

O novo acordo ortográfico e os correctores automáticos

Helena Figueira, Afonso Mendes,
Pedro Mendes, Cláudia Pinto

Priberam
Alameda D. Afonso Henriques, 41 - 2.º Esq.
1000-123 Lisboa, Portugal
Tel.: +351 21 781 72 60
Fax: +351 21 781 72 79

{hgf, amm, prm, cp}@priberam.pt

Resumo

A análise que neste artigo se pretende fazer centra-se na aplicação prática das novas regras ortográficas decorrentes do Acordo Ortográfico de 1990 em programas informáticos, como o *FLiP* ou o *Novo Corretor Aurélio*, que incluem correcção ortográfica e sintáctica, e não tanto na polémica relativa aos aspectos positivos ou negativos do referido acordo, nem no seu impacto cultural, social ou económico. Pretende-se apresentar alguns dos casos problemáticos na implementação da nova ortografia na correcção ortográfica e examinar também o impacto desta implementação em variados recursos linguísticos ou em módulos que podem ser integrados noutros produtos e serviços.

Palavras-chave

Língua portuguesa, ortografia, acordo(s) ortográfico(s), correcção ortográfica.

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, passadas quase duas décadas, e após a ratificação pelo Brasil (2004), por Cabo Verde (2006) e por São Tomé e Príncipe (2006), teve a sua ratificação por Portugal publicada em 29 de Julho de 2008 no *Diário da República*. 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 reacções que então emergiram.

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 por desconhecimento³. No entanto, no âmbito do trabalho que a Priberam vem realizando na área do processamento da língua natural (PLN)⁴, faz todo o sentido questionar em que medida as novas regras ortográficas irão afectar o comportamento de correctores ortográficos para o português de norma europeia, nomeadamente o *FLiP*, e para o português de norma brasileira, nomeadamente o *Novo Corretor Aurélio*, 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.

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 pela Priberam está disponível em <http://www.priberam.pt/docs/AcOrtog91.pdf>.

² Ver Castro *et al.* (1987) e *A Questão do "Acordo Ortográfico"* (1988).

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://www1.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 pela Priberam, ver Amaral *et al.* (2003).

Como consideração preliminar, é pertinente estabelecer conceitos operatórios de *ortografia*, *erro ortográfico* e *norma* (ver secção 1.1). Em seguida, analisa-se a problemática da correcção ortográfica em que se inserem os correctores ortográficos do *FLiP* e do *Novo Corretor Aurélio*, cujo modo de funcionamento se descreve resumidamente (ver secção 2). Segue-se uma discussão da aplicação do Acordo Ortográfico de 1990 aos correctores ortográficos da Priberam (ver secção 3), referindo-se os casos mais problemáticos na implementação (ver secção 3.1). Por fim, examina-se o impacto desta implementação em recursos linguísticos ou em módulos que podem ser individualizados e integrados noutros produtos e serviços, a necessidade de criação de uma ferramenta de conversão de texto para a nova ortografia e ainda a possibilidade de criação de produtos destinados a outras variedades de português (ver secção 3.2).

1.1. Ortografia, erro ortográfico e norma

A linguagem escrita, contrariamente à linguagem oral, não é adquirida por exposição a textos escritos, antes resulta de aprendizagem⁵, 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.

Para a definição de um conceito de *ortografia* pode partir-se da consulta de um dicionário: "forma correcta de escrever as palavras; parte da gramática que ensina a escrever correctamente as palavras de uma língua"⁶. 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 etimologia⁷, a ortografia foi objecto de avanços e recuos, aproximações e distanciamentos na sua relação com a fonética e com a etimologia⁸ e é sobre este movimento que muitas vezes recaem as opiniões e os argumentos apresentados em debates sobre o assunto. Não nos pareceu profícuo fazer aqui uma cronologia da ortografia portuguesa⁹, mas é importante verificar que muitos dos argumentos usados ao longo da história da ortografia portuguesa¹⁰ se repetem actualmente, para um e outro lado da argumentação.

Por outro lado, "a ortografia, «ciência de escrever dereitamente» [...], 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"¹¹, sendo contudo "a bitola de uma certa eficiência e categoria sociais"¹². 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 peremptórias em termos de correcção¹³. 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 aspecto em que há uma regulamentação explícita através de textos legais. Presentemente, e no caso da norma europeia, até 29 de Agosto de 2008 estavam em vigor dois textos legais, o Decreto n.º 35 228, de 8 de Dezembro de 1945¹⁴ e a sua alteração pelo artigo único do Decreto-Lei n.º 32/73 de 6 de Fevereiro de 1973¹⁵. A norma onomástica está regulamentada pelo Decreto-Lei n.º 131/95, de 6 de Junho, alterado pelo Decreto-Lei n.º 36/97, de 31 de

⁵ Ver Delgado-Martins (1996).

⁶ PRIBERAM, "ortografia", in *Priberam - Língua Portuguesa On-line* [em linha], 2008, http://www.priberam.pt/dlpo/definir_resultados.aspx?pal=ortografia [consultado em 2008-07-28].

⁷ 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".

⁸ Sobre este assunto, ver Marquilhas (1987) e Mateus (2002).

⁹ Ver Cuesta e Luz (1971), Gonçalves (2003), Marquilhas (1987, 2000) e Tavani (1987).

¹⁰ Ver Gonçalves (2003).

¹¹ Mateus (2002: pp. 92-93).

¹² Mateus (2002: p. 93).

¹³ 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 dos poucos em que podem ser categóricas.

¹⁴ O texto oficial anotado pela Priberam está disponível em http://www.priberam.pt/docs/AcOrtog45_73.pdf.

¹⁵ Disponível em http://www.priberam.pt/docs/AcOrtog45_73.pdf.

Janeiro¹⁶. No caso da norma brasileira, vigorava o *Formulário Ortográfico*¹⁷, aprovado pela Academia Brasileira de Letras em Agosto de 1943, com alterações de 1971, pela Lei nº 5.765, de 18 de Dezembro¹⁸.

Acresce a estes dados a realidade portuguesa relativamente à política linguística, uma vez que não há, em Portugal, uma instituição com funções efectivas na definição de norma ou na normalização 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"¹⁹, que não se assemelha àquelas, que "nasceram só para o estudo das [respectivas] Línguas", como refere Pina Martins, Presidente da Academia à data do seu 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."²⁰ É neste âmbito da escrita, e sobretudo na ortografia, ligada a fenómenos socioculturais, que mais facilmente se chega à determinação do erro, enquanto desvio relativamente a uma norma²¹ e ao que é considerado correcto, pois a norma ortográfica, dada a sua estabilidade, deve estar menos sujeita à variabilidade e à subjectividade, e não implica juízos linguísticos, mas apenas 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 flutuante, sendo um termo genérico que pode incluir os critérios que regulam aquilo que é considerado por um grupo maioritário como linguisticamente correcto ou uma variante linguística mais prestigiada sociocultural ou politicamente. Consultando novamente um dicionário geral de língua, pode encontrar-se uma definição vaga de *norma* como "aquilo que é de uso geral, relativamente à língua"²². 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, ainda que a norma ortográfica seja prescritiva e convencionalmente estabelecida. E, se para a escrita é relativamente mais fácil encontrar referências relativamente à norma, para a pronúncia estas referências são escassas e resumem-se aos (raros) dicionários com transcrição fonética²³ e às indicações de ortoépia em dicionários²⁴ e vocabulários²⁵.

