



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE HUMANIDADES
DEPARTAMENTO DE LETRAS VERNÁCULAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

FRANCISCO GLEIBERSON DOS SANTOS NOGUEIRA

**VOZ PASSIVA ANALÍTICA: ANÁLISE LÉXICO-FUNCIONAL COM
IMPLEMENTAÇÃO COMPUTACIONAL NO XLE**

FORTALEZA

2019

FRANCISCO GLEIBERSON DOS SANTOS NOGUEIRA

VOZ PASSIVA ANALÍTICA: ANÁLISE LÉXICO-FUNCIONAL COM
IMPLEMENTAÇÃO COMPUTACIONAL NO XLE

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará, como requisito final para obtenção do título de Mestre em Linguística.

Orientador: Prof. Dr. Leonel Figueiredo de Alencar Araripe

Fortaleza

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- D1v DOS SANTOS NOGUEIRA, FRANCISCO GLEIBERSON.
VOZ PASSIVA ANALÍTICA: ANÁLISE LÉXICO-FUNCIONAL COM IMPLEMENTAÇÃO
COMPUTACIONAL NO XLE / FRANCISCO GLEIBERSON DOS SANTOS NOGUEIRA. – 2019.
139 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Linguística, Fortaleza, 2019.
Orientação: Prof. Dr. Leonel Figueiredo de Alencar Araripe.
1. VOZ PASSIVA. 2. GRAMÁTICA LÉXICO-FUNCIONAL. 3. XLE. I. Título.

CDD 410

FRANCISCO GLEIBERSON DOS SANTOS NOGUEIRA

VOZ PASSIVA ANALÍTICA: ANÁLISE LÉXICO-FUNCIONAL COM
IMPLEMENTAÇÃO COMPUTACIONAL NO XLE

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará, como requisito final para obtenção do título de Mestre em Linguística.

Orientador: Prof. Dr. Leonel Figueiredo de Alencar Araripe

Aprovada em: _____/____/____.

BANCA EXAMINADORA

Prof. Dr. Leonel Figueiredo de Alencar Araripe (Orientador)

Universidade Federal do Ceará (UFC)

Prof. Dra. Márcia Nogueira

Universidade Federal do Ceará (UFC)

Prof. Dr. Gabriel de Ávila Othero

Universidade Federal do Rio Grande do Sul (UFRGS)

A Ernando (*in memoriam*), por haver
sonhado comigo este trabalho.

AGRADECIMENTOS

É impossível agradecer a todos que me acompanharam neste percurso, alguns, porém, devem ser nomeados pela sua importância, direta ou indireta, para a realização deste trabalho.

Primeiramente ao professor doutor Leonel Figueiredo de Alencar Araripe, exemplo de pesquisador e linguista, por todas as orientações prestadas desde a época da graduação.

À professora doutora Claudete Lima, pelas generosas contribuições durante a qualificação deste trabalho.

Aos professores participantes da banca examinadora, professora doutora Márcia Nogueira e ao professor doutor Gabriel de Ávila Othero, pela leitura atenta, pelas valiosas críticas, colaborações e sugestões.

Aos colegas do CompLin (Grupo de Pesquisa Computação e Linguagem Natural) Daniel Brasil, Jessé Mourão, Jéssica Oliveira, Juliana Barroso e Juliana Gurgel, pela ajuda em várias etapas deste trabalho.

A todos os meus alunos, pois, de uma forma ou de outra, ajudaram-me a superar os impasses gerados pela pressão e falta de tempo. Dentre eles, dirijo-me de maneira especial a Eric Dantas, a fim de agradecer-lhe pela compreensão e ajuda demonstradas ao longo da minha pesquisa.

Aos amigos, por me acompanharem nesta trajetória, sou especialmente grato a:

Tito Lívio, exemplo de intelectual, professor e tradutor, pelos ensinamentos de ordem acadêmica e pessoal, pela amizade e lealdade nos momentos mais difíceis;

Elvis Freire, pela ajuda com a revisão e formatação deste trabalho, pela presença importante e constante fora dele;

A Matheus Vieira e Liziane Menezes, legados da Universidade para a vida;

Juan Huidobro, pelas infindáveis e interessantes conversas, por ter sido ‘ouvidos’ quando precisei.

A Anderson Moraes, sem cuja ajuda e companheirismo grande parte deste trabalho não teria sido possível.

Ao CNPq, pela concessão de bolsa de pesquisa.

Last but not least, à minha família. Ao meu pai Francimar Nogueira (*in memoriam*), que estaria mais feliz do que eu com essa vitória. Ao meu irmão, Gledson

Nogueira, por ser parte da minha história. À minha mãe, Valdenice Nogueira, cuja assinatura se encontra presente em tudo o que faço, e cujo apoio incondicional me dá forças para não desistir, dedico a ela todas as minhas conquistas.

“Cada língua tem no seu gênio uma força de espontaneidade e seleção, um critério de acerto e um tipo de beleza, que se exercem, ou se enunciam, pela simplicidade e o instinto dos que a falam. É essa intuição da vernaculidade, esse como que sexto sentido, o da linguagem, que parece ter por órgão o ouvido, e do ouvido recebe o nome.”

Rui Barbosa, em *Réplica. A Crítica do Dr. Clóvis Beviláqua*.

RESUMO

Neste trabalho, propomos uma análise linguístico-computacional da voz passiva analítica em português do Brasil (doravante PB), cujo estatuto é debatido por Perini (2010), o qual afirma não haver em PB voz passiva, mas sim uma formação predicativa adjetival. Objetivamos, à luz dos princípios da Gramática Léxico Funcional (LFG) e da Linguística Computacional, verificar qual a classificação mais adequada para o particípio passivo (PART-PASS) em orações passivas perifrásticas. Defendemos a ideia de que, do ponto de vista gerativo, o PART-PASS pertence à classe dos adjetivos e apresentamos e discutimos as evidências morfológicas e sintáticas (cf. ALENCAR, 2016) que corroboram essa conjectura. A fim de testar as vantagens desse tipo de análise, propomos a implementação de duas gramáticas LFG no *Xerox Linguistic Environment* (XLE), que constitui o estado da arte para a implementação de gramáticas em LFG (CROUCH *et al.*, 2011): (i) uma gramática em que o PART-PASS é implementado como A (adjetivo), ou seja, a gramática G-A; e (ii) outra em que em que o PART-PASS é implementado como V (verbo), a gramática G-V. Nossa hipótese básica é a de que a gramática G-A apresenta um *design* mais eficiente. O desenvolvimento das gramáticas deu-se a partir da adaptação de uma gramática LFG do francês (SCHWARZE & ALENCAR, 2016), por meio da qual os fenômenos do português foram implementados de maneira incremental. A testagem de nossas gramáticas é feita mediante a análise de dois conjuntos-teste, o positivo e o negativo, de modo que nosso *parser* deve atribuir uma análise satisfatória a todas as sentenças contidas naquele, mas nenhuma àquelas contidas neste. Os resultados apontam que a gramática G-A apresenta um *design* mais eficiente quando se levam em conta fatores como o tempo de processamento de CPU e de *subtrees* geradas pelas gramáticas (cf. BUTT *et. al.*, 1999).

Palavras-chave: Voz passiva. Gramática Léxico-Funcional. XLE.

ABSTRACT

In this work, we propose a computational-linguistic analysis of the analytic passive voice in Brazilian Portuguese (hereinafter referred to as BP), whose status is discussed by Perini (2010), who stated that there is no passive voice in PB but rather a predicative adjective construction. Guided by the principles of Lexical Functional Grammar (LFG) and Computational Linguistics, we aim to verify the most suitable classification for the passive participle (PART-PASS) in passive periphrastic sentences. From a generative point of view, we defend the idea that PART-PASS belongs to the word class of adjectives and we present and discuss morphological and syntactic evidences (see ALENCAR, 2016) that corroborate this conjecture. In order to test the advantages of this type of analysis, we propose the implementation of two LFG grammars using the Xerox Linguistic Environment (XLE), which is the state of the art for the implementation of grammars in LFG (CROUCH et al., 2011): (a) a grammar in which PART-PASS is implemented as A (adjective), that is, the G-A grammar; and (b) a grammar in which PART-PASS is implemented as V (verb), the G-V grammar. Our basic hypothesis is that the G-A grammar presents a more efficient design. The development of both grammars was based on the adaptation of an existing LFG grammar for the French language (SCHWARZE & ALENCAR, 2016), through which the Portuguese phenomena were implemented incrementally. The testing of our grammars is done by analyzing two test sets, the positive and the negative, so that our parser must assign a satisfactory analysis to all the sentences contained in the first test set, but none in the second test set. The results show that the G- A grammar presents a more efficient design when some factors, such as CPU processing time and subtrees generated by the grammars, are taken into account (see BUTT et al., 1999).

Keywords: Passive voice. Lexical Functional Grammar. XLE.

ZUSAMMENFASSUNG

In der vorliegenden Arbeit schlagen wir eine computerlinguistische Untersuchung des analytischen Passivs im brasilianischen Portugiesisch (von nun an BP) vor; laut Perini (2010), der den Status dieser Passivform unter die Lupe nahm, gibt es im PB kein Passiv, sondern eher eine prädikative Adjektivkonstruktion. Basierend auf den Prinzipien der Lexikalisch-Funktionalen Grammatik (LFG) und der Computerlinguistik beabsichtigen wir herauszufinden, welche die adäquateste Klassifikation für das Partizip Perfekt Passiv (PART-PASS) in passiven periphrastischen Sätzen ist. Aus generativer Sicht vertreten wir die Annahme, dass das PART-PASS zur Wortklasse der Adjektive gehört, was wir nachstehend anhand von morphologischen und syntaktischen Beweisstücken (siehe ALENCAR, 2016) untermauern. Um die Vorteile einer solchen LFG- basierten Analyse zu testen, schlagen wir somit die Implementierung von zwei LFG-Grammatiken unter Verwendung der XLE-Software vor, die dem Stand der Technik für die Implementierung von LFG-Grammatiken entspricht (CROUCH et al., 2011): (a) eine Grammatik, in der PART-PASS als A (Adjektiv) implementiert ist, die sogenannte G-A-Grammatik; und (b) eine Grammatik, in der PART-PASS als V (Verb) implementiert ist, d. h. die G-V-Grammatik. Wir gehen von der Grundhypothese aus, dass die G-A-Grammatik ein effizienteres Implementierungsdesign bietet. Die Entwicklung beider Grammatiken basierte auf der Anpassung einer bereits bestehenden LFG- Grammatik für die französische Sprache (SCHWARZE & ALENCAR, 2016), durch die die portugiesischen Phänomene möglichst inkrementell umgesetzt wurden. Das Testen unserer Grammatiken erfolgt durch die Untersuchung eines positiven und eines negativen Testsatzes, so dass unser Parser allen Sätzen, die in dem ersten Testsatz enthalten sind, eine zufriedenstellende Analyse zuordnen muss, aber keiner in dem zweiten. Die Ergebnisse zeigen, dass die G-A- Grammatik ein effizienteres Design bietet, unter der Voraussetzung, dass einige Faktoren wie etwa die CPU-Prozessierungszeit und von den Grammatiken erzeugte *Subtrees* berücksichtigt werden (siehe BUTT et al., 1999).

Schlüsselwörter: Passiv. Lexikalisch-funktionale Grammatik. XLE.

RESUMEN

En este trabajo proponemos un análisis lingüístico-computacional de la voz pasiva analítica en portugués de Brasil (de ahora en adelante PB), cuyo estatus es debatido por Perini (2010), quien afirma que no hay voz pasiva en PB, sino una formación predicativa adjetival. A la luz de los principios de la Gramática Léxico Funcional (LFG) y de la Lingüística Computacional, tratamos de comprobar cuál es la clasificación más apropiada para el participio pasivo (PART-PASS) en oraciones pasivas perifrásticas. Defendemos la idea de que, desde el punto de vista generativo, el PART-PASS pertenece a la clase de los adjetivos y presentamos y discutimos las evidencias morfológicas y sintácticas (cf. ALENCAR, 2016) que corroboran esta conjetura. Para comprobar las ventajas de este tipo de análisis, proponemos la implementación de dos gramáticas LFG en el *Xerox Linguistic Environment* (XLE), que constituye el estado del arte para la implementación de gramáticas en LFG (CROUCH et al., 2011): (i) una gramática en la que el PART-PASS se implementa como A (adjetivo), es decir, la gramática G-A; y (ii) otra en la que el PART-PASS se implementa como V (verbo), la gramática G-V. Nuestra hipótesis básica es que la gramática G-A presenta un *design* más eficiente. El desarrollo de las gramáticas se produjo a partir de la adaptación de una gramática LFG del francés (SCHWARZE & ALENCAR, 2016), a través de la cual los fenómenos del portugués fueron implementados de manera incremental. El examen de nuestras gramáticas se realiza mediante el análisis de dos conjuntos-test, el positivo y el negativo, de modo que nuestro *parser* debe atribuir un análisis satisfactorio a todas las sentencias contenidas en aquel, pero ninguna a aquellas contenidas en este. Los resultados muestran que la gramática G-A presenta un *design* más eficiente cuando se tienen en cuenta factores como el tiempo de procesamiento de la CPU y los *subtrees* generados por las gramáticas (cf. BUTT et. al., 1999).

