

Implicações do acordo ortográfico da língua portuguesa de 1990 em programas informáticos de correção ortográfica

Helena Figueira

Afonso Mendes

Pedro Mendes

Cláudia Pinto

Resumo

A iminente ratificação por Portugal do Acordo Ortográfico de 1990 terá inevitavelmente consequências na norma ortográfica vigente em Portugal (decorrente do acordo de 1945) e no Brasil (decorrente do formulário de 1943). A análise que neste artigo se pretende fazer centra-se na aplicação prática das novas regras ortográficas a programas informáticos, como o FLiP ou o Novo Corretor Aurélio, que incluem correção ortográfica e sintática, e não tanto na polémica relativa aos aspectos positivos ou negativos do referido acordo, nem no seu impacto cultural, social ou económico.

Palavras-chave: Língua portuguesa, ortografia, acordo(s) ortográfico(s), correção ortográfica.

Abstract

The imminent ratification of the 1990 Orthographic Convention by Portugal will inevitably produce consequences in the European Portuguese standard (which presently follows the 1945 Convention) and in the Brazilian standard (which follows the 1943 Convention). This article mostly focuses on the changes the 1990 Convention implies in spelling checkers such as FLiP or Novo Corretor Aurélio, software tools that include spelling and grammar verification, and does not intend to discuss the negative and positive aspects of the Convention nor its cultural, social and economical impact.

Key-words: Portuguese, orthography, spelling reform(s), spell checking.

1. Introdução

O “Acordo Ortográfico da Língua Portuguesa”¹, assinado em 1990 pela Academia das Ciências de Lisboa, Academia Brasileira de Letras e delegações de Angola, Cabo Verde, Guiné-Bissau, Moçambique e São Tomé e Príncipe, com a adesão da delegação de observadores da Galiza, está, passadas quase duas décadas, na iminência de ser ratificado por Portugal, após a ratificação pelo Brasil (2004), por Cabo Verde

¹ Aprovado, para ratificação, na Assembleia da República Portuguesa a 4 de Junho de 1991 e publicado a 23 de Agosto de 1991 em *Diário da República*. O texto oficial anotado está disponível em <http://www.priberam.pt/docs/AcOrtog91.pdf>.

(2006) e por São Tomé e Príncipe (2006). Este documento resulta das alterações feitas ao Acordo Ortográfico de 1986 (texto à data veementemente contestado e discutido, sobretudo em Portugal²), atendendo assim a críticas e reações que então emergiram. Atualmente, o debate de há 18 anos parece querer reavivar-se na comunicação social e na sociedade, ainda que com menor intensidade³.

O presente artigo tem por intuito debater algumas questões que não são, geralmente, afloradas em discussões sobre o assunto, seja pela superficialidade ou pela paixão com que o tema é abordado, seja por manifesto desinteresse ou desconhecimento. No entanto, no âmbito do processamento da língua natural (PLN)⁴, faz todo o sentido questionar em que medida as novas regras ortográficas irão afetar o comportamento de corretores ortográficos de português, relativamente ao FLiP, para o português de norma europeia, relativamente ao Novo Corretor Aurélio, para o português de norma brasileira, ou ainda relativamente a produtos futuros, como pacotes de ferramentas linguísticas destinados a escreventes das diversas variedades africanas do português. O impacto será sentido com mais visibilidade pelos inúmeros utilizadores da versão portuguesa do Microsoft Office.

Como consideração preliminar, é pertinente fazer uma breve cronologia da ortografia portuguesa (ver secção 1.1) e esclarecer os conceitos-chave de *ortografia*, *erro ortográfico* e *norma* (ver secção 1.2). Em seguida, analisa-se a problemática da correção ortográfica em que se inserem os corretores ortográficos do FLiP e do Novo Corretor Aurélio (ver secção 2), cujo modo de funcionamento se descreve resumidamente. Segue-se uma discussão da aplicação do Acordo Ortográfico de 1990 aos corretores ortográficos da Priberam e o seu impacto em recursos linguísticos ou em módulos que podem ser individualizados e integrados noutros produtos e serviços (ver secção 3).

² A título de exemplo, veja-se *A Questão do “Acordo Ortográfico”*, de 1988, uma compilação de artigos do Movimento contra o Acordo Ortográfico ou *A Demanda da Ortografia Portuguesa – Comentário do Acordo Ortográfico de 1986 e subsídios para a compreensão da Questão que se lhe seguiu*, de 1987, livro organizado pelos linguistas Ivo Castro, Inês Duarte e Isabel Leiria.

³ A este respeito, é de referir o debate realizado a 31 de Janeiro de 2008 na Casa Fernando Pessoa, com a participação dos escritores José Eduardo Agualusa e Vasco Graça Moura, dos linguistas Ivo Castro e João Malaca Casteleiro e do editor Nelson de Matos, transmitido no programa *Escrita em Dia*, da Antena 1, nos dias 6 e 13 de Fevereiro de 2008. A audição do programa, disponível em <http://ww1.rtp.pt/multimedia/index.php?prog=1002>, permite constatar que se mantém, embora menos marcadamente, o mesmo tipo de críticas de que foi alvo o texto de 1986, as quais eram essencialmente de cariz linguístico, geopolítico, purista ou nacionalista.

⁴ Sobre o desenvolvimento de ferramentas para PLN, levado a cabo na Priberam, ver Amaral *et al.* (2003).

1.1. Breve cronologia da ortografia portuguesa

A história da ortografia (ver secção 1.2) portuguesa pode dividir-se em três períodos essenciais⁵, que de seguida se apresentam sucintamente.

O primeiro período, geralmente denominado período fonético, que vai do séc. XIII até meados do séc. XVI, caracteriza-se pela raridade de documentos escritos e pela falta de sistematicidade, dependendo a grafia das opções dos copistas para a reprodução dos sons, com uma consequente instabilidade ortográfica.

O segundo período, denominado período pseudoetimológico⁶, tem como balizas o séc. XVI e o início do séc. XX, caracterizando-se pela difusão de grafias (supostamente) etimológicas e pela ausência de uniformidade.⁷

A partir do final do séc. XIX⁸, mas sobretudo do início do séc. XX, inicia-se, finalmente, o chamado período das reformas ortográficas⁹, um período de estabilização da ortografia.

É no início do séc. XX que se assiste ao surgimento dos primeiros textos legais que visam regular a forma como se escreve a língua portuguesa. Em 1911, o governo português aprova uma reforma ortográfica e instituiu uma ortografia oficial, sem que no entanto haja contactos com o Brasil.

Em 1931 é aprovado o primeiro acordo ortográfico entre Portugal e Brasil, mas a Academia das Ciências de Lisboa ignora algumas disposições. Esta academia publica, em 1940, o *Vocabulário Ortográfico da Língua Portuguesa* e em 1943 é a vez de a Academia Brasileira de Letras publicar o seu *Pequeno Vocabulário Ortográfico da Língua Portuguesa*, precedido de um *Formulário Ortográfico*.

Em 1945, a Conferência Interacadémica de Lisboa para a unificação ortográfica da língua portuguesa dá origem aos três documentos que constituem o Acordo Ortográfico de 1945¹⁰, aprovado pelos governos português e brasileiro. A Academia das Ciências de Lisboa publica, em 1947, o *Vocabulário Ortográfico Resumido da Língua Portuguesa*. No Brasil, no entanto, a referência continua a ser o vocabulário publicado pela Academia Brasileira de Letras em 1943 e em 1951 é mesmo revogado o Acordo de 1945.

