diacrítica
Revista do Centro de Estudos Humanísticos
Vol. 40, n.º 1, 2026, pp. 195–223. doi: 10.21814/diacritica.6859

Disponibilidade lexical em aprendentes chineses de português como língua estrangeira: um estudo exploratório com listas de produção lexical

Lexical availability in Chinese learners of Portuguese as a foreign language: an exploratory study using lexical production lists

Mengru Liu[1]
0009-0002-7098-1505

[1] Departamento do Português, Faculdade de Línguas e Culturas Ocidentais, Universidade de Estudos Internacionais de Sichuan, Chongqing, China.

Autor correspondente: elsa_89928@hotmail.com

Resumo. O presente estudo tem como objetivo analisar a disponibilidade lexical em aprendentes chineses de português, considerando diferentes tempos de aprendizagem (um ano e dois anos de aprendizagem) e níveis de proficiência linguística (A1, A2 e B1), com foco em seis centros de interesse: "partes do corpo", "alimentos e bebidas", "roupa", "meios de transporte", "escola" e "profissões". O estudo baseia-se ainda nos dados de Fernández dos Santos e Silva Soares (2024) relativos a aprendentes de nível A2 de diferentes línguas maternas (LMs), para explorar o possível efeito da LM na disponibilidade lexical e na estrutura do léxico mental dos aprendentes. Os resultados revelam que, na maioria dos casos, o tempo de aprendizagem exerce um efeito positivo na expansão do léxico mental e na inclusão de vocábulos periféricos. Por seu torno, o aumento do nível de proficiência linguística, para além de ter o mesmo efeito na expansão do léxico mental, reforça a estabilidade na ativação dos vocábulos centrais e de alta frequência. Contudo, o efeito sobre a mobilização dos vocábulos periféricos ou de baixa frequência é limitado. Esta expansão e estabilidade, assim como a mobilização mais ou menos ativa de determinados vocábulos, podem estar relacionadas com o ambiente de aprendizagem, os contextos de uso e a experiência de vida dos aprendentes. Além disso, a LM exerce um efeito modulador significativo na escolha de vocábulos periféricos e culturalmente específicos. Estes três fatores interagem, influenciando conjuntamente a expansão do vocabulário, a estabilidade na sua mobilização e a diversidade lexical, aspetos que devem ser considerados no ensino do vocabulário a aprendentes de português como língua estrangeira.

Palavras-chave: Português como língua estrangeira. Disponibilidade lexical. Tempo de aprendizagem. Nível de proficiência linguística. Língua materna.

Abstract. This study aims to analyze the lexical availability of Chinese learners of Portuguese, considering different learning durations (one year and two years) and language proficiency levels (A1, A2 and B1), with a focus on six centers of interest: "body parts", "food and drink", "clothing", "means of transportation", "school" and "professions". The study also draws on data from Fernández dos Santos and Silva Soares (2024) concerning A2-level learners with different native languages, in order to explore the possible effect of native language on lexical availability and the structure of learners' mental lexicon. The results reveal that, in most cases, the length of learning exerts a positive effect on the expansion of the mental lexicon and the inclusion of peripheral vocabulary. In turn, an increase in linguistic proficiency not only has the same effect

on the expansion of the mental lexicon but also reinforces the stability in the activation of central and high-frequency lexical items. However, the effect on the mobilization of peripheral or low-frequency vocabulary is limited. This expansion and stability, as well as the more or less active mobilization of specific lexical items, may be related to the learning environment, contexts of use, and learners' life experiences. Furthermore, native language has a significant modulatory effect on the selection of peripheral and culturally specific vocabulary. These three factors interact, jointly influencing vocabulary expansion, stability in its mobilization, and lexical diversity, aspects that should be taken into account in teaching vocabulary to learners of Portuguese as a foreign language.

Keywords: Portuguese as a foreign language. Lexical availability. Learning duration. Language proficiency level. Native language.

1. Introdução

Tanto na aquisição de uma língua segunda (L2) como na aprendizagem de uma língua estrangeira (LE), a competência lexical constitui uma parte importante da competência linguística e está também estreitamente relacionada com a proficiência comunicativa dos aprendentes. De facto, no ensino de LEs, revelam-se limitações para aferir a rapidez e a eficácia com que estes conseguem mobilizar vocabulário em contextos comunicativos autênticos e mesmo nos testes tradicionais, permitindo avaliar a dimensão do léxico dominado por eles. Perante esta limitação, a lexicologia e a linguística aplicada introduziram o conceito de "disponibilidade lexical" ou "palavras disponíveis" (Michéa, 1953), concebido para analisar a capacidade dos aprendentes de mobilizar o vocabulário em contextos específicos. Este ponto de estudo ajuda a avaliar melhor o nível efetivo de competência lexical e proficiência comunicativa dos aprendentes.

As investigações existentes sobre a disponibilidade lexical continuam a centrar-se sobretudo em aprendentes de espanhol como LE (como Calero Fernández, 2025; Hernández Muñoz & Tomé Cornejo, 2017; Palapanidi & Mavrou, 2024; Trigo Ibáñez, Dono López, & Fernández Gómez, 2025) e poucos são os estudos com enfoque em aprendentes de português (Fernández dos Santos & Silva Soares, 2024). Cabe ressaltar que, ainda não existem estudos sobre a disponibilidade lexical em aprendentes chineses de português. Isto implica que, no âmbito dos estudos sobre o português como língua estrangeira (PLE) na China, não só há uma ausência de análise das características da produção lexical dos aprendentes em diferentes campos semânticos, como também uma exploração insuficiente sobre estratégias de aprendizagem do léxico ou métodos de ensino em contexto de sala de aula.

Face ao crescente interesse pelo ensino do português, tanto a nível mundial como na China, revela-se particularmente pertinente investigar a disponibilidade lexical dos aprendentes chineses. Este tipo de investigação permite não só refletir de forma mais objetiva a competência lexical dos aprendentes, como também fornecer uma base empírica para otimizar a elaboração de materiais didáticos e ajustar estratégias pedagógicas.

Assim sendo, o presente estudo tem como objetivo realizar uma exploração preliminar da disponibilidade lexical em estudantes chineses universitários aprendentes de português em diferentes campos semânticos, para o qual recorremos a listas abertas de produção lexical com limite de tempo. O estudo incide na produção de vocabulário central, na diversidade lexical e nas variações entre campos semânticos. Para além disso, analisam-se, ainda, os fatores que poderão influenciar estes desempenhos, com o intuito de fornecer possíveis contributos para o ensino-aprendizagem do vocabulário no contexto do PLE.

2. Enquadramento teórico

2.1. Disponibilidade lexical: conceito e fatores associados

A disponibilidade lexical, enquanto ferramenta científica para estudar o vocabulário e a produção linguística dos aprendentes, tem vindo a receber crescente atenção nos domínios tanto da aquisição de uma L2 como do ensino-aprendizagem de uma LE. Este conceito remonta à década de 1950 (ver secção anterior), em França, tendo sido inicialmente concebido como um recurso auxiliar no ensino do francês, com o objetivo de avaliar o conjunto de vocábulos que os aprendentes conseguiam mobilizar de forma ativa dentro de campos semânticos específicos.

Com o aprofundamento das pesquisas, os estudos sobre este tema expandiram-se gradualmente para falantes de espanhol como L2 ou LE (Ávila Muñoz, 2016; Borrego Nieto & Fernández Juncal, 2002; Sánchez-Saus Laserna, 2016; Santos Díaz, 2017; Santos Díaz, Trigo Ibáñez, & Romero Oliva, 2020; Tomé Cornejo, 2015) e falantes de inglês (Ferreira & Echeverría, 2010; Germany & Cartes, 2000; Kartsevski & Blanco, 2021). Convém ressaltar que estes estudos consolidam um conjunto de metodologias de investigação plenamente padronizadas, como listas abertas de produção lexical com limite de tempo e análise de campos semânticos.

Deste modo, torna-se evidente que, no âmbito do ensino-aprendizagem de LE, a disponibilidade lexical se consolidou como uma ferramenta fundamental na avaliação da competência lexical dos aprendentes. No ensino do léxico, para além da ênfase atribuída às palavras de alta frequência, que são essenciais para satisfazer necessidades comunicativas imediatas, torna-se igualmente relevante proporcionar, aos aprendentes, oportunidades de aprendizagem e consolidação de vocábulos de menor frequência, mas de elevada utilidade e potencial de mobilização.

O vocabulário disponível integra "palavras de baixa frequência e pouco estáveis" (Biderman, 1996, p. 31), mas que, apesar disso, são usuais e mantêm grande utilidade. Trata-se de unidades lexicais estreitamente ligadas aos interesses dos aprendentes, às escolhas temáticas e às intenções comunicativas (Biderman, 1996, pp. 27–46).

Mais especificamente, a disponibilidade lexical não se define pela frequência de uso, mas pela facilidade com que determinadas palavras podem ser evocadas e utilizadas em contextos comunicativos ou temáticos específicos (Samper Hernández, 2002, p. 8). Quando, ao abordar um determinado tema, um vocábulo surge de forma rápida e espontânea na mente do falante, podemos considerá-lo um vocábulo disponível.

Assim, os estudos sobre disponibilidade lexical centram-se na capacidade do falante de ativar, num tempo limitado, o repertório vocabular relativo a um dado campo semântico. Isto demonstra, em última análise, que "conhecer palavras" e "ser capaz de as usar efetivamente" não são realidades totalmente coincidentes. A disponibilidade lexical depende em grande medida da estrutura do léxico mental do aprendente e da sua capacidade de acesso e recuperação imediata.

Importa salientar que, no léxico mental do falante, as palavras não se organizam de forma linear segundo a ordem alfabética, mas estruturam-se numa rede de relações abstratas, que podem estar ligadas a contextos culturais ou civilizacionais específicos (Chaves, 2013, p. 186). Para os aprendentes de LEs, essa rede de relações torna-se ainda mais evidente. Nomeadamente, os aprendentes tendem a associar determinados vocábulos, de forma natural, a conceitos, situações ou representações sociais familiares, ancorados na sua própria experiência cultural, linguística e de vida.

Para além disso, o processo de geração e evocação do vocabulário disponível é, em grande medida, condicionado pelo nível de proficiência linguística. De forma geral, o sistema de vocabulário disponível dos falantes nativos distingue-se de forma significativa, em termos de quantidade e qualidade, do dos não nativos (Hernández Muñoz & Tomé Cornejo, 2017). No caso dos aprendentes de LE, na produção lexical, tendem a apoiar-se em componentes específicos do seu léxico mental e a recorrer a estratégias de associação limitadas para aceder ao vocabulário, evidenciando, assim, formas de organização distintas das dos falantes nativos. Deste modo, o nível de proficiência constitui um fator incontornável a considerar na investigação sobre disponibilidade lexical em aprendentes de LE.