Palabras clave: Voz pasiva. Gramática Léxico-Funcional. XLE.

LISTA DE TABELAS

Tabela 1 – Desvio padrão relativo de G-V	125
Tabela 2 – Desvio padrão relativo de G-A	125
Tabela 3 – Análise de variância com fator duplo	126
Tabela 4 – Quantidade de subtrees	127

LISTA DE QUADROS

Quadro 1 – Quadro comparativo entre LFG e modelos gerativos transformacionais... 57

LISTA DE FIGURAS

Figura 1 - Representação dos verbos monoargumentais.....	42
Figura 2 - Estrutura de superfície de um verbo inacusativo com AI deslocado.....	43
Figura 3 - Estrutura de superfície de verbo inacusativo com AI <i>in si</i>	44
Figura 4 - Estrutura de superfície de uma sentença passiva.....	49
Figura 5 - Estrutura de superfície de uma sentença passiva analisada como predicativa adjetiva.....	50
Figura 6 - Estrutura sintagmática da sentença <i>The two small children are chasing that dog</i>	54
Figura 7 - Estrutura sintagmática de uma sentença em warlpiri.....	55
Figura 8 - Análise IC do exemplo (i).....	58
Figura 9 - Estrutura sintagmática do exemplo (i).....	58
Figura 10 - Estrutura de valência do exemplo (i).....	60
Figura 11 - Análise da sentença (iii) por um fragmento de gramática em XLE....	62
Figura 12 - Entrada lexical de <i>procura</i>	64
Figura 13 - Estrutura F da sentença <i>O cavaleiro procura a fada</i>	65
Figura 14 - Estrutura C anotada 1.....	68
Figura 15 - Estrutura C anotada 2.....	69
Figura 16 - Estrutura C e estrutura F da sentença <i>A fada é procurada por um cavaleiro</i> geradas no XLE a partir da gramática G-V.....	73
Figura 17 - Estrutura C e estrutura F do exemplo teste negativo <i>*A fada é esperado</i>	85
Figura 18 - Estrutura F do exemplo teste negativo “A fada é esperado”.....	85
Figura 19 - Estrutura C de sentença com verbo <i>ser</i> em função de cópula.....	95
Figura 20 - Estrutura F de sentença com verbo <i>ser</i> em função de cópula.....	95
Figura 21 - Estrutura C da sentença (1).....	98
Figura 22 - Estrutura F da sentença (1).....	98
Figura 23 - Estrutura C da sentença (7).....	101
Figura 24 - Estrutura F da sentença (7).....	101
Figura 25 - Estrutura C da sentença (9).....	105
Figura 26 - Estrutura F da sentença (9).....	105
Figura 27 - Estrutura F da sentença (3) gerada pela G-A.....	112
Figura 28 - Estrutura F da sentença (3) gerada pela G-V2.....	112

Figura 29 - Estrutura C da sentença (5).....	114
Figura 30 - Estrutura F da sentença (5).....	114

LISTA DE ABREVIATURAS E SIGLAS

1	Primeira pessoa
2	Segunda pessoa
3	Terceira pessoa
A	Adjetivo
ACC	Acusativo
ADJ	Adjunto
ADJ-PP	Sintagma preposicional com função de adjunto
ADV	Advérbio
ADVP	Sintagma adverbial
AE	Argumento externo
AI	Argumento interno
AP	Sintagma adjetival
ATTRIB	Atributivo
ATYPE	Tipo de adjetivo
AUX	Auxiliar
AVM	Matriz de Atributo e Valor
C	Complementador
CASE	Caso
CFORM	Forma da conjunção
CLAUSE_TYPE	Tipo de sentença
COMP	Complemento
COMPL-PP	Complemento PP
CP	Sintagma complementador
CTRL	Controle

D	Determinante
DECLAR	Declarativa
DEF	Definido
DEG	Grau
DITRANS	Ditransitivo
DO	Objeto direto
DO-CTRL	Controle do objeto direto
DP	Sintagma determinante
<i>e</i>	<i>Épsilon</i>
FEM	Feminino
G-A	Gramática do adjetivo
G-ADV	Gramática do advérbio
GEN	Gênero
GerundP	Sintagma de gerúndio
Impf	Imperfeito
IND	Indicativo
INDEF	Indefinido
INF	Infinitivo
InfP	Sintagma infinitivo
INTERR	Interrogativo
INTRANS	Intransitivo
IO	Objeto indireto
IO-CTRL	Controle de objeto indireto
IP	Sintagma flexional
LC	Linguística Computacional

LFG	<i>Lexical-Functional Grammar</i>
M	Masculino
MOOD	Modo
N	Nome
NOM	Nominativo
NP	Sintagma nominal
NUM	Número
OBJ	Objeto
OBJ2	Segundo objeto
OBL	Oblíquo
OBL-PP	Sintagma preposicional oblíquo
OPT-DITRANS	Ditransitividade opcional
OPT-TRANS	Transitividade opcional
P	Preposição
PA	Preposição e artigo
PART	Particípio
PART_PASS	Particípio passivo
PART_PAST	Particípio passado
PASS-TRANS	Passiva transitiva
PB	Português brasileiro
PERS	Pessoa
PL	Plural
POSTNOM	Pós-nominal
PP	Sintagma preposicional
PRED	Predicado

PRED-AP	Sintagma adjetival predicativo
PREDIC	Predicativo
PRENOM	Pré-nominal
PRES	Presente
PRON	Pronome
PRON_TYPE	Tipo de pronome
PROP	Proposição
PRS	Presente
PSG	Gramática de Estrutura Sintagmática
PUNCT	Ponto
S	Sentença
SC	Small clause (mini-oracão)
SBJV	Subjuntivo
S-CTRL	Controle de S
SG	Singular
SN	Sintagma nominal
SPEC	Especificador
SUBJ	Sujeito
SV	Sintagma verbal
TENSE	Tempo
TRANS	Transitivo
UG	Gramática Universal
V'	V-barras
V-AGR	Concordância verbal
VFORM	Forma do verbo

VP	Sintagma verbal
Vtr	Verbo transitivo
XCOMP	Complemento predicativo
XLE	<i>Xerox Linguistic Environment</i>

SUMÁRIO

1	INTRODUÇÃO.....	24
1.2	OBJETIVOS.....	33
1.2.1	<i>Objetivo Geral</i>	33
1.2.2	<i>Objetivos Específicos</i>	33
1.3	Problemas da Pesquisa.....	33
1.4	Hipóteses.....	34
1.4.1	<i>Hipótese Básica</i>	34
1.4.2	<i>Hipóteses Secundárias</i>	34
2	UM PANORAMA SOBRE A VOZ PASSIVA.....	35
2.1	A Gramática de Port-Royal e as vozes verbais.....	35
2.2	A Gramática de Napoleão Mendes de Almeida e as vozes verbaus.....	38
2.3	A Gramática de Celso Cunha e Lindley Cintra e as vozes verbaus.....	39
2.4	A passiva na Gramática Gerativa.....	41
2.4.1	<i>Os verbos inacusativos</i>	42
2.4.2	<i>Sobre a passiva na GT</i>	45
2.4.3	<i>Sobre a aquisição das sentenças passivas</i>	50
3	DOS FUNDAMENTOS TEÓRICOS DA LFG.....	52
3.1	A gramática de estrutura sintagmática (PSG).....	57
3.2	A teoria de valência.....	60
3.3	A teoria das funções gramaticais.....	61
3.3.1	<i>Esquemas funcionais</i>	66
3.4	A análise da passiva na LFG.....	71
3.5	Da passiva sintética.....	74
3.6	Evidências morfológicas e sintáticas para a classificação do PART-PASS.....	76
4	METODOLOGIA.....	80
4.1	Do <i>corpus</i> da pesquisa.....	80
4.1.1	<i>Da construção do conjunto-teste e das gramáticas computacionais em LF</i>	83
4.1.2	<i>Da criação do teste negativo</i>	84

4.1.3	<i>Do material utilizado</i>	88
4.1.4	<i>Do procedimento de análise</i>	88
4.2	Da criação da gramática	89
5	A IMPLEMENTAÇÃO LINGUÍSTICO COMPUTACIONAL DA PASSIVA EM UMA GRAMÁTICA LFG/XLE	91
5.1	A Gramática G-V	92
5.1.1	<i>Implementação da regra da passiva</i>	106
5.2	A Gramática G-A	111
6	ANÁLISE DOS RESULTADOS	120
6.1	Testagem	120
7	CONCLUSÃO	128
	REFERÊNCIAS	130
	APÊNDICE A - Arquivo de teste positivo testfile.lfg	135
	APÊNDICE B - Arquivo de teste negativo negative-test.lfg	137

1 INTRODUÇÃO

O presente trabalho analisa o *status* da voz passiva em PB à luz de uma abordagem gerativista. Pretendemos, tomando por base a proposta de Perini (2010), analisar a hipótese de ser o Português Brasileiro (doravante PB) uma língua que não apresenta voz passiva¹. Segundo o autor, aquilo que convencionalmente é classificado como voz passiva e apresentado em gramáticas tradicionais, como as de Cegalla (1967), Napoleão de Almeida (1980), Rocha Lima (1996), Cunha e Cintra (1985), Luft (2002) e Bechara (2005), como parte integrante do quadro de conjugação verbal², consistiria, na verdade, em uma construção predicativa adjetival.

O estudo das vozes verbais já é *per se* um imperativo dentro dos estudos linguísticos, uma vez que trabalhos há que apontam as falhas teóricas existentes na conceituação das diferentes vozes, as quais, *grosso modo*, são apresentadas de maneira vaga e imprecisa (D'AVILLA, 2006). O problema de classificação e análise parece residir no fato de os termos usados pela gramática tradicional calcarem-se na terminologia greco-latina (COSERIU, 1987).

O latim e o grego antigo apresentavam uma forma verbal marcada morfologicamente para a voz passiva, cuja permanência em línguas românicas modernas tais como o português, o espanhol, o italiano e o francês não se evidencia. O que há em comum a este grupo de línguas, no que concerne à voz passiva, é uma formação perifrástica de um verbo auxiliar (geralmente *ser* ou *estar*)³ acompanhada da forma infinita (particípio) de outro verbo. Na formação composta a que nos referimos aqui, cujo modelo é esboçado nos exemplos abaixo, retirados de Wandruszka (1969, p. 427), é o verbo auxiliar, em sua forma finita, que recebe as informações referentes a tempo, modo, número e pessoa. Os exemplos são cotejados por Wandruszka a partir de excertos da Bíblia, especificamente, de versículos de Lucas (Lucas 6, 37-38)⁴.

¹ Perini (2010, p.174-179) traz a discussão apenas em torno das construções chamadas de passiva analítica. Sobre a chamada passiva sintética ou pronominal, o autor não tece considerações. No que diz respeito à passiva pronominal, veja-se a Fundamentação Teórica deste trabalho. Observe-se que há uma edição mais recente do mesmo livro de Perini publicada em 2016; todavia, por não haver mudanças de conteúdo em relação à edição de 2010, usamos a obra mais antiga como base para nossas referências.

² Em Cunha e Cintra (1985, p.372), por exemplo, a categoria de voz é apresentada como uma variação verbal, o que equivale dizer que aquela seria uma forma verbal.

³ Nesta dissertação, serão objeto de análise e tratamento linguístico-computacional apenas as chamadas passivas verbais, ou seja, construções que se fazem com o verbo *ser* e particípio. As construções com os verbos *estar/ficar* e particípio, as passivas adjetivais, não foram por nós trabalhadas. O tratamento desse tipo de construções, a qual envolve outras complexidades, poderá ser objeto de estudos futuros.

⁴ Os negritos são nossos.

