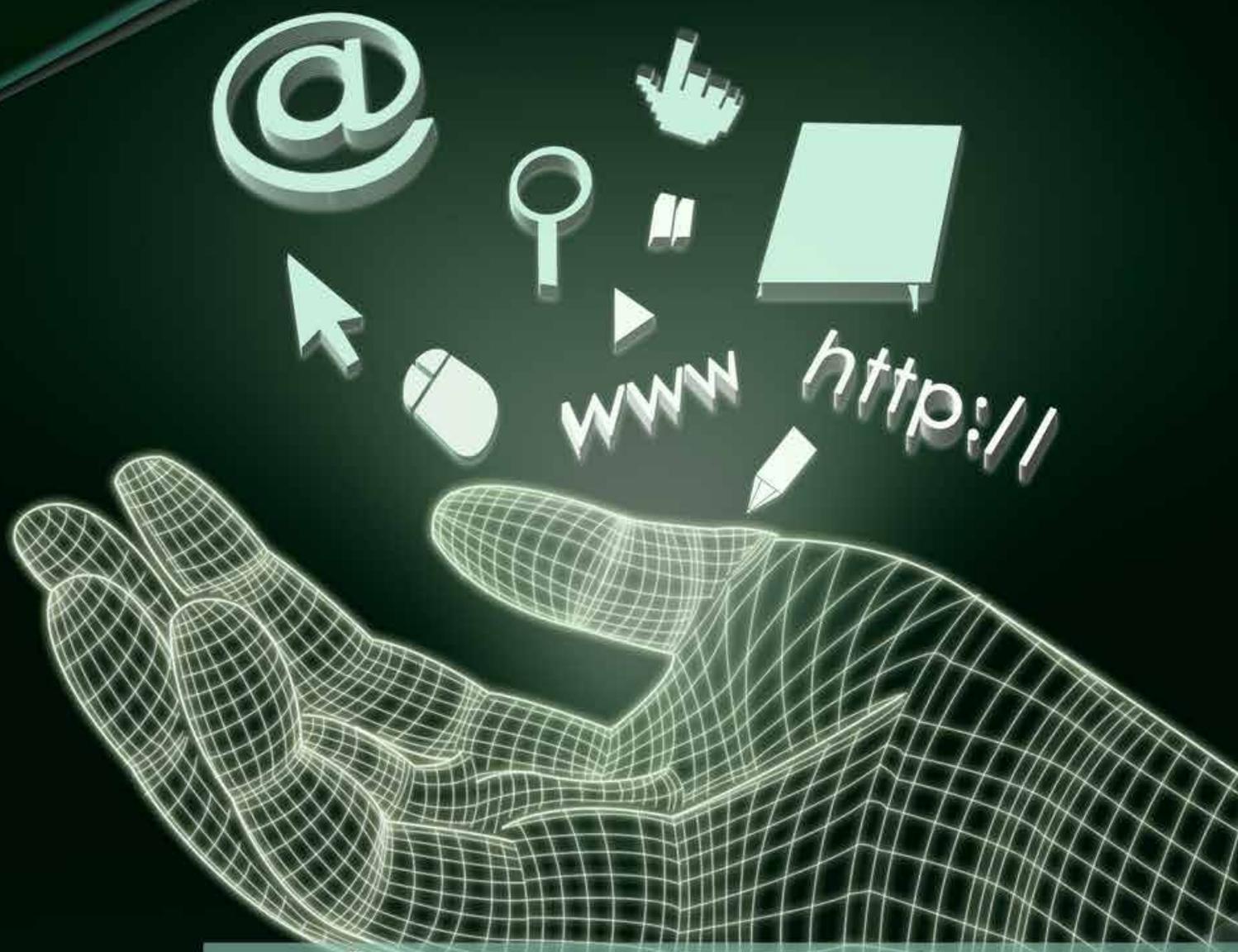


Platô

REVISTA DO INSTITUTO INTERNACIONAL
DA LÍNGUA PORTUGUESA

número
03
volume 2
2 0 1 3



A LÍNGUA PORTUGUESA NA INTERNET E NO MUNDO DIGITAL

Colóquio Internacional
Guaramiranga

Contato:

Instituto internacional da Língua Portuguesa (IILP)

Av Andrade Corvo, nº 8

Plateau, Cidade da Praia - Cabo Verde

Telefone: (238) 261 95 04

www.iilp.org.cv

www.riilp.org

e-mail: revistariilp@gmail.com

número

03

volume 2

2013

REVISTA DO INSTITUTO INTERNACIONAL
DA LÍNGUA PORTUGUESA

Colóquio Internacional de Guaramiranga:

A Língua Portuguesa na Internet e no Mundo Digital

Volume 2 Número 3 2013

Gilvan Müller de Oliveira
(organizador)**Editores**Gilvan Müller de Oliveira
Rosângela Morello**Secretaria Executiva**

Denise Fonseca

Comitê editorial

Ana Isabel Soares (Instituto Camões, Portugal)

António Branco (Universidade de Lisboa, Portugal)

Clémence Jouët-Pastré (Universidade de Havard, Estados Unidos da América)

Emir José Suaiden (Instituto Brasileiro de Informação em Ciência e Tecnologia/IBICT, Brasil)

Gregório Firmino (Universidade Eduardo Mondlane, Moçambique)

Guadalupe Teresinha Bertussi (Universidade Pedagógica do México, México)

Manuel Monteiro da Veiga (Universidade de Cabo Verde, Cabo Verde)

Maria José Grosso (Universidade de Macau, Macau)

Equipe Técnica

Alberto Gonçalves: revisor técnico

Ana Paula Seiffert: divulgação

Felipe de Almeida: design gráfico

Vanessa de Luca Bortolato: design gráfico

Ficha catalográfica (em tramitação)

Revista do Instituto Internacional da Língua Portuguesa (RIILP) - Instituto
Internacional da Língua Portuguesa (IILP) - V.2, N.3 (2013), Cidade da Praia,
Cabo Verde: Editora do IILP, 2012

Semestral

ISSN: 2311-6625 on line

Arte da capa: Felipe Almeida e Vanessa de Luca



ÍNDICE

04 APRESENTAÇÃO: A LÍNGUA PORTUGUESA NA INTERNET E NO MUNDO DIGITAL
Gilvan Müller de Oliveira

06 O ESPAÇO DAS LÍNGUAS NA INTERNET: PANORAMA E TENDÊNCIAS COM ÊNFASE NO PORTUGUÊS
Daniel Pimienta

28 A LÍNGUA PORTUGUESA FACE AO CHOQUE TECNOLÓGICO DIGITAL
António Branco

38 O USO DE *CORPORA* PARA A CONSTITUIÇÃO DE RECURSOS LEXICOGRÁFICOS DE REFERÊNCIA: O CASO DO VOC
José Pedro Ferreira, Gladis Maria de Barcellos Almeida, Margarita Correia

56 CONSIDERAÇÕES AO “COMPUTÊS”: COMO A ADOBE ADAPTA-SE ÀS NOVAS TENDÊNCIAS DO PORTUGUÊS NA INTERNET E NO MUNDO DIGITAL
Guta Ribeiro

68 A INTERNET EM MOÇAMBIQUE: IMPLANTAÇÃO, USO E DESAFIOS
Inês Machungo / Gregório Firmino

80 CONTRIBUTO GALEGO PARA A LÍNGUA PORTUGUESA NA ERA DIGITAL: AS FERRAMENTAS AO-OA DA ACADEMIA GALEGA DA LÍNGUA PORTUGUESA
Ernesto Vazquez Sousa / Concha Rousia

96 MULTILINGUISMO NO CIBERESPAÇO: A PARTICIPAÇÃO DO PORTUGUÊS NUMA B@BEL DIGITAL?
Cláudio Menezes

106 O CENTRO VIRTUAL CAMÕES: DIVULGAÇÃO E FORMAÇÃO
Rui Vaz



114 INTERNET: A NOVA GRANDE CARAVELA DA LÍNGUA PORTUGUESA NA ERA DIGITAL

Afonso João Miguel

124 CARTA DE GUARAMIRANGA



APRESENTAÇÃO

A LÍNGUA PORTUGUESA NA INTERNET E NO MUNDO DIGITAL

Gilvan Müller de Oliveira

Que desafios e perspectivas marcam a gestão da língua portuguesa na nova economia ou na assim chamada *economia do conhecimento*? Qual a situação das línguas face à internet e quais ações e inovações têm sido implementadas?

Constituindo um dos quatro eixos de discussão propostos ao Instituto Internacional da Língua Portuguesa pelo *Plano de Ação de Brasília para a Promoção, Difusão e Projeção da Língua Portuguesa* (PAB2010), estas e outras questões foram abordadas no III Colóquio “A Língua Portuguesa na Internet e no Mundo Digital” realizado em Guaramiranga, Ceará, Brasil, entre os dias 24 e 26 de abril de 2012.

Especialistas e gestores de várias partes do mundo trouxeram ao debate as múltiplas faces da gestão das línguas nesse novo paradigma comunicacional das redes digitais. Este número reúne alguns dos tópicos debatidos no evento.

Nos dois primeiros textos, ganha centralidade a análise da entrada das línguas na rede digital. Pimenta traz um breve histórico sobre a emergência da internet, discute a situação da língua portuguesa nesse ambiente e aponta a necessidade de ampliar não só o acesso dos falantes às tecnologias como, e principalmente, de gerar e disponibilizar conteúdos nas diferentes línguas. Na sequência, Branco descreve o choque tecnológico, dando destaque às ações necessárias para assegurar ao português sua posição de língua de comunicação internacional com projeção global.

No segundo bloco temos a apresentação de dois importantes projetos para a gestão da língua portuguesa. O primeiro deles é o Vocabulário Ortográfico Comum da Língua Portuguesa – VOC, concebido e realizado multilateralmente sob a coordenação geral do IILP. Os autores Ferreira, Almeida e Correia analisam os diferentes tipos de dados usados na construção de obras lexicográficas e descrevem a metodologia adotada no desenvolvimento do VOC, procurando mostrar o papel instrumental dos *corpora* para a definição de primeiras nomenclaturas de variedades em processo de standardização, para a validação de dados obtidos da tradição e para a aferição da frequência e especificidade do âmbito de uso de cada entrada. Ribeiro, por sua vez, tematiza as iniciativas da Adobe Systems, analisando a constituição das comunidades de tradutores na internet, em especial no âmbito da lusofonia. A autora descreve os processos de criação, implementação e avaliação de alguns projetos da Adobe Systems destinados a potencializar o rápido acesso à informação e menciona os desafios e planos futuros.

É um panorama sobre a implantação e usos da internet em Moçambique que nos apresenta o texto Machungo e Firmino. De acordo com os autores, assiste-se no país a uma significativa implantação e uso de tecnologias de informação e comunicação e ao recurso à internet. No entanto, consideram que há constrangimentos que impedem um mais amplo acesso dos cidadãos à rede digital relacionados à situação socioeconômica e linguística do país caracterizada entre outros aspectos, pela diversidade linguística e um elevado índice de analfabetismo.

Um terceiro conjunto de textos tematiza o desenvolvimento de iniciativas no âmbito de instituições de ensino ou promoção da língua portuguesa. Rousia e Vazquez Souza trazem uma síntese dos projetos digitais da Academia Galega da Língua Portuguesa colocando em destaque três deles: i) o Portulano de Recursos: escritório virtual e agrupador de recursos digitais sobre a língua portuguesa desenhado com o software livre Netvibes; ii) o Arquivo Digital: repositório de objetos digitais orientado à preservação e divulgação que funciona sobre um motor DSpace e gerencia metadados no standard Dublin Core; e iii) o projeto do Dicionário Eletrônico e-Estraviz em parceria com a AGAL (Associação Galega da Língua). Menezes faz uma breve descrição da situação estatística das línguas no mundo, discute metodologias e tecnologias para a inclusão e revitalização das línguas no mundo digital e conclui com uma apresentação de projetos e sugestões destinados a uma maior presença da língua portuguesa no ciberespaço. Rui Vaz coloca o foco sobre o Centro Virtual Camões e suas principais linhas de atuação o ensino e aprendizagem do português e formação dos agentes da educação, bem como para a divulgação da língua e cultura portuguesa.

Por fim, Afonso Miguel lança um olhar abrangente sobre a história da língua portuguesa, mencionando sua potencialidade face à internet e os condicionamentos socioeconômicos, científicos e tecnológicos existentes em cada país. Entre os desafios enfrentados, o autor cita o reconhecimento das especificidades sociolinguísticas de cada país membro da CPLP, que requer, por exemplo, a inserção de terminologias científicas e técnicas locais no português comum a usar na internet e noutros meios digitais.

Para fechar o número, A Carta de Guaramiranga sintetiza as principais demandas formuladas pelos participantes do III Colóquio, e nos impulsiona na busca por concretizar ações que garantam cada vez mais o pleno exercício do direito ao conhecimento sedimentado no democrático uso e consistente circulação das línguas na internet e no mundo digital, e muito especialmente do português.

O ESPAÇO DAS LÍNGUAS NA INTERNET: PANORAMA E TENDÊNCIAS COM ÊNFASE NO PORTUGUÊS¹

Daniel Pimenta

Sobre o autor

Presidente do FUNREDES², Membro do comitê executivo da MAAYA³.
Contato: pimenta@funredes.org

Resumo

Iniciando por uma breve introdução à questão das línguas na internet, apresentamos uma análise dos índices que mapeiam a sua presença trazendo dados específicos sobre o português. Desde uma perspectiva histórica, buscamos delinear tendências para o futuro das línguas mostrando a necessidade de ampliar não só o acesso dos falantes às tecnologias como, e principalmente, de gerar e disponibilizar conteúdos nas diferentes línguas.

Palavras-chave

Línguas na Internet. Português. Mapeamentos. Multilinguismo.

¹ Este artigo em português é inspirado em uma subseção de versão anterior publicada em inglês do capítulo *Content and Language* do livro *Accelerating Development Using the Web: Empowering Poor and Marginalized Populations*, G. Sadowsky Editor, W3C, 2012. A respectiva apresentação pode ser consultada na página: <http://funredes.org/presentation/IILP-Fortaleza%20Portugues.ppt> Traduzido para o português por Isis Ribeiro Berger.

² Networks & Fundação para o Desenvolvimento (FUNREDES, sigla para o nome em espanhol – <http://funredes.org>) é uma organização não governamental com sedes na República Dominicana, França, Venezuela, Espanha e Guiana Francesa que tem sido pioneira (desde 1988) no campo do uso de Tecnologias da Informação e Comunicação para o Desenvolvimento (ICT4D) e tem trabalhado, desde sua fundação, na importância transversal da diversidade linguística por meio de diversos projetos que lidam com a exclusão digital, com o impacto social da Internet e também com o letramento digital.

³ MAAYA (Rede Mundial pela Diversidade Linguística), uma organização de sociedade civil, é uma rede de agentes interessados na diversidade linguística. Embora sua missão não enfoque somente no ciberespaço, uma grande parte das atividades da MAAYA é relacionada ao mundo digital. MAAYA é um sub-produto do processo WSIS e foi lançada em 2002 pela Academia Africana de Línguas (<http://www.acalan.org>). A MAAYA organizou três Simpósios Internacionais sobre Multilinguismo e Ciberespaço, um em 2009 em Barcelona (<http://www.maaya.org/spip.php?article105>) e outro em 2011 em Brasília (<http://www.let.unb.br/simc/>) e o terceiro em Paris, em 2012. A MAAYA, que está prestes a publicar um antologia sobre os riscos e desafios do multilinguismo no ciberespaço, tem assumido a iniciativa do projeto DILINET para resolver os problemas dos indicadores da diversidade linguística no ciberespaço. A MAAYA reúne alguns 25 agentes da área entre seus membros, incluindo organizações internacionais como a UNESCO, ITU e OIF (<http://www.francophonie.org>). MAAYA vem dedicando grande parte de sua energia na organização da Reunião Mundial de Cúpula sobre a Diversidade Linguística prevista para 2015.

1 - Introdução às línguas na internet

O quão importante foi, é e será o fator linguístico na evolução da Internet? Quais repercussões poderia ter a internet na vida das línguas? Como os conteúdos da internet e as línguas se inter-relacionam? Qual a situação específica do português? Estas são questões difíceis de serem respondidas, em parte devido à escassez de dados mais sólidos.

Os pioneiros da Internet (uma grande porcentagem de fato, se abarcarmos nessa ampla definição todos que contribuem e experts desde a década de 1980 até o início da década de 1990) criaram uma grande gama de neologismos e expressões originais que tentavam capturar as dimensões e originalidade da mudança de paradigma que estava ocorrendo concomitante ao desenvolvimento de redes de infraestrutura.

Para citar apenas três exemplos:

- 1) A capacidade obtida para adequar o nível local ao nível global fez surgir o termo "glocal" com a conotação de algo que compartilha ambas as características.
- 2) Anonimidade no ciberespaço– alguns leitores devem se lembrar da famosa tirinha do nova-iorquino Peter Steiner em 1993 dizendo "Na internet, ninguém sabe que você é um cão"⁴.
- 3) Sempre foi lugar comum afirmar que "a internet não tem fronteiras".

Alguém deve se perguntar se esses recentes conceitos, que foram criados em contextos do inglês falado, são ainda pertinentes em um contexto multilíngue. No que se refere ao "glocal", o termo definitivamente não é tão relevante: para a grande maioria de falantes de línguas indígenas que não possuem o inglês como segunda língua não é fácil relacionar seu ambiente local e a cultura holística ao espaço global do inglês falado com sua cultura mais regimentada e analítica. Já no que tange à anonimidade do cão que está usando a internet, aqueles dias não mais existem, especialmente se tratamos de um cão 'socialmente conectado'. Nesse caso, muitas pessoas não estão cientes de que se trata de um cão, nem mesmo de que raça ou do que teve para o café da manhã, a não ser que, obviamente, não se trate de um cão falante de inglês. Já no que se refere à fronteira, esse é o caso mais interessante dos três. À luz da situação atual, o conceito de fronteira deve merecer ser desconstruído e reconstruído como uma nova expressão: 'por internet, as únicas fronteiras são as línguas'⁵ e há centenas delas⁶, mais do que países existentes no mapa mundi. Somente pessoas multilíngues podem ousar cruzar algumas dessas fronteiras e as ferramentas vem sendo desenvolvidas para lidar com o desafio linguístico. O uso de tradutores automáticos está começando a oferecer uma certa ilusão equivocada para aqueles que buscam cruzar territórios estrangeiros.

A "territorialidade linguística" da internet e como ela está relacionada aos conteúdos tem sido constantemente subestimada nas análises, porque as pessoas naturalmente tendem a pensar a partir de suas próprias fronteiras linguísticas. Assim, é importante descobrir e então analisar a dimensão oculta da *capacidade da internet* de incluir a fim de lidar com os desafios futuros do mais recente estágio de evolução e especialmente com a tão mencionada exclusão

⁴ http://en.wikipedia.org/wiki/On_the_Internet,_nobody_knows_you%27re_a_dog

⁵ Ver, por exemplo, <http://googleresearch.blogspot.com/2011/07/languages-of-world-wide-web.html> para ter ideia de quanto os territórios linguísticos tendem a ser rígidos.

⁶ Os índices para os números de línguas situadas (línguas que possuem uma existência digital) é considerado abaixo de 500 a contar de hoje.

Daniel Pimienta

digital

As línguas são importantes, talvez mais do que nunca, em fazer com que o mundo se torne menor. A internet aumenta a probabilidade de encontros entre pessoas de diferentes línguas e a pergunta real é: que uso relevante alguém pode fazer de seu acesso à internet se sua língua materna não é reconhecida ou se não há conteúdo em sua língua? Tais problemáticas não eram óbvias nos estágios iniciais do desenvolvimento da rede, quando a maioria dos usuários eram profissionais de Tecnologia da Informação, pesquisadores e acadêmicos ou, algumas vezes, ativistas internacionais de tecnologia. Tratava-se de pessoas altamente educadas, pelas quais o inglês era provavelmente compreendido no mínimo como segunda ou terceira língua por uma porcentagem significativa. No entanto, a rede se difundiu amplamente e as questões de ordem linguística se tornaram primordiais uma vez que seu objetivo é prover acesso a qualquer um no planeta. Dentre os sete milhões de seres humanos, menos de 20%⁷ são capazes de usar o inglês e provavelmente menos de 15% de forma eficiente. Somente essa estatística é um poderoso indicativo de que a crença de que o inglês será mantido como língua franca da rede tem seus dias contados.

No século 21, a vantagem estratégica no mundo virtual aponta para o multilinguismo. Governos têm gradualmente se tornado conscientes dessa realidade e tentado encorajar uma mudança que abarque outras línguas. O Conselho Britânico, por exemplo, alertou jovens cidadãos ingleses sobre o risco que estão correndo na nova economia da União Europeia se permanecerem monolíngues [Graddol-2006]. Dessa forma, a questão da escolha linguística e diversidade na internet é uma das chaves para abrir portas para a inclusão digital de forma mais direta e abrangente, especialmente levando em consideração a ligação complexa e indissolúvel entre as línguas, o substrato de conhecimento e o conteúdo, uma das duas facetas chaves para que a malha da rede⁸ seja melhor compreendida.

2 - Alguns índices sobre as línguas na internet

Que dados podem ser coletados neste estágio considerando a variedade e a presença proporcional das línguas na internet?

Antes de responder essa pergunta, será útil retomar alguns fatos sobre as línguas e situar o estágio em que nos encontramos atualmente. Há uma estimativa de 30.000 línguas existentes no mundo desde que os seres humanos tornaram-se capazes de produzir discurso. A partir de então, muitas línguas tornaram-se extintas e, em tempos modernos, há uma estimativa de que algo em torno de 6000 a 9000 permanecem em uso. Levando em consideração os requisitos para que todas as línguas tenham uma existência na rede⁹,

⁷ Comumente, em demolinguística, não há índices consensuais para a população total do mundo com capacidade de compreender inglês como segunda língua. A população usuária de inglês como primeira língua é estimada em menos de 375 milhões (menos de 6% da população mundial). O índice para o inglês como segunda língua possui grandes variações, dependendo em particular da definição do nível de letramento: de 470 milhões para um bilhão (que perfaz um total de falantes de inglês no mundo na casa dos 10%-20%). Graddol (2006) apresenta um índice de 508 milhões de falantes de inglês como segunda língua (citando uma referência de Nicholas Ostler, 2005, publicada em 'Empires of the Word: a language history of the world.' London, Harper Collins) - mas ele previa que esse número poderia ser maior que 1 bilhão. No livro "O futuro do inglês" ('The future of English', British Council, 2000) Graddol apresenta estimativa (em milhões) para as três seguintes categorias: inglês como primeira língua (375), inglês como segunda língua (375) e inglês como língua estrangeira (750). A questão permanece no nível em que se qualificam as pessoas da última categoria: nível de falantes eficientes.

⁸ A outra sendo a comunicação que é também mantida por meio das línguas.

⁹ Ter uma existência digital começa pela localização da língua. Esse conceito se refere a codificação de seu alfabeto e isso acarreta em uma série de outros requisitos desde a existência de *softwares* relacionados a essa língua (a exemplo dos corretores gramaticais e dicionários) à existência de conteúdos significativos. Vide Diki Kidiri [2008].

consideremos os fatos a seguir em torno das línguas (Lewis, 2009; Crystal, 2006):

- 6 línguas (inglês, árabe, espanhol, **português** e russo) são as línguas oficiais em 60% dos países.
- 1.3% das línguas (85) são faladas por mais de 10 milhões de pessoas cada e juntas representam 78% da população mundial.
- 50% de todas as línguas são faladas por menos de 10.000 pessoas.
- 25%
- Uma língua desaparece em média a cada dois meses.
- Menos de 10% das línguas possuem sua forma escrita¹⁰

A questão da presença das línguas na internet pode ser verificada pelo cruzamento de dados do *Ethnologue* sobre as línguas¹¹ com os dados do ITU referentes aos acessos à internet¹². Parece que há uma correlação estatística entre países com grande diversidade linguística e países com baixa penetração na internet¹³.

Então, o que sabemos exatamente sobre a presença das línguas na internet? A situação atual é de certa forma paradoxal e frustrante! Por muitos anos a temática alcançou somente um interesse marginal e poucos pioneiros produziram indicativos experimentais (Pimienta-2009-2). Agora o tema da diversidade linguística vem ganhando mais atenção e, enquanto alguns dos esforços passados perderam vigor, há grande estímulo para avaliar profundas mudanças que ocorreram no universo do conteúdo. Ao buscar dados sobre o número de usuários da internet por línguas, o *Globalstats*¹⁴ apresentou índices alguns anos atrás e era claro quanto à metodologia usada para coletá-los. Desde 2007, o *Internet world stats*¹⁵ o substituiu assumindo a liderança, no entanto os dados agora são limitados às "dez mais" línguas. Seguem os dados relativos ao ano de 2011:

¹⁰ http://www15.gencat.cat/pres_casa_llengues/AppJava/frontend/sabiesque_detall.jsp?id=18&idioma=5

¹¹ <http://www.ethnologue.com/web.asp>

¹² <http://www.itu.int/ITU-D/ict/statistics/>

¹³ Outras duas correlações que têm levantado questionamento são entre a grande biodiversidade e a alta diversidade linguística e outra entre quaisquer uma das duas e a questão da pobreza. Em outras palavras, a parte rica do planeta é rica em informação mas é linguisticamente pobre.

¹⁴ A fonte URL (<http://global-reach.biz/globstats/index.php3>) desapareceu mas o motor "Wayback" do archive.org permite que acessemos as últimas memórias para armazenar as mais recentes informações memorizadas de 2008 e que mostra os dados produzidos em 2004:

<http://web.archive.org/web/20041019013615/www.global-reach.biz/globstats/index.php3>

¹⁵ <http://www.internetworldstats.com/>

Tabela 1: Número de usuários da internet por língua, 31 de maio de 2011

As 10 línguas mais usadas na rede (Número de usuários da internet por língua)

As dez maiores línguas na internet	Usuários da internet por língua	Penetração na internet por língua	Crescimento na internet (2000 - 2011)	Usuários da Internet % do total	População mundial para tal língua Estimativa de 2011)
<u>inglês</u>	565,004,126	43.4 %	301.4 %	26.8 %	1,302,275,670
<u>chinês</u>	509,965,013	37.2 %	1,478.7 %	24.2 %	1,372,226,042
<u>espanhol</u>	164,968,742	39.0 %	807.4 %	7.8 %	423,085,806
<u>japonês</u>	99,182,000	78.4 %	110.7 %	4.7 %	126,475,664
<u>português</u>	82,586,600	32.5 %	990.1 %	3.9 %	253,947,594
<u>alemão</u>	75,422,674	79.5 %	174.1 %	3.6 %	94,842,656
<u>árabe</u>	65,365,400	18.8 %	2,501.2 %	3.3 %	347,002,991
<u>francês</u>	59,779,525	17.2 %	398.2 %	3.0 %	347,932,305
<u>russo</u>	59,700,000	42.8 %	1,825.8 %	3.0 %	139,390,205
<u>coreano</u>	39,440,000	55.2 %	107.1 %	2.0 %	71,393,343
As 10 maiores línguas	1,615,957,333	36.4 %	421.2 %	82.2 %	4,442,056,069
Outras línguas	350,557,483	14.6 %	588.5 %	17.8 %	2,403,553,891
<u>Total mundial</u>	2,099,926,965	30.3 %	481.7 %	100.0 %	6,930,055,154

Fonte : <http://www.internetworldstats.com/stats7.htm>

No que tange à estrutura do universo de conteúdos, em termos de línguas, a regra foi a existência de índices conflitantes entre os anos de 1997 e 2007 (Pimienta-2009-2). Algumas fontes de dados indicaram uma estável (e exagerada) presença do inglês (cerca de 80% ao longo da década). Outras fontes de dados, com mais amostras, apontaram para um declínio constante de 80% para 45% no mesmo período. Nos últimos anos, a situação tem sido extremamente complexa principalmente devido às mudanças estruturais e quantitativas da rede. Enquanto motores de busca foram capazes de indexar¹⁶ uma parte significativa da rede no período de 1996 a 2007 (cerca de 80%), desde 2007 vem apresentando declínio de, provavelmente, no mínimo 10% e está se tornando pior a cada dia. Enquanto muitas fontes tentaram quantificar o tamanho da rede até aproximadamente o ano de 2008 (vide PIMIENTA, 2008), desde então ninguém parece ousar prover mais informações, ainda que seja uma aproximação para dados qualitativos, embora dados quantitativos existam sob a forma de usuários da internet, nomes de domínios registrados, *hosts* e endereços de IP.

¹⁶ Uma página indexada é uma página cujo conteúdo foi analisado e a informação foi armazenada para posterior acesso pelos motores de busca.

Tabela 2: O tamanho da Internet

INTERNET ELEMENT FIGURE	2008	2011
Usuários da Internet - (Fonte: <i>Internet world stats</i>)	1.4 bilhões	2.1 bilhões
Domínios ativos registrados (Fonte: http://news.netcraft.com/archives/web_server_survey.html)	70 milhões	160 milhões
Domínios registrados (Fonte: http://news.netcraft.com/archives/web_server_survey.html)	140 milhões	463 milhões
<i>Internet hosts</i> (Fonte : http://www.domaintools.com/internet-statistics/country-ip-counts.html)	500 milhões	3500 milhões
Páginas da rede ¹⁷	140 bilhões ¹⁸	?????
Páginas indexadas da rede	20–40 bilhões	Provavelmente o mesmo que em 2008 ¹⁹

Além da confirmação do crescimento exponencial da internet, o que pode se depreendido dessa tabela é que o tamanho dos indexes vinha alcançando uma assíntota até 2007 e a proporção de páginas indexadas passa por um declínio constante desde então (de 80% a provavelmente, menos de 10%).

Por que esses dados relativos às ferramentas de busca são tão relevantes para a temática da diversidade linguística na Internet? Eles têm impacto direto no estudo da diversidade linguística porque alguns dos métodos para avaliar a presença das línguas na rede consistem nos motores de busca, que eram capazes de contar ocorrências de vocábulos. Eles também têm um impacto indireto porque os outros métodos de busca consistem em “rastrear a rede”²⁰ e aplicar um algoritmo de reconhecimento de línguas para cada página. Se as poderosas companhias desenvolvendo motores de busca desistem de navegar por todo o universo da internet, de que forma uma seção de busca de alguma outra organização irá investir em alguma tarefa lucrativa? Finalmente, indexar é também relevante para a temática devido a uma dialética negativa. A ausência de uma página nos índices dos motores de busca é praticamente a garantia da sua invisibilidade. Além disso, com a decrescente porcentagem de indexação, que não é aplicada de forma igualitária entre as línguas, existe uma tendenciosidade nos dados que possui implicações críticas para as línguas que não são consideradas globalmente relevantes²¹. Em síntese: a rede se tornou praticamente infinita

¹⁷ Se algo é considerado “rede invisível” (páginas dinâmicas principalmente) espera-se que seja de 100 a 500 vezes maior de acordo com Bergman, M. The Deep Web: Surfacing Hidden Value - <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104>, e então o índice poderá alcançar um trilhão

¹⁸ Em 2008 um novo motor de busca foi lançado com o nome de CUIL (<http://en.wikipedia.org/wiki/Cuil>) o qual supõe ter um indexo de tamanho próximo ao do tamanho da rede, oferecendo o índice de 120 bilhões de páginas.

¹⁹ <http://www.worldwidewebsite.com/> apresenta índices entre 20-50 bilhões de páginas indexadas em 2011. O Google, por sua vez, não tem fornecido índices públicos (<http://www.google.com/help/indexsize.html>). <http://dejanseo.com.au/measuring-size-of-googles-index/> apresenta dados de 25 bilhões.

²⁰ “Crawling the web” (Rastreamento da rede) se refere ao mecanismo usado, por exemplo, pelos motores de busca para criar suas bases de dados indexados para sistematicamente ler cada página de cada sítio visando analisar seus conteúdos. O tamanho da rede é tão grande que alguns truques são usados para evitar a leitura de cada uma das páginas (por exemplo, limitando a profundidade da inserção na árvore de relações [*links*] de um mesmo sítio).

²¹ Perceba que seria benéfico ser capaz de medir a inclinação linguística dos motores de busca e também se essa situação pudesse abrir caminhos para a criação de motores de busca dedicados às línguas ou famílias de línguas.

Daniel Pimienta

e novos métodos são necessários para que se possa compreender a sua natureza, a estrutura de seu conteúdo e particularmente em termos de línguas. No entanto, não se trata somente de tamanho: a natureza da rede por si só mudou consideravelmente à medida que o conceito de páginas da internet evoluiu de estático (como era no início) para dinâmico²² (o que se tornou a regra). Abordagens totalmente novas são necessárias para vasculhar em sua complexa estrutura e para desenvolver indicadores significativos e acurados da diversidade linguística. A pesquisa dessas ferramentas e metodologias não somente forneceria um caminho para caracterizar a vasta substância da rede e melhor compreender o universo do conteúdo da Internet; também abriria novos caminhos para novos modelos de motores de busca (Search 2.0) capazes de melhor lidar com a info-diversidade do universo de conteúdos. Isso tudo é tanto mais importante porque o universo peculiar de conteúdos está agora ameaçado por muitos fatores, incluindo o spam, cópias e ausência de uma cobertura adequada de línguas.

Nesse contexto, caracterizado pela falta de produção de indicadores estáveis, confiáveis e sustentáveis para a diversidade linguística, e até que projetos como o DILINE²³ produza indicadores necessários, a única opção é confiar nos índices fornecidos pelas três mais importantes fontes: Internet World Stats, para índices de usuários por língua; o Projeto Observatório das Línguas (Language Observatory Project)²⁴ e FUNREDES/Union Latina²⁵, para dados relativos à páginas da rede por língua.

No que se refere ao primeiro período, de 1997-2007, os estudos de referência (Pimienta, 2009-2) mostraram alguns fatos que serão descritos adiante. Em termos de prevalência de línguas, o inglês não foi tão crucial quanto indicado; a porcentagem de páginas em inglês decresceu de forma regular, de 80% em 1998 para 40% no ano de 2007. Essa tendência acompanhou, com certo grau de paralelismo, a porcentagem de usuários da internet falantes de inglês (fonte: InternetWorldStats); estimado para estar acima de 60% em 1998 e decrescendo para 32% em 2008, sendo o índice atual de 26,8% (dezembro de 2011)²⁶.

Os dados mostraram uma primeira etapa (1998-2003) quando a razão entre conteúdos e usuários teve pequena variação em torno das línguas. Esse quadro ofereceu as bases para argumentos em favor do fomento de políticas públicas para o provimento de acesso a usuários como uma forma natural para a geração de conteúdos de forma linguisticamente proporcional. Em uma segunda etapa (2004-2008), essa razão entre os conteúdos e usuários sofreu decréscimo mostrando que a última leva de usuários da internet era mais frequentemente de consumidores de conteúdos do que de produtores dos mesmos. Isso pode ser explicado pelo fato de que os primeiros usuários da internet possuíam alto grau de letramento digital. Naquele

²² Uma página estática tem a propriedade de aparecer sempre idêntica quando acessada por um *browser* (um programa como o Internet Explorer ou Mozilla) através da indicação da URL (Unified Resource Locator – Localizador de Fontes Unificado) no campo para os endereços das páginas, (uma expressão como <http://www.w3.org/Addressing/URL/url-spec.txt>). Uma página dinâmica tem seu conteúdo em evolução ao longo do tempo e dependendo do contexto (a exemplo dos acessos a base de dados).

²³ Sob a liderança da MAAYA, respeitados Centros de Pesquisa estão unindo esforços com a Unesco, ITU, OIF, União Latina e Funredes para abrir o campo para projetos de pesquisa avançada que buscariam sistematizar a criação de indicadores da diversidade linguística no mundo digital. O consórcio está atualmente em fase de definição para conseguir fundos da *European Union Reserch funds* (7th Programa Modelo) em posição preferencial devido ao tamanho do investimento solicitado e as inovações em pesquisa implicadas. Muitos dos parceiros, especialmente organizações internacionais e civis, irão manter a pressão até que os orçamentos sejam feitos, já que políticas públicas, assim como a economia digital, irá se beneficiar com o desenvolvimento e manutenção de tais indicadores com vistas ao progresso. Vide <http://dilinet.org>.

²⁴ O Projeto Observatório de Línguas : <http://www.language-observatory.org/>

²⁵ Observatório de Línguas e Culturas na Internet da Funredes:
<http://funredes.org/lc> ou União Latina http://dtil.unilat.org/LI/2007/index_fr.htm

²⁶ With Chinese very close at 24.2%.

momento, os mais recentes eram conduzidos virtualmente para a chamada *netiquette*²⁷ e para o comportamento digital. Na era atual, há um argumento válido em prol de políticas de desenvolvimento do letramento digital de novos usuários da Internet e estímulo para que esses usuários sejam produtores assim como consumidores do que está disponível na rede.

Uma mensagem muito importante tem sido estabelecida pelos primeiros estudos (da Funredes/União Latina assim como os da LOP) sobre a exclusão digital no hemisfério sul. Ela evidenciou que a lacuna de conteúdo foi bem maior que a lacuna de acesso para línguas internacionais e várias ordens de magnitude para línguas locais, conforme demonstrado nos seguintes índices:

- 4% dos acessos globais são derivados da África (Fonte: Internet world stats 2007)
- 0.6 % dos sites em francês possuem base na África (Fonte: Funredes/União Latina 2007)
- 0.6% dos sites em inglês possuem base na África (Fonte: Funredes/União Latina 2007)
- A porcentagem de páginas em línguas locais africanas varia de 0.06% a 0.0006% dependendo da língua (Fonte: LOP 2007)

A exclusão digital deve então ser muito mais uma questão de conteúdos e de línguas do que de acessos propriamente ditos. Esse ponto é um argumento extremamente importante em favor de políticas de inclusão digital que não se restringem à questão do acesso, mas que também enfoquem em conteúdos locais (e indiretamente na educação para fomentar novos produtores de conteúdos, um processo que se inicia pelo estímulo ao letramento digital)²⁹.

No que concerne ao segundo período mencionado anteriormente, há diversas descobertas que apontam para diversas tendências possíveis, apesar da ausência de indicadores de produção estáveis:

- A evolução recente da Rede mostra um número considerável de usuários chineses da internet se sobrepondo ao número total de usuários americanos em julho de 2008³⁰.
- Após o movimento revolucionário no norte da África e Oriente Médio, o espaço da língua árabe no Facebook está crescendo rapidamente e, se a tendência persistir, irá logo sobrepor-se ao inglês³¹ nessas regiões.
- A Wikipedia é o espaço da internet com maior diversidade linguística abrangendo aproximadamente 271 línguas diferentes que nela contribuem com artigos³².
- O índice máximo de 271 para a Wikipedia é comparado aos índices das 70 línguas que o Mozilla suporta, das 67 do Facebook, das 63 do Internet Explorer, 51 do Google Tradutor, 50 do Blogger, 19 do YouTube, 6 do Flickr e 4 do LinkedIn³³.