A nosso ver, para além deste fator, o tempo de aprendizagem exerce um papel significativo no desenvolvimento do léxico mental e na disponibilidade lexical dos aprendentes. Isto é, à medida que o tempo de aprendizagem se prolonga, aumentam aos aprendentes as oportunidades de exposição a *input* e de produção, o que leva frequentemente a um aumento do número de palavras no léxico mental e a uma maior complexidade da sua rede semântica.

A situação que normalmente ocorre é que, nas fases iniciais de aprendizagem, os aprendentes concentram-se sobretudo na aprendizagem de vocábulos centrais de maior disponibilidade ou de frequência relativamente elevada. Com a extensão do tempo de aprendizagem, o léxico mental tende a expandir-se gradualmente para incluir vocábulos periféricos de menor disponibilidade ou de baixa frequência. Simultaneamente, o aumento do tempo de aprendizagem tende a reforçar, de forma geral, a capacidade de acesso imediato ao vocabulário, o que promove uma maior disponibilidade lexical.

Com base na análise acima, propomos a seguinte hipótese: o nível de proficiência linguística e o tempo de aprendizagem influenciam a disponibilidade lexical dos aprendentes de LE. Convém ressaltar que, embora estes dois fatores não sejam totalmente independentes[1], eles influenciam a disponibilidade lexical dos aprendentes sob

[1] O tempo de aprendizagem mais prolongado poderá associar-se a um nível de proficiência linguística mais elevado. Contudo, nem sempre isto acontece. De facto, os dados do presente estudo mostram que alguns aprendentes que estudaram dois anos de português podem apresentar um nível semelhante ao de aprendentes com apenas um ano de aprendizagem, especialmente no caso de nível A2.

perspetivas diferentes. Especificamente, o primeiro reflete a competência global de uso da língua e do vocabulário e o segundo, por sua vez, indica o tempo de exposição acumulado à língua e ao vocabulário. Por este motivo, optamos por tratar ambos neste estudo (ver subsecção 4.3).

2.2. Estudo sobre a disponibilidade lexical no PLE

No âmbito da investigação sobre PLE, Fernández dos Santos & Silva Soares (2024) foram as primeiras a aplicar o estudo da disponibilidade lexical a aprendentes de PLE. Elas analisaram a produção lexical de 30 aprendentes de nível A2 em centros de interesse como "partes do corpo", "escola: móveis e materiais", "roupa", "meios de transporte", "alimentos e bebidas" e "profissões e ofícios". Paralelamente, recolheram, através de um questionário fechado, informações sociolinguísticas sobre os aprendentes, incluindo sexo, língua materna (LM), idade, conhecimentos de outras línguas e tempo de aprendizagem de PLE.

No processo de análise de dados recolhidos, o estudo realizado por estas investigadoras considerou principalmente o sexo e a língua materna dos aprendentes como variáveis, combinando métodos quantitativos e qualitativos, de modo a comparar os números de palavras (*tokens*) e de palavras diferentes/ vocábulos produzidos (*types*) em cada campo semântico por aprendentes de diferentes LMs.

O estudo de Fernández dos Santos & Silva Soares (2024) apresenta um valor teórico e prático relevante. Por um lado, preenche uma lacuna na investigação sobre a disponibilidade lexical no ensino do PLE. Por outro, os resultados demonstram que a metodologia estabelecida pelo *Proyecto Panhispánico de Disponibilidad Léxica* (PPHDL) é igualmente aplicável à investigação em PLE, fornecendo, assim, uma base científica para o planeamento lexical nesta área. Tal estudo contribui para a elaboração de manuais didáticos, para a otimização de estratégias pedagógicas em contexto de sala de aula e para o aprofundamento da investigação sobre o ensino-aprendizagem do léxico dos aprendentes.

Convém salientar que, embora a investigação tenha incluído dados de um falante nativo de mandarim, este consistiu apenas num caso isolado, pelo que não se realizou qualquer análise adicional com base na variável "língua materna". Neste caso específico, a produção lexical por cada centro de interesse foi a seguinte: "partes do corpo" (1 palavra), "escola: móveis e materiais" (6 palavras), "roupa" (2 palavras), "meios de transporte" (7 palavras), "alimentos e bebidas" (3 palavras) e "profissões e ofícios" (7 palavras) (cf. Fernández dos Santos & Silva Soares, 2024, p. 110). Em termos de quantidade de vocabulário, o falante nativo de mandarim apresentou uma produção lexical claramente superior nas categorias "escola", "roupa" e "profissões", em comparação com outros campos semânticos. No entanto, dada a amostra extremamente reduzida, não é possível generalizar estas conclusões.

Inspirado na investigação de Fernández dos Santos & Silva Soares (2024), o presente estudo concentrar-se-á no desempenho de aprendentes de PLE com LM chinesa, relativamente à disponibilidade lexical em centros de interesse específicos. Serão analisados quatro indicadores-chave, nomeadamente o número de palavras diferentes e

de vocábulos produzidos, a média de palavras por participante, o índice de coesão, bem como os vocábulos mais disponíveis, de modo a observar o desenvolvimento da competência lexical e as características do léxico mental dos aprendentes chineses de PLE.

3. Metodologia

3.1. Objetivos do estudo

O principal objetivo deste estudo consiste em explorar as características do léxico disponível dos aprendentes chineses de português universitários, com especial atenção aos efeitos potenciais do nível de proficiência linguística e do tempo de aprendizagem dos aprendentes sobre a disponibilidade lexical. Paralelamente, procede-se a uma comparação com dados de aprendentes de outras LMs com níveis linguísticos equivalentes, de modo a identificar de forma mais detalhada as características do léxico disponível dos aprendentes de PLE com LM chinesa nos mesmos campos semânticos, fornecendo, assim, um possível contributo para a melhoria do ensino do léxico a aprendentes chineses de PLE.

3.2. Questionário

O estudo recorreu à aplicação de questionários online, distribuídos aos informantes e com limite de tempo, utilizando a plataforma *Wenjuanxing*[2] como instrumento para recolher dados. O questionário foi estruturado em duas partes: (i) informações sociolinguísticas dos participantes e (ii) uma tarefa de associação lexical com tempo limitado.

No que concerne a tarefa de evocação lexical, os participantes receberam a seguinte instrução: para cada tema apresentado, escrevam livremente todos os vocábulos em português que lhes vierem à mente, procurando registar o maior número possível de palavras dentro de um período de 2 minutos por tema. Foram instruídos a realizar esta tarefa de forma independente, sem recorrer a dicionários, inteligência artificial ou ajuda de terceiros, e a separar cada palavra por vírgulas.

A primeira parte tem como objetivo recolher informações sociolinguísticas dos participantes, incluindo idade, sexo, ano de escolaridade, conhecimentos de outras LEs e nível de proficiência do português. Com base na discussão sobre a possível influência do nível de proficiência linguística na disponibilidade lexical (ver secção 2), o questionário incluiu uma pergunta de autoavaliação do nível de proficiência[3]. Convém reconhecer que esta autoavaliação pode introduzir um certo grau de subjetividade e que alguns aprendentes poderão não estar familiarizados com os critérios de referência do português europeu. Assim, optámos por acrescentar às opções descrições junto de cada opção, de

[2] Esta plataforma é um sistema muito utilizado na China para a distribuição e recolha de questionários *online* e permite o controle do tempo de resposta a cada pergunta.
[3] Devido à dificuldade de aplicar avaliações externas padronizadas a todos os participantes no contexto desta investigação, optámos por este método.

modo a permitir que os participantes escolhessem aquela que melhor correspondesse à sua situação.

Relativamente ao tempo de aprendizagem, outro fator mencionado na secção 2, os participantes são classificados com base no seu ano académico. Isto é, os aprendentes do 1.º ano são considerados com um ano de experiência de aprendizagem, enquanto os do 2.º são considerados com dois anos, dado que a recolha de dados decorreu no final do segundo semestre, em julho.

A segunda parte segue a mesma metodologia utilizada por Fernández dos Santos & Silva Soares (2024). Também recorrendo a listas de produção lexical com tempo limitado, orientámos os participantes, através da definição de centros de interesse específicos, a escrever, num formato aberto e num limite de 2 minutos, as palavras ativadas no seu léxico mental. Foram selecionados, igualmente, 6 centros de interesse: "partes do corpo", "alimentos e bebidas", "roupa", "meios de transporte", "escola: móveis e materiais" e "profissões".

Os mesmos centros de interesse permitem uma comparação direta com os dados de Fernández dos Santos & Silva Soares (2024), através da qual é possível observar as diferenças na disponibilidade lexical entre aprendentes chineses de português e falantes de outras LMs (espanhol e inglês), identificando as suas características específicas.

Convém salientar que, na tarefa de evocação lexical, a plataforma que selecionámos (*Wenjuanxing*, ver infra) permite que cada questão seja apresentada numa interface independente, iniciando-se a contagem de tempo assim que o participante acede à questão. Após dois minutos, o sistema passa automaticamente para a questão seguinte, não sendo possível avançar antes desse limite, garantindo, assim, uma gestão temporal rigorosa.

Estas questões foram apresentadas em chinês, por duas razões principais: (i) facilitar uma compreensão mais clara do tema por parte dos aprendentes; e (ii) evitar que o próprio vocábulo do tema funcione como pista de resposta. Por exemplo, se a questão apresentasse diretamente o vocábulo em português, como "roupa", o aprendente tenderia a produzir essa palavra como primeira resposta, comprometendo a validade dos dados.

3.3. Perfil dos participantes

Os participantes deste estudo foram, no total, 34 aprendentes do curso de Português de uma universidade localizada no interior da China. Destes, 12 foram do 1.º ano e 22 do 2.º ano, que correspondem, respetivamente, a um ano e dois anos de experiência de aprendizagem. É de referir que o seu ambiente de aprendizagem do português era, em geral, muito semelhante tanto ao nível do plano curricular e do plano de ensino como do corpo docente. Para além disso, nenhum dos aprendentes possuía experiência prolongada de aprendizagem em países de língua portuguesa, nomeadamente em Portugal e no Brasil. Quer isto dizer, os principais *input* e *output* restringiam-se ao ambiente formal de ensino, ou seja, em contexto de sala de aula.

Com base nisto, estes fatores externos semelhantes permitem que os resultados deste estudo reflitam melhor as diferenças na disponibilidade lexical em função dos fatores de tempo de aprendizagem e de nível de proficiência linguística

Ao mesmo tempo, e com o intuito de obter uma compreensão mais abrangente do perfil dos participantes, o questionário recolheu também informações sobre o sexo, a idade e o domínio de outras línguas, cuja distribuição se apresenta na Tabela 1.

