

ESTUDO SINTÁTICO-SEMÂNTICO EM GRAMÁTICA FORMAL: A TOPICALIZAÇÃO NO LXGRAM

Cid Ivan da Costa Carvalho (UFERSA/UFC)
cidivanc@gmail.com

Introdução

Dentro dos estudos linguísticos modernos, há duas perspectivas de abordagem das línguas, uma formal e outra funcional. A divergência entre elas se dá pelo fato de que: a primeira estuda as formas linguísticas básicas: fonético-fonológica, morfológica, sintática e semântica; enquanto que a segunda estuda a função que essas formas exercem no uso. Isso equivale dizer que o formalismo analisa e descreve a língua como um sistema e que o funcionalismo observa a língua dentro de um contexto de interação social. Sem desconsiderar a importância da segunda perspectiva para estudos linguísticos, este trabalho mostra o estudo sintático e semântico numa visão formal. Por considerar a língua como “um conjunto formado por sequências resultantes da concatenação de elementos extraídos de um conjunto de símbolos”. (ALENCAR, p. 19, 2011).

Desse conceito de língua formal, depreende-se duas ideias fundamentais: (1) os elementos ocorrem numa sequência não aleatória, ou seja, possuem uma ordem de apresentação no sistema – sintaxe – e (2) cada elemento possui valor ou significado próprio no sistema – léxico. Este, segundo Kenedy (2013), é como um repositório de *informações linguísticas* que dão origem as representações fonológicas e semânticas; aquele, a sintaxe, aplica operações combinatórias dos elementos lexicais de forma recursiva. Assim, cada língua humana possui esses dois componentes linguísticos na produção de sentenças.

Considerando essa relação léxico e sintática na língua portuguesa, classificam-na como uma língua SVO quanto à ordem dos termos numa oração completa, ou seja, a estrutura oracional em que o sujeito surge primeiro, seguido do verbo e por último o objeto. No entanto, há autores que discutem a rigidez com que a classificam, uma vez que se podem observar diferentes ordens dos termos nas frases. De fato, é efetiva e predominante a ordem direta (SVO) na língua, todavia outras relações sintáticas, em que se observam estruturas *topicalizadas* são amplamente usadas, em especial, na modalidade falada.

Assim, este trabalho apresenta algumas análises sintáticas e semânticas de sentenças topicalizadas que foram geradas pela gramática formal LXGram. Para isso, foram seguidos três passos: pesquisa bibliográfica sobre a temática no formalismo linguístico; seleção de sentenças do PORCUFORT – Corpus do Português Oral Culto de Fortaleza – e adaptação; e teste das sentenças simples topicalizadas e não topicalizadas no LXGram. Desse modo, procura-se responder a seguinte pergunta: o LXGram analisa sintática e semanticamente as sentenças simples que contêm objeto direto topicalizado?

O artigo está dividido em cinco subtítulos. O primeiro apresenta os três principais tipos de gramáticas formais e expõe sobre as bases sintática e semânticas do LXGram. O segundo define o conceito de topicalização dentro dos estudos linguísticos e na gramática formal. O terceiro mostra a classificação dos verbos na teoria de valência que, também, fazem parte dos fundamentos sintáticos-semânticos do sistema. O quarto é apresentado o *corpus* de onde foram tiradas as sentenças para observação do sistema. O último subtítulo expõe sobre a análise das sentenças e apresenta as mesmas em forma de árvores sintáticas-semânticas geradas pela LXGram.

1. Formalismo linguístico e o LXGram

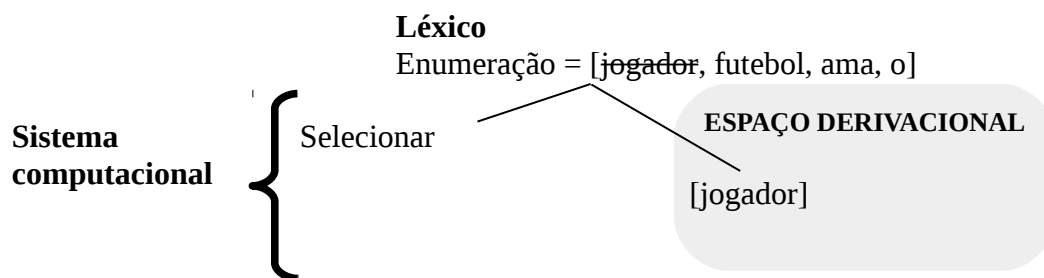
O termo *formal* é empregado na Linguística em oposição ao termo *funcional*. Ambos termos remetem à teorias linguísticas divergentes: o primeiro remete a investigações dos fatores linguísticos estruturais; o segundo está relacionado a estudos do universo pragmático, do uso da língua. Desse modo, distinguem as gramáticas formais das funcionais. A principal gramática formal

é a *Gramática Gerativa*. Porém, além da abordagem gerativa, há propostas que compartilham a concepção de língua como um sistema de representação mental.

