

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D19P92>

Title: Uma proposta de vocabulário bilíngue de pesticidas organofosforados por meio da linguística de corpus: foco no trato da variação denominativa

Creator: José victor De souza - **ORCID:** [0000-0002-4587-4073](https://orcid.org/0000-0002-4587-4073)

Affiliation: São Paulo State University (unesp.br)

Principal Investigator: José Victor de Souza

Project Administrator: Paula Tavares Pinto

Funder: São Paulo Research Foundation (fapesp.br)

Funding opportunity number: 2021/08830-9

Grant: <https://bv.fapesp.br/pt/bolsas/203005/uma-proposta-de-vocabulario-bilingue-de-pesticidas-organofosforados-por-meio-da-linguistica-de-corpu/>

Template: Digital Curation Centre (português)

Project abstract:

O Brasil é um dos maiores produtores agrícolas no mundo e, por conseguinte, um dos maiores consumidores de agrotóxicos. Dessa maneira, pesquisas científicas dentro da área da química de pesticidas tornam-se cada vez mais imperativas no país. Contudo, no que diz respeito à terminologia da área, constataram-se problemas tradutórios e terminológicos em textos científicos que tratam de pesticidas dentro da categoria dos organofosforados. Os problemas giram em torno, principalmente, da falta de consenso das traduções de termos do inglês para o

português, o que gera notável variação denominativa. Assim, o presente projeto tem como objetivo geral inventariar termos da química de pesticidas em português e em inglês, dentro da subárea dos organofosforados, tendo como objetivo específico: a) definir a microestrutura do verbete para um vocabulário especializado da área, levando em conta suas particularidades, as variações denominativas encontradas bem como suas possíveis causas. Serão extraídos termos de um corpus bilíngue comparável por meio do cálculo da chavicidade de cada item lexical. Ao serem identificadas variantes dentre os termos, estas serão classificadas quanto a seu tipo. Em seguida, serão aplicados índices estatísticos e feitas consultas a um especialista da área para definir a variante mais adequada para constar na entrada do verbete. Por fim, a partir de dados do corpus e de entrevista com o especialista, serão definidas as possíveis causas das variações denominativas encontradas. As informações serão registradas em fichas terminológicas e culminarão num vocabulário bilíngue.

Start date: 07-01-2022

End date: 04-30-2023

Last modified: 05-15-2023

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Uma proposta de vocabulário bilíngue de pesticidas organofosforados por meio da linguística de corpus: foco no trato da variação denominativa

Coleta de Dados

Que dados serão coletados ou criados?

Os dados coletados serão o conteúdo textual de artigos científicos, teses e dissertações. O formato escolhido para armazenamento foi o de texto simples (.txt). Todos os dados coletados para a pesquisa são de acesso livre estão disponíveis gratuitamente na web. Os dados criados a partir da análise serão fichas terminológicas, salvas no formato MS Access. Todos os dados estarão disponíveis no computador pessoal do pesquisador principal e na nuvem para acesso da equipe de pesquisa, tendo um volume de por volta de 100 MB.

Como os dados serão coletados ou criados?

Os dados serão coletados da web manualmente ou por web scraping. Todos os dados coletados para a pesquisa estão disponíveis na web. As fichas terminológicas serão elaboradas pelo pesquisador principal com base em trabalhos científicos publicados anteriormente.

Documentação e Metadados

Que documentação e metadados acompanharão os dados?

Os dados criados encontrar-se-ão em anexo aos relatórios e à dissertação do beneficiário, podendo ser consultados por meio do repositório da Unesp uma vez publicados. Todos os dados podem ser lidos por softwares leitores de texto ou PDF (Word, LibreOffice, Adobe e afins).

Ética e Conformidade Legal

Como você administrará qualquer questão ética?

Não há questões éticas envolvidas

Como você vai gerenciar os direitos autorais e os direitos de propriedade intelectual (IP / IPR)?

Por questões de direitos autorais, o corpus compilado em sua totalidade não será

disponibilizado para acesso livre. Ademais, durante a análise do corpus, os textos são analisados por fragmentos (linhas de concordância) e não são disponibilizados em sua totalidade.

Armazenamento e Backup

Como os dados serão armazenados e terão backup durante a pesquisa?

Todos os dados serão armazenados no computador pessoal do pesquisador principal com backup na nuvem do Sketch Engine. Não há data de validade.

Como você vai gerenciar o acesso e a segurança?

Somente membros do grupo de pesquisa têm acesso aos dados por meio das credenciais do Sketch Engine.

Seleção e Preservação

Quais dados são de valor a longo prazo e devem ser mantidos, compartilhados e / ou preservados?

Os dados gerados poderão ser utilizados para pesquisas futuras do pesquisador principal ou de seus colegas de pesquisa. Nenhum dado será destruído ou retido.

Qual é o plano de preservação a longo prazo do conjunto de dados?

A dissertação com a análise dos dados ficará disponível no repositório da Unesp. Os dados em si ficarão preservados na nuvem. Não há geração de custos.

Compartilhamento de Dados

Como você vai compartilhar os dados?

Os dados só serão acessíveis por meio da análise do pesquisador principal. Nenhum usuário terá acesso aos dados brutos a não ser membros do grupo de pesquisa do pesquisador principal. O compartilhamento dos dados brutos é feito por meio do Sketch Engine ou por meio de outros recursos da nuvem (Gmail, Drive, etc.). O glossário será disponibilizado para livre acesso.

Existem restrições ao compartilhamento de dados requeridos?

Não se aplica

Responsabilidades e Recursos

Quem será responsável pelo gerenciamento de dados?

O pesquisador principal e os orientadores serão responsáveis pelos dados.

Quais recursos você precisará para entregar seu plano?

Todos os recursos necessários são disponibilizados pelo instituto.
