

VOP
Vocabulário Ortográfico do
Português

Margarita Correia
c/ a colaboração de José Pedro Ferreira

Instituto de Linguística Teórica e Computacional - Lisboa

Plano da apresentação

- O que é o *VOP*?
- A génese
- A filosofia de base
- O resultado
- Outros recursos
 - O Lince e o Portal da Língua Portuguesa

O que é o VOP?

- O *VOP* – *Vocabulário Ortográfico do Português* é um vocabulário ortográfico, em suporte informático, de acesso livre e apto para ser constantemente atualizado e desenvolvido. Está alojado no Portal da Língua Portuguesa.

www.portaldalinguaportuguesa.org/

- O VOP é um vocabulário de larga escala, com mais de 200 000 entradas principais, feito com base no VLP de Rebelo Gonçalves.
- O que o distingue de outros vocabulários existentes é o facto de ser pensado de forma modular e de modo a permitir a sua expansão e a criação de novos recursos de modo simples.

A génese

- A ideia inicial era criar apenas um conversor. No entanto, apenas faria sentido criar um conversor tendo como base um vocabulário atualizado e moderno, da era digital.
- Foi por isso criado o VOP, usando a plataforma de edição e de disponibilização ao público em geral que vinha sendo criada no ILTEC, o Portal da Língua Portuguesa.

1. Motivação

Grandes preocupações levantadas pela aplicação do AO:

- a) Custos, morosidade, e desatualização dos recursos existentes
- b) Dificuldade técnica de aplicação das mudanças e dispersão de critérios
- c) Dificuldade de aplicação em todos os países da CPLP

a) Custos, morosidade, e desatualização dos recursos existentes

Na era digital, estes problemas têm que ser ultrapassados com um vocabulário eletrónico e por um conversor informático de larga escala e bem pensado. Para lhes fazer face, o Governo Português aprovou o financiamento do conversor Lince e do VOP.

b) Dificuldade técnica de aplicação das mudanças e dispersão de critérios

O texto do AO é omissivo quanto a vários aspectos da grafia do português e contém algumas incongruências. A aplicação do AO antes de estarem publicados recursos de base fez com que haja publicações extemporâneas e com entendimentos / interpretações diversos.

Estes problemas implicam a criação de um recurso centralizado, um VOP, que garanta uma interpretação única do AO.

c) Dificuldade de aplicação em todos os países da CPLP

A aplicação de algumas regras implica o conhecimento da língua tal como ela é usada em cada país, sem que haja, na generalidade dos casos, dados (é este o caso de Portugal). A feitura de um VOC com base numa base metodológica una permitiria ultrapassar estes problemas.

- O projeto VOP foi submetido em simultâneo com o projeto *Conversor para a Nova Ortografia* (hoje chamado *Lince*), ao Fundo da Língua Portuguesa.
- Foi aprovado e teve início em junho de 2009.
- A versão Beta foi publicada em fevereiro de 2010.
- A versão atual foi disponibilizada em junho de 2010.

A filosofia de base

- O VOP foi concebido como:
 - Um vocabulário representativo do Português Europeu;
 - Um serviço público – disponível gratuita e universalmente;
 - Uma plataforma digital, atualizável e extensível, podendo ser usada por outros, por exemplo, para a elaboração do VOC.

A metodologia

- Estabelecimento de critérios de aplicação do AO;
- Estabelecimento de critérios lexicográficos;
- Adaptação das bases de dados existentes e criação das interfaces de trabalho;
- Selecção das fontes a incluir;
- Revisão e actualização da base de dados previamente existente, a MorDebe (cerca de 150 000);
- Inserção dos novos lemas;
- Revisões e actualizações.

O resultado

- Mais que uma mera lista de palavras com indicação de categoria gramatical, apresenta:
 - a flexão completa de cada palavra,
 - a divisão em sílabas,
 - a marcação da sílaba tónica,
 - as relações funcionais estabelecidas entre classes, entre outras.
- <http://www.portaldalinguaportuguesa.org/index.php?sel=exact&query=negociar&action=simplesearch&base=form>

- Uma das propriedades que mais o destaca é o facto de indicar em todos os casos de variação, na situação ortográfica até há pouco legalmente em vigor, as variantes correspondentes à grafia do Brasil e às das restantes variedades.