Rodrigues e Augusto (2009) apresentam os principais modelos formais de gramáticas. A gramática a proposta pelo modelo gera as estruturas da língua resultando da atuação de um sistema computacional, baseado em dois níveis de representação para uma sentença da língua: Forma Fonética (FF) e Forma Lógica (FL). A gramática proposta pela LFG – *Lexical Functional Grammar* – baseia em restrições e apresenta duas características definidoras: sua orientação fortemente lexicalista e sua orientação funcional, tomando como primitivos teóricos as funções gramaticais. Além disso, ela opera com múltiplos níveis de representação, cada qual com sua própria arquitetura, vocabulário e restrições, os quais são paralelos e ligados por meio de regras de correspondência. Nesse modelo, a língua é uma relação entre forma e função linguística. A gramática proposta pela HPSG – *Head-Driven Phrase Structure Grammar* – contruída numa arquitetura representacional e se pauta numa visão lexicalista, como a LFG, uma vez que lança mão de um léxico altamente rico e complexo nas representações.

A língua formal, no programa minimalista, apresenta o léxico como *informações linguísticas* chamadas de *traços* que variam de língua para língua. Kenedy (2013) afirma que há três tipos de traços linguísticos existentes no léxico: semânticos, fonológicos e formais. Os traços semânticos remetem ao significado, os traços fonológicos referem-se à substância sonora dos itens lexicais e os traços formais dizem respeito a informações que serão acessadas pelo sistema computacional e que não se realizam numa palavra isolada. A língua também apresenta a sintaxe como combinação (concatenação) dos itens lexicais de forma recursiva, resultando em representações infinitamente novas a cada momento da produção. Na linguística forma, essa é a operação de derivação que tem como produto final a representação. Dessa forma, o sistema computacional (sintaxe) realiza duas operações básicas: seleciona e enumera os itens lexicais e Merge. A *enumeração* é um conjunto de itens retirados do léxico e a *seleção* é “a operação que retira um item lexical da enumeração e o introduz no espaço derivacional.” (IDEM, p. 130). A operação *Merge* é responsável por criar objetos sintáticos complexos, como sintagma, orações, sentenças e frases. Veja na figura 1, como ocorre o processo efetuado pelo sistema computacional.

Figura 1: Operações computacionais de enumeração e seleção e de Merge



Fonte: Kenedy (2013, p.130) – adaptado.

Nessas condições, diz-se que a “gramática de uma língua formal é um mecanismo capaz de produzir todas e somente as cadeias pertencentes a essa língua.” (ALENCAR e OTHERO, 2011, p. 22). A língua formal, apresentada na figura 1, é composta de quatro elementos no léxico e, estes, podem ser concatenados por meio de uma operação de nível mais complexo, ou seja, poderá unir dois ou mais elementos para construir uma unidade maior. Assim, a gramática dessa língua, que chamaremos L, seleciona os elementos – $\Sigma = \{\text{jogador, futebol, ama, o}\}$ – que podem ser enumerados em L e concatena-os projetando os elementos nível mais complexo. Então, uma gramática é um dispositivo capaz de gerar uma língua.

A língua formal é apresentada pela LFG sob níveis de representação a estrutura argumental (estrutura-a), a estrutura funcional (estrutura-f) e a estrutura de constituintes (estrutura-c).

A **estrutura-a** codifica informação lexical relativa ao número, tipo sintático e

organização hierárquica dos argumentos de um predicador, tendo uma natureza mista – semântica e sintática.

A **estrutura-f** codifica funções gramaticais como sujeito, objeto, e também traços funcionais como tempo, aspecto, pessoa, número, etc.

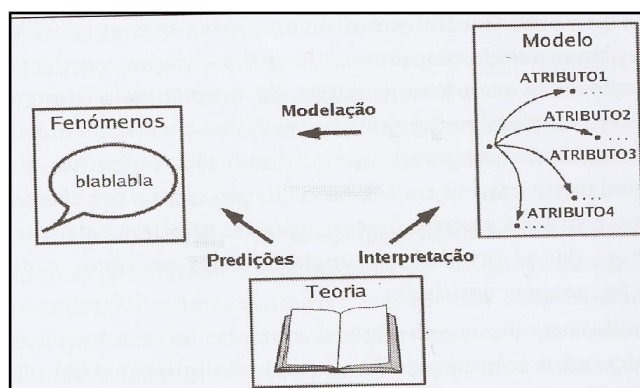
A **estrutura-c** codifica a forma sintática superficial da sentença, incluindo informação de natureza categorial, ordem de palavras e estrutura de constituintes. (RODRIGUES e AUGUSTO, 2009, p. 139, 140).

