

Semiprodutividade construcional: uma investigação empírica da complementação sentencial do português brasileiro

The partial productivity of grammatical constructions: an empirical investigation into Brazilian Portuguese clausal complementation

Dayanne de Oliveira Ximenes¹
Diogo Pinheiro²

Resumo: A literatura em Gramática de Construções Baseada no Uso – entendida aqui como a vertente cognitivo-funcional da Gramática de Construções – menciona três mecanismos possíveis para explicar como o falante evita a supergeneralização: o conhecimento gramatical, o enraizamento e o bloqueio estatístico. Este artigo busca avaliar a realidade psicológica de cada um desses mecanismos a partir da análise de uma construção específica do português brasileiro: a Construção de Complementação Sentencial (por exemplo, “Ela disse que sairia”). Após uma análise baseada em *corpus*, da qual derivamos três hipóteses relativas à representação mental da construção em pauta, relatamos um experimento de produção induzida. Os resultados confirmam a atuação de um mecanismo baseado em conhecimento gramatical e de um mecanismo baseado em conhecimento estatístico em termos de enraizamento. Por outro lado, não foi possível comprovar a atuação do bloqueio estatístico.

Palavras-chave: Complementação sentencial. Conhecimento semântico. Enraizamento. Bloqueio estatístico.

Abstract: How do speakers retreat from overgeneralization? In Usage-Based Construction Grammar, three answers have been offered: (i) speakers rely on general grammatical knowledge; (ii) speakers rely on item-based statistical knowledge in terms of entrenchment; (iii) speakers rely on item-based statistical knowledge in terms of preemption. In order to probe into the psychological validity of these explanations, this paper focuses on the Brazilian Portuguese Clausal Complementation Construction (e.g. “Ela disse que estava bem” ‘She said that she was fine’). After a corpus-based analysis, from which we derive three specific hypothesis concerning the representation of the underlying constructional network, an elicited production experiment was carried out where subjects were asked to watch short videos and answer interpretive questions. The results suggest that grammatical (semantic) knowledge and entrenchment play an important role in the retreat from overgeneralization; on the other hand, no evidence was found to support statistical preemption.

Keywords: Clausal complementation. Semantic knowledge. Entrenchment. Statistical preemption.

¹ Universidade Federal do Rio de Janeiro, Programa de Pós-Graduação em Linguística, Rio de Janeiro, RJ, Brasil. Endereço eletrônico: dayannecolosso@gmail.com.

² Universidade Federal do Rio de Janeiro, Faculdade de Letras, Programa de Pós-Graduação em Linguística, Rio de Janeiro, RJ, Brasil. Endereço eletrônico: diogopinheiro@letras.ufrj.br.

Primeiras palavras

Como toda teoria gramatical, a Gramática de Construções Baseada no Uso – entendida aqui como a vertente cognitivo-funcional da Gramática de Construções – deve lidar com o problema da geratividade linguística. Em outras palavras, o modelo deve representar o conhecimento linguístico do falante de maneira tal que ele seja capaz de licenciar todas e apenas as sentenças possíveis na língua sob investigação. Frequentemente, esse desafio se revela nada trivial, como mostram os exemplos abaixo (GOLDBERG, 1995):

- (1) a. Joe gave the earthquake relief fund \$5.
b. ?? Joe donated the earthquake relief fund \$5.

Ainda que a única diferença entre as duas sentenças sejam os verbos – que são, inclusive, semanticamente próximos –, a primeira é avaliada como natural e a segunda como pouco aceitável para falantes do inglês. Sob uma ótica construcionista, e assumindo-se que construções podem ser mais ou menos produtivas a depender da quantidade de itens da língua que podem instanciá-las, o problema colocado pelos exemplos acima pode ser referido como a questão da *semiprodutividade construcional*.

Para investigar a semiprodutividade de construções e explicar, portanto, a natureza do saber inconsciente que leva o falante a evitar usos como (1b), pesquisadores alinhados à GCBU têm evocado a existência de dois tipos de conhecimento: o gramatical, que envolve o domínio inconsciente de regularidades fonológicas, morfológicas, sintáticas, semânticas ou pragmáticas, e o estatístico, que diz respeito ao registro da experiência do falante com cada item – ou sequência de itens – particular. A relevância e o papel de cada tipo de conhecimento, no entanto, ainda não são consensuais na literatura.

Uma primeira discussão se coloca entre aqueles que enfatizam o papel do conhecimento estatístico, como Dabrowska (2014) e Rosanen, Ambridge e Pine (2016), e aqueles que chamam a atenção para a importância das generalizações gramaticais, como Goldberg e Perek (2015) e Ambridge e Goldberg (2008). Assim, ao mesmo tempo em que tem ganhado força na literatura a preocupação com o conhecimento puramente estatístico, baseado no item, é possível identificar um movimento de reação, que procura valorizar o caráter geral e regular do conhecimento linguístico do falante.

Uma segunda discussão diz respeito à atuação dos dois mecanismos ligados ao conhecimento estatístico: o enraizamento (em inglês, “entrenchment”), que diz respeito simplesmente ao grau de fixação de uma determinada construção no inventário construcional

do falante, e o bloqueio estatístico (em inglês, “statistical preemption”), que é um tipo de aprendizado baseado em erro de previsão (“error-driven learning”). Enquanto alguns estudos argumentam em favor da centralidade do bloqueio estatístico (ROBENALT; GOLDBERG, 2015; PEREK; GOLDBERG, 2017), outros sugerem que apenas o enraizamento é verdadeiramente efetivo para impedir supergeneralizações (AMBRIDGE *et alii*, 2015).

Inserindo-se nesses debates, esta pesquisa se volta para uma construção presente na gramática do português brasileiro (PB), a que nos referimos aqui como Construção de Complementação Sentencial (CCS). Trata-se de uma estrutura sintático-semântica esquemática que tem a forma SUJEITO + VERBO + SINTAGMA ORACIONAL INTRODUZIDO PELO COMPLEMENTIZADOR “QUE” e é utilizada produtivamente para, por exemplo, reportar discursos, percepções e avaliações. Exemplos da CCS podem ser vistos abaixo:

- (2) A diretora anunciou que o concurso foi adiado.
- (3) Meu irmão percebeu que ninguém estava interessado.

