

Faculdade de Letras da UFMG  
Cláudia Brandão Vieira

**A INFLUÊNCIA DE INFORMAÇÕES TEMÁTICAS NA  
PREVISÃO DE ARGUMENTOS VERBAIS**

Belo Horizonte  
2019

Cláudia Brandão Vieira

## **A INFLUÊNCIA DE INFORMAÇÕES TEMÁTICAS NA PREVISÃO DE ARGUMENTOS VERBAIS**

Tese apresentada ao Programa de Pós-graduação em Estudos Linguísticos (POSLIN) da Faculdade de Letras da Universidade Federal de Minas Gerais (FALE/UFMG) como requisito parcial para a obtenção do título de Doutora em Linguística.

Área de Concentração: Linguística Teórica e Descritiva

Linha de Pesquisa: (1D) Processamento da Linguagem

Orientador: Ricardo Augusto de Souza

Coorientadora: Mahayana Cristina Godoy

Belo Horizonte  
2019

V658i

Vieira, Cláudia Brandão.

A influência de informações temáticas na previsão de argumentos verbais [manuscrito] / Cláudia Brandão Vieira. – 2019.

163 p., enc. : il., tabs., p&b., color.

Orientador: Ricardo Augusto de Souza.

Coorientadora: Mahayana Cristina Godoy.

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Processamento da Linguagem.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: p. 130-142.

Apêndices: p. 143-163.

1. Linguística – Teses. 2. Semântica – Teses. 3. Língua portuguesa – Verbos – Teses. I. Souza, Ricardo Augusto de. II. Godoy, Mahayana Cristina. III. Universidade Federal de Minas Gerais. Faculdade de Letras. IV. Título

CDD : 410



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

PosLin

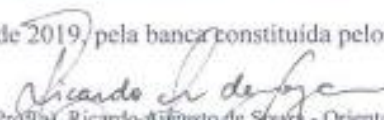
## FOLHA DE APROVAÇÃO

**A influência de informações temáticas na previsão de argumentos verbais**

**CLAUDIA BRANDÃO VIEIRA**

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Processamento da Linguagem.

Aprovada em 29 de agosto de 2019, pela banca constituída pelos membros:

  
Prof(a). Ricardo Augusto de Souza - Orientador  
UFMG

  
Prof(a). Mahayana Cristina Godoy - Coorientadora  
UFRN

  
Prof(a). Tílis Maira Machado de Sá  
UFMG

  
Prof(a). Larissa Santos Ciriaco  
UFMG

  
Prof(a). Alexandre Delfino Xavier  
UFMG

  
Prof(a). Jan Edson Rodrigues Leite  
UFPA

Belo Horizonte, 29 de agosto de 2019.

## **AGRADECIMENTOS**

Agradeço ao Professor Ricardo Souza o apoio institucional e acadêmico que me foi oferecido durante os momentos mais difíceis do desenvolvimento desta tese.

Agradeço à Professora Mahayana Godoy a atenção e o apoio acadêmico, sem os quais a realização deste trabalho não seria possível.

Agradeço a todos que durante esses quatro anos, seja de forma direta ou indireta, contribuíram para a realização deste trabalho. Sou grata pelo apoio acadêmico, pela compreensão e pela amizade de todos que me auxiliaram a permanecer trabalhando.

Agradeço à CAPES o apoio financeiro fornecido durante os quatro anos de desenvolvimento deste trabalho.

## EPÍLOGO

“Nothing is ever over  
Life breathes life in its turn  
Sometimes the people listen  
Sometimes the people learn”  
(Neil Gaiman)

## RESUMO

Neste estudo, investigamos se propriedades temáticas dos verbos podem auxiliar na previsão de traços semânticos de nomes propícios a ocupar a posição de argumento interno, durante o processamento linguístico. A previsão durante a compreensão da linguagem pode ser definida como o uso de representações internas de alto nível – conhecimentos do falante sobre o mundo real em interação com informações provenientes do contexto sentencial/discursivo – na pré-ativação de representações linguísticas de níveis mais baixos (*eg.*: fonológicas, semânticas, sintáticas) durante o desenvolvimento de um enunciado. Este trabalho fundamenta-se na proposta de Kuperberg e Jaeger (2016), para os quais a previsão linguística é um fenômeno gradual e probabilístico. Nessa perspectiva, a interação entre contexto e conhecimentos refinados sobre eventos pode facilitar a compreensão de palavras relacionadas à estrutura do evento ativado. No entanto, pouco se sabe sobre a força contextual necessária para que essa interação ocorra. Pretendemos investigar se verbos que possuem um papel temático bem delimitado (conforme a proposta de MCRAE; FERRETTI; AMYOTE, 1997) apresentam força suficiente para pré-ativar representações semânticas mais refinadas em contextos bastante simples, formados apenas pelo verbo e pelo argumento interno (*e.g.*: “derramar – leite”). Nosso objetivo é verificar se essa pré-ativação de informações semânticas facilita o processamento do argumento interno. Ademais, também pretendemos observar se há diferenças entre o comportamento preditivo de verbos com papéis temáticos bem delimitado e verbos com papéis temáticos pouco delimitados. Para compreendermos tais questões, realizamos inicialmente a construção de um *corpus* com dados de tarefas de cloze. Esse *corpus* nos permitiu observar as tendências de subseleção de argumentos de 520 verbos a partir dos julgamentos de no mínimo 50 participantes. Em seguida, realizamos um julgamento de ajuste temático no qual manipulamos a capacidade de subseleção dos verbos selecionados e a previsibilidade do argumento interno. Nosso objetivo foi observar se a capacidade de subseleção de argumentos causava comportamento temático divergente entre dois grupos de verbos. A análise dos julgamentos sugere que verbos de forte subseleção possuem papéis temáticos bem delimitados. Em comparação aos verbos de fraca subseleção, as notas atribuídas aos eventos expressos pelos verbos de forte subseleção são mais altas quando o argumento é mais previsível, e mais baixas quando o argumento é implausível. Por fim, com a finalidade de verificar o comportamento dos verbos durante o processamento on-line em contextos simples, realizamos duas tarefas de decisão lexical. Verificamos se as características temáticas dos verbos poderiam auxiliar na previsão de argumentos internos mais esperados (*e.g.*: “derramar – leite”) em comparação a argumentos internos implausíveis (*e.g.*: “derramar – casa”). As tarefas também foram aplicadas com o objetivo de observar se havia correspondências entre os resultados obtidos tanto no julgamento de ajuste temático quanto nas tarefas de decisão lexical. Foram encontradas evidências de que as informações temáticas dos verbos de forte subseleção podem contribuir para a previsão de traços semânticos dos argumentos verbais. Os resultados, em grande parte, se correlacionam com as evidências verificadas nos julgamentos de ajuste temático.

*Palavras-chave:* previsão, informações temáticas, representação semântica, conhecimentos de eventos

## ABSTRACT

This study aims at investigating whether the thematic properties of the verbs can help predicting the semantic features of names which frequently occupy the position of internal argument during language processing. Prediction during language comprehension can be defined as the use of high-level internal representations - the speaker's real-world knowledge interacting with information coming from sentence/discourse contexts - to pre-activate lower levels of linguistic representations (e.g. phonological, semantic, syntactic) during the processing of an input. Kuperberg and Jaeger (2016), the work on which this dissertation is based, posits that language prediction is a gradual and probabilistic phenomenon. According to this perspective, the interaction between context and fine-grained knowledge about events may facilitate the comprehension of words related to the structure of the activated event. However, little is known about the contextual strength necessary for this interaction to occur. We intend to investigate whether the verbs with a well-defined thematic role (according to MCRAE; FERRETTI; AMYOTE, 1997) have enough strength to pre-activate more refined semantic representations in fairly simple context formed only by a proper noun and a verb (e.g. spilled - milk). Our aim is to verify if such pre-activation of semantic information facilitates the processing of the internal argument. Furthermore, we intend to observe if there are any differences between predictive behavior of verbs with well-defined thematic roles and verbs with poorly defined thematic roles. In order to understand such questions, we initially built a corpus containing data from cloze tests. This corpus allowed for an observation of the predictive behavior of 520 verbs collected from the judgements of at least 50 participants. Subsequently, we conducted a thematic role assignment task (FERRETTI; MCRAE; HATHERELL, 2001; MCRAE; FERRETTI; AMYOTE, 1997) in which we manipulated the subselection strength of selected verbs and the previsibility of the internal arguments. We aim at observing whether the ability of subselection of arguments caused divergent thematic behavior between two groups of verbs. The analysis of judgments suggests that verbs with greater subselection strength have well-defined thematic roles. When compared to verbs with lower subselection strength, the grades attributed to the events expressed by the verbs with greater subselection strength are higher when the argument is more predictable and lower when the argument is implausible. Finally, aiming to verify the behavior of the verbs during on-line processing in simple contexts, we have run two lexical decision tests. We have verified whether the thematic characteristics of the verbs could help the most expected arguments prediction (e.g., spill milk) in comparison to implausible internal arguments (e.g., spill house). The tests have also been applied with the objective of investigating whether there were any correspondences between results obtained from both the thematic role assignment test and the lexical decision test. Evidence was found shows that the thematic information from strong subselection verbs could contribute to verbal arguments' semantic features prediction. For the most part, the results correlate with the evidence obtained from thematic role assignment test.

*Keywords:* prediction, thematic information, semantic representation, event knowledge



## LISTA DE ILUSTRAÇÕES

FIGURA 1 - QUADRO COM AS CONDIÇÕES EXPERIMENTAIS DE DELONG, URBACH E KUTAS (2005).....	37
FIGURA 2 – PROPRIEDADES DE SUBSELEÇÃO DO VERBO "DERRAMAR" .....	47
FIGURA 3 - PROPRIEDADES DE SUBSELEÇÃO DO VERBO "APARAR" .....	48
FIGURA 4 - HISTOGRAMA DE DISTRIBUIÇÃO DAS PROBABILIDADES DE CLOZE NO CORPUS.....	77
FIGURA 5- GRÁFICO COM CONCENTRAÇÃO DE ITENS POR TIPO DE SUBSELEÇÃO .....	79
FIGURA 6 - GRÁFICO DE PROPRIEDADES DE SUBSELEÇÃO DO VERBO MANOBRAR .....	81
FIGURA 7 – GRÁFICO DE PROPRIEDADES DE SUBSELEÇÃO DO VERBO AGARRAR .....	82
FIGURA 8 – QUADRO COM CONDIÇÕES EXPERIMENTAIS DO JULGAMENTO DE AJUSTE TEMÁTICO.....	83
FIGURA 9 – QUADRO COM EXEMPLOS DE FRASES DO JULGAMENTO DE AJUSTE TEMÁTICO .....	86
FIGURA 10- FORMULÁRIO DO JULGAMENTO DE AJUSTE TEMÁTICO .....	88
FIGURA 11 – QUADRO DE ANÁLISES POST-HOC DOS RESULTADOS DO MELHOR MODELO AJUSTADO: JULG~COND*TIPO+(1+COND\SUJ) +(1 ITEM) .....	90
FIGURA 12 - QUADRO DE ANÁLISES POST-HOC - CONTRASTE ENTRE VERBOS DE FORTE E FRACA SUBSELEÇÃO .....	91
FIGURA 13 - GRÁFICO DE FREQUÊNCIA DE JULGAMENTOS DOS VERBOS DE FORTE SUBSELEÇÃO .....	91
FIGURA 14 - GRÁFICO DE FREQUÊNCIA DE JULGAMENTOS DOS VERBOS DE FRACA SUBSELEÇÃO .....	92
FIGURA 15 – QUADRO COM CONDIÇÕES EXPERIMENTAIS DA TAREFA DE DECISÃO LEXICAL 1.....	103
FIGURA 16 - ORDEM DE APRESENTAÇÃO DOS ESTÍMULOS NA TAREFA DE DECISÃO LEXICAL 1 .....	104
FIGURA 17 - HISTOGRAMA DE DISTRIBUIÇÃO DOS TRS DA TAREFA DE DECISÃO LEXICAL 1 .....	106
FIGURA 18 - BOXPLOT DOS TRS NAS CONDIÇÕES DA TAREFA DE DECISÃO LEXICAL 1 .....	109
FIGURA 19 – GRÁFICO COM PROPRIEDADES DE SUBSELEÇÃO DO VERBO CONFEITAR.....	111
FIGURA 20 – GRÁFICO COM PROPRIEDADES DE SUBSELEÇÃO DO VERBO DIMINUIR .....	112
FIGURA 21 - ORDEM DE APRESENTAÇÃO DOS ITENS NA TAREFA DE DECISÃO LEXICAL 2 .....	115
FIGURA 22 - HISTOGRAMA DE DISTRIBUIÇÃO DOS TRS NA TAREFA DE DECISÃO LEXICAL 2 .....	116
FIGURA 23 – QUADRO COM ANÁLISES POST-HOC DOS RESULTADOS DO MELHOR MODELO AJUSTADO: TR~ COND*TIPO+(1 ALVO) + (1+COND SUJ).....	117

## LISTA DE TABELAS

TABELA 1 – VERBOS SELECIONADOS PARA O BANCO DE CLOZES E ARGUMENTOS MAIS CITADOS .....	143
TABELA 2 – VALORES DE CLOZE DE TODOS OS VERBOS UTILIZADOS COMO ITENS EXPERIMENTAIS (EM%) .....	150
TABELA 3 – ÍTENS EXPERIMENTAIS – JULGAMENTO DE AJUSTE TEMÁTICO .....	158
TABELA 4 – ÍTENS DE TAREFAS DE DECISÃO LEXICAL.....	160

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>12</b>
<b>2 FUNDAMENTOS TEÓRICOS.....</b>	<b>24</b>
<b>2.1 Processamento linguístico e antecipação .....</b>	<b>24</b>
2.1.1 Modelos baseados em restrições.....	24
2.1.2 Processamento antecipatório .....	32
<b>2.2 Perspectivas sobre a antecipação no processamento linguístico .....</b>	<b>38</b>
<b>2.3 Os conhecimentos de eventos como parte da representação lexical .....</b>	<b>50</b>
<b>2.4 Relações Temáticas .....</b>	<b>55</b>
<b>2.5 Priming semântico e processos de antecipação de informações on-line .....</b>	<b>60</b>
<b>3 A TAREFA DE CLOZE .....</b>	<b>69</b>
<b>3.1 Materiais .....</b>	<b>74</b>
<b>3.2 Participantes.....</b>	<b>75</b>
<b>3.3 Procedimentos.....</b>	<b>75</b>
<b>3.4 Resultados e discussão .....</b>	<b>76</b>
<b>4 JULGAMENTO DE AJUSTE TEMÁTICO .....</b>	<b>80</b>
<b>4.1 Materiais .....</b>	<b>84</b>
<b>4.2 Participantes.....</b>	<b>87</b>
<b>4.3 Procedimentos.....</b>	<b>87</b>
<b>4.4 Resultados .....</b>	<b>88</b>
<b>4.5 Discussão.....</b>	<b>93</b>
<b>5 DECISÃO LEXICAL .....</b>	<b>97</b>
<b>5.1 Tarefa de decisão lexical 1.....</b>	<b>101</b>
5.1.1 Materiais.....	101
5.1.2 Participantes.....	103
5.1.3 Procedimentos .....	103
5.1.4 Resultados .....	105
5.1.5 Discussão .....	110
<b>5.2 Tarefa de decisão lexical 2.....</b>	<b>114</b>
5.2.1 Materiais.....	114
5.2.2 Participantes.....	114
5.2.3 Procedimentos .....	114
5.2.4 Resultados .....	115
5.2.5 Discussão .....	118
<b>6. DISCUSSÃO GERAL .....</b>	<b>121</b>
<b>6.1 Considerações finais.....</b>	<b>127</b>
APÊNDICE A.....	143
APÊNDICE B.....	150
APÊNDICE C.....	158
APÊNDICE D.....	160

## 1 INTRODUÇÃO

Nesta pesquisa, dois dos principais pressupostos assumidos estão relacionados às perspectivas de que o processamento linguístico ocorre a partir da antecipação de informações de diversos tipos, e à noção de que informações contidas no léxico e construídas na experiência do falante com o mundo são fundamentais para o processamento linguístico. Por simplificação terminológica, a palavra “antecipação” será utilizada como sinônimo de “previsão”<sup>1</sup>.

Sob a perspectiva linguística, o conceito de antecipação está relacionado à capacidade de o cérebro lidar com a velocidade e a complexidade do processamento linguístico por “pensar à frente”, gerando informações sobre prováveis próximos estímulos e se preparando antecipadamente para o processamento desses estímulos em múltiplos níveis (FEDERMEIER, 2007). Em outras palavras, a previsão durante a compreensão da linguagem pode ser definida como o uso de representações internas de alto nível – conhecimentos do falante sobre o mundo real em interação com informações provenientes do contexto linguístico, discursivo/situacional – na pré-ativação de representações linguísticas de níveis que se encontram mais abaixo durante o desenvolvimento de um enunciado (KUPERBERG; JAEGER, 2016). Sob o ponto de vista linguístico, podemos afirmar que as representações de níveis mais baixos são caracterizadas como informações sensoriais, de níveis fonético, ortográfico ou mesmo fonológico. Os níveis cognitivos mais altos são aqueles em que processos mnemônicos mais complexos passam a ocorrer, como os níveis sintático, semântico e, em um sentido amplo, pragmático/discursivo.

A atuação de mecanismos de antecipação na cognição humana não se restringe ao processamento linguístico. Em sentido amplo, o processamento antecipatório refere-se a qualquer tipo de processo que incorpore ou gere não apenas informações sobre o passado ou o presente, mas também sobre estados futuros do corpo ou do ambiente. A

<sup>1</sup> Para uma distinção entre previsão e antecipação, veja Van Petten e Luka (2012), assim como Bubic, Von Cramon e Schubotz (2010).

palavra "antecipação" tem muitas definições devido ao seu uso em diferentes contextos em uma ampla variedade de domínios, como por exemplo, na biologia, na epistemologia, na psicologia, na robótica e na ciência da computação (BUBIC; VON CRAMON; SCHUBOTZ, 2010). Estudos em psicologia cognitiva e neurociência indicam que a capacidade de antecipar informações pode influenciar diversas funções cognitivas, como a atenção e o processamento de informações para tomada de decisões (BUTZ, 2004). No campo da inteligência artificial, um sistema antecipatório possui um modelo preditivo de si mesmo e/ou de seu ambiente, o que permite uma mudança do estado atual em um instante, de acordo com as previsões do modelo referentes a um instante posterior (ROSEN, 1985).

Ao longo das últimas décadas, um número consistente de pesquisas linguísticas (ALTMANN; KAMIDE, 1999; CHOW *et al.*, 2016; DELONG; URBACH; KUTAS, 2005; FEDERMEIER, 2007; FINE *et al.*, 2013; LAU, ELLEN F.; HOLCOMB; KUPERBERG, 2013; LAU, ELLEN FRANCES, 2009; VAN BERKUM *et al.*, 2005; VAN PETTEN; LUKA, 2012; WICHA; MORENO; KUTAS, 2004) sugere que os processos de antecipação são utilizados inconscientemente por falantes durante a compreensão e a produção da linguagem. Kuperberg e Jaeger (2016) sugerem que é logicamente impossível explicar os efeitos dos estudos citados acima sem assumir que o contexto influencia o estado do sistema de processamento de linguagem antes que uma entrada de nível mais baixo em direção a níveis mais altos seja observada.

Apesar de estímulos linguísticos muitas vezes serem rápidos, confusos e ambíguos, a comunicação frequentemente ocorre de maneira ágil e eficiente (CHOW *et al.*, 2016). Um dos motivos para tal facilidade de comunicação pode ser relacionado à capacidade de os falantes inconscientemente anteciparem informações linguísticas. Por exemplo, se considerarmos um fragmento de sentença como em (1), é possível que a maioria dos falantes antecipem que, no final desse fragmento de sentença, ocorreria a palavra "carne". Contudo, se considerarmos contextos mais simples, formados por apenas um nome próprio e um verbo, como em (2), a antecipação de um item lexical específico, ou mesmo de traços semânticos desse item, não seria possível devido à simplicidade do contexto discursivo e às propriedades de uso do verbo "ver", que não proporcionam informações específicas durante o processo de articulação da linguagem.

(1) O churrasqueiro fatiou generosos pedaços de ...

(2) João viu ...

Este trabalho embasa-se parcialmente nos fundamentos da Estrutura Hierárquica Gerativa do Processamento da Linguagem proposta por Kuperberg e Jaeger (2016), cuja proposta afirma que durante a compreensão da linguagem, o falante procura inferir, com o máximo de certeza, a causa subjacente das informações sensoriais que são produzidas incrementalmente ao longo do tempo. Os pesquisadores consideram a previsão um fenômeno probabilístico, gradual e sujeito a ser realizado em vários níveis representacionais.

A previsão é considerada um fenômeno probabilístico, pois, segundo os autores, durante o processamento, os falantes utilizam informações de alto nível para controlar as diferentes probabilidades<sup>2</sup> de expectativas paralelas a serem consideradas durante o processamento linguístico. Com isso, diferentes possibilidades são consideradas ao mesmo tempo e cada uma representa diferentes graus de crença. Caso uma entrada de nível mais baixo desconforme uma crença de alto grau probabilístico, o processador se readaptará para que uma nova crença seja reconsiderada probabilisticamente. No que diz respeito à gradualidade do fenômeno, é proposto que a magnitude, a força com que o efeito de antecipação ocorre, depende da maneira como determinados verbos, estruturas temáticas ou amplos contextos discursivos geram tendências contra ou a favor de uma suposta análise processual preferida. Além disso, os autores propõem que a ativação ocorre em diferentes níveis representacionais devido à influência das restrições contextuais. Contextos com restrições altamente lexicais podem influenciar os estágios iniciais do processamento de palavras de entrada, sugerindo que eles podem ser usados para pré-ativar informações em níveis sublexicais de representação, que refletem o nível fonológico, ortográfico e perceptual.

<sup>2</sup> Múltiplas previsões podem ser realizadas em um determinado nível representacional sempre com probabilidades que somam 1.0 no total. Com isso, uma previsão muito forte corresponderá a uma alta probabilidade (e.g.: 0.9) de crença em uma determinada representação futura; e uma previsão fraca corresponde a variadas probabilidades de crença em diversas possíveis representações futuras.

Nesta pesquisa, optamos pela adoção do modelo de Kuperberg e Jaeger (2016) devido às similaridades entre a Estrutura Hierárquica Gerativa do Processamento da Linguagem e os modelos de processamento baseados em restrições (MACDONALD; PEARLMUTTER; SEIDENBERG, 1994; SPIVEY-KNOWLTON; SEDIVY, 1995; TANENHAUS *et al.*, 1995; TRUESWELL *et al.*, 1993), que serão devidamente descritos no capítulo 2. Em ambas as propostas, é possível verificar a presença de estruturas de processamento probabilístico, assim como de processamento paralelo.

Existe uma proposta de que os mecanismos de previsão atuam de maneira paralela. Kuperberg e Jaeger (2016) apontam que a atualização de uma crença é consistente com a ideia de que podemos calcular preditivamente múltiplos candidatos em paralelo, cada um com diferentes pontos fortes ou graus de crença. Sob a perspectiva dos modelos baseados em restrições, a perspectiva de processamento paralelo assume que o processador linguístico calcula várias análises sintáticas em paralelo, cada uma com algum grau de suporte probabilístico. Isso não implica necessariamente que todas as análises possíveis sejam examinadas exaustivamente, mas sim que múltiplas análises suficientemente prováveis são consideradas em paralelo (CROCKER; BRANTS, 2000; JURAFSKY, 1996; LEWIS, 2000)

Kuperberg (2016) aponta que o desafio que levou ao desenvolvimento de modelos baseados em restrições nos anos 90 foi a dificuldade de se explicar uma série de fenômenos verificados durante o processamento linguísticos como será discutido no capítulo 2. Embora tais modelos lexicais baseados em restrições não sejam suficientes para explicar descobertas de alguns estudos mais atuais que utilizam medidas neurofisiológicas – a exemplo de Chow *et al.* (2016)<sup>3</sup> – a pesquisadora defende que os princípios das abordagens baseadas em restrições não devem ser abandonados.

Sob a perspectiva dos modelos baseados em restrições, propõe-se que os falantes são capazes de usar informações provenientes de seus conhecimentos e suas experiências para contrabalançar possíveis próximas entradas lexicais, antes que as informações representacionais estejam disponíveis. Por exemplo, quando lemos o fragmento de

<sup>3</sup> Detalhado no capítulo 2.

sentença “Ele espalhou no pão quente a...” construímos uma representação parcial do evento “*espalhar algo no pão*”. Antes de decidirmos qual item lexical é o melhor candidato para completar a sentença, usamos a representação de evento para prever os recursos semânticos associados com a palavra “manteiga”, o que também pode promover a provável previsão de palavras plausíveis e semanticamente associadas, como “margarina”, do que a previsão de palavras não associadas, como “meia”. (FEDERMEIER; KUTAS, 1999).

Com isso, os modelos baseados em restrições constituem um estágio importante para o desenvolvimento das perspectivas probabilísticas de processamento linguístico. Eles sugerem que, devido à força da evidência que se pode encontrar em diferentes níveis representacionais, os falantes ativam representações lexicais gradativamente, de maneira a combiná-las para constituir representações de eventos, que, por sua vez, modulam as expectativas para os próximos itens lexicais. Ao propor que a compreensão da linguagem é altamente interativa (ELMAN; HARE; MCRAE, 2004; RUMELHART; MCCLELLAND, 1982), os falantes são considerados capazes de usar rapidamente as fontes de informação provenientes de suas experiências para construir uma representação rica e de alto nível de significado.

Nesta pesquisa, possuímos a hipótese de que propriedades probabilísticas de subseleção de argumentos internos podem diferenciar o comportamento de dois diferentes grupos de verbos. Devido a essas propriedades de subseleção, acreditamos ser possível que informações inerentes a cada grupo de verbo influenciem de maneira distinta a antecipação de informações de nível semântico durante o processamento de palavras em contextos reduzidos.

Acreditamos que verbos com capacidade de subselecionar um menor número de entidades do mundo real podem possuir papéis temáticos mais bem delimitados (MCRAE; FERRETTI; AMYOTE 1997), o que facilitaria a antecipação de traços semânticos de argumentos internos muito esperados. Contudo, para nós, assim como para Ferreti, McRae e Hatherell (2001), McRae, Ferretti e Amyote (1997), Altmann e Kamide (1999) e Kuperberg (2013), os papéis temáticos de um verbo são conceitos formados a partir de experiências cotidianas, durante as quais as pessoas aprendem



sobre quem ou o que desempenha determinados papéis em situações específicas. Esses conhecimentos servem como uma importante restrição na compreensão e produção online da linguagem. McRae, Ferretti e Amyote (1997) sugerem que um papel temático não é verbalizado pelo falante sob a forma de traços gerais como: “agente”, “paciente”, “volição”, “causa”, “senciência” (DOWTY, 1991; SCHLESINGER, 2006). Os falantes, na verdade, são mais propensos a descrever os participantes de um evento sob a forma de traços mais específicos relacionados aos conhecimentos de mundo. Por exemplo, aquele que “acusa” é descrito pelos falantes como possuidor das características de “ser malvado”, “ser crítico”, “ser bem informado”. Dessa forma, como em uma teoria de protótipos, os traços são mais característicos do que definidores.

Para compreender como o comportamento de subseleção dos verbos pode influenciar a delimitação temática de argumentos internos, iniciamos nossa pesquisa observando fragmentos de frases como:

(3) Carla manobrou \_\_\_\_\_

(4) Mariana quebrou \_\_\_\_\_

A observação dos fragmentos nos fez questionar se é mais fácil para o falante antecipar uma continuação de frases como (3) do que de frases como (4). É importante enfatizar que apesar de os dois fragmentos de sentenças apresentarem uma estrutura similar, formada apenas por um nome próprio e um verbo, o fragmento (3) parece atrair um sintagma nominal específico: a palavra “carro”.

Consideramos que a capacidade de verbos subselecionarem argumentos internos não é homogênea e, com isso, características temáticas relacionadas principalmente a conhecimentos sobre eventos do mundo real influenciam de maneira distinta a antecipação de traços semânticos. Por exemplo, é possível que o número de entidades no mundo real que possuem os traços semânticos necessários para ocupar a posição de argumento interno do verbo “manobrar” seja menor do que o número de entidades que possuem os traços semânticos necessários para ser argumento interno do verbo

“quebrar”. Com isso, a pré-ativação dos traços semânticos do argumento interno do verbo “manobrar” ocorreria de maneira mais eficiente probabilisticamente – pois envolveria menos competidores – do que a ativação de traços semânticos do argumento interno do verbo “quebrar”.

Para conhecer o comportamento de subseleção dos verbos e a previsibilidade dos argumentos internos, realizamos diversas tarefas de cloze<sup>4</sup> (TAYLOR, 1953).

Em nossa tarefa de cloze, utilizamos fragmentos de sentenças bastante simples, formados por um nome próprio e um verbo, a fim de restringir a previsibilidade das informações provenientes dos verbos, como:

(5) Thaís reformou \_\_\_\_\_

O verbo “reformatar” exemplifica o comportamento mencionado anteriormente. Nas tarefas de cloze, os participantes devem citar qual a resposta mais provável para um fragmento como (5). No caso, a palavra “casa” foi a continuação considerada mais provável para o fragmento de sentença que continha tal verbo. Embora outras palavras – como “quarto”, “apartamento”, “banheiro” – tenham sido citadas com menores frequências, as respostas citadas na tarefa mostraram que o verbo possui certa propensão a preferir argumentos internos que apresentam traços semânticos bastante específicos – *e.g.*: seja um imóvel ou seja parte de um imóvel.

A tarefa de cloze é utilizada na maioria dos trabalhos empíricos que investigam as propriedades do processamento antecipatório para parametrizar o conceito de restrição lexical (DELONG; URBACH; KUTAS, 2005; FEDERMEIER, 2007; FINE *et al.*, 2013; LAU; HOLCOMB; KUPERBERG, 2013; LAU, 2009; VAN BERKUM *et al.*, 2005; VAN PETTEN; LUKA, 2012; WICHA; MORENO; KUTAS, 2004). Estudos eletrofisiológicos (KUTAS; HILYARD, 1984; VAN PETTEN; KUTAS, 1990) indicam que a probabilidade de cloze de

<sup>4</sup> A metodologia da tarefa de cloze encontra-se descrita na seção 2.4.

uma palavra está fortemente relacionada à amplitude do N400<sup>5</sup>, um componente cerebral fortemente relacionado a fenômenos de previsão de ordem semântica.

Para Federmeier *et al.* (2007), os efeitos de facilitação de palavras em contextos moderadamente e altamente previsíveis, seja em medidas de tempos de resposta, padrões de movimento ocular ou eletrofisiologia do cérebro, é mediado pelo grau de ajuste entre uma palavra específica e um determinado contexto. Tal ajuste é frequentemente definido empiricamente pelo uso de um procedimento de cloze. Kuperberg (2016) sugere que a interpretação dos dados de cloze refletem a força e as probabilidades de expectativas paralelas que o cérebro de um indivíduo calcula durante uma determinada tarefa.

Nosso objetivo nessa tarefa foi conhecer o comportamento de subseleção de um grande número de verbos, a fim de observar a existência de dois prováveis grupos que apresentassem diferentes comportamentos de subseleção. Nossa intenção era identificar se alguns verbos possuíam uma maior força de subseleção de argumentos internos do que outros verbos. Pretendíamos também verificar como as propriedades verbais de subseleção de argumentos internos poderiam diferenciar o comportamento temático dos verbos, fazendo com que existissem papéis temáticos mais bem delimitados em verbos cujo número de argumentos subselecionados fosse menor.

Nesta pesquisa, associamos o efeito de previsibilidade lexical parametrizada por meio da tarefa de cloze (TAYLOR, 1953) à capacidade de subseleção de argumentos verbais. Acreditamos que quanto menor for o número de entidades citadas como previsíveis diante de um determinado contexto, mais forte é a capacidade de subseleção de argumentos internos por parte do verbo, devido ao número limitado de itens lexicais capazes de ocupar a posição de argumento interno. Por consequência, quando um número restrito de entidades é citado na tarefa, acreditamos que as informações temáticas relacionadas a esse verbo serão mais bem delimitadas e mais restritas.

<sup>5</sup> O N400 é uma onda cerebral motivada pela apresentação de um estímulo e cujo pico acontece em torno de 400 ms após o início desse estímulo. Tal componente é bastante relevante para os estudos sobre informações semânticas. Um detalhamento mais completo do componente será apresentado no capítulo 2.

Após a análise das tarefas, verificamos a ocorrência de verbos cuja propensão a subselecionar nomes específicos na posição de argumentos internos é bastante acentuada. O fato de considerarmos que a origem desse comportamento possui relação com propriedades temáticas dos verbos deve-se à utilização de um contexto com carga informacional bastante reduzida, uma vez que os nomes próprios quase não possuem carga semântica.

Com a realização de diversas tarefas de cloze, foi possível a delimitação de dois grupos: (1) verbos de forte subseleção<sup>6</sup> – com maior tendência a subselecionar argumentos internos específicos (*e.g.*: verbo “manobrar”, cujo argumento “carro” é apontado como resposta mais provável em grande parte das respostas da tarefa de cloze); (2) verbos de fraca subseleção<sup>7</sup> – com menor tendência a subselecionar argumentos internos específicos. (*e.g.*: verbo “quebrar”, cujo argumento “perna” figura entre dezenas de outros possíveis argumentos passíveis de ocupar a posição de argumento interno).

Supomos que os verbos de forte subseleção, como o verbo “manobrar”, possuem papéis temáticos bem delimitados devido ao número restrito de traços semânticos formadores das entidades que prototipicamente ocupam a posição de argumento interno desse verbo (*e.g.*: entidades como “carro”, “moto”, “ônibus”, “caminhão”). Nossa hipótese propõe que tal característica favoreça a pré-ativação de representações semânticas dos nomes mais esperados que ocupam a posição de argumento interno durante o processamento de linguístico. Em relação aos verbos de fraca subseleção, como o verbo “quebrar”, acreditamos que os papéis temáticos não são tão bem delimitados, uma vez que o conjunto de entidades que preenchem o papel de argumento interno possui poucas regularidades em seus traços semânticos, sem que se destaque alguma em termos de frequência de execução do papel. É possível que a pré-ativação de representações semânticas em verbos de fraca subseleção seja mais frágil, devido à

<sup>6</sup> Os verbos de forte subseleção foram considerados aqueles cuja palavra mais citada (mais esperada) no teste de cloze possuía probabilidade maior do que 0.50. Mais detalhes serão expostos no capítulo 3.

<sup>7</sup> Os verbos de fraca subseleção foram considerados aqueles cuja palavra mais citada (mais esperada) no teste de cloze possuía probabilidade menor do que 0.50. Mais detalhes serão expostos no capítulo 3.

ocorrência de diferentes possíveis candidatos (*e.g.*: “perna”, “vaso”, “copo”, “pé”, “braço”, “jarra”, “cara”, “vidro”, “carro”...).

Ademais, um detalhe relevante de nosso trabalho diz respeito ao fato de utilizarmos contexto mínimos em nossas tarefas. Na literatura linguística, a definição do que pode vir a ser um contexto é bastante complexa. Segundo Dey (2001), os humanos são bem-sucedidos ao transmitir ideias uns aos outros e ao reagir adequadamente. Isso ocorre devido a muitos fatores, como as propriedades linguísticas compartilhadas, o entendimento comum de como o mundo funciona e a implícita compreensão das situações cotidianas. Quando há comunicação entre humanos, eles são capazes de usar informações situacionais implícitas, ou contexto, para aumentar a amplitude das possibilidades de comunicação. Com isso, adotamos nesta pesquisa a definição segundo a qual entendemos como contexto qualquer informação que pode ser usada para caracterizar uma situação de comunicação ou entidade comunicativa. Assim, ao tratarmos de contextos mínimos, estamos nos referindo a situações nas quais apenas uma ou duas palavras são apresentadas. Mesmo em contextos menores formados por apenas uma palavra, consideramos as informações situacionais da comunicação como componentes do contexto. Além disso, também nos referimos a contextos informativos. Nesse caso, tratamos de contextos em que está disponível uma ampla gama de informações que podem ser utilizadas para caracterizar uma entidade comunicativa, como um item lexical.

Há a possibilidade de conhecimentos temáticos menos detalhados sobre eventos do mundo real – como as relações temáticas entre verbo e argumento interno – tornarem-se mais evidentes caso a representação do contexto não seja suficientemente forte para acessar conhecimentos mais específicos sobre a estrutura de eventos. É como se, em contextos mínimos, o nível de detalhamento da estrutura de um evento ativado pelas representações de alto nível seja extremamente simples, relacionado apenas a traços semânticos de possíveis participantes.

Os resultados de Vieira (2015) sugerem que conhecimentos menos detalhados sobre eventos – preferências de subseleção aferidas a partir de teste de cloze – tornam-se evidentes quando a representação contextual é reduzida ou não fornece informações

suficientes para se sobrepôr às propriedades de subseleção dos verbos. Estudos como os de Camblin *et al.* (2007), Ditman *et al.* (2007) e Van Berkum *et al.* (2003) indicam a existência de tensões entre operações de caráter local (top-down) e processos relacionados aos conhecimentos de alto nível (bottom-up). É possível que, em situações em que o contexto não atue de maneira suficientemente informativa, as informações de caráter local (como a relação entre verbos e argumentos) podem tornar-se mais evidentes.

Huettig e Mani (2016) argumentam que grande parte das evidências sobre processamento preditivo não possui relevância para responder à pergunta sobre a importância exata da previsão para a compreensão da linguagem, visto que a maior parte dos estudos sobre antecipação utilizam frases cujas palavras são extremamente previsíveis e os contextos são bastante informativos. De fato, uma pesquisa realizada por Luke e Christianson (2016) por meio de um grande *corpus* de tarefas de cloze textuais mostrou que apenas 0.05 das palavras de conteúdo são previsíveis.

Atualmente, um pequeno número de estudos se propõe a investigar a previsão de informações de nível semântico em contextos mínimos, compostos por uma ou duas palavras. Algumas dessas exceções são os trabalhos de Delaney-Busch *et al.* (2019), Lau *et al.* (2014) e Weber *et al.* (2016), que utilizam o paradigma de *priming* juntamente com diferentes técnicas eletrofisiológicas para indicar efeitos de adaptação preditiva ao apresentarem blocos pares de palavras relacionadas semanticamente. Com isso, consideramos que a existência de poucos estudos e a ausência de debates a respeito da ocorrência de efeitos de previsão em contextos mínimos não contribuem para as discussões acerca da possibilidade de a antecipação ser um fenômeno inerente ao processamento linguístico que se realiza mesmo em níveis mais simples. Também não contribuem para o caso contrário, pois há a possibilidade de a antecipação ser um processo bastante específico, que demanda contextos altamente restritivos para que seja realizada durante o processamento linguístico.

Nesta tese, além da tarefa de cloze citada anteriormente e apresentada no capítulo 3, realizamos um julgamento de ajuste temático e duas tarefas de decisão lexical. O objetivo geral dessas tarefas foi verificar como tendências de subseleção dos verbos podem diferenciar o comportamento temático de dois grupos de verbos com relação à antecipação de traços semânticos de argumentos internos.

No Capítulo 2, encontram-se detalhados, os fundamentos teóricos que embasam todas as tarefas realizadas durante essa pesquisa. A discussão sobre os resultados revelados nas três tarefas está exposta no Capítulo 6.

No Capítulo 4, apresentamos um julgamento de ajuste temático. Essa tarefa foi realizada com o objetivo específico de verificar se o comportamento temático dos dois grupos de verbos identificados na tarefa de cloze é diferenciado conforme o julgamento dos falantes, que se baseia em conhecimentos de mundo. Por meio da tarefa, verificamos a adequação dos argumentos internos aos verbos citados no cloze.

Por fim, no Capítulo 5 realizamos duas tarefas de decisão lexical a fim de verificar, de maneira cronométrica, se os padrões de resultados identificados no julgamento de ajuste temático podem ser observados na tarefa de decisão lexical. Verificamos se o efeito de previsão de traços temáticos ocorre principalmente entre os verbos de forte subseleção. Ademais, verificamos como o tempo de apresentação dos estímulos pode influenciar o processo de ativação de traços temáticos.

## 2 FUNDAMENTOS TEÓRICOS

### 2.1 Processamento linguístico e antecipação

Este trabalho apoia-se em duas ideias centrais: a noção de que informações contidas no léxico e relacionadas à experiência do falante com o mundo são fundamentais no processamento linguístico e a perspectiva de que o processamento linguístico ocorra a partir da geração de expectativas. Ambas as noções estão expressas, em certa medida, nos modelos de processamento baseado em restrições (MACDONALD; PEARLMUTTER; SEIDENBERG, 1994; SPIVEY-KNOWLTON; SEDIVY, 1995; TANENHAUS *et al.*, 1995; TRUESWELL *et al.*, 1993) e na Estrutura Gerativa Hierárquica de Processamento da Linguagem (KUPERBERG; JAEGER, 2016). Por esse motivo, iniciaremos nosso trabalho realizando uma revisão das principais características dessas duas ideias centrais.

#### 2.1.1 Modelos baseados em restrições

Segundo McRae e Matsuki (2013), modelos baseados em restrições não são modulares, pois qualquer tipo de informação e conhecimento disponível para o falante no momento da compreensão é usado para avaliar potenciais interpretações ao longo do tempo. Tais modelos também são paralelos, pois múltiplas interpretações possíveis são consideradas, e essas interpretações muitas vezes competem entre si. O princípio que subjaz os modelos baseados em restrições propõe que múltiplas fontes de informação – ou restrições – são utilizadas para compreender sentenças e resolver ambiguidades. Essas restrições podem incluir vieses sintáticos gerais, informação probabilística específica do léxico, significado da palavra, restrições seletivas de verbos, conhecimento de eventos do mundo real, vieses pragmáticos contextuais, prosódia da fala, entre outros tipos de informações intrassentenciais e extrassentenciais, incluindo contextos linguísticos e visuais.