Tabela 1. Perfil dos participantes

Variáveis		Número por grupo	Proporção	Número total
Sexo	Masculino	7	20,59%	34
	Feminino	27	79,41%	
Idade	18	3	8,82%	34
	19	13	38,24%	
	20	16	47,06%	
	21	2	5,88%	
Ano de escolaridade	1.º ano	12	35,29%	34
	2.º ano	22	64,71%	
Domínio de outras LEs	inglês	33	97,06%	34
	inglês e russo	1	2,94%	
Nível de proficiência em português	A1	6	17,65%	34
	A2	20	58,82%	
	B1	7	20,59%	
	B2	1	2,94%	

Os 34 participantes têm idades médias compreendidas entre os 19 e os 20 anos e a maioria são do sexo feminino (27 face a 7). Todos os participantes possuem formação em inglês, sendo que apenas um domina adicionalmente o russo. Ou seja, à exceção do português, nenhum deles possui conhecimentos de outras línguas românicas.

No que diz respeito aos dois fatores mencionados anteriormente (ver secções 2 e 3), observamos que: (i) em termos do tempo de aprendizagem, 12 participantes (35,29%) têm um ano de experiência e 22 (64,71%) dois anos; (ii) quanto ao nível de proficiência do português, 6 participantes encontram-se no nível A1 (17,65%), 20 no A2 (58,82%), 7 no B1 (20,59%) e apenas 1 no B2 (2,94%). Estas distribuições permitirão comparar a disponibilidade lexical entre diferentes grupos, a qual será abordada na seguinte secção.

4. Tratamento e análise dos dados

4.1. Critérios de tratamento dos dados

Durante o processo de tratamento de dados, as palavras produzidas pelos participantes nos questionários foram inicialmente registadas em Excel. Na fase de organização e normalização dessas palavras, recorreram-se principalmente aos critérios utilizados em estudos sobre a disponibilidade lexical em espanhol (Samper Hernández, 2002, pp. 24–27; cf. também Samper Padilla, 1998). Convém ressaltar que, tendo em conta as características específicas dos participantes deste estudo, esses critérios também foram ajustados e complementados sempre que necessário, como se descreve a seguir.

(1) Eliminação de palavras repetidas. Sempre que um mesmo participante registasse mais de uma vez a mesma unidade lexical dentro de um mesmo centro de interesse,

apenas se conservava a primeira ocorrência, sendo as repetições subsequentes descartadas. Por exemplo, no centro de interesse "alimentos e bebidas", o participante 7 registou *batata* por duas vezes e o participante 21 registou *maçãs* também duas vezes; no tratamento de dados, apenas a primeira ocorrência de cada palavra foi mantida.

(2) Correção ortográfica. Tendo em conta que os participantes são aprendentes chineses de português e que o instrumento de resposta foi um telemóvel ou computador, algumas variações na escrita ou ausência de acentos podem não constituir erros ortográficos genuínos, especialmente nos casos de acentuação omitida. Assim, as palavras que puderam ser claramente identificadas foram corrigidas para a grafia padrão do português. Exemplos incluem: *cafe → café*, *chapeu → chapéu*, *cabeca → cabeça*, *medico → médico*, entre outros.

(3) Uniformização das formas ortográficas. No tratamento dos dados, foram identificadas algumas variações ortográficas em determinados vocábulos. Nestes casos, procedeu-se à sua uniformização. Por exemplo, no que respeita às diferenças ortográficas, *actor/ator* foi unificado como *ator* e *arquitecto/arquiteto* como *arquiteto*, de forma a garantir a consistência e a comparabilidade dos dados.

(4) Uniformização de número e género. Para os substantivos que surgiram apenas no singular ou apenas no plural, manteve-se a forma original. Para aqueles que apareceram tanto no singular como no plural, adotou-se a marcação combinada (ex.: *maçã/s*), refletindo integralmente o seu uso. No centro de interesse "profissões", os substantivos com distinção de género foram também representados de forma combinada (ex.: *professor/professora*), de modo a contemplar simultaneamente os géneros masculino e feminino.

(5) Preservação de palavras compostas e estrangeirismos. No caso das palavras compostas, escolheu-se manter a forma original, sem divisão ou alteração, para evitar confusão entre diferentes unidades semânticas. Por exemplo, *empregado de mesa* e *empregado de balcão* referem-se a conceitos distintos e devem ser claramente diferenciados. Quanto aos estrangeirismos considerados "neologismos universalizados", preservou-se a grafia original (normalmente em inglês), como *jeans*, de forma a refletir o uso corrente dos participantes no contexto da língua portuguesa.

(6) Tratamento da última palavra incompleta. Devido ao limite de tempo para a resposta, alguns participantes não completaram a última palavra dentro do tempo estipulado. Palavras cujo significado não possa ser inferido de forma clara foram eliminadas, como *n** ou *esc**. No entanto, se for possível deduzir razoavelmente a forma completa a partir do fragmento escrito, a palavra foi considerada válida, por exemplo: *sop → sopa*, *casac → casaco*.

De acordo com os critérios acima, os dados foram inseridos em Excel e devidamente processados. Em seguida, procedemos uma análise quantitativa dos dados utilizando o *LexPro*. O *LexPro* é uma ferramenta de análise lexical multilíngue, capaz de realizar análises hierárquicas de frequência lexical e avaliar as características psicoléxicas do vocabulário (Van Parys, De Wilde, Macken & Montero Perez, 2024). Além disso, esta ferramenta permite calcular diretamente o índice de disponibilidade lexical, sendo, por isso, o instrumento adequado para o processamento de dados neste estudo.

Todas as variáveis e os centros de interesse foram organizados e codificados de forma uniforme, em conformidade com as orientações do *Manual de Uso LexPro* (Hernández Muñoz, Tomé Cornejo, López García & Bartol Hernández, 2023). Após este processamento, os dados são importados para o software *LexPro* para análises quantitativas e qualitativas adicionais, permitindo a obtenção de indicadores-chave acima mencionados (ver secção 2).

4.2. Análise Quantitativa

Após o tratamento dos dados, foram obtidos, no total, 2085 palavras e 479 vocábulos produzidos por 34 participantes. Para apresentar de forma abrangente a produção lexical dos participantes nos 6 centros de interesse estabelecidos, calcularam-se o número total de palavras (*tokens*) e o de vocábulos produzidos (*types*) em cada centro, bem como a média de palavras por participante e o índice de coesão, como se mostra na Tabela 2.

Tabela 2. Indicadores quantitativos gerais por centro de interesse

Índice	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
Total de palavras (*tokens*)	291	403	363	313	362	353
Vocábulos (*types*)	73	101	61	55	104	85
Média por informante	8,56	11,85	10,68	9,21	10,65	10,38
Índice de coesão	0,12	0,12	0,18	0,17	0,10	0,12

Em relação ao número total de palavras produzidas, que permite avaliar a riqueza do vocabulário, o centro de interesse “alimentos e bebidas” apresenta a maior quantidade de palavras (403). Seguem-se, com valores próximos, “roupa” (363), “escola” (362) e “profissões” (353). Em contraste, “meios de transporte” (313) e, sobretudo, “partes do corpo” (291) apresentam números relativamente mais baixos. Estes dados indicam que os aprendentes produzem mais palavras nos campos semânticos relacionados com alimentos e bebidas e com atividades de aprendizagem ou sociais.

No que concerne a número de vocábulos produzidos, que reflete, por sua vez, a diversidade lexical. Os centros de interesse “escola” (104) e “alimentos e bebidas” (101) mantêm-se no topo, evidenciando uma maior diversidade lexical nestes campos semânticos. Por outro lado, nos domínios “roupa” (61) e “meios de transporte” (55) apresentam menos vocábulos, indicando que o uso do vocabulário por parte dos aprendentes é mais concentrado nestes campos semânticos.

A média de palavras por participante é obtida dividindo-se o número total de palavras pelo número de participantes. Desta forma, o centro de interesse “alimentos e bebidas” apresenta o valor mais elevado, com uma média de 11,85 palavras, seguido por “roupa” (10,68) e “escola” (10,65), sugerindo que estes campos semânticos estimulam uma maior produção lexical.

O índice de coesão reflete a frequência de repetição de um mesmo vocábulo dentro de um conjunto lexical, funcionando como um indicador do grau de "consenso" vocabular numa comunidade linguística. Neste sentido, valores elevados podem indicar que, em determinados campos semânticos, os aprendentes tendem a utilizar certos vocábulos, enquanto valores baixos sugerem um uso lexical mais disperso e diversificado.

A este respeito, os campos semânticos "roupa" (0,18) e "meios de transporte" (0,17) apresentam valores ligeiramente superiores aos de outros centros de interesse, enquanto "escola" (0,10) regista o valor mais baixo. Isto indica que, nos campos "roupa" e "meios de transporte", os aprendentes repetem os mesmos vocábulos com maior frequência, ao passo que, no campo semântico "escola", a repetição é menor e a diversidade lexical é mais evidente.

De forma geral, os aprendentes demonstram tanto uma produção lexical elevada como uma diversidade lexical significativa nos campos semânticos "alimentos e bebidas" e "escola". Por outro lado, no campo semântico "meios de transporte", tanto a quantidade de palavras produzidas como a diversidade lexical são relativamente limitadas, sugerindo que este domínio poderia beneficiar de uma atenção pedagógica reforçada para ampliar o léxico mental.

4.3. Análise qualitativa

Com base na discussão apresentada na secção 2, tanto o tempo de aprendizagem como o nível de proficiência linguística podem influenciar a disponibilidade lexical. Por isso, é necessário verificar, neste primeiro momento, se existe uma correlação elevada entre os dois fatores, de modo a determinar se estes podem ser considerados como variáveis independentes na próxima análise.

A este respeito, recorremos ao *SPSS* (*Statistical Package for the Social Science*) para realizar a análise de correlação. Visto que ambas as variáveis são contínuas, mas não seguem a distribuição normal, é utilizado o coeficiente de correlação de *Spearman*, disponível no SPSS para testar a correlação entre elas. O coeficiente obtido é de 0,356, situando-se entre 0,2 e 0,4, o que indica a existência de uma correlação positiva, mas de fraca a moderada entre as duas variáveis.

Este resultado sugere que, no caso dos participantes deste estudo, não se verifica uma relação de elevada dependência nem colinearidade significativa entre os dois fatores. Deste modo, optamos por considerá-las como variáveis relativamente independentes e por analisá-las de forma separada.

4.3.1. O tempo de aprendizagem como variável

Nesta subsecção, compararemos os dados de grupos com diferentes tempos de aprendizagem (um ano e dois anos) em 6 centros de interesse, considerando quatro indicadores acima referidos (ver Tabela 3).