Para um corrector ortográfico, no entanto, a sinalização do erro ortográfico tem de corresponder a um conceito operatório que permita um equilíbrio entre critérios linguísticos, o respeito pela norma ortográfica, as exigências dos utilizadores e as limitações do PLN. Este equilíbrio, já difícil *a priori*, torna-se mais difícil perante alterações ortográficas significativas, sobretudo se o texto que as regula for ambíguo ou incongruente, principalmente 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 actividades como a realização de ditados, por exemplo. Inerente a ela está a noção de correcção (e a de correctivo, em épocas mais distantes). A ideia da detecção de erros ortográficos e da respectiva correcção é introduzida no primeiro contacto com a alfabetização, mantendo-se mais ou menos presente ao longo da vida.

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

¹⁶ Sobre regulamentação dos nomes próprios, ver Castro (2003).

¹⁷ Disponível em <http://www.academia.org.br/abl/cgi/cgilua.exe/sys/start.htm?sid=20>.

¹⁸ Disponível em <http://www.academia.org.br/abl/cgi/cgilua.exe/sys/start.htm?infol=2453&sid=19>.

¹⁹ *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa* (2001: ix).

²⁰ Ver Leite e Callou (2002: p. 57).

²¹ Sobre variação e norma, ver Mateus (2005) e Castro (2003).

²² PRIBERAM, "norma", in *Priberam - Língua Portuguesa On-line* [em linha], 2008, http://www.priberam.pt/dlpo/definir_resultados.aspx?pal=ortografia [consultado em 2008-07-28].

²³ Para o português europeu, em dicionários monolíngues, ver *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa* (2001), *Grande Dicionário Língua Portuguesa* (2004) e *Dicionário Gramatical de Verbos Portugueses* (2007); em dicionários bilingues, ver sobretudo *Grande Dicionário Português/Francês* (1998) e *Dicionário Verbo Português-Ingês* (2000).

²⁴ Para o português europeu, ver Houaiss (2002).

²⁵ Para o português europeu, ver Gonçalves (1966) e Machado (2001).

Em termos de PLN, foi na década de 70 do séc. XX que se assistiu ao surgimento dos primeiros auxiliares informáticos de correcção ortográfica²⁶. Na sua fase inicial, os programas eram essencialmente verificadores ortográficos, pois detectavam erros mas não apresentavam sugestões de correcção²⁷, só mais tarde evoluindo para correctores ortográficos, que geralmente fornecem sugestões para os erros detectados. Em menos de uma década, os programas de correcção 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 correctores ortográficos, cada vez mais robustos e sofisticados, contêm já verificação automática e dispõem de opções de correcção automática bastante seguras, podendo também ser encontrados em *web browsers*, servidores de correio electrónico *online* ou sítios de alojamento de blogs.

Na correcção de um texto electrónico há a considerar diversos níveis: ortográfico, sintáctico e estilístico. Sumariamente, um programa de correcção ortográfica permite a detecção de diferentes tipos de erros ortográficos²⁸, a saber, linguísticos ou cognitivos (ex.: **análize / análise*), tipográficos ou de digitação (ex.: **casamneto / casamento*) e de transmissão (ex.: confusão de *ç* com *c* ou de *rn* com *m*, num texto digitalizado).

2.1. FLiP e Novo Corretor Aurélio

Comercializado desde 1995, o *FLiP – Ferramentas para a Língua Portuguesa* – tem vindo a afirmar-se nos mercados português e internacional como uma ajuda indispensável à escrita em computador ou à revisão de texto electrónico de diversas origens, como o texto resultante de digitalização por reconhecimento óptico de caracteres (ROC).

As ferramentas *FLiP* têm como antecedente o corrector ortográfico *Lince*²⁹, surgido em 1994, e que constituiu o primeiro corrector ortográfico comercializado para o português europeu. A primeira versão do *FLiP*, de 1995, continha um corrector ortográfico, um dicionário de sinónimos e um hifenizador. A segunda versão, o *FLiP 97*, passou a incluir o primeiro corrector sintáctico 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. 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*³⁰, estão incluídos, para o português europeu, um corrector ortográfico, um corrector sintáctico e estilístico, oito dicionários temáticos, um hifenizador, um dicionário de sinónimos, um conjugador de verbos, quatro auxiliares de tradução e um editor de texto, bem como correctores ortográficos para alemão, espanhol, francês, inglês e italiano.

Actualmente, o corrector ortográfico da Priberam está integrado nos blogs do Sapo³¹.

Além da correcção ortográfica e sintáctica, 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 LegiX³², na área de motores de pesquisa (através da participação nos projectos TRUST³³ e MCAST³⁴), na área da recolha de informação em respostas a perguntas (através da participação no CLEF³⁵), na área da extracção de informação (sistema de pesquisa nos *sites* da TSF³⁶ e do JN³⁷) ou na área do processamento de fala

²⁶ Ver Peterson (1980).

²⁷ O primeiro verificador ortográfico usava um dicionário com as 10 000 palavras mais comuns em língua inglesa e remonta a 1961 (cf. <http://www.ibiblio.org/pub/historic-linux/distributions/slackware/2.1/usr/doc/ispell/Contributors>); ver Earnest (1963).

²⁸ Para uma tipologia de erros ortográficos e uma breve análise das estratégias envolvidas na sua correcção em PLN, veja-se Medeiros (1995: pp. 60-107).

²⁹ Ver Andrade (1992).

³⁰ Está disponível em <http://www.flip.pt/online> uma versão de demonstração que apresenta algumas das funcionalidades do *FLiP*.

³¹ Ver <http://blogs-beta.sapo.pt>.

³² LegiX é um conjunto de ferramentas de informação jurídica da Priberam; ver www.legix.pt para mais informação.

³³ 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.

³⁴ 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/>).

³⁵ CLEF – *Cross Language Evaluation Forum* – 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 na campanha de *Question Answering* do último CLEF, veja-se Amaral *et al.* (2007).

³⁶ Ver <http://www.tsf.pt/>.

³⁷ Ver <http://www.jn.pt/>.

(através da participação no projecto Tecnovoz³⁸). O desempenho nestas áreas tem assim sido testado e avaliado exaustivamente, 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 corrector 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*³⁹, um dos maiores jornais do Brasil.

O primeiro produto específico apenas para o português do Brasil foi o *FLiP:mac Brasil*⁴⁰ (2003), seguindo-se o *FLiP Brasil* (2004) e o *Novo Corrector Aurélio* (2007). Presentemente, as ferramentas linguísticas da Priberam para o português do Brasil incluem um corrector ortográfico, um corrector sintáctico e estilístico, oito dicionários temáticos, um hifenizador, um dicionário de sinónimos e antónimos e um conjugador verbal. Ainda que os recursos linguísticos sejam diferentes, estas ferramentas baseiam-se no *software* que a Priberam vem desenvolvendo há mais de 15 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 correctores sintácticos da variedade portuguesa e da variedade brasileira. Tal acontece porque, salvaguardadas as devidas diferenças, a generalidade das regras sintácticas 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 ortográficos com implicações sintácticas (por exemplo, o uso problemático da crase).

