

DOI: <http://dx.doi.org/10.1590/1982-4017-170205-5616>

## FREQUÊNCIA DE PALAVRAS E PROCESSOS INFERENCIAIS: O QUE APRENDIZES DE UMA L2 PODEM DIZER SOBRE ISSO

Alessandra Baldo\*

Tássia Ávila Silva\*\*

Universidade Federal de Pelotas  
Centro de Letras e Comunicação  
Pelotas, RS, Brasil

**Resumo:** Este artigo apresenta os resultados de um estudo cujos objetivos foram verificar possíveis relações entre a frequência de palavras e os processos de inferência lexical de aprendizes de inglês como língua estrangeira (L2), também entre a frequência das palavras e os graus de inferências apropriadas de tais palavras. Foram analisados protocolos verbais de 15 sujeitos de nível intermediário de inglês durante a realização de uma atividade de leitura na qual lhes era solicitado inferir o significado de 21 palavras, divididas entre baixa, média e alta frequências. Constatou-se que as estratégias com maior número de ocorrências permaneceram invariáveis ao longo das três categorias de frequências das palavras e que, na direção contrária da hipótese inicial, maior percentual de inferências apropriadas foi obtido nas palavras de baixa frequência, em comparação com as de alta frequência. Os resultados são interpretados com base na literatura vigente.

**Palavras-chave:** Frequência de palavras. Inferência lexical. Aquisição de vocabulário em L2.

### 1 INTRODUÇÃO

A ampliação do léxico é uma das principais habilidades a serem desenvolvidas no processo de aquisição de uma língua estrangeira (L2), sendo fundamental tanto na compreensão quanto na produção oral e escrita. É, ao mesmo tempo, das mais complexas: enquanto os processos de apresentação e, em etapa imediatamente posterior, de apropriação das palavras que compõem a língua materna (L1) de um indivíduo ocorrem, na maioria das vezes, de forma tão natural que podem ser considerados em grande proporção automáticos, isso está bem distante de ser verdadeiro para aprendizes de uma L2. Tendo em mente o ambiente formal de ensino, fatores como o início tardio da aquisição de uma L2, a restrição de insumo e a artificialidade do ambiente de ensino-aprendizagem são desafios que se impõem ao objetivo maior de qualquer aprendiz de uma L2, ou seja, apropriar-se dessa nova língua para poder comunicar-se de forma eficiente, tanto pelo falar/ouvir como pelo escrever/ler.

---

\* Professora-associada da área de Inglês. Doutorado em Letras pela Pontifícia Universidade Católica do Rio Grande do Sul, RS. E-mail: [alessabaldo@gmail.com](mailto:alessabaldo@gmail.com).

\*\* Especialista em Letras – área de concentração Linguística Aplicada – pela Universidade Federal de Pelotas. E-mail: [tassia\\_avila@hotmail.com](mailto:tassia_avila@hotmail.com).

Esses fatores, naturalmente, valem para toda e qualquer habilidade em uma língua estrangeira. Contudo, considerando, em primeiro lugar, que o léxico perpassa todas essas habilidades, e, em segundo, que os fatores envolvidos na aquisição de um único item lexical são de natureza variada, conforme será visto adiante, parece apropriado afirmar que tais fatores têm um peso ainda maior quando o que se almeja é o desenvolvimento do léxico em uma L2.

Este artigo focaliza, então, um aspecto da aquisição lexical, a relação entre a frequência de palavras e os processos inferenciais de aprendizes de uma L2. Mais especificamente, o texto relata os resultados de uma pesquisa que investigou em que medida as frequências das palavras inferidas impactaram (i) nas estratégias inferenciais e (ii) nos escores de inferência lexical de quinze aprendizes de proficiência intermediária em inglês como L2, com base em protocolos verbais obtidos durante a aplicação de tarefas de atribuição de significado a palavras de diferentes classes gramaticais e frequências.

Ainda que a frequência das palavras tenha sido tema de um número considerável de estudos, as relações mais usualmente analisadas são relacionadas ao tempo de reconhecimento e a diferenças nas definições entre palavras de alta e de baixa frequência (LOHNAS; KHANA, 2003; MALMBERG; NELSON, 2003; MARINELLIE; JOHNSON, 2003; MARINELLIE; CHAN, 2006). Nesse contexto, este estudo busca contribuir para ampliar o escopo das relações passíveis de serem avaliadas em estudos sobre frequências de palavras.

Iniciamos o artigo com uma breve retomada de questões relativas ao conceito de palavra e com uma listagem de fatores relacionados à aquisição das palavras em uma L2, seguindo-se com uma classificação dos vocábulos em relação à sua frequência. O conceito de inferência lexical, juntamente com a definição de estratégias inferenciais, é resgatado na segunda parte da revisão dos conceitos teóricos, a fim de facilitar a compreensão do tema sob estudo.

Continuamos com a apresentação da metodologia e a análise dos dados obtidos, finalizando com reflexões relativas aos resultados advindos do estudo, à luz de pesquisas sobre léxico, estratégias inferenciais e frequência de palavras em L2.

## 2 A PALAVRA

### 2.1 DEFINIÇÃO DE PALAVRA

A concepção usual de palavra como a menor unidade de linguagem com conteúdo semântico enquanto suficiente para abordar determinadas questões linguísticas é problematizada por estudiosos do léxico, sob argumentos de natureza diversa. A título de ilustração, apresentaremos, a seguir, as ponderações de Murphy (2010) e Hanks (2014) com relação à complexidade dos conceitos que o termo *palavra* pode abarcar.

Murphy (2010, p. 5) argumenta que, no campo da semântica lexical, o termo *palavra* deve ser substituído pela expressão abstrata *lexema*, uma vez que nem todas as palavras são lexemas, e nem todos os lexemas são palavras. A natureza abstrata dos lexemas é explicada pela relação entre fones e fonemas: da mesma maneira que os primeiros são a concretização dos últimos, os lexemas são a concretização de cada instância de uso daquele lexema, chamadas de unidades lexicais e associadas com uma gama de significados.

Para a autora, uma forma linguística representa um lexema sempre que for associada com um significado não-composicional, ou seja, sempre que os significados das formas dos lexemas não forem construídos pelos significados de suas partes. Os lexemas de uma língua incluem, assim, palavras simples, ou morfemas livres, morfemas presos, como *in-* em *infeliz*, palavras morfologicamente complexas, como *ponto-agulha*, e expressões idiomáticas, como em “carregar o mundo nas costas” e “sair do controle.” (MURPHY, 2010, p. 5-6).

Além da ambiguidade entre lexema e unidade lexical, Murphy acrescenta outra dificuldade na definição do termo *palavra*, que é o fato de a noção poder ser acessada por uma variedade de formas, incluindo a ortográfica, a semântica, a fonológica e a gramatical. Expandindo a análise sobre cada uma delas, a autora afirma que a definição ortográfica tem como ponto de partida os espaços entre palavras para marcar seu início e seu fim. Porém, o problema em determinar uma palavra por sua grafia é que já devemos ter em mente o que essa significa antes mesmo de sabermos onde colocar os espaços na escrita. Já a definição semântica relaciona a palavra com o conceito completo de algo. Contudo, indaga a autora, como poderia alguém afirmar o que é um significado completo? Por exemplo, no caso da palavra *policeman* (policial, em inglês), um falante poderia entendê-la como dois conceitos – ser policial (*police*) e ser homem (*man*) –, enquanto outro poderia interpretar como um único conceito, envolvendo o estado de ser um policial do sexo masculino. Em relação à fonologia da palavra, Murphy (2010, p. 12) destaca que seria provavelmente a abordagem menos eficiente, considerando-se que uma “palavra fonológica é uma unidade da língua sujeita às restrições fonotáticas baseadas na palavra particular e nos processos fonológicos”, e testes fonológicos são específicos de cada língua. Por fim, a autora explica que uma palavra gramatical é identificada por suas inflexões morfológicas e sua distribuição na frase.

