

# Projetos de Pesquisa em Ciências da Linguagem 2024

PROJETOS DE PESQUISA EM CIÊNCIAS DA LINGUAGEM | N.º 1 | 2024  
FICHA TÉCNICA

**Periodicidade:** anual

**Capa:** Marina Salimon

**Autores:**

*Cristiana Mesquita Reis:* estudante de Ciências da Linguagem na Faculdade de Letras da Universidade do Porto 2021-2024

*Fábio Barcellos Granja:* estudante de Ciências da Linguagem na Faculdade de Letras da Universidade do Porto 2021-2024; bolsheiro de investigação da Fundação Calouste Gulbenkian

*Joana Ribeiro da Silva Gomes:* estudante de Ciências da Linguagem na Faculdade de Letras da Universidade do Porto 2021-2024

*Letícia Rodrigues Alberici:* estudante de Línguas, Literaturas e Culturas na Faculdade de Letras da Universidade do Porto 2022-2025

*Marina Araújo Salimon:* estudante de Ciências da Linguagem na Faculdade de Letras da Universidade do Porto 2021-2024

*Rafaela Gomes de Moraes Fernandes da Gama:* estudante de Ciências da Linguagem na Faculdade de Letras da Universidade do Porto 2021-2024

**Comissão editorial:** Cristiana Mesquita Reis; Fábio Barcellos Granja; Joana Ribeiro da Silva Gomes; Letícia Rodrigues Alberici; Marina Araújo Salimon; Rafaela Gomes de Moraes Fernandes da Gama

**Editores:** Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto

**Conselho Científico:**

*Ana Rita Fontes da Cunha:* Faculdade de Letras da Universidade do Porto

*António José Rodrigues Leal:* Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto

*Carlos Rogério Sousa e Silva:* Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto; Laboratory of Language Evolution da Universidade de Neuchâtel

*Isabel Margarida Duarte:* Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto

*Juliana Novo Gomes:* Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto

*Maria Alexandra Guedes Pinto:* Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto

*Maria da Purificação Moura Silvano:* Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto

*Vanessa Gomes Teixeira Anachoreta:* Faculdade de Letras da Universidade do Porto; Centro de Linguística da Universidade do Porto

**Coordenação:** Maria Alexandra Guedes Pinto

**ISBN:** 978-989-9193-62-8

**DOI:** <https://doi.org/10.54499/UIDB/00022/2020>

Este trabalho é financiado por fundos nacionais através da FCT - Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/00022/2020.

## ÍNDICE

Nota Editorial _____	4
<b>Cristiana Mesquita Reis</b>	
Questões de referência nominal num <i>corpus</i> de notícias _____	7
<b>Fábio Barcellos Granja</b>	
Contribuições de léxico-estatística para a genealogia dos crioulos de base portuguesa _____	22
<b>Joana Ribeiro da Silva Gomes</b>	
Análise de relações discursivas em Ted Talks com base na ISO 24617-8 _____	47
<b>Letícia Rodrigues Alberici</b>	
A emoção transmitida pelas palavras: escala de <i>valência, alerta e familiaridade</i> por indivíduos bilingues (L1 – Português Brasileiro e L2 – Inglês) _____	72
<b>Marina Araújo Salimon</b>	
Superdiversidade e <i>Codeswitching</i> entre português e inglês em falantes de português brasileiro como língua materna _____	99
<b>Rafaela Gomes de Moraes Fernandes da Gama</b>	
Análise da distância ideológica entre partidos políticos de um ponto de vista pragmático e discursivo _____	126

## NOTA EDITORIAL

O volume *Projetos de Pesquisa em Ciências da Linguagem N°1 2024* reúne um conjunto de cinco pequenos projetos de investigação elaborados no âmbito da unidade curricular de *Projeto de Pesquisa* do 2º semestre do 3º ano da licenciatura de Ciências da Linguagem, que tem como objetivo desenvolver nos estudantes competências no desempenho de tarefas de investigação científica ou de apoio à investigação científica, estimulando a sua participação em equipas e projetos de investigação. Simultaneamente, a UC confronta os estudantes com a aplicação prática dos conhecimentos fundamentais adquiridos ao longo da licenciatura, quer a nível teórico, quer a nível metodológico, contribuindo para o aprofundamento desses conhecimentos e competências.

O presente volume é o resultado da compilação destes trabalhos, elaborados por estudantes do ciclo de estudos, sob supervisão dos docentes de Linguística da Licenciatura.

No capítulo *Questões de referência nominal num corpus de notícias*, Cristiana Mesquita Reis desenvolve um estudo que visa contribuir para uma descrição mais detalhada dos mecanismos linguísticos utilizados em notícias de jornal de Português Europeu, mais concretamente quanto à forma como são referidos os participantes nos eventos descritos nas notícias, em particular com determinante artigo definido. Assim, este relatório contribui para o estudo da referência nominal, com resultados relativos à forma como a informação semântica dos participantes nos eventos e os mecanismos de determinação se interligam na criação de narrativas jornalísticas.

Fábio Barcellos Granja em *Contribuições de léxico-estatística para a genealogia dos crioulos de base portuguesa* estabelece uma rede genealógica de 18 crioulos de base portuguesa, com base em métodos léxico-estatísticos, que incluem as listas Swadesh transcritas foneticamente e anotadas semântica e sintaticamente com base em Bowerman (2007) para medir distâncias fonéticas. Estas foram calculadas com o software Cog (Dasipit 2015), sendo os dados ordenados para cada Crioulo, no sentido de verificar correlações com testes de Spearman. Os resultados confirmam classificações genealógicas qualitativas anteriores, formando cinco grupos: Alta Guiné, Golfo da Guiné, Norte da Índia, Sul da Índia e Sudeste Asiático. Também demonstram que, por vezes, mais do que fatores intralinguísticos, ecossistemas linguísticos socio-geográficos parecem influenciar a estabilidade lexical e o agrupamento genealógico dos crioulos.

Joana Ribeiro da Silva Gomes, no trabalho intitulado *Análise de relações discursivas em Ted Talks com base na ISO 24617-8*, investiga o uso das relações discursivas em textos de tipo narrativo em português europeu, avaliando se a proposta teórica da ISO 24617-8 é eficiente para uma anotação completa e abrangente dessas relações discursivas. Para concretizar estes objetivos, foram selecionadas três *Ted Talks* de carácter narrativo e anotadas as relações discursivas com base no

esquema de anotação da ISO 24617-8. Foram ainda identificados e classificados os conectores sempre que presentes e, simultaneamente, a aplicabilidade da ISO 24617-8 foi testada. A análise dos resultados permitiu concluir que, no *corpus* recolhido, as relações discursivas mais frequentes são *Expansion*, *Synchrony* e *Cause*, e que, na maioria das vezes, as relações não são marcadas por conectores. O estudo permitiu ainda concluir que a ISO 24617-8 apresenta alguns pontos negativos, nomeadamente no que diz respeito à indefinição de alguns conceitos de relações discursivas.

No projeto *A emoção transmitida pelas palavras: escala de valência, alerta e familiaridade por indivíduos bilíngues (L1 – Português Brasileiro e L2 – Inglês)*, Letícia Rodrigues Alberici propõe-se abordar o tema do processamento de palavras em segundas línguas (L2), que, embora tenha sido objeto de interesse constante da psicolinguística nas últimas décadas, tem patenteado conclusões muito variáveis quanto às respostas emocionais em L2. Assim, a autora realiza um estudo exploratório de classificação, para medir a forma como os leitores de Inglês como L2 respondem à *valência*, ao *alerta* e à *familiaridade* de 80 palavras do *corpus* elaborado a partir de fontes como o ANEW em Inglês (EN) e em Português do Brasil (PB), investigando como o efeito da proficiência em L2 afeta a forma como os leitores experienciam e classificam substantivos em PB e EN. Os dados indicam que falantes de PB como L1 que relataram maior proficiência em Inglês apresentaram respostas emocionais mais acuradas do que os falantes com menor proficiência. Os falantes com maior proficiência tendem a classificar as palavras negativas usando valores mais negativos; palavras positivas, usando valores mais positivos; palavras neutras, usando valores mais neutros, e, as palavras tabus são julgadas mais negativamente e com maior alerta do que os respetivos julgamentos na L2. Além disso, as diferenças de familiaridade refletiram a proficiência na L2: quanto maior a proficiência, maior a familiaridade com as palavras testadas em todos os tipos (negativas, positivas, tabus e neutras). Estes resultados realçam a importância da ecologia comportamental na aprendizagem de línguas, particularmente no que se refere ao processamento emocional de palavras.

Marina Araújo Salimon em *Superdiversidade e Codeswitching entre português e inglês em falantes de português brasileiro como língua materna* realiza uma breve síntese dos estudos sobre *codeswitching* e superdiversidade, seguida da análise de dados obtidos num inquérito, visando descrever a prática do *codeswitching* entre português e inglês em falantes nativos de português brasileiro. Observando dados quantitativos e depoimentos dos indivíduos, o estudo concluiu que pessoas jovens e mulheres realizam mais a prática do *codeswitching*, e que a proficiência em inglês não afeta necessariamente o envolvimento dos falantes com a mesma. Também conclui que, na amostra de indivíduos analisada, a prática do *codeswitching*, assim como a aprendizagem do inglês em si, está ligada ao contexto cultural do advento da internet e das redes sociais.

O projeto intitulado *Análise da distância ideológica entre partidos políticos de um ponto de*

*vista pragmático e discursivo*, da autoria de Rafaela Gomes de Moraes Fernandes da Gama, visa analisar as diferentes estratégias discursivas utilizadas no género *debate político*, mostrando que as diferenças sociopolíticas dos indivíduos, inseridos num determinado contexto social, resultam em fenómenos linguísticos relevantes no estudo pragmático-discursivo. Para tal, foram utilizados como objeto de estudo dois debates políticos televisivos entre representantes de partidos concorrentes às eleições legislativas portuguesas de 2024 que apresentam diferenças ideológicas, exemplificadas através de fenómenos linguísticos como Marcadores Discursivos, Cortesia e Descortesia, Formas de tratamento, Mecanismos de atenuação e intensificação e Atos de fala. Estas estratégias interligam-se, também, com a construção do indivíduo e da sociedade através do poder persuasivo e da ação da linguagem.

**Maria Alexandra de Araújo Guedes Pinto,**

Diretora da Licenciatura em Ciências da Linguagem

Faculdade de Letras da Universidade do Porto

Centro de Linguística da Universidade do Porto

# QUESTÕES DE REFERÊNCIA NOMINAL NUM *CORPUS* DE NOTÍCIAS

Cristiana Mesquita Reis

Orientado pelo Professor Doutor António Leal

## **Resumo**

Este projeto visa contribuir para uma descrição mais detalhada dos mecanismos linguísticos utilizados em notícias de jornal de Português Europeu, mais concretamente quanto à forma como são referidos os participantes nos eventos descritos nas notícias, em particular com determinante artigo definido. Assim, este relatório contribui para o estudo da referência nominal, com resultados relativos à forma como a informação semântica dos participantes nos eventos e os mecanismos de determinação se interligam na criação de narrativas jornalísticas.

Palavras Chave: determinante artigo definido; referência nominal; semântica; Português Europeu; notícias de jornal.

## **Abstract**

This project aims to contribute to a more detailed description of the linguistic mechanisms used in European Portuguese newspaper articles, more specifically about how the participants in the events described in the news are referred to, in particular with a definite article determinant. Thus, this report contributes to the study of nominal reference, with results relating to the way in which the semantic information of participants in events and the mechanisms of determination are interconnected in the creation of journalistic narratives.

Keywords: determiners articles; nominal reference; semantics; European Portuguese; news article writing.

## **1. Introdução**

A referência nominal é um elemento linguístico de grande relevo no que toca à produção de textos jornalísticos, pois é uma estratégia que tem impacto na precisão, clareza e coesão das informações narradas. No contexto da língua portuguesa, a análise de estratégias de referência nominal em notícias assume uma pertinência significativa para compreender como é que os mecanismos linguísticos são empregues na construção de narrativas jornalísticas.

Este projeto de pesquisa tem como objetivo principal explorar questões de referência nominal num corpus de 50 notícias produzidas pela agência Lusa. Através da anotação e classificação de sintagmas nominais introduzidos por artigos definidos, o estudo visa identificar padrões, usos e funções dos determinantes artigos definidos no contexto de produções textuais jornalísticas reais.

Ao longo deste relatório serão abordados conceitos teóricos fundamentais, desde a natureza do sintagma nominal até à especificidade dos determinantes artigos definidos, proporcionando uma base para a análise dos dados coletados. A apresentação e análise dos resultados obtidos permitirão uma compreensão mais aprofundada das estratégias de referência nominal utilizadas nas notícias em Português Europeu, contribuindo para a produção de textos jornalísticos mais eficazes e coerentes.

Assim, este trabalho será iniciado por um enquadramento teórico, onde serão apresentados conteúdos conceituais relevantes para a análise do objeto de estudo, seguido de uma apresentação do corpus e explicação das metodologias utilizadas. Posteriormente, serão debatidos alguns problemas que surgiram ao longo do processo de trabalho, acompanhados, depois, de uma exposição e análise dos resultados obtidos neste projeto de pesquisa. Com base nestes pontos, o trabalho culminará numa reflexão final que pretende sintetizar todo o processo aqui apresentado.

Espera-se que este estudo não só enriqueça o conhecimento sobre a referência nominal em contexto de jornalismo de redação, como também possa servir como uma ferramenta valiosa para a construção de manuais e a formação de profissionais da área, promovendo a qualidade da comunicação escrita e das informações veiculadas pela imprensa.

## 2. Enquadramento Teórico

Antes de iniciar uma análise mais prática sobre as questões de referência nominal, é necessário introduzir uma contextualização teórica que permita aprofundar conhecimentos acerca dos conteúdos que serão alvo de investigação. Como tal, esta síntese, baseada em Peres (2013) e Miguel e Raposo (2013), focar-se-á, inicialmente, numa análise teórica mais abrangente, começando por explorar o sintagma nominal, até uma perspetiva mais restrita sobre os conteúdos que serão pertinentes para a realização deste projeto, abordando os sintagmas nominais determinados definidos e culminando numa análise mais estreita sobre os determinantes artigos definidos.

Em primeiro lugar, é necessário compreender que o universo do discurso diz respeito ao conjunto de todas as entidades que podem ser nomeadas numa determinada língua.

Dentro deste universo, a semântica nominal explora a classificação destas entidades e as suas representações linguísticas, dividindo-as em quatro principais grupos: entidades contáveis ou não contáveis; indivíduos-grupos ou totalidades-partes; espécies-espécimes; entidades reais ou virtuais.

Quanto ao primeiro grupo, a distinção entre entidades contáveis e não contáveis considera entidades percebidas como independentes umas das outras, permitindo que sejam contadas, isto é, que se conheça o número cardinal de membros destas mesmas entidades (e.g., *pessoa*), ao contrário de entidades que não são passíveis de contabilização exata, sendo, por isso, não contáveis (e.g., *amizade*).

Relativamente ao segundo grupo, este denota a distinção entre entidades individuais, que se referem a entidades singulares distintas e independentes (e.g., *pessoa*), e entidades coletivas, que se referem ao conjunto de indivíduos (e.g., *multidão*). Este contraste difere da relação entre entidades totais e parciais na medida em que as totalidades são vistas como unidades completas (e.g., *computador*) e as partes como as entidades que compõem essas mesmas unidades (e.g., *processador*, *ecrã*, *teclado*).

A relação entre espécies e espécimes assume as espécies como sendo categorias amplas e concretas que englobam características comuns (e.g., *canário*), constituídas por casos particulares dentro da própria espécie, os espécimes (e.g., *canário da gloster*, *canário topete alemão*, *canário lizard*).

Por último, o quarto grupo distingue entidades com base na natureza da sua existência. As entidades reais possuem uma presença concreta no mundo físico (e.g., *Comprei este livro naquela loja.*), ao invés das entidades virtuais, que apenas existem no domínio da imaginação ou da previsão (e.g., *A coruja do Harry Potter trouxe-lhe uma carta.*).

É, também, importante ter em conta que o valor semântico de um sintagma nominal está veiculado à sua constituição interna e a fatores externos, o que permite classificá-los quanto à forma e quanto ao seu valor.

Quanto à forma, os sintagmas nominais podem ser determinados (quando introduzidos por determinantes), quantificados (quando incorporam elementos de quantificação, como quantificadores ou outras expressões que indiquem quantidades, tal como, por exemplo, advérbios de quantidade) ou reduzidos (quando não apresentam explicitamente um determinante ou elemento de quantificação, mantendo, na mesma, a sua estrutura nominal). Para além disso, os sintagmas podem ser mistos, isto é, quando, por meio de coordenação, é possível combinar vários destes tipos de SN.

Quanto ao valor, ou seja, quanto à sua função, os sintagmas nominais podem assumir três valores distintos: referencial, quantificacional e predicativo.

O valor referencial de um SN diz respeito à sua capacidade de identificar ou referir-se a uma ou mais entidades. Apesar de poder coexistir com outras funções, o valor referencial é, habitualmente, a principal função de um sintagma nominal, podendo ter valor particular (i.e., apontando para entidades específicas, como indivíduos, grupos, totalidades ou partes) ou valor genérico (i.e., não apontando para entidades específicas, mas para uma classe ou tipo de entidade). Assim, o valor referencial de um sintagma nominal assume forma por meio de nomes próprios – com ou sem artigo –, sintagmas nominais definidos ou indefinidos e pronomes (visto estes mesmos substituírem nomes ou SN, apontando para as mesmas entidades no contexto da comunicação).

Para além disso, os sintagmas nominais podem assumir valor predicativo, tipicamente quando esse sintagma é combinado com um verbo copulativo, expressando uma propriedade ou atributo associado à entidade referida pelo SN.

Já o valor quantificacional de um sintagma nominal permite quantificar ou especificar a quantidade das entidades que o mesmo representa.

Tendo em conta que este projeto dará mais ênfase ao determinante artigo, mais concretamente, ao definido, é necessário esclarecer como é que este se relaciona com os sintagmas nominais.

Os sintagmas nominais determinados podem ser classificados como sendo definidos - quando introduzidos por *o, a, os, as* - ou indefinidos – quando introduzidos por *um, uma, uns, umas* -, tendo em conta que o definido implica uma quantificação universal e o indefinido uma quantificação existencial. Isto significa que os SN determinados definidos exigem algum tipo de conhecimento prévio das entidades a que se referem. Estas características associam três possíveis tipos de classificação aos sintagmas nominais definidos: SN definido epistémico (i.e., quando se refere a entidades que são únicas e facilmente identificáveis no sistema de conhecimento do mundo em questão), SN definido textual (i.e., quando se refere a entidades que já foram previamente introduzidas no contexto textual por, por exemplo, sintagmas introduzidos pelos indefinidos) e SN definido situacional ou deítico (i.e., que se refere a entidades presentes no contexto situacional da enunciação).

É, também, importante ter em conta que os sintagmas nominais definidos podem referir não só entidades particulares (e.g., *O telemóvel partiu*), como também espécies (e.g., *O telemóvel é / Os telemóveis são uma invenção que alterou a forma como as pessoas se comunicam.*) ou tipos (e.g., *Não consigo viver sem o telemóvel.*). No entanto, é necessário ter em conta que a qualidade de ser genérico não é intrínseca a SN determinados, mas é antes adquirida em contexto.

De facto, o sistema de determinantes permite veicular as interpretações dos SN como tendo valor referencial (i.e., quando designa uma entidade particular no mundo) ou não referencial (i.e., quando designa uma entidade não identificada pelo falante ou quando designa uma propriedade que se aplica a uma entidade ou grupo de entidades). Como explicitado anteriormente, isto assume que o determinante artigo definido é usado quando o falante indica que tem em mente uma entidade única e específica dentro do contexto da comunicação, presumindo que o ouvinte/leitor seja capaz de a identificar (e.g., *O cão mordeu-me*). Isto pressupõe que a menção da entidade não seja nova, mas já conhecida ou reconhecível dentro da enunciação, ao contrário do que acontece com o determinante artigo indefinido, que assume uma referência mais abstrata (e.g., *Um cão mordeu-me*).

Assim, estão associadas três propriedades (intrínsecas) ao artigo definido: a condição de existência, de unicidade e de identificação. Estas condições pressupõem que, num determinado contexto situacional, existe uma e apenas uma entidade a ser tida em conta no momento da enunciação, seja por estar saliente na consciência dos interlocutores ou por ter sido mencionada anteriormente, o que permite que a mesma seja imediata e facilmente identificável. O determinante artigo definido assume, ainda, a condição da inclusividade, que considera, essencialmente, os artigos definidos plurais *os* e *as*. Esta propriedade esclarece que um sintagma nominal definido plural tem como referente o conjunto mais inclusivo das entidades assumidas pelo grupo nominal, partilhando um valor de quantificação universal com o quantificador *todos* (e.g., *Dá-me os marcadores vermelhos; Dá-me todos os marcadores vermelhos.*)

Tendo em conta que os determinantes não têm uma interpretação pré-determinada (e.g., específica/não específica), mas antes introduzem uma leitura desse tipo no sintagma nominal que especificam, é possível concluir que o artigo definido, aliado ao grupo nominal, que fornece uma descrição que determina o referente, possui como principal função assinalar ao ouvinte/leitor a existência de uma única entidade que vai ao encontro dessa mesma descrição. A este uso do artigo definido é dado o nome de 'uso (valor) epistémico'. Quanto maior a adição de informação a essas expressões, por meio de, por exemplo, complementos ou modificadores do nome, mais específica é a identificação da entidade referida, cabendo o papel de especificação do sintagma nominal à estrutura sintática da frase e não propriamente ao determinante artigo definido.

Para além do uso epistémico, o artigo definido pode ser aplicado de acordo com a relação que os sintagmas que introduz estabelecem com o seu referente. Por um lado, pode assumir um uso (valor) situacional, quando a entidade referida pelo SN é exclusivamente identificada pelo contexto da enunciação (e.g., *Chegas-me a mala?*). Por outro lado, assume um uso (valor) anafórico quando, por meio de anáfora, apresenta uma entidade que foi anteriormente introduzida linguisticamente (e.g., *Hans Zimmer deve ganhar mais um prémio para o ano. O compositor compõe para muitos filmes.*). Como referido anteriormente, o determinante artigo definido pode, também, denotar um uso (valor) genérico quando se refere a uma espécie (e.g. *O dodó está extinto.*).

De acordo com Peres (2013), o determinante artigo definido pode, ainda, possuir um uso (valor) possessivo quando introduz sintagmas nominais que se encontram numa relação semântica possessiva, por meio de, por exemplo, meronímia, holonímia, hiperonímia ou hiponímia (e.g., *O cão mordeu a cauda e, depois, lambeu as patas.*), o que corresponde, de acordo com outros autores, a casos de anáfora associativa.

De facto, o determinante artigo definido como introdutor de sintagmas nominais possui uma pluralidade de funções e possíveis usos. A revisão da literatura feita nesta secção permite compreender esta classe de palavras, criando condições para a análise de questões de referenciação nominal, mais concretamente, acerca dos mecanismos linguísticos de determinação usados na referência nominal num *corpus* de notícias.

### 3. *Corpus e Metodologia*

Este projeto assenta na análise de uma seleção de 50 notícias produzidas pela agência Lusa disponibilizadas no programa *Brat* (Stenetorp et al., 2012). Neste programa de anotação, foram registados todos os sintagmas nominais introduzidos por artigos definidos, sendo levados em consideração os sintagmas referentes a participantes (*participants*) e os que denotam eventos (*events*), embora o foco da análise posterior se tenha centrado nos primeiros. O esquema de anotação usado baseia-se na norma ISO 24617 (parte 6 e parte 9), tendo sido construído no âmbito do projeto *Text2Story* (cf. Silvano et al., 2021; Leal et al., 2022) - <https://text2story.inesctec.pt/>.

Em cada sintagma definido classificado como *participant* foram feitas as respetivas anotações no que toca ao núcleo – *Lexical\_Head*. De seguida, anotou-se o *Individuation\_Domain* como sendo *set* (nome coletivo), *individual* (nome contável) ou *mass* (nome não contável). Quanto ao domínio do SN, foi possível denominar o *Participant\_Type\_Domain* de acordo com 18 opções disponíveis, sendo elas *Per* (indivíduos humanos reais ou fictícios), *Org* (entidades com uma estrutura organizacional), *Obj* (objetos tangíveis feitos pelo Homem), *Nat* (entidades naturais, não feitas pelo Homem), *Fac* (edifícios), *Veh* (veículos), *Path* (localização em forma de percurso), *Pl* (referências locativas em variadas formas – *water, celestial, mountain, civil, country, mount\_range, capital, region, state*), *Loc* (lugares que não se enquadram nas formas disponibilizadas por *Pl*) e *Other* (entidades que não se enquadram em nenhuma das outras categorias). Por último, analisou-se o *Involvement* que, no caso do SN definido, poderia ser classificado como *1*, *>1* ou *All*. Posteriormente, foram criadas ligações entre as expressões anotadas, podendo estas ligações serem de relação de identidade, i.e., trata-se da mesma entidade extralinguística (*OLINK\_objIdentity*), de parte-todo (*OLINK\_partOf*) ou de membro-grupo (*OLINK\_memberOf*).

Para além disso, foram também considerados os sintagmas nominais introduzidos por artigo indefinido ou sem qualquer tipo de determinante, para efeitos de determinação de mecanismos anafóricos.

Em simultâneo com esta análise foram anotados em Excel os parâmetros de denotação do SN (referencial – genérico ou particular – ou predicativo), de licenciamento (epistémico, textual – e respetivo tipo de anáfora – ou situacional) e, em casos de SN introduzido por artigo definido plural, o valor quantificacional (para considerar os casos de ambiguidade - totalidade ou não totalidade).

#### 4. Limitações na anotação

Ao longo do processo de anotação surgiram alguns problemas que dificultaram o decurso deste trabalho.

Em primeiro lugar, é relevante destacar o leque limitado de ligações entre entidades que o esquema de anotação oferece. Durante a anotação dos SN definidos, surgiram muitos sintagmas que possibilitavam ligações de, por exemplo, *bem-posse* (e.g. notícia 1: “um bar” – “o proprietário”; notícia 21: “uma viatura” – “o condutor”), *relação entre* (e.g. notícia 4: “um homem” – “a ex-companheira”; notícia 7: “a mãe” – “as crianças”; notícia 31: “a vítima” – “o companheiro”) ou *localizado em* (e.g. notícia 19: “a residência da vítima” – “o distrito de Leiria”; notícia 34: “a sua casa” – “a província de Buenos Aires”). Na ausência de soluções de anotação, pode-se considerar a incompletude de possíveis conclusões associadas a ligações entre as entidades anotadas.

Para além disso, surgiram algumas questões ontológicas relativamente à classificação dos domínios dos participantes (*Participant\_Type\_Domain*), no que toca, por exemplo, a referências ao corpo das vítimas dos eventos relatados. Assim sendo, foram consideradas várias hipóteses de anotação de alguns SN como “os corpos” (notícia 14), sendo elas *Per* ou *Nat*. De forma a inferir uma decisão final, recorreu-se ao *Dicionário da Língua Portuguesa da Academia das Ciências de Lisboa* que apresenta como definição do nome “corpo” 14 entradas, das quais se destacam como relevantes para esta matéria: “1. parte material dos seres animados; o organismo humano ou animal, formado por uma série de sistemas e aparelhos”, “2. parte material do corpo humano por oposição ao espírito” e “3. o ser humano morto”. Tendo em conta estes conceitos, optou-se pela classificação final *Nat*, assumindo “corpo” como sendo um elemento constitutivo de *Per*, mas não a mesma por inteiro, que pressupõe a existência de uma identidade, o “espírito”, que existe apenas em vida.

Um debate semelhante surgiu relativamente a expressões como “a nota”. Segundo o mesmo dicionário, o nome “nota” corresponde a: “2. apontamento ou anotação que se toma sobre uma matéria ou acontecimento”, “3. papel onde são registados esses apontamentos”, “4. indicação colocada na parte inferior de uma página, na margem, no final de um documento ou de uma parte de um documento com o fim de precisar, completar, comentar ou esclarecer uma palavra ou uma determinada passagem do texto” e “5. comunicação sucinta sobre um determinado assunto”. Tendo em conta estas definições da palavra, é possível considerar, como possíveis classificações de domínio deste tipo de SN, *Obj* (no que diz respeito às entradas 2 e 3) ou *Other* (no que toca às entradas 4 e 5). Assim sendo, não se optou por uma única anotação geral, mas sim por uma análise situacional de cada caso em que expressões semelhantes surgiram, analisando se, em cada contexto, “a nota” poderia fazer referência ao objeto no qual foi registada informação relevante para o evento em causa (*Obj* – como na notícia 35) ou à informação presente nesse mesmo objeto (*Other* – na notícia 32).

Ainda relacionado com a atribuição de classificações aos SN *Participant*, surgiram várias referências numéricas que não se enquadravam em nenhuma oferta de classificação de domínios proposta pelo esquema, como é o caso de, por exemplo, “o valor máximo de 114 mil euros” e “o montante líquido de 5.400,17 euros” (notícia 46) e “os 35 e 40 anos” (notícia 50). O mesmo ocorreu com entidades não materiais como “o Facebook do clube” (notícia 26). Assim, optou-se por lhes ser atribuída a classificação *Other*.

Para além disso, surgiram problemas associados a expressões lexicalizadas, que não foram consideradas para efeitos de análise deste trabalho, como é o caso de “o largo de Olhão” (notícia 29). Neste caso, a palavra “largo” não surge como adjetivo ou nome, mas sim inserido na expressão “ao largo”, que poderia ser substituída por, por exemplo, “nas redondezas de Olhão”. Todos os casos semelhantes não foram anotados.

## 5. Apresentação e análise dos resultados

Na totalidade das 50 notícias analisadas, estão presentes um total de 9391 palavras, o que gera uma média de 187,82 palavras por notícia.

No que toca à anotação de sintagmas nominais introduzidos por artigo definido, foram assinalados 139 como eventos (*Event*) (1) e 942 como participantes (*Participant*) (2), sendo 83 desses participantes não introduzidos por artigo definido e contabilizados apenas para efeitos de licenciamento textual.

(1) *O trânsito chegou a estar cortado* (notícia 1)

(2) *disse à Lusa fonte da GNR* (notícia 1).

Relativamente aos atributos dos SN *Participant*, é possível destacar os seguintes resultados associados à individuação das expressões:

<i>Indivuduation_Domain</i>	<u>Frequência Absoluta</u>	<u>Frequência Relativa</u>
<i>Individual</i>	827	96,27%
<i>Set</i>	27	3,14%
<i>Mass</i>	5	0,58%

**Tabela 1.** Frequência absoluta e relativa das anotações quanto ao *Indivuduation\_Domain*.

No que diz respeito ao *Involvement*, tendo em conta que foram apenas considerados os SN introduzidos por artigo definido, seja singular ou plural, foram registadas as seguintes ocorrências:

<i>Involvement</i>	<u>Frequência Absoluta</u>	<u>Frequência Relativa</u>
<i>1</i>	717	83,47%
<i>&gt;1</i>	61	7,10%
<i>All</i>	81	9,43%
Leitura ambígua: <i>&gt;1 e All</i>	56	6,52%

**Tabela 2.** Frequência absoluta e relativa das anotações quanto ao *Involvement*.

É importante ressaltar que, tal como confirmado na *Tabela 2*, dos 142 SN anotados como tendo *Involvement >1* ou *All*, 56 foram assinalados como oferecendo leitura ambígua, isto é, como permitindo considerar ambas as leituras de envolvimento numérico das entidades – (3) e (4).

(3) No âmbito de uma ação de policiamento, os militares detetaram o suspeito num terreno agrícola a cultivar canábis. (notícia 35).

(4) A mulher foi detida por militares do Posto Territorial de Mértola (...). Os militares efetuaram uma busca à viatura e uma revista à mulher (notícia 47).

Em ambos os exemplos, a expressão “os militares” permite que se façam, simultaneamente, as leituras de que atuaram todos os militares mobilizados para o evento ou apenas uma parte deles.

Associado ao tipo de domínio de cada *Participant*, obtiveram-se os seguintes dados:

<i>Participant_Type_Domain</i>	<u>Frequência Absoluta</u>	<u>Frequência Relativa</u>
<i>Per</i>	237	27,59%
<i>Org</i>	275	32,01%
<i>Obj</i>	31	3,61%
<i>Loc</i>	30	3,49%
<i>Nat</i>	29	3,38%
<i>Fac</i>	54	6,29%
<i>Veh</i>	22	2,56%
<i>Path</i>	12	1,40%
<i>Pl_water</i>	16	1,86%
<i>Pl_celestial</i>	0	0%
<i>Pl_mountain</i>	0	0%
<i>Pl_civil</i>	71	8,27%
<i>Pl_country</i>	10	1,16%

**Tabela 3.** Frequência absoluta e relativa das anotações quanto ao *Participant\_Type\_Domain*.

<i>Participant_Type_Domain</i>	<u>Frequência Absoluta</u>	<u>Frequência Relativa</u>
<i>Pl_mount_range</i>	0	0%
<i>Pl_capital</i>	4	0,47%
<i>Pl_region</i>	9	1,05%
<i>Pl_state</i>	6	0,70%
<i>Other</i>	53	6,17%

**Tabela 3 (continuação).** Frequência absoluta e relativa das anotações quanto ao *Participant\_Type\_Domain*.

Estes resultados revelam padrões significativos no que diz respeito à predominância de referências a entidades individuais e organizacionais. A grande ocorrência deste tipo de entidades pode ser justificável pelo facto de as notícias fazerem referência a eventos que denotam crimes, crimes estes que pressupõe a existência de pessoas ou organizações que os cometeram e entidades organizacionais que trataram de os resolver (e.g., GNR, PSP, as autoridades locais, os Bombeiros, etc.).

Quanto às ligações (*OLINK*), foram registadas, na sua totalidade, 42 ligações, estando estas distribuídas pelos seguintes sub-tipos:

<b>Ligação</b>	<b>Frequência Absoluta</b>
<i>OLINK_partOf</i>	23
<i>OLINK_objIdentity</i>	7
<i>OLINK_memberOf</i>	12

**Tabela 4.** Frequência absoluta de ligações *OLINK* entre as expressões anotadas.

Cada um destes sub-tipos está associado a uma relação de identidade, no sentido de que cada expressão associada denota a mesma entidade extralinguística (*OLINK\_objIdentity*) (5), a uma relação de parte-todo (*OLINK\_partOf*) (6) ou de membro-grupo (*OLINK\_memberOf*) (7).

(5) *o homem; a vítima* (notícia 19).

(6) *o município de Camacupa; o Bié* (notícia 8).

(7) *os militares; a GNR* (notícia 37).

No que toca à denotação dos SN definidos classificados como *Participant*, foram consideradas as expressões completas sem divisão das mesmas entre os vários SN que as constituem (i.e. considerou-se a expressão *a unidade de Vila Real do Centro Hospitalar de Trás-os-Montes e Alto Douro* como um todo, não fazendo a separação dos vários SN que fazem parte da mesma). De um total de 554 casos, todos possuíam denotação particular à exceção de dois casos – (8) e (9) - que possuem denotação genérica. Quanto ao licenciamento, não se verificou nenhum situacional, sendo todos os SN licenciados por leitura epistémica (10), à exceção de 216 que surgem por meio de licenciamento textual. Destes 216, 110 são utilizados por meio de anáfora por substituição (11), 59 por repetição (12) e 47 por anáfora associativa (13).

(8) (...) *agradecemos aos profissionais de saúde, e a todos que endereçaram mensagens, clubes, atletas, instituições e adeptos de todo o mundo, que se juntaram a nós numa gigantesca demonstração de apreço pela vida humana* (notícia 26).

(9) *A fiscalização marítima moçambicana apreendeu 1.100 quilos de caranguejo, no centro do país, em menos de uma semana, capturado na "época de veda", quando é proibido apanhar a espécie* (...) (notícia 13).

(10) *o alto minho; a GNR* (notícia 50).

(11) *a mulher – a suspeita* (notícia 47).

(12) *o arguido – o arguido* (notícia 46).

(13) *a GNR – o comando territorial de Faro* (notícia 1).

			<u>Freq. Absoluta</u>	<u>Freq. Relativa</u>
<b>Denotação</b>	Referencial	Genérica	2	0,36%
		Particular	552	99,64%
	Predicativa		0	0%

**Tabela 5.** Frequência absoluta e relativa da classificação denotativa dos SN.

			<u>Freq. Absoluta</u>	<u>Freq. Relativa</u>
<b>Licenciamento</b>	Epistémico		338	61,01%
	Textual	Anáfora Associativa	47	8,48%
		Anáfora por repetição	59	10,65%
		Anáfora por substituição	110	19,86%
	Situacional		0	0%

**Tabela 6.** Frequência absoluta e relativa da classificação do licenciamento dos SN.

Os dados apresentados nas tabelas 5 e 6 mostram que a maioria dos SN possui um valor referencial particular, sendo estes essencialmente utilizados, na sua maioria, com recurso a um licenciamento epistémico, mas com uma ocorrência igualmente relevante por meio de licenciamento textual. Estes resultados vão ao encontro da objetividade requerida em textos jornalísticos, que devem ser claros, de modo a facilitar a leitura e compreensão do leitor. São estes mecanismos maioritariamente utilizados também objetivos e simples.

## **6. Reflexões finais**

Este trabalho permitiu uma compreensão aprofundada dos mecanismos linguísticos utilizados na construção de narrativas jornalísticas em Português Europeu, possibilitando a identificação de padrões claros no uso e função do determinante artigo definido na imprensa.

Os resultados obtidos demonstraram que a maioria dos sintagmas nominais analisados possuem um valor referencial particular, o que reflete a necessidade de clareza e especificidade na identificação das entidades mencionadas nas notícias. Para além disso, a predominância de licenciamento epistémico e textual evidencia que, de facto, os textos jornalísticos dependem, muitas vezes, do conhecimento prévio das entidades e da continuidade textual para garantir a compreensão do leitor. Estes mecanismos de referência nominal permitem construir uma narrativa coesa, onde cada entidade mencionada é claramente identificável e contextualizada, característica esta associada ao uso do determinante artigo definido aqui colocado em foco.

É, também, relevante referir que a análise das ligações entre as expressões anotadas destaca a importância das relações entre elas, reforçando a necessidade de manter uma rede coesa de referências ao longo das narrativas, coesão esta impulsionada pelo recurso a referências anafóricas.

Os problemas encontrados durante o processo de anotação refletem a complexidade deste tipo de análise e justificam a necessidade de ferramentas mais robustas para a anotação dos mais variados elementos linguísticos. No entanto, são também estas dificuldades que revelam a heterogeneidade marcante de mecanismos linguísticos de referência nominal na língua portuguesa.

## BIBLIOGRAFIA

- ISO-24617-6. (2016). Language resource management- semantic annotation framework (semaf) - part 6:principles of semantic annotation (semaf principles). Standard, Geneva, CH.
- ISO-24617-9. (2019). Language resource management- semantic annotation framework (semaf) - part 9: Reference annotation framework (raf). Standard, Geneva, CH.
- Leal, António; Silvano, Purificação; Amorim, Evelin; Cantante, Inês; Silva, Fátima; Jorge, Alípio & Campos, Ricardo (2022). The place of ISO-Space in the Text2Story multilayered annotation scheme. In Proceedings of The Eighteenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation. [https://sigsem.uvt.nl/isa18/ISA-18\\_24\\_Paper.pdf](https://sigsem.uvt.nl/isa18/ISA-18_24_Paper.pdf)
- Miguel, Matilde & Raposo, Eduardo Paiva (2013). Determinantes. In Eduardo Paiva Raposo, Fernanda Nascimento, Maria Antónia Mota, Luísa Segura & Amália Mendes (orgs.), *Gramática do Português*, Vol. 1, 819-882. Lisboa: Fundação Calouste Gulbenkian.
- Peres, João Andrade (2013). Semântica do sintagma nominal. In Eduardo Paiva Raposo, Fernanda Nascimento, Maria Antónia Mota, Luísa Segura & Amália Mendes (orgs.), *Gramática do Português*, Vol. 1, 735-818. Lisboa: Fundação Calouste Gulbenkian.
- Silvano, Purificação; Leal, António; Silva, Fátima; Cantante, Inês; Oliveira, Fátima & Jorge, Alípio (2021) Developing a multilayer semantic annotation scheme based on ISO standards for the visualization of a newswire corpus. In Proceedings of The Seventeenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation. Groningen, Netherlands. <https://sigsem.uvt.nl/isa17/SilvanoEtAl.pdf>
- Stenetorp, P., Pyysalo, S., Topic, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). Brat: a Web-based Tool for NLP-Assisted Text Annotation. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 102–107, Avignon, France, April. Association for Computational Linguistics.

# CONTRIBUIÇÕES DA LÉXICO-ESTATÍSTICA PARA A GENEALOGIA DOS CRIoulos DE BASE PORTUGUESA

Fábio Barcellos Granja

Orientado Pelo Professor Doutor Carlos Rogério Sousa e Silva

## Resumo

Neste artigo, estabelecemos uma rede genealógica de 18 crioulos de base portuguesa utilizando métodos léxico-estatísticos. Utilizamos listas Swadesh transcritas foneticamente e anotadas semântica e sintaticamente com base em Bower (2007) para medir distâncias fonéticas. Estas foram calculadas com o software Cog (Daspit 2015); os resultados foram ordenados para cada Crioulo para verificar correlações com testes de Spearman. Os resultados confirmam classificações genealógicas qualitativas anteriores, formando cinco grupos: Alta Guiné, Golfo da Guiné, Norte da Índia, Sul da Índia e Sudeste Asiático. Também demonstram que ecossistemas linguísticos socio-geográficos parecem influenciar a estabilidade lexical e o agrupamento genealógico, por vezes mais do que factores intralinguísticos.