Uma das diferenças entre modelos baseados em restrições e outros tipos de modelo de processamento, a exemplo do modelo *Garden Path* (FRAZIER, 1987; FRAZIER; RAYNER, 1982), é a assunção de que se possa existir um pequeno ou nenhum atraso no curso



temporal de ativação das informações que se encontram disponíveis durante o processamento. Contudo, é concebível que a computação de alguns tipos de informação dos modelos baseados em restrições possa levar mais tempo do que outras. Por exemplo, a computação de informações que exigem a combinação conceitual dos significados de várias palavras pode demorar mais do que a computação de informação ligada a uma única palavra, como preferências de subcategorização de verbos.

No modelo, múltiplas potenciais interpretações alternativas são ativadas ao mesmo tempo, em paralelo. Em outras palavras, em qualquer momento, os falantes constroem múltiplas interpretações relevantes das sentenças fornecida. Tais múltiplas interpretações são ponderadas probabilisticamente, uma vez que o sistema atribui valores a cada uma dessas interpretações. Maiores ativações são atribuídas a estruturas que são mais prováveis, e menores ativações são atribuídas a estruturas que são menos prováveis. Uma suposição implícita na maioria dos modelos baseados em restrições é que as estruturas sintáticas competem pela ativação.

Altmann e Steedman (1988) investigaram em inglês SPs (sintagmas preposicionais) que possuem a interpretação temporariamente ambígua, o que os torna passíveis de serem considerados modificadores de SNs (sintagmas nominais) ou fontes de SVs (sintagmas verbais) ao serem situados após SNs. Os autores consideram possível que o contexto referencial seja bastante relevante na interpretação das ambiguidades temporárias. Sob essa perspectiva, a preferência pelos SNs modificadores é favorecida pela necessidade de distinguir entre duas entidades que são denotadas por SNs. Os autores utilizaram pequenas histórias após as quais eram apresentadas sentenças alvo como:

(6) Um ladrão invadiu um banco carregando alguma dinamite. Ele planejava abrir um cofre. Uma vez lá dentro, viu que havia um cofre com uma fechadura nova, um cofre com uma fechadura antiga<sup>9,10</sup>. (Contexto usado nas sentenças alvos; Tradução nossa)

<sup>9</sup> (6) A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside, he saw that there was a safe with a new lock a safe with an old lock.

(7) The burglar blew open the safe with the dynamite and made off with the loot.

(7) Sentença alvo 1 – possui a interpretação do SP como fonte:

O ladrão abriu o cofre **com a dinamite** e fugiu com o saque.

(8) Sentença alvo 2 – possui a interpretação do SP como modificador:

O ladrão abriu o cofre **com a nova fechadura** e fugiu com o saque.

Para Altmann e Steedman (1988), devido à existência de dois possíveis referentes na pequena história – ou seja, a informatividade do contexto – há facilidade na leitura da sentença alvo 2. Sentenças alvo como 1 não continham o referente modificado, deixando implícitas informações bastante necessárias para o estabelecimento da coerência da história, o que demandava maior esforço do leitor.

Assim como o trabalho de Altmann e Steedman (1988), evidências apresentadas por outros pesquisadores sugerem que os modelos baseados em restrições podem abranger uma série de outros fenômenos relacionados a conhecimentos mais amplos do que a traços estritamente linguísticos. Um exemplo é existência dos efeitos de frequência sob as propriedades de subcategorização dos verbos, tratada por Traxler (2011) em uma revisão sobre os modelos baseados em restrições.

Conforme apresentado pelo autor, informações sobre categorias das palavras podem ser muito mais do que simples indicações sobre a formação de estrutura das sentenças. Um exemplo inspirado em Traxler (2011) é o caso das estruturas relacionadas à categoria dos verbos. Se considerarmos dois verbos como *tomar* e *colocar*, é possível verificarmos que, além das diferenças de significado, a estrutura sintática demandada por cada um desses verbos é completamente diferente. Por exemplo, *tomar* pode aparecer em estruturas como:

(8) The burglar blew open the safe with the new lock and made off with the loot.

**10** Neste trabalho, optamos por realizar a tradução dos exemplos apenas em parte dos casos, já que alguns exemplos são mais representativos na língua inglesa. Por causa disso, optamos por não traduzir todos os exemplos apresentados nesta tese.

(9) Maria tomou um suco.

Contudo, sentenças como (10) podem soar estranhas

(10) \*Maria tomou

Com isso, é possível observar que o verbo *tomar* parece demandar um complemento pós verbal. Especificamente, o verbo parece demandar um SN (sintagma nominal) como seu complemento.

Contudo, o caso do verbo *colocar* parece diferente. Podemos observar o comportamento desse verbo por meio dos exemplos a seguir:

(11) \*Maria colocou

(12) \*Maria colocou a carteira

(13) Maria colocou a carteira na bolsa.

*Colocar* de forma semelhante a *tomar* também não pode ocorrer sem um complemento pós verbal. Contudo, esse verbo além de demandar um complemento formado por um SN, também necessita de um complemento que expressa um lugar, formado por um SP. Com isso, embora ambas palavras pertençam à categoria dos verbos, elas possuem propriedades de subcategorização de complementos diferentes. Elas pertencem a subcategorias diferentes, porque possuem requisitos diferentes para suas estruturas sintáticas.

Verbos como *tomar* e *colocar* possuem uma estrutura mais rígida sobre os tipos de complementos que são demandados pelo verbo. Entretanto, existem verbos cujos complementos subcategorizados são mais flexíveis. Vejamos o exemplo do verbo *ler*:

(14) Maria estava lendo.

(15) Maria estava lendo um texto.

Assim, o verbo *ler* é capaz de não subcategorizar qualquer complemento ou pode subcategorizar um complemento formado por um SN. Tais fatos citados acima são uma demonstração de que os verbos podem ser flexíveis em sua estrutura de subcategorização.

Os exemplos citados acima demonstram as diversas possibilidades de subcategorização apresentadas pelos verbos. Os modelos baseados em restrições pressupõem que as pessoas consideram preferências de subcategorização durante o processamento incremental da sentença. As informações sobre preferências de subcategorização refletem as probabilidades de os verbos aparecerem com maior frequência relacionados a determinadas estruturas do que a outras. Por exemplo, se considerarmos o fragmento de sentença:

(16) Maria percebeu ...

É possível que o fragmento possa ser completado por um SN (sintagma nominal), como:

(17) ...o início do incêndio.

Também é possível que o fragmento seja completado por um SC (sintagma complementizador), formando uma nova sentença, como:

(18) ...que o incêndio terminou.

Neste caso, não sabemos qual é o tipo de estrutura mais comum a um fragmento de sentença composto pelo verbo *perceber*. Contudo, para os modelos baseados em restrições essa informação é bastante relevante, pois, essencialmente, o processador linguístico impõe restrições às estruturas que são menos frequentes no uso da língua. No caso, o processador ativa informações sobre o passado – por exemplo, a probabilidade de que uma determinada estrutura ocorra quando um dado verbo estiver

presente – e as utiliza para prever estruturas probabilisticamente mais prováveis. Em outras palavras, os modelos baseados em restrições pressupõem que as pessoas ativam informações sobre preferências de subcategorização. As preferências de subcategorização refletem as probabilidades de que determinadas estruturas e determinados verbo sejam utilizados em conjunto. Com isso, é possível que o analisador preveja, por exemplo, que um SC tenda a ocorrer mais frequentemente do que um SN, e, por isso, essa estrutura sintática passe a receber um peso maior. Tal peso é o que contribui para que o verbo seja geralmente complementado por um SC. Contudo, quando o fragmento de sentença fornece uma entrada para uma estrutura diferente (um SN, no caso), o analisador baseado em restrição precisa readaptar suas informações probabilísticas sobre restrições para que seja possível a formação de uma estrutura pouco frequente. Essa interpretação de que preferências de subcategorização de complementos são direcionadas por um viés frequentista está presente em diversos trabalhos que utilizam a perspectiva dos modelos baseados em restrições (GARNSEY *et al.*, 1997; MITCHELL; HOLMES, 1985; TRUESWELL; TANENHAUS; KELLO, 1993).

Preferências temáticas acerca de entidades mais propícias a ocuparem determinados papéis em eventos do mundo real também são evidências de que as experiências do falante são essenciais para o processamento linguístico. McRae, Spivey-Knowlton e Tanenhaus (1998) observaram que, em muitos casos, verbos podem aparecer tanto na voz ativa quanto na voz passiva, e, para muitos verbos, existem sintagmas nominais que são melhores ao preencher a função de agente do que o papel de paciente e vice-versa.

A hipótese de McRae *et al.* (1998) previa que, quando confrontados com fragmentos de frase que eram ambíguos entre uma leitura de verbo principal e uma leitura de relativa reduzida<sup>11</sup>, os leitores poderiam ser influenciados pelo sujeito inicial SN, caso houvesse um agente ou paciente muito provável. Caso houvesse um agente, isso deveria encorajar uma interpretação do verbo principal; caso houvesse um paciente, uma relativa reduzida deveria ser favorecida.

<sup>11</sup> Em inglês, há uma semelhança na forma do particípio verbal e do pretérito verbal. Em geral, ambos terminam em **ed ou d**. Com isso, uma forma ambígua como *arrested* pode ter os significados de *preso* e *preendeu*.

Tal fato foi precisamente o que McRae e colegas encontraram em seus resultados. Fragmentos de sentenças como *The cop arrested*<sup>12</sup> ... promoveram a interpretação do verbo principal em vez de uma interpretação de relativa reduzida. No caso de fragmentos como *The criminal arrested*<sup>13</sup>..., a probabilidade da leitura da relativa reduzida foi aumentada. Cada uma dessas leituras é motivada pelo fato de os falantes saberem que no mundo real é mais frequente *um policial prender alguém* e *um criminoso ser preso*. Tendo em vista esses resultados, McRae *et al.* concluíram que as especificações de papéis temáticos dos verbos devem ir além da simples informação categórica, como agente, paciente, instrumento, beneficiário e assim por diante. Os dados experimentais sugerem que as informações temáticas são detalhadas no que se refere às preferências de preenchimentos desses papéis, e que as preferências são específicas dos verbos. Essas preferências são expressas não apenas no que se refere ao preenchimento dos papéis nominais, mas também no que se refere a seus atributos.

Além do viés probabilístico sobre o processamento linguístico, os modelos baseados em restrições apresentaram as primeiras evidências acerca da ocorrência de processos antecipatórios no processamento linguístico. Em um dos primeiros estudos a respeito de processos antecipatórios na linguagem, Altmann e Kamide (1999) utilizaram o paradigma do mundo visual<sup>14</sup> para verificar se a interação entre informações temáticas do verbo e conhecimentos sobre eventos do mundo real permite a previsão de complementos verbais.

No experimento, os pesquisadores observaram se, ao ouvir frases como “a mulher vai beber...”, os participantes olhavam para a figura de uma taça de vinho, o único referente líquido dentre outros expostos na tela do computador – os outros referentes eram: um queijo, um batom, uma cadeira e uma planta. Para os autores, os resultados sugerem que informações temáticas provenientes de verbos e informações visuais podem ser

<sup>12</sup> *O policial prendeu* (Tradução nossa).

<sup>13</sup> *O criminoso preso* (Tradução nossa).

<sup>14</sup> O paradigma do mundo visual é uma técnica de rastreamento ocular em que “frases são apresentadas auditivamente aos participantes enquanto um rastreador ocular registra, com precisão, para onde o sujeito está olhando enquanto ouve o estímulo linguístico” (GODOY; VIEIRA; CUNHA LIMA, 2016).

usadas para antecipar a identificação de referentes sujeitos a serem mencionados. Como o olhar dos participantes era direcionado antes da menção ao referente, ainda no momento em que o verbo estava sendo processado, assume-se que o processador age de modo antecipar possíveis referentes.

É possível argumentar que os resultados verificados por Altmann e Kamide (1999) possuem relação com mecanismos de restrições seletivas puramente sintáticas, que são ativados durante o processamento incremental do verbo *beber*. Contudo, o rápido e ativo uso da informação visual para a previsão do referente correto e os resultados de um estudo de Kamide, Altmann e Haywood (2003)<sup>15</sup> demonstram que os falantes utilizam conhecimentos de eventos para prever possíveis referentes em um contexto visual durante a articulação incremental da fala.

Portanto, como pode ser demonstrado após a revisão dos estudos citados, é possível afirmar que os modelos baseados em restrições incluem em sua perspectiva a possibilidade de ocorrência de antecipações ou expectativas de estrutura e conteúdo. Os exemplos demonstram que a alocação de recursos provenientes da experiência do falante auxilia no processo paralelo, incremental e probabilístico da compreensão da sentença. Com isso, a partir da próxima seção apresentaremos uma visão mais detalhada sobre o que entendemos como processamento linguístico antecipatório e quais as propriedades do modelo de Estrutura gerativa hierárquica do processamento da linguagem adotado nesta pesquisa.

<sup>15</sup> A ser detalhado na seção 2.2.

### 2.1.2 Processamento antecipatório

A discussão a respeito da atuação de mecanismos de antecipação na cognição humana não se restringe ao processamento linguístico. Em sentido amplo, o processamento antecipatório refere-se a qualquer tipo de processo que incorpore ou gere não apenas informações sobre o passado ou o presente, mas também sobre estados futuros do corpo ou do ambiente. Historicamente, a investigação sobre mecanismos de antecipação iniciou-se quase paralelamente aos estudos sobre processamento perceptual e motor. O interesse pela inter-relação entre percepção e ação nasceu da busca por explicações para uma questão fundamental em nossa cognição: como o nosso mundo visual permanece estável, apesar do constante deslocamento de imagem introduzido pelos movimentos dos olhos e da cabeça? Uma das primeiras propostas a relacionar expectativas e ações foi formulada no século XIX dentro do princípio ideomotor, que recentemente foi atualizado por teorias que sugerem a existência de códigos compartilhados entre percepção e ação (BUBIC; VON CRAMON; SCHUBOTZ, 2010; CLARK, 2013).

Estudos em Psicologia Cognitiva e em Neurociência indicam que a capacidade de antecipar informações pode influenciar diversas funções cognitivas, como o processamento atencional e o processamento de informações para tomada de decisões. Tal direcionamento em relação ao futuro tem sido reconhecido como relevante e benéfico para diferentes aspectos do processamento de informações, tais como percepção, aprendizado implícito, teoria da mente e outros processos cognitivos em humanos, e, de uma forma mais rudimentar, em animais (c.f. BUTZ, 2004; CLARK, 2013 para revisões sobre o assunto).

Clark (2013) apresenta uma proposta bastante relevante para os modelos de processamento linguístico adotados nesta pesquisa. Sob tal perspectiva, cérebros são compreendidos essencialmente como máquinas de previsão, uma vez que são feixes de células que suportam a percepção e a ação ao combinar constantemente entradas sensoriais de níveis mais baixos – como informações sensoriais – com previsões de níveis cognitivos mais altos – como conhecimentos pragmáticos sobre o mundo real. Essas atividades parecem ocorrer por meio de um modelo gerativo hierárquico, cujo



objetivo é minimizar os erros de previsão dentro de uma cascata bidirecional de processamento cortical. Para o autor, essa perspectiva proporciona um modelo unificador da percepção e da ação, elucida o papel funcional da atenção e pode registrar a contribuição do processamento cortical para o sucesso de adaptação filogenética.

Sob a perspectiva linguística, o conceito de antecipação está relacionado à capacidade de o cérebro lidar com a velocidade e a complexidade do processamento linguístico por “pensar à frente”, gerando informações sobre prováveis próximos estímulos e preparando-se antecipadamente para o processamento desses estímulos, em múltiplos níveis (FEDERMEIER, 2007). Nesta pesquisa, assim como Kuperberg e Jaeger (2016), consideramos que a capacidade de previsão do falante está apoiada em uma representação interna do contexto. Essa representação interna é composta por informações linguísticas e não linguísticas – que podem incluir desde representações parciais inferidas de entradas contextuais previamente processadas, até representações de nível superior, como representações de eventos. Sob essa perspectiva, a representação interna do contexto do falante altera o estado do sistema de processamento linguístico antes do surgimento de novos estímulos que estejam em níveis inferiores – *e.g.*: representações mentais de níveis ortográfico, fonético, fonológico.

Diferentes métodos experimentais sugerem que previsões de informações de nível semântico ocorrem durante processamento linguístico. No desenvolvimento desta tese, consideramos fundamental evidenciar o impacto que cada um deles possui para delimitação do fenômeno da previsão linguística.

Em um dos estudos a respeito de processos preditivos na linguagem, Kamide, Altmann e Haywood (2003) utilizaram o paradigma do mundo visual para verificar se a interação entre informações temáticas do verbo e conhecimentos sobre eventos do mundo real permite a previsão de complementos verbais. No experimento, os pesquisadores observavam como participantes reagiam ao ouvir fragmentos de sentenças como “*the*

*girl will ride*<sup>16</sup>...” ou “*the man will ride*”<sup>17</sup>, enquanto olhavam para as figuras de um homem, de uma criança, de um carrossel com vários cavalinhos e de uma motocicleta. Durante a articulação da frase “*the girl will ride*” a maioria dos participantes antecipavam seu olhar para o carrossel antes mesmo da palavra “*carousel*”<sup>18</sup> fosse articulada. No caso do fragmento “*the man will ride*”, os participantes antecipavam o olhar para a figura da motocicleta, antes mesmo da articulação da palavra “*motorcycle*”<sup>19</sup>. Para os autores, os resultados sugerem que informações temáticas provenientes do conhecimento sobre eventos do mundo real dos falantes e informações visuais podem ser usadas antecipadamente na identificação de referentes sujeitos a serem mencionados. Como o olhar dos participantes era direcionado antes da menção ao referente, ainda no momento em que o verbo estava sendo processado, assume-se que o processador age de modo antecipar possíveis referentes. O ponto de principal divergência entre esse experimento e o de Altmann e Kamide (1999) diz respeito à impossibilidade de atribuição das informações antecipadas a restrições seletivas do verbo. Como os dois fragmentos de sentenças utilizavam o mesmo verbo “*ride*”, apenas o conhecimento de mundo poderia permitir ao falante associar o verbo “*ride*” ao argumento “*carousel*” quando há referência a uma criança. Essa particularidade do desenho experimental elimina a possibilidade de se atribuir aos mecanismos de restrição seletiva a escolha do referente antecipado.

Embora os resultados dos estudos de Altmann e Kamide (1999) e de Kamide, Altmann e Haywood (2003) indiquem que os participantes utilizaram ativamente seus conhecimentos de eventos para prever as entidades a serem referidas, os contextos utilizados em experimentos desse tipo são considerados bastante restritivos, o que, de certa forma, aumenta a probabilidade de antecipação de referentes. Em paradigmas experimentais como esse, a informação visual possui o poder de direcionar a compreensão do participante, reduzindo ao mínimo o número de entidades que provavelmente participariam do evento.

<sup>16</sup> No caso desse experimento, consideramos que a tradução mais aproximada seja “a garota vai montar”.

<sup>17</sup> “O homem vai montar” (tradução nossa).

<sup>18</sup> “Carrossel” (tradução nossa).

<sup>19</sup> “Motocicleta” (tradução nossa).

Estudos em rastreamento ocular durante a leitura também indicam a atuação de processos antecipatórios no processamento linguístico. Staub (2015) realiza uma extensa revisão crítica de trabalhos que utilizam rastreamento ocular para investigar propriedades de previsibilidade lexical. O autor afirma que a previsibilidade influencia medidas de rastreamento ocular bastante iniciais como a primeira fixação, a duração do primeiro olhar (MATSUKI *et al.*, 2011) e o salto de palavras. Tal influência em medidas tão iniciais é equivalente à influência de fatores considerados de níveis pré-lexicais ou lexicais de representações linguísticas, como informações visuais e ortográficas ou a frequência das palavras.

Contudo, o conjunto de resultados que trouxeram maior número de evidências sobre a ocorrência de processos preditivos durante o processamento linguístico foi obtido por meio da técnica eletrofisiológica de potenciais evocados. Nessa técnica, os potenciais elétricos produzidos no cérebro são aferidos em intervalos específicos de tempo e relacionados diretamente a manipulações experimentais pré-estabelecidas. O principal objetivo desse procedimento é relacionar as ativações eletrofisiológicas geradas pelo cérebro aos processos cognitivos desempenhados especificamente nas tarefas experimentais.

Dentre os componentes de potenciais evocados relacionados especificamente ao processamento da linguagem, o N400, um componente de polaridade negativa cujo pico acontece a partir de 400ms da apresentação de um estímulo, mostrou-se especialmente relevante para os estudos sobre antecipação. Na década de 80, o N400 tornou-se conhecido na literatura psicolinguística como um índice de utilização de informações de significado. Em geral, todas as palavras com conteúdo semântico produzem o N400. A amplitude desse componente é sensível a fatores relacionados à facilidade de acesso a informações na memória – como frequência da palavra, repetição e vizinhança ortográfica (KUTAS; VAN PETTEN; KLUENDER, 2006)

Além disso, existem evidências (KUTAS; HILLYARD, 1980, 1984) de que a amplitude do componente também é influenciada – de modo inversamente proporcional – pela expectativa de ocorrência de uma palavra em um determinado contexto discursivo. Kutas e Hillyard (1980) mostraram que durante a leitura de frases como “Eu prefiro café

com creme e...” palavras conceitualmente esperadas, como a palavra “açúcar”, provocavam a diminuição da amplitude do N400 se comparadas a palavras conceitualmente inesperadas, como a palavra “meias”.

É possível que a redução da amplitude desse componente seja motivada por informações de cunho semântico que podem ser previsíveis, pois os traços conceituais dessas palavras são pré-ativados com base no contexto (KUTAS; FEDERMEIER, 2000; LAU; PHILLIPS; POEPEL, 2008).

Tendo em vista a possibilidade de investigar o comportamento de informações de cunho semântico por meio desse componente, pesquisadores buscaram desenhos experimentais especificamente destinados a verificar a ocorrência de antecipação durante o processamento linguístico. Em um estudo seminal, DeLong, Urbach e Kutas (2005) encontraram evidências de processos de antecipação durante a compreensão da leitura. No experimento, os pesquisadores usaram manipulações que exploravam o conhecimento de mundo de leitores e a variação entre os artigos indefinidos do inglês *a* e *an*. Embora possuam significados idênticos, a ocorrência de cada um desses artigos é direcionada pelo fonema inicial da palavra seguinte. Os pesquisadores criaram frases que poderiam ser continuadas por palavras esperadas ou palavras inesperadas, porém plausíveis. O artigo adequado para a palavra esperada era sempre diferente daquele adequado para a palavra inesperada, conforme exemplificado a seguir:

<b>Continuação esperada</b>
<p><i>The day was breezy so the boy went outside <b>to fly a kite</b> in the park .20</i></p> <p>* Ventava naquele dia, então o menino saiu para empinar <b>uma pipa</b> no parque</p>
<b>Continuação inesperada</b>
<p><i>The day was breezy so the boy went outside <b>to fly an airplane</b> in the park</i></p> <p>* Ventava naquele dia, então o menino saiu para empinar <b>um avião</b> no parque.</p>

<sup>20</sup> Nesta tese, optamos pela apresentação de alguns exemplos em inglês com suas traduções em português por considerarmos que o traço latente em questão é mais visível na língua inglesa.

Figura 1 - Quadro com as condições experimentais de DeLong, Urbach e Kutas (2005)

Os resultados mostraram que a amplitude do N400 nos artigos *a* e *an* era influenciada pela expectativa por certas palavras <sup>21</sup>. Artigos adequados para palavras esperadas (*e.g.* **a** para **kite**) apresentavam um N400 significativamente menor do que o N400 encontrado em artigos inadequados para palavras esperadas, mas adequados para palavras inesperadas (*e.g.* **an** para **airplane**). Uma questão importante é o fato de que os participantes previram os sintagmas nominais esperados, antes mesmo de visualizá-los. Os resultados mostram que os participantes utilizaram a representação contextual e conhecimento sobre eventos do mundo real para prever que, em um dia de muito vento, é mais provável que um menino empine uma pipa (*kite*) do que um avião (*airplane*). Como as duas formas do artigo possuem significados idênticos, esses resultados não podem ser relacionados à facilidade/dificuldade de integrar informações.

Diante da importância do estudo de DeLong e colegas (2005) para as teorias sobre o processo de antecipação no processamento linguístico, é importante mencionar a replicação realizada por Nieuwland e Politzer-Ahles *et al.* (2018) do estudo de 2005. Os resultados do estudo, que envolveu 9 laboratórios e 334 sujeitos, replicaram com sucesso os efeitos de redução do N400 nos nomes, mas falharam em replicar os resultados da antecipação fonológica de artigos. Para os autores, os resultados de DeLong e colegas (2005) não mostram evidências suficientes de que as pessoas pré-ativam rotineiramente o fonema inicial de um substantivo próximo, ou talvez qualquer outra informação de baixo nível, a respeito da forma fonológica ou ortográfica da palavra. Sem a pré-ativação do fonema inicial, a instanciação do artigo não faz com que as pessoas revisem sua previsão sobre o significado do próximo substantivo, não havendo, portanto, qualquer impacto no processamento. Crucialmente, esta conclusão é

<sup>21</sup> Os artigos indefinidos *a* e *an* do inglês possuem um comportamento distinto dos artigos indefinidos do português, pois não realizam concordância de gênero gramatical. No caso dos artigos em inglês, se trata de um efeito puramente fonológico, motivado pela cliticização do artigo, levando um sintagma nominal como *an airplane* a ser realizado como uma única palavra fonológica. Não há informação morfológica em ação, contrariamente ao que ocorre em português.

incompatível com a visão de previsão fonológica forte, pois sugere que a pré-ativação não ocorre no nível de detalhamento que é frequentemente assumido.

Contudo, Nieuwland e Politzer-Ahles *et al.* (2018) não refutam a importância dos processos antecipatórios para a compreensão da linguagem. Também não descartam a ideia de que leitores podem eventualmente antecipar formas fonológicas de palavras previsíveis. Os autores alegam que os resultados, no entanto, não devem ser tomados como evidência contra previsão no processamento de linguagem, uma vez que os resultados relacionados aos efeitos nos nomes se aproximam de outros estudos que indicam a ocorrência de processamento preditivo (OTTEN; VAN BERKUM, 2008; VAN BERKUM *et al.*, 2005; WICHA; MORENO; KUTAS, 2004)<sup>22</sup>.

Portanto, apesar do grande número de evidências experimentais que se acumulou ao longo das últimas décadas<sup>23</sup>, é evidente que muitas questões a respeito dos processos preditivos durante o processamento linguístico ainda se encontram em aberto. Tal fato revela as necessidades de delineamentos experimentais mais cuidadosos e de investigações mais específicas sobre previsão. Na próxima seção, abordaremos brevemente a relevância de investigações sobre questões específicas – como níveis representacionais pré-ativados durante o processamento linguístico – para a delimitação de um modelo de processamento preditivo.

## **2.2 Perspectivas sobre a antecipação no processamento linguístico**

Como o trabalho de Nieuwland e colegas (2018) nos mostra, atualmente, os questionamentos acerca da existência de mecanismos preditivos na arquitetura do processamento linguístico estão dando lugar a estudos que têm por objetivo compreender as características desses mecanismos. Perguntas sobre quando a previsão ocorre e quais mecanismos são responsáveis por desencadeá-la ainda requerem

<sup>22</sup>Existem questionamentos acerca da metodologia e da análise inferencial utilizados por Nieuwland e Politzer-Ahles *et al.* (2018). Para maiores detalhes, veja as respostas de Yan, Kuperberg e Jaeger (2017) e DeLong, Urbach e Kutas (2017).

<sup>23</sup> Para revisões sobre o assunto, confira Van Pette e Luka (2012) e Kuperberg e Jaeger (2016).

investigações. Huettig (2015) afirma que embora haja progressos nas pesquisas acerca do tipo de informações utilizadas, ainda existem muitas questões a serem investigadas a respeito dos processos de previsão. Pouco se conhece sobre os níveis de representações ativadas e sobre o quanto essas ativações são dependentes do contexto.

Huettig e Mani (2016) questionam a obrigatoriedade da antecipação para a compreensão da linguagem e criticam a metodologia empregada em diversos estudos. Os pesquisadores propõem que os mecanismos de antecipação podem auxiliar na compreensão em situações específicas, mas o processamento linguístico não necessita frequentemente da previsão para ser realizado com sucesso. Uma primeira evidência apresentada por Huettig e Mani (2016) aponta que nem todas as pessoas realizam o processo de antecipação. Eles abordam o trabalho de Mishra *et al.* (2012) no qual foi comparada, por meio do paradigma do mundo visual (c.f. nota 7), a capacidade de antecipação de pessoas pouco e muito letradas. Os adultos muito letrados demonstraram a capacidade de prever objetos alvo muito maior do que a de adultos pouco letrados após ouvir um adjetivo que auxiliaria na previsão de um objeto. Pessoas muito letradas deslocaram seus olhos para os alvos aproximadamente 1000 ms antes dos pouco letrados. Os olhares das pessoas iletradas para o objeto alvo só ocorriam após o começo da palavra alvo. Para Huettig e Mani (2016), os resultados são uma indicação de que adultos pouco letrados não realizam o processo de antecipação durante o processamento linguístico, e que esse processo antecipatório é fortemente modulado por fatores como as habilidades de letramento dos ouvintes.

Ademais, o tempo necessário para a ocorrência da antecipação também é apontado pelos pesquisadores como um fator que demonstra a não obrigatoriedade dos processos de antecipação. O fator curso-temporal necessário para realização do processamento antecipatório é relevante nesta pesquisa. Isso ocorre, pois, nas duas tarefas on-line de decisão lexical realizadas durante esta pesquisa, utilizamos diferentes cursos temporais para verificar a ocorrência da antecipação. Não sabemos se o processamento de palavras em contextos mínimos demanda um maior ou menor curso-temporal. Pretendemos verificar tal questão por meio de nossos experimentos.

Huettig e Guerra (2015) verificaram que os efeitos da antecipação podem desaparecer quando os participantes não recebem informações por um tempo adequado. Participantes eram incentivados a olhar para um grupo de objetos enquanto ouviam frases como “olhe para o piano”. A pista para o processo antecipatório era fornecida pelo gênero do artigo definido. Os olhares antecipatórios para o objeto alvo só foram verificados quando os participantes puderam observar o *display* com objetos durante 4 segundos antes de ouvir a frase. Quando o tempo de apresentação do display foi de 1 segundo, olhares antecipatórios para o objeto alvo só foram identificados quando o ritmo de articulação da frase foi mais lento.

Os resultados de Huettig e Guerra (2015) estão em consonância com o trabalho de Chow *et al.* (2016), já que ambos encontraram evidências de que a previsão de verbos em final de sentenças é sensível aos papéis temáticos dos argumentos caso haja um maior curso temporal para que a informações antecipatórias estejam disponíveis. Por meio de sentenças em Mandarim, os pesquisadores manipularam a distância linear entre argumentos e verbos com a utilização de um advérbio, em sentenças cuja adaptação para o português brasileiro ocorreria na seguinte ordem:

(19) “Ontem à tarde – o policial – o bandido – prendeu” (Tradução nossa)

(20) “O policial – o bandido – ontem à tarde – prendeu”<sup>24</sup> (Tradução nossa)

Por meio de medidas de N400, foram encontrados indicativos de previsão do verbo final apenas nas sentenças com ordem semelhante a (20). Os pesquisadores argumentam que a previsão envolve cálculos que podem exigir diferentes quantidades de tempo, da mesma forma como em aritmética alguns cálculos podem ser mais fáceis ou mais difíceis de serem previstos. Os resultados de Huettig e Guerra (2015) e Chow *et al.* (2016) podem ser compreendidos como uma indicação de que o processamento preditivo

24 (a) “zuotian-xiawu {jingcha / xiaotou} ba {xiaotou / jingcha} zhua-le...  
yesterday-afternoon {cop} BA {thief} arrest...”

(b) “{jingcha / xiaotou} ba {xiaotou / jingcha} zai zuotian-xiawu zhua-le...  
{cop} BA {thief} ZAI yesterday-afternoon arrest...”



precisa atingir um certo limiar de tempo antes que uma ação comportamentalmente observável seja iniciada. Em vista disso, é possível que, para que haja antecipação de informações linguísticas, um certo patamar de tempo seja necessário.

Huettig e Mani (2016) argumentam que grande parte das evidências sobre processamento preditivo não possui relevância para responder à pergunta sobre a importância exata da previsão para a compreensão da linguagem, visto que a maior parte dos estudos sobre antecipação utilizam frases cujas palavras são extremamente previsíveis (cloze alto). Para os autores, estudos com palavras de menor previsibilidade (cloze baixo) devem ser mais bem explorados, pois é possível que a antecipação não seja um processo imprescindível para a compreensão. De fato, um estudo (LUKE; CHRISTIANSON, 2016) realizado por meio de um grande *corpus* de tarefas de cloze textuais, com 2689 palavras em 55 passagens, mostrou que apenas 0.05 das palavras de conteúdo são previsíveis.

Em geral, a maior parte dos trabalhos que tratam dos processos de antecipação mensura a previsibilidade dos itens lexicais por meio da tarefa de cloze (TAYLOR, 1953). Na literatura psicolinguística, a tarefa de cloze é amplamente utilizada para que seja verificado o quão esperado pode ser um item lexical diante de determinados fragmentos de sentenças ou de textos. Embora comumente usada para avaliar a capacidade de pequenas passagens (OLLER, 1973; ANDERSON, 2005), a tarefa é frequentemente utilizada para quantificar previsibilidade de palavras após contextos sentenciais e discursivos (para uma completa revisão sobre esse tipo de aplicação da tarefa, veja VAN PETTEN; LUKA, 2012). Por exemplo, em uma tarefa de cloze, é possível mensurar a probabilidade da palavra “bolo” ocorrer após o fragmento “A confeitadeira vai decorar...” por meio da porcentagem de respostas que citam a palavra “bolo” como a palavra que provavelmente dará seguimento ao fragmento de sentença<sup>25</sup>.

Atualmente, existem poucos estudos que utilizam palavras pouco previsíveis ou imprevisíveis, porém plausíveis. A crítica de Huettig e Mani (2016) evidencia a

<sup>25</sup> No Capítulo 3, as especificidades da tarefa de cloze serão detalhadas.

necessidade de pesquisas que utilizem contextos mínimos e palavras de tendências de subseleção, variáveis que são manipuladas neste estudo.

Como foi citado em nossa introdução, a literatura linguística possui dificuldades para determinar uma definição para o termo “contexto”. O sucesso comunicativo dos humanos deve-se à capacidade de transmitirem ideias uns aos outros e ao fato de reagirem adequadamente às informações comunicativas. Isso ocorre devido a muitos fatores como as propriedades linguísticas compartilhadas, o entendimento comum de como o mundo funciona e a implícita compreensão das situações cotidianas (DEY, 2001). Durante a comunicação, os humanos são capazes de usar informações situacionais implícitas, ou contexto, para aumentar a amplitude das possibilidades de comunicação. Diante dessa perspectiva, adotamos nesta pesquisa a definição segundo a qual o contexto é qualquer informação que pode ser usada para caracterizar uma situação de comunicação ou entidade comunicativa. Assim, ao tratarmos de contextos mínimos, estamos nos referindo a situações nas quais apenas uma ou duas palavras são apresentadas. Mesmo em contextos menores, formados por apenas uma palavra, consideramos que as informações situacionais da comunicação são componentes do contexto. Além disso, também nos referimos a contextos informativos. Nesse caso, tratamos de contextos em que está disponível uma ampla gama de informações que pode ser utilizada para caracterizar uma entidade comunicativa, como um item lexical.

É possível que haja tensões entre operações de caráter local (*top-down*) e processos relacionados aos conhecimentos de alto nível (*bottom-up*). É possível também que conhecimentos temáticos menos detalhados sobre eventos do mundo real – como as relações temáticas entre verbo e argumento interno – tornem-se mais evidentes caso a representação discursiva não seja suficientemente forte para acessar conhecimentos mais específicos sobre a estrutura de eventos. É como se, em contextos mínimos, o nível de detalhamento da estrutura de um evento ativado pelas representações de alto nível seja extremamente simples, relacionado apenas a traços semânticos de possíveis participantes. A exemplo do que consideramos contextos mínimos, podemos citar os contextos formado por um nome próprio e uma palavra, com em “João derramou” ou contextos experimentais formados por uma só palavra nas tarefas do paradigma de *priming*.

Paczynski e Kuperberg (2012) encontraram evidências eletrofisiológicas de que a semelhança semântica de uma palavra com um contexto precedente pode diminuir os custos de violações de conhecimento detalhado sobre os eventos do mundo real, mas não os custos de violações de animacidade.

(21) The pianist played his music while the bass was strummed by the drummer during the song.<sup>26</sup>

(22) The pianist played his music while the bass was strummed by the gravedigger during the song.<sup>27</sup>

(23) The pianist played his music while the bass was strummed by the drum during the song<sup>28</sup>.

(24) The pianist played his music while the bass was strummed by the coffin during the song<sup>29</sup>.

Ao comparar sentenças como (21) e (22), os pesquisadores verificaram que o N400 em *drummer* era menor do que o N400 em *gravedigger*, devido aos conhecimentos detalhados sobre eventos do mundo real. Contudo, esse padrão de resultado não acontece entre as sentenças (23) e (24). Mesmo que *drum* tenha traços semânticos em comum com o contexto discursivo, o N400 aferido nessa palavra foi semelhante ao aferido em *coffin*. Esses resultados sugerem que em situações em que o contexto não é robusto o suficiente, alguns tipos de conhecimentos de eventos relacionados à representação da palavra (*e.g.* animacidade) não são sobrepostos por processos amplos de ativação da estrutura interna do contexto, fazendo com que conhecimentos

<sup>26</sup> O pianista tocou sua música, enquanto o baixo foi dedilhado pelo baterista durante a canção (Tradução nossa).

<sup>27</sup> O pianista tocou sua música, enquanto o baixo foi dedilhado pelo coveiro durante a canção (Tradução nossa).

<sup>28</sup> O pianista tocou sua música, enquanto o baixo foi dedilhado pelo tambor durante a canção (Tradução nossa).

<sup>29</sup> O pianista tocou sua música, enquanto o baixo foi dedilhado pelo caixão durante a música (Tradução nossa).

detalhados sobre eventos do mundo real não se sobreponham a conhecimentos relacionados à representação da palavra.

Vieira (2015) sugere que conhecimentos sobre eventos do mundo real – previsões aferidas por meio de tarefa de cloze acerca de certos tipos de argumento internos que podem normalmente ser relacionados a certos tipos de eventos – tornam-se evidentes quando a representação contextual é reduzida. Ao comparar as medidas de rastreamento ocular em frases como (25), (26), (27) e (28), foram encontradas evidências de que os tempos totais de leitura e os números de fixação são menores em argumentos internos previsíveis (*e.g.*: soldados) do que em argumentos internos imprevisíveis (*e.g.*: atletas). Esse resultado ocorre independentemente de o contexto ser formado por um arranjo entre um nome de uma entidade específica (*e.g.*: coronel) e um verbo (*e.g.*: recrutar) ou por um arranjo entre um nome próprio (*e.g.*: Paulo) e um verbo.

(25) O coronel recrutou os **soldados** alemães durante a guerra.

(26) O coronel recrutou os **atletas** juvenis durante uma competição.

(27) Paulo recrutou os **soldados** alemães durante a guerra.

(28) Paulo recrutou os **atletas** juvenis durante uma competição.

Os resultados encontrados por Vieira (2015) deixam em aberto a questão do quão informativo deve ser o contexto discursivo para que conhecimentos mais refinados sobre eventos do mundo real sejam utilizados para prever argumentos verbais. Consideramos que o fato de existirem poucos estudos a respeito da ocorrência de efeitos de antecipação em contextos muito simples dificultam as discussões acerca da possibilidade de o fenômeno ser inerente ao processamento linguístico, ou seja, que ele sempre se realize, mesmo em níveis mais simples. Caso o fenômeno não possua essas características, existe a possibilidade de a antecipação ser um processo bastante específico, que demande contextos altamente restritivos para que seja realizada durante o processamento linguístico.

Por meio de uma perspectiva diferente de Huettig e Mani (2016), ao considerar que a antecipação é um fenômeno probabilístico, gradual e relacionado à diferentes níveis representacionais, nossas hipóteses baseiam-se em Kuperberg e Jaeger (2016), que propõem um *Estrutura Gerativa Hierárquica de Processamento da Linguagem*.

Nessa perspectiva, a previsão é considerada um fenômeno probabilístico, pois, segundo os autores, durante o processamento, os falantes utilizam informações de alto nível para controlar as diferentes probabilidades de expectativas paralelas a serem consideradas durante o processamento linguístico. Com isso, diferentes possibilidades são consideradas ao mesmo tempo e cada uma representa diferentes graus de crença. Caso uma entrada de nível mais baixo descon firme uma crença de alto grau probabilístico, o processador se readaptará para que uma nova crença seja reconsiderada probabilisticamente. No que diz respeito à gradualidade do fenômeno, é proposto que a magnitude, a força com que o efeito de antecipação ocorre, depende do quanto determinados verbos, estruturas temáticas ou amplos contextos discursivos geram tendências contra ou a favor de uma suposta análise processual preferida. Além disso, os autores propõem que a ativação ocorre em diferentes níveis representacionais devido à influência das restrições contextuais. Contextos com restrições altamente lexicais podem influenciar os estágios iniciais do processamento de palavras de entrada, sugerindo que eles podem ser usados para pré-ativar informações em níveis sublexicais de representação, que refletem o nível fonológico, ortográfico e perceptual. (DELONG *et al.*, 2005).

Sob essa perspectiva da compreensão da linguagem, o falante procura inferir, com o máximo de certeza, a causa subjacente às informações sensoriais que são reproduzidas incrementalmente ao longo do tempo. Para inferir essa causa, o falante faz uso de um modelo gerativo interno, um conjunto de representações organizadas hierarquicamente para poder melhor explicar as propriedades estatísticas das entradas *bottom-up*.