Tabela 3. Indicadores por centro de interesse com o tempo de aprendizagem como variável

Índice	Variável	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
Total de palavras (*tokens*)	Um ano	104	129	120	98	114	111
	Dois anos	187	274	243	215	248	242
Vocábulos (*types*)	Um ano	51	56	30	29	60	46
	Dois anos	44	89	53	47	87	73
Média por informante	Um ano	8,67	10,75	10,00	8,17	9,50	9,25
	Dois anos	8,50	12,45	11,05	9,77	11,27	11,00
Índice de coesão	Um ano	0,17	0,19	0,33	0,28	0,16	0,20
	Dois anos	0,19	0,14	0,21	0,21	0,13	0,15

Conforme ilustrado na Tabela 3, os aprendentes com dois anos de aprendizagem apresentam, em todos os centros de interesse, um número total de palavras superior ao dos com um ano, sendo o aumento mais pronunciado nos domínios "alimentos e bebidas" (129 a 274), "escola" (114 a 248) e "profissões" (111 a 242). Importa salientar que, dado que o número de participantes no grupo de dois anos é superior ao do de um ano, o número total de palavras por si só não permite refletir adequadamente as diferenças na produção lexical dos aprendentes. Por isso, é necessária a consideração da média de palavras por participante para uma avaliação mais precisa.

Assim, constatamos que, na maioria dos campos semânticos, a média de palavras por participante é também mais elevado nos aprendentes com dois anos de aprendizagem. Destacam-se novamente os três centros referidos: "alimentos e bebidas" (10,75 a 12,45), "escola" (9,50 a 11,27) e "profissões" (9,25 a 11,00). Para além disso, na maioria dos campos semânticos, os aprendentes com dois anos de aprendizagem apresentam um maior número de vocábulos em comparação com os de apenas um ano. Tais resultados sugerem o efeito positivo do tempo de aprendizagem sobre a expansão lexical e também a diversidade lexical dos aprendentes.

Contudo, existe uma exceção no domínio "partes do corpo". Observa-se que, neste campo semântico, a média de palavras por participante do grupo de dois anos é semelhante à do grupo de um ano, apresentando até uma ligeira diminuição (de 8,67 a 8,50). A comparação do número de vocábulos revela um outro fenómeno digno de nota: o grupo com um ano de aprendizagem produziu 51 vocábulos, enquanto o grupo com dois anos apenas 44. Ou seja, à medida que o tempo de aprendizagem aumenta, os vocábulos ativados neste campo semântico não se ampliam, antes pelo contrário, registam-se uma diminuição.

No nosso entender, este fenómeno pode ser explicado pelo facto de os vocábulos relacionados com "partes do corpo" apresentarem contextos de uso relativamente limitados, tanto em contexto de sala de aula como fora dele. Quer isto dizer que o aumento

do tempo de aprendizagem não implica necessariamente uma expansão do léxico mental em todos os campos semânticos, especialmente para vocábulos periféricos ou de uso menos frequente. Na ausência de contextos de uso específico, ou quando os vocábulos previamente aprendidos não são revisados nem ampliados atempadamente, é possível que a extensão do tempo de aprendizagem conduza ao esquecimento, resultando, consequentemente, a uma redução nos vocábulos ativados.

Quanto ao índice de coesão, os aprendentes com um ano de aprendizagem apresentam um valor mais elevado em "roupas" (0,33) do que os de dois anos (0,21). A redução deste índice indica que os aprendentes com menor tempo de aprendizagem tendem a concentrar no uso de poucos vocábulos centrais de uso relativamente frequente, enquanto os de dois anos recorrem a um léxico mais diversificado.

Adicionalmente, nos campos semânticos "partes do corpo" e "escola", as variações do índice de coesão são pouco significantes. Em particular, em "partes do corpo", o número de vocábulos e a média de palavras por participante não aumentam, chegando mesmo a decrescer, enquanto o índice de coesão regista um ligeiro aumento. Isto sugere que o tempo de aprendizagem tem um efeito limitado sobre o vocabulário neste domínio, possivelmente devido ao reduzido número de vocábulos centrais, o que facilita a construção de conjuntos de unidades lexicais fixas pelos aprendentes, conforme referido anteriormente.

No centro de interesse "escola", mesmo não se observando uma diferença significativa nos índices de coesão, apresentam os valores mais baixos dentro do respetivo grupo. Isto indica que a escolha das palavras dos aprendentes neste campo semântico é relativamente dispersa, refletindo uma elevada diversidade. Este fenómeno poderá estar relacionado com o ambiente de aprendizagem dos aprendentes, no qual o contacto e o uso do português abrangem objetos e conceitos diretamente ligados ao ambiente de ensino formal, havendo mais oportunidades de exposição à língua e de produção linguística relativas a este campo semântico.

Em suma, o tempo de aprendizagem exerce uma certa influência na disponibilidade lexical dos aprendentes chineses de português, embora essa influência varie consoante os diferentes campos semânticos. De um modo geral, quanto maior o tempo de aprendizagem, maior tende a ser o número total de palavras e o de vocábulos produzidos dos aprendentes, verificando-se uma expansão do léxico mental e uma maior diversidade na ativação lexical.

No entanto, influenciado pela fixidez do vocabulário central e pelos contextos de uso, o aumento do tempo de aprendizagem não se traduz de forma significativa no acréscimo do número de palavras nem na diversidade lexical em torno do campo semântico "partes do corpo". Para além disso, caso os aprendentes disponham, nas fases iniciais, de oportunidades relativamente abundantes de exposição à língua e de uso da língua, como no campo semântico "escola", o seu léxico mental poderá já ter atingido certo grau de riqueza desde cedo. Nestes casos, a prolongação do tempo de aprendizagem pode não ter um efeito relevante no aumento da diversidade lexical.

Para aprofundar a análise da influência do tempo de aprendizagem na disponibilidade lexical de aprendentes chineses de português, também contabilizámos,

separadamente, os 10 vocábulos mais disponíveis em cada centro de interesse, para os grupos com um e dois anos de aprendizagem (ver Tabelas 4 e 5).

Tabela 4. Os 10 vocábulos mais disponíveis em cada centro de interesse (Grupo de 1.º ano)

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
1	olho/s	pão	camisa	carro/s	cadeira/s	professor(es)/a(s)
2	cabeça	arroz	saia	avião	mesa	médico(s)/a(s)
3	mão	maçã/s	calça/s	ónibus	carteira	engenheiro(s)/a(s)
4	boca	leite	casaco	bicicleta	caneta	advogado(s)/a(s)
5	nariz	banana/s	blusa	metro	lápis	dentista
6	pé/s	batata	meia/s	táxi	quadro	bombeiro/s
7	cabelo	peixe	sapato/s	barco	sala/s de aula	ator/atriz
8	língua	uva	roupa/s	autocarro/s	professor/a	enfermeiro(s)/a(s)
9	lábio/s	queijo	chapéu	comboio	livro/s	polícia
10	rosto	ovo/s	vestido	moto/a	escola	arquiteto/a

Tabela 5. Os 10 vocábulos mais disponíveis em cada centro de interesse (Grupo de 2.º ano)

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
1	cabeça	pão	camisa	carro/s	cadeira/s	professor(es)/a(s)
2	olho/s	arroz	calça/s	avião	mesa	médico(s)/a(s)
3	cabelo	café	saia	metro	caneta	advogado(s)/a(s)
4	boca	batata	roupa/s	bicicleta	cama	engenheiro(s)/a(s)
5	mão	chá	sapato/s	trem	livro/s	polícia
6	nariz	água	casaco	barco	escola	dentista
7	pé/s	leite	vestido	táxi	mochila	arquiteto/a
8	corpo	peixe	chapéu	ónibus	universidade	bombeiro/s
9	joelho	carne	meia/s	autocarro/s	lápis	jornalista
10	rosto	ovo/s	blusa	moto/a	sala/s de aula	ator/atriz

Comparando os dados das duas tabelas acima, constatamos que os dois grupos apresentam uma consistência geral elevada nos vocábulos de alta disponibilidade lexical em cada centro de interesse, embora se evidenciem pequenas diferenças associadas ao aumento do tempo de aprendizagem.

Especificamente, no domínio "partes do corpo", os vocábulos mais disponíveis de ambos os grupos são relativamente consistentes em *cabeça*, *olho/s*, *mão*, *boca* e *nariz*, indicando que estas palavras constituem uma reserva do léxico central relativamente consolidada desde fases iniciais de aprendizagem. Em particular, *cabeça* ocupa 2.º lugar no grupo de um ano e 1.º lugar no grupo de dois anos, com índices de disponibilidade relativamente estáveis (0,59; 0,61). Paralelamente, o grupo de dois anos inclui também

vocábulos como *corpo* e *joelho*. Estes resultados sugerem que, com um maior tempo de aprendizagem, o uso dos vocábulos centrais tende a tornar-se mais consistente e estável, refletindo também uma expansão da rede semântica da reserva do léxico mental dos aprendentes.

De igual modo, em "alimentos e bebidas", *pão* e *arroz* ocupam posições idênticas em ambos os grupos (os dois primeiros lugares) e apresentam índices de disponibilidade superiores aos restantes vocábulos, evidenciando o seu estatuto central no léxico mental dos aprendentes. Supomos que a ativação destes vocábulos esteja, possivelmente, associada ao seu sentido cultural, enquanto alimentos básicos da população chinesa, que é, ao mesmo tempo, muito consumidos também em Portugal e no Brasil.

No que respeita a outros vocábulos centrais, *leite* também apresenta ligeiras diferenças de ordenação entre os grupos, cujos índices de disponibilidade são também muito próximos (0,28 no grupo de um ano e 0,27 no grupo de dois anos), confirmando também a sua estabilidade do núcleo lexical dos aprendentes. Tal resultado poderá ter a mesma explicação em relação ao sentido cultural. Isto é, tanto para a cultura alimentar chinesa como para a portuguesa ou brasileira, *leite* é um vocábulo que aparece frequentemente na vida quotidiana.

Em contraste, o desempenho do grupo de dois anos neste campo semântico revela uma ampliação mais significativa da mobilização lexical. Por exemplo, o vocábulo *café*, cuja disponibilidade no grupo de um ano é apenas de 0,12, ocupando o 15.º lugar, regista um aumento significativo para 0,37 no grupo de dois anos, ascendo ao 3.º lugar. De forma semelhante, *chá* passa do 23.º lugar (0,09) no grupo de um ano para o 5.º lugar (0,31) no grupo de dois anos.

Esta alteração marcante sugere que, à medida que o tempo de aprendizagem aumenta, os aprendentes expandem gradualmente o escopo de mobilização lexical, passando de ingredientes básicos para categorias semânticas mais diversificadas e culturalmente relevantes, partindo de uma base de vocabulário central de "alimentos e bebidas" já estabilizada.