Apesar de poder ser instanciada por uma ampla gama de verbos, é possível verificar que a CCS não é plenamente produtiva, como mostram os exemplos abaixo:

- (4) ?? João criticou que o jogo acabou tarde.
- (5) ? Ela opinou que a festa foi perfeita.

Para investigar a semiprodutividade da CCS, contribuindo para o debate acerca dos mecanismos cognitivos que permitem ao falante evitar a supergeneralização, este estudo foi dividido em duas partes. Na primeira, realizou-se uma análise de *corpus* com o objetivo de levantar hipóteses sobre a representação mental da CCS por parte de falantes nativos do PB, em particular no que tange aos mecanismos que bloqueiam a supergeneralização. Na segunda, buscou-se verificar, por meio de um experimento psicolinguístico de produção induzida, cada uma das hipóteses levantadas anteriormente, a fim de avaliar a realidade psicológica dos mecanismos de bloqueio da supergeneralização³.

³ Como se observa, este artigo combina dois recursos metodológicos: método observacional (análise de *corpus*) e método experimental. Essa combinação tem sido produtiva na literatura em GCBU (ver, por exemplo, Ambrige *et alii* (2015) e Ambridge *et alii* (2008)), o que se explica, a nosso ver, por duas razões. A primeira, de natureza mais geral, diz respeito ao fato de que a integração de métodos complementares permite que o ponto fraco de cada um deles (a maior dificuldade em controlar as variáveis e a alegada ausência de evidência negativa, no caso das investigações observacionais; a artificialidade da tarefa e potencialmente dos resultados, no caso das investigações experimentais) seja em alguma medida compensado pelo outro, em um movimento de beneficiamento mútuo. A

O artigo está organizado como segue. Na próxima seção, apresentamos os princípios básicos da GCBU e discutimos o problema da semiprodutividade construcional. Em seguida, apresentamos a primeira parte do estudo, baseada na análise de dois *corpora*. Na seção seguinte, sistematizamos as hipóteses sobre a representação do conhecimento construcional, que podem ser derivadas da análise de *corpus*. Na sequência, voltamos para a segunda parte do estudo, baseada em metodologia experimental. Nesse momento, detalhamos o experimento realizado, descrevemos os resultados obtidos e, sobretudo, discutimos suas implicações analítico-descritivas. As considerações finais sintetizam as principais contribuições da pesquisa.

Gramática de Construções Baseada no Uso

A Gramática de Construções Baseada no Uso (GCBU), aqui definida como a vertente funcional-cognitiva da Gramática de Construções, é um modelo gramatical que busca capturar a totalidade do conhecimento linguístico do falante de modo declarativo (isto é, não derivacional), sob a forma de uma rede de construções gramaticais (unidades simbólicas que consistem em pareamentos de forma e significado). Para que isso seja possível, o conceito de construção gramatical deve ser amplo o suficiente para abarcar uma variedade de unidades linguísticas, de palavras e expressões cristalizadas (“bola”, “Quem tudo quer tudo perde”, etc.) até padrões prosódicos e sintáticos esquemáticos (entonação ascendente; padrão SVO; etc.), passando por padrões parcialmente especificados (ex: “Vê se X”, “Que mané X”; etc).

À luz da GCBU, a capacidade gerativa da linguagem se explica a partir da integração entre construções mutuamente compatíveis. Por exemplo: uma construção abstrata, como a construção bitransitiva do inglês, cuja forma é SUJEITO VERBO OBJ₁ OBJ₂, pode ser combinada com construções mais concretas, fixas e fonologicamente preenchidas, como o verbo já flexionado “told”, o determinante “a” e os nomes “Joe”, “Mary” e “story”, resultando na sentença “Joe told Mary a story”. É importante notar, entretanto, que não são todos os verbos que podem ser instanciados nesse padrão. Na verdade, mesmo alguns verbos semanticamente próximos de “tell” são bloqueados na construção bitransitiva, como se observa em “*Joe whispered Mary a story” (GOLDBERG, 1995).

Sabemos que não é comum que um falante nativo produza sentenças gramaticalmente inaceitáveis na sua língua, a despeito do fato de que, durante o processo de aprendizado, ele tende a não ser exposto diretamente a evidências negativas, como correções explícitas ou sinais ostensivos de incompreensão por parte do ouvinte (BOYD; GOLDBERG, 2011). Diante disso,

segunda, atrelada especificamente à natureza da GCBU, diz respeito ao fato de que esse modelo teórico estuda, precisamente, o impacto do uso concreto sobre a representação mental subjacente.

como explicar o fato de eles não supergeneralizarem construções semiproductivas, produzindo enunciados inaceitáveis como “*Joe whispered Mary a story”?

Na literatura em GCBU, dois tipos de respostas têm sido propostos para esse problema. De um lado, diversos autores sugerem que a supergeneralização é evitada graças a uma exigência de compatibilidade entre propriedades gramaticais da construção abstrata e da construção lexical com potencial para instanciá-la. Segundo Goldberg (1995), é isso que se observa nos exemplos abaixo:

- (6) a. I invited Bob into the room.
- b. I asked Bob into the room.
- c. *I begged Bob into the room

Embora (6a) e (6b) sejam gramaticais, o mesmo não se pode dizer de (6c), apesar do fato de o verbo “beg” ser semanticamente próximo a “invite” e “ask”. Segundo Goldberg (1995), isso acontece porque a construção de movimento causado (que está presente nos três enunciados) apresenta uma restrição semântica segundo a qual o deslocamento do tema (no caso, “Bob”) deve ser a situação *default*, esperada. Em (6c), porém, a presença do verbo “beg” sugere que a ocorrência do movimento não pode ser presumida. Assume-se, assim, que a inserção de “beg” na construção de movimento causado produz um conflito semântico, cujo resultado é uma sentença malformada. Neste caso, portanto, a generalização indevida de (6c) seria evitada graças ao conhecimento gramatical do falante acerca da semântica da construção de movimento causado e da construção lexical “beg”.

Muitos autores sustentam ainda que, além desse conhecimento de natureza gramatical, a supergeneralização pode ser evitada graças a um conhecimento do tipo estatístico (AMBRIDGE, 2013; BOYD; GOLDBERG, 2011). Esse tipo de conhecimento é idiossincrático, na medida em que diz respeito ao registro do histórico da experiência do falante com cada item particular. Especificamente, dois mecanismos já foram propostos: o enraizamento (“entrenchment”), por meio do qual o falante evita um item A em uma construção abstrata B sempre que não experiencia essa formulação específica; e o bloqueio estatístico (“statistical preemption”), por meio do qual o falante evita um item A em uma construção abstrata B sempre que não experiencia essa formulação específica e, ao mesmo tempo, experiencia recorrentemente uma formulação alternativa funcionalmente comparável.

