

DISPONIBILIDADE LEXICAL EM ESTUDANTES DE PORTUGUÊS COMO LÍNGUA ESTRANGEIRA: UM ESTUDO PIONEIRO

MIRTA FERNÁNDEZ DOS SANTOS*
Faculdade de Letras da Universidade do Porto

CLÁUDIA SILVA SOARES**
Faculdade de Letras da Universidade do Porto

RESUMO: No enquadramento das investigações científicas sobre disponibilidade lexical em línguas estrangeiras, neste artigo são apresentados os principais resultados quantitativos e qualitativos de um estudo de investigação desbravador, dado ser o primeiro realizado com estudantes de português europeu como língua estrangeira (PLE), que faz parte de uma indagação de maior alcance e se baseia na inquirição da disponibilidade lexical de 30 informantes de PLE, posicionados no nível A2, relativamente aos centros de interesse “partes do corpo”, “escola: móveis e materiais”, “roupa”, “meios de transporte”, “alimentos e bebidas”, “profissões e ofícios”. Não existindo uma tradição metodológica própria sobre léxico disponível no âmbito da linguística lusófona, foram seguidas as pautas estabelecidas no Proyecto Panhispánico de Disponibilidad Léxica com o intuito de efetuar uma análise comparativa entre este trabalho e o estudo *Disponibilidad léxica en alumnos de español como lengua extranjera* (Samper Hernández, 2002), visto ambas as investigações partilharem os mesmos objetivos, tipologia de informantes e metodologia. Foi ainda contrastado o vocabulário produzido pelos informantes com as listagens de referência de léxico publicadas no *Referencial Camões PLE* (2017) para o nível A2.

PALAVRAS-CHAVE: disponibilidade lexical; português como língua estrangeira; estudo pioneiro; Proyecto Panhispánico de Disponibilidad Léxica

LEXICAL AVAILABILITY IN PORTUGUESE AS A FOREIGN LANGUAGE STUDENTS: A PIONEERING STUDY

Abstract: In the framework of scientific research on lexical availability in foreign languages, in this article we present the main quantitative and qualitative results

* Para correspondência dirigir-se a: Mirta Fernández dos Santos (mfernandez@letras.up.pt).

** Para correspondência dirigir-se a: Cláudia Silva Soares (soaresclaudia511@gmail.com)

of an exploratory research study, the first of which was carried out with students of European Portuguese as a Foreign Language (PFL). The study, which is part of a broader-scope investigation, was based on the inquiry into the lexical availability of 30 PFL informants at level A2 in the prompts “body parts”, “school: furniture and materials”, “means of transport”, “food and drink”, “clothing” and “professions and trades”. There being no methodological tradition on available vocabulary in the field of Lusophone linguistics, the guidelines established in the Pan-Hispanic Project on Lexical Availability were followed, which is why we intend to carry out a comparative analysis between our work and the study of Lexical availability in Spanish as a Foreign Language students (Samper Hernández, 2002), since both investigations share the same objectives, types of informants and methodology. The vocabulary produced by the informants was further contrasted with the lexical reference lists published in the Camões PFL Reference Document (2017) for level A2.

Keywords: lexical availability; Portuguese as a Foreign Language; pioneering study; Pan-Hispanic Project on Lexical Availability

1. INTRODUÇÃO

Este trabalho expõe os resultados fundamentais de uma investigação, subordinada à temática da disponibilidade lexical, que foi realizada com um grupo de aprendentes de português como língua estrangeira (PLE) de nível A2 que aprendem o idioma de Luís de Camões em contexto universitário, em concreto na Faculdade de Letras da Universidade do Porto.

Trata-se de um estudo pioneiro, uma vez que, apesar do auge alcançado pelos estudos de disponibilidade lexical nos dias de hoje, especialmente no terreno do ensino-aprendizagem de línguas estrangeiras, ainda não foi feita nenhuma indagação académica que tenha o propósito de apurar o léxico disponível de estudantes de PLE.

Pese ao exposto e de acordo com dados do Instituto Camões (2022), o português é a língua materna de aproximadamente 280 milhões de pessoas, o que a converte na 5.^a língua mais falada do mundo, na 3.^a mais falada no hemisfério ocidental e na mais falada no hemisfério sul da Terra. É ainda idioma oficial da União Europeia, do Mercosul, da União de Nações Sul-Americanas, da Organização dos Estados Americanos, da União Africana e dos Países Lusófonos. Devido ao facto de se tratar de uma língua tão difundida num mundo globalizado, há cada vez mais pessoas interessadas em aprender português como língua estrangeira, seja por motivos profissionais ou académicos, seja por outras razões, o que explica o aumento da procura que se tem vindo a notar ao longo das últimas duas décadas, sobretudo em países como Brasil e Portugal.

Por conseguinte, o crescimento gradual, mas imparável, do ensino de PLE no mundo justifica, do nosso ponto de vista, a realização da investigação cujos resultados são expostos neste artigo, com vista à obtenção de informação proveitosa que contribua para a melhoria dos processos de ensino-aprendizagem do vocabulário no âmbito de PLE, à semelhança do que já acontece com outras línguas estrangeiras de destacado protagonismo no planeta.

Para concluir este marco preliminar, resta referir que o nosso artigo se estrutura em seis partes, para além da introdução, a saber: considerações teóricas, considerações

metodológicas, apresentação e análise comparativa dos resultados¹, conclusões e referências bibliográficas.

2. CONSIDERAÇÕES TEÓRICAS

Como é do conhecimento da comunidade científica, os estudos de disponibilidade lexical - integrados no âmbito da léxico-estatística - constituem uma área de investigação recorrente e profícua no seio da investigação linguística contemporânea. O seu elevado grau de interesse científico a nível internacional resulta do “valor pedagógico-didático que su abordaje entraña para la enseñanza de la lengua” (Pacheco Carpio, Cabrera Albert & González López, 2017: 239), sem subestimar o seu potencial para uma melhor compreensão da comunicação humana de um ponto de vista cultural e psicossocial.