No centro de interesse "roupa", os dois grupos apresentam os mesmos dez vocábulos mais disponíveis, embora a ordem não seja exatamente a mesma. Tal constatação indica que, desde o início da aprendizagem, os aprendentes já possuem uma estrutura central relativamente estável no léxico mental deste campo semântico, sem alterações significativas com o aumento do tempo de aprendizagem. Nomeadamente, o vocábulo *camisa* ocupa o 1.º lugar no vocabulário lexical em ambos os grupos, com uma diferença muito reduzida no índice de disponibilidade (0,65 no grupo de um ano e 0,67 no de dois anos), evidenciando a sua posição consolidada no léxico mental dos aprendentes. Enquanto isso, os vocábulos como *calça/s* e *saia* apresentam igualmente ordenações semelhantes. Isto demonstra que o grau de mobilização destes três vocábulos é elevado e relativamente estável, sendo pouco influenciado pelo tempo de aprendizagem. Este fenómeno poderá estar relacionado com a funcionalidade destes vocábulos e a contextos de uso relativamente limitados.

De forma semelhante, os vocábulos mais disponíveis no domínio "meios de transporte" não apresentam diferenças significativas entre os dois grupos. No que diz respeito à ordenação, os vocábulos *carro/s*, *avião* e *bicicleta* na disponibilidade lexical

mantêm-se idênticos, ocupando, respetivamente, 1.º, 2.º e 4.º lugares nas listas. Estes resultados indicam que estes vocábulos ocupam uma posição relativamente estável no léxico mental dos aprendentes.

É de salientar que *carro/s* não apenas apresenta o índice de disponibilidade mais elevado dentro deste campo semântico, como também alcança o valor mais alto entre todos os seis centros de interesse analisados neste estudo (0,83). Tal evidência sugere que *carro/s* se destaca pela generalidade na ativação lexical dos aprendentes. O seu estatuto central é, muito provavelmente, influenciado pelas experiências de vida, pela frequência de exposição e produção linguística, bem como pela elevada transferibilidade em contextos interculturais, fatores que contribuem, em conjunto, para a sua ativação prioritária e constante no léxico mental.

Adicionalmente, surgem os vocábulos *trem* e *metro,* bem como *autocarro* e *ónibus*, refletindo que as formas variantes do português europeu e brasileiro coexistem no léxico mental dos aprendentes. Destaca-se ainda que, no grupo de um ano, o índice de *ónibus* (0,43, 3.º lugar) é significativamente superior ao do grupo de dois anos (0,30, 8.º lugar). Tal resultado pode estar relacionado com diferenças nos contextos de exposição linguística. Por exemplo, o grupo de um ano possui maior contacto com o português brasileiro em contexto de ensino formal (por exemplo, com docentes brasileiros), o que promove a mobilização de *ónibus* mais ativada no léxico mental.

No centro de interesse "escola", ambos os grupos se concentram em vocábulos como *cadeira/s*, *mesa*, *caneta* e *livro/s*. Estes vocábulos mais disponíveis correspondem, na sua maioria, aos objetos concretos e mais visíveis no contexto da sala de aula e, consequentemente, constituem *input* repetido desde as fases iniciais da aprendizagem, tornando-se vocabulários centrais relativamente mais estáveis.

Paralelamente, no grupo de dois anos surgem também vocábulos como *mochila* e *universidade*, que apresentam maior diversidade contextual do que os vocábulos do grupo de um ano, como *quadro* ou *carteira*. Estes resultados sugerem que, uma vez consolidado o vocabulário central em contexto da sala de aula, o léxico mental dos aprendentes tende a expandir-se gradualmente para incluir vocábulos periféricos relacionados com experiências de aprendizagem e de vida, à medida que o tempo de aprendizagem aumenta.

Importa notar que o índice de disponibilidade de *carteira* neste campo semântico apresenta uma diferença significativa. De forma concreta, no grupo de um ano atinge 0,42 (3.º lugar), enquanto no grupo de dois anos desce para 0,08 (21.º lugar). Esta alteração pode dever-se ao facto de, com o aumento do tempo de aprendizagem, os aprendentes irem progressivamente dominando um maior número de vocábulos centrais relacionados com o contexto de ensino formal e assim, vocábulos com uso ou exposição relativamente frequente neste contexto poderão ter reduzido a disponibilidade de *carteira*. Paralelamente, *carteira* em português europeu apresenta caráter polissémico, podendo referir-se tanto a *mesa a estudar* como a *bolsa de mão*. Ao entrarem em contacto com contextos de uso mais diversificados, os aprendentes poderão tender a associá-lo à aceção do significado *bolsa de mão*, diminuindo assim a sua disponibilidade no campo semântico da escola.

No centro de interesse "profissões", os vocábulos disponíveis apresentam-se relativamente mais consistentes, particularmente *professor(es)/a(s)*, *médico(s)/a(s)*,

advogado(s)/a(s) e *engenheiro(s)/a(s)*. Destaca-se que, com diferenças reduzidas nos índices entre o grupo de um ano e o de dois anos, *professor(es)/a(s)* ocupa o 1.º lugar (0,82; 0,75) e *médico(s)/a(s)* mantém-se no 2.º lugar (0,63; 0,60). Esta consistência na ordenação evidencia que a ativação de tais vocábulos é muito estável no léxico mental dos aprendentes. Comparativamente, o grupo de dois anos apresenta uma diversidade lexical ligeiramente superior neste campo semântico, incluindo, por exemplo, o vocábulo *jornalista*.

De forma geral, os dois grupos apresentam uma elevada sobreposição, o que sugere que os vocábulos centrais de elevada disponibilidade destes campos semânticos se estabelecem já nas fases iniciais da aprendizagem do português. Contudo, com o aumento do tempo de aprendizagem, observa-se uma expansão do léxico mental e a cobertura semântica mais abrangente em vários domínios. Ou seja, um maior tempo de aprendizagem possibilita a integração de vocábulos periféricos ou de baixa frequência no léxico central, promovendo, assim, a acumulação de vocabulário e o enriquecimento da rede semântica do léxico mental dos aprendentes. Paralelamente, poderá acontecer uma ligeira diminuição da eficiência na mobilização dos vocábulos centrais originais, em função do *input* linguístico recebido ou de contexto de uso destes vocábulos.

4.3.2. O nível de proficiência do português como variável

Conforme referido na secção 2, o nível de proficiência linguística constitui igualmente um fator relevante a considerar na análise da disponibilidade lexical. Nesta subsecção, será explorado este fator como variável, através da análise estatística e da comparação do desempenho dos grupos de diferentes níveis nos seis centros de interesse. Importa referir que apenas 1 participante se situava no nível B e devido a este número muito reduzido, não é possível realizar uma análise estatística fiável. Deste modo, na próxima análise em que o nível de proficiência linguística é considerado como variável, focamos nos três grupos de aprendentes A1, A2 e B1.

De forma igual à subsecção anterior, são apresentados quatro indicadores chave (ver Tabela 6).

Tabela 6. Indicadores por centro de interesse com o nível de proficiência do português como variável

Índice	Variável	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
Total de palavras (*tokens*)	A1	42	71	65	47	62	54
	A2	174	223	212	188	221	211
	B1	64	103	78	72	72	81
Vocábulos (*types*)	A1	23	42	30	25	44	33
	A2	60	74	44	41	76	66
	B1	27	58	29	20	43	43
Média por participante	A1	7,00	11,83	10,83	7,83	10,33	9,00
	A2	8,70	11,15	10,60	9,40	11,05	10,55
	B1	9,14	14,71	11,14	10,29	10,29	11,57

Índice de coesão	A1	0,30	0,28	0,36	0,31	0,23	0,27
	A2	0,14	0,15	0,24	0,23	0,15	0,16
	B1	0,34	0,25	0,38	0,51	0,24	0,27

Importa salientar que, devido às diferenças significativas no tamanho das amostras entre os níveis (6 em A1, 20 em A2 e 7 em B1), os valores relativos ao número total de palavras não podem ser considerados um indicador fiável. Por isso, a análise incide sobretudo nos três indicadores restantes.

No que diz respeito à média de palavras por participante, verifica-se que, na maioria dos campos semânticos, o grupo B1 apresenta valores superiores ao dos grupos A2 e A1. Estes dados revelam que o aumento do nível de proficiência do português exerce, de modo geral, um efeito positivo sobre a quantidade total de vocabulário ativado. Contudo, no centro de interesse "escola", o valor mais elevado é registado no grupo A2, enquanto o grupo B1 apresenta o valor mais baixo.

No número de vocábulos apresenta-se um fenómeno semelhante. Em comparação com o grupo A2, o grupo B1 apresenta uma menor diversidade no léxico mental. Por exemplo, no domínio "escola", o grupo A2 produz 76 vocábulos, ao passo que o grupo B1 regista apenas 43. Mesmo este resultado podendo estar relacionado com o número maior da amostra do nível A2, parece indicar que o aumento do nível nem sempre leva a um incremento na diversidade lexical, uma vez que esta também pode estar associada a outros fatores. Por exemplo, poderá dever-se ao facto de os aprendentes do nível A2 estarem, naquele momento, a aprender os vocábulos deste campo semântico e, consequentemente, a ter um contacto mais frequente com eles.

Relativamente ao índice de coesão, observa-se que, na grande maioria dos campos semânticos, os grupos A1 e B1 apresentam índices relativamente próximos entre si, geralmente superiores aos do grupo A2. Tal sugere que, na transição do nível A1 para A2, o capital lexical dos aprendentes tende a diversificar-se, resultando numa menor taxa de repetição. Contudo, com o aumento contínuo para o nível B1, a produção lexical passa a concentrar-se nos vocábulos centrais e de alta frequência, aumentando novamente a taxa de repetição. Importa particularmente salientar que o valor mais elevado entre todos os índices obtidos se regista no grupo B1, no domínio "meios de transporte", o que sugere que os aprendentes partilham um léxico mental relativamente próximo neste campo semântico.

Assim, concluímos que, com o aumento do nível de proficiência do português, a escala do léxico mental e a competência lexical dos aprendentes aumentam globalmente, apresentando uma tendência de crescimento na maior produção lexical. Contudo, esta evolução não se traduz necessariamente numa correlação positiva em todos os aspetos, dado que aprendentes de nível mais elevado podem apresentar uma dependência excessiva dos vocábulos centrais, fazendo com que a diversidade lexical não aumente de forma proporcional. Em consequência, o aumento do nível de proficiência promove sobretudo a estabilidade do léxico central, tendo um impacto limitado sobre a diversidade lexical e a ativação de vocábulos periféricos.

Para analisar de forma mais detalhada os vocábulos centrais de alta frequência nos níveis A1–B1, foram extraídos os 10 vocábulos mais disponíveis em cada centro de interesse (ver Tabelas 7, 8 e 9).

