

# Controle metodológico: criação de um *corpus* para estudos sobre o processamento lexical de indivíduos bilíngues e multilíngues

Nalim Barbosa Pinto<sup>a</sup>

Ana Beatriz Arêas da Luz Fontes<sup>b</sup>

## Resumo

*Estudos sobre o acesso lexical multilíngue utilizam tarefas de decisão lexical para comparar o processamento de palavras cognatas com não-cognatas e investigar se há co-ativação das línguas durante o processamento em uma língua só. Para entender como funciona esse processo em trilingües Português/Inglês/Italiano a partir de uma tarefa de decisão lexical em italiano, foi elaborado um corpus seguindo as seguintes etapas: seleção de 1129 substantivos em italiano que possuem de 4 a 8 letras, e suas frequências de ocorrência; tradução das palavras italianas para o português e para o inglês; classificação das palavras como cognatas nas três línguas, cognatas italiano-português, cognatas italiano-inglês e não-cognatas nas três línguas, de acordo com um índice de similaridade ortográfica (SO); e classificação de ambiguidade lexical no italiano. Foram retiradas da lista as palavras homônimas, restando apenas palavras sem ambiguidade ou polissêmicas. A seguir, pseudopalavras foram criadas a partir de palavras existentes em italiano. Após essa etapa, separamos as palavras por condição e calculamos as médias de frequência, número de letras e de sílabas e similaridade ortográfica para averiguar, estatisticamente, se eram iguais. Por fim, a lista de palavras que irá compor a tarefa de decisão lexical em italiano conta com três condições de 22 palavras cognatas cada, uma de 66 não-cognatas e uma de 132 pseudopalavras. Essa tarefa nos permitirá verificar a co-ativação do inglês e do português no acesso lexical de palavras em italiano.*

**Palavras chave:** Acesso lexical, multilinguismo, controle metodológico.

Recebido em: 15/05/2018

Aceito em: 08/07/2018

<sup>a</sup> Grupo de pesquisa de Processamento de Linguagem Bilíngue (ProLinGue) e do Laboratório de bilinguismo e cognição (LABICO), Universidade Federal do Rio Grande do Sul. E-mail: ana.fontes@ufrgs.br.

<sup>b</sup> Departamento de Línguas Modernas da Universidade Federal do Rio Grande do Sul. E-mail: nalim.barbosa@ufrgs.br.

## Introdução

A psicolinguística é o ramo da ciência cognitiva que estuda a relação entre a mente humana e a linguagem (FIELD, 2003). Ela explora, através de estudos de caráter experimental, o comportamento linguístico em falantes de um ou mais idiomas, tanto de maneira individual quanto em grandes grupos. Dessa forma, busca-se compreender como a estruturação da mente humana influencia a comunicação.

O que se pode notar ao realizar-se uma pesquisa no Portal de Periódicos da Capes é que o número de pesquisas feitas dentro da área da psicolinguística vem aumentando nos últimos 25 anos. Uma rápida busca no Portal, utilizando o termo “psicolinguística” no campo assunto, retorna 30 publicações entre os anos de 1991 e 2000. Entre os anos de 2001 a 2010 esse número sobe para 71, e entre 2011 e 2018 passa para 94 trabalhos (pesquisado no dia 14/05/2018). Como mostram os dados, esta linha de pesquisa está em crescimento.

No entanto, uma de suas subáreas, a que estuda o processamento de linguagem por bilíngues e multilíngues, ainda carece de mais estudos. Isso fica evidente, por exemplo, quando outra pesquisa é realizada no Portal de Periódicos da Capes. Utilizando os termos “psicolinguística” e “bilinguismo” no campo assunto, o resultado retorna o total de 50 trabalhos publicados de 2000 a 2018 (pesquisado no dia 14/05/2018); quando se usa “psicolinguística” e “multilinguismo” a pesquisa retorna 18 trabalhos publicados de 2003 a 2018 (pesquisado em 14/05/2018). É nesse contexto de escassez de estudos sobre o processamento da linguagem por sujeitos bilíngues e multilíngues que o presente estudo se insere. Mais especificamente, este trabalho se insere na área da psicolinguística que estuda o acesso lexical em indivíduos bilíngues e multilíngues.

Dá-se o nome de *acesso lexical* ao processo que nos permite recuperar itens lexicais quando necessário (FIELD, 2003), isto é, o processo pelo qual acessamos nosso léxico para reconhecer uma ou mais palavras que nos são apresentadas. Para alguns autores esse acesso lexical, no caso de indivíduos falantes de mais de uma língua, é seletivo, ou seja, o bilíngue/multilíngue é capaz de “desligar” uma das suas línguas e operar em um modo monolíngue (CARAMAZZA; BRONES, 1979, p. 212-214).

Para outros autores, entretanto, tal acesso é não-seletivo, isto é, as línguas do bilíngue/multilíngue permanecem ativas paralelamente durante o processo, sem que seja possível “desligar” uma delas (VAN HELL; DIJKSTRA, 2002, p. 780-789; VAN ASSCHE et al., 2012, p. 174; LEMHÖFER; DIJKSTRA, 2004, p. 533-550; DE GROOT et al., 2000, p. 397-428; BARCELOS, 2016; PICKBRENNER, 2017).

Para verificar se existe tal co-ativação paralela das línguas, alguns estudos utilizam tarefas de reconhecimento de palavras que têm como estímulos itens cognatos e não-cognatos. Cognatas, nesse contexto, são as traduções que possuem ortografia igual ou semelhante entre as línguas (*curva*, em português e italiano e *curve* em inglês), enquanto que as não cognatas não compartilham ortografia (*erba*, italiano; *grama*, português e *grass*, inglês). Esses estudos demonstram que bilíngues/multilíngues reconhecem palavras cognatas mais rapidamente e com maior acurácia do que as não cognatas. Isto se deve ao fato de que existe uma sobreposição semântica e ortográfica entre as línguas no caso das cognatas (*vida - vita*), enquanto que para as não cognatas a sobreposição é apenas semântica (DE GROOT, 1992a, p. 1001-1018; DE GROOT, 1992b, p. 389-412).

Para conduzir estudos sobre o acesso lexical bilíngue/multilíngue como os citados acima, é necessário que se considere questões de controle metodológico. Essas pesquisas, de caráter experimental, necessitam deste controle para que seja possível identificar os efeitos das variáveis em estudo, e assim, obter um resultado mais confiável (CHRISTENSEN; JOHNSON; TURNER, 2011). Tal controle metodológico é especialmente importante na criação de instrumentos de pesquisa que visam a avaliar o efeito de variáveis independentes em uma variável dependente.

Na criação de uma tarefa de decisão lexical para multilíngues português-inglês-italiano em que se verifica o efeito do tipo de palavra (cognata versus não cognata) no tempo de reconhecimento, por exemplo, é importante considerar e controlar a frequência de uso das palavras escolhidas para a tarefa. A frequência de uso das palavras influencia a rapidez com que os participantes as reconhecem, ou seja, palavras de alta frequência são percebidas e produzidas de forma mais eficiente do que as palavras de baixa frequência (HOWES;

SOLOMON, 1951, p. 401-410; BRYSSBAERT; NEW, 2009, p. 977-990). A frequência poderia, então, influenciar os resultados não nos permitindo avaliar apropriadamente o efeito de tipo de palavra, e por isso é controlada de forma que as palavras cognatas e não cognatas da tarefa possuam a mesma média de frequência de uso.