⁵ Ver Cuesta e Luz (1971), Marquilhas (1987) e Tavani (1987).

⁶ Ver Cuesta e Luz (1971: p. 337)

⁷ Ver Gonçalves (2003) e Marquilhas (2000).

⁸ Ver Castro e Leiria (1987b).

⁹ Ver Cuesta e Luz (1971: p. 337)

¹⁰ Aprovado em Portugal pelo Decreto n.º 35 228, de 8 de Dezembro de 1945. O texto oficial anotado está disponível em http://www.priberam.pt/docs/AcOrtog45_73.pdf.

Em 1971 elabora-se um acordo reduzido entre as academias portuguesa e brasileira que apenas é tornado oficial no Brasil. Em 1973, o governo português faz pequenas alterações ao texto legal de 1945¹¹.

No seguimento de negociações feitas em 1971 e 1975 e retomadas em 1986, é assinado neste ano um novo Acordo Ortográfico por Angola, Brasil, Cabo Verde, Guiné-Bissau, Moçambique, Portugal e São Tomé e Príncipe, mas as reações da opinião pública portuguesa inviabilizam a sua ratificação¹². Os mesmos países (a que se juntará Timor Leste, após 2002) assinam um novo texto em 1990, que pretende ser um compromisso entre o de 1945 e o de 1986 e que carece ainda de ratificação em Portugal para vigorar.

1.2. Ortografia, erro ortográfico e norma

A linguagem escrita, contrariamente à linguagem oral, não é adquirida por exposição a textos escritos, antes resulta de aprendizagem¹³, isto é, de alfabetização. Através de processos de repetição, de leitura e de memorização há uma aprendizagem formal, geralmente institucionalizada, da representação gráfica da oralidade. Tal representação obedece a regras convencionadas, algumas das quais adquiridas de forma explícita (por exemplo, o ensino da regra que estipula que os advérbios terminados em *-mente* não são acentuados graficamente).

Para a definição de um conceito de *ortografia* pode partir-se da consulta de um dicionário: “forma correta de escrever as palavras; parte da gramática que ensina a escrever corretamente as palavras de uma língua”¹⁴. Constitui então a ortografia apenas a parte da língua mais convencionada e menos natural, que corresponde à forma gráfica que esta assume nos seus suportes escritos. Esta consideração é tanto mais importante quanto é falso o argumento, tantas vezes invocado, de que um acordo que pretenda uma “ortografia unificada” pretende também uma língua unificada.

Sendo um compromisso convencionado entre a fonética e a tradição gráfica, relacionada muitas vezes com a etimologia¹⁵, a ortografia foi objeto de avanços e recuos,

aproximações e distanciamentos na sua relação com a fonética e com a etimologia¹⁶ e é sobre este movimento que muitas vezes se fazem os debates.

Por outro lado, “a ortografia, «ciência de escrever corretamente» [...], é ciência árida que todos aprendemos a desamar. Numa época em que a mentalidade está «convertida» para uma valorização quase exclusiva da inteligência analítica e criadora [...] torna-se difícil de aceitar um estudo assente na memória visual e auditiva, apresentado por vezes de uma forma agressiva e desencarnado de um contexto ideológico”¹⁷, sendo contudo “a bitola de uma certa eficiência e categoria sociais”¹⁸. Mais do que isto, a ortografia contemporânea, sendo uma convenção estabelecida e estabilizada, é uma das partes menos problemáticas da gramática, talvez a área da língua cujas dúvidas permitem respostas mais perentórias em termos de correção¹⁹. O reflexo deste facto é o espaço reduzido que a ortografia ocupa nas gramáticas, uma vez que não há margem para grande problematização sobre algo previamente convencionado.

No caso da língua portuguesa, a ortografia é o único aspeto em que há uma regulamentação explícita através de textos legais. Presentemente, e no caso da norma europeia, estão em vigor dois textos legais, o Decreto n.º 35 228, de 8 de Dezembro de 1945 e a sua alteração pelo artigo único do Decreto-lei n.º 32/73 de 6 de Fevereiro de 1973. No caso da norma brasileira, está em vigor o *Formulário Ortográfico* de 1943, com alterações de 1971.

Acresce a estes dados a realidade portuguesa relativamente à política linguística, uma vez que não há, em Portugal, uma instituição com funções efetivas na definição de norma ou na normatização linguística. A este respeito, é interessante verificar a posição da Academia das Ciências de Lisboa, que, ao contrário, por exemplo, das congéneres espanhola e francesa, é uma “instituição interdisciplinar”²⁰, que não se pode assemelhar às academias espanhola e francesa, as quais “nasceram só para o estudo das [respetivas] Línguas”, como refere Pina Martins, Presidente da Academia à data do seu

¹¹ Através do decreto-lei n.º 32/73, de 6 de Fevereiro, disponível em http://www.priberam.pt/docs/AcOrtog45_73.pdf.

¹² Ver Castro e Leiria (1987a).

¹³ Ver Delgado-Martins (1996).

¹⁴ PRIBERAM INFORMÁTICA, “ortografia”, in *Priberam Informática – Língua Portuguesa On-line* [em linha], 2008, http://www.priberam.pt/dlpo/definir_resultados.aspx?pal=ortografia [consultado em 2008-02-22].

¹⁵ Em Castro *et al.* (1987: p. 37) afirma-se que “[...] a generalidade dos especialistas que se têm ocupado de questões ortográficas considera que,

numa língua de cultura com uma longa tradição gráfica, como é o caso do português, um sistema ortográfico é, necessariamente, um compromisso mais ou menos equilibrado entre etimologia, tradição gráfica e critério fonético”.

¹⁶ Sobre este assunto, ver Marquilhas (1987) e Mateus (2002).

¹⁷ Mateus (2002: pp. 92-93).

¹⁸ Mateus (2002: p. 93).

¹⁹ Nos serviços de consultoria linguística, como o que disponibiliza a Priberam em <http://www.flip.pt/Duvidas/>, este é o âmbito em que, estatisticamente, as respostas são mais curtas e o único em que podem ser categóricas.

²⁰ *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa* (2001: ix).

prefácio ao *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa*.

A par do que foi dito, é de referir que “não se pode pensar no uso de uma língua em termos de «certo» ou «errado» e em variante regional «melhor» ou «pior» [...]. No ensino da língua escrita, contudo, procura-se neutralizar as marcas identificadoras de cada grupo social, a fim de atingir um padrão único abstrato e idealizado que seja supranacional.”²¹ É neste âmbito da escrita, e sobretudo na ortografia, que mais facilmente se determinam os erros, enquanto desvio relativamente à norma²² e ao que é considerado correto, pois a norma ortográfica, dada a sua estabilidade, deve estar menos sujeita à variabilidade e à subjetividade, e não implica juízos linguísticos, mas o respeito por uma convenção. A noção de erro ortográfico é então muito clara para o utilizador da língua, uma vez que tem como referência um sistema rígido, adquirido pela escolarização e objecto de legislação.