²⁷ *Netiquette* é um neologismo que se refere à comportamentos aceitáveis na rede. Mais detalhes podem ser encontrados em <http://www.albion.com/netiquette/>.

²⁸ Esse dado felizmente aumentou para 11.4% em 2011 mas há grande probabilidade de que a exclusão de conteúdos não mude muito.

²⁹ O que está em questão é o conceito de "pertença" da tecnologia que não deve ser confundida com a questão do acesso. Vide (Pimienta, -2009-1) para desenvolvimento mais detalhado do conceito.

³⁰ <http://www.nytimes.com/2008/07/26/business/worldbusiness/26internet.html>

³¹ <http://www.slateafrique.com/17731/sur-facebook-arabe-depasse-anglais>

³² Além disso, a Wikipedia manteve uma maravilhosa compilação de dados estatísticos sobre as línguas: http://en.wikipedia.org/wiki/Wikipedia:Multilingual_statistics. Aparentemente, outro indicativo de tempos de mudança, esta página não vem sendo atualizada desde outubro de 2008 em algumas partes e desde dezembro de 2009 para dados diretos em cada uma das línguas. No entanto, a Wikipedia está oferecendo agora uma página nova e muito interessante para dados estatísticos, (<http://stats.wikimedia.org/>) que inclui alguns dados sobre as línguas (<http://stats.wikimedia.org/#fragment-12>).

³³ Todas essas fontes correspondem ao ano de 2010 e se referem à interface (não ao conteúdo).

Daniel Pimienta

- Um estudo produzido pela União Latina em 2010³⁴ desenvolveu uma perspectiva da evolução da proporção das línguas em várias áreas da Internet, a partir da coleta de grande número de fontes virtuais demográficas e estatísticas³⁵. No entanto, havia uma lacuna no que se refere a uma unidade metodológica que desse conta das fontes e dados em níveis variados de confiabilidade. Por isso, os resultados só podem ser visualizados como indicativos das possíveis tendências.

³⁴ Este estudo não foi publicado ainda e alguns dos excertos apresentados adiante possuem a permissão de Daniel Prado.

³⁵ Segue seleção de algumas fontes relevantes usadas para tal estudo:

<http://socialmediastatistics.wikidot.com/>

http://www.w3schools.com/browsers/browsers_firefox.asp

http://www.readwriteweb.com/archives/international_blogosphere.php

<http://technorati.com/state-of-the-blogsphere/>

<http://www.mathaba.net/news/?x=408348>

<http://blobalize.typepad.com/micro/statistics/>

<http://www.sysomos.com/insidetwitter/>

<http://en.wordpress.com/stats/>

<http://www.p2pon.com/2010/02/19/10-million-downloads-of-pirated-games-in-december-alone-says-esa/>

http://journal.webscience.org/209/2/websci09_submission_115.pdf

<http://www.email-marketing-reports.com/metrics/email-statistics.htm>

<http://www.sifry.com/alerts/archives/000433.html> and <http://www.sifry.com/alerts/archives/000436.html>

<http://mashable.com/2010/02/24/half-messages-twitter-english/>

<http://www.vincegolanco.com/2010/very-interesting-social-media-statistics-facebook-twitter-flickr-linkedin-etc/>

O espaço da Internet	Inglês %	Outras línguas
Live Journal (Blog)	55%	Espanhol 6% - Russo 18% - Português 1%
WordPress (Blog)	66%	Espanhol 9% - Portugues 7%
Facebook (Rede Social)	35% ¹	Espanhol 10% - francês 10% - Português 6%
Tweeter (Rede Social)	50% ²	Japonês 14% - português 9% - Espanhol 4% - malaio 6%
Technorati (blog)	25%	japonês 25% - chinês 31% - português 3%
LinkedIn (Rede social)	50%	híndi 15% - espanhol 6% -
Second Life (Rede social)	39%	francês 17% - alemão 10% - espanhol 7% - holandês 7%
Yandex (Motor de busca) – Tráfego	1%	Russo 95%
Baidu (Motor de busca) – Tráfego	1%	Chinês 96%
Yahoo (Motor de busca) – Tráfego	36%	híndi 9% - espanhol 7% - árabe 5% - português 2%
SAPO		O português é uma das poucas línguas com um motor de busca específico
Bing (Motor de busca) – Tráfego	40%	espanhol 10% - chinês 7% - francês 6% - híndi 5% - português 2%
Google (Motor de busca) – Tráfego	38%	híndi 9% - espanhol 6% - árabe 5% - chinês 4% - Portuguese 3%
Google (Motor de busca) – Conteúdo	39%	alemão 7% - espanhol 6% - francês 6% - italiano 4% - português 3%
MSN (Motor de busca) – Tráfego	25%	espanhol 18% - português 7% - japonês 6% - francês 6% - árabe 6%
MSN (Motor de busca) – Conteúdo	50%	alemão 6% - francês 5% - espanhol 5% - italiano 3%
Exalead (Ferramenta de busca) – Tráfego	13%	híndi 19% - francês 17% - alemão 11% - espanhol 7% - chinês 6%
Exalead (Motor de busca) – Conteúdo	48%	francês 7% - alemão 4% - espanhol 4% - italiano 3%
MegaUpload (P2P)	13%	espanhol 22% - português 11% -

³⁶ Uma referência mais atualizada indica o inglês com 52% e apresenta as 10 línguas mais usadas no Facebook, surpreendendo pela posição do turco e indonésio após o espanhol e o francês. Fonte: <http://www.insidefacebook.com/2010/05/24/facebooks-top-ten-languages-and-who-is-using-them/>.

³⁷ Um estudo mais recente indica o inglês abaixo de 50% e apresenta, na sequência, o japonês, português, o malaio e o espanhol. (Vide: http://semicast.com/static/downloads/Semicast_Half_of_messages_on_Twitter_are_not_in_English_20100224_fr.pdf)

Daniel Pimienta

Google (Motor de busca) – Tráfego	38%	híndi 9% - espanhol 6% - árabe 5% - chinês 4% - Portuguese 3%
Google (Motor de busca) – Conteúdo	39%	alemão 7% - espanhol 6% - francês 6% - italiano 4% - português 3%
MSN (Motor de busca) – Tráfego	25%	espanhol 18% - português 7% - japonês 6% - francês 6% - árabe 6%
MSN (Motor de busca) – Conteúdo	50%	alemão 6% - francês 5% - espanhol 5% - italiano 3%
Exalead (Ferramenta de busca) – Tráfego	13%	híndi 19% - francês 17% - alemão 11% - espanhol 7% - chinês 6%
Exalead (Motor de busca) – Conteúdo	48%	francês 7% - alemão 4% - espanhol 4% - italiano 3%
MegaUpload (P2P)	13%	espanhol 22% - português 11% - francês 11% - japonês 8% - italiano 5%
RapidShare (P2P)	11%	espanhol 14% - árabe 14% - híndi 8%
DepositFiles (P2P)	10%	russo 26% - espanhol 16% - árabe 8% - português 6%
FileFactory (P2P)	11%	Árabe 16% - espanhol 11% - português 8% - híndi 7% - japonês 7%
HotFile (P2P)	13%	árabe 11% - japonês 11% - espanhol 8% - híndi 7%
Skype	9%	português 10% - chinês 8% - francês 5% - alemão 5%
Firefox Download	23%	alemão 11% - espanhol 8% - híndi 7% - francês 7%
Wikipedia (artigos)	23%	alemão 7% - francês 6% - polonês 5% - japonês 5% - espanhol 3.8% - português 3.6%
YouTube	27%	espanhol 12% - francês 6% - híndi 5% - alemão 5% - português 4%

- A maioria dos dados são baseados no fluxo de dados por país usando o <http://webboar.com> e não diretamente sobre línguas e conteúdos.

- As ferramentas de busca em que se indica "conteúdo" são derivadas da Funredes em 2007 e não foram publicadas porque não foram consideradas confiáveis o suficiente; no entanto são confiáveis o suficiente para uma aproximação da porcentagem de páginas em dada língua nos índices da ferramenta de busca em dada língua. (somente o catalão, inglês, francês, alemão, italiano, **português** e romeno foram computados).

- Algumas vezes os índices relativos ao russo incluem outros países como a Ucrânia e a Bielorrússia.

Esses índices confirmam uma forte tendência da internacionalização da internet e um relativo declínio do inglês o qual foi percebido no período precedente. A existência bem sucedida de novas ferramentas de busca direcionadas para línguas específicas (chinês ou russo) poderia iniciar uma nova tendência no contexto dos motores de busca genéricos que são levados a pender seus índices em direção ao inglês como consequência da crescente limitação do rastreamento do universo virtual inteiro.

3 - Mais dados referentes ao português no mundo digital

Tabela 3. Falantes de primeira e segunda língua da maioria das línguas faladas

	Primeira língua (falantes em milhões, número de países) Fonte: Ethnologue 4/2012	Primeira e segunda língua Fonte: União Latina 2008	
1	chinês (1213, 31)	Chinês	> 1000
2	espanhol (329, 44)	Ingles	> 500
3	inglês (328, 112)	Hindi	> 500
4	árabe (221, 57)	Espanhol	> 200
5	híndi (182, 20)	Russo	> 200
6	bengali (bengalês) (181, 10)	árabe	>200
7	português (178, 37)	bengali	
8		português	

O português pode ser considerado a quinta língua do mundo em termos de falantes de língua materna, ao lado do híndi e do bengali já que as diferenças são periféricas. No entanto, se considerarmos falantes de primeira e segunda língua a posição da língua segue em oitavo lugar.

Tabela 4. As dez línguas mais comuns de usuários da Internet

Língua	Usuários	Penetração por Língua	Crescimento 2000 - 2011	% Total	Usuários em 2011
inglês	565 milhões	43.4%	301.4%	26.8%	1,302 milhões
chinês	510 milhões	37.2%	1,478.7%	24.2%	1,372 milhões
espanhol	165 milhões	39.0%	807.4%	7.8%	423 milhões
japonês	99 milhões	78.4%	110.7%	4.7%	126 milhões
português	83 milhões	32.5%	990.1%	3.9%	254 milhões
alemão	75 milhões	79.5%	174.1%	3.6%	95 milhões
árabe	65 milhões	18.8%	2,501.2%	3.3%	347 milhões
francês	60 milhões	17.2%	398.2%	3.0%	348 milhões
russo	60 milhões	42.8%	1,825.8%	3.0%	139 milhões
coreano	40 milhões	55.2%	107.1%	2.0%	71 milhões
As 10 mais	1,615 milhões	36.4%	421.2%	82.2%	4,442 milhões
TOTAL	2,100 milhões	30,3%	481,7%	100,0%	6,930 milhões

Daniel Pimenta

O português é a quinta língua mais usada em termos de usuários da internet com penetração modesta - menos de um terço - o que indica a possibilidade para mais crescimento.

Tabela 5. Portuguese speaking Internet countries – Fonte: Internet World Stats

Usuários da internet falantes de português					
Países	População (Estimativa de 2011)	Usuários da internet (31/12/2011)	Penetração (% da População)	Usuários (% da Tabela)	Facebook 31/12/2011
Angola	13,338,541	744,195	5,6%	0,9%	322.300
Brasil	203,429,773	79,245,740	38,9%	91,5%	35,158,740
Cabo Verde	516,100	148,800	28,8%	0,2%	148,800
Guiné Bissau	1,596,677	37,123	2,3%	4,3%	Não disponível
Moçambique	22,948,858	975,395	4,2%	1,1%	186,460
Portugal	10,760,305	5,455,217	50,7%	6,3%	4,174,000
São Tomé e Príncipe	177,506	31,012	17,3%	0,0%	3,320
Timor Leste	1,177,834	2,361	0,2%	0,0%	Não disponível
TOTAL	253,947,594	86,639,843	34,1%	100,0%	39,993,620

A exclusão digital é claramente visível nesta tabela onde alguns dos países de língua oficial portuguesa tem índices de penetração na rede inferiores a 10%.

Tabela 6. Sites em línguas Latinas

Páginas web em	1998	2000	2001	2002	2003	2004	2005	2007
INGLÊS	75 %	60 %	52 %	50 %	49 %	45 %	45 %	45 %
ESPAÑHOL	2,53 %	4,79 %	5,50 %	5,80 %	5,31 %	4,08 %	4,60 %	3,80 %
FRANCÊS	2,81 %	4,18 %	4,45 %	4,80 %	4,32 %	4,00 %	4,95 %	4,41 %
PORTUGUÊS	0,82 %	2,25 %	2,55 %	2,81 %	2,23 %	2,36 %	1,87 %	1,39 %
ITALIANO	1,50 %	2,62 %	3,08 %	3,26 %	2,59 %	2,66 %	3,05 %	2,66 %
ROMENO	0,15 %	0,21 %	0,18 %	0,17 %	0,11 %	0,11 %	0,17 %	0,28 %
ALEMÃO	3,75 %	2,85 %	6,75 %	7,21 %	6,80 %	7,13 %	6,94 %	5,90 %
CATALÃO								0,14 %
OUTRAS LÍNGUAS	13,44 %	22,20 %	23,68 %	25,97 %	29,65 %	31,32 %	33,43 %	36,54 %

Comparado a outras línguas latinas, o português sofre uma redução da porcentagem de conteúdos no período de medição. Mais que uma questão de acesso, a lacuna a ser preenchida deve ser percebida em termos de conteúdos.

Tabela 7. Produção de páginas em português (produtividade)

	2007	2005
Brasil	71% (0.90)	71 % (0.95)
Portugal	15 % (0.98)	17 % (1.0)
EUA	4 % (5.0)	8 % (5.4)
Espanha	3.8 % (3.7)	2.3 % (1.2)

Fonte: Funredes/União Latina

É interessante notar que os índices altos e produtividade dos EUA e Espanha. Índices mais detalhados por país mostram a profundidade da exclusão de conteúdos já que a França, sozinha, está produzindo em 11/2007 mais páginas em português (0.21 % do total) que a Angola, Cabo Verde, Guiné Bissau, Moçambique, São Tomé e Príncipe e Timor Leste juntos (0,16%)!

Tabela 8. Tabela de Indicadores da Internet 2008

	Inglês	Espanhol	Francês	Italiano	Português.	RUM ?	Alemão	CAT	TOT.
Falantes (milhões)	670	400	130	60	205	30	120	9	6600
Falantes (% população mundial)	10.1%	6.1%	2.0%	0.9%	3.1%	0.5%	1.8%	0.1%	130%
Internautas (milhões)	366	102	58	31	47	5	59	2	1154
Internautas (%)	54.6%	25.4%	44.9%	52.3%	23.1%	16.5%	49.1%	23.1%	17.5%
Internautas % da população mundial	5.5%	1.5%	0.9%	0.5%	0.7%	0.1%	0.9%	0.0%	17.5%
% Internautas por língua	32%	9%	5%	3%	4%	0%	5%	0.2%	130%
% Páginas por língua.	45.0%	3.8%	4.4%	2.7%	1.4%	0.3%	5.9%	0.1%	100%
Produtividade por língua	1.42	0.43	0.87	0.98	0.34	0.66	1.16	0.74	1
Páginas por internauta	4.44	0.63	2.24	2.93	0.45	0.62	3.25	0.96	

Fonte: GlobaReach e Funredes/União Latina

Há uma clara indicação da debilidade do português em termos de conteúdos disponíveis na rede em comparação com outras línguas estudadas.

Tabela 9. Artigos e tráfego da Wikipedia 4/2012

	Línguas	# ARTIGOS Milhões	# USO Milhões Acessos por hora
1	inglês	4	10.5
2	francês	1.2	1
3	italiano	0.9	0.7
4	alemão	0.9	1.4
5	russo	0,9	1.1
6	japonês	0.8	1.5
7	português	0,7	0.5
8	holandês	0,7	0.2
9	espanhol	0,6	1.5
10	polonês	0.6	0.5

Fonte: Wikipedia³⁸

O espanhol se sobrepunha ao português há três anos atrás. Esse é um indicativo promissor de crescimento de conteúdos em português nos últimos anos.

Tabela 10. Twitter Os 5 maiores mercados globais por busca

País	Dados estatísticos Porcentagem	
Indonésia	Junho 2010	20.8%
	Dezembro 2010	19.0%
Brasil	Junho 2010	20.5%
	Dezembro 2010	21.8%
Venezuela	Junho 2010	19.0%
	Dezembro 2010	21.1%
Países baixos	Junho 2010	17.7%
	Dezembro 2010	22.3%
Japão	Junho 2010	16.8%
	Dezembro 2010	20.0%

Fonte: <http://www.billhartzer.com/pages/comscore-twitter-latin-america-usage/>

Tabela 11. As dez maiores línguas do Facebook

As 10 maiores línguas do Facebook	
Total de falantes por língua	
inglês	213,237,080
espanhol	61,209,600
francês	23,504,300
turco	21,911,00
indonésio	20,481,260
italiano	16,222,700
alemão	11,265,320
chinês	9,645,140
português	6,119,680
árabe	3,457,160

Fonte: gold. insidenetwork.com/Facebook
Dados do Facebook.com de 24 de maio de 2010

INSIDE NETWORK

³⁸ Vide <http://stats.wikimedia.org/EN/TablesCurrentStatusVerbose.htm>

Daniel Pimenta

Os dois maiores “êxitos” do português estão mais relacionados ao uso do que na produção de conteúdo.

4 - Perspectiva histórica e tendência sobre as línguas na internet

A tabela a seguir sintetiza uma perspectiva histórica da evolução das redes em relação à diversidade linguística.

Tabela 12: síntese

Era da Internet	Enquadre temporal	Usuários	características	LANGUAGE WISE
Pré-história	1970 - 1990	Até dois milhões	<ul style="list-style-type: none">- BITNET, UUCP e outras redes- Usuários de companhias de Tecnologia da Informação, pesquisadores, acadêmicos e ativistas.- De débeis terminais de acesso ao acesso ao mainframe (computadores de grande porte) até o nascimento do PC (Personal Computer – Computador Pessoal)	<ul style="list-style-type: none">- Comunidade global de profissionais com grande domínio de inglês.- ASCII não estendido significa nenhum diacrítico como em inglês.
O nascimento da Rede	1990 - 1995	Até 15 milhões	<ul style="list-style-type: none">- O índice de profissionais de internet perpassa pesquisadores e acadêmicos.- Difusão do PC (Computador pessoal)	<ul style="list-style-type: none">- MIME- Nascimento do Unicode- inglês acima de 80%
Rede 1.0	1995 - 2005	Até 1 bilhão	<ul style="list-style-type: none">- A internet se torna mídia de massa- Poderosos motores de busca- Multiplicação de pequenas empresas	<ul style="list-style-type: none">- o inglês declina para 50%- Aumento de línguas ocidentais buscando equilíbrio³⁹ com o inglês

³⁹ Equilibrado em proporção relativa de seus falantes com a variação dependendo das línguas; francês, alemão e italiano sendo representados acima da média, em oposição ao português e espanhol.

Rede 2.0	2005 - 2010	Até dois bilhões	<ul style="list-style-type: none"> - Aumento de monopólios - Redução do tamanho dos indexes dos motores de busca - Aumento dos vídeos - Aumento da internet móvel 	<ul style="list-style-type: none"> - inglês em torno de 33% - Aumento de línguas asiáticas - Aumento do árabe - Aumento de localização
A próxima era	2010 - 2015	Até 3 bilhões	<ul style="list-style-type: none"> - Aumento de motores e busca para línguas específicas. - Aumento de bibliotecas digitais para línguas específicas. - Aumento de conteúdos não textuais - Aumento de interfaces de voz. - Internet de objetos 	<ul style="list-style-type: none"> - IDN⁴⁰ generalizado - Assíntota do inglês até 25% - Aumento do árabe - Aumento de línguas indígenas com grande número de falantes⁴¹ - Aumento do uso de <i>softwares</i> de tradução. - Localização de centenas de línguas
A internet no futuro	2015 - 2030	75% da população mundial conectada	<ul style="list-style-type: none"> - Telas em toda parte - teclados dando lugar a interfaces de voz. - Disseminação de <i>softwares</i> de tradução. - Motores de busca de nova geração - Bibliotecas virtuais de nova geração 	<ul style="list-style-type: none"> - aumento de políticas linguísticas voltadas para a internet - conteúdos equilibrados em diversas línguas-saberes. - Equilibradas ferramentas linguísticas. - Localização de milhares de línguas. - Ferramentas públicas de suporte à aprendizagem de línguas e multilinguismo.

⁴⁰ *Internationalized domain name* (IDN) é nome de um domínio da internet que usa um diferente script do latim sem traços diacríticos (o mesmo usado pelo inglês), assim permite escrever URL's em, por exemplo, árabe, chinês ou francês com traços diacríticos (exemplos: çætlà.fr ou baño.es).

⁴¹ A exemplo do quechua, aymara ou guarani na América Latina; o crioulo no Caribe; o suaíle ou hause na África.

5 - Conclusão

Garantir a presença total de uma língua no mundo digital é um complexo processo que demanda uma perspectiva holística desde sua codificação, produção de teclados, desenvolvimento de softwares nessas línguas (processadores de textos e corretores gramaticais, por exemplo) até o letramento digital e produção de conteúdos (Diki Kidiri, 2008). Além desses elementos, a existência de livrarias virtuais para a literatura nas línguas e a oferta de recursos online para a aprendizagem e processamento da língua (como os dicionários, orientações gramaticais, glossários) são também extremamente importantes. Após uma língua ser identificada, a maior questão e talvez a mais difícil de lidar é como prover conteúdos nessa língua.

A que uso dar acesso se não há conteúdos na língua dos usuários; ao conceder acesso livre a quaisquer conteúdos em qualquer lugar não estamos nós somente aumentando os riscos da aculturação pela internet?

Não muitas línguas têm sido objeto de políticas públicas visando garantir sua revitalização no mundo real (catalão e hebraico são duas histórias bem sucedidas). No entanto, o objetivo nesse contexto é desenvolver políticas públicas que deem suporte para línguas especificamente para o mundo digital. Este é um novo campo de atividade com pouca experiência documentada⁴². A maioria das línguas apresenta alguma interseção com políticas públicas digitais; o que está em pauta, então, não é simplesmente a presença de novas línguas no mundo digital, mas também a promoção de línguas existentes na internet uma vez que uma forte presença dessas línguas no cyber espaço gera implicações econômicas, sociais e culturais.

Alguns especialistas advogam pela emergência de políticas linguísticas com enfoque na internet e oferecem um algoritmo para medir o peso das línguas⁴³ que abarca alguns indicadores relacionados à internet. Tais medidas podem servir de base para a criação e avaliação de políticas para as línguas na internet⁴⁴. (vide <http://www.portalingua.info/fr/poids-des-langues/>, em francês). Uma tarefa relacionada implicaria em sensibilizar responsáveis pelas línguas para a importância de tais medidas; É chegado o tempo para as línguas que estão em eminência de extinção⁴⁵.

Ao longo de sua evolução, a internet está adentrando uma nova era em que as questões de línguas e culturas estão sendo cruciais. Se as fronteiras da internet são as línguas, então a construção de uma internet sem fronteiras está se iniciando com grandes desafios que podem gerar novos tempos e fomentar a colaboração sem a qual nenhum progresso tangível pode ser facilmente alcançado.

Nessa nova era o português possui claramente um melhor papel e deve tentar obter um lugar no mínimo coerente com sua importância no mundo real e até mesmo batalhar pela referida importância social, econômica e cultural de sua representação no mundo digital.

⁴² Por exemplo, não há evidência no início de que a criação de um domínio linguístico dedicado ao catalão (em 2006, o primeiro e único desse tipo) tenha desencadeado a criação de conteúdos em tal língua. Seria interessante verificar se é o caso de que alguns anos após tenha gerado outros domínios linguísticos e culturais (Vide http://en.wikipedia.org/wiki/Proposed_top-level_domain).

⁴³ Esse aumento no momento não é favorável ao português que aparece em média na 15 posição (atrás do italiano) mesmo que seu peso relativo mude com diferentes parâmetros com vistas a melhores resultados.

⁴⁴ CALVET, 2009.

⁴⁵ Ver Tabela: *Categorização de línguas para requisitos de políticas no cyber espaço* em Pimienta, 2009-2.

Referências

Suzuki I., Mikami Y., et al.2002. 'A Language and Character Set Determination Method Based on N-gram Statistics', In ACM Transactions on Asian Language Information Processing, Vol.1, No.3, September 2002, pp.270-279.

Paolillo J., Pimenta D., Prado D., et al.2005. 'Measuring linguistic diversity on the Internet', UNESCO, 12/2005
http://portal.unesco.org/ci/en/ev.php-URL_ID=20882&URL_DO=DO_TOPIC&URL_SECTION=201.html

Crystal D.2006. 'Language and the Internet.', Second Edition, Cambridge University Press, 9/2006, (ISBN-13: 9780521868594 | ISBN-10: 0521868599)

Graddol D.2006. 'English Next', British Council
<http://www.britishcouncil.org/learning-research-english-next.pdf>

Ess, C. 2006. From Computer-Mediated Colonization to Culturally-Aware ICT Usage and Design, In P. Zaphiris and S. Kurniawan (eds.), Human Computer Interaction Research in Web Design and Evaluation. Hershey, PA: Idea Publishing.

Diki-Kidiri, M.2007. 'Securing a place for language in cyberspace', Unesco CI.2007/WS/1. 2007
<http://unesdoc.unesco.org/images/0014/001497/149786E.pdf>

Nandasara S.T. et al.2008. 'Analysis of the Asian Languages on the Web Based on N-gram Language Identification'. In The International Journal on Advances in ICT for Emerging Regions (ICTer), Volume 1, Issue 1, 2008

Pimenta D.2008. 'Accessing Contents', Chapter of Global Information Society Watch, APC, HIVOS, ITEM, 2008
<http://www.giswatch.org/gisw2008/thematic/AccessingContent.html>

Lewis, M.2009. 'Ethnologue: Languages of the World, Sixteenth edition.', Dallas, Tex.: SIL International.
<http://www.ethnologue.com/>

Pimenta D.2009-1. 'Digital divide, social divide, paradigmatic divide', in International Journal of Information Communication Technologies and Human Development, V1, N1, January-March 2009

Older version accessible at:
http://funredes.org/mistica/english/cyberlibrary/thematic/Paradigmatic_Divide.pdf

Pimenta D., Prado D., Blanco Á2009-2. 'Twelve years of measuring linguistic diversity in the Internet: balance and perspectives', Unesco,
<http://unesdoc.unesco.org/images/0018/001870/187016e.pdf>

Prado D., Pimenta D. , Lemoulinier A.2010. 'Diversité linguistique et cyberspace : état de l'art, enjeux et opportunités', Cosmopolis
http://agora.qc.ca/cosmopolis.nsf/Articles/no2010_1_Diversite_linguistique_et_cyberspace___etat_de_l?OpenDocument

Daniel Pimienta

ITU.2010. World Telecommunication/ICT Development Report 2010 – Monitoring the WSIS Targets. A mid-term review, Target 9, pp175-192
<http://www.itu.int/pub/D-IND-WTDR-2010>

Partnership on Measuring ICT for Development's Task Group on Measuring the WSIS Targets, Measuring the WSIS Targets - A statistical framework, Target 9, pp91-96, ITU, 2011
<http://www.itu.int/ITU-D/ict/partnership/wsistargets/index.html>

NET.LANG: Towards a multilingual cyberspace, C&F Editor, coordinated by MAAYA., 2012, ISBN 978-2-9158255-08-4
<http://net-lang.net/>



A LÍNGUA PORTUGUESA FACE AO CHOQUE TECNOLÓGICO DIGITAL

António Branco

Sobre o autor

Universidade de Lisboa
Contato: Antonio.Branco@di.fc.ul.pt

Resumo

No presente artigo, apresento uma breve análise da situação da língua portuguesa face ao choque tecnológico digital, os riscos e as oportunidades que neste contexto se colocam, e indico as ações imediatas que surgem como necessárias para se recuperar atrasos e assegurar para o português a sua posição de língua internacional de comunicação com projeção global.

Palavras-chave

Choque tecnológico digital. Tecnologia da linguagem. Língua portuguesa.

1 - Introdução

Tal como qualquer outro dos cerca de 6 000 idiomas existentes no planeta, a língua portuguesa é uma janela para o mundo que nos rodeia. É através da linguagem, e do seu uso em todas as situações do quotidiano, que comunicamos, aprendemos, nos entretendemos, planeamos o futuro, e fabulamos ou nos comparamos com uma história ou um poema.

Por paradoxal que à primeira vista pareça, sendo uma janela para o mundo que nos rodeia, a linguagem humana é também e cada vez mais um dos últimos obstáculos comunicacionais com que nos deparamos na era digital e num mundo globalizado.

As novas tecnologias da informação e da comunicação colocam ao nosso alcance pessoas de todo o mundo com quem será fácil interagir, assim como um acervo infindável de informação a que será possível aceder. Contudo, mesmo quando se encontram asseguradas exaustivamente as mais avançadas condições técnicas de acesso, este novo e ilimitado universo de possibilidades continua na sua esmagadora maioria inacessível, encerrado que está dentro das barreiras invisíveis das línguas que o dividem.

De igual modo, as novas tecnologias digitais colocam ao nosso alcance novos e cada vez mais poderosos dispositivos, desde eletrodomésticos até robôs pessoais, que alargarão as capacidades de cada um de nós para limites ainda por explorar. Porém, a utilização generalizada destes dispositivos continua em grande medida condicionada por interfaces idiossincráticas que restringem a sua utilização e limitam de forma drástica a concretização do seu potencial.

No presente artigo, apresento uma breve análise da situação da língua portuguesa face ao choque tecnológico digital, tendo como enfoque os riscos e as oportunidades que neste contexto se colocam.

No presente artigo, apresento uma breve análise da situação da língua portuguesa face ao choque tecnológico digital, tendo como enfoque os riscos e as oportunidades que neste contexto se colocam.

Uma discussão mais alongada e devidamente circunstanciada e referenciada do assunto abordado neste artigo pode ser encontrada em Branco *et al.*, 2012²

2 - Linguagem e tecnologia

Tal como em outras áreas da existência humana, a evolução científica e tecnológica tem alterado as condições de utilização, e da própria existência, das línguas naturais ao longo da história. Em alguns casos, estas novas condições de utilização das linguagens resultaram daquilo a que se pode chamar de choques tecnológicos, que estiveram na origem de profundas revoluções civilizacionais.

¹ Por conveniência da leitura, a referência completa a esta publicação é repetida aqui: BRANCO, António, Amália MENDES, Sílvia PEREIRA, Paulo HENRIQUES, Thomas PELLEGRINI, Hugo MEINADO, Isabel TRANCOSO, Paulo QUARESMA, Vera Lúcia Strube de LIMA & Fernanda BACELAR, 2012, *A Língua Portuguesa na Era Digital / The Portuguese Language in the Digital Age, Coleção Livros Brancos, Berlim, Springer. Este é o volume dedicado à língua portuguesa, de uma Coleção de Livros Brancos sobre as línguas europeias na era digital, que conta com 30 volumes, correspondendo a outros tantos idiomas.*

António Branco

Um dos primeiros choques tecnológicos envolvendo a linguagem humana de que há registo consistiu no advento da escrita, há cerca de seis mil anos. Esta inovação passou a permitir que os interlocutores comunicassem em linguagem natural de forma assíncrona, sem terem de estar na presença um do outro em simultâneo.

Com a escrita, pode-se dizer que se quebrou a barreira do tempo na utilização da linguagem. Por sua vez, com o advento da imprensa mecânica, há cinco séculos atrás, quebrou-se a barreira social no acesso à informação escrita. As publicações generalizaram-se e deixaram de estar acessíveis apenas para um pequeno grupo de leitores.

Há algumas décadas atrás um outro choque tecnológico para a linguagem natural teve lugar com o advento das telecomunicações. A barreira do espaço na utilização da linguagem foi quebrada, passando então a ser possível aos interlocutores comunicarem de forma síncrona apesar de não se encontrarem presentes no mesmo local.

Todos estes choques tecnológicos na utilização da linguagem natural tiveram impactos civilizacionais enormes, sobejamente assinalados pelos historiadores. O que tem sido porém muito menos assinalado são os seus impactos nas condições de existência das próprias línguas naturais. Estas mudanças tecnológicas têm proporcionado uma evolução tremenda nas condições de comunicação entre indivíduos: possibilitaram alargar as virtualidades da linguagem humana muito para além do que esta permite se restringida apenas a situações de conversa oral face a face. Mas a par dessa evolução, induziram também fortes mecanismos de involução ao nível das línguas.

A cada choque tecnológico, novas condições de utilização da linguagem têm colocado novas e mais estritas condições de existência para os diferentes idiomas. Um exemplo bem conhecido é o das línguas sem sistema de escrita, que foram desaparecendo, e continuam a desaparecer, perante a vantagem de se usar outras línguas, com escrita, que permitam tirar partido dos benefícios dessa superioridade tecnológica.

O progresso tecnológico parece assim exibir uma natureza mefistofélica no que toca à linguagem humana. Ao mesmo tempo que permite alargar a capacidade de comunicação entre indivíduos, cria também fatores de redução da diversidade linguística e do multilinguismo, e dessa forma de redução do património cultural e da pluralidade de mundividências que as diferentes línguas sustentam. Dos cerca de 6 000 idiomas existentes hoje em dia, estima-se que cerca de 2 500 correm o risco de se extinguirem nas próximas décadas, como acontece a cada ano que passa com as várias línguas que morrem com o desaparecimento do seu último falante.

Neste enquadramento, cabe atentarmos no mais recente choque tecnológico para a utilização da linguagem humana, que se encontra em curso nos dias de hoje. Este choque é provocado pela utilização das novas tecnologias digitais e resulta na expansão das condições de utilização das línguas para um novo patamar sem precedentes.

3 - Tecnologia da linguagem

A aplicação das novas tecnologias à linguagem natural, e em concreto o processamento computacional das línguas naturais, está a dar origem a uma nova área de investigação, desenvolvimento e inovação conhecida por tecnologia da linguagem. De um ponto vista genérico, esta tecnologia pode ser caracterizada como se ocupando com a obtenção da representação do significado a partir do processamento computacional de expressões

linguísticas, e vice-versa, da obtenção de expressões linguísticas a partir do processamento da representação do seu significado.

A tecnologia da linguagem engloba duas subáreas, conhecidas por processamento da linguagem e por processamento da fala. O processamento da fala ocupa-se com a obtenção de uma representação discreta a partir de um sinal analógico correspondente a enunciados orais, e vice-versa. O processamento da linguagem, por sua vez, ocupa-se em mapear entre uma sequência discreta de símbolos linguísticos e a representação do seu significado.

A tecnologia da linguagem está a abrir o leque inaudito de novas condições e de novas oportunidades para a utilização da linguagem natural:

- Ajudará as pessoas a comunicar entre si ainda que não falem uma língua comum;
- Apoiará uma nova geração de interfaces naturais e intuitivas, baseadas em linguagem natural, com todo o tipo de dispositivos, desde eletrodomésticos até robôs pessoais.

a. Aplicações

As novas oportunidades propiciadas pela tecnologia da linguagem resultam da sua exploração para o desenvolvimento de diversas aplicações computacionais, que podem afetar de modo diverso as condições de utilização da linguagem natural, e que se encontram atualmente em diferentes estados de maturidade tecnológica e de comercialização.

Algumas destas aplicações são usadas em contextos de utilização profissional especializados, como é o caso, por exemplo, dos detetores de plágio, para o trabalho de avaliação dos estudantes por parte dos professores, ou dos ambientes de apoio automatizado à tradução, para o trabalho dos tradutores.

Algumas outras aplicações são utilizadas de forma mais generalizada, fazendo parte do nosso quotidiano, como por exemplo, os corretores ortográficos.

Certas aplicações da tecnologia da linguagem, ainda que cruciais para o desempenho do sistema global em que se inserem, não são apercebidas pelos utilizadores, como é o caso, por exemplo, dos detetores de linguagem nos motores de busca.

Algumas delas, sendo de interesse para todo o tipo de utilizadores, são de especial relevo para pessoas portadoras de deficiências, como é o caso dos sintetizadores de voz no apoio à leitura por invisuais.

Há aplicações que servem propósitos circunscritos de utilização, como no caso dos sistemas de produção de documentação com base em linguagem controlada, usados na produção de manuais técnicos de manutenção de aeronaves. Outras aplicações servem uma leque mais amplo de situações de utilização, como é o caso dos sistemas de reconhecimento de fala, que podem ser integrados em dispositivos de controlo num automóvel ou suportar sistemas de legendagem automática de vídeos, entre inúmeras outras utilizações.

Face à limitação de espaço do presente artigo e à natureza do seu objetivo, não caberia fazer aqui uma listagem sistemática nem uma apresentação exaustiva de cada uma das possíveis aplicações da tecnologia da linguagem, das suas funcionalidades ou das situações da sua utilização. Para efeitos de alargar a ilustração apresentada até aqui, listam-se em seguida alguns casos mais notórios de aplicações e aqueles cuja designação será em grande medida autosexplicativa das suas funcionalidades. Para ajudar a exposição, estes exemplos serão

reunidos nos seguintes grupos temáticos:

Interação homem-máquina:

- Interfaces com dispositivos e agentes artificiais
- Detecção de linguagem, autor, domínio,...
- Classificação de textos
- Agrupamento de textos
- Busca de documentos
- Extração de informação
- Levantamento de opiniões
- Interfaces com bases de dados
- Resposta a perguntas
- Reconhecimento de fala
- Síntese de fala
- ...

Interação multilingue

- Tradução automática
- Agentes conversacionais
- Publicação multilingue
- ...

Produção e verificação de linguagem

- Correção ortográfica
- Correção gramatical
- Detecção de plágio
- Linguagens controladas e sistemas de produção de documentação
- Localização de software
- Legendagem automática
- Sistemas de ditado
- Sumarização
- Geração de relatórios
- Ambientes de apoio à tradução
- Simplificação de textos
- ...

Aprendizagem de linguagens:

- Formação
- Avaliação de competências
- ...

Web:

- Anotação de metadados
- Busca web avançada
- Gestão de ontologias
- ...

Algumas destas aplicações já encontraram não só a sua maturidade tecnológica mas também os modelos de negócio apropriados que permitem a sua ampla difusão e utilização, como acontece com os populares motores de busca de documentos.

Outras aplicações encontram-se ainda em fase de investigação científica ou de protótipo, como é o caso, por exemplo, dos sumarizadores.

Muitas outras aplicações encontrar-se-ão certamente ainda por ser desenhadas e imaginadas consoante a tecnologia da linguagem, que lhes servirá de base, vier a ser

desenvolvida, amadurecida e explorada em novas soluções para os utilizadores.

b. Tecnologia nuclear

A tecnologia da linguagem é assim uma tecnologia facilitadora e para algumas línguas que são objeto de esforços intensos e concertados de investigação, como é o caso notório da língua inglesa, encontra-se em acelerado ritmo de progresso e desenvolvimento.