Para Kuperberg e Jaeger (2016), as entradas *bottom-up* organizam-se em múltiplos níveis de representação, partindo do nível perceptual e ortográfico para o nível fonológico, sintático, semântico, até as representações contextuais/discursivas de níveis mais altos. Em geral, são relevantes para esse modelo tanto as características do

ambiente comunicativo mais amplo, quanto os objetivos comunicativos. No topo dessa hierarquia gerativa, há um conjunto de hipóteses, de diferentes graus de crença, sobre a causa subjacente à entrada sensorial encontrada.

O falante é capaz de testar suas hipóteses gerando ativamente previsões probabilísticas que são propagadas para níveis mais baixos do modelo gerativo. Tais previsões alteram a distribuição de crenças prévias nos níveis mais baixos de representação antes que novas entradas *bottom-up* tornem-se disponíveis. A intensidade com que as crenças são propagadas dos níveis mais elevados para os níveis mais baixos dependerá da certeza que as mantém. Com isso, quando novas evidências *bottom-up* tornam-se disponíveis, o falante é informado se suas previsões probabilísticas serão ou não serão mantidas. Quando ocorrem erros de previsão, as evidências são usadas para atualizar, por meio da inferência bayesiana, as crenças de alto nível acerca da causa subjacente à entrada sensorial (KUPERBERG, 2016).

O modelo gerativo hierárquico propõe, assim como o modelo baseado em restrições (MACDONALD; PEARLMUTTER; SEIDENBERG, 1994; SPIVEY-KNOWLTON; SEDIVY, 1995; TANENHAUS *et al.*, 1995; TRUESWELL *et al.*, 1993), que os mecanismos de previsão atuam de maneira paralela. Kuperberg e Jaeger (2016) apontam que a atualização de uma crença é consistente com a ideia de que podemos calcular preditivamente múltiplos candidatos em paralelo, cada um com diferentes pesos de ativação ou graus de crença. Sob a perspectiva dos modelos baseados em restrições, a perspectiva de processamento paralelo assume que o analisador calcula várias perspectivas simultaneamente, cada uma com algum grau de suporte probabilístico. Isso não implica necessariamente que todas as análises possíveis sejam examinadas exaustivamente, mas sim que múltiplas análises suficientemente prováveis são consideradas em paralelo (CROCKER; BRANTS, 2000; JURAFSKY, 1996; LEWIS, 2000)

Por exemplo, se observarmos as propriedades de subseleção do verbo “derramar”, verificamos uma alta probabilidade de a palavra “leite” ser subselecionada em casos de contextos mínimos. Isso pode levar à ativação dos traços semânticos presentes na representação dessa palavra. Contudo, é possível observamos a presença de outros múltiplos candidatos ativados em paralelo, que, por qualquer motivo, poderiam ocupar a

posição de argumento interno. Embora esses candidatos possuam traços semânticos comuns à palavra mais esperada, devido à boa delimitação do papel temático, como existe uma forte preferência probabilística por uma palavra específica, uma readaptação do sistema de processamento antecipatório poderá ser necessária.

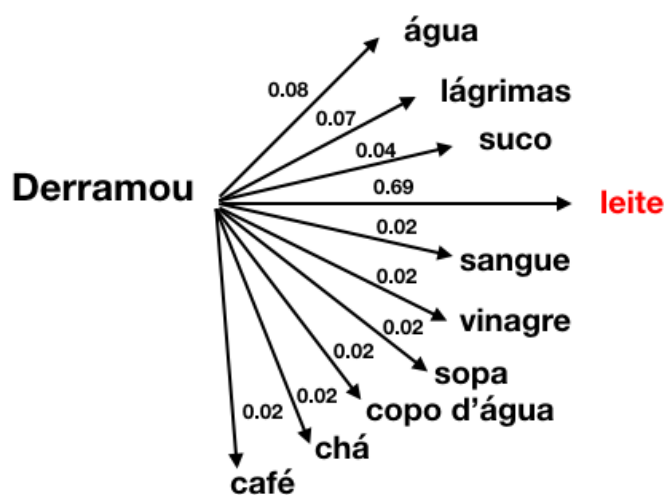


Figura 2 – Propriedades de subseleção do verbo "derramar"

Essa questão da alta probabilidade de subseleção pode não ocorrer da mesma maneira com o verbo "aparar". As probabilidades de subseleção desse verbo ativadas paralelamente são mais subdivididas, ocorrendo, inclusive, duas probabilidades muito próximas entre os argumentos com maiores probabilidades, "grama" e "cabelo". Dessa forma, embora haja uma leve preferência por um argumento específico, essa preferência não parece ser suficiente para que, em contextos mínimos, sejam antecipados os traços semânticos de nenhuma palavra específica. É possível que a ausência de um papel temático bem delimitado possivelmente tornará menos impactante, caso seja necessária, a readaptação do sistema de processamento antecipatório.

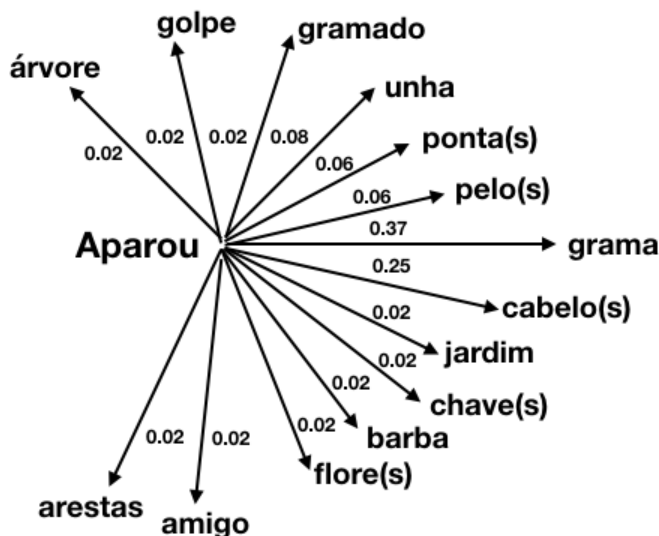


Figura 3 - Propriedades de subseleção do verbo "aparar"

O falante é capaz de testar suas hipóteses gerando ativamente previsões probabilísticas que são propagadas para níveis mais baixos do modelo gerativo. Tais previsões alteram a distribuição de crenças prévias nos níveis mais baixos de representação antes que novas entradas *bottom-up* tornem-se disponíveis. A intensidade com que as crenças são propagadas dos níveis mais elevados para os níveis mais baixos dependerá da certeza que as mantém.

É importante enfatizar que tais preferências probabilísticas verificadas em contextos mínimos podem ser alteradas conforme a força da representação contextual/discursiva. Há evidências de que a força da representação interna do contexto pode influenciar a especificidade das crenças ativadas, se sobrepondo a conhecimentos menos detalhados, mais relacionados a representações de níveis mais baixos, como informações verbais. Camblim, Gordon e Swaab (2007) investigaram em contextos bastante informativos a permanência de efeitos locais de associações semânticas entre palavras. Efeitos de facilitação motivados por associações locais foram verificados em medidas iniciais de rastreamento ocular apenas quando as palavras-alvo estavam em sentenças isoladas ou em textos incoerentes, sugerindo que a associação semântica é muito sensível à representação discursiva. Em contextos bastante informativos, os resultados mostraram



que palavras incongruentes com o contexto, porém associadas localmente, provocam um aumento significativo do N400 em comparação a palavras não associadas localmente e congruentes com o contexto. Os pesquisadores sugerem que,

Se os leitores têm um contexto pequeno à sua disposição para ajudar no processamento lexical, eles poderão confiar mais fortemente em relações mais simples de palavra com palavra. À medida que mais informação torna-se disponível para o modelo discursivo, os efeitos de associação tornam-se cada vez mais tardios. Assim, quando um modelo de discursivo coeso e mais detalhado for construído, este modelo global pode substituir a facilitação associativa *bottom-up*. (CAMBLIM; GORDON; SWAAB, 2007, p. 126; Tradução nossa)<sup>30</sup>

De modo geral, os resultados de Paczynski e Kuperberg (2012), Vieira (2015) e Camblim, Gordon e Swaab (2007) indicam que a antecipação de conhecimentos sobre eventos do mundo real durante o processamento é um recurso cognitivo dinâmico, sujeito às restrições da representação interna do contexto do falante. Também mostram que conhecimentos de eventos menos detalhados ligados a representações lexicais podem ficar mais evidentes quando o contexto não é suficientemente robusto para que a representação discursiva interaja com conhecimentos mais detalhados sobre eventos.

Na próxima seção, trataremos das características dessas representações presentes em nossa memória de longo prazo conhecidas como conhecimentos sobre eventos do mundo real. Discutiremos propostas que consideram não existir uma divisão estrita entre os conhecimentos lexicais, ou de “dicionário”, e os conhecimentos enciclopédicos, ou de mundo. Também abordaremos as influências que a admissão de um modelo lexical expandido exerce neste trabalho.

<sup>30</sup> If readers have little context at their disposal to aid in lexical processing they may rely more heavily on simple word-to-word relationships. As more information becomes available for the discourse model, the effects of association become increasingly delayed. Thus, when a cohesive and more detailed discourse model can be constructed this global model may override bottom-up associative facilitation. (CAMBLIM, GORDON, SWAAB, 2007, p. 126).

### 2.3 Os conhecimentos de eventos como parte da representação lexical

As pessoas possuem conhecimentos sobre a maneira como o mundo se organiza. Um exemplo simples é o fato de sabermos que é mais frequente “um policial prender um ladrão” do que “um ladrão prender um policial”. Não há dúvidas de que os falantes usam esse tipo de conhecimento para compreender e produzir a língua. Contudo, o modo como esse tipo de conhecimento é utilizado durante o processamento linguístico ainda é alvo de controvérsias entre as teorias linguísticas. As diferenças entre as teorias centram-se na arquitetura representacional e na prioridade temporal dos tipos de processamento. Ou seja, teorias diferem sobre se o que pode ser chamado de conhecimento pragmático ou conhecimento de mundo pode influenciar a compreensão imediatamente, ou apenas tardiamente (MCRAE; MATSUKI, 2009).

Elman (2009) afirma que os conhecimentos sobre eventos do mundo real desempenham um papel significativo nos estágios iniciais do processamento e na análise estrutural de sentenças. Para o pesquisador, isso representa um enigma para as visões tradicionais do léxico, uma vez que ele precisa ser expandido para incluir fatores que não parecem pertencer plausivelmente a ele. Caso essa expansão não ocorra, o léxico poderia tornar-se empobrecido, uma vez que praticamente toda a informação sobre o significado das palavras seria removida. Para o autor, a perspectiva adequada para a resolução desse impasse é a assunção da inexistência do léxico, transferindo para as palavras a função de sustentar a estrutura gramatical com seu rico conteúdo. Nessa perspectiva, as palavras funcionam como estímulos externos que afetam o estado interno do sistema em função do estado anterior da rede (memória) e a estrutura dinâmica da rede (gramática).

Nas últimas décadas, o interesse pelo estudo da palavra voltou a ser foco dos estudos de linguística (GOLDBERG, 2003). Atualmente, existe um movimento em direção a estudos que passam a considerar as palavras como entidades gramaticalmente ricas que são fortemente influenciadas pela sensibilidade do contexto de uso. Neste trabalho, embora concordemos com as afirmações de Elman (2009) a respeito da importância das palavras na ativação de conhecimento mais detalhado sobre o mundo real – em que estão presentes padrões de subcategorização, propriedades temáticas, estruturas de eventos, entre outros – não acreditamos na total inexistência de um aparato cognitivo

em que estejam representadas informações relevantes para o processamento, como informações estritamente linguísticas e uma parte dos conhecimentos de mundo.

Elman (2009) apresenta diversas evidências de que o léxico não deve ser considerado metaforicamente um dicionário, um repositório de informações excepcionais, que não podem ser inferidas por meio de regras. Um dos trabalhos com resultados mais contundentes para uma visão mais dinâmica do léxico é o de Ferretti, Kutas e McRae (2007) a respeito de aspecto verbal. No trabalho, os pesquisadores encontraram diversas evidências de que o aspecto verbal influencia a ativação de conhecimentos de eventos quando as pessoas leem tanto palavras isoladas quanto sentenças inteiras.

Os pesquisadores mostraram que a leitura de locais de determinados eventos (*e.g.*: trono) é facilitada quando seguida por verbos com aspecto imperfectivo (*e.g.*: estava sentando), mas não por verbos com aspecto perfectivo (*e.g.*: sentou). Além disso, em um segundo experimento, as pessoas produziram mais frases preposicionais locativas ao completarem fragmentos de sentenças com aspecto imperfectivo do que com aspecto perfectivo. Por fim, foram encontradas evidências de que a amplitude do componente eletrofisiológico N400 para nomes de localizações variou em função do aspecto e da tipicidade, sendo menor para sentenças imperfectivas com locais altamente esperados (*e.g.*: O rei estava sentado no trono) e maior para sentenças imperfectivas com locais menos esperados (*e.g.*: O rei estava sentado no palco). No que se refere às sentenças perfectivas, os participantes apresentaram menores amplitudes de N400 em sentenças imperfectivas (*e.g.*: O rei estava sentando no trono) do que em sentenças perfectivas (*e.g.*: O rei sentou no trono).

Os resultados de Ferretti, Kutas e McRae (2007) possuem para Elman (2009) duas importantes implicações. A primeira é o fato de que os efeitos de facilitação de leitura devido ao aspecto verbal não são facilmente acomodados pelos modelos que tratam mais tradicionalmente sobre o *priming* entre palavras, como os *spreading activation models*<sup>31</sup>. Nesses modelos, o *priming* é realizado através de links que conectam palavras

<sup>31</sup> Nos *spreading activation models*, quando uma palavra alvo é precedida por uma palavra associada (palavra *prime*) em tarefas de reconhecimento de palavras, os participantes parecem ter um melhor

relacionadas e que servem para passar a ativação de uma para outra. Tais links parecem não ser sujeitos a reconfigurações dinâmicas ou modulação sensíveis ao contexto (COLLINS; LOFTUS, 1975). Portanto, modelos clássicos de *spreading activation* não preveem em seu escopo efeitos de *priming* entre verbos e nomes, e, principalmente, não sugerem ocorrência desse tipo de efeito devido a qualquer configuração aspectual.

A segunda implicação está relacionada à maneira como as preferências pelo argumento interno do verbo são codificadas. Criticamente, o efeito parece ocorrer na mesma escala temporal de outras informações que afetam as expectativas por argumentos verbais. Os dados do ERP indicaram diferenças aspectuais dentro da janela temporal de 400ms a partir da apresentação da palavra alvo. A acessibilidade e o impacto imediatos dessas informações indicariam que provavelmente elas seriam incluídas na representação lexical do verbo. Contudo, é difícil propor como uma informação tão dinâmica poderia estar codificada sob a forma de informação temática (ELMAN, 2009).

Para Ferretti, Kutas e McRae (2007), seus resultados indicam que não há dúvida de que as expectativas sobre partes subsequentes de um texto podem ser geradas pelo verbo em si. Contudo, é possível que as expectativas geradas pelas informações aspectuais não estejam vinculadas estritamente ao verbo, mas sim à representação de eventos de maneira mais geral, que estão armazenados na memória de longo prazo. Embora os resultados de diversos estudos (FERRETTI *et al.*, 2001; MCRAE *et al.*, 2005) possam ser acomodados pela expansão das representações lexicais de verbos, os efeitos do aspecto verbal não parecem passíveis de serem integrados em tal solução. Os autores sugerem que, em uma série de outros resultados de estudos como o de Kamide, Altmann e Haywood (2003), os efeitos das expectativas são gerados principalmente pela estrutura do evento e por seus participantes (agentes, pacientes, instrumentos), em vez de estarem simplesmente relacionados ao próprio verbo em si. Com isso, o conceito de

desempenho na quantidade de tempo que levam para responder. Por exemplo, os sujeitos respondem mais rapidamente à palavra "médico" quando é precedida por "enfermeira", do que quando é precedida por uma palavra não relacionada como "cenoura". Esse efeito de *priming* semântico entre nomes que possuem significados próximos dentro de uma rede cognitiva parece ser ativado por uma matriz de conceitos como unidades cognitivas em nossa memória. Cada conceito consiste em um nó cujos elementos ou características associadas são conectados por arestas. Mais detalhes serão abordados na seção acerca da tarefa de Decisão Lexical.

representação de eventos pode servir como uma maneira útil para se entender como ocorrem os resultados ligados ao aspecto verbal. Para os autores, é possível que o verbo possa prover pistas, algumas bastante potentes, capazes de ativar e interagir com os conhecimentos mais complexos sobre eventos do mundo real que os falantes possuem.

McRae e Matsuki (2009) afirmam que pouco se sabe sobre a maneira como os falantes representam e ativam os conhecimentos de eventos. Contudo, está claro que a ativação desse tipo de conhecimento exige computações complexas, flexíveis e dinâmicas. Por exemplo, quando uma pessoa pensa em um nome eventivo como “almoço”, uma parte do conhecimento prototípico da palavra é ativada, sendo influenciada pelo contexto e pelos objetivos comunicativos. Em suma, ao ouvir a palavra “almoço”, uma parte dos conhecimentos sobre o que o evento envolve (participantes, atividades, sequência temporal, instrumentos) é ativada e influenciada pelo contexto da comunicação. A linguagem fornece dicas que podem ser usadas para direcionar a atenção do falante para diferentes aspectos dos eventos.

Como Elman (2009), Jackendoff (2002) também afirma que é difícil traçar uma divisão entre o que é um conhecimento puramente lexicalizado, ou de dicionário, e um conhecimento sobre eventos do mundo real, ou enciclopédico: “Por exemplo, a diferença entre *murder* e *assassinate* é que a última implica um motivo político por parte do agente.”<sup>32</sup>(p. 286; Tradução nossa). Sob uma visão bastante restritiva de conteúdos lexicais, é bastante difícil definir se um sentido complexo como “motivo político” faria parte de nossos conhecimentos de dicionário. Entretanto, caso fosse assumido que esse conteúdo pertencesse a nosso conhecimento conceitual geral, sem uma separação definitiva entre semântica e pragmática, a distinção de significados entre as duas palavras se tornaria possível. Para o pesquisador, a proposta de se diferenciar definitivamente conhecimentos de mundo de conhecimentos estritamente lexicalizados encontra, frequentemente, exemplos que consistem em anomalias para a teoria.

<sup>32</sup> “For example, the difference between murder and assassinate is that the latter implies a political motive on the part of the agent.” (JACKENDOFF, 2002, p. 286).

Kuperberg (2013) propõe que em nossa memória de longo prazo está representada uma série de “conhecimentos de eventos” que podem ser ativados durante o processamento linguístico. São informações sobre eventos bastante específicos – *e.g.* “garçoneiro serve cliente”; “mecânico conserta carros” – que permitem aos falantes prever, com probabilidade relativamente alta, propriedades semânticas mais refinadas sobre palavras adicionadas à mensagem. Embora nem todos os contextos forneçam evidências suficientes para que os falantes possam inferir hipóteses tão específicas, muitos fornecem evidências suficientes para prever de forma confiável estruturas de eventos mais gerais, como as representações que especificam propriedades semânticas mais gerais, que são necessárias para que os participantes desempenhem papéis semânticos relacionados a certos tipos de ações ou estados (DOWTY, 1991; FILLMORE, 1967; JACKENDOFF, 1972; MCRAE; FERRETTI; AMYOTE, 1997).

Inferências no nível das estruturas de eventos podem ser baseadas em vários tipos de informações contextuais, incluindo as restrições seletivas de verbos (CHOMSKY, 1965; KATZ; FODOR, 1963) e sinais proeminentes que aparecem consistentemente em diferentes línguas, como animacidade em combinação com a ordem linear de argumentos ou com as marcações de casos (BATES; MACWHINNEY, 1987; BORNKESSEL; SCHLESEWSKY, 2006; BORNKESSEL-SCHLESEWSKY; SCHLESEWSKY, 2009; MACWHINNEY; BATES; KLIEGL, 1984). Essas inferências a respeito da estrutura de eventos podem, por sua vez, levar os leitores a gerar fortes previsões acerca das propriedades semânticas mais gerais da informação futura. (KUPERBERG, 2016, p. 11; Tradução nossa)<sup>33</sup>.

Há possibilidade de que conhecimentos mais específicos sobre eventos do mundo real – como relações entre os prováveis participantes de um evento – sejam ativados na ausência de representações contextuais robustas. As representações conceituais robustas são aquelas que provêm informações discursivas suficientemente elucidativas

<sup>33</sup> Inferences at the level of event structures can be based on several different types of contextual information, including the selection restrictions of verbs (Chomsky, 1965; Katz & Fodor, 1963) and certain prominence cues that appear consistently across different languages, such as animacy in combination with either the linear order of arguments or case markings (Bates & MacWhinney, 1987; Bornkessel-Schlesewsky & Schlewsky, 2006, 2009; MacWhinney, Bates & Kliegl, 1984). These event structure inferences can, in turn, lead comprehenders to generate strong predictions for coarse-grained semantic properties of upcoming information. (KUPERBERG, 2016, p. 11).

para facilitar o processo de antecipação de traços semânticos de itens lexicais, por exemplo,

(33) The day was breezy so the boy went outside to fly\*

\*Ventava naquele dia, então o menino saiu para empinar ... (Tradução nossa)

Acreditamos que alguns tipos de verbos – especificamente, aqueles que possuem forte capacidade de subselecionar argumentos específicos com maior probabilidade – possuem maior capacidade de acionar inferências a respeito das entidades mais propícias a ocuparem a função semântica de paciente (ou a posição de argumento interno) nos eventos que denotam. Isso ocorre por esses verbos geralmente possuírem papéis temáticos bem delimitados, tornando bastante restrito o número de entidades que prototipicamente ocupam tal função. Em outras palavras, com base nas experiências de uso da língua no mundo real, alguns verbos adquirem a tendência probabilística de subselecionar argumentos internos específicos, o que pode facilitar a antecipação de informações semânticas referentes a seus argumentos internos mais prováveis. Tal efeito pode estar relacionado ao fato de que, por aceitarem um pequeno número de entidades do mundo real na posição de argumento interno, esses verbos possuem traços temáticos mais bem delimitados no que se refere às entidades mais propícias a ocupar a posição de argumento interno.

Na seção a seguir, trataremos da maneira como conhecimentos sobre eventos do mundo real podem contribuir para delimitação das representações temáticas verbais. Faremos uma pequena revisão sobre as principais teorias que discutem como as informações temáticas podem ser delimitadas e, em seguida, apresentaremos nossa proposta acerca da maneira como os papéis temáticos estão estritamente relacionados aos conhecimentos sobre eventos do mundo real.

## **2.4 Relações Temáticas**

Segundo Dowty (1991), Relações Temáticas (ou Papel Temático) são conceitos frequentemente debatidos na teoria sintática e semântica moderna, mas sobre os quais

não existe acordo quanto à sua natureza e definição. Tradicionalmente, os papéis temáticos são concebidos como o resultado do problema de interface. Eles foram propostos com o objetivo de tratar da relação entre sintaxe e semântica no que se refere à relação dos predicados e seus argumentos. São considerados uma forma de representação do significado lexical que é útil na explicação da interface entre sintaxe e semântica. *Grosso modo*, os papéis temáticos são generalizações sobre os argumentos de um predicado criadas com o intuito de demonstrar as regularidades entre a representação semântica e a expressão sintática do predicado.

O conceito de papel temático é tratado por diferentes autores na literatura linguística. Diante de tamanha diversidade, não é nosso objetivo realizar uma revisão exaustiva sobre cada uma dessas teorias. Trataremos brevemente de algumas características das teorias mais relevantes a fim evidenciar as diferenças históricas entre determinados tratamentos das relações temáticas e a perspectiva que adotamos neste trabalho.

Na linguística moderna, uma das primeiras noções de papéis temáticos, ou relações temáticas, foi proposta por Gruber (1965) em sua tese de doutorado. Sob a perspectiva da gramática gerativa, o autor propunha um nível pré-lexical em que propriedades sintáticas e semânticas eram combinadas. Entretanto, foi Fillmore (1967) quem evidenciou a relevância da discussão sobre relações temáticas para a teoria linguística. Ele apresenta sua perspectiva das relações temáticas por meio da Gramática de Casos. Segundo Lima (1982), Fillmore (1967), a partir da Teoria Padrão de Chomsky (1965), propõe a existência de uma base sintática, ligada a elementos semânticos, mais profunda do que as bases até então concebidas. A Gramática de Casos (FILLMORE, 1968) é uma teoria sobre a representação semântica fundada nas relações existentes entre uma ação (ou um estado) denotada pelo verbo e seus demais constituintes. Essas relações são chamadas de relações de casos, ou, simplesmente, casos conceituais. A Gramática de Casos rejeita a distinção entre categorias e funções gramaticais, propondo uma representação direta de funções semânticas do nome e, portanto, rejeitando também a pertinência das funções "sujeito" e "objeto direto" na estrutura profunda:

Fillmore propõe a introdução da noção de caso no componente de base de toda língua, sendo que a estrutura de base comporta termos



primitivos que são relações casuais. Sua proposta mantém a centralidade da sintaxe, como proposto por Chomsky em *Aspects...* e ressalta a importância de "relações semanticamente relevantes" (que podem estar encobertas), que envolvem nomes e as estruturas que os contêm, formando um conjunto finito específico — as relações casuais, universalmente válidas. Fillmore usa caso para "identificar a relação sintático-semântica subjacente", e "forma casual" para significar "a expressão de uma relação casual numa determinada língua" (LIMA, 1982, p. 33).

Inicialmente, Chomsky (1965) considerou que a estreita relação entre sintaxe e semântica ocorria por meio das subcategorizações. Tais fenômenos possuíam o objetivo de descrever os contextos sintáticos em que os elementos de uma categoria lexical poderiam ocorrer devido a exigências de significado – as restrições seletivas. A partir de 1981, no contexto da Teoria de Princípios e Parâmetros, foi proposta a Teoria Teta, que prevê atribuição de Papéis Teta para todos os argumentos dos predicados. O critério prevê que o núcleo do predicado atribui biunivocamente papéis aos argumentos e que cada argumento recebe um papel. Segundo Cançado (2000),

A teoria utiliza desses papéis apenas sob uma perspectiva diacrítica, isto é, o conteúdo semântico dos papéis temáticos particulares não tem relevância gramatical, o que interessa é a existência ou não desses papéis que serão percebidos como meros índices da estrutura argumental, de natureza sintática. Portanto, termos como agente, tema, etc. reduzem-se a termos descritivos de referência, sem qualquer estatuto na teoria; quando muito, servem como propriedades, primitivas ou derivadas, na descrição lexical (p. 300).

Por meio de uma perspectiva distinta de Chomsky (1981), Jackendoff (1987) propõe um tratamento semântico/conceitual das relações temáticas. Para ele, as relações temáticas são relevantes informações da estrutura conceitual, cuja importância é maior do que a de informações de interface ou sintáticas. Tais relações não são teoricamente elementares, mas são definidas por categorias ontológicas, como os operadores IR, FICAR e CAUSAR, na estrutura conceitual. O padrão dessas relações, sua natureza e distribuição, é verificado empiricamente, observando os padrões lexicais e sintáticos da língua natural em relação aos significados. Sob tal perspectiva, os papéis temáticos não obedecem ao critério teta. Portanto, alguns verbos atribuem mais de um papel para o mesmo argumento, outros atribuem o mesmo papel a dois argumentos diferentes, e

alguns verbos possuem papéis temáticos que não são atribuídos a nenhum SN. Segundo Dowty (1991), o interesse de Jackendoff em papéis temáticos surge não para indexar argumentos, mas puramente para descrever os padrões semânticos de subcategorização no léxico e na sintaxe, o que para ele revela uma estrutura conceitual. Dessa forma, Jackendoff não considera que as relações temáticas estejam localizadas em um componente independente da gramática.

Dowty (1991) sugere que os papéis temáticos não possuem o estatuto de primitivos semânticos (como ocorre em propostas que listam uma série de papéis temáticos), mas são definidos em termos dos acarretamentos lexicais que um verbo pode produzir (CANÇADO; AMARAL, 2017). As propriedades dos papéis são abstrações, propriedades que os argumentos possuem em virtude de serem os argumentos de um determinado verbo. O linguista argumenta que a melhor teoria para descrever este domínio não é um sistema tradicional de papéis discretos, composto por denominações como agente, paciente, fonte, mas uma teoria na qual os únicos papéis são dois conceitos, formados por conjuntos de características. Eles são chamados PROTO-AGENTE e PROTO-PACIENTE. Cada um desses papéis é caracterizado por um conjunto de acarretamentos verbais. Por exemplo, um argumento que for identificado como PROTO-AGENTE pode ser composto por qualquer um dos acarretamentos desse papel, como “volição”, “percepção”, “causa de mudança de estados”, “movimento” e “existência independente”. Os argumentos de um verbo podem ter qualquer um dos dois proto-papéis (ou ambos) em graus variados, de acordo com o número de vinculações que cada tipo de verbo possui.

Neste trabalho, pretendemos observar tanto por meio de uma tarefa de julgamento, quanto por meio de uma tarefa de processamento temporizada, como esses conhecimentos de eventos do mundo real podem contribuir para diferenciação de duas classes de verbos que podem apresentar comportamentos temáticos divergentes. Com isso, assumimos que o conceito de papéis temáticos inclui uma série de conhecimentos sobre os eventos do mundo real e sobre a maneira como eles se organizam. Desta forma, nos aproximamos da proposta de McRae, Ferretti e Amyote (1997), segundo a qual os conhecimentos sobre os eventos estão armazenados na memória do falante de forma linguisticamente relevante. Para os pesquisadores, os papéis temáticos são conceitos

específicos dos verbos que, internamente, são formados por um conjunto de traços relacionados a conhecimentos de mundo. São conhecimentos adquiridos ao longo da vivência do falante a respeito dos papéis que entidades e objetos desempenham em eventos reais.

Sob essa perspectiva, os conhecimentos sobre os papéis executados pelos participantes de um evento são acionados de modo eficiente no momento em que um verbo é lido ou ouvido. Contudo, os autores destacam que nem todas as instâncias de um evento são acessadas quando o verbo é encontrado.

Em outras palavras, a consciência de um leitor não é tomada por lembranças de eventos em que ocorrem acusações quando leem ACUSA. Em vez disso, o argumento é que a memória se organiza de maneira semelhante a um modelo de memória de matricial<sup>34</sup> (...), de modo que uma consequência de ouvir ou ler ACUSA é o cálculo de uma representação que corresponde a um acusador prototípico. Observe que os modelos de memória desses tipos também permitem que as representações calculadas sejam alteradas pelo contexto. Assim, ouvir ACUSA durante um filme sobre uma briga de família pode produzir representações de papéis um pouco diferentes do que durante um drama de tribunal. (MCRAE; FERRETTI; AMYOTE, 1997; Tradução nossa)<sup>35</sup>

Além disso, é importante ressaltar que alguns papéis temáticos parecem ser mais bem delimitados do que outros, devido à consistência em sua estrutura de traços. Ou seja, alguns papéis temáticos parecem ser mais claramente definidos do que outros devido a diferenças importantes nas regularidades dos traços semânticos das palavras que podem vir a ocupar a posição de argumento interno do verbo (MCRAE; FERRETTI; AMYOTE, 1997).

<sup>34</sup> Para mais detalhes, ver Hintzman (1986) e McRae e Jones (2013).

<sup>35</sup> In other words, a reader's consciousness is not flooded with memories of accusing events when she reads ACCUSE. Rather, the claim is that memory is organised in a manner akin to a matrix memory model (...), so that one consequence of hearing or reading ACCUSE is the computation of a representation that corresponds to a prototypical accuser. Note that memory models of these types also enable computed representations to be shaded by context. Thus hearing ACCUSE during a movie about a squabbling family might produce somewhat different role representations than during a courtroom drama. (MCRAE; FERRETTI; AMYOTE, 1997)

Em suma, a partir dessa breve revisão é possível presumir que os papéis temáticos são definidos como as funções desempenhadas pelos argumentos subcategorizados pelo verbo. Sob uma perspectiva que privilegie a semântica, são os papéis assumidos pelos participantes na estrutura dos eventos denotados pelo verbo. Existe na literatura linguística um vigoroso debate acerca da natureza sintática, semântica ou conceitual desses papéis<sup>36</sup>. Ademais, a taxonomia dos papéis temáticos – que inclui nomenclaturas como agente, paciente, instrumento, tema, alvo – também desperta diversas discussões. Alguns teóricos (CHOMSKY, 1981) propõem que os papéis temáticos são apenas índices sintáticos presentes na estrutura argumental e cujo conteúdo semântico não possui relevância gramatical. Outros restringem o conteúdo dos papéis a um certo número de traços semânticos gerais (DOWTY, 1991). Neste trabalho, consideramos que o conceito de papéis temáticos inclui uma série de conhecimentos sobre os eventos do mundo real e sobre a maneira como eles se organizam. Como McRae, Ferretti e Amyote (1997), propomos que os papéis temáticos são conceitos específicos dos verbos que são formados por um conjunto de traços relacionados a conhecimentos de mundo.

A seguir, discutiremos as características de um dos paradigmas experimentais utilizados nesta pesquisa: o paradigma de *priming* semântico. Buscaremos melhor compreender como informações temáticas verbais relacionadas a conhecimentos sobre eventos do mundo real podem facilitar a antecipação de informações semânticas pertencentes aos argumentos internos. Trataremos das propriedades do paradigma e revisaremos estudos que podem nos auxiliar a compreender como ocorre o processo on-line de antecipação de traços semânticos de palavras que podem ocupar a posição de argumentos verbais.

## **2.5 Priming semântico e processos de antecipação de informações on-line**

Diversos estudos sobre reconhecimento de palavras (ANTOS, 1979; BECKER, 1980; DE GROOT, 1984; DELANEY-BUSCH *et al.*, 2019; FERRETTI; MCRAE; HATHERELL, 2001; MCNAMARA, 1992; SEIDENBERG *et al.*, 1984) comprovam que o tempo de reação a uma palavra alvo é menor quando essa palavra é precedida por uma palavra semanticamente

<sup>36</sup> Para um debate aprofundado, ver Jackendoff (2002; 1972) e Dowty (1991).

relacionada. Por exemplo, uma palavra como “leite” será reconhecida – ou lida – mais rapidamente quando antecedida pela palavra “café” do que quando antecedida pela palavra “bolsa”. Esse efeito de facilitação entre palavras semanticamente relacionadas é chamado de *priming semântico*.

Dentre as propostas que explicam a ocorrência de *priming semântico*, os modelos que se baseiam em Collins e Loftus (1975) são bastante difundidos. Nesses modelos, a ativação de conceitos ocorre por propagação – esse fenômeno é também conhecido como *spreading activation*. Em geral, os modelos de *spreading activation* propõem que os conceitos estão armazenados na memória dos seres humanos sob a forma de uma rede, na qual cada conceito corresponde a um nó. A ativação de um nó possibilita a ativação de outros nós relacionados, isto é, um conceito ativa outros conceitos com os quais há interligação.

Sob essa perspectiva, o *priming semântico* ocorre porque palavras alvo são precedidas por palavras associadas, e, com isso, os nós conceituais das palavras alvo – e de outras palavras relacionadas – são parcialmente ativados antes de suas verdadeiras ocorrências. Considerando-se que certo nível de ativação do conceito é necessário para que uma palavra seja reconhecida, o tempo de reconhecimento de uma palavra alvo precedida por uma palavra associada deverá ser menor do que o tempo de ativação de uma palavra alvo precedida por uma palavra não relacionada.

Porém, os modelos que explicam o efeito de *priming* por meio de redes semânticas possuem dificuldades em prever facilitações entre verbos e argumentos que sejam desencadeadas por relações derivadas de conhecimentos de mundo. Com isso, nos modelos que preveem o *priming semântico* por meio de *spreading activation*, as informações de ligação entre verbos e argumentos se delimitam a restrições seletivas. Segundo McRae *et al.* (2005), em perspectivas tradicionais de redes semânticas (COLLINS; QUILLIAN, 1969), os efeitos de *priming semântico* ocorrem principalmente entre pares de nomes. Ao tratar de representações verbais, esses modelos incluem apenas o significado central e os links de informações temáticas bastante restritas (*e.g.* animacidade) para os nós que servem como espaços reservados para possíveis sintagmas nominais que possam preencher esses papéis nas sentenças. Tais links podem

ser bidirecionais, permitindo que a ativação se espalhe do nó do verbo para os nós da função temática. Contudo, os links e nós temáticos em tais modelos são insuficientes para explicar resultados que considerem a atuação de conhecimentos específicos sobre o mundo real, uma vez que incluem apenas conteúdo semântico mínimo, limitado a informações gerais de restrições seletivas. O link temático entre um verbo e um nó de argumento interno pode especificar, por exemplo, que o preenchimento desse nó deve ser animado, mas não são exigidas informações mais detalhadas sobre a semântica dos itens.

Existem evidências de que o efeito de *priming* é desencadeado não apenas por relações semânticas, mas também por relações originadas do conhecimento de mundo (FERRETTI; MCRAE; HATHERELL, 2001; HARE *et al.*, 2009; MCRAE *et al.*, 2005). Hare e colegas (2009) encontraram efeitos de *priming* entre nomes de eventos e participantes típicos desses eventos, em pares como “liquidação – comprador”. Também foi encontrado *priming* entre nomes de locais e objetos encontrados nesses locais, por exemplo, entre “geladeira – comida”. Os autores sugerem que esse efeito não ocorre devido a mecanismos de *spreading activation*, mas devido à experiência das pessoas com eventos do mundo real.

Ferreti, McRae e Hatherell (2001) encontraram evidências, por meio de três experimentos, de que verbos são capazes de ativar informações temáticas a respeito de agentes, pacientes e até mesmo instrumentos<sup>37</sup>. As tarefas utilizavam julgamentos de animacidade, no qual os participantes deveriam decidir se uma palavra alvo, apresentada logo após uma palavra *prime*, representava “algo vivo”. A utilização de um julgamento de animacidade foi justificada pelo fato de esse tipo de tarefa acessar exclusivamente informações semânticas, diferentemente de outros tipos que podem acessar informações semânticas, fonológicas e ortográficas. Utilizando um rápido intervalo (SOA<sup>38</sup> de 250ms) entre a apresentação das palavras *prime* e das palavras alvo,

<sup>37</sup> Utilizamos aqui a nomenclatura de agente, paciente e instrumento sem assumirmos a necessidade de uma complexa e extensa lista de papéis temáticos. Nosso objetivo ao utilizar tal nomenclatura é apenas representacional. Uma maneira de nos referirmos, no sentido amplo, a entidades que possuem conjuntos de traços prototípicos.

<sup>38</sup> SOA é a abreviatura para *Stimulus Onset Asynchrony*. É o intervalo de tempo entre o final da

os pesquisadores encontraram efeitos de *priming* entre pares de palavras como “esfregando-zelador”, “prendendo-bandido” e “mexendo-colher”.

Além disso, sugerindo que as informações associadas a vários papéis temáticos são computadas na forma de traços prototípicos, os pesquisadores encontraram evidências de *priming* entre pares de pacientes típicos de um verbo e características prototípicas desses pacientes, como no par “contratado-qualificado”. No caso desse experimento, devido à impossibilidade de se aplicar o julgamento de animacidade em adjetivos, realizou-se a tarefa de decisão lexical. Nesse tipo de tarefa, os sujeitos devem julgar se uma palavra alvo, apresentada logo após uma palavra *prime*, constitui uma palavra existente na língua em estudo ou não. A tarefa de decisão lexical é capaz de acessar ao mesmo tempo informações semânticas, ortográficas e fonológicas (FERRETI; MCRAE; HATHERELL, 2001)

Ferreti, McRae e Hatherell (2001) consideram que seus resultados demonstram que conhecimentos de eventos do mundo real devem ser considerados parte dos verbos, uma vez que houve efeito de *priming* relacionado a diversos tipos de relacionamentos entre as palavras. Além disso, as redes de *spreading activation* não incluem mecanismos que possam explicar os efeitos de *priming* presentes nos experimentos. Embora suposições *post-hoc* possam ser incorporadas, os mecanismos adicionais necessários para explicar o *priming* entre verbos e agentes, pacientes e instrumentos, a falta de *priming* para agentes e pacientes em papéis incongruentes, tornaria a perspectiva de *spreading activation* indistinguível de uma perspectiva de ativação de esquemas. Com isso, os pesquisadores propõem que, se o significado de um verbo é fortemente influenciado pelos conceitos dos nomes que preenchem seus papéis temáticos, é possível que haja sentido em visualizar informações temáticas como parte do significado dos verbos.

Uma outra perspectiva relevante para a pesquisa desenvolvida no âmbito desta tese é a do “Modelo Probabilístico Racional de Adaptação Experimental” de Delaney-Busch *et al.* (2019). No trabalho, os pesquisadores utilizam o paradigma de *priming* para comprovar apresentação da palavra *prime* e o início da apresentação da palavra alvo.

que, quando a informação semântica é ativada por um contexto antes de uma nova entrada linguística, o processamento semântico da nova palavra normalmente é facilitado, o que atenua a amplitude do potencial evocado (ERP) N400<sup>39</sup>. A modulação N400 é observada mesmo quando o contexto é apenas uma palavra *prime* semanticamente relacionada a uma palavra alvo. Tal efeito de *priming* semântico no N400 é mediado pela probabilidade de encontrar um par *prime*-alvo relacionado dentro de um bloco experimental, sugerindo que os participantes podem adaptar a força de suas previsões à validade preditiva do ambiente experimental mais amplo.