Ao contrário das noções de *ortografia* e de *erro ortográfico*, a noção de *norma* é bem mais fluante, sendo um termo genérico que pode incluir os critérios que regulam aquilo que é considerado por um grupo maioritário como linguisticamente correto ou uma variante linguística mais prestigiada socioculturalmente, ou até politicamente. Consultando novamente um dicionário, pode encontrar-se uma definição vaga de norma como “aquilo que é de uso geral, relativamente à língua”²³. Esta formulação, obviamente não específica, permite múltiplas leituras, sobretudo no que diz respeito ao que é um uso geral, se nacional, se regional, se sociocultural, pressupondo necessariamente a (co)existência de várias normas.

Para um corretor ortográfico, no entanto, a sinalização do erro ortográfico tem de corresponder a um conceito operativo que permita um equilíbrio entre critérios linguísticos, o respeito pela norma, as exigências dos utilizadores e as limitações do PLN. Este equilíbrio, já difícil *a priori*, torna-se mais instável perante alterações ortográficas significativas, sobretudo se elas forem ambíguas ou incongruentes, num contexto de ausência de obras de referência autorizadas.

2. Correção ortográfica

É nos primeiros anos de escolaridade que o aprendente se depara com a noção de erro em geral, e de erro ortográfico

em particular, em atividades como a realização de ditados, por exemplo. Inerente a ela está a noção de correção (e a de corretivo, em épocas mais distantes). A ideia da deteção de erros ortográficos e da respetiva correção é introduzida no primeiro contacto com a alfabetização, mantendo-se mais ou menos presente ao longo da vida.

A correção ortográfica, nomeadamente a tipográfica, foi durante muito tempo apenas feita manualmente e sobre papel, mas a proliferação de textos eletrónicos criou a necessidade de outras formas de correção mais potentes e mais automáticas.

Em termos de PLN, foi na década de 70 do séc. XX que se assistiu ao surgimento dos primeiros auxiliares informáticos de correção ortográfica²⁴. Na sua fase inicial, os programas eram essencialmente verificadores ortográficos (“spelling checkers”), pois detetavam erros mas não apresentavam sugestões de correção²⁵, só mais tarde evoluindo para corretores ortográficos (“spelling correctors”)²⁶, que geralmente fornecem sugestões para os erros detetados. Em menos de uma década, os programas de correção ortográfica passaram a ser disponibilizados em processadores de texto, que conheceram nas décadas de 80 e 90 do século passado franca expansão. No início do séc. XXI, os corretores ortográficos, cada vez mais robustos e sofisticados, contêm já verificação automática e dispõem de opções de correção automática bastante seguras, podendo também ser encontrados em *web browsers*, servidores de correio eletrónico *online* ou sítios de alojamento de blogues.

Na correção de um texto eletrónico há a considerar diversos níveis: ortográfico, sintáctico e estilístico. Sumariamente, um programa de correção ortográfica permite a deteção de diferentes tipos de erros ortográficos²⁷, a saber, linguísticos ou cognitivos (ex.: **análize / análise*), tipográficos ou de digitação (ex.: **casamneto / casamento*) e de transmissão (ex.: a possível confusão de ç com c, num texto digitalizado).

2.1. FLiP

Comercializado desde 1995, o *FLiP – Ferramentas para a Língua Portuguesa* – tem vindo a afirmar-se nos mercados

²¹ Ver LEITE e Callou (2002: p. 57).

²² Sobre variação e norma, ver Mateus (2005).

²³ PRIBERAM INFORMÁTICA, “norma”, in *Priberam Informática – Língua Portuguesa On-line* [em linha], 2008, http://www.priberam.pt/dlpo/definir_resultados.aspx?pal=ortografia [consultado em 2008-02-22].

²⁴ Ver Peterson (1980).

²⁵ O primeiro verificador ortográfico remonta a 1961 (<http://www.ibiblio.org/pub/historic-linux/distributions/slackware/2.1/usr/doc/ispell/Contributors>) e usava um dicionário com as 10 000 palavras mais comuns em língua inglesa; ver Earnest (1963).

²⁶ Actualmente, em inglês, as expressões “speller”, “spelling checker” e “spelling corrector” podem designar um programa que detecta e corrige erros ortográficos, i.e., um corrector ortográfico.

²⁷ Para uma tipologia de erros ortográficos e uma breve análise das estratégias envolvidas na sua correção em PLN, veja-se Medeiros (1995: pp. 60-107).

português e internacional como uma ajuda indispensável à escrita em computador ou à revisão de texto eletrónico de diversas origens, como o texto resultante de digitalização por reconhecimento ótico de caracteres (ROC).

As ferramentas FLiP têm como antecedente o corretor ortográfico *Lince*²⁸, surgido em 1994, e que constituiu o primeiro corretor ortográfico comercializado para o português europeu. A primeira versão do *FLiP*, de 1995, continha um corretor ortográfico, um dicionário de sinónimos e um hifenizador. A segunda versão, o *FLiP 97*, passou a incluir o primeiro corretor sintático para o português.

A gama FLiP é hoje constituída por um conjunto diversificado de produtos, disponíveis para várias plataformas (Windows, Mac) e aplicações (Office, InDesign, Milenium, etc.), tanto para o português europeu como para o português do Brasil e alguns dos componentes do FLiP foram licenciados pela Microsoft para serem incluídos na versão portuguesa do Microsoft Office desde a versão XP.

Na presente versão do pacote de ferramentas linguísticas *FLiP 6*²⁹, estão incluídos, para o português europeu, um corretor sintático e estilístico, um corretor ortográfico, oito dicionários temáticos, um hifenizador, um dicionário de sinónimos, um conjugador de verbos, quatro auxiliares de tradução e um editor de texto. No *FLiP 6* incluem-se também um corretor sintático e estilístico, um corretor ortográfico, oito dicionários temáticos, um hifenizador, um dicionário de sinónimos e antónimos e um conjugador verbal para o português do Brasil, bem como corretores ortográficos para alemão, espanhol, francês, inglês e italiano.

Atualmente, o corretor ortográfico da Priberam está integrado nos blogs do Sapo³⁰, o maior portal português.

Além da correção ortográfica e sintática, a tecnologia FLiP e os seus componentes têm sido usados em diversas áreas do PLN a que a Priberam se dedica, como seja na indexação do LegiX³¹, na área de motores de pesquisa (através da participação nos projetos TRUST³² e MCAST³³), na área da recolha de

informação em respostas a perguntas (através da participação no CLEF³⁴) ou na área do processamento de fala (através da participação no projecto TECNOVOZ³⁵). O desempenho nestas áreas tem assim sido testado e avaliado internacionalmente e em aplicações práticas, com vista à consequente melhoria das suas capacidades.

No final dos anos 90 do séc. XX, a Priberam adquiriu um léxico que seria o ponto de partida para as ferramentas do português do Brasil. As primeiras versões do corretor ortográfico e do hifenizador para português do Brasil foram integradas no sistema Unisys' Hermes e são usadas desde 2001 por *O Estado de São Paulo*, um dos maiores jornais do Brasil.

O primeiro produto específico apenas para o português do Brasil foi o *FLiP:mac Brasil*³⁶ (2003), seguindo-se o *FLiP Brasil* (2004) e o *Novo Corretor Aurélio* (2007). Presentemente, as ferramentas linguísticas da Priberam para o português do Brasil incluem um corretor sintático e estilístico, um corretor ortográfico e um hifenizador. Ainda que os recursos linguísticos sejam diferentes, estas ferramentas baseiam-se no *software* que a Priberam vem desenvolvendo há mais de 10 anos.

