

# A propósito do verbo *falar* no português brasileiro: uma análise em *corpus* e em bases de dados verbais

Isaac Souza de Miranda Junior<sup>1</sup>, Marcela Monteiro Lemos Couto<sup>1,2</sup>, Francimeire Leme Coelho<sup>1</sup>, Roana Rodrigues<sup>3</sup>, Oto Vale<sup>1</sup>

<sup>1</sup> Universidade Federal de São Carlos (UFSCar), São Carlos, Brasil

<sup>2</sup> Instituto Federal de São Paulo (IFSP), Boituva, Brasil

<sup>3</sup> Universidade Federal de Sergipe (UFS), São Cristóvão, Brasil

isaacmiranda@estudante.ufscar.br, mmlcouto@gmail.com,  
flcoelho@estudante.ufscar.br, r.roanarodrigues@gmail.com,  
otovale@ufscar.br

**Abstract.** *This work investigates the syntactic-semantic proprieties of the verb “falar” (to speak) of Brazilian Portuguese (BP) in a journalistic corpus and three BP verbal databases (VerbNet.Br, VerboWeb and Verbo-Brasil). The data demonstrate a polyvalence and complexity of “falar”. In addition, concerning the analysis of the databases, it was verified a need for manual reviews, made out by linguists, and/or expansions of the linguistic descriptions, especially when considering the relevance of constructions with the verb “falar” for different applications in the Natural Language Processing area.*

**Resumo.** *Neste trabalho, investiga-se o comportamento sintático-semântico do verbo “falar” no português brasileiro (PB) em um corpus jornalístico e em três bases de dados verbais do PB (VerbNet.Br, VerboWeb e Verbo-Brasil). Os dados demonstram a polivalência e complexidade de “falar”. Além disso, no que se refere à análise das bases de dados, foi possível constatar a necessidade de revisões manuais, realizadas por linguistas, e/ou ampliações das descrições linguísticas, principalmente ao considerar a relevância de construções com o verbo “falar” para diferentes aplicações na área de Processamento de Língua Natural.*

## 1. Introdução

O verbo é um elemento nuclear das línguas naturais e atua, sobretudo, como responsável pela seleção dos argumentos necessários e essenciais para a construção de uma oração. Há muitas pesquisas descritivas sobre o comportamento dos verbos do português brasileiro, doravante PB, a partir de diferentes abordagens teórico-metodológicas. Pode-se mencionar trabalhos que se dedicam à descrição morfológica dessa classe de palavras [Bassani 2009], além de investigações de cunho sintático-semântico com propostas de tipologias [Rassi e Vale 2013] e a criação de variadas obras lexicográficas para os usuários comuns da língua [Fernandes 2005 (1940)], [Borba 1990], entre outros.

As descrições formalizadas sobre as construções verbais do PB podem ainda atuar como recursos para diversos empreendimentos na área de Processamento de Língua Natural (PLN), como a sumarização e tradução automáticas e a análise de sentimentos. Em um estudo recente [Rodrigues - no prelo], foram selecionadas, analisadas e

comparadas três bases de dados verbais do PB (VerbNet.Br, VerboWeb e Verbo-Brasil)<sup>1</sup>, que, ademais de seu inegável valor descritivo em si, contribuem com os recursos disponíveis para o PLN. Em suma, o estudo enfatizou a existência de bases de dados robustas com informações sintático-semânticas sobre os verbos PB e o papel imprescindível do linguista na elaboração e revisão de dados lexicais para fins computacionais. Além disso, a pesquisa revelou a necessidade de adaptação, ampliação e, até mesmo, a criação de bases de dados verbais que considerem a polivalência e polissemia de determinados verbos e construções para além de seu uso como *pleno*, incluindo verbos auxiliares, verbos-suporte, ou verbos em expressões multipalavras.