O objetivo do presente estudo é, então, compilar um conjunto metodologicamente controlado de estímulos linguísticos para ser utilizado em estudos de reconhecimento de palavras com indivíduos bilíngues de português-italiano, e multilíngues português-inglês-italiano. Viu-se necessária a criação deste corpus devidamente controlado no que diz respeito a características como frequência de uso, número de sílabas, número de letras e similaridade ortográfica, que abrangesse estas três línguas, porque não encontramos um corpus com tais características em estudos anteriores.

Embora não tenhamos encontrado na literatura um corpus que atenda às nossas necessidades, identificamos a existência de outros criados com finalidades semelhantes a nossa. Prior, Macwhinney e Kroll (2007, p. 1029-1038), por exemplo, criaram um conjunto de normas de tradução para grandes amostras de palavras em inglês e espanhol. Tal conjunto é composto por 670 palavras em inglês e 762 palavras em espanhol, selecionadas com base em respostas de 80 participantes bilíngues submetidos a uma tarefa de tradução única.

O conjunto de palavras em inglês contém 241 substantivos, 79 verbos, e 350 itens sintaticamente ambíguos, apresentando variados níveis de frequência, imaginabilidade, concretude, familiaridade e idade de aquisição (IdA). O conjunto de palavras em espanhol, por sua vez, inclui 525 substantivos e 237 verbos, que também apresentam variados níveis de frequência, imaginabilidade, concretude e familiaridade. Classificações de IdA não foram disponibilizadas para este último grupo de palavras. Entretanto, as mesmas foram classificadas como cognatas ou não-cognatas por um grupo de 30 falantes nativos de inglês sem conhecimento de espanhol, português, francês ou italiano.

Além de criar um recurso para utilização na pesquisa com bilíngues inglês-espanhol, Prior et al (2007, p. 1029-1038) examinaram a relação entre probabilidade de tradução e uma série de variáveis lexicais, tais como nível de frequência e concretude. Dentre outros resultados, os autores verificaram que bilíngues menos proficientes tendem a realizar traduções menos prováveis com mais frequência do que bilíngues com maior proficiência.

Já no estudo de Tokowicz e Kroll (2007, p. 727-779), o objetivo era examinar a relação entre concretude de palavras e a ambiguidade em termos de número de tradução, e estudar se os bilíngues acessam os mesmos conceitos nas duas línguas. Para este estudo foram criadas e traduzidas duas listas de palavras feitas por oito estudantes bilíngues inglês-espanhol e oito estudantes bilíngues espanhol-inglês.

Para formar essas duas listas foram usadas 500 palavras já utilizadas em estudos anteriores (MILLER; KROLL, 2002, p. 614-628). As autoras controlaram uma série de variáveis relacionadas às traduções das palavras, como, por exemplo, sua forma plural. Neste caso, palavras que consistiam na forma plural de outras não foram incluídas na lista. As palavras foram divididas em duas listas combinadas de modo que não diferiram em relação ao comprimento da palavra em inglês ou espanhol. Cada participante traduziu palavras de apenas uma das listas; a atribuição da lista para a direção da tradução foi contrabalançada entre os participantes, de modo que as palavras que um participante tenha traduzido do inglês para o espanhol foram traduzidas por outro participante do espanhol para o inglês. As palavras foram listadas em ordem aleatória nos folhetos e os participantes deveriam dar apenas uma tradução para cada palavra (a primeira tradução que eles pensassem).

A lista de estímulos obtida foi utilizada em uma série de três experimentos. No primeiro, participantes bilíngues inglês-espanhol traduziram palavras com apenas uma tradução. No segundo experimento, foram traduzidas palavras com uma ou mais traduções. Finalmente, no terceiro experimento, examinou-se como a interação entre concretude e ambiguidade influencia o processamento da linguagem em uma tarefa de decisão lexical. Os resultados encontrados indicam que não há vantagem de palavras concretas sobre palavras abstratas em termos de latência de tradução quando estas palavras são associadas a apenas um significado. Em contrapartida, quando se consideram palavras com mais de um significado associado, palavras concretas são traduzidas mais rapidamente. Por fim, observou-se que, na decisão lexical, palavras abstratas são mais influenciadas pela ambiguidade do que palavras concretas, indicando que a interação entre concretude e número de significados é uma propriedade geral do sistema de processamento de linguagem.

Fraga et al (2017, p. 1-10), por outro lado, desenvolveram um banco de dados de palavras ambíguas em espanhol (SAW). Palavras polissêmicas e homônimas de sete letras ou menos foram retiradas de bases de dados espanholas anteriormente publicadas, e seus

significados verificados no dicionário *Real Academia Espanhola* (RAE) (2001). A maior parte destes significados era de substantivos, existindo ainda uma parcela pequena de adjetivos e verbos. Para chegar ao conjunto final de palavras, foi realizada uma série de classificações em relação ao número de entradas no dicionário e o número de sentidos; além de exclusões de palavras com significados regionais, em desuso, ou que só tinham uma entrada com um sentido no dicionário. Por fim, os autores selecionaram as palavras restantes com duas ou mais entradas no dicionário, mas que não foram incluídas em bancos de dados anteriores. Assim, o conjunto final incluiu 210 palavras: 34 homógrafos com 8 sentidos ou mais, 43 homógrafos com 7 sentidos ou menos, 59 palavras polissêmicas com 8 sentidos ou mais, e 74 palavras polissêmicas com 7 sentidos ou menos.

Este banco de dados foi utilizado em uma tarefa de recuperação de significados com o objetivo reunir todos os significados relatados por um grupo de participantes para o conjunto de palavras espanholas disponibilizado. Para isso, as 210 palavras foram distribuídas em sete listas de trinta palavras, gerando 21 folhetos diferentes. Cada participante recebeu um folheto contendo as palavras para as quais ele deveria fornecer diretamente todos os significados que conseguisse pensar, ou formular frases que demonstrassem tais significados.

As respostas obtidas foram atribuídas a categorias de significado seguindo três etapas. Primeiro, calculou-se o número de significados dados pelos participantes. Segundo, foram avaliados todos os significados para cada palavra, de acordo com critérios estatísticos e de dicionário. Finalmente, os regionalismos e as respostas ilegíveis foram considerados inválidos. Após essas etapas, calculou-se a correlação entre o número de sentidos do dicionário e o número total de sentidos dos participantes. Os resultados evidenciaram uma correlação significativa, demonstrando a existência de uma relação entre o número de sentidos incluídos no dicionário e os fornecidos pelos participantes. Além disso, as análises de decisão lexical e os tempos de nomeação mostraram um poder preditivo substancial para o número de sentidos dos participantes, mas não para os do dicionário.

É importante destacar que, a fim de que o banco de dados desenvolvido se adequasse à tarefa realizada, os autores criaram um método de classificação apontando a propriedade da palavra (polissêmica, homônima, homógrafa), seu número de letras, de significados (pelos participantes e pelo dicionário) e de entradas no dicionário. A existência deste controle metodológico é de extrema importância não só para a obtenção de resultados confiáveis, mas

também para permitir que o banco de dados desenvolvido seja utilizado em futuras pesquisas desenvolvidas com monolíngues do espanhol e com bilíngues, ou outros estudos relacionados à ambiguidade de tradução.