Igualmente complexa é a discussão sobre a definição de *palavra* proposta por Hanks (2013). Para o autor, o termo é ambíguo em pelo menos cinco acepções: *token*, tipo (*type*), lema, expressões com mais de uma palavra e entrada lexical. Inicialmente, ele apresenta a distinção entre *token* e tipo, o primeiro significando qualquer e cada ocorrência de “encadeamentos de letras entre dois espaços ou sinais de pontuação”, e o segundo, um único encadeamento de letras, desconsiderando a frequência de uso (HANKS, 2013, p. 26). Assim, o termo *token* faz referência ao número total de palavras em um texto, independentemente da frequência com que são repetidas, enquanto tipo faz referência ao número total de palavras distintas em um texto.

Em seguida, define lema como “uma coleção de todas as formas flexionadas do que é, conceptualmente, uma palavra” (HANKS, 2013, p. 26). O autor ilustra o conceito através do verbo *nadar*, que consiste da forma-base *nadar* e de todas as outras formas flexionadas, como *nadas*, *nadando*, *nadado*, cada uma delas sendo de um diferente tipo. Já o conjunto de nomes “nado” e “nados” (forma flexionada) são em geral considerados como lemas separados. Nesse contexto, o autor cita o sistema de classificação de palavras de dicionários clássicos da língua inglesa para mostrar que, em alguns deles, diferentes classes gramaticais do mesmo lema estão agrupadas na mesma entrada lexical, enquanto estão separadas em outros.

Por fim, Hanks (2013) também observa que dicionários monolíngues tipicamente incluem itens lexicais com múltiplas palavras sob a entrada de uma palavra simples. Desse modo, conclui, o termo palavra pode, de fato, denotar os conceitos de *token*, tipo, lema, expressão lexical ou entrada lexical, dependendo da abordagem adotada.

## 2.2 FATORES ENVOLVIDOS NO CONHECIMENTO DE UMA PALAVRA EM L2

O acesso a palavras depende do conhecimento do *vocabulário* de uma língua. Read (2009, p. 17) defende que o domínio do vocabulário por estudantes de L2 abarca bem mais do que o conhecimento de itens lexicais – o aprendiz deve ser capaz de acessar o conhecimento lexical e de utilizá-lo eficientemente ao comunicar-se na segunda língua. Segundo o autor, conhecer uma palavra significa conhecer seu valor semântico, os diversos significados associados a ela, a rede de associações entre a palavra e outras palavras na língua, o seu comportamento sintático, a sua forma-base e as derivações que podem ser feitas a partir desta, o grau de ocorrência da palavra na fala ou escrita, como também as limitações de uso de acordo com variações de função e situação (READ, 2009, p. 25).

Essa multiplicidade de fatores envolvida na aquisição das palavras de uma L2 tem relação direta com as noções de *vocabulário receptivo* e *vocabulário produtivo*, partindo da premissa de que o vocabulário que um falante adquire é maior do que o vocabulário que é efetivamente “produzido” por ele. Melka (1997, p. 85) busca explicar a distância entre a recepção e a produção de vocabulário de um aprendiz afirmando que existem graus de conhecimento, os quais são infinitos e imperceptíveis. Esses graus começam com o conhecimento mais básico da palavra, quando, por exemplo, o seu reconhecimento visual ocorre em um contexto. Neste ponto, é difícil notar se a palavra já está armazenada no léxico mental. Se estiver, é possível que esteja de forma incompleta, de modo que não permite ao sujeito reproduzir ou produzir a palavra.

Certos graus de conhecimento podem ser estabelecidos como sendo de alta familiaridade, ou seja, próximo ao conhecimento produtivo. Tanto em L1 quanto em uma L2, esse conhecimento engloba os vários significados que uma palavra pode ter. Possuir conhecimento fonológico, morfológico e sintático do léxico corresponde, em princípio, a um alto grau de familiaridade. Melka (1997) salienta a importância que teria estabelecer em que ponto a familiaridade poderia dizer que o conhecimento já não seria mais receptivo e poderia ser convertido em conhecimento produtivo. Além disso, como o processo de reconhecimento se manifesta e como uma palavra é acessada no léxico mental são alguns dos pontos que a autora busca esclarecer. Neste sentido, parece que apenas após muitas ocorrências uma palavra pode ser considerada como fazendo parte do léxico mental de um aprendiz. Assim, a distância entre vocabulário receptivo e vocabulário produtivo não é fixa ou permanente, e as fronteiras entre estas duas noções podem mover-se de acordo com fatores linguísticos ou extralinguísticos.

### 2.3 FREQUÊNCIA DE PALAVRAS E CORPUS ELETRÔNICO

Entre os fatores linguísticos que devem ser considerados no contínuo entre vocabulário receptivo e vocabulário produtivo, a frequência das palavras certamente ocupa destaque, especialmente quando consideramos que a frequência é o número de vezes em que as palavras pertencentes ao léxico de uma língua ocorrem dentro de faixas específicas. De acordo com este número de ocorrências, encontram-se *palavras de alta, média e baixa frequências*. Tais palavras podem se diferenciar de acordo com as diferentes faixas etárias, língua, cultura e escolarização (STIVANIN; SCHEUER, 2007, p. 206).

Nesse contexto, as palavras de alta frequência, como se infere pelo nome, são as habitualmente empregadas na linguagem do dia a dia, constituindo-se de um grande número de vocábulos gramaticais e de palavras de significado muito geral, como verbos e nomes de uso comum. Na realidade, a maior parte das cem palavras mais frequentes da língua inglesa são artigos, preposições, pronomes, verbos auxiliares e formas do verbo *to be* (ser/estar), como também conjunções. Já as palavras de baixa frequência, conseqüentemente, são as que têm uso bastante restrito pelos falantes de uma determinada língua; dentre as classes gramaticais principais desse grupo constam nomes, verbos, adjetivos e advérbios.

A classificação das palavras em faixas de frequência, no que diz respeito à língua inglesa, teve início na década de 50 do século passado, com a organização da *West's General Service* (MICHAEL, 1953) e o *corpus Brown* (KUCERA; FRANCIS, 1963). Atualmente é possível encontrar diversos *corpora* online, sendo que entre os mais recentes está o *Corpus of Contemporary American English* (COCA), com cerca de 460 milhões de palavras do inglês americano.<sup>1</sup>

Entre outros tantos, um dos benefícios dos *corpora* eletrônicos para o ensino-aprendizagem de línguas é o de, ao estabelecer com precisão as frequências das palavras de uma língua em contextos definidos, abrir caminho para a compreensão de fenômenos como, por exemplo, o paradoxo da frequência das palavras (LOHNAS; KAHAMA, 2003; MALMBERG; NELSON, 2003) e a influência da frequência das palavras na definição de vocábulos (REDER, 2000; MARINELLIE; JOHNSON, 2003; MARINELLIE; CHAN, 2006).

### 3 INFERÊNCIA LEXICAL: CONCEITOS E ESTRATÉGIAS

A inferência lexical pode ser concebida como um processo de atribuição de significado a palavras desconhecidas, seja na L1 ou em uma L2, concretizado pela aplicação de diferentes recursos cognitivos. Oxford (1990, p. 47) a define como "o uso de uma variedade de indícios linguísticos e não-linguísticos para adivinhar o significado de todas as palavras que o aprendiz não conhece".