Durante a tarefa, os participantes viram dois tipos de blocos em que a porcentagem de pares de palavras *prime*-alvo relacionados variava. No primeiro bloco, uma pequena porcentagem de estímulos – 10% dos estímulos – possuía itens relacionados. No segundo bloco de estímulos, a porcentagem de estímulos com pares de palavras relacionadas subia para 50%. Para garantir que os participantes processassem as palavras semanticamente e, ao mesmo tempo, não prestassem atenção explicitamente nas relações semânticas entre *primes* e alvos, eles eram instruídos a apertar um botão o mais rápido possível quando viam o nome de um animal. O intervalo de tempo entre o final da apresentação da palavra *prime* e o início da palavra alvo (SOA) foi 600ms.

Os resultados indicaram que as estimativas de se encontrar uma palavra alvo relacionada à palavra *prime* dentro de um bloco em que a frequência de pares de palavras relacionadas era muito alta (50%) influenciaram significativamente a amplitude do N400 item por item. Ou seja, quando os blocos possuíam uma maior porcentagem de palavras relacionadas, os sujeitos perceberam e adaptaram suas expectativas a respeito da porcentagem de pares de palavras relacionadas presente em cada bloco. Esses efeitos de adaptação foram demonstrados utilizando o paradigma de *priming*, por meio de medidas de potenciais evocados.

Para os autores, esses achados sugerem que os princípios bayesianos contribuem para o modo como os falantes adaptam suas previsões de traços semânticos à estrutura

<sup>39</sup> Medida de processamento semântico.



estatística de seu ambiente mais amplo, com implicações para a significância funcional do componente N400 e para a natureza preditiva do processamento da linguagem.

Segundo Delaney-Busch *et al.* (2019), no paradigma de *priming*, inicialmente, dois tipos de conhecimentos de longo prazo são usados: conhecimento do relacionamento semântico entre palavras; conhecimento sobre a frequência da palavra isolada. Contudo, outros fatores mais relevantes também contribuem para os resultados encontrados no paradigma. Em cada julgamento, o grau em que o sujeito usa cada uma das fontes de conhecimento de longo prazo sobre a palavra *prime* para gerar previsões probabilísticas sobre o alvo é ponderado pelo grau em que ele acredita que encontrará um alvo não relacionado ou relacionado, ou seja, pelo ambiente experimental em que a palavra está inserida.

Em outras palavras, o sujeito utiliza seu conhecimento de mundo a respeito do relacionamento semântico entre palavras – no caso de nosso experimento: a frequência de cloze – para produzir previsões a respeito da palavra alvo a partir da palavra *prime*. Ele também considera sua crença de que irá encontrar um alvo relacionado ou não relacionado após o *prime*. Isso equivale à sua crença de que o relacionamento semântico do *prime* será um preditor informativo do alvo. Se o participante crê com 100% de probabilidade que irá encontrar um par de palavras não relacionado, então ele pode ignorar o *prime* (e a probabilidade de cloze de qualquer palavra-alvo em potencial). Nesse caso, suas previsões sobre o alvo serão baseadas apenas em seu conhecimento distribuído acerca da probabilidade de ver uma palavra alvo em um contexto mediano ou aleatório. Isso ocorre da mesma maneira com a frequência de palavras: palavras de alta frequência são mais prováveis do que as de baixa frequência em um contexto mediano.

Em um paradigma de *priming* semântico, no qual decisões sobre pares de palavras relacionadas e não relacionadas são apresentadas em ordem aleatória, o participante nunca sabe de antemão se um alvo será relacionado ou não relacionado ao *prime*. Entretanto, em qualquer momento do experimento, ele pode ter alguma estimativa probabilística sobre o encontro de palavras relacionadas ou não relacionadas. O modelo assume que o participante usa essa estimativa como um fator que pondera a influência

relativa do relacionamento semântico entre as palavras versus o conhecimento de frequência de palavras para estimar a probabilidade final de encontrar qualquer alvo em particular. É relevante destacar que a estimativa da probabilidade de encontrar um par de palavras relacionado versus não relacionado não é estática, mas que pode mudar no decorrer de um experimento, assim como a confiança nessa estimativa. Tais crenças são atualizadas, com base no Teorema de Bayes, em julgamentos sucessivos à medida que há avanço na tarefa. Como resultado, o modelo produz uma estimativa final da probabilidade de encontrar uma palavra alvo para cada julgamento individual dos participantes da tarefa.

É importante enfatizar que outros experimentos utilizando um desenho experimental semelhante, ao dividir os blocos de palavras por porcentagem de associação e explorar o paradigma de *priming*, também foram realizados por Lau *et al.* (2014) e Weber *et al.* (2016), utilizando outras técnicas neurofisiológicas. Tais experimentos possuíam os objetivos de verificar a natureza adaptativa do processo de antecipação. Não entraremos em detalhes a respeito desses experimentos devido à complexidade das técnicas neurofisiológicas envolvidas nos estudos, reportando apenas os resultados de maneira geral. Lau *et al.* (2014) realizaram experimentos de magnetoencefalografia (MEG) em conjunção com experimentos de potenciais evocados e de ressonância magnética funcional (fMRI). Os resultados foram semelhantes a Delaney-Busch *et al.* (2019) no que se refere aos potenciais evocados. No que se refere à localização do efeito, os dados de MEG e fMRI se apoiam, mostrando facilitação contextual no giro temporal superior esquerdo anterior para somente o bloco de alta expectativa. Weber *et al.* (2016) realizaram a ressonância magnética funcional (fMRI) para examinar as redes neuroanatômicas envolvidas no processamento preditivo semântico e na adaptação. Os resultados mostraram que, em blocos de forte tendência preditiva (versus blocos de baixa tendência preditiva), foram replicadas as observações prévias de atividade reduzida em palavras semanticamente previsíveis na região anterior esquerda superior / média do córtex temporal, refletindo o processamento facilitado de alvos que são consistentes com previsões semânticas.

O estudo de Delaney-Busch *et al.* (2019) possui relevância para esta pesquisa por utilizar uma técnica que permite avaliar claramente o curso temporal do processo de

antecipação e adaptação por meio do paradigma de *priming*. Tal pesquisa apresenta uma proposta clara e detalhada sobre a forma como antecipação de traços semânticos – no caso de nosso experimento, traços temáticos – pode ocorrer durante uma tarefa que explora o contexto mínimo do paradigma de *priming*. Além disso, também explora um SOA maior, mostrando que, mesmo que as informações temáticas baseadas em conhecimentos de mundo estejam diretamente relacionadas à representação do verbo, os processos estratégicos de antecipação de informações podem demandar um curso temporal maior.

À vista das questões expostas no decorrer deste capítulo, fica clara a necessidade de estudos mais detalhados sobre os processos de antecipação de níveis representacionais durante o processamento linguístico. O impacto da representação contextual sobre a pré-ativação de representações de níveis mais altos também é uma questão que deve ser investigada. No caso deste estudo, nosso interesse está direcionado para as representações de níveis semânticos. Especificamente, esperamos poder contribuir para a compreensão da influência de propriedades temáticas provenientes da subseleção de argumentos verbais na antecipação de representações de níveis semânticos de nomes que podem ocupar a posição de argumentos internos. Também esperamos compreender se tais propriedades de subseleção podem ser suficientemente fortes a fim de auxiliar na pré-ativação de informações temáticas específicas, relacionadas a conhecimentos de mundo dos falantes.

Como foi citado anteriormente, nossa hipótese prevê que propriedades probabilísticas de subseleção de argumentos internos por parte dos verbos podem distinguir o comportamento de dois diferentes grupos de verbos. Devido a essas propriedades de subseleção, acreditamos ser possível que informações temáticas inerentes a cada grupo de verbo influenciem de maneira distinta a antecipação de informações de nível semântico durante o processamento de palavras em contextos reduzidos. Acreditamos que verbos com capacidade de subselecionar menos entidades do mundo real podem possuir papéis temáticos mais bem delimitados (MCRAE; FERRETTI; AMYOTE 1997), o que facilitaria a antecipação de traços semânticos de argumentos internos muito esperados.

Especificamente, sugerimos que alguns papéis temáticos parecem ser mais claramente definidos do que outros devido a diferenças importantes nas regularidades dos traços semânticos das palavras que podem vir a ocupar a posição de argumento interno do verbo. Consideramos ser possível que essas propriedades de delimitação dos papéis temáticos influenciem a previsão de traços semânticos de itens lexicais que podem ocupar a posição de argumentos internos.

Em outras palavras, possuímos a hipótese de que a boa delimitação de um papel temático está diretamente relacionada à força de subseleção de argumentos internos que um verbo possui. Caso um verbo possua um papel temático bem delimitado, ele terá fortes tendências a subselecionar poucas entidades na função de argumento interno, como é o caso do verbo “manobrar”, em que um número reduzido de entidades do mundo real que possuem traços semânticos limitados é capaz de ocupar a função de seu argumento interno. Acreditamos que verbos com fortes tendências de subseleção possuirão facilidades na previsão de informações semânticas de seus possíveis argumentos internos se comparados a verbos que possuem fraca tendência de subselecionar argumentos internos específicos, como é o caso do verbo “agarrar”, cujas entidades passíveis de ocupar a posição de argumento interno possuem traços semânticos bastante diversos.

Para conhecermos o comportamento dos verbos e estudarmos as possibilidades de serem divididos em dois diferentes grupos (de forte e fraca subseleção), utilizamos a tarefa de cloze (TAYLOR, 1953). Tal tarefa permite compreendermos adequadamente as propriedades de subseleção dos verbos quando inseridos em contextos mínimos, formados apenas por um nome próprio e um verbo. As propriedades da tarefa de cloze serão detalhadas no próximo capítulo. Em seguida, abordaremos todas as tarefas realizadas a fim de verificar a hipótese de que verbos de forte e fraca subseleção possuem comportamentos temáticos diferentes, e, com isso, não se assemelham no que se refere à capacidade de antecipar informações semânticas de seus argumentos internos.

### 3 A TAREFA DE CLOZE

Para investigarmos como as propriedades verbais de subseleção de argumentos internos influenciam o acesso a traços semânticos desses argumentos, realizamos uma série de tarefas de cloze (TAYLOR, 1953). Para iniciarmos nossa pesquisa, necessitávamos compreender como os verbos diferenciavam-se com relação à capacidade de subselecionar um argumento interno. Nossa intenção era identificar se alguns verbos possuíam uma maior força de subseleção de argumentos internos do que outros verbos. Para observarmos esse comportamento dos verbos em relação a argumentos internos, foi criado um *corpus* de tarefas de cloze (TAYLOR, 1953). Esse *corpus* foi produzido com o intuito de facilitar a escolha dos verbos de diferentes tendências de subseleção de argumentos que foram usados nas tarefas apresentadas nos próximos capítulos. No capítulo que segue, revisaremos propostas presentes na literatura que utilizam esse tipo de tarefa e apresentaremos os métodos e análises da construção de nosso *corpus*.

Na literatura psicolinguística, a tarefa de cloze (TAYLOR, 1953) é uma das formas utilizadas para se mensurar a previsibilidade de uma palavra em um determinado contexto. Embora tenha sido criada para avaliar e comparar a legibilidade de pequenas passagens, a tarefa é frequentemente utilizada para quantificar previsibilidade de palavras após contextos sentenciais e discursivos. Segundo Taylor (1953), o termo cloze se relaciona ao conceito de *closure*, proveniente da psicologia Gestalt. Esse conceito se refere à tendência dos seres humanos em completar padrões familiares, mas não finalizados.

Para Federmeier *et al.* (2007), os efeitos de facilitação de palavras em contextos moderadamente e altamente previsíveis é mediado pelo grau de ajuste entre uma palavra específica e um determinado contexto. Tal ajuste é frequentemente definido empiricamente usando um procedimento de cloze. Kuperberg (2016) sugere que a interpretação dos dados de cloze podem refletir a força e as probabilidades de expectativas paralelas que o cérebro de um indivíduo calcula durante uma determinada tarefa.

Estudos eletrofisiológicos (KUTAS; HILLYARD, 1984; KUTAS; VAN PETTEN, 1994; VAN PETTEN; KUTAS, 1990) indicam que a probabilidade de cloze de uma palavra está relacionada à amplitude do N400. Palavras muito previsíveis – palavras com alta probabilidade de ocorrência nas tarefas de cloze – produzem ondas N400 com amplitudes reduzidas. Porém, palavras com baixa probabilidade de cloze também são capazes de gerar redução do N400 caso compartilhem traços semânticos com palavras de alta probabilidade (FEDERMEIER; KUTAS, 1999).

Na tarefa de cloze, participantes devem criar continuações para fragmentos de sentenças isoladas, ou diversos fragmentos de texto completo. A probabilidade de uma palavra ser produzida nessas tarefas é chamada de probabilidade de cloze. Em geral, são investigadas as probabilidades de ocorrência de palavras após contextos bastante informativos e contextos neutros. Quando a maior parte dos participantes da tarefa utiliza uma única palavra para completar um mesmo fragmento, infere-se que, naquele contexto, tal palavra é bastante previsível e possui uma alta probabilidade de cloze.

Existem discussões sobre o tipo de conhecimento aferido em uma tarefa de cloze. Usualmente, as probabilidades de cloze são consideradas estimativas de previsibilidade de palavras, considerando-se o uso real da língua. Contudo, a conexão entre as informações fornecidas na tarefa e o conhecimento linguístico implícito dos falantes ainda é pouco conhecida.

Para Staub *et al.* (2015), duas perspectivas teóricas são usadas para explicar como as informações expressas durante a tarefa de cloze estão associadas aos conhecimentos probabilísticos (linguísticos e não linguísticos) dos falantes. A primeira perspectiva propõe que os resultados da tarefa expressam a previsibilidade de uma palavra para uma comunidade de falantes, mas não para os falantes individualmente. Por exemplo, considerando-se que em uma tarefa de cloze com o fragmento (29), foram aferidas as seguintes probabilidades: 0.90 para a continuação “bolo” e 0.10 para a continuação “doces”.

(29) Carla confeitou \_\_\_\_\_

Sob a perspectiva teórica acima, existem dois grupos de participantes: um grupo composto por 0.90 dos participantes, para os quais a palavra “bolo” é a melhor continuação; um segundo grupo composto por 0.10 dos participantes, para os quais a palavra “doces” é a melhor continuação. Tal diferença ocorre porque as experiências linguísticas e os conhecimentos de mundo de cada indivíduo dos grupos são diferentes. Assim, cada indivíduo desses grupos considera apenas uma palavra como a melhor continuação para o trecho.

A segunda perspectiva teórica, que é assumida nesta pesquisa, propõe que, embora cada participante produza uma única resposta durante a tarefa, outras respostas também podem fazer parte do conhecimento probabilístico do falante (ROLAND *et al.*, 2012). Em outras palavras, se considerarmos o exemplo acima, a divisão das probabilidades em 0.90 e 0.10 não indica que cada grupo de participantes preveja apenas uma palavra como melhor continuação para o trecho. É possível que a divisão esteja presente, na verdade, na mente dos falantes. Dessa forma, “bolo” seria a palavra mais previsível, mesmo para os participantes que não a tenham expressado durante a tarefa. Aspectos probabilísticos do processo de resposta à tarefa seriam os responsáveis pela variabilidade das respostas, e não as diferenças nas experiências linguísticas individuais (STAUB *et al.*, 2015).

O motivo para assumirmos a segunda perspectiva teórica deve-se ao fato de que, se considerarmos que a tarefa de cloze reproduz processos cognitivos de previsão durante a compreensão, a primeira perspectiva teórica apresentada acima suscita questionamentos acerca da relevância da restrição contextual para o processo de previsão. Conforme indicam Van Petten e Luka (2012), se o falante prevê uma única palavra durante a leitura/audição de sentenças, palavras com baixos valores de cloze possuem um impacto processual semelhante, mesmo que a restrição contextual seja diferente. Ou seja, o custo processual de se encontrar uma palavra pouco esperada – de probabilidade de 0.05, por exemplo – é igual em contextos de alta restrição, em que há forte preferência por uma palavra específica – *e.g.*: quando as probabilidades das palavras são 0.90, 0.05 e 0.05 – e em contextos de baixa restrição, em que não há preferência por uma palavra específica – *e.g.*: quando as probabilidades são de 0.30, 0.20, 0.15, 0.15, 0.15 e 0.05. Isso ocorre porque, nos dois contextos, 95% dos

participantes considerarão igualmente inesperadas as palavras de cloze 0.05 presentes no contexto.

A associação entre as informações obtidas por meio da tarefa de cloze e as estatísticas reais de uso da língua é verificada em algumas pesquisas. Smith e Levy (2011) investigaram essa questão por intermédio de grandes *corpora*. Probabilidades de cloze verificadas após contextos de quatro palavras (*4-grams cloze task*) foram comparadas a continuações frequentemente observadas em *corpora* de textos on-line e literários. Os resultados indicaram que as medidas dos *corpora* podem ser parcialmente relacionadas a probabilidades de cloze, uma vez que parte da variabilidade encontrada na tarefa de cloze não pode ser explicada por medidas de *corpora*. Para os autores, embora parte dessas diferenças possa influenciar medidas on-line do processamento linguístico, é possível que, na verdade, tais distorções reflitam erros genuínos nos modelos probabilísticos individuais dos falantes. Isto é, é possível que os modelos probabilísticos individuais sejam afetados por fatores como familiaridade e associação semântica das palavras com o contexto, o que pode ocasionar a produção de palavras com esse viés durante a tarefa de cloze.

Shaoul, Baayen e Westbury (2014) realizaram um estudo acerca da relação entre informações probabilísticas derivadas de corpus e medidas de cloze verificadas após microcontextos de três palavras (*3-grams cloze task*). As probabilidades de cloze refletiram as distribuições de frequências observadas no corpus. A probabilidade condicional de uma palavra ocorrer em determinado contexto foi associada à frequência de respostas e à ordem de múltiplas respostas fornecidas pelos participantes na tarefa. Os autores sugerem que a ligação entre as medidas verificadas em corpus e as probabilidades de cloze indicam o uso de um mesmo tipo de informação probabilística pelo sistema linguístico durante a tarefa. Tais informações são derivadas do conhecimento implícito do falante, que foi adquirido por meio do uso das palavras em diversos contextos.

Staub *et al.* (2015) verificaram a latência de respostas durante uma tarefa de cloze. Por meio de uma versão adaptada da tarefa, na qual os estímulos eram apresentados visualmente e as respostas deveriam ser pronunciadas, os pesquisadores encontraram



evidências de que a latência das respostas é influenciada pelas probabilidades de cloze. Respostas mais esperadas eram produzidas em menor tempo do que respostas menos esperadas. O nível de restrição dos contextos também influenciou o período de latência.

Estudos que utilizam rastreamento ocular comprovam a relação entre medidas de cloze e atividades cognitivas relacionadas ao processamento linguístico. Um considerável número de estudos revela que, durante a leitura, palavras com alta probabilidade de cloze<sup>40</sup> são fixadas por menos tempo e saltadas com maior frequência do que palavras com baixa probabilidade de cloze (ASHBY; RAYNER; CLIFTON, 2005; BALOTA; POLLATSEK; RAYNER, 1985; EHRLICH; RAYNER, 1981; RAYNER; WELL, 1996). Contudo, para que sejam encontradas diferenças significativas em medidas de rastreamento ocular, uma forte diferenciação entre os índices de previsibilidade das palavras parece ser necessária.

As investigações supracitadas evidenciam a relevância das tarefas de cloze para os estudos que utilizam a previsibilidade de palavras como variável independente. Sua ampla utilização em estudos sobre o impacto da previsibilidade na compreensão da língua (BICKNELL *et al.*, 2010; FEDERMEIER; KUTAS, 1999; MATSUKI *et al.*, 2011; WARREN; MCCONNELL, 2007, dentre outros) são um indicativo de sua confiabilidade. Contudo, poucos estudos possuem propostas de compreenderem as propriedades das informações ativadas durante a tarefa de cloze.

Nesta pesquisa, associamos o efeito de previsibilidade lexical parametrizada por meio da tarefa de cloze (TAYLOR, 1953) à capacidade de subseleção de argumentos verbais. Acreditamos que, quanto menor o número de entidades citadas como previsíveis diante de um determinado contexto, mais forte é a capacidade de subseleção de argumentos internos por parte do verbo, devido ao número limitado de itens lexicais capazes de ocupar a posição de argumento interno. Por consequência, quando um número restrito

<sup>40</sup> Conforme Frisson, Rayner e Pickering (2005), palavras com alta probabilidade de cloze possuem percentuais de ocorrência entre 0.70 a 0.90. Já as palavras com baixa probabilidade possuem percentuais abaixo de 0.10

de entidades é citado na tarefa, acreditamos que as informações temáticas relacionadas a esse verbo serão mais bem delimitadas e mais restritas.

A existência de um conjunto de estímulos adequadamente normatizados é de suma importância para estudos que manipulam a previsibilidade de palavras. A aplicação de tarefas cloze é um dos procedimentos necessários para o manejo seguro dessa propriedade e constitui um dos métodos mais adequados para se confirmar a intuição de pesquisadores sobre o comportamento de subseleção de um verbo em determinados contextos.

Portanto, por pretendermos conhecer o comportamento de subseleção de um grande número de verbos por meio da aplicação de tarefas de cloze, buscamos observar a existência de dois prováveis grupos com diferentes comportamentos de subseleção de argumentos. Com esse procedimento, buscávamos verificar como as propriedades verbais de subseleção de argumentos internos poderiam diferenciar o comportamento temático de cada grupo de verbos.

Por meio desses procedimentos, foi possível criar um corpus com os dados de subseleção de 520 verbos. A partir dos dados desse corpus, foi possível selecionar todos os itens utilizados nas demais tarefas deste estudo. Além disso, o corpus também nos permitiu efetuar a separação entre os dois grupos de verbos de diferentes tendências de subseleção, conforme os procedimentos explicitados na seção 3.4.

### **3.1 Materiais**

Os estímulos das tarefas consistiam em fragmentos de sentenças em que nomes próprios eram colocados na posição de argumento externo e combinados aleatoriamente com verbos flexionados no pretérito perfeito do indicativo. Uma lacuna foi posicionada após cada um dos fragmentos para que os participantes completassem as frases. A seguir, estão expostos dois exemplos de estímulos empregados na tarefa:

(30) Maria derramou \_\_\_\_\_

(31) João chutou \_\_\_\_\_

A restrição dos conteúdos semânticos dos fragmentos de sentenças possibilitou a verificação das propriedades de previsão de argumentos por parte dos verbos. A opção por se utilizar nomes próprios foi motivada pelo fato de esse tipo de nome possuir uma carga semântica mínima. A simplicidade dos contextos permitiu que os eventos denotados pelos verbos fossem salientados.

Utilizamos o *Dicionário gramatical de verbos* (BORBA, 1990) para selecionar os verbos presentes nos fragmentos de sentença. Foram selecionados 520 verbos (para a lista completa, ver APÊNDICE A) que possuem tendência a apresentar regência transitiva direta. Os nomes próprios foram selecionados com base na pesquisa realizada pelo site *BabyCenter* dos 100 nomes próprios mais comuns do Brasil em 2015. Os 520 fragmentos foram divididos em 17 diferentes listas que continham entre 30 e 33 itens. Cada fragmento ocorria em apenas uma lista.

### **3.2 Participantes**

Cada uma das 17 tarefas foi respondida por 52 estudantes de graduação e pós-graduação da Universidade Federal de Minas Gerais. Todos os participantes eram falantes nativos de Português Brasileiro, com idades entre 18 e 30 anos (média de 22 anos). Dos 884 respondentes, 646 (73%) eram mulheres.

### **3.3 Procedimentos**

As tarefas foram feitas on-line por meio das plataformas *Sona Systems* e *Google Forms* em uma única sessão de 15 minutos, em média. Os participantes realizavam a tarefa no lugar que julgassem mais adequado, sem a presença do pesquisador. Foi solicitado aos participantes que realizassem a tarefa em ritmo natural, e que outras tarefas não fossem efetuadas simultaneamente. Os participantes foram instruídos a completar os fragmentos de sentenças com “os primeiros finais que lhes viessem à cabeça”. Foi

explicado que respostas muito complexas não eram necessárias e que, caso fosse necessário, algumas respostas poderiam ser repetidas.

### 3.4 Resultados e discussão

As tarefas de 26 participantes (3% do total) foram excluídas por não possuírem no mínimo 70% das respostas ou por apresentarem, em mais de 70% dos itens, respostas inadequadas ao propósito deste estudo – com continuações sem sentido (*e.g.*: continuações implausíveis) ou com repetições injustificadas (*e.g.*: repetições da mesma palavra em mais de 30% das respostas). Após a exclusão, as 17 diferentes listas mantiveram-se com pelo menos 50 respondentes.

As respostas elaboradas para cada fragmento foram analisadas e codificadas. Passagens respondidas com continuações sem sentido ou em branco eram codificadas como SR (sem resposta). Mantivemos inalteradas as respostas que possuíam apenas adjuntos – *e.g.*: “Joana arrasou **na festa**”. Apenas o núcleo do sintagma nominal foi considerado quando as continuações eram longas, como “derramou o suco que estava em cima da mesa”. Por meio de um *script*<sup>41</sup> na linguagem *Ruby*<sup>42</sup>, os dados das tarefas foram contabilizados, e o percentual de ocorrências de todas as respostas fornecidas pelos participantes em cada passagem foi calculado. Após o cálculo automático dos percentuais, efetuamos, por meio do programa Excel, o somatório de respostas que possuíam uma mesma palavra em sua forma singular e plural.

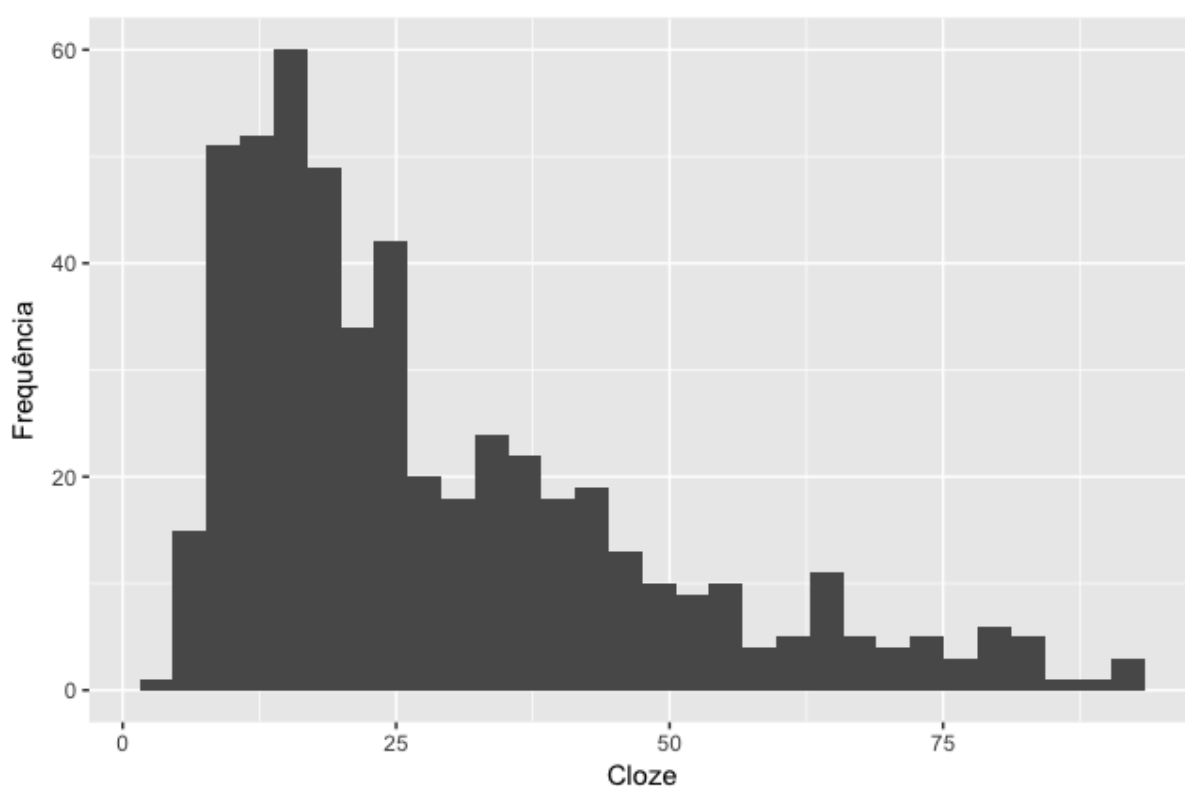
Nesta primeira etapa da pesquisa, nosso interesse direcionou-se para a probabilidade da resposta mais citada em cada item (a lista completa dos verbos e as probabilidades das respostas mais citadas encontram-se no APÊNDICE A). A constituição do *corpus* de dados de cloze permitiu-nos analisar a força da previsão de argumentos internos por

<sup>41</sup> *Scripts* são programas escritos para um sistema de tempo de execução especial que automatiza a execução de tarefas que poderiam alternativamente ser executadas uma por vez por um operador humano.

<sup>42</sup> *Ruby* (MATSUMOTO, 2013) é uma linguagem de programação criada por Yukihiro Matsumoto em 1995. Ela foi criada com o objetivo de ser uma linguagem mais legível e fácil de trabalhar. Seu código fonte é interpretado por um programa interpretador – no caso deste trabalho, o *framework Ruby on Rails* – e executado pelo sistema operacional.

parte dos verbos. Haja vista nosso interesse em verificar se a capacidade de previsão de argumentos internos implica em diferenças no comportamento processual dos verbos, uma vez que acreditamos que as probabilidades verificadas nas respostas mais citadas são as que melhor representam esse comportamento.

A distribuição das probabilidades de cloze das respostas mais citadas em cada fragmento está no histograma a seguir. A maior parte dos verbos testados possui média ou baixa capacidade de subselecionar argumentos específicos, pois estão concentrados na faixa abaixo de 0.50.



*Figura 4- Histograma de distribuição das probabilidades de cloze no corpus*

Na literatura sobre previsibilidade, existe certa dificuldade para se estabelecer os parâmetros em que as diferenças de probabilidades de cloze podem exibir efeitos processuais. Diversos estudos utilizam respostas congruentes – respostas passíveis de ocorrer – a um dado contexto com objetivo de verificar a existência de diferenças entre respostas de alta e baixa probabilidades. Van Petten e Luka (2012) revisaram 18 estudos que manipulavam as probabilidades de palavras diante de contextos congruentes. Nas

condições mais previsíveis, as probabilidades de cloze variavam entre 0.27 e 0.92. Nas condições menos previsíveis, as probabilidades variavam entre 0.01 e <0.50.

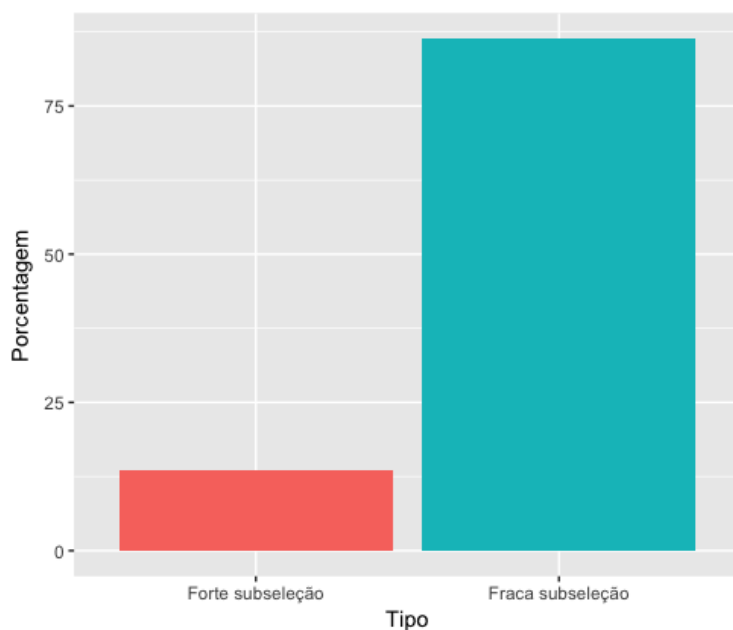
Considerando-se a diversidade dos parâmetros que diferenciam a alta e a baixa probabilidade de cloze, optamos por adotar um critério semelhante ao de DeLong *et al.* (2011), em que as palavras de alta probabilidade de cloze possuem um valor maior do que 0.50 e as palavras de baixa probabilidade possuem um valor menor do que 0.50. Especificamente, optamos por tais parâmetros, pois o estudo eletrofisiológico de DeLong *et al.* (2011) exhibe resultados bastante consistentes ao comparar manipulações de alta e baixa probabilidades de cloze. Ademais, DeLong *et al.* (2011) é uma extensão de DeLong, Urbach e Kutas (2005), um dos estudos de maior impacto para a literatura sobre antecipação no processamento da linguagem<sup>43</sup>.

Assim, em virtude da pequena restrição contextual dos itens de nosso *corpus*, acreditamos ser bastante adequado adotarmos uma divisão em que verbos de forte tendência de subseleção possuem argumentos internos mais previsíveis com probabilidades de cloze maiores do que 0.50; e verbos de fraca tendência de subseleção possuem argumentos internos mais previsíveis com probabilidades de cloze menores do que 0.50. Ou seja, os verbos com cloze acima de 0.50 serão considerados de **forte subseleção**, devido ao pequeno número de entidades passíveis de ocupar a posição de argumento interno, e os verbos de cloze menor do que 0.50 serão considerados verbos de **fraca subseleção**, devido ao grande número de entidades passíveis de ocupar a posição de argumento interno.

No gráfico de barras abaixo, estão demonstrados os diferentes parâmetros de concentração entre os verbos considerados de forte subseleção e de fraca subseleção. Como é possível verificar, a maior parte dos verbos de nosso *corpus* são caracterizados

<sup>43</sup> Embora Nieuwland e Politzer-Ahles *et al.* (2018) não tenham conseguido replicar o efeito de antecipação de formas fonológicas de artigos encontrado em DeLong, Urbach e Kutas (2005), os efeitos relacionados à antecipação de nomes foram devidamente confirmados. Ademais, existem questionamentos acerca da metodologia e da análise inferencial utilizadas por Nieuwland e Politzer-Ahles *et al.* (2018). Para maiores detalhes, veja as respostas de Yan, Kuperberg e Jaeger (2017) e de DeLong, Urbach e Kutas (2017).

como de fraca subseleção. Verificamos que, de um total de 520 verbos, há 76 verbos de forte subseleção e 444 verbos de fraca subseleção.



*Figura 5- Gráfico com concentração de itens por tipo de subseleção*

Neste estudo, diversas tarefas de cloze foram realizadas com o objetivo de mensurar a previsibilidade de argumentos internos após fragmentos sentenciais simples – formados por um nome próprio e um verbo – e de criar um corpus de possíveis itens experimentais. Essas tarefas nos permitiram acessar diretamente conhecimento de eventos do mundo real dos participantes, selecionando entidades tipicamente relacionadas aos eventos ativados pela união do verbo ao argumento interno. A seguir, detalharemos uma tarefa off-line e dois experimentos on-line realizados no decorrer da pesquisa que utilizam os itens experimentais criados a partir das tarefas de cloze.

#### 4 JULGAMENTO DE AJUSTE TEMÁTICO

Tendo em vista que as características temáticas dos verbos são muito relevantes para este trabalho, para melhor compreender as propriedades temáticas dos dois grupos de verbos estudados – verbos de forte subseleção e verbos de fraca subseleção – optamos por realizar uma tarefa off-line de julgamento de ajuste temático.

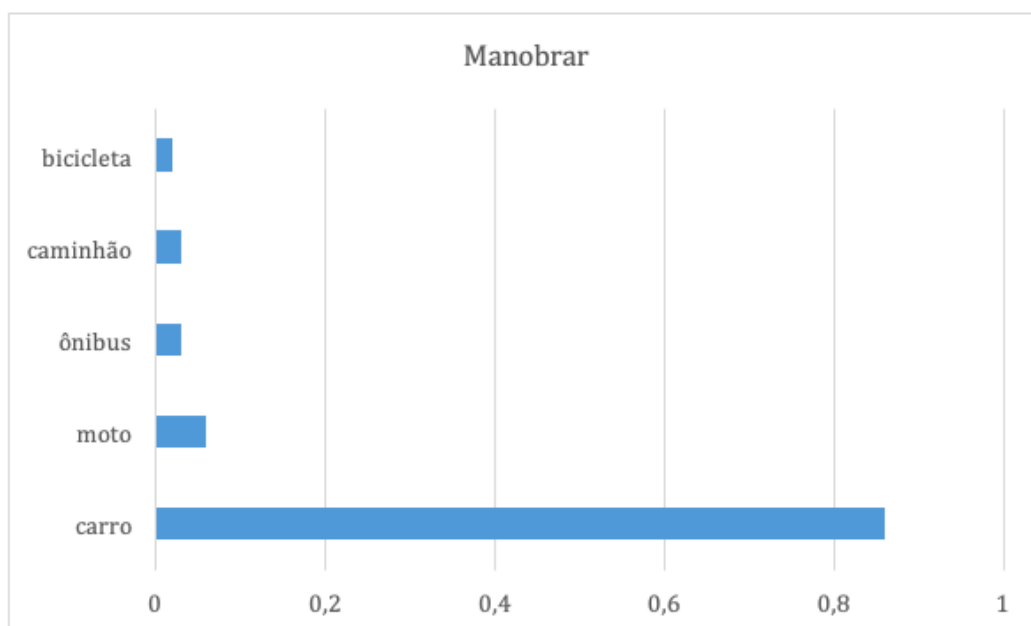
Segundo McRae e colegas (1997), a tarefa é uma boa indicação de que os papéis temáticos possuem uma estrutura interna de traços, pois mostra que alguns conceitos nominais são classificados como melhores, ou mais prototípicos, do que outros. A tarefa indica o limite até o qual um nome preenche preferências de subseleção de um dado papel. Ou seja, ele avalia, por meio de julgamentos humanos, a probabilidade de uma entidade participar de certos eventos. Nosso principal objetivo ao aplicar esta tarefa foi verificar a consistência da relação entre as medidas da tarefa cloze (TAYLOR, 1953) e a força das relações temáticas aferidas na tarefa de McRae, Ferretti e Amyote (1997).

Consideramos que pouco se sabe sobre a forma como as propriedades temáticas dos verbos podem influenciar processos antecipatórios de traços semânticos de argumentos internos. Por isso, o julgamento de ajuste temático (FERRETTI; MCRAE; HATHERELL, 2001; MCRAE; FERRETTI; AMYOTE, 1997; SAYEED; GREENBERG; DEMBERG, 2016) mostra-se adequado para se conhecer melhor o comportamento temático dos verbos. A tarefa foi desenvolvida com a finalidade de quantificar a adequação das relações temáticas entre verbos e argumentos a partir dos conhecimentos sobre o mundo real. Por exemplo, ela busca quantificar quão adequado é o argumento “bolo” para o verbo “assar”. Isto é, busca estimar o nível de adequação dos argumentos internos a partir dos conhecimentos sobre o mundo real dos participantes.

Possuímos a hipótese de que os verbos de forte subseleção apresentam papéis temáticos mais claramente definidos do que os verbos de fraca subseleção. Assim, há a possibilidade de os argumentos relacionados a verbos de forte subseleção serem considerados mais adequados do que os argumentos dos verbos de fraca subseleção. Acreditamos que uma possível explicação para a boa definição dos papéis temáticos dos



verbos está relacionada ao número de entidades citadas na tarefa de cloze. Realizamos diferentes tarefas de cloze e confirmamos que os verbos de forte subseleção são associados a um pequeno número de argumentos que geralmente possui traços semânticos em comum, conforme exposto na Figura 6 abaixo, que ilustra as propriedades de subseleção do verbo “manobrar”. Em relação aos verbos de fraca subseleção, observou-se a associação a um maior número de argumentos com traços semânticos mais variados, conforme exposto na Figura 7 a seguir, que ilustra as propriedades de subseleção do verbo “agarrar”. É possível que, durante o processamento antecipatório, a boa definição dos traços temáticos dos verbos de forte subseleção facilite a antecipação de traços temáticos dos argumentos mais esperados desse tipo de verbo.



*Figura 6 - Gráfico de propriedades de subseleção do verbo manobrar*

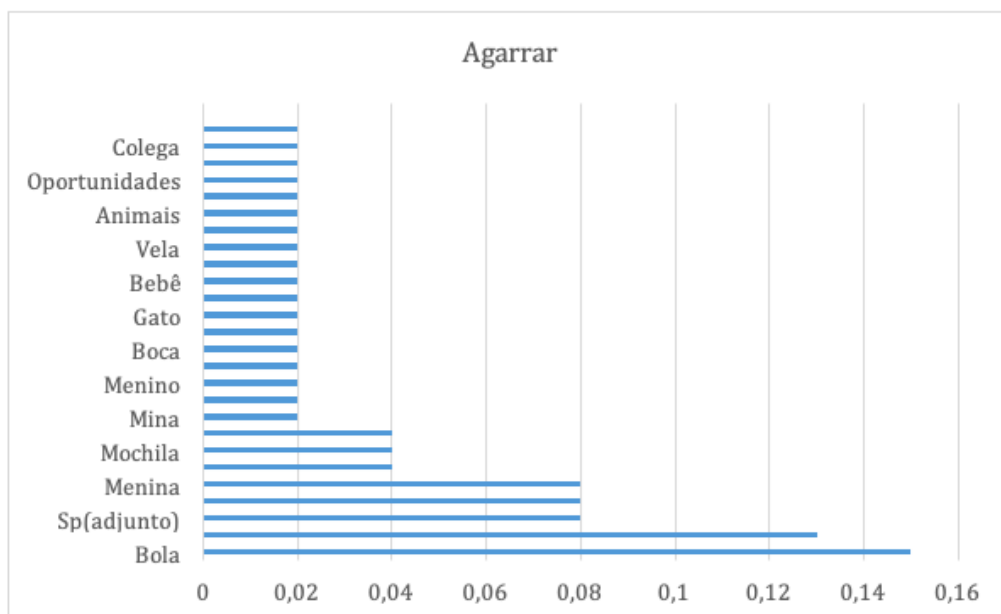


Figura 7 – Gráfico de propriedades de subseleção do verbo agarrar

Ademais, observamos a relação entre a força de subseleção dos verbos e a previsibilidade dos argumentos internos. Para isso, utilizamos argumentos cujas previsibilidades foram consideradas:

- mais previsível – palavra mais citada no teste de cloze;
- pouco previsível – palavra cuja probabilidade de ser citada no teste de cloze varia de 0.02 até 0.11;
- imprevisível, porém plausível – palavra que plausivelmente poderia ser argumento interno do verbo, mas que não foi citada no teste de cloze;
- implausível – palavra que não poderia ocupar a posição de argumento do verbo devido a restrições de plausibilidade.