A título de exemplo, considere-se a sentença do inglês “?? Explain me the story”, que soa pouco natural para um falante do inglês (BOYD; GOLDBERG, 2011). Uma explicação

baseada no conceito de enraizamento para o fato de que falantes nativos evitam esse tipo de formulação seria bastante simples. Aqui, a ideia é simplesmente a de que o falante nunca ou apenas muito raramente ouve o verbo “explain” instanciado na construção bitransitiva, o que o leva a adotar a postura linguisticamente conservadora de não arriscar essa instanciação.

Por outro lado, uma explicação baseada no mecanismo de bloqueio estatístico é um pouco mais complexa. Neste caso, a proposta é a seguinte: se o falante experiencia frequentemente o verbo “explain” na construção dativa (como em “She explained the story to me”) em situações nas quais esperava que fosse usada a construção bitransitiva (como em “?? She explained me the story”), acaba por concluir que a construção bitransitiva não é apropriada especificamente para esse verbo, ainda que outros verbos semanticamente próximos sejam possíveis em ambos os padrões (BOYD; GOLDBERG, 2011). Nesse sentido, o bloqueio estatístico é entendido como um tipo de aprendizado baseado no erro (“error-driven learning”); isto é, assume-se que o aprendizado resulta da frustração de uma expectativa linguística (ROBENALT; GOLDBERG, 2015).

Conforme observado na introdução, não existe consenso na literatura em GCBU quanto ao papel dos dois tipos de conhecimento – o gramatical e o estatístico – na prevenção da supergeneralização. Além disso, não existe consenso em relação à atuação dos dois tipos de processos de natureza estatística – o enraizamento e o bloqueio. Diante disso, este trabalho busca contribuir com novos elementos para elucidar essas questões teóricas.

Uma primeira aproximação: explorando os *corpora*

Para investigar a CCS do PB, recorreremos inicialmente aos *corpora* NILC/São Carlos e Corpus Brasileiro (linguateca.pt/ACDC). Para encontrar instâncias dessa construção, foi utilizado o comando `[pos="V"] "que"`, de modo que foram buscadas sequências que contivessem qualquer forma verbal seguida da palavra “que”. Todos os resultados foram revisados manualmente, a fim de que resultados indevidos fossem descartados.

Ao todo, foram encontradas 1664 instâncias da CCS, distribuídas por 157 verbos distintos⁴. Uma primeira aproximação dos dados revelou uma característica comum a todos os verbos que se combinam com a CCS: trata-se de verbos que exigem (na acepção que tomam na moldura sintática da CCS) complemento que denote proposição⁵. Assim, é possível explicar a

⁴ Em função das limitações de espaço, não é possível apresentar aqui a frequência token de cada verbo. Alguns dados são apresentados na próxima seção a fim de embasar as redes construcionais postuladas.

⁵ Tal como definida aqui, uma proposição é uma representação conceptual de situações, eventos, estados de coisas, etc., sendo reconhecível pela possibilidade de ser verdadeira ou falsa (LAMBRECHT, 1994).

completa ausência de muitos verbos entre 157 *types* identificados – como “comprar” e “comer”, por exemplo – pelo fato de que seus complementos não expressam proposições. A título de exemplo, contrastem-se as seguintes sentenças:

- (7) a. Na verdade esse teste, da maneira como foi construído, **sugeri**u que eles desconheciam a contagem. (*Corpus Brasileiro*)
b. E a Turma da Mônica, ops equipe econômica, **pensa** que recessão é passar o filme duas vezes: re-sessão. (*Corpus Brasileiro*)
- (8) a. Júlia **comprou** roupas novas.
b. Ontem **comi** três biscoitos diferentes.

Observa-se, nos exemplos, que os complementos de “sugerir” e “pensar” expressam proposições – já que seus *denotata* são avaliáveis como verdadeiros ou falsos –, o que justifica a compatibilidade deles com a CCS. Os complementos dos verbos “comprar” e “comer”, por outro lado, não expressam proposições; em vez disso, designam entidades – o que explica a impossibilidade de serem combinados com a CCS.

Se, por um lado, todos os 157 verbos encontrados se agrupam semanticamente pelo fato de apresentarem complemento proposicional, também é verdade que eles podem ser organizados em classes semânticas mais restritas. Especificamente, é possível identificar, em meio a esse conjunto de predicadores verbais, três subclasses semânticas: verbos de comunicação, verbos de cognição, e verbos de causação. Vejamos:

- (9) Segundo ele, as cervejarias não devem enviar comunicado para a Apas **informando** que querem voltar a negociar, como chegou a solicitar a entidade. (*Nilc/São Carlos*)
- (10) Seria ingenuidade, entretanto, **supor** que as fotos tomadas pelos internos realizam a imagem que eles fazem de si e da instituição. (*Nilc/São Carlos*)
- (11) O que **permite** que um mesmo nome se enlace nos três nós é a possibilidade homonímica de um nome sempre poder ser outro. (*Corpus Brasileiro*)

Vê-se, em (9), um exemplo em que o verbo “informar” está combinado com a CCS. Verbos como esse, assim como “sugerir”, “comunicar”, “avisar”, “propagar” e “confessar”, entre muitos outros, compõem uma das três classes semânticas nas quais é possível distribuir os verbos compatíveis com a CCS: a classe dos verbos de comunicação. Em (10), tem-se o verbo

“supor” como um representante da classe de cognição, juntamente a outros como “entender”, “pensar”, “refletir”, “imaginar”, “especular”, etc. E em (11), por fim, tem-se a combinação da CCS com “permitir”, pertencente à classe dos verbos de causação, junto a outros como “deixar” e “conseguir”.

Até aqui, temos, portanto, o seguinte quadro: a Construção de Complementação Sentencial do PB (i) exige que seu complemento sintático denote uma proposição e (ii) admite verbos pertencentes a três classes semânticas (comunicação, cognição e causação). Essa análise, no entanto, é insuficiente para caracterizar plenamente o fenômeno da complementação sentencial em termos construcionistas, já que não explica a incompatibilidade entre a CCS e uma série de verbos que em princípio deveriam estar aptos a instanciá-la, como “criticar”, “elogiar”, “exaltar”, “analisar”, “reconsiderar”, “alfinetar” e “discorrer”.

