

(Re)começando a discutir as locuções verbais

Elvis de Souza, Cláudia Freitas

Departamento de Letras
PUC-Rio – Brasil

elvis.desouza99@gmail.com, claudiafreitas@puc-rio.br

Abstract. *The need to annotate every word of every sentence in a corpus sheds light on grammatical issues not always highlighted in the linguistic description or which discussions do not come to definitive proposals. In this paper, we gather the postulations of [Câmara Júnior 1992], [Vilela and Koch 2001] and [Bagno 2012] on the issue of verbal phrases to assist us in grammar annotation of constructions such as [estar a $V_{\text{infinitive}}$], with an emphasis on the “a”, “de” and “para” particles that can figure in the center of the expressions. Despite some discrepancies between the authors’ observations, we eventually outline a treatment for the morphosyntactic annotation task.*

Resumo. *A necessidade de anotar todas as palavras de todas as sentenças em um corpus joga luz sobre questões gramaticais nem sempre destacadas na descrição linguística ou cujas discussões não chegam a propostas definitivas. Neste trabalho, reunimos as postulações de [Câmara Júnior 1992], [Vilela and Koch 2001] e [Bagno 2012] sobre a questão das locuções verbais para nos auxiliar na anotação gramatical de construções como [começar a $V_{\text{infinitivo}}$], com ênfase nas partículas “a”, “de” e “para” que podem figurar no centro das expressões. A despeito de algumas divergências entre as observações dos autores, esboçamos, no final, um tratamento para a tarefa de anotação morfosintática.*

1. Introdução

A anotação de corpora nos confronta com desafios novos a cada sentença. Assim ocorre no processo de revisão da anotação do corpus Bosque-UD [Rademaker et al. 2017], a versão em Universal Dependencies [Nivre et al. 2016] do corpus Bosque, que é parte integrante do projeto Floresta Sintá(c)tica [Afonso et al. 2002], com textos jornalísticos em português do Brasil e de Portugal.

O Universal Dependencies é um projeto de anotação de treebanks multilíngue, cujo objetivo é facilitar o desenvolvimento de parsers multilíngues. De um ponto de vista linguístico, a ideia é que o modelo possa ser compreendido e utilizado por não linguistas, e por isso as análises propostas se aproximam, em grande medida, daquelas das gramáticas tradicionais. Como em qualquer treebank, um corpus em UD deve conter, entre outros atributos morfosintáticos, uma classificação gramatical (substantivo, verbo, adjetivo etc.) e uma classificação sintática (objeto, sujeito, adjuntos etc.) para todas as palavras em todas as sentenças que compõem o corpus.

Ao longo do processo de revisão da anotação, um dos desafios com que nos deparamos foi o de anotar construções do tipo [estar a $V_{\text{infinitivo}}$]. Além da tarefa de decidir

a classe gramatical e a função sintática dos dois verbos, precisamos realizar a anotação também da partícula “a” entre a forma “estar” e o verbo no infinitivo que a segue. Comparativamente, precisamos decidir se tais ocorrências devem receber um tratamento igual ao de casos como os das assim nomeadas locuções verbais aspectuais – [acabar de $V_{\text{infinitivo}}$] – e as locuções verbais modais – [querer $V_{\text{infinitivo}}$]. No corpus, atualmente, as primeiras são tratadas como um caso de locução (*aux*) e, as segundas, como um caso de subordinação entre orações, em que a segunda é uma oração substantiva objetiva direta reduzida de infinitivo (*xcomp*) da segunda (figuras 1 e 2).