Contrariamente ao que sucede com o léxico, que é individualizado para o português europeu e para o português do Brasil, a estrutura básica da gramática é partilhada pelos corretores sintáticos da variedade portuguesa e da variedade brasileira. Tal acontece porque, salvaguardadas as devidas diferenças, a generalidade das regras sintáticas para a escrita do português do Brasil é semelhante à do português europeu e porque os mesmos fenómenos ocorrem em ambas as variedades do português, embora por vezes em diferentes registos de língua. As especificidades do português do Brasil foram no entanto consideradas, nomeadamente construções lexicais fixas e erros fonéticos com implicações sintáticas (por exemplo, o uso problemático da crase).

²⁸ Ver Andrade (1992).

²⁹ Está disponível em <http://www.flip.pt/online> uma versão de demonstração que apresenta algumas das funcionalidades do FLiP.

³⁰ Ver <http://blogs-beta.sapo.pt>.

³¹ LegiX é um conjunto de ferramentas de informação jurídica da Priberam; ver www.legix.pt para mais informação.

³² TRUST – *Text Retrieval Using Semantic Technologies* – foi um projecto co-financiado pela Comissão Europeia (IST-1999-56416), com vista ao desenvolvimento de um motor de pesquisa multilingue, capaz de responder a perguntas em língua natural, formuladas em inglês, francês, italiano, polaco e português.

³³ M-CAST – *Multilingual Content Aggregation System based on TRUST Search Engine* – foi um projecto co-financiado pela Comissão Europeia (EDC 22249 M-CAST), com vista ao desenvolvimento de uma infra-estrutura

multilingue que permitisse aos produtores de conteúdos pesquisar, consultar e integrar recursos de vastas colecções multilingues de textos (e multimédia), tais como bibliotecas digitais, recursos de editoras, agências de notícias e bases de dados de publicações científicas (<http://www.m-cast.infovide.pl/>).

³⁴ CLEF – *Cross Language Evaluation Fórum* – é um projecto de avaliação internacional, financiado pelo 6º Programa Quadro da Comissão Europeia, que visa promover a investigação na área dos sistemas multilingues (<http://www.clef-campaign.org/>). Sobre a participação da Priberam Informática na campanha de *Question Answering* do último CLEF, veja-se Amaral *et al.* (2007).

³⁵ O projecto TECNOVOZ pretende desenvolver novas competências nas tecnologias da fala (<http://www.tecnovoz.pt/>).

³⁶ Ver <http://www.flip.pt/tabid/252/Default.aspx>.

2.1.1. Corrector ortográfico: modo de funcionamento

O corretor ortográfico tem como objectivo detetar erros de ortografia e apresentar sugestões para a sua correção. O funcionamento do corretor ortográfico do FLiP baseia-se na comparação das palavras utilizadas num documento com uma lista de palavras (“dicionário”) reconhecidas pelo módulo de correção ortográfica³⁷. O conjunto das palavras conhecidas pelo corretor ortográfico é constituído pela reunião da lista de palavras geral, das listas de palavras dos dicionários temáticos seleccionados e das palavras adicionadas aos dicionários de utilizador ativos. Se uma determinada palavra não for reconhecida, esta é assinalada como errada e são apresentadas sugestões para a sua correção.

Nas versões mais recentes da maioria das aplicações da Microsoft, a verificação ortográfica é feita automaticamente enquanto se escreve. Neste caso, os erros ortográficos aparecem sublinhados a vermelho. Clicando sobre a palavra assinalada com o botão do lado direito do rato, o utilizador poderá seleccionar uma das sugestões de substituição, ignorar a palavra ou adicioná-la ao dicionário de utilizador. Normalmente será seleccionada uma das sugestões apresentadas para substituir a palavra incorreta que se encontra no documento. Sempre que possível, as sugestões são apresentadas por ordem decrescente de probabilidade de se tratar da palavra correta. Na ordenação das sugestões são considerados fatores como a semelhança fonética, a semelhança gráfica (importante em textos digitalizados com ROC), distâncias/adjacências no teclado e frequência das palavras.

O corretor ortográfico não se limita a assinalar erros ortográficos, mas veicula opções no reconhecimento de palavras. Assim, o corretor pode assinalar como desconhecidas palavras dicionarizadas que deliberadamente não quer ver reconhecidas, como palavras obsoletas ou de baixa frequência cujo reconhecimento pode mascarar erros em palavras muito comuns (ex.: *demonstração, gostar, treinador*) ou palavras tabuísticas ou ofensivas (ex.: *cagar, merda, puta*)³⁸.

O corretor ortográfico contém ainda algoritmos de correção automática de erros. Certos erros em que não há dúvidas quanto à palavra que o utilizador pretende escrever são corrigidos automaticamente, com base na sugestão dada pelo corretor ortográfico (por exemplo, a palavra *francês* é imediatamente corrigida para *francês*).

Algumas verificações e sugestões ortográficas são detetadas pelo corretor sintático e não pelo módulo de correção ortográfica, dado que este não dispõe de informação contextual. Um caso típico é o das palavras compostas ligadas por hífen. Para expressões como *fim de semana*³⁹ ou *pequeno almoço* é sugerida a respectiva substituição por *fim-de-semana* e *pequeno-almoço*. Outro caso é o da confusão entre as palavras *à* (contração da preposição *a* com o artigo *a*) e *há* (forma do verbo *haver*). Em qualquer dos casos referidos não existe erro ortográfico se as palavras forem consideradas isoladamente; apenas a análise do contexto de ocorrência permite decidir sobre a correção ortográfica do texto em análise.

3. Aplicação do Acordo Ortográfico de 1990 em corretores ortográficos

A ortografia de um adulto escolarizado é, ou deveria ser, em geral, estável e quase automática, com pequenas áreas problemáticas que irão constituir hesitações ou dúvidas. Qualquer reforma ortográfica vai desequilibrar essa estabilidade, criando resistências à descaracterização gráfica das palavras cuja ortografia é alterada.

As principais resistências de um escrevente às alterações ortográficas relacionam-se, então, não tanto com a memória etimológica das palavras, mas com a imagem gráfica memorizada durante a aprendizagem e repetida enquanto escrevente (as formas antigas *condemnar, contracto, fructo, orthographia, pharmácia, psalmo, sancto, sciência* ou *victória* são casos em que essa memória se perdeu, sem que os utilizadores da língua atual sintam qualquer desconforto gráfico porque não têm delas memória).

Num momento de alterações ortográficas significativas, como será o caso aquando da entrada em vigor do novo acordo,

³⁷ A inclusão de novos vocábulos, seja por sugestão de utilizadores, seja por pesquisas em *corpora*, quando não atestados, é efectuada se a sua formação respeitar as regras ortográficas e morfológicas da língua portuguesa e se consagrados pelo uso. Os neologismos, estrangeirismos e aportuguesamentos polémicos ou muito recentes são, em geral, excluídos, até se tornarem consagrados pelo uso e/ou registo lexicográfico.

³⁸ Na apreciação que fez do FLiP 5, Fernando Venâncio critica esta opção: “Quando os dicionários já contêm vulgarismos sexuais e escatológicos, a Priberam mantém-se invulgarmente virginal. Por convicções morais? Longe disso. Trata-se de uma exigência da Microsoft, que não deseja ter de apresentar alternativas disfóricas. E, assim, tendo você grafado «cavalho»,

o programa supõe-lhe a intenção de redigir «cavalo», ou «cavalgo», ou «carvalho», mas exclui (o que noutro caso não faria) que você se enganou na terceira letra. É uma cedência europeia ao puritanismo transatlântico.” (*Expresso* de 10-12-2005, suplemento “Actual”, p. 76).