Quer integrada em sistemas mais amplos quer suportando aplicações autónomas, esta tecnologia é crucial para a revolução tecnológica em curso: ajudará a ultrapassar as últimas fronteiras comunicacionais tanto no que toca à comunicação entre os seres humanos e os agentes e dispositivos artificiais, como no que toca à própria comunicação dos seres humanos entre si.

A investigação em tecnologia da linguagem pode ser vista como se desenrolando em duas dimensões principais inter-relacionadas. Por um lado, envolve o desenvolvimento das chamadas ferramentas de processamento computacional e, por outro, de aquilo que é usual designar-se por recursos linguísticos.

Os recursos linguísticos para uma dada língua são os conjuntos de dados, de diferentes tipos, que são cruciais para apoiar quer a investigação científica sobre essa língua, quer o desenvolvimento e a avaliação de ferramentas de processamento para a mesma.

Como exemplo de um tipo de conjunto de dados dos mais simples, pode-se mencionar os corpora de texto corrido, que consistem em coleções de textos, de domínios diversos, tal como eles foram publicados. Mas mais comumente, os recursos linguísticos são conjuntos de dados altamente complexos que são laboriosamente produzidos de modo a registar e a compilar os aspetos mais sofisticados dos fenómenos linguísticos tal como estes ocorrem ou são instanciados em enunciados usados pelos falantes. Por exemplo, no caso dos léxicos, cada expressão pode ser pormenorizadamente classificada de acordo com as suas várias características linguísticas, desde os aspetos fonéticos até aos semânticos. No caso dos chamados *treebanks*, num outro exemplo, são as frases que são alvo de caracterização exaustiva em termos das relações sintáticas que se encontram instanciadas entre as suas palavras.

Exemplos de recursos linguísticos:

- Corpora anotados
- Corpora multilingues paralelos e alinhados
- Bases de dados de fala
- Listas de palavras
- Abreviaturas, nomes próprios, palavras funcionais,...
- Vocabulários
- Léxicos
- Ontologias lexicais
- Terminologias
- Treebanks
- Propbanks
- DeepBanks
- ...

As ferramentas de processamento para uma dada língua, por sua vez, realizam um leque de diferentes tarefas que, quando encadeadas, contribuem para executar o mapeamento entre forma e significado nessa língua. Essa tarefas incluem desde a funcionalidade mais básica de determinar o início e o fim de frases, por exemplo, até funcionalidades altamente complexas e sofisticadas, como acontece nas gramáticas de processamento linguístico profundo, que permitem obter a representação semântica em forma lógica da frase de entrada.

A título de ilustração, pode-se também mencionar os etiquetadores morfossintáticos — que anotam as expressões num texto com a categoria morfossintática (verbo, nome, advérbio, etc) que lhe cabe em cada uma das suas ocorrências —, os lematizadores — que associam a cada palavra a sua forma flexionada canónica, e.g. a forma infinitiva no caso dos verbos —, ou os analisadores de dependências gramaticais — que identificam as funções gramaticais (Sujeito, Objecto Direto, etc) entre as expressões constituintes de uma frase—, entre vários outros exemplos de diferentes tipos e funcionalidades de ferramentas para o processamento linguístico, que lidam com a estrutura, a ambiguidade e a criatividade das línguas naturais.²

Exemplos de ferramentas de processamento:

- Separador de frases
- Separador de palavras
- Etiquetador morfossintático
- Lematizador
- Analisador morfológico
- Reconhecedor de nomes de entidades
- Desambiguador de aceções de palavras
- Analisador de constituência sintática
- Analisador de dependências gramaticais
- Etiquetador de papéis semânticos
- Gramática para processamento linguístico profundo (análise semântica)
- ...

4 - Desafios

Depois das barreiras do tempo e do espaço e da barreira social terem sido quebradas em anteriores choques tecnológicos, é a própria barreira da linguagem, e da diversidade linguística, que se encontra agora a ser ultrapassada com a revolução resultante da tecnologia da linguagem.

Tal como nos choques tecnológicos anteriores, o impacto civilizacional será tremendo e de uma dimensão ainda difícil de antever na sua totalidade. De modo análogo, a par das novas oportunidades abertas, também novos desafios e riscos se colocam agora aos idiomas e à sua existência.

No passado, línguas que em consequência do seu contexto económico e histórico particular, não foram estudadas e, por exemplo, para as quais não foi desenvolvido um sistema de escrita viram-se votadas ao beco sem saída da extinção.

² Para experimentar e melhor compreender a funcionalidade de algumas destas ferramentas, é possível recorrer o LX-Center, que apresenta vários demonstradores e serviços linguísticos online gratuitos para a língua portuguesa, em <http://lxcenter.di.fc.ul.pt>.

Hoje em dia, esse desafio joga-se ao nível da ciência e da tecnologia da linguagem. Línguas que, em consequência do seu contexto económico e histórico, não venham a ser alvo de estudo científico e para as quais não sejam desenvolvidas as soluções tecnológicas apropriadas, são línguas que a prazo arriscam a sua progressiva irrelevância, em detrimento de outras melhor preparadas, e eventualmente a sua própria extinção.

Neste enquadramento, importa analisar a actual situação no que diz respeito à língua portuguesa.

Existem 236 milhões de falantes do português em quatro continentes. Este é um número que crescerá para cerca de 335 milhões em 2050 e que faz desta língua a quinta com maior número de falantes no mundo, depois do chinês, castelhano, inglês e árabe. Entre os falantes das suas diferentes variantes existe compreensão mútua generalizada, e se se atender ao critério da compreensão mútua entre falantes para a circunscrição de um idioma, isso pode fazer do português a terceira língua com maior número de falantes no mundo, depois do inglês e do castelhano. Cabe assinalar também que o português é atualmente um dos idiomas de trabalho de 27 organizações internacionais.

A língua portuguesa reúne assim condições ímpares para assegurar o estatuto de língua de comunicação internacional com projeção global.

Esta oportunidade encontra-se porém em preocupante contraste com os riscos que decorrem do baixo nível da preparação do português para a era digital em termos de ciência e tecnologia da linguagem.

No quadro da rede científica europeia de excelência META-NET,³ foi publicada uma Coleção de Livros Brancos cujos volumes analisam a situação de diferentes línguas na era digital, tendo em atenção a sua preparação em termos de tecnologia da linguagem. Cada volume, elaborado por um grupo de especialistas, é dedicado a uma de 30 línguas europeias. No âmbito desta iniciativa, foi elaborado um estudo comparativo do estado dessas diferentes línguas, tendo sido feita uma classificação apoiada numa escala de cinco níveis, nomeadamente "Apoio excelente", "Apoio bom", "Apoio médio", "Apoio fragmentário" e "Pouco/nenhum apoio".

Em resultado desse estudo comparativo, em termos de processamento da linguagem, a língua portuguesa surge classificada como tendo "Apoio fragmentário": surge em situação mais vantajosa que apenas 8 das outras 29 línguas, e em situação igual ou pior que as restantes, entre as quais se inclui o Alemão, o Castelhana, o Francês, o Italiano e o Neerlandês, avaliada como gozando de "Apoio médio", e o Inglês, com "Apoio bom" (Branco *et al.*, 2012, p.36).

Para enfrentar o choque tecnológico digital, a língua portuguesa encontra-se muito menos preparada tecnologicamente que as línguas com que compete por protagonismo no sistema mundial — e inclusive muito menos preparada que outras línguas com muito menor projeção.

Face a estes desafios, são necessárias ações imediatas para que se possam obter progressos importantes para recuperar atrasos e assegurar para o português a sua posição de língua internacional de comunicação com projeção global. Neste enquadramento, cabe assinalar que há uma boa comunidade de centros de investigação, em Portugal e no Brasil,

³ <http://www.meta-net.eu>

que cooperam ativamente entre si e que, de momento, têm capacidade instalada para fazer avançar a tecnologia da linguagem para a língua portuguesa. É porém necessário garantir o incremento estratégico do esforço aplicado nesta área para segurar e incrementar esta capacidade e se vir a alcançar um patamar de progresso sustentado.

De modo similar ao que tem vindo a ser feito nas últimas décadas para outras línguas, e em particular e de forma notória para a língua inglesa, este esforço deve ser articulado de acordo com as seguintes medidas estratégicas:

- estabelecer programas de estímulo à **investigação especificamente inter e multidisciplinar na área do processamento computacional da língua portuguesa**, em particular, e da ciência e tecnologia da linguagem, em geral, nas Universidades dos países de língua portuguesa;
- estabelecer programas de investigação e desenvolvimento de longo alcance que fomentem a **construção de recursos linguísticos e de ferramentas para o processamento computacional do português**, e de **aplicações da tecnologia da linguagem capacitadas especificamente para a língua portuguesa**;
- fomentar programas de **cooperação entre os países de língua portuguesa** que promovam a partilha e a transferência de conhecimento e contribuam para um estado de desenvolvimento tecnológico das diferentes variedades do português mais equilibrado;
- promover a **adesão a infra-estruturas de investigação internacionais dedicadas especificamente à ciência e tecnologia da linguagem natural**, como é o caso do CLARIN, a primeira infra-estrutura internacional para a área, criada em fevereiro de 2012.

Referências

BRANCO, António, Amália MENDES, Sílvia PEREIRA, Paulo HENRIQUES, Thomas PELLEGRINI, Hugo MEINEDO, Isabel TRANCOSO, Paulo QUARESMA, Vera Lúcia Strube de LIMA & Fernanda BACELAR. 2012. *A Língua Portuguesa na Era Digital / The Portuguese Language in the Digital Age*, Coleção Livros Brancos, Berlim, Springer, ISBN 978-3-642-29592-8.



O USO DE CORPORA PARA A CONSTITUIÇÃO DE RECURSOS LEXICOGRÁFICOS DE REFERÊNCIA: O CASO DO VOC

José Pedro Ferreira, Gladis Maria de Barcellos Almeida,
Margarita Correia

Sobre o autor

José Pedro Ferreira
Instituto de Linguística Teórica e Computacional, Lisboa, Portugal.
zpperreira@gmail.com

Gladis Maria de Barcellos Almeida
Universidade Federal de São Carlos e Núcleo Interinstitucional de Linguística Computacional,
São Carlos, Brasil.
gladis@ufscar.br

Margarita Correia
Instituto de Linguística Teórica e Computacional e Universidade de Lisboa, Lisboa, Portugal.
margarita.correia@gmail.com

Resumo

Este artigo analisa os diferentes tipos de dados usados na construção de obras lexicográficas, com particular incidência no uso de dados extraídos de *corpora*, centrando-se na sua utilidade para a construção de recursos de referência, nomeadamente no âmbito dos trabalhos em curso para a criação do Vocabulário Ortográfico Comum da Língua Portuguesa (VOC). Começamos por fazer um apanhado geral das diferentes práticas na lexicografia de outras línguas, descrevendo depois os poucos casos representativos do uso de *corpora*, em sentido contemporâneo, nas obras dicionarísticas do português, centradas historicamente na tradição lexicográfica e na introspeção – sendo usados dados de corpora sobretudo para a extração de abonações e obtenção de aceções. Descrevendo a metodologia adotada no desenvolvimento VOC, procuramos mostrar o papel instrumental dos *corpora* para a definição de primeiras nomenclaturas de variedades em processo de standardização, para a validação de dados obtidos da tradição e para a aferição da frequência e especificidade do âmbito de uso de cada entrada.

Palavras-chave

Corpus. Lexicografia. Dicionário. Vocabulário ortográfico.

1 - Introdução

A constituição do Vocabulário Ortográfico Comum da Língua Portuguesa (VOC) é um desiderato da Comunidade dos Países de Língua Portuguesa (CPLP) e uma determinação do Acordo Ortográfico da Língua Portuguesa de 1990 (AO90). O projeto para a sua execução encontra-se atualmente em curso sob coordenação do Instituto Internacional de Língua Portuguesa (IILP). Tentativas anteriores de elaborar um vocabulário ortográfico comum esbarraram nas diferentes interpretações dadas ao texto das reformas ortográficas que pretendiam aplicar, que se refletiram no desenvolvimento de vocabulários ortográficos nacionais com divergências (Brasil e Portugal).¹

A mais recente reforma ortográfica da língua portuguesa, em curso desde 2009, que visa a unificação da ortografia nos vários países da CPLP, apenas unificou as regras de escrita, que preveem a coexistência de variantes nacionais em determinados contextos, e não exige a fusão das tradições lexicográfica e ortográfica. Com efeito, continuarão a existir casos de divergência que dependem não da aplicação de regras, mas da tradição, como 'úmido' no Brasil e 'húmido' nos restantes países. Outro dos problemas que se colocam deve-se ao facto de os diferentes países se encontrarem em momentos diferentes da consolidação e normalização das suas variedades nacionais. Neste contexto, o desenvolvimento do VOC nas condições descritas neste trabalho constitui não só a base da aplicação do AO90 em todos os países da CPLP, como um avanço sem precedentes na lexicografia de língua portuguesa.

Por estas razões, a solução adotada no VOC é uma abordagem mista, na qual a tradição lexicográfica e o uso real em contextos de uso das variedades-padrão são tidos em conta para obter uma fatia representativa do léxico da língua. Por um lado, a tradição desempenha um papel fundamental na determinação das formas existentes e na das que não são alteradas pela reforma ortográfica; por outro lado, para que o VOC seja representativo do português como um todo, língua pluricêntrica de oito países, deve registar pelo menos as formas mais frequentes em uso em todos os países, dando especial atenção àqueles com menor ou inexistente tradição lexicográfica. Com esse objetivo, todos os países da CPLP estão a constituir, no âmbito do VOC, recursos lexicográficos representativos do português tal como é usado no seu espaço nacional, seguindo uma metodologia comum, visando a constituição de recursos com desenho e dimensões comparáveis.

O presente artigo aborda o papel dos *corpora* na lexicografia moderna e no estabelecimento da norma, nomeadamente em línguas pluricêntricas como o português, e apresenta os princípios, critérios, metodologia, recolha de dados, processo de compilação e funções dos *corpora* no VOC. Deste modo, num primeiro momento, discutir-se-á o papel dos *corpora* em lexicografia, para, num segundo momento, se apresentarem as principais características dos *corpora* em geral e suas especificidades relativamente ao pluricentrismo do português, e, finalmente, se explanar o seu uso no trabalho desenvolvido no âmbito do VOC.

¹ Para uma descrição detalhada do projeto VOC e da estrutura, princípios e metodologia geral seguidas na prossecução dos seus objetivos, leia-se Almeida et al. (2013) e Ferreira et al. (2012a).

2 - O papel dos corpora na lexicografia

Quando falamos de *corpus* atualmente, imediatamente pensamos em *corpus* eletrônico. Mas o conceito de *corpus* não é um conceito novo em linguística – nem dependente exclusivamente deste suporte. Svensén (2009: 43) define *corpus* como “a collection of language material selected according to certain criteria and used as basis for linguistic investigation”. No entanto, *corpus* é entendido atualmente como “a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research”². Esse formato digital, promovido pelo advento do computador, interferiu diretamente não só na concepção que se tem de *corpus*, como também na sua forma de armazenamento e exploração, já que os recursos oferecidos pela máquina permitiram que grandes quantidades de textos pudessem ser processadas em questão de segundos, fazendo com que muitas hipóteses sobre determinados processos linguísticos pudessem ser testadas rápida e eficientemente.

Em lexicografia, o *corpus* diz respeito às fontes do trabalho lexicográfico. De acordo com Atkins & Rundell (2012: 45-46):

‘A reliable dictionary is one whose generalizations about word behaviour approximate closely to the ways people normally use (and understand) language when engaging in real communicative acts (such as writing novels or business reports, reading newspapers, or having conversations).’ But how can we feel confident that we know how people normally use words, and hence that the description given in our dictionary is reliable? Reliability depends on the kind of evidence that underpins our account of the language – and evidence comes in several forms.

As fontes lexicográficas podem ser primárias ou secundárias. As primárias são constituídas por dados autênticos, que podem ser resultantes da introspeção, de ficheiros de citações (citation files) e de *corpora* propriamente ditos. As secundárias são descrições da língua previamente existentes, em especial dicionários, gramáticas e outros estudos (SVENSÉN, 2009: 39).

Definido o conceito de *corpus*, refira-se o que se entende por ficheiro de citações. Estes são recolhidos a partir de um determinado conjunto de textos estrategicamente escolhidos que se consideram representativos da língua ou da variedade a ser descrita; desses textos são extraídos itens lexicais ou citações de material que são considerados adequados para serem incluídos no dicionário (SVENSÉN 2009: 41).

A utilização de *corpora* em sentido lato não é prerrogativa do mundo moderno. Há muito que se elaboram *corpora* constituídos por citações, que designaremos como “*corpora* de citações”, para a validação ou definição de uma norma lexical. Na tradição lexicográfica do português e do inglês – para exemplificar com duas grandes línguas pluricêntricas – estes *corpora* de citações viriam a ser organizados e publicados em forma de grandes dicionários de língua.

² SINCLAIR, 2005

Na língua portuguesa, pode citar-se o *Vocabulário Portuguez e Latino*, de Rafael Bluteau, publicado em oito volumes entre 1712-1728, o primeiro dicionário para o qual foi fixado um *corpus* de citações³, que foi utilizado como exemplário de uso linguístico para as palavras que constavam da nomenclatura⁴. Outro exemplo, já no século XIX, é o *Diccionario da Lingua Portuguesa*, 2ª edição, de António de Moraes Silva, publicado em 1813, que também se valeu de um *corpus* de citações como fonte para a recolha de exemplos⁵.

Na língua inglesa, pode referir-se *A Dictionary of the English Language*, publicado em 1755, para o qual Samuel Johnson reuniu um *corpus* de textos para poder observar as palavras em seu uso autêntico e como fonte de exemplos para os verbetes. Outra importante obra é *The Oxford English Dictionary*, cujo projeto lexicográfico também se valeu de uma grande quantidade de citações, obtidas a partir de uma carta de apelo enviada pelo próprio Murray, responsável pela obra, em 1879, pedindo colaborações. Mais de 800 leitores voluntários passaram a enviar ao editor tiras de papel com citações que deveriam conter as palavras a eles atribuídas (BIBER, CONRAD & REPPEN, 1998).⁶

Os dicionários produzidos com base em *corpora* de citações são feitos pelo método de citações (*citation method*) (SVENSÉN, 2009: 41-42). Os *corpora* de citações destes dicionários são apenas constituídos, num primeiro momento, por excertos de obras literárias do cânone, as que se consideram adequadas à exaltação da língua do Estado pelos primeiros dicionários das Academias, para, posteriormente, incluírem também excertos de textos que representem o padrão, incluindo o discurso científico e o discurso institucional, nas lexicografias inglesa e norte-americana, e mais tarde excertos de textos de épocas mais remotas, na realização de dicionários históricos⁷.

O século XX marca efetivamente o dealbar da linguística como ciência, e o dicionário incorporou as realizações da linguística e o seu discurso científico próprio, tendo-se transformado, sobretudo ao longo da segunda metade do século passado, num objeto distinto daquele que conhecíamos: os dicionários são hoje predominantemente descritivos do uso da língua e da norma em uso⁸. É na segunda metade do século XX que assistimos ao nascimento da lexicografia baseada em *corpus*.

Comparando o método de *corpus* (*corpus method*) com o método de citações, Svensén (2009: 44-45) destaca as seguintes diferenças fundamentais:

- o método de *corpus* permite a quantificação, ao passo que, ao utilizar o método de citações, existe uma seleção quer dos textos, quer das citações que são extraídas e dos próprios itens lexicais;

José Pedro Ferreira, Gladis Maria de Barcellos Almeida, Margarita Correia.

- o método de *corpus* fornece informação sobre os aspetos típicos e centrais da língua, dado o grande número de fenómenos linguísticos frequentes que o *corpus* geralmente inclui, enquanto o método de citações, criado mediante seleção deliberada de dados, pode tender a enfatizar usos invulgares e periféricos das palavras.

Também Atkins & Rundell (2008: 52), ao enumerarem as desvantagens do método de citações, colocam a nota no carácter subjetivo que preside à seleção das citações e referem que já Noah Webster e James Murray tinham observado que "human readers tend to notice what is remarkable and ignore what is typical, and this creates a bias towards the novel or the idiosyncratic usages which inevitably catch the reader's eye."

Deste modo, o método de *corpus* permite que o dicionário seja representativo da língua em uso e a sua representatividade passa a assentar na própria representatividade do *corpus*, permitindo em simultâneo adotar critérios de maior objetividade no processo lexicográfico.

Em nosso entender, são três as grandes mudanças que a lexicografia baseada em *corpus* trouxe aos dicionários: a representatividade do vocabulário descrito, a natureza da exemplificação usada e, por fim, a descrição da variação linguística e da norma.

A primeira grande mudança tem a ver com a representatividade do vocabulário descrito. Numa lexicografia baseada em *corpus*, as palavras não podem ser inseridas na nomenclatura do dicionário apenas com base na introspeção, nos dados dos *corpora* de citações ou das fontes secundárias, mas sim com base em critérios mais objetivos e passíveis de serem aplicados com rigor, que vão da frequência de ocorrência de um item no *corpus* à sua representatividade no tipo de discurso cujo vocabulário se pretende retratar. Deste modo, os bons dicionários de hoje são representativos de fatias bem delimitadas do léxico de uma língua, primeiramente balizadas pelo próprio *corpus*, pela sua representatividade e balanceamento.

A segunda mudança prende-se com o tipo de exemplificação do uso das unidades lexicais que é feita. Os dicionários modernos não apresentam apenas abonações (excertos retirados do *corpus* de citações destinados a validar ou a indicar o "bom uso" das palavras), mas, sim, exemplos de uso, retirados das mais diversas fontes incluídas no *corpus* e destinados, sobretudo, a dar conta do uso efetivo, do uso real da língua. Esses exemplos podem ser completamente autênticos (retirados do *corpus* como tal, como acontece no projeto COBUILD), ou podem ser editados e trabalhados pelo lexicógrafo para melhor servirem os seus objetivos. Os exemplos de uso são um complemento das demais informações associada à unidade lexical, veiculando aspetos gramaticais, colocacionais, enciclopédicos e culturais, entre outros, facilitando o uso efetivo dos itens lexicais em novos contextos.

A terceira diferença importante entre os dicionários baseados em *corpus* e os seus antepassados é que, sendo o *corpus* constituído por textos representativos de diversos registos linguísticos, os dicionários de hoje têm a preocupação de retratar a variação lexical, não apenas a variação diatópica e a diacrónica – como já ocorria, de resto, ainda que de forma assistemática –, mas também a variação diastrática e a diafásica. Se, nas nossas sociedades, os dicionários veiculam a norma e se os dicionários baseados em *corpus* apresentam as características anteriormente enunciadas (representatividade, exemplificação e respeito pela variação), é o próprio conceito de norma que é veiculado pelo dicionário que diverge da dos seus antepassados: a norma veiculada pelo dicionário baseado em *corpus* é a norma baseada no uso efetivo da língua – embora apenas nos registos incluídos no *corpus* tomado como referência – e já não uma norma idealizada, centrada apenas numa seleção feita pelo lexicógrafo, tipicamente com critérios subjetivos e feita apenas a partir dos registos literário e formal.

Como temos vindo a demonstrar, lexicografia e linguística de *corpus* estão intimamente ligadas. Por outro lado, existe também uma íntima relação entre lexicografia baseada em *corpus* e lexicografia de aprendizagem (L2 *lexicography*), visto que os primeiros grandes projetos de lexicografia baseada em *corpus* se destinavam em primeiro lugar a estudantes de língua, particularmente de língua estrangeira; dicionários de língua materna e dicionários bilingues usaram corpora um pouco mais tarde⁹. O exemplo pioneiro e mais típico do dicionário baseado em *corpus* é o *Collins COBUILD English Language Dictionary*, publicado pela primeira vez em 1987.

Apesar dos muitos desenvolvimentos que a lexicografia baseada em *corpus* tem conhecido a nível internacional, a verdade é que existe a este respeito um longo caminho a percorrer pela lexicografia do português. Para a língua portuguesa, o único dicionário inteiramente baseado em *corpus* de que temos conhecimento é o *Dicionário de Usos do Português do Brasil*, de Francisco da Silva Borba, também conhecido como DUP, com primeira edição em 2002, que, de acordo com a informação disponível, foi confeccionado com base num largo *corpus* escrito de mais de 200 milhões de palavras do português brasileiro contemporâneo.

Em Portugal, verifica-se que o *Dicionário da Língua Portuguesa Contemporânea* usa exemplos retirados do *Corpus de Referência do Português Contemporâneo*, embora a extensão e a forma desse processo não sejam explicitadas na informação existente sobre o dicionário. Além do *corpus*, este dicionário foi construído a partir de fontes secundárias e de ficheiros de citações elaborados especificamente para completar a informação recolhida no *corpus*.

Os atuais dicionários de língua portuguesa usam sobretudo fontes secundárias, particularmente dicionários existentes, cujos dados são completados e atualizados com recurso seja a *corpora* de citações, seja a *corpora* em sentido pleno. Porém, na maioria dos grandes dicionários da língua portuguesa, quer portugueses, quer brasileiros, não é jamais explicitada a relação que mantêm com os corpora. É o caso de grandes dicionários como o *Michaelis* (*Michaelis – Moderno Dicionário da Língua Portuguesa*, 1998), o *Aurélio* (*Dicionário Aurélio da Língua Portuguesa*, já na 5.^a edição, publicada em 2010, que traz versão impressa e eletrónica), o *Houaiss* (*Dicionário Houaiss da Língua Portuguesa*, 2009, também em versão impressa e eletrónica), ou o *Grande Dicionário da Língua Portuguesa*, 2.^a ed., da Porto Editora, de 2010¹⁰.

Também os vocabulários ortográficos oficiais mais recentes recorrem sobretudo a fontes secundárias, particularmente vocabulários já existentes, socorrendo-se dos *corpora* apenas para atualização e completamento dos dados. É o caso do *Vocabulário Ortográfico da Língua Portuguesa*, 5.^a ed., da Academia Brasileira de Letras (VOLP), e também do *Vocabulário Ortográfico do Português* (VOP), baseado principalmente no *Vocabulário da Língua Portuguesa*, de Rebelo Gonçalves, na sua edição de 1966. Ambos os vocabulários, VOLP e VOP, estão a ser integrados no VOC.

A realização do VOC com a metodologia esclarecida neste trabalho constituirá um avanço na lexicografia baseada em *corpus* de língua portuguesa.

⁹ Cf. TEUBERT, 2001 e SVENSÉN, 2009: 45-47.

¹⁰ Para uma retrospectiva da lexicografia de língua portuguesa, cf. Correia 2009.

3 - O papel dos corpora para a avaliação e o estabelecimento do padrão

Além de os corpora terem um importante papel em grandes projetos lexicográficos orientados por uma metodologia empírica, considera-se que a palavra seja a principal unidade de análise dos estudos realizados em linguística de corpus, dada a facilidade de sua identificação pelas ferramentas computacionais (VIANA, 2011). Em projetos envolvendo o léxico, a palavra é praticamente a porta de entrada para a análise de corpus, ou, como afirma Calzolari (1997), "all Language Engineering applications require knowledge about words". Não é sem razão, pois, que projetos e atividades de pesquisa situados no âmbito do léxico tenham no corpus seu maior aliado.

A moderna noção de corpus também carrega consigo requisitos que devem ser tidos em conta num projeto de corpus. São eles: representatividade (*representativeness*), balanceamento (*balance*), diversidade (*diversity*) e tamanho (*size*)¹¹.

De entre todos esses requisitos, a representatividade é crucial, tendo em conta que um corpus representativo tende a ser bem balanceado, ter boa diversidade e tamanho adequado aos objetivos da pesquisa. Assim, para a construção do VOC¹², especial atenção foi dada à representatividade, já que o vocabulário ortográfico deverá servir como importante instrumento de normalização lexical do português.

Quanto ao caso específico das variedades menos representadas do português, há que levar em conta a discrepância entre as condições linguísticas do português no conjunto dos oito países membros da CPLP: Angola, Brasil, Cabo Verde, Guiné-Bissau, Moçambique, Portugal, São Tomé e Príncipe e Timor-Leste.

Com exceção de Portugal e Brasil, que já têm corpora, grandes dicionários tradicionais (instrumentos linguísticos que permitem a construção e são muitas vezes a base de vocabulários ortográficos) e que são países onde o português é a língua mais usada, os demais países (Angola, Cabo Verde, Guiné-Bissau, Moçambique, São Tomé e Príncipe e Timor-Leste) vivem uma situação linguística bastante diferente, uma vez que grande parte dos seus cidadãos não tem o português como primeira língua.

Observe-se, na Tabela 1, a quantidade de línguas que convive com o português nos Países Africanos de Língua Oficial Portuguesa (PALOP) e em Timor-Leste, numa lista necessariamente incompleta dada a inexistência de consenso em relação às línguas faladas em cada país e, em alguns casos, à falta de dados e de estudos recentes. Apresenta-se, também, na mesma tabela, a população aproximada de cada país e o ano de sua independência, para se ter uma noção um pouco mais ampla do cenário linguístico, demográfico e político no qual se insere este projeto do VOC.

¹¹ MCENERY & WILSON, 1996; KENNEDY, 1998; BIBER et al., 1998; RENOUF, 1998; SINCLAIR, 2005.

¹² A constituição dos corpora para a elaboração do VOC será detalhada na parte 4.2.

Estados-membros	Línguas faladas (além do português)	População	Ano da independência
Angola	quicongo, quimbundo, chóque, umbundo, cuaniama, nhaneca, fiote, nganguela, entre outras	18.498.000	1975
Cabo Verde	cabo-verdiano (ou crioulo de Cabo-Verde)	499.796	1975
Guiné-Bissau	crioulo da Guiné-Bissau, balanta, diola, papel, mandinga, manjaco, fula, bijagó, entre outras	1.499.230	1974
Moçambique	suáíli, maconde, macua, nianja, chope, cóti, sena, xona, xítsua, xindau, nhúngue, chuabo, lome, suázi, changana, rongga, entre outras	20.366.795	1975
São Tomé e Príncipe	forro, principense (ou lunguyé) e angolar.	187.356	1975
Timor-Leste	tétum, entre várias línguas austronésias e algumas papuanas	1.066.582	1975 ¹³ e 2002 ¹⁴

Tabela 1: Línguas faladas nos PALOP e em Timor-Leste, além do português

Tendo em vista, pois, este cenário, antes de se pensar em constituir vocabulários nacionais de cada um desses países, é preciso primeiro constituir os *corpora* que representem, de facto, a norma lexical de cada um deles, *corpora* que incluam tanto palavras comuns a todos ou a alguns países como aquelas que são específicas de apenas um deles.

Além desse aspeto, foi preciso ter em conta que os *corpora* compilados fossem comparáveis em tamanho, em distribuição por géneros e em relação ao peso de cada um dos géneros no conjunto final¹⁶, de modo que o VOC, depois de pronto, evidenciasse uma gestão política de língua descentralizada, despolarizada (saindo do eixo Portugal e Brasil) e com participação ativa de todos os estados-membros da CPLP.

Crítérios iniciais que orientaram a seleção dos textos a incluir nos *corpora* para o VOC foram: i) textos escritos e em contexto formal; ii) prioridade para textos já em formato digital. A partir daí, iniciou-se um trabalho coletivo, contando com a cooperação das equipas nacionais de cada país, de forma a estabelecer as características dos *corpora*. A partir dessas primeiras decisões, foram-se desenhando os *corpora*, de modo a atender o principal requisito, a representatividade.

¹³ Independência de Portugal.

¹⁴ Fim da ocupação indonésia.

¹⁵ As informações foram obtidas no portal da CPLP (<http://www.cplp.org/>); nos portais dos respectivos governos (Angola - <http://www.governo.gov.ao/>; Cabo Verde - <http://www.governo.cv/>; Guiné-Bissau - <http://www.anpguinebissau.org/>; Moçambique - <http://www.portaldogoverno.gov.mz/>; São Tomé e Príncipe - <http://www.presidencia.st/>; Timor-Leste <http://timor-leste.gov.tl/>) e na Wikipédia, especificamente para conferir os dados populacionais. No caso de Timor-Leste, foi também consultado Batoré e Casadinho (2009).

¹⁶ ALMEIDA & FERREIRA, 2012.

Ser representativo significa que um *corpus* deve ser elaborado de forma a espelhar determinadas características linguísticas da comunidade cuja língua está sob análise¹⁷. Daí a importância de se fazerem escolhas adequadas para que o *corpus* possa realmente refletir comportamentos linguísticos. Algumas questões devem ser postas durante a seleção dos textos: que documentos, que tipos de textos, que gêneros textuais? Enfim, o que de facto representa os usos linguísticos de uma comunidade? No caso dos *corpora* para o VOC, a pergunta teve de ser ainda mais específica: quando se escreve em português num contexto formal nesses países, que palavras são empregadas?

Diretamente associado à representatividade, está o conceito de balanceamento. Embora seja um conceito vago, de acordo com Sinclair (2005), é preciso ter em mente que o *corpus* deve ter um equilíbrio entre gêneros discursivos (informativo, científico, etc.), tipos de textos incluídos (artigo, editorial, entrevista, dissertação, etc.), temas (pesca, agricultura, saúde, educação, etc.), ou até mesmo títulos, ou autores. O ideal seria que se conseguisse levar em conta todas essas categorias, mas sempre atendendo às demandas do objetivo que se pretende atingir.

Se o *corpus* for cuidadosamente balanceado, ele terá uma boa diversidade. A propósito disso, Biber *et al.* (1998:248) assinalam que dado que “there are important differences in the use of lexical, grammatical and discourse features across different varieties of language”, há argumentos para afirmar que o conceito de “língua geral” é hipotético, dado que cada gênero discursivo tem seus próprios padrões de uso. Nesse sentido, um *corpus* representativo deve conter uma grande diversidade de gêneros, tipos de textos e assuntos, pois a frequência de muitas palavras pode variar de acordo com o assunto¹⁸.

Um *corpus* pensado para ser representativo, balanceado e diversificado tem, em geral, um tamanho adequado ao tipo de pesquisa que se vai realizar e à metodologia a ser adotada na pesquisa¹⁹. Quando se fala em tamanho de *corpus*, não se trata somente do número total de ocorrências ou átomos (*tokens*) e de tipos (*types*, palavras diferentes), mas com quantas categorias (gêneros discursivos, tipos de textos, assuntos, títulos, autores, etc.) um *corpus* deve contar, quantas amostras de cada categoria e quantas palavras existem dentro de cada amostra²⁰. Assim, os *corpora* que estão sendo compilados nos PALOP e em Timor-Leste atendem aos requisitos de representatividade, balanceamento, diversidade e tamanho acima especificados).²¹

Todavia, para a gestão descentralizada de uma língua pluricêntrica, tão importante quanto esses requisitos é o envolvimento de todas as equipas nacionais²² no projeto; afinal, quem mais sabe responder às questões que devem ser feitas durante a seleção de textos para compor um *corpus* são as equipas representativas desses países.

¹⁷ SINCLAIR, 2005.

¹⁸ BIBER *et al.*, 1998.

¹⁹ SINCLAIR, 2005.

²⁰ KENNEDY, 1998.

²¹ Cf. detalhamento destes aspetos do projeto na parte 4.2.

²² Cf. ALMEIDA *et al.*, 2013.

4 - O uso de corpora no VOC

Tendo em conta o que foi dito, os *corpora* assumem necessariamente um papel central na criação de novos recursos lexicográficos para a língua portuguesa quando tomada como um todo, como é o caso do VOC, sendo o garante da validação dos dados constantes de obras existentes e de que os dados novos são realmente representativos da língua em uso nos vários países.

Por um lado, pela sua natureza, herança e princípios metodológicos, os dicionários existentes para o português são mais uma perpetuação da tradição lexicográfica, ainda que em alguns casos parcialmente atualizada, do que um verdadeiro repositório atualizado das palavras de facto em uso nos países de língua portuguesa, o que afeta a sua representatividade. Por isso, uma primeira função dos *corpora* no VOC é a de validação dos dados constantes da tradição lexicográfica do português.

Além disto, dado que a reforma ortográfica não torna homogênea a escrita de muitas palavras com formas ortográficas já antes divergentes (e.g. 'úmido' no Brasil, mas 'húmido' nos restantes países), qualquer repositório lexical multinacional para o português deve identificar, marcar e ligar por meio de remissões as diversas variantes, o que só é possível executar num curto espaço de tempo recorrendo à análise semiautomática das divergências entre *corpora* representativos de cada variedade.

Por último, atendendo à falta de cobertura lexicográfica atual para muitas das variedades nacionais do português e à inexistência de *corpora* representativos dessas variedades, a criação e uso de tais recursos de base é essencial para captar a realidade da língua escrita em contexto formal em cada país e proceder à sua integração num recurso que a represente, fixando a ortografia do seu léxico. Neste apartado, os *corpora* servem, pois, de fonte primária para as entradas da obra em criação.

4.1 - O uso de corpora com recursos existentes

O VOC não será, no entanto, e como já afirmado, constituído unicamente por informação constituída por *corpora*. De facto, como apontado na secção 2, o português dispõe já de uma longa tradição lexicográfica, construída sobretudo nos últimos dois séculos em Portugal e no Brasil, que deu origem a um acervo extenso, ainda que não comparável ao de outras línguas de idêntica relevância mundial.

Por isso, o VOC será constituído por entradas com duas origens, que representarão duas partes distintas da obra: uma, correspondente à *memória lexicográfica do português*, composta pelas entradas dos vocabulários ortográficos oficiais já existentes; outra, correspondente ao léxico obtido através da mineração de *corpora*, representativa do léxico efetivamente em uso em todos os países. A memória lexicográfica do português será constituída por um subconjunto equilibrado das entradas do *Vocabulário Ortográfico do Português* (VOP), produzido, sob a responsabilidade de Margarita Correia, pelo Instituto de Linguística Teórica e Computacional e oficial em Portugal, e do *Vocabulário Ortográfico da Língua Portuguesa*, 5.^a ed. (VOLP), coordenado por Evanildo Bechara, da Academia Brasileira de Letras. Ambas as obras são baseadas nas nomenclaturas de dicionários existentes²³, ainda que os *corpora* venham a desempenhar um papel importante na integração de ambos no VOC, como se verá adiante.

Em qualquer obra desta natureza, que se proponha integrar informação provinda de fontes várias lexicográficas e com características distintas, uma das principais tarefas passa pela homogeneização das entradas e pelo controlo da representatividade de cada uma delas, equilibrando o resultado final. Um primeiro problema que a execução do VOC levanta passa pelo facto de que as obras a integrar na memória lexicográfica do português têm características diferentes quanto aos critérios lexicográficos de inclusão e de identidade lexical, além de diferentes dimensões e propriedades (o VOLP tem uma nomenclatura muito mais extensa que o VOP, mas inclui muito menos informação para cada entrada). Este facto impossibilita, desde logo, a integração direta das duas obras lexicográficas.