No contexto nacional, Siqueira, Gil e Melo (2010, p. 127-145) também desenvolveram um estudo sobre controle metodológico que busca contribuir para a elaboração de testes psicolinguísticos. Neste trabalho as autoras construíram uma lista com 135 sentenças envolvendo material linguístico metafórico e não metafórico. Para isto elas realizaram dois estudos que envolveram familiaridade e convencionalidade (estudo 1), e níveis de familiaridade, alerta e valência (estudo 2). Para o primeiro estudo foram elaborados dois questionários, um que averiguava o grau de convencionalidade das expressões e outro que avaliava a sua familiaridade. Em cada questionário havia oito metáforas nas quais os participantes deveriam responder em uma escala o quão convencional cada metáfora era, assim como o quão familiar elas lhe pareciam.

Foram calculadas as médias dos dados obtidos e os resultados mostraram que existe relação entre convencionalidade e familiaridade, ou seja, quando o participante considera um mapeamento conceitual sendo convencional, ele irá considerar que as expressões linguísticas metafóricas que derivam desses mapeamentos também são familiares.

Para o segundo estudo também foram elaboradas tabelas específicas que envolviam familiaridade, alerta e valência. Para a realização deste estudo, foram elaboradas 135 sentenças, sendo 67 literais e 68 metafóricas. O tamanho das sentenças foi controlado e pareado em relação ao número de palavras e caracteres, assim, a lista de sentenças ficou dividida entre sentenças pequenas com frases de quatro até cinco palavras, com 15 a 25 caracteres; e sentenças grandes para aquelas de cinco a sete palavras, com 26 a 35 caracteres. As sentenças foram também controladas quanto à complexidade sintática por dois juízes, no intuito de tornar parelha as estruturas e evitar que a complexidade gramatical influenciasse nos resultados de testes envolvendo outras funções cognitivas. As 135 frases foram divididas em quatro blocos, três de 34 sentenças e um de 33. Os tipos de sentenças (literal ou metafórica), o tamanho das frases de cada grupo e a valência presumida (positiva, negativa e neutra) foram distribuídos nos blocos de forma equitativa. Após a realização da tarefa, foram calculadas as médias e desvios padrão de todas as sentenças. Assim, as sentenças obtiveram as classificações

para a variável alerta: frases de alerta médio, alerta baixo e alerta alto. Para a valência, após os cálculos das médias, foi elaborada a seguinte escala: sentenças negativas, neutras ou positivas. Por fim, como produto deste trabalho, obteve-se uma lista de 135 frases classificadas quanto ao tipo (literal ou metafórico) e ao tamanho (grande ou pequeno).

Os estudos acima citados desenvolveram listas de palavras e métodos de controle conforme a necessidade que encontraram. A maioria deles criou as listas através dos participantes que os auxiliaram nas traduções das palavras ou em referenciar os significados de cada palavra. No presente estudo, porém, a lista de palavras e suas características foram desenvolvidas com a retirada de palavras de um *corpus* e suas traduções com o auxílio de dicionários, o que o difere dos anteriores. Estes estudos nos mostram que o controle metodológico é uma peça importante no decorrer de uma pesquisa porque garante resultados mais confiáveis e apurados. Dessa forma, o presente estudo visa a elaborar um conjunto metodologicamente controlado de estímulos linguísticos para ser utilizado em estudos de reconhecimento de palavras com indivíduos bilíngues de português-italiano, e multilíngues português-inglês-italiano. Para isso, descreveremos cada passo e cada decisão metodológica tomada no processo de compilação do conjunto.

## Metodologia e resultados

### *Procedimentos de elaboração de estímulos para tarefa de decisão lexical em português-inglês-italiano*

Para a elaboração de estímulos de uma tarefa de decisão lexical para multilíngues português-inglês-italiano foi feita uma seleção de palavras em italiano através do *Corpus e Lessico di Frequenza dell'Italiano Scritto*. O CoLFIS é um banco de dados lexical da escrita italiana, baseado em um corpus de mais de 3 milhões de palavras. Deste corpus, extraímos, aproximadamente, 1100 substantivos que possuem de 4 a 8 letras e suas frequências de ocorrência. Foram selecionados aqueles com frequências entre 100 a 2700. Por uma decisão metodológica, escolhemos substantivos para manter apenas uma classe gramatical, e porque substantivos são palavras de base dos quais outras palavras podem derivar, assim como os verbos (BRYSAERT; NEW, 2009, p. 977-990). Além disso, optou-se por



substantivos porque o seu significado depende menos do contexto (PRIOR; MACWHINNEY; KROLL, 2007, p. 1029-1038).

Foram descartadas da lista de 1129 substantivos as palavras que estavam repetidas por ter mais de um significado, ou mais de um uso. Quando a palavra se repetia, foi selecionada a de maior frequência. Quando a mesma palavra aparecia na lista, uma vez no singular e outra no plural, foi descartada a palavra no plural, restando assim a palavra singular do par.

Estas palavras foram então traduzidas para o português e para o inglês com a ajuda de um dicionário do italiano (DASTOLI, 2007) e um dicionário online do inglês ([www.wordsmyth.net](http://www.wordsmyth.net)). Após essa tradução, foram classificados como cognatos os pares de traduções com similaridade ortográfica (SO) de 0,6 ou mais, e como não-cognatos os pares com SO inferior a 0,6. Esse critério foi desenvolvido a partir de um algoritmo criado por Van Orden (1987, p. 181-198) que calcula essa similaridade ortográfica e que foi utilizado anteriormente por Barcelos (2016) e Pickbrenner (2014). A partir desta classificação obtivemos as listas de itens para cada condição: (a) cognatas nas três línguas (*curva* para italiano e português e *curve* para inglês), (b) cognatas italiano-português (*infanzia* para italiano, *infância* para português e *childhood* para inglês), (c) cognatas italiano-inglês (*storie* para italiano, *stories* para inglês e *histórias* para português); e (d) não-cognatas nas três línguas (*uova* para italiano, *ovos* para português e *eggs* para inglês), de acordo com um índice de similaridade ortográfico.

A equação utilizada para calcular a similaridade ortográfica, conforme descrito por Van Orden (1987, p. 181-198), segue abaixo:

$$SO = 10 * ((50F + 30V + 10C) / A) + 5T + 27B + 18E$$

No qual:

F = Número de pares de letras adjacentes na mesma ordem compartilhada por pares de palavras.

V = Número de pares de letras adjacentes na ordem inversa compartilhada por pares de palavras.

C = Número de letras simples compartilhadas por pares de palavras.

A = Número médio de letras nas duas palavras.

T = Proporção de número de letras da palavra mais curta para o número mais longo.

B = 1 se a primeira letra nas duas palavras for a mesma, de outra forma B = 0

$E = 1$  se a última letra nas duas palavras é a mesma, de outra forma,  $E = 0$

Por exemplo: LETTURA/LEITURA

$$SO = 10 * ([50(4) + 30(0) + 10(6) / 7] + 5(1) + 27(1) + 18(1)) / 10 * ([50(6) + 30(0) + 10(7) / 7] + 5(1) + 27(1) + 18(1)) = 871,43 / 1028,6 = 0,85$$

Cada condição foi criada seguindo a seguinte regra: a condição cognata italiano-português não podia ser cognata com o inglês; assim como a condição cognata italiano-inglês não podia ser cognata com o português.