³⁹ Os exemplos apresentados dizem respeito ao português europeu na versão actual do FLiP 6 e seguem as disposições dos textos legais em vigor à data.

sobretudo se o texto publicado em 1991 não for alterado (a percentagem de palavras a alterar é pequena, mas a frequência de algumas palavras ou famílias de palavras é muito alta), e dado o volume de textos atualmente produzido em formato eletrónico, a confiança num corretor ortográfico é essencial. Por outro lado, e atendendo ao facto de que o corretor verifica a ortografia e oferece sugestões de correção, constitui também um instrumento privilegiado para criar novos mecanismos de tomada de conhecimento, aprendizagem e memorização das alterações ortográficas convencionadas.

Aparentemente, o texto legal tem preocupações específicas com o PLN, na “Nota Explicativa”, aquando da explicitação das razões ponderadas para a manutenção dos acentos gráficos nas proparoxítonas e paroxítonas, recuando assim no estipulado pelo texto de 1986:

“5.3. e) Alargamento, com a abolição dos acentos gráficos, dos casos de homografia, do tipo de *análise* (s.)/*analise* (v.), *fábrica* (s.)/*fabrica* (v.), *secretária* (s.)/*secretaria* (s. ou v.), *vária* (s.)/*varia* (v.), etc., casos que, apesar de dirimíveis pelo contexto sintático, levantariam por vezes algumas dúvidas e constituiriam sempre problema para o tratamento informatizado do léxico”.

Este argumento não parece, no entanto, ser pertinente, dado o número de ambiguidades entre classes gramaticais que a língua contém e com as quais o PLN tem obrigatoriamente de lidar. Mais do que isso, a eliminação das consoantes ditas “mudas” (ver *Base IV*) vai aumentar o número de palavras homógrafas (ex.: *a(c)to/ato*, substantivo e flexão verbal de *atar*; *ó(p)tico/ótico*, adjetivo relativo à vista e ao ouvido, respetivamente). A mesma consequência terá a eliminação do acento gráfico para distinguir palavras graves que são homógrafas de palavras proclíticas (ver *Base IX*, 9.^o). Veja-se como o caso particular da omissão do acento agudo na 3^a pessoa do singular do indicativo do verbo *parar* (*pára*) irá criar um par homógrafo no português europeu (mas homónimo no português do Brasil), dificultando o reconhecimento e a leitura de frases como “Por vezes para para garantir que não é poeta” ou como “Para para se oferecer como testemunha, embora já vá atrasado para um encontro”⁴⁰. Acresce que em muitos processadores de texto não são permitidas repetições de palavras. Pelo que, à luz do que já acontece em frases como “Se se tivesse esquecido, o mundo lho teria lembrado”⁴¹, nos exemplos acima mencionados será

assinalado um erro na palavra repetida, com a indicação de que uma delas deve ser eliminada. Para além disso, esta opção parece ser inconsistente com o estipulado no n^o 3 da *Base VII* para o caso do verbo *pôr* e da preposição *por*:

“3.^o Prescinde-se de acento gráfico para distinguir palavras oxítonas homógrafas, mas heterofónicas/heterofónicas, do tipo de *cor* (*ô*), substantivo, e *cor* (*ó*), elemento da locução *de cor*; *colber* (*ê*), verbo, e *colber* (*é*), substantivo. Excetua-se a forma verbal *pôr*, para a distinguir da preposição *por*.”

Repare-se como o critério que é válido para *pôr/por* não parece ser suficiente no caso de *pára/para*, o que é revelador de falta de sistematicidade.

O texto do Acordo de 1990 não prevê soluções para muitos dos problemas que cria e é lacunar ou ambíguo em alguns aspectos, pelo que será necessário definir linhas gerais explícitas e fornecer ao utilizador a explicação de algumas opções tomadas pela Priberam. Estas opções têm de ser tanto mais explícitas quanto mais sentida for a ausência de instrumentos lexicográficos autorizados, nomeadamente o “vocabulário ortográfico comum da língua portuguesa”, previsto no Acordo (ver *art.º 2.º*), que permita substituir as obras tidas como referência, de que são exemplo o *Tratado de Ortografia*⁴² e o *Vocabulário*⁴³ de Rebelo Gonçalves ou o *Vocabulário Ortográfico* da Academia Brasileira de Letras⁴⁴.

É forçoso que exista um instrumento comum, como um *thesaurus* da língua portuguesa, mas será útil ir mais longe e permitir a criação de vários vocabulários ortográficos autorizados consoante a norma de cada país, de que o “vocabulário comum” poderia ser a súpula. Estariam então, de alguma forma, organizadas *subortografias nacionais*, como propõe Ivo Castro, no seguimento da sua defesa de uma “versão fraca de unificação”⁴⁵, para evitar uma grande variação interna dentro de cada espaço nacional. A “Nota explicativa” do Acordo indicia uma solução semelhante no ponto 4.4 quando afirma que “os dicionários da língua portuguesa, que passarão a registar as duas formas em todos os casos de dupla grafia, esclarecerão, tanto quanto possível, sobre o alcance geográfico e social desta oscilação de

⁴⁰ Exemplos extraídos do *corpus* CETEMPúblico, disponível em <http://www.linguateca.pt/>.

⁴¹ *Idem*.

⁴² Ver Gonçalves (1947).

⁴³ Ver Gonçalves (1966).

⁴⁴ Ver Academia Brasileira de Letras (1981).

⁴⁵ Ver Castro *et al.* (1987: p. 8). Sobre este assunto, ver também a intervenção de Ivo Castro em “2ª parte do debate sobre o Acordo Ortográfico” do programa *Escrita em Dia*, emitido na Antena 1 da RTP em 2008-02-13, disponível em http://mp3.rtp.pt/mp3/envia_file.php?file=wavrss/at1/194713_22309-0802202249.mp3&name=Escrita%20em%20Dia

pronúncia”. É neste âmbito que se podem inserir também os corretores ortográficos em geral, e o FLiP e o Novo Corretor Aurélio em particular, uma vez que se trata de produtos comerciais que se constituem como auxiliares de escrita e de revisão de texto eletrónico.

3.1. Implementação

A ratificação do Acordo de 1990 implicará variadas alterações ao módulo de correção ortográfica do FLiP, mas sobretudo grandes opções de fundo que se podem constituir em vários cenários diferentes e implicar diferentes soluções técnicas.

As principais alterações serão ao nível dos léxicos (bases de dados relacionais onde se codifica informação lexical variada⁴⁶) relativos à variedade do português europeu e do português brasileiro.

De entre as hipóteses para implementação das alterações ortográficas, pode considerar-se a criação de um léxico único que reúna todas as variantes ortográficas possíveis de uma mesma palavra, atendendo às facultatividades possíveis (ex.: *tectónico/tectônico/tetónico/tetônico*). Esta opção parece ser claramente desvantajosa, uma vez que multiplica as variantes possíveis e não tem em conta as práticas ortográficas nacionais de Portugal e do Brasil.