Essas estruturas gramaticais e os *traços* de gênero, de número, de pessoa, etc. são atribuídos por meio de uma matriz atributo-valor, onde os atributos estão sempre do lado esquerdo e os atributos do lado direito da matriz.

A HPSG – *Head-Driven Phrase Structure Grammar* – é uma estrutura linguística para a qual há uma quantidade substancial de trabalho publicado. Isso permite a implementação direta de análises gramaticais bem conhecidos, que são linguisticamente fundamentados e foram submetidos a escrutínio científico. Ela tem, também, um impacto positivo na capacidade de reutilização e extensibilidade na análise linguística. A literatura dessa gramática formal produziu análises muito precisas de dependências de longa distância, e um ponto forte geral desse formalismo computacional é justamente a implementação desse tipo de fenômeno.

Branco (2011) diz que a HPSG associa representações gramaticais a expressões de linguagem natural, incluindo a representação formal do seu significado. Isso é feito por meio de *estrutura de modelação*, em que a gramática e as entidades são dominados por uma estrutura matemática que serve de modelo para as estruturas. Desse modo, a “estrutura e suas partes estão em correspondência com as entidades observáveis por forma a que estas se encontrem representadas por entidades da estrutura simbólicas que a modela.” (BRANCO, 2011, p. 223).

Figura 2 – Diagrama esquemático das relações entre modelo e teoria



Fonte: Branco (2011, p. 224).

Para representar os observáveis do modelo, são usadas as hierarquias de tipos e as estruturas de traços. O primeiro é “um grafo acíclico em que os nós são etiquetados com etiquetas de tipos”; e o segundo “é um grafo em que os nós são etiquetados com etiquetas de tipos e arcos com etiquetas de tributos”.

Uma propriedade especial da estrutura de traços é que dois caminhos distintos no grafo podem levar a um mesmo nó. Nesta medida, a partilha de estrutura é possível, ou seja, é possível dois caminhos no grafo terem uma mesma estrutura como valor comum. (IDEM, p.225).

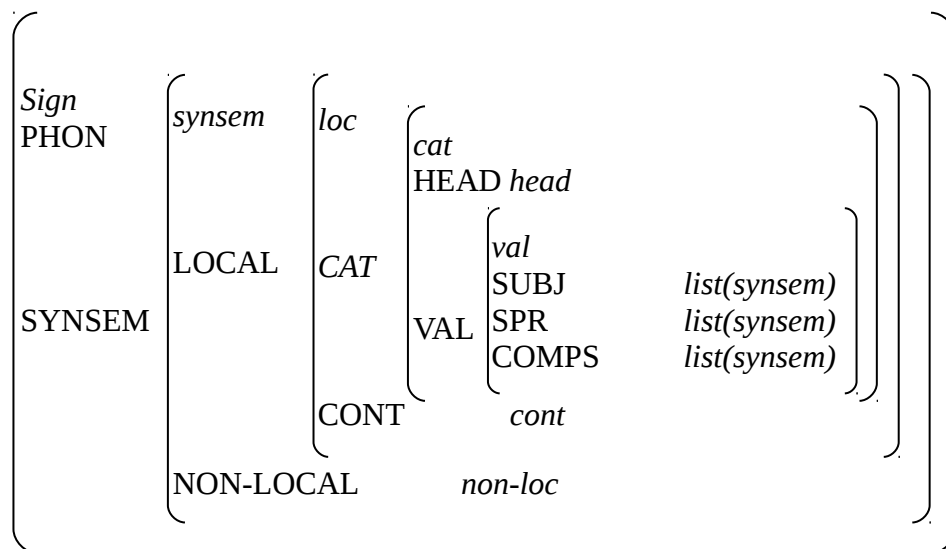
A teoria é constituída de especificações que devem ser interpretadas pelo modelo e que devem está previstas no fenômeno linguístico. Desse modo, as especificações são um conjunto de restrições que estabelecem a predição de que entidade do modelo pertence a entidade da língua natural. Nesse caso, a teoria é uma gramática formal de língua e a gramática é uma conjunção das restrições, que a entidade linguística deve satisfazer.