Para fazer face a estes problemas, o VOC foi planeado sobre a utilização de uma plataforma centralizada de edição lexicográfica, que permite a homogeneização das obras, recorrendo a *corpora* para selecionar subconjuntos de dados. Foi escolhida pelo IILP a plataforma usada pelo ILTEC na construção do VOC, o OSLIN²⁴, pelo que as operações a executar incidem maioritariamente sobre o VOLP, ainda não adaptado a essa plataforma.

O OSLIN (sigla de *Open Source Lexical Information Network*, “rede de informação lexical de código aberto”) é um conjunto de ferramentas e estruturas organizadas numa plataforma digital desenhada com base em serviços alojados na nuvem que permite a rápida criação de recursos lexicais e a sua gestão e alargamento²⁵. A plataforma contém ferramentas dedicadas a diferentes funções, pensadas para serem usadas por lexicógrafos, que facilitam a gestão e a criação, de forma integrada, a partir de um par *forma de citação – classe gramatical*, de entradas lexicais com outra informação formal associada: paradigma flexional, todas as formas flexionadas, divisão silábica (para fins de translineação), acentuação, remissões para entradas relacionadas (entre elas as variantes diatópicas decorrentes da aplicação do AO90, muito relevantes no caso desta obra), relações funcionais com outras entradas (e.g., todos os nomes deverbais eventivos, nomes de qualidade e advérbios deadjetivais são associados explicitamente) e, dentro em breve, informação sobre a morfologia e a fonética²⁶.

²³ Para uma descrição do VOP, disponível em <http://www.portaldalinguaportuguesa.org/?action=vop>, consulte-se Correia e Ferreira [no prelo]. As fontes do VOLP não são referidas explicitamente na obra, embora, em comunicações públicas dos seus autores, seja ponto assente que teve como base obras lexicográficas do português.

²⁴ JANSSEN, 2005.

²⁵ Para uma descrição mais detalhada do sistema multivalente de edição e manutenção dos recursos lexicais do OSLIN, consulte-se Ferreira et al. (2008) e a secção 4.2 do presente trabalho.

²⁶ ASHBY & FERREIRA, 2010.

A integração faseada do VOLP nesta plataforma encontra-se em fase de conclusão. A nomenclatura daquela obra foi primeiro importada para o sistema OSLIN, seguindo operações de equivalência, desdobramento e sobreposição de classes gramaticais, tornando-as (e às entradas que as contêm) compatíveis com as do VOC. Cada entrada do VOLP, agora já adaptada aos critérios de identidade lexical e ao *tagset*²⁷ do OSLIN, foi depois marcada como estando ou não presente também no VOC, sendo validada automaticamente a sua inclusão no sistema nos casos em que tal se verificava.

O restante material lexical do VOLP está, neste momento, a ser sistematicamente cruzado e integrado no VOC com base em interseções entre a obra e corpora existentes, de modo a fasear o trabalho e a validar os dados. Em primeiro lugar, o VOLP foi cruzado com um léxico computacional construído a partir do cópulo NILC²⁸. Esse léxico foi primeiramente, tal como o VOLP, adaptado à estrutura lexical e categorial do VOC, depois de conformado com o AO90, tarefa executada com recurso ao conversor *Lince*²⁹ e a verificações manuais por padrão através de expressões regulares correspondentes aos contextos em que, no português do Brasil, existem potenciais mudanças.

Posteriormente, o VOLP foi também cruzado com o *Corpus Brasileiro*³⁰; para tal, além das operações indicadas acima, é adicionalmente necessário proceder à lematização dos átomos constantes do *corpus*, à sua etiquetagem morfossintática e à sua contagem e ordenação por frequência³¹.

Depois destas operações preparatórias, que permitiram a criação de bases de dados lexicais MySQL independentes para cada fonte, é executado para cada uma delas o mesmo pré-processamento computacional para os pares *lema-categoria gramatical* que foi feito para o VOLP, usando ferramentas disponíveis no sistema de administração do OSLIN, de modo a tornar os dados compatíveis com os das outras bases de dados lexicais. A base de dados contendo o VOLP já de acordo com o formato OSLIN é então fornecida, através de cruzamento, com informação de frequência obtida a partir do *corpus*, de modo a obter um subconjunto de mais de 200 000 entradas do VOLP que estejam atestadas com suficiente frequência. As entradas a inserir no VOC estão, pois, sempre atestadas na fonte lexicográfica de referência, o VOLP, e em *corpora*.

4.2 - O uso de corpora para constituição de novos recursos

Para as restantes variedades do português, no entanto, não é possível adotar este procedimento acima descrito, dado que não existem ainda recursos lexicográficos de referência que as representem. Cada um destes países optou por uma de três opções:

- criar autonomamente um recurso lexicográfico representativo, a ser depois tratado de acordo com o procedimento explicitado na parte 4.1 deste trabalho;
- constituir uma lista de palavras características de cada país a adicionar à base comum entretanto constituída (que seria usada como *corpus* de exclusão para pesquisa das formas características de uma dada variedade);

²⁷ Conjunto de etiquetas classificativas das palavras quanto às categorias gramaticais a que pertencem.

²⁸ PINHEIRO & ALUÍSIO, 2003.

²⁹ FERREIRA et al., 2012b.

³⁰ SARDINHA et al., 2009.

³¹ O léxico do NILC já tinha sido alvo destas operações, pelo que não foram, nesse caso, necessárias.

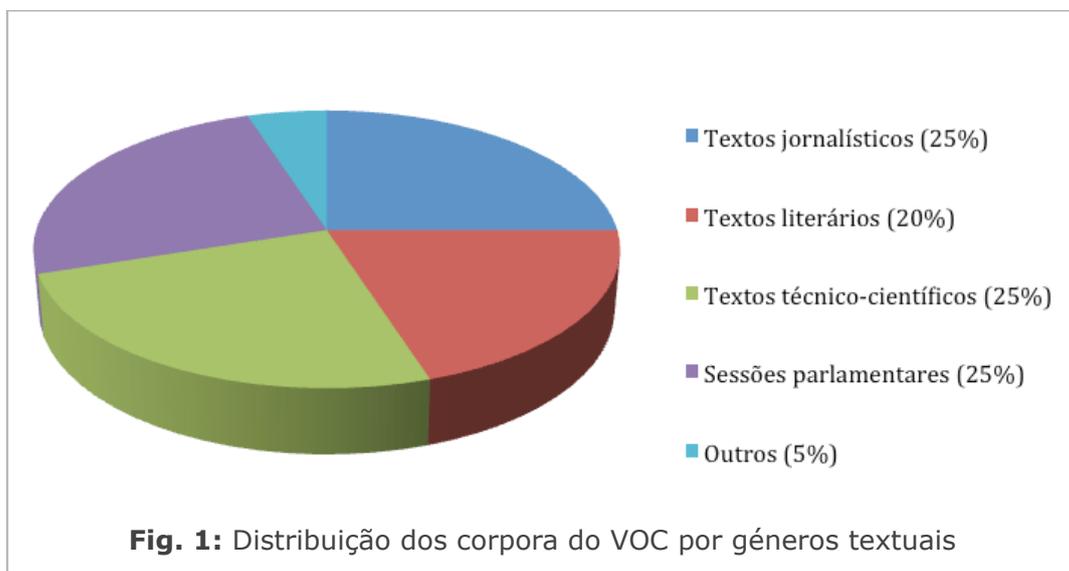
- executar o trabalho de levantamento dos *corpora*, construindo de forma integrada no OSLIN vocabulários ortográficos nacionais, tomando partido das suas ferramentas para a rápida construção de recursos lexicais.

Seguindo as decisões, no seio do IILP, dos representantes de todos os países da CPLP, que refletem, em boa medida, as conclusões do exposto na parte 3 do presente texto, o trabalho está a ser executado em alguns países de acordo com a segunda proposta enunciada acima, e noutros com base em corpora, seguindo a última das opções apresentadas. É sobre o trabalho destas equipas que se debruça esta secção. Seguindo uma metodologia comum³², cada estado-membro da CPLP assegura a existência de uma equipa nacional que executa, em coordenação com a equipa central do projeto, as tarefas lexicográficas necessárias à prossecução do objetivo final.

Um primeiro problema para que isto seja possível prende-se, no entanto, com a inexistência de *corpora* de referência representativos das variedades nacionais desses países. O pouco trabalho de desenvolvimento feito nesse sentido dá em geral conta, ademais, de variedades sociais que não se coadunam com o caráter dos recursos a criar, que se pretendem representativos de variedades diafasicamente relacionadas com a escrita em contexto formal e com a norma de cada país.

A primeira e essencial tarefa do projeto tem passado, por isso, pela constituição de um *corpus* de referência para cada país da CPLP que dele não disponha e se proponha executar este trabalho. Como se justifica na parte 3 deste artigo, tal deverá ser feito salvaguardando a representatividade, o balanceamento, o tamanho e a diversidade, tendo ainda atenção à exequibilidade e custos, essenciais para permitir a rápida execução do pretendido sem demasiado elevados gastos com recursos humanos. Foi elaborado e aprovado pelos representantes técnicos dos vários países, em sede própria, um conjunto de metas quanto a tipos textuais, sua proveniência e o peso no resultado final.

Cada *corpus* nacional terá no mínimo 30 milhões de átomos, distribuídos por sub-corpora (cf. Fig. 1) com texto literário (20%), jornalístico (25%), legislativo e de sessões parlamentares (25%), técnico (saúde, educação, ambiente, pescas e agricultura: 25%) e de proveniência variada (5%).



³² ALMEIDA & FERREIRA, 2012.

Esta distribuição assegura que todos os países possam ter as mesmas fontes, todas passíveis de serem obtidas desde logo em formato digital, de modo a reduzir o tempo de execução e o custo com recursos humanos que seria necessário para processar texto noutros suportes. Além disso, garantiu-se que o material necessário existe em todos os países que seguiram esta via, o que determinou os domínios de especialidade do sub-*corpus* técnico indicados acima.

Os textos são obtidos por cada equipa nacional em formato digital, renomeados consoante o tipo e a origem da fonte e colocados numa estrutura de pastas hierarquizada num servidor comum. Cada ficheiro é depois processado pela equipa central, que o converte para texto simples e sugere elementos não textuais a serem limpos, e pela equipa nacional, que procede à validação desse processamento e à limpeza, num conjunto de operações recursivas em que cada equipa nacional é a responsável última pela qualidade do *corpus* representativo da sua variedade. Posteriormente, cada texto é dividido em extratos, que são anotados quanto à proveniência e tipo, atomizado, lematizado, etiquetado e finalmente o seu material lexical é convertido num léxico de frequência contendo forma de citação (correspondente ao lema), classe de palavra e índice de frequência³³.

Cada uma destas listas lexicais de frequência é depois inserida numa base de dados própria no sistema OSLIN, sendo então assumida como lista de candidatos à inserção no VOC. Todas as entradas já registadas no sistema, provindas da tradição lexicográfica a partir dos vocabulários oficiais existentes, ou obtidas a partir dos vocabulários ortográficos nacionais já construídos no âmbito do projeto, são marcadas como estando atestadas para a variedade nacional do país cujo *corpus* está a ser processado, sendo atribuído a cada entrada um índice de frequência nacional. Depois de definida uma linha de corte com base na frequência e na distribuição por sub-*corpus*, as entradas que não estão ainda atestadas são finalmente inseridas no VOC por cada equipa nacional usando a plataforma de gestão do sistema.

Para este efeito, foi integrado na plataforma de gestão do OSLIN um módulo de inserção em massa de entradas lexicais que permite gerar automaticamente todas as propriedades formais de cada palavra, já indicadas na parte 4.1 do presente texto, a partir da forma de citação e da categoria gramatical de cada candidato. Os lexicógrafos da equipa nacional de cada país, tal como acontece com as entradas obtidas a partir do VOP e do VOLP, verificam a legitimidade da entrada e a correção das informações geradas automaticamente pelo sistema, definindo manualmente apenas o paradigma flexional de cada entrada, informação que não pode ser aduzida pelo sistema³⁴.

Cada palavra constante do VOC é marcada explicitamente quanto à sua proveniência e quanto às fontes e países em que se encontra atestada, sendo as propriedades formais de cada entrada acompanhadas desta informação e, nos casos em que esteja disponível, do seu índice de frequência e das variantes já atestadas no Vocabulário. Está em desenvolvimento um módulo adicional do OSLIN que permite a identificação semiautomática de variantes, que facilitará a difícil tarefa de registar todas as remissões correspondentes à variação nacional existente entre países, quer a resultante da aplicação das regras do AO90, nos casos em que permite variação, quer a que não decorre da sua aplicação, dado que este não unifica a grafia do português, mas apenas a enunciação das suas regras³⁵.

³³ Para uma descrição detalhada do processo seguido, consultar Almeida & Ferreira (2012) e Ferreira et al. (2012a).

³⁴ Para uma descrição detalhada deste módulo, consultar Janssen (2011).

³⁵ Para uma descrição detalhada do processo de constituição do VOC e das funções das suas equipas, cf. Almeida et al. (2013) ou Ferreira et al. (2012a).

6 - Referências bibliográficas

ALMEIDA, G. M. B.; FERREIRA, J. P. 2012. *Manual para a elaboração de corpora. Com vista à organização dos Vocabulários Ortográficos Nacionais dos países integrantes da CPLP*. Lisboa (Portugal): ILTEC; São Carlos (SP, Brasil): NILC.

ALMEIDA, G. M. B.; FERREIRA, J. P.; CORREIA, M.; OLIVEIRA, G. M. 2013. Vocabulário Ortográfico Comum (VOC): constituição de uma base lexical para a língua portuguesa. In: *Estudos Linguísticos*, 42 (1): p. 204-215.

ASHBY, S.; FERREIRA, J. P. 2010. The Role of Morphology in Generating High-Quality Pronunciation Lexica for Regional Variants of Portuguese. In: Branco, A., Klautau, A.; Vieira, R.; Lima, V. L. S. de; Pardo, T. A. S. (eds.). *Computational Processing of the Portuguese Language – Lecture Notes in Artificial Intelligence*, v. 6001. Berlin Heidelberg Springer-Verlag: 162-165.

ATKINS, B. T.S.; RUNDELL, M. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

BATORÉO, H.; CASADINHO, M. 2009. O português, uma língua pluricêntrica: o caso de Timor-Leste. In: *Revista Portuguesa de Humanidades, Estudos Linguísticos*, 13-1. Braga: Universidade Católica Portuguesa de Braga, 63-79. Disponível em: <http://www.catedraportugues.uem.mz/lib/docs/bib_timor/Batoreo_Casadinho_2009.pdf>. Acesso em 12 ago 2012.

BIBER, D.; CONRAD, S.; REPPEN, R. 1998. *Corpus Linguistics – Investigating Language Structure and Use*. Cambridge, U.K.: Cambridge University Press.

CALZOLARI, N. 1997. Lexicon and Corpus: a multi-faceted interaction. In: *Lèxic, corpus i diccionaris*. Cicle de conferencies 95-97. Barcelona: Institut Universitari de Lingüística Aplicada.

CORREIA, M. 2008. Lexicografia no início do século XXI – novas perspectivas, novos recursos e suas consequências. In: Júnior, M. A. (coord.) *Lexicon – Dicionário de Grego-Português, Actas de Colóquio*. Lisboa: Centro de Estudos Clássicos / FLUL, pp. 73-85.

CORREIA, M. 2009. *Os dicionários portugueses*. Lisboa: Caminho.

CORREIA, M.; FERREIRA, J. P. 2013. Dicionários e vocabulários ortográficos na constituição da norma. In: Moita Lopes, L. P. (ed.). *O Português no Século XXI*. São Paulo: Parábola.

CORREIA, M.; FERREIRA, J. P. [no prelo]. Vocabulário Ortográfico do Português Correia, Margarita e Ferreira, José Pedro (no prelo). In: *Actas del 5.º Congreso Internacional de Lexicografía Hispánica*, Madrid, 25-27 de junho de 2012.

FERREIRA, J. P.; BARBOSA, S.; JANSSEN, M. 2008. Mordebe Admin: A Lexical Management System. In: *Proceedings of the XIII EURALEX International Congress*. Barcelona: Institut Universitari de Lingüística Aplicada - Universitat Pompeu Fabra; Documenta Universitaria: 351-357.

FERREIRA, J. P.; JANSSEN, M.; ALMEIDA, G. M. B.; CORREIA, M.; OLIVEIRA, G. M. 2012. The Common Orthographic Vocabulary of the Portuguese Language: a set of open lexical resources for a pluricentric language. In: CALZOLARI, N.; CHOUKRI, K.; DECLERCK, T.; DOĞAN, M. U.; MAEGAARD, B.; MARIANI, J.; ODIJK, J.; PIPERIDIS, S. (eds.). *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA): 1071-1075.

FERREIRA, J. P.; LOURINHO, A.; CORREIA, M. 2012. Lince, an End User Tool for the Implementation of the Spelling Reform of Portuguese. In: Caseli, H. M.; Villavicencio, A.; Teixeira, A. J. S.; Perdigão, F. (eds.). *Computational Processing of the Portuguese Language - 10th International Conference, PROPOR 2012 Proceedings*. Lecture Notes in Computer Science 7243. Springer: 46-55.

JANSSEN, M. 2005. Open Source Lexical Information Network. In: BOUILLON, P.; KANZAKI, K. (eds.). *Proceedings of the Third International Workshop on Generative Approaches to the Lexicon*, May 19-21 2005. Genebra: École de Traduction et d'Interpretation – Université de Genève.

JANSSEN, M. 2011. Computer-Aided Inflection for Lexicography Controlled Lexica. In: KOSEM, I.; KOSEM, K. *Electronic Lexicography in the 21st Century New Applications for New Users – Proceedings of eLex 2011*. Liubliana: Trojina, Institute for Applied Slovene Studies: 96-105.

KENNEDY, G. 1998. *An Introduction to Corpus Linguistics*. Londres / Nova Iorque: Longman.

LARA, L. F. 1997. *Teoría del diccionario monolingüe*. México: El Colegio de México.

MCENERY, T.; WILSON, A. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.

MURAKAWA, C. A. A. 2001. Tradição lexicográfica em língua portuguesa. In: OLIVEIRA, A. M. P.; ISQUERDO, A. N. (eds.). *As ciências do léxico: lexicologia, lexicografia e terminologia*, 2ª. ed., Campo Grande: Ed. UFMS, 153-159.

MURAKAWA, C. A. A. 2006. *António de Morais Silva: lexicógrafo da língua portuguesa*. Araraquara: Laboratório Editorial FCL/UNESP; São Paulo: Cultura Acadêmica Editora.

PINHEIRO, G. M.; ALUÍSIO, S. M. 2003. *Córpus NILC: descrição e análise crítica com vistas ao projeto Lacio- -TR-03-03*. Disponível em: <<http://www.nilc.icmc.usp.br/lacioweb/downloads/NILC-TR-03-03.zip>>. Acesso em 8 set 2013.

Rebelo Gonçalves, F. 1966. *Vocabulário da Língua Portuguesa*. Coimbra: Coimbra Editora.

RENOUF, A. (ed.) 1998. *Explorations in Corpus Linguistics*. Amsterdão: Rodopi.

SARDINHA, T. B.; MOREIRA FILHO, J. L.; ALAMBERT, E. 2009. The Brazilian Corpus: A one-billion word online resource. In: MAHLBERG, M.; GONZÁLEZ-DÍAZ, V.; SMITH, C. (Eds.). *Proceedings of the Fifth Corpus Linguistics Conference, CL2009*, University of Liverpool, 20-23 July 2009.

SINCLAIR, J. 2005. Corpus and Text – Basic Principles. In: WYNNE, M. (Ed.). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow: 1-16. Disponível em: <<http://ota.ahds.ac.uk/documents/creating/dlc/chapter1.htm>>. Acesso em 12 ago 2013.

**O USO DE CORPORA PARA CONSTITUIÇÃO DE RECURSOS
LEXICOGRÁFICOS DE REFERÊNCIA: O CASO DO VOC**

José Pedro Ferreira, Gladis Maria de Barcellos Almeida, Margarita Correia.

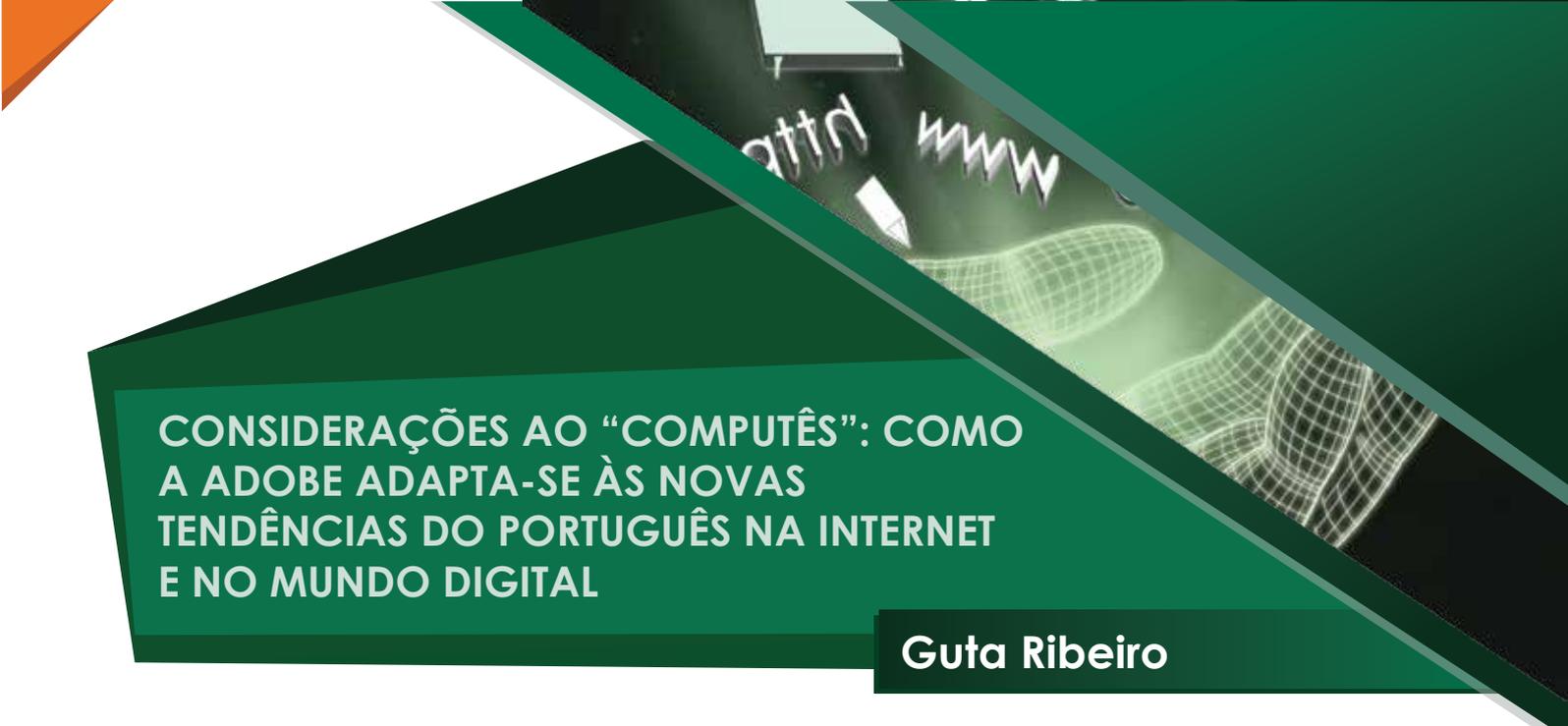
SVENSÉN, B. 2009. *A Handbook of Lexicography – The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

TEUBERT, W. 2001. Corpus Linguistics and Lexicography. *International Journal of Corpus Linguistics*, 6 (special issue), 125–153.

VIANA, V. 2011. Linguística de corpus: conceitos, técnicas & análises. In: Ed. VIANA, V.; TAGNIN, S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial.

**O USO DE CORPORA PARA CONSTITUIÇÃO DE RECURSOS
LEXICOGRÁFICOS DE REFERÊNCIA: O CASO DO VOC**

José Pedro Ferreira, Gladis Maria de Barcellos Almeida, Margarita Correia.



CONSIDERAÇÕES AO “COMPUTÊS”: COMO A ADOBE ADAPTA-SE ÀS NOVAS TENDÊNCIAS DO PORTUGUÊS NA INTERNET E NO MUNDO DIGITAL

Guta Ribeiro

Sobre o autor

Adobe Systems, Inc., Sr. Group Program Manager.
Contato: gutarib@adobe.com

Resumo

Uma das áreas de pesquisa mais interessantes para o grupo de globalização da Adobe Systems é o impacto das redes sociais e da internet no nosso trabalho de localização. E, para mim, as consequências para o português e as novas tendências da língua na internet. O “computês” seria uma referência ao português usado na área da informática, uma variação usada pelas comunidades de tradutores na internet para atender a necessidade de acesso rápido à informação num mundo tecnológico em constante mudança. Neste trabalho, pretendo 1) mostrar algumas alternativas em localização que a Adobe tem lançado com o fim de alcançar o melhor resultado possível para nossos clientes lusófonos em tempo real; 2) abordar as estratégias de implementação desses projetos; 3) fazer uma avaliação dos mesmos e uma análise do impacto, negativo e positivo, que esses projetos já demonstraram para a Adobe e, finalmente, 4) mencionar planos futuros.

Palavras-chave

Internetês, Computês. Adobe. Adobe systems. Globalização. Localização.

Abstract:

Computês’: How Adobe Adapts to the New Varieties of Portuguese Found on the Internet in a Digital World

One of the most interesting areas of research for the Globalization team at Adobe Systems is the impact that social networks and the Internet have on our localization work. Furthermore, for me, the consequences for the Portuguese language and the new varieties of Portuguese found on the Internet. “Computês” refers to the Portuguese language found on the Internet; variations used by the communities of translators online, aiming to keep up with quick access to information in an ever changing technological world. On this paper, I intend to 1) demonstrate some localization alternatives used by Adobe seeking to get the best possible results for our Portuguese-speaking customers in real time; 2) speak about the strategies used to implement these projects; 3) evaluate the projects and analyze how these projects have already impacted our work positively and negatively and, finally, 4) touch on some next steps.

1 - Introdução

A Adobe Systems realiza diversos tipos de trabalho de localização, utilizando diferentes tipos de ferramentas e técnicas, algumas tradicionais e outras modernas. Hoje, vou discutir apenas alguns dos nossos processos, que devem ser de interesse particular. Tenham em mente, entretanto, que esta é apenas uma faceta do processo completo de localização da Adobe.

Somos membros da TAUS (Translation Automation User Society) e membros fundadores da TAUS Data Association, cujo foco é melhorar a qualidade da tradução, baseando-se em dados estatísticos. Nosso grupo está à frente dos novos desenvolvimentos nesta área e tentamos liderar o mercado com a adoção de novas tecnologias.

Atualmente, a Adobe Systems trabalha com cinco máquinas de tradução: ProMT, Language Weaver, Systran, Moses e CCID (para Chinês Simplificado). Moses é uma nova tecnologia para desenvolver máquinas de tradução. As máquinas desenvolvidas através do pacote Moses são chamadas máquinas Moses. Utilizamos todas estas máquinas para pré-traduzir aplicativos, documentação, páginas de web, material de marketing etc., em vários idiomas.

Além do uso de máquinas de tradução acima citadas, terceirizamos o trabalho de tradução, pós-edição, teste, trabalhando com os nossos parceiros de localização mundialmente usados pela indústria de localização.

A Adobe, como líder no fornecimento de soluções que permitem aos clientes produzir, distribuir e lucrar com a criação de conteúdo para a mídia e publicações ou para marketing digital, tem investido nos setores educacional, serviços financeiros e governo nas comunidades dos países de língua portuguesa.

A língua portuguesa na internet é hoje a quinta língua mais utilizada na rede mundial de computadores. O desafio da Adobe é conseguir manter a consistência linguística dos termos entre todos os produtos da Adobe e, ao mesmo tempo, atender a demanda do mercado que exige produtos localizados o mais rapidamente possíveis, com qualidade, e baixo custo.

Guta Ribeiro

Atualmente, a Adobe produz 20 aplicativos de SaaS (Software as a Service, ou On Demand software – aplicativos como serviço ou aplicativo por demanda). Estes produtos têm um ciclo de lançamento bastante rápido – até semanais – e são na maioria das vezes gratuitos.

No intuito de cumprir o cronograma de lançamento dos produtos e, ao mesmo tempo, preencher os nossos requisitos de internacionalização e preservar a consistência linguística entre os produtos da Adobe, usamos várias ferramentas próprias, outras terceirizadas, e estamos sempre pesquisando novas estratégias em localização ágil.

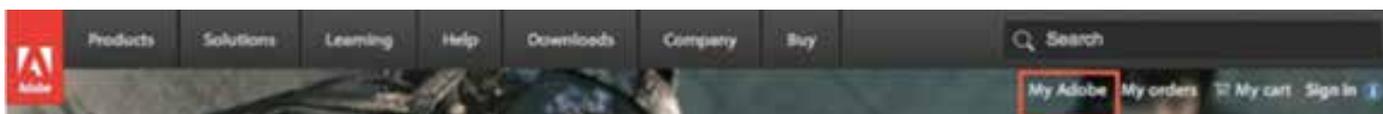
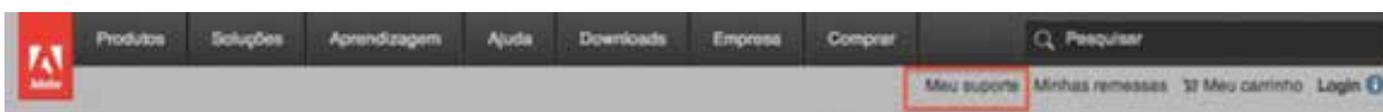
Uma das áreas de pesquisa mais interessantes para o grupo de globalização da Adobe Systems é o impacto das redes sociais e da internet no nosso trabalho de localização. E, para mim, as consequências para o português e as novas tendências da língua na internet.

Neste trabalho, pretendo 1) mostrar algumas alternativas em localização que temos lançado com o fim de alcançar o melhor resultado possível para nossos clientes lusófonos em tempo real; 2) abordar as estratégias de implementação desses projetos; 3) fazer uma avaliação dos mesmos e uma análise do impacto que esses projetos já demonstraram para a Adobe e, finalmente, 4) mencionar planos futuros.

2 - Considerações Iniciais

Minha apresentação não tratará dos assuntos da transcrição. Na transcrição, os tradutores possuem autoria criativa e plena liberdade sobre o texto. Este método é usado quando o departamento de marketing deseja transmitir uma mensagem específica com clareza, independente de considerações culturais e linguísticas. Este tipo de linguagem é muito usado na localização de publicidade, piadas, slogans ou em situações onde é necessário o entendimento claro com o cliente.

Veja abaixo como o sítio da Adobe Brasil faz a transcrição do termo “My Adobe”. “Meu suporte” foi escolhido pelo escritório da Adobe Brasil como o termo mais apropriado para este mercado.



Tampouco vou abordar temas como as palavras novas do tipo ‘blog’, ‘post’, ‘face’, ‘chat’ etc., assim como as contrações que apareceram com os chats, como ‘d+’, ‘vc’, ‘pls’, ‘qq’ etc., ou a tendência de retirada de acentos, cedilhas, e até das pontuações em geral, relacionados ao “internetês”. Para mim, como linguista, as consequências para o português e as novas tendências da língua na Internet são temas fascinantes¹.

¹ Cf. RIBAS et al., 2007.

Abaixo, alguns exemplos interessantes do “internetês” encontrados na web:



Yahoo Brasil



Google Brasil

Note-se, em 'Chat Salas de Bate-Papo', a redundância no uso de ambos, o termo em inglês 'chat' e a sua tradução.

Na Adobe, acreditamos que a compreensão, com clareza, é vital. Se a forma original do termo for melhor compreendida quando mantida no original, então devemos usar a forma original. 'Bitmap', 'hamburger', 'hot dog', 'download', 'display', 'flat', 'lobby', 'home theater', 'Flex', são termos tão comuns que se começarmos a traduzir ficam incompreensíveis. A língua é viva! Os computadores e a comunicação global só fizeram acelerar a evolução.

Adotamos o que gostaria de chamar o “computês”: uma referência ao português usado na área da informática, uma variação usada pelas comunidades locais de usuários e de tradutores na internet para atender a necessidade de acesso rápido à informação num mundo tecnológico em constante mudança.

Alguns exemplos do “computês” seriam:

- Erro de **script** do Internet Explorer para Imprimir > Revisão da impressão se o navegador permitir **applets** no Windows XP/Vista

IBM Brasil

O que é **Deployment** ?

Resumindo: Deployment é a tarefa de instalar um software (Windows, Office ou qualquer outro) em diversas estações de maneira simples e eficiente visando organizar, facilitar e agilizar a manutenção da rede local após a sua implementação.

www.baboo.com.br/

Finalmente alguns exemplos retirados dos glossários da Adobe Systems:

Inglês	Português
anti-aliasing	antialiasing
anti-spyware	antispyware
multi-threading	multithreading
multi-spin	multispin
pre-release	pré-release
pre-roll	pré-roll

1. *pré-roll* = tem a ver com a veiculação online de anúncios antes/durante/depois da exibição de algum vídeo.
2. *antialiasing* = suavização de borda - se refere a Photoshop, Illustrator, Flash, Fireworks.
3. *multicore* = em geral se refere a um computador com vários núcleos de processamento (dual core, por exemplo).
4. *multithreading* = em geral, tem a ver com processamento do computador.

Note-se que, apesar dos termos terem sido mantidos no inglês, eles obedecem as regras de acentuação e da reforma ortográfica de 9 de maio de 2011.

3 - Desenvolvimento

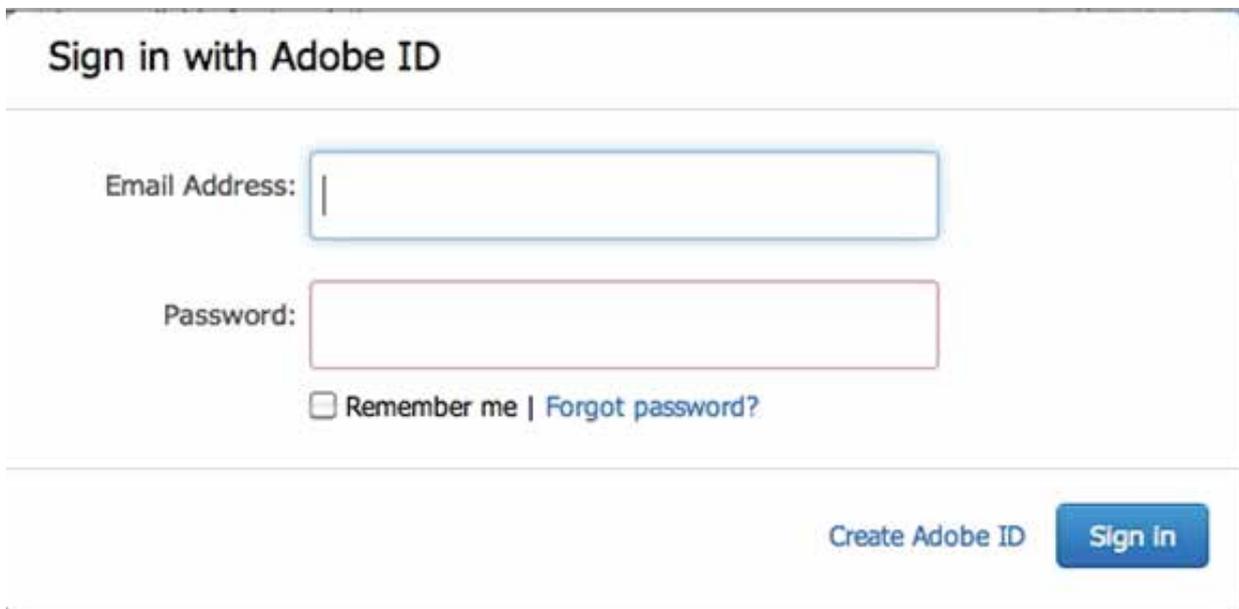
Para demonstrar algumas iniciativas de sucesso envolvendo as comunidades locais como tradutores, damos preferência a estes indivíduos da comunidade 1) por serem usuários experts dos softwares da Adobe; 2) por serem voluntários; 3) pelo seu comprometimento em colaborar rapidamente com traduções e/ou comentários que melhoram a qualidade dos produtos da Adobe. Abordarei três exemplos de ferramentas alternativas de tradução usadas na web e mostrarei como tornamos simples para a comunidade ter acesso a estes sítios da Adobe e contribuir com traduções ou colaborar com suas ideias.

O Business Catalyst, um aplicativo SaaS, que tem desenvolvimento ágil, é uma plataforma de hospedagem na web, um catalisador que ajuda aumentar o seu negócio através da construção, marketing, venda e suporte de páginas na web. BC é traduzido através do projeto Adobe Translator - <https://community.translate.adobe.com>. CTT é uma ferramenta de tradução comunitária, que permite que tradutores voluntários em todo o mundo possam traduzir a interface do aplicativo.

Em junho de 2011, Business Catalyst foi colocado no Adobe Translator em cinco línguas, para efeito de revisão pela comunidade local de usuários. Além de analisar as línguas já incluídas no produto, a comunidade teve a oportunidade de fornecer traduções para outros idiomas – dinamarquês, italiano, holandês, português do Brasil, romeno e esloveno. O que aconteceu ao longo dos meses foi um exemplo clássico de contribuições surpreendentes e sólidas. A Adobe pretende abrir, cada vez mais, projetos para a comunidade, permitindo não só revisão, mas também tradução para um número cada vez maior de idiomas.