Palavras Chave: Crioulos, Estabilidade Lexical, Tipologia, Fonologia, Semântica.

## Abstract

In this paper, we aim to establish a genealogical network of 18 Portuguese-based Creoles using lexico-statistical methods. We have used phonetically transcribed Swadesh lists and further annotated them semantically and syntactically based on Bower (2007) to measure phonetic distances. They were calculated using Cog software (Daspit 2015), and the results were ordered for each Creole so we could check correlations using Spearman's tests. The results confirm previous qualitative genealogical classifications into five groups: Upper Guinea, Gulf of Guinea, Northern India, Southern India, and Southeast Asia. They highlight that socio-geographical language ecosystems seem to influence lexical stability and genealogical grouping, sometimes more than intra-linguistic factors.

Keywords: Creoles, Lexical Stability, Typology, Phonology, Semantics.

## 1 Introdução

Desde o início do estudo científico das línguas Crioulas, não parece haver consenso sobre se são ou não tipologicamente diferentes das não-Crioulas: por um lado, autores como Bickerton (1981) e Bakker et al. (2011) defendem que formam uma classe distinta de línguas; por outro lado, outros consideram-nas como descendentes da língua lexicalizadora, estando o seu desenvolvimento relacionado com a Aquisição de Segunda Língua (DeGraff 2001, Mufwene 2001).

Avanços tecnológicos recentes incrementaram o estudo destas línguas, permitindo a utilização de métodos computacionais para investigar agrupamentos e relações entre elas (Silva 2023). Adicionalmente, observamos que, no caso dos Crioulos de base portuguesa, há uma carência de estudos genealógicos que os analisem numa perspetiva quantitativa (Clements 2021).

Neste estudo, pretendemos contribuir para a análise genealógica deste grupo linguístico, aplicando métodos léxico-estatísticos (Starostin 2016) a uma lista Swadesh (Swadesh 1955) multilíngue com cognatos foneticamente transcritos e alinhados em 18 Crioulos de base portuguesa e a língua lexicalizadora. Em primeiro lugar, atribuímos cada a glosa da lista Swadeshum dos 19 domínios sintático-semânticos propostos por Bower (2007). Com base nestes domínios, calculámos o grau de proximidade fonética entre as palavras crioulas pertencentes a cada domínio sintático e semântico e os correspondentes cognatos portugueses através de uma análise assistida por computador. Assim, pudemos avaliar a ordem de estabilidade global de cada Crioulo, ou seja, quais os domínios sintático-semânticos que apresentavam maior grau de semelhança fonética com a língua lexicalizadora e quais os que se encontravam mais distantes desta. Com isto, pretendemos responder às seguintes questões de investigação:

- Os Crioulos de base portuguesa seguem caminhos diferentes de diversificação linguística?
- Como podemos medir e representar as distâncias entre estas línguas? (Daspit 2015, Starostin 2016, Daval-Markussen, Bøegh & Bakker 2017)
- Quais são os factores linguísticos e não linguísticos que desempenham um papel na diversificação linguística destes Crioulos? (Mufwene 2001, Silva 2023)

### 1.1 Classificações anteriores dos Crioulos de base portuguesa

Tendo em conta os factores tipológicos clássicos (Chaudenson 1979), os Crioulos de base portuguesa podem ser analisados, entre outros (Schuchardt 1889, Cardoso 2020):

- numa perspetiva sócio-histórica (e.g., *Crioulos de plantação, maroon e de forte*)

- numa perspetiva demográfica (e.g., *Crioulos endógenos e exógenos*)
- de uma perspetiva substratista (e.g., *substrato Indo-Europeu, Banto ou Malaio*)

Apesar destas diferentes abordagens tipológicas, estes Crioulos são tradicionalmente agrupados geograficamente (Cardoso 2020), seguindo os grupos representados na Tabela 1 (adaptado de Cardoso 2020, Clements 2021, Silva 2023).

África	Alta Guiné	<i>Crioulo Cabo-verdiano, Kriyol da Guiné Bissau, Casamancês</i>
	Golfo da Guiné	<i>Fá d’Ambô, Principense, Santomé, Angolar</i>
Ásia	Norte da Índia	<i>Daman, Diu, Korlai</i>
	Sul da Índia	<i>Kannur, Sri Lanka</i>
	Sudeste Asiático	<i>Macau, Malacca</i>

Tabela 1: Divisão geográfica dos Crioulos de base portuguesa

As línguas crioulas escolhidas para serem analisadas ao longo deste estudo foram seleccionadas tendo em conta dois critérios principais:

- Representam os cinco subgrupos de Crioulos de base portuguesa (Clements 2021);
- Existe material descritivo adequado sobre eles, nomeadamente uma lista Swadesh transcrita foneticamente (Swadesh 1955).

Assim, o primeiro subgrupo, o da Alta Guiné, engloba: o Kriyol da Guiné-Bissau, o Casamancês e o Crioulo Cabo-Verdiano e as suas variedades. Abrangendo os atuais territórios de Guiné-Bissau, Senegal e Cabo Verde, estes foram os primeiros Crioulos de base portuguesa a surgir (Biagui 2012), mantendo-se geralmente em contacto com a língua lexicalizadora durante três a cinco séculos (Silva 2023). Enquanto os estudos que analisam as diferentes variedades destes Crioulos são abundantes para o Cabo-Verdiano (Baptista 2019, Moreira 2012, Lopes et al. 2014), são ainda muito recentes para o Kriyol da Guiné-Bissau (Mendes 2023), e escassos para o Casamancês. Por esta razão, optámos também por incluir no nosso estudo as variedades de Barlavento e Sotavento do Crioulo Cabo-Verdiano: Santo Antão e São Vicente representam as primeiras, e Santiago, Fogo e Brava as segundas. Os Crioulos da Alta Guiné estão representados nas Figuras 1 e 2.

O grupo seguinte era constituído pelos Crioulos do Golfo da Guiné: Principense, Santomé, Angolar e Fá d’Ambô. Falados em São Tomé e Príncipe e na ilha de Annobón (que pertence à Guiné

Equatorial), são todos Crioulos exógenos - ou seja, a sua formação envolveu a migração para um território anteriormente desabitado. Tiveram origem na ilha de São Tomé, espalhando-se depois para outras partes do arquipélago (Bandeira 2017). Passaram por um processo de reestruturação gramatical comparativamente mais drástico, resultando em Crioulos *radicais*, distantes do Português e mutuamente ininteligíveis (Carvalho & Lucchesi 2016). A sua dispersão geográfica é ilustrada na Figura 3.



Figura 1: Variedades cabo-verdianas Figura 2: Variedades continentais do subgrupo da Alta Guiné subgrupo da Alta Guiné



Figura 3: Crioulos do subgrupo do Golfo da Guiné

O grupo do Norte da Índia envolveu os crioulos de Damão, Diu e Korlai. Foram formados em antigas colónias portuguesas na costa norte da Índia e têm todas as línguas Indo-Europeias como substrato (e adstrato, no caso do Hindi). Damão e Diu partilham uma longa história de contacto com a língua lexicalizadora (mais de 400 anos), enquanto Korlai teve contacto direto com os portugueses desde meados do século XVI até 1740. (Clements & Koontz-Garboden 2002). Assim, as duas primeiras línguas reflectem mais a influência do superstrato, enquanto o Korlai tende a ser mais influenciado pelo seu substrato Marathi (I. Smith 2012, Cardoso 2012, Clements 2013). Elas estão representadas geograficamente na Figura 4.

Em contraste, o grupo do Sul da Índia engloba os Crioulos que têm como substrato principal as línguas dravídicas: faladas na região do Malabar, estes Crioulos foram outrora numerosos e dispersos pelo Sul da Índia e chegaram ao Sri Lanka, tendo-se provavelmente formado a partir de um *Pidgin* de base portuguesa falado numa área onde a multietnicidade era a norma. Os Crioulos da nossa amostra foram o Kannur, falado em Cannanore e transmitido *pelo menos até à década de 1950* (Cardoso 2019), e o português do Sri Lanka, cujo número atual de falantes é desconhecido, mas que é muito provavelmente uma língua moribunda (I. R. Smith 2013). São apresentados na Figura 5.

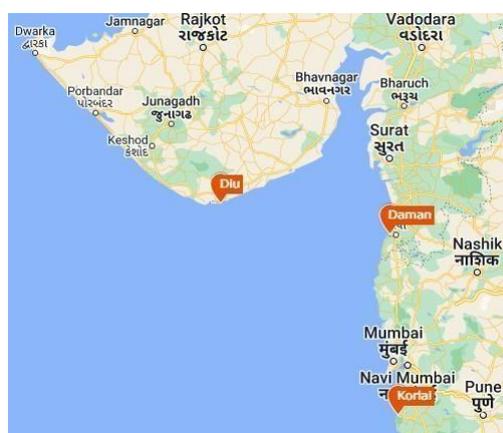


Figura 4: Os três Crioulos do Norte da Índia representados ao longo da costa indiana (Clements 2000, Cardoso 2019).



Figura 5: Crioulos do Sul da Índia falados na Índia e no Sri Lanka

O último grupo é o do Sudeste Asiático, que inclui os Crioulos de Macau e de Malaca (este último também sendo designado por *Papiá Kristang*). São falados, respetivamente, na região autónoma chinesa de Macau e na cidade de Malaca, na Malásia, ambas alcançadas pelos portugueses no século XVI, onde estabeleceram entrepostos comerciais. O *Papiá Kristang* formou-se numa zona de forte contacto linguístico, tendo como substrato línguas de diferentes famílias (cf. Tabela 2) e

sendo ainda influenciado por *Pidgins* indo-portugueses (Baxter 2013). O Crioulo de Macau, por sua vez, foi inicialmente influenciado pelo *Papiá Kristang* (Batalha1988) e teve o Cantonês como principal língua de substrato. O local onde ambas as línguas são faladas está representado na Figura 6.

Em oposição a esta abordagem geográfica da classificação dos Crioulos de base portuguesa, estes podem também ser divididos em seis grupos distintos, seguindo uma proposta substratista.



Figura 6: Crioulos do Sudeste Asiático falados na Malásia e em Macau

Por outras palavras, tem em conta a(s) língua(s) de sub/adstrato envolvida(s) na formação de cada Crioulo. Isto modifica rigorosamente a classificação dos Crioulos asiáticos, em que os Crioulos do Norte da Índia corresponderiam ao Gauro-Português (substrato Indo-Ariano), os do Sul da Índia ao Dravido-Português (substrato Dravídico), e os Crioulos do sudeste asiático seriam ainda divididos em Malaio-Português (substrato Malaio, i.e. Crioulo de Malaca) e SinoPortuguês (substrato Sinítico, i.e. Crioulo de Macau). A classificação dos Crioulos africanos permanece a mesma (Clements 1996, Cardoso 2020).

Finalmente, a Tabela 2 ilustra as principais línguas de substrato envolvidas na génese dos Crioulos da nossa amostra, bem como as duas diferentes propostas de agrupamento tipológico baseadas na geografia e nas línguas de substrato.

Classificação geográfica	Classificação substratista	Crioulo	Principais línguas de substrato
Alta Guiné	Alta Guiné	CV São Vicente	Wolof, Temne, Mandinka
		CV Santiago	Wolof, Temne, Mandinka
		CV Brava	Wolof, Temne, Mandinka
		CV Santo Antão	Wolof, Temne, Mandinka
		CV Fogo	Wolof, Temne, Mandinka
		Casamancese	Wolof, Mandinka, Nyun
		Guinea-Bissau	Balanta, Fula, Mandinga, Manjaku, Mankanya, Pepel
Golfo da Guiné	Golfo da Guiné	Santomé	Edo, Kikongo
		Principense	Edo, Kikongo, Yoruba
		Angolar	Kimbundu
		Fá dAmbô	Edo, Kikongo, Yoruba
Norte da Índia	Gauro-Português	Diu	Gujarati
		Daman	Gujarati
		Korlai	Marathi
Sul da Índia	Dravido-Português	Sri Lanka	Tamil, Cingalês
		Kannur	Malayalam
Sudeste Asiático	Malaio-Português	Malacca	Malaio, Chinês Hokkien
	Sino-Português	Macau	Cantonês, Malaio, Cingalês

Tabela 2: Tabela com as principais línguas de substrato de cada Crioulo da nossa amostra (Michaelis et al. 2013) contrastando as classificações geográfica e de substrato (Clements 1996)

## 2 Métodos

Nesta secção, descreveremos os passos que seguimos relativamente à recolha e ao tratamento de dados, bem como os métodos quantitativos que foram utilizados para analisar a proximidade entre Crioulos. Todos os dados utilizados neste estudo, bem como os materiais suplementares, estão disponíveis online no GitHub: <https://github.com/CreoPhon/CreoPhonPt/blob/main/Swadesh.csv>

### 2.1 Recolha de dados

Os nossos dados foram recolhidos, na sua maioria, a partir de listas de palavras transcritas foneticamente dos Crioulos seleccionados e de dicionários destas línguas, sendo estes os resultados finais do trabalho de campo realizado por vários autores. O principal objetivo ao seleccionar as palavras a recolher destes recursos era constituir uma lista Swadesh de 207 palavras (Swadesh 1955) para cada Crioulo. Como nem todos os Crioulos estão igualmente documentados, o número efetivo de palavras para cada lista dependia dos materiais que os descreviam. Assim, o número de palavras recolhidas para cada Crioulo variou entre 120 (Angolar) e 202 (Damão), com uma média de 170 palavras por lista.

Depois que as 18 listas de palavras foram completadas, estas foram dispostas numa tabela .csv em que cada coluna corresponde a uma língua e cada linha a uma palavra. Após este alinhamento manual dos cognatos, acrescentámos uma coluna de glosa, uma com os cognatos portugueses alinhados e transcritos foneticamente e outra com o domínio sintático-semântico de cada palavra.

Assim, a nossa lista Swadesh foi também anotada com categorias que correspondem às usadas por Bower (Bower 2007). Isto foi feito para analisar posteriormente quais os domínios sintático-semânticos mais e menos estáveis diacronicamente em relação à língua lexicalizadora.

A figura 7 ilustra o número de palavras em cada domínio.

### 2.2 Processamento de dados

Depois de cada lista Swadesh ter sido constituída, o passo seguinte foi trabalhar com palavras que não estavam transcritas foneticamente: uma vez que alguns dos nossos recursos apenas forneciam formas ortográficas das palavras, era crucial converter os grafemas em caracteres IPA. Neste processo, utilizámos as convenções PHOIBLE para garantir a interoperabilidade (Moran & McCloy 2019). Para o fazer, baseámo-nos em gramáticas destes Crioulos ou nos mesmos recursos, que frequentemente continham uma tabela de mapeamento com correspondências grafema-fonema. Utilizámos este método de substituição de caracteres para cada lista de palavras não transcrita foneticamente, seguindo conversões como as ilustradas na Tabela 3.

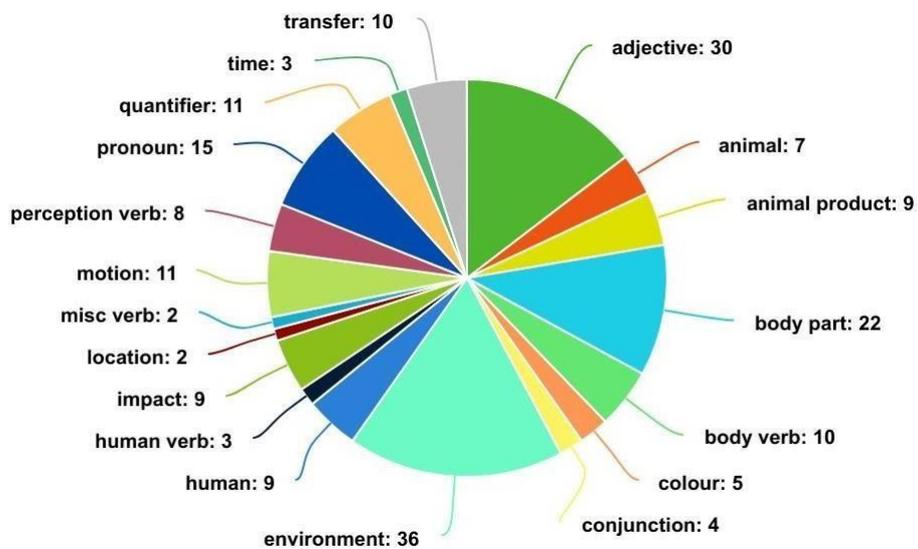


Figura 7: Gráfico de cotação dos campos sintático-semânticos

Ordem	Grafema	Fonema
1	dj	dZ
2	y	j
3	x	S
4	kh	x
5	aa	a
6	ee	E:
7	oo	O:
8	uu	u:
9	ii	i:
10	êê	e:
11	ôô	o:

Tabela 3: Ordem de conversão de grafemas do FÁ d'Ambô em caracteres IPA (Hagemeijer, Maurer-Cecchini & Segorbe 2020)

### 2.3 Análise de dados e medições

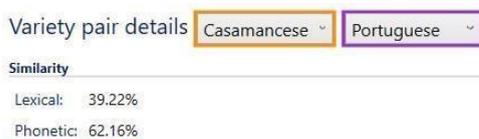
Após o tratamento dos dados, a nossa análise começou com a criação de 19 novas tabelas .csv, cada uma correspondendo a um domínio sintático-semântico e contendo colunas com os 18 Crioulos mais a forma portuguesa e a glosa. Cada linha correspondia a cognatos alinhados dentro do campo semântico da tabela.

Estas tabelas foram analisadas individualmente com o software COG (Daspit 2015, que utiliza o método lexicostático de Blair para comparar lexemas de diferentes variedades transcritos foneticamente. É sensível aos traços, atribuindo assim a cada par de fonemas contrastados um valor de 1 a 3. 1 corresponde à partilha de ponto e modo de articulação, 2 à coincidência em um ou outro e 3 à diferença em ambos ou à supressão de segmentos, como ilustrado na Figura 8. Com base nestes valores, é atribuída uma percentagem de semelhança a um par de variedades. Assim, ao calcular a média dos valores de semelhança de todas as palavras em duas listas de palavras em contraste, o algoritmo produz valores de proximidade fonética para duas línguas, como mostra a Figura 9. Como este estudo analisa as relações diacrónicas entre os Crioulos de base portuguesa e a língua lexicalizadora, o resultado final consistiu em 18 valores de semelhança fonética para cada domínio sintático-semântico. Estes valores tinham como objetivo quantificar a proximidade entre os Crioulos e o português do século XVI.



heart  
| k o r - s o ŋ |  
| k u r e s e ã w |  
1 2 1 3 1 2 3  
Similarity: 66.99%

Figura 8: Resultado da análise COG de duas palavras (respetivamente, do Casamancês e do Português)



Variety pair details Casamancese Portuguese  
Similarity  
Lexical: 39.22%  
Phonetic: 62.16%

Figura 9: Valor de semelhança fonética de duas variedades analisadas através do COG (também do Casamancês e do Português)

Para além dos valores de semelhança fonética, o COG também produz uma Neighbor-net das línguas analisadas derivada dos valores de similaridade entre as listas Swadesh de cada variedade. Este modelo de diversificação baseado na distância é particularmente vantajoso para a análise genealógica das línguas Crioulas, uma vez que também mostra as influências laterais, ou seja, as que estão envolvidas no contacto linguístico (Daval-Markussen, Bøegh & Bakker 2017). Em contrapartida, os dendrogramas, que o COG também gera, não apresentam essas relações laterais, daí a opção pelo primeiro modelo. Na nossa Neighbor-Net, o comprimento das linhas representa a distância genética estimada ou a divergência evolutiva entre as línguas. Por outras palavras, quanto mais longa for a linha que une dois Crioulos contrastados, mais distantes são um do outro.

Com os valores de semelhança fonética de cada domínio para cada Crioulo em mãos, o passo seguinte foi classificar a ordem de estabilidade para cada Crioulo. Ou seja, para cada língua foi criada uma lista em que os primeiros itens correspondiam aos domínios sintático-semânticos mais estáveis e os últimos aos menos estáveis: para isso, cada valor de semelhança fonética foi convertido num número que representava a ordem de estabilidade relativa num dado Crioulo: os valores mais altos (alta estabilidade) ocupavam as primeiras posições do ranking e os valores mais baixos (baixa estabilidade) as últimas. Assim, cada lista apresentava a classificação da estabilidade dos domínios de um Crioulo, como na Figura 10. Esta mostra, por exemplo, que em Korlai o domínio sintático-semântico *misc[ellaneous] verb* é altamente estável (portanto, mais próximo da língua lexicalizadora), enquanto *location* não o é.

Class	Korlai_order	Daman_order	Diu_order	Sri.Lanka_orde	Kannur_order	Malacca_order
adjective	12	13	12	10	10	4
animal	3	3	5	12	4	10
animal product	17	17	18	1	2	2
body part	7	10	10	7	5	3
body verb	6	6	6	9	9	11
colour	8	15	13	3	7	5
conj	9	2	2	18	16	18
environment	11	7	7	8	12	8
human	2	1	4	2	1	7
human verb	5	12	9	15	14	17
impact	15	11	15	13	13	15
location	19	19	19	19	19	19
misc verb	1	5	1	6	6	6
motion	18	18	16	17	18	16
perception verb	10	8	3	14	15	12
pro	14	14	14	11	11	13
quant	4	4	8	4	8	9
time	13	16	17	5	3	1
transfer	16	9	11	16	17	14

Figura 10: Tabela que representa a classificação numérica da estabilidade de cada domínio sintático-semântico num determinado Crioulo

A última etapa consistiu em mapear as correlações entre as ordens de estabilidade de cada língua. Para este fim, foram aplicados testes de Spearman à amostra numérica. O objetivo final era obter o coeficiente de correlação de Spearman ( $\rho$ ) para cada Crioulo relativamente a todos os outros Crioulos da nossa amostra. Essencialmente, os testes de Spearman permitiram-nos quantificar este valor de correlação de -1 a +1, correspondendo -1 à correlação negativa perfeita, 0 à ausência de correlação e +1 à correlação positiva perfeita (Daniel 1990). Para isso, seguimos a seguinte fórmula, em que  $d_i$  corresponde à diferença entre dois valores de classificação (por exemplo, 15 - 12) de duas variedades comparadas e  $n$  ao número de observações (no nosso caso, os 19 domínios sintático-semânticos):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

As implicações linguísticas da obtenção destes valores relacionam-se principalmente com a quantificação de relações estatisticamente significativas entre línguas através da análise de padrões de estabilidade lexical. Isto significa que, para duas línguas negativamente correlacionadas, os domínios mais estáveis de uma tendem a corresponder aos domínios menos estáveis da outra (uma correlação negativa mais forte implica uma correspondência mais forte). Do mesmo modo, para duas línguas correlacionadas positivamente, os domínios mais estáveis de ambas tendem a coincidir, sendo esta tendência mais forte com um valor de  $\rho$  mais próximo de

1.

Quanto à forma como estes dados foram apresentados, optámos por uma matriz de correlação gerada pelo software de computação estatística RStudio (RStudio Team 2020) utilizando o pacote *rstatix* (Kassambara 2023). O resultado final foi ainda aprimorado com o software de design gráfico Canva para aperfeiçoar a visualização dos dados. A matriz representa essencialmente o valor de correlação entre duas variantes contrastadas (no nosso caso, a classificação de estabilidade de cada língua crioula), de modo a que cada variante seja comparada com todas as outras. Na nossa matriz, é atribuído um código de cores a cada valor de correlação, estando os tons de azul associados a uma correlação positiva e os tons de vermelho a uma correlação negativa. Assim, quanto mais escuro for um tom de azul numa matriz, mais forte é a correlação entre duas variantes, e vice-versa.

Por último, uma distinção fundamental entre os dois modelos que escolhemos para representar a diferenciação e a afinidade linguísticas reside no tipo de dados gramaticais que os alimentam: a Neighbor-Net é alimentada apenas com dados de proximidade fonética derivados de comparações entre cognatos alinhados em diferentes variedades (representando, portanto, a Fonologia e o Léxico); por seu lado, a Matriz de Correlação é gerada a partir de dados de estabilidade lexical obtidos a partir da análise dos diferentes graus de mudança dos

domínios sintático-semânticos em diferentes Crioulos (reflectindo uma interface Sintaxe-Semântica).

### 3 Resultados e discussão

Nesta secção, analisaremos criticamente os resultados obtidos com a nossa investigação, centrada na representação de agrupamentos de línguas estatisticamente quantificáveis que partilham afinidades cognitivas e fonéticas. Os nossos resultados podem ser divididos em duas partes principais: uma matriz de correlação que mostra as conexões positivas ou negativas entre os Crioulos em termos da sua ordem de estabilidade e um modelo Neighbor-Net que mostra as relações de cognatos entre as línguas. Abordaremos os resultados nesta mesma ordem.

#### 3.1 Análise dos resultados decorrentes da Matriz de Correlação

Com algumas excepções, a matriz de correlação (Figura 11) está organizada geograficamente e quase não apresenta correlações negativas. Isto significa que, embora apresente conjuntos claros e distinguíveis que reúnem línguas que têm ordens de estabilidade positivamente correlacionadas, a maioria dos Crioulos ou está positivamente correlacionada entre si ou não está

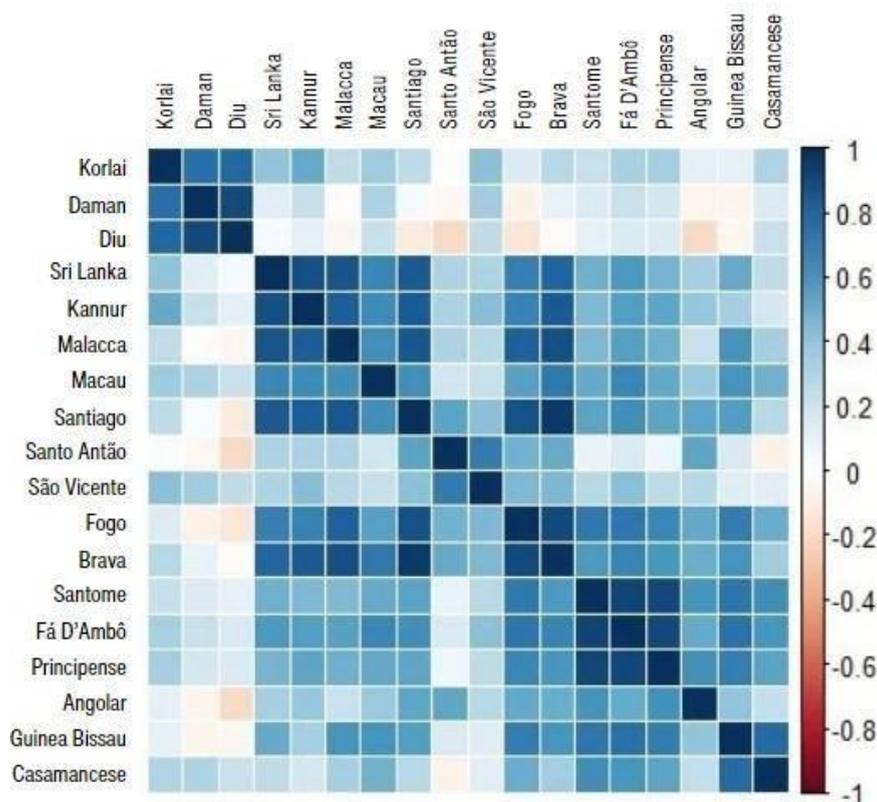


Figura 11: Matriz que mostra as correlações positivas e negativas entre as ordens de estabilidade dos Crioulos de base portuguesa correlacionada de todo.

Podemos observar um primeiro grupo de correlações positivas partilhadas com os Crioulos de Korlai, Damão e Diu: este corresponde ao subgrupo do Norte da Índia dos Crioulos indo-portugueses (Cardoso 2020). Apesar de não partilharem correlações positivas fortes com nenhum outro Crioulo, podemos observar graus muito elevados de correlação positiva entre eles, especialmente entre Damão e Diu. Com efeito, estes dois Crioulos partilham uma história de proximidade e contacto mútuo entre as suas respectivas comunidades de falantes, bem como um período semelhante de contacto com a língua lexicalizadora - respetivamente, 426 e 402 anos (Clements & Koontz-Garboden 2002). O Crioulo de Korlai, no entanto, partilha uma correlação positiva comparativamente mais baixa com Damão e Diu, apesar de pertencer ao mesmo subgrupo. Esta disparidade parece, mais uma vez, encontrar a sua explicação na história: O Korlai teve um período de contacto com o português consideravelmente mais curto (216 anos), um dos factores que provavelmente explicam a influência substancial do substrato Marathi na fonologia, sintaxe, morfologia e léxico do Korlai (Silva 2023, Clements 2013). Isso não é observado em Damão e Diu, que são relativamente mais influenciados pela língua lexicalizadora (I. Smith 2012, Cardoso 2012, Clements & Koontz-Garboden 2002).

Também se nota um grau semelhante de coesão entre Kannur e Sri Lanka: isto não é surpreendente, uma vez que ambos são Crioulos do Sul da Índia que partilham um substrato dravídico principal. O que é menos esperado, no entanto, é o facto de estes Crioulos partilharem correlações positivas elevadas com os de Malaca, Macau (que também estão altamente correlacionados entre si, correspondendo ao grupo do Sudeste Asiático) e os cabo-verdianos do Sotavento (Santiago, Brava e Fogo).

A primeira correspondência pode ser analisada à luz dos movimentos populacionais históricos que envolvem o subcontinente indiano e o sudeste asiático. Com efeito, Baxter (2013) salienta que a génese do Papiá Kristang está relacionada com a importação de *modelos de português pidginizado* da Índia para Malaca. Neste sentido, este contexto histórico poderá esclarecer a ligação entre os Crioulos do Sul da Índia e Malaca. Quanto à relação entre Macau e os Crioulos do Sul da Índia, parece estar em ação uma ligação histórica semelhante. 31,75% da população de Macau entre 1601 e 1624 era constituída por escravos malaqueses e indianos (Teixeira 1961): o contacto desta parcela da população com as restantes durante a formação do Crioulo de Macau poderá ser uma explicação para as semelhanças observadas. De facto, tanto Batalha (1988) como Dalgado (1903) referem a proximidade linguística (desta vez baseada em factores qualitativos) entre Macau e os Crioulos indo-portugueses.

A segunda correspondência parece ter uma explicação sociolinguística: as ilhas do Sotavento (Brava, Fogo, Santiago e Maio, esta última não incluída na nossa amostra) caracterizaram-se por casamentos interétnicos entre colonos portugueses e populações nativas

(no caso de Cabo Verde, escravos africanos trazidos para as ilhas) (Lopes et al. 2014, Baptista 2019). Um processo semelhante ocorreu em toda a Ásia (Clements & Koontz-Garboden 2002), conduzindo em muitos casos a uma fase de *homestead* permanente, isto é, de contacto próximo entre variedades crioulas e variedades nativas portuguesas (Silva 2023, Mufwene 2007). Em ambos os casos, estas condições análogas de contacto podem ser factores relevantes para a caracterização destes Crioulos como *leves*, implicando que têm um maior grau de inteligibilidade mútua e proximidade com o lexificador do que os Crioulos *radicais*, como os do Golfo da Guiné (Carvalho & Lucchesi 2016).

Partindo deste grupo de correlação positiva maior, o seguinte engloba os Crioulos caboverdianos do Barlavento (Norte) de Santo Antão e São Vicente: estas línguas estão altamente correlacionadas entre si, mas diferem da maioria dos outros Crioulos, partilhando correlações positivas, embora mais fracas, com os Crioulos cabo-verdianos do Sotavento (Brava, Fogo e Santiago). Por conseguinte, a matriz de correlação parece também representar a divisão histórica, geográfica e linguística entre as ilhas do Sul e do Norte de Cabo Verde: estas ilhas passaram, de facto, por processos de colonização diferentes, sendo as do Sul as primeiras a terem sido alcançadas e ocupadas pelos portugueses. Apesar de terem sido todas descobertas na segunda metade do século XV, as ilhas do Sotavento foram ocupadas desde o início, maioritariamente por colonizadores portugueses e por populações escravizadas trazidas de África, pertencentes a diferentes etnias; em contrapartida, as ilhas do Barlavento atraíram uma quantidade significativa de migrantes europeus oriundos de vários países, e apenas a partir do século XVI (Baptista 2019). Estas diferentes origens históricas, bem como a geografia, parecem ser factores que, pelo menos parcialmente, explicam a divisão dos Crioulos cabo-verdianos em Crioulos do Barlavento e do Sotavento na matriz.

O grupo seguinte é constituído por Santomé, Fá d'Ambô e Principense. Estes Crioulos partilham também uma correlação mais fraca com o Angolar, que em conjunto correspondem ao grupo do Golfo da Guiné. Os valores de correlação mais fortes entre Santomé, Fá d'Ambô e Principense quando comparados com o Angolar, um parente linguístico, podem ser atribuídos a diferentes factores: i) enquanto o Angolar tem um substrato Kimbundu significativo, os primeiros Crioulos estão mais relacionados com o Edo, o Kikongo e o Yorubá (Tjerk Hagemeijer 2011); ii) tipologicamente, o Santomé, o Fá d'Ambô e o Principense são Crioulos de plantação, e o Angolar é o único Crioulo de base portuguesa, de origem maroon (ou quilombola), tendo surgido numa sociedade formada por escravos fugidos (Lorenzino 1998).

Por último, podemos observar uma forte correlação positiva entre o Casamancês e o Kriyol da Guiné-Bissau. Estes Crioulos também pertencem tipologicamente ao Grupo da Alta Guiné e partilham correlações positivas significativas com outros Crioulos luso-atlânticos,

nomeadamente Santomé, Fá d'Ambô, Principense, Fogo e Brava. Por esta razão, parece que a ordem de estabilidade do Casamancês e do Kriyol não está apenas correlacionada com os Crioulos de Cabo Verde, como seria de esperar dada a sua proximidade histórica, mas também com os do Golfo da Guiné. Este facto fornece mais uma evidência para o agrupamento destes dois subgrupos num macrogrupo atlântico que representa os Crioulos africanos de base portuguesa, por oposição aos formados na Ásia (Cardoso 2020).

Em suma, a maioria dos agrupamentos de correlações positivas apresentados na matriz de correlações corresponde à classificação tradicional dos Crioulos de base portuguesa referida na subsecção 1.1: as correlações entre as classificações de estabilidade lexical destes Crioulos parecem apontar na direção de que a sua estabilidade é sobretudo orientada pela geografia, mas também influenciada por factores sócio-históricos. Estes resultados também devem ser lidos com cautela, pois há exceções à hipótese do agrupamento geográfico que, após uma inspeção mais profunda - como tentámos fazer nesta subsecção - podem ser explicadas por elementos não linguísticos, como o período/extensão do contacto com a língua lexicalizadora.

### 3.2 Análise dos resultados decorrentes da Neighbor-Net

Apesar de a nossa matriz de correlação sugerir relações genealógicas entre os Crioulos de base portuguesa, ainda apresenta alguns resultados inesperados, como o grupo do Sudeste Asiático-Cabo Verde. A Neighbor-Net na Figura 12 parece ser uma melhor opção para explorar as ligações filogenéticas entre as línguas da nossa amostra.

Como mencionado na secção Métodos, este modelo computacional utiliza dados de cognatos para constituir uma rede linguística em que o comprimento das linhas que ligam duas variedades corresponde à distância entre elas. Assim, o que podemos ver são agrupamentos de línguas que partilham traços fonéticos comuns e um percurso de distanciamento semelhante relativamente a uma língua lexicalizadora comum: neste caso, o Português do século XVI.

Assim sendo, o primeiro grupo linguístico é muito coeso e inclui os Crioulos do Golfo da Guiné: Fá d'Ambô, Principense, Santomé e Angolar. Optámos também por incluir na nossa amostra a língua hipotética do Proto-Crioulo do Golfo da Guiné, com base na investigação realizada por (Bandeira 2017). Este grupo está razoavelmente distante do português, dada a extensão das linhas que os separam: isto corresponde à distinção Crioulo Radical vs. Crioulo Leve avançada por Carvalho e Lucchesi (2016), em que os Crioulos deste grupo se qualificariam como Radicais, dado o elevado grau de desvio linguístico relativamente à sua língua lexicalizadora. Isto parece ser especialmente verdade para o Angolar, o único Crioulo *maroon* de base

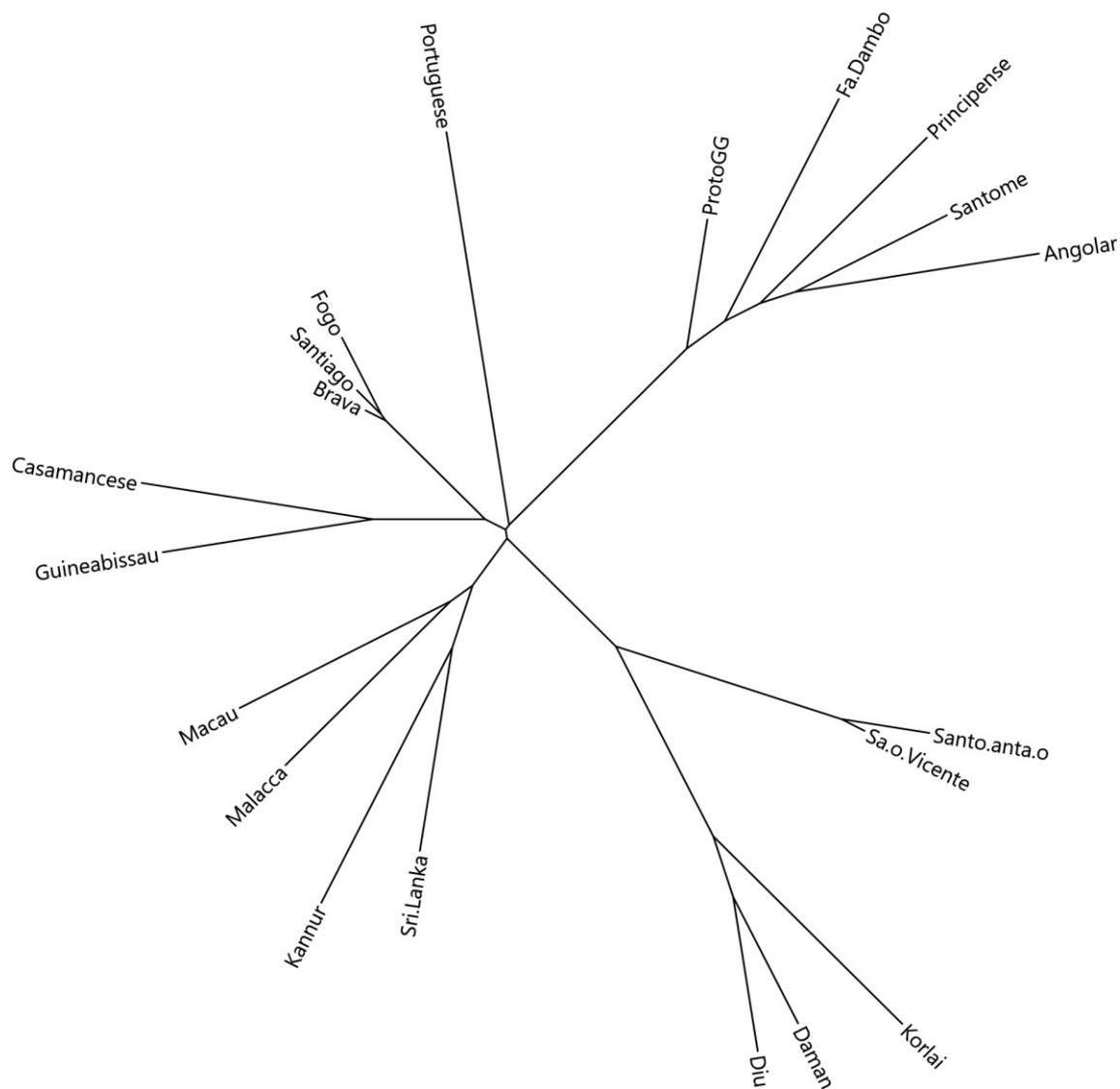


Figura 12: Neighbor-Net gerada a partir de dados fonéticos e de cognatos de Crioulos de base portuguesa.

portuguesa e simultaneamente um Crioulo com uma duração muito curta de contacto com o Português (Lorenzino 1998): na Neighbor-Net, isto reflecte-se no longo comprimento das linhas que o separam (e aos seus parentes do Golfo da Guiné) da língua lexicalizadora.

O ramo seguinte agrupa dois grupos de variedades diferentes. Em primeiro lugar, os Crioulos de Barlavento de Cabo Verde, que estão separados dos de Sotavento (que se encontram no extremo oposto da Neighbor-Net): esta divisão parece refletir os diferentes percursos históricos que se seguiram nestes grupos de ilhas - o primeiro foi ocupado principalmente por portugueses e outros migrantes europeus, e o segundo por colonos portugueses que trouxeram consigo escravizados de várias etnias da África Ocidental (Baptista 2019). Concomitantemente, a relativa distância dos Crioulos do Barlavento em relação ao Português, quando comparada com a dos Crioulos do Sotavento (representados muito mais perto do lexificador), corrobora a

oposição de Baptista (2019) à afirmação de que os primeiros tendem a ser acrolectais e os segundos basilectais.

Juntamente com os Crioulos do Barlavento encontramos, surpreendentemente, os Crioulos do Norte da Índia: dentro deste grupo, porém, as distâncias linguísticas correspondem à classificação tradicional: Damão e Diu, partilhando o mesmo substrato Gujarati, estão mais próximos uns dos outros do que de Korlai, que tem um substrato Marathi (Clements & Koontz-Garboden 2002).