2.1.1. Corrector ortográfico: modo de funcionamento

O corrector ortográfico tem como objectivo detectar erros de ortografia e apresentar sugestões para a sua correcção. O funcionamento do corrector ortográfico do *FLiP* baseia-se na comparação das palavras utilizadas num documento com uma lista de palavras (designada por *dicionário*) reconhecidas pelo módulo de correcção ortográfica⁴¹. O conjunto das palavras conhecidas pelo corrector ortográfico é constituído pela reunião da lista de palavras geral, das listas de palavras dos dicionários temáticos⁴² seleccionados e das palavras adicionadas aos dicionários de utilizador⁴³ activos. Se uma determinada palavra não for reconhecida, esta é assinalada como errada e são apresentadas sugestões para a sua correcção.

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, geralmente 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 incorrecta que se encontra no documento. As sugestões são apresentadas por ordem decrescente de probabilidade de se tratar da palavra correcta. Na ordenação das sugestões são considerados factores como a semelhança fonética, a semelhança gráfica (importante em textos digitalizados com ROC), adjacências no teclado e frequência das palavras.

O corrector ortográfico não se limita a assinalar erros ortográficos, mas veicula opções no reconhecimento de palavras. Assim, o corrector 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/demonstração, gostar/gostar, treinador/treinador*) ou palavras tabuísticas ou ofensivas (ex.: *cagar, merda, puta*)⁴⁴.

³⁸ O projecto Tecnovoz pretende desenvolver novas competências nas tecnologias da fala (<http://www.tecnovoz.pt/>).

³⁹ Ver <http://www.estadao.com.br/>.

⁴⁰ Ver <http://www.flip.pt/tabid/252/Default.aspx>.

⁴¹ 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.

⁴² Os dicionários temáticos são listas de palavras agrupadas por domínios técnicos ou científicos; ver <http://www.flip.pt/Modulos/DicionariosTematicos/tabid/277/Default.aspx>.

⁴³ Um dicionário de utilizador permite adicionar palavras não reconhecidas à lista de palavras reconhecidas pelo corrector ortográfico do FLiP.

⁴⁴ 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

O corrector ortográfico contém ainda algoritmos de correcção 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 corrector ortográfico (por exemplo, a palavra *françês* é imediatamente corrigida para *francês*).

Algumas verificações e sugestões ortográficas são detectadas pelo corrector sintáctico e não pelo módulo de correcção 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*⁴⁵ 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 correcção ortográfica do texto em análise.

3. Aplicação do Acordo Ortográfico de 1990 em correctores ortográficos

A ortografia de um adulto escolarizado é, ou deveria ser, em geral, estável e quase automática, com pequenas áreas problemáticas, onde surgem inevitavelmente 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 sua imagem gráfica, memorizada durante a aprendizagem e repetida enquanto escrevente (as formas antigas *absorção*, *condemnar*, *contracto*, *fructo*, *orthographia*, *pharmácia*, *psalmo*, *sancto*, *sciência* ou *victória* são casos em que essa memória se perdeu, sem que os utilizadores da língua actual sintam qualquer desconforto gráfico porque não têm delas memória; alterações como a supressão de acentuação gráfica nos advérbios terminados em *-mente*, como em *praticamente*, estão porventura mais próximas de muitos utilizadores, até pela facilidade de encontrar edições ainda com esta grafia).

Num momento de alterações ortográficas significativas⁴⁶ (a percentagem de palavras a alterar é pequena, mas a frequência de algumas palavras ou famílias de palavras é muito alta), sobretudo se o texto estabelecido em 1990 não for alterado, e dado o volume de textos actualmente produzido em formato electrónico, a confiança num corrector ortográfico é essencial. Por outro lado, e atendendo ao facto de que o corrector verifica a ortografia e oferece sugestões de correcção, constitui também um instrumento privilegiado para criar novos mecanismos de tomada de conhecimento, aprendizagem e memorização das alterações ortográficas convencionadas⁴⁷.

Aparentemente, como explicitado na "Nota Explicativa do Acordo Ortográfico da Língua Portuguesa", aquando da exposição das razões ponderadas para a manutenção dos acentos gráficos nas proparoxítonas e paroxítonas, recuando assim no estipulado pelo texto de 1986, o texto legal tem preocupações específicas com o PLN (sublinhado nosso):

"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 sintáctico, 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*, adjectivo relativo à vista e ao ouvido, respectivamente). 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 3ª 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 "Para para se oferecer como testemunha, embora já vá atrasado para um encontro"⁴⁸. Acresce que em muitos processadores de texto

intenção de redigir «cavalgo», 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).

⁴⁵ 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 desta versão.

⁴⁶ Para a visualização das alterações do texto de 1990 relativamente ao de 1945 e respectivas observações, ver <http://www.flip.pt/AcordoOrtografico/Introducao/tabid/514/Default.aspx>.

⁴⁷ É sintomático que um guia para o uso da nova grafia afirme que "alguma prática da nova ortografia e o recurso, hoje de utilização tão comum, a um corretor ortográfico atualizado resolverão rapidamente as dificuldades" (CASTELEIRO e CORREIA, 2007: p.5).

⁴⁸ Exemplos extraídos do *corpus* CETEMPúblico (disponível em <http://www.linguateca.pt/>), que sofreram adaptação da forma *pára*.

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"⁴⁹, 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* (sublinhado nosso):

"3.º Prescinde-se de acento gráfico para distinguir palavras oxítonas homógrafas, mas heterofónicas/heterofônicas, do tipo de *cor* (*ô*), substantivo, e *cor* (*ó*), elemento da locução *de cor*; *colher* (*ê*), verbo, e *colher* (*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* (e vice-versa), 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, ambíguo ou incoerente 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 maior, de que são exemplo o *Tratado de Ortografia*⁵⁰ e o *Vocabulário*⁵¹ de Rebelo Gonçalves ou o *Vocabulário Ortográfico* da Academia Brasileira de Letras⁵².

É forçoso que exista a curto prazo 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 Castro⁵³, no seguimento da sua defesa de uma "versão fraca de unificação"⁵⁴, 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 pronúncia" (sublinhado nosso). É neste âmbito que se pretendem inserir o *FLiP* e o *Novo Corretor Aurélio*, uma vez que se trata de produtos comerciais que constituem auxiliares de escrita e de revisão de texto.

3.1. Implementação das alterações

A implementação das alterações ortográficas preconizadas pelo Acordo de 1990 é uma empresa complexa e morosa, implicando modificações nos módulos de correcção, mas também opções de fundo que podem reflectir vários cenários e implicar diferentes soluções técnicas. É necessário avaliar como falaciosas ou veleiosas quaisquer considerações que procurem fazer crer na rapidez e na facilidade desta tarefa.