- (1) Latim: *Nolite iudicare, et non iudicabimini ... Date, et dabitur vobis.*
 Alemão: *Richtet nicht, so werdet ihr auch nicht gerichtet ... Gebet, so wird euch gegeben*
 Inglês: *Judge not, and you will not be judged ... Give, and it will be given to you.*
 Francês: *Ne jugez pas, et vous ne serez pas jugés ... Donnez, et l'on vous donnera.*
 Italiano: *Non giudicate e non sarete giudicati ... Date e vi sarà dato.*
 Espanhol: *No juzguéis y no seréis juzgados ... Dad y se os⁵ dará.*
 Português: *Não julgueis, e não sereis julgados ... Dai, e dar-se-vos-á.*

Analisando-se os exemplos acima, é possível notar que, enquanto o latim apresenta uma forma verbal com marcação morfológica que expressa a voz passiva⁶, as demais línguas utilizam-se de construções perifrásticas com o verbo auxiliar. Wandruszka (*id. ib.*) apresenta também exemplos em que aparece o “causador” [agente, *Verursacher*]⁷:

- (2) Inglês: *I have been insulted by everyone here.*
 Alemão: *Ich bin hier von jedermann beleidigt worden.*
 Francês: *J'ai été insultée par tout le monde ici.*
 Italiano: *Qui io sono stata insultata da tutti.*
 Espanhol: *He sido insultada por todos los presentes.*
 Português: *Já fui insultada por todos aqui.*

Normalmente, baseando-se em um senso comum, espera-se que o PB, por ser uma língua proveniente da matriz neolatina e por também ter recebido forte influência do grego antigo, tenha os mesmos parâmetros que as outras línguas românicas elencadas e examinadas por Wandruszka. Para nos acercarmos melhor desta questão, tomemos, p. ex., o trabalho de Tarallo e Kato (1989) acerca da variação da ordem VS (verbo-sujeito) em PB, em que os autores asseveram que as línguas podem

⁵ Os exemplos também trazem, no espanhol e no português, a presença da passiva sintética; esta, porém, como já mencionado, não será objeto específico da pesquisa.

⁶ Chamamos a atenção para o fato de que mesmo no latim clássico, o uso das formas sintéticas da passiva estavam restritos aos tempos do *infectum*. Nos tempos do *perfectum*, o latim clássico usava formas analíticas compostas por participio passado passivo e o auxiliar *esse*.

⁷ Wandruszka retira os exemplos e as respectivas traduções da obra *Arms and the Man*, de Bernard Shaw.

diferir qualitativamente desde que, em relação a uma dada propriedade paramétrica, estejam marcadas positiva ou negativamente, ou seja, que uma dada língua apresente ou não um dado parâmetro. Sendo assim, o PB, ainda que seja uma língua neolatina, pode diferir, e, de fato, é que o nos parece, no tocante à voz passiva do latim e grego antigos.

A escolha pelo tema da pesquisa parece-nos oportuna, haja vista que, embora tradicionalmente a existência da voz passiva tenha sido um fato quase inquestionável na maioria das gramáticas de abordagem mais conservadora, o assunto não é consenso entre linguistas, mostrando-se não trivial. Discussões acerca da existência ou não de voz passiva em línguas românicas podem ser encontradas desde a década de 1970. Câmara Jr. (1979), por exemplo, põe em xeque a existência da voz passiva em português, por essa não gozar de caracterização morfológica.

Para a abordagem gerativista, adotada neste trabalho, as construções em voz passiva constituem um tema relevante e profícuo para pesquisas. Em muitos trabalhos consultados, para línguas diversas, o fenômeno da voz passiva é tratado como sendo formado por verbo auxiliar mais particípio passado (SCHWARZE & ALENCAR (2016), LUNGUINHO (2011), SIMIONI (2010), MIOTO *et al.* (2007), CARNIE (2002)).

Não obstante, a partir dos avanços nos estudos linguísticos, sobretudo dentro do quadro teórico da Gramática Léxico-Funcional (LFG), trabalhos há em que ressalvas são feitas quanto ao caráter desse particípio, sugerindo-se a possibilidade de tratar-se de um adjetivo (citem-se os trabalhos de Alencar (2015a) e Santos (2014) para o PB; Patejuk e Przepiórkowski (2014) para o polonês).

A presente pesquisa tem como ponto desencadeador uma lacuna teórica apresentada na *Gramática do Português Brasileiro*, de Perini (2010), para o qual, em PB, o verbo *ser*, em construções passivas, não é visto como mero auxiliar, senão como o próprio portador da diátese⁸ verbal, isto é, da voz verbal. Perini (*op. cit.*) postula que o chamado particípio passado, tido como verbo principal na análise tradicional, é, de fato, um adjetivo, representando assim, numa análise gerativa, o núcleo de um sintagma adjetival (e não verbal, como decorrente da análise tradicional). Os exemplos abaixo apresentariam a mesma estrutura composicional:

⁸ O termo *diáthesis* corresponde, em grego, ao conceito de voz verbal. Em latim, temos o chamado *genus verbi*.

1. O cavaleiro é valente.
2. O cavaleiro é esperado.

A forma ‘esperado’, presente no exemplo (2), é, na acepção defendida por Perini (*op. cit.*), um particípio nominal (cuja forma em alguns verbos é distinta: *entregue/entregado; impresso/imprimido*), o qual não seria parte da diátese verbal do verbo *esperar*. A conclusão a que chega Perini é a de que a frase (2) representa, na verdade, uma diátese do verbo *ser*, semelhante à construção em (1), quer dizer, do verbo *ser* com um complemento qualitativo.

Ainda que a hipótese defendida pelo autor acerca da voz passiva nos pareça válida, a argumentação apresentada em sua obra não é de todo definitiva, fato que ele próprio reconhece: “Não quero dar a impressão (*sic*) que a questão está decidida de uma vez por todas; mas a evidência disponível indica que a construção passiva não é uma diátese do verbo morfologicamente relacionado ao particípio nominal” (*id. ib.* p.179). É necessário, desse modo, um estudo mais aprofundado da hipótese levantada, bem como um cotejo de *corpora* diversos, a fim de que sejam apuradas evidências morfológicas e sintáticas que validem a conjectura levantada.

Saliente-se que a presente dissertação também deve importante influência teórica ao trabalho de Alencar (2015a), orientador desta pesquisa, em que, feitas as ressalvas à argumentação de Perini, são apresentadas evidências mais sólidas que corroboram a classificação de perífrase adjetival para a voz passiva. Alencar (*id.*) elenca evidências morfológicas e sintáticas que melhor fundamentam a análise do particípio passado como de caráter adjetival. Contudo, o pesquisador ressalta que ainda há poucos estudos que tratam dessa temática, sendo ainda necessário um escrutínio de tal proposta.

Nosso trabalho, por tratar-se de uma pesquisa no âmbito da Linguística Computacional (LC), pretende contribuir para mitigar o problema da carência no Brasil de estudos para a construção de uma gramática computacional do PB, gramática essa “que tenha grande cobertura e se baseie nas pesquisas mais recentes em gramática gerativa” (cf. ALENCAR, 2013a, p.183). A LC, desse modo, mediante seus pontos de vista teóricos e práticos, representa um instrumento para que se atinjam as metas aqui propostas.

Se se pergunta qual a relevância desta dissertação dentro dos estudos linguísticos, queremos crer que nosso trabalho apresenta uma contribuição, ao testar a hipótese da não existência de voz passiva analítica em PB, na medida em que propomos

a implementação das duas análises opostas, isto é, de uma (a) em que o particípio passivo é analisado como pertencente à categoria V (verbo), e de outra (b) em que o particípio passivo é analisado como sendo do tipo A (adjetivo). Para isso, lançaremos mão de um dos modelos não derivacionais de gramática gerativa, a saber, a Gramática Léxico-Funcional (LFG, do inglês *Lexical-Functional Grammar*) (BRESNAN, 2001; FALK, 2001), e a implementação de ambos os *parsers* no *Xerox Linguistics Environment*, o XLE⁹ (CROUCH *et al*, 2011), um *software* de eficácia comprovada, que constitui o estado da arte para a implementação de gramáticas no modelo gerativo da LFG.

A LFG, dentro dos diversos modelos gramaticais que constituem hoje o nível mais avançado de desenvolvimento da pesquisa em linguística, pertence ao grupo daqueles que permitem implementação computacional e possibilitam, ainda, uma austera revisão das análises gramaticais. A LFG é, por isso, uma ferramenta bastante utilizada por estudantes e pesquisadores de Informática e de Processamento de Linguagem Natural (PLN), de linguísticas computacionais e, do mesmo modo, por linguistas teóricos (SCHWARZE & ALENCAR, 2016).

Ainda que mencionadas algumas das vantagens e possibilidades que se podem obter através do estudo de fenômenos linguísticos por meio da LFG, qualquer um que se disponha a estudar esse formalismo deparar-se-á com uma lacuna de artigos, livros, pesquisas, dissertações e teses, no Brasil, sobre o PB. Uma busca no repositório de teses e dissertações da CAPES, por exemplo, apresenta como resultado unicamente dois trabalhos: a tese de Santos (2014) supramencionada¹⁰ e o trabalho de Morais (1988), este último não leva em conta questões de implementação computacional.

Em estudos de LC de línguas como o inglês e o alemão, o estado da arte da implementação de gramáticas para PLN é bastante robusto. Müller (2016), entretanto, apresenta dados do estado da arte de implementação em LFG e HPSG para o português que chamam atenção, infelizmente, por sua escassez¹¹. Esta pesquisa justifica-se, portanto, como tentativa de contribuir para expandir esse tipo de estudo no Brasil e em língua portuguesa, já que comparativamente os Estados Unidos, Alemanha e alguns

⁹ Agradeço a Valéria de Paiva, John Maxwell e Daniel Bobrow pela licença do XLE concedida ao PPGL da UFC.

¹⁰ Situação que admite, em se tratando de outro formalismo linguístico, a HPSG (do inglês, *Head-Driven Phrase Structure Grammar*), algumas exceções. Citem-se os trabalhos de Fritsch (2013), Marruche (2012), Lawall (2014), Neto (2007), dentre outros.

¹¹ Para trabalhos em PB que utilizam o aporte teórico da LFG, apenas Alencar (2004, 2013) é citado. Abordagens em HPSG, por seu turno, estão representadas pelos trabalhos de Branco & Costa (2008a, 2008b) e Costa & Branco (2010), os quais analisam o português europeu.

países asiáticos, tais como Japão e Indonésia, já se encontram na dianteira nos estudos de PLN. Acreditamos não ser mais cabível, em meio à sofisticação tecnológica em que já nos encontramos, fazer linguística, leia-se ciência, sem uma abordagem computacional para o fenômeno linguagem.

Como mencionado, propomos a implementação em LFG/XLE das duas formas de análise de nosso objeto de pesquisa. Nossa intenção, ao implementar as duas versões de análise do fenômeno voz passiva em PB com seus respectivos *parsers*¹², é a de verificar que análise se mostra computacionalmente mais eficiente. Até o momento, o que encontramos na literatura da área é o resultado de um projeto PIBIC a que faremos referência a seguir, no qual a análise realizada indicaria ser mais eficaz a análise do participio como sendo um adjetivo.

Trata-se de um projeto de iniciação científica sob orientação do Prof. Dr. Leonel Figueiredo de Araripe, realizada por bolsista de iniciação científica e por membros do CompLin¹³ (Grupo de Pesquisa Computação e Linguagem Natural), em que a proposta de Perini para o PB seria analisada na esteira da teoria gerativa e da LC. O trabalho realizado trouxe, como mencionado acima, resultados preliminares que vão ao encontro de nossa proposta de ser a classificação do participio passivo como adjetivo mais eficiente.

Não obstante, devido à complexidade de adaptar as estruturas da gramática

¹² Em ciência da computação, o termo *parser* faz referência ao algoritmo ou programa que executa o parsing, processo de analisar uma sequência de entrada (lida de um arquivo de computador ou do teclado, por exemplo) para determinar sua estrutura gramatical segundo uma determinada gramática formal existente. Essa análise faz parte de um compilador, englobando a análise lexical e a análise semântica. [Esta definição foi resumida a partir deste trecho original em inglês: "Parsing, also referred to as syntax analysis, has been and continues to be an essential part of computer science and linguistics. Parsing techniques have grown considerably in importance, both in computer science, i. e. advanced compilers often use general CF parsers, and computational linguistics where such parsers are the only option. They are used in a variety of software products including Web browsers, interpreters in computer devices, and data compression programs; and they are used extensively in linguistics." Tal trecho faz parte da divulgação do livro *Parsing Techniques: A Practical Guide (Monographs in Computer Science)*, da autoria de Dick Grune e Ceriel J.H. Jacobs (2010)]. Na introdução do mesmo livro, os autores trazem mais uma definição de 'parsing', a saber: "Parsing é o processo de estruturação de uma representação linear de acordo com uma determinada gramática. Essa definição foi mantida de forma abstrata propositadamente, para permitir o maior grau de abertura que uma interpretação possa permitir. Essa "representação linear" pode ser uma sentença, um programa de computador, um padrão de tricô, uma sequência de estratos geológicos, uma peça musical, ações realizadas conforme comportamentos rituais, em resumo, qualquer sequência linear em que os elementos precedentes de alguma maneira restrinjam o próximo elemento." O trecho original correspondente é o seguinte: "Parsing is the process of structuring a linear representation in accordance with a given grammar. This definition has been kept abstract on purpose to allow as wide as an interpretation is possible. The 'linear representation' may be a sentence, a computer program, a knitting pattern, a sequence of geological strata, a piece of music, actions in ritual behaviour, in short any linear sequence in which the preceding elements in some way restrict the next element." (GRUNE; JACOBS, 2010, p. 1).