Neste artigo, serão discutidos, mais especificamente, alguns dos aspectos sintático-semânticos do verbo *falar* no PB. Pode-se mencionar ao menos três motivações principais para a elaboração do presente estudo, a saber: (i) o fato de *falar* ser um verbo polivalente e bastante frequente na língua; (ii) o valor de *falar* em construções (diáteses) de comunicação [Couto 2017], que se constituem por um emissor, na posição de sujeito agente, uma mensagem e um destinatário, relevantes para o PLN ao se considerar, principalmente, sua atuação na introdução de discursos reportados que expressam *falas* (*opiniões*); e (iii) o comportamento de *falar* no PB, contrapondo-o ao verbo *dizer*, demarcando especificidades de usos se comparados a outras línguas naturais, como o espanhol [Humblé 2006] e o inglês [Dehaspe e Eynde 1991]<sup>2</sup>.

Portanto, esta pesquisa se dedica a discutir as seguintes questões: (i) quais os comportamentos sintático-semânticos de *falar* no PB, a partir de uma análise em *corpus*; (ii) como o verbo *falar* está descrito em três bases de dados verbais do PB: VerbNet.Br, VerboWeb e Verbo-Brasil; e (iii) quais informações sobre *falar* poderiam ser incluídas nas bases analisadas. Sendo assim, espera-se propor uma descrição sintático-semântica de *falar*, além de divulgar, analisar e contribuir com as bases de dados verbais disponíveis na língua.

Desse modo, este trabalho se organiza da seguinte maneira: na seção 2, são apresentados os usos de *falar*, encontrados no *corpus* de notícias do jornal Folha de São Paulo [Santana 2019], com a utilização do software Unitex. Na seção 3, descreve-se e compara-se o lexema *falar* nas três bases de dados selecionadas nesta investigação. Por fim, na seção 4, são apresentadas as considerações finais e os encaminhamentos para pesquisas futuras.

## 2. O verbo *falar* em *corpus*

A fim de verificar como o verbo *falar* se comporta em uso, além dos conhecimentos introspectivos dos pesquisadores e da classificação proposta por Borba (1990), foram analisadas as suas colocações no *corpus* do jornal Folha de São Paulo [Santana 2019], que consta de 167.053 notícias. Os dados foram processados pelo Unitex<sup>3</sup>, uma

---

<sup>1</sup> As bases de dados selecionadas no trabalho - VerbNet.Br [Scarton 2013], VerboWeb [Cançado *et al.*, 2018] e Verbo-Brasil [Duran *et al.* 2013] - sofreram atualizações na última década, estão disponíveis online e de maneira gratuita e apresentam a descrição sintático-semântica de, pelo menos, 1.000 lexemas verbais do PB.

<sup>2</sup> As pesquisas citadas destacam especificidades de uso de *hablar* e *decir*, em espanhol, e *to tell*, *to say*, *to speak* e *to talk*, em inglês, contrapondo-as aos casos em que *falar* atua como sinônimo de *dizer*, em alguns contextos do PB que introduzem um discurso reportado.

<sup>3</sup> Unitex/GramLab. Disponível em: <<https://unitexgramlab.org/pt>>. Acesso em: jun. de 2021.

ferramenta plurilingue, de código aberto, que possui dicionários eletrônicos e gramáticas locais para processamento e análise de dados textuais.