A Priberam tem acompanhado atentamente todo o processo relativo ao Acordo Ortográfico, no sentido de não tomar decisões precipitadas e de produzir ferramentas que correspondam qualitativamente às expectativas dos utilizadores, não só em Portugal, mas também no Brasil e em outros países da CPLP, nomeadamente Angola e Moçambique. Pelo facto de a Priberam ter adquirido uma grande responsabilidade devido aos produtos que desenvolve, às marcas a que se tem associado e ao número de utilizadores que recorrem aos seus produtos e serviços, é exigido aos seus correctores um alto grau de qualidade e a resolução pronta das dúvidas ortográficas (e não a criação de mais áreas de dúvida ou hesitação).

Com as devidas distâncias decorrentes do tipo de registo em causa, a prova de que não se trata de uma tarefa rudimentar, ou sequer fácil, é a existência de duas obras lexicográficas⁵⁵ cuja ortografia é publicitada como estando já de acordo com as novas disposições ortográficas, mas que,

⁴⁹ Exemplo extraído do *corpus* CETEMPúblico (disponível em <http://www.linguateca.pt/>).

⁵⁰ Ver Gonçalves (1947).

⁵¹ Ver Gonçalves (1966).

⁵² Ver Academia Brasileira de Letras (1981).

⁵³ Sobre este assunto, ver 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.

⁵⁴ Ver Castro *et al.* (1987: p. 8).

⁵⁵ *Novo Dicionário da Língua Portuguesa Conforme Acordo Ortográfico* (Texto Editores, 2007) e *Novo Dicionário Editora da Língua Portuguesa 2009 – Acordo Ortográfico* (Porto Editora, 2008).

independentemente dos modelos lexicográficos diferentes, tomam decisões diferentes na grafia de um número considerável de palavras⁵⁶.

As principais alterações aos correctores da Priberam foram efectuadas ao nível dos léxicos (bases de dados relacionais onde se codifica informação lexical variada⁵⁷) relativos à variedade do português europeu e do português brasileiro.

De entre as hipóteses para implementação das alterações ortográficas, chegou a considerar-se a criação de um léxico único que reunisse todas as variantes ortográficas possíveis de uma mesma palavra, atendendo, por exemplo, às facultatividades e acentuações possíveis (ex.: *tectónico/tectônico/tetónico/tetônico*). Esta opção parece ser claramente desvantajosa, uma vez que multiplicaria as variantes e não teria em conta as práticas ortográficas nacionais de Portugal e do Brasil, não sendo esclarecedora nem pedagógica quanto às variantes preferenciais em cada espaço nacional. Para restringir as desvantagens desta hipótese, poderia ser implementado no *software* um conjunto de opções de selecção de formas exclusivas ou preferenciais pertencentes a cada uma das variedades nacionais, à semelhança da selecção feita com os dicionários temáticos. Esta alternativa apresentava, 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*), sintácticas (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 correcção de erros de concordância (ex.: *a/o disquete*), na desambiguação morfossintáctica (ex.: categoria gramatical da palavra *cara*) ou na desambiguação semântica (ex.: *bala=projéctil*; *bala=rebuçado*). Paralelamente, o léxico correspondente a certas áreas técnicas e científicas é específico e em muitos casos exclusivo de cada uma das variedades do português.

O cenário que se afigurou então como mais produtivo e mais seguro, nomeadamente para satisfação das exigências dos utilizadores e para evitar uma desnecessária variação interna dentro de cada espaço nacional, foi 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. Mantêm-se, assim, as bases de dados lexicais utilizadas actualmente no *FLiP* e no *Novo Corrector Aurélio*, nas quais se procedeu às devidas alterações preconizadas pelo Acordo Ortográfico. A necessidade de manutenção destes dois léxicos justifica-se pelo facto de, no entender da Priberam, 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 fazem 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 das diferenças ortográficas que provêm da tradição lexicográfica dos dois países (ex.: *alforge/alforje*, *missanga/miçanga*) ou ainda o da flexão preferencial, numa e noutra norma, de verbos terminados em *-guar* e *-quar* ou o da discordância nas duplas consoantes *-mm-* ou *-nn-* (ex.: *connosco/conosco*, *comummente/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 (ver *Base II*), 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".

O módulo sintáctico, que, como se viu anteriormente, possibilita também correcção 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 outro argumento válido para a manutenção de dois léxicos individualizados.

A seguir apresentam-se quatro dos vários casos problemáticos em que a Priberam considera indispensável explicitar os critérios para as suas opções.

⁵⁶ A título de exemplo, veja-se o registo de *côa/coa*, *géiser/geiser* ou *forma/fôrma*, a par de outros que veremos nas subsecções seguintes.

⁵⁷ Ver Amaral *et al.* (2003).

3.1.1. Casos problemáticos

3.1.1.1 Nomes próprios

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 licencia a admissão de formas de origem estrangeira (ex.: *Zimbabwe*) e 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.: *Bahia/Baía, Irão/Irã, Madrid/Madri*), o que torna premente a produção de documentos de referência a este nível e nitidamente vantajosa a adopção de léxicos independentes.

3.1.1.2 Possibilidade de dupla acentuação

Deve referir-se que um dos principais argumentos utilizados na defesa do texto legal em análise é o da aproximação da escrita à pronúncia. De facto, por diversas vezes ao longo do texto legal se refere as "pronúncias cultas da língua"⁵⁸ como critério norteador das prescrições ortográficas, e no ponto 3 da "Nota Explicativa", essa escolha é mesmo explicitada (sublinhado nosso):

"Pode dizer-se ainda que, no que respeita às alterações de conteúdo, de entre os princípios em que assenta a ortografia portuguesa se privilegiou o critério fonético (ou da pronúncia) com um certo detrimento para o critério etimológico.

[...]

É ainda o critério da pronúncia que conduz à manutenção da dupla acentuação gráfica do tipo de *económico* e *econômico*, *efémero* e *efêmero*, *género* e *gênero*, *génio* e *gênio*, ou de *bónus* e *bônus*, *sêmen* e *sêmen*, *tênis* e *tênis*, ou ainda de *bebê* e *bebê*, ou *metro* e *metrô*, etc."

Este preceito, de simplicidade aparente, esbarra no facto de a pronúncia culta ser de aferição difícil, dada a falta de instrumentos prescritivos ou normalizadores.