Mais uma vez reflectindo o que já foi observado na Figura 11, um outro ramo engloba dois grupos: os Crioulos do Sul da Índia (Sri Lanka e Kannur) e os Crioulos do Sudeste Asiático (Macau e Malaca). Mais uma vez, uma possível explicação para esta junção parece estar nas intensas vagas migratórias da Índia para Malaca e Macau nos respectivos períodos de colonização durante os séculos XVI e XVII (Baxter 2013, Teixeira 1961). Ao mesmo tempo que esta influência histórica aparece discernível no modelo de diferenciação, a manutenção da distinção tipológica destes grupos (Sudeste Asiático e Sul da Índia) evidencia a manutenção do Princípio Fundador: *características da variedade fundadora tinham frequentemente vantagem selectiva* sobre as das línguas envolvidas em contactos posteriores (Mufwene 1996).

Finalmente, o último ramo incorpora, por um lado, os Crioulos vizinhos de Casamança e da Guiné-Bissau e, por outro, os Crioulos cabo-verdianos do Sotavento. Esta proximidade é esperada, uma vez que estas línguas constituem uma parte significativa dos Crioulos da Alta Guiné: como já foi referido, observamos uma separação entre os Crioulos do Sotavento, que se apresentam próximos dos seus parentes linguísticos, e os Crioulos do Barlavento, que ocupam uma posição mais distante. Este distanciamento, como já foi referido, pode muito bem ser atribuído às diferentes condições de contacto a que estes grupos linguísticos estiveram sujeitos. Simultaneamente, não deve deixar de ser referido que os Crioulos da Alta Guiné são amplamente considerados os mais próximos do lexificador, tanto geográfica como linguisticamente (Kihm 1994).

Em conclusão, a Neighbor-Net mostra, de facto, uma distribuição maioritariamente geográfica das suas variedades, baseada apenas em critérios fonéticos e de cognatos. As variações em relação à classificação tradicional dos Crioulos de base portuguesa não devem ser interpretadas como divisões linguísticas recém-descobertas, mas antes como uma indicação de que os dados das listas Swadesh podem reflectir realidades históricas e sociológicas partilhadas (Baptista 2019). Assim, a estabilidade lexical nestes Crioulos parece encontrar uma explicação mais clara na ideia de ecossistemas linguísticos geograficamente específicos (Mufwene 2001).

## 4 Conclusões

### 4.1 Constatações principais

Os dados apresentados na secção anterior parecem dar uma resposta parcial às questões de investigação colocadas na introdução:

- Os Crioulos de base portuguesa seguem caminhos diferentes de diversificação linguística?
- Como podemos medir e representar as distâncias entre estas línguas?
- Quais são os factores linguísticos e não linguísticos que desempenham um papel na diversificação linguística destes Crioulos?

Em primeiro lugar, tanto os resultados da Neighbor-Net como os da Matriz de Correlação apontam para a formação de agrupamentos de línguas que partilham características fonéticas e sintático-semânticas: a ordem de estabilidade é diferente de Crioulo para Crioulo, mas há claras tendências de partilha que correspondem à classificação geográfica. Assim, a estabilidade lexical dos itens da lista Swadesh dos Crioulos de base portuguesa parece indicar que os percursos de diversificação linguística não são os mesmos em todos eles, o que levanta a questão: quais são os factores que influenciam a emergência destas diferentes trajetórias de evolução linguística?

Dar uma resposta definitiva a esta questão implicaria a investigação de temas ontológicos e profundamente teóricos no âmbito da génese dos Crioulos, o que ultrapassa o âmbito deste artigo. É, no entanto, possível apontar factores ligados às condições de contacto que parecem explicar os agrupamentos que relatámos. A geografia aparenta desempenhar um papel importante: sabemos pelos Sprachbunds que a mudança linguística também pode emergir de factores areais (Matasovic 2012) e, de facto, os nossos resultados mostram que as línguas se organizam principalmente pelas áreas onde são faladas. O efeito da geografia na mudança e na diversidade linguísticas também foi discutido por Nash *et al.* (2020), que referem que as ilhas são geralmente consideradas zonas de contacto linguístico intensivo e demonstram que *o isolamento geográfico relativo pode ter um impacto na língua*. De facto, muitas das línguas de contacto de base portuguesa investigadas neste trabalho são faladas em ilhas etnicamente diversas, e outras em áreas isoladas, como fortes e antigos entrepostos comerciais.

No entanto, algumas ligações apresentadas nos nossos resultados certamente não se devem à geografia, como é o caso da forte correlação positiva entre as ordens de estabilidade dos Crioulos de Macau e de Kannur. Em vez disso, os factores socio-históricos são mais adequados para clarificar tais relações: neste caso específico, salientámos o grande afluxo de escravizados do Sul indiano para Macau durante a formação do seu Crioulo (Teixeira 1961). Os papéis que as condições económicas, históricas e sociológicas desempenham na génese dos Crioulos foram extensamente descritos por Mufwene (2001) e, de facto, os diferentes ecossistemas linguísticos

de onde estas línguas emergiram parecem clarificar padrões de estabilidade (tanto os esperados como a maioria dos inesperados).

Com isto em mente, apesar de não haver consenso acadêmico sobre se os Crioulos são tipologicamente diferentes de outras línguas ou não (Hagemer 2015), podemos pelo menos notar que os vários caminhos de divergência linguística em relação à estabilidade lexical crioula discutidos neste estudo são comparáveis aos observados na evolução do Latim para as línguas Românicas (Dworkin 2016, Bassetto 2016). Neste sentido, a formação de Crioulos de base portuguesa não parece ser, no que diz respeito à estabilidade lexical, completamente diferente da transferência do Latim para as línguas Românicas, algo que já foi sugerido por Silva (2023) para o domínio da Fonologia dos Crioulos e DeGraff (2009) para a Sintaxe dos Crioulos.

Ainda assim, estas afirmações não devem ser entendidas como a exclusão de influências intra-linguísticas na formação dos diferentes percursos de divergência relacionados com a estabilidade lexical: uma das características dos Crioulos de base portuguesa que desempenha um papel importante na sua classificação é a diferença das suas línguas de substrato (Cardoso 2020). Este fator parece explicar em parte, por exemplo, a clara separação entre os Crioulos do Norte da Índia e os do Sul da Índia, tanto na Matriz de Correlação como na Neighbor-Net.

Para além disso, como Silva (2023) indica, pode dar-se o caso de, por vezes, a mudança linguística ser motivada por tendências universais: padrões não inerentes a todas as línguas, mas que estão presentes na maioria delas e que indicam que alguns processos de mudança são mais prováveis do que outros. No entanto, se este é um fator que determina a estabilidade lexical nos Crioulos de base portuguesa, não deve ser o principal, e muito menos o único: os resultados da Neighbor-Net e da Matriz de Correlação ainda seguem um arranjo principalmente geográfico.

## 4.2 Limitações e direcções futuras

As limitações deste estudo relacionam-se principalmente com a escassez de dados de campo: para alguns dos Crioulos que analisámos, os materiais disponíveis eram escassos, sendo por vezes constituídos por uma lista de vocábulos básicos com menos de 150 palavras que não foram transcritas foneticamente. Outros Crioulos, como os da Batávia (também designado por *Tugu*) e de Timor-Leste, tiveram de ser retirados da nossa amostra final, uma vez que não existiam listas de vocabulário básico publicamente disponíveis nestas línguas, o que significa que não havia material suficiente para efetuar uma análise léxico-estatística delas. Este facto não é exclusivo destes Crioulos do Sudeste Asiático, uma vez que, como refere Cardoso (2020), muitos Crioulos do Sul da Índia estão também subdocumentados. Para mitigar este problema,

é fundamental que se realizem mais estudos de campo sobre os Crioulos de base portuguesa, de modo a documentar línguas que, em muitos casos, estão moribundas e quase extintas (I. R. Smith 2013, Cardoso 2020).

Estudos futuros que derivem dos resultados deste artigo podem, em primeiro lugar, beneficiar-se da abordagem do aparente excepcionalismo do Norte da Índia: olhando para os Crioulos de Damão, Diu e Korlai representados na figura 11, nota-se que partilham fortes correlações positivas entre si, mas não com qualquer outro Crioulo. Para além disso, o Crioulo de Diu está correlacionado negativamente com pelo menos cinco outras línguas de contacto. Isto leva-nos a perguntar: que factores poderiam explicar esta falta de correlação com os outros Crioulos de base portuguesa? Uma resposta possível poderia residir no facto de estes serem os únicos

Crioulos de língua lexicalizadora portuguesa que têm como substrato línguas Indo-Europeias (Gujarati para Damão e Diu e Marathi para o Korlai): tratar-se-ia, portanto, de investigar em que medida o contacto entre línguas e culturas Indo-Europeias poderia explicar as diferenças linguísticas acima mencionadas relativamente aos outros Crioulos do grupo.

Além disso, poderia ser benéfico abordar as inovações que estas línguas poderiam partilhar com o Português do Brasil e outras variedades. De facto, a língua portuguesa no Brasil esteve em contacto com línguas Bantas (e.g. Kimbundu, Iorubá) e Tupi (e.g. Nheengatu), entre outras. Apesar de não ser considerado um Crioulo, o Português Brasileiro difere do Português Europeu em aspectos fonológicos, sintácticos e morfológicos, algo que nalguns casos poderá ser atribuído à influência do substrato (Bossaglia 2019). No domínio da Fonologia, Silva (2021) traça paralelos relevantes entre os processos de mudança fonológica nos Crioulos de base portuguesa e os que afectaram o Português do Brasil, demonstrando que algumas mudanças observadas nos primeiros estão também presentes no segundo. No domínio da estabilidade lexical, no entanto, há ainda uma lacuna de conhecimento sobre este assunto.

Por último, é importante abordar o conceito de mistura linguística: a ideia de que ela não se restringe aos Crioulos está presente pelo menos desde a afirmação de Schuchardt (1884) de que *es gibt keine völlig ungemischte Sprache* (ou seja, *não existe uma língua totalmente não mista*). De facto, pode argumentar-se que mesmo o Proto-Indo-Europeu tinha características gramaticais que poderiam ser atribuídas a um hipotético Sprachbund ancestral com línguas Caucásicas (Matasovic 2012). Numa abordagem semelhante, Thomason e Kaufman (1988) analisam ainda numerosos casos de contacto e mistura linguística que envolvem línguas não Crioulas, como os dialectos do Norte da Rússia e o Inglês. Por conseguinte, a abordagem de *quão* misturados são os Crioulos de base portuguesa e outras línguas pode ser um próximo passo promissor para a Criolística e a Tipologia Linguística.

## Referências Bibliográficas

- Bakker, Peter et al. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1). 5–42. <https://doi.org/10.1075/jpcl.26.1>.
- 02bak.
- Bandeira, Manuele. 2017. *Reconstrução fonológica e lexical do protocioulo do Golfo da Guiné*. Universidade de São Paulo Ph.D. dissertation. <https://doi.org/10.11606/t.8.2017.tde-05042017-134159>.
- Baptista, Marlyse. 2019. Variation in creole languages: Insights from a Swadesh list. *Journal of Ibero-Romance Creoles* 9(1). 251–277. <http://www.acblpe.com/revista/volume-9-2019/variation-in-creole-languages-insights-from-aswadesh-list>.
- Bassetto, B. F. 2016. *Elementos de Filologia Românica - Volume 2: História Interna das Línguas Românicas*. Editora da Universidade de São Paulo.
- Batalha, Graciete. 1988. *Glossário do dialecto macaense*. Macao: Instituto Cultural de Macao.
- Baxter, Alan N. 2013. Papiá Kristang. Em Susanne Maria Michaelis et al. (eds.), *The survey of pidgin and creole languages. Vol. 2: Portuguese-based, Spanish-based and French-based languages*, 122–130. Oxford: Oxford University Press.
- Biagui, Noël Bernard. 2012. *Description générale du créole afro-portugais parlé à Ziguinchor (Sénégal)*. Université Cheik Anta Diop Ph.D. dissertation. <http://www.theses.fr/2012INAL0017>.
- Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma Press.
- Bossaglia, Giulia. 2019. *Linguística Comparada e Tipologia*. Parábola Editorial.
- Bowern, Claire. 2007. *Linguistic Fieldwork*. London: Palgrave Macmillan.
- Cardoso, Hugo. 2012. Luso-Asian comparatives in comparison. Em Hugo Cardoso, Alan Baxter & Mário Pinharanda-Nunes (eds.), *Ibero-Asian Creoles: Comparative Perspectives*, 81–124. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.46>.
- Cardoso, Hugo. 2019. The Indo-Portuguese creoles of the Malabar: Historical cues and questions. en. Em Pius Malekandathil, Lotika Varadarajan & Amar Farooqi (eds.), *India, the Portuguese, and maritime interactions: Religion, language and cultural expression*, vol. 2, 345–373. Delhi: Primus Books.
- Cardoso, Hugo. 2020. Contact and Portuguese-Lexified Creoles. Em *The Handbook of Language Contact*, cap. 23, 469–488. John Wiley & Sons, Ltd. <https://doi.org/https://doi.org/10.1002/9781119485094.ch23>.
- Carvalho, Ana Maria & Dante Lucchesi. 2016. Portuguese in Contact. Em W. Leo Wetzels, João Costa & Sergio Menuzzi (eds.), *The Handbook of Portuguese Linguistics*, 41–55. Oxford:

- John Wiley & Sons. <https://doi.org/10.1002/9781118791844.ch3>.
- Chaudenson, Robert. 1979. *Les Créoles Français*. Fernand Nathan.
- Clements, J. Clancy. 1996. *The Genesis of a Language*. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.16>.
- Clements, J. Clancy. 2000. Evidência para a existência de um pidgin português asiático. pt. Em Maria Antónia Pereira & Ernesto Mota (eds.), *Crioulos de base lexical portuguesa*, 185–200. Braga: Associação Portuguesa de Linguística.
- Clements, J. Clancy. 2013. Korlai. Em Susanne Maria Michaelis et al. (eds.), *The survey of pidgin and creole languages. Volume 2: Portuguese-based, Spanish-based, and French-based Languages*. Oxford: Oxford University Press. <https://apics-online.info/surveys/40>.
- Clements, J. Clancy. 2021. Portuguese-Lexified Creoles. Em *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.452>.
- Clements, J. Clancy & Andrew Koontz-Garboden. 2002. Two Indo-Portuguese creoles in contrast. *Journal of Pidgin and Creole Languages* 17. 191–236.
- Dalgado, Sebastião. 1903. Dialecto indo-português de Damão. *Ta-Ssi-Yang-Kuo* 3(6).
- Daniel, Wayne. 1990. Spearman Rank Correlation Coefficient. en. Em Wayne Daniel (ed.), *Applied Nonparametric Statistics*, 358–365. 2nd edition. Boston, MA: PWS-Kent.
- Dasplit, Damian. 2015. *Cog software*. <https://software.sil.org/cog/>.
- Daval-Markussen, Aymeric, Kristoffer Bøegh & Peter Bakker. 2017. Chapter 7. West African languages and creoles worldwide. Em 141–174. <https://doi.org/10.1075/z.211.07dav>.
- DeGraff, Michel. 2001. Morphology in creole genesis: Linguistics and ideology. en. *Current Studies in Linguistics Series* 36. 53–122.
- DeGraff, Michel. 2009. Language Acquisition in Creolization and, Thus, Language Change: Some Cartesian-Uniformitarian Boundary Conditions. *Language and Linguistics Compass* 3(4). 888–971. <https://doi.org/https://doi.org/10.1111/j.1749-818X.2009.00135.x>.
- Dworkin, Steven. 2016. Lexical stability and shared lexicon. Em *The Oxford Guide to the Romance Languages*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199677108.003.0032>.
- Hagemeijer, T. 2015. Perspectives on historical syntax. Em Carlotta Viti (ed.), *The Gulf of Guinea creoles: A case-study of syntactic reconstruction*. John Benjamins. <https://doi.org/10.1075/slcs.169>.
- Hagemeijer, Tjerk. 2011. The Gulf of Guinea Creoles. *Journal of Pidgin and Creole Languages*

26(1). 111–154. <https://doi.org/10.1075/jpcl.26.1.05hag>.

Hagemeyer, Tjerk, Philippe Maurer-Cecchini & Armando Zamora Segorbe. 2020. *A Grammar of Fa d'Ambô*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/doi:10.1515/9783110766264>.

Kassambara, Alboukadel. 2023. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.7.2. <https://rpkgs.datanovia.com/rstatix/>.

Kihm, Alain. 1994. *Kriyol Syntax*. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.14>.

Lopes, Raimundo et al. 2014. *A Variação Geográfica do Crioulo Caboverdiano*. Jürgen Lang (ed.). FAU University Press.

Lorenzino, Gerardo. 1998. *The Angolar Creole Portuguese of São Tomé: its Grammar and Sociolinguistic History*. Lincom Europa.

Matasovic, Ranko. 2012. Areal Typology of ProtoIndoEuropean: The Case for Caucasian Connections. *Transactions of the Philological Society* 110. <https://doi.org/10.1111/j.1467-968X.2012.01309.x>.

Mendes, Ronaldo. 2023. *Variação interna em Kriol (crioulo da Guiné-Bissau): Fonologia e aspetos sociolinguísticos*. Universidade do Porto MA dissertation.

Michaelis, Susanne Maria et al. (eds.). 2013. *APiCS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://apics-online.info/>.

Moran, Steven & Daniel McCloy (eds.). 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. <https://phoible.org/>.

Moreira, Ana Karina Tavares. 2012. *Documentation et description grammaticale et lexicale du créole afro-portugais de île de Fogo (République du Cap-Vert, Afrique de l'Ouest)*. Institut National des Langues et Civilisations Orientales Ph.D. dissertation. <https://theses.fr/2020INAL0028>.

Mufwene, Salikoko. 1996. The Founder Principle in Creole Genesis. *Diachronica* 13. 83–134. <https://doi.org/10.1075/dia.13.1.05muf>.

Mufwene, Salikoko. 2001. *The Ecology of Language Evolution*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511612862>.

Mufwene, Salikoko. 2007. Population Movements and Contacts in Language Evolution. *Journal of Language Contact* 1(1). 63–92. <https://doi.org/10.1163/000000007792548332>.

Nash, Joshua et al. 2020. On languages on islands. *Acta Linguistica Hafniensia* 52(1). 81–116. <https://doi.org/10.1080/03740463.2020.1736747>.

RStudio Team. 2020. *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA. <http://www.rstudio.com/>.

- Schuchardt, Hugo. 1884. *Slawo-deutsches und Slawo-italienisches*. Leuschner & Lubensky.
- Schuchardt, Hugo. 1889. Beiträge zur Kenntnis des creolischen Romanisch. V. Allgemeineres über das Indoportugiesische. *Zeitschrift für Romanische Philologie* 13. 476–516.
- Silva, Carlos. 2021. The representation of Portuguese palatal sonorants through the eyes of Portuguese-based Creoles. en. Em *Language Issues. Proceedings of the 8th conference Language issues: A young scholars perspective*, 19–39. Moscow: Institute of Linguistics/Kantsler.
- Silva, Carlos. 2023. *Consonant stability of Portuguese-based creoles*. Available at <https://hdl.handle.net/10216/157550>. University of Porto PhD thesis.
- Smith, Ian. 2012. Measuring substrate influence: Word order features in Ibero-Asian Creoles. Em *Ibero-Asian Creoles: Comparative Perspectives*, 125–148. Amsterdam: John Benjamins. <https://doi.org/10.1075/cll.46>.
- Smith, Ian R. 2013. Sri Lanka Portuguese. Em Susanne Maria Michaelis et al. (eds.), *The survey of pidgin and creole languages. Vol. II: Portuguese-based, Spanish-based and French-based languages*, 111–121. Oxford: Oxford University Press.
- Starostin, G. 2016. From wordlists to proto-wordlists: reconstruction as optimal selection. *Faits de Langues* 47. 177–200.
- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2). 121–137.
- Teixeira, Manuel. 1961. Macao e a sua Diocese. pt. Em *As ordens e congregações religiosas em Macao*, vol. 3. Macao: Tipografia da Missão do Padroado.
- Thomason, Sarah & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press. <https://doi.org/10.1525/9780520912793>.

# ANÁLISE DE RELAÇÕES DISCURSIVAS EM TED-TALKS COM BASE NA ISO 24617-9

Joana Ribeiro da Silva Gomes

Orientado pela Professora Doutora Purificação Silvano

## Resumo

Este trabalho tem como objetivos investigar o uso das relações discursivas em textos de tipo narrativo em português europeu e avaliar se a proposta teórica da ISO 24617-8 é eficiente para uma anotação completa e abrangente dessas relações discursivas. Para concretizar estes objetivos, foram selecionadas três *Ted Talks* de carácter narrativo e anotadas as relações discursivas com base no esquema de anotação da ISO 24617-8. Foram ainda identificados e classificados os conectores sempre que presentes. Simultaneamente, analisou-se a aplicabilidade da ISO 24617-8. A análise dos resultados permitiu concluir que no nosso corpus as relações discursivas mais frequentes são *Expansion*, *Synchrony* e *Cause*, e que, na maioria das vezes, as relações não são marcadas por conectores. O estudo realizado permite ainda concluir que a ISO 24617-8 apresenta alguns pontos negativos, nomeadamente no que diz respeito à indefinição de alguns conceitos de relações discursivas.

Palavras Chave: Relações Discursivas, Conectores, Texto Narrativo, ISO 24617-8

## Abstract

This work aims to investigate the use of discourse relations in narrative texts in European Portuguese and to assess whether the theoretical proposal of ISO 24617-8 is efficient for a complete and comprehensive annotation of these discourse relations. Three narrative *Ted Talks* were selected to achieve these objectives and the discourse relations were annotated using the ISO 24617-8 annotation scheme. The connectors were also identified and classified whenever they were present. At the same time, the applicability of ISO 24617-8 was analysed. Regarding the results, we concluded that the most frequent discourse relations in our corpus are *Expansion*, *Synchrony* and *Cause*, and that most of the time the relations are not marked by connectors. The study also led us to conclude that ISO 24617-8 has some negative points, namely concerning the undefinition of some concepts of discourse relations.

Keywords: Discourse Relations, Connectors, Narrative Texts, ISO 24617-8

## 1. Introdução

As relações discursivas, entendidas como relações de sentido entre unidades discursivas, desempenham um papel fundamental nos estudos da área Semântica, nomeadamente na compreensão da estrutura e do significado do texto (Oliveira *et al.* 2010).

Existem várias propostas teóricas, para a anotação de relações discursivas, como *Rhetorical Structure Theory* (RST), (Mann & Thompson, 1987), *Segmented Discourse Representation Theory* (SDRT) (Asher & Lascarides, 2003) e a *Penn Discourse Treebank* (PDTB) (Prasad *et al.* 2008) (cf. Silvano, 2010), o que dificulta a comparação de datasets anotados. Nesse sentido, tal como refere Tomaszewska *et al.* (2024), é importante encontrar uma proposta que estipule um esquema de anotação que seja interoperável e que possa ser usado em várias línguas e géneros textuais. É neste contexto que a *ISO 25617-8 Semantic relations in discourse, core annotation schema (DR-core)* (ISO, 2016) é relevante, pois é uma proposta teórica adotada para a anotação de relações discursivas a nível global, com o objetivo de uniformizar o mais possível a anotação.

Neste estudo de carácter exploratório, os principais objetivos são os seguintes: (i) investigar as relações discursivas em monólogos de natureza narrativa em português europeu; e (ii) determinar se a *ISO 25617-8* é adequada para anotar essas relações discursivas. Procuraremos perceber de que maneira são estabelecidas as relações discursivas, destacando ainda os conectores que lhes estão associados, bem como a frequência com que cada relação discursiva surge no texto. Para tal, foram selecionadas três Ted Talks de carácter maioritariamente narrativo, sendo que a primeira foi uma Ted Talk de treino, portanto, de menor extensão. A anotação dos dados foi feita num documento em Excel seguindo vários parâmetros, como a identificação dos argumentos e dos respetivos papéis semânticos (“semantic roles”). De seguida, procedeu-se a uma análise qualitativa e quantitativa dos resultados.

O artigo encontra-se estruturado da seguinte forma: na Secção 2 será descrito o conceito de relação discursiva (2.1.), apresentadas as propostas teóricas mais relevantes que integram as relações discursivas (2.2.) e as propostas que usam a proposta da ISO 24617-8 em português europeu (2.3.). Na Secção 3, será descrito o estudo, começando com as questões de investigação (3.1.), o corpus e a metodologia (3.2.), passando à apresentação e discussão dos resultados (3.3.) e, por último, à reflexão sobre os desafios da anotação. Por fim, a Secção 4 é dedicada às considerações finais.

## **2. Relações discursivas**

Nesta secção será descrito o conceito de relação discursiva, bem como as propostas teóricas existentes para a anotação das mesmas.

### **2.1. Conceito de relação discursiva**

As relações discursivas, ou relações retóricas, são descritas como “relações de sentido entre unidades discursivas” (Oliveira *et al.* 2010), constituindo um elemento fundamental para a compreensão da forma como o significado é construído.

Existem várias designações para este mesmo conceito, nomeadamente "relações de coerência", "relações de coesão", "relações lógicas" e "relações pragmáticas" (Silvano, 2010). No entanto, cada um destes termos pode ter implicações teóricas específicas, dependendo do autor ou da proposta em questão. Esta variação terminológica pode gerar confusão e dificultar não só a comparação das diferentes abordagens, como também dos corpora anotados. É neste contexto que se considera que seria produtiva uma sistematização das designações, o que levaria a uma terminologia mais clara e unificada que pudesse ser adotada para melhorar a consistência na descrição e análise das relações discursivas (Silvano, 2010).

### **2.2. Propostas teóricas**

Várias teorias têm integrado as relações discursivas nas suas propostas. Algumas das propostas mais (re)conhecidas são as seguintes: *Rhetorical Structure Theory* (RST), (Mann & Thompson, 1987), *Segmented Discourse Representation Theory* (SDRT) (Asher & Lascarides, 2003) e a *Penn Discourse Treebank* (PDTB) (Prasad *et al.* 2008) (cf. Silvano, 2010 para uma revisão sistemática destas propostas). Estas propostas variam na sua natureza semântica e pragmática e diferem em alguns dos postulados estabelecidos para a inferência e representação das relações discursivas.

A RST, de Mann & Thompson (1987), procura explicar como partes de um texto se relacionam entre si de maneira a formar uma unidade coesa e coerente, descrevendo como os textos são organizados em termos de relações retóricas, ou seja, relações de sentido entre segmentos do texto.

A SDRT, proposta por Asher & Lascarides (2003), consiste numa abordagem formal, com o objetivo de modelar a estrutura e a interpretação de discursos e diálogos, através da interface semântica/pragmática. Nesta proposta teórica, as relações discursivas são entendidas como relações que estabelecem ligações entre os significados das frases. Esta proposta divide-se nos seguintes pontos essenciais (Asher & Lascarides, 2003):

1. Segmentação do Discurso: O discurso é dividido em segmentos, cada um representando uma unidade semântica, como uma frase ou um grupo de frases. Esses segmentos são chamados de *Discourse Units* (DUs).
2. Relações de Coerência: As DUs são conectadas por relações de coerência, que são relações semânticas ou pragmáticas que indicam como uma DU se relaciona com outra. Exemplos de relações de coerência incluem *Narration*, *Explanation*, *Elaboration*, *Contrast*, e *Result*.

Nesta proposta a inferência das relações discursivas depende de fontes linguísticas, como o léxico e a semântica composicional, mas também de fontes não linguísticas, como o conhecimento do mundo dos falantes (cf. Silvano, 2010).

A proposta do PDTB é uma das mais usadas para fazer a anotação de corpora em várias línguas. Tal como refere Oliveira *et al.* (2010), esta abordagem “considera as relações retóricas como predicados binários, podendo o primeiro argumento localizar-se na frase que contém o conector, na frase anterior, na frase posterior, na frase não adjacente e podendo corresponder a uma frase simples, a uma frase parcial ou a um conjunto de frases”. A anotação das relações discursivas baseia-se na informação lexical fornecida pelos conectores (Prasad *et al.* 2008).

## **2.3. ISO 24516-8**

### **4.2.1. Os princípios da ISO 24617-8**

A ISO 24617 *Language resource management – Semantic annotation framework* (*SemAF*) propõe um esquema de anotação semântica e está dividida em várias partes que cobrem diversos aspetos da anotação, nomeadamente a parte referencial, a parte temporal, a parte das funções semânticas, entre outras. Uma dessas partes a parte 8 - *Semantic relations in discourse, core annotation schema* (*DR-core*) (ISO, 2016) aborda a anotação das relações discursivas.

A ISO 24617-8 é uma proposta utilizada para a anotação de relações discursivas a nível global, desenvolvida com o objetivo de tornar a anotação mais eficaz e uniforme, alcançando assim a abordagem interoperável.

As relações discursivas agrupam-se em dois grandes grupos, as relações assimétricas, e as relações simétricas, como se pode ver na Tabela 1.

Relações discursivas	
Assimétricas	Simétricas
<i>Cause</i>	<i>Conjunction</i>
<i>Expansion</i>	<i>Contrast</i>
<i>Asynchrony</i>	<i>Synchrony</i>
<i>Concession</i>	<i>Similarity</i>
<i>Elaboration</i>	<i>Disjunction</i>
<i>Exemplification</i>	<i>Restatement</i>
<i>Manner</i>	
<i>Condition</i>	
<i>Negative Condition</i>	
<i>Purpose</i>	
<i>Exception</i>	
<i>Substitution</i>	
<i>Functional dependence</i>	
<i>Feedback dependence</i>	

**Tabela 1: Relações discursivas**

Segundo esta proposta teórica, as relações de sentido são estabelecidas entre dois ou mais segmentos que descrevam situações/ eventualidades distintas. Cada segmento que constitua cada uma destas relações discursivas é chamado de argumento. O primeiro argumento é representado como Arg.1 e o segundo como Arg.2, e a cada argumento é atribuído um papel semântico. No caso das relações assimétricas a cada argumento é atribuído um papel semântico distinto e no caso das relações simétricas cada argumento recebe o mesmo papel semântico. Os exemplos (1) e (2) ilustram esta distinção.

(1) “Se eu gostava de um rapaz, pedia sempre a opinião a Okoloma.” (TT3; linha 5)<sup>1</sup>

**Arg.1:** Se eu gostava de um rapaz, (Antecedent)

**Arg.2:** pedia sempre a opinião a Okoloma. (Consequent)

(2) : “Okoloma morava na rua e cuidava de mim como um irmão mais velho” (TT3; linha 4)

**Arg.1:** Okoloma morava na rua (Synchrony 1)

**Arg.2:** e cuidava de mim como um irmão mais velho. (Synchrony 2)

No exemplo (1), é estabelecida a relação discursiva assimétrica *Condition*, em que os argumentos recebem papéis semânticos diferentes, neste caso Arg.1 desempenha o papel

<sup>1</sup> Esta notação diz respeito à Ted Talk (TT) em que o exemplo se encontra, assim como a linha da folha de Excel em que pode ser encontrada.

semântico de *Antecedent* e Arg.2 o de *Consequent*. Esta relação é inferida sempre que um dos argumentos descreve uma situação não realizada, que quando realizada conduz à situação descrita no outro argumento. No exemplo (2), é estabelecida a relação discursiva simétrica *Synchrony*, pois há sobreposição temporal e, neste caso, os dois argumentos recebem o mesmo papel semântico: Arg.1 recebe o papel semântico de *Synchrony 1* e Arg.2 o de *Synchrony 2*.

#### 4.2.2. Aplicação da ISO 24617-8 em Português Europeu

A ISO 24617-8 não tem sido muito aplicada à anotação de dados nas diversas línguas. Em português europeu, há apenas dois estudos que usam esta proposta para a análise das relações discursivas em diferentes tipos de dados.

O primeiro estudo resultou na criação do *Discourse Relations In Perfect Participial Sentences Corpus* (DRIPPS<sup>2</sup>) (Silvano *et al.* 2023). Esta primeira versão do DRIPPS reúne 993 frases com orações adverbiais com o gerúndio composto em quatro variedades do português (português europeu, português do Brasil, de Moçambique e de Angola) e em inglês britânico, anotadas com relações discursivas de acordo com a norma da ISO 24617-2-8.

O segundo estudo, realizado no âmbito da iniciativa *Multilingual Multilingual Discourse Annotation Initiative* (MDAI) (Tomaszewska *et al.* 2024), tem como objetivo examinar a distribuição das relações discursivas em polaco, inglês e português europeu, bem como os desafios de aplicar a norma a estes dados. Neste estudo, os autores concluem que a ISO-24617-8 é eficaz para uniformizar as relações discursivas entre diferentes idiomas, mas enfrenta alguns desafios, como definir o escopo dos argumentos e anotar tipos específicos de relações, como a relação *Expansion*. As conclusões destacam a necessidade de definições mais claras para certas relações discursivas e diretrizes mais precisas para a delimitação dos argumentos, especialmente no que diz respeito à inclusão de conectores. Além disso, os autores enfatizam a importância de esforços colaborativos contínuos para expandir a inclusão de idiomas e dados mais abrangentes, com o objetivo de ampliar o alcance da pesquisa de discurso multilingue guiada pela ISO 24617-8 (Tomaszewska *et al.* 2024).

---

<sup>2</sup> Este corpus está disponível em <https://github.com/johnycordeiro/DRIPPS/>

### **3. O estudo**

Nesta secção, inicialmente, serão apresentadas as questões às quais este estudo se propõe responder, de seguida, será descrita a forma como se procedeu à recolha dos dados, assim como os parâmetros que foram tidos em consideração para análise. Por fim, serão apresentados e discutidos os resultados.

#### **3.1 Objetivos e questões de investigação**

Os dois objetivos do estudo foram os seguintes: (i) investigar as relações discursivas em monólogos de natureza narrativa em português europeu, procurando perceber de que forma são estabelecidas as relações discursivas, destacando os conectores que lhes estão associados, assim como a frequência com que cada relação discursiva ocorre no texto e (ii) determinar se a *ISO 25617-8* é adequada para anotar essas relações discursivas. Assim, este estudo de carácter exploratório procura responder a quatro questões centrais:

1. Quais são as relações discursivas mais frequentes em monólogos de carácter narrativo?
2. Quais são os conectores mais utilizados para assinalar as relações discursivas?
3. Quais são as relações discursivas que surgem mais associadas ao uso de conector?
4. Será a proposta da ISO adequada para a anotação das relações discursivas de forma completa e abrangente?

#### **3.2. Corpus e metodologia**

O corpus para este estudo exploratório foi constituído tendo como base os seguintes critérios:

1. Monólogos de carácter maioritariamente narrativo;
2. Monólogos em português europeu;
3. Monólogos com mais de 600 palavras.

Estes foram os critérios definidos, pois tipicamente os textos narrativos têm mais relações discursivas e mais variadas. Para além disso, com textos de maior extensão esperava-se também encontrar mais variedade e uma frequência maior de relações discursivas.

Tendo em conta estes critérios foram seleccionadas três Ted Talks, identificadas na Tabela 2.

	<b>Título</b>	<b>N.º de palavras</b>	<b>Fonte</b>
Ted Talk 1	“A história do gato”	685	<a href="#">Portuguese CAT Annotation 13 March.xlsx (sharepoint.com)</a>
Ted Talk 2	“O perigo da história única”	2802	<a href="https://www.ted.com/talks/chimamanda_ngozi_adichie_the_danger_of_a_single_story">https://www.ted.com/talks/chimamanda_ngozi_adichie_the_danger_of_a_single_story</a>
Ted Talk 3	“Todos devemos ser feministas”	3877	<a href="#">Chimamanda Ngozi Adichie: We should all be feminists   TED Talk</a>

**Tabela 2: Informação sobre as Ted Talks analisadas**

Formado o corpus, foram definidos os parâmetros de análise. Tendo em consideração as questões de investigação e os objetivos do estudo, considerou-se que na análise devia ser anotada não só a relação discursiva, como também os conectores que ocorressem a ligar argumentos das relações discursivas. Por isso, na anotação dos dados, realizada num documento de Excel, foram seguidos os seguintes passos:

- 1.º Identificação do segmento de texto em que está presente uma relação discursiva.
- 2.º Identificação do argumento 1 e argumento 2.
- 3.º Identificação da relação discursiva.
- 4.º Identificação do papel semântico dos argumentos 1 e 2.
- 5.º Identificação e classificação do conector que ligasse os argumentos da relação discursiva.

De notar que, para este estudo, só foram considerados como conectores as conjunções e locuções conjuncionais adverbiais e advérbios conectivos.

A Figura 1 ilustra como foi feita a anotação dos dados.

	A	B	C	D	E	F	G	H
	Segmento do texto	Connective		Argument 1	Argument 2	Discourse relation	Argument 1 role	Argument 2 role
1	Gostaria de começar a falar de um dos meus maiores amigos, Okoloma Maduwesi. Okoloma morava na minha rua e cuidava de mim como um irmão mais velho.			Gostaria de começar a falar de um dos meus maiores amigos, Okoloma Maduwesi.	Okoloma morava na minha rua e cuidava de mim como um irmão mais velho.	Expansion	Narrative	Expander
2	Okoloma morava na minha rua e cuidava de mim como um irmão mais velho.	E	Conjunção coordenativa copulativa	Okoloma morava na minha rua	e cuidava de mim como um irmão mais velho.	Synchrony	Synchrony 1	Synchrony 2

**Figura 1: Print screen da folha de Excel onde foram anotados os dados**

Definidos os passos da anotação, seguiu-se a fase do treino. Assim, com base nas definições e exemplos apresentados na ISO 24617-8, foi anotada a Ted Talk 1, primeiramente em conjunto com uma anotadora especialista em relações discursivas, e depois, individualmente. Nesta fase de treino, houve várias reuniões para esclarecer dúvidas e uniformizar alguns critérios de anotação. De seguida, passou-se à anotação das Ted Talks 2 e 3, sempre acompanhada de reuniões frequentes para discutir casos problemáticos e chegar a um consenso.

Terminada a anotação, foi realizada uma análise qualitativa e quantitativa dos resultados.

### 3.3. Apresentação e discussão dos resultados

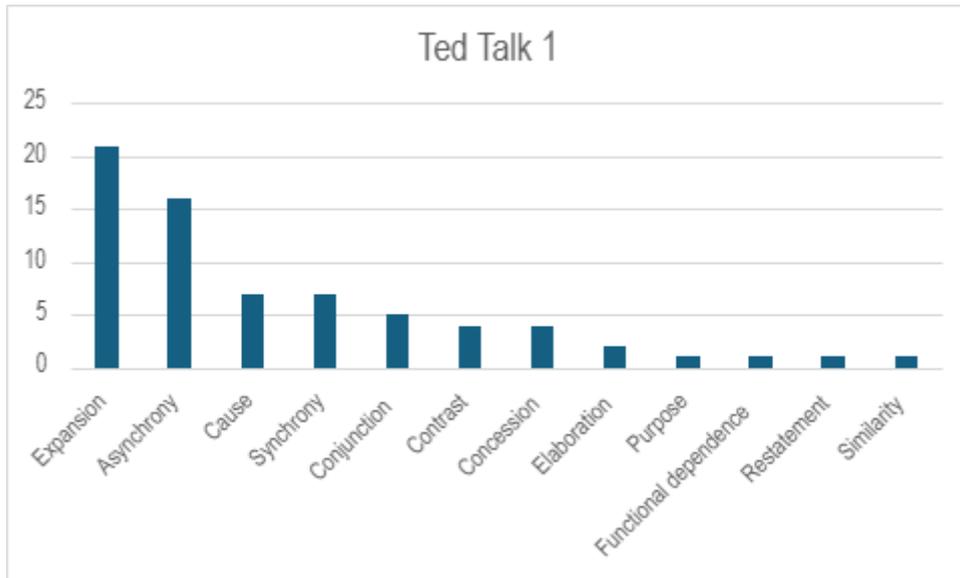
A Tabela 3 sistematiza alguns dos resultados gerais da anotação.