---

<sup>1</sup> Como será visto na seção sobre a metodologia, o COCA foi o *corpus* utilizado no estudo aqui descrito para a classificação das frequências das 21 palavras analisadas.



Em conformidade com essa definição, Pulido (2007, p. 66) destaca a complexidade da realização de inferências lexicais em L2: além da integração de indícios contextuais com o conhecimento linguístico e extralinguístico a fim de iniciar o processo de inferência, é necessário também que o leitor atente para as conexões entre as novas formas lexicais e seus significados, e integre a nova informação linguística ao significado global do texto que está em construção.<sup>2</sup> Na seção dedicada à revisão dos estudos sobre o tema, a autora nomeia os tipos de conhecimento empregados na elaboração da inferência lexical. Entre os conhecimentos linguísticos, cita, por exemplo, o uso da informação morfológica para obter indícios sobre a classe da palavra, os componentes de significado da palavra, as funções gramaticais e os papéis semânticos.

Observadas tanto a importância como a complexidade da inferência lexical, um número considerável de estudos tem se voltado a definir como esse processo ocorre (HAASTRUP, 1991; PARIBAKHT; WESLSCH, 1999; NASSAJI, 2003; 2006; COMER, 2012). Ainda que os objetivos específicos desses estudos diverjam do apresentado neste texto, há dois critérios metodológicos em comum na grande maioria deles: (i) a necessidade de definição e classificação das estratégias de inferência lexical, na medida em que é a partir da análise das estratégias que o processo de inferência pode ser avaliado, e (ii) o uso de protocolos verbais como o instrumento mais adequado para a obtenção dos dados. Por fazerem parte da metodologia do trabalho de pesquisa aqui relatado, seguimos com a descrição da aplicação desses critérios em dois estudos, Nassaji (2003) e Paribakht e Welsche (1999).

Nassaji<sup>3</sup> (2003) acrescentou a noção de fontes de conhecimento ao conceito de estratégias inferenciais em seu estudo sobre inferência lexical em L2. Para o autor, trata-se de uma distinção necessária na medida em que entende estratégias como atividades cognitivas ou metacognitivas conscientes que o leitor emprega “para obter o controle de um problema, ou resolvê-lo, sem qualquer apelo explícito a qualquer fonte de conhecimento como forma de auxílio”, as quais podem ser relativas ao conhecimento gramatical, morfológico, discursivo, de mundo ou da L1 (NASSAJI, 2003, p. 655). Desse modo, sua classificação prevê seis tipos de estratégias de inferência lexical – repetição, verificação, analogia, monitoramento, autoquestionamento e analogia – e cinco tipos de fontes de conhecimento – gramatical, morfológico, discursivo, de mundo e da L1.

Valendo-se de uma classificação menos detalhada, Paribakht e Welsche<sup>4</sup> (1999) trabalham com dois grupos de estratégias de inferência lexical, extralinguísticas e linguísticas. Nas estratégias extralinguísticas os autores incluem todo e qualquer tipo de conhecimento prévio; nas linguísticas, constam conhecimento gramatical no nível da sentença, morfologia da palavra, pontuação, conhecimento textual/discurso, relações de homonímia, associações entre palavras de qualquer natureza, como sinonímia e antonímia, e entre palavras cognatas.

---

<sup>2</sup> O estudo de Pulido (2007) analisa as inferências lexicais em leitura. De qualquer modo, entendemos que a inferência lexical no discurso oral ocorre basicamente da mesma forma, reservadas as particularidades de cada modalidade discursiva.

<sup>3</sup> O autor investigou o nível de sucesso inferencial de 21 estudantes de nível intermediário em inglês como L2 a partir da leitura de um texto, estabelecendo relações entre tipos de estratégias, fontes de conhecimento e eficácia das inferências lexicais empregadas pelos aprendizes.

<sup>4</sup> Os autores analisaram os tipos de conhecimento – linguístico e extralinguístico – empregados na inferência lexical por aprendizes de inglês como L2.

Com relação ao instrumento de coleta de dados, ambos os estudos empregaram protocolos verbais, ou “protocolos de pensar em voz alta”, os quais também foram utilizados no estudo aqui descrito. A técnica consiste em solicitar que indivíduos, após orientação específica, verbalizem o seu pensamento durante o processamento de uma tarefa específica. Sendo um instrumento de análise introspectivo, os protocolos verbais têm uma importância ímpar para estudos que investigam processos cognitivos na medida em que, ao permitirem acesso a um traço direto da informação em foco, possibilitam evidência indireta dos estágios internos do processo cognitivo (ERICSON; SIMON, 1993, p. 220). Em uma revisão da literatura sobre os protocolos verbais, Afflerbach (2000, p. 166) observou que as contribuições que eles têm oferecido podem ser percebidas tanto em estudos visando a um único processo, como análise da ideia principal do texto, geração de inferências, predição dos conteúdos e monitoramento cognitivo, com em estudos visando a vários processos de forma conjunta, como, por exemplo, a totalidade de uma tarefa de leitura.

#### 4 METODOLOGIA

Os instrumentos da pesquisa foram o *corpus* obtido de pesquisa de Baldo et al. (2016), sobre inferência lexical por aprendizes de língua inglesa como L2, e o *corpus* eletrônico COCA – *Contemporary Corpus of American English*, desenvolvido por Mark Davies, da Universidade de Brigham Young.

Como dados auxiliares, também foram utilizados os resultados obtidos no estudo de Baldo et al. (2016) no que se refere ao número de estratégias de inferência lexical mais comumente empregados pelos participantes, como também aos escores nas tarefas de inferência lexical. O *corpus* do estudo é constituído por protocolos verbais de quinze aprendizes adultos de língua inglesa como L2, de nível intermediário de proficiência, graduandos do curso de Licenciatura em Letras, habilitação em Português e Inglês, de uma Universidade do Rio Grande do Sul. A tarefa dos participantes consistia na leitura de dois textos com palavras desconhecidas – o primeiro com 10 palavras, e o segundo, com 11, como apresentado na Tabela 1 a seguir. Um deles consistia na parte inicial do romance *1984*, de George Orwell, em língua inglesa, e o segundo, na parte inicial de um conto do escritor australiano Arthur Davies, intitulado *Dad and the Donovans*. Ambos eram versões adaptadas das obras para aprendizes de nível intermediário da série “Readers”, e os trechos selecionados possuíam em média mil palavras cada um. Após a leitura silenciosa, solicitava-se aos participantes que buscassem inferir o significado das palavras selecionadas por meio da técnica dos protocolos verbais (conforme seção 3), demonstrada no início da sessão. Os protocolos eram gravados em áudios, para posterior transcrição e análise das estratégias.

**Tabela 1 – Palavras a serem inferidas**

<i>Texto 1</i>	<i>Texto 2</i>
<i>Fonte: Romance</i>	<i>Fonte: Conto</i>
<i>Blew</i> (Soprou)	<i>Curl</i> (murchar)
<i>Overalls</i> (macacão – uniforme)	<i>Wipe</i> (enxugar)
<i>Fair</i> (loiro)	<i>Bucket</i> (balde)
<i>Rough</i> (áspero)	<i>Curse</i> (amaldiçoar)
<i>razor blades</i> (lâminas de barbear)	<i>Unwillingly</i> (a contragosto)
<i>Figures</i> (cifras)	<i>Crops</i> (plantações)
<i>to stare</i> (olhar fixamente)	<i>Enquire</i> (questionar)
<i>Surroundings</i> (arredores)	Receipts (recibos)
<i>Canteen</i> (cantina)	<i>Grin</i> (sorrir)
<i>to pour</i> (servir-se)	<i>Paddock</i> (cercado)
_____	<i>Echoed</i> (ecoou)

Para a classificação das estratégias de inferência lexical, avaliamos inicialmente a proposta de Nassaji (2003), e, com base nos dados dos protocolos verbais produzidos pelos participantes do estudo, propusemos uma classificação com um maior número de estratégias, mas, por outro lado, sem a diferenciação entre estratégias e fontes de conhecimento proposta pelo autor.

