ANÁLISE DIALETOMÉTRICO COM PRODIS – UM PASSO A MAIS NOS ESTUDOS PROSÓDICOS DO AMPER

Citation: Fernández Planas, A.M.; Elvira-García, W.; Balocco, S.; Roseano, P. (2019). Análisis dialectométrico con ProDis: Un paso más en los estudios prosódicos de AMPER. En: J. Dorta (ed.) Investigación geoprosódica. Amper: análisis y retos. Madrid/Frankfurt am Main, Iberoamericana Vervuert. pp. 119-135 ISBN 978-84-9192- 039-7.

* Trabalhar os dados do AMPER para abordar as diferenças geoprosódicas e ver como se distribuem e se agrupam a partir de critérios quantitativos, levando em conta os parâmetros de F0, duração e intensidade.
* Dialetometria:

1. uma disciplina que iniciou nos anos 70 a partir dos estudos de Séguy (1973) e Goebl (1981) e definem:
2. É uma aliança metodológica entre a geolingüística e a taxonomia numérica como disciplina matemática.

* Objetivo:

1. Estudos dialetométricos pretendem utilizar a enorme quantidade de dados gerados através de estudos dialetológicos e nos Atlas Linguísticos para estabelecer agrupamentos entre a base de dados empíricos disponíveis e obter uma distribuição do espaço virtual dos dados.
2. Levar a cabo um tipo de análise estatístico para o cálculo das distâncias dialetais entre distintas línguas ou variedades de uma mesma língua, para o qual a dialetometria funciona operando com grandes quantidades de dados de forma quantitativa e objetiva.

* Estudos do AMPERCAT:

1. Baseou sua metodologia ferramenta Stat-Distance (Romano et al, 2011) Moutinho et al, 2011, Rilliard et al, 2008)

* A Dialetometria oferece algumas vantagens inegáveis em relação a dialetometria clássica:

1. Permite questionar de forma relativamente simples quantidades enormes de dados que de forma humana não seria possível e partir dele extrair conclusões que são estatisticamente mais fiáveis.
2. Não há, a priori, um tratamento de dados posto que se considera grandes quantidades de dados sem ponderar quantitativamente a função de um parâmetro e outro.
3. Oferece os resultados em forma de gráficos, tabelas, dendogramas, que resultam de análise de Clúster, e gráficos de distribuição em um espaço virtual que resultam de processos de escalamento multidimensional.
   * Programas para trabalhar a dialetometria: (usaram dados alfabéticos)
4. Visual Dialetometry (Goebl, 2004)
5. Gabmap (Nerbonne, 2011)
6. Diatech (Aurrekaetxea, 2013)

* Para os dados do AMPER é preciso de um programa que :

1. Seja capaz de trabalhar com dados numéricos representativos dos parâmetros que se estudem, estando os dados normalizados em semitons.
2. Que seja capaz de considerar 3 repetições de cada frase por informante
3. Que possa ponderar por duração, por intensidade ou por ambos parâmetros, se assim desejar.
4. Que seja capaz de trabalhar com frases que para a mesma estrutura sintática possuam diferentes números de sílabas entre diferentes línguas.
5. Que permita manter a comparação entre todas as que se desejar mesmo quando não compartilham as mesmas estruturas sintáticas.
6. Que ofereça, além dos gráficos com resultados (dendogramas, MDS...) as matrizes de correlações resultantes de todos os cálculos
7. Que incorpore a significação estatística de todos os gráficos resultantes.
   * No AMPER começou-se a desenvolver um primeiro conjunto de rotinas – Stat- Distances (Romano e Miotti, 2008), Rilliard e Lai, 2008), (Romano et. Al, 2011) a partir de uma ideia inicial pautada na tese de Antonio Romano (1999) com o objetivo de trabalhar a dialetometria no AMPER, porém resultou numa ferramenta inacabada posto que não satisfazia os últimos 3 pontos acima.
   * UB- Laboratório de Fonética – desenvolveu um sistema de trabalho alternativo que utiliza uma série de rotinas para calcular distâncias de uma parte do corpus com a distância euclidiana no Praat, excel e SPSS – chamado de calcu-Dista (Roseano et al, 2017). Aprimorou-se para o Pro-Dis (Elvira Garcia et al 2018)
   * PRODIS :
8. Realiza a media e a mediana de correlação (Perarson) por informantes e por ponto de inquérito. A partir desses dados realiza uma análise de clúster e uma análise MDS que permite classifica-los em diferentes grupos segundo sua semelhança.

Com base nisso, pode-se ponderar a correlação pela intensidade (Hermes, 1991) pela duração ou ambos.

1. Funcionamento:

A partir de uma base de dados numéricos em arquivos txt, é preciso calcular a matriz de correlações que lhes corresponde e refletir os resultados em forma de tabelas e principalmente gráficos cuja apresentação visual facilita a compreensão dos agrupamentos.

Uma vez computadas as correlações, se passa a criar grupos de informantes e de pontos de inquéritos que tenham correlação que tenham correlações altas e se realiza uma análise de Clúster a partir de uma técnica de classificação não supervisionada para que o investigador não defina previamente quantos grupos deseja obter. Se trata de uma análise de Clúster aglomerativa hierárquico com “lincaje” completo que agrupa os elementos a partir do elementos mais distante como referência.

Os resultados podem estabelecer-se em forma de distribuição de potnos de informantes em um espaço virtual a partir das técncias de escalamento Multidimensional MDS)

* + - * Gráfico de Shepard
      * Dendorgama bidimensional e tridimensional com valor de stress em ambos os casos
      * Mapa geográfico com pontos de inquérito e sua descrição dialetal.