LXGram é uma gramática formal baseada em codificações linguísticas gerais e suplementadas com um modelo estocástico para análise de resolução de ambiguidade. Além disso, é

um sistema disponível que segue o quadro gramatical de *Head-Driven Structure Grammar* (HPSG), uma das teorias linguísticas mais importantes utilizados no processamento da linguagem natural.

Esse sistema computacional representa as entidades, na **representação sintática**, organizadas num sistema de tipos. Cada tipo está ligado a um conjunto de traços, os quais são pares formados por atributo e valor. Veja na figura 3 como se apresenta a matriz com os tipos e traços.

Figura 3 – Matriz de atributo-valor dos tipos em HPSG



Fonte: Branco (2011, p. 239).

Na figura acima, os atributos do tipo sign são PHON (Phonology) e SYNSEM (sintático e semântico). O primeiro atributo guarda informações fonológica e ortográfica das entidades; o segundo contém informações categorias e semânticas. O atributo SYNSEM do tipo sign tem como valor uma *synsem*, no qual há dois atributos LOCAL E NON-LOCAL. O atributo LOCAL contém informação acerca da sintaxe e da semântica, onde se encontra os valores de CAT (category) e CONT (content), respectivamente. O atributo NON-LOCAL contém informações que acomodam as dependências de longa distâncias. As informações sintáticas estão no atributo CAT que tem por tipo HEAD, que codifica informação sobre classe de palavras e a valência e que guarda informações sobre a valência dos subtraços: SUB (subject), SPR (specifier) e COMPS (complements).

Além dessa representação sintática, o LXGram usa *Minimal Recursion Semantic* (MRS) para a **representação de significado**. Uma representação MRS é uma descrição de um conjunto de possíveis fórmulas lógicas que diferem apenas no âmbito relativa das relações presentes nestas fórmulas. Em outras palavras, ele suporta âmbito sob especificação. Representações semânticas fornecer um nível adicional de abstração, como ordem em que palavra completamente abstrato e linguagem restrições gramatical específicos. Além disso, o formato de MRS de representação semântica que é empregado pelo sistema é bem definida no sentido de que ela é conhecida como mapeamento das representações de MRS e as fórmulas de segunda ordem lógica, para os quais existe um conjunto de interpretação teórica. Devido às limitações de espaço, impossível fornecer um relato detalhado das representações MRS aqui. O capítulo **oito** do livro *Abordagens computacionais da teoria da gramática* fornece uma descrição muito clara dele.

Branco (2010) afirma, também, que o LXGram está em desenvolvimento ativo, mas já suporta uma ampla gama de fenômenos linguísticos, como dependências de longa distância, coordenação, subordinação, modificação e muitos quadros de subcategorização. Além disso, léxico desse sistema contém 25.000 entradas; 64 regras lexicais, 101 regras de sintaxe e cerca de 850 tipos de folhas lexicais, que determinam as propriedades sintáticas e semânticas de entradas lexicais, e 35K linhas de código (excluindo o léxico).

Outra característica desse sistema é o fato de que ele apresenta bom desempenho para o Português Europeu (PE) e Português do Brasil (PB). Ele contém entradas lexicais que são

específicos para qualquer uma delas, e que abrange tanto a nível europeu e sintaxe brasileira. Isso é feito por meio de um mecanismo implementado, a fim de controlar o comportamento gramática sobre variação entre PE e PB, segundo Branco e Costa (2008). Esse mecanismo é usado com os propósitos de: (1) restringir e aceitar apenas as sentenças do PE e o PB, mas não mostra traços específicos de ambas variações e; (2) usar a gramática para detectar qual variedade é usada em algum texto.

Além disso, há um modelo de desambiguação estatística que foi treinado, a fim de selecionar automaticamente a análise mais provável de uma frase quando a gramática produz múltiplas soluções. Este modelo foi treinado a partir de um conjunto de dados composto por 2.000 sentenças de texto de jornal, usando um algoritmo de máxima entropia. As análises linguísticas que são implementadas na gramática são documentadas em um relatório que é atualizado e ampliado com cada versão da gramática.

O LXGram é uma gramática formal implementada sobre as bases da HPSG, modelo de gramática descreve e explica os fenômenos linguísticos. Nesse sentido, o sistema gera sentenças com descrições sintáticas e semânticas formais da língua portuguesa. Porém, esse sistema analisa sintática e semanticamente as sentenças simples que contêm objeto direto topicalizado?

2. Topicalização

No tópico anterior, observou-se que as gramáticas formais são baseadas em traços. Um dos maiores problemas que surge para essas abordagens é a variação na ordem das entidades (palavras) dentro do sistema, ou seja, a topicalização do objeto.