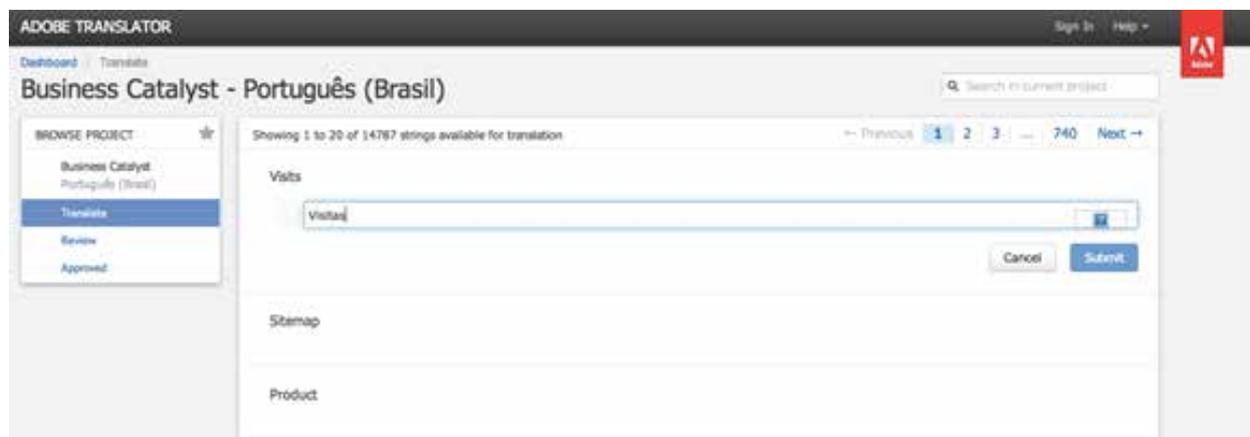
A metodologia usada é a seguinte: após internacionalização do produto e integração com o CTT, usamos tradução automática (feita através das máquinas citadas acima) e oferecemos os termos da interface do produto à comunidade através do sítio:
<https://community.translate.adobe.com/>:

Em primeiro lugar, é necessário criar um Adobe ID:



<https://community.translate.adobe.com>

Depois de selecionar o produto e a língua, se pode começar o trabalho de tradução:

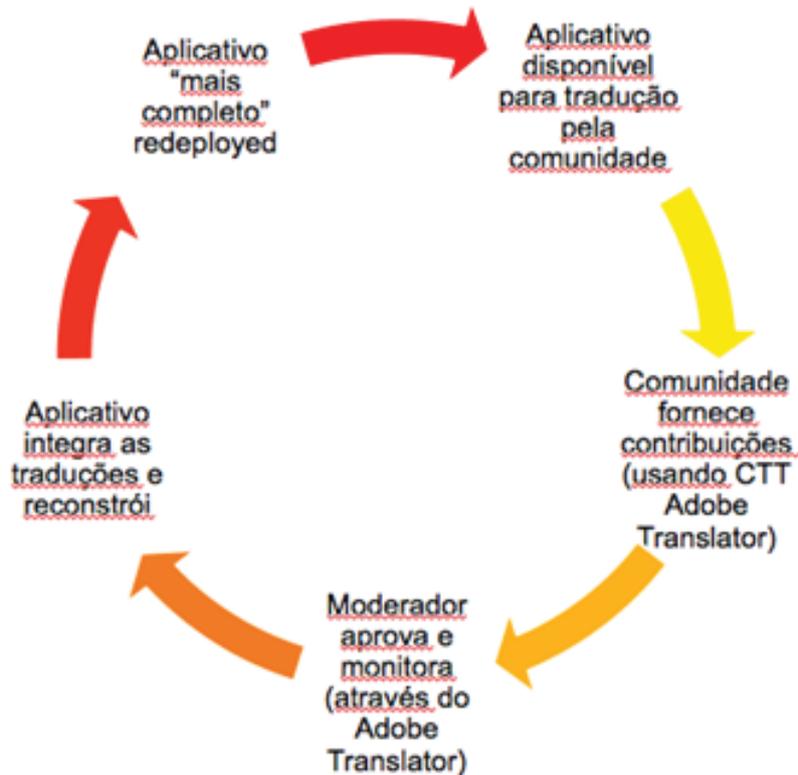


<https://community.translate.adobe.com>

Guta Ribeiro

Para cada termo ou frase usada na interface, se abre um campo onde o tradutor da comunidade oferece sua tradução. Os termos entrados pela comunidade serão indicados para revisão pelo líder ou moderador da comunidade para este produto. Após revisão, serão incorporados aos glossários da Adobe.

O ciclo de contribuição pela comunidade é o seguinte:



Ciclo de contribuição pela comunidade

A visão da Adobe é transformar CTT numa plataforma aberta à comunidade local, alinhando a localização de produtos da Adobe com as necessidades reais do mercado. Num futuro bem próximo, será a solução padrão para a localização dos aplicativos da Adobe chamados “Touch Apps” – Photoshop Touch, Adobe Ideas etc.

Atualmente, os produtos que estão disponíveis para tradução através do CTT são:

- Adobe Nav;
- Adobe Story;
- Adobe Translator;
- Business Catalyst;
- Photoshop (apenas para a comunidade de pré-release);
- Flex SDK;
- Muse.

Para maiores informações sobre os últimos acontecimentos do CTT, pode-se consultar o grupo no Facebook: <http://www.facebook.com/adobe.community.translation>.

O segundo exemplo de como a Adobe envolve a comunidade nos projetos de localização é o projeto Adobe TV – <http://tv.adobe.com/translations/> – de tradução comunitária, que amplia o alcance do conteúdo, permitindo que tradutores voluntários em todo o mundo possam traduzir vídeos instrutivos em qualquer idioma.



<http://tv.adobe.com/translations/>

Assim que o usuário se inscreve como “tradutor” da Adobe, passa a escolher os programas que deseja traduzir no idioma selecionado.

Alguns exemplos de traduções feitas pelas comunidades do português do Brasil abaixo:



Desenvolvendo Revistas Eletrônicas com InDesign CS5.5 e Adobe Digital Publishing Suite

Show: [CS5.5 Design Premium Feature Tour](#)

Neste vídeo, Terry White demonstra como você pode criar uma revista eletrônica para tablets utilizando o InDesign CS5.5. Ele mostrará para você incríveis interatividades que poderão ser adicionadas em suas publicações, como uma sequencia de imagens e a possibilidade de ampliação do conteúdo, também mostrará como você poderá pré visualizar em seu desktop o conteúdo utilizando o visualizador de conteúdo para desktop.



Crie eBooks mais atraentes com o InDesign CS5.5

Show: [CS5.5 Design Premium Feature Tour](#)

Neste vídeo, Terry White mostra a você todos os novos recursos do InDesign CS5.5 - incluindo texto vinculado, painel artigos e a capacidade de incorporar áudio e vídeo - isto vai ajudar você a criar livros mais atraentes do nunca.



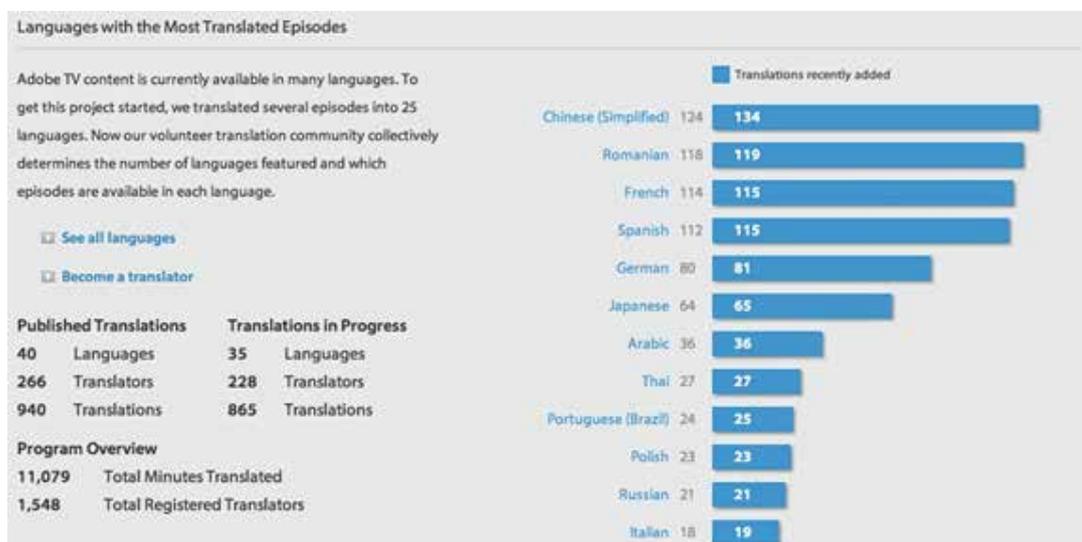
Edição de assinatura do Creative Suite 5.5: uma nova e conveniente opção de pagamento

Show: [CS5.5 Production Premium Feature Tour](#)

Saiba mais sobre a nova opção de assinatura da Adobe. Ela oferece os mais recentes recursos e funcionalidades do Creative Suite por um custo mensal acessível. Você descobrirá seus benefícios e verá como é fácil adquirir, usar e gerenciar seu software de assinatura.

<http://tv.adobe.com/translations/>

Para demonstrar o sucesso da Adobe TV, veja abaixo o número de participações na Adobe TV até 8 de abril de 2012:



<http://tv.adobe.com/translations/>

Como mostra o quadro, o português do Brasil está incluído entre os dez idiomas mais traduzidos na Adobe TV.

Meu terceiro e último exemplo é o nosso Blog de Globalização. A Adobe promove a difusão da língua portuguesa como veículo de informação e acesso ao conhecimento tecnológico através do nosso Blog de Globalização -<http://blogs.adobe.com/globalization/pt/>. Este é um fórum internacional que permite o diálogo entre a Adobe e as comunidades dos países lusófonos.

Como exemplo de tradução automática crua, sem revisão profissional, veja o exemplo de uma página do nosso Blog de Globalização <http://blogs.adobe.com/globalization/pt/>:

Está chegando: Série mitos de Globalização

por [LEANDRO REIS](#) publicado em [ABRIL 6, 2012](#)

Este artigo foi escrito originalmente em Inglês. Texto em outros idiomas é fornecido através de tradução automática.

A partir deste mês, vamos publicar uma série de artigos sobre mitos de globalização.

Se você já trabalhou na indústria da Globalização o tempo suficiente, provavelmente já ouviu muitos dos mitos que vamos descrever.

Se você é novo ou de fora da indústria, então espero que esta série vai ajudar a esclarecer a situação..

O primeiro artigo será publicado em breve. Fique ligado.

Blog de Globalização da Adobe Systems

Vale ressaltar que aqui colocamos a nota “Este artigo foi escrito originalmente em Inglês. Texto em outros idiomas é fornecido através de tradução automática.” No caso do Blog de Globalização, a tradução automática crua, sem revisão profissional, é aceitável pela comunidade porque o valor mais importante do Blog é a disseminação rápida de informação para e pela comunidade de usuários da Adobe.

4 - Resultados

Os projetos exemplificados acima oferecem resultados positivos e negativos para o nosso grupo de globalização na Adobe. Produtos SaaS, com uma interface de usuário relativamente simples, que têm um ciclo de lançamento bastante rápido e são, na maioria das vezes, gratuitos, seriam os melhores candidatos ao CTT.

Um aspecto negativo do CTT é que, atualmente, oferece em primeiro lugar, os termos da interface do produto traduzidos automaticamente, crus, sem revisão, o que frustra alguns dos nossos usuários. A Adobe planeja melhorar este aspecto do CTT, oferecendo termos pré-traduzidos através de nossas máquinas de tradução que usam os nossos bancos de termos revistos e aprovados.

Guta Ribeiro

A Adobe TV permite que tradutores voluntários traduzam vídeos instrutivos em qualquer idioma. Em alguns casos, as traduções tornam os vídeos mais difíceis de compreender. Na maioria dos casos, as contribuições são relativamente boas e a resposta da comunidade é positiva, como mostramos acima.

O Blog de Globalização tem como propósito principal informar, o mais rapidamente possível, todas as notícias referentes a Globalização. Neste caso, a tradução automática crua, sem revisão profissional, é aceitável pela comunidade. Não se trata de um aplicativo, mas um fórum onde o objetivo principal é a comunicação entre participantes e a criação e fortalecimento de ‘networks’ sociais.

5 - Considerações Finais

Uma pesquisa que tenho interesse em fazer na Adobe seria mostrar, de 5 milhões de termos retirados dos glossários da Adobe, quantos termos teriam sido mantidos no original, em inglês. Quantos teriam sido usados nas duas formas, ou seja, teriam “dupla nacionalidade” – inglês e português, dependendo do aplicativo e do contexto.

Por que teria interesse em analisar os termos que são mantidos no original? Acredito que estes termos serão possivelmente “incorporados” à língua, não só a escrita, mas também preservando a pronúncia, com sotaque português, é claro. Todos falam /mausi, egui, xis, daunlodi/. Já outros termos simplesmente não são incorporados, como ‘keyboard’. Estamos acostumados a usar ‘teclado’, há muitos anos. Ninguém usa ‘file’. Todos preferem ‘arquivo’. Através da parceria com as comunidades locais, poderíamos possivelmente chegar a um glossário mais realista, um português para produtos da Adobe, uma variação usada pelas comunidades locais de nossos usuários – uma variação do computês.

Em conclusão, a Adobe tenta desenvolver e implementar alternativas em localização com o objetivo de alcançar o melhor resultado possível para nossos clientes lusófonos em tempo real. As novas tendências de softwares desenvolvidos rapidamente, com atualizações até semanais, lançam um desafio para o trabalho de localização. Precisamos contar cada vez mais com a contribuição das comunidades locais e, ao mesmo tempo, temos que fazer um trabalho consciente de manutenção da consistência da terminologia e respeito ao idioma, tudo isso em tempo real.

Abraçamos todas e quaisquer parcerias e esperamos poder contar com os ensinamentos do IILP e de todos os participantes deste Colóquio para que a Adobe possa ter sucesso diante das comunidades de usuários lusófonos. Ao mesmo tempo, nos colocamos a disposição para ajudar a divulgar os seus trabalhos através dos nossos networks sociais.

Referências

Adobe.com no Brasil, <http://www.adobe.com/br>.

Adobe Community Translation Tool, <https://community.translate.adobe.com>.

Adobe Community Translation Facebook Page,
<http://www.facebook.com/adobe.community.translation>. Moderator: Dirk Meyer.

Adobe TV, Adobe Systems, <http://tv.adobe.com/translations>.

Blog de Globalização da Adobe Systems, <http://blogs.adobe.com/globalization/pt/>.
Moderator: Leandro Reis.

Baboo, <http://www.babo.com.br>.

Dicionário de computês
http://hugotavio.multiply.com/journal/item/1431?&show_interstitial=1&u=%2Fjournal%2Fitem.

IBM Brasil, <http://www.ibm.com/br/pt/>.

Lusofonia, <http://www.orkut.co.in/Main#Community?cmm=57996&fr=1>
Location: Portugal.

PORTUGUÊS X COMPUTÊS Comunidade ORKUT com 7.971 membros
<http://www.orkut.com/Main#CommMsgs?tid=99398&cmm=23331&hl=pt-BR>
Created: Mar 18, 2004; Location: Brazil; Owner: Vinícius Ramon Fontanela;
Moderators: Miguel Afonso Linhares, SC., Luciano Caseli, Rosi., André Simões.

RIBAS, E. et. al. 2007. A influência da linguagem virtual na linguagem formal de adolescentes
In: IX CICLO DE PALESTRAS SOBRE NOVAS TECNOLOGIAS NA EDUCAÇÃO. Anais. Rio Grande
do Sul: CINTED - Centro Interdisciplinar de Novas Tecnologias na Educação. UFRGS: 2007. p.
1-10. Disponível em:
<http://www.cinted.ufrgs.br/ciclo9/artigos/8dElisangela.pdf> .Consultado em: 23/06/2011.

Google Brasil, <http://www.google.com.br/>.

Yahoo Brasil, <http://br.yahoo.com/>.



A INTERNET EM MOÇAMBIQUE: IMPLANTAÇÃO, USO E DESAFIOS

Inês Machungo / Gregório Firmino

Sobre o autor

Inês Machungo
Universidade Eduardo Mondlane.
machungoines@aol.com.

Gregório Firmino
Universidade Eduardo Mondlane.
gregoriofirmino@gmail.com

Resumo

As tecnologias de informação e a internet em particular constituem no mundo de hoje um meio de interacção e comunicação das sociedades facilitando o seu desenvolvimento socioeconómico e cultural. Este poderoso meio de comunicação permite e facilita o diálogo intracomunitário, o diálogo governo, empresariado e comunidade, de que resulta a circulação de informação relevante para o desenvolvimento e integração das comunidades no mundo globalizado. Em Moçambique, apesar de enfrentar diversos constrangimentos, assiste-se a uma significativa implantação e uso de tecnologias de informação e comunicação e ao recurso à internet. Assim, na presente comunicação abordaremos aspectos relacionados com a implantação e uso da internet em Moçambique, dando particular atenção aos meios disponíveis de acesso à rede da internet. Destacaremos os constrangimentos que impedem um mais amplo acesso dos cidadãos à rede digital, e por fim mostraremos alguns desafios inerentes ao uso da internet no contexto moçambicano, mormente tendo em conta a situação socioeconómica e linguística do país, caracterizada entre outros aspectos, pela diversidade linguística e um elevado índice de analfabetismo.

Palavras-chave

Recursos digitais; Internet; Desenvolvimento; Multilinguismo.

1 - Introdução¹

Moçambique está situado na costa Sul Oriental de África e é linguisticamente caracterizado por ser heterogéneo; coexistem no país línguas autóctones de raiz bantu, a língua portuguesa e diversas línguas minoritárias particularmente de origem asiática. Segundo dados do último censo, realizado em 2007, a língua portuguesa é falada por cerca de 39% da população, residindo a sua maioria nas zonas urbanas; apenas 6% dos falantes do português têm esta língua como língua materna. Os restantes têm-no como língua segunda e o grau de proficiência é muito variável. É neste contexto linguístico que abordaremos as questões relacionadas com a difusão e uso das Tecnologias de Informação e Comunicação (TICs) no país.

Nesta comunicação apresentaremos um breve historial do nascimento e implantação da rede de internet em Moçambique, mostraremos o estado actual do uso da internet dando particular enfoque aos avanços e constrangimentos encontrados no processo de expansão da rede e por fim mencionaremos os desafios a enfrentar para que no país o acesso e uso da internet seja uma realidade para todos os moçambicanos.

2 - Breve historial do uso da internet em Moçambique

A história da internet em Moçambique está ligada à Universidade Eduardo Mondlane, mais particularmente ao Centro de Informática (CIUEM). Este é o provedor pioneiro da internet e correio electrónico em Moçambique; foi a partir de um computador do CIUEM que a 3 de Fevereiro de 1993 foi enviada a primeira mensagem electrónica para a Universidade de Rhodes na África do Sul; nesse ano a internet tinha cerca de 100 utilizadores, sendo na sua maioria docentes universitários. O CIUEM é até a data o responsável pelo domínio MZ. Em 1997, o CIUEM passa a oferecer o serviço completo de internet (*full internet service*) e era o único ISP (*Internet Service Provider*) em Moçambique. Novos ISP's foram surgindo e hoje é possível garantir serviços de internet a grande parte do país.

As iniciativas desencadeadas pelo CIUEM, a expansão e divulgação do conceito de TIC em Moçambique consubstanciam-se nos vários projectos criados visando a massificação e deselitização da informática no país. Destacamos a acção determinante do MAAC (Mozambique Acacia Advisory Committee) baseado no CIUEM que idealizou e liderou grande parte dos projectos ligados às TICs.

Pelo grande contributo que têm dado à divulgação e desenvolvimento das TICs no país mencionamos em particular os seguintes projectos/acções, que tiveram um certo impacto no desenvolvimento do uso dos meios digitais em vários sectores de actividade no país:

- a) Telecentros, cujo objectivo era a expansão das TICs no meio rural;
- b) Schoolnet visando a integração das TICs no processo de ensino-aprendizagem no Sistema Nacional de Educação;

¹ Para a preparação desta comunicação, contámos com a colaboração de diversas pessoas e entidades. Agradecemos, especialmente, o apoio recebido de:

Augusto Nunes - Director da INTIC
Claudino Dias - do CIUEM
Francisco Mabila - Director do CIUEM
Venâncio Massingue - Ministro da Ciência e Tecnologia .

A INTERNET EM MOÇAMBIQUE: IMPLANTAÇÃO, USO E DESAFIOS

Inês Machungo / Gregório Firmino

- c) Evolução pela comunicação e informática que tinha como objectivo reforçar o uso das TICs na governação através da formação e assistência técnica em TICs a funcionários do Governo;
- d) Tecnologia de Comunicação e Gestão Ambiental de Recursos Naturais centrado na gestão da conservação e maneio dos recursos naturais;
- e) Informação, Educação e Networking para Mulheres através da Internet. Este programa dedica-se à recolha, processamento e disseminação de informação sobre mulher e género, capacita mulheres no desenho de websites, em navegar na internet, para garantir o acesso da mulher à internet;
- f) Programa para a pesquisa e desenvolvimento de uma cultura de paz através das TICs.

O número de iniciativas e projectos em todos os quadrantes da vida social e económica foi crescendo, o que exigiu do Governo e da sociedade em geral uma atitude mais interventiva, no sentido de normalizar as diferentes iniciativas. É neste contexto que no ano de 2000 o Governo aprova a Política Nacional de Informática² que indica

o quadro de princípios e objectivos que permitirão que as tecnologias de informação e comunicação sejam o motor impulsionador dos vários aspectos do desenvolvimento nacional, contribuindo para a erradicação da pobreza absoluta e melhoria geral da vida dos moçambicanos, para a mais ampla participação dos cidadãos na Sociedade Global de Informação, para a elevação da eficácia e eficiência na prestação de serviços, para a melhoria da governação e aprofundamento da democracia, para fazer do país produtor e não mero consumidor das tecnologias de informação e comunicação, e para a sua participação na economia mundial, cada vez mais assente na informação e no conhecimento (218 – (19/10))..

São objectivos específicos da Política de Informática entre outros:

- a) Elevar a consciência nacional sobre o papel e o potencial das tecnologias de informação e comunicação no desenvolvimento sustentável de Moçambique;
- b) Contribuir para a erradicação da pobreza absoluta e para a melhoria das condições de vida dos moçambicanos;
- c) Proporcionar acesso universal à informação a todos os cidadãos para melhorar o seu nível e desempenho na educação, ciência e tecnologia, saúde, cultura, entretenimento e nas suas actividades em geral;
- d) Expandir e desenvolver o ensino da Informática no Sistema Nacional de Educação;
- e) Contribuir para aumentar a eficiência e eficácia dos sectores público e privado;
- f) Contribuir para que o país seja produtor e não apenas consumidor das tecnologias de informação e comunicação
- g) Criar um clima favorável à indústria, negócios e investimentos na área das tecnologias de informação e comunicação;

² Cf. Boletim da República, 3º. Suplemento, Política de Informática, disponível em http://www.portaldogoverno.gov.mz/docs_gov/fold_politicas/tecInform/politica_informatica_br.pdf

- h) Assegurar que os planos e projectos de desenvolvimento em todos os sectores possuam uma componente de tecnologias de informação e comunicação;
- i) Contribuir para reduzir e gradualmente eliminar as assimetrias regionais, as diferenças entre a cidade e o campo, e entre os vários segmentos da sociedade, no acesso às oportunidades de desenvolvimento;
- j) Potenciar e facilitar a integração do país na economia mundial e na Sociedade Global de Informação.

Para dar corpo a estes princípios e objectivos, foi criado o **Mozambique Information and Communication Technology Institute (MICTI)**, um programa e um plano de acção para a implementação da Política de Informática do país. Com este programa pretende-se "criar competências técnicas e capacidades em TIC em Moçambique facilitadas pela combinação de energias e contribuições do Governo, Universidades, sector privado interno e internacional, bem como de agências de desenvolvimento e de cooperação.

3 - Expansão e uso da Internet

Moçambique, apesar de ser um dos países com um índice de desenvolvimento mais baixo³, tem desenvolvido esforços para se colocar no mundo da tecnologia digital, e em particular no da internet. Contudo, à excepção da rede de telefonia móvel que se encontra em franca expansão (existem neste momento três companhias de telefonia móvel no país), são ainda insuficientes os recursos oferecidos à população para o uso das tecnologias de informação. De acordo com o Censo Populacional de 2007, o último realizado até a data, a percentagem de utentes da internet era de cerca de 0,9%. Dados estatísticos fornecidos pela Internet World Stats indicam que até Dezembro de 2011 o grau de penetração da internet em Moçambique era de 4,3%. Portanto, há indícios claros de que apenas uma percentagem muito reduzida de moçambicanos tem acesso à internet.

Num estudo intitulado *Mozambique ICT Sector Performance Review*, Mabila et al.⁴ indicam que em Moçambique existem dois meios principais de acesso à internet: através do computador (laptop ou computador de mesa) e através do telefone móvel, e em qualquer dos casos verificou-se que o uso da internet é muito baixo por variadas razões de que se destacam os custos elevados dos equipamentos, a deficiente cobertura da rede de electricidade, a reduzida largura de banda, barreiras linguísticas, entre outras.

3.1 - A telefonia móvel e a internet

Com a expansão da rede de telefonia móvel (existem neste momento no país três companhias oferecendo estes serviços), e tendo em conta que os aparelhos de telefone modernos dispõem de dispositivos que permitem o acesso à internet, era de se esperar que o número de utentes da internet aumentasse. Contudo, a situação é diversa.

Dados do estudo supracitado revelam que a percentagem de utentes de telefonia móvel em Moçambique é de 25,7%, dos quais 17,3% vivem em zonas rurais e 53% em zonas

³ Dados do Relatório Mundial de Desenvolvimento Humano, lançados pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) em 2011, indicam que o IDH de Moçambique é de 0,322.

⁴ MABILA et al. 2010.

urbanas; os números apresentados permitem inferir que 74,3% da população não tem telefone celular ou SIM Card activo. Por outro lado, os serviços de GSM e 3G não abrangem todo o país e os aparelhos de preço mais baixo têm tecnologia limitada. O mercado está pois longe de estar saturado. Assim, podemos afirmar que a maior parte da população não usa a internet disponibilizada por via de telefonia móvel, quer por ser analfabeta,⁵ quer por falta de recursos financeiros (o custo deste serviço ainda é muito elevado), quer ainda porque desconhece os serviços.

O número de pessoas que acedem à internet através do telefone celular é muito inferior ao número dos que o fazem por via do computador. E isto deve-se ao facto de que muitos dos que estão conectados à rede fazem-no no emprego, na escola, no internet café etc. O acesso através do celular é apenas um complemento, muitas das vezes usado em situações de emergência quando o acesso a partir do ponto fixo não é possível.

3.2 - Dados estatísticos sobre o uso da internet

O acesso à internet significa a possibilidade de acesso a uma informação variada sobre economia, negócios, cultura etc. Ter acesso à internet é sinónimo de poder. Contudo, este poder é dado a quem já o possui, pois para se ter acesso à internet, tem de se estar num local em que haja acesso a essa tecnologia, e tem de se possuir os recursos financeiros que permitam o seu uso, como se pode notar no quadro que se segue.

No quadro 1 apresenta-se a capacidade informática do país no ano de 2000, por província.

	C.Del	Niassa	Namp	Zamb	Tete	Man	Sof	Ibane	Gaza	MPr	MCid	Total
Comp.	65	172	402	337	348	203	880	79	118	712	8021	11516
Email	18	16	66	90	13	122	110	13	17	768	4024	5257
Intern.	9	15	50	36	13	17	78	11	14	64	2229	2536

Legenda: **Comp.** Computadores existentes – **Email** Acesso email – **Intern.** Acesso Internet

Quadro 1 - Capacidade informática do país no ano de 2000, por província

O quadro mostra que Maputo-cidade, que tem o estatuto de uma província, possui 71% dos computadores existentes no país, mais de 85% dos outros recursos, ou seja, Maputo cidade tem mais recursos que todas as outras províncias combinadas. Em Maputo está concentrada a classe média-alta do país, detentores do poder político e financeiro; em Maputo, também estão sediados a maior parte dos projectos de desenvolvimento económico-social e um número considerável de organizações internacionais. Também é em Maputo que há maior possibilidade de acesso à internet, mesmo para os cidadãos de recursos modestos.

⁵ A taxa de analfabetismo no país é de 51,9%. Note-se que apesar das inúmeras promoções feitas pelas companhias de telefone móvel para reduzir o preço das SMS, muitos dos detentores de telefones celulares, sobretudo as zonas rurais, não beneficiam destas promoções pois, não sabendo ler nem escrever, usam o telefone apenas para transmissão de voz.

De salientar que mais de 65% da população moçambicana vive nas áreas rurais. Por outras palavras, a grande maioria da população do país não tem acesso aos recursos e benefícios que as TICs oferecem.

Gillwald & Stork⁶ realizaram um estudo referente ao uso da internet no país tendo em conta algumas variáveis de que destacamos, ponto de acesso à internet, razões para o fraco uso da internet, frequência de uso e factores que limitam o uso da internet. O estudo forneceu os seguintes dados:

a) Pontos de acesso

Grande parte dos que usam a internet declararam que o faziam a partir de uma instituição educacional ou biblioteca ou no local de trabalho

Tabela 1 - Pontos de acesso

Ponto de acesso	% ¹
Instituição educacional ou biblioteca	35,4%
Local de trabalho	33,3%
Casa ou em casa de outros	32,1%
Internet-café	27,9%
Telefone-móvel	21,72%

Fonte: GILLWAD & STORK, 2008.

b) Fraco uso da internet

As razões apresentadas para o fraco uso da internet por parte daqueles que sabem o que é a internet, foram no geral ou a falta de computador ou a falta de domínio de uso do computador.

Tabela 2 - Razões do fraco uso da internet

Razões do fraco uso da internet	%
Não tem computador	68,5%
Não sabe usar computador	33,6%
Não tem com quem comunicar via email	20,6%
Não tem acesso à internet	13,8%
Não dispõe de recursos financeiros	7,4%
Não quer usar a internet	7,1%
Não sabe ler ou escrever	0,3%

Fonte: GILLWAD & STORK, 2008.

⁶ GILLWALD & STORK, 2008.

⁷ Era permitido que os inquiridos indicassem mais do que uma opção de resposta. Daí que a soma dos valores percentuais possa ser superior a 100%.

c) Frequência de uso da internet

Os utentes da internet indicaram, na sua maioria, que usavam a internet todos os dias ou quase todos os dias.

Tabela 3 - Frequência do uso da internet

Frequência do uso da internet	%
Todos os dias ou quase todos os dias	60,0%
Pelo menos uma vez por semana	23,7%
Menos de uma vez por mês	8,3%
Pelo menos uma vez por mês	7,9%

Fonte: GILLWAD & STORK, 2008.

d) Factores limitativos

São vários os factores limitativos, no uso da internet, podendo-se destacar a falta de tempo, conteúdo sem interesse, falta de acesso a um computador com acesso à internet, ou o custo elevado do acesso à internet.

Tabela 4 - Razões do fraco uso da internet

Razões do fraco uso da internet	%
Falta de tempo	29,7%
Não tem conteúdo interessante	22,4%
Não tem acesso computador com ligação à internet	21,1%
Custo de acesso elevado	19,9%
Velocidade baixa da internet	14,6%
Não sabe como usar a internet	1,6%
Não há conteúdos em línguas locais	0,8%

Fonte: GILLWAD & STORK, 2008.

Se adicionarmos as percentagens relativas às variáveis **instituição educacional** e **trabalho**, teremos cerca de 68% de utentes da internet usando os serviços disponibilizados pelos locais de trabalho e estudo, que como sabemos não imputam custos ao uso da internet. Isto pode significar que a maioria dos utentes não tem recursos financeiros para adquirir um computador ou para pagar o custo dos serviços. De notar que parte dos que afirmaram usar a internet em casa, fazem-no também em casa de outros. Esta interpretação é sustentada pelos dados obtidos nas respostas na tabela 2, em que cerca de 68% dos utentes afirma não usar a internet por não possuir computador.

Se compararmos as respostas apresentadas na tabela 2 (em que lugar se acede à internet) com as respostas à pergunta 4 (factores que limitam o uso da internet) poderemos concluir que a falta de tempo manifestada pelos utentes se deve ao facto de estes usarem

a internet no local de trabalho e que a informação que procuram não está relacionada com o trabalho. Veja-se que ainda nas respostas na tabela 4 os inquiridos indicam que a internet não tem conteúdo interessante. O que procuram os moçambicanos na internet? Esta é uma pergunta cuja resposta nos ajudaria a esclarecer o facto de uma boa parte dos utentes da internet acharem que nesta não são veiculados conteúdos interessantes.

Assim, pode-se inferir que actualmente:

- a) são poucos os que usam a internet em casa; grande parte usa ou no local de trabalho ou na escola ou no internet-café.
- b) Os que a usam fazem-no todos os dias ou quase todos os dias.
- c) Os que não usam não o fazem porque não têm computador ou não sabem usar.
- d) De entre os que usam há factores limitativos como falta de tempo, conteúdo pouco interessante, falta de acesso à internet, custo e baixa velocidade da internet.

3.3 - Instituições que gerem o uso da internet

Falar da internet em Moçambique é falar dos meios que permitem o seu uso e das instituições que gerem e promovem a sua divulgação. Várias são as instituições no país que fornecem meios de acesso à internet, sendo as instituições do ensino superior, pela sua natureza, as que oferecem mais opções e a preços reduzidos ou grátis na maioria dos casos. Contudo, na presente comunicação destacaremos o Instituto Nacional de Tecnologias de Informação e Comunicação (INTIC) por ser uma instituição com representação em todo o país e que presta serviços mais abrangentes.

O INTIC é uma instituição que gere e administra a Rede Electrónica do Governo (GovNet) através da qual funcionários de cerca de 250 instituições públicas de todo o país, acedem à internet. Esta instituição disponibiliza o Portal do Governo através do qual, a partir de qualquer parte do Mundo, se pode aceder a informação sobre o Governo de Moçambique. Este Portal dispõe também de informação sobre os diversos serviços prestados ao cidadão pela função pública.

O Portal do Governo é também usado como meio de:

- a) acesso do cidadão ao correio electrónico, através do webmail;
- b) participação do cidadão na governação, contribuindo para o reforço da democracia participativa.

O INTIC actua a nível central, provincial e distrital.

3.3.1 - Província

Em cada capital provincial do país existe um *Centro Provincial de Recursos Digitais* (CPRD) ligado à rede central e que tem por objectivo servir a função pública na província. Os CPRD estão também vocacionados para servir o cidadão comum, massificando assim o uso da internet e mais particularmente do correio electrónico. Os CPRD oferecem entre outros:

- a) Sala de internet, em que sendo subsidiada, a internet é oferecida ao cidadão comum a preços mais baixos que o de outras instituições afins;

- b) Serviços de formação, apoio e assistência técnica;
- c) Nestes centros, através dos voluntários, também se faz sensibilização aos futuros utilizadores sobre as vantagens do uso da internet.

Cerca de 80% das províncias do país têm um CPRD em funcionamento.

3.3.2 - Distrito

Em nível das sedes dos distritos, foram criados *Centros Multimédia Comunitários* (CMC), onde o cidadão pode ter 'acesso partilhado' à internet; nestes centros, para além dos pontos de acesso à internet, funcionam também rádios comunitárias. Os CMC são de grande relevância para a vida do cidadão, pois a partir destes são veiculados assuntos de interesse comunitário; por exemplo, são lançadas campanhas de divulgação de assuntos ligados à saúde (vacinações, epidemias etc), à educação, às campanhas agrícolas entre outros. Os documentos são baixados por via da internet e são traduzidos e divulgados em línguas locais através da rádio comunitária. Equipes de voluntários, maioritariamente estudantes, dão apoio aos utentes da internet, para além de fazerem campanhas de divulgação e formação, visando o uso mais frequente e adequado dos recursos digitais disponíveis. Os CMC estão já instalados em cerca de 40% dos distritos do país encurtando assim as distâncias entre as capitais provinciais e os distritos e entre os distritos entre si; desta forma se reduzem significativamente os custos e outros recursos que seriam usados em transportes quer de pessoas quer de materiais.

Ainda no âmbito da divulgação e facilitação de comunicação por via da internet e dentro do Programa das TICs para o desenvolvimento, foram criadas *Unidades Móveis de Tecnologias de Informação e Comunicação* (UNTIC); estas têm por função formar residentes dos distritos, promover conhecimento e serviços na área das novas tecnologias de informação.

O INTIC gere também o Sistema Centralizado de correio electrónico do governo meio seguro e fiável de comunicação electrónica para membros do Governo e outros titulares de cargos públicos, sendo os maiores utilizadores deste serviço os Ministérios e Governos Provinciais.

4 - Língua e uso da internet

Tendo em conta que mais de 70% dos utentes da internet no país são professores, estudantes e funcionários públicos, a título de amostragem, procurámos identificar a língua por eles usada nas páginas da internet visitadas e a língua usada nos browsers utilizados. Para tal analisámos dados fornecidos pelo INTIC e pelo CIUEM referentes ao período que vai de 1 a 31 de Janeiro de 2012, tendo-se apurado que:

- a) Os browsers mais usados estão em português;
- b) Há uma grande procura de sites em inglês mas o tempo de permanência de visita é inferior ao dos sites em português.
- c) Em relação aos sites em português os mais visitados são os brasileiros e principalmente as páginas de divulgação de trabalhos académicos;

Estes dados demonstram claramente que, no contexto moçambicano, o português é fundamental no acesso à internet, o que se harmoniza com o cenário linguístico do país. Apesar de o inglês ser importante, de estar entre as línguas usadas no meio digital, o português é a língua melhor dominada pelos moçambicanos para o acesso a internet.

5 - Constrangimentos no acesso e uso da internet

Muitos são os avanços verificados ao longo dos últimos 10 anos, mas existem igualmente muitos constrangimentos. Dados recolhidos dos relatórios apresentados pelas CPRD e CMC indicam que algum do equipamento informático instalado nos centros não está em uso, por se encontrar danificado devido à falta de manutenção; os centros não dispõem de técnicos especializados para garantirem o seu funcionamento adequado. As oscilações de corrente eléctrica e a falta de equipamento adequado de protecção são outros dos factores que contribuem para o desgaste do equipamento. Pode-se pois inferir que a disponibilização de equipamento não é suficiente, outros recursos, particularmente os humanos, têm de estar disponíveis.

O elevado preço dos computadores e telefones celulares constitui também um entrave a ter em conta. Moçambique tem um PIB de \$480 USD e um computador custa cerca de \$700 USD; o preço do telefone celular com capacidade instalada para aceder à internet custa, no mínimo, o equivalente a \$30 USD.

A largura da banda é limitada, do que resulta uma deficiente qualidade na transmissão de dados. O elevado custo dos serviços é outro factor que reduz a possibilidade de o cidadão comum usar a internet.

A rede de distribuição de energia eléctrica ao longo do país é deficiente, o que afecta grandemente o fornecimento de serviços de internet em grande parte dos distritos e localidades.

Factores de outra índole contribuem também para o fraco acesso e uso da internet pelos moçambicanos. Referimo-nos particularmente à taxa elevada de analfabetismo, às dificuldades linguísticas particularmente no uso do inglês, e também do português que como se sabe, apesar de ser língua oficial, não é dominada por cerca de 75% da população, à iliteracia digital e ao desconhecimento em geral das vantagens oferecidas por estes recursos do mundo digital.

6 - Desafios e oportunidades

As TICs são um poderoso meio de inclusão social. Ao estabelecerem redes de comunicação possibilitam a interacção entre as pessoas, facilitam transacções quer económicas quer sociais, fortificam e consolidam as relações entre os cidadãos, governo, empresariado e outras forças vivas dentro e fora do país. São um motor de desenvolvimento económico e social das sociedades. Ter acesso às TICs significa ter acesso ao mundo global da informação e conhecimento. O panorama demográfico, linguístico, económico e social de Moçambique no presente momento revela que as TICs propiciam a exclusão social, impedindo os cidadãos de adquirirem conhecimentos que lhes permita integrarem-se no mundo globalizado em que vivemos.

