

RELATÓRIO DE PESQUISA

Normas de concretude para 521 palavras do português e sua relação com aquisição da linguagem

Thayná Cristina ANANIAS 

Universidade Federal do Rio Grande do Norte (UFRN)



OPEN ACCESS

EDITADO POR

- Raquel Freitag (UFS)

AVALIADO POR

- Guilherme Duarte Garcia
(ULaval)

- Cândido Samuel Fonseca de
Oliveira (CEFET-MG)

SOBRE OS AUTORES

- Thayná Cristina Ananias
Conceptualização,
Investigação, Metodologia,
Escrita – rascunho original.

- Mahayana Cristina Godoy
Análise Formal, Metodologia,
Escrita – análise e edição.

DATAS

- Recebido: 12/04/2024

- Aceito: 13/05/2024

- Publicado: 23/05/2024

COMO CITAR

Ananias, T. C.; Godoy, M. C.
(2024). Normas de concretude
para 521 palavras do português e
sua relação com aquisição da
linguagem. *Revista da Abralín*, v.
23, n. 1, p. 1-18, 2024.

Mahayana Cristina GODOY 

Universidade Federal do Rio Grande do Norte (UFRN)

RESUMO

A concretude é uma métrica frequentemente presente em estudos normativos de diversas línguas (Bonin et al., 2003; Della Rosa et al., 2010; Soares et al. 2016). Todavia, ainda não há trabalhos em português brasileiro, que saibamos, envolvendo a relação entre concretude e idade de aquisição por classe gramatical. Nesta lacuna, então, objetivamos inserir este trabalho. O conceito de concretude adotado diz respeito a um atributo de uma palavra relacionado à emergência de uma experiência sensorial (Lima e Buratto, 2021). Realizamos um experimento on-line de coleta para as notas de concretude. Ao final, lidamos com o total de notas para 521 palavras dadas por 772 participantes. Entre as palavras tidas como mais concretas, notou-se que os substantivos são consideravelmente mais concretos em comparação com todas as outras classes. Em oposição, as menos concretas tendem a ser onomatopeias, interjeições ou palavras funcionais. Por fim, nossos resultados indicam que não parece haver uma correlação significativa entre concretude e idade de aquisição, contrariando achados do Português Europeu reportados por Soares et al. (2016).

ABSTRACT

Concreteness measurements are often present in normative studies of different languages (Bonin et al., 2003; Della Rosa et al., 2010; Soares et al. 2016). However, to the best of our knowledge there are still no studies in

Brazilian Portuguese that investigates the relationship between concreteness and age of acquisition by grammatical class. The present work fills this gap. Concreteness was primarily defined as an attribute of a word related to the emergence of a sensory experience (Lima and Buratto, 2021). We carried out an online experiment to gather concreteness scores for 521 words. These words were assessed by 772 participants who were native speakers of Brazilian Portuguese. Nouns received higher concreteness scores when compared to words from all other grammatical classes. In contrast, words rated as the less concrete tend to be onomatopoeia, interjections or functional words. Finally, we could not find any significant correlation between concreteness and age of acquisition, contradicting previous findings from European Portuguese by Soares et al. (2016).

PALAVRAS-CHAVE

Concretude. Idade de aquisição. Classe Gramatical. Português.

KEYWORDS

Concreteness. Age of acquisition. Part of speech. Portuguese.

RESUMO PARA NÃO ESPECIALISTAS

A concretude é um atributo relacionado à emergência de uma experiência sensorial a partir de uma palavra. Este atributo, na companhia de alguns outros, permite o avanço da pesquisa sobre a percepção e sobre o comportamento humano, no que tange, por exemplo, memória e linguagem. Neste trabalho, relacionamos a concretude à idade de aquisição de certas palavras do português em prol de saber se palavras mais concretas tendem – ou não – a ser adquiridas mais cedo. Para tanto, realizamos um experimento on-line de coleta da percepção de concretude por falantes nativos do português brasileiro. Ao final, lidamos com o total de notas para 521 palavras dadas por 772 participantes. Entre as palavras tidas como mais concretas, notou-se que os substantivos, como *lápiz*, *mochila* e *chinelo*, são consideravelmente mais concretos em comparação com todas as outras classes. Em oposição, as palavras menos concretas tendem a ser onomatopeias, interjeições ou palavras funcionais – como *tão*, *na* e *tanto*. Por fim, nossos resultados indicam que não parece haver uma correlação significativa entre concretude e idade de aquisição, contrariando resultados já reportados com dados do Português Europeu.

Introdução

Estudos normativos contribuem para a comunidade científica mediante o controle de aspectos que podem influenciar a variável dependente mensurada em um trabalho. Em estudos da área da Psicolinguística e da Psicologia Cognitiva, a concretude percebida de uma palavra é um dos fatores que podem afetar o seu processamento e aprendizagem (Gilhooly e Logie, 1980; Van Hell e Candia Mahn, 1997), o que tem motivado a publicação de diversos trabalhos normativos com métricas para concretude em português (e.g. Soares et al., 2016; Vasconcelos et al., 2019; Lima e Buratto, 2021) e outras línguas (e.g. Della Rosa et al., 2010).

Este artigo contribui para o corpo de estudos normativos da concretude em português, destacando sua relação com classe gramatical e métricas de aquisição da linguagem. Tanto quanto sabemos, este é o primeiro trabalho em português a averiguar a concretude de palavras para além das classes de nomes, verbos e advérbios, além de ser o primeiro em português brasileiro a correlacionar concretude com métricas de aquisição.

Na seção a seguir, levantamos um estado da arte sobre a concretude e explicitamos a lacuna em que este trabalho se insere. Nas seções seguintes, apresentamos os materiais, métodos utilizados e, para finalizar, os resultados e possíveis conclusões a partir dos dados coletados e das análises arroladas.

1. Estado da arte

Para este trabalho, estamos entendendo a concretude como um atributo de uma palavra relacionado à emergência de uma experiência sensorial (Lima e Buratto, 2021). Entre *mesa* e *justiça*, por exemplo, é certamente mais fácil materializar uma ideia que represente *mesa*, o que, conseqüentemente, a torna perceptivamente mais concreta que *justiça*.