**Quadro 1. Verbo *falar* em corpus**

Colocações	Especificidade	Estrutura <sup>4</sup>	Papéis Temáticos	Exemplo
∅	Construção intransitiva	N <sub>0</sub> V	N <sub>0</sub> = Agente	<i>Camila <b>fala</b> tranquilamente.</i>
Preposição	sobre	N <sub>0</sub> V Prep Nnr	N <sub>0</sub> = Agente Nnr = Tópico	<i>O deputado não <b>fala</b> sobre o assunto</i>
	de			<i>Aquele cara só <b>fala</b> de economia.</i>
	em			<i>Dória <b>fala</b> em alterar a lei.</i>
	com <sup>5</sup>	N <sub>0</sub> V Prep NHum	N <sub>0</sub> = Agente NHum = Co-Agente	<i>Machado <b>fala</b> com a Folha.</i>
	a	N <sub>0</sub> V Prep NHum	N <sub>0</sub> = Agente NHum = Destinatário	<i>A candidata <b>falou</b> a uma multidão.</i>
	para			<i>Ele <b>fala</b> para uma plateia de 300 brasileiros.</i>
Complemento direto	Discurso reportado	N <sub>0</sub> V Que F <sub>1</sub> Prep NHum <sub>2</sub>	N <sub>0</sub> = Agente F <sub>1</sub> = Mensagem NHum <sub>2</sub> = Destinatário	<i><b>Falei</b> para os jogadores que teríamos um desafio grande.</i>
Complemento direto	Nome restrito	N <sub>0</sub> V Nr	N <sub>0</sub> = Agente Nr = Objeto [língua]	<i>Ana <b>fala</b> (inglês, português, árabe)</i>
Complemento direto	Nome restrito	N <sub>0</sub> V Nr	N <sub>0</sub> = Agente Nr = Objeto [qualidade da mensagem]	<i>Rui <b>fala</b> (bobagem, palavrão, besteira, amenidades, inverdades).</i>
Expressões multipalavras	Diferentes estruturas	-	-	<i>Carlos <b>fala</b> (grego, pelos cotovelos, a mesma língua que Ana)</i>

<sup>4</sup> Notação da estrutura sintática: N<sub>0</sub>, N<sub>1</sub> N<sub>2</sub>: argumentos na posição de sujeito, primeiro complemento e segundo complemento; Nr: argumento preenchido por um nome restrito; Nnr: argumento preenchido por um nome sem restrição; NHum: argumento preenchido por nome humano; F<sub>1</sub>: oração; Prep: preposição.

<sup>5</sup> Além da construção com a preposição *com*, foram encontrados, no *corpus*, construções pronominais com o verbo *falar* sempre construídas com sujeito plural. Neste trabalho, entende-se que tais construções são resultado de um processo transformacional de pronominalização, devido à simetria encontrada em: “A [não] fala com B” e “B [não] fala com A”, assim “A e B [não] se falam”, como no exemplo retirado do corpus: *Os irmãos não se falam desde o fim do Oasis.*

Ao todo, foram encontradas 49 mil ocorrências com *falar*<sup>6</sup> e investigadas 1.470 colocações. No Quadro 1, apresentam-se as construções analisadas, assim como uma proposta de descrição sintática (estrutura) e semântica (papéis temáticos)<sup>7</sup>, seguindo a terminologia adotada pelo modelo teórico-metodológico do Léxico-Gramática [Gross 1975].

Conforme se verifica no Quadro 1, o verbo *falar* aparece em orações intransitivas, que, por sua vez, geralmente selecionam um circunstancial de modo. Além disso, encontraram-se casos em que *falar*: (i) seleciona uma preposição, desencadeando diferentes construções sintáticas e semânticas; (ii) seleciona complemento direto e indireto, constituindo um discurso reportado; (iii) seleciona um complemento direto com nome restrito (de idioma/dialeto), que, conforme aponta Borba (1990:740), se refere a uma construção estática, sinonímia de “ter a capacidade de”; (iv) seleciona um complemento com nome restrito, fazendo referência à qualidade da mensagem; e (v) é um constituinte nuclear de expressões multipalavras<sup>8</sup>.

É interessante mencionar que, mesmo em se tratando de um *corpus* de textos jornalísticos, *falar* como verbo constituinte de construções comunicativas apresenta caráter informal, sendo encontrado em discursos reportados (marcado por aspas) e/ou em seções com menor formalidade na escrita, como o “caderno de esportes”.

### **3. O verbo *falar* nas bases de dados verbais do PB**

Conforme mencionado, nesta seção, serão apresentadas as descrições de *falar*, juntamente com uma análise comparativa com o Quadro 1, na VerbNet.Br, Verbo-Brasil e VerboWeb, bases de dados verbais relevantes do PB, que podem ser utilizadas em diferentes empreendimentos na área de PLN.

A VerbNet.Br [Scarton 2013] é um recurso léxico-computacional (RLC) que agrupa verbos do PB em diferentes classes semânticas. O recurso foi criado de maneira semiautomática a partir do RLC do inglês, a VerbNet [Schuler 2005]. As classes semânticas, inspiradas em Levin (1993), agrupam os verbos de acordo com semelhanças semânticas compartilhadas e suas alternâncias sintáticas.