A Tabela 2 descreve os doze tipos de estratégias de inferência lexical identificados em nosso estudo: releitura, repetição, uso do contexto, análise gramatical, análise sintático-semântica, análise morfossemântica, analogia com forma, analogia semântica, analogia com L1, conhecimento extratextual, monitoramento e recuperação automática. Entre eles, cabe notar que a última deles, “recuperação automática”, não é uma estratégia no mesmo sentido das demais. Sua função no estudo foi indicar os casos em que o sujeito já conhecia a palavra, não tendo recorrido, portanto, a processos inferenciais para encontrar o significado correspondente da palavra.

**Tabela 2 – Classificação das Estratégias**

<i>N.</i>	<i>Estratégia</i>	<i>Definição</i>
1	Releitura	Releitura da frase ou parágrafo no qual a palavra ocorre.
2	Repetição	Repetição da palavra e/ou frase na qual a palavra ocorre.
3	Uso do contexto	Uso dos indícios textuais próximos à palavra a ser inferida.
4	Análise gramatical	Uso do conhecimento das funções gramaticais ou categorias sintáticas como verbo, adjetivos ou advérbios.



5	Análise morfossemântica	Tentativa de descobrir o significado da palavra pela análise de suas partes ou componentes.
6	Análise sintático-semântica	Tentativa de descobrir o significado da palavra/expressão com base na posição em que ocupa na frase.
7	Analogia com forma	Inferência com base na similaridade de som ou forma com outras palavras, ou palavras da mesma família com morfemas distintos.
8	Analogia semântica	Tentativa de transpor o significado conhecido da palavra na L2 – entrada lexical 1 - para um contexto em que este não se aplica – entrada lexical 2.
9	Analogia com L1	Inferência a partir de palavra similar na L1.
10	Conhecimento extratextual	Inferência através do conhecimento adquirido extratextualmente.
11	Monitoramento	Exibição de consciência do problema ou da facilidade da tarefa.
12	Recuperação automática	Lembrança do significado da palavra a partir de recuperação automática das informações textuais (e/ou do conhecimento de mundo).

A pesquisa relativa à frequência das palavras na língua inglesa realizada no COCA teve os seguintes parâmetros: foram selecionadas apenas as ocorrências das palavras em linguagem escrita, excluindo-se registros orais, entre os anos de 2000 e 2012. No campo da linguagem escrita, foram consideradas somente a linguagem acadêmica, de ficção, e de jornais e revistas. A restrição a esses contextos justifica-se pelo fato de os textos em que estavam as palavras a serem inferidas pelos participantes pertencerem aos gêneros literários romance e conto, respectivamente, o que motivou os parâmetros empregados na busca – ou seja, linguagem escrita em contextos formais de uso. Os resultados das buscas realizadas no *Corpus* podem ser visualizados nos Gráficos 1 e 2, apresentados na seção seguinte.

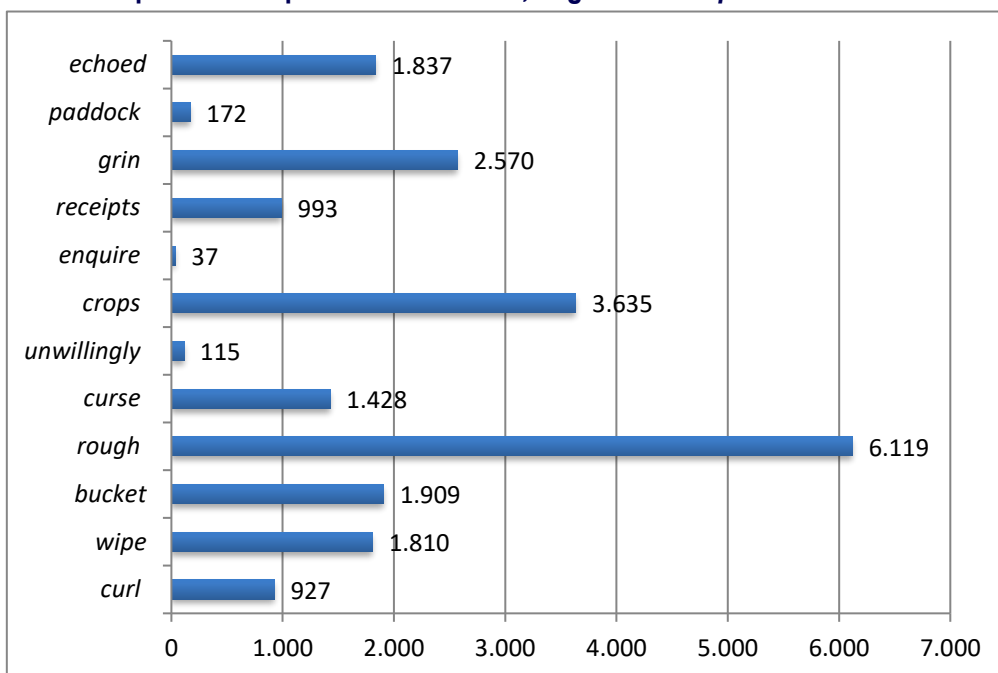
## 5 APRESENTAÇÃO E ANÁLISE DOS DADOS

Esta seção tem três partes. Primeiramente descrevemos o resultado da classificação das 21 palavras entre as três faixas de frequência estipuladas pelo estudo. Em seguida, analisamos as relações verificadas entre a frequência das palavras e o uso das estratégias de inferência lexical. Na parte final da seção, avaliamos a relação entre os escores de inferência lexical e a frequência das palavras.

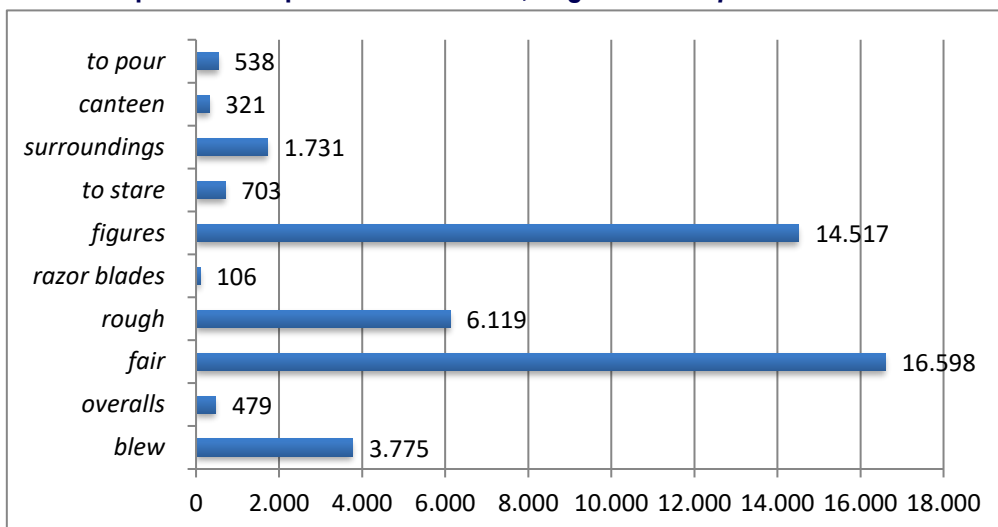
## 5.1 FREQUÊNCIA DAS PALAVRAS

Foram consideradas de baixa frequência as palavras com menos de 800 ocorrências; de média frequência, as palavras entre 801 e 3.000 ocorrências; e de alta frequência, as palavras que apresentaram mais de 3.000 ocorrências, conforme apresentado nos Gráficos 1 e 2. Assim, de um total de 21 palavras, 5 foram consideradas de alta frequência, contabilizando 24% do total, 8 foram classificadas como de média frequência, com 38%, e as oito restantes, como de baixa frequência, representando, igualmente, 38% do total. A Tabela 2 apresenta as palavras divididas por níveis de frequência.