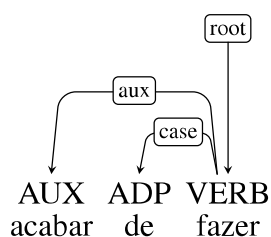


Figura 1. A anotação de *acabar de fazer* no Bosque-UD 2.4

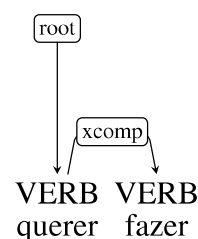


Figura 2. A anotação de *querer fazer* no Bosque-UD 2.4

Na gramática UD, a anotação sintática das partículas “de”, “a” e “para” entre dois verbos pode variar drasticamente caso consideremos que 1) elas fazem parte de uma locução verbal, indicando que os dois verbos estariam sendo unidos pela partícula e as três palavras formam uma unidade, ou 2) elas iniciam uma oração, sendo que o segundo verbo é, via de regra, complemento do primeiro verbo, que não é auxiliar. No caso das locuções verbais de tempo composto – [ter/haver $V_{\text{participio}}$] – há consenso nas diretivas do projeto (e nas gramáticas) de que se trata de uma unidade verbal – e o verbo ter/haver deve ser anotado como de função sintática “auxiliar” e deve ser dependente do segundo verbo, indicando uma unidade verbal. Por outro lado, se construções como “gostar de cantar” devem ser consideradas uma locução ou uma subordinação entre orações não está claro. Especificamente, se a palavra “de” deve ser anotada como parte de uma locução verbal modal, ou como conjunção subordinativa (iniciando uma oração reduzida de infinitivo que complementa o primeiro verbo) é um ponto de discussão.

A fim de sustentar nossa decisão, lançamos mão dos dicionários e gramáticas específicos do português, de tal modo que o corpus seja anotado de maneira linguisticamente embasada, mas sem deixar de manter uma proximidade com abordagens das chamadas gramáticas tradicionais.

Tendo esse objetivo em mente, buscamos em [Câmara Júnior 1992], [Vilela and Koch 2001] e [Bagno 2012] o que se postulou sobre o assunto, dados que sistematizaremos na seção 2, apontando também algumas lacunas e divergências entre os autores. Por fim, na seção 3, pensaremos em algumas soluções para nossas questões a partir do que observamos nas gramáticas e tentaremos encarar a anotação de algumas sentenças complexas de nosso corpus com base nelas.

2. Revisão da literatura

As construções [estar a $V_{\text{infinitivo}}$] estão presentes apenas na parte portuguesa do corpus Bosque-UD, totalizando 80 sentenças, como em (1) e (2). Há também uma sentença do tipo [estar para $V_{\text{infinitivo}}$], na parte brasileira (3).

1. CP3-4 – «**Estamos a dotar** os computadores de um novo sentido» disse Steve d’Averio, director de marketing para a Europa da Logitech.
2. CP285-4 – António Guterres foi o primeiro convidado de uma série de debates com líderes políticos que o Inesc **está a promover**.
3. CF835-12 – Cursando economia na Faap, Kiko espera ansioso o seu telefone celular, que «**está para sair**».

Para dar direcionamento às questões, precisamos procurar referência em diferentes seções das gramáticas, sendo que alguns fenômenos são interpretados e nomeados diferentemente entre os autores. Entre as seções que contemplam o que procuramos, estão *preposição, verbo auxiliar, locução verbal, verbo modal e conjunção*.

2.1. Sobre as locuções verbais

Na *Gramática da Língua Portuguesa* [Vilela and Koch 2001], os autores enquadram as construções [estar a/para $V_{\text{infinitivo}}$] em duas categorias simultaneamente, sem distinção: verbos copulativos e verbos auxiliares de aspecto. Ambas as categorias estão dentro da seção *Verbos plenos e verbos auxiliares*, de tal modo que o primeiro verbo se configura como um verbo auxiliar, e o segundo, pleno. Segundo os autores, portanto, nas construções acima (1)-(3), a forma *estar* é a forma:

“(...) em que o peso gramatical é preponderante, ou porque o verbo se deslexicalizou e reforçou o seu peso gramatical (gramaticalizando-se) e necessita de um verbo pleno para poder funcionar como predicado ou porque o núcleo predicativo é constituído por um nome (*ter consideração por*), por um adjetivo (*ser inteligente*)” ([Vilela and Koch 2001], p. 72).