Se tomarmos a subortografia como conceito operativo, deixa de ser problemática a manutenção das sequências *-éma-*, *-éme-*, *-émi-*, *-émo-*, *-ému-*, *-éna-*, *-éne-*, *-éni-*, *-éno-*, *-énu-*, *-óma-*, *-óme-*, *-ómi-*, *-ómo-*, *-ómu-*, *-óna-*, *-óne-*, *-óni-*, *-óno-*, *-ónu-* para palavras da variedade europeia do português (ex.: *fenómeno*, *sêmen*, *ténue*) e das sequências *-êma-*, *-ême-*, *-êmi-*, *-êmo-*, *-êmu-*, *-êna-*, *-êne-*, *-êni-*, *-êno-*, *-ênu-*, *-ôma-*, *-ôme-*, *-ômi-*, *-ômo-*, *-ômu-*, *-ôna-*, *-ône-*, *-óni-*, *-ôno-*, *-ônu-* para palavras da variedade brasileira do português (ex.: *fenômeno*, *sêmen*, *tênue*). Entende-se assim a dupla acentuação prevista para estas sequências na Base IX, 2.º *Obs.* e 3.º *Obs.* e na Base XI, 3.º como uma possibilidade de incluir variantes das variedades portuguesa e brasileira na ortografia oficial, mas considera-se que um corrector ortográfico para o mercado do português europeu deverá incluir apenas as variantes portuguesas, tal como um corrector ortográfico para o mercado do português brasileiro deverá incluir apenas as variantes brasileiras.

Esta opção justifica-se pela manutenção das práticas ortográficas das duas normas e evita excepções que possam confundir os utilizadores e gerar ainda mais dúvidas ortográficas. No ponto 5.2.1. da "Nota Explicativa" prevê-se que, neste aspecto, "existe uma ou outra excepção a esta regra, como, por exemplo, *cômore* e *sêmola*, mas estes casos não são significativos." Apesar disso, as duas obras lexicográficas recentemente publicadas que publicitam aplicar já as regras ortográficas do Acordo Ortográfico de 1990 tomam opções diferentes neste ponto (ver Tabela 1), o que permite verificar a dificuldade de sistematização das excepções apresentadas, ainda que se trate de um número pouco significativo.

Texto Editores ⁵⁹	Porto Editora ⁶⁰
cômore/cômore	cômore
fêmea	fêmea/fêmea
estômago	
sêmola	

Tabela 1: Exemplos de dicionários para a possibilidade de dupla acentuação.

⁵⁸ Ver, por exemplo, a Base IV, 1.º, alíneas a) e b) ou Base IX, 2.º, alíneas a) e b).

⁵⁹ *Novo Dicionário da Língua Portuguesa Conforme Acordo Ortográfico*, Lisboa: Texto Editores, 2007.

⁶⁰ *Novo Dicionário Editora da Língua Portuguesa 2009 – Acordo Ortográfico*, Porto: Porto Editora, 2008.

Por outro lado, nas opções acima apresentadas como concordantes nos dois dicionários, uma respeitando a excepcionalidade prevista na nota explicativa (*sêmola*), mas não as pronúncias de muitos falantes⁶¹, outra (*estômago*) não prevista como exceção, o que tornaria lícita a grafia *estômago*, que corresponderia a uma pronúncia efectiva em alguns dialectos do português europeu (por exemplo, no dialecto madeirense⁶²).

3.1.1.3 Grafia de grupos consonânticos

O caso apontado por muitos críticos do Acordo de 1990 como mais problemático é o da facultatividade das chamadas "consoantes mudas" (ver *Base IV*). Também aqui o critério orientador apresentado pelo legislador é o das chamadas "pronúncias cultas da língua", como explanado no ponto 3 da "Nota Explicativa":

"É o critério da pronúncia que determina, aliás, a supressão gráfica das consoantes mudas ou não articuladas, que se têm conservado na ortografia lusitana essencialmente por razões de ordem etimológica.

É também o critério da pronúncia que nos leva a manter um certo número de grafias duplas do tipo de *caráter* e *carácter*, *facto* e *fato*, *sumptuoso* e *suntuoso*, etc."

Ainda que a existência de duplas ortografias não seja inédita, sobretudo no português do Brasil (ex.: *acessível/accessível*, *aspecto/aspeto*, *sanguinário/sangüinário*⁶³), mas também no português europeu (ex.: *abcissa/abscissa*, *cisto/quisto*, *cobarde/covarde*, *loíça/louça*, *rotura/ruptura*), algumas interpretações deste novo texto legal poderão aumentar exponencialmente o número de casos de múltiplas grafias (conforme a variação em cada país, em cada dialecto e, se a essa última análise se chegasse, em cada falante) e dificultar as decisões a tomar pela lexicografia e também pelos ensinantes da língua portuguesa, uma vez que este princípio introduz maior variabilidade e subjectividade no ensino da ortografia e da noção de erro ortográfico.

A lexicografia pode, pela sua natureza descritiva, registar quaisquer variantes, como aconselha o texto do Acordo⁶⁴:

"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 pronúncia."

Esta é uma indicação explícita de que as diferenças de pronúncia equivalerão diferenças de grafia com determinados alcances geográficos, isto é, indica que as práticas ortográficas nacionais serão necessariamente diferentes, o que reforça a argumentação para a escolha de subortografias nacionais.

Também neste ponto, e como referido anteriormente, as duas obras lexicográficas já publicadas segundo a nova ortografia divergem amiúde, como exemplificado na Tabela 2, o que denuncia a dificuldade de sistematização ortográfica neste e noutros tópicos⁶⁵.

Texto Editores	Porto Editora
anfractuoso/anfratuoso	anfractuoso
excepção/exceção	exceção
manufatura	manufactura/manufatura
perfeccionista	perfeccionista/perfeccionista
seção/secção	
tacto/tato	tato
tectónico/tectônico/tetónico/tetônico	tectónico/tectônico
tumefacto	tumefacto/tumefato

Tabela 2: Exemplos para a grafia de consoantes, retirados de dicionários.

Em outros casos, algumas opções convergentes nos dois dicionários (ver, por exemplo *seção/secção* na Tabela 2) são discutíveis, uma vez que se trata de dicionários destinados ao mercado do português

⁶¹ Esta é uma constatação empírica, não fundamentada em dados estatísticos.

⁶² Esta é uma constatação empírica, não fundamentada em dados estatísticos.

⁶³ A título de exemplo, compare-se também opções discordantes no *Novo Dicionário Aurélio* (DA) e na versão brasileira do *Dicionário Houaiss da Língua Portuguesa* (DH), nos verbetes *céptico* (remissivo no DA) /*cético* (remissivo no DH), *manufactura* (não registado no DH e registado no DA, a par de *manufatura*).

⁶⁴ Ver "Nota Explicativa", ponto "4.4 - Casos de dupla grafia [base IV, 1.º, c) e d), e 2.º]".

⁶⁵ Este facto não é inédito na lexicografia portuguesa, mesmo com o texto de legal de 1945, mas estas divergências ocorrem justamente em aspectos que o texto legal pretende regularizar.

européu e todos os instrumentos lexicográficos⁶⁶ desta variedade preconizam a prolação da consoante *c* nesta palavra.