**Gráfico 1 – Frequência das palavras do texto 1, segundo o corpus COCA**



**Gráfico 2 – Frequência das palavras do texto 2, segundo o corpus COCA**



As estratégias de inferência lexical empregadas pelos sujeitos em suas tentativas de atribuição de significado às novas palavras encontram-se na Tabela 3. Para fins de análise, foram desconsideradas as estratégias com menos do que 5% do total de ocorrências possíveis,<sup>5</sup> o que foi obtido multiplicando-se o número de palavras com o número de sujeitos, conforme pode ser visualizado na Tabela 3.

**Tabela 3 – Divisão das palavras em alta, média e baixa frequências**

<i>Alta freq.</i>	<i>Freq.</i>		<i>Média freq.</i>	<i>Freq.</i>		<i>Baixa freq.</i>	<i>Freq.</i>
<i>Fair</i>	16.598		<i>Grin</i>	2.570		<i>to stare</i>	703
<i>Figures</i>	14.517		<i>Bucket</i>	1.909		<i>to pour</i>	538
<i>Rough</i>	6.119		<i>Echoed</i>	1.837		<i>Overalls</i>	479
<i>Blew</i>	3.775		<i>Wipe</i>	1.810		<i>Canteen</i>	321
<i>Crops</i>	3.635		<i>Surroundings</i>	1.731		<i>Paddock</i>	172
			<i>Curse</i>	1.428		<i>Unwillingly</i>	115
			<i>Receipts</i>	993		<i>Razor blades</i>	106
			<i>Curl</i>	927		<i>Enquire</i>	37

## 5.2 RELAÇÃO ENTRE FREQUÊNCIA DE PALAVRAS E USO DE ESTRATÉGIAS DE INFERÊNCIA LEXICAL

Os percentuais de uso das estratégias inferenciais com relação à classificação das palavras por sua frequência – baixa, média e alta – estão apresentados na Tabela 4. Entre parênteses está a posição de cada estratégia em relação às demais, seguida pelo número de ocorrências, na coluna seguinte.

O primeiro dado que se pode observar é que, independentemente da frequência da palavra, as estratégias de inferência lexical mais empregadas foram uso do contexto, em primeiro lugar, e releitura, em segundo. Tantos nas palavras de baixa, como nas de média e de alta frequências, essas estratégias foram empregadas de modo bastante significativo, cabendo notar, contudo, as variações dos percentuais entre as faixas de frequência de palavras. O menor número de ocorrências das duas estratégias citadas se deu nas palavras de frequência média, quando comparadas com as de alta e de baixa frequência: 82 e 77 ocorrências, correspondentes a 68,3% e 64,2% do total de uso possível, respectivamente. Já os vocábulos de alta frequência, e diferentemente do que era esperado, foram os que apresentaram o maior número de uso das estratégias, ou seja, 73 e 72, correspondentes a 97,3% e 96% do total de usos possíveis para cada estratégia.

<sup>5</sup> A exceção foi o uso da recuperação automática para as palavras de alta frequência, incluída porque o uso desse recurso era igual ou superior a 5% para as palavras de baixa e média frequências.

Tabela 4 – Percentuais, Posição e Ocorrências de Estratégias Utilizadas

Estratégias	Palavras Baixa Frequência		Palavras Média Frequência		Palavras Alta Frequência	
Uso do contexto	81,6% (1 <sup>a</sup> )	98	68,3% (1 <sup>a</sup> )	82	97,3% (1 <sup>a</sup> )	73
Releitura	73,3% (2 <sup>a</sup> )	88	64,2% (2 <sup>a</sup> )	77	96% (2 <sup>a</sup> )	72
Análise gramatical	19,1% (3 <sup>a</sup> )	23	15,8% (4 <sup>a</sup> )	19	18,6% (4 <sup>a</sup> )	14
Analogia com forma	15,8% (4 <sup>a</sup> )	19	15% (3 <sup>a</sup> )	18	12%	9
Monitoramento	6,6%	8	7,5%	9	22,6% (5 <sup>a</sup> )	17
Verificação	14,1%	17	13,3% (5 <sup>a</sup> )	16	26,6% (3 <sup>a</sup> )	20
Recuperação automática	5%	6	5,8%	7	2,6%	2
Analogia com L1	15% (5 <sup>a</sup> )	18	12,5%	15	10,6%	8
Total de estratégias		120		120		75
N. palavras		8		8		5
N. sujeitos		15		15		15

O maior percentual das estratégias de uso do contexto e releitura nos vocábulos de alta frequência, em comparação com os de baixa e de média frequência, pode ser explicado por características particulares presentes em três das cinco palavras de alta frequência analisadas – *fair* (loiro), *figures* (cifras) e *crops* (plantações) –, como múltiplas entradas lexicais semelhantes às da L1, pela forma, e ausência de pistas contextuais.

A questão das múltiplas entradas lexicais está diretamente relacionada à palavra *fair*, frequentemente, associada com o significado de “justo”, entrada lexical mais conhecida, e bem menos como sinônimo de “loiro”, entrada que deveria ser acionada no contexto. Assim, os participantes liam a palavra utilizando seu significado mais frequente, mas não conseguiam encontrar sentido na leitura realizada, o que trazia como consequência a necessidade de reler a frase e tentar se apropriar do contexto para formular uma nova inferência relativa ao vocábulo – ou seja, a necessidade de fazer uso das estratégias de releitura e de uso do contexto.

*Figures* também apresenta mais de uma entrada lexical, e uma delas foi frequentemente acionada pelos participantes na medida em que possuía quase a mesma forma que a palavra correspondente em português – ou seja, *figures* como pessoa importante – ainda que o significado apropriado da palavra, no texto, fosse “cifras”. Diferentemente do que ocorreu durante a tentativa de inferência da palavra *fair*, quando o significado de “justo” era acessado e em seguida rechaçado, o significado de “pessoa

importante” era acessado e aceito, ainda que não fizesse sentido na frase. Entendemos que essa aceitação por parte dos sujeitos se deveu em parte à proximidade formal das palavras da L1 e da L2, o que levou os participantes a terem mais convicção, ao formularem sua inferência.

Como ilustração desses fatos, há os exemplos 1 e 2, relativos aos protocolos dos sujeitos 13 e 9, respectivamente. As intervenções do entrevistador estão em itálico, e são iniciadas com a letra “E”.

Exemplo 1

Palavra: *fair* (loiro)

Eu sei o significado, é justo, né. Alguma coisa assim, *fair*.

*E: Tá. E você acha que encaixa o significado de justo aí, nesse contexto?*

Complicado. “*His hair was fair*”. Justo não seria a melhor tradução.

*E: Sim.*

Pra esse contexto aqui, né.

*E: Isso.*

*His hair was fair. I think it's kind of... The hair was kind of okay, or no problems with that.*<sup>6</sup>

*E: Entendo. E essa ideia de okay, você faz alguma relação com a ideia de fair ser justo?*

Isso, assim. O cabelo tá *okay*, assim, sem problemas nenhum com ele, assim. Não sei se tá certo.