Para restringir as desvantagens desta hipótese poderia ser implementado no *software* um conjunto de opções de seleção de formas exclusivas ou preferenciais pertencentes a cada uma das variedades nacionais, à semelhança da seleção feita com os dicionários temáticos. Esta alternativa apresenta, contudo, inconvenientes quando o utilizador não explora as potencialidades da aplicação e se limita a utilizar as opções predefinidas. Mais importante do que isso, um léxico único ignoraria importantes diferenças no que diz respeito às informações de cada item lexical, como sejam variações lexicais (ex.: *autocarro/ônibus, boleia/carona*), morfológicas (ex.: género de *disquete*, feminino em Portugal, masculino no Brasil), semânticas (ex.: *bala, concreto*), sintáticas (ex.: *no meio de/em meio a*) ou estatísticas (ex. a frequência de *cara* com determinado género interfere no treino de algoritmos de desambiguação), que não têm nenhuma relação com o sistema ortográfico. Este tipo de informação é, no entanto, usado no PLN, por exemplo, na correção de erros de concordância (ex.: *a/o disquete*), na desambiguação morfossintáctica (ex.: categoria gramatical da palavra *cara*) ou na

desambiguação semântica (ex. *bala=projétil; bala=rebuçado*).

O cenário aparentemente mais produtivo, nomeadamente para satisfação das exigências dos utilizadores de cada mercado, parece ser aquele em que se mantêm dois léxicos distintos para as variedades portuguesa e brasileira, cada um deles incluindo, para além das formas graficamente distintas das duas variedades, também as formas comuns. Manter-se-iam, assim, as bases de dados lexicais utilizadas actualmente no FLiP, nas quais se procederia às devidas alterações preconizadas pelo Acordo Ortográfico. A necessidade de manutenção destes dois léxicos justifica-se pelo facto de a unificação da ortografia oficial portuguesa não invalidar a manutenção de práticas ortográficas distintas nas diferentes variedades do português. A este respeito, refira-se novamente o conceito de subortografias nacionais, como norma de cada país para ensino nas escolas, difusão em gramáticas e dicionários e veiculação na comunicação social. Os dois léxicos referidos fariam parte da subortografia de cada uma das variedades do português para as quais a Priberam tem actualmente recursos. Um exemplo claro da necessidade desta separação diz respeito a pontos em relação aos quais o texto do Acordo de 1990 é omissivo, nomeadamente o da flexão de verbos terminados em *-guar* e *-quar* ou o da discordância nas duplas consoantes *-mm-* ou *-nn-* (ex.: *connosco/conosco, comumente/comumente*). Da mesma forma, o Acordo é omissivo sobre a divergência específica em formas como *húmido/úmido*, pelo que se deve manter esta distinção nas práticas ortográficas das duas variedades. De facto, o texto permite justificar ambas as grafias, a do português europeu, com *h* “por força da etimologia”, e a do português brasileiro, sem *h* “quando, apesar da etimologia, a sua supressão está inteiramente consagrada pelo uso” (ver *Base II*).

Se tomarmos a *subortografia* como um conceito operativo, não parece ser muito problemática a manutenção das sequências *-éma-, -éme-, -émi-, -émo-, -ému-, -éna-, -éne-, -éni-, -éno-, -énu-, -ôma-, -ôme-, -ômi-, -ómo-, -ómu-, -óna-, -óne-, -óni-, -óno-, -ónu-* para palavras da variedade europeia do português (ex.: *fenómeno, ténue*) e das sequências *-êma-, -ême-, -êmi-, -êmo-, -êmu-, -êna-, -êne-, -êni-, -êno-, -ênu-, -ôma-, -ôme-, -ômi-, -ómo-, -ômu-, -ôna-, -ône-, -ôni-, -ôno-, -ônu-* para palavras da variedade brasileira do português (ex.: *fenômeno, ténue*).

O caso mais problemático apontado por todos os críticos do Acordo de 1990 é o da facultatividade das chamadas “consoantes mudas” (ver *Base IV*). Ainda que a existência de duplas ortografias não seja inédita, sobretudo no português do Brasil (ex.: *acessível/accessível, aspecto/aspeto, sanguiná-*

⁴⁶ Ver Amaral et al. (2003).

rio/sanguinário⁴⁷), mas também no português europeu (ex.: *abcissa/abscissa, cisto/quisto, cobarde/covarde, loiça/louça, rotura/ruptura*), este novo texto legal vem aumentar exponencialmente o número de casos (dentro de cada país e na soma de todas as facultatividades potenciais) e dificultar as decisões a tomar pela lexicografia e também pelos ensinantes da língua portuguesa, uma vez que introduz maior variabilidade e subjectividade no ensino da ortografia e da noção de erro ortográfico. Estas decisões serão mais complicadas relativamente ao léxico técnico e científico (que corresponde aos dicionários temáticos do corretor), uma vez que se trata de palavras menos comuns, sobre as quais há pouca informação lexical (nomeadamente de carácter fonético ou ortoépico relativo à prolação ou não das consoantes) em dicionários e vocabulários de língua, não se encontrando geralmente informação de carácter linguístico em obras técnico-científicas especializadas.

Uma das áreas do léxico a poder sofrer alterações ortográficas é a dos nomes próprios, cuja grafia poderá manter-se em antropónimos, sociónimos ou marcas registadas (ver *Base XXI*), o que implicará o reconhecimento de nomes próprios com grafias não admitidas em nomes comuns (ex.: *Baptista*). A recomendação de substituição de topónimos estrangeiros “tanto quanto possível” por formas vernáculas (ver *Base I*) é, porém, suficientemente vaga para poder admitir os topónimos actualmente em uso pela comunicação social, nomeadamente diferentes grafias (ex.: *Pequim/Beijing*) para um mesmo topónimo estrangeiro; da mesma forma, o texto legal tanto pode licenciar a admissão da substituição de formas de origem estrangeira (ex.: *Zimbabwe*), como de variantes aportuguesadas (ex.: *Zimbábue/Zimbabué*). O texto do Acordo é também omissivo quanto à variação da toponímia entre as duas variedades portuguesa e brasileira (ex.: *Madrid/Madri; Irão/Irã*), o que torna premente a produção de documentos de referência a este nível e vantajosa a adoção de léxicos independentes.

Uma vez que o hífen é de utilização problemática em português⁴⁸, é importante referir as alterações que a *Base XVI* introduz relativamente ao texto de 1945, que vão no sentido de uma simplificação do uso do hífen “nas formações por

prefixação, recomposição e sufixação”. Relativamente a muitas das chamadas “formações por recomposição”, a prática ortográfica actual advém do disposto nas 16 alíneas da *Base XXIX* de 1945, complementadas pelos esclarecimentos do *Tratado de Ortografia*, de Rebelo Gonçalves, e, nos casos omissos, da tradição lexicográfica ou de relações analógicas com outros prefixos. Por exemplo, o elemento *bigio-*, de *bigiossanitário*, à falta de indicações específicas, seguirá o paradigma de outros como *hidro*⁴⁹. Segundo o texto de 1990, estes casos serão tratados segundo regras contextuais e já não seguindo listas de prefixos, que não poderão prever neologismos. Do ponto de vista do tratamento computacional do léxico, este tipo de regras contextuais é claramente vantajosa.