Por conta da abordagem inter-linguística utilizada na criação desse RLC, as classes presentes na VerbNet.Br são alinhadas diretamente com as classes do inglês. Cada uma delas apresenta as seguintes informações: (i) os membros que a compõem; (ii) os papéis temáticos, que foram herdados da versão do inglês; (iii) as restrições seletivas, que são impostas aos papéis temáticos (como animacidade, concretude etc.), também herdadas do inglês; os frames sintáticos do PB e do inglês, que descrevem a transitividade verbal e, também, itens lexicais selecionados em alternâncias em particular; e (iv) os predicados semânticos, que fornecem informações sobre as relações entre os participantes e o evento da ação verbal.

---

<sup>6</sup> Este número abarcou também as formas nominais *a fala* e *o falar*. Neste trabalho, analisaram-se somente as construções verbais.

<sup>7</sup> Os papéis temáticos utilizados na pesquisa baseiam-se no estudo de Santos (2014).

<sup>8</sup> Em Vale (2002) estão descritas e classificadas, sintático-semânticamente, 22 expressões multipalavras com o verbo *falar* no PB, tais como: *falar da boca pra fora* ou *falar de barriga cheia*.

**Quadro 2. Verbo *falar* na VerbNet.Br**

FALAR		
Classes	Papéis temáticos [restrições seletivas]	Alternâncias
<i>amuse-31.1</i>	Experiencer [+animate] Stimulus Result	V; V_NP; V_NP_PP [com]
<i>chit_chat-37.6</i>	Agent [+animate   +organization] Co-Agent [+animate   +organization] Topic [+communication]	V; V_PP[com]; V_PP[sobre]; V_PP[sobre]_PP[com]
<i>correspond-36.1</i>	Agent [+animate   +organization] Co-Agent [+animate   +organization] Theme	V; V_PP[com]_PP[sobre]; V_PP[sobre]
<i>dub-29.3</i>	Agent [+animate   +organization] Theme [+concrete   +organization] Result	V_NP
<i>establish-55.5</i>	Agent [+animate   +organization] Theme	V_NP
<i>lecture-37.11</i>	Agent [+animate   +organization] Topic Recipient [+animate   +organization]	all
<i>talk-37.5</i>	Agent [+animate   +organization] Co-Agent [+animate   +organization] Topic [+communication]	V; V_PP[com]; V_PP[com]_PP[sobre]; V_PP[para]; V_PP[para]_PP[sobre]; V_PP[sobre]_PP[com]; V_PP[sobre]_PP[de]; V_PP[sobre]_PP[para]
<i>transfer_mesg-37.1.1</i>	Agent [+animate   +organization] Topic Recipient [+animate   +organization] Source	V; V_NP; V_NP_PP[a]; V_NP_PP[para]

Nota. Fonte: Adaptado de Scarton (2013).

Conforme se observa no Quadro 2, o verbo *falar* está presente em oito classes da VerbNet.Br. e muitos *frames* sintáticos estão presentes em mais de uma classe. É importante salientar que a VerbNet.Br não exemplifica as alternâncias com dados reais do PB ou mesmo criados por introspecção do falante, o que dificulta, em alguns casos, a interpretação dos dados descritos. Para a alternância sintática **V\_PP[sobre]\_PP[de]**, da classe *talk-37.5*, por exemplo, não foi localizada nenhuma ocorrência compatível no *corpus* de dados jornalísticos como se verifica nos dados dispostos no Quadro 1. A pista para o PB é que se trata de um *frame* com duas preposições (*sobre/de*) e com a restrição seletiva [+communication] imposta ao papel temático Topic. Note-se que as preposições *sobre* e *de* precedidas do verbo *falar* selecionam potencialmente argumentos com papel temático assunto (Tópico) ou *conteúdo da mensagem* (Mensagem), conforme defende Couto (2017). No entanto, no *corpus* não há ocorrência do verbo *falar* com essas duas preposições antecedendo o *conteúdo da mensagem*.