Após essas seleções, houve a contagem de letras e sílabas das palavras em italiano, pois tamanho de palavra é uma variável muito importante nos estudos de decisão lexical - palavras mais longas levam mais tempo para serem reconhecidas do que palavras mais curtas (BRYSSBAERT; NEW, 2009, p. 977-990). Além disso, foram calculadas as médias de frequência, número de letras, número de sílabas e SO para que elas fossem equivalentes estatisticamente entre as condições da tarefa. A Tabela 1 apresenta tais médias. Em contrapartida, as médias das palavras não-cognatas devem ser diferentes para que exista uma polarização entre cognatas e não-cognatas. Desta forma, poderemos observar a facilitação cognata entre as línguas, ou seja, se a palavra cognata facilitou o reconhecimento na tarefa de decisão lexical, ou se ela atrapalhou o participante.

**Tabela 1** - Médias de frequência, número de letras, número de sílabas e similaridade ortográfica por condição.

Condição Cognata	Frequência	Número de letras	Número de sílabas	SO
Português-Ingês-Italiano	335,82	6,41	2,77	0,78
Italiano-Ingês	266,5	6,32	2,55	0,73
Italiano-Português	318,79	6,55	2,73	0,77
Condição Não-Cognata Português-Ingês-Italiano	279,48	6,5	2,65	0,21

Fonte: do autor.

A partir das classificações dos 1.129 substantivos italianos com o português e com o inglês, a condição cognata italiano-inglês foi a que teve menor número de itens dentro dos critérios entre os pares de palavras, ficando com 25 itens. Com isso, para futuras análises estatísticas da tarefa de decisão lexical, era necessário que todas as condições da tarefa tivessem o mesmo número de itens. Assim, as outras condições cognatas (português-inglês-italiano e italiano-português) ficaram com a mesma quantidade de itens da condição italiano-inglês. Para que o controle metodológico fosse ainda maior, foi necessário retirar os itens com  $SO = 0,6$ , pois estas seriam o limite entre a classificação cognata e não cognata. Após estas etapas, foi realizada a classificação de ambiguidade lexical do italiano. Foram retiradas as palavras homônimas em italiano *aiuto* (que pode significar o substantivo ajuda ou o verbo *aiutare* (ajudar) conjugado na primeira pessoa singular, por exemplo, *Io aiuto mia madre* - Eu ajudo minha mãe). Devido a este controle, em cada condição, restaram 22 itens com as médias estatisticamente equivalentes, ou seja, 22 palavras cognatas português-inglês-italiano, 22 cognatas italiano-português, 22 cognatas italiano-inglês, e 66 não-cognatas português-inglês-italiano. As Tabelas 2-5 apresentam exemplos de tais conjuntos de palavras, suas frequências, número de letras, número de sílabas e similaridade ortográfica entre os pares de palavras.

**Tabela 2** - Conjunto de palavras cognatas italiano-português.

Cognatas italiano-português	Frequência	Nº de letras	Nº de sílabas	SO italiano-português
<i>francese/francês</i>	103	8	3	0,82
<i>tessuto/tecido</i>	105	7	3	0,61

Fonte: do autor.

**Tabela 3 - Conjunto de palavras cognatas italiano-inglês.**

Cognatas italiano- inglês	Frequência	Nº de letras	Nº de sílabas	SO italiano- inglês
<i>pepe/pepper</i>	102	4	2	0,7
<i>traccia/trace</i>	105	7	2	0,6

Fonte: do autor.

**Tabela 4- Conjunto de palavras cognatas italiano-português-inglês.**

Cognatas italiano- português- inglês	Frequência	Nº de letras	Nº de sílabas	SO italiano- português	SO italiano- inglês	SO português- inglês	SO geral
<i>curva/curva/curve</i>	101	5	2	1,00	0,70	0,70	0,80
<i>clinica/clínica/clinic</i>	105	7	3	1,00	0,79	0,79	0,86

Fonte: do autor.

**Tabela 5 - Conjunto de palavras não cognatas italiano-português-inglês.**

Não Cognatas italiano-português- inglês	Frequência	Nº de letras	Nº de sílabas	SO italiano- português	SO italiano- inglês	SO português- inglês	SO geral
<i>erba/grama/grass</i>	102	4	2	0,24	0,08	0,55	0,29
<i>giacca/jaqueta/ jacket</i>	102	6	2	0,25	0,21	0,52	0,33

Fonte: do autor.

A partir do número total de 132 itens das três condições cognatas e não cognatas, criou-se o mesmo número de pseudopalavras (exemplos na Tabela 6) para que a tarefa não se torne mecânica, dado que o participante lerá somente palavras. É necessário que a tarefa apresente pseudopalavras para que o participante de fato tome uma decisão quanto ao status do estímulo que é apresentado na tarefa. Ou seja, para decidir se um estímulo forma uma palavra que existe em italiano, ele precisa também ser exposto a estímulos que não formam

palavras existentes no italiano. Para isso, pseudopalavras foram criadas a partir da troca de uma letra de palavras existentes em italiano, formando palavras não existentes, porém com ortografia e fonologia possíveis. As palavras utilizadas para criar as pseudopalavras não foram utilizadas na tarefa.

Nas Tabelas 2-6 apresentamos alguns dos itens selecionados para a tarefa de decisão lexical com trilíngues. No anexo A encontram-se as listas completas de estímulos utilizados na tarefa.

**Tabela 6-** Conjunto de pseudopalavras.

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>appello/appella</i>	213	7	3	0,62	0,68
<i>caffè/ceffe</i>	164	5	2	0,75	0,78

Fonte: do autor.

### Conclusão

O presente estudo teve como objetivo desenvolver um método que possibilite a criação de um conjunto de palavras que pode ser usado em uma tarefa de decisão lexical. Este conjunto precisava envolver as línguas português, inglês e italiano. Portanto, 1.129 palavras italianas foram classificadas quanto à frequência de uso (retirada do *corpus*), médias de número de sílabas e número de letras, em italiano; e status cognatos e similaridade ortográfica, nestas três línguas. Mas para manter o controle estatístico as palavras foram classificadas por condições. São elas: condição cognata nas três línguas, cognata português-italiano, cognata italiano-inglês, não cognata nas três línguas e pseudopalavras. Após esta classificação obtivemos uma lista com 264 palavras e pseudopalavras para serem utilizadas no experimento.

Ambas as listas podem ser utilizadas em pesquisas com bilíngues e multilíngues que envolvam português-inglês-italiano. A lista com 264 palavras e pseudopalavras pode ser

utilizada em tarefas de reconhecimento de palavras/decisão lexical, pois ela está estatisticamente controlada. Já a lista original e completa com 1.129 palavras italianas, classificadas conforme descrito anteriormente, pode ser usada em futuras pesquisas sobre o processamento da linguagem em italiano, especialmente em estudos que visam a um certo nível de controle metodológico. Como dito anteriormente, os estudos, que serviram como referência, como os de Prior et al (2007, p. 1029-1038), Tokowicz et al (2007, p. 727-779), Fraga et al (2017, p. 1-10) e Siqueira et al (2010, p. 127-145), produziram listas a partir dos participantes da pesquisa e o uso de dicionários. Neste estudo, porém, a lista foi produzida pela retirada de substantivos de um *corpus* do italiano escrito, e, a partir desta lista, as palavras foram classificadas como mostrado acima. Este estudo busca contribuir com os estudos que investigam como se dá o acesso lexical em bilíngues e multilíngues, visto que ainda é um ramo da psicolinguística pouco investigado, como mostrado na pesquisa por artigos no Portal de Periódicos da Capes. Além disso, a disponibilização destas listas tem como objetivo auxiliar os pesquisadores que realizam tarefas de processamento de palavras com bilíngues e multilíngues português-inglês-italiano.