Na perspectiva da gramática gerativa, Perini (2002, p. 212) definiu de grupo de correspondência. Para este autor, o grupo de correspondência é o conjunto de estruturas que se corresponde total ou parcialmente. As relações de correspondência se aplica a estruturas individuais, e não a estruturas generalizadas; as definições, por exemplo, permitem verificar que as sentenças:

- a) Leo comeu o peixinho.
- b) O peixinho, Leo comeu.

São sentenças totalmente correspondentes; porém as definições não informam nada diretamente sobre a correspondência entre estruturas das formas: SVO¹ e OSV, respectivamente. Nos casos de correspondência total, as frases têm sempre as mesmas funções sintáticas, ocupadas pelos mesmos elementos léxicos. Porém, a correspondência parcial pode haver mudanças de funções de certos termos.

Dentre os grupos de correspondência apresentados pelo autor, pode-se apresentar a *topicalização*. Esse autor define a topicalização² da seguinte forma: “Sempre que em uma oração houver um termo marcado pelo [+Ant] (anterioridade), haverá uma oração idêntica a essa, com a diferença de que o termo em questão está no início da frase.” (PERINI, 2002, p. 213). Em português, a ordem em que a sentença apresenta como não marcada está presente no exemplo *a*, ou seja, apresenta a estrutura SVO.

Os constituintes podem movimentar-se na sentença, criando outras estruturas, mas com mesma correspondência. Exemplos:

- b) O peixinho, Leo comeu. [OSV]
- c) Comeu o peixinho, Leo [VOS]
- d) Leo o peixinho comeu. [SOV]
- e) Comeu Leo o peixinho [VSO]

As variações de *b* a *e* na construção sintática da língua são chamadas de ordens marcadas, ou seja, trata-se de variantes posicionais marcadas e restrita a determinados contextos. A movimentação de constituintes no domínio da sintaxe mostra que, um componente linguístico, há

¹ Perini utilizou a seguinte abreviatura a) Suj + NdP + OD; b) OD, + Suj + NdP. Em que Suj corresponde ao sujeito da oração, NdP é o núcleo do predicado e OD é o Objeto Direto da oração. Aqui utilizaremos a estrutura SVO, para corresponder a cada item apresentado pelo autor, respectivamente.

² Neste trabalho, os termos “topicalização”, “marcada” e “deslocamento à esquerda” serão sinônimos. Isso ficará mais claro na análise das sentenças, no quinto tópico.

troca de expressões (movimento físico) e a migração de traço (movimento fictício). A *topicalização* é, portanto, um tipo de movimentação física, ou seja, os constituintes da sentença mudam de lugar. Na sintaxe, o objeto direto movimentar-se para o início da sentença, dando-lhe um realce na interpretação semântica.

No estudo formal, a estrutura informacional da sentença é descrita pela oposição entre *fundo* e *foco*. Nesse caso, as suposições compartilhadas pelo falante e pelo ouvinte no momento físico constituem uma espécie de “pano de fundo” ou informação pressuposta, diante da qual a informação nova, o foco, destaca, veja Gabriel (2011).

Na sentença *b*, por exemplo, o deslocamento à esquerda ocorrido do objeto – O – põe como informação de fundo o sujeito – S – e o verbo – V e o O se eleva a posição de foco o qual preenche o lugar vazio da informação pressuposta. Linguisticamente, o foco chama mais atenção do que o fundo. Gabriel (2011) afirma que o destaque do material focado é realizado ou por meio de construções sintáticas especiais ou por meio da prosódia, ou seja, por meio da colocação do acento sentencial ou nuclear. Considerando os exemplos *a*, *b* e *c* das sentenças acima.

A sentença *a*, em um contexto, tenta responder a pergunta: “quem como o peixinho?” a resposta seria: [LEO] comeu o peixinho, dando destaque ao sujeito da oração. A sentença *b* é uma resposta a pergunta “o que Leo comeu?”, Leo comeu [o PEIXINHO], focando no objeto da sentença. A sentença *c* apresenta como informação nova o predicado, Leo [COMEU O PEIXINHO], pois é uma resposta a pergunta “o que aconteceu?”. Esses exemplos mostram que o destaque recai sobre parte da sentença quando falante responde a perguntas diferentes.

Num quadro formal da LXGram, são atribuídos aos nós individuais traços relativos à estrutura informacional que são condicionados pragmaticamente, veja a figura 3, de forma que se um constituinte recebe o traço de foco (destaque), então se desloca para cima na estrutura arbórea.

3. Valência verbal

A análise do sistema LXGram foi implementada sobre o fundamento da teoria de valência, ou seja, o verbo como suporte de categorias de tempo, modo, número e pessoa que passa-se ao levantamento das propriedades sintáticas-semânticas das sentenças.