	Número de segmentos	Número de tipos diferentes de relações discursivas	Número de conectores
Ted Talk 1	70	12	11
Ted Talk 2	293	18	17
Ted Talk 3	534	18	14

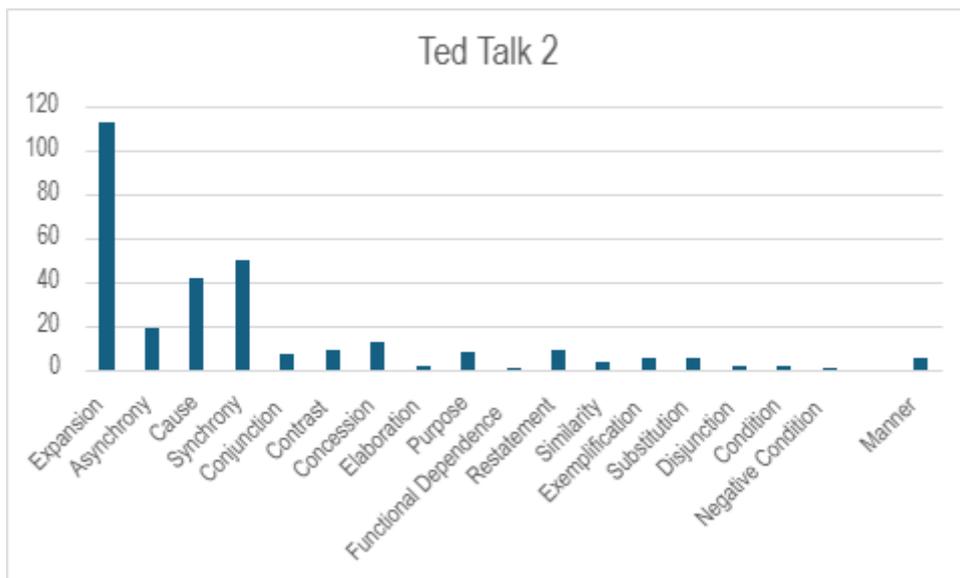
**Tabela 3: Resultados gerais da anotação**

#### 3.3.1. Relações discursivas

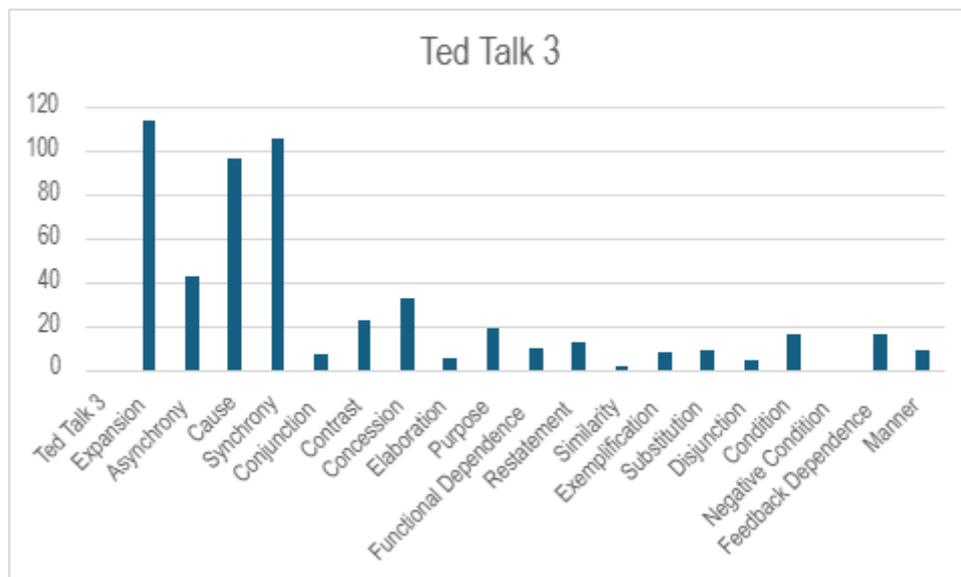
A comparação da distribuição das relações discursivas pelas três Ted Talks, apresentadas nos gráficos 1, 2 e 3 permite-nos concluir que as duas últimas têm características mais semelhantes do que a primeira, o que é expectável dada a sua extensão.



**Gráfico 1: Relações discursivas na Ted Talk 1**



**Gráfico 2: Relações discursivas na Ted Talk 2**



**Gráfico 3: Relações discursivas na Ted Talk 3**

A observação dos gráficos permite verificar que nas três Ted Talks analisadas a relação discursiva mais frequente é a *Expansion*, tanto na segunda como na terceira Ted Talk seguem-se as relações discursivas *Synchrony* e *Cause*. No entanto, na primeira Ted Talk observa-se uma maior frequência da relação discursiva *Asynchrony*, o que indicia que esta Ted Talk apresenta mais segmentos de narração do que as restantes, uma vez que a relação discursiva *Asynchrony* se caracteriza por representar uma relação de sucessividade temporal.

Já a grande frequência da relação discursiva *Expansion* nas três Ted Talks pode ser explicada por estarmos perante monólogos de carácter narrativo, sendo, por isso, comum a presença de uma relação discursiva cuja função seja fornecer mais detalhes acerca de uma dada situação. No entanto, o facto da relação discursiva *Expansion* ser tão abrangente e geral e dos critérios para a sua anotação não estarem bem definidos contribuí

para que esteja tão presente em textos narrativos. A grande frequência da relação discursiva *Cause* na Ted Talk 3 pode explicar-se por esta Ted Talk se tratar de um texto narrativo no qual a autora narra e explica acontecimentos ocorridos ao longo da sua vida.

Os exemplos (3), (4), (5), (6) ilustram as relações discursivas mais frequentes nas três Ted Talks.

(3) “Encontraram um sobrevivente inesperado um gato preto e branco agarrado a uma tábua que flutuava” (*Expansion*); (TT1; linha 6).

**Arg.1:** Encontraram um sobrevivente inesperado (*Narrative*)

**Arg.2:** um gato preto e branco agarrado a uma tábua que flutuava. (*Expander*)

(4) “Nos meses que se seguiram este gato caçou ratos e levantou o espírito dos britânicos até que foram atingidos por um torpedo que despedaçou o casco e afundou o navio.” (*Asynchrony*); (TT1; linha 10).

**Arg.1:** Nos meses que se seguiram este gato caçou ratos e levantou o espírito dos britânicos (*Before*).

**Arg.2:** até que foram atingidos por um torpedo que despedaçou o casco e afundou o navio. (*After*).

(5) “Todas as minhas personagens eram brancas e de olhos azuis. Brincavam na neve. Comiam maçãs.”(*Synchrony*); (TT2; linha 12).

**Arg.1:** Todas as minhas personagens eram brancas e de olhos azuis. (*Synchrony 1*).

**Arg.2:** Brincavam na neve. Comiam maçãs. (*Synchrony 2*).

(6) “as feministas são mulheres infelizes porque não conseguem arranjar um marido.” (*Cause*); (TT3; linha 41).

**Arg.1:** as feministas são mulheres infelizes (*Result*)

**Arg.2:** porque não conseguem arranjar um marido. (*Reason*)

Em (3) é possível inferir a relação discursiva *Expansion*, uma vez que esta se caracteriza por o Arg.2 fornecer mais detalhes para a compreensão do Arg.1, recebendo o Arg.1 o papel semântico *Narrative* e o Arg.2 *Expander*. Em (4), estamos perante uma relação de sucessividade temporal marcada pelo conector “até que”, o Arg.2 ocorre posteriormente ao Arg.1, recebendo o papel semântico *After* e o Arg.1 de *Before*. Em (5) é estabelecida mais uma vez uma relação temporal, no entanto, de sobreposição temporal. É inferido que as duas situações ocorrem em simultâneo, recebendo, por isso, o Arg.1 o papel semântico *Synchrony 1* e o Arg.2 *Synchrony 2*. Em (6) é inferida uma relação de causalidade, em que o Arg.2 representa a causa da situação representada por o Arg.1, que é o resultado.

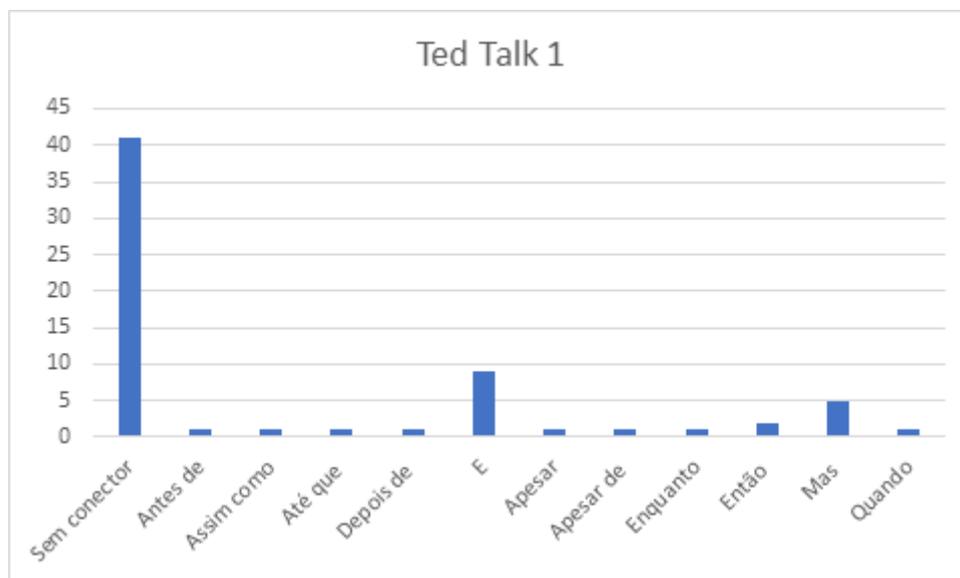
### 3.3.2. Conectores

No nosso estudo, analisamos ainda o uso de conectores nos três textos que constituem o nosso corpus. A tabela 4 ilustra o número de conectores que existe em cada Ted Talk.

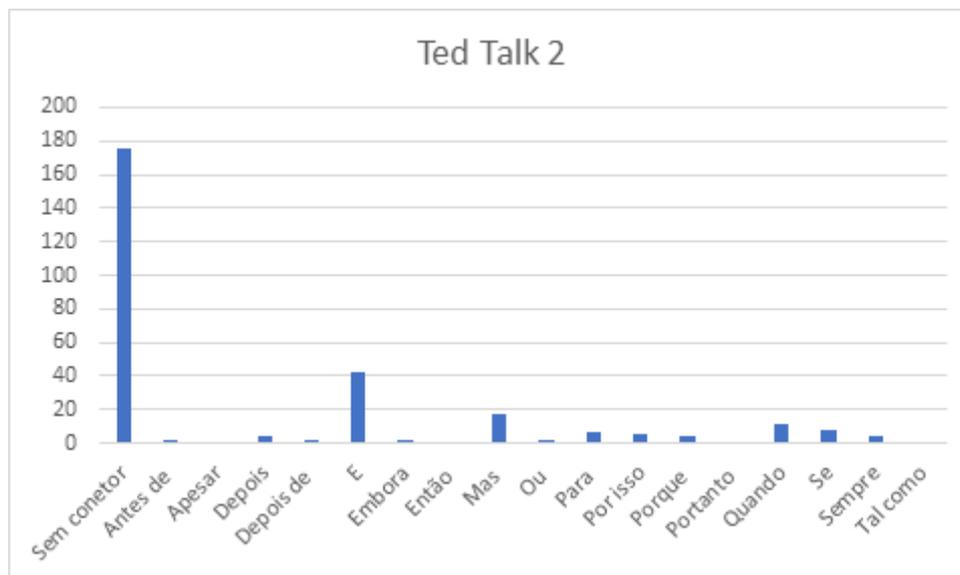
Ted Talk 1	Ted Talk 2	Ted Talk 3
11	17	14

**Tabela 4: Número de diferentes conectores presentes em cada Ted Talk**

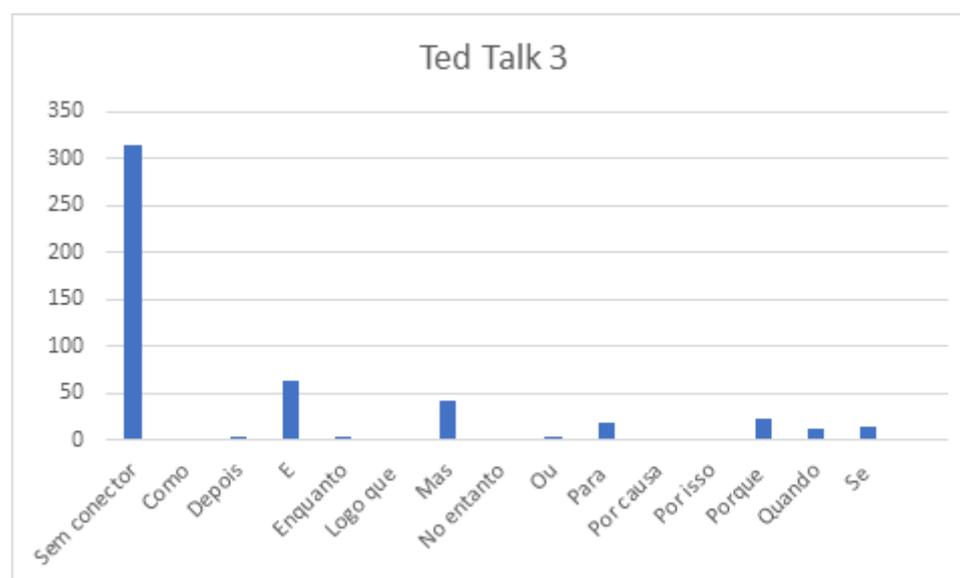
Nos gráficos 4, 5 e 6 é possível observar os conectores presentes nas três Ted Talks.



**Gráfico 4: Conectores presentes na Ted Talk 1**



**Gráfico 5: Conectores presentes na Ted Talk 2**



**Gráfico 6: Conectores presentes na Ted Talk 3**

Como é possível observar nos gráficos apresentados, nas três Ted Talks o conector mais frequente é a conjunção coordenativa copulativa “e”, como ilustrado em (7), (8), (9) e (10):

(7) “Nos meses que se seguiram este gato caçou ratos e levantou o espírito dos britânicos.” (*Conjunction*); (TT1; linha 9).

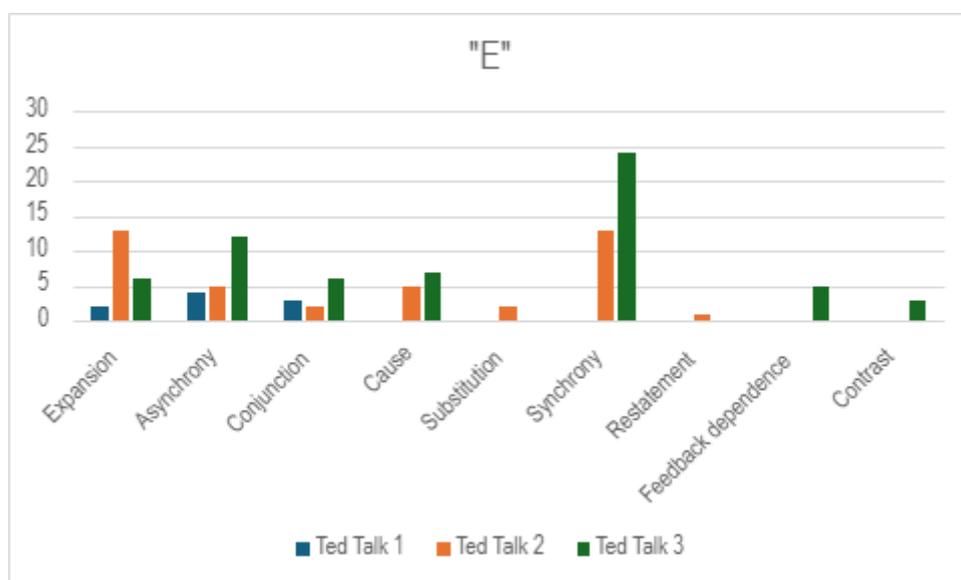
(8) “Não havia muitos disponíveis. **E** não eram tão fáceis de encontrar como os livros estrangeiros.” (*Expansion*) (TT2; linha 34).

(9) “Okolma morava na minha rua **e** cuidava de mim como um irmão mais velho.” (*Synchrony*) (TT3, linha 3).

(10) “Os homens têm testosterona **e**, em geral são fisicamente mais fortes que as mulheres.” (*Cause*) (TT3; linha 126).

Em (7), entre os Arg.1 e Arg.2 não é evidente que haja uma relação de sobreposição, ao contrário de (9), em que essa relação temporal ocorre. Por isso, num caso é inferida a relação de *Conjunction* e noutro a de *Synchrony*. Já em (8) a conjunção “e” introduz mais detalhes sobre a situação do primeiro argumento, daí a relação de *Expansion*. É de frisar que no exemplo (10) a relação discursiva *Cause* introduzida pela conjunção coordenativa copulativa “e” pode facilmente confundir-se com a relação discursiva *Expansion*, uma vez que também está a fornecer mais detalhes sobre a entidade “homens”. No entanto, a relação *Cause* torna-se mais marcada, uma vez que o Arg.2 (“e, em geral são fisicamente mais fortes que as mulheres”) é o resultado do Arg.1 (“Os homens têm testosterona”), recebendo, portanto, o papel semântico de *Result* e Arg.1 o de *Reason*.

Como é possível observar pelos exemplos apresentados, e como ilustra o gráfico 7, o conector “e” está associado a uma grande diversidade de relações discursivas:



**Gráfico 7: Relações discursivas associadas ao conector “e”**

Verificou-se também que as relações discursivas mais associadas ao uso de conector são as relações *Concession* e *Asynchrony*. Tal resulta da Relação *Concession* estar intrinsecamente ligada ao uso de conector, uma vez que é difícil e pouco usual expressar uma concessão sem o auxílio de um conector, como se pode ver no exemplo (11):

(11) “Claro que não era permitido usar o ponteiro, **mas** era uma hipótese excitante para mim, aos 9 anos.” (TT3; linha 74).

Neste caso, seria possível retirar o conector, mas a relação de concessão não seria tão fácil de inferir. O conector mais utilizado para marcar esta relação discursiva é o “mas”, não a ligar orações coordenadas, mas sim frases.

Também a Relação Discursiva *Asynchrony* pode não estar associado nenhum conector, mantendo-se o valor de sucessividade temporal como em (12).

(12) ” Abri a bolsa, meti a mão dentro da bolsa” (TT3; linha 108).

No entanto, mais uma vez a relação é mais explícita se estiver realizado o conector como em (13).

(13) “e prestou o mesmo serviço noutras três embarcações - uma das quais também afundou - **antes de** se reformar e ir para a Casa dos marinheiros, em Belfast.” (TT1; linha 17).

Neste caso a relação de sucessividade temporal é claramente marcada pela locução subordinativa adverbial temporal “antes de”.

A relação discursiva *Disjunction* nunca ocorre sem a presença da conjunção coordenativa disjuntiva “ou”, uma vez que sem esta não seria possível manter o valor de disjunção, como ilustrado em (14).

(14) “Se uma mulher se arranja para uma reunião de negócios, tem que se preocupar em não parecer demasiado feminina **ou** se vai ser levada a sério ou não.” (TT 3; linha 422).

Já as relações discursivas *Purpose* e *Conjunction*, apesar de poderem ocorrer sem a presença de um conector, o seu sentido é mais marcado quando associadas à conjunção

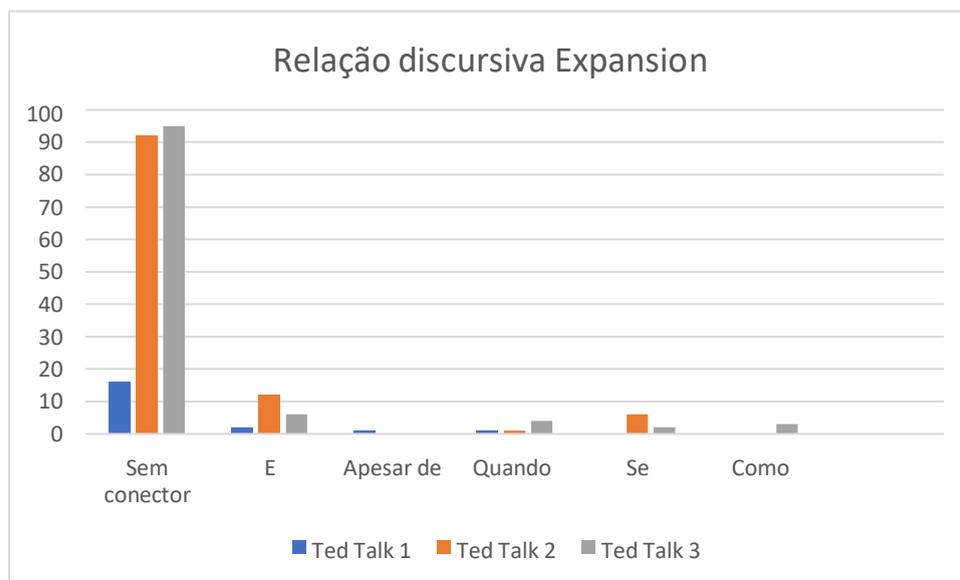
subordinativa adverbial final “para” e à conjunção coordenativa copulativa “e” respetivamente, como ilustrado em (15) e (16).

(15) “espera-se sempre que seja o rapaz a pagar, **para** provar a sua masculinidade” (TT3; linha 211).

(16) “Brincavam na neve. Comiam maçãs. **E** falavam muito do tempo, como era maravilhoso o sol ter aparecido” (TT2; linha 422).

Note-se que em (16) a relação discursiva *Conjunction* pode ser facilmente confundida com a relação discursiva *Synchrony*, no entanto, pelo contexto não nos é possível perceber que as duas situações ocorram ao mesmo tempo.

Por sua vez a relação discursiva *Expansion*, é a relação discursiva menos associada ao uso de conector, o que se explica pelo facto de se caracterizar por uma relação discursiva que fornece mais detalhes para a compreensão da situação descrita, não estabelecendo, portanto, uma relação que tenha associado um conector como pista para a sua inferência. Assim, esta relação está associada a uma grande variedade de conectores, como mostra o gráfico seguinte e como é ilustrado em (17), (18), (19) e (20):



**Gráfico 8: Conectores associados à relação discursiva *Expansion***

(17) “Gostaria de começar a falar de um dos meus maiores amigos, Okoloma Maduewesi. Okoloma morava na minha rua e cuidava de mim como um irmão mais velho.”(TT3; linha 2).

**Arg.1:** Gostaria de começar a falar de um dos meus maiores amigos, Okoloma Maduewesi.

**Arg.2:** Okoloma morava na minha rua e cuidava de mim como um irmão mais velho.

(18) “Por isso, eu fui uma leitora precoce. Lia livros para crianças, britânicos e americanos.” (TT2; linha 5).

**Arg.1:** Por isso, eu fui uma leitora precoce.

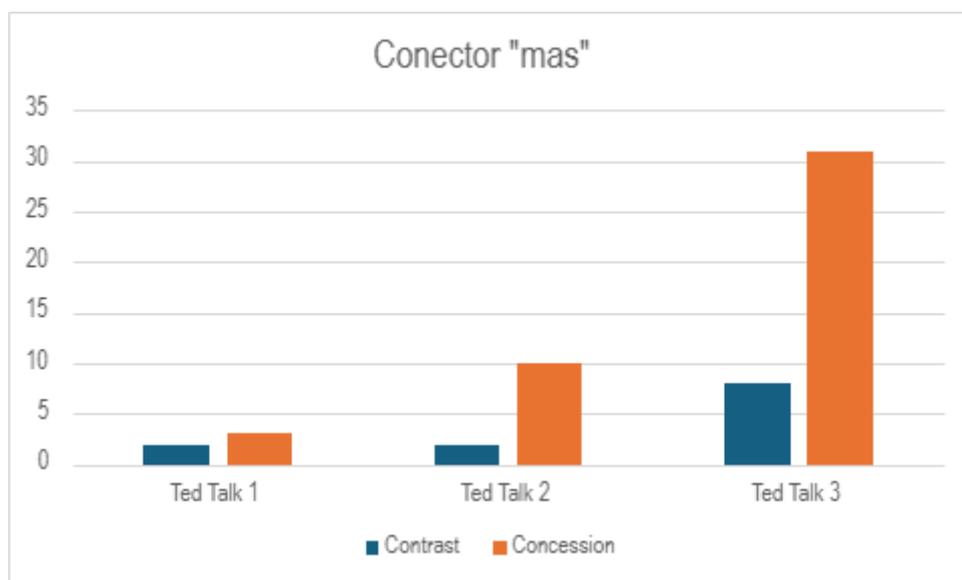
**Arg.2:** Lia livros para crianças, britânicos e americanos.

(19) “Não havia muitos disponíveis. **E** não eram tão fáceis de encontrar como os livros estrangeiros.” ( TT2; linha 34).

(20) “**Quando** os homens dizem isso, em geral é sobre algo que não deviam ter feito.” (TT3; linha 281).

Em (17) e (18) é possível inferir a relação discursiva *Expansion* sem a presença de um conector. Em ambos os exemplos, o Arg.2 fornece mais informação relativa ao que foi mencionado no Arg.1, não sendo necessário a presença de um conector que marque essa relação. Em (19) o Arg.2 fornece mais detalhes sobre a situação descrita no Arg.1, no entanto, com recurso à conjunção coordenativa copulativa “e”, marcando um valor de adição de nova informação. No exemplo (20) a relação discursiva é marcada pela conjunção subordinativa adverbial temporal “quando” em Arg.1 que enquadra o contexto em que a situação descrita em Arg.2 ocorre. Fornecendo mais detalhes para a sua compreensão.

A seguir ao conector “e”, o “mas” é o conector mais frequente em todas as Ted Talks. A conjunção coordenativa adversativa “mas” apenas pode estar associada a dois tipos de relações discursivas: *Contrast* e *Concession*, como mostra o gráfico seguinte e como é ilustrado em (21) e (22):



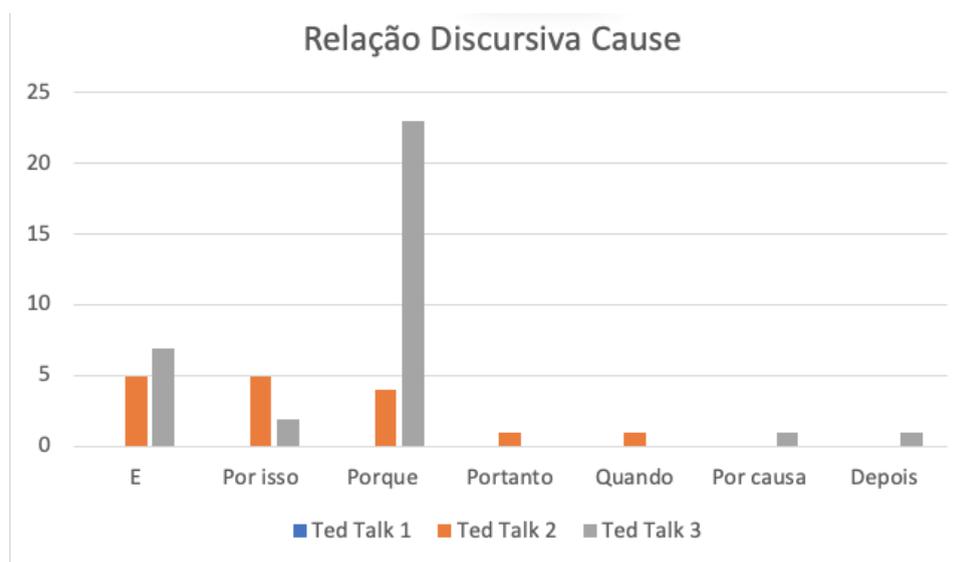
**Gráfico 9:** Frequência do conector “mas” nas relações discursivas *Contrast* e *Concession*

(21) “Durante anos, eu tive o desejo desesperado de provar cerveja de gengibre. **Mas** isso é outra história” (*Contrast*) (TT2; linha 26).

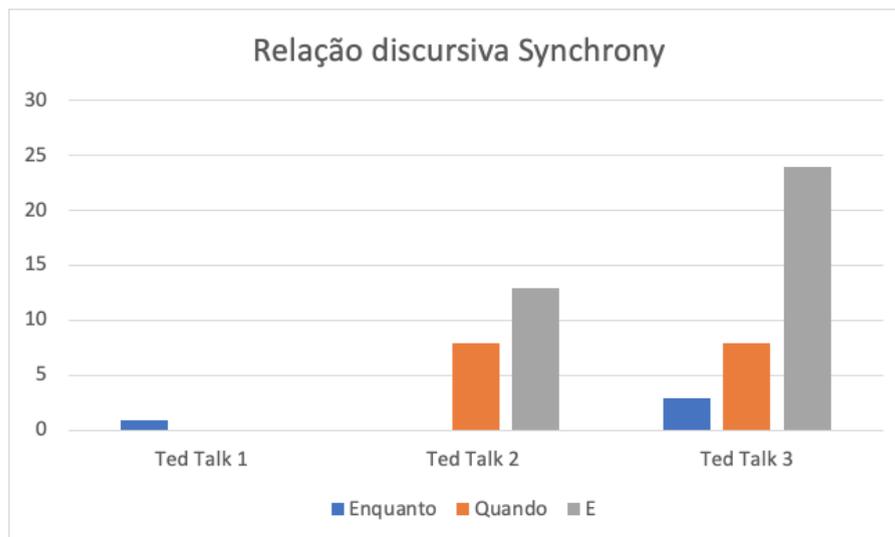
(22) “Devo dizer que, antes de ir para os Estados Unidos, eu não me identificava conscientemente como africana. **Mas** nos EUA, sempre que África surgia, as pessoas voltavam-se para mim.” (*Concession*) (TT2, linha 86).

Em (21) inferimos a relação discursiva *Contrast*, uma vez que o Arg.2 apresenta uma diferença relativamente ao Arg.1, estabelecendo um contraste entre as duas situações representadas pelos dois argumentos. Já em (22) é estabelecida não uma diferença entre os argumentos, mas uma quebra de expectativa, estabelecida pela conjunção coordenativa adversativa “mas”, face ao Arg.1. Note-se que seria possível trocar o “mas” pela conjunção subordinativa adverbial concessiva “embora” e a relação discursiva *Concession* seria igualmente estabelecida. Apesar da conjunção coordenativa adversativa “mas” ser o conector associado à relação discursiva *Contrast* por excelência, ocorre com maior frequência associado à relação discursiva *Concession*.

As relações discursivas *Cause* e *Synchrony* estão associadas a uma grande diversidade de conectores face a outras relações discursivas, como ilustram os gráficos 10 e 11 e os exemplos em (23), (24), (25), (26), (27) e (28).



**Gráfico 10: Conectores associados à relação discursiva *Cause***



**Gráfico 11: Conectores associados à relação discursiva *Synchrony***

(23) “que usa saltos altos **porque** gosta.” (TT3; linha 57).

(24) “e, por isso, ficou muito desapontada **quando** eu desencantei a minha cassete da Mariah Carey.” (TT2; linha 75).

(25) ”Eu tinha cedido à história única dos mexicanos **e** não podia sentir mais vergonha de mim.” (TT2; linha 146).

(26) “eu convenci-me que os livros pela sua própria natureza tinham de incluir estrangeiros, **e** tinham de ser sobre coisas com que eu não me identificava” (TT2; linha 30).

(27) “As coisas mudaram **quando** descobri livros africanos.” (TT2, linha 32).

(28) “E **enquanto** os gatos aprendiam a tolerar a presença de humanos e de outros gatos, na horasdas refeições, pensamos que os agricultores também toleravam os gatos em troca do controlo de pragas,” (TT1, linha 39).

Em (23) a relação de causalidade é estabelecida pela conjunção subordinativa adverbial causal “porque”, representando o Arg.2 a causa de Arg.1. Em (24), apesar de estarmos perante a conjunção subordinativa adverbial temporal “quando” no Arg.2, que pelo seu valor primário indicaria uma relação de sobreposição temporal ou o fornecimento de detalhes, neste caso exprime de forma mais marcada a causa da situação descrita no Arg.1. O mesmo ocorre em (25): é possível observar a presença da conjunção coordenativa copulativa “e”, cujo valor habitual é também o de adição, mas que neste exemplo exprime o resultado da situação descrita no Arg.1.

Já os três exemplos seguintes ilustram a relação *Synchrony*. Em (26) é possível perceber pelo contexto que a conjunção coordenativa copulativa “e” estabelece

meramente uma relação de sobreposição temporal. Em (27), apesar de poder ser inferido que o Arg.2 estabelece a causa de Arg.1, a leitura mais marcada, neste contexto, é a de sobreposição temporal. Em (28) a inferência da relação de sobreposição temporal é mais imediata, uma vez que o conector “enquanto” já contém o significado de ocorrência de mais do que uma situação em simultâneo.

### 3.4. Desafios da anotação

Na presente secção serão apresentadas algumas questões às quais a proposta teórica da ISO 24617-8 não consegue dar uma resposta cabal.

Durante a investigação foi possível perceber que várias relações discursivas geram especial ambiguidade, sendo dificilmente distinguidas, cabendo ao anotador seleccionar a que lhe parecerá mais evidente. Tal facto deve-se em parte à existência de relações discursivas muito abrangentes, passíveis de representar vários tipos de relações desentido, como é o caso, por exemplo, da relação discursiva *Expansion*, cujos critérios para a sua anotação não estão muito claros. Atentemos no seguinte exemplo em (29):

(29) “Quando o João era criança, partiu a perna.”

**Arg.1:** Quando o João era criança

**Arg.2:** partiu a perna

A relação discursiva estabelecida entre estes dois segmentos tanto encaixa na definição proposta pela ISO 24617-8 da relação *Expansion*, “*Arg2 is a situation involving some entity/entities in Arg1, expanding on the setting relevant for interpreting Arg1. The Arg1 and Arg2 situations are distinct.*” (ISO, 2016), como encaixa na definição também proposta pela ISO da relação *Synchrony* “*Some degree of temporal overlap exists between Arg1 and Arg2. All forms of overlap are included.*” É, portanto, pertinente fixar critérios mais rigorosos que permitam identificar de forma mais adequada as diferentes relações discursivas.

Por vezes, as relações discursivas de *Asynchrony* e *Cause* também podem ser difíceis de distinguir, como em (30) e (31):

(30) “Fazemos com que elas sintam, só por terem nascido mulheres, que já são culpadas de qualquer coisa. Depois, as raparigas crescem e tornam-se mulheres que não veem que têm desejos.” (TT3; linha 324).

**Arg.1:** Fazemos com que elas sintam, só por terem nascido mulheres, que já são culpadas de qualquer coisa.

**Arg.2:** Depois, as raparigas crescem e tornam-se mulheres que não veem que têm desejos.”

(31) “Depois de ter estado vários anos nos Estados Unidos, como africana comecei a perceber a reação da minha companheira de quarto para comigo.” (TT2; linha 95). **Arg.1:**

Depois de ter estado vários anos nos Estados Unidos, como africana

**Arg.2:** comecei a perceber a reação da minha companheira de quarto para comigo.

De acordo com a *ISO 24617-8* a proposta para a definição destas relações discursivas são respetivamente:

*Asynchrony:* “*Arg1 temporally precedes Arg2*”

*Cause:* “*Arg2 is an explanation for Arg1.*”

Como é possível observar, ambas as definições coincidem com o significado dos argumentos em cima apresentados, cabendo, portanto, ao anotador optar pela relação discursiva que lhe pareça mais marcada. No caso do exemplo (30), a relação discursiva mais marcada será a de *Cause*, apesar da presença do advérbio conectivo “depois”, uma vez que *Arg.2* apresenta de forma mais explícita, o resultado de *Arg.1*. Já no exemplo (31), apesar de continuarmos a ter “depois”, o contexto, a semântica composicional, faz com que a relação discursiva mais evidente seja a de *Asynchrony*.

Outro aspeto importante a referir é o facto de que, segundo a proposta teórica adotada para este estudo, apenas constituem argumentos situações expressas por um verbo. No entanto, muitas vezes temos situações expressas por um nome sem a presença de um verbo. Atentemos em (32) e (33):

(32) “Quando uma mulher se arranja **para uma reunião de negócios.**” (TT3; linha 422)

(33) “**A abundância de presas nos celeiros infestados de roedores** atraiu estes animais normalmente solitários.” (TT1; linha 37)

No exemplo (32), “para uma reunião de negócios” constitui um evento, no entanto, sem a presença de um verbo a proposta teórica da *ISO 24617-8* não prevê que esta situação seja anotada como um evento, logo não poderá ser ligado à situação “a mulher arranjar-se” por uma relação discursiva. Em (33), mais uma vez no primeiro argumento não está

presente nenhum verbo, mas sim um nome que representa um evento, e, por isso, não se anota qualquer relação discursiva.

Seria importante rever estes critérios de anotação, de forma a tornar mais produtiva e abrangente a anotação de relações discursivas, à semelhança do que foi concluído em (Tomaszewska *et al.* 2024) nomeadamente no que diz respeito aos critérios de anotação da relação discursiva *Expansion*.

## 1. Conclusões

Os dois objetivos principais deste estudo exploratório foram perceber quais as relações discursivas mais frequentes em monólogos de caráter narrativo, assim como aquelas que estão mais associadas ao uso de conector e quais são os mais frequentes e se a proposta da ISO 24617-8 é adequada para uma anotação eficaz e uniforme das relações discursivas. Para isso, construímos um corpus de três monólogos de natureza narrativa e fizemos a sua anotação no que diz respeito às relações discursivas, papel dos argumentos dessas relações discursivas e aos conectores usados.

Realizado o estudo foi possível responder às questões de investigação formuladas. Quanto à primeira questão de investigação, “Quais são as relações discursivas mais frequentes em textos narrativos?”, foi possível concluir que a relação discursiva *Expansion* é a relação discursiva mais frequente, seguida da *Synchrony* e *Cause*. A elevada frequência destas relações discursivas pode ser explicada pelo facto de ser comum numa narrativa haver muitos segmentos cuja função seja fornecer detalhes adicionais para a compreensão de dada situação (*Expansion*), assim como estabelecer relações de temporalidade simultânea (*Synchrony*) e de causalidade (*Cause*).

Quanto à segunda questão de investigação, “Quais são os conectores mais utilizados para assinalar as relações discursivas?”, os conectores mais utilizados para assinalar as relações discursivas são a conjunção coordenativa adversativa “mas”, mais associada à relação discursiva *Concession* e a conjunção coordenativa copulativa “e”, cujo uso é mais abrangente, estando associado a uma grande variedade de relações discursivas.

Relativamente à questão de investigação três, “Quais as relações discursivas é que surgem mais associadas ao uso de conector?”, as relações discursivas mais associadas ao uso de conector são a relação *Purpose*, *Concession*, *Conjunction*, *Disjunction* e *Contrast*. No entanto, na sua maioria as relações discursivas não estão associadas ao uso de

conector, sendo a relação discursiva inferida com base noutras fontes de informação linguística, como o léxico e a semântica composicional e o conhecimento do mundo.

Por fim, no que diz respeito à questão de investigação quatro, “Será a proposta da ISO adequada para a anotação de relações discursivas de forma completa e abrangente?”, podemos concluir que, apesar de a proposta da *ISO 24617-8* ser bastante produtiva para anotar de maneira mais eficiente as relações discursivas, foram verificadas algumas falhas resultantes dos critérios definidos para estabelecer quais os segmentos que podem constituir-se como argumentos de uma relação discursiva. Para além disso, há também a necessidade de rever a definição de algumas relações discursivas, uma vez que existem relações discursivas representadas por determinados segmentos que podem corresponder a mais do que uma definição de diferentes relações discursivas.

### Referências Bibliográficas

- Asher, Nicholas and Alexander Lascarides (2003) *Logics of Conversation*. Cambridge: University Press.
- Mann, W. C. and S. A. Thompson (1988) “Rhetorical Structure Theory: A Theory of Text Organization”, in *Text*, 8, 243-281.
- Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Oliveira, F.; L. F. Cunha e P. Silvano (2010) “Rhetorical Relations in Texts: the Contribution of Aspect”, in *Estudos Linguísticos*, n. 5, Julho 2010, Lisboa: Edições Colibri, 277-292.
- Silvano, Purificação (2010) *Temporal and Rhetorical Relations: The Semantics of Sentences with Adverbial Subordination in European Portuguese*. PhD dissertation. Porto: Faculdade de Letras da Universidade do Porto
- Silvano, P. and L. F. Cunha (2009) “O papel das restrições aspectuais nas relações retóricas: o caso das frases complexas com quando”, in *Textos Seleccionados do XXIV Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa: Associação Portuguesa de Linguística, 239-250.
- Silvano, Purificação (2013), “A Semântica das Frases com Subordinação Adverbial: o Contributo das Relações Retóricas”, in *Textos Seleccionados do XXVIII Encontro Nacional da APL*, Lisboa: APL, p. 595-614

Silvano, Purificação; Cordeiro, João; Leal, António & Pais, Sebastião. (2023). DRIPPS: a Corpus with Discourse Relations in Perfect Participial Sentences. In Proceedings of the 4th Conference on Language, Data and Knowledge, pages 470–481, Vienna, Austria. NOVA CLUNL, Portugal.

Tomaszewska, A.; Silvano, P; Leal A & Amorim, E. (2024) “ISO 24617-8 Applied: Insights from Multilingual Discourse Relations Annotation in English, Polish and Portuguese”

ISO. (2016). Language resource management- Semantic annotation framework (SemAF) - Part 8 - Semantic relations in discourse, core annotation schema (DR-core). Standard, Geneva, CH.

# A EMOÇÃO TRANSMITIDA PELA PALAVRAS: ESCALA DE VALÊNCIA, ALERTA E FAMILIARIDADE POR INDIVÍDUOS BILINGUES (L1 – PORTUGUÊS BRASILEIRO E L2 – INGLÊS)

Letícia Rodrigues Alberici

Orientado pela Professora Doutora Juliana Novo Gomes

## Resumo

O tema do processamento de palavras em segundas línguas (L2) tem sido um constante objeto de interesse da psicolinguística nas últimas décadas. No entanto, as conclusões sobre as respostas emocionais em L2 têm sido muito variáveis. Realizamos um estudo exploratório de classificação para medir explicitamente a forma como os leitores de Inglês como L2 respondem à *valência*, ao *alerta* e à *familiaridade* de 80 palavras do *corpus* elaborado a partir de algumas fontes, como o ANEW em Inglês (EN) e em Português do Brasil (PB). Investigamos como o efeito da proficiência em L2 afeta a forma como os leitores experienciam e classificam substantivos em PB e EN. Os dados indicam que falantes de PB como L1 que relataram maior proficiência em Inglês apresentaram respostas emocionais mais acuradas do que aquelas respostas dadas por falantes com menor proficiência em Inglês. Ou seja, falantes com maior proficiência tendem a classificar as palavras negativas usando valores mais negativos; palavras positivas, mais positivamente; palavras neutras, de maneira mais neutra, e, as palavras tabus são julgadas mais negativamente e com maior alerta do que os respectivos julgamentos na L2. Além disso, as diferenças de familiaridade refletiram a proficiência na L2 - quanto maior a proficiência, maior a familiaridade com as palavras testadas em todos os tipos de palavras (negativas, positivas, tabus e neutras).