Verificamos se existem interações entre o comportamento de subseleção dos verbos e a relação com a previsibilidade dos argumentos internos medida por meio de tarefas de cloze. Nossa hipótese sugere que verbos de fraca subseleção não apresentem grandes diferenças nos julgamentos temáticos nos níveis de previsibilidade dos argumentos (i)

mais previsíveis, (ii) pouco previsíveis e (iii) imprevisíveis, porém plausíveis, devido à má definição de seus traços temáticos. Também, optamos por utilizar os níveis de previsibilidade (ii) pouco previsível e (iii) imprevisível, porém plausível devido ao fato de um estudo de Smith e Levy (2013) mostrar que os efeitos de cloze em medidas de rastreamento ocular – logo, efeitos de previsibilidade – possuem uma distribuição logarítmica. Assim, é possível que diferenças significativas entre essas duas faixas de baixa previsibilidade sejam visualizadas.

O quadro a seguir sumariza as condições experimentais utilizadas em nosso julgamento de ajuste temático.

<b>Grupo</b>	<b>Condição</b>	<b>Verbo</b>	<b>Previsibilidade do argumento</b>	<b>Probabilidade cloze</b>
Forte subseleção	Mais previsível	Manobrou	Carro	0.86
Forte subseleção	Pouco previsível	Manobrou	Moto	0.06
Forte subseleção	Imprevisível, porém plausível	Manobrou	Barco	0.00
Forte subseleção	Implausível	Manobrou	Quarto	0.00
Fraca subseleção	Mais previsível	Achatou	Massa	0.19
Fraca subseleção	Pouco previsível	Achatou	Lata	0.02
Fraca subseleção	Imprevisível, porém plausível	Achatou	Uvas	0.00
Fraca subseleção	Implausível	Achatou	Curva	0.00

*Figura 8 – Quadro com condições experimentais do julgamento de ajuste temático*

Durante o julgamento de ajuste temático, os participantes mensuram, por intermédio de notas de 1 a 7, o quanto os eventos denotados pela união dos verbos aos argumentos internos são comuns no mundo real. A nota 1 representava um evento incomum, a nota 7 representava um evento muito comum. A título de exemplo, vejamos o caso de eventos denotados pelo verbo de fraca subseleção “provocar”. Inicialmente, a tarefa apresenta a pergunta “Quão comum é?” ao participante. Em seguida, são expostos quatro eventos na forma passiva como: (1) “uma briga ser provocada” (mais previsível); (2) “danos serem provocados” (pouco previsível); (3) “uma crise ser provocada” (imprevisível, porém plausível); (4) “uma pedra ser provocada” (implausível). Uma nota que representa a frequência de ocorrência dos eventos no mundo real deve ser atribuída a cada evento. Nas tarefas, é esperado que eventos com forte subseleção de argumentos mais previsíveis – ou seja, com papéis temáticos bem delimitados – tenham maior probabilidade de receber notas maiores do que os eventos com verbos de fraca subseleção de argumentos mais previsíveis. Também esperamos encontrar poucas diferenças entre as notas dos eventos denotados pelos verbos de fraca subseleção de acordo com o nível de previsibilidade dos argumentos.

Portanto, a partir da proposta para interpretação dos papéis temáticos de McRae, Ferretti e Amyote (1997) e das tarefas julgamentos de ajuste temático (FERRETTI; MCRAE; HATHERELL, 2001; MCRAE; FERRETTI; AMYOTE, 1997; SAYEED; GREENBERG; DEMBERG, 2016), consideramos ser possível mensurar de maneira cautelosa as possíveis diferenças entre a força de subseleção dos verbos e os níveis de previsibilidade dos argumentos internos. A partir de tais resultados, acreditamos ser possível realizar um delineamento cuidadoso de nossos experimentos on-line, uma vez que os julgamentos de ajuste temático não são suficientes para mostrar se o conhecimento temático está diretamente relacionado à representação do verbo.

#### **4.1 Materiais**

Após a definição do critério que diferenciaria o comportamento preditivo dos verbos nas tarefas de cloze, iniciamos a seleção dos itens utilizados em nosso julgamento a respeito de eventos do mundo real. Selecionamos 24 verbos de forte subseleção (cloze do argumento mais previsível > 0.50) e 24 verbos de fraca subseleção (cloze do argumento

mais previsível < 0.50). Os verbos possuíam duas, três ou quatro sílabas. Para selecionar esses 48 verbos, observamos o conjunto de respostas fornecidas nos fragmentos de sentença de cada verbo (APÊNDICE B). Nos conjuntos de respostas relacionadas a cada verbo, selecionamos respostas com argumentos internos pouco previsíveis, cujas probabilidades de cloze variavam entre 0.018 e 0.11 (média: 0.03; desvio padrão: 0.02). Selecionamos, para possíveis continuações de nossas sentenças, argumentos não citados nas tarefas para compor duas condições experimentais: argumentos internos imprevisíveis, porém plausíveis (cloze 0) e argumentos implausíveis (cloze 0; c.f. Figura 7). Os argumentos foram submetidos a uma pré-tarefa com 10 pessoas, na qual elas efetuavam julgamentos semelhantes ao da tarefa planejada. Por limitações metodológicas na seleção dos itens<sup>44</sup>, optamos por utilizar verbos cujos argumentos internos possuem duas ou três sílabas. Controlamos o número de sílabas dos argumentos internos ligados a um mesmo verbo. Os argumentos poderiam ter duas ou três sílabas. Assim, nos quatro níveis da variável previsibilidade – (1) mais previsível; (2) pouco previsível; (3) imprevisível, porém plausível; (4) implausível – os argumentos internos sempre possuíam a mesma quantidade de sílabas: duas ou três sílabas. Desta forma, a média de sílabas dos nomes argumentos internos em todos os níveis da variável previsibilidade era semelhante – média de 2.42 em todos os níveis de previsibilidade dos verbos de forte subseleção; e média de 2 sílabas em todos os níveis de previsibilidade dos verbos de fraca subseleção.

A frequência dos nomes que ocupavam a posição de argumento interno foi verificada no *corpus ptTenTen*, disponível no *SketchEngine*<sup>45</sup> e foi transformada em valores *Zipf*, conforme a proposta de Van Heuven *et al.* (2014). A média dos valores *Zipf* transformados na condição em que os argumentos possuíam cloze >0.50 foi 4.67; o desvio padrão foi 0.5. A média dos valores transformados na condição em que os argumentos possuíam cloze <0.50 foi 4.59; o desvio padrão foi 0.39.

<sup>44</sup> Os verbos que subselecionavam argumentos acima de 0.50 eram poucos, a restrição no tamanho das palavras dificultava a seleção dos verbos.

<sup>45</sup> *SketchEngine* é um *Corpus Query System*, ou sistema de pesquisa de corpus. Está disponível no site: <http://www.sketchengine.co.uk>

Os itens experimentais consistiam em 48 grupos de julgamentos sobre a habitualidade de eventos no mundo real. Os grupos de eventos eram formados pela união de um verbo específico a quatro diferentes argumentos internos. Os eventos eram expressos na forma passiva, *e.g.* “leite ser derramado”. Uma escala de julgamentos que possibilitava notas de 1 a 7 foi utilizada. A nota 1 representava um evento incomum, e a nota 7 representava um evento muito comum. Os eventos eram divididos em dois diferentes níveis de subseleção – forte e fraca subseleção – e, além disso, possuíam quatro diferentes níveis de previsibilidade em cada nível de subseleção, sendo classificados como: (1) muito previsíveis; (2) pouco previsíveis; (3) imprevisíveis, porém plausíveis e (4) implausíveis. Por exemplo:

<b>Níveis de previsibilidade</b>	<b>Verbos de forte subseleção</b>	<b>Verbos de fraca subseleção</b>
<b>Mais previsível</b>	Ordem ser <b>acatada</b>	Carne ser <b>moída</b>
<b>Pouco previsível</b>	Regra ser <b>acatada</b>	Queijo ser <b>moído</b>
<b>Imprevisível, porém plausível</b>	Norma ser <b>acatada</b>	Ervas serem <b>moídas</b>
<b>Implausível</b>	Carne ser <b>acatada</b>	Muro ser <b>moído</b>

*Figura 9 – Quadro com exemplos de frases do julgamento de ajuste temático*

Os níveis de previsibilidade e os tipos de verbos foram classificados a fim de verificar o critério de adequação temática. Eventos mais previsíveis e expressos por verbos com maior força de subseleção possuiriam o papel temático mais bem caracterizado. Por isso, era esperado que, por possuírem um papel temático bem definido, haveria uma diferença clara entre os julgamentos conforme o nível de previsibilidade. Já nos verbos de fraca subseleção, devido ao papel temático pouco definido, não haveria uma diferença clara entre os julgamentos conforme o nível de previsibilidade. Também era esperada uma clara diferenciação entre julgamentos dos níveis de previsibilidade dos verbos de forte subseleção e de fraca subseleção. Em outras palavras, nossa hipótese previa que,

conforme a previsibilidade dos eventos e a força de subseleção dos verbos, haveria maior diferenciação entre as notas atribuídas pelos participantes.

Os 48 grupos de eventos foram divididos em duas listas de 24 grupos (todos os itens estão presentes no APÊNDICE C). Os participantes poderiam julgar apenas uma das duas listas. Cada grupo de eventos só poderia ocorrer uma única vez. 50% de cada lista é formado por grupos de verbos de forte subseleção, o restante da lista é formado por grupos de verbos fraca subseleção.

## **4.2 Participantes**

Participaram do experimento 52 estudantes de graduação e pós-graduação da Universidade Federal de Minas Gerais. Todos os participantes deram consentimento formal, por escrito, autorizando a utilização dos dados nesta pesquisa. Todos os participantes eram falantes nativos de Português Brasileiro, com idades entre 18 e 30 anos (média de 20 anos). Dos 52 respondentes, 46 (88%) eram mulheres.

## **4.3 Procedimentos**

Os questionários foram aplicados de modo on-line por meio das plataformas *Google Forms* em uma única sessão de 15 minutos, em média. Os participantes eram solicitados a atribuir uma nota em uma escala de 1-7 sobre o quão comum no mundo real seriam os eventos citados na tarefa. Na escala, a nota 1 representava algo incomum e a nota 7 representava algo muito comum. A seguir, apresentamos uma imagem da tela de julgamento visualizada pelos participantes.

**1. QUÃO COMUM É**

um vinho ser derramado? \*

1 2 3 4 5 6 7

Incomum        Muito comum

um pato ser derramado? \*

1 2 3 4 5 6 7

Incomum        Muito comum

um leite ser derramado? \*

1 2 3 4 5 6 7

Incomum        Muito comum

um suco ser derramado? \*

1 2 3 4 5 6 7

Incomum        Muito comum

Figura 10 – Formulário do julgamento de ajuste temático

Antes do início da tarefa, havia instruções detalhadas sobre o que realizar e um exemplo do tipo de julgamento que o participante deveria realizar. Os julgamentos foram divididos em duas listas, com isso, todos os participantes realizaram julgamentos dos dois tipos de verbos – forte subseleção e fraca subseleção – em todos os níveis de previsibilidade – (1) mais previsível; (2) pouco previsível; (3) imprevisível, porém plausível; (4) implausível. Cada julgamento só poderia pertencer a uma das listas. Os participantes só realizaram os julgamentos de uma das listas. A ordem dos julgamentos foi aleatorizada, mas todos os julgamentos relacionados a um determinado verbo sempre apareciam na mesma tela – como exemplificado acima.

#### 4.4 Resultados

Para explorar a influência da capacidade de subseleção dos verbos e da previsibilidade dos argumentos internos nos julgamentos de ajuste temático, foi utilizado o pacote *cumulative link mixed models (CLMM)* do *software* R, equipado com a aproximação de Laplace (CHRISTENSEN, 2015). O CLMM facilita a análise de dados ordinais – como os



juízos realizados em nossas tarefas – por meio de modelos mistos de links cumulativos<sup>46</sup>. Em primeiro lugar, o modelo de link cumulativo se adequa à natureza ordinal da variável resposta. Uma variável ordinal é uma variável categórica cujos níveis têm uma ordenação natural. Apesar das diversas discussões sobre como se deve analisar dados provenientes de escalas de avaliação<sup>47</sup>, não existe muita consonância sobre o assunto. Acreditamos que, se adotássemos um tratamento conservador e utilizássemos métodos não paramétricos tradicionais (teste Mann-Whitney, teste de Friedman, entre outros), o poder estatístico de nossa análise seria diminuído, pois tais testes não permitem a análise simultânea de diversos efeitos aleatórios.

Em segundo lugar, o uso de modelos mistos é apropriado para desenhos experimentais que utilizam dados pareados e medidas repetidas, como as utilizadas neste estudo. Nesse tipo de modelo, a análise dos dados do juízo de ajuste temático se justifica pela possibilidade de manipular variáveis de efeito fixo e efeito aleatório.

Em nossa análise, ajustamos o modelo linear com os juízos de ajuste temático como variável resposta, e o nível de previsibilidade dos argumentos internos por parte dos verbos e a força de subseleção como variáveis preditoras. A comparação de dois modelos aninhados mostrou que a interação entre a capacidade da seleção de argumentos internos por parte dos verbos e a previsibilidade dos argumentos melhor contribuem para explicar a variação entre os juízos ( $\chi^2 [3] = 11.797, p < 0.01$ , ajustado por FDR<sup>48</sup>). No modelo, também estava incluído um fator aleatório de juízo das condições em razão dos sujeitos (*random slope*), além de um intercepto aleatório por itens e por sujeito.

<sup>46</sup> Tradução nossa para *Cumulative Link Mixed Models*.

<sup>47</sup> Para uma revisão sobre o assunto, ver Harpe (2015).

<sup>48</sup> Benjamini e Hochberg (1995) propuseram controlar a FDR, definida como a proporção de hipóteses nulas  $H_0$  verdadeiras entre as hipóteses nulas rejeitadas, ou seja, a proporção de erros devido à rejeição incorreta de  $H_0$ . Diferentemente do nível de significância, o qual é pré-estabelecido antes de iniciar as análises, o FDR é calculado a partir das informações presentes nos dados após a realização dos múltiplos testes de hipóteses. (GONÇALVES, 2013).

Uma análise post-hoc foi feita por meio do pacote *Lsmeans*<sup>49</sup> (LENTH, 2016). Foram realizadas 28 comparações entre os fatores presentes no teste. O p-valor foi ajustado pelo método Tukey<sup>50</sup>. A seguir, está exposto o quadro com as 28 comparações:

julg ~ cond* tipo + (1 + cond   suj) + (1   item)							
Fraca				Forte			
Contraste	$\beta$	Z	P	Contraste	$\beta$	Z	P
Mais - Pouco	1.62135	3.704	0.0052	Mais - Pouco	2.93677	6.499	<.0001
Imprevisível - Pouco	-0.69649	-1.779	0.6346	Imprevisível - Pouco	-0.03720	0.095	1.0000
Imprevisível - Mais	-2.31784	-5.312	<.0001	Imprevisível - Mais	-2.97396	-6.578	<.0001
Implausível - Mais	-8.39567	-14.979	<.0001	Implausível - Mais	-10.34714	-17.561	<.0001

Figura 11 – Quadro de análises post-hoc dos resultados do melhor modelo ajustado:  $julg \sim con * tipo + (1 + cond \setminus suj) + (1 | item)$

<sup>49</sup> *Lsmeans* é um pacote que possibilita a comparação de valores entre grupos em modelos lineares, em modelos de links cumulativos e em outros tipos de modelos. Além disso, esse pacote permite as comparações entre interações de fatores. As médias de mínimos quadrados são médias de preditores que são ajustados para médias de outros fatores no modelo. Ao realizar múltiplas comparações, é indicado o ajuste do p-valor.

<sup>50</sup> O teste de Tukey é uma forma robusta de se realizar comparações entre diversos pares. Também é conhecido como Tukey HSD (Teste de Tukey da Diferença Honestamente Significativa). O teste de Tukey foi desenvolvido por John Wilder Tukey e apresentado em 1949 no artigo titulado “Comparing Individual Means in the Analysis of Variance”. Esse teste consiste na comparação dos possíveis pares de médias e se baseia na diferença mínima significativa (D.M.S.), considerando os percentis do grupo.

CONTRASTE FORTE - FRACO			
Previsibilidade	$\beta$	Z	P
Mais	0.98616	2.371	0.0236
Implausível	-0.96529	-2.240	0.0319

Figura 12 – Quadro de análises post-hoc - contraste entre verbos de forte e fraca subseleção

Os resultados indicaram que os participantes avaliaram como mais improváveis – atribuíram valores menores – os eventos na condição implausível quando o tipo de verbo era de forte subseleção do que os eventos na mesma condição em que o tipo de verbo era de fraca subseleção [ $\beta = -0.96528901$ ;  $Z = -2.240$ ,  $p < 0.05$ ].

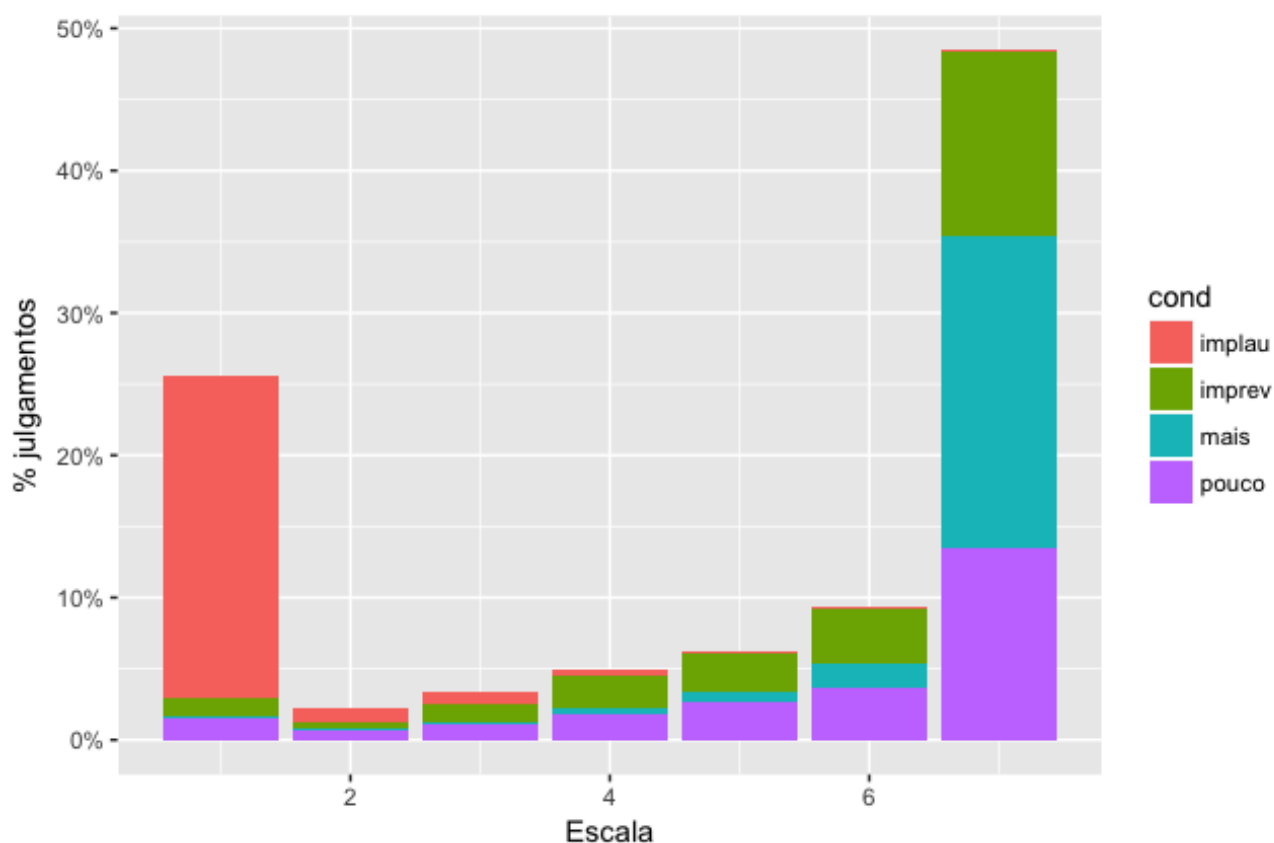


Figura 13 - Gráfico de frequência de julgamentos dos verbos de forte subseleção

Na comparação entre as condições em que o argumento interno era mais previsível, os julgamentos foram mais altos quando os verbos eram de forte subseleção. Ou seja, os participantes consideraram mais habituais os eventos em que o argumento interno era mais previsível e o tipo de verbo era de forte subseleção [ $\beta = 0.98616466$ ;  $Z = 2.371$ ,  $p < 0.05$ ].

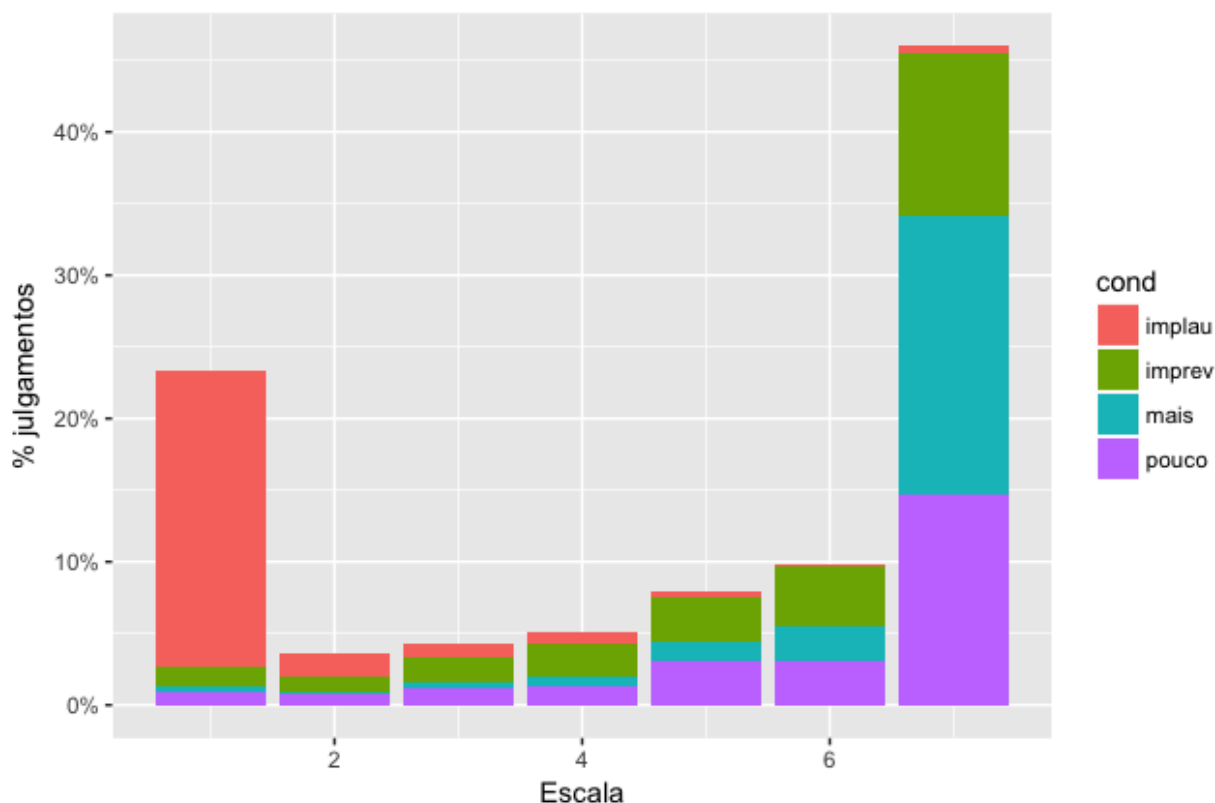


Figura 14 - Gráfico de frequência de julgamentos dos verbos de fraca subseleção

Nos verbos de fraca subseleção, as condições em que o argumento interno era mais previsível receberam julgamentos significativamente maiores do que as condições em que o argumento era pouco previsível [ $\beta = 1.62134605$ ;  $Z = 3.704$ ,  $p < 0.01$ ]. Os julgamentos eram significativamente maiores nas condições em que o argumento interno era mais previsível e os verbos eram de forte subseleção [ $\beta = 2.93676738$ ;  $Z = -6.499$ ,  $p < .0001$ ] do que nas condições em que argumento interno era pouco previsível e os verbos eram de forte subseleção.

Por fim, é importante ressaltar que nos dois tipos de verbos – forte e fraca subseleção – não houve diferenças significativas entre os julgamentos das condições imprevisíveis e

pouco previsíveis. Com isso, nos verbos de forte subseleção, a comparação entre argumentos internos imprevisíveis e pouco previsíveis mostrou os seguintes resultados [ $\beta = -0.03719629$ ;  $Z = -0.095$ ,  $p = 0.9242$ ]. Nos verbos de fraca subseleção, a comparação entre argumentos internos imprevisíveis e pouco previsíveis [ $\beta = -0.69649313$ ;  $Z = 1.779$ ,  $p = 0.6346$ ] não mostrou diferenças.

#### 4.5 Discussão

Nossas hipóteses para este experimento previam que os argumentos mais relacionados aos verbos de forte subseleção seriam considerados mais comuns do que os argumentos dos verbos de fraca subseleção. Isso ocorreria porque os verbos de forte subseleção apresentariam papéis temáticos mais claramente definidos do que os verbos de fraca subseleção. Ademais, prevíamos que os verbos de fraca subseleção não apresentariam grandes diferenças nos julgamentos de ajuste temático de acordo com níveis de previsibilidade dos argumentos.

A primeira hipótese se confirmou em certa medida. Os resultados de nosso julgamento de ajuste temático sugerem que os argumentos internos mais previsíveis de forte subseleção eram julgados mais comuns do que os argumentos internos mais previsíveis de fraca subseleção. Ou seja, os argumentos mais previsíveis de forte subseleção recebiam notas mais altas do que os argumentos mais previsíveis de fraca subseleção. Por exemplo, os participantes julgavam mais comuns os eventos do tipo (36) do que os eventos do tipo (37):

\* Quão comum é:

(36) Uma **ordem** ser acatada? [ Condição muito previsível de forte subseleção]

(37) Uma **carne** ser moída? [Condição muito previsível de fraca subseleção]

Quando os argumentos internos eram implausíveis de forte subseleção, eles eram julgados menos comuns do que os argumentos internos implausíveis de fraca subseleção. Ou seja, os argumentos implausíveis de forte subseleção recebiam notas mais baixas do que os argumentos implausíveis de fraca subseleção. Por exemplo, os participantes julgavam menos comuns os eventos do tipo (38) do que os eventos do tipo (39):

\* Quão comum é:

(38) Uma **nuvem** ser guiada? [ Condição implausível de forte subseleção]

(39) Um **sangue** ser quebrado? [Condição implausível de fraca subseleção]

Esses resultados podem ser compreendidos como indicações de que as propriedades temáticas dos grupos de verbos diferenciam-se com relação à delimitação de um protótipo de papel temático. Contudo, essa diferenciação clara entre os dois grupos foi encontrada apenas entre argumentos mais previsíveis e argumentos implausíveis. É possível que, por estar restrito aos limites extremos da previsibilidade, tal efeito esteja primordialmente ligado aos eventos expressos por verbos de forte poder de subseleção de argumentos internos.

Entre os itens pouco previsíveis e imprevisíveis, porém plausíveis não houve diferenciação entre os julgamentos nas condições de forte subseleção e fraca subseleção. Por exemplo, nos verbos de fraca subseleção, os participantes julgaram que eventos como os exemplos abaixo não apresentavam diferenças significativas.

\*Quão comum é:

(40) Uma **marca** ser promovida? [ Condição pouco previsível de fraca subseleção]

(41) Um **hotel** ser promovido? [Condição imprevisível, porém plausível de fraca subseleção]

Os participantes também julgaram da mesma forma os eventos nos verbos de forte subseleção, como no exemplo a seguir:

\*Quão comum é:

(42) Uma **regra** ser acatada? [ Condição pouco previsível de forte subseleção]

(43) Uma **norma** ser acatada? [Condição imprevisível, porém plausível de forte subseleção]

Uma possível explicação para essa ausência de diferenciação deve-se à probabilidade de os falantes julgarem todos os eventos imprevisíveis ou pouco previsíveis da mesma forma, já que são considerados inesperados de maneira geral. Entretanto, seria relevante investigar se tal diferença se mantém em experimentos de dados on-line, uma vez que existem evidências de que o cloze – logo, a previsibilidade – possui efeito logarítmico (SMITH; LEVY, 2013). Ou seja, é possível que ele afete mais as comparações entre medidas de processamento de itens que possuem índices de previsibilidade bastante baixos – de cloze próximo a 0.

Contrariando uma de nossas hipóteses, os verbos de fraca subseleção não apresentaram julgamentos semelhantes entre as condições de previsibilidade do argumento interno. Eventos mais previsíveis foram julgados significativamente mais comuns do que eventos pouco previsíveis; imprevisíveis, porém plausíveis e implausíveis. Ou seja, receberam notas mais altas. Como demonstrado acima, eventos pouco previsíveis e imprevisíveis, porém plausíveis foram julgados de maneira semelhante. Além disso, os eventos implausíveis foram julgados significativamente menos comuns do que eventos mais previsíveis; pouco previsíveis e imprevisíveis, porém plausíveis. Ou seja, receberam notas mais baixas.

Esse comportamento sugere que tais verbos sofrem influência de preferências temáticas relacionadas ao teste de cloze, que podem se tornar mais evidentes em contextos reduzidos. O fato de o efeito do julgamento dos níveis de previsibilidade ser semelhante entre os eventos dos verbos de forte subseleção e fraca subseleção pode indicar um padrão de influência de conhecimentos sobre o mundo real nos julgamentos exercidos a partir das respostas do teste de cloze. Como o efeito de diferenciação só ocorre entre os níveis extremos de previsibilidade – entre os níveis mais previsíveis e implausíveis – é possível que, em contextos reduzidos, os verbos de fraca subseleção exibam leves preferências por argumentos mais previsíveis, direcionadas por conhecimentos mais

detalhados sobre o mundo real. Além disso, os julgamentos dos argumentos implausíveis sugerem ser influenciados por conhecimentos mais gerais sobre o mundo real, pois parecem ser claramente afetados pela plausibilidade dos eventos expressos em contextos reduzidos.

Tendo em vista que os julgamentos de ajuste temático não são suficientes para mostrar se o conhecimento temático está diretamente relacionado à representação do verbo, consideramos relevante realizar experimentos que utilizem medidas on-line do processamento da linguagem. As tarefas têm o objetivo de verificar se as diferenças no delineamento do papel temático estão principalmente relacionadas aos verbos de forte subseleção devido às diferenças verificadas entre os verbos de forte e fraca subseleção. Além disso, verificaremos se essas diferenças estão principalmente relacionadas aos extremos dos níveis de previsibilidade – mais previsível e implausível. Acreditamos que a verificação de uma possível relação entre os julgamentos temáticos e o processamento lexical contribui para os estudos a respeito dos tipos de traços representacionais que são ativados durante o processamento antecipatório da linguagem. Para verificar esta hipótese, foram realizados dois experimentos de decisão lexical.



## 5 DECISÃO LEXICAL

Até o momento, verificamos, por meio de tarefas de julgamentos off-line, que os eventos expressos por verbos e argumentos considerados mais previsíveis possuem probabilidades de julgamento divergentes de qualquer outro nível de previsibilidade, ao serem considerados eventos mais comuns. Tal comportamento parece refletir a percepção dos falantes sobre o quão estereotípicos são certos argumentos relacionados a verbos específicos, uma vez que os verbos de forte subseleção foram considerados mais comuns do que os verbos de fraca subseleção. Além disso, sugerimos que a plausibilidade dos argumentos também parece ser relacionada a verbos específicos, uma vez que os eventos implausíveis expressos por verbos de forte subseleção foram considerados menos comuns do que os eventos implausíveis expressos por verbos de fraca subseleção. Contudo, como os resultados do julgamento de ajuste temático e da tarefa de cloze se referem a tarefas off-line, é relevante para esta pesquisa verificar como a relação entre verbos de forte e fraca subseleção, e argumentos mais esperados e implausíveis se comportam em tarefas de processamento on-line.

Com o objetivo de verificarmos, em uma tarefa de processamento on-line, se o comportamento temático dos verbos de forte e fraca subseleção se assemelha aos padrões de resultados verificados no julgamento de ajuste temático realizado anteriormente, efetuamos duas tarefas de decisão lexical. Ademais, essas tarefas foram realizadas com o objetivo de observar como os verbos, quando em contextos mínimos, realizam a previsão de informações temáticas relacionadas aos argumentos internos. Os experimentos, que foram baseados nas propostas de Ferretti, McRae e Hatherell (2001) e Delaney-Busch *et al.* (2019), consistiram em tarefa que utilizou os verbos de forte e de fraca subseleção, e os argumentos internos mais previsíveis e implausíveis que coletamos na primeira fase deste estudo.

Cada uma das tarefas realizadas neste trabalho reflete, em certa medida, a perspectiva de dois diferentes trabalhos. Por um lado, na primeira tarefa, verificamos, como Ferretti e colegas (2001), se o efeito de facilitação entre verbos e argumentos (*e.g.*: “prender” – “bandido”) é extremamente rápido e pode ser usado para mostrar que o conhecimento de mundo está fortemente ligado à atribuição de papéis temáticos on-line e, portanto,

deve ser considerado como parte do conhecimento temático ativado durante o uso dos verbos no processamento linguístico.

Por outro lado, na segunda tarefa, como Delaney-Busch *et al.* (2019), observamos se o efeito de *priming* pode estar relacionado a processos estratégicos de previsão que podem ocorrer durante a tarefa de decisão, o que pode demandar um processamento semântico mais profundo e temporalmente mais tardio. É importante destacar que, embora Delaney-Busch *et al.* (2019) tratem de processos estratégicos de previsão, a ativação de conhecimentos de mundo presentes na memória de longo prazo também é considerada no processamento de informações derivadas dos verbos. Neste trabalho, acreditamos ser possível que efeitos contextuais provenientes do ambiente experimental afetem a ativação de conhecimentos sobre o mundo real durante a ativação dos verbos, tornando os efeitos de *priming* um pouco mais tardios. Além disso, ponderamos que, devido ao tamanho das palavras do português brasileiro, o processamento de informações de níveis mais baixos – como ortográfico e o fonológico – é realizado em períodos mais curtos do que os demandados por informações semânticas (HINO; LUPKER.; PEXMAN; 2002)

Com isso, por meio das duas tarefas de decisão lexical, pretendemos investigar duas questões. A primeira questão se relaciona à ativação automática ou estratégica dos mecanismos de *priming* que relacionam verbos à argumentos internos. Observaremos se a ocorrência do efeito de *priming* está relacionada a ativações automáticas, por meio de tempos de ativação de palavras mais curtos. Também verificaremos se a ocorrência do efeito de *priming* é influenciada por processos estratégicos, que demandam um tempo maior de ativação de palavras.

A segunda questão refere-se ao tipo de verbo sujeito aos efeitos de *priming*. Observaremos se, como no julgamento de ajuste temático, haverá um benefício geral para pares de verbos e argumentos mais esperados ou relacionados independentemente do tipo de verbo, ou se o efeito de *priming* será manifestado apenas entre os pares em que os verbos possuem forte poder de subseleção. Se considerarmos a hipótese principal deste trabalho, não faremos nenhuma previsão específica a respeito do processo de ativação das palavras. Contudo, com relação à ativação relacionada ao tipo

de verbo, esperamos encontrar efeitos de *priming* apenas entre os verbos de forte subseleção, devido à facilidade de ativação de traços temáticos de verbos que possuem probabilisticamente o papel temático mais bem delimitado.

No delineamento das tarefas dos dois experimentos, verificamos se existiam diferenças no processo de decisão lexical após a apresentação de pares de palavras formados por verbos com diferentes forças de subseleção e argumentos internos de diferentes previsibilidades. Na tarefa, utilizamos pares de palavras formados pelas seguintes condições: (i) argumentos relacionados/mais esperados, verbos de forte subseleção; (ii) argumentos relacionados/mais esperados, verbos de fraca subseleção; (iii) argumentos não relacionados/implausíveis, verbos de forte subseleção; (iv) argumentos não relacionados/implausíveis, verbos de fraca subseleção.

A tarefa de decisão lexical consiste em julgar se um estímulo alvo apresentado após um estímulo *prime* – em nosso trabalho, o argumento interno apresentado após verbo – consiste em uma palavra que existe na língua portuguesa. Os participantes são orientados a realizar essa tarefa de julgamento o mais rapidamente e corretamente possível.

Nas tarefas de decisão lexical, o relacionamento semântico entre a palavra *prime* (palavra antecedente) e uma palavra alvo é uma das variáveis preditoras manipuladas. Nas condições relacionadas, a palavra *prime* é geralmente relacionada semanticamente à palavra alvo. Nas condições controle, o *prime* é geralmente uma palavra não relacionada semanticamente ao alvo. No caso de nossos experimentos, utilizamos os argumentos implausíveis nessas condições.

Ademais, o intervalo entre apresentação do *prime* e o início da apresentação da palavra alvo – o *Stimulus Onset Asynchrony* (SOA) – pode ser uma das variáveis a serem manipuladas. Neste trabalho, o SOA foi manipulado apenas entre os dois experimentos. Como estamos interessados em processos de pré-ativação de traços semânticos, acreditamos ser possível que esses processos exijam um SOA maior, devido à influência do contexto experimental mínimo que pode afetar a ativação dos traços temáticos.

Também consideramos que o tamanho das palavras do português brasileiro necessita de SOAs maiores. Outrossim, acreditamos que a ativação de informações de alto nível, como as informações conceituais, pode ter um curso temporal um pouco maior, devido à ativação inicial de informações de baixo nível, como as informações ortográficas e fonológicas. Com isso, em nosso primeiro experimento o SOA foi de 250ms, mas já no segundo esse tempo foi aumentado para 600ms. Os esquemas com a apresentação dos estímulos utilizados na tarefa serão apresentados na próxima seção.

No primeiro experimento, possuíamos a hipótese de que, caso os julgamentos dos argumentos internos previsíveis ocorressem de forma mais rápida do que os julgamentos de argumentos imprevisíveis após a apresentação de um verbo de forte subseleção, os resultados estariam em consonância com Ferretti, McRae e Hatherell (2001), cuja proposta previa que, se o significado de um verbo com papel temático bem delimitado é fortemente influenciado pelos conceitos dos nomes que preenchem a função de argumento interno, é possível que informações temáticas façam parte do significado dos verbos. Para tais autores, o acesso a informações temáticas on-line ocorreria de maneira automática e extremamente rápida, o que demandaria o uso de um paradigma de SOA curto – no caso, 250ms. O uso de um SOA curto supostamente indicaria que os prováveis efeitos de facilitação temporal nas respostas da tarefa são fruto da íntima relação entre os verbos e a representação de informações derivadas do conhecimento de mundo.

No segundo experimento, com o propósito de verificar se a relação temática entre os verbos de forte subseleção e os argumentos muito previsíveis pode ocasionar estratégias de previsões de palavras, mesmo em contextos mínimos, utilizamos um SOA longo, de 600ms. A opção por utilizar tal duração do SOA foi influenciada pelo trabalho de Delaney-Busch *et al.* (2019), que identificou uma atenuação e adaptação da amplitude do N400 de acordo com informações probabilísticas do contexto experimental, em circunstâncias semelhantes às encontradas em uma tarefa de decisão lexical. Essa atenuação ocorreu devido à influência da validade preditiva do ambiente experimental, que permitiu a antecipação de informações sobre a frequência de ocorrência de pares de palavras que viriam a ocorrer na tarefa. No trabalho, o SOA possuía a duração de 600ms.

Portanto, a fim de avaliar a natureza da ativação on-line de informações temáticas provenientes de verbos de forte e fraca subseleção, realizamos as tarefas de decisão lexical cujas hipóteses foram expostas acima. Nas seções subsequentes, serão detalhados os materiais, os procedimentos e os resultados dessas tarefas.

## 5.1 Tarefa de decisão lexical 1

### 5.1.1 Materiais

O experimento possuía 96 itens experimentais que foram divididos em duas listas. Os participantes foram expostos a uma das duas diferentes listas, que eram compostas por 96 itens no total. Cada lista possuía 48 itens experimentais e 48 itens distratores. Os itens distratores consistiam em 24 pares formados por palavras (*primes*) e pseudopalavras (alvos), e em 24 pares formados por pseudopalavras (*primes*) e pseudopalavras (alvos). Foram usados os mesmos distratores nas duas listas exibidas aos participantes. Dos 48 itens experimentais, cada lista possuía (i) 12 pares experimentais compostos por verbos de forte subseleção e argumentos relacionados mais previsíveis; (ii) 12 pares experimentais compostos por verbos de forte subseleção e argumentos não relacionados implausíveis; (iii) 12 pares experimentais compostos por verbos de fraca subseleção e argumentos relacionados mais previsíveis; (iv) 12 pares experimentais compostos por verbos de fraca subseleção e argumentos não relacionados implausíveis. Na tarefa, todos os verbos eram usados como *prime* e todos argumentos internos eram usados como alvo.