A definição de concretude pode, à primeira vista, parecer simplória, porém, este aspecto linguístico é comumente associado a várias outras métricas implicadas na linguagem e até mesmo à subjetividade humana. Ela tem sido associada à experiência sensorial (Calais et al., 2012); emoção (Wang e Yao, 2016); familiaridade e imageabilidade (Yao et al., 2016), frequência, iconicidade (Perry, Perlman e Lupyan, 2015), etc. Entretanto, por fatores como tais não exercerem, de fato, influência sobre a percepção de concretude, sugere-se que a concretude seja um fator independente das palavras (Paivio, 1968; Nelson e Schreiber, 1992; Janczura et al., 2007; Calais et al., 2012).

No que tange o aprendizado de línguas – tanto primeira quanto segunda língua – a concretude também demonstra um papel considerável (Janczura et al., 2007):

Na aprendizagem da língua nativa observa-se que palavras concretas são adquiridas com maior facilidade (Gorman, 1961; Paivio, 1968), mais cedo e melhor do que palavras abstratas (Schwanenflugel, 1991; Van Hell & Candia Mahn, 1997), e na aprendizagem de uma segunda língua, palavras concretas são mais

fáceis de adquirir e menos prováveis de serem esquecidas do que palavras mais abstratas (de Groot & Keijzer, 2000), além da latência para traduzir palavras entre duas línguas ser também influenciada pela concretude (Habuchi, 2003).

Ao fazer um recorte para aspectos linguísticos, é comum que a concretude seja associada a métricas como imageabilidade e frequência. Tal correlação tem sido estudada há décadas, tanto em amostras de palavras do português europeu (PE) quanto do português brasileiro (PB). Apesar de não ser a variedade foco deste trabalho, em Marques (2005), há a proposição de disponibilizar a medida de imageabilidade e de concretude para 250 substantivos do PE. O autor explicita que as medidas estão positivamente correlacionadas, ou seja, quanto mais concretas, mais imagináveis. Em Soares et al. (2016), posteriormente, já há a expansão do *corpus* para 3800 palavras do PE (nomes, adjetivos, advérbios, verbos e interjeições) e das métricas coletadas, uma vez que apresenta valores normativos de imageabilidade, de concretude, de frequência subjetiva e – apesar de não parecer um foco do trabalho – idade de aquisição. Como resultado, os autores indicaram que imageabilidade e frequência não apresentam uma correlação estatística significativa; já concretude e frequência apresentam, apesar de fraca, uma correlação negativa, ou seja, quanto maior a frequência, menor a nota de concretude. E, quanto à última métrica, os autores pincelam que houve uma correlação positiva entre idade de aquisição e concretude.

Quanto aos trabalhos realizados com amostra de palavras do PB, a concretude aparece também sendo associada a algumas dessas métricas já citadas. Em Stein e Gomes (2009), por exemplo, há a apresentação de medidas normativas para coleta de diferentes métricas, como concretude, frequência e associação semântica. Tais medidas são coletadas a partir de listas temáticas com, pelo menos, 15 palavras semanticamente associadas. No total, os autores utilizaram 660 palavras do PB. Como resultado, indicaram que concretude, frequência e emotividade parecem estar positivamente correlacionadas, isto é, quanto maior a concretude, maior a frequência e maior a emotividade da palavra.

Vasconcelos et al. (2019), por sua vez, propõem um método para um teste de leitura para investigar memória de trabalho. A concretude entra como uma variável de seleção para as palavras-alvo do trabalho, dada a tendência de que palavras concretas apresentam mais chances de serem recuperadas pela memória de trabalho quando comparadas a palavras abstratas. Utilizaram-se, então, de 120 palavras já tidas como concretas do *corpus* Léxico do Português Brasileiro (Estivalet e Meunier, 2015) e selecionaram as 100 mais concretas para compor o *corpus* do experimento. O interesse de Lima e Buratto (2021), assim como Vasconcelos et al. (2019), recai no estudo da memória. No trabalho mais recente, há a coleta da nota de concretude para 80 palavras do português brasileiro com o objetivo de estimar a similaridade entre alguns pares de palavras do PB e do Swahili e investigar a precisão de memória em uma tarefa de aprendizado a partir desses pares.

Apesar de já termos, então, notas de concretude para palavras do português europeu (Marques, 2005; Soares et al., 2016) e brasileiro (Janczura et al., 2007; Stein e Gomes, 2009; Calais et al., 2012; Vasconcelos et al., 2019; Lima e Buratto, 2021), nenhum dos estudos em português brasileiro correlaciona essa métrica com idade de aquisição. Em outras línguas, também há trabalhos em que tal relação se mostra possível, como é o caso de Gilhooly e Logie (1980), com palavras em inglês, Van

Hell e Mahn (1997), com palavras em espanhol e em alemão, Bonin et al. (2003), em francês, e Della Rosa et al. (2010), em italiano. De modo geral, estes trabalhos indicam que palavras mais concretas parecem ser adquiridas mais cedo, com exceção de Bonin et al. (2003), cujos autores explicitam a correlação da concretude com a imageabilidade e com a valência emocional, mas não com idade de aquisição. Em Gilhooly e Logie (1980), há indicativos de que as palavras que são adquiridas mais cedo tendem a ser mais concretas, mais familiares e menos ambíguas. Para Van Hell e Mahn (1997), a relação aparece por apontarem um papel importante de palavras concretas no aprendizado de línguas estrangeiras. Ademais, de acordo com Della Rosa et al. (2010), os participantes dos experimentos, aparentemente, contam com o mesmo tipo de informação ao avaliar concretude, imageabilidade, familiaridade e idade de aquisição, mesmo em línguas distintas.

O presente trabalho contribui para o estudo da métrica de concretude ao apresentar dados referentes a um *corpus* de aquisição da linguagem do português. Dois pontos na escolha de nosso *corpus* em relação a estudos anteriores merecem destaque. Em primeiro lugar, teremos uma análise da correlação entre concretude e idade de aquisição para contrastar os dados aqui coletados com o trabalho de Soares et al. (2016). A escolha do nosso *corpus* permite também a análise de concretude em diferentes classes gramaticais. Sobre esse ponto, destacamos que a maioria dos estudos sobre concretude em português investiga um *corpus* formado por substantivos. Dentre eles, Soares et al. (2016) é o trabalho com maior variedade de categorias gramaticais, sendo 3800 palavras do PE distribuídas entre nomes, adjetivos, verbos, advérbios e interjeições. Como acréscimo, nosso artigo apresenta notas de concretude para onomatopeias, pronomes, conjunções, preposições e numerais, além das classes já investigadas por Soares et al. (2016).

