Piwek, Mihai Lintean, Svetlana Stoyanchev, e Cristian Moldovan. 2012. A Detailed Account of The First Question Generation Shared Task Evaluation Challenge. *Dialogue & Discourse*, 3(2):177–204.
- Santos, Diana. 2011. Relatório da Linguateca relativo ao ano de 2011. Relatório técnico, Linguateca/FCCN. <http://www.linguateca.pt/documentos/Relatorio2011LinguatecaFinal.pdf>.
- Santos, Diana. 2012. Porquê o Páxico? Razões para uma avaliação conjunta. *Linguamática*, 4(1), Abril, 2012. Neste volume.
- Santos, Diana e Nuno Cardoso, editores. 2007. *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca, Linguateca, 12 de Novembro, 2007.

Simões, Alberto, Luís Costa, e Cristina Mota. 2012. Tirando o chapéu à Wikipédia: A coleção do Págico e o Cartola. *Linguamática*, 4(1), Abril, 2012. Neste volume.

Simões, Alberto, Paulo Rocha, e Rúben Fonseca. 2009. Webpaper — más perguntas e boas respostas: a arte de interrogar. Em Paulo Dias, António José Osório, e Altina Ramos, editores, *O digital e o currículo*. Centro de Competência da Universidade do Minho, pp. 227–238, Maio, 2009.