O corrector ortográfico tem, como a sua designação indicia, uma natureza prescritiva, e não é de todo desejável que o corrector crie mais dúvidas ortográficas, mas que as resolva de forma inequívoca e assertiva. Por este motivo, as opções tomadas para a grafia destas consoantes não visam reflectir as opções individuais de um falante ou de um grupo de falantes, mas aquelas que foram consideradas como pertencentes à chamada "norma culta". Entendemos como referências da norma culta a pronúncia registada em dicionários com transcrição fonética ou as indicações de ortoépia em dicionários e vocabulários. As opções resultam, então, sobretudo da consulta de obras de referência e, na ausência de opções nestas obras, de analogias feitas pelo linguista, de acordo com o seu conhecimento dessas mesmas obras (de contrário, o juízo de cada falante poderá proporcionar uma plêiade de grafias que pode subverter o próprio conceito de ortografia). No caso das chamadas "consoantes mudas", as decisões tomaram em consideração a actualidade das obras de referência relativamente à pronúncia e foi estabelecida uma ordem de importância, relativamente à sua pertinência. Um exemplo paradigmático da necessidade desta hierarquização é a ortoépia essencialmente conservadora e etimológica do *Vocabulário* de Rebelo Gonçalves⁶⁷, por contraponto ao de José Pedro Machado⁶⁸ (ver Tabela 3).

	R.Gonçalves	J.P.Machado
acção	k	àç
acepção	p	èç
eléctrico	k	étri
óptimo	p	ót

Tabela 3: Exemplos de indicações ortoépicas em R. Gonçalves e J. P. Machado.

As decisões são mais complicadas relativamente ao léxico técnico e científico (que corresponde aos dicionários temáticos do corrector), 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.

Em qualquer caso, esta opcionalidade na grafia de consoantes, conforme a sua articulação ou não na pronúncia, coloca problemas técnicos adicionais, pela necessidade de encontrar uma maior personalização do corrector ortográfico para que a coerência ortográfica seja maior. Assim, a possibilidade de haver opções pessoais ou institucionais para a grafia de determinadas palavras (por exemplo, *jacto/jato*) ou grupos de palavras (por exemplo, *sector/setor*, *sectorial/setorial*, *sectorizar/setorizar*) é um desafio suplementar.

3.1.1.4 Uso do hífen

Uma vez que o hífen é de utilização problemática em português⁶⁹, é importante referir as alterações que a *Base XVI* introduz relativamente ao texto de 1945, que vão aparentemente 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 advinha 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 elementos análogos. Por exemplo, o elemento *higio-*, de *higiossanitário*, à falta de indicações específicas, seguia o paradigma de outros como *hidro*⁷⁰. Segundo o texto de 1990, estes casos serão tratados segundo regras contextuais e já não seguindo pequenas listas de prefixos, que não poderão prever neologismos. Do ponto de vista do tratamento computacional do léxico, este tipo de regras contextuais é claramente vantajoso.

As regras contextuais esbarram, contudo, em excepções e lacunas que tornam problemática, em alguns casos, a sua aplicação. À semelhança do que foi feito para os outros tópicos apontados anteriormente, é sobretudo neste que se apresenta mais profícua a comparação (ver Tabela 4) entre as opções dos dois

⁶⁶ Ver Gonçalves (1966), *Grande Dicionário Português/Francês* (1998), *Dicionário Verbo Português-Inglês* (2000), Machado (2001), *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa* (2001) e *Grande Dicionário Língua Portuguesa* (2004).

⁶⁷ Ver Gonçalves (1966).

⁶⁸ Ver Machado (2001).

⁶⁹ Nos serviços de consultoria linguística disponibilizados pela Priberam (<http://www.flip.pt/duvidas>), as questões sobre o uso do hífen têm uma alta frequência.

⁷⁰ Ver Gonçalves (1947: pp. 251-252).

dicionários publicados com a nova ortografia, para verificarmos que não é linear a aplicação do disposto na *Base XV* e na *Base XVI*.

Texto Editores ⁷¹	Porto Editora
abrogação	ab-rogação
céu da boca	céu-da-boca
euroasiático*	euro-asiático
indoeuropeu	indo-europeu
maria vai com as outras	maria-vai-com-as-outras
para quedas	paraquedas
re-entrar	reentrar
sobre-excitar	sobreexcitar
sotopor	
subreptício	sub-reptício
tão somente	tão-somente

Tabela 4: Exemplos para o uso do hífen, retirados de dicionários.

Como exemplo, podemos verificar regras contextuais em relação aos "prefixos" ou nos "falsos prefixos" terminados na vogal *o* (os exemplos apontados no texto legal são *aero-*, *agro-*, *auto-*, *bio-*, *co-*, *eletro-*, *geo-*, *hidro-*, *macro-*, *pró-*, *proto-*, *pseudo-* e *retro-*) nas alíneas a) e b) do ponto 1.º da *Base XV*. Mas logo de imediato surge uma exceção para o elemento *co-* na Observação à alínea b) do referido ponto, e na alínea e) para os prefixos *soto-*⁷² e *vizo-* e na alínea f) para o prefixo *pró-*. Esta formulação aproxima-se então, neste aspecto, da de 1945, pois invalida o critério contextual sem apresentar outro argumento no ponto 6.3 da "Nota Explicativa" que não "as restantes regras são formuladas em termos de unidades lexicais, como acontece com oito delas (*ex-*, *sota-* e *soto-*, *vice-* e *vizo-*; *pós-*, *pré-* e *pró-*)."

Neste campo, e retomando os elementos terminados na vogal *o*, a principal dificuldade é a de determinar a diferença clara entre as palavras que devem seguir o disposto na *Base XV* e o disposto na *Base XVI*. Veja-se na Tabela 4, a título de exemplo, a diferença de opções para o registo de *euroasiático/euro-asiático* e *indoeuropeu/indo-europeu*. Neste dois casos, como em outros análogos, poderá argumentar-se que se trata de "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" (*Base XV*) para justificar o uso do hífen ou poderá argumentar-se que não se justifica o uso do hífen uma vez que se trata de "formações com prefixos (como, por exemplo: *ante-*, *anti-*, *circum-*, *co-*, *contra-*, *entre-*, *extra-*, *hiper-*, *infra-*, *intra-*, *pós-*, *pré-*, *pró-*, *sobre-*, *sub-*, *super-*, *supra-*, *ultra-*, etc.) e em formações por recomposição, isto é, com elementos não autónomos ou falsos prefixos, de origem grega e latina (tais como: *aero-*, *agro-*, *arqui-*, *auto-*, *bio-*, *eletro-*, *geo-*, *hidro-*, *inter-*, *macro-*, *maxi-*, *micro-*, *mini-*, *multi-*, *neo-*, *pan-*, *pluri-*, *proto-*, *pseudo-*, *retro-*, *semi-*, *tele-*, etc.)" (*Base XVI*).