Borba (1996, p. 46, 47) afirma que o verbo em português comporta de zero a quatro argumentos. Os verbos impessoais que têm *valência zero* são classificados como *avalentes*. Estes verbos indicam fenômenos meteorológicos ou estado físico como, por exemplo, nevar, chover, trovejar.

Os verbos que exigem apenas um argumento são classificados como *monovalentes*. Estes verbos indicam atividades fisiológicas (tossir, respirar), atividades dos seres (comer, beber), movimento dos seres (ajoelhar-se, acocorar-se) e processos que passam com os seres (sonhar, dormir). Os verbos que selecionam dois argumentos são classificados como *bivalentes*. Os verbos que indicam movimento direcional (ir, vir, chegar), verbos que indicam apreciação (gostar de, apreciar), verbos de percepção (ver, ouvir, sentir) verbos de afetividade (amar, namorar, odiar). Os verbos que exigem três argumentos são chamados de *trivalentes*. Os verbos de atribuição (dar, doar, atribuir), os de declaração (dizer, afirmar, contar), os de união (unir, atar, combinar) são exemplos de verbos de três lugares. Por último, os verbos precisam de quatro argumentos são chamados de *tetravalentes*. Os verbos de transferência de localização (transferir, passar, levar) e os de transposição (traduzir, verter, transpor).

Os verbos bivalentes possuem os traços de movimento físico do objeto direto e apresenta dois argumentos: o argumento externo – o sujeito – e o argumento interno – objeto. As sentenças em que esse tipo de verbo é objeto de estudo desse trabalho.

4. Procedimentos metodológicos

As sentenças analisadas que fundamentam a investigação do fenômeno linguístico da topicalização linguística são provenientes do PORCUFORT – Corpus do Português Oral Culto de

Fortaleza. Esse *corpus* é constituído de três tipos de registros: a Elocução Formal (EF), em que contém textos orais de palestras e aulas; os Diálogos entre Informante e Documentador (DID), que apresenta textos constituídos de entrevistas, e os Diálogos entre Dois Informantes (D2), que são diálogos informais entre dois informantes. Esses registros se distinguem pelo grau de formalidade. O primeiro apresenta o maior grau de formalidade, porque o ambiente em que ocorre exige que o sujeito planeje a fala; o segundo apresenta um nível intermediário de formalidade, uma vez que não intimidade entre o entrevistador e o entrevistado; o último é o menos informal, pois os informantes são familiares e amigos.

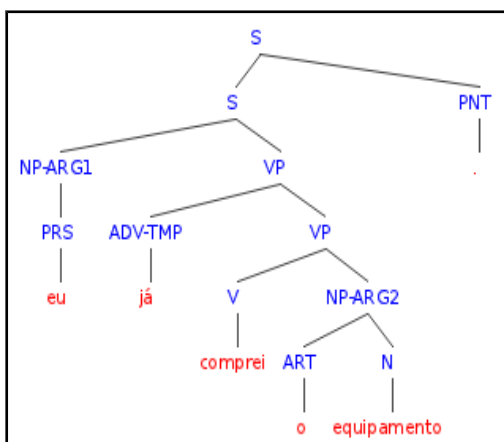
A escolha desse *corpus* se deve ao fato de ser representativo da fala culta de Fortaleza e apresentar sentenças linguísticas marcadas – topicalizadas. Após a leitura, foram coletadas algumas sentenças dos Inquéritos 21 e 54. Por questões de análise no sistema, foram suprimidas as reticências que havia no texto na transcrição original. Feitos esses procedimentos, copiamos uma-a-uma as sentenças para a sistema LXGram, porém foram selecionadas apenas duas sentenças para este trabalho. programa que analisa sintática e semanticamente sentenças do português europeu e brasileiro.

5. Geração e análise das sentenças

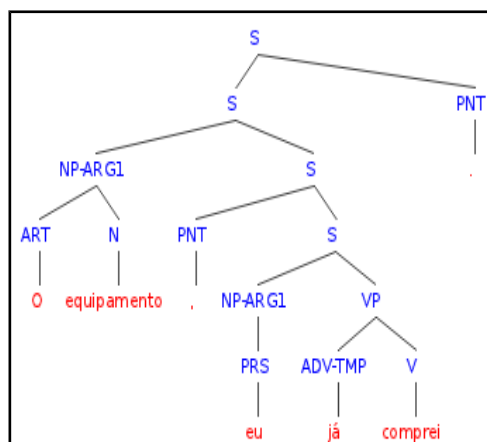
No *output* das sentenças, são apresentadas etiquetas que são ligadas aos *nós* da árvore. Se se verificar de baixo para cima, verá que há etiquetas representativas das **categorias morfossintáticas** – são as etiquetas que estão próximas dos termos, como P (preposição), ART (artigo) e N (nome), etc. - há também etiquetas representativas dos **constituintes imediatos**, como: NP (sintagma nominal - *Noun Phrase*) PP (sintagma preposicional - *Preposition Phrase*) VP (sintagma verbal - *Verb Phrase*) e a há, além disso, as etiquetas representativas dos argumentos, por exemplo: NP-AR G1 (argumento que preenche a condição de sujeito), NP-AR G2 (argumento que preenche a condição de objeto).