Em cada lista, 72 itens constituíam pares de palavras não associadas e 24 itens constituíam pares de palavras associadas. Em 50% dos itens, os participantes efetuavam uma decisão lexical a respeito de uma pseudopalavra e, nos 50% restantes, a decisão era a respeito de uma palavra. Todos os verbos e argumentos ocorriam apenas uma vez em cada uma das listas. Isto é, caso o verbo de forte subseleção “derramar” ocorresse com juntamente com a palavra “leite” – argumento relacionado mais esperado – em uma das listas, tal verbo ocorreria com a palavra “casa” – argumento não relacionado implausível – na outra lista. Portanto, cada verbo era visualizado uma única

vez por cada participante. 50% dos verbos eram de forte subseleção e 50% dos verbos eram de fraca subseleção. Todos os argumentos não relacionados implausíveis foram constituídos por uma reorganização dos pares em que os argumentos relacionados mais esperados passavam a ser não relacionados implausíveis. Dessa forma, as palavras que constavam na condição controle eram as mesmas palavras que estavam presentes na condição alvo. Todas as palavras na função de argumento eram visualizadas uma única vez pelos participantes.

Por conseguinte, os itens de cada lista se organizavam nas condições<sup>51</sup> listadas abaixo:

Condição	<i>Prime</i>	Alvo
Distratores palavra/pseudopalavra	Fósforo	Mopra
Distratores pseudopalavra/pseudopalavra	Lometa	Cobe
Experimentais argumento relacionado/mais previsível, verbos de forte subseleção	Derramar	Leite
Experimentais argumento relacionado/mais previsível, verbos de fraca subseleção	Balear	Ladrão

<sup>51</sup> Todos os itens presentes nas duas listas utilizadas na tarefa estão expostos no APÊNDICE D.

Experimentais Argumento não relacionado/implausível, verbos de forte subseleção	Pescar	Leite
Experimentais argumento não relacionado/implausíveis, verbos de fraca subseleção	Moer	Ladrão

*Figura 15 – Quadro com condições experimentais da tarefa de decisão lexical 1*

### 5.1.2 Participantes

Participaram da tarefa 35 alunos de graduação e pós-graduação da Universidade Federal de Minas Gerais. Todos os participantes deram consentimento formal, por escrito, autorizando a utilização dos dados nesta pesquisa. Todos eram falantes nativos de Português Brasileiro, com idades entre 18 e 30 anos (média de 22,4 anos). Dos 35 participantes, 30 eram mulheres (85,71%). Dois participantes eram canhotos (5,71%).

### 5.1.3 Procedimentos

O experimento, realizado por meio de um notebook, consistiu em uma única sessão de aproximadamente 20 minutos de duração. O software *PsychoPy Builder* (PEIRCE *et al.*, 2019) foi utilizado para a apresentação pseudoaleatorizada dos estímulos, o registro dos tempos de reação (TR) e a anotação das respostas dos julgamentos. Antes do início do experimento, os participantes foram instruídos a ler as duas combinações de letras que apareceriam na tela do computador e julgar se a segunda combinação formava uma palavra do Português Brasileiro. Foi explicado que a tecla L – que se encontrava marcada com um adesivo verde – deveria ser pressionada caso o participante julgasse que a combinação de letras formava uma palavra; já a tecla A – que se encontrava marcada com um adesivo vermelho – deveria ser pressionada caso o participante julgasse que a

combinação de letras não formava uma palavra. O tempo de reação (TR) e as respostas de cada um dos julgamentos foram registrados durante o experimento pelo software.

Cada item começava com um intervalo de 1200ms (1.20s) de tela em branco e, logo em seguida uma cruz era posicionada no centro da tela por 300ms (0.30s). Em seguida, a palavra prime era apresentada em letras minúsculas por 200ms (0.20s), logo após, uma máscara com os símbolos “&&&” era exibida por 50ms (0.05s) e a palavra alvo era apresentada em letras minúsculas por 2000ms (2.00s). Como detalhado anteriormente, optamos por um SOA mais baixo, de 250ms (0.25s), para verificarmos a automatização do processo. Após a realização do julgamento, a tela permanecia em branco por mais 1200ms (1.20s) e a cruz era novamente posicionada no centro da tela por 300ms (0.30s). Se o participante não fizesse o julgamento no intervalo em que a palavra alvo estava disposta na tela, o tempo de reação (TR) era computado como nulo.

Antes de iniciarem a tarefa, os participantes eram submetidos a uma sessão de treinamento com 10 pares de itens distratores para que pudessem se familiarizar com os procedimentos de decisão lexical. Após a realização do treinamento e a solução de eventuais dúvidas, era iniciada a tarefa.

Logo, a ordem de apresentação dos pares de palavras, durante a tarefa, obedeceu ao seguinte esquema:



Figura 16 - Ordem de apresentação dos estímulos na tarefa de decisão lexical 1

As duas listas de apresentação de estímulos necessitaram ser pseudoaleatorizadas devido a limitações do software. Cada um dos dois tipos de listas foi aleatorizado dez vezes por meio do software *Excel*, criando-se diferentes formas de exibição dos estímulos. Havia um total de 20 listas aleatorizadas que foram exibidas aos



participantes. Das 20 listas aleatorizadas, 14 listas foram exibidas para 2 participantes e 7 listas foram exibidas para apenas 1 participante. Tal procedimento foi realizado com o objetivo de não criar qualquer viés devido à ordem de apresentação dos estímulos.

#### 5.1.4 Resultados

Foram descartados os dados de uma participante por não atingir o patamar mínimo de 80% de acertos na tarefa de decisão lexical. Com isso, foram analisados os dados de tempo de reação (TR) de 34 participantes – 29 mulheres e 5 homens. Apenas os TR dos julgamentos que se referiam às condições experimentais foram analisados. 0,15% dos itens experimentais (5 itens) foram eliminados por ultrapassarem o limite de tempo para realização do julgamento (2000ms). 0,64% dos itens experimentais (21 itens) foram eliminados por erros na tarefa de decisão. No total, foram analisadas 1606 observações de itens experimentais de um total de 1632. O histograma a seguir mostra a distribuição dos dados de TR:

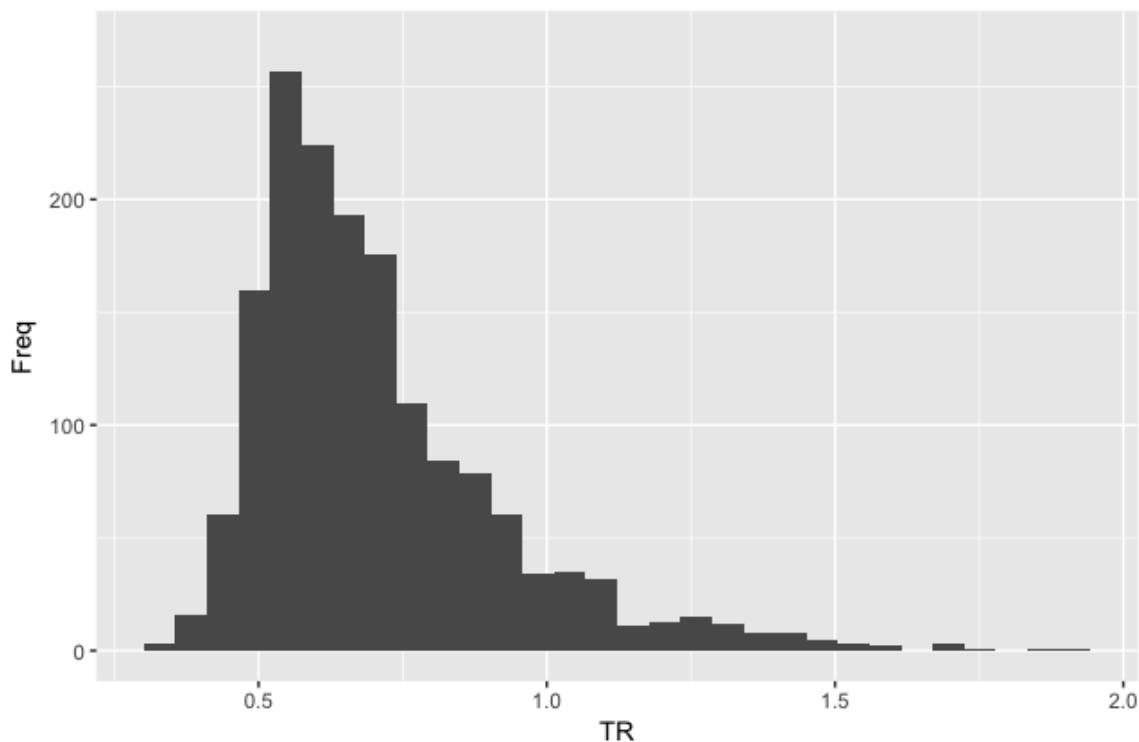


Figura 17 - Histograma de distribuição dos TRs da tarefa de decisão lexical 1

Na análise dos dados, seguimos a proposta de Lo e Andrews (2015), para os quais os dados de TR contêm idiosincrasias que demandam dos pesquisadores o uso de múltiplos participantes e de medidas repetidas. A distribuição desses TRs, que geralmente é positivamente inclinada, leva ao uso de técnicas estatísticas que podem ocasionar problemas interpretativos. Os autores afirmam que a técnica de transformação de dados, para evitar problemas provenientes de violação da distribuição normal, não é a mais adequada quando se refere a variáveis respostas como TRs.

Segundo Lo e Andrews (2015), diferenças entre os TRS das condições experimentais são compreendidas como um reflexo direto das diferenças na quantidade de tempo necessário para realizar certas operações mentais. Contudo, a transformação dos dados de uma variável resposta não apenas normaliza os resíduos, mas também distorce propriedades da escala de razão em medidas de tempo. Em outras palavras, as diferenças entre processos mentais que são medidos diretamente a partir do tempo podem ser distorcidas, ou até mesmo eliminadas, a partir da transformação dos dados.

Muitos trabalhos utilizam modelos lineares de efeito misto (LMMs) para analisar TRs provenientes de tarefas de decisão lexical, devido à sofisticação do modelo em capturar relações cruzadas entre grupos de estímulos contrabalanceados que são apresentados a diferentes sujeitos. LMMs são métodos poderosos, pois permitem investigar, em um nível bastante refinado, interações de efeitos experimentais com diferenças por indivíduos e por itens, de uma maneira que métodos tradicionais como ANOVA não são capazes, devido ao fato de as variáveis serem agregadas. Contudo, a aplicação desses modelos não é adequada em dados de variáveis respostas com distribuição assimétrica, como os TRs. A recomendação mais comum é a de que os dados devem ser transformados para evitar que poucos *outliers* obscureçam as principais tendências que caracterizam o conjunto de dados (BAAYEN; DAVIDSON; BATES, 2008).

Lo e Andrews (2015) propõem em vez de o uso da técnica de transformação de dados a aplicação de modelos lineares generalizados de efeito misto (GLMMs). Para os autores, os GLMMs permitem aos pesquisadores o uso da variável dependente mais apropriada à sua questão de pesquisa, ao mesmo tempo em que satisfazem o critério matemático de normalização dos resíduos na regressão linear ao relaxar o critério exigido nos LMMs. Como os GLMMs não assumem o padrão de distribuição Gaussiana, é possível que o pesquisador especifique a distribuição apropriada. A distribuição *Gamma* possibilita uma aproximação adequada da distribuição observada em TRs. Ademais, os GLMMs permitem definir exatamente a natureza da relação esperada entre preditor e resposta observada, o que não é possível quando os dados são transformados<sup>52</sup>. A *link function* define a natureza da relação entre a resposta observada e o construto ao qual os fatores fixos estão relacionados. No caso dos experimentos de decisão lexical, nenhuma função em especial parece ser requerida, uma vez que se assume que a resposta observada reflete diretamente o construto latente – em nosso experimento, o tempo requerido para o acesso lexical. Em casos como esse, a função que ligará os valores resposta à variável preditora no modelo será a *link function identity*.

<sup>52</sup> Quando os dados são transformados, as escalas transformadas e originais são conectadas por uma *link function* diferenciável monotônica que permite a retransformação para a métrica original, fornecendo um mapeamento um-para-um entre a faixa de valores ajustados produzidos pelo preditor linear na métrica transformada e a faixa de valores observados na métrica original. (LO; ANDREWS, 2015).

Portanto, diante da discussão acima, optamos por adotar em nossa análise de dados um modelo linear generalizado de efeito misto (GLMM). Para explorar a influência da capacidade de subseleção dos verbos e do relacionamento/previsibilidade dos argumentos internos nos TRs da tarefa de decisão lexical, foi utilizada a função *Generalized Linear Mixed-Effects Models (GLMER)*, equipada com a aproximação de Laplace, do pacote *LME4* (BATES; MAECHLER & BOLKER, 2012) do *software R*. No modelo, assumimos que a distribuição de nossos dados pertence à família *Gamma*, com *link function identity*. Foram realizadas três comparações entre modelos aninhados. Não houve efeito das variáveis relacionamento/previsibilidade dos argumentos e o tipo de subseleção dos verbos para o ajuste do modelo ( $\chi^2 [1] = 0.1005$ ,  $p = 1$ , ajustado por FDR); ( $\chi^2 [1] = 0.0469$ ,  $p = 1$ , ajustado por FDR); ( $\chi^2 [0] = 0.0$ ,  $p = 1$ , ajustado por FDR).

Os resultados não mostraram qualquer efeito das variáveis preditoras nos tempos de reação (TRs). Isto é, não foram encontradas diferenças significativas entre os tempos de reação de nenhuma das condições<sup>53</sup>. Uma análise *post-hoc* foi feita por meio do pacote *Lsmeans* (LENTH, 2016). Foram realizadas seis comparações entre os fatores presentes no teste. O p-valor foi ajustado pelo método *Tukey*. A seguir, está exposto um quadro com as quatro comparações de boxplot dos TRs:

<sup>53</sup> Detalhes no quadro a seguir.

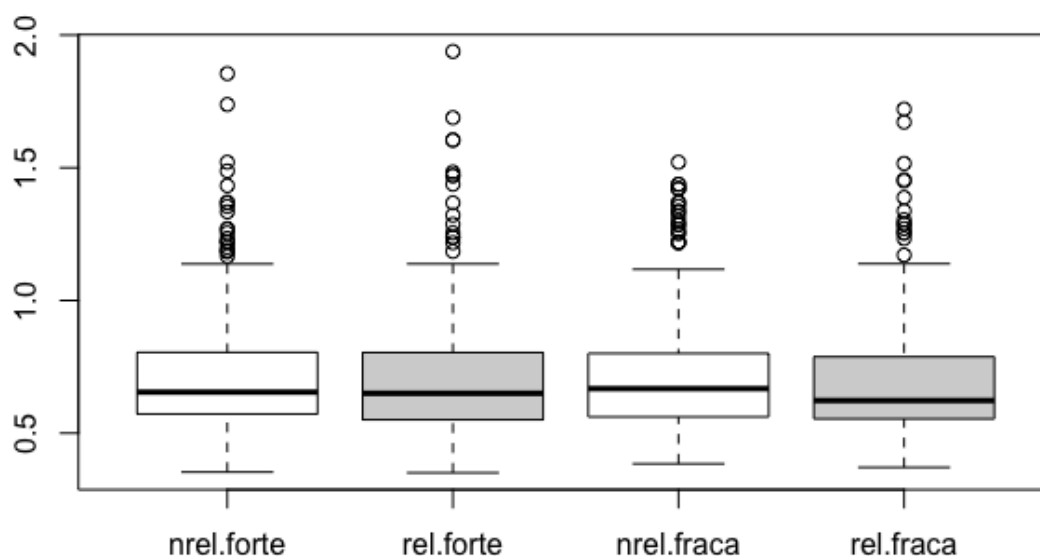


Figura 18 - Boxplot dos TRs nas condições da tarefa de decisão lexical 1

O gráfico acima ilustra a ausência de efeitos nos TRs. As medianas das condições e as distribuições dos TRs são praticamente iguais. Não foram encontradas diferenças entre os TRs das condições em que os argumento - alvos - eram (i) não relacionados/implausíveis, de forte subseleção e (ii) relacionados/mais esperados, de forte subseleção ( $\beta = 0.016188373$ ,  $z = 1.187$ ,  $p = 0.6353$ ). Além disso, não foram encontradas diferenças entre as palavras alvo (i) não relacionadas/implausíveis, de fraca subseleção e (ii) relacionadas/mais esperadas, de fraca subseleção ( $\beta = 0.016188373$ ,  $z = 1.187$ ,  $p = 0.6353$ ). Por fim, também não foram encontradas diferenças entre as palavras alvo (i) relacionadas/mais esperadas, de forte subseleção e (ii) relacionadas/mais esperadas, de fraca subseleção ( $\beta = 0.002683316$ ,  $z = 0.148$ ,  $p = 0.9988$ ); nem mesmo foram encontradas diferenças entre argumentos (i) não relacionados/implausíveis, de forte subseleção e (ii) não relacionados/implausíveis, de fraca subseleção ( $\beta = -0.002683316$ ,  $z = 0.148$ ,  $p = 0.9988$ )

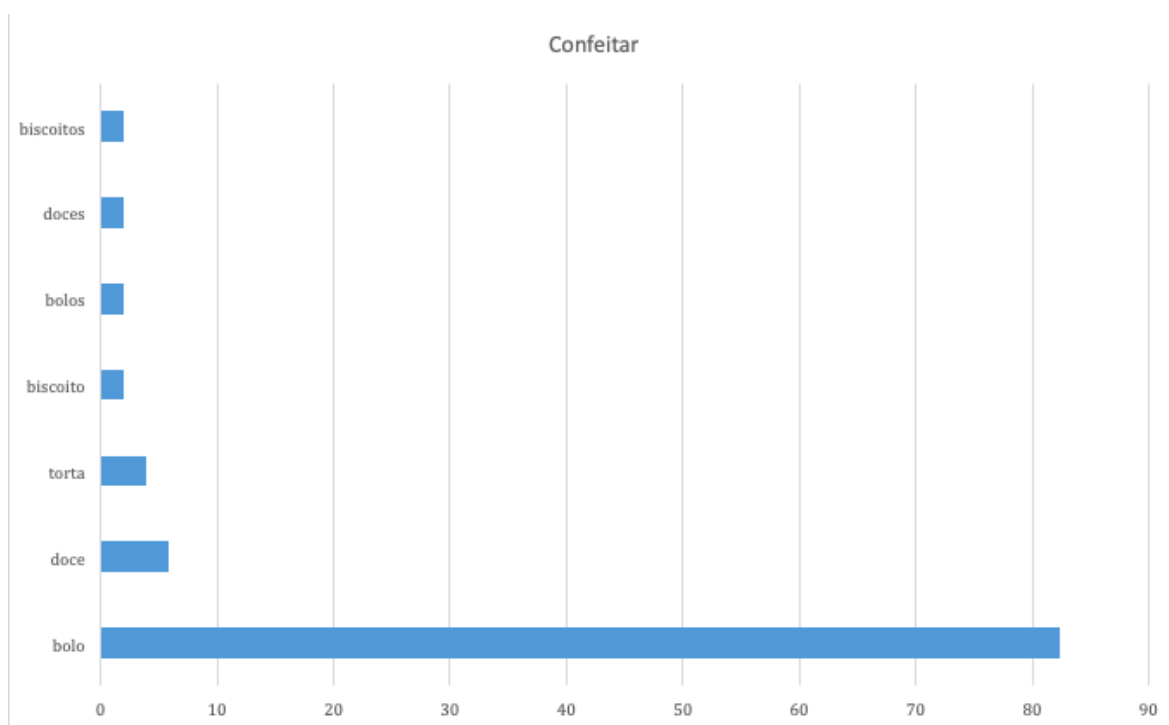
### 5.1.5 Discussão

Os resultados de nossa tarefa de decisão lexical sugerem que, sob um SOA de 250ms, os participantes não reagem temporalmente durante a tarefa de decisão lexical de maneira semelhante aos padrões de julgamentos verificados na tarefa de ajuste temático realizada anteriormente. Também sugerem que, em contextos mínimos, o efeito de facilitação de traços temáticos entre verbos de forte subseleção e argumentos muito esperados é inexistente. Não foi encontrado nenhum efeito de facilitação de processamento nas condições em que o alvo era uma palavra relacionada/mais esperada de forte subseleção em comparação a palavras não relacionadas/ implausíveis, de forte subseleção. Além disso, os tempos de reação a palavras não relacionadas/ implausíveis de forte subseleção não foram significativamente maiores dos que os tempos de reação a palavras não relacionadas/ implausíveis de fraca subseleção.

Esse resultado difere dos achados verificados no trabalho de Ferretti, McRae e Hatherell (2001), em que um efeito de *priming* semântico foi verificado entre verbos e argumentos internos cuja relação temática, por meio de julgamento de ajuste temático, foi considerada forte. Algumas diferenças entre nosso trabalho e o de Ferretti, McRae e Hatherell (2001) podem ter ocasionado resultados divergentes. O primeiro ponto refere-se ao tipo de julgamento utilizado na tarefa. Os pesquisadores utilizaram julgamentos de animacidade, em que os sujeitos julgavam se a palavra alvo era “algo vivo”, dissemelhantemente de nossos julgamentos que se referiam à existência de palavras na língua portuguesa. Supostamente, os julgamentos de animacidade correlacionam-se melhor com a ativação de informações semânticas do que as tarefas de decisão lexical, que parecem ativar informações de níveis ortográfico, fonológico e semântico. Contudo, a tarefa de animacidade impõe muitas restrições aos itens experimentais, dificultado sua aplicação em itens experimentais tão específicos como os nossos, em que a força de subseleção e a previsibilidade dos argumentos eram variáveis preditoras.

Também é importante notar que, no trabalho de Ferretti, McRae e Hatherell (2001), os julgamentos de ajuste temático foram utilizados apenas para confirmar a intuição dos pesquisadores a respeito de verbos que pareciam ter um papel temático bem definido.

Ou seja, não foi utilizado nenhum critério – como as tarefas de cloze – para confirmar o comportamento de subseleção dos verbos, inclusive para conhecer as entidades mais esperadas na posição ou o número de entidades passíveis de ocupar a posição, o que indicaria uma tendência a possuir traços temáticos bem limitados ou não. Diferenças de subseleção como as verificadas entre os verbos expostos abaixo foram utilizadas a partir da intuição do pesquisador:



*Figura 19 – Gráfico com propriedades de subseleção do verbo confeitar*

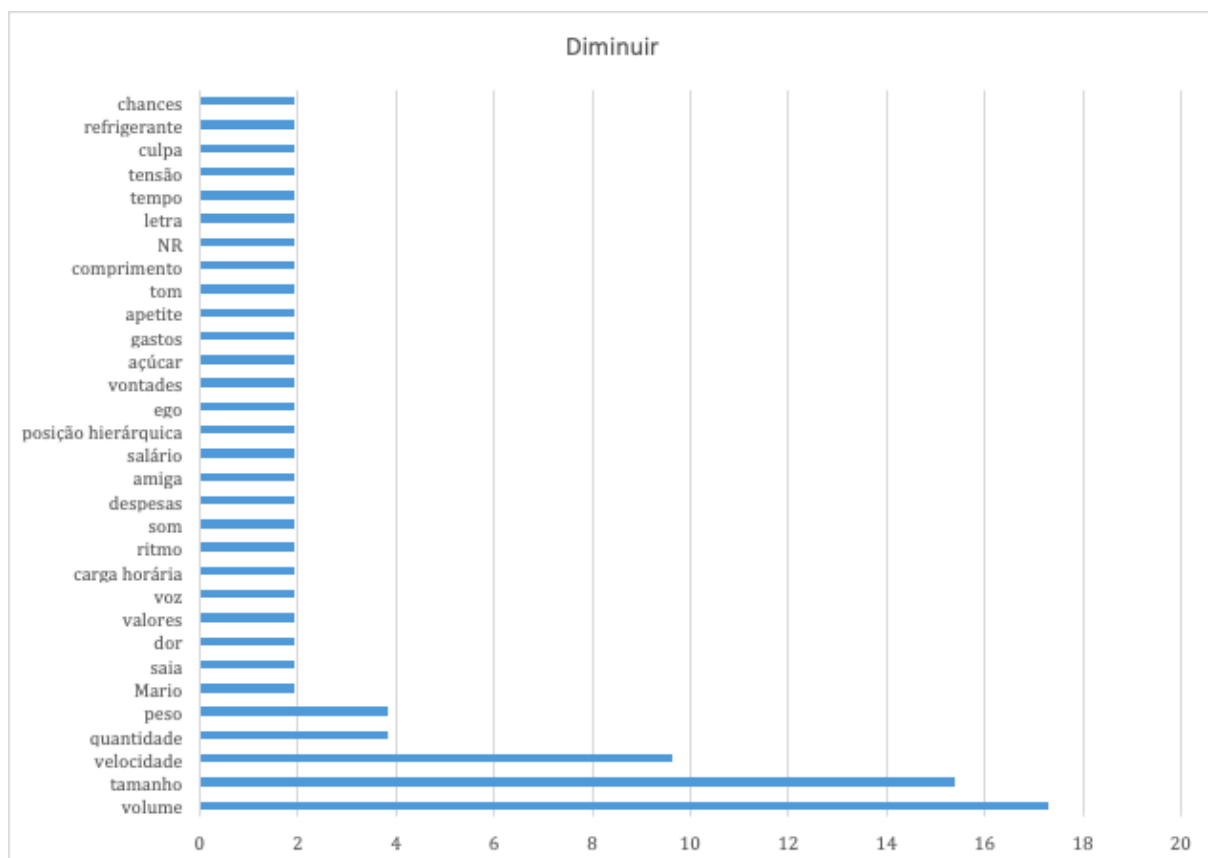


Figura 20 – Gráfico com propriedades de subseleção do verbo diminuir

Em nossos julgamentos de ajuste temático, foi notado que verbos de fraca subseleção também possuíam a tendência de receber julgamentos maiores quando o argumento era mais esperado. Assim, tendo em vista que os verbos de Ferretti, McRae e Hatherell (2001) foram selecionados conforme a intuição dos pesquisadores, não há garantias de que a tarefa de decisão lexical desses autores foi capaz de diferenciar verbos de fraca e forte subseleção. Embora essas diferenças não sejam suficientes para explicar a divergência nos resultados, é necessário abordá-las a fim de evidenciar as diferenças entre os desenhos experimentais.

Um outro ponto a ser citado diz respeito à análise estatística utilizada no artigo de 2001. Os autores reportam utilização da análise de variância (ANOVA), uma técnica que pode ser bastante problemática para a análise de dados de TRs – de natureza positivamente inclinada –, uma vez que utiliza médias das variáveis aleatórias por itens e por sujeitos.



Segundo Lo e Andrews (2015), nas análises tradicionais que utilizam média, como a ANOVA, questões relativas à falta de normalidade dos dados são geralmente ignoradas porque o método tem sido mostrado repetidamente como robusto em relação a violações da normalidade (GLASS *et al.*, 1972; LIX *et al.*, 1996). No entanto, muitas teorias cognitivas foram desenvolvidas e validadas ao questionarem a prática da utilização de médias de TRs, evidenciando muitos problemas nas interpretações de resultados (SPEELMAN; MCGANN, 2013). Os modelos lineares mistos têm impactado os estudos psicolinguísticos por preencherem uma lacuna de modelos como ANOVA, que não são capazes de registrar relacionamentos cruzados entre grupos de estímulos linguísticos contrabalanceados quando apresentados a diferentes sujeitos – o tipo de desenho experimental utilizado em nossas tarefas de decisão lexical. Com isso, esses modelos têm oferecido uma solução estatística para esse problema, tornando-se uma forma bastante recomendada de análise (BAAYEN; DAVIDSON; BATES, 2008). Portanto, as diferenças nos tratamentos estatísticos dos dois estudos são pontos importantes a se considerar ao se refletir sobre as diferenças de resultados.

Por fim, deve-se refletir sobre a questão do SOA de 250ms. Acreditamos que a ativação de traços temáticos está relacionada ao uso de conhecimentos probabilísticos sobre o mundo real. É possível que a ativação desses conhecimentos específicos ocorra de maneira estratégica – como proposto por Delaney-Busch *et al.* (2019) – demandando um curso temporal maior para que haja antecipação dos traços temáticos dos argumentos mais esperados de verbos de forte subseleção. Por conseguinte, nossos resultados parecem indicar que a ativação de traços temáticos relacionados a conhecimentos de mundo não se trata de um processo rápido e automático, como proposto por Ferretti, McRae e Hatherell (2001), mas sim de algo que parte de um processo estratégico e probabilístico de antecipação de informações.

A fim de verificar essa hipótese, realizamos uma nova tarefa de decisão lexical seguindo os mesmos procedimentos da primeira tarefa, porém utilizando um SOA de 600ms. Os detalhes dessa nova tarefa serão expostos a seguir.

## 5.2 Tarefa de decisão lexical 2

### 5.2.1 Materiais

Os materiais utilizados na tarefa de decisão lexical 2 foram exatamente os mesmos materiais utilizados na tarefa de decisão lexical 1.

### 5.2.2 Participantes

Participaram da tarefa 23 alunos de graduação e pós-graduação da Universidade Federal de Minas Gerais. Todos os participantes deram consentimento formal, por escrito, autorizando a utilização dos dados nesta pesquisa. Todos eram falantes nativos de Português Brasileiro, com idades entre 18 e 30 anos (média de 23,5 anos). Dos 23 participantes, 18 eram mulheres (78,26%). Um participante era canhoto (4,34%).

### 5.2.3 Procedimentos

Os procedimentos foram semelhantes aos realizados na tarefa de decisão lexical 1. Contudo, foram realizadas mudanças no SOA. Na tarefa de decisão lexical 2, o SOA adotado foi de 600ms.

Dessa maneira, cada item começava com um intervalo de 1200ms (1.20s) de tela em branco e, logo em seguida, uma cruz era posicionada no centro da tela por 300ms (0.30s). Em seguida, a palavra *prime* era apresentada em letras minúsculas por 500ms (0.50s), logo após, uma máscara com os símbolos “&&&” era exibida por 100ms (0.10s) e a palavra alvo era apresentada em letras minúsculas por 2000ms (2.00s). Como detalhado anteriormente, optamos nesta segunda tarefa por um SOA de 600ms (0.60s). Após a realização do julgamento, a tela permanecia em branco por mais 1200ms (1.20s) e a cruz era novamente posicionada no centro da tela por 300ms (0.30s). Se o participante não fizesse o julgamento no intervalo em que a palavra alvo estava disposta na tela, o tempo de reação (TR) era computado como nulo.

Antes de iniciarem a tarefa, os participantes eram submetidos a uma sessão de treinamento com 10 pares de itens distratores para que pudessem se familiarizar com os procedimentos de decisão lexical. Após a realização do treinamento e a solução de eventuais dúvidas, era iniciada a tarefa.

Logo, a ordem de apresentação dos pares de palavras, durante a tarefa, obedeceu ao seguinte esquema:



Figura 21 - Ordem de apresentação dos itens na tarefa de decisão lexical 2

As duas listas de apresentação de estímulos necessitaram ser pseudoaleatorizadas devido a limitações do software. Cada um dos dois tipos de listas foi aleatorizado 10 vezes por meio do software *Excel*, criando-se diferentes formas de exibição dos estímulos. Havia um total de 20 listas aleatorizadas que foram exibidas aos participantes. Das 20 listas aleatorizadas, 3 listas foram exibidas para 2 participantes e 10 listas foram exibidas para apenas 1 participante. Tal procedimento foi realizado com o objetivo de não criar qualquer viés devido à ordem de apresentação dos estímulos.

#### 5.2.4 Resultados

Foram descartados os dados de uma participante por não atingir o patamar mínimo de 80% de acertos na tarefa de decisão lexical. Com isso, foram analisados os dados de tempo de reação (TR) de 22 participantes – 17 mulheres e 5 homens. Apenas os TRs dos julgamentos que se referiam às condições experimentais foram analisados. Um item experimental (0.01% dos itens experimentais) foi eliminado por ultrapassar o limite de tempo para realização do julgamento (2000ms). Seis itens experimentais (0,59% dos itens experimentais) foram eliminados por erros na tarefa de decisão. Dois itens experimentais (0.02% dos itens experimentais) foram eliminados por terem tempos de

reação menor do que 300ms, tempo menor do que o necessário para o processo de decisão e realização do julgamento. No total, foram analisadas 1047 observações de itens experimentais de um total de 1056. O histograma a seguir mostra a distribuição dos dados de TR:

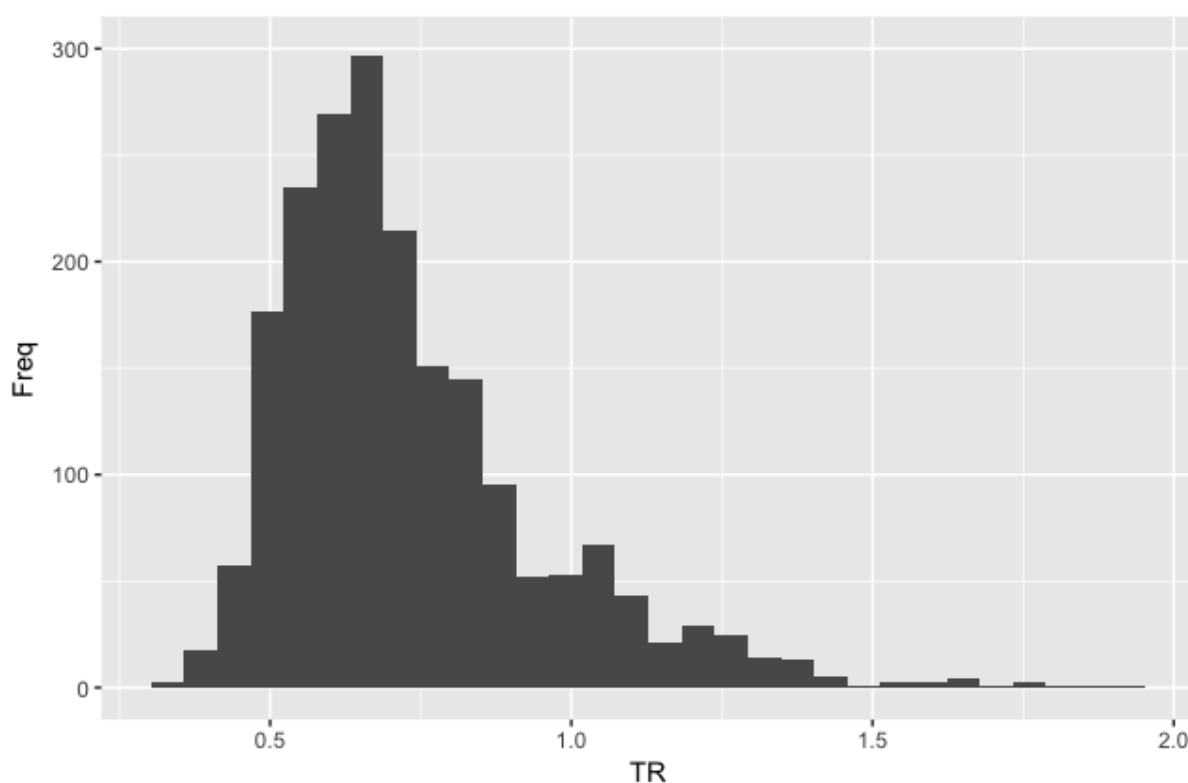


Figura 22 - Histograma de distribuição dos TRs na tarefa de decisão lexical 2

Novamente, optamos por adotar em nossa análise de dados um modelo linear generalizado de efeito misto (GLMM). Para explorar a influência da capacidade de subseleção dos verbos e do relacionamento/previsibilidade dos argumentos internos nos TRs da tarefa de decisão lexical, foi utilizada a função *Generalized Linear Mixed-Effects Models (GLMER)*, equipada com a aproximação de Laplace, do pacote *LME4* (BATES; MAECHLER & BOLKER, 2012) do software *R*. No modelo, assumimos que a distribuição de nossos dados pertence à família *Gamma*, com *link function identity*. Foram realizadas três comparações entre dois modelos aninhados. No melhor modelo ( $\chi^2 [1] = 11.357$ ,  $p < 0.001$ , ajustado por FDR), consideramos o relacionamento/previsibilidade dos argumentos e o tipo de subseleção dos verbos, com interação entre as variáveis. Como fatores aleatórios, nós utilizamos *intercepts* para sujeitos e itens, e também usamos *random slopes* para o relacionamento/previsibilidade

dos argumentos em razão dos sujeitos.

Os resultados mostraram que nas condições (1) relacionados/mais esperados, de forte subseleção, os tempos de reação (TRs) eram significativamente menores do que nas condições (2) não relacionados/implausíveis, de forte subseleção ( $\beta = -0.035655429$ ,  $z = -2.595$ ,  $p = 0.00945$ ). Uma análise *post-hoc* foi feita por meio do pacote *Lsmeans* (LENTH, 2016). Foram realizadas seis comparações entre os fatores presentes no teste. O p-valor foi ajustado pelo método *Tukey*. A seguir, está exposto o quadro com as seis comparações:

TR ~ cond*tipo + (1 alvo)+(1+cond suj)			
CONTRASTE	$\beta$	Z	P
rel,forte - nrel,forte -	-0.038997269	-3.208	0.0073
rel,forte - rel,fraca	0.009415774	0.610	0.9289
nrel,forte - nrel,fraca	0.019988609	1.275	0.5787
rel,fraca - nrel,fraca	-0.028424434	-2.340	0.0891

Figura 23 – Quadro com análises *post-hoc* dos resultados do melhor modelo ajustado: TR~ cond\*tipo+(1|alvo) + (1+cond|suj)

Após as comparações, é possível verificar que não houve diferenças significativas entre nenhuma das outras condições. Curiosamente, não foi encontrada diferença significativa nos tempos de reação (TRs) da comparação entre as condições (1) relacionados/mais esperados, de forte subseleção e (2) relacionados/mais esperados, de fraca subseleção ( $\beta = 0.003893626$ ,  $z = -0.214$ ,  $p = 0.9965$ ). Também não foram encontradas diferenças significativas na comparação entre as condições (1) relacionados/mais esperados, de fraca subseleção e (2) não relacionados/implausíveis, de fraca subseleção ( $\beta = -0.027977448$ ,  $z = -2.103$ ,  $p = 0.1520$ ).

### 5.2.5 Discussão

Os resultados de nossa segunda tarefa de decisão lexical sugerem que os pares de verbos de forte subseleção e argumentos internos relacionados/mais previsíveis desencadeiam tempos de reação (TRs) significativamente menores na tarefa de decisão lexical, quando comparados a pares formados por verbos de forte subseleção e argumentos internos não relacionados/implausíveis. Uma análise das comparações *post-hoc* sugere que não houve efeitos significativos em nenhuma das outras condições.

Nossa hipótese de que os verbos de forte subseleção e argumentos mais previsíveis apresentariam um efeito de *priming* se confirmou. Tal efeito pode ser relacionado aos julgamentos de ajuste temático em que os eventos com verbos de forte subseleção e argumentos mais esperados foram considerados mais comuns do que os eventos com verbos de fraca subseleção e argumentos mais esperados. Uma possível análise para o fato de o efeito de facilitação ter sido encontrado exclusivamente nos verbos de forte subseleção pode estar relacionada à força das relações temáticas entre verbos de forte seleção e argumentos mais esperados. É possível que, durante a atuação de mecanismos de antecipação, apesar da influência das informações probabilísticas presentes no ambiente experimental, a força das associações temáticas dos verbos de forte subseleção tenha conseguido sobressair-se. A presença desses mecanismos é evidenciada perante a presença de um SOA mais longo, ou seja, que possua um tempo maior de apresentação entre o início da palavra *prime* e o início da palavra alvo, como o utilizado neste segundo experimento – 600ms. O curso temporal maior entre a apresentação da palavra *prime* e da palavra alvo parece facilitar o processo de antecipação de traços semânticos da palavra alvo. Ademais, mesmo explorando um paradigma experimental que não aciona exclusivamente a ativação de informações semânticas (tarefa de decisão lexical), o efeito de facilitação foi verificado entre os verbos de forte subseleção.

Para Delaney-Busch *et al.* (2019), o relacionamento entre antecipação e adaptação ao ambiente experimental ocorre pelo fato de a meta principal do sujeito de uma comunicação ser a identificação de uma causa latente que melhor explique as estatísticas de ocorrência das entradas linguísticas. Nossos resultados sugerem que a associação temática entre os verbos de forte subseleção e os argumentos relacionados/mais esperados é forte o suficiente para desencadear um efeito de *priming*. Tal efeito parece estar relacionado a fortes propriedades de antecipação de traços semânticos comuns entre as palavras do par *prime*/alvo ponderadas pela adaptação do sujeito às informações presentes no contexto experimental. No paradigma da tarefa de decisão, os conhecimentos provenientes do contexto experimental parecem contribuir para a ocorrência do efeito de facilitação<sup>54</sup>. Em cada julgamento realizado durante a tarefa, o grau em que o sujeito usa seu conhecimento de longo prazo a respeito do relacionamento entre a palavra *prime* e um possível alvo para gerar previsões probabilísticas sobre o alvo é moderado pelo grau em que ele acredita que encontrará um alvo não relacionado ou relacionado.

Em nossa tarefa de decisão 2, dentre todos os pares apresentados, apenas 25% possuía verbos de forte subseleção. Dentre esses 25%, apenas 12,5% ocorriam com argumentos relacionados/mais esperados. Assim, considerando-se as informações probabilísticas de nosso ambiente experimental, podemos inferir que, na maior parte da tarefa, os sujeitos eram expostos a pares em que a palavra *prime* não possuía forte tendência probabilística de ocorrer juntamente à específica palavra alvo – 75% dos pares. Com isso, apenas o aparecimento de um verbo de forte subseleção pode provocar efeitos fortes o suficiente no reajuste do mecanismo probabilístico de antecipação, fazendo com que diferenças significativas entre os argumentos relacionados/mais esperados e não relacionados/implausíveis realizassem-se. Em outras palavras, em um ambiente experimental em que grande parte das palavras *prime* não teriam relacionamento – 50% dos casos – ou teriam um fraco relacionamento – 25% dos casos – com a palavra alvo, apenas a presença de verbos de forte subseleção faria com que os sujeitos atualizassem seus conhecimentos probabilísticos sobre o provável relacionamento entre palavra

<sup>54</sup> Neely (2012) apresenta uma revisão sobre o fato de a proporção de itens relacionados afetar a magnitude do efeito de *priming*.