Exemplo 2

Palavra: *figures* (cifras)

Pessoas que se destacaram na produção de ferro no último ano.

*E: Figures aí então está como pessoas, as pessoas que se destacaram?*

Figuras. Isso, figuras que se destacaram no último ano na produção de ferro. Na produção de ferro no último ano.

Com relação à palavra *crops* (plantações), o fato que melhor parece explicar o uso frequente das estratégias de releitura e uso de contexto é a ausência de pistas contextuais que pudessem guiar os sujeitos em seus processos inferenciais. Conforme Laufer (1997, p. 28), um dos obstáculos para a formulação de inferências bem-sucedidas sobre palavras da L2 está na ausência de pistas contextuais, e uma ilustração desse fato pode ser constatada no protocolo do sujeito 8.

Exemplo 3

Palavra: *crops* (plantações)

Vou para a sétima palavra, *crops*. Vou ler no texto, é um substantivo, mas essa palavra não é conhecida pra mim, não sei o significado, nem pelo contexto eu consegui identificar. (sujeito volta ao texto). É, é desconhecida pra mim e o texto acho que não traz assim muitas informações pra ajudar.

<sup>6</sup> *His hair was fair* (participante relê frase do texto em inglês). Eu acho que é tipo...seu cabelo era tipo OK, ou sem problemas com o cabelo dele. (tradução nossa)

Outro dado que vale notar, ainda sobre as palavras de alta frequência, é o do uso da verificação como terceira estratégia mais utilizada, com 26,6% do total de usos possíveis da estratégia sendo registrado, e o monitoramento como a quarta mais frequentemente empregada, com 22,6% do total. O uso significativo dessas duas estratégias, em nosso entender, está de acordo com a observação feita anteriormente sobre as peculiaridades de três das cinco palavras de alta frequência inferidas pelos sujeitos da pesquisa. Considerando que duas delas tinham mais de uma entrada lexical e que uma delas não apresentava dados contextuais suficientes para desencadear o processo inferencial, tanto a verificação como o monitoramento parecem ser escolhas naturais.

Voltando-nos agora às palavras de média e de baixa frequência, observa-se pela Tabela 4 que as estratégias mais frequentemente empregadas foram analogia com forma e análise gramatical (15 e 15,8%, respectivamente) nas palavras de média frequência, enquanto que, para as de baixa frequência, ocorreu o inverso: a terceira estratégia mais usada foi análise gramatical (19,1%), seguida de analogia formal (15,8%). É interessante observar que são as mesmas estratégias que ocupam o terceiro e quarto lugares tanto para as palavras de média como para as de baixa frequência, e que ambas dizem respeito aos aspectos estruturais e formais das palavras, e não aos semânticos. Uma explicação para esse resultado pode estar no fato de que, por se tratar de palavras menos frequentes – e, portanto, com as quais os participantes estariam menos familiarizados – era somente esse tipo de estratégia que os sujeitos conseguiam empregar, devido a seu caráter mais geral, de estabelecimento de relações e analogias com aspectos relacionados mais ao conhecimento da língua estrangeira, enquanto sistema do que com aspectos relacionados, especificamente, ao léxico da L2.

### 5.3 RELAÇÃO ENTRE FREQUÊNCIA DAS PALAVRAS E ESCORES DE INFERÊNCIA LEXICAL

A fim de verificar uma possível relação entre a natureza das palavras com relação à sua frequência e aos escores de inferência lexical obtidos pelos participantes, as respostas dos participantes relativas ao significado das novas palavras foram classificadas em cinco categorias: (i) apropriada, (ii) inapropriada, (iii) parcialmente apropriada, (iv) não realização da inferência, e (v) conhecimento prévio da palavra. As inferências lexicais foram consideradas apropriadas quando contemplavam o significado da palavra em sua acepção completa; parcialmente apropriadas, quando o significado era contemplado em determinados aspectos, mas não em outros; e inapropriadas, quando o significado inferido diferia por completo do significado da palavra.<sup>7</sup> A quarta categoria foi utilizada para os casos em que os sujeitos não conseguiam atribuir qualquer significado à palavra, enquanto a última dizia respeito às palavras conhecidas pelos participantes previamente à atividade de leitura.

---

<sup>7</sup> Ilustramos o critério de classificação das inferências consideradas parcialmente apropriadas através da palavra “*overalls*” (macacão): definições “vestimenta” ou “peça de roupa” foram consideradas parcialmente apropriadas por contemplarem o aspecto mais geral da palavra, mas não totalmente apropriadas, por não darem conta do aspecto mais específico, ou seja, o tipo de vestimenta.



A Tabela 5 apresenta os resultados dessa classificação, com os percentuais sendo apresentados primeiramente, seguidos do número de ocorrências para cada caso, e, no final, do número total de inferências verificado para as palavras de baixa, média e alta frequências.

**Tabela 5 – Inferências Versus Frequência das Palavras**

<i>Inferências</i>	<i>Palavras de Baixa Frequência</i>		<i>Palavras de Média Frequência</i>		<i>Palavras de Alta Frequência</i>	
Inferências Apropriadas	41,7%	48	22,2%	37	16,8%	16
Inferências Inapropriadas	30,4%	35	22,8%	38	46,3%	44
Inferências Parcial. Apropriadas	1,7%	2	1,8%	3	1,8%	3
Não realização de inferências	20,8%	24	16,2%	27	31,5%	30
Conhecimento prévio da palavra	5,2%	6	4,2%	7	2,1%	2
Total de inferências		115		166		95

Ao observarmos a Tabela 5, verificamos as inferências realizadas pelos sujeitos da pesquisa de acordo com a frequência das palavras. As palavras de baixa frequência apresentam predominantemente 41,7% de as inferências apropriadas e 30% de as inferências inapropriadas, enquanto 20,8% dos sujeitos não realizam qualquer inferência. As palavras de média frequência apresentam 22,2% de as inferências apropriadas e 22,8% de as inapropriadas, enquanto 16% dos sujeitos não realizam inferências. Já para as palavras de alta frequência, apenas 16,8% dos sujeitos realizaram a inferência apropriada, enquanto 46,3% dos sujeitos fizeram uma inferência inapropriada e 31,5% não realizaram inferência alguma.

Podemos perceber que as palavras de baixa frequência tiveram maior índice de inferências apropriadas do que as palavras de média e alta frequência. Este resultado merece atenção na medida em que, pelo menos teoricamente, se esperaria um maior número de inferências apropriadas nas palavras de alta frequência, com as quais, teoricamente, os sujeitos estariam mais familiarizados.

No entanto, uma leitura cuidadosa das respostas dos participantes é suficiente para mostrar que esse achado tem coerência, e pode ser explicado pelas mesmas razões levantadas para explicar o padrão de uso das estratégias de inferência lexical, ou seja, (i) o fato de que os sujeitos, com frequência, conheciam apenas uma das entradas lexicais das palavras *fair* e *figures*, (ii) a semelhança na forma da palavra *figures* com a L1, e (iii) a ausência de pistas contextuais para a inferência da palavra *crops*. Como ilustração, apresentamos o exemplo 4, do sujeito 6, que mostra a inferência da palavra *fair* com base

no significado de “justo”, e não de “loiro”, o que influenciou negativamente o resultado, e os exemplos 5 e 6, que mostram o conhecimento de apenas uma entrada lexical da palavra *figures*, ou seja, “figuras”, dos protocolos verbais dos Sujeitos 2 e 5, respectivamente.

Exemplo 4

Palavra: *fair* (loiro)

(Sujeito relê) *His... his hair was fair....* “curto”. Não sei, pra mim eu vejo “curto”.