¹³ Grupo cadastrado na plataforma do CNPq: <http://dgp.cnpq.br/dgp/espelhogrupo/2542059310124116>.

do francês às existentes no português, o projeto inicial de discutir o estatuto da voz passiva em PB foi alterado, e, por fim, trabalhou-se unicamente com um excerto do francês, adaptado de Schwarze e Alencar (2016). Ademais, identificamos algumas incongruências e falhas metodológicas na execução do projeto. Dentre os problemas a que nos referimos, apontam-se: (i) a não manutenção da teoria sintática escolhida durante a adaptação para o português da gramática do francês proposta por Schwarze e Alencar (op. cit.) para a análise da voz passiva: a solução *ad hoc* para implementar subcategorização dos APs foi a de recorrer-se à teoria X-barra, com dois níveis de projeções sintagmáticas (A' e A''), unicamente para o APs, o que não se mostra consistente do ponto de vista linguístico-teórico; (ii) os dados obtidos não foram tratados estatisticamente de uma maneira mais apurada, não se podendo afirmar categoricamente que o desempenho no processamento teve ganho realmente significativo.

Resumindo, no que concerne à metodologia utilizada no projeto PIBIC, verificou-se que não houve um controle rigoroso durante a compilação das duas gramáticas e não se trabalhou com o PB, mas sim com o francês. Ou seja, ainda que haja evidências de que o custo de processamento tenha sido menor nos casos em que os participios foram analisados como adjetivos, um desenho metodológico mais rigoroso deve ser praticado.

Propomo-nos, então, a desenvolver uma gramática em LFG do PB a fim de melhor verificar os resultados do fenômeno linguístico foco de nosso estudo. No caso desta pesquisa, o grande diferencial será que trabalharemos com uma gramática do PB, por nós desenvolvida, que, por sua vez, já será mais ampla do que o fragmento da gramática LFG para o francês utilizado no projeto PIBIC. A isso, acrescente-se que faremos um maior controle na parte experimental da pesquisa, de cuja descrição detalhada falaremos na seção de metodologia.

Nossa hipótese, ainda a ser testada, seguindo na esteira dos resultados preliminares do projeto PIBIC, é que o processamento computacional será menos dispendioso, ou seja, implicará em menos gasto de tempo pela CPU, quando se considera o participio passado, em construções com “verbo cópula + participio”, pertencente à categoria dos adjetivos. Haveria, assim, evidências linguístico-computacionais que confirmariam nossas conjecturas.

Ressalte-se ainda que uma das aplicações práticas desse trabalho, embora não seja seu escopo principal, é a de desenvolver ferramentas que possam ser utilizadas

em PLN, como a anotação automática de *corpora* linguísticos¹⁴, a título de exemplo. Através da implementação em XLE de uma gramática capaz de analisar sentenças que contenham o que vem sendo tradicionalmente classificado como voz passiva, almejamos avançar em trabalhos como aqueles realizados por Alencar (2013a, 2013b), em sua gramática BrGram¹⁵. Pretendemos, dessa maneira, ajudar a preencher a lacuna em PB a que Alencar (2015b) chama de “paisagem de *corpora*”. O termo é a tradução para o alemão *Korpuslandschaft* utilizado por Lemnitzer e Zinsmeister (2006) e faz referência ao fato de o português não dispôr de um *parser* de grande extensão baseado em LFG.

Do exposto, é preciso que cada vez mais pesquisas em linguística computacional sejam realizadas, cujo objeto de estudo seja o PB, do contrário, correremos o risco de permanecermos amadores no *modus operandi* científico e tecnológico em comparação com outras universidades ou centros de pesquisa espalhados pelo mundo.

Pode-se questionar qual a utilidade de pesquisas em gramática gerativa e, do mesmo modo, como apresentado em Alencar (2008, p.2), por que se utilizar de formalismos matemáticos para a descrição e análise de um objeto como a língua, que parece pertencer a uma esfera diametralmente oposta ao das ciências naturais, às quais a matemática se presta de maneira eficaz para a análise. Muitos pesquisadores e estudantes de Letras/Linguística, influenciados por questões idiossincráticas e pela dificuldade em lidar com ciências exatas, rechaçam tais tipos de investigação formalista baseando-se em um argumento superficial de que métodos de investigação exatos e formais não se harmonizariam com um objeto tipicamente cultural, nomeadamente, a língua/linguagem. Entretanto, linguistas que trabalham com gerativismo defendem que somente uma análise formal em que se levem em conta os processos de uma língua particular e as propriedades da linguagem em geral permitiria testar, com precisão, hipóteses sobre os fenômenos referentes a esse objeto. O uso das teorias gerativas aplicadas à informática permitiria ainda mecanismos de testagem mais rigorosos, uma vez que estes podem ser feitos de maneira automática, evitando-se, por exemplo,

¹⁴ “A anotação de *corpus* é o procedimento de acrescenta informação (lexical, morfológica, semântica, sintática, discursiva, estrutural etc.) a um *corpus*” (INDURKHYA e DAMERAU, 2010, p.167).

¹⁵ A BrGram (ALENCAR, 2013) é uma gramática computacional do português desenvolvida em LFG/XLE seguindo os moldes do Projeto ParGram. O ParGram (2012), por seu turno, é um projeto que visa a produzir gramáticas de ampla cobertura para línguas tipologicamente diferentes (húngaro, inglês, alemão, espanhol etc.). Estas são desenvolvidas, de maneira colaborativa, utilizando-se o modelo da LFG através de um conjunto de traços gramaticais e aparato conceitual acordados em comum.

procedimentos que estejam mais ligados à intuição, sujeitos a falhas.

Ao testarmos as hipóteses de Perini (2010) sobre a formação da voz passiva em PB, entendemos que o presente estudo pretende preencher lacunas já mencionadas na linguística descritiva e teórica, assim como contribuir com mais pesquisas nas áreas de linguística computacional e linguística de *corpus* em PB, cujos resultados poderão ser utilizados, por exemplo, na criação de *treebanks*¹⁶. A condição *sine qua non* para a construção de *treebanks* é justamente um *parser*, ou seja, a existência de uma ferramenta que gere de forma automática a análise das sentenças de um dado *corpus*.

Alencar (2015b), no texto no mesmo Projeto de Iniciação Científica já mencionado, registrado na Pró-Reitoria de Graduação da UFC, destaca que atualmente apenas dois *corpora* desse tipo estão livremente disponíveis no âmbito da língua portuguesa: o CORDIAL-SIN (CARRILHO ET AL., 2011; CORDIAL, 2012) e o *Corpus* Histórico do Português Tycho Brahe (GALVES; FARIA, 2010). O pesquisador também adianta que os dois *treebanks* foram concebidos segundo o modelo gerativo transformacional da Teoria da Regência e da Ligação (TRL), contemplando dois contextos distintos. Enquanto o CORDIAL-SIN é composto de transcrições de entrevistas realizadas com falantes de diferentes dialetos de Portugal, o *Corpus* Histórico do Português Tycho Brahe toma por base textos de cunho histórico. Como se pode depreender do aqui exposto, não existe um *treebank* de textos do português contemporâneo do Brasil com análises sintáticas profundas baseadas em um modelo gerativo como o da LFG.

Em suma, consideramos o presente estudo como uma contribuição para uma análise e uma descrição mais apuradas do fenômeno da voz passiva em PB, e, ao mesmo tempo, um estímulo para que outros linguistas realizem investigações semelhantes neste vasto – ainda pouco explorado - campo da LC. Também queremos contribuir para que estudantes e pesquisadores do PB tenham acesso ao formalismo LFG/XLE.

¹⁶ Sobre *treebanks*, veja-se esta afirmação: “Bancos de árvores são *corpora* estruturalmente anotados que representam (além de anotações morfológicas e de *part of speech*) relações sintáticas, semânticas e até intersentenciais. A palavra *árvore* se refere à estrutura básica ou típica da anotação, que corresponde a noção de *árvore* como definido na teoria formal dos grafos” (INDURKHAYA & DAMERAU, 2010, p.170).

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo geral desta pesquisa é identificar, à luz dos princípios da Gramática Léxico- Funcional e da Linguística Computacional, qual das duas propostas de classificação apresentadas é mais eficiente em PLN para o PART-PASS (particípio passivo) em orações passivas perifrásticas.

1.2.2 Objetivos Específicos

Com o intuito de melhor analisar as construções que tradicionalmente são chamadas de voz passiva analíticas no PB, tanto do ponto de vista da linguística gerativa, como também tendo por base a linguística computacional, buscaremos esclarecer os seguintes pontos, objetivos específicos da pesquisa:

- Línguístico: Testar a hipótese de que o PART-PASS é um elemento de categoria A;
- Computacional: Testar a hipótese de que a gramática em que o PART-PASS é implementado como A (adjetivo), ou seja, a gramática G-A, é mais eficiente do que a G-V, aquela em que o PART-PASS é implementado como V (verbo).

1.3 Problemas da Pesquisa

O presente trabalho de dissertação surge da observação de que as descrições convencionais do fenômeno da voz passiva no PB não são consensuais. Acreditamos que o questionamento acerca do *status* da passiva no PB é justificável, e uma vez que um dos focos desta investigação é a aplicação computacional de teorias linguísticas, furtamo-nos, por questões metodológicas e pragmáticas, de uma discussão teórica mais extensa a respeito do estatuto do particípio em construções passivas. Uma discussão teórica mais detalhada, porém, pode ser alvo de trabalhos futuros.

Levando-se em conta o que foi até aqui apresentado e discutido, tomamos por plausível a proposta de Perini (2010) de que as construções tradicionalmente classificadas como voz passiva seriam, do ponto de vista gerativo, formações predicativas adjetivais. Nesse contexto, acreditamos poder então formular a questão

basilar a que esta pesquisa se propõe a responder:

- Qual é, no caso da implementação de uma gramática computacional, a classificação mais eficiente¹⁷ para o PART-PASS?

1.4 Hipóteses

Diante da discussão apresentada até o momento, podemos formular as hipóteses que nortearam nossa pesquisa acerca da voz passiva em PB.

1.4.1 Hipótese Básica

A hipótese básica aqui levantada é a de que a classificação mais consistente (do ponto de vista linguístico-gerativo) e mais eficiente (do ponto de vista computacional) para o PART-PASS é a de pertencente à categoria A, ou seja, que as construções de voz passiva analítica em PB são, na verdade, formações predicativas adjetivais.

1.4.2 Hipóteses Secundárias

A gramática G-A resulta no *parsing* (processamento) mais eficiente.

¹⁷ Por eficiente, entende-se aqui qual a análise menos dispendiosa para o processamento computacional e que, ao mesmo tempo seja mais elegante e concisa do ponto de vista linguístico. *Grosso modo*, baseamos no princípio epistemológico da Navalha de Occam, o qual afirma que a melhor análise para um fenômeno é aquela que pressupõe a menor quantidade de premissas, defendemos ser a classificação que aglutina construções adjetivais e passivas verbais (com verbo ‘ser’) num mesmo grupo mais eficiente tanto do ponto de vista de processamento computacional como do de elegância teórica.

2 UM PANORAMA SOBRE A VOZ PASSIVA

Este trabalho apresenta uma reflexão a respeito da razoabilidade da hipótese de Perini (2010) sobre a inexistência da chamada voz passiva analítica em PB e, mais particularmente, sobre o caráter adjetival do particípio passado. A pesquisa adota uma perspectiva gerativista, mais especificamente pautada no arcabouço teórico da LFG, e busca evidências tais que corroborem nossa forma de ver o objeto linguístico aqui investigado. Por fim, dentro do âmbito da linguística computacional, pretendemos apresentar qual das duas análises divergentes contidas na literatura representa a forma de tratamento mais eficiente para as chamadas construções passivas analíticas.

Exaurir, em uma introdução teórica, as diversas teorias linguísticas existentes acerca das vozes verbais, mesmo se elegendo a voz passiva como foco, parece-nos tratar-se de uma tarefa de Sísifo. Por isso, optamos por trazer à baila, ainda que de maneira sucinta e, em alguns pontos, generalizada, os problemas que a chamada voz passiva tem suscitado nos estudos linguísticos. Não se pretende aqui, entretanto, traçar um panorama *ab ovo* da voz passiva, senão apresentar as principais análises propostas para lidar com o fenômeno.

Feita esta ressalva, discutiremos, inicialmente, como a voz passiva é analisada em algumas gramáticas tradicionais. Uma das primeiras dificuldades com a qual nos deparamos ao estudar as vozes verbais é conseguir uma definição clara do conceito de voz. Visando-se a este alvo, cumpre investigar criticamente o que alguns gramáticos de renome disseram, ao longo do tempo, sobre vozes verbais.