## REFERÊNCIAS

BARCELOS, L. *O acesso lexical em trilíngues brasileiros falantes de português, inglês e francês*. 2016. Dissertação (Mestrado em Linguística Aplicada) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul.

BERTINETTO, P. M.; BURANI, C.; LAUDANNA, A.; MARCONI, L.; RATTI, D.; ROLANDO, C.; THORNTON, A. M. *Corpus e lessico di frequenza dell'italiano scritto (CoLFIS)*. Scuola Normale Superiore di Pisa, 2005.

BRYBAERT, M.; NEW, B. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, 2009, 41(4) 977-990.

CARAMAZZA, A.; BRONES, I. Lexical access in bilinguals. *Bulletin of the Psychonomic Society*, v.13, n. 4, p. 212-214, 1979.

CHRISTENSEN, L. B.; JOHNSON, B.; TURNER, L. A. *Research methods, design, and analysis*.2011.

DASTOLI, C. A., et al. *Parola Chiave: dizionario di italiano per brasiliani*. São Paulo: M. Fontes, 2007.

DE GROOT, A. M. B. Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 18, n. 5, p. 1001-1018, 1992a.

DE GROOT, A. M. B. Bilingual lexical representation: A closer look at conceptual representations. *Advances in psychology*, v. 94, p. 389-412, 1992b.

DE GROOT, A. M. B.; DELMAAR, P.; LUPKER, S. J. The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology Section A*, v. 53, n. 2, p. 397-428, 2000.

FIELD, J. *Psycholinguistics: A resource book for students*. Psychology Press, 2003.

FRAGA, I.; PADRON, I.; PEREA, M.; COMESAÑA, M. I saw this somewhere else: The Spanish Ambiguous Words (SAW) database. *Lingua*, v. 185, p. 1-10, 2017.

HOWES, D. H.; SOLOMON, R. L. Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, v. 41, p. 401-410, 1951.

LEMHÖFER, K.; DIJKSTRA, T. Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, v. 32, n. 4, p. 533-550, 2004.

MILLER, N. A.; KROLL, J. F. Stroop effects in bilingual translation. *Memory&Cognition*, v. 30, p. 614-628, 2002.

PICKBRENNER, M. B. *O acesso lexical em multilíngues inglês (l2) e alemão (l3) no reconhecimento de palavras tipologicamente similares*. 2014. Tese (Doutorado em Linguística Aplicada) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul.

PRIOR, A.; MACWHINNEY, B.; KROLL, J. F. Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, 2007, 39(4) 1029-1038.

SIQUEIRA, M.; GIL, M.; MELO, T. Contribuições para a elaboração de testes psicolinguísticos: construção de uma lista de sentenças. *Gragoatá*, v. 15, n. 29, p. 127-145, 2010.

TOKOWICZ, N.; KROLL, J. F. Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, v. 22, n. 5, p. 727-779, 2007.

VAN ASSCHE, E.; DUYCK, W.; HARTSUIKER, R. J. Bilingual word recognition in a sentence context. *Frontiers in psychology*, v. 3, p. 174, 2012.

VAN HELL, J. G.; DIJKSTRA, T. Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic bulletin & review*, v. 9, n. 4, p. 780-789, 2002.

VAN ORDEN, G. C. A ROWS is a ROSE: Spelling, sound, and reading. *Memory & cognition*, v. 15, n. 3, p. 181-198, 1987.

WORDSMYTH. *The premier Educational Dictionary-Thesaurus*, 2018. Disponível em <<https://www.wordsmyth.net/>>. Acesso em 14/05/2018.



## ANEXO A

Tabela 2 - Conjunto de palavras cognatas italiano-português.

Cognatas italiano-português	Frequência	Nº de letras	Nº de sílabas	SO italiano-português
<i>francese/francês</i>	103	8	3	0,82
<i>tessuto/tecido</i>	105	7	3	0,61
<i>sostegno/sustento</i>	122	8	3	0,63
<i>calore/calor</i>	127	6	3	0,77
<i>mondiale/mundial</i>	138	8	3	0,64
<i>infanzia/infância</i>	147	8	3	0,87
<i>piatto/prato</i>	148	6	2	0,75
<i>lettura/leitura</i>	163	7	3	0,85
<i>pesce/peixe</i>	181	5	2	0,66
<i>esercito/exército</i>	192	8	4	0,81
<i>filo/fio</i>	210	4	2	0,73
<i>velocità/velocidade</i>	249	8	4	0,60
<i>pericolo/perigo</i>	250	8	4	0,76
<i>bilancio/balanço</i>	269	8	3	0,73
<i>anima/alma</i>	271	5	3	0,69
<i>peso/peso</i>	318	4	2	1,00
<i>denaro/dinheiro</i>	320	6	3	0,61
<i>sede/sede</i>	323	4	2	1,00
<i>vittoria/vitória</i>	347	8	3	0,96
<i>accordo/acordo</i>	419	7	3	0,91
<i>film/filme</i>	1215	4	1	0,73
<i>modo/modo</i>	1771	4	2	1,00

Fonte: do autor.

**Tabela 3** - Conjunto de palavras cognatas italiano-inglês.

Cognatas italiano- inglês	Frequência	Nº de letras	Nº de sílabas	SO italiano- inglês
<i>pepe/pepper</i>	102	4	2	0,7
<i>traccia/trace</i>	105	7	2	0,6
<i>scene/scenes</i>	127	5	2	0,8
<i>autunno/autumn</i>	139	7	3	0,6
<i>sanità/sanity</i>	141	6	3	0,7
<i>visione/vision</i>	148	7	3	0,8
<i>stazione/station</i>	167	8	3	0,6
<i>reazione/reaction</i>	180	8	4	0,6
<i>funzione/function</i>	185	8	3	0,6
<i>villa/villa</i>	194	5	2	1
<i>studenti/students</i>	210	8	3	0,7
<i>scienza/science</i>	215	7	2	0,7
<i>unità/unit</i>	244	5	3	0,8
<i>storie/stories</i>	246	6	2	0,8
<i>tono/tone</i>	251	4	2	0,7
<i>campione/champion</i>	266	8	3	0,7
<i>passione/passion</i>	277	8	3	0,8
<i>sport/sport</i>	319	5	1	1
<i>stile/style</i>	322	5	2	0,8
<i>ottobre/october</i>	348	7	3	0,6
<i>pace/peace</i>	401	4	2	0,8
<i>persone/people</i>	1276	7	3	0,6

Fonte: do autor.