Salienta-se, ainda, o comportamento do verbo *falar* descrito na classe *lecture-37.11*, o qual possui a descrição *all* no que diz respeito às alternâncias sintáticas. Isso quer dizer que o verbo *falar* admite todas as alternâncias sintáticas do PB. Essa é uma

informação difícil de ser atestada, porque não existe uma delimitação teórico-metodológica de quais seriam essas alternâncias ou uma lista das alternâncias do PB para realizar buscas e testes com maior rigor metodológico. Novamente, a ausência de exemplos é um outro grande limitador dessa verificação.

A VerboWeb [Cançado *et al.* 2018] é uma base digital de consulta lexical, que, devido sua organização e metodologia, pode ser utilizada como RLC. Nela, os verbos estão classificados em função de suas características sintático-semânticas também com base nas classes de Levin (1993). A descrição proposta foi realizada de maneira manual por linguistas e a tipologia verbal ocorre em duas etapas de categorização: classes e subclasses. A classificação em *classes* consiste em uma divisão dos verbos em função de quatro características: *Estrutura sintática básica* (disposição sintática dos constituintes: Sintagma Nominal+Verbo, etc.); *Papéis temáticos* (relação semântica entre um predicado e seus argumentos: agente, paciente, etc.); *Decomposição de predicado* (demonstração de um predicado em função de predicados primitivos, ou seja, a descrição dos predicados irredutíveis que compõem um predicado complexo: *O palestrante falava muito* ⇒ [X DO <EVENT>]); *Aspecto lexical* (classificação dos predicados em função de duração, telicidade e dinamicidade nas classes de Vendler (1967)). Por sua vez, a organização das subclasses consiste no arranjo interno dos verbos de uma classe em função de suas características sintático-semânticas em comum.

**Quadro 3. Verbo *falar* na VerboWeb**

<p><b>Verbo:</b> <i>Falar</i>  <b>Exemplo:</b> <i>O palestrante falava muito</i></p>
<p><b>Classe:</b> Atividade: verbos internamente causados (inergativos)  <b>Propriedades da Classe:</b>          Conteúdo semântico recorrente na classe: x faz/produz um evento em si mesmo          – Estrutura sintática básica: [SN V] (verbo intransitivo)          – Estrutura de papéis temáticos: {Agente}          – Estrutura de decomposição de predicados: [X DO &lt;EVENT&gt;]          – Aspecto lexical básico: atividade          – Licencia um objeto cognato: <i>O palestrante falou uma fala bonita.</i>          – Licencia um adjunto equivalente ao objeto cognato: <i>O palestrante falou muito bonito.</i></p>
<p><b>Subclasse:</b> Verbos de expressão (modo de fala)  <b>Propriedades da Subclasse:</b>          Denota um evento de fala          – Licencia mensagem comunicada na posição de objeto: <i>O palestrante falava muita besteira.</i>          – Licencia a mensagem e o destinatário nas posições de objeto: <i>O palestrante falava muita besteira para o público jovem.</i>          – Licencia um objeto sentencial : <i>O palestrante falava que os jovens são o futuro do país.</i>          – Licencia destinatário na posição de objeto indireto: <i>O palestrante falava animado para o público jovem.</i></p>

Nota. Fonte: Cançado M., Amaral, L., e Meireles, L. (2021). VerboWeb. Falar. [http://www.lettras.ufmg.br/sistemas/verboweb\\_cliente/ver\\_verbo.php?id=1220](http://www.lettras.ufmg.br/sistemas/verboweb_cliente/ver_verbo.php?id=1220), junho de 2021.

O verbo *falar* é descrito dentro da VerboWeb como parte da subclasse dos *verbos de expressão* dentro da classe *verbos de atividade* internamente causados (inergativos), como é replicado no Quadro 3.

Com o Quadro 3, pode-se notar que a base descreve dois usos de *falar*, o primeiro como verbo intransitivo, fazendo parte da classe dos verbos inergativos, e o segundo como verbo de comunicação, referente à subclasse dos verbos de expressão, desconsiderando, no entanto, os casos em que *falar* ocorre precedido das preposições *sobre*, *de*, *em* e *com*, presentes no Quadro 1.