Analisando ainda outros casos problemáticos no uso do hífen, verificamos que o texto legal prevê o hífen nas formações em que o prefixo termina em *-r*, quando combinado com elementos iniciados por *r* (como em *super-requintado*), mas não prevê os casos em que o primeiro elemento termina em consoante oclusiva (como *ab-* ou *sub-*) e segundo elemento se inicia por *r* (como em *ab-reacção* ou *sub-reptício*)⁷³. É isto que explica a divergência verificada na Tabela 4 na grafia de *abrogação/ab-rogação* e *subreptício/sub-reptício*. No entanto, para que seja mantida a pronúncia [R] do segundo elemento, terá de manter-se o hífen, pois os casos de *ab-r*, *ad-r*, *ob-r*, *sob-r*, *sub-r* e afins são os únicos casos na língua em que há os grupos *br* ou *dr* (que se podiam juntar a *cr*, *fr*, *gr*, *pr*, *tr* e *vr*) sem que a consoante <r> seja uma vibrante alveolar. De outro modo, o *r* ligar-se-á à consoante que o precede e passará de vibrante velar (ex.: *ab*[R], *sub*[R]) a vibrante alveolar (ex.: *ab*[r], *sub*[r]). Não se pode, por isso, alterar a fonética por causa da ortografia, nem alterar a grafia, criando uma exceção ortográfica, só porque o legislador/relator ou afim escamoteou este caso. O argumento de que a opção de manter o hífen nestes casos segue o espírito do acordo pode reforçar-se se olharmos, por exemplo, para os casos de *circum-* e *pan-*, onde não se criam exceções à estrutura silábica, nem à pronúncia (cf. *circum-anal* e não *circumanal*; *pan-óptico* e não *panóptico*). Este é um exemplo das inferências que têm de ser feitas na tomada de decisões, devido à qualidade do texto legal.

⁷¹ As palavras marcadas com asterisco constam do *Novo Grande Dicionário da Língua Portuguesa Conforme Acordo Ortográfico*, em 2 volumes, da mesma editora e com a mesma data de publicação, e não do *Novo Dicionário da Língua Portuguesa Conforme Acordo Ortográfico*.

⁷² Repare-se na Tabela 4 que nenhum dos dicionários respeita o disposto relativamente ao prefixo *soto-* no registo de *sotopor*.

⁷³ Ao contrário do texto de 1945, que o o previa na *Base XXIX*, 6.º e 7.º.

Uma outra área especialmente controversa é a da distinção entre palavras hifenizadas e locuções (ver, na Tabela 4, as grafias de *céu-da-boca/céu da boca*, *maria-vai-com-as-outras/maria vai com as outras* e *tão-somente/tão somente*). As alterações a realizar nos correctores da Priberam obrigam a alterações no módulo de correcção sintáctica, uma vez que é o corrector sintáctico que detecta erros contextuais. Estas alterações dizem respeito, por exemplo, à distinção entre palavra hifenizada e locução em casos como *fim-de-semana/fim de semana*, mas, para opções mais gerais neste aspecto, haverá necessidade de definir critérios mais explícitos, para contornar as ambiguidades do texto legal (*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 europeu⁷⁴; e, como contraponto, a palavra hifenizada está praticamente ausente das obras de referência do português do Brasil⁷⁵, sendo mais comum o registo da locução *fim de semana*⁷⁶. O registo lexicográfico⁷⁷, neste caso, poderia constituir exemplo de "exceções já consagradas", o que validaria qualquer uma das opções (como palavra hifenizada e como locução).

Como exemplo adicional desta (in)definição de critérios para uso do hífen, pode mencionar-se ainda o facto de o texto legal referir apenas o emprego do "hífen nas palavras compostas que designam espécies botânicas e zoológicas" (*Base XV*, 3.º), não se reportando a outros domínios técnicos específicos em que é comum o uso de palavras compostas que entram no mesmo paradigma. Será que este critério pode ser interpretado de maneira a separar, através de grafias diferentes (com e sem hífen), os significados de palavras que têm mais do que uma acepção, sendo uma delas da Botânica ou da Zoologia? Por exemplo, faria sentido grafar *flor-de-lis* para a acepção da Botânica e *flor de lis* para a acepção da Heráldica)? E como se combinará o uso do hífen em vocábulos de Botânica e Zoologia em contextos que impedem o uso do hífen através da *Base XVI* (por exemplo, *contra-erva*, se se considerar uma palavra composta e designar uma espécie botânica, deverá manter o hífen, mas se se considerar uma formação com o prefixo *contra-*, não deverá ser grafada com hífen).

3.2. Outras implicações

O Acordo Ortográfico terá o seu impacto mais visível em todos os utilizadores da versão portuguesa do Microsoft Office, uma vez que, como referido anteriormente, houve um licenciamento de componentes do *FLiP* pela Microsoft.

Não é de importância menor o impacto que as alterações ortográficas poderão ter em recursos linguísticos⁷⁸, nomeadamente em ontologias, *thesauri* ou dicionários bilingues, que são menos perceptíveis 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 reequacionado até haver um *corpus* estável e de qualidade na nova ortografia.

Adicionalmente, qualquer alteração significativa num componente do *FLiP*⁷⁹ pode ter consequências noutros produtos e serviços da Priberam, *on-line* ou *off-line*. Estes módulos, que podem ser individualizados e integrados em aplicações diversas, incluem um corrector ortográfico, um corrector sintáctico, 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 correcção ortográfica e sintáctica integrada em serviços de *webmail*, em programas de entrada de dados ou em motores de pesquisa. É de salientar que a

⁷⁴ Nomeadamente em Gonçalves (1966), Machado (2001), *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa* (2001) ou *Grande Dicionário Língua Portuguesa* (2004).

⁷⁵ Não consta, por exemplo, do *Vocabulário Ortográfico da Língua Portuguesa*, da Academia Brasileira de Letras.

⁷⁶ Nomeadamente no *Dicionário Houaiss da Língua Portuguesa*, o *Novo Dicionário Aurélio da Língua Portuguesa* ou o *Aulete Digital - Dicionário Contemporâneo da Língua Portuguesa*.

⁷⁷ 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".

⁷⁸ Sobre alguns destes recursos, ver Amaral *et al.* (2003).

⁷⁹ Para mais informações sobre os componentes *FLiP*, ver <http://www.flip.pt/ComponentesFLiP/tabid/296/Default.aspx>.

experiência da Priberam, com o *Dicionário da Língua Portuguesa On-Line*⁸⁰ e o *LegiX.pt*⁸¹, mas sobretudo com os sistemas de pesquisa nos *sites* da TSF e do JN, revela que muitos resultados nulos obtidos nos motores de pesquisa se devem a erros ortográficos ou grafias divergentes.

As implicações do Acordo Ortográfico em motores de pesquisa são muito importantes, uma vez que os resultados encontrados dependem da capacidade de o sistema indexar palavras que sofreram alterações ortográficas e relacioná-las com a sua grafia anterior (e vice-versa), tendo também em conta os erros existentes ou digitados. É monumental o volume de informação disponível que não será com certeza actualizado, não estando os motores de pesquisa devidamente preparados para estabelecer essa relação entre ortografias.

Por outro lado, a necessidade inevitável de adaptação de grande quantidade de textos em formato electrónico durante e após o período de transição de 6 anos legalmente previsto, levou a Priberam a desenvolver ferramentas de conversão automática de texto para a nova ortografia. Estas ferramentas contemplarão no essencial as mesmas opções definidas para os correctores ortográficos, poderão ser integradas em diferentes aplicações ou recursos e terão em conta configurações pessoais ou institucionais prévias.