Por outro lado, também haverá necessidade de realizar alterações no módulo de correção sintática, uma vez que é o corretor sintático que deteta erros contextuais. Estas alterações dizem respeito, por exemplo, à distinção entre palavras hifenizadas e locuções (ex.: *fim-de-semana/fim de semana*). Neste caso, haverá necessidade de definir critérios mais explícitos, para contornar as ambiguidades do texto legal (ver *Base XV*):

“1.º Emprega-se o hífen nas palavras compostas por justaposição que não contêm formas de ligação e cujos elementos, de natureza nominal, adjetival, numeral ou verbal, constituem uma unidade sintagmática e semântica e mantêm acento próprio, podendo dar-se o caso de o primeiro elemento estar reduzido [...]”

“6.º Nas locuções de qualquer tipo, sejam elas substantivas, adjetivas, pronominais, adverbiais, prepositivas ou conjuncionais, não se emprega em geral o hífen, salvo algumas exceções já consagradas pelo uso [...]”

O caso de *fim-de-semana/fim de semana* é paradigmático, pois *fim de semana* é, no texto legal, um exemplo de locução em que não se emprega o hífen. O registo de *fim-de-semana* como palavra hifenizada é, no entanto, feito pela esmagadora maioria das obras de referência do português europeu⁵⁰; e, como contraponto, a palavra hifenizada é praticamente ausente das obras de referência do português do

⁴⁷ A título de exemplo, compare-se também opções discordantes no *Novo Dicionário Aurélio* (DA) e no *Dicionário Houaiss da Língua Portuguesa* (DH), nos verbetes *céptico* (remissivo no DA) /*cético* (remissivo no DH), *manufatura* (não registado no DH e registado no DA, a par de *manufatura*).

⁴⁸ Nos serviços de consultoria linguística disponibilizados pela Priberam (<http://www.flip.pt/duvidas>), as questões sobre o uso do hífen têm uma alta frequência.

⁴⁹ Ver Gonçalves (1947: pp. 251-252).

⁵⁰ Nomeadamente o *Vocabulário da Língua Portuguesa*, de Rebelo Gonçalves, o *Grande Vocabulário da Língua Portuguesa*, de José Pedro Machado, o *Dicionário da Língua Portuguesa On-line*, o *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa* ou o *Grande Dicionário Língua Portuguesa*, da Porto Editora.

Brasil⁵¹, sendo mais comum o registo da locução *fim de semana*⁵². O registo lexicográfico⁵³, neste caso, poderia constituir exemplo de “exceções já consagradas”, o que validaria qualquer uma das opções (como palavra hifenizada e como locução). Como exemplo desta (in)definição de critérios, pode mencionar-se ainda o facto de o texto referir apenas o emprego do “hífen nas palavras compostas que designam espécies botânicas e zoológicas”, não se reportando a outros domínios técnicos específicos em que é comum o uso de palavras compostas que entram no mesmo paradigma.

O módulo sintático, que, como se viu anteriormente, possibilita também correção ortográfica, faz uso de relações lexicais como a homografia, a homofonia e a paronímia para alertar o utilizador para possíveis confusões na escrita de algumas palavras. Estas relações lexicais, sobretudo a homofonia e a paronímia, dependem da realização fonética mais comum em cada variedade do português e em alguns casos devem ser exclusivas de cada variedade (ex.: *mal/mau* ou *cessão/seção/sessão* no português do Brasil), o que constitui argumento para a manutenção de dois léxicos individualizados.

3.2. Outras implicações

É importante reiterar a referência feita anteriormente ao licenciamento de componentes do FLiP pela Microsoft para referir o impacto do Acordo Ortográfico de 1990 em todos os utilizadores da versão portuguesa do Microsoft Office.

Adicionalmente, qualquer alteração significativa num componente do FLiP⁵⁴ pode ter consequências em outros produtos e serviços da Priberam, *on-line* ou *off-line*. Estes módulos, que podem ser individualizados e integrados em aplicações diversas, incluem um corretor ortográfico, um corretor sintático, um analisador morfológico e um conjugador de verbos e permitem incorporar tecnologia robusta, testada por dezenas de milhares de utilizadores, em produtos variados. Como exemplos de utilização destes

componentes podem citar-se a correção ortográfica e sintática integrada em serviços de *webmail*, em programas de entrada de dados ou em motores de pesquisa. É de salientar que a experiência da Priberam, com o *Dicionário da Língua Portuguesa On-Line*⁵⁵ e o *LegiX.pt*⁵⁶, revela que muitos resultados nulos obtidos nos motores de pesquisa se devem a erros ortográficos na pesquisa.

Não é de importância menor, no entanto, o impacto que as alterações ortográficas poderão ter em recursos linguísticos⁵⁷, nomeadamente em ontologias, *thesauri* ou dicionários bilingues, que são menos visíveis para o utilizador, mas não deixam de ter uma importância crucial no desenvolvimento do PLN. Da mesma forma, o trabalho com *corpora* terá de ser re-equacionado até haver um *corpus* de qualidade com a nova ortografia.

Aos recursos de que a Priberam dispõe atualmente poderão juntar-se futuramente e a médio prazo léxicos pertencentes às subortografias do português das variedades africanas, nomeadamente de Moçambique e de Angola, países onde a Priberam já estabeleceu parcerias que pretende desenvolver. As variedades africanas como objeto de projetos futuros colocam questões adicionais: as que dizem respeito às práticas ortográficas nacionais (originadas por influência anglo-saxónica ou por motivos políticos⁵⁸), cujas soluções não estão previstas no texto legal, nomeadamente a utilização das consoantes *k*, *w* e *y* em nomes comuns e não exclusivamente em nomes próprios ou derivados de nomes próprios estrangeiros (antropónimos, sociónimos ou topónimos, normalmente; ver *Base I*, 2^o) ou no uso do apóstrofo em casos não previstos nas bases. Também aqui o conceito de subortografias nacionais parece ser muito útil, embora com problemas acrescidos devido às práticas ortográficas, mas sobretudo porque sempre foram as academias portuguesa e brasileira os motores das alterações ortográficas, pelo peso tradicional das duas normas.

4. Referências

⁵¹ Não consta, por exemplo, do *Vocabulário Ortográfico da Língua Portuguesa*, da Academia Brasileira de Letras.

⁵² Nomeadamente o *Dicionário Houaiss da Língua Portuguesa*, o *Novo Dicionário Aurélio da Língua Portuguesa* ou o *Aulete Digital – Dicionário Contemporâneo da Língua Portuguesa*.

⁵³ É interessante, neste aspecto, comparar a edição brasileira (Houaiss: 2001) e a portuguesa (Houaiss: 2002) do *Dicionário Houaiss da Língua Portuguesa* e verificar que apesar de *fim de semana* estar registado nas duas edições como locução e não como palavra hifenizada, na edição portuguesa foi acrescentada a observação “também se escreve com hífen”.

⁵⁴ Ver <http://www.flip.pt/ComponentesFLiP/tabid/296/Default.aspx>.

A Questão do “Acordo Ortográfico”, s/l: Movimento contra o Acordo Ortográfico, 1988.

ACADEMIA BRASILEIRA DE LETRAS, *Vocabulário Ortográfico da Língua Portuguesa*, 1981: Bloch Editores S.A., [Rio de Janeiro] [versão actualizada em linha em <http://www.academia.org.br/abl/cgi/cgilua.exe/sys/start.htm?sid=23>].