Após essa breve apresentação, mostramos como a gramática gerativa vem tratando o fenômeno da voz passiva nas suas principais vertentes.

2.1 A Gramática de Port-Royal e as vozes verbais

Apostando na chancela e no valor atribuídos por Noam Chomsky e Ernest Renan à Gramática de Port-Royal¹⁸, acreditamos ser ela digna de menção na parte

¹⁸ Sobre a importância atribuída por Chomsky à Gramática de Port-Royal, vejam-se os comentários tecidos pelos tradutores brasileiros da obra em questão: “Para finalizar, destacando mais uma vez o valor dessa obra, chamamos a atenção do leitor para Noam Chomsky, com seu *Cartesian Linguistics* e seu subtítulo *A Chapter in the History of Rationalist Thought*, de 1966, de que existe uma tradução brasileira (Vozes-Edusp, Petrópolis 1972). No capítulo “Estrutura profunda e Estrutura de superfície”, Chomsky declara que a maioria das chamadas “gramáticas universais” (ou “gerais”) posteriores segue as linhas da de Port-Royal; essas linhas partem do modo pelo qual os conceitos se combinam para a formação do

teórica introdutória deste trabalho.

No ano de 1660, foi publicada na França a célebre *Grammaire générale et raisonnée contenant les fondemens de l'art de parler, expliqués d'une manière claire et naturelle*, conhecida no Brasil, notadamente mediante a tradução de Bruno Fregni Bassetto e Henrique Graciano Murachco (cf. Arnauld; Lancelot, 2001), como *Gramática de Port-Royal ou Gramática Geral e Razoada contendo os fundamentos da arte de falar, explicados de modo claro e natural; as razões daquilo que é comum a todas as línguas e das principais diferenças ali encontradas*. Essa gramática encontra-se dividida em duas partes principais. A Primeira Parte tem como título “Onde se fala das letras e dos caracteres da escrita”, e a Segunda Parte, “Onde se fala dos princípios e dos motivos sobre os quais se baseiam as diversas formas da significação das palavras” (*id.*).

Na Segunda Parte, em que se esmiúçam as diferentes classes gramaticais, há uma seção em que os autores se dedicam às vozes verbais sob o seguinte título: *Dos verbos que podem ser chamados adjetivos, e de suas diferentes espécies, ativos, passivos e neutros*. Sobre essa temática, Arnauld e Lancelot (2001, p. 100), fazem as seguintes considerações:

Já dissemos que os homens, tendo reunido em uma infinidade de situações algum atributo particular à afirmação, haviam feito esse grande número de verbos diferentes do substantivo, que se encontra em todas as línguas e que se poderia chamar *adjetivos*, para mostrar que a significação, que é própria de cada um, é acrescentada à significação comum a todos os verbos, que é a da afirmação. Trata-se, porém, de um erro comum crer que todos esses verbos significam ações ou paixões, porque nada existe que um verbo não possa ter como seu atributo, se for do agrado dos homens juntar a afirmação com esse atributo. Vemos que até o verbo substantivo *sum, eu sou*, muitas vezes é adjetivo, pois, em vez de tomá-lo como significando apenas a afirmação, junta-se-lhe o mais geral de todos os atributos, que é o ser, como quando digo: *Je pense, donc je suis – Penso, logo existo (sou), je suis* aí significa *sum ens*, sou um ser, uma coisa; *existo* significa também *sum existens*, eu sou, eu existo. Isso, porém, não impede que se possa conservar a divisão comum desses verbos em ativos, passivos e neutros. São chamados propriamente ativos os que significam uma ação à qual se opõe uma paixão, como *bater, ser batido; amar, ser amado*, quer essas ações terminem num substantivo, o que se chama ação real, como *bater, romper, matar, escurecer* etc., quer terminem apenas num objeto, o que se chama ação intencional, como *amar, conhecer, ver*.

Interessa-nos, aqui, com base nas afirmações dos dois estudiosos de Port-Royal, destacar o fato de que, na visão dos autores, “nada existe que um verbo não possa ter como seu atributo, se for do agrado dos homens juntar a afirmação com esse

juízo, segundo as três operações do espírito: conceber, julgar e raciocinar.” (ARNAULD; LANCELOT, 2001, p. XXX.)

atributo”. Ao darem espaço para uma discussão sobre verbos e atributos/adjetivos, Arnauld e Lancelot logo fazem referência às – expressas em termos atuais – vozes verbais, sublinhando, em primeira instância, ser “um erro comum crer que todos esses verbos significam ações ou paixões”. Acresça-se, nesta oportunidade, que o termo “paixão”, contido nessa afirmação, encerra o sentido de “voz passiva”. “Paixão”, que chega ao português através do latim, é, em sua origem, derivado do vocábulo grego *páthos*, que significa, em princípio, “tudo aquilo por que passamos [aquilo que sofremos], em oposição àquilo que fazemos (...)”¹⁹. Se, nas gramáticas do português em geral, não mais se recorre a esse termo como sinônimo de “voz passiva”, é relevante observar que, na língua alemã, paralelamente ao termo *Passiv*, também existe uma forma, atualmente pouco usada, chamada *Leideform*, que, numa tradução literal, significaria “forma do sofrimento” e, numa tradução ajustada à temática em que é empregada, “passividade” ou “voz passiva”.

Aquilo que mencionamos no início deste trabalho a respeito das diferenças existentes entre línguas românicas modernas e latim e grego antigos é também discutida quando, prosseguindo em seu texto, os dois gramáticos franceses de Port-Royal explicitam a existência de diferenças morfológicas entre a voz passiva em línguas antigas e línguas europeias de seu tempo:

Isso teve como consequência que, em muitas línguas, os homens se serviram da mesma palavra, dando-lhe diversas inflexões, para significar um e outro, chamando verbo ativo àquele que tem a inflexão²⁰, pela qual assinalaram a ação, e verbo passivo, àquele que dispõe de inflexão, para a qual marcaram a paixão: *amo* – *amor* (“amo” – “sou amado”); *verbero* – *verberor* (“chicoteio” – “sou chicoteado”). Isso esteve em uso em todas as línguas antigas, latina, grega e orientais; e o que é mais, essas últimas dão três ativos a um mesmo verbo, cada qual com seu passivo e um recíproco, que contém algo de um e de outro, como seria *amar-se*, que significa a ação do verbo sobre o próprio sujeito desse verbo. Mas as línguas usuais da Europa não têm passivo e, em seu lugar, se servem de um particípio formado do verbo ativo e tomado no sentido passivo, com o verbo substantivo *eu sou*, como *eu sou amado* etc.

Como vimos acima, Arnauld e Lancelot dão-nos uma explicação²¹, à luz da

¹⁹ Cf. *Le Grand Bailly – Dictionnaire Grec Français*: “ce qu’on éprouve, p[ar] opp[osition] à ce qu’on fait (...)”.

²⁰ Na tradução brasileira, há a seguinte nota de rodapé à guisa de explicação para “àquele [verbo] que tem a inflexão”: “O que os autores chamam de *ativo* nós denominamos *transitivo* ou *intransitivo*, isto é, o que expressa ação que pode ou não “transitar” para um objeto” (ARNAULD; LANCELOT, 2001, p. 101).

²¹ Esta pesquisa deverá também se ocupar de detectar onde os termos apresentados por Arnauld e Lancelot têm sua origem, notadamente em contextos linguísticos. Em primeira linha, sabe-se que o termo *páthos*, num contexto mais geral referente à ideia de passividade, remonta a Aristóteles e passa também por Descartes. Cf., sobre este tema, o editorial da Revista Latino-Americana de Psicopatologia Fundamental (Acesso em 17 jun.2017.): <http://www.fundamentalpsychopathology.org/uploads/files/revistas/volume05/n2/editorial_v5n2.pdf>.

comparação interlinguística, já nos idos de 1660, que sempre voltamos a encontrar em tradicionais gramáticas de diferentes línguas e, obviamente, também nas gramáticas do PB, como veremos a seguir mediante dois exemplos de gramáticas amplamente usadas e aceitas como tradicionais no Brasil.

2.2 A gramática de Napoleão Mendes de Almeida e as vozes verbais

Em sua *Gramática Metódica da Língua Portuguesa*, Napoleão Mendes de Almeida, que além de gramático era filólogo e professor de português e latim, faz estas considerações sobre as vozes verbais:

§388 – **Voz ativa:** O verbo de uma oração está na *voz ativa* quando a ação é evidentemente *praticada* pelo sujeito; este, em tal caso, é o *agente da ação verbal*. Exemplos: “O caçador matou o tigre” – “Ele passou de ano” – “Pedro voltará amanhã”.

Nessas orações, os verbos *matou*, *passou* e *voltará* indicam ações praticadas pelos respectivos sujeitos: *caçador*, *ele* e *Pedro*.

§389 – **Voz passiva:** Acontece muitas vezes que a pessoa ou coisa, a que se atribui a ação verbal, recebe a ação em vez de praticá-la. Na primeira das orações acima (“O caçador *matou* o tigre”) o *caçador* é o sujeito de *matou*; nesta outra oração: “O caçador *foi morto* pelo tigre”, o sujeito continua sendo o mesmo (Quem foi morto?) – *O caçador*, pois é a ele que se atribui o fato de ser morto, mas, agora, o sujeito não pratica, e, sim, recebe a ação verbal.

Por que essa diferença? – Porque no primeiro caso o verbo está na *voz ativa* e, no segundo, na *voz passiva*.

Voz passiva é, pois, a que expressa uma ação *sofrida*, *recebida* pelo sujeito; o sujeito, nesse caso, é *paciente* ou *recipiente* da ação. (ALMEIDA, 1980, p. 208).

Constatamos, no excerto acima citado, que Napoleão Mendes de Almeida mantém-se firmemente arraigado à concepção tradicional, conforme a qual o verbo é o centro dos sintagmas tanto nas estruturas de voz ativa quanto nas de voz passiva apresentadas em seus exemplos. Recorre, ademais, à já mencionada ideia de “paixão”, embora não empregue esse termo que, já de longuíssima data, está em desuso na língua portuguesa (salvo em outra situação temática, que é o caso específico da *paixão* [= sofrimento] *de Cristo*), ao se referir a “uma ação sofrida, recebida pelo sujeito”. A passividade de que é portador o sujeito das orações na voz passiva reflete-se nas explanações do renomado filólogo, quando este, confirmando suas ideias anteriormente expostas, reitera que “o sujeito, nesse caso, é *paciente* ou *recipiente* da ação. Sobre a formação da voz passiva no PB, Napoleão Mendes de Almeida explica:

§391 – A voz passiva é em português indicada de duas maneiras:

1ª mediante os verbos auxiliares *ser* e *estar* e o *particípio* de certos verbos ativos: *ser visto*

(*sou visto, és visto, é visto ...*); *estar preso* (*estou preso, estás preso, está preso ...*).

Notas: a) Também o verbo *ficar* se presta, às vezes, para indicar a voz passiva; na oração “*Ele foi preso*” – podemos sem sacrifício do sentido passivo da oração, substituir o *foi* por *ficou*: “*Ele ficou preso*”.

b) O português não possui flexões verbais sintéticas para a voz passiva; em latim e em grego, a passiva pode expressar-se por uma única palavra, ao passo que o português necessita de duas. Quer isso dizer que, propriamente, não possuímos *verbos passivos*, mas *voz passiva*.

2ª Mediante o pronome *se*, que então se diz *pronome apassivador*; este caso se dá sempre que o sujeito é ente inanimado, conseqüentemente incapaz de praticar a ação verbal, ou quando o sentido da oração mostra que o sujeito é apenas paciente.

Na oração: “*Alugam-se casas*” – *casas* não pratica a ação de alugar, e sim, recebe, sofre tal ação, o que equivale a dizer que *casas* não é o agente, mas o paciente da ação verbal. O verbo é passivo, e essa passividade é indicada pelo pronome *se*. A oração “*Alugam-se casas*” é idêntica à oração “*Casas são alugadas*”; em ambas o sujeito é *casas*, que, pelo fato de estar no plural, deverá levar também para o plural o verbo; dizer “*Aluga-se casas*” é erro igual a “*casas é alugada*”.

Constituem, conseqüentemente, erros inomináveis, construções como: “*Vende-se livros usados*” – “*Conserta-se relógios*” – “*Reforma-se chapéus*”. (ALMEIDA, *id.* p. 210)

2.3 A gramática de Celso Cunha e Lindley Cintra e as vozes verbais

Nessa gramática, encontram-se as seguintes explicações básicas sobre as vozes em geral, aqui extraídas da versão portuguesa da gramática em epígrafe:

Vozes.