A coleta realizada insere-se em um contexto maior de trabalho de dissertação de mestrado que buscava relacionar a iconicidade com outras métricas, como concretude, frequência e idade de aquisição (Ananias, 2024). Porém, neste artigo, são apresentadas análises centradas na concretude, e, portanto, contribuições não presentes na dissertação.

Assim, considerando a lacuna identificada, objetivamos, principalmente, relacionar a concretude e a idade de aquisição. Em segundo plano, ainda propomos averiguar sua correlação com classe gramatical em prol de identificar se nossos resultados seguem o padrão dos outros estudos já realizados. Por fim, disponibilizamos à comunidade científica nossos dados brutos e uma tabela resumindo média e desvio-padrão de cada uma das 521 palavras analisadas, de modo a contribuir para futuros estudos que precisem controlar a concretude desses itens.

2. Materiais

O *corpus* utilizado foi aproveitado do *MacArthur-Bates Communicative Developmental Inventories* (MCDI), um banco de dados que disponibiliza a proporção de crianças (entre 18 e 30 meses) que produzem determinadas palavras e sentenças em diversas línguas, inclusive em português – no caso,

PE¹. Na Figura 1, a seguir, é possível visualizar essa relação em forma gráfica a partir das medidas de ‘mãe’, ‘aqui’ e ‘tirar’.

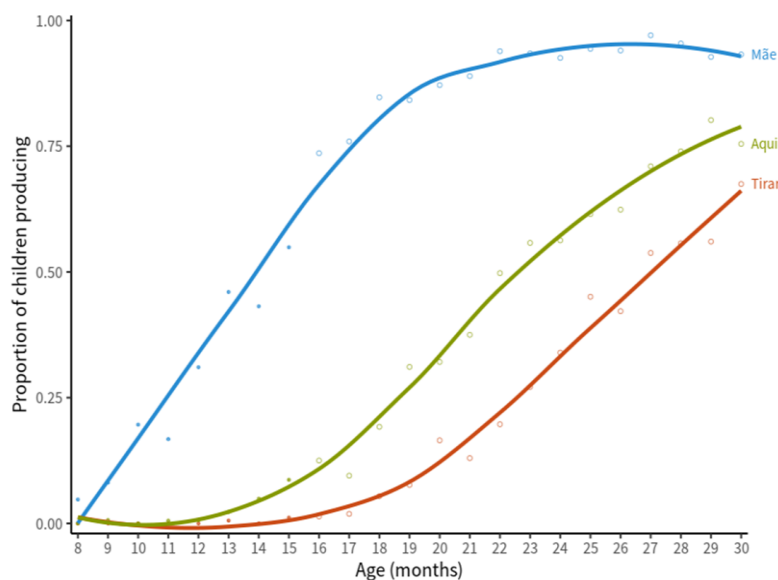


Figura 1 – Exemplo de dados do MCDI

Fonte: Elaboração própria a partir dos dados do MCDI, 2024.

A Figura 1 nos permite exemplificar a contribuição dos dados do MCDI ao perceber que a proporção de crianças produzindo, aos 18 meses, ‘mãe’ é extremamente mais alta do que ‘aqui’ e, mais ainda, do que ‘tirar’. Com o objetivo de utilizar um valor específico de produção para cada palavra, valemo-nos de sua respectiva porcentagem de produção aos 30 meses, a exemplo de estudos anteriores (Perry, Perlman e Lupyan, 2015).

Este corpus contempla substantivos, adjetivos, verbos, numerais, onomatopéias, interjeições, preposições, pronomes, conjunções e advérbios. Perry, Perlman e Lupyan (2015) propõem um padrão de agrupamento para essas classes ao tratar dos dados do MCDI para espanhol e inglês; ao lidar com os dados em português, optamos por seguir o mesmo padrão para este trabalho (cf. Tabela 2).

Com o intuito de realizar o experimento com falantes do PB, foram feitas adaptações e exclusões no corpus original. Quando, por exemplo, era possível perceber uma semelhança considerável entre a forma do PE apresentada e a forma escrita do PB, visualizamos uma abertura para adaptação (e.g. de ‘cocoró’ para ‘cocoricó’). Além disso, percebemos que o corpus varia no número das palavras, ora apresenta palavras no singular, ora no plural, sem um critério específico, como é o caso de ‘ameixas’ aparecer no plural e ‘banana’ no singular, mesmo as duas sendo frutas. Por esse motivo, quando não são palavras referentes a objetos que usualmente aparecem em pares (e.g. ‘meias’ e ‘olhos’),

¹ Idealmente, o *corpus* de idade de aquisição, ao ser relacionado com o de concretude, seria constituído pela proporção de aprendizado de crianças brasileiras. Porém, como não há um corpus equivalente ao MCDI com dados do PB, consideramos uma válida saída utilizarmos o do PE. Esse empréstimo pode representar uma influência nos resultados, o que nos motiva a incentivar a criação de um corpus como este para o PB em pesquisas futuras.

adaptamos a fim de apresentar palavras no singular (de ‘ameixas’ para ‘ameixa’). Além disso, os itens do corpus que apresentam flexão de gênero (e.g. ‘este’/ ‘esta’/ ‘estes’/ ‘estas’) também foram padronizados em apenas um item masculino singular (‘este’).

Alguns dos itens foram descartados com base em seis critérios, elencados na tabela 1:

	Critérios	Exemplos	Quantidade
1	Palavra específica do PE	Tinoni; autocarro; fixe	42
2	Palavra estrangeira	Puzzle, Shopping	3
3	Expressão	Onde está?; não mexa!	49
4	Palavra composta	coca-cola; guarda-chuva	6
5	Pessoal	Nome próprio; nome do(a) educador(a)	3
6	Verbo conjugado	Queres; vamos	2
Total			105

Tabela 1 - Critérios de palavras excluídas
 Fonte: Elaboração própria, 2024

No que se refere ao critério (i), as palavras foram excluídas por não serem evidentemente semelhantes o suficiente para serem adaptadas. O segundo critério indica palavras de origem estrangeira, como ‘puzzle’, do inglês. Os critérios (iii) e (iv) foram estabelecidos dado o foco do trabalho ser, especificamente, palavras de forma unitária, então, caso o item fosse constituído por mais de uma, retiramos. Além disso, por ser constituído para identificar idade de aquisição, o corpus dispõe de alguns sintagmas específicos utilizados por crianças, como “xixi!”, que, apesar de ser apenas uma palavra, é uma expressão infantil específica. Nesta mesma linha de raciocínio, o critério (v) é utilizado para descartar sintagmas específicos da aquisição, como o nome de parentes. Por fim, o sexto e último critério determina o descarte de itens com verbos conjugados, uma vez que optamos por apresentar apenas verbos no infinitivo, de forma a padronizar a classe de palavras.