**Tabela 4 - Conjunto de palavras cognatas italiano-português-inglês.**

Cognatas italiano-português- inglês	Frequência	Nº de letras	Nº de sílabas	SO italiano-português	SO italiano- inglês	SO português- inglês	SO geral
<i>curva/curva/curve</i>	101	5	2	1,00	0,70	0,70	0,80
<i>clinica/clínica/clinic</i>	105	7	3	1,00	0,79	0,79	0,86
<i>commedia/comédia/ comedy</i>	127	8	3	0,96	0,65	0,69	0,77
<i>bomba/bamba/bomb</i>	128	5	2	1,00	0,75	0,75	0,83
<i>evento/evento/event</i>	139	6	3	1,00	0,78	0,78	0,85
<i>merito/mérito/merit</i>	141	6	3	1,00	0,77	0,77	0,85
<i>disco/disco/disc</i>	148	5	2	1,00	0,75	0,75	0,83
<i>metodo/método/ method</i>	169	6	3	1,00	0,69	0,69	0,79
<i>prodotto/produto/ product</i>	185	8	3	0,90	0,60	0,67	0,73
<i>compagno/ companheiro/ companion</i>	186	8	3	0,73	0,64	0,61	0,66
<i>simbolo/símbolo/ symbol</i>	194	7	3	1,00	0,66	0,66	0,77
<i>colore/cor/color</i>	212	6	3	0,63	0,78	0,86	0,76
<i>gesto/gesto/gesture</i>	226	5	2	1,00	0,61	0,61	0,74
<i>errore/erro/error</i>	237	6	3	0,82	0,83	0,73	0,79
<i>novembre/novembro/ november</i>	269	8	3	0,75	0,74	0,69	0,73
<i>termine/término/term</i>	277	7	3	0,74	0,66	0,63	0,68
<i>proposta/porposta/ proposal</i>	318	8	3	1,00	0,72	0,72	0,81
<i>dollari/dólares/ dollars</i>	321	7	3	0,66	0,75	0,85	0,75
<i>nord/norte/north</i>	334	4	1	0,60	0,60	0,70	0,63
<i>presenza/presença/ presence</i>	435	8	3	0,88	0,70	0,70	0,76
<i>caso/caso/case</i>	1254	4	2	1,00	0,66	0,66	0,77
<i>governo/governo/ government</i>	1492	7	3	1,00	0,64	0,65	0,76

Fonte: do autor.

**Tabela 5 - Conjunto de palavras não cognatas italiano-português-inglês.**

Não Cognatas italiano-português-inglês	Frequência	Nº de letras	Nº de sílabas	SO italiano-português	SO italiano-inglês	SO português-inglês	SO geral
<i>erba/grama/grass</i>	102	4	2	0,24	0,08	0,55	0,29
<i>giacca/jaqueta/jacket</i>	102	6	2	0,25	0,21	0,52	0,33
<i>rapina/roubo/theft</i>	102	6	3	0,33	0,04	0,05	0,14
<i>fianco/lado/side</i>	105	6	2	0,26	0,06	0,08	0,13
<i>orologio/relogio/clock</i>	105	8	4	0,59	0,19	0,15	0,31
<i>scudetto/escudete/escutcheon</i>	105	8	3	0,43	0,21	0,52	0,38
<i>giallo/amarelo/yellow</i>	120	6	2	0,31	0,3	0,22	0,28
<i>tedeschi/alemães/german</i>	120	8	3	0,14	0,05	0,16	0,12
<i>uova/ovos/eggs</i>	121	4	2	0,27	0,05	0,24	0,19
<i>polvere/poeira/dust</i>	125	7	3	0,44	0,03	0,03	0,17
<i>decina/dez/ten</i>	126	6	3	0,48	0,07	0,09	0,21
<i>raccolta/coleção/collection</i>	137	8	3	0,28	0,19	0,59	0,36
<i>chiusura/encerramento/closing</i>	138	8	3	0,11	0,34	0,06	0,17
<i>rumore/ruído/noise</i>	141	6	3	0,46	0,26	0,07	0,26
<i>alba/alvorada/dawn</i>	142	4	2	0,55	0,08	0,2	0,28
<i>mestiere/ofício/occupation</i>	145	8	3	0,05	0,11	0,37	0,18
<i>foglie/folhas/leaves</i>	147	6	2	0,45	0,08	0,28	0,27
<i>negozio/loja/shop</i>	149	7	3	0,05	0,05	0,08	0,06
<i>consigli/conselhos/advice</i>	150	8	3	0,53	0,07	0,06	0,22
<i>lacrime/lágrimas/tears</i>	162	7	3	0,56	0,09	0,25	0,3
<i>cibo/comida/food</i>	167	4	2	0,36	0,07	0,07	0,17
<i>stelle/estrelas/stars</i>	172	6	2	0,28	0,44	0,35	0,36
<i>scoperta/descoberta/discovery</i>	179	8	3	0,55	0,27	0,51	0,44
<i>finestra/janela/window</i>	181	8	3	0,33	0,14	0,07	0,18

Tabela 5 - continuação

Não Cognatas italiano-português- inglês	Frequência	Nº de letras	Nº de sílabas	SO italiano- português	SO italiano- inglês	SO português- inglês	SO geral
<i>attesa/esperança/ hope</i>	182	6	3	0,32	0,06	0,12	0,17
<i>vettura/carro/car</i>	182	7	3	0,11	0,13	0,59	0,28
<i>spesa/gasto/expense</i>	189	5	2	0,15	0,21	0,05	0,14
<i>tensione/tensão/ voltage</i>	200	8	3	0,57	0,26	0,1	0,31
<i>gioia/alegria/joy</i>	204	5	2	0,34	0,06	0,02	0,14
<i>riunione/reunião/ meeting</i>	204	8	3	0,51	0,12	0,13	0,25
<i>indagine/pesquisa/ search</i>	209	8	4	0,08	0,07	0,12	0,09
<i>nascita/nascimento/ birth</i>	209	7	3	0,58	0,07	0,05	0,23
<i>sfida/desafio/ challenge</i>	209	5	2	0,2	0,04	0,06	0,1
<i>maestro/professor/ teacher</i>	215	7	3	0,22	0,15	0,21	0,19
<i>giovedì/quinta-feira/ thursday</i>	217	7	3	0,06	0,06	0,08	0,06
<i>droga/especiaria/ spice</i>	228	5	2	0,22	0,05	0,23	0,17
<i>giardino/jardim/ garden</i>	238	8	3	0,3	0,51	0,26	0,36
<i>giornale/jornal/ newspaper</i>	240	8	3	0,39	0,08	0,07	0,18
<i>sorella/irmã/sister</i>	241	7	3	0,25	0,36	0,08	0,23
<i>edizione/edição/ edition</i>	225	8	4	0,5	0,59	0,52	0,54
<i>legno/madeira/wood</i>	248	5	2	0,05	0,06	0,05	0,05
<i>gambe/pernas/legs</i>	251	5	2	0,08	0,09	0,25	0,14
<i>malattia/doença/ disease</i>	251	8	4	0,23	0,07	0,35	0,22
<i>venerdì/sexta-feira/ friday</i>	258	7	3	0,05	0,13	0,06	0,08
<i>viso/rostro/face</i>	269	4	2	0,33	0,05	0,04	0,14
<i>lunedì/segunda-feira/ monday</i>	271	6	3	0,18	0,08	0,24	0,17