[http://www.portaldalinguaportuguesa.org/index.php?action=lemma&lemma=89852&highlight=^an%F3nimo\\$](http://www.portaldalinguaportuguesa.org/index.php?action=lemma&lemma=89852&highlight=^an%F3nimo$)

<http://www.portaldalinguaportuguesa.org/index.php?sel=exact&query=facto&action=simplesearch&base=form>

- A arquitetura de dados foi pensada para a reusabilidade, permitindo a expansão rápida (as 158 000 formas do VLP foram inseridas em cerca de um ano) e a construção fácil de recursos derivados, disponibilizados imediatamente no Portal da Língua Portuguesa, tais como
 - Vocabulário de Mudança;
 - Lince;
 - Outros dicionários e outras ferramentas
 - www.portaldalinguaportuguesa.org

Lince

- O Lince é uma ferramenta que converte o texto de ficheiros informáticos para a nova grafia. Funciona em todos os sistemas operativos, suporta ambas as versões antigas de ortografia do português, converte todos os formatos de ficheiro de texto mais comuns e permite converter em simultâneo um grande número de ficheiros com grande rapidez.
- Permite definir as preferências do utilizador nos casos de variação gerada pelo AO (o que não se aplica, no caso do Brasil), escolher o destino do ficheiro convertido e seleccionar excertos de texto que não se quer converter. Dentro de semanas, até ao fim do ano, permitirá marcar no texto convertido as alterações efetuadas aquando da conversão.

O Portal da Língua Portuguesa

O PLP é uma plataforma em linha que permite a disponibilização universal de recursos linguísticos ao público em geral. Todos os recursos desenvolvidos pelo grupo do Léxico e Modelização Computacional do ILTEC vão sendo disponibilizados gratuitamente no Portal à medida que são concluídos. No Portal, além do VOP e do Lince, estão neste momento os seguintes recursos:

- Dicionário de Gentílicos e Topónimos
- Dicionário de Estrangeirismos
- Vocabulário de Mudança
- Dicionário de Nomes Deverbais
- Dicionário de Nomes de Qualidade
- Dicionário de Advérbios Deadjetivais
- a legislação passada e presente, como o texto de todas as reformas ortográficas e as nomenclaturas gramaticais em vigor.

- Estes recursos são mantidos e expandidos constantemente pela equipa do grupo, neste momento com 11 pessoas. Toda essa gestão dos recursos, bem como os pouco morosos processos que os criaram, tem como base um ambiente de administração.
- É essa interface de administração que sustenta centralizadamente todos os recursos, incluindo o VOP e o Lince. Embora seja a parte menos visível dos nossos recursos, é na verdade essa a nossa ferramenta mais valiosa e que maior contributo pode dar no futuro:
 - É de conceção modular, sendo possível adaptá-la com relativa rapidez a novas tarefas e é operada apenas por linguistas especializados em lexicografia, morfologia, fonética, etc.

Neste momento, está pensada para gerir as bases de dados da seguinte forma:

- extrai automaticamente neologismos a partir de fontes selecionadas, como jornais de referência. Quando atingem elevada frequência e estabilidade, esses neologismos são tratados lexicograficamente, sendo depois inseridos, de acordo com critérios uniformes.

- integra algoritmos de verificação contínua dos dados, de modo a garantir a sua idoneidade e homogeneidade;

- integra verificações que asseguram que os dados de cada base estão de acordo com as restantes;

- permite a resposta a dúvidas de utilizadores;

- contém um módulo, feito propositadamente para o VOP, que permite a criação muito rápida de vocabulários a partir de fontes ou de *corpora*, previamente tratados segundo metodologias específicas.

Foram estas ferramentas e metodologias que permitiram criar com tão grande rapidez a maior base de dados lexicais, digital e pública, do português, em tão pouco tempo.