Ademais, pode-se reparar também que o “licenciamento de mensagem comunicada na posição de objeto” proposto na VerboWeb faz referência ao mesmo uso (iv) no Quadro 1, porém, o exemplo utilizado (*O palestrante falava muita besteira*) não se refere ao *conteúdo da mensagem* como proposto na VerboWeb, mas sim, como descrito na seção 2, à *qualidade da mensagem*. *Besteira* no exemplo, não seria a fala em discurso reportado, mas sim uma qualidade/característica associada a ele pelo locutor.

Por último, a Verbo-Brasil [Duran *et al.* 2013] é um RLC em que, ao contrário das bases anteriores, os verbos não estão distribuídos em classes maiores de categorização, mas sim descritos individualmente em função dos diferentes sentidos (*roleset*) que podem assumir. Mesmo que exista dentro da base uma possível classificação (*vncls: verbnet class*) em função das classes da VerbNet [Schuler 2005], existem sentidos que são vazios quanto essa classificação por não estarem contemplados na VerbNet.

Assim como as outras bases, a Verbo-Brasil é orientada pela descrição das classes de Levin (1993), em que cada verbo é classificado de acordo com suas propriedades sintáticas e aspectos semânticos em comum. Além disso, na Verbo-Brasil, cada sentido é descrito e classificado em função dos seus argumentos e vinculado a um sentido existente no PropBank [Palmer *et al.* 2005] quando possível. Cada argumento recebe uma identificação de papel temático (*vnrole: verbnet role*) em função dos papéis temáticos descritos na VerbNet [Schuler 2005].

O verbo *falar* é descrito dentro da Verbo-Brasil [Duran *et al.* 2013] com apenas um sentido contendo quatro argumentos que não necessariamente precisam ocorrer juntos em uma sentença. Dentro da descrição de *falar*, encontra-se também o sentido de duas expressões multipalavras: *falar mal* e *dar o que falar* e os seus respectivos argumentos e papéis temáticos. A descrição do sentido de *falar* pode ser verificada no Quadro 4:

**Quadro 4. Verbo *falar* na Verbo-Brasil**

	Sentido	Estrutura Argumental
Falar	<b>Roleset id:</b> falar.01, dizer; declarar; <b>vncls:</b> 37.7; <b>Mapeamento para o inglês:</b> say.01, talk.01, speak.01	<b>Arg0:</b> <i>falante</i> (vnrole: 37.7-agent) <b>Arg1:</b> <i>idioma falado; assunto (falar de, falar sobre, falar contra, falar a favor de) ou elocução</i> (vnrole: 37.7-topic) <b>Arg2:</b> <i>ouvinte, interlocutor (falar a, falar para, falar com)</i> (vnrole: 37.7-recipient) <b>Arg3:</b> <i>atributo do arg0 (falar como)</i> <sup>9</sup>

Nota. Fonte: Duran, M. S., Martins J. P., Coimbra, M., Patire, P. A., Hartmann N., e Aluísio, S. (2014). Verbo-Brasil. Framefile-falar-v. <http://143.107.183.175:21380/verbobrasil/textoFrames/falar-v.html>, junho de 2021.

<sup>9</sup> A construção *falar como* foi encontrada no corpus mas, por não se tratar de um argumento selecionado pelo verbo, não foi inserida no Quadro 1.

A forma como *falar* está descrito no Quadro 4, mesmo que abrangente, pode gerar alguns problemas de descrição, uma vez que o **roleset falar.01** apresenta 32 exemplos<sup>10</sup> diferentes de construções com o verbo, em que cada construção tem especificidades sintáticas distintas, apresentando números distintos de argumentos. Vale ressaltar também que em alguns exemplos da Verbo-Brasil, como o **roleset falar.01** exemplo 2 (*Não ouvi falar nada sobre arbitragens*) e o **roleset falar.01** exemplo 30 (*Takuo Hirano e Tetsuyuki Hirano, pai e filho, vieram falar sobre o novo conceito de design que estão desenvolvendo*), são representados na base com estruturas argumentais idênticas (contando apenas com **arg1**), o que é um problema considerando que ambos os exemplos apresentam estruturas distintas, enquanto que no exemplo 2 não há a presença explícita de um **arg0**, no exemplo 30 *Takuo Hirano e Tetsuyuki Hirano* é **arg0** não somente de *vir*, mas como também de *falar*.