Por outro lado, o verbo no infinitivo nas construções (1)-(3) é verbo pleno, o que ocorre quando:

“(...) o conteúdo se dirige diretamente para a configuração da processualidade existente no mundo extralinguístico e que gramaticalmente pode funcionar como predicado da frase sem qualquer apoio ou suporte” ([Vilela and Koch 2001], p. 72).

Dentro da categoria de auxiliares de aspecto, esta que contém a construção em foco [estar a/para $V_{\text{infinitivo}}$], há também as construções do tipo [começar/continuar a $V_{\text{infinitivo}}$]. É possível concluir, portanto, que, de um ponto de vista formal, as construções são análogas e devem receber o mesmo tratamento morfossintático. Ainda dentro da seção de verbos auxiliares, Vilela e Koch inserem os verbos auxiliares de tempos compostos [ter/haver $V_{\text{participio}}$] e auxiliares de modo [querer¹ $V_{\text{infinitivo}}$].

Embora as construções com o “estar” sejam mais marcadamente auxiliares, em outras construções, entretanto, julgar se um verbo está mais ou menos gramaticalizado, isto é, se funciona ou não como um verbo auxiliar em um dado contexto não é tarefa

¹São alguns dos poucos verbos inseridos na categoria *auxiliares de modo*, junto com [ter de/que, dever e poder $V_{\text{infinitivo}}$].

simples. Depende, por exemplo, de uma análise que compare o uso/função do verbo em uma sentença com o uso/função, do mesmo verbo, em outros contextos, que evidenciem o chamado sentido pleno desse verbo. A categorização como verbo pleno ou auxiliar pode, portanto, divergir para diferentes anotadores, assim como diverge para diferentes gramáticas.

Vejamos, como exemplo, a sentença (1) abaixo. Nela, o verbo “começar” está, indubitavelmente, exercendo sua função plena: é predicado verbal da sentença cujo sujeito é “A corrida sucessória”. No entanto, nas frases (2) e (3), embora as gramáticas nos digam que as formas de “começar” indicam apenas o aspecto do segundo verbo – aspecto inceptivo –, aceitamos como absolutamente possível uma leitura que assume o sentido pleno de “começar”, assim como em (1). Conseguimos, inclusive, parafrasear (2) e (3) como “começou a coordenação” e “começa a preparação”, respectivamente.

É possível argumentar que a diferença é sintática: em (1) o verbo é intransitivo e, em (2)-(3), possivelmente transitivo. A observação da frase (4), porém, desfaz a tese, pois trata-se de um “começar” transitivo com sentido pleno.

1. CF288-3 – A corrida sucessória **começa** esta semana com um quadro mais claro e definido do que o da semana passada.
2. CF28-1 – Pela segunda vez desde quando **começou** a coordenar as ações no Rio, há duas semanas, o Exército mudou o nome das operações.
3. CF118-2 – O escritório de Júlio Neves já **começa** a preparar novos estudos para o prolongamento deste corredor além do shopping Morumbi, em direção à ponte do Socorro.
4. CF39-2 – Diniz **começou** sua carreira automobilística em 1989, no Brasileiro de Fórmula Ford, campeonato em que obteve a sexta posição na classificação final.

Mesmo considerando que haveria um certo consenso em dizer que [começar a $V_{\text{infinitivo}}$] é uma locução verbal aspectual, Vilela e Koch não deixam claro se a noção de aspecto inclui as partículas “a”, “de” e “para” no escopo da locução verbal aspectual ou se elas não estão auxiliando na noção de aspecto junto ao verbo auxiliar. Em consequência, nós, na tarefa de anotação, precisamos decidir se tais partículas também podem ser consideradas de função auxiliar, e qual sua classe gramatical.