*E: Fair como “curto”... sim...*

É, assim, meio curto. Não longo, nem curto demais nem longo demais. Aquela... aquela coisa, quase na medida.

*E: Entendo. E por que essa associação com o fair, como “curto”...*

Eu pensei em... tipo, “*it isn't fair*”, tipo que “não é justo”, então é uma coisa que não... não fica adequada... né... então um cabelo que não ... tipo “como assim?” Não, então deve ser uma coisa... envolvendo o cabelo... um cabelo... fair... hum... tipo, quase na medida. Não sei, ... nem curto nem comprido...

*E: Sim, está bem. A associação com fair, aqui, é no sentido de justo.*

De justo.

Exemplo 5

Palavra: *figures* (cifras)

Acredito que sejam “figuras” mesmo.

*E.: Sim. E te parece que fecha, assim, quando você lê... Quando você leu pela primeira vez, era isso? Isso te chamou a atenção...*

Deixa eu ver. Sim.

*E: Ok. São figures...*

Sim.

*E: Tá bem. E você já conhecia essa palavra? Ou é a primeira vez que vê?*

Já, já tinha visto, sim, lido em outros contextos.

*E: E sempre foi com o significado de “figuras”?*

Sim.

Exemplo 6

Talvez imagens ou...coisas, eu conheço a palavra, mas não bate o significado.

*E.: Qual é o significado?*

Eu conheço como imagens, coisas, formas, figuras, dentro do apartamento flat dele, a voz estava lendo. Ah, uma lista de imagens. Estava lendo uma lista de imagens do último ano. Eu acho que são imagens.

*E.: Imagens. Sim.*

Nesse caso, eu acho.

*E.: Você acha que esse significado fecha com a frase? Você conhece outro significado para figures que poderia fazer sentido aí também?*

Fotos? Mas seria *picture* ou *photos*. Ou de repente uma página, mas pra mim são imagens, figuras.

Outro fator que também interferiu determinando o menor percentual de acertos em relação ao significado das palavras de alta frequência foi a criação de falsos cognatos entre palavras de grafia semelhante, fato ilustrado pelos exemplos 7 e 8. No primeiro caso relatado, advindo do protocolo verbal do sujeito 9, temos uma analogia entre diferentes formas – *curse* / *course* – com escrita semelhante, a primeira significando “amaldiçoar”, e a segunda, “curso”. No segundo, o sujeito 7 entende a palavra *heat*, “calor”, que serviria como indício contextual para interpretação da palavra-chave, como *hat*, “chapéu”, o que acarretou a inferência inapropriada da palavra *curled* (*murchava*).

#### Exemplo 7

Palavra: *curse* (amaldiçoar)

*curse, curse* na linha 13 ... *curse* quando eu olhei rapidamente me lembrou *course*, de curso em português, traduzindo, a gente não deve fazer muito isso (traduzir), mas ... que ele estava dizendo pros próprios ... como é o nome dos pássaros pretos? Que a gente tem em português, é os corvos, os próprios corvos saiam sigam o seu caminho, saiam daqui. Ele estava dizendo é uma expressão, “*curse you*”, siga o seu caminho no sentido de saiam daqui.

E: *Sim, entendi, a ideia de curso mesmo*, sigam o seu curso.

Sim. O pai dele via o cavalo e dizia isso “*curse you, curse you*”, como se dissesse para os pássaros saírem e seguirem o caminho deles.

#### Exemplo 8

Palavra: *curled* (*murchava*)

Eu acho que a primeira é um verbo, mas o que significa exatamente eu não sei, mas, eu acho que é um verbo.

E: *E por que você acha que é um verbo?*

Por que tem um substantivo antes, que alguma coisa *everything* (tudo), então eu acho que por isso é um verbo e tem *-ed*, final *-ed* na palavra, né, que seria no passado no caso. Acho que é por isso.

E: *Tá, mas o contexto não te ajuda a imaginar o que poderia ser?*

Talvez, eu não sei se esta palavra (aponta para *heat*) é o que eu tô pensando.

E: *O que está pensando?*

Alguma coisa que (pensando) não sei se isto (*heat*) é chapéu. Se fosse, é que eu não lembro exatamente, seria uma coisa que abrange, cobre tudo, alguma coisa assim. Eu pensei assim, não sei se está certo.

Por outro lado, um fator que auxiliou os sujeitos a realizarem inferências apropriadas em relação às palavras de baixa frequência, em que, a princípio, se esperava um número maior de inferências incorretas, foi a utilização bem-sucedida do contexto. Nos textos lidos pelos participantes, algumas dessas palavras continham pistas contextuais que facilitavam a sua inferência (LAUFER, 1997, p. 31), conforme se pode ver nos exemplos que seguem, relativos aos protocolos dos participantes 1 e 10.

#### Exemplo 9

Palavra: *paddock* (cercado)

(Sujeito relê). Aqui, eu acho que (volta a reler em silêncio), ah, eu não sei a tradução específica, mas acho que é algum espaço nesse próprio lugar, nessa fazenda, onde tu colocas os animais à noite para eles não fugirem. Eu não sei a palavra específica, eu não sei nem em português.

*E.: E essa ideia veio do contexto? A palavra não lhe é familiar?*

Não, não é, foi pelo contexto.

Exemplo 10

Palavra 10: *unwillingly* (a contragosto)

Acho que isso significa sem querer, porque aqui fala que a mãe dele chamou porque ela queria falar com ele e aí ele entrou, e aí tipo parece que ele não queria fazer isso.

*E: E esse significado veio do contexto ou você se guiou pelas partes da palavra para chegar a isso?*

Pelo contexto, porque parece que aconteceu alguma coisa que ele não queria.

Esperamos ter conseguido mostrar, a partir dos exemplos aqui selecionados, as motivações subjacentes ao resultado da análise de dados a que se procedeu neste estudo, quanto à relação entre a frequência das palavras e o grau de inferências apropriadas a respeito do significado dessas palavras. Como foi verificado ao longo desta seção, o fato de os aprendizes apresentarem um maior número de inferências apropriadas justamente nas palavras de menor frequência pôde ser explicado por características próprias de cada uma das palavras, entre as quais se destacaram (i) as múltiplas entradas lexicais, (ii) a semelhança com a forma de palavras na L1, e as características do contexto no qual estavam inseridas, bem como (iii) a ausência de pistas contextuais. Além disso, a criação de falsos cognatos em palavras que compunham o contexto imediato no qual a palavra estava inserida também foi uma das causas de inferências inapropriadas.

Complementarmente, foi possível ilustrar um dos motivos responsáveis pelo número maior de inferências apropriadas nas palavras de baixa frequência, ou seja, a presença de pistas contextuais, conforme os últimos dois exemplos demonstraram. Explicitando melhor, o que de fato resultou em inferências apropriadas foi o emprego eficiente das pistas contextuais pelos aprendizes mais do que a simples presença das pistas contextuais, uma vez que de nada adianta haver pistas contextuais se elas não forem reconhecidas.

Como conclusão geral, parece correto afirmar, com relação às palavras sob análise neste estudo, que os escores de inferência lexical obtidos relacionaram-se mais a características particulares das palavras a serem inferidas, considerando-se tanto o contexto em que se encontravam como a língua materna dos aprendizes, e menos ao grau de frequência de uso na língua.

## 6 CONSIDERAÇÕES FINAIS

Neste texto, relatamos os resultados de um estudo que tinha por objetivo verificar possíveis variações no padrão de uso de estratégias de inferência lexical em relação à frequência das palavras, como também possíveis relações entre frequência de palavras e escores de inferência lexical dessas palavras. Verificamos, primeiramente, que as duas

estratégias inferenciais mais usadas foram contexto e releitura, tanto para palavras de baixa, quanto de média e de alta frequência. Em segundo lugar, foi possível observar que as palavras de baixa frequência obtiveram um maior percentual de inferências apropriadas, quando relacionadas às de média e de baixa frequência, tendo tal resultado sido relacionado a fatores de natureza diversa, e não especificamente à frequência de uso das palavras.