Como visto na seção, cada base exprime uma cobertura distinta dos usos de *falar*, isso ocorre devido à polivalência e polissemia de *falar*, reforçando a necessidade da revisão e ampliação das descrições disponíveis sobre esse verbo nas bases de dados analisadas.

#### 4. Considerações finais

Por se tratar de um verbo bastante frequente, o comportamento do verbo *falar* apresenta um número considerável de empregos que merecem ser considerados em qualquer trabalho descritivo. Esse é um fato comum em itens de alta frequência que apresentam, via de regra, uma polissemia que nem sempre é detectável à primeira vista.

De fato, retomando as questões que nortearam esta pesquisa, nota-se que o verbo *falar* no PB apresenta um comportamento sintático-semântico variado, que requer estudos aprofundados. Além dos casos recenseados no Quadro 1, destaca-se ainda a necessidade de análises dedicadas à polivalência e polissemia de *falar* em construções como: *falar em* (*seu discurso + casamento + inglês...*), *falar* (*como + na condição de + na qualidade de*) (*presidente + professor*) e *falar* (*por + no lugar de*) *alguém*. Tais complementos poderiam ser atribuídos à estrutura argumental do verbo? Esse tipo de pergunta poderia ser estendida a outras construções de comunicação, considerando-se o comportamento de diferentes *verba dicendi* [Costa e Freitas 2017] [Baptista 2010].

No que se refere ao estudo de *falar* nas bases de dados verbais, verificou-se que todas as bases enfatizam o uso de *falar* em construções de comunicação. Constatou-se ainda que as bases verbais construídas para fins computacionais (VerbNet.Br e Verbo-Brasil) necessitam de uma revisão linguística especializada, seja para incluir exemplos reais da língua, seja para refinar a descrição e propor classificações mais granulares dos dados. Por sua vez, a base que se constrói sob descrições linguísticas manuais (VerboWeb) parece ser a que possui maior coerência teórico-metodológica, embora apresente um número reduzido de construções analisadas, podendo ampliar o estudo de tal verbo em futuras atualizações da base.

Sendo assim, acredita-se que as informações sobre o verbo *falar* levantadas no *corpus* e descritas neste trabalho podem impulsionar não só ao incremento e atualização das bases de dados verbais elencadas, como também à realização de novos estudos sintático-semânticos sobre esse verbo.

---

<sup>10</sup> Os exemplos da Verbo-Brasil foram extraídos do *corpus* PLN-Br [Bruckschen *et al.* 2008].

## Agradecimentos

Os autores agradecem à CAPES (Código Financeiro 001) e ao Centro de Inteligência Artificial (C4AI) da Universidade de São Paulo, apoiado pela IBM e FAPESP (nº 2019/07665-4).