2.2. Sobre as preposições

Sem muito pensar no assunto, costumamos associar as formas das palavras “a”, “de” e “para” à de preposições, que, sendo (ilusoriamente) poucas, decoramos². Suponhamos que se possa dizer que as partículas “a”, “de” e “para”, que aparecem nas construções [comecei a $V_{\text{infinitivo}}$], [acabei de $V_{\text{infinitivo}}$] e [estar para $V_{\text{infinitivo}}$], são preposições. Em seu *Dicionário de Linguística e Gramática*, o autor [Câmara Júnior 1992] considera que preposições são:

“vocábulos que servem de morfema de relação para subordinar um substantivo como: adjunto a outro substantivo ou como complemento a um verbo. Esse processo de subordinação tem o nome de regência” ([Câmara Júnior 1992], p. 198).

²[Bagno 2012] nos alerta para a inadequação da tradição gramatical no tocante às preposições: decoramos, em média, 17, entre as quais poucas ainda são utilizadas atualmente, e deixamos de fora outras tantas que são mais usuais.

No entanto, ao postular que a partícula “a” em [começar a $V_{\text{infinitivo}}$] é preposição, fazemos diferente do que Mattoso definiu, pois no caso das locuções verbais não há um substantivo sendo relacionado, mas dois verbos; ou seja, o segundo verbo, quando muito, seria complemento do primeiro, e preposições não fazem relação entre verbos. Não é à toa que, no verbete “Aspecto”, Mattoso descreve as conjugações perifrásticas³ com “estar” sem nomear a partícula “a”: “o verbo auxiliar *estar*, conjugado com um gerúndio ou um infinitivo regido de *a*” ([Câmara Júnior 1992], p. 61).

Ao lidar com verbos modais, em sua *Gramática pedagógica do português brasileiro*, o autor [Bagno 2012] afirma que os verbos modais são auxiliares e os verbos que os seguem, seu complemento:

“a construção com os verbos modais se faz sempre com infinitivos na posição de verbo principal. Ao mesmo tempo, os verbos principais se constituem o complemento direto do verbo modalizador” ([Bagno 2012], p. 572).

Nesse ponto, a proposta vai ao encontro de Mattoso, pois, fazendo vista grossa e assumindo que preposições podem introduzir complementos que são verbos (orações subordinadas substantivas objetivas indiretas reduzidas de infinitivo), o segundo verbo, nessas construções, é complemento do primeiro.

Essa proposta, no entanto, traz algumas incongruências. Em primeiro lugar, do ponto de vista da anotação no ambiente UD, há uma contradição se levarmos ao cabo a observação de Bagno: não podemos considerar que o segundo verbo em [querer/poder/precisar $V_{\text{infinitivo}}$] é, ao mesmo tempo, complemento do primeiro verbo e verbo principal de uma locução verbal, pois, em UD, verbos auxiliares não podem ter complemento. Se encaramos que o segundo verbo é complemento do primeiro, ambos devem ser plenos.

Em segundo lugar, porque [Câmara Júnior 1992] argumenta, no verbete de conjugações perifrásticas, que:

“É má técnica de descrição gramatical considerar formas perifrásticas a combinação de dois verbos numa única oração em que ambos guardam a sua significação verbal e a significação total é uma das significações (**quero sair - vamos conversando** até a casa - já **tenho** uma carta **escrita**) e não houve a gramaticalização do primeiro verbo” ([Câmara Júnior 1992], p. 80).

Com a afirmação, além de partir do pressuposto de que seja fácil identificar quando um verbo guarda sua significação total – já vimos que não o é –, o autor utiliza o mesmo exemplo prototípico de “locuções verbais modais” – *quero sair* – para dizer que não concorda que se trate de uma locução verbal, na contramão tanto de [Vilela and Koch 2001] quanto de [Bagno 2012], que consideram tais construções como locuções verbais.