O facto expresso pelo verbo pode ser representado de três formas:

2.3.1 como *praticado* pelo sujeito:

João **feriu** Pedro.

Não **vejo** rosas neste jardim.

2.3.2 como *sofrido* pelo sujeito:

Pedro **foi ferido** por João.

Não **se vêem** [= são vistas] rosas neste jardim.

2.3.3 como *praticado e sofrido* pelo sujeito:

João **feriu-se**.

Dei-me pressa em sair.

No primeiro caso, diz-se que o verbo está na VOZ ACTIVA; no segundo, na VOZ PASSIVA; no terceiro, NA VOZ REFLEXIVA.

Como se verifica dos exemplos acima, o objecto directo da VOZ ACTIVA corresponde ao sujeito da VOZ PASSIVA; e, na VOZ REFLEXIVA, o objecto directo ou indirecto é a mesma é a mesma pessoa do sujeito. Logo, para que um verbo admita a transformação de voz, é necessário que ele seja TRANSITIVO. (CUNHA; CINTRA, 1985b, p. 382s.)

Como se pode depreender das explicações acima, Cunha e Cintra detalham, com exemplos ilustrativos, as três vozes admitidas na língua portuguesa. Além disso, ressaltam a necessidade da existência de verbos transitivos para que possa ocorrer a

transformação de voz. Ao entrarem em pormenores sobre a voz passiva especificamente, escrevem os autores:

Voz passiva.

Exprime-se a VOZ PASSIVA:

a) Com o VERBO AUXILIAR *ser* e o PARTICÍPIO do verbo que se quer conjugar: Pedro **foi ferido** por João.

b) Com o PRONOME APASSIVADOR *se* e uma terceira pessoa verbal, singular ou plural, em concordância com o sujeito:

Não **se vê** [= é vista] **uma rosa** neste jardim. Não **se vêem** [são vistas] **rosas** neste jardim. (CUNHA; CINTRA, 1985b, p. 383s.)

Verifica-se que, ao explicarem a formação da voz passiva em português, Cunha e Cintra, mediante suas considerações, constroem um texto muito próximo das ideias apresentadas por - Napoleão Mendes de Almeida na seção anterior. Como tal fato se repete em outras gramáticas do português, reputadamente tradicionais e em uso no Brasil, deixaremos, com o fito de evitarmos redundâncias, de apresentar considerações idênticas ou muito semelhantes apresentadas por outros gramáticos brasileiros, como ocorre, à guisa de ilustração, em Cegalla (1967), Rocha Lima (1996), Luft (2002) e Bechara (2005).

A fim de deixar claro o recorte do fenômeno linguístico com os quais trabalharemos nesta pesquisa, salientamos que aquilo que a gramática tradicional chama de passiva analítica constitui-se na verdade de dois grupos: passivas verbais e passivas adjetivais. Não implementamos em nosso estudo passivas adjetivais, formadas com os verbos *estar/ficar* em PB, sobretudo por haver evidências psicolinguísticas e sintáticas de que se trata de diferentes tipos de passivas o que exigiria provavelmente soluções e implementações distintas.

Na verdade, dentro de cada um dos dois grupos de passivas, há dois subgrupos: aquelas com verbos de ação e aquelas com verbos de não ação. Apresentamos o exposto de modo resumitivo abaixo:

- (i) *ser* (passiva verbal):
- O castelo será trancado (pelo cavaleiro)/ O castelo será trancado / O castelo trancado está... (verbo de ação);
 - A fada foi vista (pelo cavaleiro)/ A fada foi vista/ *A fada vista está... (verbo de não ação).
- (ii) *estar/ficar* (passiva adjetival):

- O castelo estava trancado (pelo cavaleiro)/ O castelo estava trancado/ O castelo trancado;
- O castelo ficava trancado (pelo cavaleiro)/ O castelo ficava trancado/ O castelo trancado;
- A fada estava vista (pelo cavaleiro)/ *A fada estava vista/ *A fada vista está...;
- A fada ficava vista (pelo cavaleiro)/ *A fada ficava vista/ *A fada vista está...²²

2.4 A passiva na Gramática Gerativa

Nesta subseção, abordamos as principais análises do fenômeno da voz passiva dentro do quadro teórico da teoria gerativa, passando, no próximo capítulo, a analisá-la dentro do escopo específico da LFG. Iniciamos apresentando um grupo específico de verbos monoargumentais: os verbos inacusativos. A razão pela qual os trazemos à baila é o fato de que vários autores descreverem o processo de passivização como sendo nada mais do que um processo de inacusativização de um verbo transitivo (MIOTO *et al.*, 2007).

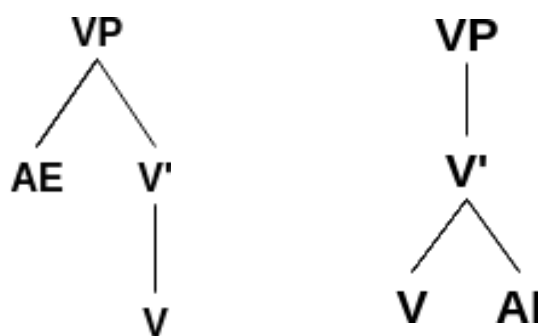
2.4.1 Os verbos inacusativos

Quando se fala de valência verbal²³, pode-se classificar os verbos em diferentes tipos a depender do número de complementos que ele exija. Em PB, há desde verbos a-valentes (e.g. *chover, ventar, amanhecer, nevar*) que, normalmente, não exigem nenhum argumento interno ou externo, até verbos, para alguns autores, tetravalentes (e.g. *traduzir: A traduz alguma coisa de X para Y*). No caso dos verbos monovalentes ou monoargumentais, verbos que selecionam um único argumento, aqueles a que a gramática tradicional classifica como verbos intransitivos, uma subdivisão faz-se necessária, uma vez que há pelo menos duas possibilidades de seleção desse argumento: como argumento externo (AE) ou interno (AI).

²² Agradeço aos apontamentos, indicações e discussão de exemplos feitos pelo Prof. Dr. Gabriel de Ávila Othero, membro da banca de arguição deste trabalho.

²³ Ver seção 3.2 sobre teoria da valência

FIGURA 1: Representação dos verbos monoargumentais



Fonte: Elaborada pelo autor.

A consequência a que se chega analisando-se a FIGURA 1 acima é a de que a chamada classe de verbos intransitivos é, na verdade, formada por dois grupos distintos de verbos. Ao grupo de verbos que seleciona um argumento interno dá-se o nome de verbos inacusativos. O grupo que seleciona como único argumento um argumento externo recebe o nome de inergativos. Uma divisão clássica de verbos classifica como sendo intransitivos tanto os verbos que não possuem objeto (os inergativos) ou verbos cujo argumento interno não pode ser alçado a sujeito da passiva (inacusativos).

As estruturas mostradas na FIGURA 1 diferem tanto do ponto de vista sintático, como visto na seleção de AI ou AE, como também do ponto de vista semântico, uma vez que codificam com o verbo diferentes papéis semânticos (KATO, 2015).

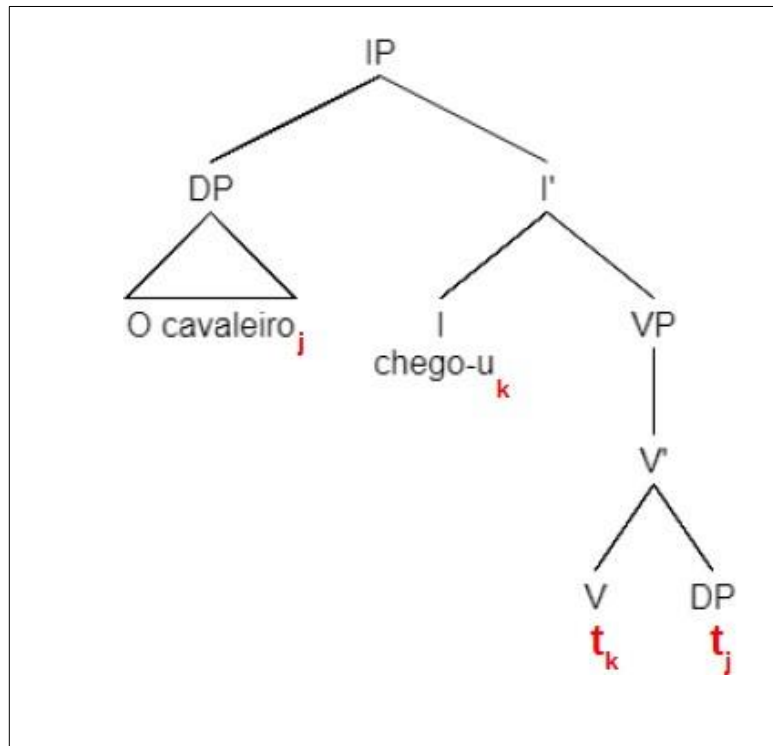
Kato (*op. cit.*) afirma ainda que o argumento interno de um verbo inacusativo pode tanto aparecer *in situ* (vide FIGURA 1 acima) ou pode deslocar-se para a posição de argumento externo, ou seja, a posição de sujeito da oração (vide FIGURA 2 na página seguinte). Uma vez deslocado, o DP argumento interno recebe Caso Nominativo, o que desencadeia a concordância com o verbo.

A incapacidade de os verbos inacusativos atribuírem Caso a seus complementos internos é consequência da chamada generalização de Burzio que afirma: “se um verbo não tem um argumento externo, então ele não pode atribuir caso acusativo (estrutural)” (MÜLLER, 2016, p.106)²⁴. Em outras palavras, um verbo só pode atribuir Caso a seu AI se atribuir papel temático a seu AE. Consequentemente, aplicando-se o

²⁴ No original: “If V does not have an external argument, then it does not assign (structural) accusative case”.

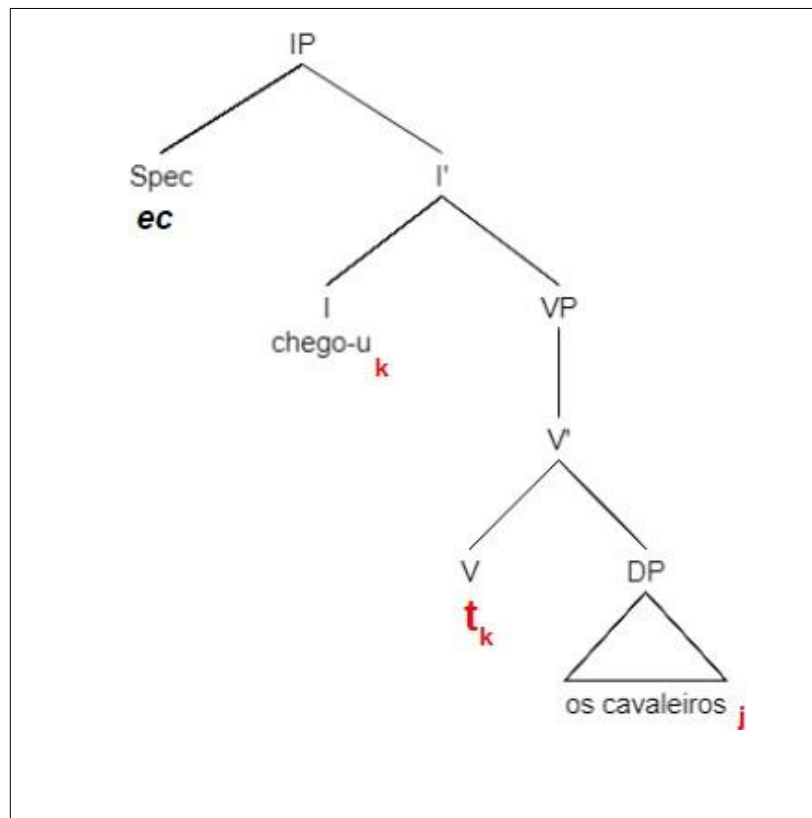
raciocínio de modo inverso, um verbo só atribui papel temático a seu AE se atribuir Caso a seu AI. Uma vez que os verbos inacusativos não selecionam AE, não têm a capacidade de atribuir Caso Acusativo a seu AI.

FIGURA 2: Estrutura de superfície de um verbo inacusativo com AI deslocado.



Fonte: Elaborada pelo autor.

Há também a possibilidade de o DP argumento interno não ser alçado para a posição de SPEC de IP, permanecendo *in situ*. Nesse caso, semelhantemente ao que ocorre com os chamados verbos impessoais, há a inserção de um pronome nulo expletivo na posição de SPEC.

FIGURA 3: Estrutura de superfície de verbo inacusativo com AI *in situ*.

Fonte: Elaborada pelos autores.