Ao final da exclusão das palavras coerentes com os critérios elencados, restaram 521 palavras para coleta de notas de concretude. A quantidade de palavras por categoria gramatical pode ser vista na Tabela 2:

Categoria	Quantidade
Substantivo	304
Verbo	83
Adjetivo	44
Palavra funcional + pronome + advérbio	66
Onomatopeia + interjeição	14
Numeral	10
Total	521

Tabela 2 – Quantitativo final de palavras para o experimento
 Fonte: Elaboração própria, 2024

Após a criação das listas, elaboramos um script com o auxílio dos pacotes de software gratuitos do Jspsych, um framework JavaScript que possibilita a criação de instrumentos executáveis em

navegador de *web* (De Leeuw, 2015). Em seguida, testamos e hospedamos o experimento no *Mind-probe*, servidor gratuito europeu.

3. Procedimentos e Participantes

O experimento para coleta de notas de concretude foi realizado em formato *on-line*, de forma que os participantes foram recrutados a partir de e-mail enviado pela SINFO/UFRN. A coleta de dados foi realizada entre 14 de abril e 11 de maio de 2023. Como critério de seleção, o participante deve ser maior de idade e falante nativo do PB.

Antes de iniciar o experimento, destacamos que os participantes são apresentados ao Termo de Consentimento². Ao aceitar participar, o voluntário via, em sua tela, as instruções e uma definição para o que precisaria julgar naquela palavra, sendo, neste caso, a concretude. Para tanto, explicitava-se o conceito utilizado (cf. introdução) e solicitava-se que o voluntário desse uma nota entre muito concreta (7) e nada concreta (1) (Janczura *et al.*, 2007; Stein e Gomes, 2009; Calais *et al.*, 2012) para uma lista de 21 ou 22 palavras. Depois de dar a nota de concretude que considerava adequada para cada palavra, o participante ainda relatava se a conhecia ou não, pois, mesmo tendo feito uma seleção inicial (cf. seção de materiais), devíamos confirmar se a palavra era, realmente, conhecida por falantes brasileiros.

Em um segundo momento, após responder à tarefa de julgamento, perguntamos gênero, idade e língua materna do participante. Não foi solicitado nome ou outra informação pessoal.

Dos 784 participantes que finalizaram a tarefa de julgamento, 10 (1,27%) foram descartados: 2 deles por serem menores de idade e 8 por não serem falantes nativos de português brasileiro. A média de idade dos participantes restantes foi de 33,82.

4. Resultados

A média de concretude das palavras variou entre 1.42 ($S = 3.73$) e 7.00 ($S = 3.24$). As palavras percebidas como menos concretas se mostraram, majoritariamente, pertencentes às classes das palavras funcionais e das interjeições. Como exemplo, podemos citar as três menos concretas: *tão* ($\bar{X} = 1.42$, $S = 3.73$); *na* ($\bar{X} = 1.45$, $S = 2.67$) e *oh* ($\bar{X} = 1.52$, $S = 1.91$). Ao investigar as 100 palavras menos concretas, só se encontra um substantivo, sendo *horas* ($\bar{X} = 2.73$), as outras 99 são palavras funcionais, interjeições, onomatopeias, verbos e, em menor número, adjetivos. Para conferir mais especificamente, por cada categoria, na tabela 3, apresentamos as médias brutas de concretude por classe gramatical (antes de unir as classes):

² O experimento foi revisado e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Norte (CAAE: 63342122.0.0000.5537).

Classe gramatical	Média de concretude	Desvio-padrão
Substantivo	6.30	1.36
Numeral	3.75	2.27
Verbo	3.55	2.08
Adjetivo	3.55	2.13
Pronome	2.87	2.02
Onomatopeia	2.83	2.15
Advérbio	2.69	1.93
Interjeição	2.33	2.00
Palavra funcional	2.18	1.83

Tabela 3 – Médias brutas de concretude por classe gramatical
 Fonte: Elaboração própria, 2024

Na Tabela 4, listamos as 10 das palavras mais concretas de nosso experimento. Uma lista completa com média e desvio-padrão de todas as palavras analisadas está disponível como material suplementar deste artigo.

Item	Média de concretude	Desvio padrão
Lápis	7.00	3.24
Mochila	6.96	2.63
Chinelos	6.94	2.51
Alface	6.94	2.66
Cenoura	6.94	3.41
Mesa	6.93	2.11
Bicicleta	6.93	2.47
Portão	6.93	2.97
Uvas	6.93	2.98
Ambulância	6.92	2.78

Tabela 4 – Amostra de palavras mais concretas do experimento
 Fonte: Elaboração própria, 2024

Apenas por essa amostra, conseguimos perceber o seguinte fato: as palavras mais concretas tendem a apresentar palavras da classe de substantivos. Por outro lado, quando vemos as 10 palavras menos concretas (cf. Tabela 5), percebemos que boa parte dessas palavras são onomatopeias, interjeições ou palavras funcionais.

Item	Média de concretude	Desvio padrão
Tão	1.42	0.96
Na	1.45	0.82
Oh	1.52	1.30
Tanto	1.59	1.16
Desse	1.75	1.34
Uau	1.83	1.67
Ah	1.84	1.68
Da	1.85	1.69
Porque	1.95	1.51
Já	1.97	1.51

Tabela 5 – Amostra de palavras menos concretas do experimento
 Fonte: Elaboração própria, 2024

A Figura 2 mostra a relação entre idade de aquisição e concretude para cada classe de palavras, o que nos permite perceber alguns padrões.