- (12) a. ?? João criticou que o jogo acabou tarde.
b. João criticou o fato de que o jogo acabou tarde.
- (13) a. ?? Ela elogiou que a festa foi perfeita.
b. “A festa foi perfeita”, elogiou ela.

Os exemplos (12a) e (13a) sugerem que os verbos “criticar” e “elogiar” não são plenamente aceitáveis na CCS, uma intuição que é corroborada pelo fato de que, dentre todas as instâncias da construção encontradas nos amplos *corpora* analisados, nenhuma delas continha qualquer um desses dois verbos⁶. Ao mesmo tempo, os exemplos (12b) e (13b) mostram que, diferentemente do que ocorre com exemplos como (8a) a (8b), a estranheza não pode ser explicada, neste caso, com base na restrição semântica de que o objeto deve denotar uma proposição. Afinal, os exemplos (b) evidenciam que os verbos “criticar” e “elogiar” apresentam objetos diretos com semântica proposicional – eles apenas não admitem que tais proposições sejam formalmente expressas por meio da moldura sintática da CCS. Nesse sentido, usos do tipo (12a) e (13a) constituiriam *supergeneralizações* indevidas da CCS.

Dado que a restrição a complementos não-proposicionais não explica a impossibilidade (ou menor naturalidade) de tais usos, como dar conta do fato de que o falante do PB tende a evita-los? É dessa questão que tratamos na próxima seção.

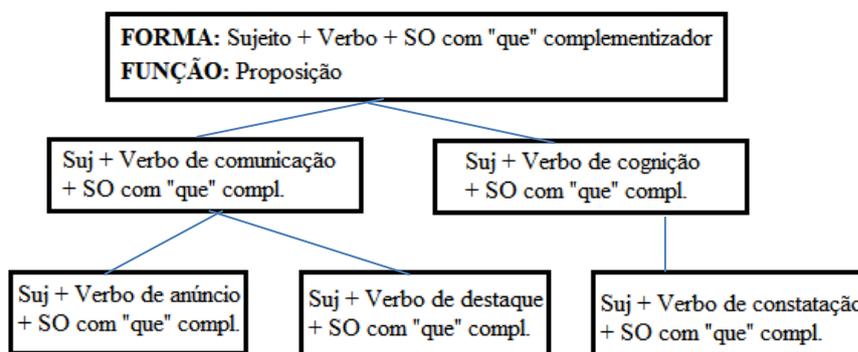
⁶ Certamente, a instanciação de “criticar” e “elogiar” na CCS não produz enunciados fortemente agramaticais. No entanto, tanto as evidências oriundas da análise de *corpus* (relatadas nesta seção) quanto as evidências experimentais (a serem apresentadas na próxima seção) sugerem que tais enunciados tampouco são inteiramente aceitáveis.

Como representar a rede construcional da CCS (a fim de evitar a supergeneralização)?

Conforme exposto acima, pesquisadores alinhados à GCBU apontam que falantes podem evitar supergeneralização utilizando-se de um conhecimento de natureza gramatical e/ou de um conhecimento de natureza estatística – que pode ser do tipo enraizamento e/ou do tipo bloqueio. Aplicando essas hipóteses ao objeto deste estudo, obtemos três possibilidades distintas para o desenho da rede construcional do falante, que representa parte do seu conhecimento acerca da complementação sentencial no PB.

A primeira hipótese é a de que o falante emprega na CCS verbos pertencentes a determinadas classes semânticas ainda mais específicas (que as de comunicação, cognição e causação mencionadas na seção anterior), ao mesmo tempo em que evita verbos de outras classes. Nesse caso, assume-se que a generalização é bloqueada graças a um conhecimento de natureza gramatical – mais especificamente, semântica. À luz dessa hipótese, o conhecimento do falante acerca da CCS pode ser representado em uma rede como a seguinte:

Figura 1 - Rede construcional da complementação sentencial: conhecimento gramatical



Fonte: elaborado pelos autores.

No nível mais alto da rede, tem-se a CCS mais esquemática, com suas especificações de forma e de significado. Em um nível intermediário, temos construções definidas pelas três grandes classes semânticas abrangidas pela CCS⁷. Por fim, em um nível mais baixo, há subconstruções, nas quais estão especificadas classes verbais mais restritivas passíveis de serem combinadas com a CCS. Na rede desenhada acima, destacamos três dessas classes, a título de exemplificação: as classes dos verbos de anúncio, de destaque e de constatação.

Segundo a primeira hipótese, portanto, o falante inconscientemente agruparia, em subclasses semânticas específicas, os verbos passíveis de instanciar a CCS – e, a partir daí,

⁷ Por uma questão de clareza visual, representamos apenas a classe dos verbos de comunicação e de cognição.

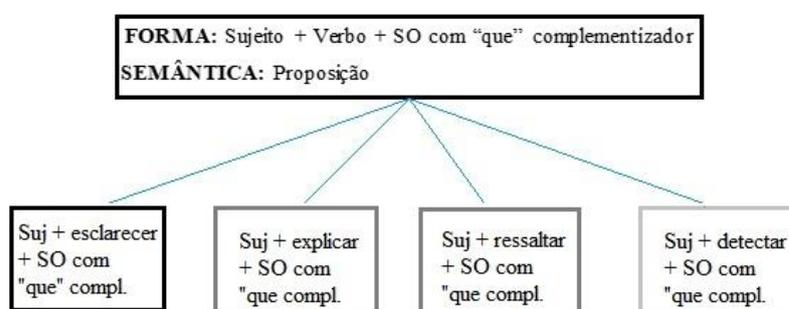
apenas aceitaria na construção verbos pertencentes a essas classes. Nesse sentido, a rejeição a enunciados como (12a) e (13a) resultaria de uma *generalização semântica*.

A segunda hipótese, por seu turno, sugere que o falante emprega na CCS verbos que frequentemente experiencia sendo empregados nela, ao mesmo tempo em que evita aqueles verbos com os quais não tem nenhuma experiência na construção. O falante estaria, portanto, sendo guiado por um *conhecimento estatístico do tipo enraizamento*. Nesse caso, as sentenças exemplificadas em (12a) e (13a) seriam evitadas porque o falante simplesmente não experiencia esses verbos na CCS (como mostra a análise de *corpus* apresentada acima).