Para reverter esta situação, desafios se impõem à sociedade, alguns já identificados na Política Nacional de Informática, nomeadamente:

- a) Proporcionar acesso universal à informação a todos os cidadãos para melhorar o seu nível e desempenho na educação, ciência e tecnologia, saúde, cultura, entretenimento e nas suas actividades em geral;
- b) Expandir e desenvolver o ensino da Informática no Sistema Nacional de Educação;
- c) Reduzir os custos de acesso à internet (um pacote básico de internet custa cerca de \$35USD mensais);
- d) Identificar conteúdos virados para as aspirações dos moçambicanos;
- e) Identificar formas de ultrapassar as barreiras linguísticas (os moçambicanos são fluentes nas línguas locais e/ou português).

7 - Considerações finais

O presente estudo mostrou que os objectivos plasmados na Política de Informática traçada pelo Governo e que visavam fundamentalmente o acesso dos cidadãos moçambicanos às TICs estão longe de ser concretizados. Vários factores que anteriormente mencionámos e que repetimos concorrem para esta situação:

- a) A taxa elevada de analfabetismo, que impede os cidadãos de terem acesso à escrita recurso por excelência utilizado nas TICs; associada ao analfabetismo está a barreira linguística, pois como se mencionou acima uma percentagem reduzida da população tem domínio do português ou do inglês, línguas usadas nos meios digitais.
- b) O preço elevado dos computadores, e as taxas de acesso à internet que são incompatíveis com os recursos financeiros dos cidadãos;
- c) O desconhecimento das TICs, facto este associado também às condições socio-culturais dos moçambicanos;
- d) Deficiente cobertura da rede de electricidade e internet.

Para que Moçambique entre de facto no mundo da internet e no mundo digital em geral é preciso que:

- a) Se desenvolvam campanhas de combate à iliteracia incluindo a digital;
- b) Se garanta a massificação do acesso ao computador, ao telefone, aos serviços da internet;
- c) Se garanta que os conteúdos dos sites disponíveis sejam relevantes e no formato linguístico adequado;
- d) Se criem pacotes de incentivos para quem investe na TIC;

e) O país disponha de recursos financeiros e técnicos que permitam a implementação de qualquer decisão que seja tomada relativamente às TICs.

Em suma, é necessário que se envidem esforços para que sejam verdadeiramente implementados os princípios plasmados na política de informática.

Referências

GILLWALD, A.; STORK, C.2008. *Towards Evidence-based ICT Policy and Regulation: ICT Access and Usage in Africa*.Vol.1, Policy Paper.

MABILA, F. et al.2010. Mozambique ICT Sector Performance Review 2009/2010. In: *Towards Evidence- based ICT Policy and Regulation*, Vol. 2, Policy Paper 16, 2010.

MATUSSE, R.2003. *História da Informática em Moçambique*. Maputo : MAAC.

Portal do Governo de Moçambique, <http://www.portaldogoverno.gov.mz>



CONTRIBUTO GALEGO PARA A LÍNGUA PORTUGUESA NA ERA DIGITAL: AS FERRAMENTAS AO-OA DA ACADEMIA GALEGA DA LÍNGUA PORTUGUESA

Ernesto Vazquez Sousa / Concha Rousia

Sobre o autor

Ernesto Vazquez Souza
Academia Galega da Língua Portuguesa; AGLP-Arquivo Digital.

Concha Rousia
Academia Galega da Língua Portuguesa.

Resumo

Apresentação sintética dos projetos digitais que a Academia Galega da Língua Portuguesa está a desenvolver para a promoção da Língua Portuguesa na Galiza e para o conjunto da Lusofonia. Em destaque o Portulano de Recursos: escritório virtual e agrupador de recursos digitais sobre a língua portuguesa desenhado com o software livre Netvibes; o Arquivo Digital: repositório de objetos digitais orientado à preservação e divulgação que funciona sobre um motor DSpace e gerencia metadados no standard Dublin Core; e o projeto do Dicionário eletrônico e-estraziz em parceria com a AGAL (Associação Galega da Língua).

Palavras-chave

AGLP. Dspace. Repositórios. Redes sociais. Agrupador de recursos.

1 - Introdução

A **Academia Galega da Língua Portuguesa** é uma entidade privada que se define estatutariamente como “instituição científica e cultural ao serviço do Povo galego” que pretende “promover o estudo da Língua da Galiza para que o processo da sua normalização e naturalização seja congruente com os usos que vigoram no conjunto da Lusofonia”.

A Academia Galega da Língua Portuguesa foi constituída em 20 de setembro de 2008. Desde a sua fundação e na atualidade é presidida pelo Professor Doutor José-Martinho Montero Santalha e apresenta-se como uma continuação histórica da ideia de unidade do galego-português que representaram vultos como Ernesto Guerra da Cal, Ricardo Carvalho Calero, Manuel Rodrigues Lapa ou Luís Filipe Lindley Cintra, que em 1984 incluíra os dialetos da língua galega como parte dos do português europeu (CINTRA, 1971) e o consagrou na Gramática que editou junto de Celso Cunha (CUNHA; CINTRA, 1984).

A ideia de criação de uma Academia de caráter lusófono na Galiza foi do Professor Doutor Ricardo Carvalho Calero (1910-1990). A proposta de criação da Academia foi defendida pelo e tarefas para o futuro”¹ e, mais recentemente, numa intervenção em Bragança, em outubro de 2006, na realização do V Colóquio da Lusofonia.

Criada seguindo a tradição das academias, mas como uma iniciativa da sociedade civil, independente dos organismos políticos galegos com o intuito de efetivar esta proposta, tem sua origem ligada ao nascimento, em 1 de Dezembro de 2007 da Associação Pró-Academia Galega da Língua Portuguesa em Compostela integrada por pessoas de diferentes âmbitos da defesa da língua e com o apoio de organizações lusófonas como a Associação de Amizade Galiza-Portugal, a Associação Galega da Língua, a Associação Sócio-Pedagógica Galego-Portuguesa ou o Movimento Defesa da Língua.

Posteriormente, em 7 de abril de 2008, foi confirmada pelo presidente da Associação Cultural Pró Academia Galega da Língua Portuguesa durante a sua intervenção na Assembleia da República de Portugal, na Conferência Internacional de Lisboa, em 7 de abril de 2008 sobre o Acordo Ortográfico da Língua Portuguesa por meio do seu Presidente, Ângelo Cristóvão², com um comunicado a respeito da posição galega e o papel da futura Academia Galega. Neste ato também o resto de associações lusófonas galegas acima citadas estiveram representadas por meio do Presidente da AGAL (Associação Galega da Língua), Alexandre Banhos. Estas intervenções tiveram larga repercussão nos meios de comunicação.

Finalmente, a Academia foi constituída em 20 de setembro de 2008, realizando a sua sessão inaugural em 6 de outubro de 2008 contando com a participação e respaldo de instituições como a Academia das Ciências de Lisboa, a Academia Brasileira de Letras, o Reitor da Universidade Aberta, a Universidade de Santiago de Compostela, a Associação Internacional de Lusitanistas e a Junta da Galiza (Governo Autónomo da Galiza), que enviaram representantes oficiais.

A Academia é um órgão estatutário da Fundação Academia Galega da Língua Portuguesa, entidade de competência estatal inscrita no Ministério da Cultura da Espanha (Ordem CUL/1075/2011, de 1 de março).

¹ MONTERO SANTALHA, 1994.

² CRISTOVÃO, 2008.

Em 17 de maio de 2010 recebeu o Prémio Meendinho, atribuído pela fundação com o mesmo nome, "por existir um antes e um depois, desde a sua constituição, a respeito da projeção e da realidade da língua galega na Lusofonia toda, e como um elemento mais dela".

A Academia Galega da Língua Portuguesa tem estabelecido **Protocolos de Cooperação** com 25 entidades portuguesas, brasileiras e angolanas:

Em Angola:

- Fundação António Agostinho Neto (Luanda) (<http://www.agostinhoneto.pt/>)

No Brasil:

- Instituto Cultural Brasil-Galiza (Florianópolis-Santiago) <http://www.brasil-galiza.org>
- Instituto de Investigação e Desenvolvimento em Políticas Linguísticas (IPOL) (Florianópolis - Sta. Catarina) <http://www.ipol.org.br/>
- Instituto Histórico e Geográfico Brasileiro (Rio de Janeiro) <http://www.ihgb.org.br/>
- Associação Brasileira de Linguística (Curitiba - Paraná) <http://www.abralin.org/>
- Academia Brasileira de Música (Rio de Janeiro) <http://www.abmusica.org.br/>
- Real Gabinete Português de Leitura (RJ) <http://www.realgabinete.com.br/portaWeb/>

Em Portugal:

- Colóquios da Lusofonia (Açores) <http://www.lusofonias.net/>
- Academia das Ciências de Lisboa <http://www.acad-ciencias.pt>
- Academia Portuguesa da História (Lisboa) <http://www.aph.com.pt/>
- Academia Internacional de Cultura Portuguesa (Lisboa) <http://www.portaldacultura.gov.pt/organismos/Pages/academias.aspx>
- Sociedade de Geografia de Lisboa (Lisboa) <http://www.socgeografialisboa.pt/>
- Universidade Aberta (Lisboa) <http://www.univ-ab.pt/>
- Centro de Estudos de História do Atlântico (Funchal, Madeira) <http://www.ceha-madeira.net/>
- Priberam Informática, s.a. (Lisboa) <http://www.priberam.pt/>
- Porto Editora <http://www.portoeditora.pt>
- Instituto de Filosofia Luso-Brasileira (Lisboa) <http://iflb.webnode.com/>
- Sociedade da Língua Portuguesa <http://www.slp.pt>

- ISCTE - Instituto Universitário de Lisboa <http://www.iscte.pt>
- Instituto Luso-Árabe para a Cooperação <http://www.luso-arabe.org/>
- Centro Nacional de Cultura <http://www.cnc.pt/>

A Academia Galega da Língua Portuguesa colabora ativamente com o resto de academias da língua, especialmente através da sua Comissão de Lexicologia e Lexicografia, que elaborou um vocabulário de léxico galego para ser incorporado no próximo Vocabulário Ortográfico Comum, decorrente da implementação do Acordo Ortográfico, e cuja coordenação é responsabilidade do Instituto Internacional da Língua Portuguesa.

A Sessão Interacadémica realizada em 14 de abril de 2009 em Lisboa sob a presidência das três academias foi a primeira cerimónia conjunta, apresentando-se publicamente o Léxico da Galiza a ser integrado no Vocabulário Comum.

Mais de 800 entradas do Léxico da Galiza foram incorporadas no *Vocabulário Ortográfico da Língua Portuguesa*³ editado em outubro de 2009 pela Porto Editora. Em março de 2010 a Academia Galega da Língua Portuguesa participou da Conferência Internacional sobre o Futuro da Língua portuguesa no Sistema Mundial realizada no Itamaraty, Brasília, por convite do Ministério de Relações Exteriores do Brasil.

Junto à atividade institucional, a publicação do *Léxico da Galiza* e a sua inclusão nos vocabulários ortográficos da Porto Editora e da Priberam marcam uma linha de atuação que vai ser intensificada nos próximos anos.

Em 25 de setembro de 2012 decorreu em Santiago o **II Seminário de Lexicologia**, com a participação institucional de 12 especialistas, com especial destaque para os professores Evanildo Bechara e Malaca Casteleiro. Participaram as seguintes entidades: Academia das Ciências de Lisboa (ACL), Academia Brasileira de Letras (ABL), Associação Docentes de Português na Galiza (DPG), Instituto Camões, Porto Editora e Priberam Informática.

Neste evento foi apresentado o programa informático FLiP 8⁴, que incorpora na sua configuração a opção de uso do Léxico da Galiza, além dos pertencentes aos outros países de língua portuguesa. Deste modo, o nosso contributo lexical é divulgado junto do grande público de uma forma prática. O arquivo foi disponibilizado na página da Academia (www.aglp.net).

A Academia Galega da Língua Portuguesa também articula uma Coleção de Clássicos Galegos com obras de autores como Rosalia de Castro, Eduardo Pondal, João Vicente Biqueira, Manuel Leiras Pulpeiro, Lois Amado Carvalho, editando também de forma periódica anual um *Boletim da Academia Galega da Língua Portuguesa* (ISSN 1888-9763) e outras publicações científicas como o *Anexo Galiza: Língua e Sociedade* (2009).

³ O Vocabulário Ortográfico da Língua Portuguesa da Porto Editora com mais de 250 000 vocábulos representativos do vasto património lexical da língua portuguesa é acessível em: <http://www.infopedia.pt/vocabulario/>.

⁴ O FLiP 8 é a oitava versão do pacote de ferramentas de revisão e auxílio à escrita da Priberam que inclui um corretor ortográfico, um dicionário e um conversor para o Acordo Ortográfico.

2 - As Ferramentas digitais da AGLP

Moramos num mundo definido pelo uso generalizado de tecnologias que passaram em pouco tempo de dar acesso em tempo real à informação a uma procura da plena comunicação entre **comunidades de utentes** que geram e produzem não apenas conteúdos quanto que personalizam, modificam, definem ou gerem as próprias ferramentas ou os meios em que se produz a comunicação.

A rede, as redes, são naturalmente sociais, mas definem novos agrupamentos humanos, novas fidelidades, novas necessidades individuais e grupais, que não têm a ver com os espaços geográficos ou com as fronteiras nacionais ou educativas quanto com as preferências culturais, e as identidades ideológicas, os valores, as fidelidades.

No decurso deste desenvolvimento a língua portuguesa passou de ser a língua nacional de uns quantos países em via de desenvolvimento a canal de comunicação pelo que circulam e do que fazem uso estas novas comunidades. Uma língua que se movimenta num espaço global como uma das possíveis línguas de uso internacional avalizadas por uma das grandes potências emergentes do presente: o Brasil.

Nesse espaço internacional e social que se define pela internet, os conceitos de Lusofonia e Acordo Ortográfico tomam uma forma e valor que vai definir e no que podem entrar os espaços galegos criativos, comunicativos, económicos e sociais.

Desde a década de 90 do século XX, por causa do processo de fagocitação das grandes editoras científicas e pela implementação dos sistemas de avaliação baseados nos índices de impacto e de citação dos textos académicos, produziu-se uma constante redução do espaço de publicação científico e o temor ao controlo deste por parte de grupos editoriais movimentados apenas pelos seus interesses económicos e estratégias editoriais.

Segundo Hélio Kuramoto⁵, a partir desta perceção surgiu o movimento *Open Access* cuja primeira declaração foi elaborada e lançada como resultado de uma reunião realizada nos dias 1 e 2 de dezembro de 2001, organizada pela Open Society Institute (OSI), o Budapest Open Access Initiative (BOAI).

É um movimento que defende o acesso aberto a artigos de pesquisas, principalmente através de meios digitais. Acesso sem restrições, *online*, livre de qualquer cobrança de taxa ou necessidade de assinatura ou pagamento de licenças.

A declaração Budapest Open Access Initiative recomendava duas estratégias complementares:

1. Autoarquivamento; e
2. Revistas Científicas de Acesso Livre.

A primeira estratégia considera que os pesquisadores e académicos precisam de ferramentas e assistência para depositar os seus artigos publicados em revistas arbitradas por revisão de pares (peer-review) em arquivos eletrónicos abertos, prática também chamada de autoarquivamento. Considerando que esses arquivos estejam em conformidade com as

⁵ KURAMOTO, 2012.

normas criadas pelo Open Archives Initiative (OAI), em seguida, os motores de busca e outras ferramentas podem tratá-los separadamente como sendo um só. Desta forma, os usuários, não precisarão saber quais arquivos existem ou onde eles se localizam, a fim de encontrar e fazer uso de seu conteúdo.

O acesso aberto através da rede é uma excelente solução para a escassez de recursos, os custos elevados para manter e atualizar grandes bibliotecas, a dificuldade em manter um número elevado de assinaturas e a existência de poucos exemplares de obras para muitos leitores.

Deve considerar-se também que hoje é cada dia menos necessário o acesso físico ao documento ou a propriedade, primando a facilidade (e ainda a maneira) de acesso à informação e as facilidades de pesquisa em seu conteúdo. Acompanhado isto com os baixos custos editoriais resulta a possibilidade de difusão e intercâmbio de resultados de pesquisa a um público incomparavelmente maior.

O impacto potencial que o acesso às redes de comunicações eletrónicas pode ter sobre a concorrência e a ampla disponibilidade de redes e serviços da rede é um princípio fundamental que sustenta o enquadramento normativo europeu. Na Europa, os operadores tradicionais receberam o mandato de facilitar o acesso aos operadores que competem com o fim de reduzir as barreiras à entrada e para promover a prestação competitiva de serviços de comunicações eletrónicas em benefício dos consumidores europeus⁶.

No caso galego a aparição e espalhamento da internet e a intervenção e comunicação dos galegos nos espaços de língua portuguesa provoca uma reestruturação da imagem de nós e da língua como múltipla, global e ferramenta internacional e local de comunicação (registros – âmbitos – ambientes – comunidades utentes).

No eixo desta dupla AO-OA (Acordo Ortográfico, Movimento Open Access)⁷ a Academia Galega da Língua Portuguesa, desde a sua fundação em 2008, dentro da sua definição antes citada como instituição científica e cultural ao serviço do Povo galego que pretende promover o estudo da Língua da Galiza para que o processo da sua normalização e naturalização seja congruente com os usos que vigoram no conjunto da Lusofonia, tem ido desenvolvendo e colaborando como parceiro em diversas ferramentas e recursos digitais.

1. Academiagalega.org.- desde 2009 e em parceria com a Associação Cultural Pro-Academia Galega da Língua Portuguesa, o sítio foi desenhado pela empresa Eomática para informar das iniciativas, atividades da Academia Galega da Língua Portuguesa, informações e notícias sobre a Lusofonia. Permite aceder a todas as ferramentas, sites e publicações da AGLP, nomeadamente, Boletim em pdf e Léxico da Galiza para ser integrado no Vocabulário Ortográfico Comum da Língua Portuguesa.⁸

⁶ BEREC, 2011.

⁷ O acesso aberto (Open Access) à ciência é o movimento que persegue a disponibilidade livre e gratuita dos conteúdos científicos em internet. Apesar de tratar de uma iniciativa conhecida, existem ainda ideias confusas e também más interpretações sobre alguns dos seus objetivos, características ou funcionamento. Os fundamentos e antecedentes do acesso aberto devem entender-se como uma mudança radical no funcionamento do sistema de comunicação científica, envolvendo uma análise da situação das revistas científicas e os repositórios (que constituem as duas vias estabelecidas para conseguir esta mudança de modelo), as feições legais relacionados com a informação científica, a atitude dos cientistas a respeito do acesso aberto e as políticas desenvolvidas para facilitar a extensão deste modelo. (cf. Abadal, 2012).

⁸ A inclusão de léxico da Galiza no FLiP 8 da Priberam responde ao trabalho da Comissão de Lexicologia e Lexicografia da Academia Galega da Língua Portuguesa. O Protocolo de Cooperação assinado com a Priberam implica que mais conteúdos diferenciais do português galego serão acrescentados à língua comum através dos produtos desta empresa; ver: <<http://www.academiagalega.org/info-atualidade/portugues-da-galiza-no-flip-8.html>>.

O léxico da Galiza é um documento elaborado pela Comissão de Lexicologia e Lexicografia da Academia Galega da Língua Portuguesa, que foi apresentado na Academia das Ciências de Lisboa o dia 14 de abril de 2009, e atualmente integra 2000 vocábulos de uso frequente na Galiza, ainda não integrados nos dicionários gerais da língua portuguesa.

Este contributo foi acrescentado ao *Vocabulário Ortográfico da Língua Portuguesa* da Porto Editora, lançado em outubro de 2009, e no programa FLiP-8 da empresa Priberam, que foi disponibilizado em agosto de 2010. Os utentes podem utilizar o corretor ortográfico com as diversas variedades do português

2. Web-sites associados, em parceria com diversas entidades e associações civis, académicas e culturais galegas e internacionais.

2.1. Ricardo Carvalho Calero.- homenagem no centenário (2010).

2.2. Ernesto Guerra da Cal.- homenagem no centenário (2011).

2.3. Signum.- suporte ao site da Associação de académicos e especialistas internacionais para promover estudos da Marca.

3. Escritórios Netvibes⁹ - agrupadores de recursos e ferramentas formativas compartilháveis livremente dentro da filosofia da Web 2.0 e Web 3.0. Desenhados com Netvibes, desktop online baseado na tecnologia AJAX com um alto nível de personalização.

3.1. Escritório AGLP. informações e links institucionais, com ligações institucionais, parcerias e informações. http://www.netvibes.com/aglp#AGLP_-_Apresentação

3.2. Survival Kit - estojo virtual que visa oferecer de forma rápida e intuitiva os mais úteis recursos de apoio para a aprendizagem e o uso correto do Português.

É um escritório muito simples e útil, que pode integrar-se em qualquer computador com ligação à rede, desenhado para neofalantes, rapazes de liceu e gente que sabendo galego se acheque ao português.

Com melhor estética e muito mais simples em 2-3 abas reúne numa principal os melhores e mais efetivos recursos. Inseriu-se uma caixinha de multipesquisador (Google, Wiki pt...), outra de vídeos, um FLiP corretor, um conjugador de verbos, um dicionário de pt, o Estraviz, o Léxico, uma ligação a alguma biblioteca de clássicos em Português (BN e B Domínio Público...) uma ligação ao PGL e outra ao Portulano (AGLP). http://www.netvibes.com/academia-galega-da-lingua-portuguesa#Survival_Kit

3.3. Portulano de Recursos - na imagem das velhas Cartas portulanos, este escritório é uma solução para navegar pela rede num momento em que a expansão de usos, sites e TIC, está a produzir uma absoluta mudança na percepção do conhecimento.

Agrupa recursos em linha, enlaces, catálogos, dicionários, corretores, bibliografias, ligações a bibliotecas virtuais, repositórios, e-books e informações de utilidade para bom uso e aprendizagem da língua portuguesa. Ao jeito dos velhos Portulanos o escritório desenha percursos que permitem navegar a língua com a maior certeza por rotas de destino a portos cheios de mercadorias.

⁹ Apresentaram-se estas ferramentas no II Seminário de Lexicologia.

É uma ferramenta desenhada pela Academia Galega da Língua Portuguesa como agrupador de recursos e ferramentas formativas online, compartilháveis livremente dentro da filosofia da Web 2.0 e Web 3.0.

Desenhado para uso da sociedade dentro dos princípios fundacionais da AGLP, está especialmente dirigido a todos os galegos que têm interesse na língua portuguesa e na cultura lusófona e para todos os lusófonos com curiosidade sobre a Galiza.

Está estruturado nesta ordem sequencial:

- Recursos em língua portuguesa: dicionários, corretores ortográficos, conjugação verbal, recursos em tradução, gramáticas, dicas para um português correto como o *falar com jeito*; buscadores e ferramentas para a pesquisa lexicográfica.
- Recursos gerais de informação; catálogos bibliográficos, repositórios científicos, revistas e bibliotecas online; enciclopédias em linha, glossários especializados, toponímia, onomástica, etimologia e muito mais...
- Vária: RSS, debates, blogues e sites, jogos de língua, destaques.

Além disto também incorpora informação sobre a Galiza e o seu projeto reintegrante de incorporação lusófona. http://www.netvibes.com/portulano#Portulano_de_recursos

4. Arquivo Digital-Repositório da AGLP - desde 2011, é um repositório de objetos digitais ou acervo de coleções de informação em suportes digitais desenhado com o motor de Software livre DSpace, criado pelo MIT, e com interface Manakin criada pela Biblioteca da Universidade do Texas.

A interface pode ser extensivamente modificada pelos "aspetos" e "temas" baseado em XSL. O DSpace é disponibilizado livremente às instituições de investigação, sob a forma de um produto de código aberto, que pode ser livremente adaptado e expandido funcionalmente, nos termos da BSD Open source license.

Aceita todas as formas de materiais digitais, arquivos de texto, imagem, vídeo e áudio, o que possibilita custodiar os mais variados tipos de conteúdos.

Os repositórios digitais são coleções de informação em suporte digital (não apenas documentos mas objeto / item). E estão associados às ideias de:

- Suportes digitais (descarregáveis)
- Tecnologias digitais
- Produção e transmissão digital (não meros documentos tradicionais digitalizados)
- Conteúdos audiovisuais
- Objetos de aprendizagem
- Metadados, procura e etiquetagem automática

- Tags: arquivo compartilhar formatos e-learning

Um repositório visa ser parte da Web semântica, a web “inteligente”, capaz de conceder um significado a um arquivo que será disponibilizado para outros utilizadores, podendo ser usado como fonte de pesquisa.

Os documentos são organizados em comunidades e coleções e integrados por meio das aplicações de Dublin Core, esquema de metadados desenvolvido pela Dublin Core Metadata Initiative (DCMI) para promover a adoção de padrões de metadados, que visa descrever objetos digitais utilizando XML (Extensible Markup Language) e o RDF (Resource Description Framework).

Os metadados são as informações que ficam embutidas nos formatos digitais e permitem aos autores e aos editores adicionarem dados sobre tipologia e características, como, quando e quem gerou os documentos. São informações sobre a forma e conteúdo que facilitam na WEB as buscas, o compartilhamento e a captura de arquivos entre sistemas. São desenvolvidos a partir (e em função) de dados, por isto que é designado como “dados sobre dados” ou “informação sobre a informação”.

A importância dos metadados para a web semântica está basicamente ligada à facilidade de recuperação e interpretação dos dados por parte de humanos e de computadores.

As possibilidades de uso são:

- Capturar e descrever documentos digitais de acordo com um *workflow* adaptável aos processos específicos de uma comunidade de usuários.
- Distribuir os documentos digitais da instituição na Web, possibilitando a pesquisa e obtenção de cópias aos utilizadores,
- Preservar os documentos digitais a longo prazo.
- Possibilidade de enriquecer o documento com metadados facilitando a procura.

O arquivo organiza-se hierarquicamente em categorias e coleções que agrupam conjuntos documentais que integram por sua vez registros (simples/completos) e que se completa com uma caixa de procura.

- Categorias no repositório
 - Coleções
 - Conjuntos documentais
 - Registro (simple/completo)
 - Arquivos no item (ficha do registro)
 - Relações com outros itens

A “ficha catalográfica” de um objeto (os metadados que serão acrescentados a ela) é um registro eletrônico que contém descrições desta e que permitem que se saiba do que se trata sem ter que se ler ou ouvir todo o seu conteúdo. O registro seria uma representação do documento com as suas relações com outros documentos.

Os conteúdos gravações audiovisuais, áudios, fotografias e documentos pdf apresentam-se sob Licença Creative Commons. (Pode ser copiado, distribuído, exibido e executado com atribuição da fonte, desde que seja para fins não comerciais, não podendo criar derivações da mesma).

As Categorias no repositório são:

1. Publicações da Academia
2. Trabalhos académicos
3. Outros textos
4. Eventos comunicativos
5. Informativos
6. Vários

Sendo 1. Publicações da Academia, 4. Eventos comunicativos e 5. Informativos, as categorias que atualmente contêm a maior parte dos itens e documentos em pdf. Nomeadamente a Categoria 1 contém os pdf dos 4 Boletins da Academia Galega da Língua Portuguesa editados mais o Anexo *Galiza, língua e Sociedade*. Estes textos podem descarregar-se livremente como conjunto ou como textos individuais.

Os eventos compõem-se de áudios, vídeos, fotografias e outros documentos relacionados que se arquivam conforme à sua procedência tratando de recolher todo o dado de origem, autoria, localidade e data de produção, reprodução, características técnicas. Está conformado por palestras, conferências, debates, atos poéticos e literários gerados dentro do movimento associativista galego em defesa da Língua.

O arquivo digital da Academia Galega da Língua Portuguesa foi desenhado com três objetivos: **divulgação, formação e preservação:**

- Divulgação
 - Intervenção programática na Sociedade Civil Galega
 - Projeto reintegrante – associativismo
 - Construção do português da Galiza
- Formação
 - Carências educativo – formativas da sociedad
 - Suporte da AGLP

- Preservação
 - Documentação pública e privada da AGLP
 - Acervo documental do associativismo
 - Memória do galego oral nos 1970-2000
 - Qualidade dos materiais e do seu tratamento

Origina-se na constatação do espalhamento e mesmo perda (por obsolescência de formatos por desaparecimento de sites, rotura de links) de conteúdos audiovisuais importantes para a história contemporânea da Língua da Galiza e do Reintegracionismo¹⁰, e da necessidade de dar corpo ao arquivo e produção divulgativa académica da Academia Galega da Língua Portuguesa.

Com poucos recursos materiais e humanos¹¹ o processo exemplifica as possibilidades do Software livre e do Movimento Open Access combinado com as possibilidades do Português como língua internacional de uso crescente.

A montagem e instalação do Dspace teve um demorado processo que se articulou no ano 2011. Após a instalação e processo informático e localização no site procedeu-se à adaptação às necessidades da Academia Galega da Língua Portuguesa com trabalhos e reuniões:

- Tradução de conteúdos e sistema de metadados ao português (Pt_Gz)
- Elaboração dos registos tipo e análise dos metadados apropriados ao nível mais detalhado contidos nos próprios documentos e nos registos.
- Estrutura do arquivo, implementação de pesquisas e conjuntos
- Provas de catalogação

O resultado, em processo, é um repositório normalizado que disponibiliza materiais textuais e audiovisuais de alta qualidade.

Os pontos fortes do AD são:

- A qualidade do sistema, do material e padrão nas ferramentas suportadas numa rede internacional de desenvolvimento institucional e aprimoramento, o que garante estabilidade, preservação, intercâmbio padrão e sustentabilidade ante as evoluções técnicas
- O detalhe e cuidado no tratamento do material digital pela escolha dos melhores (qualidade e uso) programas e qualidade da cópia conservada (gráfica, audiovisual)

¹⁰ O reintegracionismo, lusismo ou movimento lusófono galego é o nome com que se conhece o movimento social na Galiza que pretende a reintegração da língua galega no âmbito linguístico português e de adaptar a ortografia galega à desse âmbito. Também é a corrente linguística que defende essas teses.

¹¹ Uma equipa formada por: três académicos Dr. Celso A. Cáccamo, professor de Linguística Geral da Universidade da Crunha; Vitor M. Lourenço, Diretor do Portal Galego da Língua e responsável informático do E-estraviz que foi o coordenador; Dr. Ernesto V. Souza, na altura Bibliotecário da Universidade de Valhadolid e José António Cidre Bardelás, arquiteto, técnico em informática. Colaboram também regularmente no visionado, indizado e resume documental M^a José Castelo Lestom e o académico Prof. Francisco Paradelo

- O esforço na implementação de uma estrutura estável e duradoira no tempo

Pontos fracos:

- Falta de recursos económicos e humanos
- O voluntarismo do trabalho dos académicos coordenadores e do pessoal eventual de apoio que fornece os recursos e contribui à sua catalogação e revisão
- A descontinuidade do próprio trabalho, efeito da não profissionalização e da ausência de orçamentos e por lógica da disponibilidade do trabalho.

Próximos passos do arquivo:

- Continuar a ritmo que permitam os recursos para integração de mais objetos no arquivo previsivelmente nestes conjuntos:
 - Publicações da AGLP
 - Eventos comunicativos
 - Documentos para a história contemporânea
 - Outras publicações académicas
- Melhorar o aprimoramento dos existentes, para facilitar a sua descarga, audição e integração dos documentos e dados em outros ficheiros e sistemas de gestão de bibliografias
- Campanha divulgativa para dar a conhecer o projeto e solicitarmos colaboração e incorporação de recursos humanos e fundos documentais (áudio e audiovisuais de carácter "histórico")¹².

5 - Iniciativas promovidas pela AGAL nas que colabora AGLP

A Associação Galega da Língua (AGAL) é uma associação sem ânimo lucrativo, legalmente constituída em 1981, que visa a plena normalização do Galego-Português da Galiza e a sua reintegração no âmbito linguístico a que historicamente pertence: o galego-luso-brasileiro.

A AGAL busca a incorporação do galego ao âmbito linguístico galego-luso-brasileiro. Considera que o galego é a denominação que o português tem na Galiza e assume que o nome internacional é português.

O local da AGAL na Internet é conhecido como "Portal Galego da Língua".

5.1. E-estraviz - o mais completo dicionário galego em linha, constantemente atualizado e melhorado por uma equipa de até quinze pessoas, disponibilizado pela AGAL sob a direção do próprio lexicógrafo e académico Isaac Alonso Estraviz.

¹² O que se está a desenvolver com uma Campanha da Pro-AGLP de Informação pelas Escolas Oficiais de Idiomas, Universidades e espaços Associativistas culturais da Galiza.

Sem qualquer dúvida, este é um dos mais completos manuais das línguas românicas publicados em internet, e o mais contrastado no que à variante galega da nossa língua diz respeito, pois conta com um total de 127.316 entradas.

Como bom dicionário manual, o e-Estraviz possui o léxico mais fundamental e completo possível, quer dos ramos científico-técnicos quer das expressões mais populares. Por isso, inclui abundante léxico científico e popular de mais uso.

Este dicionário vai destinado para ser utilizado por toda a lusofonia, e, nomeadamente, pela cidadania galega.

Desde 2011 prepara com apoio da Academia Galega da Língua Portuguesa a sua adaptação ao Acordo Ortográfico.

5.2. PGL - O Portal Galego da Língua - desde 2002, é um site informativo e de debate da Associação Galega da Língua que oferece notícias e opinião desde Galiza para o espaço lusófono. Tem-se convertido no verdadeiro espaço de conformação e associativismo de valores civis do reintegracionismo contemporâneo¹³.

O Portal Galego da Língua visa oferecer aos seus utentes (galegos, portugueses, brasileiros e outros) todas as notícias da língua, assim como uma especial preocupação com o sempre precário estado da língua oral na Galiza.

O Portal Galego da Língua é uma das mais populares janelas que o Movimento Lusófono Galego tem na Internet. O PGL, que tem as suas origens em 1999 com os primeiros trabalhos da Associação Galega da Língua (AGAL), nos últimos anos foi deixando num segundo lugar o seu papel como sítio corporativo desta organização para ser um dos referentes da atualidade linguística do País.

O portal geral de notícias tem hoje em dia uma média de 30.000 visitas mensais; a comunidade no seu conjunto (e-Estraviz, Blogues agal-gz, Planeta NH) tem chegado a picos de até 100.000 visitas/mês.

Algumas das suas seções são "Isso não é galego, é português" (de título irónico), "Falar com jeito" (destinado aos galegos que, dada a sua situação, metem muitos castelhanismos nos seus textos e falas), o dicionário e-Estraviz da língua "galego-portuguesa", a GZ-editora etc.

Outros sites da AGAL vinculadas ao Portal Galego da Língua:

1. Planeta NH - ludo site de divertimento com perguntas e concursos. Projeto lúdico criado na Galiza para toda a Lusofonia. Para participar elaborando Jogos sobre os teus temas de interesse.
2. Blogues AGAL - site da AGAL que fornece software e ajuda para a criação e concepção de blogues pessoais e coletivos.
3. Imperdível - loja on-line da AGAL onde é possível adquirir livros, música e publicações galegas e da Lusofonia.
4. Wiki-Faq do Reintegracionismo - esta wiki pretende fornecer informações teóricas e

¹³ DURÃO, 2008.

práticas sobre o reintegracionismo, tanto sobre a filosofia como sobre o movimento social. Se bem que esta wiki seja da Associação Galega da Língua, e logicamente preste especial atenção a esta organização lusófona, pretende dar uma visão geral sobre o reintegracionismo.

6 - Conclusão

Nestes momentos está a ser, e no futuro imediato o português será ainda mais, uma das línguas com mais projeção internacional. Dizer português e galego é o mesmo que dizer espanhol e castelhano.

O português da Galiza, como variedade europeia do português, começa a ser conhecido e reconhecido no espaço lusófono (desde a sua participação nas negociações do Acordo Ortográfico em 1986 e 1990), o que abre novas possibilidades para uma evolução positiva da questão da língua na Galiza.

Um dos âmbitos pioneiros e onde se tornou evidente esta realidade tem sido a internet. Foros, correios, sites, colaborações e contatos gerados pelos pioneiros num tecido de intercâmbio de informação e conhecimentos que começam a dar os seus primeiros frutos em forma de desenvolvimento de ferramentas, colaborações e programas.

A integração e participação com naturalidade da Academia Galega da Língua Portuguesa nos espaços internacionais de Língua Portuguesa, por meio dos seus produtos e ferramentas, é ao mesmo tempo um anúncio da mudança de narrativa que a respeito da questão galega e da sua língua se abre na Galiza e uma evidência da pujante realidade internacional e de intercâmbio que é a língua portuguesa para todos os lusófonos.

No decurso deste desenvolvimento a língua portuguesa passou a ser canal de comunicação pelo que circulam e do que fazem uso estas novas comunidades. Nesse espaço internacional e social que se define pela internet, os conceitos de Lusofonia e Acordo Ortográfico tomam uma forma e valor que vai ser definitiva, e em que podem entrar os espaços galegos criativos, comunicativos, económicos e sociais.

6 - Conclusão

Advice to Authors: A guide to copyright and local deposit of full text. UCD *Library* v., n. 2011. <<http://www.ucd.ie/t4cms/Guide75.pdf>>

ABADAL, Ernest. 2012. [e-Book] *Acceso abierto a la ciencia Barcelona*, Editorial UOC. <<http://eprints.rclis.org/bitstream/10760/16863/1/2012-acceso-abierto-epi-uoc-vfinal-autor.pdf>>

Academia Galega da Língua Portuguesa
<http://www.netvibes.com/aglp#AGLP_-_Apresentação>

Academia Galega da Língua Portuguesa. Arquivo Digital:
<<http://arquivo.academiagalega.org/xmlui>>

Academia Galega da Língua Portuguesa (2009) *Galiza: Língua e Sociedade (XIV Ensaio)* Anexo 1 do Boletim da Academia Galega da Língua Portuguesa.

Academia Galega da Língua Portuguesa. Comissão de Lexicologia e Lexicografia. *Léxico da Galiza para ser integrado no Vocabulário Ortográfico Comum da Língua Portuguesa*, Santiago de Compostela, novembro de 2011

<http://www.academiagalega.org/images/stories/2011/20111110_lexico_aglp.pdf>

ADNI, Rhiza. 2011. The Impact of Open Access Initiative on Knowledge Sharing. *E-LIS. E-prints in Library and Information Science*. <<http://eprints.rclis.org/handle/10760/16778>>

AGAL (Associação Galega da Língua) <<http://www.ir.gl/3fc44e>>

BEREC 2011. [e-Book] *BEREC Report on "Open Access" 2011*.