Estes resultados realçam a importância da ecologia comportamental na aprendizagem de línguas, particularmente no que se refere ao processamento emocional de palavras.

Palavras Chave: emoção; valência; alerta; familiaridade; bilinguismo; Português

Brasileiro; Inglês

## **Abstract**

The subject of word processing in second languages (L2) has been a constant object of interest in psycholinguistics over the last few decades. However, findings on emotional responses in L2 have been highly variable. We conducted a pilot classification study to explicitly measure how non-native readers of English respond to the valence, arousal and familiarity of 80 words from the *corpus* drawn from some sources, such as ANEW in English (EN) and in Brazilian Portuguese (BP). We investigate how the effect of L2 proficiency, study time and formal or informal exposure affect the way readers experience and classify nouns in BP and EN. The data indicate that speakers of BP as L1 who reported higher English proficiency had more accurate emotional responses than those responses given by speakers with lower English proficiency. In other words, speakers with higher proficiency tend to classify negative words using more negative values; positive words, more positively; neutral words, more neutrally; and taboo words are judged more negatively and with greater arousal values than the respective judgments in L2. In addition, the differences in familiarity judgment reflects proficiency levels in L2 - the higher the proficiency, the greater the familiarity with the words tested in all types of words (negative, positive, taboo and neutral). These results highlight the importance of behavioral ecology in language learning, particularly with regard to the emotional processing of words.

Keywords: emotion; valence; arousal; familiarity; bilingualism; Brazilian Portuguese; English

## INTRODUÇÃO

Uma questão importante no âmbito das ciências da linguagem é a forma como a produção, a compreensão, a aprendizagem e a memória linguística podem ser moduladas por fatores extralinguísticos, como a emoção. Quando uma pessoa ouve ou lê uma palavra, ao longo do processamento linguístico, estudos recentes (Ku et al., 2020, Silva-Nasser, 2023, Kaczmarek & Harmon-Jones, 2024) mostram que também uma carga emocional é associada àquilo que está sendo interpretado, que pode variar de indivíduo para indivíduo, mas que em muitos casos esta interpretação é compartilhada por grupos ou conjunto de indivíduos. Ou seja, quando lemos uma palavra, como *amor*, assim como somos capazes de perceber a sua forma e o seu significado, também somos capazes, mesmo nunca tendo experienciado o sentimento de amar, de perceber algumas características básicas, como se esta é ou não uma palavra positiva; se traz ou não alguma forma de alerta ou, ainda, se esta é uma palavra muito ou pouco comum na nossa língua. Mas como estudar a emoção provocada pelas palavras, algo inicialmente tão subjetivo e impreciso, de maneira a produzir análises científicas relevantes? E, mais especificamente, como investigar a emoção transmitida nas palavras, quando expressas na primeira língua (L1) de um indivíduo? Haverá diferenças entre a emoção veiculada por palavras na primeira língua ou numa língua estrangeira (L2)?

Nos últimos anos, foram criadas normas para várias características lexicais, como *emocionalidade* (Kristensen, Gomes, Justo, & Vieira, 2011; Oliveira, Janczura, & Castilho, 2013), *concretude* (Janczura et al., 2007), *associação livre* (Janczura, Castilho, Keller & Oliveira, 2017) e *familiaridade* (Leitão, Figueira, & Almeida, 2010) em diversas línguas, incluindo o Português Brasileiro (PB). Embora alguns estudos tenham avaliado diferentes características das palavras, nenhum deles avaliou índices de valência, alerta e familiaridade para o mesmo conjunto de estímulos em Português Brasileiro (PB) e Inglês (EN), como L1 e L2, respectivamente.

Com este estudo procuramos preencher essa lacuna da literatura. Portanto, um dos objetivos deste trabalho foi obter classificações de valência (*valence*), alerta (*arousal*) e familiaridade (*familiarity*) para um conjunto único de palavras em PB e em EN.

## 1.1 Um pouco de história

A ciência começa a olhar para as emoções na segunda metade do século XIX, com estudos experimentais de Charles Darwin e de G. B. Duchenne de Boulogne, e logo depois com William James e Walter B. Cannon. Entretanto, a temática da emoção é deixada de lado no século seguinte, dada a sua dubiedade metodológica e inexatidão conceitual e estrutural. Nas últimas décadas, porém, pesquisadores debruçaram-se no estudo da emoção nos seus mais diversos domínios, sendo a psicolinguística um deles, privilegiando a pesquisa no âmbito do processamento linguístico, especialmente nas primeiras línguas (L1) e, mais recentemente, no âmbito das segundas línguas (L2).

A forma encontrada pela ciência (Bradley & Lang, 1999) para perscrutar a emoção no que diz respeito às propriedades afetivas da língua, em específico à carga emocional que as palavras evocam, consiste na avaliação de um conjunto de palavras com base em duas dimensões<sup>3</sup> (*valência* e *alerta*), que pressupõem a ideia de que uma emoção responde a dois estímulos: os apetitivos e os aversivos. A primeira dimensão, a *valência* (em Inglês, *valence*), consiste na sondagem do quão positiva ou negativa, i.e. quão agradável ou desagradável, uma palavra é. Palavras como *amor* e *feliz* têm a tendência de serem reconhecidas como palavras positivas e, de fato, apresentam uma valência alta, enquanto palavras como *podre* e *assassino* apresentam uma baixa valência.

A outra dimensão avaliada é, em Inglês, *arousal*. Traduzido geralmente como *alerta*, *excitabilidade* ou *intensidade*, ele corresponde ao quão relaxada ou estimulada uma pessoa fica ao se deparar com uma palavra. Ou seja, é o grau de agitação que uma palavra pode causar. Palavras como *vida* e *morte*, embora com valências opostas, apresentam um nível parecido de alerta, enquanto que *raiva* e *paz* apresentam níveis de alerta muito discrepantes.

Embora ainda não haja consenso, alguns estudos (Kim, Park, Lee e Sohn, 2001, citado por Oliveira et al., 2013) tentam identificar regiões cerebrais envolvidas no processamento das informações de valência e alerta. No geral, a valência parece ativar a amígdala direita, enquanto o córtex pré-frontal está relacionado com o alerta.

---

<sup>3</sup> No estudo original há ainda a dominância (*dominance*), que se refere a quão dominados ou em controle os falantes acham que estão sobre as próprias emoções face a uma palavra.

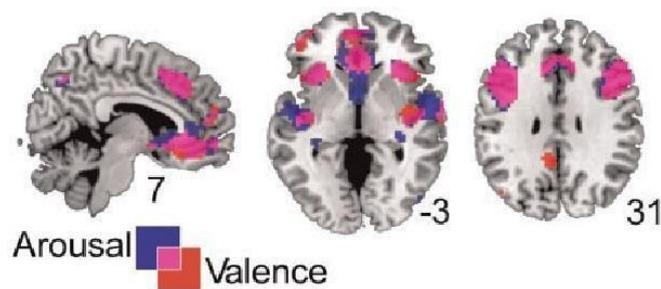


FIGURA 1: Valência (vermelho) e alerta (azul) no córtex cerebral.<sup>4</sup>

## 1.2 Normas de Medidas Afetivas (ANEW)

O ANEW (*Affective Norms for English Words*), ou Normas de Medidas Afetivas, foi o estudo que desenvolveu a metodologia de avaliação de valência e alerta. O ANEW foi desenvolvido pelos pesquisadores M. M. Bradley e P. J. Lang do National Institute of Mental Health (NIMH), Center for the Study of Emotion and Attention (CSEA), na Universidade da Flórida em 1999, e, desde então, tem sido reproduzido em diversas línguas, incluindo o Português nas duas variedades, PB (Kristensen et al., 2011) e PE (Soares et al., 2012). A versão em Inglês do ANEW (Bradley & Lang, 1999) contém 1.034 palavras, que foram avaliadas por participantes nas duas dimensões citadas anteriormente e em dominância por meio do *Self-Assessment Manikin* (SAM).

A metodologia SAM (Bradley & Lang, 1994) é composta por duas escalas: uma para a valência (Figura 1A) e outra para o alerta (Figura 1B). Ambas as escalas apresentam originalmente nove pontos de avaliação, sendo cinco deles intercalados, representados por bonecos com a figura humana. Para a valência, a representação do rosto do boneco varia de totalmente sorridente para uma figura descontente, passando por um boneco com rosto neutro. Quanto ao alerta, a variação se dá principalmente na parte central do corpo do boneco. Quanto mais estimulante é a palavra, independentemente da carga mais positiva ou negativa que ela carregue, mais motivado e energizado é o boneco, com olhos bem abertos, enquanto uma palavra menos estimulante deixa-o mais inerte e sonolento.

<sup>4</sup> A Figura 1 foi retirada do estudo de Bestelmeyer et al., 2017. Neste estudo, os autores fizeram uma análise neurolinguística das regiões do cérebro associadas à valência e ao alerta, sem, porém, apoiar uma segregação completa das estruturas associadas.

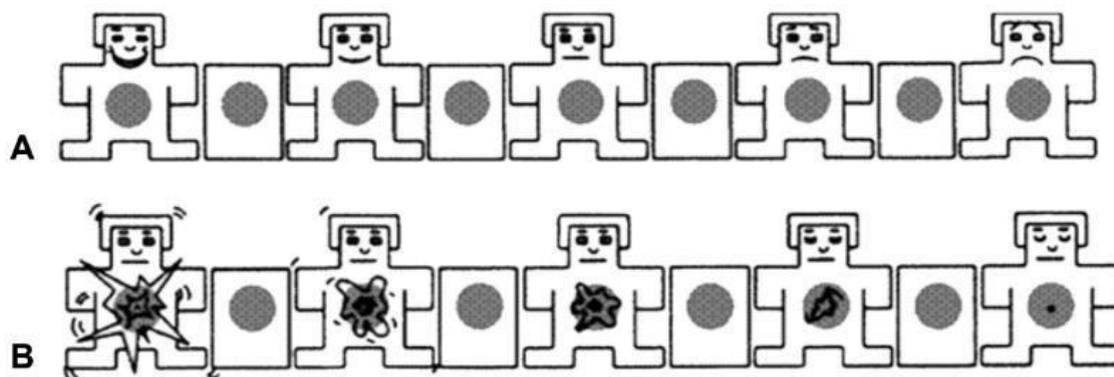


FIGURA 2: Escala do SAM para valência (A) e alerta (B).

Entretanto, como foi mencionado, o ANEW (Bradley & Lang, 1999) já foi traduzido e adaptado, completa ou parcialmente, para diversas línguas. O ANEW-Br (Kristensen et al., 2011), que é o ANEW adaptado ao PB, contou com julgamentos de valência e alerta para 1.046 palavras em PB. As palavras, traduzidas do Inglês para o Português por um linguista, foram traduzidas novamente para o Inglês para fins de verificação. Após ajustes e adaptações, a lista com 1.046 foi julgada por 755 universitários por meio do SAM e obteve indicações emotivas similares ao ANEW americano (Bradley & Lang, 1999), o que indica uma adequação do método.

No ANEW-Br (Kristensen et al., 2011), a valência da palavra *assassino* é 1,16. Já a palavra *liberdade* obteve 8,80 de valência. As valências de *assassino* e *liberdade* são os dois extremos do estudo (1,16 e 8,80). A média de valência é 5,16, com um desvio padrão de 2,20. Já os valores de alerta ficaram compreendidos entre 2,22 (*tranquilo*) e 7,67 (*assalto*). A média foi de 4,57 (DP = 1,04).

Estudos normativos como este servem de *corpus* para muitos outros estudos sobre como as emoções e a linguagem interagem entre si. Por exemplo, Nascimento et al., 2018, identificou, a partir do *corpus* anotado do ANEW-Br (Kristensen et al., 2011), um conjunto de palavras relacionadas à depressão, através do qual é possível identificar sinais de comportamento depressivo nas redes sociais.

Conforme apontado, as palavras avaliadas no ANEW-Br (Kristensen et al., 2011) apresentaram índices de valência e alerta similares às palavras do ANEW em EN (Bradley & Lang, 1999). Contudo, os estudos costumam elicitar respostas por falantes de primeiras línguas. O que aconteceria se um falante julgasse as mesmas palavras em L1 e L2? Poderíamos aventar alguma diferença caso os resultados fossem obtidos através da comparação entre L1 e L2? Poderíamos pensar que o nível de proficiência na L2 interferiria com o julgamento lexical da emoção transmitida?

### 1.3 Bilinguismo

O segundo objetivo deste trabalho é, a partir de um conjunto de classificações de valência (*valence*), alerta (*arousal*) e familiaridade (*familiarity*) para um único conjunto de palavras em PB e EN, verificar se indivíduos com diferentes níveis de proficiência em L2 atribuem valores similares ou distintos para palavras correspondentes na L1 e na L2. Ou seja, coletamos as classificações obtidas para as palavras nas três dimensões (valência, alerta e familiaridade) na primeira (PB) e na segunda língua (EN) com a intenção de percebermos as diferenças no processamento da carga emocional de uma palavra em decorrência do nível de proficiência em L2 dos falantes.

A questão central do presente trabalho é, então, a comparação de como as dimensões valência, alerta e familiaridade são *sentidas* na primeira (L1, PB) e na segunda língua (L2, EN). Buscamos então responder: as palavras carregam o mesmo peso emocional ou são sentidas mais intensamente na L1 ou na L2?

Por exemplo, muitas pessoas, principalmente aquelas que estão em relacionamentos afetivos multilíngues, relatam que falar *eu te amo* em sua L1 parece ser “mais forte” do que na L2. De igual modo, em momentos de grande emoção, como sustos, iminência de morte e excitação, as pessoas tendem a recorrer à sua L1 - uma vez que as palavras parecem não ter o mesmo *peso emocional* na L2. Isso se deve à familiaridade com a língua ou essa sensação de maior peso emocional na L1, que também se reflete na valência e no alerta das palavras, continua existindo mesmo quando o grau de familiaridade com a L2 é alto?

Há um crescente consenso de que as palavras são sentidas com mais intensidade na L1 do que na L2. Isso se expressa em decorrência de uma tendência neutralizante. Pesquisas mostram que os falantes tendem a neutralizar a carga emocional das palavras na L2 em comparação com a sua L1. Um exemplo é o estudo de Harris, Ayçiçeği & Gleason (2003), que examinou a reação de falantes diante de reprimendas e frases tabus na sua primeira língua, neste caso o turco, e, na L2 (EN).

Um estudo semelhante foi realizado por Margarida V. Garrido e Marília Prada (2018), no qual o Português Europeu (PE) era a L1 dos participantes, e o Inglês, a L2. Em ambas as investigações os resultados foram os mesmos: há maior atribuição de emoção às palavras e expressões quando ditas ou escritas na L1 do indivíduo. De acordo com o estudo de Garrido & Prada (2018), a valência e a intensidade emocional (alerta) foram significativamente amplificados na L1 do participante em comparação à L2.<sup>5</sup>

---

<sup>5</sup> É válido acrescentar que, ainda de acordo com o mesmo estudo de Garrido & Prada, e em consonância com Sheikh & Titone (2015), os participantes foram mais propensos a neutralizar expressões negativas e tabus na segunda língua do que palavras e expressões positivas.

É interessante notar que mesmo falantes com alto nível de proficiência na L2 reportam um sentimento menos intenso em relação a L2. Certamente uma pessoa com pouco conhecimento da língua sentiria menos emoção nas suas palavras e expressões, mas uma pessoa com grande familiaridade com o idioma também estaria sujeita à neutralização emocional?

Outros estudos (Degner et al, 2012 e Segalowitz, et al, 2008) parecem sugerir que a emoção é independente do nível de proficiência, visto que pessoas com grande competência na L2 também demonstram menos emotividade e naturalidade nesta língua. Entretanto, segundo a *Lexical Quality Hypothesis* (Perfetti, 2007), o nível do conhecimento linguístico de um indivíduo pode causar a neutralização emocional, posto que a falta de familiaridade com palavras menos frequentes pode constituir uma barreira para o pleno entendimento e, conseqüentemente, para o aprofundamento de emoção em um texto, por exemplo<sup>6</sup>.

#### **1.4 O estudo sugerido e o estudo modelo**

Com base na literatura, pretende-se no presente trabalho investigar se, de fato, os valores de valência e alerta são modulados pelo nível de proficiência e amplificados na L1 em comparação com a L2. Ademais, analisa-se se a familiaridade com uma determinada palavra na L2 desempenha um papel relevante nesta discussão. Propõe-se então um estudo com palavras em PB (L1) e em Inglês (L2), a fim de testar a relação entre valência, alerta e familiaridade nas duas línguas. Para tal, pretende-se uma reprodução em menor escala do estudo de Garrido & Prada (2018) em PB (L1) e em Inglês (L2).

Garrido & Prada (2018) investigaram a questão por meio de um estudo em Português Europeu (PE), que contava com um total de 230 participantes (majoritariamente mulheres, com média etária de 23 anos), todos voluntários e estudantes, provenientes de universidades portuguesas públicas e privadas. Todos relataram possuir o PE como L1 e o Inglês como L2. Fatores como a idade de aquisição da L2, a forma de aprendizado, a frequência de uso da L2 e os objetivos desse uso também foram verificados.

Quanto ao *corpus*, as autoras selecionaram 640 palavras (320 em Inglês e as suas correspondentes traduções em PE). As palavras subdividiam-se em negativas, positivas, neutras e tabus. Em cada língua, havia 110 palavras negativas (80 adjetivos e 30 substantivos), 110 palavras positivas (novamente 80 adjetivos e 30 substantivos), 60 palavras neutras (30 adjetivos e 30 substantivos) e 40 tabus. As palavras foram selecionadas a partir de algumas fontes, nomeadamente do ANEW (Bradley & Lang, 1999), da sua adaptação para o PE (Soares et al.,

---

<sup>6</sup> Além do nível de proficiência, alguns estudos (Clark & Paivio, 1991, e Harris, Gleason e Aycıçegi, 2006) se debruçaram a analisar outros aspectos que podem influenciar como as palavras são sentidas na segunda língua, tais como a concretude e a abstração das palavras e o contexto de aprendizagem ao qual o falante foi exposto.

2011), além de um banco de dados de palavras portuguesas avaliadas quanto à valência e familiaridade (Garcia-Marques, 2003) e de um dicionário de sinônimos; as palavras tabus foram escolhidas a partir de Janschewitz (2008). As palavras foram traduzidas do Inglês para o Português por dois especialistas independentes e, então traduzidas outra vez para o Inglês por outro especialista independente.

O procedimento do experimento deu-se presencialmente, e as avaliações foram feitas pelos participantes em papel. As 640 palavras foram separadas em listas de 80 palavras (40 em Inglês e as suas correspondentes traduções) para evitar a fadiga dos participantes. Todas as listas incluíram as quatro categorias de palavras (negativas, positivas, neutras e tabus) e as palavras formavam as listas de maneira aleatória. Uma mesma pessoa, porém, avaliou sempre o par de palavras correspondentes, embora elas não estivessem juntas a fim de evitar influências no julgamento. Uma mesma palavra foi julgada por, no mínimo, 26 e, no máximo, 32 participantes.

Os participantes julgaram cada palavra nas três categorias (valência, alerta e familiaridade) por meio de uma escala de 7 pontos, sendo 1, respectivamente, muito negativo, não muito intenso e não muito familiar; e 7 muito positivo, muito intenso e muito familiar. Então foi solicitado aos participantes que traduzissem as palavras do Inglês para o PE, a fim de que fosse garantido o entendimento do significado das palavras. Por fim, os participantes forneceram dados sócio-demográficos e algumas informações pessoais que poderiam influenciar os resultados.

## **2. O presente estudo**

### **2.1 Corpus**

Construiu-se um *corpus*<sup>7</sup> com 262 palavras (131 em PB e as suas correspondentes em EN). Das 131 palavras, há 30 negativas, 30 positivas, 31 neutras e 40 tabus. As palavras foram escolhidas a partir do *corpus* de Garrido & Prada (2018), selecionando-se, no caso das palavras negativas, positivas e neutras, apenas os substantivos. As palavras tabus foram retomadas na íntegra.

As palavras em língua inglesa selecionadas por Garrido & Prada (2018) foram mantidas, porém muitas traduções para PE careceram de adaptação para o PB, e para isso olhou-se para o ANEW-Br (Kristensen et al., 2011). O estudo de Kristensen et al., 2011, porém, usou a escala de 9 pontos e não observou dados da familiaridade das palavras.

Dessa maneira, as palavras em PE contidas no estudo de Garrido & Prada (2018) que também estavam no ANEW-Br (Kristensen et al., 2011) foram retomadas sem adaptações. Entretanto,

---

<sup>7</sup> A tabela com o *corpus* elaborado está disponível nos anexos.

as palavras que não foram investigadas no ANEW-Br (Kristensen et al., 2011) ou que são próprias do PE precisaram de adaptação para o PB. Para isso, traduziu-se as palavras a partir do Inglês por meio do *Cambridge Dictionary*, com a confirmação feita pelo *Longman Dictionary of Contemporary English Online*, e, quando necessário, usou-se também um dicionário de sinônimos. As traduções ou sinônimos ou ainda palavras no mesmo campo semântico, se fosse o caso, encontrados no ANEW-Br (Kristensen et al., 2011) foram usados no lugar das traduções literais em PE. Quando não encontrou-se tradução ou sinônimo apropriado no ANEW-Br (Kristensen et al., 2011), foi feita uma tradução livre com os dicionários supracitados.

No que tange às palavras tabus, o procedimento teve o mesmo ponto de partida, porém seguiu outro rumo. Procurou-se inicialmente palavras correspondentes no ANEW-Br (Kristensen et al., 2011), porém poucas foram encontradas. Olhou-se então para o estudo de Claudiane Silva-Nasser (2023), que analisou, entre outros aspectos, a valência, o alerta e a familiaridade de palavras tabus no Português Brasileiro, dialeto carioca (da região do Rio de Janeiro). Silva-Nasser replicou o estudo de Janschewitz (2008) e selecionou as palavras tabus por meio de um questionário que teve como base, principalmente, a familiaridade dos participantes com as palavras tabus.

Para o presente estudo, as palavras de Garrido & Prada (2018) com correspondências perfeitas no estudo de Silva-Nasser (2023) foram retomadas sem complicações. Aquelas que não foram encontradas, em decorrência da variação linguística entre o PE e o PB, foram adaptadas para o PB por meio dos dicionários citados e do site de sinônimos e, então, localizadas no estudo de Silva-Nasser (2023). Entretanto, as palavras cuja correspondência não foi encontrada de nenhuma forma (mesmo com adaptações ou sinônimos) foram traduzidas livremente por meio dos dicionários citados.

## ***2.2 Metodologia***

Realizou-se, no presente trabalho, um estudo exploratório a fim de testar a aplicabilidade do estudo sugerido acima.

### ***2.2.1 Participantes***

Para o estudo exploratório, foram convidados 16 participantes, 7 homens e 9 mulheres com média etária de 29,56 anos, que voluntariamente participaram no presente estudo. Nascidos nas regiões 5 do Brasil (9 sudeste, 3 nordeste, 2 sul, 1 centro e 1 norte), todos relataram possuir o PB como L1. Três participantes declararam possuir o Ensino Médio concluído, e o restante, 13 participantes, declararam ter o Ensino Superior.

Quanto à proficiência em Inglês, 4 participantes relataram não ter conhecimentos de Inglês, 4 são do nível básico, 4 do nível intermediário e 4 do nível avançado. Dos que têm conhecimentos da língua, 0 tiveram contato com o idioma antes dos 6 anos de idade, 8 entre os 6 e os 12 anos, 2 entre os 13 e os 17 anos e 2 a partir dos 18 anos. 4 nunca tiveram contato com o Inglês.

### 2.2.2 Estímulos

A lista de estímulos foi constituída a partir do *corpus* anteriormente descrito, elaborado no presente estudo a partir de, principalmente, Garrido & Prada (2018), Kristensen et al. (2011), Bradley & Lang (1999) e Silva-Nasser (2023). Das 262 palavras (131 em PB e as suas correspondentes traduções em Inglês), foram selecionadas 80 palavras (40 em EN e as suas correspondentes traduções em PB). A seleção foi feita a partir das palavras em Inglês, visto que as palavras em PE, de Garrido & Prada (2018), muitas vezes foram adaptadas para PB, e as palavras em PB não continham, por vezes, dados de valência, alerta e tampouco de familiaridade.

Assim, dos estímulos em Inglês, escolheu-se 10 palavras negativas, 10 positivas, 10 neutras e 10 tabus. Das negativas, selecionou-se as cinco palavras com menores indicativos de valência (respectivamente: *murderer, abuse, terrorist, hate, bomb*) e as cinco palavras com maiores indicativos de alerta (respectivamente: *war, murderer, panic, fear, extermination*). Quando houve repetição, substituiu-se a palavra repetida pela seguinte na lista (como é o caso de *murderer*, que foi substituída por *death*). Das palavras positivas, escolheu-se as cinco com maior valência (*paradise, truth, smile, kiss, friend*) e as cinco com menor alerta (*flower, profit, treasure, miracle, fame*). Das neutras, selecionou-se três palavras com menor valência (*handkerchief, cord, standard*) e duas com maior valência (*paper, hydrant*); duas com menor alerta (*bus, chair*) e três com maior alerta (*board, locker, chin*). O modelo das palavras neutras foi seguido para as palavras tabus (respectivamente: *bitch, slut, whore, ejaculation, orgasm, tits, snorf, cocksucker, sex, fornication*).

Dessa forma, os 80 estímulos foram os seguintes:

NEGATIVAS		POSITIVAS		NEUTRAS		TABUS	
ING	PB	ING	PB	ING	PB	ING	PB
abuse	abuso	fame	fama	board	quadro	bitch	vadia
bomb	bomba	flower	flor	bus	ônibus	cocksucker	boqueteira

NEGATIVAS		POSITIVAS		NEUTRAS		TABUS	
ING	PB	ING	PB	ING	PB	ING	PB
death	morte	friend	amigo	chair	cadeira	ejaculation	ejaculação
extermination	chacina	kiss	beijo	chin	queixo	fornication	fornicação
fear	medo	miracle	milagre	cord	corda	orgasm	orgasmo
hate	ódio	paradise	paraíso	handkerchief	tecido	sex	sexo
murderer	assassino	profit	lucro	hydrant	hidrante	slut	vagabunda
panic	pânico	smile	sorriso	locker	cofre	snorf	pum
terrorist	terrorista	treasure	tesouro	paper	papel	tits	tetas
war	guerra	truth	verdade	standard	molde	whore	puta

TABELA 1: Estímulos utilizados no estudo exploratório.<sup>8</sup>

### 2.2.3 Procedimentos

Para a realização deste estudo exploratório, foi utilizada a plataforma *web-based* PCIBex (Schwartz & Zehr, 2018). Esta plataforma é administrada pelo *Controlador PennController* da Universidade da Pennsylvania (UPenn) e recolhe e guarda os dados de cada participante recrutado para participar no estudo criado na plataforma. Os estudos criados são programados diretamente na Plataforma PCIBex através de *scripts* experimentais em linguagem *Javascript* (Anexo X).

A primeira etapa consistiu em criar um *script* no PCIBex em 7 blocos, descritos abaixo:

- (1) Termo de Consentimento Livre e Esclarecido (TCLE)
- (2) Coleta de Dados dos Participantes
- (3) Instruções
- (4) Treino e Familiarização com a Tarefa de Classificação em Escala
- (5) Tarefa de Classificação de Palavras em Português em Escala de 5 pontos
- (6) Tarefa de Classificação de Palavras em Inglês em Escala de 5 pontos
- (7) Agradecimento e Encerramento do Estudo

<sup>8</sup> Em anexo, encontram-se as tabelas completas com os dados de valência, alerta e familiaridade das respectivas palavras.

(1) TCLE

A primeira tela exibida aos participantes foi o *Termo de Consentimento Livre e Esclarecido*, no qual era apresentado o objetivo e o detalhamento da pesquisa, bem como a tarefa dos participantes e as consequências da participação anônima. Também, a equipe de pesquisadoras, os contatos e o registro da pesquisa eram apresentados. Ao final do TCLE, era pedido ao participante que confirmasse a compreensão do Termo e o aceite da participação voluntária no estudo.

(2) Coleta de Dados dos Participantes

Em seguida, os participantes viam uma tela para coleta de informações sócio-linguístico-demográficas, como: o dispositivo no qual estavam realizando a atividade, email, idade, lateralidade, escolaridade e alguns dados sobre o conhecimento e o uso da L2 (Inglês).

(3) Instruções

Na sequência, eram apresentadas as instruções detalhadas do estudo e um exemplo.

(4) Treino e Familiarização com a Tarefa de Classificação em Escala

Para que o participante se familiarizasse com a tarefa de classificação em escala de 5 pontos, um treino foi criado com 4 palavras para além das palavras de teste. Apenas após a familiarização no treino, o participante passava ao estudo.

(5) Tarefa de Classificação de Palavras em Português em Escala de 5 pontos

Após o treino, o participante passava ao estudo com palavras na L1, em Português. A tarefa era ler uma palavra na tela e julgar o grau de valência, alerta e familiaridade desta palavra em escalas de 5 pontos.

(6) Tarefa de Classificação de Palavras em Inglês em Escala de 5 pontos

Após a classificação de palavras na L1 (Português), o participante passava ao estudo com palavras na L2 (Inglês). O procedimento era exatamente o mesmo - ler e julgar o grau de positividade, excitabilidade e familiaridade da palavra em escalas de 5 pontos, porém as palavras estavam em Inglês.

(7) Agradecimento e Encerramento do Estudo

Ao final do teste, os dados eram salvos e, na tela, aparecia um agradecimento pela participação seguido do aviso de que o participante deveria fechar a tela do computador.

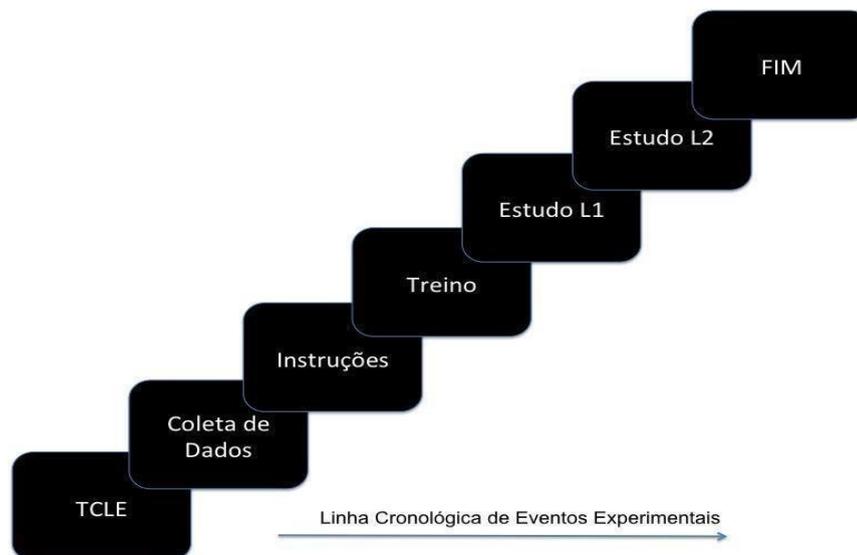


FIGURA 3: Sequência de Eventos Experimentais apresentados aos participantes do estudo.

Uma vez que o *script* estava criado, um teste piloto com dois participantes foi executado para assegurar que o programa estava funcionando, que não havia erros no *script* e que os dados estavam sendo gravados corretamente. Além disso, as instruções e estímulos foram avaliados uma vez mais para garantir que não havia erros e que todas as informações estavam claras. Uma vez que essas avaliações foram feitas, a funcionalidade do teste é assegurada. Os dados dos dois testes pilotos foram descartados e novos participantes foram recrutados.

Para este estudo exploratório, decidimos analisar as respostas de 4 participantes em cada nível de proficiência (*sem conhecimento da L2; domínio básico da L2, intermediário e avançado*).

Cada participante recebeu um *link* por meio do qual iniciava o estudo na Plataforma PCIBex. Uma vez que as etapas (1 - 4) da sequência acima eram terminadas, o participante iniciava o teste (etapas 5 e 6).

A tarefa experimental consistia na classificação de palavras (estímulos) em três dimensões: valência, alerta e familiaridade. Cada participante julgou 80 palavras (40 em PB e as suas correspondentes traduções em EN, descritas em *Estímulos*) em dois blocos de língua (L1 e L2). Cada palavra era exibida no centro da tela e seguida por cada uma das escalas: (A) *valência*, (B) *alerta* e (C) *familiaridade*, conforme Figuras 4, 5 e 6, respectivamente<sup>9</sup>.

<sup>9</sup> As escalas continham 5 pontos graduados e adaptados da escala de 9 pontos do SAM (Bradley & Lang, 1994), uma vez que a escala de 7 pontos utilizada no estudo de Garrido & Prada (2018) não estava disponível.

Na primeira escala, Figura 4 abaixo, o participante classificava a valência da palavra lida, em cinco gradações: (1) muito negativa, (2) pouco negativa, (3) nem negativa nem positiva, (4) pouco positiva (5) muito positiva.

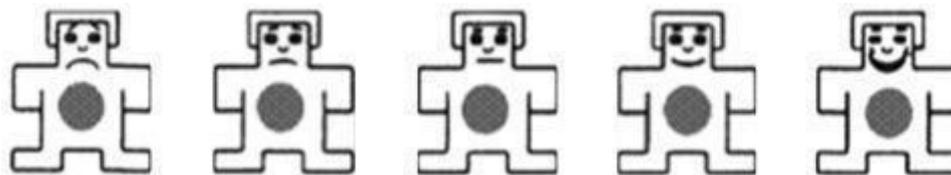


FIGURA 4: Escala visual de cinco pontos para classificação da dimensão Valência.

Na sequência, o participante classificava o grau de alerta da palavra lida, em cinco gradações, conforme Figura 5: (1) muito suave, (2) pouco suave, (3) nem suave nem intensa, (4) pouco intensa (5) muito intensa.

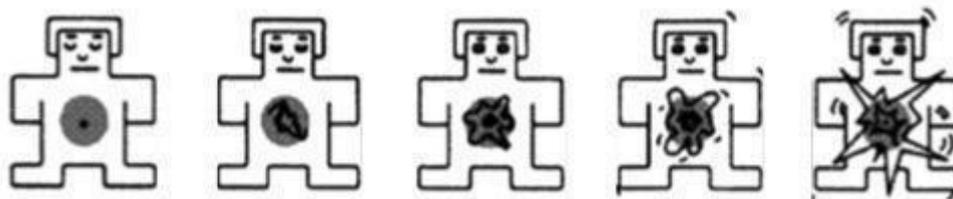


FIGURA 5: Escala visual de cinco pontos para classificação da dimensão Alerta.

Por último, o participante classificava a sua familiaridade com a palavra lida, em cinco gradações (Figura 6): (1) muito não familiar, (2) pouco não familiar, (3) nem não familiar nem familiar, (4) pouco familiar (5) muito familiar.



FIGURA 6: Escala visual de cinco pontos para classificação da dimensão Familiaridade.

Ao final das etapas experimentais em L1 e L2, os dados de resposta (classificações em três dimensões) eram salvos, o participante lia um agradecimento na tela e o estudo era encerrado.

O estudo completo levava em torno de 15 minutos e os dados gravados e coletados foram posteriormente analisados.

### 3. Resultados

Foram analisadas respostas de um total de 16 participantes distribuídos em quatro grupos de acordo com a proficiência declarada<sup>10</sup> em L2 (Inglês): nível de proficiência ((i) *sem conhecimento da L2*; (ii) *domínio básico da L2*, (iii) *intermediário* e (iv) *avançado*). Dados de quatro (4) participantes foram excluídos com base nos critérios de exclusão<sup>11</sup>.

Conforme amostra descrita abaixo, foram analisadas 880 respostas (classificações) em cada dimensão estudada (valência, intensidade, i.e. alerta, e familiaridade).

Descriptives		
	Tipo_Julgamento	Julgamento
N	Valência	880
	Intensidade	880
	Familiaridade	880
Missing	Valência	0
	Intensidade	0
	Familiaridade	0

TABELA 2: Descritivo do número total de pontos de dados analisados (N) e pontos de dados ausentes (*Missing*) em cada dimensão (valência, intensidade e familiaridade).

A média de respostas para as dimensões de valência, alerta e familiaridade foram analisadas e processadas para cada grupo de palavra do *corpus* criado: (i) Negativas, (ii) Positivas, (iii) Tabus e (iv) Neutras, conforme Tabela 3 abaixo.

<sup>10</sup> Estudos sobre a proficiência em segundas línguas mostram uma correlação entre a Proficiência Declarada e a Proficiência Testada (Flores, Zhou & Eiras, 2022). Portanto, neste estudo exploratório, utilizamos a Proficiência Declarada como medida de Proficiência em L2.

<sup>11</sup> Foram excluídos dados acima e abaixo de três desvios padrões (SDs); dados de respostas obtidos após 2000 milissegundos e/ ou dados de respostas até 200 milissegundos. Além disso, os participantes com idades maiores do que 35 anos foram excluídos da amostra, bem como aqueles abaixo dos 18 anos de idade.

Mean	Valência	Portugues	Negativa	1.32
			Positiva	4.54
			Tabu	2.51
			Neutra	3.06
		Ingles	Negativa	1.46
			Positiva	4.14
			Tabu	2.75
			Neutra	3.05
	Intensidade	Portugues	Negativa	4.68
			Positiva	3.46
			Tabu	4.16
			Neutra	2.88
		Ingles	Negativa	4.41
			Positiva	3.42
			Tabu	3.93
			Neutra	2.85
	Familiaridade	Portugues	Negativa	4.50
			Positiva	4.85
			Tabu	4.24
			Neutra	4.58
		Ingles	Negativa	3.98
			Positiva	4.20
			Tabu	3.21
			Neutra	3.46

TABELA 3: Descritivo da média de classificação (1-5 pontos) para cada dimensão analisada (valência, intensidade e familiaridade) em cada uma das línguas (L1 e L2) para cada grupo de palavra analisado (Negativas x Positivas x Tabus x Neutras).

Ainda conforme a Tabela 3, vemos que este estudo foi capaz de captar as nuances de classificação nas escalas de 5 pontos nas três dimensões investigadas. Ou seja, as palavras negativas foram, de fato, classificadas na escala de valência de 1 a 5 pontos como palavras negativas, obtendo na média uma classificação de 1,32 (sendo 1 muito negativa e 5 muito positiva). As palavras positivas obtiveram uma classificação de 4,54 na escala de positividade. Já as palavras tabus, 2,51, e as neutras, 3,06.

Na segunda língua, este teste também se mostrou eficaz em coletar as classificações lexicais nas dimensões estudadas – palavras negativas, 1,46; palavras positivas, 4,14; palavras tabus, 2,75; e as neutras, 3,05.

Na dimensão de alerta (sendo 1 muito suave e 5 muito intensa), as palavras negativas em PB obtiveram uma média de 4,68; as positivas, 3,46; as palavras tabus, 4,16; e as palavras neutras, 2,88. Em Inglês, a classificação das palavras negativas teve uma média de 4,41; das positivas, uma média de 3,42; das palavras tabus, 3,93; e das palavras neutras, 2,85.

Na dimensão de familiaridade (sendo 1 muito pouco familiar e 5 muito familiar), as palavras negativas em PB mostram uma média de 4,50; as positivas, 4,85; as palavras tabus, 4,24; e as palavras neutras, 4,58. Já na segunda língua (Inglês), a familiaridade das palavras negativas

obteve uma média de 3,98; das positivas, uma média de 4,20; das palavras tabus, 3,21; e das palavras neutras, 3,46.

Ainda sobre a capacidade do teste em refletir a percepção dos falantes sobre as três dimensões extra-linguísticas estudadas, podemos ver no gráfico abaixo que, na média, as classificações nas três escalas (Valência, Intensidade, i.e. alerta, e Familiaridade) parecem mostrar que a percepção dos falantes sobre essas dimensões reflete um conhecimento metalinguístico sofisticado. Os falantes são capazes de perceber e classificar as diferentes palavras (Negativas x Positivas x Tabus x Neutras) nas diferentes dimensões de maneira acurada – palavras negativas de maneira negativa, positivas de maneira positiva e as neutras como neutras. Já as palavras tabus são, na maior parte das vezes, classificadas como negativas.