*prime* e palavra alvo, mediando os conhecimentos probabilísticos sobre o ambiente experimental e o provável aparecimento de uma palavra tematicamente relacionada.



## 6. DISCUSSÃO GERAL

Este trabalho fundamentou-se no modelo de compreensão proposto por Kuperberg e Jaeger (2016), segundo o qual o falante procura inferir, com o máximo de certeza, a causa subjacente às informações sensoriais que são reproduzidas incrementalmente durante a compreensão da língua. Para inferir essa causa, o falante faz uso de um conjunto de representações organizadas hierarquicamente que podem melhor explicar as propriedades estatísticas das entradas lexicais. Em geral, são relevantes para esse modelo interno tanto as características do ambiente comunicativo mais amplo, quanto previsões probabilísticas que são propagadas para níveis mais baixos de representação. No topo dessa hierarquia, há um conjunto de hipóteses, de diferentes graus de crença, sobre a causa subjacente à entrada sensorial encontrada. A intensidade com que as crenças são propagadas dos níveis mais elevados para os níveis mais baixos depende da certeza que as mantém. Com isso, quando novas evidências *bottom-up* tornam-se disponíveis, o falante é informado se suas previsões probabilísticas serão mantidas. Quando ocorrem erros de previsão, as evidências são usadas para atualizar, por meio da inferência bayesiana, as crenças de alto nível acerca da causa subjacente da entrada sensorial.

Neste estudo, investigamos se verbos com forte tendência a subselecionar argumentos internos específicos possuem um comportamento temático divergente de verbos que possuem uma fraca tendência a subselecionar argumentos internos específicos. Observamos se esse comportamento temático divergente poderia estar relacionado ao nível de previsibilidade dos argumentos internos. Por fim, verificamos se, em contextos mínimos como o de uma tarefa de decisão lexical, as diferenças entre o comportamento temático dos dois grupos de verbos permaneceriam da mesma forma.

Os resultados de um julgamento de ajuste temático indicaram que os participantes consideraram mais habituais os eventos expressos por verbos de forte subseleção e argumentos internos mais previsíveis (*e.g.*: derramar – leite) do que os eventos expressos por verbos de fraca subseleção e argumentos internos mais previsíveis (*e.g.*: quebrar – perna). Os participantes também julgaram menos habituais os eventos na condição em que o verbo era de forte subseleção e o argumento interno era implausível

(e.g.: derramar – pato) do que na condição em que o verbo era de fraca subseleção e o argumento interno era implausível (e.g.: quebrar – sangue). Também foi verificado que, nos verbos de fraca subseleção, os eventos em que ocorria um argumento interno mais previsível eram considerados mais habituais do que os eventos expressos por argumentos internos de diferentes previsibilidades (e.g.: pouco previsíveis; imprevisíveis, porém plausíveis; implausíveis). Tal fato foi considerado relevante, pois era esperado que verbos de fraca subseleção apresentassem julgamentos semelhantes quando os argumentos eram mais previsíveis e pouco previsíveis.

Esses resultados podem ser compreendidos como indicações de que as propriedades temáticas dos grupos de verbos diferenciam-se com relação à delimitação de um protótipo de papel temático. Contudo, essa diferenciação clara entre os dois grupos foi encontrada apenas entre argumentos mais previsíveis e argumentos implausíveis. É possível que, por estar restrito aos limites extremos da previsibilidade, tal efeito esteja primordialmente ligado aos eventos expressos por verbos de forte poder de subseleção de argumentos internos. Com relação aos verbos de fraca subseleção que exibem preferências por argumentos mais previsíveis, é possível que esse comportamento demonstre que tais verbos carregam o mínimo de preferências temáticas que se tornam mais evidentes em contextos reduzidos.

Após verificarmos que as diferenças mais significativas ocorreram nos julgamentos em que os argumentos são mais previsíveis e também quando os argumentos são implausíveis, foi proposta a realização de uma tarefa de decisão lexical para verificarmos se os padrões de comportamentos observados na tarefa de ajuste temático seriam mantidos em uma tarefa on-line. Também pretendíamos verificar se, diante de um contexto mínimo e com diferentes probabilidades de cloze, conforme a necessidade apontada por Huettig e Mani (2016), a antecipação de traços temáticos dos argumentos internos seria feita pelos verbos.

A primeira tarefa de decisão lexical foi baseada no trabalho de Ferretti, McRae e Hatherell (2001), em que um efeito de *priming* semântico foi verificado entre verbos e argumentos internos cuja relação temática, por meio de julgamento de ajuste temático, foi considerada forte. O trabalho dos pesquisadores mostrou, a partir de um SOA curto

(250ms), que o acesso a informações temáticas on-line ocorreria de maneira automática e extremamente rápida. No trabalho de Ferretti, McRae e Hatherell (2001), de maneira diferente de nosso trabalho, a tarefa realizada pelos participantes demandava julgamentos de animacidade, nos quais deveria ser decidido se a palavra alvo representava algo vivo ou não.

Não encontramos resultados semelhantes aos de Ferretti, McRae e Hatherell (2001). Não encontramos nenhum efeito de *priming* entre pares de verbos e argumentos mais previsíveis, seja entre verbos de forte subseleção (*e.g.*: derramar – leite), seja entre verbos de fraca subseleção (*e.g.*: quebrar – perna). Uma de nossas hipóteses para a ocorrência desses diferentes resultados se refere ao tipo de tarefa utilizada. Devido a características intrínsecas a nossos itens experimentais, o julgamento de animacidade não pode ser realizado. É possível que durante a utilização da tarefa de decisão lexical, que acessa informações de caráter ortográfico, fonológico e semântico, as informações semânticas demandem um SOA maior para serem acessadas, ao contrário do julgamento de animacidade que demanda, principalmente, informações de caráter semântico.

Dentre nossas hipóteses para as diferenças nos resultados, também abordamos o fato de que os julgamentos de ajuste temáticos foram utilizados apenas para confirmar a intuição dos pesquisadores a respeito de verbos que pareciam ter um papel temático bem definido. Critérios, como as tarefas de cloze, não foram utilizados para confirmar o comportamento de subseleção dos verbos, inclusive nem para conhecer as entidades mais esperadas na posição de argumento interno ou o número de entidades passíveis de ocupar a posição, o que indicaria uma tendência a possuir traços temáticos bem delimitados ou não. A falta de critérios bem definidos para a seleção dos verbos é relevante, pois, conforme verificamos, verbos de fraca subseleção – cujos resultados indicaram a existência de um papel temático mais mal delimitado – também exibem leves preferências por argumentos mais previsíveis, apesar da probabilidade de cloze baixa. Tal fato pode gerar a impressão errônea de que um verbo de fraca subseleção possa ter um papel temático bem delimitado. Consideramos que, embora esse fato não explique totalmente as diferenças nos resultados, ele expõe as diferenças entre o desenho experimental deste trabalho e o de Ferretti, McRae e Hatherell (2001)

Também foi apontada como possível razão de divergência entre os resultados a metodologia estatística empregada na análise dos dados. O trabalho de Ferretti, McRae e Hatherell (2001) foi realizado por meio da análise de variância, uma técnica que pode ser bastante problemática para TRs de natureza positivamente inclinada. Para Lo e Andrews (2015), nas análises tradicionais que utilizam média, como a ANOVA, questões relativas à falta de normalidade dos dados são geralmente ignoradas porque o método tem sido mostrado repetidamente como robusto em relação a violações da normalidade. Em alguns casos, os dados passam por transformações para que possam assumir a distribuição normal, o que também constitui um recurso problemático. Segundo Lo e Andrews (2015), o adequado seria o uso de um método estatístico que respeitasse a natureza dos dados da variável resposta, permitindo ao mesmo tempo que fosse satisfeito o critério matemático de normalização dos resíduos.

Em nosso trabalho, para evitar qualquer tipo de transformação nos dados, utilizamos modelos lineares generalizados de efeito misto (GLMMs). Como os GLMMs não assumem o padrão de distribuição Gaussiana, é possível que o pesquisador especifique a distribuição apropriada. Em nosso caso, a distribuição *Gamma* possibilitou uma aproximação adequada da distribuição observada em TRs. Além disso, os GLMMs permitiram que definíssemos exatamente a natureza da relação esperada entre a variável preditora e a resposta observada, o que não é possível quando os dados são transformados<sup>55</sup>. A *link function* define a natureza da relação entre a resposta observada e o construto ao qual os fatores fixos estão relacionados. No caso de nossos experimentos de decisão lexical, nenhuma função em especial parece ser requerida, uma vez que se assume que a resposta observada reflete diretamente o construto latente. Ou seja, em nossos experimentos, consideramos que o tempo gasto para fornecer a resposta na tarefa de decisão lexical é correspondente ao tempo necessário para o acesso lexical.

Por fim, conjecturamos que a ausência de efeitos de facilitação entre os pares de verbos e argumentos mais previsíveis deve-se ao curso temporal de processos estratégicos, como os previstos por Delaney-Busch *et al.* (2019), que indicam a necessidade de um curso temporal maior para que haja antecipação dos traços temáticos dos argumentos

<sup>55</sup> Para maiores detalhes, veja nota 38.

mais esperados de verbos de forte subseleção em tarefas que exploram o efeito de *priming* semântico. Tal hipótese possui alguma consonância com as propostas de Huettig e Guerra (2015) e Chow *et al.* (2016), para os quais um certo patamar de tempo necessita ser atingido para que haja antecipação de informações linguísticas. É possível que a ativação de traços temáticos seja relacionada a conhecimentos probabilísticos sobre o mundo real capazes de interagir com informações do contexto ambiental imediato – neste trabalho, o contexto experimental da tarefa de decisão lexical – mediando a previsão de informações semânticas. Por acreditarmos que esses processos demandam um curso temporal um pouco maior do que 250ms, realizamos um segundo experimento de decisão lexical baseado na proposta de Delaney-Busch *et al.* (2019) a respeito dos processos estratégicos de antecipação responsáveis por produzir os efeitos de *priming* semântico. Nessa segunda tarefa de decisão lexical, a única modificação realizada em relação ao primeiro experimento foi a implementação de um SOA de 600ms.

Para Delaney-Busch *et al.* (2019), o processo de antecipação durante a tarefa de *priming* semântico envolve a adaptação ao ambiente experimental. Isso ocorre pelo fato de a meta principal do sujeito de uma comunicação ser a identificação de uma causa latente que melhor explique as estatísticas de ocorrência das entradas linguísticas. Sob essa perspectiva, os autores propõem que os efeitos de *priming* semântico são maiores em tarefas que possuem um maior número de pares de palavras relacionadas semanticamente, do que em tarefas cujo número de palavras relacionadas semanticamente é menor. Isso ocorre porque, nas tarefas cujo número de pares de palavras relacionadas é maior, os participantes são mais propensos a usar a palavra *prime* para gerar previsões léxico-semânticas mais fortes sobre a palavra alvo. Em tarefas cuja proporção de pares relacionados é menor, os participantes tendem a subestimar as informações probabilísticas fornecidas pela palavra *prime*. Essa perspectiva encontra-se em consonância com o modelo de compreensão proposto por Kuperberg e Jaeger (2016), segundo o qual o falante procura inferir a causa subjacente às informações sensoriais que são reproduzidas incrementalmente durante a compreensão da linguagem.

Os resultados da segunda tarefa de decisão lexical indicaram que pares de verbos de forte subseleção e argumentos internos relacionados/mais previsíveis desencadeiam tempos de reação (TRs) significativamente menores na tarefa de decisão lexical, quando comparados a pares formados por verbos de forte subseleção e argumentos internos não relacionados/implausíveis. Esses resultados podem ser relacionados, em certa medida, a Huettig e Guerra (2015) e Chow *et al.* (2016), cujos resultados indicam que o processamento preditivo precisa atingir um certo limiar de tempo antes que uma ação comportamentalmente observável seja iniciada. Em nosso trabalho, foi necessário um SOA de 600ms para que o efeito de facilitação proveniente dos verbos de forte subseleção fosse identificado.

Tais resultados também sugerem que a força das relações temáticas entre verbos de forte subseleção e argumentos mais esperados foi a responsável pelo efeito de facilitação encontrado exclusivamente nos verbos de forte subseleção. Em nosso experimento, dentre todos os pares apresentados na tarefa, apenas 25% possuía verbos de forte subseleção. Dentre esses 25%, apenas 12,5% ocorriam com argumentos relacionados/mais esperados. É possível que, durante a atuação de mecanismos de antecipação, apesar das informações probabilísticas presentes no ambiente experimental não incentivarem o uso das informações da palavra *prime* na maior parte dos itens, a força das associações temáticas dos verbos de forte subseleção tenha conseguido sobressair-se.

Porém, tendo em vista as limitações técnicas da tarefa de decisão lexical, não descartamos a possibilidade de efeitos de facilitação ocorrerem entre verbos de fraca subseleção e argumentos internos mais esperados, conforme a tendência apontada no julgamento de ajuste temático. A utilização de técnicas mais refinadas, como o rastreamento ocular ou um experimento de potenciais evocados, proporcionaria condições de apurar tais questões mais a fundo. Um experimento de decisão lexical, em que durante a tarefa fosse manipulada a quantidade de estímulos semanticamente relacionados também poderia fornecer informações relevantes sobre o funcionamento dos processos estratégicos de previsão relativos à mediação entre contexto experimental e informações temáticas provindas dos verbos.

Segundo McRae *et al.* (2005), os verbos possuem um papel primordial ao ativarem a estrutura de eventos das sentenças. Como portadores de funções predicativas, exibem restrições regulares e sistemáticas sobre seus argumentos. A informação disponível a partir de verbos é notavelmente refinada tanto estruturalmente quanto semanticamente, e essa informação interage de formas complexas com outras informações. Diversos trabalhos já demonstraram como a atribuição de papéis temáticos envolve uma complexa interação entre os verbos e as informações derivadas de conhecimentos de mundo (ALTMANN; KAMIDE, 1999; MCRAE; FERRETTI; AMYOTE, 1997). Dificilmente seria possível codificar a natureza refinada das expectativas por informações temáticas específicas a partir de propostas simbólicas tradicionais nas quais a representação inclui apenas um conjunto finito de papéis temáticos gerais, como as que utilizam conceitos limitados de agente, paciente, entre outros (FILLMORE, 1968; GRUBER, 1965).

Em vista da discussão que se estendeu ao longo deste capítulo, consideramos que nossos resultados evidenciam que os conhecimentos gerais sobre eventos do mundo real relacionados aos verbos permitem que, em contextos mínimos, os sujeitos gerem expectativas sobre conceitos que futuramente ocorrerão no fluxo da linguagem – no caso, traços semânticos pertencentes a argumentos internos frequentemente relacionados aos verbos. Acreditamos que o efeito de previsibilidade identificado em nossos resultados indica que o contexto experimental não conseguiu justapor-se às relações locais entre verbos de forte subseleção e argumentos internos muito esperados. Esses resultados são passíveis de serem interpretados como evidências de que existem relações de subseleção suficientemente fortes para acessarem conhecimentos mais refinados sobre eventos do mundo real.

## **6.1 Considerações finais**

Julgamos que os resultados encontrados nesta pesquisa podem contribuir para o debate a respeito das características do processo de antecipação de traços semânticos no processamento linguístico. Ademais, consideramos que o curso temporal demandado pelo processo de antecipação de informações de nível semântico ainda é pouco conhecido. Por meio da tarefa de decisão lexical, colaboramos para a compreensão do

desenvolvimento temporal da antecipação de informações semânticas de argumentos verbais muito esperados, em contextos mínimos. Por fim, enfatizamos a relevância da obtenção de informações sobre antecipação por meio de diferentes tipos de técnicas experimentais, o que permitiria o delineamento de um quadro mais amplo do fenômeno.

A confirmação das análises que propusemos nesta tese está sujeita à realização de novos experimentos. Para futuras investigações, consideramos essencial a realização de novos experimentos que utilizem rastreamento ocular ou a técnica de potenciais evocados. A utilização de técnicas diferentes poderá proporcionar maior conhecimento sobre os processos cognitivos evidenciados na tarefa de decisão lexical, além de permitir que o acesso a conhecimentos sobre eventos do mundo real seja mais bem compreendido. Ademais, também consideramos de grande importância a realização de uma tarefa de decisão lexical em que verbos de fraca subseleção sejam expostos em blocos maiores de itens de forte relacionamento semântico. A realização de tal experimento poderia fornecer informações mais detalhadas sobre o relacionamento dos verbos de fraca subseleção e argumentos muito esperados.

Contudo, embora nossos resultados ainda estejam sujeitos à realização de novos experimentos, acreditamos ser essencial enfatizar a importância das indicações identificadas em nossas tarefas. A literatura sobre antecipação ainda possui diversos debates (c.f. HUETTING; MANI, 2016; HUETTING, 2015) acerca da relevância da força contextual necessária para o estabelecimento de processos de antecipação de informações de diferentes níveis representacionais. Nossos resultados indicaram que contextos simples, formados apenas por verbos com características temáticas específicas, são capazes de facilitar o processamento de traços semânticos de argumentos internos muito prováveis. Além disso, apresentamos indicações de que o acesso antecipatório a informações temáticas, relacionadas ao conhecimento de mundo, parece demandar um curso temporal mais prolongando, diferentemente do que indica outro estudo (FERRETI; MCRAE; HATHERELL, 2001). Por fim, também julgamos essencial enfatizar a relevância de diversificação da aplicação de técnicas experimentais nos estudos sobre antecipação. Para nós, a utilização de um variado número de técnicas experimentais, que vão além do uso de potenciais evocados, auxilia um melhor



delineamento do fenômeno da antecipação, além de contribuir para o conhecimento de características ainda pouco conhecidas do fenômeno.

Dessa forma, diante da ausência de estudos que investiguem, em Português Brasileiro, a utilização de conhecimentos de eventos durante o processamento antecipatório da linguagem, esperamos ter cooperado para o avanço de discussões teóricas e metodológicas sobre o tema, além de termos contribuído para o desenvolvimento da literatura psicolinguística brasileira. Acreditamos que a existência de várias questões para investigações futuras é indicativa de que ajudamos a ampliar um tema de pesquisas ainda tão incipiente. Com isso, desejamos que nossos resultados incentivem outros pesquisadores a realizarem investigações sobre processamento que privilegiem abordagens mais amplas, cujo tratamento ultrapasse os conhecimentos considerados estritamente linguísticos.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALTMANN, G. T.; KAMIDE, Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, v. 73, n. 3, p. 247–264, 17 dez. 1999.

ANTOS, Stephen J. Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human perception and performance*, v. 5, n. 3, p. 527, 1979.

ANDERSON, N. J. L2 strategy research. In: HINKEL, E. (Org.). *Handbook of research in second language teaching and learning*. Mahwah: L. Erlbaum Associates, 2005.

ASHBY, Jane; RAYNER, Keith; CLIFTON, Charles. Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, v. 58, n. 6, p. 1065–1086, 2005.

ANTOS, Stephen J. Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human perception and performance*, v. 5, n. 3, p. 527, 1979.

BALOTA, David A.; POLLATSEK, Alexander; RAYNER, Keith. The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive psychology*, v. 17, n. 3, p. 364–390, 1985.

BAAYEN, R.H.; DAVIDSON, D.J.; BATES, D.M. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, v. 59, n. 4, p. 390–412, 2008.

BATES, Elizabeth; MACWHINNEY, Brian. Competition, variation, and language learning. In: MACWHINNEY, Brian (Org.). *Mechanisms of language acquisition*. Hillsdale: Erlbaum, 1987. p. 157–193.

BECKER, Curtis A. Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & cognition*, v. 8, n. 6, p. 493–512, 1980.

BENJAMINI, Yoav; HOCHBERG, Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, p. 289–300, 1995.

BERKUM, Jos JA Van; HAGOORT, Peter; BROWN, Colin M. Semantic integration in

sentences and discourse: Evidence from the N400. *Journal of cognitive neuroscience*, v. 11, n. 6, p. 657–671, 1999.

BICKNELL, Klinton; ELMAN, Jeffrey L.; HARE, Mary; MCRAE, Ken; KUTAS, Marta. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, v. 63, n. 4, p. 489–505, nov. 2010.

BORNKESSEL, Ina; SCHLESEWSKY, Matthias. The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, v. 113, n. 4, p. 787, 2006.

BORNKESSEL-SCHLESEWSKY, Ina; SCHLESEWSKY, Matthias. The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Language and Linguistics Compass*, v. 3, n. 1, p. 19–58, 2009.

BROTHERS, Trevor; SWAAB, Tamara Y.; TRAXLER, Matthew J. Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, v. 136, p. 135–149, 2015.

BROWN, Colin; HAGOORT, Peter. The processing nature of the N400: Evidence from masked priming. *Journal of cognitive neuroscience*, v. 5, n. 1, p. 34–44, 1993.

BUBIC, Andreja; VON CRAMON, D. Yves; SCHUBOTZ, Ricarda I. Prediction, Cognition and the Brain. *Frontiers in Human Neuroscience*, v. 4, 22 mar. 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904053/>>. Acesso em 18 set. 2019.

BUTZ, Martin V. Anticipation for learning, cognition and education. *On the Horizon*, v. 12, n. 3, p. 111–116, set. 2004.

CAMBLIN, C. Christine; GORDON, Peter C.; SWAAB, Tamara Y. The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, v. 56, n. 1, p. 103–128, 2007.

CANÇADO, Márcia. O papel do léxico em uma teoria dos papéis temáticos. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, v. 16, n. 2, p. 297–321, 2000.

CANÇADO, Márcia; AMARAL, Luana. *Introdução à semântica lexical: papéis temáticos, aspecto lexical e decomposição de predicados*. Petrópolis: Editora Vozes, 2017.

CHOMSKY, Noam. *Aspects of the Theory of Syntax*. Cambridge: MIT Press, 1965.

CHOMSKY, Noam. *Lectures on government and binding: The Pisa lectures*. Berlin: Walter de Gruyter, 1981.

CHOW, Wing-Yee; MOMMA, Shota; SMITH, Cybelle; LAU, Ellen F.; PHILLIPS, Colin. Prediction as memory retrieval: timing and mechanisms. *Language, Cognition and Neuroscience*, v. 31, n. 5, p. 617–627, 27 maio 2016.

CHRISTENSEN, Rune Haubo B. *A Tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal Package*. [S.l.]: The Comprehensive R Archive Network, 2015.

CLARK, Andy. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, v. 36, n. 3, p. 181–204, 2013.

COLLINS, Allan M; LOFTUS, Elizabeth F. A spreading-activation theory of semantic processing. *Psychological review*, v. 82, n. 6, p. 407, 1975.

COLLINS, Allan M.; QUILLIAN, M. Ross. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, v. 8, n. 2, p. 240–247, abr. 1969.

CROCKER, Matthew W; BRANTS, Thorsten. Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, v. 29, n. 6, p. 647–669, 2000.

DELANEY-BUSCH, Nathaniel; MORGAN, E.; LAU, E.; KUPERBERG, G. R. Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, v. 187, p. 10–20, jun. 2019.

DELONG, Katherine A.; URBACH, Thomas P.; GROPE, David M.; KUTAS, Marta. Overlapping dual ERP responses to low cloze probability sentence continuations: Dual ERPs to low probability sentence continuations. *Psychophysiology*, v. 48, n. 9, p. 1203–1207, set. 2011.

DE GROOT, Annette MB. Primed lexical decision: Combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target. *The Quarterly Journal of Experimental Psychology*, v. 36, n. 2, p. 253–280, 1984.

DELONG, Katherine A; URBACH, Thomas P; KUTAS, Marta. Concerns with Nieuwland et al. multi-lab replication attempt. *[S.l.: s.n.]*, 2017.

DELONG, Katherine A; URBACH, Thomas P; KUTAS, Marta. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, v. 8, n. 8, p. 1117–1121, ago. 2005.

DEY, Anind K. Understanding and Using Context. *Personal and Ubiquitous Computing*, v. 5, n. 1, p. 4–7, 28 fev. 2001.

DIKKER, Suzanne; PYLKKANEN, Liina. Before the N400: Effects of lexical–semantic violations in visual cortex. *Brain and Language*, v. 118, n. 1–2, p. 23–28, 2011.

DOWTY, David. Thematic proto-roles and argument selection. *Language*, v. 67, n. 3, p. 547–619, 1991.

EHRlich, Susan F.; RAYNER, Keith. Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, v. 20, n. 6, p. 641–655, 1981.

ELMAN, Jeffrey L. Finding structure in time. *Cognitive Science*, 14, p. 179–211. 1990.

ELMAN, Jeffrey L. On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science*, v. 33, n. 4, p. 547–582, jun. 2009.

FEDERMEIER, Kara D. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, v. 44, n. 4, p. 491–505, jul. 2007.

FEDERMEIER, Kara D.; KUTAS, Marta. A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, v. 41, n. 4, p. 469–495, nov. 1999.

FERRETTI, Todd R.; MCRAE, Ken; HATHERELL, Andrea. Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language*, v. 44, n. 4, p. 516–547, maio 2001.

FILLMORE, Charles J. The case for case. In: BACH, E.; HARMS, R. T. (Org.). *Universals in Linguistic Theory*. London: Holt, Rinehart and Winston, 1967. p. 1-88.

FINE, Alex B.; JAEGER, Florian T.; FARMER, Thomas A.; QIAN, Ting. Rapid Expectation Adaptation during Syntactic Comprehension. *PLoS ONE*, v. 8, n. 10, p. e77661, 30 out. 2013.

GARNSEY, Susan M.; PEARLMUTTER, Neal J.; MYERS, Elizabeth; LOTOCKY, Melanie A. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, v. 37, n. 1, p. 58–93, 1997.

GODOY, Mahayana Cristina; VIEIRA, Cláudia Brandão; CUNHA LIMA, Maria Luiza. Antecipação e processamento linguístico: uma revisão das principais evidências a favor da previsão. *Signo*, v. 41, n. 71, p. 63, 17 out. 2016.

GOLDBERG, Adele E. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, v. 7, n. 5, p. 219–224, maio 2003.

GONÇALVES, Jaciane Coelho. *Influência do número de repetições na identificação de genes diferencialmente expressos em experimentos de RNA-SEQ*. Tese de doutorado. Universidade Federal de Viçosa, 2013.

GRUBER, Jeffrey Steven. *Studies in lexical relations*. Tese de doutorado. MIT, 1965.

HAGOORT, Peter. MUC (memory, unification, control): A model on the neurobiology of language beyond single word processing. In: HICKOK, Gregory; SMALL, Steven L. (Org.). *Neurobiology of language*. Cambridge: Elsevier, 2015. p. 339–347.

HARE, Mary; JONES, Michael; THOMSON, Caroline; KELLY, Sarah; MCRAE; Ken. Activating event knowledge. *Cognition*, v. 111, n. 2, p. 151–167, maio 2009.

HALE, John. A probabilistic Earley parser as a psycholinguistic model. In: NAACL, 2., 2001, Pittsburg. Anais [...]. Pittsburg: Association for Computational Linguistics, 2001. p. 1–8.

HARPE, Spencer E. How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, v. 7, n. 6, p. 836–850, 1 nov. 2015.

HINO, Yasushi; LUPKER, Stephen J.; PEXMAN, Penny M. Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between

orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 28, n. 4, p. 686, 2002.

HINTZMAN, Douglas L. "Schema abstraction" in a multiple-trace memory model. *Psychological review*, v. 93, n. 4, p. 411, 1986.

HUETTIG, Falk. Four central questions about prediction in language processing. *Brain Research*, v. 1626, p. 118–135, nov. 2015.

HUETTIG, Falk; GUERRA, Ernesto. Testing the limits of prediction in language processing: Prediction occurs but far from always. In: ANNUAL CONFERENCE ON ARCHITECTURES AND MECHANISMS FOR LANGUAGE PROCESSING (AMLAP), 21., 2015, Valetta. Anais [...]. Valetta: University of Malta, 2015.

HUETTIG, Falk; MANI, Nivedita. Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, v. 31, n. 1, p. 19–31, 2 jan. 2016.

JACKENDOFF, Ray. *Foundations of language: brain, meaning, grammar, evolution*. Oxford; New York: Oxford University Press, 2002.

JACKENDOFF, Ray. *Semantic interpretation in generative grammar*. Cambridge: MIT Press, 1972.

JACKENDOFF, Ray. *Foundations of language: brain, meaning, grammar, evolution*. Oxford; New York: Oxford University Press, 2002.

JACKENDOFF, Ray. The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry*, v. 18, n. 3, p. 369–411, 1987.

JURAFSKY, Daniel. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive science*, v. 20, n. 2, p. 137–194, 1996.

KAMIDE, Yuki; ALTMANN, Gerry T.M; HAYWOOD, Sarah L. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, v. 49, n. 1, p. 133–156, jul. 2003.

KATZ, Jerrold J.; FODOR, Jerry A. The structure of a semantic theory. *Language*, v. 39, n. 2, p. 170–210, 1963.

KUPERBERG, Gina R. Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, p. 1–15, 2016.

KUPERBERG, Gina R.; JAEGER, T. Florian. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, v. 31, n. 1, p. 32–59, 2 jan. 2016.

KUPERBERG, Gina R. The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In: MILLER, Brett; CUTTING, Laurie E.; MCCARDLE, Peggy (Org.). *Unraveling reading comprehension: behavioral, neurobiological, and genetic components*. Baltimore: Paul H. Brookes Publishing Co., 2013. p. 176–192.

KUTAS, Marta; FEDERMEIER, Kara D. Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, v. 4, n. 12, p. 463–470, 2000.

KUTAS, Marta; HILLYARD, Steven A. Brain potentials during reading reflect word expectancy and semantic association. *Nature*, v. 307, n. 5947, p. 161, 1984.

KUTAS, Marta; HILLYARD, Steven A. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, v. 207, n. 4427, p. 203–205, 1980.

KUTAS, Marta; VAN PETTEN, Cyma K. Psycholinguistics electrified: Event-related potential investigations. In: GERNSBACHER, M. A. (Org.). *Handbook of psycholinguistics*. Cambridge: Academic Press, 1994. p. 83–143.

KUTAS, Marta; VAN PETTEN, Cyma K.; KLUENDER, Robert. Psycholinguistics electrified II (1994–2005). In: TRAXLER, M.; GERNSBACHER, M. A. (Org.). *Handbook of Psycholinguistics*. 2 ed. Cambridge: Elsevier, 2006. p. 659–724.

LAU, Ellen; NAMYST, Anna; FOGEL, Allison; DELGADO, Tania. A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabra: Psychology*, v. 2, n. 1, p. 1–19, 2016.

LAU, Ellen F.; HOLCOMB, Phillip J.; KUPERBERG, Gina R. Dissociating N400 effects of prediction from association in single-word contexts. *Journal of cognitive neuroscience*, v. 25, n. 3, p. 484–502, 2013.



LAU, Ellen F.; PHILLIPS, Colin; POEPPPEL, David. A cortical network for semantics: (de) constructing the N400. *Nature Reviews Neuroscience*, v. 9, n. 12, p. 920, 2008.

LAU, Ellen F.; WEBER, K.; GRAMFORT, A.; HÄMÄLÄINEN, M. S.; KUPERBERG, G. R. Spatiotemporal signatures of lexical-semantic prediction. *Cerebral Cortex*, p. bhu219, 2014.

LAU, Ellen F. The predictive nature of language comprehension. Tese de Doutorado. College Park: University of Maryland, 2009. Disponível em: <<http://drum.lib.umd.edu/handle/1903/9599>>. Acesso em: 4 nov. 2016.

LENTH, Russell V. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, v. 69, n. 1, p. 1-33, 2016. Disponível em: <<http://www.jstatsoft.org/v69/i01/>>. Acesso em: 14 ago. 2018.

LEVY, Roger. Expectation-based syntactic comprehension. *Cognition*, v. 106, n. 3, p. 1126-1177, 2008.

LEWIS, Richard L. Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, v. 29, n. 2, p. 241-248, 2000.

LIMA, Maria Cecília Pires Barbosa. A gramática dos casos e o "dativo". *ALFA: Revista de Linguística*, v. 26, 1982.

LO, Steson; ANDREWS, Sally. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, v. 6, p. 1-16, 7 ago. 2015. Disponível em: <<http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.01171/abstract>>. Acesso em: 8 mar. 2019.

LUKE, Steven G.; CHRISTIANSON, Kiel. Limits on lexical prediction during reading. *Cognitive Psychology*, v. 88, p. 22-60, 2016.

MACWHINNEY, Brian; BATES, Elizabeth; KLIEGL, Reinhold. Cue validity and sentence interpretation in English, German, and Italian. *Journal of verbal learning and verbal behavior*, v. 23, n. 2, p. 127-150, 1984.

MATSUKI, Kazunaga; CHOW, Tracy; HARE, Mary; ELMAN, Jeffrey L.; SCHEEPERS, Christoph; MCRAE, Ken. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 37, n. 4, p. 913–934, 2011.

MATSUMOTO, Yukihiro. *Ruby Programming Language*. 2013 [1995]. Disponível em: <<http://www.ruby-lang.org>>. Acesso em: 18 set. 2019.

MCNAMARA, Timothy P. Priming and constraints it places on theories of memory and retrieval. *Psychological Review*, v. 99, n. 4, p. 650, 1992.

MCRAE, Ken; FERRETTI, Todd R.; AMYOTE, Liane. Thematic Roles as Verb-specific Concepts. *Language and Cognitive Processes*, v. 12, n. 2–3, p. 137–176, 1 mar. 1997.

MCRAE, Ken; HARE, Mary; ELMAN, Jeffrey L.; FERRETTI, Todd R.; A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, v. 33, n. 7, p. 1174–1184, 2005.

MCRAE, Ken; JONES, M. N. Semantic memory. In: REISBERG, Daniel (Org.). *The Oxford handbook of cognitive psychology*. New York: Oxford University Press, 2013. p. 206–219.

MCRAE, Ken; MATSUKI, Kazunaga. People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, v. 3, n. 6, p. 1417–1429, nov. 2009.

MISHRA, Ramesh K.; SINGH, Niharika; PANDEY, Aparna; HUETTIG, Falk. Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research*, v. 5, n. 1, p. 1-10, 2012.

MITCHELL, Don C; HOLMES, Virginia M. The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language*, v. 24, n. 5, p. 542–559, 1985.

NIEUWLAND, Mante S.; BARR, Dale J.; BARTOLOZZI, Federica; BUSCH-MORENO, Simon; DARLEY, Emily; DONALDSON, David I.; FERGUSON, Heather J.; FU, Xiao; HEYSELAAR, Evelien; HUETTIG, Falk; HUSBAND, E. Matthew; ITO, Aine; KAZANINA, Nina; KOGAN, Vita; KOHÚT, Zdenko; KULAKOVA, Eugenia; MÉZIÈRE, Diane; POLITZER-AHLES, Stephen; ROUSSELET, Guillaume; RUESCHEMEYER, Shirley-Ann; SEGAERT, Katrien; TUOMAINEN, Jyrki; WOLFSTHURN, Sarah Von Grebmer Zu. Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale

study using brain potentials. *bioRxiv*, p. 267815, 2018.

NIEUWLAND, Mante S.; POLITZER-AHLES, Stephen; HEYSELAAR, Evelien; SEGAERT, Katrien; DARLEY, Emily; KAZANINA, Nina; WOLFSTHURN, Sarah Von Grebmer Zu; BARTOLOZZI, Federica; KOGAN, Vita; ITO, Aine; MÉZIÈRE, Diane; BARR, Dale J.; ROUSSELET, Guillaume; FERGUSON, Heather J.; BUSCH-MORENO, Simon; FU, Xiao; TUOMAINEN, Jyrki; KULAKOVA, Eugenia; HUSBAND, E. Matthew; DONALDSON, David I.; KOHÚT, Zdenko; RUESCHEMEYER, Shirley-Ann; HUETTIG, Falk. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, v. 7, p. 1-24, 3 abr. 2018.

NIEUWLAND, Mante S.; VAN BERKUM, Jos JA. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience*, v. 18, n. 7, p. 1098–1111, 2006.

OLLER JR, John W. Cloze tests of second language proficiency and what they measure 1. *Language learning*, v. 23, n. 1, p. 105-118, 1973.

OTTEN, Marte; NIEUWLAND, Mante S; VAN BERKUM, Jos JA. Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, v. 8, n. 1, p. 89, 2007.

OTTEN, Marte; VAN BERKUM, Jos J. A. Discourse-based word anticipation during language processing: prediction or priming? *Discourse Processes*, v. 45, n. 6, p. 464–496, 11 nov. 2008.

PACZYNSKI, Martin; KUPERBERG, Gina R. Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, v. 67, n. 4, p. 426–448, nov. 2012.

PEIRCE, Jonathan; GRAY, Jeremy R.; SIMPSON, Sol; MACASKILL, Michael; HÖCHENBERGER, Richard; SOGO, Hiroyuki; KASTMAN, Erik; LINDELØV, Jonas Kristoffer. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, v. 51, n. 1, p. 195-203, 7 fev. 2019. Disponível em: <<http://link.springer.com/10.3758/s13428-018-01193-y>>. Acesso em: 8 mar. 2019.

RAYNER, Keith; WELL, Arnold D. Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, v. 3, n. 4, p. 504–509, 1996.

ROLAND, Douglas; YUN, Hongoak; KOENIG, Jean-Pierre; MAUNER, Gail. Semantic similarity, predictability, and models of sentence processing. *Cognition*, v. 122, n. 3, p. 267–279, mar. 2012.

ROSEN, Robert. *Anticipatory systems: philosophical, mathematical and methodological foundations*. Oxford: Pergamon Press, 1985.

SAYEED, Asad; GREENBERG, Clayton; DEMBERG, Vera. Thematic fit evaluation: an aspect of selectional preferences. In: WORKSHOP ON EVALUATING VECTOR SPACE REPRESENTATIONS FOR NLP, 1., 2016, Berlin. *Anais [...]*. Berlin: Association for Computational Linguistics, 2016. p. 99–105.

SCHLESINGER, Izchak M. *Cognitive space and linguistic case: Semantic and syntactic categories in English*. Cambridge: Cambridge University Press, 2006.

SEIDENBERG, Mark S.; WATERS, Gloria S.; SANDERS, Michael; LANGER, Pearl. Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, v. 12, n. 4, p. 315–328, 1984.

SHAOUL, Cyrus; BAAYEN, Harald; WESTBURY, Chris. N-gram probability effects in a cloze task. *The Mental Lexicon*, v. 9, p. 1-43, 2014.

SMITH, Nathaniel J.; LEVY, Roger. Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In: ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY, 33., 2011, Austin. *Anais [...]*. Austin: Cognitive Science Society, 2011. p. 1637–1642. Disponível em: <<https://vorpus.org/papers/smith-levy-2011-cogsci.pdf>>. Acesso em: 4 nov. 2016.

SMITH, Nathaniel J.; LEVY, Roger. The effect of word predictability on reading time is logarithmic. *Cognition*, v. 128, n. 3, p. 302–319, set. 2013.

SPEELMAN, Craig P.; MCGANN, Marek. How Mean is the Mean? *Frontiers in Psychology*, v. 4, p. 1-12, jul. 2013. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00451/abstract>>. Acesso em: 11 mar. 2019.

STAUB, Adrian. The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, v. 9, n. 8, p. 311–327, 2015.

STAUB, Adrian; GRANT, Margaret; ASTHEIMER, Lori; ANDREW, Cohen. The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, v. 82, p. 1–17, jul. 2015.

STOWELL, Timothy Angus. *Origins of phrase structure*. Tese de doutorado. Cambridge: Massachusetts Institute of Technology, 1981.

VAN BERKUM, Jos J. A.; BROWN, Colin M.; ZWITSERLOOD, Pienie; KOOIJMAN, Valesca; HAGOORT, Peter. Anticipating upcoming words in discourse: evidence from ERPS and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 31, n. 3, p. 443–467, 2005.

TRAXLER, Matthew J. *Introduction to psycholinguistics: Understanding language science*. Hoboken: John Wiley & Sons, 2011.

TRUESWELL, John C.; TANENHAUS, Michael K.; KELLO, Christopher. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 19, n. 3, p. 528–553, 1993.

VAN HEUVEN, Walter J. B.; MANDERA, Pawel; KEULEERS, Emmanuel; BRYLSBAERT, Marc. SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, v. 67, n. 6, p. 1176–1190, 2014.

VAN PETTEN, Cyma; KUTAS, Marta. Interactions between sentence context and word frequency in event-related brain potentials. *Memory & cognition*, v. 18, n. 4, p. 380–393, 1990.

VAN PETTEN, Cyma; LUKE, Barbara J. Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, v. 83, n. 2, p. 176–190, fev. 2012.

WARREN, Tessa; MCCONNELL, Kerry. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic bulletin & review*, v. 14, n. 4, p. 770–775, 2007.

WEBER, Kirsten; LAU, Ellen F.; STILLERMAN, Benjamin; KUPERBERG, Gina R. The yin and the yang of prediction: an fMRI study of semantic predictive processing. *PLOS ONE*, v. 11, n. 3, p. e0148637, 24 mar. 2016.

WICHA, Nicole Y. Y.; MORENO, Eva M.; KUTAS, Marta. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, v. 16, n. 7, p. 1272–1288, set. 2004.

WLOTKO, Edward W.; FEDERMEIER, Kara D. Two sides of meaning: The scalp-recorded N400 reflects distinct contributions from the cerebral hemispheres. *Frontiers in psychology*, v. 4, p. 181, 2013.

WLOTKO, Edward W.; FEDERMEIER, Kara D.; KUTAS, Marta. To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, v. 27, n. 4, p. 975–988, 2012.

YAN, Shaorong; KUPERBERG, Gina R.; JAEGER, T. Florian. Prediction (or not) during language processing. A commentary on Nieuwland et al. (2017) and DeLong et al. (2005). *bioRxiv*, p. 143750, 30 maio 2017.