3. Desbravando o Bosque-UD

Para lidar com as partículas “a”, “de” e “para” no centro das locuções verbais, nenhuma consulta a gramática nos foi especialmente relevante. No entanto, uma exposição de

³“Conjugações perifrásticas”, como definidas em [Câmara Júnior 1992], têm definição idêntica à de locução verbal com que estamos lidando.

verbos na *Gramática Pedagógica do Português brasileiro* [Bagno 2012] lançou luz sobre um dado que nos parece esclarecedor (Tabela 1).

Tabela 1. Verbos auxiliares (6 primeiras entradas) em [Bagno 2012], p. 604

Verbo auxiliar	Exemplo
acabar	Ana acabou desistindo de viajar em julho.
acabar de	Ana acaba de desistir de viajar em julho.
acabar por	Ana acabou por viajar em julho.
andar	Ana anda pensando em viajar em julho.
cessar de	Ana ainda não cessou de sofrer com a separação.
começar	Ana começou falando dos pais.

A opção que Bagno faz por colocar as partículas “de” e “por” junto ao verbo na primeira coluna, mesmo que sem qualquer comentário sobre essa colocação, nos diz algo. A noção de auxiliaridade, de fato, comparece quando o verbo é acompanhado por tais partículas, evidenciando assim, por exemplo, que “acabar” seria diferente de “acabar de”; “vir” seria diferente de “vir a”, e, do mesmo modo, “começar a” seria diferente de “começar”.

A colocação de partículas (ou preposições) próximas ao verbo nos faz lembrar do conceito de *phrasal verbs* em inglês, quando uma preposição se junta a um verbo, criando uma entrada diferente tanto do verbo de origem quanto da preposição originária. Ainda que tenhamos dificuldade em chamar tais partículas de preposição, indicar que estamos diante de fenômenos semelhantes ao de *phrasal verbs* nos parece adequado. Para lidar com a anotação morfossintática das partículas “a”, “de”, “para” e “por”, portanto, assumiremos que elas são exigidas pelos verbos auxiliares, quando queremos torná-los auxiliares.

Continuando com a análise, no contexto UD, diríamos então que “começar a”, “vir a” e “acabar de” são expressões multi-palavras (MWEs) e, neste caso, a partícula associada se une ao verbo auxiliar pela relação de dependência *compound* (relação usada para os *phrasal verbs* do inglês). Em consequência, temos uma MWE do tipo “verbo auxiliar” e dependente do verbo principal, como anotado no formato UD nas figuras (3)-(7).



Figura 3. CP3-4 – «Estamos a dotar os computadores de um novo sentido» disse Steve d’Averio, director de marketing para a Europa da Logitech.



Figura 4. CF835-12 – Cursando economia na Faap, Kiko espera ansioso o seu telefone celular, que «está para sair».

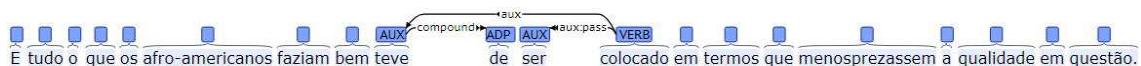


Figura 5. CF27-5 – E, assim, tudo o que os afro-americanos faziam bem *teve de ser colocado* em termos que menosprezassem a qualidade em questão.



Figura 6. CF28-1 – Pela segunda vez desde quando *começou a coordenar* as ações no Rio, há duas semanas, o Exército mudou o nome das operações.



Figura 7. CF118-2 – O escritório de Júlio Neves já *começa a preparar* novos estudos para o prolongamento deste corredor além do shopping Morumbi, em direção a ponte do Socorro.

Como resultado e de forma a contribuir para a discussão, listamos as expressões multi-palavras que indicam aspecto⁴ no corpus Bosque-UD de duas maneiras: na tabela 2, por ordem alfabética, e na tabela 3, por frequência. Este novo tipo de *compound* verbal (o *compound* “phrasal verb”) aparece em 570 ocorrências no Bosque, distribuídas em 31 combinações distintas.