Tabela 7. Os 10 vocábulos mais disponíveis em cada centro de interesse (Grupo de A1: 6 participantes)

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
1	mão	pão	camisa	carro/s	carteira	professor(es)/a(s)
2	cabeça	arroz	casaco	ónibus	cadeira/s	médico(s)/a(s)
3	olho/s	suco	calça/s	autocarro/s	mochila	advogado(s)/a(s)
4	cabelo	água	blusa	táxi	giz	dentista
5	corpo	batata	vestido	barco	quadro	ator/atriz
6	prima	ovo/s	saia	avião	cama	engenheiro(s)/a(s)
7	língua	peixe	gravata	moto/a	mesa	padeiro
8	boca	feijão	chapéu	metro	dormitório	profissão
9	palma	laranja	meia/s	bicicleta	caderno	cozinheiro/a
10	pele	camarão	jeans	cavalo	caneta	camponês

Tabela 8. Os 10 vocábulos mais disponíveis em cada centro de interesse (Grupo de A2: 20 participantes)

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
1	cabeça	arroz	camisa	carro/s	cadeira/s	professor(es)/a(s)
2	olho/s	pão	calça/s	avião	mesa	médico(s)/a(s)
3	boca	batata	saia	metro	caneta	advogado(s)/a(s)
4	nariz	maçã/s	roupa/s	bicicleta	livro/s	engenheiro(s)/a(s)
5	cabelo	leite	meia/s	ónibus	lápis	polícia
6	mão	chá	casaco	trem	escola	arquiteto/a
7	pé/s	comida	blusa	barco	sala/s de aula	dentista
8	rosto	peixe	sapato/s	táxi	cama	bombeiro/s
9	língua	carne	vestido	autocarro/s	carteira	enfermeiro(s)/a(s)
10	dente	café	chapéu	comboio	universidade	ator/atriz

Tabela 9. Os 10 vocábulos mais disponíveis em cada centro de interesse (Grupo de B1: 7 participantes)

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Meios de transporte	Escola	Profissões
1	cabeça	pão	camisa	carro/s	cadeira/s	professor(es)/a(s)
2	mão	café	calça/s	bicicleta	escola	médico(s)/a(s)
3	olho/s	arroz	saia	avião	mesa	bombeiro/s
4	pé/s	leite	sapato/s	táxi	caneta	advogado(s)/a(s)
5	boca	queijo	camisola	metro	cama	dentista
6	corpo	chá	chapéu	barco	universidade	engenheiro(s)/a(s)
7	nariz	sopa	casaco	ónibus	lápis	enfermeiro(s)/a(s)
8	coração	bolo	blusa	comboio	mochila	ator/atriz

9	língua	água	luva/s	trem	sala/s de aula	psicólogo/a
10	joelho	peixe	vestido	autocarro/s	quadro	cantor

Como se pode observar nas tabelas acima, o grupo A1 concentra-se sobretudo no uso de vocábulos centrais e básicos, enquanto os grupos A2 e B1 não só mantêm um uso estável destes vocábulos, como também conseguem recorrer a um maior número de vocábulos periféricos ou de uso menos frequente, resultando numa produção lexical mais rica. No entanto, esta expansão do léxico mental não se manifesta de forma uniforme, apresentando variações entre diferentes centros de interesse.

Especificamente, no domínio "partes do corpo", os vocábulos *cabeça*, *olho/s* e *mão* apresentam índices de disponibilidade relativamente elevados nos três grupos (A1: 0,53/0,47/0,57; A2: 0,64/0,61/0,45; B1: 0,62/0,47/0,54). Tal implica que estes vocábulos centrais são consistentes na disponibilidade entre os diferentes níveis de proficiência, refletindo que foram solidamente aprendidos nas fases iniciais de aprendizagem e ocupam posições centrais no léxico mental dos aprendentes.

Por outro lado, vocábulos como *boca* e *nariz* apresentam um aumento dos índices de disponibilidade à medida que o nível de proficiência linguística se eleva. Nomeadamente, *boca* tem um índice de disponibilidade de apenas 0,14 (8.º lugar) no grupo A1, subindo para 0,53 (3.º lugar) no A2 e mantendo-se em 0,40 (5.º lugar) no B1; *nariz* passa de 0,12 (12.º lugar) no A1 para 0,49 (4.º lugar) no A2. Estes dados sugerem que, embora a estabilidade dos vocábulos centrais e de alta frequência se mantenha, o aumento do nível de proficiência promove progressivamente a utilização de vocábulos periféricos ou de baixa frequência, enriquecendo a rede do léxico mental dos aprendentes.

Tal como acontece nos centros de interesse "partes do corpo" e "alimentos e bebidas" os aprendentes apresentam, independentemente do nível de proficiência, não apenas uma ativação dos vocábulos centrais e de uso relativamente frequente, como também uma incorporação de vocábulos periféricos ou de baixa frequência. Alguns vocábulos básicos mantêm uma posição central altamente consistente no léxico mental dos aprendentes entre os diferentes grupos (como *cabeça*, *olho/s*; *pão*, *arroz*), enquanto outros apresentam variações na frequência de utilização à medida que o nível de proficiência aumenta.

Mais concretamente, os vocábulos relativos a alimentos básicos, como *pão* e *arroz*, mantêm-se sempre entre os três primeiros, com índices de disponibilidade relativamente elevados em todos os três grupos (0,59 e 0,54 para A1; 0,45 e 0,46 para A2; 0,68 e 0,42 para B1). Os dados sugerem que estes vocábulos já se consolidam precocemente no léxico mental dos aprendentes e mantêm uma vantagem de ativação constante ao longo da aprendizagem. Este fenómeno, muito provavelmente, é influenciado pelo contexto cultural, dado que "arroz" pertence ao vocabulário de alta frequência na vida quotidiana dos aprendentes chineses e desempenha um papel importante na cultura alimentar da China.

Uma semelhante influência também se manifesta nos vocábulos relativos a bebidas. Por exemplo, *café* regista no grupo A1 um índice de disponibilidade de apenas 0,14 (16.º lugar), mas aumenta para 0,26 no grupo A2 (10.º lugar) e atinge 0,51 no grupo B1 (2.º

lugar). *Chá* passa de 0,04 em A1 (39.º lugar) para 0,29 em A2 (6.º lugar), mantendo-se em 0,30 no grupo B1 (6.º lugar). O crescimento indica que, com a progressão do nível de aprendizagem, estes vocábulos vão sendo gradualmente incorporados no vocabulário central dos aprendentes, alargando o vocabulário de alimentos básicos para incluir vocábulos com maior relevância cultural e comunicativa.

Adicionalmente, há outros vocábulos cuja ativação varia conforme o nível de proficiência. Por exemplo, *maçã/s* regista 0,11 em A1 (23.º lugar), sobe para 0,32 em A2 (4.º lugar), mas desce para 0,21 em B1 (12.º lugar). Por sua vez, *ovo/s* mantém-se estável entre os dez primeiros em A1 e A2 (0,23 e 0,25, respetivamente), mas cai significativamente em B1 para 0,14 (19.º lugar). Outro caso de variação considerável é *água*, cujo índice de disponibilidade é elevado em A1 (4.º lugar, 0,38), cai para o 17.º lugar em A2 (0,13) e volta a subir para o 9.º lugar em B1 (0,23).

Estas variações relevam que, à medida que o vocabulário se expande, alguns vocábulos periféricos ou mais facilmente substituíveis podem perder proeminência. Mesmo sendo vocábulos de uso quotidiano, a invocação pode ser influenciada pelo contexto de exposição e pela competição semântica, não mantendo necessariamente posição estável no léxico mental dos aprendentes.

No centro de interesse "roupa", os três grupos mantêm um desempenho estável relativamente a *camisa* e *calça/s*, tanto em termos de ordenação como de índice de disponibilidade lexical. Em particular, *camisa* ocupa o 1.º lugar em todos os grupos (A1: 0,73; A2: 0,65; B1: 0,62), evidenciando o seu estatuto de vocábulo central representativo deste campo semântico. *Calça/s* mantém-se consistentemente entre os três primeiros, no 3.º lugar no grupo A1 (0,42) e no 2.º lugar nos grupos A2 e B1 (0,60/0,57). Os dados demonstram que estes vocábulos continuam a ser ativamente mobilizados em diferentes níveis de proficiência.

Em contraste, alguns vocábulos apresentam variações com o aumento do nível de proficiência. Por exemplo, *casaco* regista um índice de disponibilidade elevado no grupo A1 (0,59, 2.º lugar), mas desce no grupo A2 para 0,38 (6.º lugar) e no B1 para 0,32 (7.º lugar), indicando que o seu estatuto como vocábulo central e mais disponível tende a ser substituído por outros vocábulos mais frequentes e genéricos, como *roupa*. De forma semelhante, *meia/s* atinge 0,41 (5.º lugar) no grupo A2, mas cai significativamente para 0,10 (20.º lugar) no B1. Esta variação deve-se, possivelmente, ao facto de serem vocábulos de baixa frequência ou termos que se referem a objetos menos visíveis na vida quotidiana.

Por outro lado, alguns vocábulos periféricos aumentam a sua disponibilidade nos níveis relativamente mais elevados. Por exemplo, *chapéu* apresenta um índice de 0,25 (10.º lugar) no A2 e aumenta significativamente para 0,39 (6.º lugar) no B1. Este incremento poderá estar associado ao facto de, à medida que a competência linguística se desenvolve, os aprendentes terem mais contacto com contextos culturalmente relacionados com o vestuário, tornando estes vocábulos mais disponíveis.

No centro de interesse "meios de transporte", a estabilidade dos vocábulos centrais e de alta frequência é igualmente evidente, destacando-se especialmente *carro/s*. Este vocábulo apresenta um índice de disponibilidade lexical de 1,00 no grupo A1, indicando que todos os participantes de nível A1 o ativaram em primeiro lugar neste campo

semântico. Para além disso, nos grupos A2 e B1 o índice mantém-se elevado (0,76 e 0,87, respetivamente), continuando a ocupar a primeira posição, o que evidencia o seu estatuto de vocábulo central muito representativo neste campo semântico, independentemente do nível de proficiência.

Simultaneamente, alguns vocábulos apresentam variações significativas entre os diferentes níveis. Por exemplo, *ónibus* apresenta um índice de 0,37 no grupo A1 (2.º lugar), desce para 0,35 no grupo A2 (5.º lugar) e, embora o índice aumente para 0,42 no B1, a ordenação cai para o 7.º lugar. Estas variações indicam que a sua frequência de utilização se mantém relativamente estável, mas a centralidade dentro do campo semântico diminui gradualmente.

Em contraste, *táxi* apresenta um índice de 0,24 no grupo A1 (4.º lugar), sobe para 0,32 no grupo A2 (8.º lugar) e continua a aumentar no B1 para 0,45 (4.º lugar). O incremento constante deste vocábulo poderá dever-se ao facto de os aprendentes com níveis relativamente mais elevados terem acesso a contextos mais variados, como, por exemplo, expressões relacionadas com o turismo ou com a vida quotidiana, reforçando assim a competência de mobilização mais ativa deste vocábulo.