Por fim, a terceira hipótese, que se fundamenta em um *conhecimento estatístico do tipo bloqueio*, é a de que o falante usa na CCS verbos que não são frequentemente empregados em uma construção funcionalmente comparável. De acordo com essa ideia, as sentenças em (12a) e (13a) são evitadas porque o falante não experiencia essas formulações específicas e, ao mesmo tempo, experiencia esses verbos em construções alternativas.

Tanto para a segunda quanto para a terceira hipótese, a rede postulada é a seguinte:

Figura 2 - Rede construcional da complementação sentencial: conhecimento estatístico



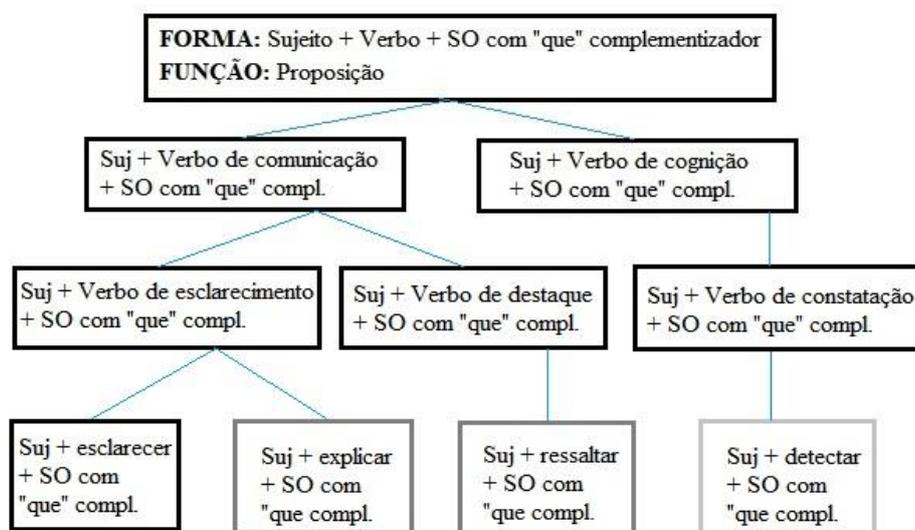
Fonte: elaborado pelos autores.

Nesse caso, no nível mais concreto, estão combinados com a CCS os verbos que o falante experiencia sendo utilizados nela. É importante destacar que as construções de nível mais baixo podem ter graus distintos de enraizamento, conforme representado pelas cores das linhas que delimitam os retângulos (quanto mais escura a linha, maior o grau de enraizamento). Segundo a GCBU, esses graus decorrem da frequência de exposição do falante a cada construção; sendo assim, assume-se que o nível de enraizamento (e, portanto, de acessibilidade cognitiva) de uma construção é diretamente proporcional à sua frequência *token*. A representação acima ilustra essa ideia com os verbos “esclarecer”, “explicar”, “ressaltar” e “detectar”: como, nesse grupo, “esclarecer” é o mais frequente (com 125 ocorrências),

“detectar” é o menos frequente (com apenas uma ocorrência) e “explicar” e “ressaltar” se situam em uma faixa intermediária de frequência (com 20 e 21 ocorrências, respectivamente), assume-se, correspondentemente, que “esclarecer” está mais enraizado que “explicar” e “ressaltar”, os quais, por sua vez, estão mais enraizados que “detectar”.

Há, ainda, a possibilidade de o conhecimento linguístico do falante incluir, a um só tempo, informações de natureza gramatical e estatística. Observe-se:

Figura 3 - Rede construcional da complementação sentencial: conhecimento gramatical + estatístico



Fonte: elaborado pelos autores.

No nível mais abstrato da rede acima, há a CCS geral com suas especificações de forma e significado; um pouco mais abaixo, em um primeiro nível intermediário, estão especificadas construções definidas pelas três grandes classes semânticas associadas à CCS; em seguida, em um segundo nível intermediário, estão presentes cada uma das classes semânticas mais específicas compatíveis com a CCS; e, por fim, em um nível mais concreto, cada um dos verbos que: (i) pertencem às classes indicadas acima (ii) são diretamente experienciados pelo falante na construção. Esses verbos também exibem níveis distintos de enraizamento, conforme indicado, mais uma vez, pelas linhas que delimitam os retângulos (e conforme sugerido pelos valores de frequência *token* apresentados no Apêndice).

Todas as propostas apresentadas até aqui são igualmente compatíveis com os dados dos *corpora*, o que sugere que o método observacional, neste caso, não é suficiente para decidir em relação à representação apropriada do conhecimento subjacente. Diante disso, desenvolvemos

um experimento psicolinguístico com o objetivo de testar empiricamente cada uma das três hipóteses apresentadas ao longo desta seção.

O estudo experimental

Visão geral e previsões experimentais

Foi desenvolvido um experimento de produção induzida, em que os participantes deveriam assistir a uma sequência de vídeos curtos e, na sequência, responder oralmente a uma pergunta interpretativa relacionada ao seu conteúdo (ex: *O que o homem anunciou?*). A resposta deveria se iniciar com uma sequência obrigatória, especificada como parte do estímulo (ex: *O homem anunciou*). Como a resposta podia ou não consistir em uma instância da CCS (dado que a sequência inicial não incluía o complementizador), a variável dependente era o índice de respostas com emprego da CCS (em oposição a outras molduras sintáticas).

Duas variáveis independentes foram controladas: a classe semântica do verbo (presente tanto na pergunta interpretativa quanto na sequência inicial da resposta) e a familiaridade do falante com o verbo. As classes semânticas escolhidas foram as de *verbos de anúncio* (como, por exemplo, “anunciar”) e *verbo de crítica* (como, por exemplo, “criticar”). Essas classes foram selecionadas por representarem, respectivamente, valores semânticos compatíveis e incompatíveis com a CCS⁸.

Em relação ao grau de familiaridade, foram utilizados verbos reais (como “anunciar” e “criticar”) e verbos inventados (como “lomar” – ver próxima subseção). Esse expediente foi utilizado para anular possíveis efeitos de familiaridade. Em outras palavras, a ideia é a de que, se um verbo inventado de anúncio ocorre mais frequentemente na CCS do que um verbo inventado de crítica, essa diferença só pode ser atribuída à classe semântica, dado que a familiaridade dos falantes com ambos os verbos é igual a zero.