Na FIGURA 3, o sujeito da sentença não é o DP *os cavaleiros*, mas sim o pronome expletivo, o que explica a não concordância entre o verbo e o argumento interno. Por vezes, entretanto, sobretudo na língua padrão escrita, há concordância do verbo com o DP plural *os cavaleiros*. Isso se deve ao fato (cf. KATO, 2015) de o DP estar associado ao expletivo, associação que explica a concordância com as marcas de número plural. É também devido a essa concordância que a gramática tradicional analisa o argumento interno como sendo o sujeito dos verbos inacusativos.

A associação acima mencionada entre o expletivo e o DP AI é explicitada por Miotto *et al.* (2007, p.202) quando o autor nos apresenta a Hipótese da Transmissão de Caso: “o expletivo nulo e o DP pós-verbal formam uma cadeia A, excepcional, já que o DP lexical está na cauda, de tal modo que o expletivo [+K] em SPEC IP transmite o NOMINATIVO para o DP lexical”. É desse modo que a TRL em sua versão revisitada (CHOMSKY, 1986) resolve o problema de sentenças passivas (e de verbos inacusativos), pois do contrário, uma condição essencial para o licenciamento de DPs seria violada. Trata-se das determinações apontadas pela Teoria Θ e da Teoria do Caso,

segundo as quais um DP tem de ter papel e, caso seja pronunciado, tem também de ter CASO. O DP *os cavaleiros* é gerado como AI do verbo *chegar*, onde recebe seu papel temático de TEMA, obedecendo a Teoria Θ , mas, se permanece *in situ*, vai de encontro à Teoria do Caso, uma vez que, nesse caso, teríamos um DP com papel temático, mas sem Caso atribuído, haja visto que tanto nos verbos inacusativos como nas sentenças passivas, o predicador VP perde a capacidade de atribuir Caso Acusativo a seu AI. A situação na qual o DP permanece *in situ* é tratada como excepcional porque, a priori, teríamos uma DP pronunciado com papel Θ , mas sem Caso. A outra situação, quando O DP se move para uma posição não-temática mais elevada na cadeia sintática (SPEC de IP) resolve as exigências no que concerne ao Caso (NOMINATIVO atribuído por IP) e temático (de TEMA atribuído pelo verbo ou pelo particípio).

2.4.2 Sobre a passiva na Teoria Gerativa

Como visto acima, podem-se dividir os verbos em subgrupos a depender da quantidade de argumentos que eles exigem, mas também dependendo das propriedades que esses argumentos precisam preencher. Vimos o caso dos verbos monoargumentais, os quais selecionam um elemento interno ou externo, podendo, assim, serem classificados como inacusativos ou inergativos.

Os verbos biargumentais, por seu turno, também apresentam peculiaridades que passam despercebidas quando se utilizam conceitos da gramática tradicional. A terminologia pode, por vezes, gerar mal-entendidos mesmo entre linguistas renomados: tome-se como exemplo a crítica feita a Chomsky por Culicover e Jackendoff (2005 *apud* MÜLLER, 2015), no tocante ao estudo da passiva. Chomsky afirma que, em inglês, a combinação do auxiliar *be* + verbo com morfologia passiva apenas podia ser usado com verbos transitivos. Culicover e Jackendoff (2005 *apud* MÜLLER, 2015), entretanto, afirmam que a generalização não se aplica a todos os verbos, e citam os exemplos com *weigh* e *cost*, cujo comportamento também se verifica em português. Vejam-se os exemplos abaixo adaptados de Müller (2015, p.42):

- (1) O livro pesa 200 gramas/ custa 50 reais
*200 gramas são pesados pelo livro/ 50 reais são custados.

A confusão dá-se pelo fato de Culicover e Jackendoff usarem o termo transitivo para qualquer verbo que requeira dois argumentos, hiperinterpretando a afirmação de Chomsky. Dizer que um verbo precisa exigir, pelo menos, dois

argumentos para poder ser passivizado não equivale a dizer que todo verbo com dois argumentos pode passar por esse processo. Desse modo, optamos, neste trabalho, na esteira de Müller (2016), a chamar de transitivos verbos com um objeto (AI) que pode ser alçado à posição de sujeito na passivização. Verbos que como os do exemplo (1) acima, com dois argumentos, seriam classificados como bivalentes.

Definidos o que entendemos como sendo verbos inacusativos e transitivos, passamos agora a discutir a análise das construções passivas dentro do quadro teórico do gerativismo. Aqui, concordamos com a ideia de que verbos passivizados são semelhantes aos verbos inacusativos na medida em que apresentam um sujeito que tem propriedades de um objeto.

Grosso modo, a passivização poderia ser vista como um processo de inacusativização. A razão para isso deve-se ao fato de que sentenças passivas perdem a capacidade de atribuir Caso Acusativo a seu complemento interno, tal qual como ocorre com os verbos inacusativos. Isso fica evidente se se comparam exemplos como (2a) e (2b) abaixo:

- (2) a. O cavaleiro a procurou.
b. *Foi procurada-a.

A impossibilidade de atribuição de Caso seria consequência da morfologia passiva presente em sentenças passivas. O morfema passivo absorve o Caso Acusativo, e, conseqüentemente, o DP argumento interno necessita ser movido (cf. HAEGMAN, 1994) para uma posição em que possa receber Caso, ou seja, movido para a posição de Spec de IP. Vejamos outro exemplo:

- (3) A rainha é esperada pelo cavaleiro.

Dentro da teoria gerativa, uma análise recorrente em vários autores considera o verbo *ser* em frases como (3) como sendo um chamado verbo de alçamento (*raising verb*), ou seja, um verbo que não seleciona um argumento externo. Na prática, equivale dizer que o sujeito das sentenças em que esse verbo está presente não é argumento do verbo, mas sim do predicado. Como exemplo para esse grupo de verbos, podem-se citar os chamados verbos cópula ou alguns outros verbos inacusativos, cujo exemplo protótipo fica a cargo do verbo *parecer*.

Nas construções de alçamento, um DP é movido, alçado, a uma posição em

que possa receber Caso. É justamente o fato já mencionado de que esses verbos não selecionam um argumento externo que possibilita o deslocamento do DP para a posição vazia de Spec de IP.

Predicadores infinitivos (como participípios), bem como predicadores adjetivais não podem atribuir Caso a seus argumentos, conforme reza o Filtro de Caso (CHOMSKY, 1981). Na TRL, assume-se que i) o sujeito tem Caso atribuído pelo I (finito) (núcleo de IP); e que (b) o Caso dos demais argumentos da sentença provém do V. Há uma distinção necessária a ser feita entre dois tipos de casos: o caso estrutural e o caso lexical. Müller (2016, p.106s) define caso estrutural como sendo o caso dos argumentos que é dependente do ambiente sintático em que o núcleo desse argumento ocorre, ou seja, um caso abstrato, não necessariamente marcado morfologicamente. O ambiente sintático possibilita a interpretação temática, uma vez que um dado DP não será nominativo ou acusativo a depender de uma semântica lexical, mas sim de acordo com a relação gramatical que esse DP estabelece com os demais componentes da sentença (cf. MIOTO *et al.*, 2007).

Desse modo, na frase (4) abaixo, o predicador adjetival *machucadas* não pode atribuir Caso a seu argumento *as crianças*. Tampouco o verbo *parecer* em (4) pode atribuir Caso Acusativo ao DP argumento interno *as crianças* por tratar-se de um verbo inacusativo. Sem receber Caso de nenhum predicador, o DP *as crianças* tem de ser movido para uma posição (*Spec* de IP) à esquerda do verbo para receber Caso, neste exemplo, o Caso Nominativo. Em (4), o DP *as crianças* deixa o vestígio *t* (do inglês *trace*) coindexado por *i*, índice que marca a posição em que o DP é gerado e recebe o papel temático.

(4) [IP *As crianças*_{*i*} [vp *parecem t_i machucadas*]].

Como verbo de alçamento, *parecer* não seleciona semanticamente o DP que ocupará a posição de argumento externo, sendo essa restrição feita pelo adjetivo *machucado/machucadas*. Ou seja, verbos inacusativos não s-selecionam seus sujeitos²⁵. Esse argumento é atestado pela agramaticalidade da sentença (5), pois o predicado *machucado* não aceita *livros* como argumento:

²⁵ Pode-se falar que alguns verbos considerados inacusativos c-selecionam seus AIs, como no caso dos verbos modais que selecionam um InfP ou dos verbos aspectuais *estar* e *ficar* que selecionam frequentemente um GerundP. De qualquer modo, nenhum desses verbos s-seleciona os Dps que aparecem como sujeitos, aceitando inclusive a troca de seus sujeitos por de tipos semânticos variados (MIOTO *et al.*, 2007)

- (5) *Os livros parecem machucados.

O tipo de relação temática que ocorre em (6), que se manifesta dentro do complemento do verbo *parecer*, é um tipo de relação de predicação que se chama SC (*small clause*, em português, mini-orção), uma vez que temos um argumento (*as crianças*) e um predicador (*machucadas*):

- (6) [_{sc}as crianças [_{machucadas}]]

O DP *as crianças* acima, sujeito da SC, ao combinar-se com verbos inacusativos como *parecer*, desloca-se para a posição de sujeito à esquerda do VP, deixando um traço em sua posição de origem:

- (7) [_{IP}as crianças parecem_i [_{sc} [_i] [_{machucadas}]]]

Mecanismo semelhante acontece com os chamados verbos de ligação, os quais também selecionam como argumento interno uma estrutura que contém predicação, a já referida *small clause*:

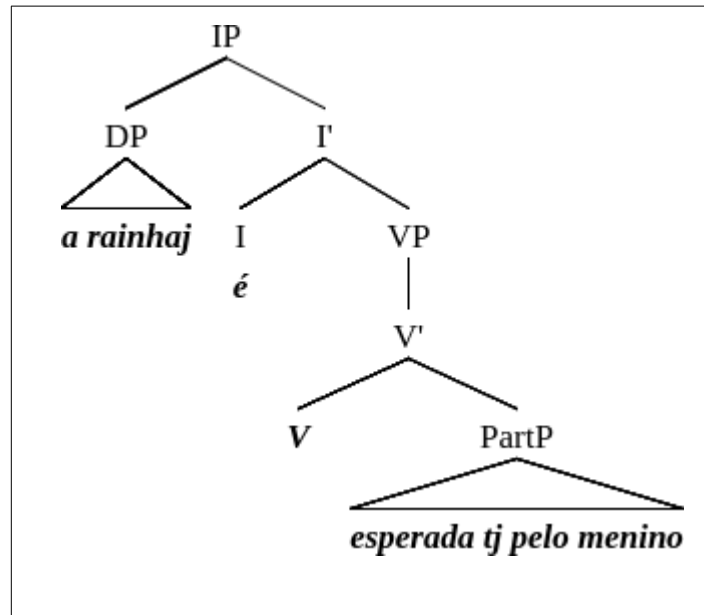
- (8) [_{IP}as crianças são [_{sc} [_i] [_{machucadas}]]]

Diz-se, tanto para as sentenças com verbos inacusativos como para aquelas com verbos de ligação, que o sujeito da sentença foi alçado para esta posição a partir da estrutura de predicação contida no complemento interno do verbo. Verbos de alçamento são, então, aqueles cujo sujeito é argumento de um outro predicador (KATO, 2015, p.52). A esta classe pertencem tanto os chamados verbos cópula ou de ligação como também os ditos verbos auxiliares.

As construções passivas assemelham-se às construções com verbos inacusativos, pois nelas o argumento interno do verbo assume a função sintática de sujeito. Na passiva analítica, por exemplo, em que há o verbo auxiliar *ser*, o verbo pleno, que contém a grade temática, permanece *in situ*, como núcleo de VP. O verbo *ser* aparece no núcleo I de IP, uma vez que não é predicador, carregando apenas as marcas de número, pessoa e tempo. É o auxiliar também que é responsável por atribuir Caso nominativo ao sujeito que é alçado da posição de complemento verbal para o Spec de IP (KATO, 2015). Veja-se o exemplo (9) abaixo, representado na FIGURA 4:

- (9) A rainha_i é *t_i* esperada pelo cavaleiro.