<[http://www.irg.eu/streaming/BoR%20\(11\)%2005%20BEREC%20report%20on%20Open%20Access_final.pdf?contentId=547137&field=ATTACHED_FILE](http://www.irg.eu/streaming/BoR%20(11)%2005%20BEREC%20report%20on%20Open%20Access_final.pdf?contentId=547137&field=ATTACHED_FILE)>.

BOSO, Augiza Karla. 2011. Repositórios de Instituições Federais de Ensino Superior e suas políticas: análise sob o aspecto das fontes informacionais. Florianópolis. <<http://eprints.rclis.org/bitstream/10760/16156/1/Disserta%20a7%20a3o%20-%20Augiza%20Karla%20Boso.pdf>>.

CINTRA, Luís Filipe Lindley. 1971. Nova proposta de classificação dos dialectos galego-portugueses. *Boletim de Filologia*, Lisboa, Centro de Estudos Filológicos, 22, 81-116 <<http://www.instituto-camoes.pt/cvc/hlp/biblioteca/novaproposta.pdf>> (no PDF, p. 7).

Creative Commons <<http://creativecommons.org/>>

CRISTOVÃO, Ângelo. 2008. "Intervenção de Ângelo Cristóvão, Representante da Associação Cultural Pró Academia Galega da Língua Portuguesa"

<<http://www.academiagalega.org/info-atualidade/interven-na-assembleia-da-repca.html>>.

CUNHA, Celso; CINTRA, Luís Filipe Lindley. 1984. *Nova Gramática do Português Contemporâneo*. Lisboa: Edições João Sá da Costa, p. 10.

Dicionário Estraviz On-line e-estraziviz <<http://www.estraziviz.org/>> e <<http://www.estraziviz.org/acerca-de.php>>.

DSpace <www.dspace.org>

MIT DSpace <<http://dspace.mit.edu/>>

Dublin Core Metadata Initiative <<http://dublincore.org/>>

DURÃO, Carlos. 2008. Síntese do Reintegracionismo contemporâneo. *Boletim da Academia Galega da Língua Portuguesa*, nº 1, 35-56

<<http://arquivo.academiagalega.org/xmlui/handle/123456789/154>>.

E-prints: <<http://www.eprints.org/openaccess/>>

Escritório AGLP <http://www.netvibes.com/aglp#AGLP_-_Apresentação>

KENNAN, Mary Anne. 2011. Learning to share: mandates and open access. *Library Management*. vol. 32, n. 4, 302-318. <<http://dx.doi.org/10.1108/01435121111132301>>

KURAMOTO, Hélio. 2012. Acesso Livre: como tudo começou. Blog do Kuramoto, <<http://kuramoto.blog.br/2012/02/27/acesso-livre-como-tudo-comecou/>>

Manakin <<https://wiki.duraspace.org/display/DSPACE/Manakin>>

MONTERO SANTALHA, J. Martinho.1994. A lusofonia e a língua portuguesa da Galiza: dificuldades do presente e tarefas para o futuro. *Temas de O Ensino de Linguística, Sociolinguística e Literatura*, Ponte Vedra-Braga, Vol. VII-IV, n. 27-38 (1991-1994), 137-149. <<http://www.lusografia.org/mmontero.htm>>

Netvibes <<http://www.netvibes.com/>>

PGL Portal Galego da Língua <<http://www.pglingua.org/>>

Portulano de Recursos <<http://www.netvibes.com/portulano>>

PRO-AGLP <<http://www.academiagalega.org/quem-somos/quem-somos.html>>

Site AGLP <<http://www.academiagalega.org>>

Site Carvalho Calero <<http://carvalhocalero.academiagalega.org/>>

Site Guerra da Cal <<http://guerradacal.academiagalega.org/index.php>>

Site Signum <<http://signum.academiagalega.org/>>

SK - Survival Kit

<http://www.netvibes.com/academia-galega-da-lingua-portuguesa#Survival_Kit>



MULTILINGUISMO NO CIBERESPAÇO: A PARTICIPAÇÃO DO PORTUGUÊS NUMA B@BEL DIGITAL?

Cláudio Menezes

Sobre o autor

Universidade de Brasília, Brasil.
Email: claudiomenezes@unb.br
Webpage: <http://let.unb.br/lea>

Resumo

Para maior participação de uma língua nos dias de hoje não se pode fechar os olhos para uma realidade: o mundo digital é um dos principais espaços para a vitalidade de uma língua em suas múltiplas dimensões. Por outro lado, graças às aplicações das tecnologias da informação e comunicação, as fronteiras para uso de uma língua (no passado, exclusivamente geográficas) se dissolveram. Daí a necessidade de alcançar e mobilizar falantes de outras línguas para que tenham interesse pelo português e vice-versa. Nesse contexto, a presente comunicação se propõe a apresentar o ambiente no qual uma língua se desenvolve no mundo digital, com uma breve descrição da situação estatística das línguas no mundo, concluindo com algumas sugestões que poderão contribuir para uma maior presença da língua portuguesa no ciberespaço.

Palavras-chave

Português. Língua. Mundo digital. Vitalidade linguística. Ciberespaço.

1 - Contexto

Fenômeno iniciado nos anos 70 do século passado, nos dias de hoje testemunha-se um acelerado processo de substituição de tecnologias analógicas por tecnologias digitais. Esse processo, chamado de convergência tecnológica, viabilizou a transformação de texto, imagem e som em arquivos digitais que puderam inicialmente ser tratados por computador e, posteriormente, por diversos tipos de novos aparelhos digitais (i-Phones, i-Pads, Tablets etc.). Acompanhado pela transformação das telecomunicações, fator que permitiu alargar a dimensões extremamente elevadas a banda de transmissão de dados, esse processo nos levou à situação que hoje vivemos no início do século XXI: a "invasão" do nosso cotidiano pelas aplicações de computador e de outros instrumentos digitais e pelo acesso à informação em tempo real, intermediado pelas redes de telemática.

No campo das línguas, cujo uso era anteriormente limitado a território circunscrito pela geografia, esse fenômeno conduziu a uma expansão da fronteira alcançada pelas línguas tanto a novos campos de aplicação como também à sua utilização em qualquer lugar do mundo servido por uma infraestrutura de informação.

Pode-se observar que esse novo ambiente tecnológico tem colaborado para uma certa "impopularidade" do monolinguismo tradicional. Nos dias de hoje, não se pode imaginar um jovem com pretensões monoglotas, e até mesmo adultos formados em uma cultura monoglota se interessam pelo conhecimento de outras línguas. De certa forma, esse "boom" faz com que a comunicação intermediada pela transmissão por satélite e pelos aparelhos digitais induza à necessidade de uso de mais de uma língua no círculo profissional e pessoal. Não será injusto dizer que vivemos um momento de decadência do monolinguismo e de ampliação do plurilinguismo.

É, portanto, no horizonte desse novo paradigma tecnológico e social que se pode pretender uma maior participação da língua portuguesa nessa "B@bel digital". Se se pretende ampliar a vitalidade da língua portuguesa em um contexto de multilinguismo, é mandatório formular um novo enfoque no relacionamento com as outras línguas, totalmente oposto ao anteriormente adotado nos períodos de colonização.

2 - Situação das línguas no ciberespaço

Há diversos parâmetros para avaliar a situação de uma língua tanto no mundo real como no virtual. Um conceito que já goza de certo reconhecimento internacional é o de "vitalidade linguística". Estudo da UNESCO a esse propósito indica que há 9 critérios ("fatores") para se determinar o grau de vitalidade/perigo e para desenvolver medidas para sua sobrevivência ou para a revitalização linguística:

1. Transmissão intergeracional da língua
2. Número absoluto de falantes
3. Proporção de falantes em relação ao total de uma população
4. Mudanças dos domínios de uso da língua

5. Resposta a novos domínios e à mídia
6. Disponibilidade de materiais para educação e alfabetização numa língua
7. Atitudes governamentais e institucionais e políticas, incluindo seu estatuto e uso oficial
8. Atitudes dos membros de uma comunidade em torno da sua própria língua
9. Quantidade e qualidade da documentação (na e sobre uma dada língua)

Dois estudos sistematizados têm servido como referência para conhecer a situação de algumas línguas no mundo digital: o estudo elaborado por Daniel Pimienta, Daniel Prado e Álvaro Blanco e o censo desenvolvido periodicamente pelo "Language Observatory Project", coordenado pelo Prof. Yoshiki Mikami, da "Technical University Nagaoka", Japão.

O primeiro desses estudos, com resumo editado pela atual "Divisão para a Sociedade do Conhecimento" da UNESCO, foi publicado sob o título "Douze années de mesure de la diversité linguistique sur l'Internet: bilan et perspectives". Sua edição em francês e inglês está disponível em:

<http://unesdoc.unesco.org/images/0014/001421/142186f.pdf> (Francês)

<http://unesdoc.unesco.org/images/0014/001421/142186e.pdf> (Inglês).

Duas tabelas desse estudo podem ajudar a orientar políticas linguísticas voltadas para assegurar a vitalidade do português no mundo digital. Senão, vejamos:

Tabela 1 - Indicadores para as línguas na Internet (2007)

	EM	ES	FR	IT	PO	RO	AL	CAT	Total
Falantes (milhões) ¹	670	400	130	60	205	30	120	9	6.607 ²
Falantes em % da população mundial	10,1%	6,1%	2,0%	0,9%	3,1%	0,5%	1,8%	0,1%	130% ³
Internautas por língua (milhões) ⁴	366	102	58	31	47	5	59	2	1.154 ⁵
Internautas em % de falantes	54,6%	25,4%	44,9%	52,3%	23,1%	16,5%	49,1%	23,1%	17,5% ⁶
Internautas em % da população mundial	5,5%	1,5%	0,9%	0,5%	0,7%	0,1%	0,9%	0,0%	17,5%
% de internautas por língua	32%	9%	5%	3%	4%	0%	5%	0,2%	130%
% de páginas Web por língua ⁷	45,0%	3,8%	4,4%	2,7%	1,4%	0,3%	5,9%	0,1%	100%
Produtividade linguística por língua ⁸	1,42	0,43	0,87	0,98	0,34	0,66	1,16	0,74	1
Páginas Web por internautas em uma língua determinada	4,44	0,63	2,24	2,93	0,45	0,62	3,25	0,96	

¹ Fonte: União Latina (2000).

² Estimativa da população mundial. O número total de falantes deveria, no entanto, ser mais elevado, se contadas as pessoas falando mais de uma língua.

³ 30% seria uma estimativa intuitiva da população que fala mais de uma língua. Este dado é provavelmente próximo da realidade nos países da OCDE, porém não na maior parte dos países em desenvolvimento, onde uma pessoa fala em média 2 ou 3 línguas (como na África).

⁴ Fonte: Internet Word Stats (2005).

⁵ Estimativa do número total de internautas.

⁶ Porcentagem da população mundial dispoendo de acesso à Internet.

⁷ Fonte: Funredes/União Latina (2005).

⁸ Relação entre a % de páginas Web por língua e a % de usuários da Internet por língua.

Tabela 2 - Dados sobre a Diversidade Linguística

(Fonte: Pimienta, Daniel et alii,
in Twelve years of measuring linguistic diversity in the Internet)

Table 6: Web presence of studied languages compared to English

	SP	FR	IT	PO	RO	GE	CAT
09/98	3.37%	3.75%	2.00%	1.09%	0.20%		
08/00	8.41%	7.33%	4.60%	3.95%	0.37%	11.00%	
01/01	9.46%	7.89%	4.93%	4.44%	0.33%	11.43%	
10/01	11.36%	9.14%	6.15%	5.61%	0.36%	14.08%	
02/02	11.60%	9.60%	6.51%	5.62%	0.33%	14.41%	
02/03	10.83%	8.82%	5.28%	4.55%	0.23%	13.87%	
02/04	10.30%	10.18%	6.09%	4.36%	0.41%	15.35%	
03/05	10.23%	11.00%	6.77%	4.15%	0.37%	15.42%	
11/07	8.45	9.80%	5.92%	3.09%	0.63%	13.12%	0.30%

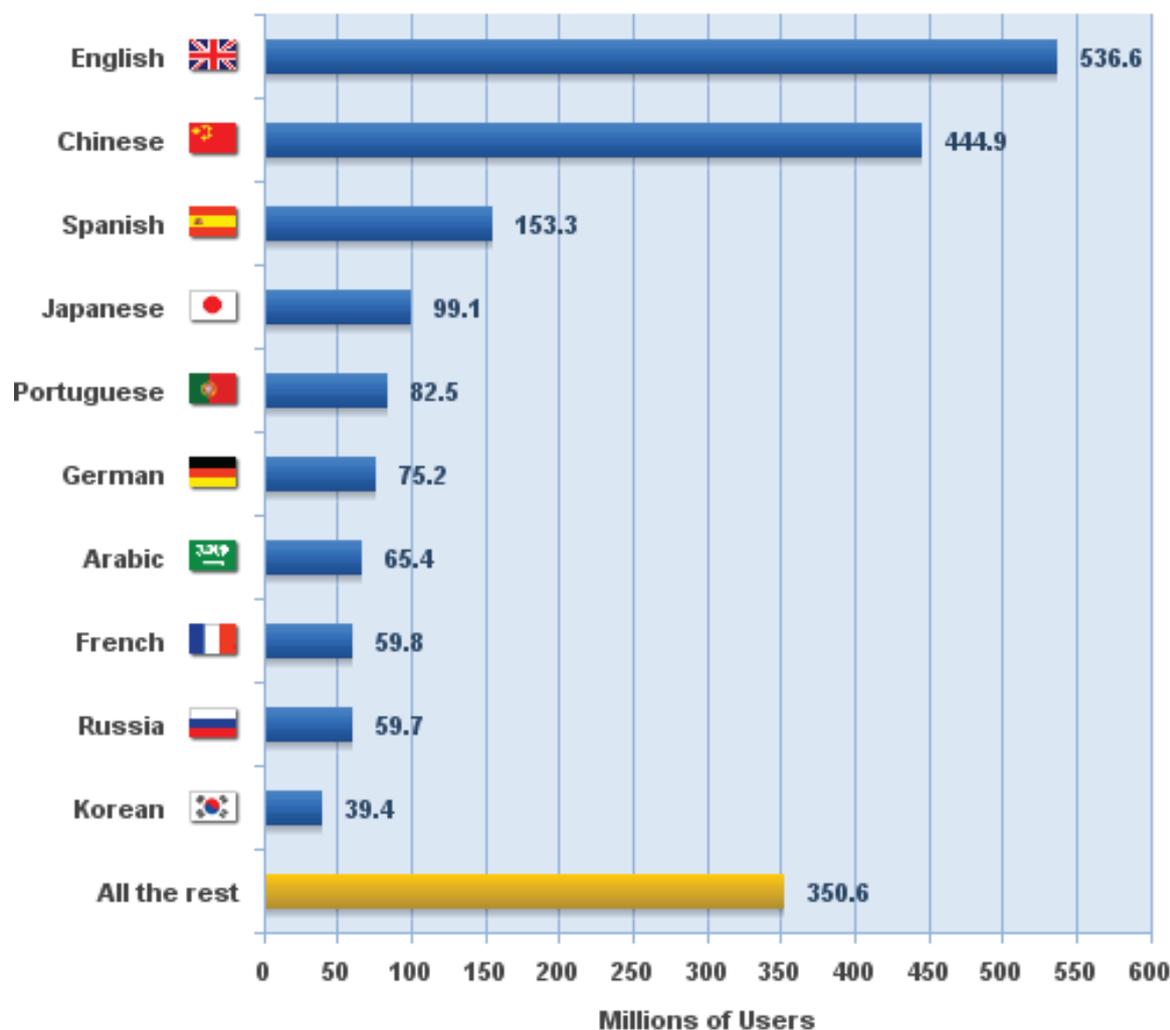
Na primeira das tabelas acima (Tabela 1), vê-se que em 2007 estimava-se em 47 milhões o número de internautas de língua portuguesa, correspondendo a 23% do total de falantes (205 milhões à época).

Já no estudo de Pimienta (Tabela 2), a posição do português oscila entre 1 e 5% em relação ao inglês, durante as diversas medições efetuadas a partir de setembro de 1998.

Já o estudo do "Internet World Statistics" (Gráfico 1) aponta em 2010 a existência de 82,5 milhões de usuários da língua portuguesa na Internet, o que nos situa em 5º lugar no ranking das línguas mais usadas na Internet.

Gráfico 1 - Estudo do "Internet World Statistics"

Top Ten Languages in the Internet 2010 - in millions of users



Source: Internet World Stats - www.internetworldstats.com/stats7.htm

Estimated Internet users are 1,966,514,816 on June 30, 2010

Copyright © 2000 - 2010, Miniwatts Marketing Group

Os indicadores acima nos permitem considerar que a participação do Português no mundo digital é promissora.

É importante considerar que a manutenção da vitalidade de uma língua requer uma preocupação e ação permanente, melhor dizendo, requer uma política de planejamento linguístico.

De acordo com o sociólogo Pierre Bourdieu⁹, uma relação de comunicação linguística é

⁹ Ver "A Economia das Trocas Linguísticas", Reproduzido de BOURDIEU, P. "L'économie des échanges linguistiques". *Langue Française*, 34, Maio 1977. Traduzido por Paula Monteiro, disponível em http://www.taddei.eco.ufrj.br/AntCom/Bourdieu_1983_EkonomiadasTrocasLinguisticas.pdf

essencialmente uma relação de força simbólica, determinada pela estrutura do grupo social em que ocorre a comunicação. No entanto, há que considerar que a língua opera tanto no campo da economia como em outros campos da sociedade. Ademais, com o processo de desterritorialização linguística causado pelas tecnologias de informação e comunicação, o antigo paradigma de comunicação monolíngue foi abalado, visto que há uma tendência evidente de bi- ou até mesmo de plurilinguismo no mundo digital.

Por isso, já não é mais pertinente que uma política de planejamento linguístico esteja voltada exclusivamente para o monolingüismo. Tais políticas deveriam levar em conta tanto a tradição em favor do monolingüismo que perdurou bastante tempo, como também o relacionamento entre as línguas faladas num espaço geopolítico e o relacionamento entre as línguas no mundo digital.

3 - Metodologias e tecnologias para inclusão e revitalização de línguas no mundo digital

Por meio da Carta de Maputo, os participantes do Colóquio de Maputo sobre a Diversidade Linguística nos Países da CPLP, realizado de 12 a 14 de setembro de 2011 na capital moçambicana, emitiram uma série de recomendações ao Instituto Internacional da Língua Portuguesa sobre esse tema. De acordo com aquele documento, 5% da diversidade linguística mundial está situada nos países da CPLP, correspondendo a mais de 300 línguas, além do português.

Dentro desse cenário, pode-se dizer que os países da CPLP constituem um microcosmo da diversidade linguística mundial, o que nos permite a realização de experimentos de planejamento e política linguística com bastante generalidade.

Embora não haja tempo para explorar em detalhe as metodologias, simbologias, tecnologias e políticas de desenvolvimento de um multilingüismo no espaço lusófono, vale realçar, a título de informação, que a inclusão de uma nova língua no ciberespaço requer as seguintes etapas:

Escolha de um sistema de escrita (“script”) para a língua (ou dialeto) oral;

Elaboração um sistema ortográfico (diversas línguas podem usar o mesmo “script”);

Mapeamento dos símbolos em uma tabela de códigos digitais (Tabelas do Consórcio UNICODE, www.unicode.org);

Padronização de fontes de caracteres;

Desenvolvimento de software e hardware e produção de conteúdos;

Importante: Desenvolvimento de conteúdos locais digitalizados (incluindo identificadores de conteúdos, tais como os “IDNs – Internationalized Domain Names”).

Cada uma dessas etapas envolve certo nível de complexidade para assegurar a evolução da “língua oral” ao “objeto digital”!

4 - A promoção da vitalidade do Português

Inúmeras ações de política e planificação da Língua Portuguesa têm sido implementadas com o objetivo de permitir a difusão internacional do nosso idioma. Um trabalho recente das professoras Simone da Costa Carvalho e Margarete Schlatter intitulado "Ações de Difusão Internacional da Língua Portuguesa"¹⁰, detalha com bastante propriedade cada uma dessas ações em diversos países. Cabe destacar entre tais ações as seguintes:

- Centros Culturais do Brasil (CCB), Institutos Culturais Bilaterais (IC) e Leitorados Brasileiros;
- Programa de Estudantes Convênio de Graduação (PEC-G) e de Pós-Graduação (PEC-PG);
- Certificado de Proficiência em Língua Portuguesa para Estrangeiros (CELPE-BRAS).
- Grupos de Ensino e Pesquisa de Português Língua Adicional no Brasil;
- Comunidade de Países de Língua Portuguesa (CPLP) e o Instituto Internacional da Língua Portuguesa (IILP);
- Políticas de Promoção da(s) Língua(s) no MERCOSUL;
- Ações de Cooperação na Área Educacional para a construção de uma região bilíngue;
- Projeto Escola Intercultural Bilíngue de Fronteira (PEIBF);
- Associação de Universidades - Grupo de Montevideu (AUGM);
- Universidade Federal da Integração Latino-Americana (UNILA);
- Curso de Espanhol-Português para Intercâmbio (CEPI).

5 - A atuação da UnB

O Instituto de Letras da Universidade de Brasília completa 50 anos em 2012 e sempre teve uma atuação bastante intensa na promoção da língua portuguesa, através dos seus 3 Departamentos:

- Departamento de Linguística, Português e Línguas Clássicas - LIP: www.unb.br/il/liv;
- Departamento de Línguas Estrangeiras e Tradução - LET: www.let.unb.br;
- Departamento de Teoria Literária e Literaturas - TEL: www.telunb.com.br

Ao lado da formação de professores e das atividades sobre literatura portuguesa, há dois destaques interessantes a mencionar aqui: o PEPPFOL - *Programa de Ensino e Pesquisa em Português para Falantes de Outras Línguas* (<http://www.let.unb.br/peppfol>) e o PBSL - *Português do Brasil como Segunda Língua*.

¹⁰ Publicado em "Cadernos do IL", Porto Alegre, n. 42, junho de 2011, pag. 260-284

O primeiro desses programas, o PEPPFOL, oferece cursos de Português a toda a comunidade estrangeira da capital federal e a alunos conveniados. Desenvolve, também, projetos de Formação Inicial Continuada para os alunos brasileiros dos cursos de Letras da UnB, além de realizar pesquisas nas áreas de ensino-aprendizagem da língua portuguesa e da cultura brasileira, de formação de professores e da inserção de novas tecnologias da comunicação (TIC) para o ensino de português a falantes de outras línguas.

A Licenciatura denominada Português do Brasil como Segunda Língua (PBSL) tem por meta a formação de professores de Língua Portuguesa para ensinar o Português do Brasil – língua, literatura e cultura – a falantes e usuários de outras línguas. Para maiores detalhes, ver <http://ovpunb.blogspot.com.br/2010/08/profissiografia-portugues-do-brasil.html>.

No que diz respeito às línguas faladas no espaço lusófono, a UnB desenvolve diversas iniciativas tanto no tocante a línguas estrangeiras como às línguas locais. São mantidas diversas licenciaturas e bacharelados em língua e literatura estrangeira e programas de Mestrado e Doutorado em Linguística Aplicada e Literaturas. No campo das línguas locais, vale realçar o trabalho desenvolvido pelo LALI - Laboratório das Línguas Indígenas (<http://www.unb.br/il/lali>), uma referência de qualidade nesse campo de estudo no Brasil.

Um aspecto a destacar nessas atividades do Instituto de Letras da UnB é o novo Bacharelado em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação (LEA-MSI, <http://let.unb.br/lea>), criado no Departamento de Línguas Estrangeiras e Tradução.

Criado na esfera do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (Reuni, <http://reuni.mec.gov.br>), esse Bacharelado, além da plataforma linguística e suas respectivas culturas, se compõe de três eixos temáticos: áudio-visual, terminologia e multilinguismo digital.

O eixo áudio-visual inclui disciplinas de legendagem, áudio-descrição, acessibilidade, línguas de sinais, entre outras.

O eixo multilinguismo digital aborda simbologias, metodologias, tecnologias e políticas para inclusão e/ou revitalização de línguas no mundo digital. Contém também diversas técnicas e uso de programas informáticos orientados para facilitar o uso de línguas estrangeiras no mundo digital (por exemplo: programas de auxílio à tradução, programas para elaboração de páginas web multilíngues etc.).

Detalhes sobre o Bacharelado LEA-MSI estão disponíveis nas páginas <http://let.unb.br/lea> e <http://let.unb.br/simc>.

6 - Conclusões

Ao nosso ver, a participação e influência do Português no mundo digital poderão ser asseguradas por três eixos de ação, a saber:

a) ações circunscritas ao espaço dos países de língua portuguesa: de certa forma, tais ações são análogas às que já foram citadas anteriormente (PEPPFOL, Escolas de Fronteira, PBSL, etc.);

b) ações novas, destinadas a fortalecer um processo de reciprocidade com as línguas faladas no espaço lusófono, nos termos preconizados pela Carta de Maputo;

c) ações destinadas a permitir o uso de metodologias de inclusão digital para as comunidades de línguas faladas no espaço lusófono, por meio do uso das metodologias e tecnologias da informação e comunicação.

Finalmente, gostaria de realçar a necessidade de colaboração com as Academias de Línguas dos países de língua oficial portuguesa, e, em particular, com a Academia Africana de Línguas (ACALAN, www.acalan.org), no sentido de estimular a convivência harmônica do Português com as 339 línguas faladas nos nossos países.

Referências

BOURDIEU, P. 1977. A Economia das Trocas Linguísticas. Reproduzido de BOURDIEU, P. L'économie des échanges linguistiques. *Langue Française*, 34. Traduzido por Paula Monteiro, disponível em http://www.taddei.eco.ufrj.br/AntCom/Bourdieu_1983_EkonomiadasTrocasLinguisticas.pdf

Carta de Maputo. 2012. *RIILP*, V. 1, N.1, 99-100.

CARVALHO, S. da C. & M. SCHLATTER. 2011. Ações de Difusão Internacional da Língua Portuguesa. *Cadernos do IL*. Porto Alegre, n. 42, 260-284.



O CENTRO VIRTUAL CAMÕES: DIVULGAÇÃO E FORMAÇÃO

Rui Vaz

Sobre o autor

Camões, Instituto da Cooperação e da Língua, I.P.; Coordenador do Centro Virtual Camões. Contato; rvaz@instituto-camoes.pt)

Resumo

O Centro Virtual Camões é o sítio na Internet do Camões, Instituto da Cooperação e da Língua, I.P. para apoio ao ensino e aprendizagem do português e formação dos agentes da educação, bem como para a divulgação da língua e cultura portuguesa. Numa era de comunicação digital, em que as distâncias geográficas são encurtadas pelo recurso às tecnologias de informação, o ensino a distância torna-se ferramenta central das políticas de ensino da língua e cultura, sendo prioridade a sua ampliação. Nesta comunicação, serão apresentadas as principais linhas de atuação do Camões, I.P. para, através do Centro Virtual Camões, potenciar a promoção externa da língua e da cultura portuguesas na Internet, assim como promover a qualificação do ensino de português no estrangeiro.

Palavras-chave

Instituto Camões. E-learning. Conteúdos digitais.

1 - Introdução

O Centro Virtual Camões é o sítio na Internet do Camões, Instituto da Cooperação e da Língua, I.P. para apoio ao ensino e aprendizagem do português e formação dos agentes da educação, bem como para a divulgação da língua e cultura portuguesa. Disponibilizando gratuitamente conteúdos nesta área, tem uma média de 500 mil acessos mensais.

Vou começar por contextualizar muito brevemente a ação do Instituto Camões através da Internet e conteúdos digitais, apresentando de seguida os projetos desenvolvidos pelo Centro Virtual Camões.

O Português é uma das seis línguas mais faladas do mundo, o que se reflete na posição que ocupa nas línguas mais usadas na internet – a 5ª, de acordo com dados de 2011¹. Contudo, um elemento do estudo da União Latina/FUNREDES não pode deixar de motivar a nossa preocupação. Apesar do crescimento na utilização, o índice de produtividade das páginas web por utilizador coloca-nos muito abaixo de outras línguas, longe do nosso potencial.

Nesse sentido, os esforços congregados de todos os países de língua portuguesa para prossecução das linhas definidas no Plano de Ação de Brasília tornam-se prementes, reforçando a necessidade de um maior investimento público nesta área.

Em particular nesta comunicação, teremos presentes as estratégias propostas pelo Plano de Ação de Brasília de promoção e difusão do ensino da língua portuguesa, nomeadamente para a atualização de bibliotecas virtuais e de utilização das tecnologias de informação e comunicação no ensino presencial e a distância.

2 - Contribuição Camões, I.P.

O Camões, I.P. de acordo com a sua atual Lei Orgânica (Decreto-Lei n.º 21/2012, de 30 de janeiro), tem como atribuições, entre outras:

- Promover, coordenar e desenvolver a realização de cursos de língua portuguesa e outros conteúdos culturais, quer em sistema presencial, quer por recurso a tecnologias de informação e comunicação;
- Conceber, promover, propor, apoiar e executar a produção de obras e projetos de divulgação da língua e da cultura portuguesas no estrangeiro;
- O desenvolvimento e promoção da utilização de plataformas para o ensino e a aprendizagem do português à distância e a divulgação da cultura portuguesa.

Assim, especialmente através do Centro Virtual Camões, o Instituto Camões dá resposta à concretização dos seus objetivos de:

- Ampliação da **oferta da aprendizagem, a distância**, da Língua e da Cultura Portuguesa, nomeadamente através de cursos de português para estrangeiros e de conteúdos destinados a aprendentes avançados da língua portuguesa;

¹ (<http://www.portalingua.info/pt/observato>).

- Promoção de **nova imagem do património cultural português** proporcionando maior visibilidade das novas formas de expressão e novos valores.

As tecnologias de informação e comunicação são o meio privilegiado para a implementação de programas de:

- Criação/atualização de **conteúdos e serviços em linha** nas áreas da informação cultural e do **ensino a distância**, incluindo conteúdos para apoio dos docentes de PLE.
- Promoção de uma **imagem externa de qualidade** da cultura portuguesa.

O Centro Virtual Camões responde ao desafio expresso na Comunicação da *Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões*, intitulada *Multilinguismo: uma mais-valia para a Europa e um compromisso comum*, designadamente quando se afirma:

“Face a uma economia cada vez mais globalizada e eletrónica e a uma massa crescente de informação em todas as línguas imagináveis, é importante que os cidadãos possam **aceder e utilizar essa informação e todos os serviços, ultrapassando as barreiras nacionais e linguísticas**, através da Internet e dos disponíveis móveis. As tecnologias da informação e da comunicação (TIC) precisam de estar conscientes e de promover a **criação de conteúdos em diferentes linhas.**”

E também tem presente que o investimento no Multilinguismo se faça em interação com a Escola, a Universidade, espaços, por excelência, do conhecimento, da memória cultural ativa, da socialização, e que terão de assumir plenamente uma dupla função: a de, sob o ponto de vista ético, capaz de formar em função das ligações necessárias e oportunas entre o **ser e o saber**, de atualizar a dimensão do **saber-ser**, sem prejuízo do **saber-fazer**; a de dotar os cidadãos de uma maior capacidade científica, que lhes assegure a confiança necessária a serem mais autênticos e mais livres e, logo, a tornarem-se exemplos de maior dimensão humanista.

Ao mesmo tempo tem presente o conceito de **flexieducation**, proposto pela Dinamarca em 2009, em resposta aos **desafios das alterações demográficas e do progresso tecnológico**, bem como à questão do contributo do novo quadro estratégico proposto para a cooperação europeia em matéria de educação e formação, o qual combina **flexibilidade e educação, formal ou não formal**. Trata-se de reconhecer competências, (afirma) aproveitando a experiência e as capacidades das pessoas.

Neste contexto, tem vindo o IC a estabelecer **parcerias** com universidades ou a chamar investigadores para a coordenação dos seus projetos, que poderemos agrupar em três eixos:

Eixo 1	Eixo 2	Eixo 3
Espaço de Saber — vivo e circulante —	Espaço de Criação de saber — interativo e Em progressão —	Espaço de Formação — contínua e Ao longo da vida —

2.1 - Espaço de Saber

Enquanto *Espaço de Saber*, o **CVC** investe, investirá nos seguintes projetos:

Biblioteca digital	Bases temáticas	Exposições virtuais
---------------------------	------------------------	----------------------------

A **Biblioteca Digital** do **CVC** dispõe para leitura **gratuita** títulos de 12 áreas do saber, da Arquitetura à Música, passando, por exemplo, do Cinema à Literatura, da Educação à Geografia, da Etnologia à História.

Com cerca de 2.400 títulos, a Biblioteca Digital Camões tem tido um **crescimento de 10%/ano**, objetivo que nos propomos manter em 2012, principalmente através do reforço da oferta, com a digitalização das edições da Comissão Nacional para a Comemoração dos Descobrimientos Portugueses e o alargamento de protocolos com editoras. Outra tendência passará pela otimização de obras existentes para formatos amigáveis de **leitura em dispositivos móveis**, correspondendo aos desejos dos nossos utilizadores.

Às **Bases Temáticas** do **CVC** agora disponíveis, relativas a Navegações Portuguesas, Ciência em Portugal, Figuras da Cultura Portuguesa, Literatura Portuguesa, Cinema Português e História da Língua Portuguesa, Música Portuguesa dos séculos XIX, XX e XXI, Arte e Artistas em Portugal, em permanente atualização, acrescerá brevemente uma nova sobre **Teatro Português** (Centro de Estudos de Teatro).

As **Exposições Virtuais** do **CVC**, compostas com carácter também pedagógico, constituem formas de divulgação do saber, nomeadamente sobre a Língua Portuguesa e a Literatura Portuguesa e as Literaturas que em Língua Portuguesa se escrevem, como, a título exemplificativo, a exposição *Novas Textualidades*, que integra Vozes dos OITO da CPLP, querendo, como o dizem as suas comissárias, dar *a palavra aos novos... e o lugar às novas textualidades e utopias, porque, se a modernidade se construiu com base em grandes*

meta-narrativas (...) A pós-modernidade traduzirá o princípio inverso: da incredulidade e da fragmentação de significados face ao passado recente. Serão publicadas novas exposições virtuais de exposições produzidas pelo Camões, I.P. para circulação presencial, como a que agora se inicia itinerância sobre Azulejo Português.

2.2 - Espaço de Criação de Saber

Enquanto **Espaço de Criação de Saber**, o CVC aposta, quer sistematicamente apostar:

a) na divulgação do **Novo Saber** produzido pela investigação *i)* das **Cátedras** criadas em parceria com o Instituto Camões; *ii)* dos **bolseiros** do Instituto Camões; *iii)* dos **professores e investigadores** que, participando em Congressos apoiados pelo Instituto Camões ou proferindo Conferências em Leitorados do IC venham a disponibilizar as suas investigações e comunicações, a serem publicadas na Biblioteca Digital Camões;

b) na promoção e disponibilização de **Tecnologias de Informação Linguística do Português**, hoje indispensáveis para a promoção da diversidade linguística, reforçando a identidade de cada uma delas pela via dos sistemas de conhecimento.

Além dos projetos **TemaNet** e **Lextec**, está em desenvolvimento uma *wordnet* que inclua todas as variedades do Português. Pretende-se, como resultado tangível, uma *wordnet* do Português que integre 10 000 nós (conceitos), com as correspondentes lexicalizações de todas as variedades do Português, devidamente identificadas.

2.3 - Espaço de Formação

Enquanto **Espaço de Formação**, o CVC desenvolve Cursos de Aprendizagem e Formação nas áreas da Língua Portuguesa, das Literaturas em Língua Portuguesa e da Cultura Portuguesa.

Assim, na área da aprendizagem, o CVC oferece os seguintes **Cursos**:

Português LE, Língua do Quotidiano	Português LE, Língua para Fins Específicos	Português LE, Língua de Criação
Português I, II e III (níveis A1 a B2)	Português para Negócios Português para Medicina Escrita Jornalística	Laboratório de Escrita Criativa

De destacar ainda o curso de **Intercompreensão Português / Espanhol / Francês**, que visa proporcionar uma abordagem alternativa às línguas estrangeiras, destinada a jovens adultos e a adultos que pretendam sobretudo ler nessas línguas (jornalistas, tradutores, agentes culturais, investigadores nas mais diferentes áreas do saber, estudantes do ensino médio-avançado e superior, etc.).

Se pensarmos que cerca de 798 milhões de pessoas têm como língua materna ou segunda uma língua românica, este projecto poderá dar um pequeno contributo para a intercompreensão de vários povos.

Outros cursos de Português para Fins Específicos estão em desenvolvimento, nomeadamente para português para diplomatas.

Na área da Formação propriamente dita, o CVC dispõe de Cursos em três áreas:

Ensino	Tradução interpretação	Cultura
Ensino e Aprendizagem do Português Língua Segunda	Tradução e Tecnologias de Informação Linguística	Pós-graduação em Cultura Portuguesa Contemporânea
Materiais Interativos para Português Língua Segunda na Web 2.0	Interpretação de Conferências	Literatura Dramática Portuguesa Contemporânea
A Nova Norma Ortográfica da Língua Portuguesa		A Novíssima Poesia Portuguesa e a Experiência Estética Contemporânea
Pragmática Linguística e Ensino do Português: A Comunicação Oral e Escrita		Literaturas Africanas de Língua Portuguesa
Formação de professores na e para a intercompreensão através de práticas colaborativas on-line		Estudos Pós-Coloniais: Atlânticos Sul
Meio século de literatura portuguesa: 1880-1930		Patrimónios de Influência Portuguesa
		Primeira República e Republicanismo

Rui Vaz

O Centro Virtual Camões realizou nos últimos 5 anos cerca de **150 cursos**, abrangendo cerca de **3.000** alunos. Fruto de parcerias com Universidades e cátedras do Camões, I.P. ou através de encomendas a especialistas e investigadores, que chamamos a colaborar connosco, um aspeto fundamental destes cursos prende-se com a sua **creditação**. Consequentemente, os nossos cursos *online* são creditados quer com **ECTS** (*European Credit Transfer System*) pelas universidades parceiras quer diretamente pelo **Centro de Formação de Professores** do Instituto Camões, acreditado pelo CCPFC (Conselho Científico para a Formação Contínua-Ministério da Educação e Ciência).

Tão importante quanto a quantidade, a perceção de qualidade dos serviços prestados nesta área e a satisfação dos utilizadores reforçam a ideia de que estamos no caminho certo. De acordo com os questionários sobre qualidade média da oferta formativa no ensino a distância:

Classificação	Porcentagem
Muito Bom	41%
Bom	37%
Satisfatório	16%
Insatisfatório	4%
Mau	2%
TAL	100%

Na área do ***b-learning***, apoiamos a formação contínua, em regime misto, presencial e à distância (*b-learning*) de intérpretes da União Africana (**UA**), deputados e funcionários do Parlamento da África do Sul, bem como aos funcionários da **CEDEAO** (Comunidade Económica para o Desenvolvimento da África Ocidental), estando a ser desenvolvido outro projeto na área da tradução e interpretação com a **Universidade Pedagógica de Maputo**, Moçambique.