## Referências

- Baptista, J. (2010). Verba dicendi: a structure looking for verbs. In: Nakamura, T., Laporte, E., Dister, A., Fairon, C. *Les Tables – La grammaire du français par le menu*. Mélanges en hommage à Christian Leclère. Cahiers du CENTAL, 6, 11-20.
- Bassani, I. S. (2009). *Formação e interpretação dos verbos denominais do português do Brasil*. Dissertação (Mestrado em Semiótica e Linguística). São Paulo, Universidade de São Paulo.
- Borba, F. S. (coord.). (1990). *Dicionário gramatical de verbos do português contemporâneo do Brasil*. São Paulo: Editora UNESP.
- Bruckschen, M., Muniz, F., Souza, J. G. C., Fuchs, J. T., Infante, K., Muniz, M. e Aluísio, S. M. (2008). Anotação linguística em XML do corpus PLN-BR. *Série de relatórios do NILC, NILC-ICMC-USP*.
- Cançado, M., Amaral, L., e Meirelles, L. L. (2018). Verboweb: Uma proposta de classificação verbal. *Revista da Anpoll*, 1(46), 123-141.
- Cançado M., Amaral, L., e Meireles, L. (2021). VerboWeb. Falar. [http://www.letras.ufmg.br/sistemas/verboweb\\_cliente/ver\\_verbo.php?id=1220](http://www.letras.ufmg.br/sistemas/verboweb_cliente/ver_verbo.php?id=1220), Junho de 2021.
- Costa, B. F. S., e Freitas, C. (2017). Verbos de elocução em português: um estudo descritivo com base em grandes corpora e motivado pela linguística computacional. *Fórum Linguístico*, 14(3), 2266-2285.
- Couto, M. M. L. (2017). *O estudo das valências verbais aplicado às construções de comunicação do português brasileiro*. Dissertação (Mestrado em Estudos Linguísticos). Belo Horizonte, Universidade Federal de Minas Gerais.
- Dehaspe, L. e van den Eynde, K. (2012). The Pronominal Approach to Verbal Valency: A formal description of speak, say, tell, and talk. In *Betriebslinguistik und Linguistikbetrieb* (pp. 273-280). Max Niemeyer Verlag.
- Duran, M. S., Martins, J. P., e Aluisio, S. M. (2013). Um repositório de verbos para a anotação de papéis semânticos disponível na web. In: *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*. <<https://www.aclweb.org/anthology/W13-4820.pdf>>, acesso em junho de 2021.
- Duran, M. S, Martins J. P., Coimbra, M., Patire, P. A., Hartmann N. e Aluísio, S. (2014). Verbo-Brasil. [Framefile-falar-v. http://143.107.183.175:21380/verbobrasil/textoFrames/falar-v.html](http://143.107.183.175:21380/verbobrasil/textoFrames/falar-v.html)>, acesso em junho de 2021.
- Fernandes, F. (2005 [1940]). *Dicionário de verbos e regimes*. 45 ed. Porto Alegre: Globo.

- Gross, M. (1975). *Méthodes en Syntaxe: régime des constructions complétives*. Paris: Hermann,
- Humblé, P. (2006). Falsos cognados. Falsos problemas. Un aspecto de la enseñanza del español en Brasil, *Revista de Lexicografía*, 2005-2006, 12: 197-207 <<https://ruc.udc.es/dspace/handle/2183/5510>>, acesso em junho de 2021.
- Levin, B. (1993). *English Verb Classes And Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Palmer, M., Gildea, D., e Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1), 71-106.
- Rassi, A. P. e Vale, O. A. (2013). Tipologia das construções verbais em PB: uma proposta de classificação do verbo dar. *Caligrama*, Belo Horizonte, v. 18, n. 2.
- Rodrigues, R., Lemos-Couto, M. M., Coelho, F. L., Miranda Jr, I. e Vale, O. (2021) Bases de dados verbais do português brasileiro (artigo em vias submissão).
- Santana, M. (2019). “News of the Brazilian Newspaper - 167.053 news of the site Folha de São Paulo (Brazilian Newspaper)”, <<https://www.kaggle.com/marlesson/news-of-the-site-folhauol>>, acesso em junho de 2021.
- Santos, R. P. T. (2014). *Automatic Semantic Role Labeling for European Portuguese*. Dissertação (Mestrado em Ciências da Linguagem). Universidade do Algarve, Faro.
- Scarton, C. E. (2013). *VerbNet.Br: construção semiautomática de um léxico verbal online e independente de domínio para o português do Brasil*. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional). Universidade de São Paulo, São Carlos.
- Schuler, K. K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.
- Vale, O. A. (2002). *Expressões cristalizadas do português do Brasil: uma proposta de tipologia*. Tese (Doutorado em Linguística e Língua Portuguesa). Universidade Estadual Paulista Júlio de Mesquita Filho, Araraquara.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, NY: Cornell Univ. Press.