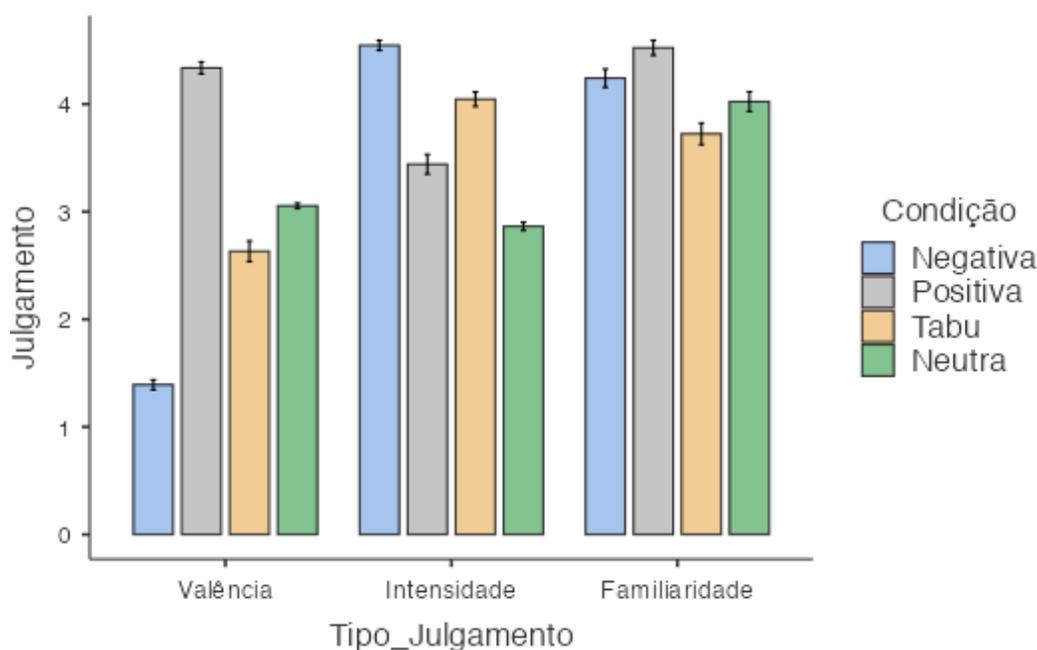


GRÁFICO 1: Média de respostas de classificação (Eixo Y: escala de 5 pontos) em cada uma das dimensões analisadas (Eixo X: valência, intensidade e familiaridade) para cada grupo de palavra analisado (Negativas x Positivas x Tabus x Neutras).

Entretanto, a proficiência parece ser um fator importante nesta distribuição: a maior proficiência na língua converge com classificações mais assertivas, como se pode ver na distribuição dos dados mostrada no gráfico abaixo.

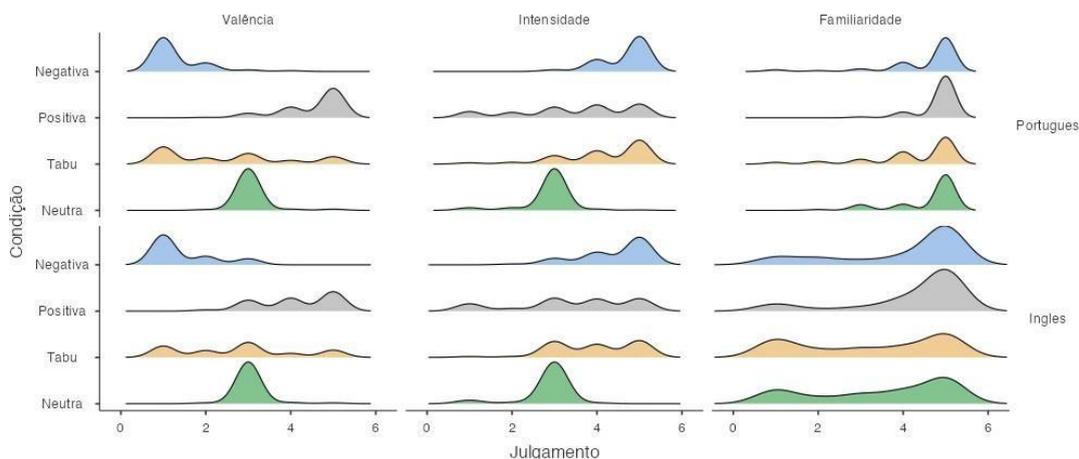


GRÁFICO 2: Distribuição dos dados de respostas de classificação (Eixo X: escala de 5 pontos) para cada grupo de palavra analisado (Eixo Y: Negativas x Positivas x Tabus x Neutras) em cada uma das dimensões analisadas (Eixo Z: valência, intensidade e familiaridade). O gráfico está dividido, no Eixo W, pela língua investigada (Português, quadrante superior; Inglês, quadrante inferior).

Ao compararmos a L1 e a L2 nas diferentes dimensões, podemos ver que a distribuição dos dados em L1 é menos esparsa, ou seja, em Português, as classificações feitas pelos participantes convergem e os dados estatísticos mostram picos bem definidos e poucos *valores discrepantes*<sup>12</sup> – palavras negativas obtém classificações entre 1-2; positivas, classificações entre 4-5; e as neutras, em torno de 3. Já as palavras tabus são, na maior parte das vezes, classificadas como negativas.

Na segunda língua, estas classificações não são tão convergentes, há, portanto, mais *outliers* e a distribuição não é tão convergente quanto em L1.

### 3.1 Análise

Foi utilizada a regressão linear para testar se as dimensões de valência, alerta e familiaridade previam significativamente diferenças na classificação lexical em L1 e L2. A regressão global foi estatisticamente significativa ( $R^2 = 0.209$ ,  $F(6, 2633) = 116$ ,  $p < .001$ ).

Verificou-se que a análise da Classificação ou Julgamento das palavras (Negativa x Positiva x Tabu x Neutra) previu significativamente as respostas nas três dimensões estudadas (valência, intensidade e familiaridade):  $\beta_{\text{Valencia}} = 2,7625$ ,  $p < .001$ ;  $\beta_{\text{Intensidade}} = 1,8705$ ,  $p < .001$  e  $\beta_{\text{Familiaridade}} = 1,2739$ ,  $p < .001$ .

<sup>12</sup> Em estatística, *outlier*, ou valor discrepante, é uma observação que apresenta um grande afastamento das demais da série. São, portanto, valores numericamente distantes da maior parte do conjunto de dados, ou seja, valores extremos. Normalmente, é considerado um *outlier* um ponto entre os 20% dos valores fora do intervalo de confiança (Soares & Siqueira, 2002). Aqui, utilizamos um intervalo de confiança de três desvios padrões para baixo e para cima.

## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>	Overall Model Test			
			F	df1	df2	p
1	0.458	0.209	116	6	2633	<.001

### Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
Tipo_Julgamento	746.0	2	373.00	242.2	<.001
Condição	256.6	3	85.54	55.6	<.001
Língua	70.7	1	70.69	45.9	<.001
Residuals	4054.4	2633	1.54		

Note. Type 3 sum of squares

[3]

### Model Coefficients - Julgamento

Predictor	Estimate	SE	t	p
Intercept <sup>a</sup>	2.7625	0.0639	43.23	<.001
Tipo_Julgamento:				
Intensidade – Valência	0.8705	0.0592	14.71	<.001
Familiaridade – Valência	1.2739	0.0592	21.53	<.001
Condição:				
Negativa – Neutra	0.0788	0.0683	1.15	0.249
Positiva – Neutra	0.7864	0.0683	11.51	<.001
Tabu – Neutra	0.1530	0.0683	2.24	0.025
Língua:				
Ingles – Portugues	-0.3273	0.0483	-6.78	<.001

<sup>a</sup> Represents reference level

TABELA 4: Resultados da Análise Estatística com Modelos de Regressão Linear.

Vamos agora analisar as variáveis investigadas:

Uma previsão direta para a discrepância entre a sensibilidade aos traços ligados à emotividade na L1 e na L2 determinada pela proficiência é que a discrepância seria mais forte em bilíngues com um nível mais baixo de proficiência na L2. Testamos essa hipótese examinando a forma como a diferença entre as classificações de valência em L1 e L2 é influenciada pelas estimativas de proficiência em Inglês dos participantes do estudo.

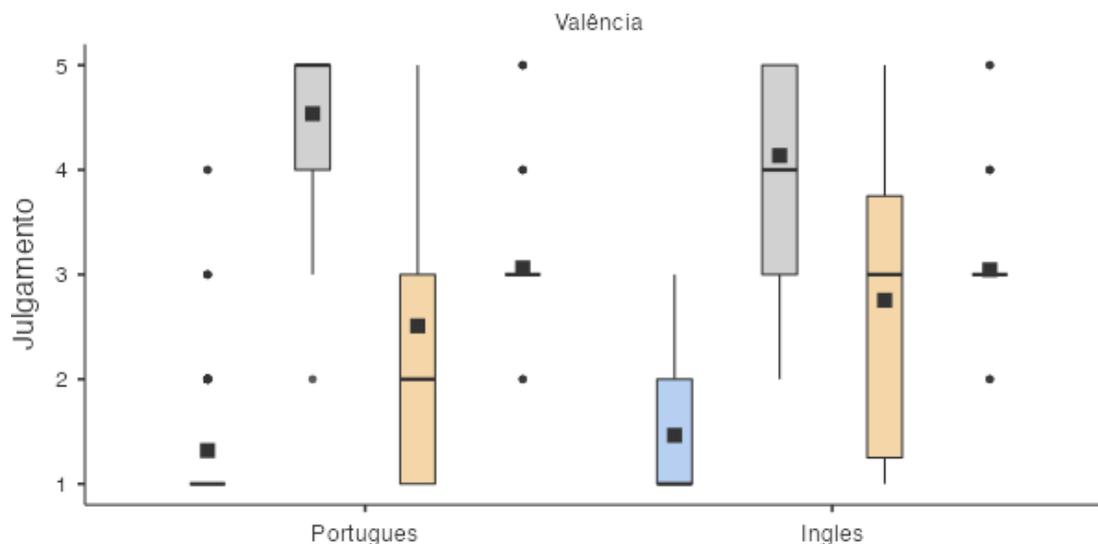


GRÁFICO 3: Média de respostas de classificação (Eixo Y: escala de 5 pontos) na dimensão analisada (Eixo X: valência) para cada Língua analisada (L1, PB x L2, Inglês). Cada coluna de dados corresponde a uma categoria de palavra (Negativa x Positiva x Tabu x Neutra).

Como podemos ver no gráfico, o julgamento da valência varia em relação à língua – os julgamentos na L1 (PB) contém menos valores discrepantes e tendem aos valores mais extremos da escala. Enquanto que os julgamentos na L2 (EN) são mais neutralizados.

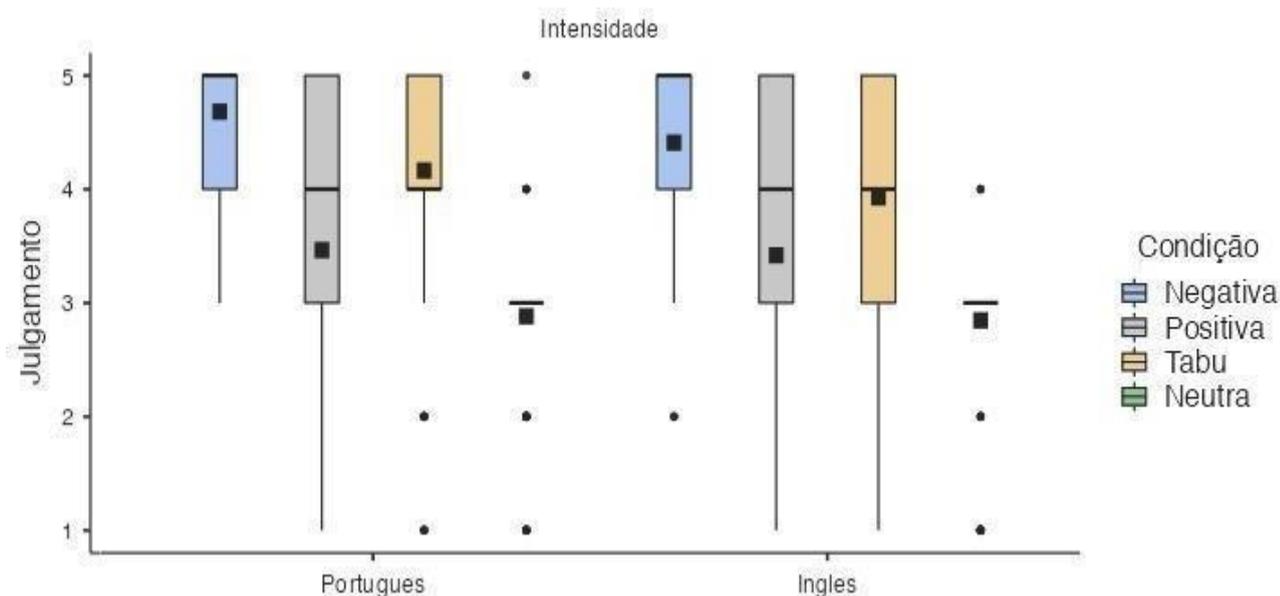


GRÁFICO 4: Média de respostas de classificação (Eixo Y: escala de 5 pontos) na dimensão analisada (Eixo X: intensidade) para cada Língua analisada (L1, PB x L2, Inglês). Cada coluna de dados corresponde a uma categoria de palavra (Negativa x Positiva x Tabu x Neutra).

Como podemos ver no gráfico acima, o julgamento da intensidade varia em relação à língua. Especialmente aqueles sobre a intensidade das palavras negativas e tabus variam entre a L1 (PB) e a L2 (EN). Sendo os primeiros mais intensos que os segundos.

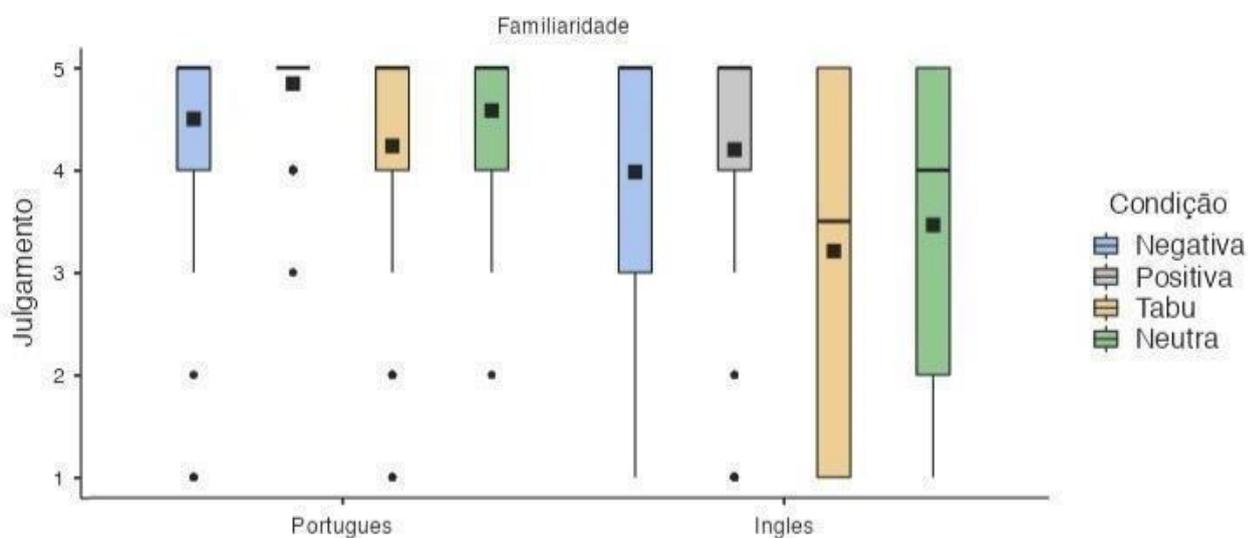


GRÁFICO 5: Média de respostas de classificação (Eixo Y: escala de 5 pontos) na dimensão analisada (Eixo X: familiaridade) para cada Língua analisada (L1, PB x L2, Inglês). Cada coluna de dados corresponde a uma categoria de palavra (Negativa x Positiva x Tabu x Neutra).

O gráfico acima mostra que o julgamento da familiaridade é modulado pelo tipo de palavra e varia em relação à língua (L1 x L2). Podemos observar que há variação no julgamento do tipo de palavra (Negativa x Positiva x Tabu x Neutra). Sendo as palavras de uma maneira geral menos familiares na L2. Havendo maior variação nas palavras tabus.

#### 4. Discussão

No presente trabalho obtivemos classificações de valência, alerta e familiaridade para um *corpus* de 80 palavras (40 em PB e as suas correspondentes traduções em Inglês). Na hipótese inicial, em consonância com o estudo de Garrido & Prada (2018), os valores de valência e alerta para palavras negativas, positivas e, principalmente, tabus seriam mais extremos na L1 do que na L2. Não eram esperadas alterações significativas nas classificações em palavras neutras. E, naturalmente, seriam maiores os valores de familiaridade para as palavras na L1 do que os valores para as mesmas palavras na L2.

Assim como no estudo de Garrido & Prada (2018), a pressuposição se mostrou assertiva. Entretanto, nosso estudo se mostra ainda mais sensível em perceber as diferenças na classificação das mesmas palavras em L1 e L2. Ou seja, neste estudo, observamos diferenças estatísticas entre a primeira e a segunda língua entre todos os tipos de palavras, inclusive nas

classificações de valência e alerta das palavras neutras. Também as palavras positivas, negativas e tabus apresentaram resultados mais expressivos em Português Brasileiro (L1) do que em Inglês (L2). Como observado nos resultados apresentados, enquanto a maioria dos participantes assinalou os pontos 1 e 2 na escala de valência para palavras negativas e tabus na L1, na L2, as classificações se concentraram nos pontos 1, 2 e 3 da escala. Sugerindo uma atenuação dos julgamentos em L2. Semelhantemente, as classificações obtidas para as palavras positivas na L1 ficaram entre 4 e 5, enquanto na L2 ficaram nos pontos 3, 4 e 5.

Isso mostra, primeiramente, que os falantes classificam de maneira precisa a valência das palavras: as negativas e tabus, com baixos valores de valência; as positivas, com altos valores; e as neutras, com valores em torno de 3. A segunda consideração é que, como esperado, os valores de valência para palavras negativas, positivas e tabus foram mais categóricos na L1 do que na L2. Isso corrobora com a ideia de neutralização na L2, visto que, de fato, foram atenuadas as classificações de valência em Inglês. Isto parece indicar que os falantes consideram uma mesma palavra como mais negativa na L1 e menos negativa na L2. O mesmo para uma palavra positiva; e que os falantes consideram uma palavra tabu menos ofensiva na L2 do que na L1.

Quanto ao alerta, nossos resultados foram diferentes do estudo de Garrido & Prada (2018). No estudo de 2018, as autoras mostram que embora os valores de alerta tenham sido diferentes na L1 e na L2 para palavras tabus, não houve discrepância na dimensão de alerta para palavras negativas e positivas. Em oposição, no presente estudo, tivemos variação na dimensão de alerta na L1 e na L2 tal como na valência, ainda que a disparidade entre os valores de alerta na L1 e na L2 tenha sido consideravelmente maior nas palavras tabus.

Em relação à familiaridade, as palavras na L1 foram classificadas como mais familiares que as palavras na L2, o que era o esperado. Como argumenta Garrido & Prada (2018), a neutralização de valência e alerta na L2 pode simplesmente derivar da falta de compreensão das palavras em Inglês que foram classificadas.

Contrasts - Proficiência				
	Estimate	SE	t	p
Intermediário - Avançado	0.2618	0.0482	5.433	<.001
Avançado - SemConhecimento	0.1833	0.0546	3.355	<.001
SemConhecimento - Básico	-0.0625	0.0631	-0.991	0.322

Contrasts - Língua				
	Estimate	SE	t	p
Portugues - Ingles	0.364	0.0397	9.18	<.001

TABELA 5: Resultados estatísticos com base nos Contrastes de Proficiência na L2.

Para aferir o grau de interferência da familiaridade na classificação de valência e alerta das palavras na L2, verificou-se os diferentes níveis de proficiência em Inglês dos participantes, como pode ser observado na Figura 7 acima. Participantes sem conhecimentos de Inglês e participantes do nível básico demonstraram maior neutralização de valência e alerta na L2 em comparação com a L1. Entretanto, uma vez que o nível de proficiência aumenta, os falantes de nível intermediário e avançado apresentam classificações de valência e alerta menos discrepantes entre a L1 e L2. Contudo, as classificações de valência e alerta na L2 realizadas pelos falantes de nível avançado ainda não se igualaram às classificações na L1. Isso demonstra que, mesmo que a familiaridade na L2 represente um papel crucial na sensibilidade do falante na dimensão emocional das palavras na segunda língua, mesmo falantes de nível avançado sentem as palavras com menos intensidade na segunda língua do que na primeira.

## 5. CONCLUSÕES

O objetivo principal deste estudo era perceber de que modo o processamento de palavras em primeiras (L1) e segundas línguas (L2) pode ser impactado pelas características lexicais de valência, alerta e familiaridade. Para tal, desenvolvemos um estudo exploratório de classificação de palavras em Português Brasileiro e Inglês com escalas de 5 pontos para cada uma das dimensões estudadas (valência, alerta e familiaridade). Foram testados 16 falantes de Português Brasileiro (L1) e Inglês (L2). Além da coleta e análise das respostas de classificação lexical nas dimensões de valência, alerta e familiaridade, pudemos investigar a interação entre a variável *proficiência em L2* e a forma como os leitores experienciam e classificam palavras em PB e EN.

Começamos por elaborar um *corpus* de 80 palavras (40 em PB e 40 em EN). Este corpus foi elaborado a partir do estudo de Garrido & Padra (2018), do ANEW em Inglês (Bradley & Lang, 1999) e em Português do Brasil (Kristensen et al., 2011) em 4 grupos de palavras (Negativas x Positivas x Tabus x Neutras), sendo os grupos negativo, positivo e neutro compostos por substantivos. Na sequência, essas 80 palavras foram submetidas à classificação com escalas de 5 pontos nas três dimensões estudadas num estudo *web-based* realizado através da Plataforma PCIbex.

Os nossos resultados estão de acordo com outros estudos e mostram que:

1. A sensibilidade às informações emocionais no nível das palavras pode ser medida em duas dimensões: valência e alerta;
2. Escalas de 5 pontos servem metodologicamente para aferir a sensibilidade às informações lexicais nas dimensões de valência, alerta e familiaridade;

3. A sensibilidade às informações lexicais nas dimensões de valência, alerta e familiaridade parece ser maior na L1 do que na L2;
4. A sensibilidade às informações lexicais nas dimensões *valência* e *alerta* interagem diretamente com o nível de proficiência na L2;
5. A proficiência na L2 é capaz de modular a sensibilidade emocional das palavras em três dimensões: valência, alerta e familiaridade;
6. Além disso, a dimensão de familiaridade lexical parece interagir com a proficiência na L2 de maneira a modular as dimensões de valência e alerta.

Podemos concluir que as características lexicais, como a valência, o alerta e a familiaridade, estão de alguma forma representadas na estrutura das palavras, uma vez que os participantes convergem na classificação destas características. Esta percepção parece ainda ser mais específica e assertiva na primeira língua (L1) do que na segunda (L2), uma vez que a proficiência na L2 parece influenciar a classificação, especialmente de palavras tabus e neutras. Ou seja, os falantes são mais perceptivos em relação às características lexicais das palavras em sua língua materna, enquanto que esta percepção parece ser mais superficial quando se trata de uma segunda língua. Além disso, a percepção em L2 parece ser ainda modulada de acordo com o grau de proficiência – falantes em níveis mais avançados mostram maior percepção sobre a valência, intensidade e familiaridade do que aqueles falantes sem nenhum conhecimento ou em níveis básicos de conhecimento. Foi também possível observar que a familiaridade linguística influencia a classificação da valência e intensidade dos itens lexicais.

Estes resultados sublinham a importância da ecologia comportamental no estudo e na aprendizagem de línguas, particularmente no que se refere ao processamento de características emocionais de palavras.

Esperamos, assim, ter contribuído para a literatura teórico-experimental do tema abordado, apresentando reflexões diversas cujo estudo pode ainda ser alargado, através, por exemplo, da criação de um *corpus* maior.

## Referências Bibliográficas

- Bestelmeyer, P. E., Kotz, S. A., & Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social cognitive and affective neuroscience*, 12(8), 1351-1358.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.
- Bradley, M.M., & Lang, P.J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- Garrido, M. V., & Prada, M. (2018). Comparing the valence, emotionality and subjective familiarity of words in a first and a second language. *International Journal of Bilingual Education and Bilingualism*, 24(2), 275-291. 10.1080/13670050.2018.1456514
- Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, 24(4), 561-579.
- Imbault, C., Titone, D., Warriner, A. B., & Kuperman, V. (2020). How are words felt in a second language: Norms for 2,628 English words for valence and arousal by L2 speakers. *Bilingualism: Language and Cognition*, 24, 281-292. <https://doi.org/10.1017/S1366728920000474>
- Kaczmarek, Ł. D., & Harmon-Jones, E. (2024). Motivational intensity and valence are married, but they are not the same person: Commentary on Campbell et al.(2021).
- Kristensen, C. H., de Azevedo Gomes, C. F., Justo, A. R., & Vieira, K. (2011). Normas brasileiras para o Affective Norms for English Words. *Trends Psychiatry Psychother*, 33(3), 135-146.
- Ku, L. C., Chan, S. H., & Lai, V. T. (2020). Personality traits and emotional word recognition: An ERP study. *Cognitive, Affective, & Behavioral Neuroscience*, 20, 371-386.
- Nascimento, H., Santos, C., Lima, F., & Silva, A. (2018). *A Study on the Use of the ANEW-Br Corpus for Identifying Words Related to Depression*. *Revista de Psicologia e Comportamento*, 14(3), 215-230.
- Oliveira, N. R., Janczura, G. A., & de Castilho, G. M. (2013, Abr-Jun). Normas de Alerta e Valência para 908 Palavras da Língua Portuguesa. *Psicologia: Teoria e Pesquisa*, 29(2), 185-200.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, 11(4), 357-383.
- Silva-Nasser, C. G. A. (2023). Uma análise psico e sociolinguística das propriedades dimensionais das palavras tabu no português carioca. *Rev. Estud. Ling.*, 31(2), 809-860.

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the Affective Norms for English words (ANEW) for European Portuguese. *Behaviorresearch methods*, 44, 256-269.

Soares, J. F., & Siqueira, A. L. (2002). *Introdução à estatística médica: Descrição e apresentação de dados* (1ª ed.). Belo Horizonte: COOPMED.

## ANEXOS

**1. Tabela com as Palavras do *Corpus* criado para o estudo sugerido:**

**Formato Excel:** [LINK 1](#)

**Formato PDF:** [LINK 2](#)

**2. Tabela com as Palavras do *Corpus* criado para o estudo exploratório:**

**Formato Excel:** [LINK 3](#)

**Formato PDF:** [LINK 4](#)

**3. Script criado para Teste de Classificação:** [LINK 5](#)

# SUPERDIVERSIDADE E CODESWITCHING ENTRE PORTUGUÊS E INGLÊS EM FALANTES DE PORTUGUÊS BRASILEIRO COMO LÍNGUA MATERNA

Marina Araújo Salimon

Orientado pela Professora Doutora Alexandra Pinto

Coorientado pela Doutora Vanessa Anachoretta e Ana Rita Fontes de Cunha

## Resumo

O artigo realiza uma breve síntese dos estudos sobre *codeswitching* e superdiversidade, seguida da análise de dados obtidos em um inquérito, a fim de descrever a prática do *codeswitching* entre português e inglês em falantes nativos de português brasileiro. Observando dados quantitativos e depoimentos dos indivíduos, conclui-se que pessoas jovens e mulheres realizam mais a prática, e que a proficiência em inglês não afeta necessariamente o engajamento, ou não, com a mesma. Também se conclui que na amostra de indivíduos analisada a prática, assim como o aprendizado do inglês em si, está ligada ao contexto cultural do advento da internet e das redes sociais.

Palavras Chave: *Codeswitching*, Sociolinguística, Análise do Discurso, Superdiversidade, Indexação.

## Abstract

The paper makes a short synthesis of the studies on codeswitching and superdiversity, followed by an analysis of the data collected through a survey to describe the practice of codeswitching between Portuguese and English in native speakers of Brazilian Portuguese. Observing the quantitative data along with testimonials from individuals, it is concluded that young people and women carry out the practice more, and that proficiency in English does not necessarily affect whether they engage in it or not. It is also concluded that in the collected sample the practice of codeswitching, as well as the acquisition of English itself, are linked to the cultural context of the emergence of the internet and social media.

Keywords: Codeswitching, Sociolinguistics, Discourse Analysis, Superdiversity, Indexicality.

## Introdução

Neste trabalho, pretende-se fazer uma abordagem sociolinguística, por meio da descrição do fenómeno linguístico da alternância de códigos em um episódio discursivo chamado *codeswitching* e sua relação com o contexto cultural da superdiversidade, que descreve as identidades híbridas que surgem nas últimas décadas, derivadas do uso intensivo da internet durante a infância e a adolescência. Realiza-se, por meio de um inquérito curto composto por uma combinação de perguntas abertas e fechadas, uma descrição etnográfica e um levantamento das funções linguísticas e expressivas do fenómeno em questão em falantes nativos do português brasileiro, assim como variações de interpretação em relação às funções e contextos de uso de acordo com idade, gênero e sexualidade ou grupo social a que pertencem os indivíduos que o praticam. A partir dos dados obtidos se conclui, como semente para futuros estudos mais aprofundados, que a prática é mais realizada por mulheres, pessoas mais jovens e pessoas que se consideram parte da comunidade *Queer*.

### 1. Enquadramento teórico

Considera-se *codeswitching* a alternância entre línguas, neste caso nomeadamente português do Brasil e inglês, em um episódio discursivo. Em tradução livre, Nílep (2006:17) descreve *codeswitching* como a prática discursiva de participantes para sinalizar mudanças de contexto usando sistemas ou subsistemas gramaticais alternantes, ou seja, códigos. De acordo com o autor:

*“The ‘context’ so signaled may be very local (such as the end of a turn at talk), very general (such as positioning vis-à-vis some macro-sociological category), or anywhere in between. Furthermore, it is important to recognize that this signaling is accomplished by the action of participants in a particular interaction”* (Nílep 2006:17)

Usaremos o termo *codeswitching* (daqui em diante referido como CS) como um termo “guarda-chuva”, que engloba diferentes termos e definições mais à frente discutidos, em oposição a uma noção generalizada de “multilinguismo”, que implica uma fronteira clara entre a alternância da utilização de uma língua e da outra, enquanto na prática do CS encontraremos uma justaposição mais fluida e recorrente das línguas ao longo do discurso.

Interessa-nos aqui descrever um contexto etnográfico em que essa prática ocorre, partindo da hipótese de que indivíduos mais novos, ao crescerem num contexto de alta interculturalidade e contato com a L2 numa idade muito permeável (Gumperz, 1982:27) realizam mais o CS. Prevê-se, também, de acordo com os estudos realizados por Dewaele e Wei (2013), que ainda permaneçam juízos de valor negativos sobre o uso híbrido de variedades linguísticas, assim como uma maior adesão à prática por parte de mulheres.

Andrautsopoulos (2007:208) refere a propósito desta temática: “internet users engage in linguistic bricolage on their homepages” representando concisamente o uso estilístico e lúdico da linguagem. Neste caso os usuários se alimentam de uma série de repertórios linguísticos, dentro e fora da sua zona de conhecimento, em um espaço comum que é a internet, que surge como um berço cultural altamente diverso.

### 1.1 Contextualização histórica

Os estudos sobre o *codeswitching* têm se alargado exponencialmente e tem se conhecido uma série de novas perspectivas, aqui resumidas para nos oferecer uma base de entendimento para a investigação realizada.

Interessa-nos observar o modelo proposto por Myers-Scotton (1983), denominado *Markedness Model*, onde cada língua em uma determinada comunidade é associada a papéis sociais particulares chamados *rights and obligations sets*. Neste contexto, a escolha do código passa a assinalar estes papéis, ou seja, marcá-los. No entanto, esta teoria implica presumir o estado interior dos falantes, determinando também fronteiras muito rígidas sobre a identidade e os papéis sociais de cada indivíduo em um contexto.

Também são relevantes as descrições de Gumperz (1982) a respeito de um “*we-code*” e um “*they-code*”, que, similarmente à teoria de Myers-Scotton, associam um significado social a cada variedade em um contexto de troca, diferenciando entre um código-eles, por exemplo associado a forças normativas e culturas dominantes de centros políticos e culturais, e um código-nós, respectivamente associado a minorias. O autor também introduz o conceito de *contextualization cues*, por meio das quais o CS exerce funções que em contextos monolíngues poderiam ser expressas pela prosódia, guiando o interlocutor à correta interpretação de seu enunciado por meio de informação contida além do conteúdo referencial.

Como surgem antes da expansão das novas mídias, as teorias propostas por Gumperz (1982) e Myers-Scotton (1983), além de associarem significados fixos a cada código, apresentam como foco apenas grupos de imigrantes ou variedades associadas a regiões fronteiriças, não contemplando nas análises as identidades plurais do mundo globalizado. É interessante, no entanto, pensar o CS com inglês como escolha marcada e código-nós, que indica um contexto cultural e uma identificação com a cultura contemporânea, por exemplo, face às gerações mais velhas e forças da norma-padrão.

Giles e Powesland (1975) descrevem a chamada *speech accommodation theory*, defendendo a ideia de que os falantes acomodam sua produção linguística e adaptam-na levando em consideração o interlocutor, com o objetivo de fazer divergir ou convergir seu estilo linguístico com o do seu par.

Rampton (1995) introduz o conceito de *crossing*, um tipo de CS praticado por falantes

adolescentes numa comunidade de imigrantes de diferentes países que atravessa os limites de etnia e comunidade linguística. O autor analisa as dinâmicas destes contatos e suas complexas implicações sociais, dando início a novas perspectivas sobre intersecções culturais e *codeswitching*.

Ao longo de todos os estudos, salienta-se que se mantém no centro da relevância o princípio cooperativo de Grice (1975), um dos pilares universais da pragmática. Este prevê que, para uma boa comunicação, o falante deve procurar se expressar da forma mais explícita e clara possível perante seu interlocutor, cooperando para que seus enunciados sejam interpretados de acordo com a sua intenção comunicativa. No contexto aqui estudado, este princípio toma relevância na escolha de realizar ou não CS em cada interação, tendo em conta a capacidade de compreensão das línguas em questão por parte do seu par.

Uma série de autores dão continuidade aos estudos sobre CS observando o avanço da tecnologia e das *new media*. Jaquemet (2005) descreve o que chama de *transidiomatic practices*, ou seja, interações que se servem de múltiplas línguas por exemplo em ambientes profissionais ou no dia a dia como manifestação do que chama de *recombinant identities*, *identidades* que surgem a partir de referências culturais comuns da cultura global como filmes ou músicas. De maneira semelhante, englobando este contexto Vertovec (2007) introduz o termo *superdiversity*, para descrever o estado cultural de intensa migração e interação, assim como intensificação da subjetividade de cada indivíduo.

Estabelece-se mais recentemente a teoria da “indexação”, que tem como objeto o significado conotativo dos signos, neste caso da produção linguística (Blomhaert, Rampton 2011:5). Interessa a implicação de significado social que cada variedade ou estilo linguístico carrega em cada dado momento de interação. De acordo com Hall (2014:262), esta corrente considera que: “Identity is constituted through the creation of semiotic links between linguistic form and social meaning a process known as **indexicality**”. É defendido também que a análise do discurso permite concluir que o processo de interação é ao mesmo tempo produtor da identidade, como também consequência do contexto particular de cada sujeito e da sua história social.

Partindo desta lente, o inglês “global” é uma língua franca que se posiciona não só como língua de poder e viabilizadora de ascensão cultural, mas também como um berço cultural para novas gerações que cresceram em contato com a língua inglesa por meio da televisão, músicas e internet. Como esclarece Hall (2014:264): “it is also an attitude, a style, a way of being in the world, a projection of progressive ideas about gender and sexuality, a declaration of cosmopolitan subjectivity”.

No berço da internet e das redes sociais, em que o contato linguístico e a ludicidade são intensos, o CS se desenvolve rapidamente e a indexação toma uma forma muito produtiva e de rápida evolução. A questão deixa de ser a competência em determinada língua e passa a ser a

capacidade de manter-se a par de sua variação, dentro do contexto socio-virtual e do panorama da cultura de massas: “In the flux of indexical relations that characterize contact zones of late capitalism, semiotic stability is at best a fleeting achievement. But this instability also explains why speakers become so invested in defending it” (Hall 2014:264)

Em 2013 Andrautsopoulos introduz o termo *networked multilingualism*, após duas décadas investigando o discurso virtual, a fim de descrever essa prática particular com características próprias que surge nos ambientes virtuais, o autor assim diferencia o ponto de vista face às teorias mais antigas sobre o CS que se baseiam principalmente nos movimentos de migração e zonas geográficas de contato entre línguas. Durante as últimas décadas, cada vez mais a internet deixou de ser um espaço fixo (um computador pessoal, por exemplo, ou um estabelecimento como *cybercafé*) e passou a estar constantemente acessível. Portanto, os costumes das relações virtuais e não virtuais influenciam-se mutuamente, formando uma rede também de costumes linguísticos, incluindo o CS.

Por final, é relevante também mencionar o processo de Aquisição da Linguagem: consideramos como hipótese que quanto mais cedo ocorrer a aquisição do inglês, mais propenso o indivíduo estaria a realizar CS, tendo em conta que terá mais confiança e conforto na sua produção.

## **1.2 Funções**

Tendo em conta a perspectiva descrita, destacam-se abaixo algumas funções da prática do CS observadas a partir de dados obtidos por meio da aplicação do inquérito do estudo. Destaca-se, no entanto, que não é possível estabilizar uma lista objetiva e exaustiva, tendo em conta que é um fenómeno de escolha interno ao falante, e o contexto de cada interação emerge da mesma, podendo o CS cumprir uma infinidade de funções.

Em primeiro lugar, interessa-nos analisar a indexação de significado na prática do CS, como definida por Rampton & Blommaert (2011). O inglês, neste caso, exerce uma função de identificação e expressão de individualidade, sendo simultaneamente uma porta de abertura a novas ideias, como também um elo de conexão entre indivíduos do mundo todo. A partir destas conexões emergem identidades, que são expressas também por meio da produção linguística e inerentes à mesma. Niley (2006:13) pontua que “crossing in various directions also serves to forge a common adolescent group, to dissociate from parents or elders, and to resist endemic stereotypes”.

Em seguida, também de certo modo ligado à formação de identidade e socialização, é relevante observar como a língua pode exercer a função de objeto de diversão, sendo utilizada para gerar humor por trocadilhos fonéticos ou semânticos: um alargamento do repertório linguístico certamente dá origem a combinações de códigos para o efeito lúdico. Nas palavras de Rampton & Blommaert (2011):

*“Instead of focusing on communicative inequalities in institutional and instrumental settings, there is emphasis on creativity and **linguistic profusion** when sociolinguistic research focuses on non-standard mixed language practices that appear to draw on styles and languages that aren’t normally regarded as belonging to the speaker, especially in recreational, artistic and/or oppositional contexts (and often among youth)”*

Outra função da prática de CS é também o posicionamento social. Hall (2014) descreve um estado chamado *Hypersubjectivity*, ilustrando como as exigências do capitalismo tardio e das identidades híbridas também originam um estado ansioso e altamente fatigante, em que as competências social e linguística são profundamente relacionadas e se tornam exigências para a competência profissional. Segundo a autora: “Language is likewise reinterpreted in this process as a technical skill that can enhance one’s individual position in the new global economy” (Hall 2014:263)

Por último, observaremos o que talvez seja considerada a função mais geral e comum do CS, que é complementar significado. Seja por meio de *contextualization cues* ou de vocabulário e estruturas gramaticais ausentes em uma das línguas, o indivíduo serve-se do seu repertório linguístico consoante o contexto para melhor expressar uma ideia específica e ilustrar aquilo que pretende.

## 2. Metodologia

A metodologia escolhida para a realização deste trabalho tem como base as ideias de Rampton & Blommaert (2011:11). Os autores defendem que, tendo em conta que o CS se trata de um fenômeno interno e subjetivo do falante, uma boa investigação a respeito da temática deve contemplar dois eixos de análise: o primeiro seria uma descrição exaustiva do discurso, a fim de levantar padrões linguísticos, e o segundo consistiria numa descrição detalhada do trajeto etnográfico, linguístico e cultural em associação com testemunhos e intuições dos falantes. Ambos em conjunto permitiriam um levantamento bem-sucedido das funções e significados desta prática. Devido às limitações associadas à natureza desta pesquisa, optou-se por focar no segundo eixo de descrição, resguardando o primeiro para estudos futuros.

A recolha de dados consistiu na realização de um inquérito, com perguntas abertas e fechadas. O inquérito foi elaborado em semelhança ao realizado por Cunha (2021), que desenvolveu uma pesquisa precursora sobre o mesmo tema em relação ao português Europeu. Não sendo possível realizar entrevistas prolongadas, juntou-se ao inquérito também perguntas abertas.

Nas perguntas fechadas, pretendeu-se adquirir os dados sociométricos para um cruzamento de informações que nos permitisse descrever os contextos sociais e etários do fenómeno em questão,

assim como alguns dados sobre a aquisição do inglês como L2 de cada um dos falantes. Por outro lado, foram incluídas perguntas abertas que possibilitassem o contato com diversas perspectivas a respeito das experiências dos falantes e suas intuições individuais. O inquérito esteve aberto entre dia 16/04/2024 e 10/05/2024, obtendo um total de 90 respostas, das quais 84 foram selecionadas para análise por cumprirem os requisitos quanto ao tratamento de dados, maioria e nacionalidade.

Segundo Rampton & Blommaert (2011:13) a superdiversidade trata de alta mutabilidade e mobilidade, e para investigá-la é necessário sempre que possível comparar diferentes gerações de estudos assim como diferentes gerações de falantes. A partir desta ideia, não foi delimitado um limite etário ou de gênero para a pesquisa, sendo comparados os hábitos e perspectivas de diferentes gerações, gêneros e sexualidade.