Os resultados aqui apresentados não podem ser generalizados, dado o número relativamente pequeno de participantes. Entendemos, de qualquer forma, que eles são válidos na medida em que podem motivar novos estudos, tanto de natureza qualitativa como quantitativa, sobre o mesmo tema, abrindo caminho para a ampliação do escopo das pesquisas sobre frequência de palavras, tradicionalmente investigadas com relação à influência do tempo de lembrança, nomeação e reconhecimento de palavras (REDER *et al.*, 2000; LOHNAS; KAHANA, 2003; MALMBERG; NELSON, 2003; MARINELLIE; JOHNSON, 2003; MARINELLIE; CHAN, 2006) e ao desenvolvimento da leitura e escrita (STIVANIN; SCHEUER, 2007; MONAGHAN *et al.*, 2017), e bem menos comumente com relação à investigação de padrões de inferência lexical.

## REFERÊNCIAS

- AFFLERBACH, P. Verbal reports and protocol analysis. In: KAMIL, M. K; MOSENTHAL, P. B; PEARSON, P. B.; BARR, R. (org.). *Handbook of Reading Research*. Vol. III. New Jersey: Lawrence Erlbaum Associates, p. 163-180, 2000.
- BALDO, A.; SOUZA, L.; SILVA, V. T. C. *Relatório Final de Pesquisa: Processos de Inferência Lexical em L2*. Pelotas: Universidade Federal de Pelotas, 2016.
- COMER, W. J. Lexical inferencing in reading L2 Russian. *Reading in a Foreign Language*, v. 24, n.2, p. 209-230, set. 2012.
- ERICSON, K.A; SIMON, H. *Protocol analysis: verbal report as data*. Cambridge, MA: MIT Press, 1993.
- HANKS, P. *Lexical Analysis: norms and exploitations*. Cambridge, Massachusetts: The MIT Press, 2013.
- HAASTRUP, K. *Lexical Inferencing Procedures or Talking about words: Receptive Procedures in Foreign Language Learning with Special Reference to English*. Tubingen, Germany: Gunter Narr, 1991.
- LAUFER, B. The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In: COADY, J.; HUCKIN, T. (Org.). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press, p.20-34,1997.
- LOHNAS, L. J.; KAHANA, M. J. Parametric effects of word frequency in memory for mixed frequency lists. *Memory & Cognition*, v. 29, n. 6, p. 1943-1946, 2003.
- MALMBERG, K. J.; NELSON, T. O. The word frequency effect for recognition and the elevated-attention hypothesis. *Memory & Cognition*, v. 31, n. 1, p. 31-45, 2003.
- MARINELLIE, S.A., JOHNSON, C.J. Adjective definitions and the influence of word frequency. *Speech Lang Hear Res.*, v. 46, n. 5, p. 1061-1076, 2003.
- MARINELLIE, S.A., CHAN, Y.L The effect of word frequency on noun and verb definitions: a developmental study. *Speech Lang Hear Res.*, v. 49, n. 5, p. 1001-1021, 2006.
- MELKA, F. Receptive vs. productive aspects of vocabulary. In: SCHMITT, N.; M. McCarthy, M. (Org.). *Vocabulary: Description, acquisition and pedagogy*. New York, NY: Cambridge University Press, p. 84-102, 1997.
- MONAGHAN, P; CHANG, Y.N.; WELBOURNE, S.; BRYSSBAERT, M. Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, v. 93, p. 1-21, abril 2017.
- MURPHY, M. L. *Lexical Meaning*. Cambridge, Cambridge University Press, 2010.
- NASSAJI, H. The Relationship Between Depth of Vocabulary Knowledge and L2 Learners' Lexical Inferencing Strategy Use and Success. *The Modern Language Journal*, v. 90, p. 387-401, 2006.

- \_\_\_\_\_. L2 vocabulary learning from context: strategies, knowledge sources and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, v. 27, n. 4, p. 645-669, 2003.
- OXFORD, R. L. *Language learning strategies: What every teacher should know*. New York: Newbury House, 1990.
- PARIBAKHT, T. S.; WESCHE, M. "Reading and "Incidental" L2 Vocabulary Acquisition: An Introspective Study of Lexical Inferencing". *Studies in Second Language Acquisition*, v. 21, p. 195-224, 1999.
- PULIDO, D. Effects of Topic Familiarity and Passage Sight Vocabulary on L2 Lexical Inferencing and Retention through Reading. *Applied Linguistics*, v. 28, n. 1, p. 66-86, 2007.
- READ, J. *Assessing Vocabulary*. Cambridge, Cambridge University Press, 2009.
- REDER, L. M.; NHOUYVANISVON, A.; SCHUNN, C. D.; AYERS, M. S.; ANGSTADT, P.; HIRAKI, K. A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 26, n. 2, p. 294-320, 2000.
- STIVANIN, L.; SCHEUER, C. I. Tempo de latência para a leitura: influência da frequência da palavra escrita e da escolarização. *Revista da Sociedade Brasileira de Fonoaudiologia*, v. 12, n. 3, p. 206-213, set. 2007.

**Recebido em: 21/11/16. Aprovado em: 23/05/17.**

**Title:** *Word frequency and inferential processes: what learners of an L2 can tell about it*

**Authors:** *Alessandra Baldo; Tássia Ávila Silva*

**Abstract:** *This article presents the main findings of a study whose objectives were to verify relationships between word frequency and lexical inferential processes of learners of English as an L2, as well as between word frequency and lexical inferential appropriateness levels. Verbal protocols of fifteen intermediate students of English as an L2 collected while they completed a 21-word inferential task were analyzed. The words were divided into three groups, according to their frequency status. Data showed that (i) the most often used lexical strategies remained the same along the three-word frequency levels, as well as that (ii) a larger proportion of appropriate inferences were seen in the low frequency word category, rather than in the medium and high frequency ones, as initially hypothesized. The findings are discussed in light of the specialized literature in the field.*

**Keywords:** *Word frequency. Lexical inference. L2 vocabulary acquisition.*

**Título:** *Frecuencia de palabras y procesos inferenciales: el que aprendices de una L2 pueden decir sobre ello*

**Autoras:** *Alessandra Baldo; Tássia Ávila Silva*

**Resumen:** *Este artículo presenta los resultados de un estudio cuyos objetivos fueron verificar posibles relaciones entre la frecuencia de ocurrencias de palabras y los procesos de inferencia léxica de aprendices de inglés como lengua extranjera (L2), también entre la frecuencia de palabras y los grados de inferencias apropiadas. Fueron analizados protocolos verbales de 15 sujetos de nivel intermedio de inglés durante la realización de una actividad de lectura, adónde les fue solicitado inferir el significado de 21 palabras, divididas entre baja, media y alta frecuencias. Se constató que las estrategias con mayor número de ocurrencias permanecieron invariables, independientemente de la frecuencia de las palabras, y que en la dirección contraria de la hipótesis inicial, mayor porcentaje de inferencias apropiadas fue obtenido en las palabras de baja frecuencia, en comparación con las de alta frecuencia. Los resultados son interpretados con base en la literatura vigente sobre el tema.*

**Palabras-clave:** *Frecuencia de palabras. Inferencia léxica. L2 adquisición de vocabulario.*



Este texto está licenciado com uma Licença Creative Commons Atribuição 4.0 Internacional.