No centro de interesse "escola", a produção lexical nos três níveis de proficiência privilegia vocábulos referentes a objetos concretos, como *cadeira/s*e *mesa*, cujos índices de disponibilidade se mantêm relativamente elevados em todos os grupos (A1: 0,37/0,23; A2: 0,62/0,58; B1: 0,56/0,43). Importa ainda referir que estes vocábulos estão diretamente relacionados com o ambiente de sala de aula, o que poderá explicar a sua posição central no léxico mental dos aprendentes.

Para além disso, *escola* apresenta um incremento considerável no índice de disponibilidade. Nomeadamente, apresenta apenas 0,15 (17.º lugar) no grupo A1, sobe para 0,25 (6.º lugar) em A2 e atinge 0,43 (2.º lugar) em B1. Este resultado poderá refletir que a rede semântica do léxico mental dos aprendentes está alargada de termos concretos para vocábulos mais conceptuais.

Pelo contrário, determinados vocábulos apresentam uma diminuição significativa da disponibilidade em diferentes níveis de proficiência. Por exemplo, *carteira* apresenta o índice mais elevado no grupo A1 (0,39, 1.º lugar), mas desce para 0,19 (9.º lugar) no grupo A2 e para 0,11 (22.º lugar) no grupo B1. Como referido anteriormente, esta variação poderá estar relacionada com a competição semântica associada a este vocábulo. Isto é, à medida que a competência linguística se desenvolve, os aprendentes começam a contactar o sentido de *carteira* como *bolsa de mão*, estabelece-se uma competição semântica no léxico mental, afetando a sua frequência de mobilização no campo semântico "escola".

No centro de interesse "profissões", *professor(es)/a(s)*, *médico(s)/a(s)* e *advogado(s)/a(s)* apresentam elevada estabilidade tanto na ordenação como no índice de disponibilidade lexical. Nomeadamente, *professor(es)/a(s)* regista índices de 0,75 (A1), 0,76 (A2) e 0,81 (B1), mantendo-se sempre no 1.º lugar. *Médico(s)/a(s)* também ocupa consistentemente o 2.º lugar, com índices de 0,56 (A1), 0,65 (A2) e 0,65 (B1). Em termos de *advogado(s)/a(s)*, mantém-se uma posição igualmente estável, com uma diminuição muito ligeira do 3.º lugar nos A1 (0,43) e A2 (0,45) para o 4.º lugar no B1 (0,36). Por

conseguinte, estes dados demonstram, em conjunto, o estatuto destes vocábulos centrais representativos deste campo semântico.

Ao mesmo tempo, verifica-se ainda que outros vocábulos são mais disponíveis à medida que o nível de proficiência se eleva. Por exemplo, *bombeiro/s* apresenta um índice de disponibilidade de apenas 0,12 no A1 (12.º lugar), sobe para 0,22 no A2 (8.º lugar) e aumenta significativamente para 0,44 no B1 (3.º lugar). Este incremento contínuo pode dever-se ao facto de, com a progressão do nível linguístico, os aprendentes irem ampliando, no seu léxico mental, o repertório vocabular relacionado com os vocábulos relativos a profissões.

Por outro lado, alguns vocábulos exibem variações de mobilização entre diferentes níveis. Por exemplo, o índice de disponibilidade de *polícia* é de apenas 0,08 no A1 (25.º lugar), sobe para 0,36 no A2 (5.º lugar), mas desce para 0,21 no B1 (12.º lugar). De forma semelhante, *dentista* ocupa o 4.º lugar no A1 (0,39), desce para o 7.º lugar no A2 (0,26) e sobe ligeiramente para o 5.º lugar no B1 (0,32). De facto, estas variações resultem, muito provavelmente, da influência da frequência de exposição linguística e da experiência de vida dos aprendentes. Dito de outra forma, alguns vocábulos, apesar de já terem sido aprendidos, podem apresentar uma diminuição na frequência de evocação quando a exposição da língua não é suficiente ou pouco constante.

Conforme a análise acima, concluímos que o desenvolvimento da competência linguística promove principalmente a expansão do léxico mental dos aprendentes e a estabilidade na utilização dos vocábulos centrais. Exerce, porém, um efeito limitado sobre a diversidade lexical e a ativação de vocábulos periféricos, podendo isto estar relacionado com o contexto de uso, do ambiente de aprendizagem e da experiência de vida dos aprendentes.

4.3.3. Análise comparativa com aprendentes de outras línguas maternas

Com base no exposto na secção 2, embora a LM não tenha sido considerada como variável no questionário, é possível realizar uma comparação com os dados provindos da investigação de Fernández dos Santos & Silva Soares (2024), o que permite, até certo ponto, inferir a possível influência da LM na mobilização lexical efetuada pelos participantes. Nesse sentido, recorremos aos dados de 20 participantes com nível de proficiência A2 deste estudo e realizaremos uma análise comparativa com os dados de aprendentes de LM espanhola e inglesa apresentados por Fernández dos Santos & Silva Soares (2024)[4]. Os centros de interesse considerados são "partes do corpo", "alimentos e bebidas", "roupa" e "escola"[5] (ver Tabelas 10, 11 e 12).

[4] Convém referir que, no estudo de Fernández dos Santos & Silva Soares (2024), os informantes não se limitavam a falantes nativos de espanhol e inglês, incluindo também falantes de outras línguas maternas, como o francês, o alemão, o mandarim, entre outras. Contudo, visto que os grupos de falantes de outras línguas maternas eram relativamente pequenos, não foram discutidos na análise de disponibilidade do estudo delas e, por isso, também não foram incluídos na nossa análise comparativa.

[5] Em relação a outros dois centros de interesse, como não foram apresentados dados relativos a palavras mais disponíveis no estudo de Fernández dos Santos & Silva Soares (2024), não nos é possível fazer uma comparação aqui com os nossos dados recolhidos.

Tabela 10. Os 10 vocábulos mais disponíveis em cada centro de interesse em aprendentes A2 (Grupo de LM chinesa: 20 participantes)[6]

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Escola
1	cabeça	arroz	camisa	cadeira/s
2	olho/s	pão	calça/s	mesa
3	boca	batata	saia	caneta
4	nariz	maçã/s	roupa/s	livro/s
5	cabelo	leite	meia/s	lápis
6	mão	chá	casaco	escola
7	pé/s	comida	blusa	sala/s de aula
8	rosto	peixe	sapato/s	cama
9	língua	carne	vestido	carteira
10	dente	café	chapéu	universidade

Tabela 11. Os 10 vocábulos mais disponíveis em cada centro de interesse em aprendentes A2 (Grupo de LM espanhola, Fernández dos Santos & Silva Soares, 2024)

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Escola
1	olho/s	água	sapatos	professor/a/es
2	cabeça	leite	calças	caneta/s
3	perna/s	pão	t’shirt	aula/s
4	mão/s	vinho	camisa	quadro
5	braço/s	batata	camisola	estudante/s
6	cabelo	feijão	vestido	mesa/s
7	pé/s	café	casaco	cadeira/s
8	boca	cerveja	saia	livro/s
9	nariz	portoferreira	casaca	computador
10	cara	bolodearroz	cuecas	ensino

Tabela 12. Os 10 vocábulos mais disponíveis em cada centro de interesse em aprendentes A2 (Grupo de LM inglesa, Fernández dos Santos & Silva Soares, 2024)

Ordem por dispon.	Partes do corpo	Alimentos e bebidas	Roupa	Escola
1	cabeça	francesinha	sapatos	aula/s
2	olho/s	água	camisa	professor/a/es
3	pé/s	pão	vestido	cadeira/s
4	boca	cerveja	camisola	estudante/s
5	mão/s	vinho	meias	exame/s
6	coração	batata	t’shirt	turma/s
7	nariz	café	casaco	escrever
8	orelha/s	sumo	roupainterior	livro/s
9	braço/s	jantar	calças	ensinar

[6] Os dados apresentados na Tabela 10 são extraídos da Tabela 8.

10	dedo/s	almoço	calçasdedemin	mesa/s

Conforme se pode observar nas tabelas acima, constatamos que os diferentes grupos de aprendentes, independentemente da sua LM, apresentam de forma semelhante um conjunto de vocábulos centrais em cada campo semântico. Por exemplo, no centro de interesse "partes do corpo", *cabeça* e *olho/s* ocupam consistentemente os dois primeiros lugares em todos os grupos (no grupo de LM chinesa e inglesa classificam-se consistentemente no 1.º e 2.º lugar; no grupo de LM espanhola no 2.º e 1.º lugar). Nos centros de interesse "alimentos e bebidas", *pão* encontra-se sempre nos três primeiros (2.º lugar no grupo chinês; 3.º lugar nos grupos espanhol e inglês). No centro de interesse "roupa", *camisa* mantém uma posição elevada em todos os grupos (1.º lugar no grupo chinês, 4.º no espanhol e 2.º no inglês).

Para além disso, verificam-se outros vocábulos disponíveis comuns nos três grupos de LM. Especificamente, no centro de interesse "partes do corpo" incluem-se *mão/s*, *boca*, *nariz*, *pé/s*, em "alimentos e bebidas", *batata* e *café*; em "roupa", *camisa*, *calça/s*, *sapatos* e *vestido*; e em "escola", *cadeira/s*, *mesa* e *livro/s*. Estes resultados indicam que, no nível A2, os vocábulos centrais dos aprendentes, relativos a partes do corpo, objetos básicos relativos ao vestuário e materiais escolares, são relativamente estáveis e apresentam consistência entre LMs diferentes.

Constatamos também que as diferenças culturais e de contexto de uso da língua influenciam significativamente a seleção do vocabulário. No campo semântico "alimentos e bebidas", os aprendentes de LM chinesa produzem com maior disponibilidade vocábulos básicos como *arroz*, *pão* e *batata* (os três primeiros). Comparativamente, os aprendentes de LM espanhola e inglesa produzem vocabulário relativo a bebidas ou alimentos regionais, como *vinho* nos grupos de espanhol (4.º lugar) e inglês (5.º lugar), e *francesinha* no grupo de língua inglesa (1.º lugar). Esta divergência poderá refletir a influência do contexto linguístico e da cultura alimentar na mobilização lexical. Paralelamente, importa reconhecer que tal poderá estar relacionado com o contexto de aprendizagem do português, uma vez que os participantes deste estudo utilizam maioritariamente o português em contexto de sala de aula, diferendo do seu uso em ambientes de imersão (por exemplo, com experiência de aprendizagem em Portugal e no Brasil).