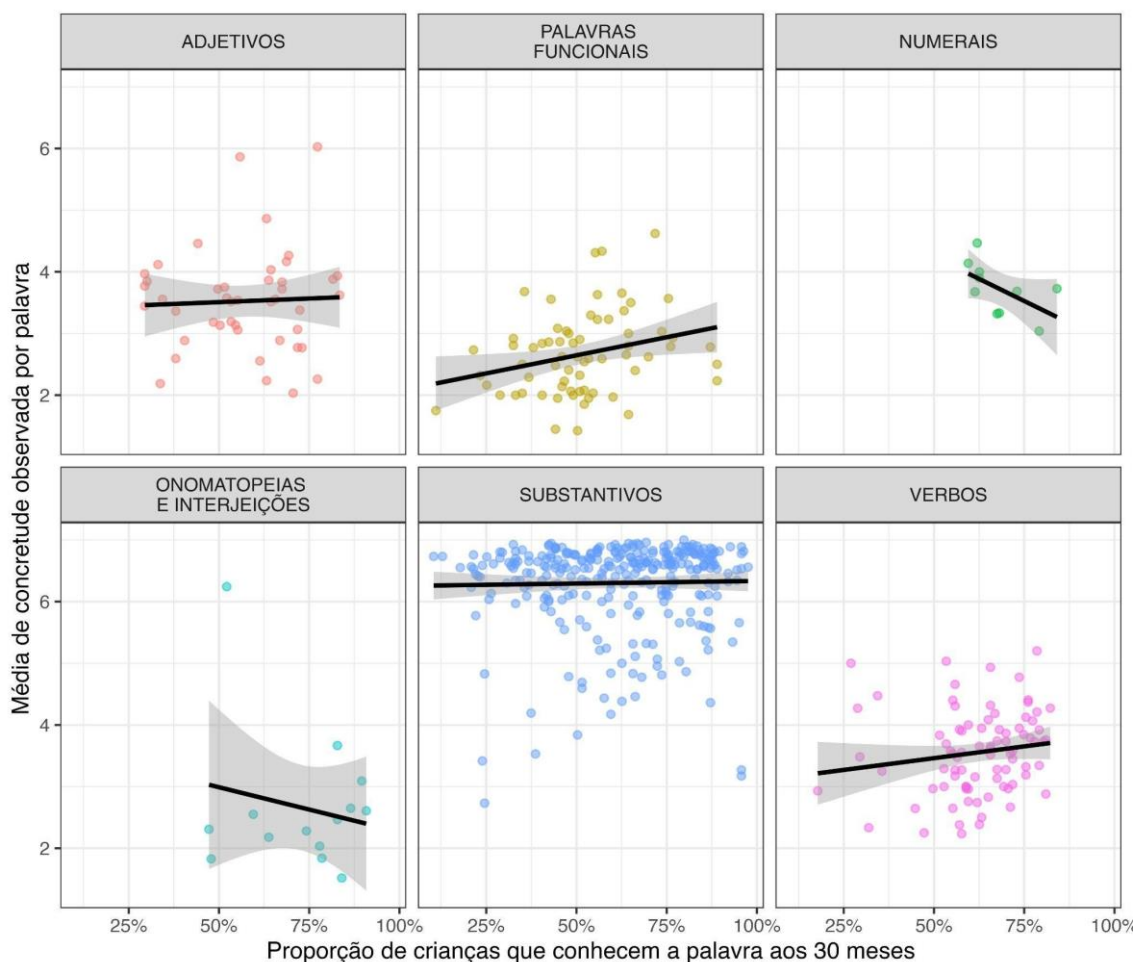


Figura 2 – Relação entre média de concretude e idade de aquisição para cada classe gramatical
 Fonte: Elaboração própria, 2024.

Primeiramente, vemos que a classe dos substantivos parece ter, comparativamente às outras classes, as médias mais altas de concretude. Esse padrão é esperado, uma vez que estamos lidando com itens lexicais presentes na fala de crianças de 2 anos e meio. Em outras palavras, é de se esperar que, nessa idade, boa parte dos substantivos usados tenham relação com as experiências concretas dos falantes.

Vemos também que as correlações entre idade de aquisição e concretude parecem variar para cada classe gramatical: há uma indicação de correlação negativa para numerais, por exemplo, correlações positivas para palavras funcionais e verbos, e, para substantivos, parece não haver nenhuma tendência mais clara. É preciso destacar, entretanto, que os dois casos em que a tendência negativa é percebida são os grupos com menos itens (numerais e onomatopeias/interjeições) e, consequentemente, o efeito não apresenta peso estatístico significativo.

5. Relação entre concretude e classe gramatical

Todas as análises estatísticas deste trabalho foram conduzidas por meio do R versão 4.3.2 (R Core Team, 2023), com o uso dos pacotes *tidyverse* (Wickham et al., 2019) para processamento dos dados e os pacotes *lme4* (Bates et al., 2015) e *ordinal* (Christensen, 2022) para análise estatística.

Para investigar a correlação entre nota de concretude e categoria gramatical, ajustamos uma série de testes por meio de um modelo linear misto com nota de concretude como variável resposta e a categoria lexical como variável preditora. Item e participante foram considerados como efeitos aleatórios. Através de uma análise por modelos aninhados notou-se que a categoria lexical tem, de fato, influência na percepção de concretude ($\chi^2(5) = 867.57$; $p < 0.0001$). Em uma análise *post-hoc*, comparações pareadas mostraram as diferenças entre todos os níveis da variável preditora.

Classes comparadas		<i>p</i> -valor	Efeito	SE
Substantivos	Verbos	<0.0001	2.726	0.088
Adjetivos	Palavras funcionais	<0.0001	0.995	0.136
Adjetivos	Numerais	1.0000	-0.039	0.249
Adjetivos	Onomatopeias + Interjeições	0.0001	0.993	0.218
Adjetivos	Substantivos	<0.0001	-2.609	0.111
Adjetivos	Verbos	0.9446	0.118	0.129
Palavras funcionais	Numerais	0.0003	-1.034	0.243
Palavras funcionais	Onomatopeias + Interjeições	1.0000	-0.002	0.211
Palavras funcionais	Substantivos	<0.0001	-3.603	0.096
Palavras funcionais	Verbos	<0.0001	-0.877	0.117
Numerais	Onomatopeias + Interjeições	0.0065	1.032	0.297
Numerais	Substantivos	<0.0001	-2.569	0.229
Numerais	Verbos	0.9860	0.157	0.239
Onomatopeias + Interjeições	Substantivos	<0.0001	-3.601	0.196
Onomatopeias + Interjeições	Verbos	0.0003	-0.875	0.207

Tabela 6 – *p*-valor da comparação entre categorias lexicais: valor ajustado pelo método de Tukey

Fonte: Elaboração própria, 2024

A Tabela 6 mostra a comparação pareada das notas de concretude entre classes gramaticais. De forma específica, temos que: os substantivos são consideravelmente mais concretos que todas as outras classes; os adjetivos, assim como numerais e verbos (classes que não apresentam diferença de percepção significativa entre si), são mais concretos que palavras funcionais, onomatopeias e interjeições.

