



Simpósio de Integração Acadêmica

"Ciências Básicas para o Desenvolvimento Sustentável"

SIA UFV 2023



The language of the Universe

Pesquisa na área de linguística | Palavras-chave: Verbo de alta frequência, Fraseologia, Corpus de aprendiz
Sarah Samilly d'Lucas Gonçalves | Bolsista BIC Jr em FAPEMIG | samillysarah9@gmail.com
Valdênia Carvalho e Almeida | Orientadora | valdenia@ufv.br

Introdução

O telescópio espacial James Webb (JWST) desde julho de 2022 tem surpreendido o mundo com imagens impressionantes do Universo. Ele é o telescópio mais poderoso já construído, marcando uma nova era de pesquisas científicas. Um exemplo notável é a Nebulosa de Órion, que pode revelar segredos sobre o nascimento das estrelas e a evolução da Via Láctea.

O projeto visou construir e explorar o "The Universe Corpus," um conjunto de textos em inglês sobre o Universo, para analisar traços linguísticos relacionados ao discurso espacial. O corpus foi organizado em três seções e utilizado para identificar colocações linguísticas. No futuro, pode ser usado para criar um glossário especializado em terminologia espacial.

Objetivos

O principal objetivo deste trabalho foi criar um corpus eletrônico chamado "The Universe Corpus" para apoiar pesquisas linguísticas, especialmente em Terminologia e Ensino de Línguas. Esse corpus abrangeu uma variedade de textos relacionados à teoria cosmológica, física do universo, constituição do universo, mecânica quântica, e teoria geral da relatividade de Einstein. Além da construção do corpus, o estudo visou identificar os termos mais comuns, focando principalmente em substantivos e adjetivos, e analisar as colocações linguísticas formadas por esses termos.

Material e Método

O estudo seguiu a metodologia da Linguística de Corpus para construir o "Corpus do Universo" a partir de fontes como a NASA, CSA e vídeos de cosmologia e astronomia no YouTube. As etapas incluíram a limpeza e adequação dos dados textuais, a identificação dos textos e metadados, a geração de dados usando o programa AntConc, e a análise linguística baseada nos princípios da Linguística de Corpus.

Resultados e Discussão

Este tópico apresenta os resultados da criação e análise do "Universe Corpus" (CorUni), segundo a metodologia. O "Universe Corpus" contém 106.298 palavras. Analisamos os itens lexicais mais frequentes, identificando 100 substantivos e adjetivos em cada seção do corpus. Há semelhança entre as seções da NASA e CSA devido à origem das fontes. Seleccionamos palavras com alta frequência em todas as seções, identificamos colocações que

revelam terminologia técnica espacial e física teórica. Destaque para o nome "James Webb" com 49 ocorrências no corpus, enriquecendo nossa compreensão da pesquisa espacial.

Número total de palavras em cada seção do corpus:

NASA	CSA	Youtube
29389	21757	55152

5 substantivos mais frequentes em cada seção do corpus:

NASA	sun	earth	planet	solar	system
CSA	earth	sun	solar	planet	space
Youtube	time	quantum	space	world	universe

5 colocações ou combinações multi-palavras mais frequentes nas três seções do corpus: solar system; black hole; quantum mechanics; magnetic field; human beings.

Conclusões

A análise do corpus construído possibilitou a identificação de várias colocações ou multi-combinação de palavras que fazem parte da terminologia ou vocabulário técnico da área estudada, tais como *solar system*, *black holes*, *Milky Way*, *magnetic field*. Os resultados apontam o impacto do telescópio James Webb na geração de dados astronômicos e o seu papel na divulgação científica, o que pode ser percebido pela alta frequência em que o nome James Webb ou apenas Webb aparece no corpus.

Em conclusão, a pesquisa proporcionou uma melhor compreensão da linguagem do Universo, contribuindo para o conhecimento sobre o assunto.

Bibliografia

- ANTHONY, L. (2021), [Computer Software]. Tokyo, Japan: Waseda University. Disponível em: <<http://www.antlab.sci.waseda.ac.jp/>>. Acesso em: 19 de setembro de 2022.
- BERBER SARDINHA, T. Linguística de Corpus. São Paulo: Manole, 2004.
- SINCLAIR, J. M. Corpus, concordance, collocation. Oxford: Oxford University Press, 1991.

Agradecimentos

Agradeço a FAPEMIG pelo financiamento e ao Colégio de Aplicação, CAP-COLUNI pelo espaço para a realização do projeto.

Apoio financeiro