Tabela 5 - continuação

Não Cognatas italiano-português-ínglês	Frequência	Nº de letras	Nº de sílabas	SO italiano-português	SO italiano-ínglês	SO português-ínglês	SO geral
<i>tavolo/mesa/table</i>	271	6	3	0,06	0,46	0,09	0,2
<i>bambina/menina/girl</i>	281	7	3	0,43	0,05	0,05	0,18
<i>dubbio/dúvida/doubt</i>	281	6	2	0,45	0,48	0,34	0,42
<i>dolore/dor/pain</i>	296	6	3	0,52	0,03	0,04	0,2
<i>giudizio/juízo/judgment</i>	300	8	3	0,44	0,15	0,41	0,33
<i>sviluppo/desenvolvimento/development</i>	303	8	3	0,27	0,1	0,51	0,29
<i>azienda/empresa/company</i>	314	7	3	0,25	0,08	0,16	0,16
<i>serata/noite/night</i>	316	6	3	0,08	0,06	0,38	0,17
<i>occhio/olho/eye</i>	322	6	2	0,56	0,02	0,04	0,21
<i>sguardo/olhada/look</i>	333	7	2	0,08	0,05	0,12	0,08
<i>fratello/irmão/brother</i>	342	8	3	0,26	0,09	0,07	0,14
<i>stanza/cômodo/convenient</i>	348	6	2	0,04	0,05	0,32	0,14
<i>palazzo/prédio/building</i>	350	7	3	0,52	0,06	0,14	0,24
<i>speranza/esperança/hope</i>	351	8	3	0,58	0,15	0,12	0,28
<i>spalle/costas/back</i>	396	6	2	0,08	0,06	0,07	0,07
<i>luogo/lugar/place</i>	507	5	2	0,48	0,07	0,09	0,21
<i>gioco/jogo/game</i>	706	5	2	0,26	0,34	0,07	0,22
<i>moglie/esposa/wife</i>	897	6	2	0,08	0,26	0,05	0,13
<i>padre/pai/father</i>	1254	5	2	0,51	0,15	0,05	0,24
<i>giorno/dia/day</i>	2077	6	2	0,05	0,03	0,41	0,16

Fonte: do autor.

**Tabela 6 - Conjunto de pseudopalavras.**

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>appello/appella</i>	213	7	3	0,62	0,68
<i>caffè/ceffe</i>	164	5	2	0,75	0,78
<i>collega/colluga</i>	144	7	3	0,79	0,65
<i>offerta/offerra</i>	113	7	3	0,79	0,64
<i>passato/passabo</i>	617	7	3	0,86	0,61
<i>scandalo/scannalo</i>	174	8	3	0,61	0,68
<i>secondo/semondo</i>	102	7	3	0,76	0,61
<i>tecnici/teccini</i>	147	7	3	0,65	0,60
<i>valore/valove</i>	412	6	3	0,66	0,75
<i>agenzia/amenzia</i>	132	7	4	0,69	0,43
<i>alleanza/alleonza</i>	115	8	4	0,60	0,52
<i>articolo/armicolo</i>	252	8	4	0,61	0,50
<i>autorità/autoriga</i>	268	8	4	0,60	0,52
<i>banche/bancce</i>	173	6	2	0,63	0,55
<i>capitano/capituno</i>	195	8	4	0,82	0,51
<i>cronaca/cropaca</i>	105	7	3	0,81	0,40
<i>difesa/disesa</i>	418	6	3	0,70	0,43
<i>effetto/effutto</i>	333	7	3	0,68	0,53
<i>esame/elame</i>	214	5	3	0,78	0,50

Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>gruppo/grupto</i>	971	6	2	0,86	0,58
<i>macchina/maccrina</i>	516	8	3	0,74	0,56
<i>massimo/massiro</i>	124	7	3	0,62	0,39
<i>oggetto/oggotto</i>	258	7	3	0,65	0,34
<i>pazienza/pazienna</i>	118	8	3	0,71	0,55
<i>progetto/provetto</i>	456	8	3	0,82	0,50
<i>punto/panto</i>	1389	5	2	0,78	0,50
<i>rischio/riscio</i>	445	7	2	0,82	0,55
<i>rispetto/risputto</i>	599	8	3	0,68	0,42
<i>ritorno/ritarno</i>	353	7	3	0,69	0,44
<i>segreto/segruto</i>	156	7	3	0,76	0,44
<i>soldati/soldaci</i>	127	7	3	0,61	0,56
<i>sorpresa/sorpreda</i>	286	8	3	0,75	0,52
<i>sospetto/sospetro</i>	112	8	3	0,63	0,47
<i>testo/tusto</i>	245	5	2	0,66	0,36
<i>codice/codace</i>	204	6	3	0,53	0,78
<i>emozioni/emofioni</i>	143	8	4	0,45	0,62
<i>lettere/lettebe</i>	238	7	3	0,55	0,70
<i>messaggi/messaghi</i>	114	8	3	0,54	0,69
<i>muscoli/musceli</i>	102	7	3	0,56	0,64



Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>persona/permona</i>	633	7	3	0,59	0,61
<i>qualità/qualiga</i>	414	7	3	0,57	0,66
<i>realtà/realto</i>	475	6	3	0,50	0,60
<i>società/sociela</i>	1078	7	3	0,54	0,66
<i>specie/spemie</i>	473	6	2	0,52	0,60
<i>spirito/sparito</i>	286	7	3	0,55	0,60
<i>varietà/variega</i>	124	7	3	0,54	0,66
<i>versione/versionu</i>	251	8	3	0,58	0,79
<i>abitanti/alitante</i>	247	8	4	0,38	0,29
<i>abiti/abimi</i>	167	5	3	0,25	0,27
<i>accusa/accuta</i>	292	6	3	0,51	0,59
<i>affari/asfari</i>	288	6	2	0,34	0,15
<i>aprile/aprule</i>	348	6	3	0,37	0,57
<i>aree/iree</i>	158	4	3	0,20	0,20
<i>attività/attiviga</i>	653	8	4	0,59	0,59
<i>attore/attoce</i>	295	6	2	0,46	0,48
<i>auto/euto</i>	101	4	2	0,23	0,04
<i>azioni/agione</i>	204	6	3	0,37	0,51
<i>benzina/bennina</i>	105	7	3	0,42	0,24
<i>cause/ciuse</i>	130	5	2	0,41	0,56

Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>civiltà/civilte</i>	124	7	3	0,56	0,58
<i>colori/comori</i>	306	6	3	0,55	0,50
<i>colpi/calpi</i>	248	5	2	0,07	0,07
<i>compagni/compagri</i>	257	8	3	0,57	0,57
<i>compito/cospito</i>	279	7	3	0,12	0,06
<i>comunità/romunita</i>	356	8	4	0,30	0,38
<i>concerto/connerto</i>	203	8	3	0,04	0,04
<i>conferma/conterma</i>	167	8	3	0,52	0,54
<i>coraggio/loraggio</i>	252	8	3	0,29	0,22
<i>corsi/corli</i>	139	5	2	0,33	0,39
<i>costo/losto</i>	217	5	2	0,49	0,34
<i>danni/danvi</i>	191	5	2	0,58	0,38
<i>decenni/decenvi</i>	112	7	3	0,49	0,55
<i>decine/demine</i>	222	6	3	0,45	0,41
<i>elezione/alezione</i>	119	8	4	0,21	0,32
<i>esempio/esespio</i>	959	7	3	0,50	0,35
<i>faccia/maccia</i>	317	6	2	0,18	0,18
<i>facoltà/farolta</i>	109	7	3	0,38	0,51
<i>farmaci/farbaci</i>	105	7	3	0,12	0,15
<i>fatti/fitti</i>	320	5	2	0,36	0,36

Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>fattori/fatturi</i>	112	7	3	0,51	0,44
<i>fiducia/fidufia</i>	346	7	3	0,37	0,18
<i>fiori/fioru</i>	296	5	2	0,45	0,35
<i>forza/firza</i>	597	5	2	0,56	0,36
<i>frasi/frami</i>	119	5	2	0,49	0,14
<i>frutto/frunto</i>	139	6	2	0,57	0,57
<i>funzioni/funnioni</i>	113	8	3	0,49	0,59
<i>gara/gura</i>	368	4	2	0,03	0,02
<i>genitori/genirori</i>	576	8	4	0,04	0,15
<i>gestione/gestiole</i>	224	8	3	0,08	0,20
<i>giro/biro</i>	429	4	2	0,06	0,08
<i>giudici/giufici</i>	331	7	3	0,12	0,07
<i>giugno/giugno</i>	438	6	2	0,26	0,06
<i>grandi/grundi</i>	102	6	2	0,51	0,43
<i>guai/ruai</i>	165	4	2	0,05	0,04
<i>idee/udee</i>	300	4	3	0,17	0,20
<i>immagini/immogini</i>	314	8	4	0,42	0,42
<i>impianto/impiunto</i>	107	8	3	0,46	0,30
<i>intesa/intesu</i>	139	6	3	0,04	0,07
<i>linee/lonee</i>	160	5	3	0,34	0,48

Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>luglio/luggio</i>	425	6	2	0,33	0,14
<i>marzo/mirzo</i>	309	5	2	0,56	0,36
<i>metri/motri</i>	602	5	2	0,47	0,33
<i>miracolo/mirafolo</i>	139	8	4	0,44	0,56
<i>misure/misute</i>	197	6	3	0,34	0,42
<i>montagne/montarne</i>	111	8	3	0,59	0,54
<i>nomi/lomi</i>	296	4	2	0,20	0,06
<i>nozze/nozzu</i>	175	5	2	0,33	0,32
<i>olio/ulio</i>	287	4	2	0,29	0,32
<i>onore/onire</i>	200	5	3	0,21	0,24
<i>ordine/ormine</i>	599	6	3	0,48	0,46
<i>ospedale/ospefale</i>	409	8	4	0,29	0,29
<i>paese/paose</i>	1196	5	3	0,50	0,44
<i>pagine/pagune</i>	296	6	3	0,51	0,57
<i>paio/jaio</i>	358	4	2	0,07	0,23
<i>piani/piuni</i>	167	5	2	0,34	0,36
<i>piante/piaste</i>	174	6	2	0,44	0,43
<i>poteri/potezi</i>	108	6	3	0,39	0,45
<i>prezzi/prezza</i>	275	6	2	0,53	0,45
<i>prodotti/prodatti</i>	305	8	3	0,55	0,55

Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>profumo/profuco</i>	117	7	3	0,44	0,44
<i>prove/prote</i>	235	5	2	0,17	0,17
<i>racconto/racconno</i>	196	8	3	0,08	0,07
<i>rapporti/rapporpi</i>	445	8	3	0,35	0,36
<i>reato/reatu</i>	118	5	3	0,09	0,09
<i>reazioni/reañioni</i>	116	8	4	0,49	0,59
<i>regia/remia</i>	127	5	3	0,18	0,14
<i>ricordi/ricormi</i>	145	7	3	0,21	0,14
<i>rifuto/rimiuto</i>	111	7	3	0,34	0,34
<i>riforme/liforme</i>	131	7	3	0,30	0,33
<i>risorse/ricarse</i>	154	7	3	0,40	0,39
<i>rottura/rottira</i>	101	7	3	0,56	0,33
<i>scala/stala</i>	112	5	2	0,52	0,54
<i>scarpe/scarpo</i>	160	6	2	0,10	0,07
<i>scena/svena</i>	500	5	2	0,52	0,48
<i>scopo/scipo</i>	165	5	2	0,23	0,07
<i>segnale/segale</i>	159	7	3	0,44	0,44
<i>signi/sesni</i>	162	5	2	0,39	0,46
<i>senso/sanso</i>	827	5	2	0,52	0,48
<i>sesso/susso</i>	320	5	2	0,55	0,35

Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>settori/sattori</i>	126	7	3	0,47	0,51
<i>sistemi/sistumi</i>	147	7	3	0,59	0,40
<i>soccorso/soccarso</i>	144	8	3	0,08	0,08
<i>soggetti/soggetta</i>	105	8	3	0,39	0,35
<i>sostanza/sostanzu</i>	144	8	3	0,52	0,54
<i>spazi/spavi</i>	167	5	2	0,25	0,54
<i>stadio/stedio</i>	189	6	2	0,53	0,51
<i>stilista/stilosta</i>	103	8	3	0,57	0,55
<i>storia/stiria</i>	1414	6	2	0,54	0,18
<i>strage/strame</i>	112	6	2	0,29	0,29
<i>svolta/svosta</i>	135	6	2	0,08	0,06
<i>tasse/tassu</i>	104	5	2	0,52	0,48
<i>temi/zemi</i>	150	4	2	0,20	0,21
<i>tempi/tumpi</i>	799	5	2	0,45	0,38
<i>tifosi/tafosi</i>	165	6	3	0,17	0,16
<i>umanità/umanito</i>	111	7	4	0,30	0,44
<i>unione/unaone</i>	113	6	3	0,57	0,57
<i>vetture/vetrure</i>	106	7	3	0,07	0,05
<i>viale/viulo</i>	109	5	3	0,07	0,08
<i>vicenda/vimenda</i>	354	7	3	0,15	0,15

Tabela 6 - continuação

Forma - Pseudopalavra	Frequência	Nº de letras	Nº de sílabas	SO pseudopalavras com português	SO pseudopalavras com inglês
<i>voci/loci</i>	258	4	2	0,06	0,15
<i>voti/toti</i>	238	4	2	0,39	0,26
<i>zona/vona</i>	573	4	2	0,26	0,26

Fonte: do autor.

### Abstract

#### **Methodological control: creation of a *corpus* for studies on the lexical processing of bilingual and multilingual individuals**

*Studies about multilingual lexical access use lexical decision tasks comparing the processing of cognate and non-cognate words to investigate whether there is co-activation of languages when multilinguals process words in a single language. To understand how this process works for Portuguese-English-Italian trilinguals using a lexical decision task in Italian, a corpus was elaborated according to the following steps: selection of 1129 Italian nouns varying from 4-8 letters in length, along with their frequencies; translation of the Italian words to English and Portuguese; classification of these words as cognates in the three languages, Italian-Portuguese cognates, Italian-English cognates and non-cognates in the three languages according to an index of orthographic similarity; and classification of words according to lexical ambiguity in Italian. Homonyms were removed, leaving only polysemous and non-ambiguous words in the list. Next, pseudowords were created from existing Italian words. After that step, the words were separated by test condition and their frequencies, length, number of syllables and orthographic similarity were compared statistically to verify their similarity. Last, the list of words to be used in the lexical decision task in Italian is composed of three test conditions of 22 cognate words each, a list of 66 non-cognate words, and a set of 132 pseudowords. This task will allow us to investigate the co-activation of English and Portuguese during lexical access in Italian.*

**Keywords:** *lexical access, multilingualism, methodological control.*