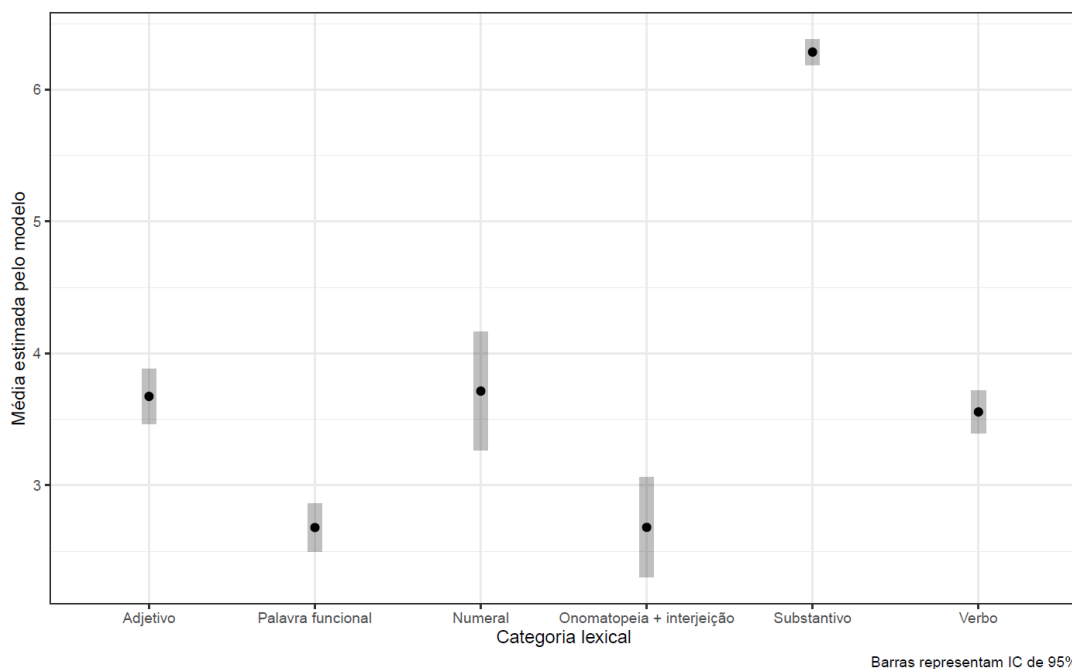


Figura 3 – Comparação entre média de nota de concretude e categorias lexicais
 Fonte: Elaboração própria, 2024

A Figura 3 possibilita a melhor visualização para indicarmos, com segurança, que a classe dos substantivos é de maior média de concretude em comparação a todas as outras. Na sequência, visualizamos uma certa equivalência de médias entre adjetivo, numerais e verbos. E, como representantes dos itens menos concretos, as palavras funcionais, onomatopeias e interjeições.

Embora o uso de modelos lineares mistos seja recomendado para tratar de dados de escala Likert (Norman, 2010; Gibson, Piatandosi e Fedorenko, 2011), há também recomendação de se usar um modelo ordinal tendo em vista a natureza dos dados. Realizamos as mesmas análises descritas acima por meio de um modelo misto para dados ordinais. Do ponto de vista inferencial, os resultados obtidos foram os mesmos.

6. Relação entre concretude e idade de aquisição para cada classe gramatical

Para testar se essas correlações vistas na Figura 2 eram significativas, ajustamos um modelo linear misto ao subconjunto de dados de cada classe gramatical. Os modelos tinham a nota de concretude como variável resposta, a idade de aquisição como variável preditora e, sempre que permitido pelos dados, interceptos aleatórios por participante e palavra. Os *p*-valores dos seis testes realizados foram ajustados usando o método *False Discovery Rate* (FDR). A Tabela 7 traz um resumo das estatísticas e valores estimados para cada modelo:

Classe	<i>b</i>	Erro-padrão	<i>valor-t</i>	<i>p-valor</i>
Adjetivos	0.2360	0.7976	0.296	0.76
Palavras funcionais	1.1286	0.4786	2.358	0.12
Numerais	-3.077	1.755	-1.753	0.16
Onomatopeias e Interjeições	-1.491	1.985	-0.751	0.69
Substantivos	0.07212	0.20423	0.353	0.76
Verbos	0.8653	0.4713	1.836	0.16

Tabela 7 – estatísticas estimadas do efeito de idade de aquisição nas notas de concretude para cada classe gramatical; *p*-valores corrigidos por FDR

Fonte: Elaboração própria, 2024

Como se pode ver, não houve nenhuma correlação significativa entre concretude e idade de aquisição. No caso das palavras funcionais, houve inicialmente um resultado que indicava correlação estatisticamente significativa, mas essa significância desaparece com o ajuste do *p*-valor. Considerando que o *p*-valor deve ser interpretado em conjunto com outras estatísticas informadas pelo modelo, é possível que estejamos aqui diante de um caso de erro do tipo II, um falso negativo. Contudo, essa interpretação deve ser vista com cautela, uma vez que o método de correção de *p*-valor utilizado foi empregado justamente para controlar as ocorrências de erro do tipo I (falsos positivos) sem incorrer em grandes taxas de erro do Tipo II (Benjamini e Hochberg, 1995). Com base nos nossos dados, portanto, não é possível afirmar correlação significativa entre idade de aquisição e notas de concretude para palavras funcionais.

As mesmas análises descritas nesta seção foram realizadas também por meio de um modelo misto para dados ordinais, e o resultado foi o mesmo no que diz respeito às inferências que podem ser feitas a partir dos dados.