⁵⁵ Ver <http://www.priberam.pt/dlpo/>.

⁵⁶ Ver <http://www.legix.pt>.

⁵⁷ Sobre alguns destes recursos, ver Amaral *et al.* (2003).

⁵⁸ Roth (2004: pp. 45-46).

- AMARAL, Carlos, Helena FIGUEIRA, Afonso MENDES, Pedro MENDES, Cláudia PINTO, “A Workbench for Developing Natural Language Processing Tools” in *Proceedings of the 1st Workshop on International Proofing Tools and Language Technologies*, Patras, Grécia, 2003. <http://www.priberam.pt/docs/WorkbenchNLP.pdf>.
- AMARAL, Carlos Adan CÁSSAN, Helena FIGUEIRA, André MARTINS, Afonso MENDES, Pedro MENDES, Cláudia PINTO, Daniel VIDAL, “Priberam’s Question Answering System in QA@CLEF 2007” in *Working Notes for the CLEF 2007 Workshop (CLEF 2007)*, Budapeste, Hungria, 19-21 Setembro, 2007. http://www.clef-campaign.org/2007/working_notes/AmaralCLEF2007.pdf
- ANDRADE, Ernesto d’, Helena SOARES, Isabel FRAÚSTO, “Lince, um corrector ortográfico português” in *Actas do 1º Encontro de Processamento da Língua Portuguesa – Escrita e Falada – EPLP’93*, Lisboa: s/e, 1992, pp. 97-100.
- ASSEMBLEIA DA REPÚBLICA, “Resolução da Assembleia da República n.º 26/91” in *Diário da República*, n.º 193/91 Série I-A, 23 de Agosto de 1991, pp. 4370-4388 in *Priberam Informática* [em linha], 2008, <http://www.priberam.pt/docs/AcOrtog91.pdf> [descarregado em 2008-02-29].
- Aulete Digital – Dicionário Contemporâneo da Língua Portuguesa* [descarregado em 2007, de <http://www.auletedigital.com.br/auletedigital/>].
- CASTRO, Ivo e Isabel LEIRIA, “A questão ortográfica de 1976”, in CASTRO, Ivo *et alii*, (orgs.), *A Demanda da Ortografia Portuguesa*, Lisboa: Edições João Sá da Costa, 1987, pp. 219-275.
- CASTRO, Ivo e Isabel LEIRIA, “As reformas ortográficas: do romantismo à actualidade”, in CASTRO, Ivo *et alii*, (orgs.), *A Demanda da Ortografia Portuguesa*, Lisboa: Edições João Sá da Costa, 1987b, pp. 204-218.
- CASTRO, Ivo, Inês DUARTE, Isabel LEIRIA (orgs.), *A Demanda da Ortografia Portuguesa*, Lisboa: Edições João Sá da Costa, 1987.
- CUESTA, Pilar Vásquez e Maria Albertina Mendes da LUZ, *Gramática da Língua Portuguesa*, Lisboa: Edições 70, 1971.
- DELGADO-MARTINS, Maria Raquel, “Representações da linguagem verbal” in FARIA, Isabel Hub *et alii* (orgs.), *Introdução à Linguística Geral e Portuguesa*, Lisboa: Caminho, 1996, pp. 85-102.
- Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa*, Lisboa: Editorial Verbo, 2001.
- EARNEST, Les, “Machine Recognition of Cursive Writing,” in *Information Processing 62*, (Proc. IFIP Congress 1962, Munich), North-Holland, Amsterdam, 1963.
- FERREIRA, Aurélio Buarque de Holanda, *Novo Dicionário Aurélio da Língua Portuguesa*, 3.ª ed., Curitiba: Positivo, 2004.
- GONÇALVES, Maria Filomena, *As ideias ortográficas em Portugal. De Madureira Feijó a Gonçalves Viana (1734-1911)*, Lisboa: Fundação Calouste Gulbenkian/Fundação para a Ciência e a Tecnologia, 2003.
- GONÇALVES, Rebelo, *Tratado de Ortografia da Língua Portuguesa*, Coimbra: Atlântida – Livraria Editora, 1947.
- GONÇALVES, Rebelo, *Vocabulário da Língua Portuguesa*, Coimbra: Coimbra Editora, 1966.
- Grande Dicionário Língua Portuguesa*, Porto, Porto Editora, 2004.
- HOUAISS, António, VILLAR Mauro, *Dicionário Houaiss da Língua Portuguesa*, Instituto António Houaiss de Lexicografia e Banco de Dados da Língua Portuguesa S/C Ltda., Rio de Janeiro: Editora Objetiva, 2001.
- HOUAISS, António, VILLAR Mauro, *Dicionário Houaiss da Língua Portuguesa*, Lisboa: Círculo de Leitores, 2002.
- LEITE, Yonne e Dinah CALLOU, *Como falam os brasileiros*, Rio de Janeiro: Jorge Zahar Editor, 2002.
- MACHADO, José Pedro, *Grande Vocabulário da Língua Portuguesa*, Lisboa: Âncora Editora, 2001.
- MARQUILHAS, Rita, “O acento, o hífen e as consoantes mudas nas Ortografias antigas portuguesas”, in CASTRO, Ivo *et alii*, (orgs.), *A Demanda da Ortografia Portuguesa*, Lisboa: Edições João Sá da Costa, 1987, pp. 103-116.
- MARQUILHAS, Rita, *A Faculdade das Letras – Leitura e escrita em Portugal no século XVII*. Lisboa: IN-CM, 2000.
- MATEUS, Maria Helena Mira, “A mudança da língua no tempo e no espaço”, in MATEUS, Maria Helena e NASCIMENTO, Fernanda Bacelar do (orgs.), *A Língua Portuguesa em Mudança*, Lisboa: Caminho, 2005, pp. 13-30.
- MATEUS, Maria Helena Mira, “O problema da ortografia. Métodos de ensino” in *A Face Exposta da Língua Portuguesa*, Lisboa: Imprensa Nacional-Casa da Moeda, 2002, pp. 91-135.
- MEDEIROS, José Carlos Dinis, *Processamento Morfológico e Correção Ortográfica do Português*, tese de mestrado, Instituto Superior Técnico, Universidade Técnica de Lisboa, 1995. [também disponível em <http://www.linguateca.pt/documentos/Medeiros95.doc>].
- MINISTÉRIO DA EDUCAÇÃO NACIONAL, “Decreto n.º 35 228”, de 8 de dezembro de 1945, in *Priberam Informática* [em linha], 2008, http://www.priberam.pt/docs/AcOrtog45_73.pdf [descarregado em 2008-02-22].
- PETERSON, James L., *Computer Programs for Spelling Correction*, Berlin: Springer-Verlag (LNCS, nº 96), 1980.
- PRIBERAM INFORMÁTICA, *Priberam Informática – Língua Portuguesa On-line* [em linha], 2006, <http://www.priberam.pt/dlpo/> [consultado em 2009-02-22].
- ROTH, Phillip, *A Postmodern Nationalist – Truth, Orality, and Gender in the Work of Mia Couto*, Lewisburg: Bucknell University Press, 2004.
- TAVANI, Guisepppe, “Antecedentes históricos: a ortografia da língua portuguesa”, in CASTRO, Ivo *et alii*, (orgs.), *A Demanda da Ortografia Portuguesa*, Lisboa: Edições João Sá da Costa, 1987, pp. 201-203.