O amplo escopo bibliográfico produzido em torno desta temática um pouco por todo o mundo evidencia a natureza multidimensional e complexa do objeto de estudo que se tem disseminado em trabalhos de fim de licenciatura (Paolini, 2017; Trujillano Peña, 2015), dissertações de mestrado (López Rivero, 2008; Samper Hernández, 2002); teses de doutoramento (Delgado Fernández, 2024; Zhou, 2022); livros (Paredes García, 2022; Herranz Llácer, 2020); capítulos de livro (Gómez Molina, 2021; Samper Padilla, 2021); artigos de revistas científicas (Palapanidi & Mavrou, 2024; Santos Díaz, Trigo Ibáñez & Romero Oliva, 2020) e comunicações em congressos (Hernández Muñoz & López García, 2023; Checa García, 2023), entre outras produções publicadas desde o virar do milénio.

Um olhar rápido sobre os trabalhos anteriormente citados, que são apenas uma pequena amostra, revela que a disponibilidade lexical, área de estudo surgida em França em meados do século XX (Gougenheim, Michéa, Rivenc & Sauvageot, 1956), é na atualidade um campo particularmente produtivo no contexto da linguística hispânica. A maior parte destes estudos, que cobrem ambas as margens do Atlântico, deriva do audacioso Proyecto Panhispánico de Disponibilidad Léxica, dirigido por Humberto López Morales, cuja finalidade é “conocer el léxico disponible de estudiantes cuya lengua materna es el español y que aún no han comenzado la etapa universitaria” (Samper Hernández, 2002: 7).

Com o passar do tempo, porém, este objetivo geral foi-se desdobrando em outros e a metodologia foi sendo adaptada, conseqüentemente, para dar resposta a novas interrogantes científicas. Nesse sentido, são cada vez mais frequentes as investigações sobre léxico disponível realizadas não com nativos, mas com informantes forâneos que aprendem uma determinada língua estrangeira.

No entanto, apesar de as investigações sobre disponibilidade lexical aplicada ao ensino de línguas estrangeiras terem vindo a crescer nos últimos anos, sobretudo no âmbito da aprendizagem do espanhol e do inglês (Akbarian & Farrokhi, 2020; Carcedo González, 2000; Jiménez Catalán & Dewaele, 2017; López Rivero, 2008;

¹ Esta análise comparativa tem por base o nosso trabalho e os resultados obtidos por Marta Samper Hernández (2002) nos centros de interesse objeto de averiguação: “partes do corpo”, “escola: móveis e materiais”, “roupa”, “meios de transporte”, “alimentos e bebidas” e “profissões e ofícios”.

Samper Hernández, 2002), e se terem publicado alguns trabalhos que assentam na análise dos dados obtidos tendo por base uma amostra composta por estudantes de espanhol como língua estrangeira, estudantes estes cuja língua materna era o português (Fernández dos Santos, 2014; Fernández Gómez, 2021; Gamazo Carretero, 2014; Mas Álvarez & Santos Palmou, 2017), a verdade é que ainda não foi realizado nenhum estudo sobre a disponibilidade lexical associada à aprendizagem do Português como Língua Estrangeira (PLE).

O caráter pioneiro da nossa investigação justifica, por isso, a sua pertinência no processo de equiparação científica das metodologias aplicadas ao ensino e aprendizagem do vocabulário de línguas estrangeiras modernas tão difundidas e procuradas na atualidade como o espanhol, o inglês, o francês² e, claro, o português.

Portanto, averiguar o léxico disponível dos alunos de PLE inquiridos norteia assim o nosso trabalho de pesquisa e a nossa pretensão de contribuir quer para a identificação de possíveis carências formativas e para a análise dos procedimentos didáticos em curso, quer para a eventual implementação de novos métodos que otimizem a aquisição de vocabulário por estes estudantes.

3. CONSIDERAÇÕES METODOLÓGICAS

3.1. Pautas gerais

Como acontece com a maioria dos trabalhos que assentam nas diretrizes estipuladas pelo Proyecto Panhispánico de Disponibilidad Léxica, para a exploração do léxico disponível da amostra de informantes selecionada no presente estudo aplicamos as pautas metodológicas mais habituais neste tipo de investigações. A referida metodologia consiste na exposição dos inquiridos a uma prova associativa de caráter psicolinguístico, a partir de núcleos temáticos concretos, também chamados de *centros de interesse*, sobre os quais os informantes devem escrever, em listas abertas e num tempo limite de dois minutos por cada um desses núcleos, os elementos léxicos que forem sendo ativados pelo seu léxico mental.

De acordo com autores como Hernández & Borrego (2004: 47), este tipo de procedimento estandardizado favorece não apenas a comparação dos resultados de diferentes estudos, mas também o desenvolvimento de uma teoria firme e coerente de aplicação geral no terreno da investigação linguística sobre o léxico.

A mencionada prova associativa apresentada aos informantes constava de duas partes: a primeira era um questionário fechado que requeria informações de caráter sociocultural, necessárias para a posterior estratificação da amostra, e a segunda, por sua vez, era composta por três folhas em branco, cada uma segmentada em duas colunas, colunas estas precedidas do nome de cada um dos centros de interesse que

² Se bem que os estudos de disponibilidade lexical começaram em França, atualmente não existem demasiadas investigações sobre esta temática ligadas à aprendizagem do francês como língua estrangeira (FLE). Destacamos, porém, a de Díez Abadie (2022), que analisa a evolução da disponibilidade lexical de um grupo de estudantes espanhóis de educação secundária que são aprendentes de FLE, e a de Santos Díaz, intitulada *El léxico bilingüe del futuro profesorado. Análisis y pautas para estudios de disponibilidad léxica* (Santos Díaz, 2020).

iriam operar como estímulo para a evocação de unidades lexicais e que, no contexto do presente estudo, como já foi referido, eram os seguintes: “partes do corpo”, “escola: móveis e materiais”, “roupa”, “meios de transporte”, “alimentos e bebidas” e “profissões e ofícios”.

De seguida, procedemos à transcrição e lematização da informação obtida nos inquéritos, de acordo com os critérios de edição³ previamente estabelecidos, para as quais foi utilizado um documento de Excel. Posteriormente, realizamos o processamento informático dos dados recolhidos através do *upload* do ficheiro de Excel para o programa informático *LexPro*, uma ferramenta para a análise de dados lexicais e construção de redes complexas que foi recentemente desenvolvida pela Universidade de Salamanca no âmbito do projeto de investigação *Dispogram*. Após o cálculo efetuado pelo *LexPro*, foram extraídos os índices quantitativos e qualitativos mais relevantes para a nossa investigação, dados que seguidamente interpretámos e comparámos primeiro entre si e, depois, com os obtidos por Samper Hernández (2002).