7. Discussão

Tomando por base os estudos que lidam com variadas classes de palavras, nenhum deles parece ter feito uma comparação entre as classes envolvendo a concretude. Em Soares *et al.* (2016), por exemplo, cuja língua em foco é o PE, apesar de lidar com a maior diversidade de classes entre os trabalhos citados, não explicita uma comparação entre elas e nem como essa característica da palavra estaria relacionada à concretude, imageabilidade, frequência e idade de aquisição. Os estudos com variadas classes em PB também seguem o mesmo raciocínio, pois nem Stein e Gomes (2006) nem Janczura *et al.* (2007) – estudos que lidam com outras classes além de substantivos – estabelecem essa comparação entre classes de palavras. Nesse cenário, os achados deste estudo quanto à relação da concretude e idade de aquisição por classes de palavras são uma contribuição.

Assim, retomamos que, de acordo com nossos dados, as correlações entre idade de aquisição e concretude parecem variar a depender da classe gramatical: há uma indicação de correlação negativa para numerais; correlações positivas para palavras funcionais e verbos; e, nenhuma correlação clara para substantivos. No entanto, nenhuma dessas correlações é significativa do ponto de vista estatístico. É relevante ressaltar que os dados utilizados são de uma natureza específica: o de palavras conhecidas e produzidas por crianças de até 2 anos e meio. Tal especificação faz com que algumas de nossas notas de concretude coletadas sejam bem maiores do que em outros estudos, principalmente no caso dos substantivos, já que a aquisição vocabular tende a refletir a experiência da criança no mundo (Tomasello, 2003). Como consequência, palavras de concretude mais baixa, especialmente entre os substantivos, estão pouco presentes em nossa amostra. Essa característica de nossos dados pode fazer com que não se capture a relação entre concretude e idade de aquisição que possa existir caso se considere uma amostra cujas notas de concretude por palavra estejam mais espalhadas entre os valores de 1 a 7 da escala usada.

Acreditamos que, por esse motivo, nossos resultados vão de encontro ao resultado levantado para o português europeu em Soares *et al.* (2016). Os autores indicam uma correlação significativa entre idade de aquisição e concretude, evidenciando que palavras mais concretas tendem a ser aprendidas mais cedo. Contudo, essa correlação não foi percebida em nossas análises. Uma primeira explicação para esse contraste pode ser encontrada na diferença de métrica da variável idade de aquisição. Soares *et al.* (2016) operacionalizam idade de aquisição como a idade estimada de aquisição de uma palavra. Essas métricas foram obtidas por outros estudos que realizaram questionários perguntando a falantes de português europeu quantos anos eles estimavam ter quando aprendiam certas palavras (Marques *et al.*, 2007; Cameirão e Vicente, 2010). Nosso trabalho, por outro lado, é baseado em um vocabulário restrito a crianças de 2 anos e meio, que encaminha para a sobre-representação de palavras muito concretas dada a sobre-representação de substantivos. Estudos que incluam métricas de crianças mais velhas podem ter mais ocorrências de palavras com notas menores de concretude, permitindo perceber uma correlação entre essas variáveis.

Além disso, há uma diferença considerável no método estatístico empregado. Soares *et al.* (2016) realizaram suas análises por meio de testes de correlação de Pearson, que são modelos lineares simples

que assumem independência dos dados. Não é possível saber se as análises foram realizadas no conjunto bruto de dados, com as notas de concretude dadas por cada participante, ou se a correlação foi feita entre a média de concretude e a média de idade de aquisição de cada palavra. Em todo caso, qualquer uma dessas análises desconsidera efeitos aleatórios de itens experimentais e participantes, medidas que costumam ser consideradas na construção de modelos lineares de efeitos mistos em estudos de psicolinguística (Godoy; Nunes, 2020). Com a inserção de efeitos aleatórios, é possível controlar melhor a ocorrência de erros do Tipo I, e efeitos antes identificados em modelos lineares simples tendem a desaparecer (Lima Jr; Garcia, 2021). A título de comparação, realizamos em nossos dados um teste de correlação de Pearson entre média de idade de aquisição e média de concretude de palavra, também ignorando os efeitos aleatórios de participantes e itens. Nessa análise, a exemplo de Soares et al. (2016), os resultados foram significativos: palavras mais concretas tendem a ter um número maior de crianças de até 30 meses que as conhecem ($t(519) = 2,58, p = 0,01, r = 0,11$). Portanto, é possível que a diferença entre os dados reportados se deva ao emprego, no presente estudo, de métodos estatísticos que controlem melhor a ocorrência de erros do Tipo I.

Apesar das limitações apontadas, reiteramos a novidade deste estudo por relacionar a percepção de concretude por falantes brasileiros e a idade de aquisição por categoria gramatical. A partir de nossos dados, a comunidade científica agora conta com métricas de concretude e idade de aquisição para 521 palavras do português.

Informações complementares

Avaliação e resposta dos autores

Avaliação: <https://doi.org/10.25189/rabralin.v23i1.2268.R>

Conflito de Interesse

As autoras não têm conflitos de interesse a declarar.

Protocolo e Pré-Registro de Pesquisa

Avaliamos os roteiros da Equator Network e percebemos que o trabalho submetido não se encaixa nos tipos elencados. Além disso, a pesquisa conduzida não foi pré-registrada em um repositório institucional independente.

Declaração de Disponibilidade de Dados

Os dados, códigos e materiais que suportam os resultados deste estudo estão disponíveis como 'arquivos suplementares' no site da revista da Abralín. Além disso, também podem ser resgatados no seguinte link: <https://osf.io/f8ntq/>

Ética

O experimento foi revisado e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Norte (CAAE: 63342122.0.0000.5537).

Fontes de financiamento

Thayná Cristina Ananias, primeira autora, agradece à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela bolsa concedida nos anos de desenvolvimento do mestrado (2022-2024), uma vez que o presente trabalho consiste em um desdobramento da dissertação.

REFERÊNCIAS

ANANIAS, T.C. **A iconicidade e a aquisição de palavras do português**. Dissertação (Mestrado em Estudos da Linguagem) - Centro de Ciências Humanas, Letras e Artes, Universidade Federal do Rio Grande do Norte, Natal, 2024.

BATES, D., MAECHLER, M., BOLKER, B., WALKER, S. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 1-48, 2015. DOI: <https://doi.org/10.18637/jss.v067.i01>

BENJAMINI, Y., HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 57, n.1, p. 289-300, 1995. DOI: <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

BONIN, P., MÉOT, A., AUBERT, L. F., MALARDIER, N., NIEDENTHAL, P., & CAPELLE-TOCZEK, M. C. Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. **L'année Psychologique**, v. 103, n. 4, p. 655-694, 2003.