## APÊNDICE A

*Tabela 1 – Verbos selecionados para o banco de clozes e argumentos mais citados*

<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>	<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>
Eliminou	Sujeira	4.00	Impulsionou	Vendas	9.62
Categorizou	Livros	5.26	Inflamou	Causa	9.62
Invalidou	Documentos	5.56	Acusou	Irmão	9.80
Empregou	Amigo	5.77	Benzeu	Filha	9.80
Esculhambou	Tudo	5.77	Complementou	Frase	9.80
Inspirou	Irmão	5.88	Democratizou	Sistema	9.80
Detalhou	Trabalho	6.12	Doutorou	Se	9.80
Ensanguentou	Camisa	6.78	Camuflou	Carro	10.00
Deportou	Imigrante	7.14	Desvalorizou	Trabalho	10.00
Formulou	Pergunta	7.27	Ferrou	Amigo	10.00
Menosprezou	Amigo	7.27	Pulverizou	Casa	10.00
Atribulou	Vida	7.41	Desclassificou	Time	10.17
Cristianizou	Família	7.41	Enlatou	Sardinha	10.17
Imitou	Palhaço	7.41	Desintoxicou	Corpo	10.20
Seduziu	Mulher	7.41	Desocupou	Lugar	10.20
Insultou	Colegas	7.55	Desparafusou	Cama	10.20
Amparou	Amigo	7.69	Dopou	Amiga	10.20
Censurou	Filha	7.69	Engordurou	Comida	10.20
Cobiçou	Marido	7.69	Esculachou	Amigo	10.20
Instigou	NR	7.69	Espancou	Irmão	10.20
Internalizou	Dor	7.84	Estragou	Carro	10.20
Abordou	Estranho	8.00	Desenferrujou	Portão	10.34
Coagiu	Pai	8.00	Esfumaçou	Desenho	10.71
Esfaqueou	Ladrão	8.00	Desenrolou	Fio	14.55
Especificou	NR	8.00	Desprotegeu	Casa	10.91
Engoliu	Moeda	8.16	Inventou	Fórmula	10.91
Esquartejou	Porco	8.16	Destrinchou	Frango	11.11
Estressou	Se	8.16	Abateu	Galinha	11.54
Desidratou	Pele	8.33	Advertiu	Amiga	11.54
Divulgou	Notas	8.33	Alfinetou	Dedo	11.54
Estereotipou	Pessoas	8.33	Beliscou	Braço	11.54
Comercializou	Roupas	8.47	Boicotou	Evento	11.54
Apelidou	Amiga	8.62	Cadastrou	Se	11.54
Escoltou	Presidente	8.62	Carregou	Celular	11.54
Espiritualizou	Alma	8.70	Cobriu	Rosto	11.54
Causou	Acidente	8.77	Desconvidou	Amigo	11.54
Banalizou	Situação	8.93	Enlameou	Sapatos	11.54
Civilizou	Índios	9.09	Bombardeou	Cidade	11.76

<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>	<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>
Fabricou	Sapatos	9.09	Cancelou	Festa	11.76
Incentivou	Amigo	9.09	Consolou	Mãe	11.76
Internacionalizou	Empresa	9.43	Convenceu	Mãe	11.76
Assediou	Chefe	9.62	Emitiu	Nota	11.76
Criou	Jogo	9.62	Humilhou	Amigo	11.76
Evangelizou	Família	9.62	Inibiu	NR	11.76
Irritou	Irmão	11.76	Entediou	Se	14.29
Torturou	Gato	11.76	Imobilizou	Perna	14.55
Descabelou	Se	11.86	Reajustou	Relógio	14.55
Estabilizou	Situação	11.86	Teceu	Roupa	14.55
Concursou	NR	12.00	Desqualificou	Candidato	14.58
Matou	Fome	12.00	Documentou	Ocorrido	14.58
Salvou	Cachorro	12.00	Esvaziou	Caixa	14.58
Sapecou	NR	12.00	Inflacionou	Preço	14.89
Desempacotou	Presente	12.07	Decorou	Quarto	15.28
Esterilizou	Alicate	12.24	Abusou	Sorte	15.38
Desacompanhou	Mãe	12.50	Agarrou	Bola	15.38
Estraçalhou	Vidro	12.73	Alvejou	Roupas	15.38
Indicou	Caminho	12.73	Cruzou	Rua	15.38
Lambuzou	Cara	12.73	Enforcou	Si mesmo	15.38
Manipulou	Remédio	12.73	Instituiu	Lei	15.38
Espaçou	Texto	12.96	Ultrajou	Se	15.38
Molestou	Criança	13.04	Acidentou	Se	15.52
Induziu	NR	13.21	Cassou	Mandato	15.52
Perturbou	Vizinhos	13.21	Comentou	Sobre ...	15.52
Alcançou	Meta	13.46	Adulou	Filho(s)	15.69
Aprontou	NR	13.46	Baleou	Ladrão	15.69
Congelou	NR	13.46	Contratou	Pedreiro	15.69
Destratou	Mãe	13.46	Divertiu	NR	15.69
Enquadrou	Quadro	13.46	Autuou	Infrator	16.00
Enrolou	NR	13.46	Contemplou	Vista	16.00
Intoxicou	Gato	13.46	Desenrugou	Testa	16.00
Desorientou	Mãe	13.56	Desgastou	Sapato	16.00
Confortou	Amigo	13.73	Furou	Pneu	16.00
Constrangeu	Irmão	13.73	Rachou	NR	16.00
Enganou	Pai	13.73	Rifou	Bicicleta	16.00
Internou	Pai	13.73	Taxou	Produtos	16.00
Intimou	Réu	13.73	Descalçou	Sapatos	16.07
Custeou	Viagem	13.79	Enfileirou	Cadeiras	16.07
Contorceu	Corpo	14.00	Elaborou	Plano	16.33
Explorou	Mundo	14.00	Deletou	Arquivo	16.36



<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>	<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>
Guardou	Dinheiro	14.00	Desabrigou	Cachorro	16.36
Interligou	Pontos	14.00	Inferiorizou	Colega	16.36
Quebrou	Perna	14.00	Injetou	Remédio	16.36
Regulou	Temperatura	14.00	Mencionou	Nome	16.36
Colonizou	País	14.04	Perdoou	Namorado	16.36
Desequilíbrio	Se	14.04	Rimou	Poema	16.36
Descontrolou	Carro	14.29	Chateou	Se	16.67
Desrespeitou	Mãe	14.29	Eletrocutou	Gato	16.67
Engavetou	Papéis	14.29	Esburacou	Rua	16.67
Entediou	Se	14.29	Esquematizou	Plano	16.67
Imobilizou	Perna	14.55	Industrializou	Cidade	16.67
Reajustou	Relógio	14.55	Interrompeu	Conversa	16.67
Teceu	Roupa	14.55	Consertou	Carro	16.95
Desqualificou	Candidato	14.58	Contornou	Rua	16.95
Documentou	Ocorrido	14.58	Alimentou	Gato	17.31
Esvaziou	Caixa	14.58	Armou	Barraca	17.31
Inflacionou	Preço	14.89	Catou	Lixo	17.31
Decorou	Quarto	15.28	Chamuscou	Cabelo	17.31
Abusou	Sorte	15.38	Conduziu	Carro	17.31
Agarrou	Bola	15.38	Contabilizou	Dinheiro	17.31
Alvejou	Roupas	15.38	Curou	Doença	17.31
Cruzou	Rua	15.38	Difamou	Colega	17.31
Enforcou	Si mesmo	15.38	Diminuiu	Volume	17.31
Instituiu	Lei	15.38	Executou	Tarefa	17.31
Ultrajou	Se	15.38	Estruturou	Casa	17.54
Acidentou	Se	15.52	Empolgou	NR	17.65
Cassou	Mandato	15.52	Empossou	Cargo	17.65
Comentou	Sobre ...	15.52	Pajeou	Bebê	17.65
Adulou	Filho(s)	15.69	Recrutou	Soldado	17.65
Baleou	Ladrão	15.69	Asilou	Avó	17.78
Contratou	Pedreiro	15.69	Desencapou	Caderno	17.86
Divertiu	NR	15.69	Estrangulou	Gato	17.86
Autuou	Infrator	16.00	Engarrafou	Suco	18.00
Contemplou	Vista	16.00	Minimizou	Janela	18.00
Desenrugou	Testa	16.00	Namorou	NR	18.00
Desgastou	Sapato	16.00	Operou	Nariz	18.00
Furou	Pneu	16.00	Passou	Vestibular	18.00
Rachou	NR	16.00	Reconheceu	Amigo	18.00
Rifou	Bicicleta	16.00	Liderou	Grupo	18.18
Taxou	Produtos	16.00	Encapuzou	Cabeça	18.33
Descalçou	Sapatos	16.07	Despedaçou	Bolo	18.37

<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>	<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>
Enfileirou	Cadeiras	16.07	Equipou	Carro	18.37
Elaborou	Plano	16.33	Bajulou	Chefe	18.52
Deletou	Arquivo	16.36	Intensificou	Estudos	18.52
Desabrigou	Cachorro	16.36	Magoou	Namorado	18.52
Inferiorizou	Colega	16.36	Abençoou	Filho	18.64
Injetou	Remédio	16.36	Desenfeitou	Árvore	18.64
Mencionou	Nome	16.36	Detestou	Comida	18.75
Perdoou	Namorado	16.36	Dissolveu	Pó	18.75
Rimou	Poema	16.36	Entulhou	Lixo	18.75
Chateou	Se	16.67	Atualizou	Status	18.97
Eletrocutou	Gato	16.67	Desencorajou	Amigo	18.97
Esburacou	Rua	16.67	Achatou	Massa	19.23
Adorou	Presente	19.23	Atropelou	Cachorro	23.08
Adquiriu	Carro	19.23	Chocou	Ovo	23.08
Ativou	Conta	19.23	Dobrou	Roupa	23.08
Balançou	Cabeça	19.23	Derrotou	Inimigo	23.08
Cercou	Casa	19.23	Concluiu	Curso	23.53
Cremou	Pai	19.23	Falsificou	Documento	23.64
Desencaixotou	Livros	19.23	Grampeou	Telefone	23.64
Agradou	Mãe	19.61	Moderou	Bebida	23.64
Assombrou	Casa	20.00	Restituiu	NR	23.64
Desligou	Televisão	20.00	Saboreou	Sorvete	23.64
Encaixotou	Livros	20.00	Desempregou	Funcionários	23.72
Fortaleceu	Músculos	20.00	Desenroscou	Parafuso	23.73
Girou	NR	20.00	Resgatou	Gato	24.00
Machucou	Joelho	20.00	Desembalou	Presente	24.07
Despreocupou	Mãe	20.41	Desamassou	Papel	24.14
Deformou	Rosto	20.69	Elegeu	Presidente	24.49
Amansou	Cachorro	20.75	Encurtou	Saia	24.49
Materializou	Sonho	20.75	Esbanjou	Dinheiro	24.49
Drogou	Amigo	20.83	Assustou	Se	24.53
Descarçou	Azeitona	21.15	Fascinou	Se	24.53
Embalou	Presente	21.15	Amarrou	Cadarço	25.00
Encheu	Copo	21.15	Arrasou	(na)festa	25.00
Comoveu	Todos	21.57	Bloqueou	Celular	25.00
Editou	Texto	21.57	Desorganizou	Quarto	25.00
Esfumou	Olhos	21.57	Destampou	Pote	25.00
Projetou	Casa	21.57	Empoeirou	Casa	25.00
Esculpiu	Estátua	21.67	Loteou	Terreno	25.00
Condenou	Réu	21.82	Examinou	Paciente	25.45
Costurou	Roupa	21.82	Flagrou	Namorado	25.45

<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>	<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>
Enxaguou	Cabelo	21.82	Fraturou	Braço	25.45
Estendeu	Mão	21.82	Frequentou	Escola	25.45
Gelou	NR	21.82	Inocentou	Réu	25.45
Incrementou	Receita	21.82	Sublinhou	Texto	25.45
Rolou	Bola	21.82	Xerocou	Livro	25.49
Incensou	Casa	22.00	Enxugou	Lágrimas	26.00
Maquiou	Amiga	22.00	Guiou	Cego	26.00
Pilhou	NR	22.00	Melou	Dedos	26.00
Tombou	NR	22.00	Provocou	Briga	26.00
Desamarrotou	Camisa	22.22	Recebeu	Dinheiro	26.00
Burocratizou	Processo	22.45	Tocou	Violão	26.00
Endividou	Se	22.92	Acariciou	Gato	26.92
Escavou	Buraco	22.92	Desencadernou	Livro	26.92
Anulou	Questão	23.08	Adubou	Horta	27.27
Apertou	Botão	23.08	Fritou	Batata	27.27
Parcelou	Compra	27.27	Despejou	Água	34.00
Rabiscou	Caderno	27.27	Processou	Empresa	34.00
Refogou	Couve	27.27	Promoveu	Festa	34.00
Transcreveu	Texto	27.27	Espremeu	Laranja	34.55
Ofendeu	Amigo	27.45	Grifou	Texto	34.55
Lavou	Carro	28.00	Pavimentou	Rua	34.55
Liquidou	Dívida	28.00	Peneirou	Areia	34.55
Moeu	Carne	28.00	Capturou	Pokémon	34.62
Moldou	Escultura	28.00	Deprimiu	Se	34.62
Pressionou	Botão	28.00	Esgotou	Paciência	34.62
Mapeou	Cidade	28.57	Destravou	Porta	34.69
Colecionou	Selos	28.81	Empilhou	Livros	34.69
Desarrumou	Quarto	28.81	Empanou	Frango	35.19
Apontou	NR	28.85	Avisou	NR	35.29
Cimentou	Calçada	29.09	Fatiou	Bolo	35.29
Imprimiu	Trabalho	29.09	Marcou	Encontro	36.00
Esborrachou	Chão	29.17	Derreteu	Chocolate	36.36
Espatifou	Chão	29.17	Invadiu	Casa	36.36
Autografou	Livro	29.31	Alfabetizou	Crianças	36.54
Afrouxou	Calça(s)	29.41	Apimentou	Relação	36.54
Algemou	Ladrão	29.41	Arquivou	Documentos	36.54
Convidou	Amigos	29.41	Esboçou	Sorriso	36.73
Contundiu	Joelho	29.63	Castigou	Filho	37.04
Estatelou	Chão	30.36	Encerou	Chão	37.04
Encharcou	Roupa(s)	30.61	Aparou	Gramma	37.25
Apaixonou	Se	30.61	Digitou	Texto	37.25

<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>	<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>
Acendeu	Luz	30.91	Ensopou	Roupa	37.50
Descascou	Batata	30.91	Orgulhou	Se	37.74
Rondou	Casa	30.91	Descongelou	Frango	37.93
Anotou	Recado	31.37	Cortou	Cabelo	38.00
Especializou	Se	31.67	Espetou	Carne	38.00
Engessou	Braço	32.00	Incendiou	Casa	38.00
Ganhou	Na loteria	32.00	Torrou	Pão	38.00
Intrigou	Se	32.08	Manchou	Camisa	38.18
Assanhou	Se	32.61	Tosou	Cachorro	38.18
Aguentou	NR	32.69	Acionou	Polícia	38.46
Atingiu	Meta	32.69	Chutou	Bola	38.46
Castrou	Gato	32.69	Enfaixou	Braço	38.78
Decifrou	Código	32.73	Irrigou	Plantas	38.89
Demitiu	Funcionário	32.73	Iludiu	Namorado	39.22
Distraiu	Se	32.73	Encapou	Caderno	39.58
Estatizou	Empresa	32.73	Fatigou	Se	39.58
Desensaboou	Corpo	33.33	Apagou	Luz	40.00
Embaraçou	Cabelo	33.33	Demoliu	Casa	40.00
Perfurou	Parede	33.33	Emplacou	Carro	40.00
Ensaboou	Corpo	40.00	Decotou	Vestido	49.09
Fechou	Porta	40.00	Podou	Árvore	50.00
Saldou	Dívida	40.00	Cronometrou	Tempo	50.00
Salgou	Carne	40.00	Desembaralhou	Cartas	50.00
Depilou	Pernas	40.38	Inquietou	Se	50.00
Medicou	Paciente	40.74	Assou	Frango	50.91
Escreveu	Carta	41.18	Folheou	Livro	50.91
Subornou	Policial	41.18	Narrou	História	50.91
Engatou	Marcha	41.82	Acordou	Cedo	50.98
Estourou	Balão	41.82	Acatou	Ordem	51.92
Rebateu	Bola	41.82	Fumou	Cigarro	52.73
Resenhou	Livro	41.82	Repicou	Cabelo	52.73
Tostou	Pão	41.82	Autenticou	Documento	52.94
Nauseou	NR	42.00	Contentou	Se	53.19
Poupou	Dinheiro	42.00	Trançou	Porta	53.85
Azulejou	Banheiro	42.11	Pescou	Peixe(s)	72.73
Alugou	Apartamento	42.31	Desacelerou	Carro	54.24
Cantou	Música	42.31	Pinçou	Sobrancelha	54.55
Checou	E-mail	42.31	Urbanizou	Cidade	54.55
Esmurrou	Porta	42.86	Serrou	Madeira(s)	54.55
Afligi	Se	43.59	Infestou	Casa	54.90
Lustrou	Movel(éis)	43.63	Adoçou	Café	55.15

<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>	<b>Verbo</b>	<b>Arg. mais citado</b>	<b>P. Cloze</b>
Incendiou	Casa	43.64	Malhou	NR	56.00
Ninou	Bebê	43.64	Sovou	Pão	56.36
Engatilhou	Arma	43.75	Deleitou	Se	56.52
Admitiu	NR	44.23	Mimou	Filho	57.41
Chifrou	Namorado	44.23	Calçou	Sapato(s)	57.76
Esfarelou	Biscoito	44.90	Capinou	Lote	58.18
Revisou	Texto	45.10	Desengatou	Marcha	59.62
Revoltou	Se	45.10	Aguou	Plantas	60.00
Indignou	Se	45.28	Depositou	Dinheiro	60.00
Curtiu	Foto	45.45	Estacionou	Carro	60.00
Descoloriu	Cabelo	45.45	Dosou	Remédio(s)	61.22
Roou	Unhas	45.45	Sacou	Dinheiro	61.82
Rufou	Tambores	45.45	Cacheou	Cabelos	63.64
Bicou	Conversa	46.15	Compôs	Música	63.64
Emoldurou	Quadro	46.94	Dedetizou	Casa	63.64
Amassou	Papel(éis)	47.27	Estufou	Peito	63.64
Ensaio	Peça	47.27	Redigiu	Texto	63.64
Privatizou	Empresa	47.27	Trançou	Cabelo	63.64
Assalariou	Empregado	47.92	Desembrulhou	Presente(s)	64.40
Dedetizou	Casa	48.08	Embaçou	Vidro	65.31
Infernizou	Vida	48.15	Amolou	Faca	65.38
Estilhaçou	Vidro	48.28	Arrombou	Porta	65.38
Assinou	Contrato	49.02	Coou	Café	65.45
Mobiliou	Casa	67.27	Embaralhou	Cartas	78.18
Resolveu	Problema(s)	67.28	Conjugou	Verbo	78.84
Escovou	Dentes	67.35	Atravessou	Rua	80.00
Derramou	Leite	68.63	Asfaltou	Rua	80.39
Educou	Filhos	68.75	Afiou	Faca	80.39
Receitou	Remédio	69.09	Calibrou	Pneu	80.76
Abasteceu	Carro	69.09	Engraxou	Sapatos	81.25
Palitou	Dentes	69.09	Recitou	Poema(s)	81.82
Ferveu	Água	70.91	Confeitou	Bolo	82.35
Capotou	Carro	72.73	Quitou	Dívida(s)	83.63
Sancionou	Lei	74.55	Despenteou	Cabelo	83.67
Embrulhou	Presente(s)	74.55	Manobrou	Carro	85.45
Untou	Forma	74.55	Assoou	Nariz	87.50
Lixou	Unha(s)	76.36	Sonegou	Imposto(s)	90.91
Alisou	Cabelo(s)	76.36	Afivelou	Cinto	90.91
Reformou	Casa	76.47	Esubalhou	Olhos	92.86









<b>Escreveu</b>		<b>Adubou</b>		<b>Encharcou</b>		<b>Achatou</b>		<b>Resgatou</b>	
Carta	41.18	Horta	27.27	Roupa(s)	30.61	Massa	19.23	Gato	24
Livro	23.53	Terra	27.27	Blusa	8.16	Nariz	7.69	Cachorro	16
NR	9.8	Planta(s)	20	Chão	6.12	Bolo	5.77	Gatinho(s)	12
Artigo	5.88	Jardim	10.91	Pano(s)	6.12	Almofada	5.77	Filhote(s)	6
Poema	3.92	Canteiro	3.64	Piso	4.08	Roupas	3.85	Amigo	4
Bilhete	1.96	Mente	1.82	Cabelo	4.08	Mala	3.85	Criança	4
Email	1.96	Flores	1.82	Calça	2.04	Bola	3.85	Cachorrinho	2
Errado	1.96	Campo	1.82	Caderno	2.04	Latinha	3.85	Número	2
Nome	1.96	Quintal	1.82	Casaco	2.04	Madeira	1.92	Rodrigo	2
Ontem	1.96	Terreno	1.82	Cachorro	2.04	Cara	1.92	Pessoa	2
Quadro	1.96	Plantação	1.82	Rua	2.04	Brincos	1.92	Vestido	2
Tese	1.96			Olhos	2.04	Biscoitos	1.92	Seguro	2
Texto	1.96			Sangue	2.04	Lata	1.92	Cheque	2
				Sala	2.04	Chocolate	1.92	Dinheiro	2
				Pastéis	2.04	Colher	1.92	Animal	2
				Carne	2.04	Prego	1.92	Cãozinho	2
				Esponja	2.04	Massinha	1.92	Amizades	2
				Tênis	2.04	Bolacha	1.92	Origens	2
				Toalha	2.04	Perna	1.92	Refugiados	2
				Banheiro	2.04	Inseto	1.92	Afogado	2
				Pia	2.04	Pão-de-queijo	1.92	Esperança	2
				Botas	2.04	Bolinha	1.92	Arquivo	2
				Calcinha	2.04	Lataria	1.92	Irmão	2
				Batata	2.04	Pão	1.92	Senha	2
				Cachecol	2.04	Machado	1.92		
				Lençóis	2.04	Moto	1.92		
						Crânio	1.92		
						Balde	1.92		
						Chapéu	1.92		
						Formiga	1.92		
						Hambúrguer	1.92		
						Brinquedo	1.92		

<b>Baleou</b>		<b>Alcançou</b>		<b>Fatiou</b>		<b>Desenrolou</b>	
Ladrão	15.69	Meta	13.46	Bolo	35.29	Fio(s)	14.55
Bandido(s)	11.76	Objetivo(s)	21.16	Queijo	17.65	Papel	10.91
Assaltante	5.88	Metas	3.85	Presunto	15.69	Linha	9.09
Irmão	5.88	Topo	3.85	Pão	15.69	Pano	7.27
Alguém	3.92	Amigos	3.85	Carne	5.88	Rolo	7.27
Baleia	3.92	Concorrentes	3.85	Cenoura	1.96	Cobertor	3.64
Criminoso	3.92	Limite	1.92	Rosca	1.96	Barbante	3.64
Vizinho	3.92	Glória	1.92	Pudim	1.96	Novelo	3.64
Amigo(s)	3.92	Pote	1.92	Pizza	1.96	Tapete	3.64
Inocente	1.96	Nota	1.92	Melancia	1.96	Carne	1.82
Policial	1.96	Felicidade	1.92			Carta	1.82
Irmã	1.96	Mãe	1.92			Conversa	1.82
Vereador	1.96	NR	1.92			Corda	1.82
Marido	1.96	Primo	1.92			Fita	1.82
Filho	1.96	Porta	1.92			Fta	1.82
Vizinha	1.96	Maria	1.92			Massa	1.82
Casa	1.96	Bola	1.92			Relação	1.82
Gato	1.96	Vitória	1.92			Cigarro	1.82
NA	1.96	Teto	1.92			Embrulho	1.82
Homem	1.96	Ônibus	1.92			Fone	1.82
Cidadão	1.96	Recorde	1.92			Nó	1.82
Pedro	1.96	Escopo	1.92			Presente	1.82
Pessoa	1.96	Amigo	1.92			Relacionamento	1.82
Pés	1.96	Irmão	1.92			Sapato	1.82
Garota	1.96	Lugar	1.92			Lençóis	1.82
Cabeça	1.96	Maçaneta	1.92			Nós	1.82
Tira	1.96	Sucesso	1.92			Cabelo	1.82
Alvo	1.96	Ponto	1.92			Cachos	1.82
Vítima	1.96	Média	1.92			Vida	1.82
Rapaz	1.96	Colegas	1.92				
		Prateleira	1.92				
		Posição	1.92				



<b>Reformou</b>		<b>Podou</b>		<b>Derramou</b>		<b>Calibrou</b>	
Casa	76.47	Árvore	50	Leite	68.63	Pneu	80.76
Quarto	11.76	Arbusto(s)	10	Água	7.84	Aparelho	3.85
Apartamento	3.92	Gramma	8	Lágrimas	7.84	Equipamento	1.92
Abrigo	1.96	Jardim	8	Suco	3.92	Máquina	1.92
Banheiro	1.96	Planta(s)	8	Sangue	1.96	Espingo	1.92
Cozinha	1.96	Cachorro	4	Vinagre	1.96	Olhos	1.92
Sala	1.96	Flores	4	Sopa	1.96	Arma	1.92
		Pêlos	2	Copo d'água	1.96	Peso	1.92
		Rosa	2	Chá	1.96	Revolver	1.92
		Mato	2	Café	1.96	Bicicleta	1.92
		Nr	2				

<b>Sonegou</b>		<b>Alisou</b>		<b>Embrulhou</b>		<b>Calçou</b>	
Imposto(s)	90.91	Cabelo(s)	76.36	Presente(s)	74.55	Sapato(s)	59.68
Nr	1.82	Franja	3.64	Sanduíche	3.64	Botas	7.68
Nota fiscal	1.82	Roupa	1.82	Camisa	1.82	Tênis	7.68
Verba	1.82	Filha	1.82	Caneca	1.82	Sapatilhas	3.84
Ajuda	1.82	Parede	1.82	Carne	1.82	Meia	3.84
Contas	1.82	Situação	1.82	Farinha	1.82	Sandália(s)	3.84
		Mãos	1.82	Doce	1.82	Par de calças	1.92
		Roupas	1.82	Estômago	1.82	Nr	1.92
		Cachorro	1.82	Pacote	1.82	Sapatilha	1.92
		Gatinho	1.82	Pão	1.82	Chinelo	1.92
		Namorado	1.82	Filho	1.82	Luvras	1.92
		Pés	1.82	Tomate	1.82	Sonhos	1.92
		Braço	1.82	Salgados	1.82	Vergonha	1.92
				Bolo	1.82		

<b>Quitou</b>		<b>Lixou</b>		<b>Capinou</b>		<b>Resolveu</b>	
Dívida(s)	83.63	Unha(s)	76.36	Lote	58.18	Problema(s)	67.27
Do jogo	3.64	Parede(s)	5.46	Quintal	18.18	Questão(ões)	16.36
Débito	3.64	Madeira	3.64	Mato	5.45	NR	5.45
Apartamento	1.82	Pintura	1.82	Terreno	3.64	Exercício	3.64
Trabalho	1.82	Porta	1.82	Gramma	3.64	Tarefa	1.82
Partida	1.82	NR	1.82	Roça	3.64	Questionário	1.82
Curso	1.82	Portas	1.82	Rua	1.82	Enigma	1.82
Casa	1.82	Carro	1.82	Jardim	1.82	Equação	1.82
		Portão	1.82	Campo	1.82		
		Dentes	1.82	Horta	1.82		
		Pés	1.82				

<b>Escovou</b>		<b>Embaçou</b>		<b>Serrou</b>		<b>Manobrou</b>	
Dentes	67.35	Vidro	65.31	Madeira(s)	54.55	Carro	85.45
Cabelo	20.41	Óculos	8.16	Árvore	10.91	Moto	5.45
Língua	2.04	Espelho	4.08	Tábua	10.91	Ônibus	3.44
Cachorro	2.04	Janela	4.08	Porta	5.45	Caminhão	3.44
Chão	2.04	Lugar	2.04	Tora	1.82	Bicicleta	1.82
Unha	2.04	Vista	2.04	Torneira	1.82		
Cavalo	2.04	Dois	2.04	Armário	1.82		
Buraco	2.04	Tudo	2.04	Galho	1.82		
		Poça	2.04	Caixa	1.82		
		Esquemas	2.04	Lenha	1.82		
		Lente(s)	2.04	Com a serra	1.82		
		Espeho	2.04	Banco	1.82		
		Esquema	2.04	Dente	1.82		
				Ferro	1.82		

## APÊNDICE C

Tabela 3 – Itens experimentais – Julgamento de ajuste temático

Verbo	Comp	Tipo	Cloze	Condição	Verbo	Comp	Tipo	Cloze	Condição
Aparar	Gramma	fracas	37.25	mais	Aguar	Plantas	forte	60	mais
Aparar	Folha	fracas	0	imprev	Aguar	Gramas	forte	0	imprev
Aparar	Honra	fracas	0	implaus	Aguar	Bolsas	forte	0	implaus
Aparar	Flores	fracas	1.96	pouco	Aguar	Flores	forte	5.45	pouco
Digitar	Texto	fracas	37.25	mais	Acatar	Ordem	forte	51.92	mais
Digitar	Frase	fracas	0	imprev	Acatar	Norma	forte	0	imprev
Digitar	Morte	fracas	0	implaus	Acatar	Carne	forte	0	implaus
Digitar	Senha	fracas	1.96	pouco	Acatar	Regra	forte	11.54	pouco
Promover	Festa	fracas	34	mais	Arrombar	Porta	forte	65.38	mais
Promover	Hotel	fracas	0	imprev	Arrombar	Caixa	forte	0	imprev
Promover	Grade	fracas	0	implaus	Arrombar	Sombra	forte	0	implaus
Promover	Marca	fracas	2	pouco	Arrombar	Banco	forte	1.92	pouco
Afrouxar	Calça	fracas	29.41	mais	Educar	Filho	forte	68.75	mais
Afrouxar	Colar	fracas	0	imprev	Educar	Neto	forte	0	imprev
Afrouxar	Ave	fracas	0	implaus	Educar	Quadro	forte	0	implaus
Afrouxar	Tênis	fracas	1.96	pouco	Educar	Irmão	forte	2.08	pouco
Ensaboar	Corpo	fracas	40	mais	Receitar	Remédio	forte	69.09	mais
Ensaboar	Braço	fracas	0	imprev	Receitar	Vacina	forte	0	imprev
Ensaboar	Sono	fracas	0	implaus	Receitar	Carteira	forte	0	implaus
Ensaboar	Rosto	fracas	2	pouco	Receitar	Repouso	forte	5.45	pouco
Moer	Carne	fracas	28	mais	Pescar	Peixe	forte	72.73	mais
Moer	Ervas	fracas	0	imprev	Pescar	Lula	forte	0	imprev
Moer	Muros	fracas	0	implaus	Pescar	Nariz	forte	0	implaus
Moer	Queijo	fracas	2	pouco	Pescar	Sujeira	forte	1.82	pouco
Estilhaçar	Vidro	fracas	48.28	mais	Untar	Forma	forte	74.55	mais
Estilhaçar	Pires	fracas	0	imprev	Untar	Prato	forte	0	imprev
Estilhaçar	Tinta	fracas	0	implaus	Untar	Suco	forte	0	implaus
Estilhaçar	Vasos	fracas	5.17	pouco	Untar	Corpo	forte	1.82	pouco
Lustrar	Móveis	fracas	43.63	mais	Sonegar	Imposto	forte	90.91	mais
Lustrar	Anel	fracas	0	imprev	Sonegar	Valor	forte	0	imprev
Lustrar	Jantar	fracas	0	implaus	Sonegar	Teclado	forte	0	implaus
Lustrar	Taça	fracas	1.82	pouco	Sonegar	Ajuda	forte	1.82	pouco
Lambuzar	Cara	fracas	12.73	mais	Recitar	Poema	forte	81.82	mais
Lambuzar	Mesa	fracas	0	imprev	Recitar	Romance	forte	0	imprev
Lambuzar	Fogo	fracas	0	implaus	Recitar	Tomada	forte	0	implaus
Lambuzar	Bancos	fracas	1.82	pouco	Recitar	Música	forte	1.82	pouco
Quebrar	Perna	fracas	14	mais	Reformar	Casa	forte	76.47	mais

Quebrar	Dente	frac	0	imprev	Reformar	Móvel	forte	0	imprev
Quebrar	Sangue	frac	0	implaus	Reformar	Limão	forte	0	implaus
Quebrar	Bola	frac	2	pouco	Reformar	Sala	forte	1.96	pouco
Agarrar	Bola	frac	15.38	mais	Podar	Árvore	forte	50	mais
Agarrar	Chapéu	frac	0	imprev	Podar	Raízes	forte	0	imprev
Agarrar	Lago	frac	0	implaus	Podar	Caderno	forte	0	implaus
Agarrar	Corda	frac	1.92	pouco	Podar	Arbusto	forte	10	pouco
Desenrolar	Fio	frac	14.55	mais	Derramar	Leite	forte	68.63	mais
Desenrolar	Saco	frac	0	imprev	Derramar	Vinho	forte	0	imprev
Desenrolar	Olho	frac	0	implaus	Derramar	Pato	forte	0	implaus
Desenrolar	Fone	frac	1.82	pouco	Derramar	Suco	forte	3.92	pouco
Encurtar	Saia	frac	24.49	mais	Calibrar	Pneu	forte	80.76	mais
Encurtar	Ciclo	frac	0	imprev	Calibrar	Motor	forte	0	imprev
Encurtar	Fiat	frac	0	implaus	Calibrar	Nuvem	forte	0	implaus
Encurtar	Prazo	frac	2.04	pouco	Calibrar	Arma	forte	1.92	pouco
Provocar	Briga	frac	26	mais	Alisar	Cabelo	forte	76.36	mais
Provocar	Danos	frac	0	imprev	Alisar	Barriga	forte	0	imprev
Provocar	Pedra	frac	0	implaus	Alisar	Fórmula	forte	0	implaus
Provocar	Crise	frac	2	pouco	Alisar	Parede	forte	1.82	pouco
Guiar	Cego	frac	26	mais	Embrulhar	Presente	forte	74.55	mais
Guiar	Barco	frac	0	imprev	Embrulhar	Vestido	forte	0	imprev
Guiar	Canto	frac	0	implaus	Embrulhar	Floresta	forte	0	implaus
Guiar	Moto	frac	2	pouco	Embrulhar	Camisa	Forte	1.82	pouco

## APÊNDICE D

Tabela 4 – Itens de tarefas de decisão lexical

<b>cond. previs.</b>	<b>tipo subsele.</b>	<b>prime</b>	<b>alvo</b>
relac/+ esp	forte	acatar	ordem
relac/+ esp	forte	aguar	plantas
relac/+ esp	forte	alisar	cabelo
relac/+ esp	forte	arrombar	porta
relac/+ esp	forte	calçar	sapato
relac/+ esp	forte	calibrar	pneu
relac/+ esp	forte	capinar	lote
relac/+ esp	forte	derramar	leite
relac/+ esp	forte	educar	filho
relac/+ esp	forte	embaçar	vidro
relac/+ esp	forte	embrulhar	presente
relac/+ esp	forte	escovar	dentes
relac/+ esp	forte	lixar	unha
relac/+ esp	forte	manobrar	carro
relac/+ esp	forte	pescar	peixe
relac/+ esp	forte	podar	árvore
relac/+ esp	forte	quitar	dívida
relac/+ esp	forte	receitar	remédio
relac/+ esp	forte	recitar	poema
relac/+ esp	forte	reformatar	casa
relac/+ esp	forte	resolver	problema
relac/+ esp	forte	serrar	madeira
relac/+ esp	forte	sonegar	imposto
relac/+ esp	forte	untar	fôrma
nrel/ impl	forte	acatar	peixe
nrel/ impl	forte	aguar	dívida
nrel/ impl	forte	alisar	ordem
nrel/ impl	forte	arrombar	pneu
nrel/ impl	forte	calçar	vidro
nrel/ impl	forte	calibrar	plantas
nrel/ impl	forte	capinar	dentes
nrel/ impl	forte	derramar	casa
nrel/ impl	forte	educar	sapato
nrel/ impl	forte	embaçar	imposto
nrel/ impl	forte	embrulhar	árvore
nrel/ impl	forte	escovar	presente



<b>cond. previs.</b>	<b>tipo subsele.</b>	<b>prime</b>	<b>alvo</b>
nrel/ impl	forte	lixar	filho
nrel/ impl	forte	manobrar	cabelo
nrel/ impl	forte	pescar	leite
nrel/ impl	forte	podar	poema
nrel/ impl	forte	quitar	madeira
nrel/ impl	forte	receitar	porta
nrel/ impl	forte	recitar	fôrma
nrel/ impl	forte	reformatar	problema
nrel/ impl	forte	resolver	unha
nrel/ impl	forte	serrar	remédio
nrel/ impl	forte	sonegar	carro
nrel/ impl	forte	untar	lote
relac/+ esp	fraca	achatar	massa
relac/+ esp	fraca	adubar	horta
relac/+ esp	fraca	afrouxar	calça
relac/+ esp	fraca	agarrar	bola
relac/+ esp	fraca	alcançar	meta
relac/+ esp	fraca	amassar	papel
relac/+ esp	fraca	aparar	grama
relac/+ esp	fraca	baleiar	ladrão
relac/+ esp	fraca	desenrolar	fio
relac/+ esp	fraca	digitar	texto
relac/+ esp	fraca	encharcar	roupa
relac/+ esp	fraca	encurtar	saia
relac/+ esp	fraca	ensaboar	corpo
relac/+ esp	fraca	escrever	carta
relac/+ esp	fraca	estilhaçar	vidro
relac/+ esp	fraca	fatiar	bolo
relac/+ esp	fraca	guiar	cego
relac/+ esp	fraca	lambuzar	cara
relac/+ esp	fraca	lustrar	móveis
relac/+ esp	fraca	moer	carne
relac/+ esp	fraca	promover	festa
relac/+ esp	fraca	provocar	briga
relac/+ esp	fraca	quebrar	perna
relac/+ esp	fraca	resgatar	gato
nrel/ impl	fraca	achatar	briga
nrel/ impl	fraca	adubar	texto
nrel/ impl	fraca	afrouxar	móveis
nrel/ impl	fraca	agarrar	festa
nrel/ impl	fraca	alcançar	corpo

<b>cond. previs.</b>	<b>tipo subsele.</b>	<b>prime</b>	<b>alvo</b>
nrel/ impl	frac	amassar	fio
nrel/ impl	frac	aparar	carta
nrel/ impl	frac	bale	massa
nrel/ impl	frac	desenrolar	horta
nrel/ impl	frac	digitar	carne
nrel/ impl	frac	encharcar	bola
nrel/ impl	frac	encurtar	gato
nrel/ impl	frac	ensaboar	grama
nrel/ impl	frac	escrever	cego
nrel/ impl	frac	estilhaçar	saia
nrel/ impl	frac	fatiar	calça
nrel/ impl	frac	guiar	roupa
nrel/ impl	frac	lambuzar	meta
nrel/ impl	frac	lustrar	perna
nrel/ impl	frac	moer	ladrão
nrel/ impl	frac	promover	cara
nrel/ impl	frac	provocar	bolo
nrel/ impl	frac	quebrar	papel
nrel/ impl	frac	resgatar	vidro
Não palv.	Não palv.	basfro	alido
Não palv.	Não palv.	bastre	arito
Não palv.	Não palv.	benva	bapata
Não palv.	Não palv.	bimbo	botife
Não palv.	Não palv.	bistla	claristo
Não palv.	Não palv.	bisto	cangola
Não palv.	Não palv.	boma	vonlita
Não palv.	Não palv.	bonho	osmita
Não palv.	Não palv.	borracha	godele
Não palv.	Não palv.	bosma	cosvata
Não palv.	Não palv.	brono	crantada
Não palv.	Não palv.	bupa	dilaba
Não palv.	Não palv.	cabide	josta
Não palv.	Não palv.	caderno	osfa
Não palv.	Não palv.	café	lipa
Não palv.	Não palv.	caneta	orgrapa
Não palv.	Não palv.	chapéu	lojoma
Não palv.	Não palv.	chave	tostra
Não palv.	Não palv.	chela	doslido
Não palv.	Não palv.	cigarro	bacola
Não palv.	Não palv.	copo	lotrope
Não palv.	Não palv.	estojo	mosbela

<b>cond. previs.</b>	<b>tipo subsele.</b>	<b>prime</b>	<b>alvo</b>
Não palv.	Não palv.	fogão	tufre
Não palv.	Não palv.	fósforo	mopra
Não palv.	Não palv.	fruta	sagola
Não palv.	Não palv.	garrafa	fida
Não palv.	Não palv.	gaveta	pador
Não palv.	Não palv.	golomba	fenho
Não palv.	Não palv.	gravata	rasgrava
Não palv.	Não palv.	histola	perjo
Não palv.	Não palv.	holafa	sorbru
Não palv.	Não palv.	hostido	forda
Não palv.	Não palv.	infopa	posbe
Não palv.	Não palv.	ispate	foti
Não palv.	Não palv.	jalambra	fotre
Não palv.	Não palv.	janela	rebros
Não palv.	Não palv.	jintoba	relpo
Não palv.	Não palv.	josmila	frenha
Não palv.	Não palv.	lápiz	mabe
Não palv.	Não palv.	lasmitu	drenio
Não palv.	Não palv.	lometa	cobe
Não palv.	Não palv.	lupa	mesbota
Não palv.	Não palv.	lustre	mibo
Não palv.	Não palv.	macore	frosno
Não palv.	Não palv.	mapa	dofena
Não palv.	Não palv.	medalha	somalu
Não palv.	Não palv.	panela	presca
Não palv.	Não palv.	trava	latraca