3.2. A amostra

À luz da metodologia adotada por Carcedo González (2000) e Samper Hernández (2002) cujos trabalhos assentam numa amostra reduzida relativamente à que suporta o Proyecto Panhispánico de Disponibilidad Léxica também nos pareceu mais adequado alicerçar a nossa investigação numa amostra de pequena dimensão - 30 estudantes de PLE da Faculdade de Letras da Universidade do Porto posicionados no nível A2 de proficiência - que nos permitisse, face ao carácter pioneiro da mesma e aos nossos objetivos, fazer um estudo rigoroso e promotor de uma base estatística sólida para realizar a posterior seleção objetiva das palavras a ensinar aos alunos.

No que concerne às variáveis tidas em conta, e no seguimento do também defendido pelos autores mencionados no parágrafo anterior, consideramos que seria mais pertinente privilegiar o sexo dos informantes, a sua língua materna⁴, idade, conhecimento de outras línguas estrangeiras e tempo de estudo de português como língua estrangeira, em detrimento de outras variáveis, como o nível sociocultural, o carácter público ou privado do estabelecimento de ensino que frequentam ou a sua procedência geográfica.

Na seguinte tabela é apresentada a estratificação da amostra por variáveis e número de estudantes:

³ Seguindo o procedimento de outros estudos semelhantes ao nosso (Carcedo González, 2000; Samper Hernández, 2002), os principais critérios de edição aplicados foram: a eliminação de palavras repetidas, a eliminação de determinantes artigos, a unificação de variantes flexivas e, sobretudo, a correção e uniformização ortográficas.

⁴ Além das línguas elencadas no inquérito (francês, espanhol, inglês, alemão e mandarim), deixámos em aberto a opção “Outra língua”, que apresentou os seguintes resultados: 4 falantes nativos de russo, 3 de ucraniano, 2 de grego, 1 de coreano, 1 de polaco, 1 de turco e 1 de hália.

Variáveis		Nº de estudantes
Sexo	Homem	11
	Mulher	19
Língua materna	Francês	2
	Espanhol	5
	Inglês	6
	Alemão	3
	Mandarim	1
	Outra	13
Idade	Menos	8
	Média	14
	Mais	8
Conhecimento de outras línguas	Sim	26
	Não	4
Tempo de estudo de PLE	Menos de 1 ano	22
	Entre 1 e 2 anos	1
	Mais de 2 anos	7

Tabela 1. *Estratificação da amostra*

Em termos de caracterização, conclui-se que a maioria dos nossos informantes são do sexo feminino (19 face a 11), têm entre 26 e 39 anos, possuem conhecimentos de outras línguas estrangeiras, para além do português, estudam PLE há menos de um ano e são nativos de uma grande variedade de línguas maternas, entre as quais se destacam o inglês e o espanhol.

Ao que conseguimos apurar, a maioria estuda português por motivos profissionais ou académicos e por necessidade de integração sociocultural em Portugal.

3.3. *Hipótese central do estudo*

Considerando a distribuição da amostra em função da variável “língua materna” - que “no ha sido investigada prácticamente en los trabajos de disponibilidad léxica aplicados a estudiantes extranjeros, lo que incrementa su importancia” (Samper Hernández, 2002: 69) -, e baseando-nos também na *praxis* da disponibilidade lexical, propusemos a seguinte hipótese para nortear a nossa investigação: o número médio de palavras (*tokens*) e vocábulos (*types*) e o índice de coesão do léxico disponível dos informantes cuja língua materna é o espanhol serão superiores aos obtidos no âmbito dos outros idiomas nativos, tendo em conta que o português e o espanhol são línguas muito próximas etimologicamente.

4. APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

4.1. Índices quantitativos

Na tabela abaixo são apresentados os dados quantitativos gerais da investigação implementada:

Número	Centro de interesse					
	Corpo	Escola	Transporte	Alimentos	Roupa	Profissões
Total de palavras	359	400	316	493	258	275
Palavras diferentes	62	151	80	198	79	121
Media por inquirido	11,97	13,33	10,53	16,43	8,6	9,17
Índice de coesão	0,9	0,09	0,13	0,08	0,11	0,08

Tabela 2. Índices quantitativos gerais

Analisando globalmente os dados procedentes das provas de disponibilidade lexical, a que foi exposta a nossa amostra no passado mês de maio de 2024, verificamos que a quantidade de palavras obtida totaliza 2101, quantidade esta distribuída, em termos de centro de interesse, da seguinte forma: 359 relativas a “partes do corpo”, 400 inerentes a “escola: móveis e materiais”, 316 referentes a “meios de transporte”, 493 respeitantes a “alimentos e bebidas”, 258 alusivas a “roupa” e 275 apontadas para “profissões e ofícios”. Este valor corresponde a uma média de respostas por aluno de 70,03 e por centro de interesse, de 350,16.

No seguimento dos dados anteriores, constatamos que o núcleo temático mais produtivo é o dos “alimentos e bebidas”, seguido do núcleo da “escola: móveis e materiais” e das “partes do corpo”, que ocupam o 2.º e 3.º lugar.

Quanto ao número de palavras diferentes, este ascende a 691, refletindo, de novo, a superioridade numérica dos centros de interesse “alimentos e bebidas” e “escola: móveis e materiais”. A 3.ª posição, no entanto, deixa agora de ser ocupada pelo núcleo “partes do corpo”, que foi substituído pelo das “profissões e ofícios”.

Por último, dividindo a média de respostas pelo número de vocábulos apresentados em cada campo lexical - índice de coesão -, concluímos que a escassa coincidência das referidas respostas por parte dos informantes explica o facto de os centros de interesse se apresentarem predominantemente dispersos ou abertos, sendo o mais compacto o de “partes do corpo” (índice de coesão de 0,19).