## **2.1 Discussão e resultados**

Primeiramente, é importante pontuar que o inquérito não foi disseminado publicamente em grande escala, portanto o recorte social obtido na amostra é relativamente limitado em termos de região e classe social. Nota-se, por exemplo, que a grande maioria dos participantes frequentou total ou parcialmente seus estudos na rede privada, o que não retrata a realidade<sup>1</sup> da maioria da população brasileira. Também devemos pontuar que a amostra é pequena, e, portanto, certas observações devem servir apenas como ponto de partida para futuras investigações. Em seguida, é importante pontuar que o nível de fluência obtido assim como a idade de aquisição da L2 foi autodeterminado pelos participantes, tendo em conta as limitações da pesquisa, e, portanto, a resposta dada pelos participantes tem implicações nos resultados. Por último, também assinalamos que a percepção dos participantes, principalmente com níveis baixos de fluência, sobre o que é ou não *codeswitching* pode variar, e futuras pesquisas mais aprofundadas sobre a diferenciação ente o uso de *codeswitching* e empréstimos, por exemplo, juntamente de instrumentos de análise mais robustos podem permitir resultados mais fiáveis.

### **2.1.1 Respostas Fechadas**

Começando por observar os dados obtidos de modo geral, nota-se que a maioria dos participantes afirma praticar o *codeswitching*, correspondendo a 66.6% dos falantes.

Em seguida, cruzamos alguns dados que consideramos mais interessantes para analisar as hipóteses antes mencionadas. No quadro 1 podemos observar a relação entre o gênero com o qual o falante se identifica e a resposta à pergunta “Você mistura português e inglês quando fala?”. A amostra obtida foi bastante equilibrada, com 36 participantes que se identificam como mulheres e 47 como homens, assim como apenas 1 participante que não compartilhou essa informação. Não

---

<sup>1</sup> MEC e Inep divulgam resultados do Censo Escolar 2023. (n.d.). *Gov.br Ministério Da Educação*. Retrieved February 2, 2024

obtivemos nenhum participante que se identificasse como pessoa não binária, mas aponta-se que seria um estudo interessante coletar percepções linguísticas, futuramente, focando nestes indivíduos, tendo em conta que o inglês oferece mais opções expressivas neutras.

Os dados obtidos dialogam com os resultados apontados pelos estudos de Dewaele e Wei (2013), e Cunha (2021), sendo as mulheres o grupo com maior adesão à prática do CS, com 80.55% das participantes afirmando que o fazem. Apesar de a maioria dos homens também optar por ‘sim’, os números já são consideravelmente mais equilibrados neste caso, com apenas 55.32% escolhendo esta opção.

1 - Você mistura português e inglês quando fala?

Gênero		Frequência	Porcentagem
Feminino (36)	Sim	29	80.55%
	Não	7	19.44%
Masculino (47)	Sim	26	55.32%
	Não	21	44.68%
Não determinado (1)	Sim	1	100%
	Não	0	0%
<b>Geral (84)</b>	Sim	56	66.6%
	Não	28	33.3%

Em relação à idade dos participantes também foram obtidos dados, apresentados no quadro 2, que corroboram com a teoria de Gumperz (1982) que propõe que são as pessoas mais jovens que mais utilizam o CS.

No primeiro grupo, constituído por jovens de 18 a 25 anos, 83.8% praticam CS. A maioria dos adultos entre os 25 e os 40 anos – equivalente a 67.8% - praticam o CS, apesar de uma percentagem ligeiramente menor em relação aos mais novos.

O grupo acima dos 40 anos já representa uma minoria de sujeitos aderentes à prática, com apenas 44% dos falantes indicando a opção ‘sim’.

Presume-se que estes resultados se relacionam com a diferença de cultura geracional: tendo em conta que as pessoas mais novas tiveram maior contato com mídias estrangeiras e as práticas transidiomáticas da internet em idades formativas, elas podem se sentir mais confortáveis e

confiantes em praticá-las. Seria interessante, em futuros trabalhos, observar mais a fundo essas diferenças geracionais e possíveis juízos de valor em relação à idade dos falantes, que podem motivar, por exemplo, falantes fluentes em inglês de não praticar o CS. Também se observa que a prática do CS não está necessariamente condicionada pela competência numa dada língua, mas também relacionada ao estilo e às dinâmicas sociais, como esclarecem Rampton & Blommaert (2011:5): “Contrast and counter-valorization play an integral part in linguistic socialization, and people develop strong feelings about styles and registers that they can hardly reproduce (if at all)”.

2 – Você mistura inglês e português quando fala?

Idade		Frequência	Porcentagem
Entre 18 e 25 (31)	Sim	26	83.8%
	Não	5	16.1%
Entre 25 e 30 (18)	Sim	12	66.6%
	Não	6	33.3%
Entre 30 e 40 (10)	Sim	7	70%
	Não	3	30%
Acima de 40 (25)	Sim	11	44%
	Não	14	56%
<b>Geral (84)</b>	Sim	56	66.6%
	Não	28	33.3%

No quadro 3, avançamos para um cruzamento da mesma pergunta “Você mistura inglês e português quando fala?” com os dados de idade de aquisição. Nota-se que a pergunta “Com que idade você considera ter se tornado fluente em inglês?” leva em conta a avaliação subjetiva do participante a respeito da sua proficiência na língua.

Apesar de ser um número absoluto pequeno, nota-se que todos os sujeitos que adquiriram o inglês até os 11 anos realizam CS. A grande maioria das pessoas aprendeu inglês entre os 11 e 18 anos, uma idade de formação em que a produção linguística é muito plástica, pois os indivíduos estão em formação escolar e atravessando os processos formação da personalidade e socialização,

estando mais propensos a adquirir novos hábitos linguísticos. Os indivíduos que se tornaram fluentes em inglês após os 18 anos e praticam CS ainda são a maioria, apesar de ser em menor proporção em relação ao grupo anterior.

Outro ponto interessante a observar, apesar de também ser um dado marginal e não permitir conclusões concretas, é que dos oito indivíduos que não se consideram fluentes em inglês, 3 afirmaram praticar CS, apontando mais uma vez para a observação de que a fluência pode não ser um fator decisivo na prática ou não de *codeswitching*. No entanto, pontua-se mais uma vez que cada falante afirmou seu próprio nível e fluência, uma vez que o escopo da pesquisa não permitiu a aplicação de ferramentas mais precisas de análise neste aspecto.

### 3- Você mistura inglês e português quando fala?

Idade com a qual o participante considera ter se tornado fluente		Frequência	Porcentagem
Até os 7 anos (1)	Sim	1	100%
	Não	0	0%
Entre 7 e 11 (2)	Sim	2	100%
	Não	0	0%
Entre 11 e 18 (48)	Sim	35	72%
	Não	13	27%
Após os 18 (25)	Sim	15	60%
	Não	10	40%
Não se consideram fluentes (8)	Sim	3	37.5%
	Não	5	62.5%

Partindo desta observação, nota-se no quadro 4 a relação entre o nível de fluência - também aqui autodeterminado - e a prática ou não de CS. As maiores porcentagens de indivíduos que praticam se encontram nos níveis proficientes C1 e C2, e independentes B2. No entanto, ainda há uma quantidade significativa de praticantes nos níveis B1 e A2, apesar de compreenderem pouco a língua.

4- Você mistura inglês e português quando fala?

Nível de Fluência		Frequência	Porcentagem
A1-iniciante (1)	Sim	0	0%
	Não	1	100%
A2 - Básico (9)	Sim	3	33.3%
	Não	6	66.6%
B1-Intermediário (7)	Sim	2	28.5%
	Não	5	71.4%
B2-Independente (13)	Sim	7	53.8%
	Não	6	46.15%
C1-Proficiência eficaz (36)	Sim	29	80.55%
	Não	7	19.44%
C2-Domínio pleno (18)	Sim	15	83.3%
	Não	3	16.6%
<b>Geral (84)</b>	Sim	56	66.6%
	Não	28	33.3%

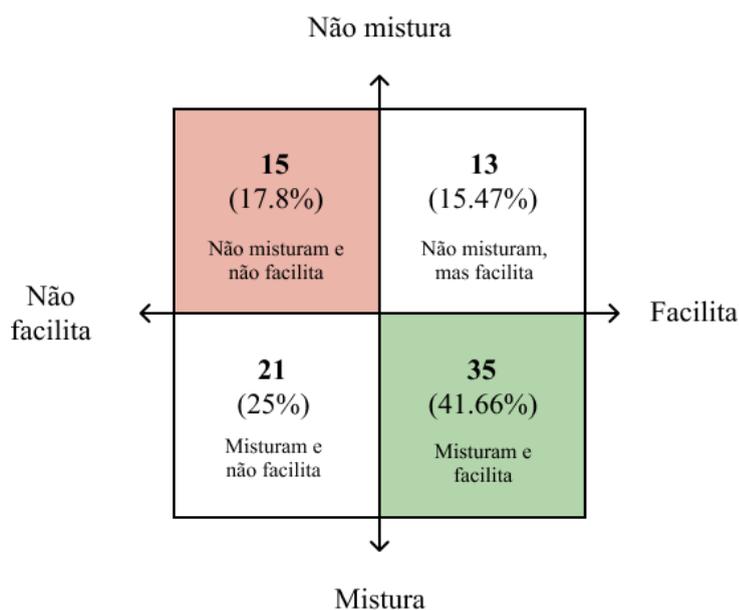
No quadro 5 observamos a relação entre orientação sexual e a prática do CS. A grande maioria das pessoas que se identifica como Queer, correspondendo a 86.2%, engaja em CS, enquanto, entre os indivíduos heterossexuais, apenas ligeiramente acima da maioria, 54.70%, o fazem. Tais dados dialogam com as ideias de Cunha (2021), que explica em seu estudo que os indivíduos Queer realizam mais CS com o português europeu e o inglês do que seus pares que se identificam como heterossexuais. No entanto, observa-se no contexto desta investigação, cujos participantes utilizam a variedade brasileira da língua, maior disparidade entre os dois grupos, o que é um ponto muito interessante para futuros estudos comparativos entre o Brasil e Portugal.

5 - Você mistura português e inglês quando fala?

Orientação sexual		Frequência	Porcentagem
Heterossexual (53)	Sim	29	54.70%
	Não	24	45.28%
Queer/LGBTQIAP + (29)	Sim	25	86.2%
	Não	4	13.7%
Prefiro não dizer (2)	Sim	2	100%
	Não	0	0%
<b>Geral (84)</b>	Sim	56	66.6%
	Não	28	33.3%

Na pergunta “Você considera que trocar de línguas durante o discurso facilita a comunicação no geral?” a maioria dos participantes (48 indivíduos - 57.14%) considerou que sim. Deste quantitativo, 35 indivíduos (41.66%) marcaram na questão anterior que praticavam CS e 13 indivíduos, apesar de não praticarem CS, concordaram que pode facilitar a comunicação. Aqui, partimos para um gráfico que melhor ilustra os diferentes posicionamentos:

6- Você mistura português e inglês quando fala?/ Você considera que trocar de línguas durante o discurso facilita a comunicação no geral?



Na pergunta “Como você desenvolveu a fluência em inglês?” foram oferecidas algumas opções em *checkbox* (escola, curso auxiliar, séries e filmes, internet e videogames), dentre as quais os indivíduos poderiam escolher mais de uma, e uma opção aberta onde poderiam complementar com outras respostas. Tendo em conta que a maior parte das respostas foi uma combinação dos dois tipos (aberta e múltipla-escolha), a maioria dos participantes afirma ter aprendido por meio de mídias. Nas demais opções, a maioria indicou cursos auxiliares, e novamente destacamos aqui as limitações do recorte social obtido. Residualmente foi apontado estudo por conta própria, literatura, etc. Abaixo inclui-se um quadro para ilustrar em a frequência de escolha das principais opções:

7 – Como vocês desenvolveu a fluência em inglês?

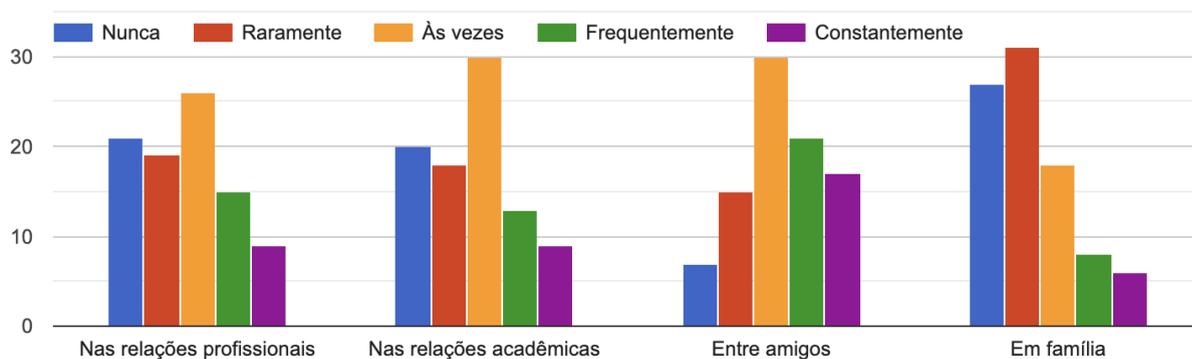
Meio de aprendizado		Frequência	Porcentagem da amostra total (84)
Mídia	Internet	46	54,76%
	Filmes e séries	44	52,38%
	Videogames	28	33,3%
	Música	3	3,5%
Outros	Curso auxiliar	53	63%
	Escola	36	42,8%
	Experiência no exterior	12	14,2%
	Trabalho	2	2,3%

Em seguida, foram apresentados diferentes contextos sociais e os participantes deveriam selecionar - dentre as opções “Nunca”, “Raramente”, “Às vezes”, “Frequentemente” e “Constantemente - com qual frequência realizavam CS em cada um dos cenários.

Os dados apresentados no gráfico abaixo mostram que os diferentes contextos sociais podem influenciar na escolha do indivíduo em praticar ou não CS: observa-se que o CS é mais praticado em contextos informais do que em contextos mais formais em que a adequação à norma padrão exerce um peso maior sobre a produção linguística. Isso aponta, mais uma vez, para a hipótese de que a prática seja um indício de maior proximidade e/ou identificação entre os falantes, e de que há o discernimento do falante em relação ao contexto social em que se insere a interação. A maioria das escolhas “Às vezes” na opção “entre amigos”, em oposição ao “frequentemente” ou “constantemente”, também nos aponta para o fato de que cada círculo social ou grupo de amigos

é gerido por suas próprias regras e práticas, e, dependendo da negociação estabelecida e da fluência dos participantes, um mesmo indivíduo pode praticar CS em um grupo social e não em outro.

8 - Nos seguintes contextos, mistura inglês e português?



Quanto aos motivos pelos quais os indivíduos realizam CS, numa resposta fechada foram apresentadas as afirmações acima, a partir das quais obtivemos os seguintes dados:

- A maioria dos participantes, (57 indivíduos - 67.8%) concordou parcial ou fortemente com a opção “Troco de línguas consoante o assunto em meu círculo social”

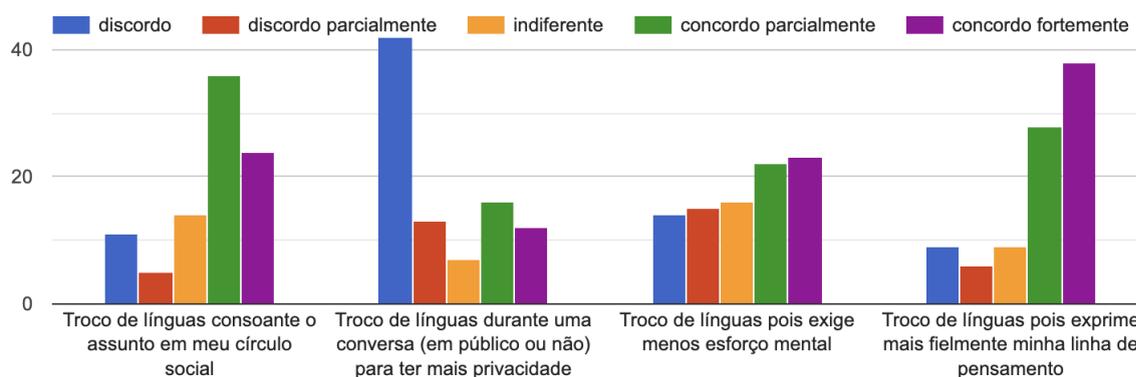
- 52 participantes (60.7%) discordaram ou discordaram parcialmente “Troco de línguas durante uma conversa (em público ou não) para ter mais privacidade”, o que demonstra que o inglês, talvez pela sua maior difusão nos últimos anos, parece ter deixado de atuar como código de privacidade;

- as respostas foram relativamente homogêneas em relação à opção “Troco de línguas pois exige menos esforço mental”, o que pode relacionar-se com a prática do CS por pessoas que tem um nível baixo de fluência e a adaptabilidade perante diferentes dinâmicas sociais e o compromisso com a cooperação;

- 61 indivíduos (72.61%) concordaram parcial ou fortemente com a afirmação “Troco de línguas pois exprime mais fielmente minha linha de pensamento”

A partir dessas respostas mais uma vez podemos corroborar a função expressiva e de contextualização e clarificação semântica do CS.

9 - Resposta de acordo com o que é mais verdadeiro para você



### 2.1.2 Respostas Abertas

Aqui, elencamos algumas das funções propostas anteriormente e pontos interessantes sobre a prática do CS, a partir da análise das respostas abertas recolhidas no inquérito relativamente às perguntas “Há algum outro contexto em que você mistura as línguas?” e “Há algum outro motivo pelo qual você troca de línguas?”. Pontuamos o uso humorístico da língua, o preenchimento de lacunas semânticas e *contextualization cues*, o posicionamento social e a indexação, assim como uma nova função não prevista que chamaremos de “*self-talk*”.

Começamos por analisar<sup>2</sup> o uso humorístico e lúdico da língua, nas respostas a seguir:

1.a “*Existem algumas piadas, expressões e histórias que só funcionam, digamos assim, no inglês.*”

1.b “*Por motivos jocosos em alguns contextos*”

Nas respostas 1.a e 1.b, podemos ressaltar a ideia de que um maior repertório linguístico é também um maior repertório lúdico, ou seja, o sujeito pode praticar o CS para fazer referência a um *meme* específico, ou até um trocadilho com palavras que só funciona em inglês ou com a combinação do português e do inglês.

1.c “*Humor, como em ‘humor’, não como em ‘mood’*”

Nota-se que na resposta 1.c o sujeito faz uso do próprio CS para clarificar o significado que pretende expressar.

A seguir, observamos a prática do CS para preenchimento de lacunas semânticas e como *contextualization cues*:

<sup>2</sup> As respostas aqui apresentadas, assim como subsequentemente neste trabalho, foram reproduzidas tal como surgiram originalmente

2.a *“Melhor expressão de uma ideia”*

2.b *“Para me expressar as vezes com palavras e expressões que não existem em português ou em inglês.”*

2.c *“Existem algumas piadas, expressões e histórias que só funcionam, digamos assim, no inglês.”*

2.d *“Acho que a facilidade é o menor esforço mental”*

2.e *“Interesse em me fazer entendido da forma mais clara possível.”*

2.f *“Quando há um termo específico onde não me recordo de uma palavra/contexto e é necessário haver essa mistura de línguas.”*

2.g *“Quando uma palavra ou expressão faz mais sentido em uma língua específica.”*

Observamos que por um lado, o preenchimento semântico pode se referir a palavras que não existem em uma das línguas, como refere 2.b, mas também pode se relacionar com o peso semântico ou outras minúcias do significado, como refere 2.g ao dizer que uma expressão “faz mais sentido”. Esta função especificamente será um terreno fértil para estudos mais aprofundados de semântica.

Partindo para uma análise do CS enquanto ferramenta de posicionamento social e profissional, temos em conta as seguintes respostas:

3.a *“No mundo acadêmico, alguns conceitos são ensinados pelo seu nome inglês, sendo mais natural referir-se ao conceito por este idioma.”*

3.b *“Termos profissionais em inglês muitas vezes são mais aceitos no meio de trabalho.”*

3.c *“Eu trabalho como jornalista (...) Tudo o que envolve o meu trabalho tem algum nível de inglês.”*

3.d *“Ao falar sobre termos técnicos de administração e marketing.”*

A partir das respostas podemos notar a exigência da compreensão do inglês dentro do ambiente profissional, assim como sua omnipresença. Na resposta 3.d, se ressalta o uso do CS pelo próprio motivo que é referido.

Foi observada também uma função que não foi prevista na pesquisa teórica, o que chamaremos de uso introspectivo ou “*self-talk*”, onde os indivíduos relatam usar a alternância de códigos quando estão sozinhos:

4.a *“Quando eu estou sozinho estudando ou fazendo um trabalho e estou pensando em voz alta,*

*eu costumo variar entre as línguas na minha cabeça”*

*4.b “(misturo as línguas) Na escrita, quando preciso escrever algo”*

*4.c “(misturo as línguas) Na hora de pensar e falar sozinha”*

*4.d “Quando estou em situação de estresse.”*

*4.e “(misturo as línguas) Em meus pensamentos.”*

*4.f “Gosto de falar inglês. Falo comigo mesmo em inglês”*

*4.g “(misturo as línguas)...quando escrevo no diário, poesias ou crônicas.”*

Esse comportamento nos aponta para vários tópicos interessantes que podem alimentar futuras pesquisas. Primeiramente, os indivíduos apresentam gosto em trocar de línguas, sendo então uma motivação para a prática a mera diversão do falante consigo mesmo. Também nos direcionam as respostas 4.c, 4.d e 4.g, por exemplo, que evidenciam que o uso do CS também poderá auxiliar acesso e expressão de emoções, e auxílio no raciocínio.

Por último, destacamos algumas respostas à pergunta final do inquérito: “Nesta última etapa, fale um pouco mais sobre a sua experiência com a aprendizagem do inglês, a internet e o *codeswitching* (misturar línguas durante o discurso). Fique à vontade para compartilhar seus relatos e pontos de vista.”. Aqui, focamos em levantar o uso da indexação, assim como o contexto cultural e o papel fulcral da internet na instauração do hábito linguístico de praticar *codeswitching*:

*5.a “Meu primeiro contato com ingles foi aos 8 anos. Com o passar dos anos vivenciais a língua em diferentes contextos, como livros, blogs, vídeos, filmes, músicas. Isso se tornou parte do meu dia a dia e foi crescendo junto com a minha personalidade. Então também fez parte de quem eu sou e de como me expresso. Além de ter a oportunidade de conhecer pessoas de diferentes culturas e também aprender coisas novas e diferentes com elas, o que foi essencial para minha identidade.”*

*5.b “Uso muito o code switching na Internet quando falo com amigos estrangeiros. Não faço isso pessoalmente porque vivo em um ambiente muito monolinguê, mas faria sem nenhum problema se houvesse necessidade ou oportunidade “*

*5.c “Comecei a aprender inglês na adolescência e foi um processo gradual ao longo de vários anos. Foi uma aprendizagem principalmente autodidata reforçada pela presença universal do inglês atualmente. A internet permite um amplo contato com o inglês, seja pelo conteúdo disponível nesta língua quanto pela possibilidade de encontrar parceiros para comunicação. Como pesquisador, o inglês também é útil para mim pela necessidade de ler e escrever artigos acadêmicos e me comunicar com colegas de outros países seja através de emails ou na*

*participação de eventos internacionais como congressos e simpósios. O code-switching acaba se tornando comum em ambientes onde os falantes são nativos do português mas compartilham o inglês como segunda língua e geralmente um assunto abordado principalmente em inglês, seja em áreas acadêmicas, jogos ou outros contextos.”*

*5.d “Como tive contato com o inglês durante a infância e desenvolvi o interesse em aprender por conta de videogames, nunca tive de frequentar cursos para aprimorar. Em conversas com estrangeiros ainda tenho problemas com vocabulário, mas consigo ser entendido facilmente. Em relação ao codeswitching, antes tinha uma certa aversão em misturar línguas em conversas com falantes fluentes em português e uma visão um pouco “protecionista” sobre o português. No último ano mudei meu ponto de vista e deixei de focar nesses aspectos e, hoje em dia, me considero neutro em relação a esses hábitos, resultado de debates sobre assuntos técnicos em que grande parte da terminologia utilizada não tem tradução direta para o português. Mesmo com esse ponto de vista que tenho atualmente, ainda tenho por hábito tentar evitar o codeswitching.”*

*5.e “Comecei a aprender inglês na infância pois queria entender o que estava acontecendo quando jogava Pokémon em meu gameboy, iniciou por meio de associação e às vezes buscava as palavras que não conhecia no dicionário. Isso continuou nos próximos anos e quando entrei num curso de inglês me nivelaram muito acima da minha idade (uns 3 anos). Terminei o curso, formando com proficiência, e percebi que preferia o inglês por ser uma língua mais “fácil”. Continuei praticando a escrita e a fala pois jogava com muitos gringos, o que faço até hoje. (...) Sobre codeswitching, minha “bolha” costuma usar muitas gírias em inglês, e isso é muito frequente e natural, já que só é necessário conver uma palavra para emular todo um significado mais abrangente”*

A partir destas respostas conseguimos coletar uma série de pontos interessantes, como por exemplo o uso do inglês como ferramenta de socialização por meio da internet e *gaming*. Também observamos em a percepção do inglês como língua de cultura, como ilustrado no exemplo 5.d. Por outro lado, constata-se nesse mesmo exemplo a presença de juízos de valor negativos em relação ao uso do CS: o participante em questão se identifica como homem, está no grupo etário entre os 25 e 30 anos, e tanto o próprio quanto os pais possuem ensino superior completo, fatores que podem estar relacionados à visão “protecionista” sobre o português. Em futuros estudos, seria interessante analisar a fundo a correlação entre níveis de escolarização, idade e permeabilidade ao CS.

## **Conclusões**

Através desta pesquisa foi possível corroborar grande parte da teoria reunida aplicando-a a

falantes nativos do português do Brasil, assim como levantar algumas funções e características da prática do *codeswitching*. Notou-se que as mulheres, pessoas mais jovens e pessoas que se identificam como parte da comunidade LGBTQIAP+ praticam CS significativamente mais do que seus pares, e que a proficiência não é um fator necessário para a prática. Foi possível observar também que a internet e a reconfiguração mundial das *mass media* têm grande influência no uso do CS, fortalecendo o inglês como língua franca, permitindo a exposição de indivíduos ao inglês desde a infância e os conectando com diversas culturas. Esse contexto é o berço da superdiversidade, que se relaciona intimamente com o hábito de praticar *codeswitching*. Finalmente, realça-se o facto de este ser um terreno fértil para futuros estudos mais alargados, tanto no campo da sociolinguística quanto da semântica, e comparações entre diversas variedades do português.

## Bibliografia

Androutsopoulos, J. (2013). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19(2), 185–205. <https://doi.org/10.1177/1367006913489198>

Agha, A. (2007). *Language and Social Structure*. Cambridge: Cambridge University Press.

Androutsopoulos, J. (2007) Bilingualism in the mass media and on the internet. In M. Heller (ed) *Bilingualism: A Social Approach*. Basingstoke: Palgrave. 207-232

Blommaert, J., & Rampton, B. (2011). Language and Superdiversity. *Diversities*, 13(2), 1-21. Retrieved [07/06/2024] from [https://newdiversities.mmg.mpg.de/?page\\_id=2056](https://newdiversities.mmg.mpg.de/?page_id=2056)

Cunha, A. (2021). *Codeswitching entre português e inglês em falantes de português como língua materna: valores e funções* (dissertation). Faculdade de Letras da Universidade do Porto, Porto.

Dewaele, J., & Wei, L. (2013). Is multilingualism linked to a higher tolerance of ambiguity? *Bilingualism: Language and Cognition*

Giles, H., & Powesland, P. (1975). *Speech style and social evaluation*. Academic Press.

Grice, H. (1975). Logic and conversation. In *Syntax and semantics, vol. 3: Speech acts* (pp. 41-58). New York City: Academic Press.

Gumperz, J. (1982). *Discourse Strategies*. Cambridge University Press.

Hall, K. & Nilep, C. (2015). Code-switching, identity, and globalization. In Deborah Tannen, Heidi E. Hamilton & Deborah Schiffrin (eds.), *Handbook of Discourse Analysis*. Blackwell. pp. 597-619

Hall, K. (2014). Hypersubjectivity: language, anxiety, and indexical dissonance in globalization. *Journal of Asian Pacific Communication*, 24(2), 259–70.

Jacquemet, M. (2005). *Transidiomatic practices: language and power in the age of globalization*.

Myers-Scotton, C. (1983). The negotiation of identities in conversation: a theory of markedness and code choice. *International Journal of the Sociology of Language*, 44, 115–36.

MEC e Inep divulgam resultados do Censo Escolar 2023. (n.d.). *Gov.br Ministério Da Educação*. Retrieved February 2, 2024, from <https://www.gov.br/inep/pt-br/assuntos/noticias/censo-escolar/mec-e-inep-divulgam-resultados-do-censo-escolar-2023>

Nilep, C. (2006). “Code Switching” in Sociocultural Linguistics. *Colorado Research in Linguistics*, 19.

Rampton, B. (1995). *Crossing: Language and Ethnicity among Adolescents*. London: Longman.

Vertovec, Steven (2006). *The Emergence of Super-diversity in Britain*, Working Paper, 25. Oxford Centre on Migration, Policy and Society.

# Anexo

## Inquérito

1. Por favor, confirme ambas as afirmações abaixo: \*

*Check all that apply.*

- Consinto às condições acima descritas para o tratamento das respostas dadas a este formulário
- Confirmando que sou maior de idade e falante nativo de português do Brasil

2. Qual a sua idade? \*

*Mark only one oval.*

- Entre 18 e 25
- Entre 25 e 30
- Entre 30 e 40
- Acima de 40

3. Identifica-se como \*

*Mark only one oval.*

- Mulher
- Homem
- Não-Binária
- Prefiro não dizer

4. Identifica-se como \*

*Mark only one oval.*

- Heterossexual
- Queer/LGBTQIAP+
- Prefiro não dizer

5. Qual é o seu grau de escolaridade? \*

*Mark only one oval.*

- Ensino fundamental incompleto
- Ensino fundamental completo
- Ensino médio completo
- Ensino Superior completo
- Pós-graduação completa
- Mestrado
- Doutorado

6. Qual é o grau de escolaridade máximo que seu **pai/mãe** possui? \*

*Mark only one oval.*

- Ensino fundamental incompleto
- Ensino fundamental completo
- Ensino médio completo
- Ensino Superior completo
- Pós-graduação completa
- Mestrado
- Doutorado
- Não tenho certeza

7. Você realizou o **ensino fundamental** na rede pública ou privada? \*

*Mark only one oval.*

- Pública
- Privada
- Ambos

8. Você realizou o **ensino médio** na rede pública ou privada? \*

*Mark only one oval.*

- Pública
- Privada
- Ambos

9. Você mistura inglês e português quando fala? \*

*Mark only one oval.*

- Sim  
 Não

10. Com que idade você considera ter se tornado fluente em inglês? \*

*Mark only one oval.*

- Até os 7 anos  
 Entre os 7 e 11 anos  
 Entre os 11 e os 18 anos  
 Após os 18 anos  
 Other: \_\_\_\_\_

11. Qual considera ser seu nível de fluência em inglês? \*

*Mark only one oval.*

- A1 - Iniciante  
 A2 - Básico  
 B1 - Intermediário  
 B2 - Independente  
 C1 - Proficiência eficaz  
 C2 - Domínio pleno

12. Como você desenvolveu a fluência em inglês? \*

*Check all that apply.*

- Escola
- Curso auxiliar
- Internet
- Videogames
- Séries e filmes
- Other: \_\_\_\_\_

dados de produção

13. Você considera que trocar de línguas durante o discurso denota: \*

*Mark only one oval.*

- Pouco conhecimento do português
- Médio conhecimento do português
- Pleno conhecimento do português

14. Você considera que trocar de línguas durante o discurso denota: \*

*Mark only one oval.*

- Pouco conhecimento do inglês
- Médio conhecimento do inglês
- Pleno conhecimento do inglês

15. Você considera que trocar de línguas durante o discurso facilita a comunicação no geral? \*

*Mark only one oval.*

- Sim
- Não

16. Justifique a resposta anterior \*

---

---

---

---

---

17. Nos seguintes contextos, mistura inglês e português? \*

*Mark only one oval per row.*

	Nunca	Raramente	Às vezes	Frequentemente	Constantemente
<b>Nas relações profissionais</b>	<input type="radio"/>				
<b>Nas relações acadêmicas</b>	<input type="radio"/>				
<b>Entre amigos</b>	<input type="radio"/>				
<b>Em família</b>	<input type="radio"/>				

18. Há algum outro contexto no qual você mistura as línguas? \*

---

---

---

---

---

19. Responda de acordo com o que é mais verdadeiro para você \*

Mark only one oval per row.

	discordo	discordo parcialmente	indiferente	concordo parcialmente	concordo fortemente
<b>Troco de línguas consoante o assunto em meu círculo social</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Troco de línguas durante uma conversa (em público ou não) para ter mais privacidade</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Troco de línguas pois exige menos esforço mental</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Troco de línguas pois exprime mais fielmente minha linha de pensamento</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

20. Há algum outro motivo pelo qual você troca de línguas? \*

---

---

---

---

---

21. Nesta última etapa, fale um pouco mais sobre a sua experiência com a aprendizagem do inglês, a internet e o *codeswitching* (misturar línguas durante o discurso). Fique à vontade para compartilhar seus relatos e pontos de vista. \*

---

---

---

---

---

# ANÁLISE DA DISTÂNCIA IDEOLÓGICA ENTRE PARTIDOS POLÍTICOS DE UM PONTO DE VISTA PRAGMÁTICO E DISCURSIVO

Rafaela Gomes de Morais Fernandes da Gama

Orientado pela Professora Doutora Isabel Margarida Duarte

## Resumo

Este trabalho visa analisar as diferentes estratégias discursivas utilizadas no discurso político, em específico no género de texto *debate político*, mostrando que as diferenças sociopolíticas dos indivíduos, inseridos num determinado contexto social, resultam em fenómenos linguísticos relevantes no estudo pragmático-discursivo. Para tal, foram utilizados como objeto de estudo dois debates políticos televisivos entre representantes de partidos concorrentes às eleições legislativas portuguesas que apresentam diferenças ideológicas, exemplificadas através de fenómenos linguísticos como Marcadores Discursivos, Cortesia e Descortesia, Formas de tratamento, Mecanismos de atenuação e intensificação e Atos de fala. Estas estratégias interligam-se, também, com a construção do indivíduo e da sociedade através do poder persuasivo e da ação da linguagem.

Palavras Chave: Estratégia discursiva; Discurso político; Pragmática; Debate televisivo.

## Abstract

This work aims to analyse the various discursive strategies used in political discourse, specifically in the text genre of political debate. Proving that the sociopolitical differences of individuals inserted in given social context are a relevant factor in pragmatic-discursive study. To this end were used as the object of study two televised political debates between representatives of parties competing in the Portuguese legislative elections, demonstrating ideological differences, exemplified through linguistic phenomena such as discourse markers, politeness and impoliteness, address terms, mechanisms of attenuation and intensification, and speech acts. These strategies are also connected to the construction of the individual and society through the persuasive power and action of language.

Keywords: Discursive strategy; Political discourse; Pragmatics; Televised debate.

## **1. Introdução**

Para cada tipo de discurso é possível relacionar traços linguístico-discursivos que o caracterizam com o respetivo género e, neste caso, que permitem analisar o género de discurso “debate eleitoral” de um ponto de vista pragmático.

Estudos como os de Kerbrat-Orecchioni (1990; 2004), no panorama de estratégias linguístico-discursivas no debate político, evidenciam a importância da análise de tópicos como formas de tratamento, marcadores discursivos e movimentos argumentativos, mecanismos de atenuação e intensificação, fenómenos de cortesia e descortesia, atos de fala, etc.

Há alguns aspetos pragmáticos que são centrais para a comunicação: O que é dito; o modo como é dito; a intenção com que é dito; onde é dito; quando é dito; para quem é dito; os papéis sociais dos interlocutores e atitudes e comportamentos dos participantes.

Num debate político televisivo, os candidatos têm como objetivo “vencer” o debate, transportando os ideais do seu partido. Com isto, gera-se uma distância não só social mas também linguística entre cada debatedor, consequência das suas diferenças político-sociais.

Os partidos políticos em Portugal identificam-se com contextos sociais e históricos distintos, o que os aproxima ou afasta. Portanto, quando colocados num confronto verbal frente a frente, por via dos seus representantes, há notórias diferenças linguísticas. São esses diferentes usos dos mecanismos linguísticos que andam lado a lado com a distância política e mostram a forma como cada debatedor tenta persuadir o votante, descredibilizando o seu adversário.

### **1.1. Objetivos**

Este estudo constrói-se sobre a ideia de que quanto mais afastados ideológica, histórica e socialmente dois partidos políticos são, mais diferenças irão ser visíveis nas suas interações, principalmente dentro de meios de comunicação social como a televisão, onde há possibilidade de confronto e espetacularização da política (CHARAUDEAU, 2011, MARQUES, 2013). Coloca-se como hipótese a crescente agressividade<sup>13</sup> e decrescente discussão de ideias no debate graças a diferentes utilizações de mecanismos linguísticos.

Para analisar o debate político eleitoral coloca-se, como crucial para a construção da intertextualidade e reflexão teórica deste trabalho (MARQUESI, 2021), a seguinte pergunta de partida:

---

<sup>13</sup> O conceito de presidencialização mostra que, cada vez mais, os debatedores se veem como candidatos a primeiro-ministro e a líderes dos partidos, o que gera maior distância entre cada um deles e uma maior agressividade verbal

- **Como se materializam, no debate televisivo, as diferenças entre partidos políticos mais próximos e mais afastados em ideologias, de um ponto de vista pragmático e discursivo?**

## **2. Metodologia e *corpus***

O objeto de estudo foram dois debates eleitorais, alvo de uma análise do ponto de vista pragmático e discursivo.

Dia 5 de fevereiro de 2024, deu-se início a 30 debates televisivos entre os representantes dos partidos concorrentes às eleições legislativas, em Portugal. Os debates frente-a-frente, que aconteceram até dia 23 de fevereiro, envolveram os candidatos dos partidos com representação parlamentar (BE, CDU, Livre, PS, PAN, PSD, IL e CH).<sup>14</sup> Transmitidos pela televisão portuguesa em horários não fixos, os debates tiveram uma duração, em média, de 30 minutos, com o pressuposto de repartição equilibrada de tempo.

Dos dois debates políticos selecionados fazem parte Paulo Raimundo, Mariana Mortágua e André Ventura. Paulo Raimundo apresenta-se como secretário-geral do Partido Comunista Português na Coligação Democrática Unitária (CDU), Mariana Mortágua como coordenadora do Bloco de Esquerda (BE) e André Ventura como presidente do Chega (CH).

O critério de seleção usado para escolher estes dois debates teve por base o objetivo de facilitar a resposta à questão elaborada inicialmente. Partindo do princípio de que os partidos com valores políticos distantes tendem a usar diferentes e mais agressivos tipos de mecanismos linguístico-discursivos do que os partidos mais próximos usam, foram selecionados debates que melhor se enquadram nessa visão prototípica de distância de ideais. É possível considerar o BE e a CDU como partidos de esquerda, mais próximos a nível de ideologias, e o BE e o CH como dois partidos totalmente afastados, que se posicionam como esquerda e extrema-direita, respetivamente.

Assim, estes antagonismos ideológicos entre os partidos selecionados funcionam como ferramentas ideais para analisar as diferenças linguísticas que existem no género de debate político.

Divididos pelos confrontos BE x CDU e CH x BE, os debates ocorreram dia 11 e dia 13 de fevereiro, respetivamente.

---

<sup>14</sup> Bloco de Esquerda (BE); Coligação Democrática Unitária (CDU), Livre (LIVRE), Partido Socialista (PS), Pessoas-Animais-Natureza (PAN), Partido Social Democrata (PSD), Iniciativa Liberal (IL) e Chega (CH).

O primeiro (BE x CDU) <sup>15</sup>foi transmitido às 22 horas pela SICNotícias e teve como moderador a pivô e jornalista Rosa de Oliveira Pinto. Este teve um total de 73.000 espectadores e uma duração de 31:52 minutos.

O segundo debate <sup>16</sup>de dia 13 de fevereiro transmitido pela RTP3 às 22 horas e moderado pelo jornalista João Adelino Faria, teve como partidos convidados Chega (CH) e Bloco de Esquerda (BE). Durou 36:55 minutos e foi visto por 246.442 pessoas.

O corpus deste trabalho consiste numa transcrição<sup>17</sup> feita com ajuda do serviço de transcrição LX-Transcriber presente na Infraestrutura de Investigação para a Ciência e Tecnologia da Linguagem Portulan Clarin.

A transcrição do debate entre os partidos Bloco de Esquerda e CDU tem uma extensão de 6232 palavras. No caso dos partidos Chega e Bloco de Esquerda, o texto tem um total de 10.182 palavras.

É importante referir que, apesar de ser possível trabalhar a análise apenas através da audição dos debates, a sua transcrição escrita dá uma maior materialidade ao texto, o que facilita a deteção de mecanismos e conseqüente análise, através da qual é possível observar, mais pormenorizada e objetivamente, os efeitos de sentido presentes nos textos.

Destaca-se ainda que, para a análise do debate nas eleições legislativas em Portugal, compreende-se o peso fundamental das campanhas no cenário político português, onde se destaca a progressiva presidencialização e a imagem do líder ganha cada vez mais importância.

### **3. Contexto e Discurso**

Como forma de enquadrar o trabalho de análise proposto anteriormente, realça-se como fundamental expor a definição e relação entre o contexto e o discurso.