4. Considerações finais

Para nos auxiliar na tarefa de anotar construções como [estar a/para V_{infinitivo}] e [começar a V_{infinitivo}] analisamos duas gramáticas ([Vilela and Koch 2001] [Bagno 2012]) e um dicionário de linguística e gramática ([Câmara Júnior 1992]) da Língua Portuguesa. Elucidativos em alguns pontos, a análise das partículas “a”, “de” e “para” nas locuções verbais não foi explicitamente abordada por nenhuma das obras. Forçados a tomar uma decisão devido à tarefa de anotação, encontramos semelhanças com os chamados *phrasal verbs*, o que nos satisfaz do ponto de vista da análise e, simultaneamente, encontra alinhamento com a decisão do inglês (e talvez de outros treebanks de UD), critério relevante no contexto de um projeto de anotação multilíngue.

O próximo passo é implementar a decisão e investigar se ela traz impactos na consistência interna da anotação, o que pode ser verificado, de maneira indireta, por uma diminuição das confusões entre as relações de auxiliaridade (*aux*) e subordinação (*ccomp/xcomp*), por exemplo. Em caso positivo, teremos bons argumentos não apenas para incluir esta proposta de classificação na discussão sobre locuções verbais, mas também para defender a produtividade de uma descrição linguística que se articula com o PLN.

Agradecimentos

Elvis de Souza é bolsista de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) no projeto *Construção de datasets para o PLN de língua portuguesa*. Número do processo da bolsa: 128693/2019-3.

⁴Lembramos aqui que não interferimos no que foi considerado locução verbal, acatando a decisão original realizada pelo parser PALAVRAS [Bick 2000].

MWE	Frequência
acabar de	11
acabar por	30
andar a	3
chegar a	22
começar a	58
começar por	6
continuar a	57
continuar por	1
deixar de	30
dever a	1
estar a	124
estar para	1
estar por	1
ficar a	4
ficar de	1
haver a	1
haver de	2
haver que	1
ir a	3
ir de	1
parar de	3
passar a	43
poder a	3
ser de	3
tender a	1
ter a	9
ter de	62
ter que	3
tornar a	1
vir a	42
voltar a	42

Tabela 2. Lista das 31 expressões multi-palavras que indicam aspecto no Bosque-UD por ordem alfabética.

MWE	Frequência
estar a	124
ter de	62
começar a	58
continuar a	57
passar a	43
vir a	42
voltar a	42
acabar por	30
deixar de	30
chegar a	22
acabar de	11
ter a	9
começar por	6
ficar a	4
andar a	3
ir a	3
parar de	3
poder a	3
ser de	3
ter que	3
haver de	2
continuar por	1
dever a	1
estar para	1
estar por	1
ficar de	1
haver a	1
haver que	1
ir de	1
tender a	1
tornar a	1

Tabela 3. Lista das 31 expressões multi-palavras que indicam aspecto no Bosque-UD por ordem de frequência.

Referências

- Afonso, S., Bick, E., Haber, R., and Santos, D. (2002). Floresta sintá (c) tica: a treebank for portuguese. In *quot; In Manuel González Rodrigues; Carmen Paz Suarez Araujo (ed) Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)(Las Palmas de Gran Canaria Espanha 29-31 de Maio de 2002) Paris: ELRA. ELRA.*
- Bagno, M. (2012). *Gramática pedagógica do português brasileiro*. Parábola Ed.
- Bick, E. (2000). The parsing system palavras. *Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*.
- Câmara Júnior, J. M. (1992). Dicionário de linguística e gramática: referente à língua portuguesa.
- Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., et al. (2016). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1659–1666.
- Rademaker, A., Chalub, F., Real, L., Freitas, C., Bick, E., and de Paiva, V. (2017). Universal dependencies for portuguese. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pages 197–206.
- Vilela, M. and Koch, I. (2001). Gramática da língua portuguesa: Gramática da palavra, gramática da frase, gramática do texto/discurso. *Coimbra: Almedina*.