4.2. Índices qualitativos

4.2.1. Analisados por variáveis “sexo” e “língua materna”

Analisados os resultados gerais, examinaremos, de seguida, o modo como as variáveis sexo e língua materna dos informantes - em concreto, neste último caso, a língua materna dos dois grupos de estudantes que demonstram uma maior disponibilidade

lexical, a saber: os falantes nativos de espanhol e de inglês - exercem a sua influência nas palavras apresentadas.

Número	Centro de interesse					
	Corpo	Escola	Transporte	Alimentos	Roupa	Profissões
Total de palavras	125	142	125	180	90	102
Palavras diferentes	41	81	50	111	46	61
Media por inquirido	11,36	12,91	11,36	16,36	8,18	9,27
Índice de coesão	0,28	0,16	0,23	0,15	0,18	0,15

Tabela 3. Resultados no subgrupo “Homem”

Número	Centro de interesse					
	Corpo	Escola	Transporte	Alimentos	Roupa	Profissões
Total de palavras	234	258	191	313	168	173
Palavras diferentes	52	119	53	143	60	88
Media por inquirido	12,32	13,58	10,05	16,47	8,84	9,11
Índice de coesão	0,24	0,11	0,19	0,12	0,15	0,1

Tabela 4. Resultados no subgrupo “Mulher”

Dada a diferença numérica que caracteriza a nossa amostra quanto à variável sexo (63% de mulheres e 37% de homens), não surpreende que em todos os centros de interesse se verifique um maior número de palavras e de vocábulos apresentado pelo grupo “Mulher”. Esta discrepância numérica reflete uma realidade observada na maioria das turmas dos diferentes níveis de aprendizagem de PLE, corroborando, em paralelo com a prevalência de informantes anglo-parlantes no âmbito da nossa amostra, um universo também registado por Samper Hernández no seu estudo (2002).

No que diz respeito à variável “língua materna”, estes são os resultados mais significativos:

Número	Língua materna	Centro de interesse					
		Corpo	Escola	Transporte	Alimentos	Roupa	Profissões
Total de palavras	Francês	25	22	19	38	15	20
	Espanhol	78	89	71	101	60	61
	Inglês	65	79	64	103	57	45
	Alemão	45	25	27	47	25	24
	Mandarim	1	6	7	3	2	7
	Outra	145	179	128	201	99	118

Tabela 5. Resultados por variável “língua materna”

Tendo por base a língua materna dos informantes e o número total de palavras indicado por cada grupo, constatamos que, tal como havíamos previsto, os estudantes hispano-falantes apenas são superados no centro de interesse “alimentos e bebidas” pelos seus colegas anglo-parlantes, ainda que a posição de 2.º lugar se deva a uma

diferença mínima de 2 palavras, evidenciando assim a sua maior competência lexical. A proximidade etimológica entre o espanhol e o português leva-nos a acreditar que esta seja a razão da superioridade numérica registada. Não deixa de ser curioso, contudo, que os inquiridos ingleses, no seguimento do exposto, ocupem o 2.º lugar nos restantes núcleos temáticos, dado integrarem a família das línguas germânicas. Seria expectável, no nosso entender, que esse lugar fosse preenchido pelos estudantes franceses devido à irmandade etimológica com o português e o espanhol.

4.2.2. *Analisados por centro de interesse*

Face ao carácter pioneiro da investigação aqui descrita e à nossa pretensão de fazer um estudo rigoroso e promotor de uma base estatística sólida que permita realizar a posterior seleção objetiva das palavras a ensinar aos alunos, decidimos centrar esta última parte do trabalho na apresentação das dez palavras mais disponíveis registadas em quatro núcleos temáticos por nós trabalhados, a saber: “partes do corpo”, “escola: móveis e materiais”, “alimentos e bebidas” e “roupa”.

Começamos a análise pelo núcleo temático “partes do corpo” em relação com as variáveis “sexo” e “língua materna”:

Grupo de estudo “Homem”

N.º	Palavra	Disponibilidade
1	pé/s	0,5085
2	cabeça	0,4834
3	olho/s	0,4499
4	mão/s	0,3801
5	boca	0,3609
6	cabelo	0,3218
7	coração	0,2920
8	perna/s	0,2824
9	braço/s	0,2702
10	nariz	0,2376

Grupo de estudo “Mulher”

N.º	Palavra	Disponibilidade
1	mão/s	0,6734
2	cabeça	0,6074
3	olho/s	0,5595
4	perna/s	0,4317
5	nariz	0,4137
6	pé/s	0,3833
7	cabelo	0,3796
8	boca	0,3672
9	orelha/s	0,2918
10	dedo/s	0,2787

Tabela 6. *Palavras mais disponíveis no centro de interesse “partes do corpo” por “sexo”*

Língua materna “espanhol”

1	olho/s	0,6007
2	cabeça	0,4922
3	perna/s	0,4563
4	mão/s	0,4000
5	braço/s	0,3938
6	cabelo	0,3919
7	pé/s	0,3814
8	boca	0,3616
9	nariz	0,3292
10	cara	0,3123

Língua materna “inglês”

1	cabeça	0,7237
2	olho/s	0,5161
3	pé/s	0,4842
4	boca	0,4123
5	mão/s	0,3983
6	coração	0,3966
7	nariz	0,3840
8	orelha/s	0,2737
9	braço/s	0,2680
10	dedo/s	0,2220

Tabela 7. *Palavras mais disponíveis no centro de interesse “partes do corpo” por “língua materna”*

No que concerne a diferenças qualitativas, tendo por base as dez primeiras unidades do centro de interesse “partes do corpo” apresentadas pelos grupos de estudo “Homem”

e “Mulher”, verificamos que se regista uma coincidência exata nas posições 2 e 3 e uma convergência noutras unidades que, embora não se encontrem nos mesmos lugares da lista, são indicadas por ambos os grupos - nomeadamente as palavras *cabeça, mão, olho, pé, boca, perna, cabelo e nariz* -, assumindo-se, por isso, como as mais disponíveis. Constatamos também que as unidades indicadas em primeiro lugar correspondem a extremidades do corpo humano, membros inferior e superior, ainda que depois não se registre uma continuidade ou ordem lógica a nível de palavras apresentadas. Estes resultados não surpreendem, contudo, quando comparados com os obtidos por Samper Hernández (2002) ou por Díez Abadie (2022: s/p):

[les] trois premiers mots font tous référence à des extrémités et sont tous cités avant «cheveux». Ensuite les mots se recoupant sont essentiellement constitués de parties du visage comme «bouche», «yeux» ou encore «oreille». Ce sont d'ailleurs ceux qui sont enseignés en premier.