Kerbrat-Orecchioni (2002) interliga língua, discurso e sociedade quando afirma que: o linguista está particularmente interessado nas regras que regem as línguas e nas suas manifestações discursivas, visto que a língua não é apenas o “conjunto de regras ou regularidades que fundamentam a produção e interpretação dos enunciados” mas também “aquelas que controlam o funcionamento de fenómenos como atos de fala diretos e indiretos, mecanismos inferenciais, sistema de turnos de fala, sequência de intervenções e trocas, conectores pragmáticos e conversacionais, marcadores de relacionamento interpessoal e rituais de descortesia, etc. Integra, portanto, dimensões linguísticas, textuais, psicológicas e sociais”.

---

<sup>15</sup> <https://www.youtube.com/watch?v=Hi-zMh-BTzg> [acedido a 03/03/2024].

<sup>16</sup> <https://www.youtube.com/watch?v=XObCSMpLH9Q> [acedido a 03/03/2024].

<sup>17</sup> Esta transcrição não segue nenhum modelo, porém, é o mais próxima possível de um texto escrito legível.

Segundo Dominique Maingueneau (2014), o discurso é uma forma de ação. Existe, dentro dele, uma troca oral, onde dois ou mais indivíduos organizam os seus enunciados em função da atitude do outro e compreendem o efeito que as suas palavras têm no outro (interatividade). É, também, contextualizado: “Não diremos que o discurso intervém num contexto, como se o contexto fosse somente uma moldura, um cenário; na realidade, não existe discurso senão contextualizado” (MAINGUENEAU, 2014 p.26). E é regido por normas, visto que cada ato implica normas particulares, tal como qualquer comportamento social. Entende-se, portanto, que o discurso é portador de um conjunto de normas que cumprem as expectativas dos sujeitos envolvidos na atividade verbal e depende do contexto em que decorre.

Como indissociável do discurso, o contexto é uma noção cuja definição é considerada ainda em progresso que implica várias perspetivas como as de: Halliday & Hassan (1987 [1976]); van Dijk (2003, 2004); Kerbrat-Orecchioni (2002) e Koch, Morato & Bentes (2011), entre outros.

Para Halliday & Hassan (1976), a importância do contexto no estudo da língua toca na sua exterioridade, “é um estado de coisas que em parte está organizado a priori, e em parte está associado a uma significação que emerge da sua própria organização (HANKS, 2008, p. 67). Quando analisado, atenta-se ao processo de construção da comunicação dos interlocutores.

The main thesis of my theory of context is that contexts should not be defined in terms of some kind of social situation in which discourse takes place, but rather as a mental representation, or model constructed by the speech participants of or about such a situation. (VAN DIJK, 2004, p.349)

A relação entre interlocutores, contexto e discurso é fundamental para a consideração dos géneros discursivos e para a análise proposta.

#### **4. Género “Debate Político Televisivo”**

O carácter fluído e transitório dos discursos e a sua irrecusável dependência contextual deu origem a uma problemática na tipologização dos discursos e à criação de inúmeras propostas de classificação.

A instabilidade dos géneros – produto socio-histórico – é fator e ao mesmo tempo consequência da mudança. (...) E desta forma, a variabilidade social e cultural, a mudança em geral que se opera nas sociedades fará necessariamente emergir novas formas, mesmo novos géneros, onde os discursos ganharão também novas configurações. (MARQUES, 2003 p.194)

Contudo, e apesar das dificuldades de classificação, seguir-se-á a proposta do debate político televisivo como género de texto, dentro do tipo de discurso “debate político” marcado pela interseção entre o debate e o discurso mediático televisivo (MARQUES, 2017).

Tendo como referência a proposta de Adam (2001) delimita-se, então, o texto em análise como um exemplo prototípico do tipo de *discurso político* e género textual *debate político televisivo* marcado pela interseção entre o debate e o discurso mediático televisivo (MARQUES, 2017). Orecchioni (1990) afirma que o debate se baseia numa discussão mais organizada e menos informal do que uma conversa.

Pode-se caracterizar o debate político através da composição do respetivo *quadro participativo* (1990). Este quadro resume o debate político à existência de quatro entidades sendo elas dois ou mais participantes ou debatedores que funcionam como adversários, um moderador e um público.

O papel do moderador dentro do funcionamento do debate consiste em regular e estruturar as intervenções dos participantes, e assume, ocasionalmente, papel de participante regular. Tem como função especificar as regras do debate, gerir a alternância e o tempo dos turnos de fala e conduzir o debate concedendo e retirando a palavra, suavizar possíveis agressividades através de interrupções, e defender os interesses informativos do público.

#### **Especifica as regras:**

(1) “Moderador: Vamos então a este debate, como é habitual, fizemos o sorteio e é a Mariana Mortágua que vai começar.” (BE x CH)

#### **Gere a alternância dos turnos de fala e conduz o debate concedendo e retirando a palavra:**

(2) “Moderador: [André Ventura, vou passar a palavra.]” (BE x CH)

Gestão do tempo concedido a cada participante:

(3) “Moderador: [Deixe só equilibrar e já lhe dou a palavra.]” (BE x CDU)

(4) “Moderador: [Vou ter que equilibrar o tempo] Mariana Mortágua, responder a isto.” (BE x CH)

#### **Suaviza possíveis agressividades com interrupções:**

(5) “AV: (...) Hipocrisia e cara de pau, mas eu agora vou-lhe dizer mais, já que quer falar sobre isso, agora vamos até ao fim.

Moderador: [André Ventura, (( )) para que fique claro...]” (BE x CH)

#### **Funciona como porta-voz do público ao defender os seus interesses informativos:**

(6) “Moderador: Proponho começarmos por assuntos concretos e medidas concretas. Mariana Mortágua, crise de habitação em Portugal” (BE x CH)

O público atua também como um elemento importante pois é ele que orienta a argumentação dos debatedores ao obrigá-los a agir conforme a sua opinião (P. BRETON, 1996, 1998). Aliás, embora não esteja presente fisicamente, é um dos elementos centrais do género debate, pois é a ele que os interlocutores querem convencer.

(7) “[...] eu quero que os *portugueses* percebam isto, (...)” (BE x CH)

Por último, os debatedores apresentam, teoricamente, papéis equilibrados e simétricos construindo estrategicamente o seu *ethos*<sup>18</sup>. Maingueneau (2008) dá também importância ao conceito de *ethos pré-discursivo*, que mostra a imagem que o outro constrói em relação ao enunciador, a partir de conhecimentos adquiridos anteriormente. Isto é um fator de grande relevância no debate político, pois os debatedores entram na interação com uma opinião social já formada sobre o adversário. O mesmo, aliás, acontece com o público. Relaciona-se, também, com o conceito de persuasão: “Diremos, portanto, que o orador que mostra em seu discurso um carácter honesto parecerá mais digno de crédito aos olhos de seu auditório” (EGGS, 2005 p.29).

O debatedor procura mostrar ao público os seus valores, afirmando *eu sou isto* e causar boa impressão, o que funciona como um evento crucial para a construção das relações candidato-eleitor.

Devido à espetacularização e contexto de debate televisivo, há um carácter pré-fixado como o tema, a duração do debate, ordem de intervenções, linguagem formal (com o objetivo de trazer uma certa credibilidade ao debatedor (*ethos discursivo*)) e um conjunto de estratégias linguísticas associadas a este tipo de discurso, acabando por se interligar com o lado argumentativo, que funcionará, então, como uma arma verbal.

## 5. Análise do corpus

### 5.1. Marcadores discursivos e Movimentos argumentativos

Como consequência da dificuldade em estabelecer critérios que definam claramente os contornos da classe dos MDs, o termo *marcadores discursivos* continua a não ser completamente evidente e consensual. Porém, reconhecem-se os estudos de Fraser (1999), Schiffrin (1987) e Fuentes Rodríguez (2022) como fundamentais nesta área e afirma-se que marcadores discursivos (MD) atuam como estratégia de persuasão dos votantes e controlo do diálogo, contribuindo para a força argumentativa do discurso político.

---

<sup>18</sup> Por *ethos*, designa-se a imagem que o locutor constrói de si no discurso para exercer uma influência sobre o seu alocutário. Apresenta uma dimensão social, permitindo ao orador exprimir-se de acordo com o seu carácter criado. (EGGS, 1999)

Os MDs podem subdividir-se em organizadores textuais, ajudando na estruturação do discurso e permitindo ao debatedor transmitir de forma eficiente e clara uma grande quantidade de informações, facilitando a compreensão do público. Pode assumir a função de organizadores enumerativos (ADAM, 2008), permitindo a manutenção da linha de raciocínio:

(8) “AV: *Primeiro* sobre esta questão do despedimento é falsa, aliás, o Chega foi o primeiro partido a comprometer-se com a [...]” (BE x CH)

(9) “PR: [...] o que o PCP não exclui, *primeiro*, para nós a questão da forma como disse, não é uma questão determinante [...]” (BE x CDU)

(10) “AV: *Segundo*, sim, João, eu não vou mentir, a primeira coisa que nós vamos fazer, se tivermos essa oportunidade [...]” (BE x CH)

Os reformuladores discursivos, de acordo com Duarte (2022), apresentam de igual forma um carácter persuasivo, visto que é uma forma de facilitar a compreensão do que foi dito anteriormente:

(11) “AV: [Foi o Chega, foi o Chega. Não, não. Agora deixe-me explicar], *isto é*, a Mariana Mortágua tem que perceber [...]” (BE x CH)

A argumentação é, também, marcada pelas sequências metadiscursivas, que ajudam a reforçar o que o enunciador pretende dizer, mas muitas vezes pode transmitir agressividade e afetar a imagem do locutor:

(12) “AV: Hipocrisia e cara de pau, mas eu agora *vou-lhe dizer mais*, já que quer falar sobre isso, agora vamos até ao fim.” (BE x CH)

Em debates com maior distância política, observam-se marcadores conversacionais típicos de registo oral formal, como forma de regulação discursiva (DUARTE, 2022):

(13) “AV: Mas *ouça*. O Chega tem, é público.” (BE x CH)

(14) “AV: *Olhe, olhe*, vou-lhe dar dados, vou lhe dar dados que não são meus.” (BE x CH)

Observa-se uma maior necessidade de interrupção e organização por parte do moderador:

(15) “Moderador: André Ventura, *deixe concluir, por favor*.” (BE x CH)

O uso dos conetores adversativos contrastivos como *mas* e *agora* que, geralmente, antecedem atos assertivos de discordância ou movimentos argumentativos possibilita ao enunciador manter o seu turno de fala:

(16) “AV: *Mas responda-me*, diga-me só se faz sentido. Entrar qualquer pessoa de qualquer maneira. Acha que isto faz sentido?” (BE x CH)

(17) “AV: Foi o Chega, foi o Chega. Não, não. *Agora deixe-me explicar*.” (BE x CH)

Por outro lado, em debates entre debatedores próximos politicamente, encontram-se menos sequências metadiscursivas e os elementos de concordância são mais explícitos:

(18) “PR: É como a Mariana, afirmou. *Estou perfeitamente de acordo*, nós convergimos em muita coisa, divergimos em outras.” (BE x CDU)

Desta forma, verifica-se a relação entre o contexto, estratégias argumentativas e a força persuasiva dos marcadores discursivos.

## **5.2. Fenómenos de (des)cortesia e depreciação linguística**

Apesar de o conceito de cortesia ter vindo a ser tratado sob diferentes perspetivas, parte-se da definição de Kerbrat-Orecchioni (2006) que designa *cortesia* como “todos os aspetos do discurso que são regidos por regras, cuja função é preservar o carácter harmonioso da relação interpessoal” e de Seara (2014): “um princípio que rege a dinâmica interacional e contribui para o equilíbrio social”.

Trata-se, então, de várias estratégias discursivas utilizadas pelos interlocutores que permitem construir e manter a autoimagem positiva e ser aceite pelo público, para assim o persuadir. As estratégias de cortesia seguintes partem da perspetiva de Brown & Levinson (1987).

**Inclusão do alocutário e dos ouvintes na interação, mostrando cooperação, explicitando reciprocidade:**

(19) “MM: Boa noite, *boa noite ao Paulo Raimundo*.” (BE x CDU)

(20) “PR: Olhe, boa noite. *Boa noite à Mariana*, boa noite a quem nos está a ouvir.” (BE x CDU)

(21) “AV: Boa noite. *Boa noite à Mariana também*.” (BE x CH)

**Reivindicação de pertença ao mesmo grupo do ouvinte, usando marcas de identidade coletivas:**

(22) “MM: [...] E foi a força do Bloco de Esquerda nesse momento, como *outras forças de esquerda*, [...]” (BE x CDU)

(23) “PR: [...] *nós* estaremos lá, com a força que *o povo* nos der [...]” (BE x CDU)

**Reivindicação de aspetos em comum entre o falante e ouvinte (pontos de vista, opiniões, atitudes, empatia, conhecimentos):**

(24) “MM: [...] *o Paulo Raimundo lembrava* hoje comemoram-se 17 anos do Referendo pela despenalização do aborto, é uma vitória de todas as mulheres em Portugal [...]” (BE x CDU)

(25) “PR: *Nós* estamos a trocar opiniões, ideias programáticas, umas convergem, outras divergem e mesmo assim e estamos a construir pá, estamos todos a construir resultados também eleitorais é isto também.” (BE x CDU)

(26) “MM: O Bloco de Esquerda e o Partido Comunista Português têm um percurso de convergência em áreas essenciais [...]” (BE x CDU)

Por outro lado, o fenómeno da descortesia, neste contexto, é uma estratégia que tem como objetivo o desequilíbrio social e interacional, produzindo um carácter socioemocional negativo. A linguagem política cria, dependendo das circunstâncias históricas, sociais e culturais<sup>19</sup>, um conjunto de procedimentos que servem para denegrir o adversário, objetivo central do universo político, caracterizando então a dinâmica do discurso político como indelicada.

Quando descortês, o indivíduo vê-se como uma figura autoritária “*Sou descortês, provooco medo e submissão.*”, está carregado de valores ideológicos (BOLÍVAR, 2005) e em inter-relação com a agressividade verbal, apesar de esta não ser sempre necessária.

Silva (2013) afirma que a descortesia tem como objetivo desconstruir a imagem do candidato adversário e a busca pessoal de conquista de votos dos possíveis eleitores (aqui está presente a notória importância do público dentro do discurso político). O locutor usa estratégias de persuasão que variam de acordo com o público, a sociedade e a época em que o discurso persuasivo acontece e é considerado descortês se desrespeitar as regras pré-estabelecidas pelo contexto do debate político.

Para Blas Arroyo (2010), a *descortesia institucional* cria-se com a espetacularização do conflito e crescente polemicidade do debate eleitoral. Pode-se dizer que a sociedade está marcada pela procura de polémica como forma de entretenimento e que é dessa forma que se obtém audiência.

**Construção de uma imagem positiva do falante**, caracterizando o oponente como incompetente. Este procedimento, de *sedução por intensificação*, é utilizado para acentuar críticas e ganhar credibilidade (*ethos* de credibilidade) ao mostrar que o locutor é um candidato confiável e competente:

(27) “AV: Boa noite. Boa noite à Mariana também. Bom, antes de mais é preciso dizer isto, o o *Chega no seu programa de habitação toca nas duas vertentes*, a oferta e a procura, *coisa que o Bloco de Esquerda ignora por completo.*” (BE x CH)

**Sobreposição de voz**, que deveria ser rara em debates políticos, de acordo com as regras pré-estabelecidas por esse tipo de discurso, é um indicador de discordância do adversário, usado para intensificar a imagem de fragilidade do outro, interrompendo o seu discurso e o seu raciocínio. Porém, o enunciador põe em risco a sua própria imagem, pois está a transgredir as

---

<sup>19</sup> Sublinhe-se que, para estudar a cortesia como elemento da pragmática é necessário ter em atenção fatores pessoais, sociais e culturais do interlocutor para não se cair num etnocentrismo que resultaria numa análise errada do fenómeno.

regras do programa televisivo e pode ser criticado pelo público. Interromper o outro é um processo descortês cujo objetivo é dominar o debate e ter vantagem sobre o outro:

(28) “AV: [Ó Mariana, ainda bem, ainda bem.]

MM: [E o medo que coloca às pessoas.]

AV: [Eu quero vos é longe. Eu quero-vos é longe o mais possível.]

MM: [Tenha calma, eu ouvi-o. Tenha calma. Por causa das posições.]” (BE x CH)

**Provocação por desqualificação** do outro partido, apontando falhas:

(29) “AV: Ele não se preocupa em construir mais, apesar de Portugal ter construído na última década menos que na última década do Estado Novo. *Não se preocupa com isso. Não se preocupa em dar benefícios fiscais a quem já sente completamente atolado em impostos.*” (BE x CH)

(30) “AV: [Sabe que está a *mentir*.]” (BE x CH)

**Insulto ao interlocutor** com o intuito de denegrir a imagem deste (KAUL DE MARLANGEON, 2005):

(31) “AV: [...] Quem foi contra o aumento dos prazos de prescrição dos crimes? A Mariana Mortágua. *A Mariana Mortágua está a fazer hoje o exercício da maior hipocrisia política do nosso tempo.*” (BE x CH)

(32) “AV: Sabe o que é que se chama isso? *Hipocrisia e cara de pau* [...]” (BE x CH)

(33) “AV: [*Eu quero-vos é longe o mais possível. Eu quero-vos é longe o mais possível. Ainda bem que não se querem sentar ao meu lado.*]” (BE x CH)

**Ataques Ad personam** como estratégia fundamental na argumentação discursiva, caracterizam-se como um ataque pessoal ao indivíduo, à sua dignidade, à sua imagem pessoal e não apenas às suas posições políticas. Ernotte & Rosier (2001) mostram que, muitas vezes, a fronteira entre o insulto político e o insulto individual se apresenta como muito ténue. Aqui, o enunciador pretende extremar a disputa, intensificando as emoções do outro (descontentamento, indignação, ódio):

(34) “AV: [Agora aa Mariana Mortágua,] sobre imigração é um bocado como a história, desculpe referir-me a sua avó, do despejo da sua avó. [...] *É é uma fantasia. A Mariana Mortágua vive nesta fantasia, tal como o despejo da sua avó, que não podia ter existido.*” (BE x CH)

AV continua a insultar o adversário (MM) com base numa história pessoal, onde mostra claramente que pretende desvalorizar o outro através de um ataque *ad personam*.

Observa-se, então, uma relação entre a descortesia e distância política e cortesia e proximidade política.

### 5.3. Formas de tratamento

As formas de tratamento, no debate, são um recurso linguístico pelo qual os debatedores se interpelam mutuamente no decorrer da interação e revelam, através de escolhas linguísticas, o nível de intimidade/aproximação ou distância entre eles. Num debate político, assumem diferentes valores semântico-pragmáticos e contribuem diretamente para a construção e preservação da imagem do locutor.

Nas seguintes tabelas são apresentadas as formas de tratamento predominantes no debate e o número de ocorrências de cada uma com apoio de exemplos.

Formas de tratamento AV → MM	Nº de vezes	Exemplos
A Mariana	39	“A Mariana vê isso como um convite à corrupção.”
Ó Mariana	8	“Ó Mariana Mortágua. Sabe que isso não é verdade. [...]”
Você	3	“Você quer mais impostos sobre as pessoas.”
Vocês	17	“É que foi o regime que vocês criaram.”
Formas de tratamento MM → AV	Nº de vezes	Exemplos
Doutor André Ventura	25	“E o <i>Doutor Ventura</i> é contra.”
O André Ventura	0	-----
Você	0	-----
Vocês	0	-----

Tabela 1. Formas de tratamento no diálogo entre André Ventura e Mariana Mortágua.

Formas de tratamento PR → MM	Nº de vezes	Exemplos
A Mariana	2	“[...] é como a Mariana, afirmou [...]”
Você	0	-----
Vocês	0	-----
Formas de tratamento MM → PR	Nº de vezes	Exemplos
O Paulo Raimundo	1	“[...] o Paulo Raimundo lembrava hoje comemoram-se [...]”
Você	0	-----
Vocês	0	-----

Tabela 2. Formas de tratamento no diálogo entre Mariana Mortágua e Paulo Raimundo.

O candidato AV utiliza 8 vezes um tratamento com função de vocativo (*Ó Mariana*), maioritariamente para chamar a atenção do oponente e marcar grande discordância, o que o torna descortês e revela agressividade. Também AV trata o adversário pelo nome próprio, simulando uma proximidade que não existe e, ao mesmo tempo, mostrando desprezo. Quanto ao uso do *você*, FT não aceitável nas normas do debate político, AV usa-o 3 vezes, o que demonstra falta de controlo sobre o seu próprio discurso e desrespeito para com o interlocutor. De seguida, o uso da FT *vocês* funciona, num debate político, como uma forma de referência ao grupo ou partido político a que pertence o interlocutor. Porém, no discurso de AV, existe com o objetivo de agrupar negativamente o adversário e os seus votantes, criticando-os, mas pondo em risco a sua imagem:

(35) “AV: [...] *Vocês* querem porque, é fácil, *vocês* querem destruir a economia toda e ficar o Estado a mandar em tudo.” (BE x CH)

Juntamente com o uso do *nós* inclusivo, onde contrastam assim os seus ideais com os do adversário, como uma distinção entre o bem (*nós*) e o mal (*vocês*):

(36) “AV: [...] há uma diferença, é que *nós* não queremos destruir a economia e o emprego. *Vocês* querem.” (BE x CH)

O *nós* também funciona como uma forma de criar cumplicidade entre o candidato e os eleitores que pretende persuadir.

Mortágua, por sua vez, apenas utiliza *Doutor André Ventura* e *Doutor Ventura* como forma de tratamento, revelando um carácter mais cortês e formal, adequado para um debate político televisivo, e, ao mesmo tempo, mantendo-se afastada do seu adversário.

No debate entre PR e MM, quase não ocorrem tratamentos diretos nem o uso do *vocês*, o que demonstra uma relação de respeito, cortesia e conseqüente proximidade entre os interlocutores.

#### **5.4. Mecanismos de atenuação**

A atenuação, mecanismo analisado a partir de estudos como da cortesia de Briz, da pragmática de Schneider e da sociolinguística de Cestero, por exemplo, mostra ter como finalidade manter e proteger a imagem positiva do interlocutor<sup>20</sup> com um conjunto de perceções como valor dos papéis sociais e culturais, como somos vistos e assim ajudá-lo a desempenhar um papel cortês na interação, ou seja, como afirma Albelda (2016), a atenuação é capaz de cuidar da imagem do locutor, evitando atos ameaçadores e caso sejam proferidos, repará-los.

---

<sup>20</sup> De acordo com a perspectiva de Goffman sobre a ligação do *self* (identidade pessoal).

Albelda mostra que as funcionalidades da atenuação se concentram tanto na autoproteção do enunciador como na relação enunciador-interlocutor.

Porém, de acordo com Briz (2013), a sua função é mais do que reparar atos ameaçadores. Em certos contextos como o debate político, visto que é um fenómeno altamente contextual, a atenuação é um fenómeno linguístico-discursivo estratégico. É não só uma forma de compensação, mas também uma estratégia usada para alcançar objetivos específicos na interação verbal.

A estratégia autoprotetora da atenuação consiste numa forma de minimizar a força ilocutória de determinados atos “trata-se de uma estratégia para “velar por si mesmo”, isto é, para salvaguardar o “eu”, isto é, buscar não se responsabilizar ou minorar as responsabilidades, ou ainda, ser politicamente correto. [...] É assim um mecanismo autocêntrico.” (BRIZ, 2013, p. 286)

Existem diversas estratégias que funcionam como atenuadores através de diferentes recursos (BRIZ, 2013, p. 111):

**Uso de verbos modais:**

(37) “PR: [há uma outra questão que foi aqui colocado eu *parece-me* que *é*/ tem uma certa importância [...]]” (BE x CDU)

(38) “MM: [...] mas *parece-me* que...ah... essa é a questão essencial [...]]” (BE x CDU)

**Procedimentos de reformulação:**

(39) “AV: *Ou seja*, foram pessoas que de facto deram um grande contributo à sociedade.” (BE x CH)

(40) “AV: [Foi o Chega, foi o Chega. Não, não. Agora deixe-me explicar], *isto é*, [...]]” (BE x CH)

**Controladores de contacto:**

(41) “AV: *Olhe*, pelo menos dois candidatos seus foram condenados por terrorismo.” (BE x CH)

(42) “AV: É evidente, *olhe*, eu acho que igualdade de género é a luta justíssima entre homens e mulheres num país [...]]” (BE x CH)

A atenuação, como fenómeno pragmático que joga com a imagem e responsabilidades dos falantes (ALBELDA, 2016) e atua como estratégia discursiva essencial para a moderação da agressividade verbal, crucial para a manutenção da eficácia comunicativa em contextos de interação intensa e de natureza agressiva como debates políticos.

Ao observar-se uma maior frequência de uso de atenuadores no debate entre AV e MM, conclui-se, portanto, que existe uma relação mais distante e propensa a agressividade.

## 5.5. Mecanismos de intensificação

Aqui, a intensificação revela a implicação do locutor no enunciado e o compromisso do mesmo com o que é dito, o que produz efeitos a nível social e comunicativo.

Para além de fenómeno socio-pragmático (Hermann, 1988), a intensificação, de acordo com Cafii & Janney (1994), está relacionada com o conceito de capacidade emotiva. Entende-se que capacidade emotiva se baseia em habilidades comunicativas afetivo-relacionais que permitem aos falantes interagir e negociar potenciais conflitos interpessoais e atingir certos objetivos na fala. Para além disso, esta ligação entre intensificação e capacidade emotiva, permite interpretar índices de sentimentos, atitudes e crenças dos falantes em relação aos tópicos abordados em diferentes situações.

Albelda (2005) explica que se intensifica a atitude a partir de vários procedimentos de intensificação com uma finalidade pragmática, procurando alcançar determinados objetivos conversacionais.

Abordando a intensificação proposicional, encontram-se maioritariamente os seguintes tipos de mecanismos linguísticos:

### **Intenso uso de advérbios:**

(43) “AV: [Não, deixe-me só dizer isto, é  *muito* simples, não, é  *muito* simples.]” (BE x CH)

(44) “AV: [Quando é que ele falou em despedimentos?  *Nunca, nunca.*]

MM: Reduzir currículos, não, mas eu quero-lhe dizer.

AV: [ *Nunca.*]

### **Repetição:**

(45) “AV: [Ah, mas está a julgar que me está a  *intimidar*. Julga que me  *intimida* a mim.]” (BE x CH)

(46) “AV: [Ó desculpa lá ó João, eu tenho de responder a isto porque  *isto é grave. Isto é grave.*]

(47) “AV: A Mariana Mortágua não se interessa com isso, interesse por  *perseguir, perseguir, perseguir e perseguir.*” (BE x CH)

### **Adjetivos:**

(48) “MM: [Com uma lei  *cruel*, com uma lei  *cruel*,] com uma lei  *cruel* e por isso eu defendo e defenderei o direito à habitação de todos.” (BE x CH)

A intensificação é, portanto, um fenómeno que tem como fim produzir certos efeitos na comunicação. É uma estratégia argumentativa que, a partir de mecanismos como o uso de advérbios e adjetivos a nível lexical e repetição a nível sintático, permite ao locutor persuadir o ouvinte, pois, segundo Briz (1998), a intensificação procura obter determinados efeitos e/ou impactos comunicativos.

## 5.6. Atos de fala

Entende-se por ato de fala (ato linguístico) um tipo de ação envolvida num determinado uso da linguagem.

Austin (1962) afirma que há sempre uma dimensão acional na linguagem e distingue o significado (o que as palavras dizem) e *força ilocutória* (o que as palavras fazem).

Austin propõe uma teoria pragmática da linguagem e afirma que ao dizer algo, palavras são sempre enunciadas com dado significado. É isto que Austin designa, na sua tricotomia, por ato locutório, que corresponde ao ato em que se produz significado.

De seguida, Austin chama ato ilocutório ao poder de fazer; com o objetivo de realizar algo.

Por ato perlocutório, Austin refere-se a um ato que produz efeitos no alocutário, como assustá-lo, ameaçá-lo, etc. É um ato de fala que, em virtude da sua força ilocutória<sup>8</sup>, produz no(s) recetor(es), num determinado contexto, o efeito pretendido pelo emissor (intimidar, persuadir, seduzir, etc.) Estes três atos estão presentes em simultâneo.

Em 1969, J. R. Searle retoma o projeto de Austin (1962) na sua obra *Speech Acts* e apresenta a Teoria dos Atos de Fala. Para Searle, o ato de fala simboliza a unidade mínima da comunicação linguística e resume-se à junção do conteúdo proposicional com a força ilocutória<sup>21</sup>. Este é a base da teoria de Searle.

O autor define também a seguinte taxinomia dos atos ilocutórios: Atos compromissivos; Atos assertivos; atos diretivos; atos expressivos e atos declarativos.

Porém, serão tratados, neste estudo, apenas os atos de fala compromissivos, assertivos e diretivos, existindo em maior número e distanciando mais visivelmente os dois debates analisados.

### 5.6.1. Atos compromissivos

Os atos *compromissivos*, de acordo com o *Dicionário Terminológico* (2008), são os atos ilocutórios nos quais o locutor se compromete a adotar um comportamento futuro.

(49) “AV: A primeira coisa que nós *vamos fazer*, se tivermos essa oportunidade, é fiscalizar o dinheiro.” (BE x CH)

(50) “MM: E *queremos determinar* a governação porque *queremos resolver* o problema da habitação, porque *queremos resolver* o problema da saúde, porque *queremos* escolas com professores, porque *queremos* um país melhor e porque *queremos* um país sem medo.” (BE x CH)

---

<sup>21</sup> Função que um enunciado assume num dado contexto (promessa, pedido, ordem, etc.);

(51) “PR: Nós *temos que começar* e aos lucros da banca primeiro para, olhe, a partir das comissões, desde logo nas comissões. Esta é uma questão fundamental [...]” (BE x CDU)

O uso do futuro *queremos + infinitivo* e a presença frequente da estrutura *temos de + infinitivo* como ato de prometer é central no pensamento político, visto que mostra o compromisso dos candidatos políticos com a satisfação do público, da sociedade. É através de promessas com valor deôntico que os candidatos mostram a sua seriedade e constroem o seu *ethos*, aumentando a sua credibilidade, estratégia usada por todos os candidatos presentes nos debates em análise.

### 5.6.2. Atos assertivos

Nos atos *assertivos*, o locutor compromete a sua responsabilidade sobre a verdade da proposição. A argumentação através dos atos assertivos é marcada pela presença da crítica e da denúncia com o objetivo de provocar e ofender o adversário, muitas vezes atenuada com verbos modais.

Partindo de uma dimensão narrativa, o uso de atos assertivos com função descritiva pretende relatar e avaliar ações e estados de coisas existentes e, simultaneamente, projetar ações futuras através de atos compromissivos. Aqui, o locutor parte do formato “se critica x ações então promete x mudanças”.

(52) “AV: O Chega no seu programa de habitação toca nas duas vertentes, a oferta e a procura, coisa que o Bloco de Esquerda ignora por completo.” (BE x CH)

Este encadeamento descrição-crítica-promessa, apesar de parte estrutural do discurso político, é mais frequente em grupos políticos opostos, ocorrendo com maior frequência no debate CH x BE.

Interligando o uso de atos assertivos e compromissivos, o candidato procura construir a sua credibilidade através de um *ethos* de competência e persuasão do eleitor, atribuindo ao adversário um *ethos* de incompetência. Contudo, um exagero de tentativas de criação de um *ethos* competente tem como consequência uma imagem negativa e perda de credibilidade.

### 5.6.3. Atos diretivos

Atos *diretivos* consistem na tentativa de levar o interlocutor a agir de acordo com o conteúdo proposicional do ato de fala enunciado.

O debate entre AV e MM está repleto de atos diretivos. Porém, estes são direcionados principalmente um ao outro, mirando na persuasão do eleitor através do ataque ao outro.

(53) “AV: Já sabemos que é assim, portanto, da próxima vez que falar disso *faça* só uma coisa, leia um bocadinho antes, em vez de vir para um debate.” (BE x CH)

(54) “MM: [*Deixe os truques.*]” (BE x CDU)

(55) “MM: [*Acalme-se, acalme-se. Deixe-me terminar.*]” (BE x CH)

Por outro lado, no debate entre PR e MM, os atos diretivos são direcionados a um *nós inclusivo* e, aqui, o objetivo não é atacar o outro mas exclusivamente motivar os eleitores, acabando também por funcionar como uma estratégia de persuasão.

(56) “PR: Nós *temos que começar* e aos lucros da banca primeiro para, olhe, a partir das comissões, desde logo nas comissões [...]” (BE x CDU)

O discurso (dizer) e a ação (fazer) relacionam-se funcionando como base para as trocas sociais (CHARAUDEAU, 2005) e os atos de fala, por sua vez, são a base para a relação entre o uso da linguagem e a realização de ações.

O discurso político, como campo em que se estabelecem relações entre *linguagem, ação e poder*, mostra que estas ações estão ligadas a jogos de poder na sociedade. Considera-se, então, que os atos de fala, no contexto político, atuam como estratégias que interligam argumentação e a sociedade e, de acordo com o contexto em que estão inseridos, têm o poder de produzir reações no alocutário, visto que são, frequentemente, proferidos com o objetivo de seduzi-lo e/ou persuadi-lo.

A análise da utilização de atos de fala é, por isso, fundamental para a avaliação crítica das mensagens de cada candidato, realçando as diferenças ideológicas de cada partido político, compreendidas através das diferenças nas escolhas de estratégias argumentativas de cada indivíduo presente na interação.

## 6. Conclusão

Existem aspetos pragmáticos que, como foi referido inicialmente, são cruciais para a comunicação humana e que nos revelam as atitudes, intenções e objetivos conversacionais dos locutores.

Estudos como os de Kerbrat-Orecchioni (1990; 2004), mostram que a análise de tópicos como os que foram abordados é crucial para o desenvolvimento da área da pragmática e da pergunta de pesquisa colocada neste estudo.

Verifica-se, através da análise elaborada ao longo do trabalho com o apoio de exemplos retirados da transcrição dos debates selecionados, a confirmação da hipótese colocada, isto é: que quanto mais afastados ideológica, histórica e socialmente dois partidos políticos são, mais diferenças de comportamento linguístico irão ser visíveis nas suas interações verbais, o que gera uma crescente agressividade e decrescente discussão de ideias.

Observa-se, também, que esta situação linguística é consequência da crescente presidencialização, já que os candidatos políticos têm como objetivo “vencer” o debate a

qualquer custo, mesmo que isso implique o ataque ao outro. Com isto, é criada uma distância não só social, onde dois ou mais candidatos não exprimem concordância a nível político-social, mas também linguística, onde os interactantes não conseguem manter um equilíbrio interacional e uma comunicação eficaz.

Como se pode constatar, ao longo da análise realizada, o debate entre André Ventura (CH) e Mariana Mortágua (BE) contém mais fenómenos linguísticos que promovem a agressividade e o desacordo entre os candidatos. Contudo, no debate entre Paulo Raimundo (CDU) e Mariana Mortágua (BE), para além de ser um debate onde, de facto, se partilham mais ideias e opiniões, é um debate com zero estratégias manipuladoras e concentra-se apenas numa interação com o objetivo de oferecer algo ao público.

Sublinha-se que, neste estudo, todas as estratégias discursivas referidas anteriormente são uma parte importante do discurso político e o seu uso gira à volta da tentativa de persuasão do público, pois todas são portadoras de um carácter argumentativo e espelho das crenças do locutor e das práticas ou necessidades da sociedade.

A linguagem é um instrumento de poder e construção social, o que torna o discurso uma forma de perpetuar e promover ideologias e resistir ou incentivar mudanças sociais. Hoje mais do que nunca, graças à *direitização* política e ao crescimento de discursos fascistas, o estudo de discursos políticos é importantíssimo, não só para o desenvolvimento da área da pragmática discursiva, mas também para a construção justa e consciente da sociedade, pois com ferramentas para compreender as estratégias utilizadas nos discursos ao nosso redor, será possível resistir a possíveis manipulações políticas.

Realço a hipótese de uma expansão desta análise e apelo à comunidade académica a continuação de estudos dentro deste tema.

## 7. Bibliografia

Adam, J.-M. (2001). En finir avec les types de textes. In M. Ballabriga (Org.), *Analyse des discours. Types et genres: communication et interprétation* (pp. 25-43). EUS.

\_\_\_\_\_. (2008). *A linguística textual. Introdução à análise textual dos discursos*. São Paulo: Cortez Editora.

Albelda, M. (2005) *La intensificación en el español coloquial*. Tesis doctoral. Universitat de Valencia.

\_\_\_\_\_. (2016). Sobre la incidencia de la imagen en la atenuación pragmática. *Revista Internacional de Lingüística Iberoamericana*, 27(1), 19-32.

Blas Arroyo, J. L. (2010). La descortesía en contextos de telerrealidad mediática. Análisis de un corpus español. In F. Orletti & L. Mariottini (Eds.), *(Des)cortesía en español. Espacios teóricos y metodológicos para estudio* (pp. 183-209). Roma-Estocolmo: Università degli Studi Roma Ter-EDICE.

Bolívar, A. (2005). *Estudios de la (des)cortesía en español*. Buenos Aires: Dunken.

Breton, P. (1996, 1998). *A Argumentação na Comunicação*. Lisboa: Publicações Dom Quixote.

Briz, A. (1998). *El español coloquial en la conversación – Esbozo de pragmatogramática*. Barcelona: Ariel.

\_\_\_\_\_. (2013). A atenuação e os atenuadores: estratégias e táticas. In *Linha d'Água*, 26(2).

Brown, P., Levinson, S. C. (1987). *Politeness: some universals in language use*. Cambridge: Cambridge University Press.

\_\_\_\_\_. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Ed.), *Questions and politeness: strategies in social interaction* (pp. 56-289). Cambridge: Cambridge University Press.

Caffi, C. (1999, 2007). On Mitigation. *Journal of Pragmatics*, 31, 881-909.

Charadeau, P. (2011). Réflexions pour l'analyse du discours populiste. *Mots. Les langages du politique*, 97, 101-116.

\_\_\_\_\_. (2005). *Le discours politique. Les masques du pouvoir*. Paris: Vuibert. Duarte,

I. M. (2022). Marcadores Discursivos y Argumentación en Debates Electorales.

Eggs, E. (1999, 2005). Ethos Aristotélico, Convicção e Pragmática Moderna. In R. Amossy (Org.), *Imagens de si no Discurso: a Construção do Ethos* (pp. 29-56). São Paulo: Contexto.

Ernotte, P., Rosier, L. (2001). *Le lexique clandestin*. Louvain-La-Neuve: Duculot.

Halliday, M. A. K., Hassan, R. (1984). *Cohesion in English*. Longman. (Original work published 1976)

Hanks, W. F. (2008). O que é contexto. In A. C. Bentes, R. C. Rezende, M. R. Machado (Orgs.), *Língua como prática social: das relações entre língua, cultura e sociedade a partir de Bourdieu e Bakhtin* (pp. 75-90). São Paulo: Cortez.

Koch, I., Morato, E., Bentes, A. (2011). Ainda o contexto: algumas considerações sobre as relações entre contexto, cognição e práticas sociais na obra de Teun van Dijk. *ALED*, 11(1), 79-91.

Kerbrat-Orecchioni, C. (1990). *Les interactions verbales*. Paris: Armand Colin.

\_\_\_\_\_. (1999, 2002). *L'énonciation. De la subjectivité dans le langage*. Paris: Armand Colin.

\_\_\_\_\_. (2002). Système linguistique et ethos communicatif. *Cahiers de Praxématique*, 38, 35-57.

\_\_\_\_\_. (2006). *Análise da Conversação: princípios e métodos* (C. Piovenazi Filho, Trad.). São Paulo: Parábola.

Maingueneau, D. (2008). Problemas de ethos. In S. Possenti, M. C. P. de Souza-E-Silva (Orgs.), *Cenas da enunciação* (pp. 25-45). São Paulo: Parábola Editorial.

\_\_\_\_\_. (2014). *Discours et analyse du discours – Introduction*. Paris: Armand Colin.

Marques, M. A. (2003). Renovação dos Discursos – Novas Formas de Interação e Legitimação dos Interlocutores. In *Diacrítica, Ciências da Linguagem*, 17(1), 189-219. Braga: Centro de Estudos Humanísticos da Universidade do Minho.

\_\_\_\_\_. (2013). Politique, humour et campagne électorale. Les enjeux d'une politique-spectacle. *Mots. Les langages du politique*, 10, 61-75.

\_\_\_\_\_. (2017). Debate Eleitoral Português: Presidencialização e Estratégias de Atenuação Linguística em Situação de Confronto Político. *Linha D'Água*, 30(1), 9-33.

Marquesi, S. C. (2021). Originalidade na escrita acadêmica: a pergunta de pesquisa em foco. *Lingvarvmarena*, 12, 115-129.

Martín Zorraquino, M. A., Portóles Lázaro, J. (1999). Los marcadores del discurso. In I. Bosque, V. Demonte (Dir.), *Gramática descriptiva de la lengua española* (t. 3, cap. 63, pp. 4051-4213). Madrid: Espasa Calpe.

Meyer-Hermann, R. (1988). Atenuación e intensificación (análisis pragmático de sus formas y funciones en español hablado). *Anuario de Estudios Filológicos*, XI, 275-290.

Seara, I. R. (Coord.). (2014). *Cortesia: Olhares e (Re)Invenções*. Lisboa: Isabel Roboredo Seara e Chiado Editora.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Silva, L. A. (2013). Descortesia e (des)construção da imagem pública. In D. Preti & M. Q. Leite (Orgs.), *Comunicação na fala e na escrita* (pp. 93-120). São Paulo: Humanitas.

Van Dijk, T. (2004). Text and Context of Parliamentary debates. In P. Bayley (Ed.), *Cross-Cultural Perspectives on Parliamentary Discourse* (pp. 339-372). John Benjamins.