No grafo 1, encontra-se a sentença “eu já comprei o equipamento” na forma não marcada. O grafo apresenta uma sentença bem formada do ponto de vista sintático e semântico. No nível sintático, o *nó* raiz liga a dois *nós* NP e VP. As entidades linguísticas foram etiquetados corretamente dentro de sua categoria correspondente, ou seja cada palavra recebeu a etiqueta de sua categoria morfossintática adequada. Os constituintes imediatos foram incluídos dentro do *nó* específico, ou seja, o sintagma que possui aspectos de sujeito e de locução adverbial foram especificados com a etiqueta pertencente ao constituinte a que corresponde, veja o caso do advérbio “já” que foi incluído dentro do VP. No nível semântico, o NP “eu” recebeu a etiqueta de argumento externo, ARG1, e o NP “o equipamento” foi classificado como primeiro argumento interno, ARG2.

Grafo 1 – Sentença não topicalizada



Grafo 2 – Sentença topicalizada

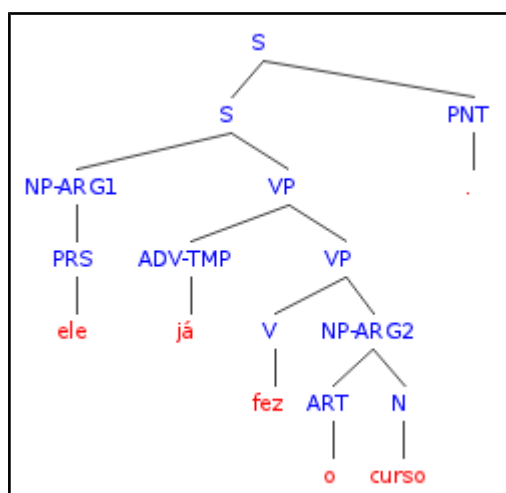


O grafo 2 mostra a sentença do grafo 1 na forma marcada, ou seja, como objeto deslocado à esquerda. No nível sintático, o LXGram gerou uma sentença simples como se fosse uma sentença complexa. Observe que o *nó* raiz está ligado a outra sentença e ao ponto final. A sentença interna ao

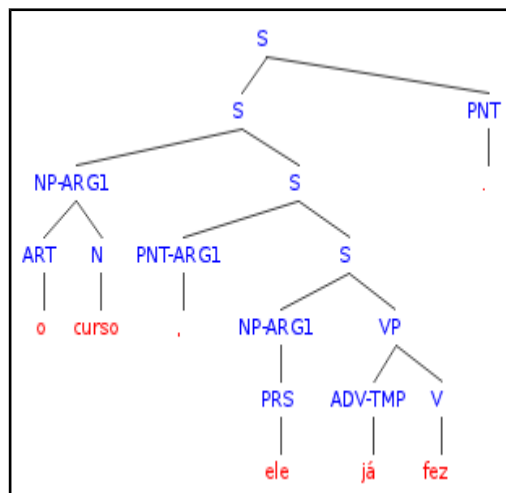
nó raiz está ligado a um NP e a outra sentença. Essa estrutura sentencial é típica de uma oração encaixada. No nível semântico, o argumento interno “o equipamento” perdeu o traço de ARG2 e passou a ARG1 na sentença marcada. Numa oração topicalizada, o que se desloca é a posição do componente e não o traço, então, posição o objeto deveria receber o traço de argumento interno.

A geração das sentenças “ele já fez o curso” e “o curso, ele já fez” estão apresentadas nos grafos 3 e 4, respectivamente. Elas possuem as mesmas estrutura sintático-semântica das sentenças exibidas pelos grafos 1 e 3. O LXGram gerou as mesmas notações no nível sintático e semântico para o par de sentenças marcadas e de sentenças não marcadas.

Grafo 3 - Sentença não topicalizada



Grafo 4 - Sentença topicalizada



As duas sentenças topicalizadas acima possuem uma particularidade, ambas estão com o objeto direto deslocado, logo, são verbos que exigem de dois lugares. Nesses exemplos, o verbo *comprar* e *fazer* não exige dois lugares, mas apenas um lugar, pois se trata de um termo que expressa movimento do ser.