Ou seja, os nossos resultados seguem a mesma linha dos alcançados pelos trabalhos realizados no âmbito da disponibilidade hispânica.

Por último, quanto à presença de disfemismos ou vulgarismos, se considerarmos o número total de vocábulos, verificamos que estes ocupam uma posição mais dianteira na lista de palavras indicadas pelo sexo masculino (a unidade *cu*, por exemplo, aparece em 18.º lugar) face à observada no grupo feminino (44.ª posição).

Empreendendo agora a análise dos resultados dos dois grupos que apresentam um maior número de informantes quanto à variável “língua materna”, verificamos que, à luz do observado no caso dos grupos de estudo “Homem” e “Mulher”, há uma grande homogeneidade a nível das palavras mais disponíveis. De facto, sete das dez unidades que ocupam os primeiros lugares de ambas as listas são coincidentes e correspondem a algumas das palavras ensinadas no nível A1 de aprendizagem, tendo por base o *Referencial Camões PLE* (2017) - referencial de conteúdos usado pelos docentes na conceção e organização de cursos de PLE -, em concreto as noções específicas do subtema “Partes do corpo”. *Cabelo, nariz, boca, cabeça, coração e olho* são palavras dadas a conhecer neste nível, às quais se acrescem, no nível seguinte, onde está posicionada a nossa amostra, as unidades *costas, barriga, cara, braço, mão, dedo, perna, pé, dente, pescoço, garganta, coluna, peito, estômago, ouvido e pele*. As dez primeiras palavras indicadas pelos alunos nativos de espanhol e inglês não correspondem maioritariamente às aprendidas em último lugar, no seguimento do previamente exposto, o que seria expectável dada a proximidade temporal, mas às estudadas no nível A1. Significará isso, no entanto, que a formação lexical dos alunos estagnou? Acreditamos que a falta de diversidade nos lugares iniciais não evidencia uma paragem no processo evolutivo do vocabulário disponível, mas, sim

[une] réponse à un stimulus isolé (...) [que] ne peut être que le reflet très partiel et imparfait de la compétence communicative (qui comprend, entre autres, la compétence lexicale) d'un locuteur dans une situation de communication ordinaire. (Díez Abadie, 2022 : s/p).

As palavras aprendidas no nível A2 aparecem em lugares posteriores, o que demonstra a sua assimilação e a evolução lexical dos alunos através da aprendizagem. Por último,

no que concerne a uma possível ordem aquando da apresentação das referidas palavras, ambos os grupos iniciam pela parte superior do corpo humano, *olho* e *cabeça*, embora depois não haja uma sucessão lógica, como se verifica através das unidades *perna* e *pé* que se encontram em 3.º lugar e são precedidas das palavras *mão* e *boca*.

Mostramos seguidamente a lista dos termos mais disponíveis no centro de interesse “escola: móveis e materiais” por “sexo” e “língua materna”:

Grupo de estudo “Homem”

Grupo de estudo “Mulher”

1	professor/a/es	0,6121
2	aula/s	0,5145
3	estudante/s	0,3553
4	aluno/a/s	0,2454
5	caneta/s	0,2325
6	livro/s	0,2122
7	exame/s	0,2077
8	biblioteca	0,1852
9	cadeira/s	0,1814
10	trabalhoparacasa	0,1498

1	professor/a/es	0,5436
2	aula/s	0,4817
3	aluno/a/s	0,3428
4	livro/s	0,3405
5	estudante/s	0,3182
6	caneta/s	0,2827
7	exame/s	0,2711
8	estudar	0,2608
9	lápiz	0,2381
10	aprender	0,2141

Tabela 8. Palavras mais disponíveis no centro de interesse “escola: móveis e materiais” por “sexo”

Língua materna “espanhol”

Língua materna “inglês”

Nº	TOKEN	DISPONIBILIDAD
1	professor/a/es	0,8108
2	caneta/s	0,5116
3	aula/s	0,4273
4	quadro	0,4000
5	estudante/s	0,3083
6	mesa/s	0,3046
7	cadeira/s	0,2353
8	livro/s	0,2305
9	computador	0,2109
10	ensino	0,2066

Nº	TOKEN	DISPONIBILIDAD
1	aula/s	0,6467
2	professor/a/es	0,5019
3	cadeira/s	0,4154
4	estudante/s	0,3960
5	exame/s	0,3335
6	turma/s	0,2758
7	escrever	0,2425
8	livro/s	0,2416
9	ensinar	0,2359
10	mesa/s	0,2347

Tabela 9. Palavras mais disponíveis no centro de interesse “escola: móveis e materiais” por “língua materna”

No seguimento dos resultados obtidos no âmbito do centro de interesse “partes do corpo”, também comprovamos que as palavras mais disponíveis para os grupos “Homem” e “Mulher” no campo lexical “escola: móveis e materiais” são coincidentes e ocupam, nalguns casos, as mesmas posições na lista (1.º, 2.º e 7.º lugares). Às sete unidades partilhadas -*professor*, *aula*, *estudante*, *aluno*, *livro*, *caneta* e *exame*- seguem-se outras que não se encontrando nos dez primeiros lugares de ambas as listas as integram também. Analisando as sete palavras comuns e indicadas pelos informantes, verificamos que estas se relacionam com os agentes principais do processo de ensino e aprendizagem -*professor* e *aluno*-, com o contexto de interação -*aula*-, com o material necessário para o registo dos conteúdos e realização das atividades propostas -*livro*, *caneta*- e com o modo de avaliação -*exame*-. Este centro de interesse é o 2.º mais produtivo relativamente ao número de palavras e ao número de vocábulos, tanto do ponto de vista da influência da variável “sexo”, como da incidência da variável língua

materna dos estudantes. O facto de corresponder a uma atividade -estudar/ frequentar um estabelecimento de ensino- que faz parte da rotina diária dos inquiridos justifica, no nosso entender, a posição ocupada.