Cruzadas as duas possibilidades de cada variável independente, obtêm-se quatro condições experimentais: verbos de anúncio reais, verbos de anúncio inventados, verbos de críticas reais e verbos de crítica inventados. Juntas, elas permitem avaliar três previsões experimentais distintas, cada uma delas associada a uma das hipóteses disponíveis na literatura construcionista para explicar como o falante evita a supergeneralização (conforme descrito na segunda seção). Observe-se:

⁸ A avaliação de (in)compatibilidade decorre tanto do nosso julgamento intuitivo quanto da coleta de dados em *corpora* – como se viu, nenhum dos itens encontrados veicula semântica de crítica, mas muitos expressam ideia de anúncio.

Quadro 1 - Previsões experimentais derivadas de cada uma das três hipóteses

Hipótese	Previsão experimental
Usamos na CCS verbos pertencentes a determinadas classes semânticas, e não a outras (conhecimento gramatical)	Verbos de anúncio inventados terão maior incidência de respostas com a CCS do que verbos de crítica inventados.
Usamos na CCS verbos que frequentemente experienciamos empregados nela (conhecimento estatístico – enraizamento)	Verbos de anúncio reais terão maior incidência de respostas com a CCS do que verbos de anúncio inventados.
Usamos na CCS verbos que não são frequentemente utilizados em uma construção funcionalmente equivalente (conhecimento estatístico – bloqueio)	Verbos de crítica inventados terão maior incidência de respostas com a CCS do que verbos de crítica reais.

Fonte: elaborado pelos autores.

Materiais e estímulos

O experimento foi montado no software Open Sesame e contou com desenho intrassujeitos, de modo que todos os participantes foram expostos à totalidade dos estímulos.

Ao todo, cada sujeito teve acesso a 16 estímulos críticos, sendo quatro para cada condição experimental. A. O quadro abaixo apresenta os verbos utilizados em cada condição:

Quadro 2 - Verbos utilizados conforme condição experimental

	Verbos de anúncio	Verbos de crítica
Verbos reais	anunciar, revelar, informar, alertar	ironizar, criticar, debochar, elogiar
Verbos inventados	lomar, dunaflar, colifar, blumar	ravifar, roguilar, bonirar, tumenar

Fonte: elaborado pelos autores.

Cada estímulo era composto por um vídeo curto – o mais longo deles tinha duração de 34 segundos – seguido de uma pergunta interpretativa. No caso das perguntas com verbos inventados, os sujeitos eram expostos a três vídeos nos quais esses verbos eram empregados (e não apenas um). Isso lhes dava oportunidade de inferir contextualmente o significado dos verbos inventados. Além disso, eles tinham acesso a um glossário, o que permitia ainda a apreensão direta desse mesmo significado.

O significado dos verbos inventados, conforme indicado no glossário a que os participantes tinham acesso, está especificado nos quadros a seguir:

Quadro 3 - Significado dos verbos de anúncio inventados

VERBO	SIGNIFICADO
Lomar	Dizer que fez alguma coisa, mesmo sem ter feito, por estar sob pressão
Dunaflar	Transmitir uma informação antes do tempo previsto.
Colifar	Transmitir uma informação com má vontade.
Blumar	Dizer o mesmo fato várias vezes.

Fonte: elaborado pelos autores.

Quadro 4 - Significado dos verbos de crítica inventados

VERBO	SIGNIFICADO
Ravifar	Expressar opinião depreciativa sobre alguém, de modo que apenas a pessoa afetada perceba.
Roguilar	Expressar uma opinião positiva sobre algo ou alguém, sem que esse seja seu real pensamento.
Bonirar	Pensar nos defeitos de algo ou alguém, mas não verbalizar essa opinião.
Tumenar	Expressar uma informação negativa sobre algo ou alguém, apesar de sua real opinião ser positiva.

Fonte: elaborado pelos autores.

Para além dos 16 vídeos que integraram os estímulos críticos, foram produzidos outros 32 vídeos, utilizados para comporestímulos distratores. Todos os estímulos foram organizados em ordem pseudoaleatória, de maneira que os estímulos críticos de mesma classe semântica e grau de familiaridade não fossem exibidos consecutivamente e de modo que não fosse possível detectar qualquer padrão na apresentação dos estímulos críticos e distratores.

Durante a gravação dos vídeos, tomou-se o cuidado de que os atores não utilizassem em suas falas instâncias da CCS, a fim de evitar a ocorrência de efeitos *depriming* e de saciação sintática. Caso esse cuidado não fosse tomado, o participante poderia ter escolhido responder a uma pergunta relacionada a algum dos verbos utilizando-se da CCS simplesmente pelo fato de ter se deparado, ao assistir ao vídeo, com essa combinação.

Participantes

Participaram voluntariamente do estudo 28 sujeitos (21 homens e sete mulheres), todos falantes nativos do português brasileiro e estudantes universitários cursando o primeiro período de um curso de Letras. Os participantes assinaram termos de consentimento e receberam, em retribuição, horas de Atividades acadêmico-científico-culturais (AACCs).

Procedimentos

Para realizar o experimento, o participante inicialmente recebia explicações orais sobre a tarefa e a dinâmica do teste. Em seguida, tinha acesso, através do software Open Sesame, às

mesmas instruções dadas oralmente, desta vez por escrito, as quais poderiam ser lidas quantas vezes quisessem e por quanto tempo julgassem necessário. Após esse momento, eles eram instruídos a apertar uma tecla que levava à exibição do primeiro vídeo.

Após a exibição de cada vídeo, surgia na tela uma pergunta interpretativa (ex: *O que a menina criticou?*). Logo abaixo, aparecia o início de uma resposta (nesse caso: *A menina criticou...*). Esse artifício foi necessário para que o voluntário fosse levado a utilizar um verbo na resposta. O participante tinha o tempo que julgasse necessário para responder oralmente, e sua fala e tempo de resposta eram gravados. Após a resposta, ele deveria pressionar a tecla indicada, que levava automaticamente à exibição do estímulo seguinte.