Em preparação encontra-se um projeto para Ensino de Português no Estrangeiro, 3º ciclo e secundário (14-18 anos), que complementarmente a presença dos cerca de 500 docentes de Língua Portuguesa que o Camões tem atualmente na sua rede de ensino básico e secundário em Espanha, Andorra, França, Reino Unido, Suíça, Alemanha, Luxemburgo, Bélgica, Holanda, África do Sul, Namíbia e Suazilândia, assim como a qualificação desta rede nos Estados Unidos, Canadá e Venezuela.

4 - Conclusão

Através do Centro Virtual Camões, o Instituto Camões dá um contributo direto à criação e produção própria de conteúdos em língua portuguesa, e, indiretamente, apoia o seu desenvolvimento através de protocolos com instituições de investigação, oferecendo, finalmente, visibilidade a produtos de investigação, desta forma colocados ao dispor da comunidade.

É necessário portanto prosseguir individualmente e em colaboração com os restantes países de língua portuguesa, no quadro das organizações oficiais, na concretização de medidas de criação e de apoio à criação, de forma a reforçar a presença da língua portuguesa na internet.



INTERNET: A NOVA GRANDE CARAVELA DA LÍNGUA PORTUGUESA NA ERA DIGITAL

Afonso João Miguel

Sobre o autor

Instituto Superior de Ciências de Educação (ISCED), Luanda, Angola.
Contato: manafonso@hotmail.com

Resumo

Este texto invoca as origens da nossa “magna língua” portuguesa, para se compreender a sua importância no mundo actual, dominado pelas TIC, e lançar um olhar para o seu futuro, que se quer mais promissor e actuante. Como se sabe, a língua levada pelos navegadores portugueses e acolhida pelos autóctones constitui hoje o elemento aglutinador dos nossos povos na CPLP. Se, no passado, esta língua portuguesa viajou de caravelas, hoje meios mais eficazes são-nos fornecidos pela ciência e pela técnica, dentre as quais a internet. E é um facto inquestionável que, em pleno século XXI, a internet facilita a divulgação da língua portuguesa, une os falantes desta língua e anula as distâncias temporais e geográficas, criando um espaço virtual comum. A dimensão da língua portuguesa no mundo virtual parece estar condicionada, entretanto, ao desenvolvimento socioeconómico, científico e tecnológico das nossas sociedades, o que passa necessariamente pelo fortalecimento das instituições nacionais. Neste caso, o conhecimento das potencialidades particulares de cada país, nos domínios em apreço, afigura-se como uma premissa fundamental. Contrariamente ao passado, nos nossos dias, deve-se respeitar a identidade de cada povo, que se revê na língua portuguesa, refreando os interesses de carácter meramente hegemónico e comercial. Este facto pressupõe, portanto, o reconhecimento das especificidades sociolinguísticas de cada país membro da CPLP, o que passa, por exemplo, pela inserção das terminologias científicas e técnicas locais no português comum a usar na internet e noutros meios digitais.

Palavras-chave

Língua portuguesa. Valor e difusão. Internet e ferramentas digitais.

1 - Ponto prévio

A minha comunicação, inserida no "Colóquio Internacional sobre a Língua Portuguesa na Internet e no Mundo Digital", intitula-se *A internet: a nova grande caravela da língua portuguesa na era digital*, e dá origem a este texto. O título em si é uma metáfora que invoca o passado da nossa "magna língua", como dignamente a caracterizou Pessoa, e uma verdadeira exaltação da sua atualidade, para se compreender a sua importância no mundo de hoje, dominado pelas tecnologias da informação e comunicação (TIC), lançando-se assim um olhar para o seu futuro, que se quer mais promissor e actuante.

Como se sabe, a língua levada pelos navegadores portugueses naquelas míticas embarcações, sob o signo dos "descobrimientos marítimos", e acolhida amigavelmente pelos autóctones, constitui hoje o elemento fundamental que aglutina os nossos povos na CPLP, sob o princípio da "unidade na diversidade".

Tomo aqui como ponto de referência o meu País, Angola, para apresentar cerca de quatro pontos que constituem o essencial da minha abordagem, nomeadamente: i) um pouco de história, no qual faço uma breve incursão histórica da LP; ii) a hora da internet, procurando enfatizar a importância deste e de outros meios tecnológicos na (re)afirmação global à LP; iii) constatar a realidade para fazer uma breve análise do contexto angolano no que tange ao uso da internet e outras ferramentas digitais; iv) alguma prudência, chamando a atenção para a necessidade de respeito pelas diferenças socioculturais dos nossos países, na hora das decisões para a concepção e elaboração de uma plataforma comum para o uso do português na internet.

1.1 - Um pouco de história

Nunca se falou tanto da língua portuguesa como agora. Com estes ecos, ressurgem sempre na memória as imagens do longínquo século XV (1482), quando as caravelas, idas do Tejo, sulcaram as águas do Atlântico, em busca de novos portos dourados. Ventos ocidentais sopraram as velas rumo às terras do Oriente, onde abundavam preciosidades, mas levavam, também, consigo uma outra riqueza: a língua portuguesa. Então, quis o fado (destino) ou o acaso que um desses primeiros portos fossem as terras de "Mwene Kongo" (manicongo). Do Atlântico, a armada de Diogo Cão avista a foz de "Nzad'ya Kongo" (rio Zaire), onde viria a atracar definitivamente.

Depois de ultrapassadas as desconfianças mútuas, estabeleceram-se os primeiros contactos, troca de embaixadas, enfim, a amizade. E neste contexto, pode-se perceber que as línguas de ambos os povos (kikongo e português) foram o elemento fulcral nessa interacção, fosse o que o fosse o nível e a performance do seu uso nesses primeiros contactos.

A par do cristianismo e da cultura, a língua portuguesa foi, no princípio, voluntária e amavelmente acolhida pelos nativos congueses, partindo do desejo dos próprios soberanos do Kongo, com a colaboração da monarquia portuguesa. Por exemplo, mesmo ainda em finais do século XV, a pedido do rei do Kongo, Nzinga Nvemba (Nkuvu), D. João II, o rei de Portugal, apoia o seu homólogo em vários domínios, dos quais se destaca o ensino e a aprendizagem da língua, conforme se ilustra nos seguintes testemunhos históricos:

Afonso João Miguel

O rei do Congo apercebeu-se logo da distância que havia entre a cultura europeia e a dos africanos. Por isso teve o cuidado de pedir ao rei de Portugal que lhe mandasse padres, mestres de letras e oficiais mecânicos... (Martins dos Santos; apud Zau, 2007: 30).

O século XVI conheceu poucos monarcas tão desejosos de que o seu reino fosse católico e o seu povo fosse educado e letrado como Afonso I. Interessante neste desejo real, é que este reino não era na Europa Católica, mas em África. Afonso I era rei do Congo¹. (Samuels, 2011: 25).

Mas a tarefa de inculcar o português no seio dos nativos pode estar ligada, inicialmente, ao trabalho dos missionários que, na sua tarefa de difundir o Evangelho e a Fé, foram também ensinando a língua portuguesa; depois, com a criação das primeiras escolas, primeiro no reino do Kongo, no século XVI, depois no reino do Ndongo, um século mais tarde². Esta aura espalhou-se, pouco a pouco, pelo vasto território que é hoje Angola.

Desenha-se uma nova arquitetura da língua portuguesa com a emancipação política das antigas colónias portuguesas, em África, aquando das suas independências. Angola, tal como outros novos países, adota o português como língua oficial, uma decisão soberana, tomada em função da realidade sócio-histórica da época. Realça-se aqui a criação, posteriormente, de uma organização que aglutinaria os cinco países africanos de expressão portuguesa: PALOP – Países Africanos de Língua Oficial Portuguesa –, tendo como base uma língua e um passado comuns.

O retoque final desse grande edifício é dado com a constituição da CPLP, em 1996, cujos arquitetos já não são os caraveleiros, mas os governantes dos nossos estados soberanos. Com esta nova estrutura, novos ventos sopraram para a língua portuguesa e ela é hoje um espaço comum, um dos elementos identitários para todos aqueles que nela se revêm; uma língua que hoje congrega povos e culturas, aglutina interesses e vontades, articula pensamentos e ideias, entrecruza vivências e sentimentos; inflama paixões e emoções; enfim, um património comum de todos os seus falantes, independentemente da sua localização espacial.

2 - A hora da internet

No âmbito do colóquio “A língua portuguesa na internet e no mundo digital” muitas coisas já foram ditas e outras há ainda por dizer sobre esse assunto (abordagens mais técnicas e refinadas). Por isso, não vou ser repetitivo e procurar, talvez, reconhecer apenas o valor da internet e das ferramentas digitais no desenvolvimento e divulgação de qualquer língua, nesse século das novas tecnologias.

Se no passado a língua foi levada para muitos pontos do Globo, de caravelas, hoje outros meios de difusão mais eficazes e dinâmicos são-nos fornecidos pela ciência e pela técnica, dentre os quais a internet e outras ferramentas digitais; quer dizer, hoje, os ventos virtuais que sopram de e para todas as direcções, impulsionados pelo desenvolvimento tecnológico de um mundo em constantes e céleres mutações, constituem uma ferramenta ideal para que a língua portuguesa continue a trilhar o bom caminho.

Pelos dados disponíveis, podemos reconhecer uma certa pujança do uso do português na internet, conforme o testemunho seguinte:

¹ Afonso I é o nome adoptado pelo soberano do Kongo Nzinga Mbemba ou *manicongo*, após à sua conversão ao cristianismo e ao consequente baptismo.

² cf. ZAU, 2007: 33.

Uma das virtudes da internet para a língua portuguesa foi o feito de unir, pela rede, os milhões de falantes lusófonos espalhados pelo globo. Com essa união, o idioma ganhou força – e valor virtual. De 7.^a língua mais falada na Web em 2007, o português é hoje a 5.^a, ficando atrás apenas do inglês, chinês, espanhol e japonês, que, com exceção do Japão, possuem muito mais falantes do que os países da CPLP. (in <http://iilp.wordpress.com/2012/03/31/lingua-portuguesa-vale-46-por-cento-do-pib-mundial>).

É consabido que o desenvolvimento multisectorial de todo e qualquer país passa, nos dias de hoje, pela implantação e uso das tecnologias de informação e comunicação, tendo como vector a língua, que deverá servir de elemento de “charneira” na articulação desses meios com a sociedade. Entretanto, para que a língua possa desempenhar o seu papel central/mediador, precisa de acompanhar essa dinâmica, o que pressupõe necessariamente ajustá-la aos desafios do momento.

Inúmeros são os meios ou recursos tecnológicos que viabilizam esse propósito da língua portuguesa e de outras, destacando-se os seguintes: uso da língua em espaços virtuais (sítios da internet, *blogs*; *chats*; redes sociais), ensino à distância, conteúdos multimédia (videoconferência, DVD, CD...), telecomunicações, investigação científica, etc. Em suma, hoje, a internet e todos estes meios ajudaram a transformar as relações linguísticas unilaterais do passado em relações multilaterais/transversais.

2.1. - Um facto, uma constatação

Diz-se que “os factos falam por si” e é verdadeiro esse aforismo. No caso em apreço, tenho constatado um facto interessante que pode ajudar a ilustrar o valor das novas tecnologias na interação dos falantes da língua portuguesa, física e virtualmente. Trata-se de algumas emissões televisivas via satélite da DSTV/*Multichoice* Angola (também presente noutros PALOP), através dos seus canais desportivos em português (“Super Sport Maximo” 1 e 2).

O extraordinário nisso é o facto de este meio ter a capacidade de juntar comentadores (geralmente antigos praticantes) e jornalistas de vários países de língua portuguesa e não só: nos seus debates sobre o futebol, fazem um grande exercício de difusão da língua portuguesa. Angolanos (Jesus), moçambicanos (Tico-Tico; Trigo; Henrique Aly), portugueses (Manich, Rui Almeida, Luís Boa Morte) e sul-africanos, provavelmente, descendentes de portugueses, todos irmanados no espírito futebolístico celebram a universalidade da língua portuguesa, evento propiciado pelo recurso a meios digitais.

Creio que, desde as competições africanas de clubes e os Campeonatos Africanos de Futebol das Nações (CAN), passando pelos jogos do futebol europeu, até as Copas do Mundo (realce para a da África do Sul), esta televisão, a par das outras aqui não mencionadas, presta um grande serviço à língua portuguesa, que está sempre presente, unindo provavelmente milhões de aficionados ao futebol, espalhado em todos os cantos do mundo.

No meio de tudo isso, o que ganha com todo este exercício linguístico é a língua portuguesa, porque tudo é feito nesta língua, sem recurso a mediação alguma. Por isso, é um facto inquestionável que, em pleno século XXI, estes meios facilitam a expansão e a divulgação da língua e interacção entre os falantes de português, ao mesmo tempo que anula as distâncias, quer temporais quer espaciais, reduzindo a descontinuidade geográfica, que caracteriza a nossa comunidade linguística, criando um espaço virtual comum.

3 - Constatar realidades

É mais do que evidente que a vitalidade da língua portuguesa na internet e no mundo digital está condicionada ao desenvolvimento socioeconómico, científico e tecnológico das nossas sociedades, o que passa necessariamente pelo fortalecimento das instituições nacionais, tal como o reconheceu o Professor João Caetano, da Universidade Aberta, que “uma língua não pode ser mais forte que os países que a sustém. Na medida em que os países crescerem em importância no palco mundial, também o português crescerá”³. Mário Filipe da Silva, do Camões, realça também esse facto ao considerar que:

... investigadores reconhecem que, no futuro, o peso de uma língua deverá ser avaliado principalmente pela força da economia, do progresso científico, da qualidade institucional e não apenas pelo número de falantes⁴.

Assim, urge mobilizar todos os recursos necessários para o feito, com a consciencialização e o envolvimento activo dos governos e de todos os utentes da língua. Nesse caso, o conhecimento das potencialidades particulares de cada país nos domínios em apreço, afigura-se como uma premissa fundamental, para garantir a solidez desse edifício comum que se pretende construir.

Na realidade angolana, em particular, longe dos níveis alcançados por outros países da CPLP, talvez mais para o Brasil e Portugal, o acesso à internet e a outras ferramentas digitais vai registrando um crescimento considerável, a julgar pelos serviços e produtos oferecidos nesse domínio e pelo número dos seus consumidores/beneficiários, sobretudo nas zonas rurais⁵.

Sublinhem-se aqui a preocupação e o esforço do Estado Angolano, através do seu Executivo, na criação de políticas e condições necessárias, que garantam o desenvolvimento sustentável do País no domínio das TIC, procurando dotar o cidadão de ferramentas indispensáveis para enfrentar os desafios da era digital. E nesse âmbito, destacam-se seguintes aspectos:

- Criação do Ministério das Telecomunicações e Tecnologias de Informação (MTTI; cf. <http://www.mtti.gov.ao/>); destaca-se nessa instituição o Centro Nacional das Tecnologias de Informação (CNTI; cf. www.cnti.gov.ao);
- Criação do Ministério do Ensino Superior e da Ciência e Tecnologia (MESCT; cf. <http://www.mesct.gov.ao/>), entidade reguladora do ensino superior no País, que tem como uma das incumbências principais o desenvolvimento e a extensão do ensino superior para todo o País. É assim que, no quadro das políticas de redimensionamento

³ in <http://iilp.wordpress.com/2012/03/31/lingua-portuguesa-vale-46-por-cento-do-pib-mundial/>; replicado de <http://observatorio-lp.sapo.pt/pt/noticias/quanto-vale-um-idioma>.

⁴ Ver <http://observatorio-lp.sapo.pt/pt/noticias/quanto-vale-um-idioma>; acesso em 31/03/2012..

⁵ Deve referir-se que, apesar de um certo crescimento nesse domínio, apenas um número reduzido da população tem acesso à internet e, mesmo naqueles casos em que existe, esse serviço é ainda deficiente/incipiente. Por exemplo, dados do Banco Mundial, no âmbito do estudo do Índice do Desenvolvimento Humano (IDH) indicam que, em 2009, apenas 3,9% da população angolana (num universo de cerca de 16 milhões de pessoas) tinha acesso à internet (cf.

[Mas hoje tenho a certeza de que a realidade é completamente diferente, embora não tenha dados atualizados para ilustrar essa realidade.](http://www.google.com/publicdata/explore?ds=d5bncppjof8f9_&met_y=it_net_userp2&idim=contry:AGO&dl=pt-P T&hl=ptPT&q=uso+de+internet+em+em+angola#!ctype=l&strail=false&bcs=d&nselm=h&met_y=it_net_user_p2 &scale_y=lin&ind_y=false&rdim=country&idim=country:AGO&ifdim=country&hl=pt_PT&dl=pt_PT&ind=false; http://www.indexmundi.com/g/g.aspx?c=ao&v=118&l=pt).</p></div><div data-bbox=)

do ensino superior em Angola, o País conta hoje com seis universidades públicas, além de vários institutos superiores, aos quais se junta um número considerável de instituições de ensino superior privadas (universidades e institutos politécnicos, sobretudo). Como era de se esperar, quase todos estes estabelecimentos têm cursos de licenciatura em Informática e em Engenharia de Telecomunicações, áreas indispensáveis para a “digitalização” massiva da sociedade.

Neste aspeto, interessa introduzir uma nota particular para falar um pouco da realidade do Instituto Superior de Ciências de Educação de Luanda (cf. <http://isced.ed.ao/>). Como instituição de ensino Superior vocacionada fundamentalmente para a formação de professores, incluiu nos seus *currícula* a cadeira de Tecnologias de Educação, cujo funcionamento é assegurado por uma ação interdepartamental, que envolve os Departamentos de Ciências de Educação, Ciências Exatas e Informática. Atualmente está a decorrer o curso de pós-graduação (Mestrado) em Tecnologias Educacionais, em cooperação com o Instituto Superior de Educação da Universidade de Lisboa. O objetivo de tudo isto é dotar o futuro professor de ferramentas necessárias para enfrentar os desafios que a atual sociedade tecnológica coloca.

- Criação de domínios oficiais de internet para Angola (co.ao e it.ao), com a existência no País de um servidor de DNS, facilitando a concepção e elaboração de sítios de internet com identidade angolana. (cf. [www.info-angola.ao/...](http://www.info-angola.ao/)). Por exemplo, a partir de página oficial do Governo de Angola (cf. www.governo.gov.ao), todos os outros ministérios e outras instituições possuem as suas páginas na internet, facilitando assim a informação necessária, de modo virtual.
- Construção de mediatecas nas 18 províncias do País, processo já em curso, por forma a facilitar o acesso da população (com realce para a estudantil) aos meios multimédia e outros meios tecnológicos afins;
- Cresce o número de sítios de internet, quer institucionais quer particulares, geridos exclusivamente por angolanos. Realça-se também, nesse caso, um acesso quase massivo às famosas redes sociais, por parte de angolanos, de todos os extratos sociais (www.twitter.com; www.facebook.com; www.youtube.com; www.sonico.com...).
- Surgimento de empresas vocacionadas para a elaboração de produtos informáticos locais, criando condições para a afirmação de uma “indústria digital”, genuinamente angolana (*Sistec Soluções, NCR Angola, TEC Informática, Tech Industrial, InforNet...*);
- Crescimento exponencial do número de usuários de telefonia móvel, televisão digital, internet sem fio, e outros serviços afins, cujos serviços são assegurados por duas companhias nacionais de telecomunicações (*Angosat, Angola Telecom, Movicel e Unitel; Zap; Uau TV, DSTV/Multichoice Angola, Net Kuya...*), o que demonstra a importância desse sector na vida da população;
- Introdução dos serviços electrónicos/digitais no domínio da banca, destacando-se as caixas electrónicas, designadas, em Angola, por “Multicaixa”. Por exemplo, no mercado angolano, surgem assim cartões de crédito *VISA Mwangolé, VISA Kamba, VISA Kumbu*, para as operações financeiras internacionais, para além os de uso interno.

Nesta era digital, é evidente que a internet e outras ferramentas do mundo virtual jogam um papel crucial para o desenvolvimento sustentável de qualquer país, inserido nesta “aldeia global”, onde a partilha de informação e do conhecimento constituem uma premissa fundamental.

Afonso João Miguel

Creio, portanto, perante o quadro acima traçado, que o que ganha com todo este exercício é a língua portuguesa, projetando-se cada vez mais, porque tudo é feito nesta língua, sem recurso a línguas estrangeiras. Por isso, é possível que pesquisas futuras sobre o uso do português na internet, incluindo todos os países da CPLP, podem ditar outros resultados, cuja tendência é a subida da percentagem no "ranking" global.

3.1 - Uma experiência

Esta experiência tem a ver com o projeto de criação do Centro de Ensino a Distância da Universidade Agostinho Neto (CEAD-UAN; cf. <http://cead-uan.com>), do qual fiz parte; fui convidado como docente do ISCED de Luanda para montar a cadeira de Língua Portuguesa (da qual fui Regente) e para a elaboração dos respectivos conteúdos.

A referência particular desse centro neste trabalho resulta da sua relevância na implementação das TIC na sociedade angolana, podendo servir mesmo de modelo nesse domínio. Pode considerar-se como uma das iniciativas pioneiras na área do ensino à distância, em sistema de aulas virtuais, em Angola.

Trata-se de um centro dotado de meios de alta tecnologia, criado com o apoio da cooperação portuguesa e da Universidade Aberta de Lisboa. Neste âmbito, o centro utiliza essencialmente as seguintes ferramentas tecnológicas para o ensino a distância: *Mikogo*; *Skype*; *Office Communication*; *Plataforma moodle/e-learning* (<http://moodle.uan.ao/moodle/>); PSP; IPOD; MP3; *smartboard*⁶, dispo de uma sala de videoconferência, uma sala Web, estúdio de gravação audiovisual digital e sistema de conversão em vídeo, para além dos serviços de apoio editorial. No domínio dos conteúdos, destaca-se a produção dos seguintes materiais, elaborados nas cadeiras de Língua Portuguesa, Estatística, Química e Informática: DVD, CD, gravações de videoconferências em suporte digital, PowerPoint assistido em computador, Manual de Apoio e Guia de Estudo informatizados.

Essencialmente, a "CEAD-UAN utiliza a VC como meio de comunicação principal nas suas atividades de ensino à distância. Estas atividades reproduzem o modelo da aula clássica, ministrada a distância por recurso à VC e complementada por diversos materiais de apoio e pela possibilidade de interagir posteriormente via audioconferência, *e-learning* ou correio eletrónico" (*in* Folheto informativo da CEAD-UAN, Luanda, 2009).

3.2 - Uma amostra:

Como amostra de trabalho que realizamos nessa instituição, foi exibido um vídeo de uma aula virtual de língua portuguesa, no âmbito do ensino a distância.

Sendo hoje uma realidade, a CEAD-UAN é membro da: *Global Development Learning Network* (GDLN), Rede Africana de Ensino a Distância, Rede Lusófona de Ensino a Distância⁷, Rede Angolana de Ensino a Distância e afigura-se como um grande parceiro do Executivo angolano no desenvolvimento do país, conforme os propósitos da sua criação, que assentam fundamentalmente no seguinte objetivo global:

⁶ JOSÉ, 2009.

⁷ Conforme se pode ler no projeto de criação do CEAD-UAN, neste momento existem apenas 4 centros GDLN em países lusófonos: Brasil: Centro GDLN de Brasília (GDLN Brasil); Moçambique: Centro de Informática da Universidade Eduardo Mondlane (CIUEM); Portugal: Centro de Ensino a Distância da Universidade Aberta (CEAD-UAB); Timor-Leste: Centro de Ensino a Distância - Dili (CEAD-Dili).

O objectivo global deste projecto consiste em apoiar o processo de desenvolvimento em Angola e, mais especificamente, as estratégias de redução da pobreza, através da promoção de programas de ensino, de actividades de (in)formação ou de simples troca de experiências e conhecimentos por recurso às TIC e às metodologias de ensino a distância, por forma a contribuir para a qualificação de recursos humanos e instituições e para o apoio técnico à resolução de problemas relevantes ou à implementação de projectos de desenvolvimento. Uma atenção especial será direccionada para a melhoria do ensino superior publico nas Províncias. (in CEAD-UAN. Projecto da Rede de Ensino a Distância de Angola. Luanda, s/d, p. 2).

Uma das vertentes específicas do referido centro consiste em "permitir dinamizar a difusão global de conteúdos de formação em português e a partilha de boas práticas de desenvolvimento entre instituições dos Países de Língua Portuguesa⁸. Nesse caso, o centro pode funcionar, portanto, como mais uma das fontes de difusão e expansão desta língua, à escala global, a julgar pelas suas potencialidades tecnológicas, conforme descrito acima.

4 - Alguma prudência e solidariedade

Tal como no passado, parece-nos que, também hoje, as razões económicas estão subjacentes nessa aventura da universalização da língua portuguesa. Mas, contrariamente à época das caravelas, nos nossos dias, que se respeite a identidade de cada povo, que se revê comumente na língua portuguesa, em detrimento de interesses hegemónicos e comerciais ou económicos⁹. Este facto pressupõe, portanto, o reconhecimento das diferenças ou especificidades sociolinguísticas de cada país membro da CPLP, o que passa, por exemplo, pela inserção de termos locais, decorrentes ou não das línguas autóctones, no português comum a usar na internet e noutros meios digitais. Uma das maneiras é, sem dúvida, o uso de glossários terminológicos que venham a contribuir para a elaboração do Vocabulário Técnico e Científico da Língua Portuguesa.

Um trabalho conjunto será a chave de sucesso para esta grande empreitada. E penso que, mais uma vez, Brasil e Portugal, pela experiência acumulada nesta matéria, devem auxiliar técnica e cientificamente os outros países que necessitarem de apoio.

Por outro lado, é necessário que se acautelem os custos; ou seja, que se analisem os aspetos económicos decorrentes de uma plataforma comum para o uso do português na internet (despesas e lucros), numa era em que tudo é vendível.

5 - Nótulas finais

1. A língua portuguesa, inicialmente falada em Portugal, é levada pelos navegadores portugueses para outras partes do Mundo. Em contacto com realidades socioculturais e linguísticas locais, esta "língua magna" ganha novas matizes, novas colorações, que se tornam cada vez mais nítidas, à medida que se consolida a sua universalidade, guindada pelos ventos da era digital; uma vez mais "internetalizada", rompe as barreiras do tempo e territoriais, tornando-se assim uma língua atemporal e do mundo.

⁶ COSTA, 2009: 4.

⁷ Não se pretende aqui expressar, entretanto, uma visão utópica da questão, quando hoje dados objetivos permitem aferir o valor económico real da língua, veja-se o caso particular de Portugal, no qual a língua portuguesa representa 17% do PIB desse país.

(cf. <http://iilp.wordpress.com/2012/03/31/lingua-portuguesa-vale-46-por-cento-do-pib-mundial/>).

Afonso João Miguel

Creio, portanto, perante o quadro acima traçado, que o que ganha com todo este exercício é a língua portuguesa, projetando-se cada vez mais, porque tudo é feito nesta língua, sem recurso a línguas estrangeiras. Por isso, é possível que pesquisas futuras sobre o uso do português na internet, incluindo todos os países da CPLP, podem ditar outros resultados, cuja tendência é a subida da percentagem no "ranking" global.

2. Apesar de certas limitações, existem condições necessárias em Angola para o uso das tecnologias de informação, em todos os domínios sociais, tendo como meio de expressão a língua portuguesa; entretanto, é necessário que se continue a mobilizar recursos tecnológicos e humanos, para se atingirem níveis satisfatórios nesse campo. Realça-se aqui o engajamento e a preocupação das autoridades locais nas políticas de desenvolvimento centradas nas TIC.

3. A concepção de estratégias comuns para o uso da língua portuguesa na internet e nos meios digitais deve partir de um trabalho diagnóstico em cada país membro da CPLP, para se ter uma ideia objetiva da realidade, aferindo assim as suas potencialidades nesse domínio.

4. A criação de espaço comum do português na internet passa pela uniformização de certas linguagens, o que pressupõe o uso de glossários terminológicos que venham a contribuir para a elaboração do Vocabulário Técnico e Científico da Língua Portuguesa, tendo em conta as especificidades dos nossos países.

Referências

AEDO, R. F. *et al.* 2008. Introdução das Tecnologias da Informação e o Conhecimento em Angola. O papel do professor e o estudante nessas mudanças. *Revista Udesc Virtu@l*. Vol. 1, No 2.

<http://revistas.udesc.br/index.php/udescvirtual/article/viewFile/1649/1328>>.

CEAD- UAN (s/d) *Projecto da Rede de Ensino a Distância de Angola*. Luanda.

CEAD-UAN (s/d) *Folheto de Apresentação do Centro de Ensino a Distância da Universidade Agostinho Neto*. Luanda.

CEAD-UAN. 2009. *Folheto informativo da CEAD-UAN*. Luanda.

COSTA, José Venâncio da. 2009. O Ensino à Distância: uma Nova Estratégia de Ensino e Aprendizagem na Universidade Agostinho Neto. *II Seminário Científico-Metodológico sobre o Ensino a Distância na UAN*. Luanda, de 15 a 16 de set. de 2009.

SAMUELS, Michael Anthony. 2011. *Educação ou Instrução: A História do Ensino em Angola [1878-1914]*. Luanda: Mayamba Editora.

ZAU, Maria Ermelinda (Coord.). 2007. *Uma Revolução Silenciosa em Angola. O Ensino Técnico-Profissional antes e depois da Independência: Uma Contribuição para a História da Educação*. Salvador, Bahia-Brasil, MED (Angola)/Raimundo João (Editor).

Sites

<http://iilp.wordpress.com/2012/03/31/lingua-portuguesa-vale-46-por-cento-do-pib-mundial/>
<http://www.gov.ao/>

<http://www.isced.ed.ao/>

<http://www.mtti.gov.ao/Institucionais/Missao.aspx>

<http://www.mesct.gov.ao/Institucionais/Atribuicoes.aspx>

<http://www.indexmundi.com/g/g.aspx?c=ao&v=118&l=pt>

<http://observatorio-lp.sapo.pt/pt/noticias/quanto-vale-um-idioma>

http://www.google.com/publicdata/explore?ds=d5bncppjof8f9_&met_y=it_net_userp2&idim=contry:AGO&dl=pt-PT&hl=ptPT&q=uso+de+internet+em+em+angola#!ctype=l&strail=false&bcs=d&nselm=h&met_y=it_net_user_p2&scale_y=lin&ind_y=false&rdim=country&idim=country:AGO&ifdim=country&hl=pt_PT&dl=pt_PT&ind=false



CARTA DE GUARAMIRANGA

COLÓQUIO INTERNACIONAL DE GUARAMIRANGA A LÍNGUA PORTUGUESA NA INTERNET E NO MUNDO DIGITAL

Os participantes do Colóquio Internacional de Guaramiranga sobre o Português na Internet e no Mundo Digital, organizado pelo Instituto Internacional da Língua Portuguesa (IILP), reunidos em Fortaleza, Redenção e Guaramiranga, Ceará, Brasil, de 23 a 26 de abril de 2012,

considerando que:

- é fundamental a presença das línguas na Internet e no mundo digital para a cultura, a economia, a educação e para o desenvolvimento da sociedade;
- é necessário manter e ampliar a vitalidade da língua portuguesa na Internet, no mundo digital e na sociedade do conhecimento em geral;
- a Internet pode proporcionar acesso ao conhecimento e a oportunidades de aprendizagem e formação;
- a revolução tecnológica estimula a criação de novos conteúdos, novas metodologias e padrões relacionados com a busca e avaliação de informação, formação do público leitor e intercâmbio do conhecimento;
- os suportes digitais – através da Internet em particular – são um meio menos oneroso da publicação e disseminação de conhecimento;
- o português é atualmente a quinta língua com mais utilizadores da Internet, mas tem um índice de produtividade (número de produtores de conteúdos por falante) inferior ao das outras línguas mais usadas;
- a Internet pode funcionar como espaço de confluência para a língua portuguesa, atenuando os efeitos da dispersão geográfica dos países que a adotaram como língua oficial;
- a criação de novos recursos linguísticos e de meios digitais de difusão do conhecimento pode permitir a instituição de um modelo aperfeiçoado de gestão das políticas da língua portuguesa, comum e mais representativo de todos os países;

- a evolução da Web mundial com a inclusão de tecnologias semânticas, possibilitando novas hiperligações de conteúdos na rede carregados de sentidos, trará novos desafios para a inclusão da língua portuguesa no estabelecimento de padrões e metadados na Internet global;
- existem diversas iniciativas – educativas e de difusão de ciência em português – mas que as plataformas que as alojam não estão ligadas e nem sempre permitem a interoperabilidade, ocasionando redundância e elevação de custos da criação de recursos;
- existem assimetrias e situações diversas entre os diferentes países quanto aos recursos existentes;
- o sucesso das línguas no domínio digital não depende apenas do seu uso efetivo na Internet, mas também da sua utilização nas ferramentas de processamento computacional;
- a inclusão das línguas nas ferramentas mais avançadas depende, por sua vez, da existência de conhecimento e de recursos linguísticos de base adequados às necessidades do processamento computacional;
- é necessário ampliar as tecnologias da informação e comunicação para o uso e difusão da língua portuguesa na sociedade do conhecimento;
- combinada com aplicações inteligentes, a tecnologia da linguagem contribuirá para concretizar o potencial da sociedade da informação: ajudará as pessoas a comunicarem entre si, mesmo que não falem uma língua comum; servirá de apoio a interfaces intuitivas com todo o tipo de dispositivos, desde eletrodomésticos até robôs pessoais;
- iniciativas de criação de recursos linguísticos em andamento e futuras, como o Vocabulário Ortográfico Comum (VOC), podem ajudar a reduzir o fosso na representatividade das variedades do português nos países em que é falado;
- há centros de investigação com capacidade instalada para fazer avançar a tecnologia da linguagem para a língua portuguesa, sendo porém necessário garantir foco estratégico para o esforço aplicado a esta área, de modo a manter a capacidade atual e alcançar um patamar de produtividade sustentada;

e considerando, ainda, as recomendações da Carta de Maputo sobre a diversidade linguística no espaço da Comunidade dos Países de Língua Portuguesa (CPLP) e da Carta da Praia sobre o português nas diásporas, bem como as virtualidades das novas tecnologias da informação e comunicação que apoiam estas recomendações, os participantes do Colóquio de Guaramiranga recomendam ao Instituto Internacional da Língua Portuguesa que:

- sejam desenvolvidos novos e mais aprofundados estudos sobre a presença do português na Internet, no âmbito de um observatório específico e que o IILP incentive a formação de redes orientadas para a investigação e desenvolvimento de pesquisas na área;
- os projetos internacionais de partilha de informação sejam estimulados, em ambiente colaborativo, a par das iniciativas nacionais;
- se crie um grupo de trabalho no IILP para identificar e estudar os projetos e portais de publicação de acervos de equipamentos culturais como bibliotecas, museus e repositórios digitais de produção científica das diversas instituições dos países da CPLP, para propor padrões e metadados comuns que proporcionem maior interoperabilidade internacional dos acervos e catálogos da língua portuguesa;

- estimule o desenvolvimento de uma gestão conducente à disponibilização digital através da Internet dos acervos e da produção científica de universidades e outras instituições;
- defina políticas públicas integrais (incluindo acesso, conteúdo e alfabetização informacional) para a presença no mundo digital das mais de 300 línguas faladas no espaço da CPLP;
- a Revista do Instituto Internacional da Língua Portuguesa, meio potencialmente importante para a presença do IILP na Internet, tenha arbitragem científica e seja indexada;
- incentive os governos a valorizarem a produção de ciência em português, concebendo a internacionalização da ciência não apenas no sentido de os falantes de português produzirem no exterior, mas também no de atrair falantes de outras línguas para a produção de ciência em português;
- dê seguimento à proposição do Plano de Ação de Brasília e crie o Portal do Professor, plataforma na Internet integrada no portal do IILP, para o ensino e aprendizagem do português comum a todos os países da CPLP;
- colabore para que os portais oficiais dos governos dos países da CPLP sejam escritos em português e, se necessário, traduzidos em outras línguas.

Os participantes do Colóquio de Guaramiranga recomendam igualmente a incorporação, na discussão da Segunda Conferência Internacional sobre o Futuro do Português no Sistema Mundial, a realizar-se em outubro de 2012 em Portugal, e a inclusão no Plano de Ação de Lisboa para a Promoção, Difusão e Projeção da Língua Portuguesa dos seguintes pontos:

- desenvolver políticas que atenuem o fosso existente entre países quanto ao acesso ao mundo digital e com a largura de banda apropriada;
- estimular a interoperabilidade entre recursos existentes, nomeadamente portais, podendo alargar o âmbito de pesquisa dos portais nacionais existentes a todos os países de língua portuguesa;
- provocar a criação de recursos educativos digitais sobre a língua portuguesa, tanto como língua materna como segunda;
- criar centros de recursos multimédia em linha nos países atualmente com menor acesso à internet, que possam servir de apoio à implementação de programas de formação, particularmente em Língua Portuguesa e em tecnologias da informação e comunicação quer a distância quer em b-learning (semipresencial);
- fomentar programas de pesquisa sobre processamento computacional da língua portuguesa, linguística computacional e inteligência artificial nas Universidades dos países da CPLP;
- originar programas especificamente inter e multidisciplinares para a ciência e a tecnologia da linguagem e o processamento computacional da língua portuguesa;
- fomentar o desenvolvimento de recursos linguísticos e de ferramentas, necessariamente padronizados, abertos e de distribuição livre, para o processamento computacional do português e a implementação de aplicações da tecnologia da linguagem aplicadas à língua portuguesa;

- estimular programas de cooperação entre os países de língua portuguesa que promovam a partilha e a transferência de conhecimento e providenciem um igual estado de desenvolvimento tecnológico das diferentes variedades nacionais da língua;
- promover a adesão a infraestruturas de investigação internacionais dedicadas especificamente à ciência e tecnologia da linguagem natural, como é o caso do CLARIN, a primeira infraestrutura internacional para a área, criada em fevereiro de 2012;
- criar, no âmbito do IILP, uma comunidade virtual que se constituirá em um fórum para debate e acompanhamento das ações sobre a promoção da língua portuguesa no universo digital;
- possibilitar o acompanhamento e a participação do IILP na iniciativa em curso Ambiente Colaborativo em Governo Eletrônico da CPLP <http://cplp.governoeletronico.gov.br/>
- estimular, através da difusão de conteúdos televisivos, a presença da língua portuguesa nos meios internacionais de comunicação e em países falantes de outras línguas; igualmente promover o desenvolvimento e acesso a conteúdos educativos em português através dos meios de comunicação digitais;

Essas recomendações, extraídas do conhecimento partilhado no Colóquio de Guaramiranga sobre a Língua Portuguesa na Internet e no Mundo Digital representam um entendimento comum dos participantes e visam contribuir para a atuação do IILP e da CPLP.

Guaramiranga, 26 de abril de 2012