No que concerne ao impacto da variável “língua materna” neste núcleo temático, aquilo que se destaca face à variável anterior, é a maior presença de termos relativos a material e mobiliário escolar, que aparecem, em concreto, na lista das dez unidades mais disponíveis dos alunos hispano-falantes (*computador e quadro*).

Tendo por base o *Referencial Camões PLE (2017)*, seria expectável, contudo, que o nome das disciplinas e/ou dos espaços escolares figurasse em grande número nas listagens de palavras recolhidas. Contudo, o seu aparecimento é pontual, o que nos leva a concluir que, embora “une langue se compose d’un ensemble de mots, quasi illimités, sémantiquement très concrets [, ces mots sont] mobilisés uniquement si le thème de la conversation le permet” (Diez Abadie, 2022).

De seguida, são apresentados os resultados qualitativos mais relevantes (em termos de índice de disponibilidade lexical) no centro de interesse “alimentos e bebidas”, ordenados pelas variáveis “sexo” e “língua materna”, respetivamente:

Grupo de estudo “Homem”			Grupo de estudo “Mulher”		
1	vinho	0,4731	1	água	0,5269
2	cerveja	0,3946	2	vinho	0,4580
3	batata	0,2839	3	cerveja	0,3787
4	água	0,2488	4	carne	0,3241
5	café	0,2468	5	sumo	0,2501
6	bacalhau	0,2214	6	peixe	0,2488
7	francesinha	0,2119	7	francesinha	0,1963
8	carnede porco	0,2117	8	café	0,1948
9	sumo	0,2020	9	chá	0,1942
10	pão	0,1921	10	legumes	0,1728

Tabela 10. *Palavras mais disponíveis no centro de interesse “alimentos e bebidas” por “sexo”*

Língua materna “espanhol”			Língua materna “inglês”		
Nº	TOKEN	DISPONIBILIDAD	Nº	TOKEN	DISPONIBILIDAD
1	água	0,4584	1	francesinha	0,4810
2	leite	0,2641	2	água	0,3875
3	pão	0,2393	3	pão	0,3536
4	vinho	0,2393	4	cerveja	0,3533
5	batata	0,2324	5	vinho	0,3514
6	feijão	0,2284	6	batata	0,3130
7	café	0,2210	7	café	0,2109
8	cerveja	0,2000	8	sumo	0,1965
9	Portoferreira	0,2000	9	jantar	0,1748
10	bolodearroz	0,2000	10	almoço	0,1698

Tabela 11. *Palavras mais disponíveis no centro de interesse “alimentos e bebidas” por “língua materna”*

Começando pela análise da incidência da variável “sexo” neste centro de interesse, esta não parece ter uma grande influência a nível das dez palavras mais disponíveis para os grupos de estudo “Homem” e “Mulher”, que coincidem em seis respostas:

vinho, cerveja, água, café, francesinha e sumo. Para além do referido, e apesar de as posições iniciais indicarem a existência de uma ordem -neste caso, começar, em primeiro lugar, pelas bebidas-, a partir da 3.^a e 4.^a posição essa sequência é quebrada com a introdução de alimentos que se vão alternando, até ao final, com outros termos referentes a bebidas.

Ainda que o centro de interesse em estudo seja o mais produtivo de todos a nível do número de palavras e de vocábulos, resultado também obtido por Samper Hernández (2002), as dez unidades mais disponíveis são termos comuns que não deixam de ter, no entanto, a sua relevância no processo de averiguação da consolidação do vocabulário aprendido no nível de proficiência anterior (A1).

Abordando agora a análise comparativa das dez palavras mais disponíveis para os estudantes falantes nativos de espanhol e de inglês, verificamos, em primeiro lugar, que se registam várias coincidências, em concreto na indicação de alimentos e bebidas que integram a refeição da maioria das pessoas, a saber: *água, pão, vinho, batata, café e cerveja*, coincidências estas também salientadas no seu trabalho por Samper Hernández (2002). A “banalidade” das respostas -isto é, o facto de serem termos básicos a nível de noções específicas deste centro de interesse- não permite identificar a origem dos informantes, nem os distingue culturalmente quanto aos hábitos alimentares. À semelhança do observado igualmente pela investigadora anteriormente referida, também não figuram entre as respostas recolhidas nomes de marcas comerciais, excetuando a referência ao *Portoferreira*, uma marca de vinho do Porto. Quanto à ordem das mencionadas respostas, esta não parece existir, uma vez que os alimentos e bebidas se vão intercalando em ambas as listas. Interessa salientar outro dado que se destaca e corrobora o descrito no *QECR*, nomeadamente:

[...] a organização cognitiva do vocabulário e do armazenamento de expressões, etc. depende, entre outras coisas, das características culturais da comunidade ou comunidades nas quais o indivíduo [...] [realizou] a sua aprendizagem” (Conselho da Europa, 2021: 35).

O conteúdo da anterior citação explicaria o facto de figurarem entre as referidas respostas unidades como *Portoferreira, bolodearroz e francesinha*, característicos da culinária do Porto, cidade onde se situa a faculdade frequentada pelos informantes da nossa amostra. Por último, verificamos que há um pequeno desvio entre os dois grupos quanto ao campo lexical proposto, nomeadamente através das palavras *jantar e almoço* que ocupam o 9.^o e 10.^o lugar na lista de respostas dos estudantes ingleses e não aparecem na lista das dez palavras mais disponíveis para os informantes hispano-falantes.