No centro de interesse "escola", para os aprendentes de LM chinesa, os vocábulos mais disponíveis são *cadeira/s*, *mesa*, *caneta* e *livro/s*. Por sua vez, nos grupos de LM espanhola e inglesa surgem com maior disponibilidade vocábulos como *professor/a/es*, *aula/s* e *estudante/s*. Esta diferença poderá dever-se ao facto de os aprendentes de LM chinesa dependerem maioritariamente dos materiais didáticos ou do contexto de ensino formal. Em contraste, os falantes nativos de espanhol ou inglês poderão ter maior exposição a atividades comunicativas tanto em sala de aula como fora dela[7]. Dito de outra forma, poderão ter experiência de aprendizagem em ambientes de imersão, o que facilita

[7] No estudo de Fernández dos Santos & Silva Soares (2024), os informantes eram estudantes de PLE da Faculdade de Letras da Universidade do Porto, ou seja, ficavam num ambiente de imersão em Portugal.

a ativação de vocábulos ligados à interação social e de uso mais frequente nas atividades comunicativas.

Adicionalmente, no centro de interesse "roupa", entre os dez vocábulos com maior índice de disponibilidade, o grupo de LM chinesa inclui *blusa* e *chapéu*, o grupo de LM espanhola contém *cuecas*, e o grupo de LM inglesa apresenta *roupa interior*. Estas diferenças indicam que a mobilização de vocábulos periféricos apresenta características diferenciadas, possivelmente relacionadas com a experiência de vida, a exposição a contextos específicos, bem como o próprio grau de formalidade ou informalidade da LM dos aprendentes.

Em síntese, os aprendentes de diferentes LMs apresentam certa uniformidade na mobilização dos vocábulos centrais que surgem frequentemente na vida quotidiana, como, por exemplo, vocábulos alimentares básicos (*arroz*, *pão* e *batata*). Ao mesmo tempo, a LM exerce uma influência significativa sobre a utilização de vocabulário culturalmente marcado, tais como *vinho* e *francesinha* para aprendentes de LM espanhola e inglesa. Por outro lado, é de salientar que, o ambiente de aprendizagem também influencia a mobilização dos vocábulos. Especificamente, num ambiente de imersão, os aprendentes recorrerão mais frequentemente a vocábulos respeitantes à interação social ou conceptuais (*professor*, *estudante/s; aula/s*). No caso de falta deste, dependem principalmente de materiais didáticos e privilegiarão mais vocábulos referentes a objetos concretos. Deste modo, a experiência cultural e a perceção do contexto respeitantes à LM poderão exercer um efeito modulador na mobilização do léxico mental dos aprendentes.

5. Conclusões

Neste artigo procedemos uma análise pontual da disponibilidade lexical em aprendentes chineses de PLE, considerando o tempo de aprendizagem e diferentes níveis de proficiência linguística, em seis centros de interesse: "partes do corpo", "alimentos e bebidas", "roupa", "meios de transporte", "escola" e "profissões". Adicionalmente, procedeu-se a uma análise comparativa entre os dados de aprendentes de LM chinesa ao nível A2 e os dados de Fernández dos Santos e Silva Soares (2024) relativos a aprendentes com LM espanhola e inglesa, com o intuito de explorar a possível influência da LM na disponibilidade lexical.

Verificámos que o tempo de aprendizagem exerce, na maioria dos casos, um efeito positivo na expansão do léxico mental e na diversidade lexical. Quanto maior o tempo de aprendizagem, a utilização de vocábulos periféricos e de baixa frequência torna-se mais ativa, e, simultaneamente, a rede semântica do léxico mental releva-se mais complexa. Todavia, este efeito varia em função de outros fatores. Por exemplo, quando os contextos de uso dos vocábulos são relativamente limitados ou quando a exposição destes vocábulos nas fases iniciais de aprendizagem já foi significativa, esta influência positiva poderá não ser muito significante.

Para além deste fator, o aumento do nível de proficiência contribui igualmente para a expansão do léxico mental e, simultaneamente, para a maior estabilidade na ativação dos vocábulos centrais e de alta frequência, como *cabeça*, *pão*, *arroz*, *carro/s*, *camisa*, *cadeira/s*, *professor(es)/a(s)*, entre outros. Mesmo que os aprendentes de níveis mais

avançados consigam mobilizar mais palavras, o efeito sobre a mobilização ativa de vocábulos periféricos ou de baixa frequência, tais como *casaco*, *meia/s*, é limitado. Este fenómeno poderá estar relacionado com fatores como o ambiente de aprendizagem, os contextos de uso e a experiência de vida do próprio aprendente.

Importa, por fim, referir que a LM exerce um efeito significativo na mobilização de vocabulário periférico e de vocábulos culturalmente específicos. Os aprendentes com LM chinesa tendem a privilegiar vocábulos relativos a alimentos básicos (*arroz*, *pão* e *batata*), enquanto os aprendentes de LM espanhola ou inglesa utilizam com maior frequência vocábulos de bebidas ou de pratos/alimentos regionais (*vinho* e *francesinha*). Para além disso, os aprendentes chineses concentram-se em vocábulos relativos a objetos concretos, enquanto os aprendentes de LM espanhola e inglesa recorrem mais frequentemente a vocábulos inerentes à interação social ou conceituais (*professor*, *estudante/s*, *aula/s*), evidenciando que a experiência cultural e a perceção do contexto influenciam a mobilização lexical.

Estas conclusões poderão fornecer referências empíricas para o ensino de vocabulário em PLE, recomendando que os docentes considerem, de forma integrada, a consolidação dos vocábulos centrais, a expansão dos periféricos e os fatores culturais da LM, para otimizar o desenvolvimento do léxico mental dos aprendentes. É importante referir que, sendo este um estudo de natureza exploratória, pode ter algumas limitações. As futuras investigações poderão incluir dados relativos a níveis de proficiência mais avançados (B2, C1 e C2), permitindo uma análise mais abrangente em relação à disponibilidade lexical em aprendentes chineses de PLE.

Financiamento: Esta pesquisa foi financiada pela Universidade de Estudos Internacionais de Sichuan, sob a concessão número SISUWYJY202313 (projeto de investigação científica de 2023 da disciplina de excelência de Línguas e Culturas Estrangeiras do Município de Chongqing).

Disponibilização de dados: Os dados que suportam os resultados deste estudo estão disponíveis mediante solicitação ao autor correspondente (e-mail: liumengru@sisu.edu.cn).

Referências

Ávila Muñoz, A. M. (2016). El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos. *Journal of Spanish Language Teaching*, *3*(1), 31–43. https://doi.org/10.1080/23247797.2016.1163038

Biderman, M. T. C. (1996). Léxico e vocabulário fundamental. *Alfa: Revista de Linguística (São Paulo)*, *40*, 27–46. https://periodicos.fclar.unesp.br/alfa/article/view/3994/3664

Borrego Nieto, J., & Fernández Juncal, C. (2002). Léxico disponible: aplicaciones a los estudios dialectales. Em M. D. Muñoz Núñez et al. (Eds.), *IV Congreso de Lingüística General* (pp. 297–306). Universidad de Cádiz y Universidad de Alcalá.

Calero Fernández, M. Á. (2025). Léxico disponible frente a competencia activa en estudiantes de ELE de origen chino. Caplletra. *Revista Internacional De Filologia*, *78*, 5–31. https://doi.org/10.7203/caplletra.78.30263

Chaves, R. (2013). Organização do léxico. In E. P. Raposo, et al. (Orgs.), *Gramática do português* (Vol. 1, pp. 185–212). Fundação Calouste Gulbenkian.

Fernández dos Santos, M., & Silva Soares, C. (2024). Disponibilidade lexical em estudantes de português como língua estrangeira: um estudo pioneiro. *Lenguas Modernas*, *64*(2), 103–119. https://hdl.handle.net/10216/164594

Ferreira, R., & Echeverría, M. S. (2010). Redes semánticas en el léxico disponible de inglés L1 e inglés LE. *Onomázein*, *21*(1), 133–153. https://doi.org/10.7764/onomazein.21.05

Germany, P., & Cartes, N. (2000). Léxico disponible en inglés como segunda lengua en instrucción formalizada. *Estudios Pedagógicos (Valdivia)*, *26*(1), 39–50. http://dx.doi.org/10.4067/S0718-07052000000100003

Hernández Muñoz, N., & Tomé Cornejo, C. (2017). Léxico disponible en primera y segunda lengua: bases cognitivas. Em F. del Barrio de la Rosa (Ed.), *Palabras vocabulario léxico: la lexicología aplicada a la didáctica ya la diacronía* (pp. 99–122). Edizioni Ca' Foscari.

Hernández Muñoz, N., Tomé Cornejo, C., López García, M., & Bartol Hernández, J. A. (2023). *Manual de LexPro*. Universidad de Salamanca. https://dispogram.usal.es/

Kartsevski, M., & Blanco, O. (2021). Léxico disponible en aprendientes con un nivel básico de inglés como segunda lengua: un estudio escolar chileno. *Literatura y lingüística*, *44*, 291–313.

Michéa, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les langues modernes*, *47*(4), 338–344.

Palapanidi, K., & Mavrou, I. (2024). La naturaleza gramatical del léxico disponible en diferentes niveles lingüísticos en español como lengua extranjera. *Revista signos*, *57*(114), 151–167. http://dx.doi.org/10.4067/S0718-09342024000100151

Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera.* Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE).

Samper Padilla, J. A. (1998). Criterios de edición del léxico disponible: sugerencias. *Lingüística*, *10*, 311–333. https://dialnet.unirioja.es/servlet/articulo?codigo=1400825

Sánchez-Saus Laserna, M. M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera de las universidades andaluzas*. Editorial Universidad de Sevilla.

Santos Díaz, I. C. (2017). Selección del léxico disponible: propuesta metodológica con fines didácticos. *Porta Linguarum*, *27*, 122–139. https://digibug.ugr.es/handle/10481/53965

Santos Díaz, I. C., Trigo Ibáñez, E., & Romero Oliva, M. F. (2020). La activación del léxico disponible y su aplicación a la enseñanza de lenguas. *Porta Linguarum, 33*, 75–93. http://hdl.handle.net/10481/62807

Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE* [Tese de doutoramento, Universidad de Salamanca]. https://hdl.handle.net/10366/128287

Trigo Ibáñez, E., Dono López, P., & Fernández Gómez, I. V. (2025). Aproximación al

léxico disponible en ELE de 42 estudiantes de la Universidade do Minho. Em A. M. Ávila Muñoz & I. C. Santos Díaz (Eds.), *Avances y desarrollo de los estudios sobre el léxico disponible. Una aproximación desde la variación léxica* (pp.167–194). Peter Lang. https://doi.org/10.3726/b22163

Van Parys, A., De Wilde, V., Macken, L., & Montero Perez, M. (2024). LexPro: a plurilingual lexical profiling tool for research and materials development. Em *EuroCALL 2024 Trnava, Abstracts* (pp. 97–98). EuroCALL.

[received in September 26 and accepted for publication in January 6 2026]