Para os estímulos com verbos inventados, antes da exibição do vídeo, era exibido um alerta de que o participante veria uma palavra nova. Em seguida, surgia o verbo inventado seguidos de dois vídeos produzidos com o objetivo de permitir a apreensão contextual do seu significado. Antes de finalmente ser exposto ao terceiro vídeo – aquele que viria acompanhado da pergunta-estímulo interpretativa –, o participante lia a seguinte mensagem: *após o próximo vídeo, você estará pronto para responder uma pergunta*. Junto a essa informação, a definição do verbo era exibida novamente (por exemplo: *Lembre-se: “lomar” = dizer que fez alguma coisa, mesmo sem ter feito, por estar sob pressão*).

O experimento foi rodado com cada um dos 28 sujeitos pelo mesmo pesquisador, o que garantiu que as instruções orais e os procedimentos fossem igualmente explicitados a todos eles. A aplicação era feita individualmente e o pesquisador permanecia na sala junto ao voluntário, durante a leitura das instruções e o início do experimento, para sanar eventuais dúvidas e problemas. Após esse momento, retirava-se do local.

Resultados

Como observado acima, a variável dependente do estudo foi a quantidade de respostas com e sem a moldura sintática da CCS. Lembramos que a tarefa experimental consistia em completar a resposta a uma pergunta interpretativa – por exemplo, para a pergunta “O que a professora alertou?”, eles deveriam completar a frase “A professora alertou...”. Sendo assim, as respostas poderiam conter ou não um complementizador, como nos exemplos abaixo:

- (14) A professora alertou que o tempo estava acabando
- (15) A professora alertou sobre a falta de tempo
- (16) A professora alertou o tempo de se preparar

Em (14), vemos um exemplo de resposta com a moldura sintática da CCS, isto é, com complemento oracional. As sentenças (15) e (16), em contrapartida, exemplificam respostas em que o participante preferiu a CCS em favor de uma estrutura não-oracional (introduzida por preposição, no primeiro caso, e não introduzida por preposição, no segundo).

Após a análise de todas as respostas, elas foram codificadas e classificadas de forma binária – isto é, de um lado, respostas que se constituíam como instâncias da CCS e, de outro, respostas que *não* se constituíam como instâncias da CCS. Nas tabelas a seguir, indicamos o número de enunciados produzidos com e sem a CCS para cada uma das quatro condições experimentais. Começamos pela condição *verbos de anúncio reais*:

Tabela 1 - Resultado das respostas do experimento com verbos de anúncio reais

Verbos de anúncio reais	Número de respostas com a CCS	Número de respostas com outras molduras sintáticas
Alertar	22	6
Informar	27	1
Anunciar	28	0
Revelar	28	0
Total	105	7

Fonte: elaborada pelos autores.

Como se observa, com verbos de anúncio reais, os participantes utilizaram quase exclusivamente a CCS – 105 respostas com essa moldura sintática contra apenas sete com outras estruturas. Entretanto, quando se trata de verbos inventados dessa mesma classe, o resultado é bastante distinto:

Tabela 2 - Resultado das respostas do experimento com verbos de anúncio inventados

Verbos de anúncio inventados	Número de respostas com a CCS	Número de respostas com outras molduras sintáticas
Colifar	6	22
Lomar	20	8
Blumar	13	15
Dunaflar	13	15
Total	52	60

Fonte: elaborada pelos autores.

As formulações escolhidas sugerem que, para verbos de anúncio inventados, a CCS não é privilegiada em relação a outras molduras sintáticas possíveis. Deve-se notar, entretanto, que está sendo comparada a quantidade de uso desses verbos na CCS com a quantidade de uso deles

em todas as outras construções possíveis; por isso, 52 usos de verbos de anúncio inventados com a CCS ainda parece um número elevado.

As respostas referentes aos verbos de crítica podem ser vistas nas tabelas a seguir:

Tabela 3 - Resultado das respostas do experimento com verbos de crítica reais

Verbos de crítica reais	Número de respostas com a CCS	Número de respostas com outras molduras sintáticas
Ironizar	4	24
Criticar	7	21
Elogiar	0	28
Debochar	5	23
Total	16	96

Fonte: elaborada pelos autores.

Tabela 4 - Resultado das respostas do experimento com verbos de crítica inventados.

Verbos de crítica inventados	Número de respostas com a CCS	Número de respostas com outras molduras sintáticas
Tumenar	9	19
Roguilar	7	21
Ravifar	2	26
Bonirar	7	21
Total	25	87

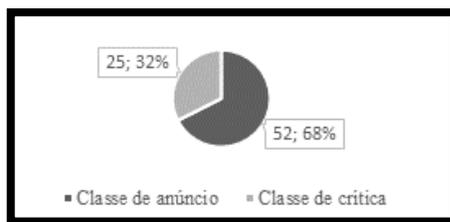
Fonte: elaborada pelos autores.

Os resultados revelam que os participantes não privilegiaram o uso de verbos de crítica na CCS. Isso vale tanto para os verbos reais (16 X 96) quanto para os inventados (25 X 87). Esses números, comparados com os alcançados com os verbos de anúncio, sugerem a atuação de um conhecimento gramatical de natureza semântica. Para avaliar essa possibilidade, assim como a possível atuação de algum mecanismo de natureza estatística, todos os resultados acima foram submetidos a um teste qui-quadrado de homogeneidade.

Verificação da hipótese 1

Como dissemos, para confirmar nossa hipótese 1, seria necessário que verbos de anúncio inventados tivessem maior incidência de respostas com a CCS do que verbos de crítica inventados. Das 224 respostas pertinentes a verbos inventados, 77 são instâncias da CCS. Dessas 77, 52 foram com verbos de anúncio e 25 com verbos de crítica, o que resulta em 67,53% contra 32,46% respectivamente, conforme demonstra o gráfico a seguir:

Gráfico 1 - Uso da CCS com verbos inventados



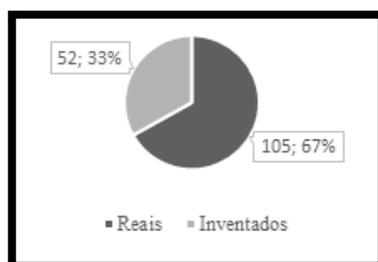
Fonte: elaborado pelos autores.

O teste qui-quadrado de homogeneidade revelou que a diferença numérica entre o total de verbos inventados de anúncio e de crítica é estatisticamente significativa ($p = 0.0021$). Esse resultado confirma, portanto, a atuação do conhecimento gramatical.