Por último, são abordados os resultados qualitativos mais destacados relativamente ao centro de interesse “roupa”, em correlação com as variáveis “sexo” e “língua materna”:

Grupo de estudo “Homem”			Grupo de estudo “Mulher”		
1	camisola	0,4798	1	vestido	0,4598
2	sapatos	0,4673	2	camisola	0,4569
3	calças	0,4410	3	camisa	0,4476
4	t’shirt	0,3243	4	sapatos	0,4012
5	camisa	0,2612	5	calças	0,3532
6	meias	0,2365	6	meias	0,2921
7	casaco	0,1483	7	casaco	0,2625
8	calções	0,1338	8	chapéu	0,2522
9	roupainterior	0,1272	9	saia	0,2447
10	chapéu	0,1009	10	t’shirt	0,2410

Tabela 12. *Palavras mais disponíveis no centro de interesse “roupa” por “sexo”*

Língua materna “espanhol”			Língua materna “inglês”		
1	sapatos	0,5779	1	sapatos	0,4600
2	calças	0,5565	2	camisa	0,4594
3	t’shirt	0,4595	3	vestido	0,4314
4	camisa	0,4469	4	camisola	0,4007
5	camisola	0,3263	5	meias	0,3641
6	vestido	0,2592	6	t’shirt	0,2726
7	casaco	0,2564	7	casaco	0,2661
8	saia	0,2231	8	roupainterior	0,2218
9	casaca	0,2000	9	calças	0,1991
10	cuecas	0,1922	10	calçasdedenin	0,1667

Tabela 13. *Palavras mais disponíveis no centro de interesse “roupa” por “língua materna”*

O primeiro aspeto que chama a atenção é que a incidência da variável “sexo” nas respostas apresentadas para o centro de interesse “roupa” se faz notar. De facto, as palavras que encabeçam ambas as listas -*camisola* e *vestido*- e o conjunto de outros termos que remetem para peças de vestuário mais prototípicas do guarda-roupa de cada um dos sexos (calças, que aparece em 3.º lugar, ou calções, no caso dos homens, frente a *vestido* e *saia*, para as mulheres) leva-nos a conjecturar que os inquiridos, consciente ou inconscientemente, começaram por indicar as peças que associam ao seu género. Para além do descrito, oito das respostas dadas são comuns (*camisola*, *sapatos*, *calças*, *t’shirt*, *camisa*, *meias*, *casaco* e *chapéu*), ocupando duas delas, inclusive, as mesmas posições nas listas (6.º e 7.º lugar). Apesar de a nossa amostra ser ligeiramente inferior à da investigação de Samper Hernández (2002), os resultados obtidos vão ao encontro dos seus, quer quanto ao número de palavras coincidentes, quer relativamente ao facto de os termos indicados englobarem calçado e vestuário. Também não encontramos nas dez palavras mais disponíveis para ambos os sexos, tal como a autora anterior, a referência a acessórios, o que seria expectável sobretudo na listagem das informantes femininas.

Examinando agora a possível influência da variável “língua materna” nas dez palavras mais disponíveis para os estudantes hispanofalantes e anglófonos, verificamos que não se registam grandes diferenças face ao descrito previamente. Senão vejamos: sete respostas são coincidentes (*sapatos*, *calças*, *t’shirt*, *camisa*, *camisola*, *vestido* e *casaco*) e duas também ocupam os mesmos lugares nas listas (1.ª e 7.ª posição).

Os termos indicados remetem igualmente para vestuário genérico e não induzem o investigador quanto à origem linguística dos inquiridos. Não se registam, por outro lado, nomes de marcas no âmbito das dez palavras mais disponíveis para os estudantes falantes nativos de espanhol e de inglês, mas, contrariamente aos dados analisados por Samper Hernández (2002), figura entre as dez respostas do último grupo a referência a um material *-denin* (tecido resistente, sarja)- que distingue os dois tipos de calças apontadas. Por último, resulta curioso que a palavra *casaca* figure entre as dez mais disponíveis dos informantes nativos de espanhol, uma vez que se trata de uma peça de roupa mais formal e não tão habitual no vocabulário de um falante nativo quer de Portugal, quer de Espanha, se bem que talvez a associação ao género feminino do vocábulo seja por causa do termo “cazadora”, que, sim, é habitual em língua espanhola, sobretudo na variante castelhana.

5. CONCLUSÕES

Chegados a este ponto, consideramos ter dado devido cumprimento ao objetivo a que nos propusemos no início do presente estudo: realizar uma primeira aproximação à disponibilidade lexical de estudantes de português como língua estrangeira (PLE), um campo de investigação até hoje inexplorado pela comunidade científica da área da léxico-estatística. Sendo o português uma das línguas mais faladas no mundo, constata-se o crescimento paulatino, mas irrefreável, do interesse pela aprendizagem de PLE, o que justifica a pertinência da nossa investigação, tendo por base a necessidade da equiparação das metodologias aplicadas ao ensino e aprendizagem do vocabulário de línguas estrangeiras modernas, com vista a uma uniformização procedimental que facilite a ulterior comparação de resultados alcançados por diferentes estudos.

Atendendo à estratificação da amostra deste estudo no que diz respeito à variável “língua materna”, os resultados quantitativos apurados levam-nos a confirmar a hipótese central formulada: os estudantes hispanofalantes produziram mais palavras (*tokens*) e mais vocábulos diferentes (*types*) do que os informantes nativos de outras línguas maternas em cinco dos seis núcleos temáticos objeto de inquirição (“partes do corpo”, “a escola: móveis e materiais”, “meios de transporte”, “roupa” e “profissões”), tendo apenas sido ultrapassados, em termos de produção lexical pelos informantes anglófonos no centro de interesse “alimentos e bebidas”. Esta prevalência poder-se-á explicar pelo facto de o português e o espanhol serem as duas línguas da rama românica etimologicamente mais semelhantes.

Por último, é importante referir que, sendo este um estudo pioneiro, é provável que sofra de algumas limitações. No entanto, é precisamente o seu caráter desbravador que multiplica as possibilidades de o mesmo ser ampliado e aperfeiçoado no futuro quer por nós, quer por outros investigadores.

