

Vocabulário Ortográfico Comum da Língua Portuguesa (VOC)

Após a apresentação prévia de esboço da plataforma do VOC, em outubro de 2013, na II Conferência Internacional sobre o Futuro da Língua Portuguesa no Sistema Mundial, seguida de validação pelo Conselho Científico do Instituto Internacional da Língua Portuguesa, reunido em maio de 2014 na cidade da Praia, e posterior incorporação no património da CPLP, por ocasião da Cimeira de Chefes de Estado e de Governo da Comunidade, que teve lugar em Díli em julho de 2014, foi formalmente lançada a 19 de fevereiro, na sede da CPLP em Lisboa, a plataforma do VOC. A plataforma está disponível publicamente através da página do IILP (<http://iilp.cplp.org>) e de endereço próprio (<http://voc.cplp.org>).

O VOC incorporará a partir deste momento, em modelo de inserção gradual, os VON (Vocabulário Ortográfico Nacional) de Brasil, Moçambique, Portugal e Timor-Leste, incluindo palavras representativas do português escrito nesses países, tanto do léxico comum, como das áreas críticas da ortografia que representam a toponímia e as formas provindas de outras línguas e cuja grafia não se encontra adaptada à do português. Até maio, os trabalhos lexicográficos de inserção e revisão de dados e de melhoramento da plataforma continuarão, com o objetivo de até lá atingir 250 000 entradas principais, mais de 10 000 topónimos e cerca de 2500 formas não adaptadas, mas, sobretudo, de garantir que os dados então presentes na plataforma sejam suficientemente confiáveis e representativos de todos os países participantes.

De acordo com a metodologia acordada no começo do projeto, o VOC albergará VON com diferentes processos genéticos: os de Brasil e Portugal resultam da incorporação controlada e gradual na plataforma, ainda em curso, dos dados do VOLP, da ABL, e do VOP, do CELGA-ILTEC, vocabulários previamente existentes, oficiais nos respetivos países, que representam a *memória lexicográfica da língua portuguesa*, constituindo a base lexical do recurso que será alojado na plataforma ora dada a conhecer ao público. Outros, como o Moçambicano e o de Cabo Verde foram construídos validando com recurso a *corpora*, uma parte da memória lexicográfica ativa em cada país, depois enriquecida com palavras não registadas naquelas obras, obtidas de fontes várias, sobretudo lexicográficas e literárias. Por último, os vocabulários de Timor-Leste e de São Tomé e Príncipe foram construídos através da recolha sistemática de léxico específico, a que foi depois acrescentado como representativo o léxico comum a todos os restantes países.

A plataforma do VOC determina, através de um processo computacional, o país de proveniência de cada visitante, provendo dados adequados à localização através do chamado *locale* (por ex. pt-BR). A localização selecionada é indicada na plataforma por bandeira, à esquerda, sendo possível alterá-la para outro país ou selecionar, clicando no símbolo do IILP, a versão *comum*, que apresenta todos os dados e toda a variação existente. A cada consulente serão, pois, apresentados os dados referentes ao seu país de origem. Não são, no entanto, apresentados em cada caso apenas os dados registados explicitamente para cada país: são dadas ao utilizador todas as palavras cujas formas sejam recomendáveis no seu espaço político. Por exemplo, não será apresentada a forma

“anónimo” no Brasil, nem a forma “anistia” em Moçambique, mas será apresentada a forma “tais”, masculino, um tradicional pano leste-timorense, em Portugal ou em qualquer outro país. As formas atestadas e não atestadas em cada país podem ser identificadas pelo consulente através de uma marca junto de cada entrada, na listagem: um quadrado cheio nos casos de palavras atestadas para aquele VON; um quadrado vazio quando provêm da memória lexicográfica do português, sem registo nas fontes do VON sob consulta.

Neste momento, as formas podem ser listadas por letra ou por pesquisa na caixa de busca à esquerda da página, que usa o texto inserido como sendo um padrão; por exemplo, para uma pesquisa por “casa” será dado como resultado não apenas “casa”, mas também “casar”, “casaca” e muitas outras. Até maio, o motor de busca será atualizado para permitir encontrar formas mesmo em caso de erro de digitação, e passará a ser possível aos consulentes transferir listagens de dados. Também passaremos a mostrar, já dentro de semanas, a divisão silábica e a marcação da sílaba tónica de cada palavra e, um pouco mais tarde, as remissões explícitas entre entradas relacionadas (nomeadamente, na versão comum as variantes geograficamente delimitadas). Serão também apresentadas sob forma simplificada as regras ortográficas, de acordo com a *Sistematização* consensuada pela Equipa Central e pelo Corpo Internacional de Consultores, bem como textos de documentação do projeto.

Com os trabalhos em curso até maio, inserindo gradualmente mais entradas providas do VOLP e do VOP, bem como muitas formas providas de outros VON e que pela primeira vez figurarão em instrumentos lexicográficos da língua portuguesa, contamos ultrapassar as 250 000 entradas lexicais principais, a que corresponderão cerca de 2 milhões de formas flexionadas da língua portuguesa. Os cerca de 10 000 topónimos já verificados e validados, que agora inseriremos na plataforma, são representativos das mais relevantes divisões administrativas dos países participantes, esperando-se que o número possa subir para além das 50 000 entradas com a inclusão de níveis administrativos mais baixos. Por último, para cada uma das 2500 formas não adaptadas será inserida não apenas a forma, em si, mas também a língua de proveniência e, caso os tenha, o domínio de especialidade em que se inscreve e a forma adaptada em português. O trabalho em curso de maior importância, no entanto, passa pela verificação aturada dos dados inseridos, garantindo que são lexicograficamente sãos e adequados a cada espaço geográfico que pretendem representar no VOC; a versão atualmente disponibilizada na plataforma, como é fácil de detetar, reflete ainda uma versão incipiente dos dados, carente de cuidados lexicográficos mais aturados.

Com efeito, no seu estado atual, a plataforma e os dados do VOC apresentam ainda fragilidades várias, que terão que ser corrigidas nos próximos meses, pelo que não marcaremos desde já todas as alterações de forma explícita. A partir de maio, no entanto, todas as alterações às bases de dados ficarão marcadas com a data e as propriedades da última alteração. Nessa data, será também possível obter, através de transferência, listas de palavras específicas de um país ou correspondentes a uma dada pesquisa feita pelo consulente. Por último, contamos poder adicionar nessa data uma função de comentários e sugestões que nos permitirá obter retorno dos consulentes, ajudando a detetar

informação carente de revisão e a suprir dúvidas junto do público em geral e especializado.

Fevereiro de 2015

A Equipa Central do VOC (texto adaptado)