Verificação da hipótese 2

Para confirmar a hipótese 2, seria necessário que verbos de anúncio reais tivessem maior incidência de respostas com a CCS do que verbos de anúncio inventados. Os resultados são numericamente compatíveis com a hipótese, como se pode ver no gráfico abaixo:

Gráfico 2 - Uso da CCS com verbos de anúncio



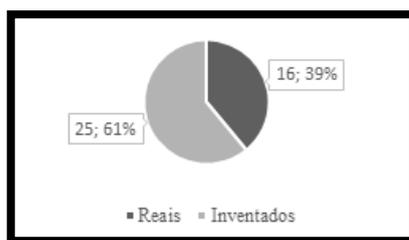
Fonte: elaborado pelos autores.

Também aqui a diferença é estatisticamente significativa ($p\text{-valor} < 0.0001$). Logo, a hipótese de que o processo de enraizamento atua sobre CCS também foi confirmada.

Verificação da hipótese 3

Para confirmar a hipótese 3, seria necessário que verbos de crítica reais tivessem maior incidência de respostas com a CCS do que verbos de crítica inventados. O gráfico com os resultados sugere a confirmação da hipótese:

Gráfico 3 - Uso da CCS com verbos de crítica



Fonte: elaborado pelos autores.

A diferença numérica apresentada acima não alcançou, contudo, significância estatística ($p = 0.15$). Isso nos impede de afirmar que os falantes utilizem um conhecimento estatístico do tipo bloqueio para evitar supergeneralização na CCS.

Discussão

Os resultados apontam para a atuação dos conhecimentos de natureza gramatical e de natureza estatística do tipo enraizamento sobre a produtividade da CCS. Isso significa que o falante evita produzir enunciados como (12a), por exemplo, por duas razões complementares: (i) porque o verbo é incompatível semanticamente com a CCS, isto é, pertencente a uma classe que não está prevista na rede construcional que representa o conhecimento do falante acerca da complementação sentencial no PB, e (ii) porque o verbo não é, na prática, experienciado nessa formulação específica. Dito de outra maneira, os resultados sugerem que o conhecimento do falante acerca da complementação sentencial no PB pode ser representado, em termos construcionistas, em uma rede como a proposta na figura 3.

Por outro lado, não foi possível comprovar estatisticamente a atuação do processo de bloqueio, o que impede a afirmação que alguns verbos não seriam utilizados na CCS devido à experiência que o falante teria com esses mesmos verbos em construções funcionalmente comparáveis. Apesar disso, a diferença numérica que foi verificada (conforme tabelas 3 e 4) sugere que pode haver atuação desse mecanismo. Para verificar essa possibilidade, é necessário que seja realizado um experimento com um número mais elevado de observações.

Palavras finais

Como dissemos introdução deste trabalho, não há consenso na literatura em GCBU quanto ao papel dos dois tipos de conhecimento – o gramatical e o estatístico – na prevenção da supergeneralização. Alguns autores enfatizam o papel do conhecimento estatístico, ao passo que outros chamam atenção para o conhecimento gramatical.

A presente pesquisa contribui para essa discussão ao sugerir que ambos os tipos de conhecimento coexistem e atuam conjuntamente no sentido de impedir que o falante recaia em supergeneralizações em relação às estruturas de complementação sentencial. Em outras palavras, o conhecimento linguístico do falante acerca da CCS envolve conhecimento de natureza gramatical (especificamente, semântica) e também conhecimento de natureza estatística do tipo enraizamento (“item-based knowledge”).

Referências

- AMBRIDGE, B. How do children restrict their linguistic generalizations?: an (un-) grammaticality judgment study. **Cognitive Science**, v. 37, n. 3, p. 508-543, dez. 2013.
- AMBRIDGE, B.; BIDGOOD, A.; TWOMEY, K. E.; PINE, J. M.; ROWLAND, C. F.; FREUDENTHAL, D. Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. **PLoS ONE**, v. 10, n. 4, p. 1-20, abr. 2015.
- AMBRIDGE, B.; PINE, J. M.; ROWLAND, C. F.; YOUNG, C. R. The effect of verb semantic class and verb frequency (entrenchment) on children’s and adults’ graded judgements of argument-structure overgeneralization errors. **Cognition**, v. 106, n. 1, p. 87-129, jan. 2008.
- AMBRIDGE, B.; GOLDBERG, A. E. The island status of clausal complements: Evidence in favor of an information structure explanation. **Cognitive Linguistics**, v. 19, n. 3, p. 349–382, set. 2008.
- BOYD, J. K.; GOLDBERG, A. E. Learning what not to say: The role of statistical preemption and categorization in a-adjective production. **Language**, v. 87, n. 1, p. 55-83, mar.2011.
- DABROWSKA, E. Recycling utterances: A speaker’s guide to sentence processing. **Cognitive Linguistics**, v. 25, n. 4, p. 617-653, out. 2014.
- GOLDBERG, A. E. **Constructions: A construction grammar approach to argument structure**. Chicago: University Press, 1995.
- HILPERT, M. **Construction Grammar and its application to English**. Edinburgh University Press, 2014.
- PEREK, F.; GOLDBERG, A. E. Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. **Cognition**, v. 168, p. 276-293, nov. 2017.
- ROSANEN, S. H. M., AMBRIDGE, B.; PINE, J. M. An elicited-production study of inflectional verb morphology in child Finnish. **Cognitive Science**, v. 40, n. 7, p. 1-35, nov. 2016.
- ROBENALT, C.; GOLDBERG, A. E. Judgment evidence for statistical preemption: It is relatively better to vanish than to disappear a rabbit, but a lifeguard can equally well

backstroke or swim children to shore. **Cognitive Linguistics**, v. 26, n. 3, p. 467-504, jun. 2015.

Sobre os autores

Dayanne de Oliveira Ximenes

Mestra em Linguística pela Universidade Federal do Rio de Janeiro (UFRJ); graduada em Letras - Português/Espanhol pela mesma instituição.

Diogo Pinheiro (Orcid iD: <http://orcid.org/0000-0003-2403-5040>)

Doutor em Linguística pela Universidade Federal do Rio de Janeiro (UFRJ); mestre em Língua Portuguesa pela mesma instituição; graduado em Letras - Português/Latim pela UFRJ. É professor da Faculdade de Letras e do Programa de Pós-Graduação em Linguística da UFRJ.

Recebido em junho de 2020.

Aprovado em setembro de 2020.