6. REFERÊNCIAS

- AKBARIAN, I. H. & FARROKHI, J. 2020. Lexical availability and writing ability of EFL learners. *International Journal of Applied Linguistics*, 31(3): 48-64.
- CARCEDO GONZÁLEZ, A. 2000. *Disponibilidad léxica en español como lengua extranjera: el caso del finlandés (estudio de nivel preuniversitario y cotejo con tres fases de adquisición)*. Turku: Universidad de Turku.
- CHECA GARCÍA, I. 2023. Complementando la disponibilidad léxica con índices de riqueza léxica: advertencias metodológicas. Comunicação apresentada nas III Jornadas Científicas sobre la Investigación en Disponibilidad Léxica. Málaga.
- CONSELHO DA EUROPA. 2001. Quadro Europeu Comum de Referência para as Línguas: aprendizagem, ensino, avaliação. Lisboa: Edições Asa.
- DELGADO FERNÁNDEZ, R. 2024. *Edad y disponibilidad léxica: la variación social en el léxico disponible adulto de Salamanca*. Tese de Doutoramento. Salamanca: Universidad de Salamanca.
- DÍEZ ABADIE, G. 2022. Analyse de l'évolution de la disponibilité lexicale d'un échantillon d'élèves apprenant le FLE durant le cycle secondaire en Espagne. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 48(2).
- FERNÁNDEZ DOS SANTOS, M. 2014. Disponibilidad léxica en alumnos de español lengua extranjera del distrito de Oporto (Portugal). Em Tobar E. & M.E. Mañas (eds.), *El español como lengua extranjera en Portugal: retos de la enseñanza de lenguas cercanas*, Tomo I, pp. 92-104. Madrid: Ministerio de Educación, Cultura y Deporte.
- FERNÁNDEZ GÓMEZ, M. I. V. 2021. *Estudio de Disponibilidad Léxica en estudiantes universitarios portugueses de ELE*. Dissertação de mestrado. Braga. Universidade do Minho.
- GAMAZO CARRETERO, E. 2014. Estereotipos en el léxico disponible de universitarios portugueses. Em Tobar E. & M.E. Mañas (eds.), *El español como lengua extranjera en Portugal: retos de la enseñanza de lenguas cercanas*, Tomo I, pp. 28-41. Madrid: Ministerio de Educación, Cultura y Deporte.
- GÓMEZ MOLINA, J. R. 2021. La disponibilidad léxica y sus aplicaciones cognitivas. Em Serrano Zapata, M. & Calero Fernández, M. A. (eds.), *Aplicaciones de la disponibilidad léxica*, pp. 207-230. Valencia: Tirant lo Blanch.
- GOUGENHEIM, G., MICHÉA, R., RIVENC, P. & SAUVAGEOT, A. 1956. *L'élaboration du français élémentaire: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- HERNÁNDEZ MUÑOZ, N. & LÓPEZ GARCÍA, M. 2023. LexPro: herramienta para el cálculo de la disponibilidad y la generación de redes léxicas. Comunicação apresentada nas III Jornadas Científicas sobre la Investigación en Disponibilidad Léxica. Málaga.
- HERNÁNDEZ, N & BORREGO, J. 2004. Cuestiones metodológicas sobre los estudios de disponibilidad léxica. Em Villayandre Llamazares, M. (ed.), *Actas del V Congreso de Lingüística General*, pp. 1519-1527. Madrid: Arco Libros.
- HERRANZ LLACÉR, C. V. 2020. *Palabra de maestro: análisis del léxico disponible de los futuros docentes*. Berna: Peter Lang.

- INSTITUTO CAMÕES. (05 de maio de 2022). Português no mundo. Instituto Camões em Praga. Disponível em: <https://bit.ly/3ROP8Wq>.
- INSTITUTO CAMÕES. 2017. Referencial Camões PLE. Disponível em <https://bit.ly/4cGcjdj>.
- JIMÉNEZ CATALÁN, R.M. & DEWAELE, J.M. 2017. Lexical Availability of Young Spanish EFL Learners: Emotion words versus Non-Emotion Words. *Language, Culture and Curriculum*, 30(3): 283-299.
- LÓPEZ RIVERO, E. 2008. *Estudio de disponibilidad léxica en 43 estudiantes de ELE*. Dissertação de Mestrado. Madrid: Universidad Antonio de Nebrija.
- MAS ÁLVAREZ, I. & SANTOS PALMOU, X. 2017. Léxico disponible en el alumnado de español en la enseñanza superior en Oporto: propostas para su análisis. Em Ruiz Martínez, A. M., Martí Sánchez, M. & Fernández López, M. C. (eds.), *Investigaciones actuales en Lingüística, Vol. VI: Aplicaciones de la Lingüística*, pp. 93-111. Madrid: Universidad de Alcalá de Henares.
- PACHECO, C. R., CABRERA, J. S. & GONZÁLEZ, I. 2017. Incidencia de la variable 'sexo' en la disponibilidad léxica de estudiantes de preuniversitario en Pinar del Río, Cuba. *Íkala Revista de Lenguaje y Cultura*, 22(2): 237-253.
- PALAPANIDI, K. & MAVROU, I. 2024. La naturaleza gramatical del léxico disponible en diferentes niveles lingüísticos en español como lengua extranjera. *Revista Signos: Estudios de Lingüística*, 114: 151-167.
- PAOLINI, A. 2017. *Estudios sobre disponibilidad léxica en alumnos de ELE en la Universidad de Padua*. Trabalho de Fim de Licenciatura. Padova: Università degli Studi di Padova.
- PAREDES GARCÍA, F. (DIR). 2022. *Léxico disponible de los estudiantes de ESO y Bachillerato de Alcalá de Henares y su comarca (Madrid)*. Madrid: Universidad de Alcalá de Henares.
- SAMPER HERNÁNDEZ, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE).
- SAMPER PADILLA, J. A. 2021. Disponibilidad léxica y sociolingüística. Em Serrano Zapata, M. & Calero Fernández, M. A. (eds.), *Aplicaciones de la disponibilidad léxica*, pp. 207-230. Valencia: Tirant lo Blanch.
- SANTOS DÍAZ, I. 2020. *El léxico bilingüe del futuro profesorado. Análisis y pautas para estudios de disponibilidad léxica*. Berna: Peter Lang.
- SANTOS DÍAZ, I., TRIGO IBAÑEZ, E. & ROMERO OLIVA, M. F. 2020. La activación del léxico disponible y su aplicación a la enseñanza de lenguas. *Porta Linguarum*, 33: 75-93.
- TRUJILLANO PEÑA, M. 2015. *Disponibilidad léxica de estudiantes de 1.º de secundaria del curso 2014/2015 en el I.E.S Marismas (Cantabria)*. Trabalho de Fim de Licenciatura. Comillas: Centro Internacional de Estudios Superiores del Español CIESE-Comillas.
- ZHOU, G. 2022. *Léxico disponible en estudiantes chinos universitarios de español como lengua extranjera*. Tese de Doutoramento. Madrid: Universidad de Alcalá de Henares.