Acresce dizer que reflexão feita nesta secção se baseia quase exclusivamente em argumentos relativos às normas europeia e brasileira do português, por não dispormos ainda de um conjunto de instrumentos descritivos que nos permitam aferir características das variedades africanas do português e das áreas mais problemáticas para os utilizadores da língua. Através dos seus serviços de consultoria linguística, a Priberam consegue, ainda assim, avaliar algumas especificidades que seriam necessárias para pacotes de ferramentas destinados a essas variedades. Aos recursos de que a Priberam dispõe actualmente poderão juntar-se, 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. E, do mesmo modo que no Microsoft Office estão previstas 15 variedades de francês, 18 variedades de inglês ou 20 variedades de espanhol, é desejável que, futuramente, surjam disponibilizadas outras variedades do português (ver Ilustração 1).

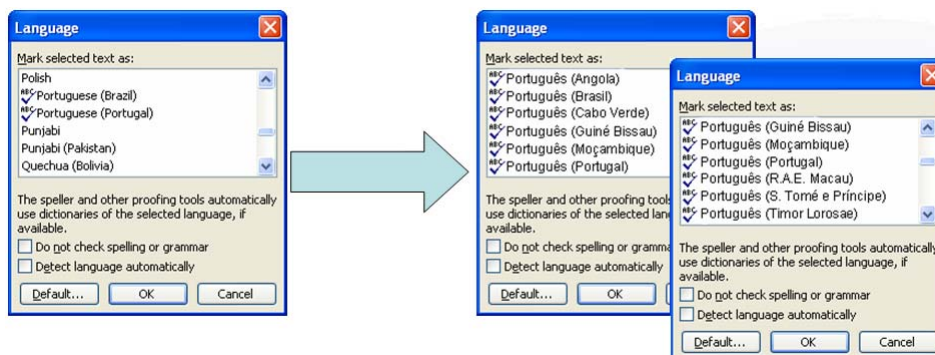


Ilustração 1: Variedades do português no Microsoft Office (à esquerda, estado actual).

Prever variedades africanas como objecto de projectos futuros coloca questões adicionais: as que dizem respeito às práticas ortográficas nacionais (originadas por influência anglo-saxónica ou por motivos políticos⁸²), 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*, 2º) 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 desses países, mas sobretudo porque sempre foram as academias portuguesa e brasileira os motores das alterações ortográficas, pelo peso tradicional das duas normas.

4. Conclusão

O "Acordo Ortográfico da Língua Portuguesa" de 1990 tem consequências não despidiendas no PLN, nomeadamente em instrumentos de correcção ortográfica e noutros recursos linguísticos e ferramentas associados. Este facto inegável não torna, contudo, evidentes as dificuldades na implementação das

⁸⁰ Ver <http://www.priberam.pt/dlpo/>.

⁸¹ Ver <http://www.legix.pt>.

⁸² Roth (2004: pp. 45-46).

alterações, que advêm tanto da qualidade do texto legal como da ausência (que se espera ser apenas a curto prazo) de instrumentos reguladores autorizados (vocabulário ou vocabulários, mas também, por exemplo, diretrizes do Ministério da Educação para o ensino da ortografia).

Uma das consequências da implementação do novo Acordo Ortográfico é a manutenção de vários correctores, mas com um maior grau de personalização.

Neste artigo, procurou ainda analisar-se alguns dos casos problemáticos, mais especificamente aqueles em que é evidente a necessidade de explicitação de opções e de critérios, uma vez que a posição da Priberam no respeitante à aplicação da nova ortografia não se limita às suas responsabilidades no mercado dos correctores automáticos, mas tem também um efeito nas escolhas de grupos editoriais, instituições e utilizadores individuais. No limite, um corrector automático pode condicionar o impacto que o Acordo Ortográfico tem na sociedade, nomeadamente como contributo pedagógico para a aplicação de novas regras ortográficas.

5. Referências

- A *Questão do "Acordo Ortográfico"*, s/l: Movimento contra o Acordo Ortográfico, 1988.
- ACADEMIA BRASILEIRA DE LETRAS, *Vocabulário Ortográfico da Língua Portuguesa*, [Rio de Janeiro]: Bloch Editores S.A., 1981 [versão actualizada em linha em <http://www.academia.org.br/abl/cgi/cgilua.exe/sys/start.htm?sid=23>].
- 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 [disponível em <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 [disponível em 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* [em linha], 2008, <http://www.priberam.pt/docs/AcOrtog91.pdf> [descarregado em 2008-07-28].
- Aulete Digital - Dicionário Contemporâneo da Língua Portuguesa* [descarregado em 2007, de <http://www.auletedigital.com.br/auletedigital/>].
- AZEVEDO, Domingos de, *Grande Dicionário Português/Francês*, 11.ª ed., Venda Nova: Bertrand Editora, 1998.
- CASTELEIRO, João Malaca (dir.), *Dicionário Gramatical de Verbos Portugueses*, Lisboa: Texto Editores, 2007.
- CASTELEIRO, João Malaca e Pedro CORREIA, *Atual*, Lisboa: Texto Editores, 2007.
- CASTRO, Ivo, "O Linguista e a Fixação da Norma", in MENDES e FREITAS (orgs.), *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa: APL, 2003, pp. 11-24.
- CASTRO, Ivo, Inês DUARTE, Isabel LEIRIA (orgs.), *A Demanda da Ortografia Portuguesa*, Lisboa: Edições João Sá da Costa, 1987.
- CHORÃO, João Bigotte (dir.), *Dicionário Verbo Português-Inglês*, Lisboa: Editorial Verbo, 2000.
- CUESTA, Pilar Vásquez e Maria Albertina Mendes da LUZ, *Gramática da Língua Portuguesa*, Lisboa: Edições 70, 1971.
- DELGADO-MARTINS, Maria Raquel, "Representações da ligação 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.
- Dicionário Editora da Língua Portuguesa 2009 – Acordo Ortográfico*, Porto: Porto Editora, 2008.
- 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, Antônio, VILLAR Mauro, *Dicionário Houaiss da Língua Portuguesa*, Instituto Antônio Houaiss de Lexicografia e Banco de Dados da Língua Portuguesa S/C Ltda., Rio de Janeiro: Editora Objetiva, 2001.
- HOUAISS, Antônio, 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 Bacerlar 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.linguatca.pt/documentos/Medeiros95.doc>].
- MINISTÉRIO DA EDUCAÇÃO NACIONAL, "Decreto n.º 35 228", de 8 de Dezembro de 1945, in *Priberam* [em linha], 2008, http://www.priberam.pt/docs/AcOrtog45_73.pdf [descarregado em 2008-07-28].

- PETERSON, James L., *Computer Programs for Spelling Correction*, Berlin: Springer-Verlag (LNCS, nº 96), 1980.
- PRIBERAM, *Priberam - 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, Guisepe, "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.
- TEXTO EDITORES, *Novo Dicionário da Língua Portuguesa Conforme Acordo Ortográfico*, Lisboa: Texto Editores, 2007.
- TEXTO EDITORES, *Novo Grande Dicionário da Língua Portuguesa Conforme Acordo Ortográfico*, Lisboa: Texto Editores, 2007.