CALAIS, L.L. LIMA-GREGIO, A. M., ARANTES, P., GIL, D., & BORGES, A. C. L. D. C. Um julgamento de concreitude de palavras. **Jornal da Sociedade Brasileira de Fonoaudiologia**, v. 24, p. 262-268, 2012. DOI: <https://doi.org/10.1590/S2179-64912012000300012>

CAMEIRÃO, M. L.; VICENTE, S. G. Age-of-acquisition norms for a set of 1,749 Portuguese words. **Behavior research methods**, v. 42, n. 2, p. 474-480, 2010.

CHRISTENSEN, R. H. B. **ordinal**: Regression Models for Ordinal Data. R package version 2022.11-16, 2022.

DELLA ROSA, P. A., CATRICALÀ, E., VIGLIOCCO, G., & CAPPA, S. F. Beyond the abstract–concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. **Behavior Research Methods**, v. 42, p. 1042-1048, 2010. DOI: <https://doi.org/10.3758/BRM.42.4.1042>

DE LEEUW, J. R. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. **Behavior Research Methods**, v. 47, n. 1, p. 1-12, 2015. DOI: <https://doi.org/10.3758/s13428-014-0458-y>

GIBSON, E., PIANTADOSI, S., & FEDORENKO, K. Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments: Linguistic Acceptability on Mechanical Turk. **Language and Linguistics Compass**, v. 5, n. 8, 509-524, 2011.

GILHOOLY, K. J., & LOGIE, R. H. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. **Behavior Research Methods & Instrumentation**, v. 12, n. 4, p. 395-427, 1980. DOI: <https://doi.org/10.3758/BF03201693>

GODOY, M. C.; NUNES, M. A. Uma comparação entre ANOVA e modelos lineares mistos para análise de dados de tempo de resposta. **Revista da ABRALIN**, [S. l.], v. 19, n. 1, p. 1-23, 2020. DOI: <https://doi.org/10.25189/rabralin.v19i1.1388>

JANCZURA, G. A., CASTILHO, G. M. D., ROCHA, N. O., VAN ERVEN, T. D. J. C., & HUANG, T. P. Normas de concreteness para 909 palavras da língua portuguesa. **Psicologia: teoria e pesquisa**, v. 23, p. 195-204, 2007. DOI: <https://doi.org/10.1590/S0102-37722007000200010>

LIMA, M. F. R. D., & BURATTO, L. G. Norms for familiarity, concreteness, valence, arousal, wordlikeness, and recall accuracy for Swahili-Portuguese word pairs. **Sage Open**, v. 11, n.1, 2021. DOI: <https://doi.org/10.1177/2158244020988524>

LIMA JR, R. M.; GARCIA, G. D. Diferentes análises estatísticas podem levar a conclusões categoricamente distintas. **Revista da ABRALIN**, [S. l.], v. 20, n. 1, p. 1-19, 2021. DOI: <https://doi.org/10.25189/rabralin.v20i1.1790>

MARQUES, J. F. Normas de imagética e concreteness para substantivos comuns. **Laboratório de Psicologia**, v. 3, n. 1, p. 65-75, 2005.

MARQUES, J. F., FONSECA, F. L., MORAIS, S., & PINTO, I. A. Estimated age of acquisition norms for 834 Portuguese nouns and their relation with other psycholinguistic variables. **Behavior Research Methods**, v.39, p.439-444, 2007.

NORMAN, G. Likert scales, levels of measurement and the “laws” of statistics. **Advances in Health Sciences Education**, v. 15, n.5, p. 625-632, 2010.

PAIVIO, A., SMYTHE, P. C., & YUILLE, J. C. Imagery versus Meaningfulness of Nouns in Paired-Associated Learning. **Canadian Journal of Psychology**, v. 22, p. 427-441, 1968. DOI: <https://psycnet.apa.org/doi/10.1037/h0082782>

PERRY, L.; PERLMAN, M.; LUPYAN, G. Iconicity in English and Spanish and its relation to lexical category and age of acquisition. **PloS one**, v. 10, n. 9, p. 137147, 2015. DOI: <https://doi.org/10.1371/journal.pone.0137147>

SOARES, A. P., COSTA, A. S., MACHADO, J., COMESAÑA, M., & OLIVEIRA, H. M. The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. **Behavior Research Methods**, v. 49, p. 1065-1081, 2016. DOI: <https://doi.org/10.3758/s13428-016-0767-4>

STEIN, L. M.; GOMES, C. F. A. Normas brasileiras para listas de palavras associadas: associação semântica, concretude, frequência e emocionalidade. **Psicologia: teoria e pesquisa**, v. 25, p. 537-546, 2009. DOI: <https://doi.org/10.1590/S0102-37722009000400009>

TOMASELLO, M. **Origens culturais da aquisição do conhecimento humano**. São Paulo: Martins Fontes, 2003.

VAN HELL, J. G., & MAHN, A. C. Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. **Language learning**, v. 47, n.3, p. 507-546, 1997. DOI: <https://doi.org/10.1111/0023-8333.00018>

VASCONCELOS, L. F., ALMEIDA, P. D. A., ESTIVALET, G. L., & FERRARI-NETO, J. Teste de Memória de Trabalho de Leitura: Versão Computadorizada Padronizada do Reading Span Test para o Português Brasileiro. Conference: **VII Jornada de Descrição do Português**, 2019. p. 303-311.

WANG Z.; YAO Z. Concreteness Effects of Emotional Noun Words: Evidences from ERP. **Acta Psychologica Sinica**, v. 44, n. 2, p. 154, 2012. DOI: <http://dx.doi.org/10.3724/SP.J.1041.2012.00154>

WICKHAM H., AVERICK M., BRYAN J., CHANG W., MCGOWAN L.D., FRANÇOIS R., GROLEMUND G., HAYES A., HENRY L., HESTER J., KUHN M., PEDERSEN TL., MILLER E., BACHE S.M., MÜLLER K., OOMS J., ROBINSON D., SEIDEL D.P., SPINU V., TAKAHASHI K., VAUGHAN D., WILKE C., WOO K., YUTANI H. "Welcome to the tidyverse." **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 2019. DOI: <https://doi.org/10.21105/joss.01686>

YAO, Z., WU, J., ZHANG, Y., & WANG, Z. (2016). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. **Behavior Research Methods**, v. 49, n.4, p. 1374-1385, 2016. DOI: <https://doi.org/10.3758/s13428-016-0793-2>