Outra particularidade é que os termos foram etiquetados morfossintaticamente dentro de sua categoria correspondente. Em uma leitura da direita para esquerda, a primeira sentença apresentou uma oração simples como sendo uma oração complexa, com mais de um verbo. Isso pode-se ver na ramificação do nó raiz S em outras. Isso levou a sistema a considerar o objeto direto *o equipamento* e *o curso* como sendo um argumento que preenche a condição de sujeito. Nessa mesma sentença, observa-se que o sujeito da oração *eu* foi etiquetado pelo argumento que lhe convinha, que é o primeiro argumento. Mas, o sistema atribuiu a mesma classificação semântica ao argumento interno – objeto.

No entanto, as sentença topicalizadas foram postas no sistema na ordem natural (ou canônica) da língua, o resultado é que o LXGram conseguiu gerar uma sentença sintático-semântica adequada a língua portuguesa. Observe que ambas apresentam a mesma estrutura – *eu já comprei* e *ele já fez* – essas partes das sentenças receberam os argumentos, as categorias morfossintáticas e os constituintes corretamente. Isso mostra que as sentenças simples não marcadas, ou seja, as sentenças de SVO, são geradas adequadamente tanto no nível sintático quanto semântico.

Conclusão

Com base na análise das sentenças geradas pelo LXGram, o sistema computacional da gramática formal ainda lida com o fenômeno da topicalização com deslocamento de traço e, conseqüentemente, de papel temático. Nas duas sentenças topicalizadas, quando o objeto foi deslocado para a posição marcada, adquiriu o traço semântico de argumento externo (NP-ARG1), porém, deveria receber o de argumento interno (NP-ARG2). O modelo computacional utilizado por essa gramática deveria prever movimentos físicos de seus constituintes sem a perda dos traços semântico. Outro ponto relevante é que a topicalização diz respeito ao realce ou enfoque especial

que se dá a determinado constituinte pelo interesse imediato que ele tem na comunicação.

Houve, também, inadequações no nível da sentença. As sentenças que foram posta para geração eram simples, possuíam apenas um núcleo verbal, e foram geradas como se fossem sentenças complexas.

Outra observação é que houve regularidade na geração das sentenças topicalizadas e não topicalizadas. Para as primeiras, o sistema gerou erros sintáticos e semânticos semelhantes, mostrando que são fáceis de serem impletados; para as segundas, a gramática gerou as sentenças do forma correta.

Referências bibliográficas

ALENCAR, Leonel Figueiredo de. Línguas formais, gramáticas e autômatos no processamento automático das palavras. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila. *Abordagens computacionais da teoria da gramática*. Campinas-SP: Mercado de Letras, 2011.

BRANCO, António e Francisco Costa. A Computational Grammar for Deep Linguistic Processing of Portuguese: LX-Gram, version A.4.1, *Relatório Técnico*, Universidade de Lisboa, Departamento de Informática, 2008. Disponível em: <http://nlxgroup.di.fc.ul.pt/lxgram/LXGramA41ImplementationReport.pdf>.

_____. A Deep Linguistic Processing Grammar for Portuguese. In: Pardo et al. (orgs.), *Computational Processing of Portuguese*, LNAI 6001, Springer, 2010, p.86–89.

BRANCO, António. HPSG: arquitetura; Representação sintática em HPSG; Representação significado em HPSG. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila. *Abordagens computacionais da teoria da gramática*. Campinas-SP: Mercado de Letras, 2011.

BORBA, Francisco da Silva. *Uma gramática de valência para o português*. São Paulo: Editora Ática, 1996.

GABRIEL, Christoph. Questões recentes da teoria minimalista. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila. *Abordagens computacionais da teoria da gramática*. Campinas-SP: Mercado de Letras, 2011.

KENEDY, Eduardo. Arquitetura da linguagem. In: _____. *Curso básico de linguística gerativa*. São Paulo: editora Contexto, 2013.

PERINI, Mário A. *Gramática descritiva do português*.4. ed. São Paulo: Editora Ática, 2002.

RODRIGUES, Érica dos S.; AUGUSTO, Marina R. A.. Modelos formais de gramática: o programa minimalista vs. Gramáticas baseadas em restrições – HPSG e LFG. **Matraga**, v.16, n.24, p. 133-149, jan./jun. 2009.

Site da gramática formalidade

<http://lxcenter.di.fc.ul.pt/services/en/LXSemanticRoleLabeller.html>