FIGURA 4: Estrutura de superfície de uma sentença passiva

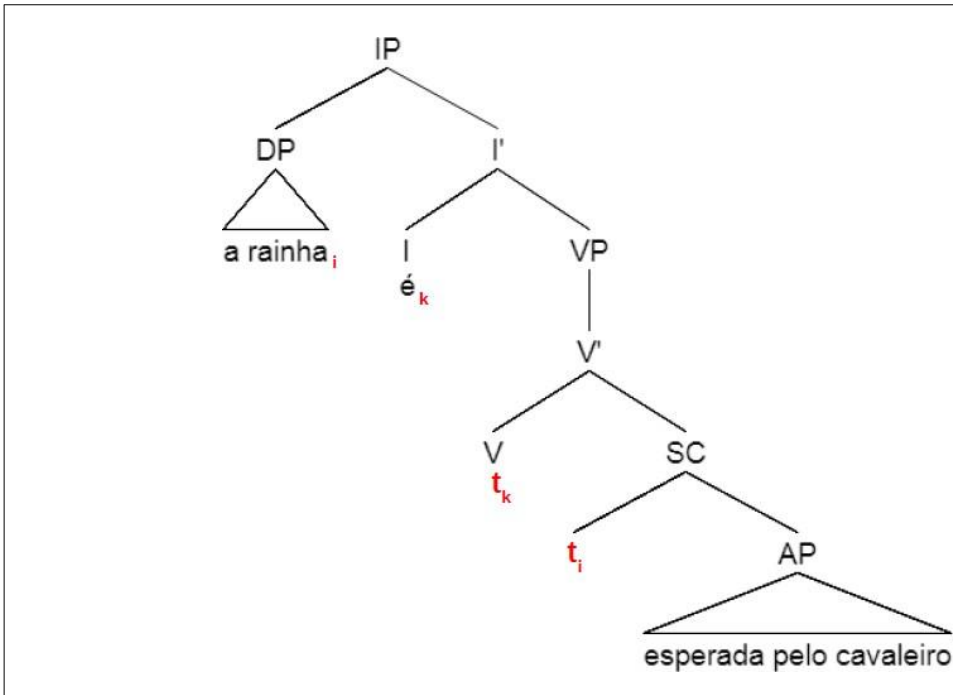


Fonte: elaborada pelo autor.

O verbo auxiliar *é* seleciona como argumento interno um sintagma participial PartP (na nomenclatura de MIOTO et al. (2007)) 'esperada', o qual, como visto, não pode atribuir Caso ao seu argumento externo 'a rainha'. Em (8), o DP argumento externo *a rainha* que ocupa a posição de sujeito do verbo, *é*, na verdade, um sujeito pertencente a grade temática do verbo *esperar*, onde *é* gerado e interpretado semanticamente, recebendo deste último papel temático. No exemplo, o sintagma *a rainha* não é argumento do predicado *é*, mas sim do predicado contido na SC.

No exemplo analisado, o particípio pertence à classe dos verbos. Numa visão não canônica que nos servirá de base para a implementação da Gramática G-A, a mesma sentença apresentará uma estrutura semelhante àquelas com verbo de ligação e uma SC. Veja-se a FIGURA 5 abaixo:

FIGURA 5: Estrutura de superfície de uma sentença passiva analisada como predicativa adjetival



Fonte: Elaborada pelo autor.

2.4.3 Sobre a aquisição das sentenças passivas

Teixeira e Othero (2018), tratando da aquisição tardia de passivas verbais com verbos de não ação (tais como *ver*, *ouvir*, *gostar*) - o chamado *efeito Maratsos* - apresentam uma das explicações mais aceitas para o fenômeno, segunda a qual essa demora ocorreria sobretudo devido ao amadurecimento tardio de um princípio linguístico, o de formação de cadeias-A, sem o qual a criança seria incapaz tanto de compreender quanto de produzir alguns tipos de sentenças passivas.

Por cadeias-A, entende-se uma cadeia sintática formada pelo deslocamento de um NP (ou DP, em abordagens gerativas mais modernas) de uma posição argumental para outra posição argumental. Uma posição argumental, também chamada de posição-A, é “uma posição na sentença que é identificada como uma função gramatical argumental (como sujeito, objeto direto, objeto de preposição etc.)” (TEIXEIRA & OTHERO, 2018, p.3243). O leitor já deve ter notado que as sentenças passivas são um exemplo de sentenças contendo as referidas cadeias-A, cujo movimento se justifica, dentro do escopo da TRL, pelo Filtro de Caso. O exemplo (9) acima contém uma cadeia-A, cuja obrigatoriedade é atestada pela agramaticalidade do exemplo (10) abaixo:

(10) * É a rainha esperada pelo cavaleiro.

Sendo assim, a sentença (9) presta-se como exemplo de movimento-A com conseqüente formação de cadeia-A, a qual é necessária para conferir gramaticalidade a uma sentença passiva verbal, uma vez que a não realização do movimento e não formação da cadeia implica em uma seqüência agramatical.

Dentro de uma teoria racionalista e maturacional de aquisição de linguagem (cf. BORER & WEXLER, 1987) as crianças não entenderiam, até certa idade, sentenças passivas do tipo verbal porque sua GU ainda não estaria equipada com princípios para formar cadeias-A.

As passivas adjetivas, entretanto, não apresentam cadeias-A, razão pela qual são encontradas mais cedo na fala de crianças, não entrando no rol das sentenças afetadas pelo *efeito Maratsos*. A razão para isso é que sentenças passivas adjetivas (em PB realizadas com os auxiliares *estar/ficar*) seriam interpretadas como adjetivos, ao passo que sentenças passivas verbais não permitiriam essa interpretação. Teixeira e Othero (2018, p.3244) ilustram essa diferença com os seguintes exemplos:

(11) a) O brinquedo foi estragado (pelo menino).

b) O brinquedo estragado.

(12) a) O gato foi visto (por Maria).

b) *O gato visto.

Na alimentação de nosso léxico, não consideramos uma distinção formal entre verbos de ação e verbos de não ação. Encorajamos que outros trabalhos sejam feitos com o intuito de implementar de modo mais robusto esse tipo de distinção entre as duas classes de verbos.

3 DOS FUNDAMENTOS TEÓRICOS DA LFG

Os principais pressupostos teóricos que subjazem nossa pesquisa e representam a base para a análise de dados e para a proposta de implementação serão apresentados na seguinte seção. Inicialmente, apresentaremos o formalismo da LFG e, mais adiante, sua implementação no ambiente do XLE. A discussão acerca de como se vem tratando o fenômeno da passiva na LFG fecha a seção de Fundamentação deste trabalho com o objetivo de mostrar os pontos sobre os quais serão testadas nossas hipóteses de análise.

Devido ao fato de a Gramática Léxico-Funcional não gozar, no Brasil, de uma tradição e divulgação em meio acadêmico, mesmo dentro do âmbito do gerativismo, achamos pertinente fazer um breve esboço dessa teoria, trazendo à luz aspectos importantes de sua origem e bases epistemológicas, pontuando, sobremaneira, aquilo que a diferencia de outras teorias linguísticas. A respeito da história desse formalismo, vejamos:

Lançada em 1982 por uma equipe multidisciplinar das áreas de Linguística, Matemática e Computação ligada ao MIT, à Universidade Stanford e ao Centro de Pesquisa de Palo Alto da Xerox (*Xero Palo Alto Research Center - PARC*), a Gramática Léxico-Funcional (...) é um dos modelos de formalismo gramatical de maior divulgação, cujo uso é fortemente empregado não apenas na Linguística Computacional, mas também na Tipologia Linguística (ALENCAR, 2004, p. 174).

Ainda que lançada oficialmente nos anos 80, as bases do modelo da LFG surgem na década de 70 com a colaboração de estudiosos das áreas de linguística descritiva, computação e psicolinguística, sobretudo, dos trabalhos de dois estudiosos, Joan Bresnan e Ronald M. Kaplan (FALK, 2001, p.3). A primeira era uma sintaticista e ex-aluna de Chomsky, a qual se inquietava em razão de evidências psicolinguísticas que pareciam apontar para algo de errado nas teorias transformacionais. Não obstante, a LFG é uma teoria que, como veremos, não admite transformações sintáticas, como aquelas propostas por Chomsky (1957) em seu trabalho seminal, ou movimentação de constituintes, presentes em modelos posteriores do gerativismo, tais como P&P e PM.

A LFG foi (cf. SCHWAZRE & ALENCAR, 2016, p.1), de antemão, tão bem elaborada, à época de sua criação, que nenhuma mudança radical teve de ser feita desde sua primeira formulação no início dos anos 80, garantindo-se, assim, que o conhecimento científico pudesse progredir continuamente.

Segundo Bersnan (2001) e Falk (2001), a LFG representa um modelo de

teoria linguística gerativa formulado em uma linguagem computacional. Trata-se de um formalismo elegante, pois sua base constitui-se de regras gramaticais simples; aspectos mais complexos são tratados em quase sua completude em nível lexical.

Esse modelo, diferentemente de outros modelos gerativos da corrente chomskiana, tais como a Teoria de Princípios e Parâmetros (PP), surgida nos anos 80, e o Programa Minimalista (PM), mais recente, lançado em 1995, é baseado em teorias não-derivacionais, ou ainda, em teorias lexicalistas calcadas na unificação (*unification-based*) ou teorias que tomem por base restrições (*constraint-based*).

Transformações, no âmbito da teoria gerativa, podem ser entendidas como “operações que derivam uma estrutura sintagmática a partir de uma outra estrutura sintagmática já formada. Tipicamente a transformação de uma estrutura em outra resulta do movimento de constituintes” (ALENCAR, 2006, p. 6). As transformações sintáticas foram necessárias na teoria gerativa de vertente chomskyana, devido ao fato de que unicamente as regras de estrutura sintagmática não eram suficientes para dar conta de todos os fenômenos e sentenças de uma língua. Recorreu-se pois as transformações de modo que novas sentenças pudessem ser derivadas por meio de movimento ou apagamento de constituintes, bem como relocação de sintagmas.

Dizer ser a LFG um modelo gerativo não transformacional deve-se ao fato de que, nesse modelo, toda ênfase é dada ao léxico, isto é, a LFG está embasada em uma teoria conhecida como lexicalista. Este fato a diferencia bastante dos modelos de TRL e PM acima mencionados, uma vez que, não havendo transformações sintáticas, as informações gramaticais distribuem-se em níveis de representação paralelos que estão relacionados através de um sistema de correspondências, sendo dois os níveis mais importantes: a estrutura constitucional (estrutura C) e a estrutura funcional (estrutura F) (*op. cit.*). Entretanto, é mister notar que a LFG constitui, sem sombra de dúvidas, uma teoria gerativa, pois “rejeita as premissas da teoria transformacional, mas não seus objetivos” (FALK, 2001, p.2)²⁶. A LFG pretende, assim como sintaxe gerativa de Chomsky, descobrir a natureza da faculdade da linguagem humana, especialmente da Gramática Universal (GU). A respeito da teoria gerativa, Carnie (2002, p. 5), por exemplo, afirma:

A tese que subjaz à gramática gerativa é que sentenças são geradas por um conjunto de procedimentos subconscientes (como programas de

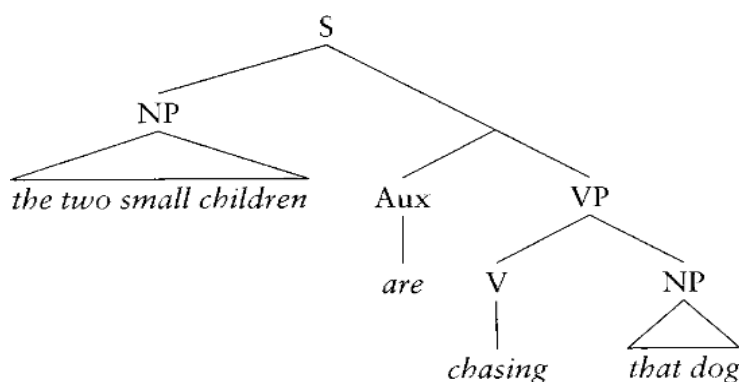
²⁶ No original: “LFG rejects the assumptions of transformational theory, nor its goals”.

computador). Esses processos são parte de nossas mentes (...). O objetivo da teoria sintática é modelar esses processos.²⁷

O objetivo imediato da abordagem gerativa é o de desenvolver modelos formais de vários aspectos da linguagem. Para Falk (2001), é justamente por meio de tais modelos que concepções formais sobre a linguagem podem ser expressas e testadas.

Ainda que a LFG mantenha uma base epistemológica calcada na teoria gerativa, achamos oportuno discutir quais características diferenciam aquela desta abordagem. Em primeiro lugar, podemos considerar o *design* configuracional da GU, problematizado por Bresnan (2001) através do exemplo abaixo:

FIGURA 6 - Estrutura sintagmática da sentença *The two small children are chasing that dog.*



Fonte: Bresnan (2001, p.5).

O exemplo acima é analisado pela autora a fim de se constatar um problema no *design* configuracional da GU, sobretudo no que tange a grande variedade de modos de expressão presentes em diferentes línguas, as quais diferem largamente na maneira pela qual agrupam ideias similares em palavras ou sintagmas (*id.*, *ib.*). A ideia de que “duas crianças pequenas estão perseguindo um cachorro” é expressa, em língua inglesa, através do significado obtido pela estrutura da FIGURA 6, em que unidades menores se agrupam para formar sintagmas, cujos elementos, por sua vez, não podem ser simplesmente separados e reorganizados de maneira aleatória na sentença sem que daí resulte agramaticalidade atestada nos exemplos abaixo (op. cit., p. 5):

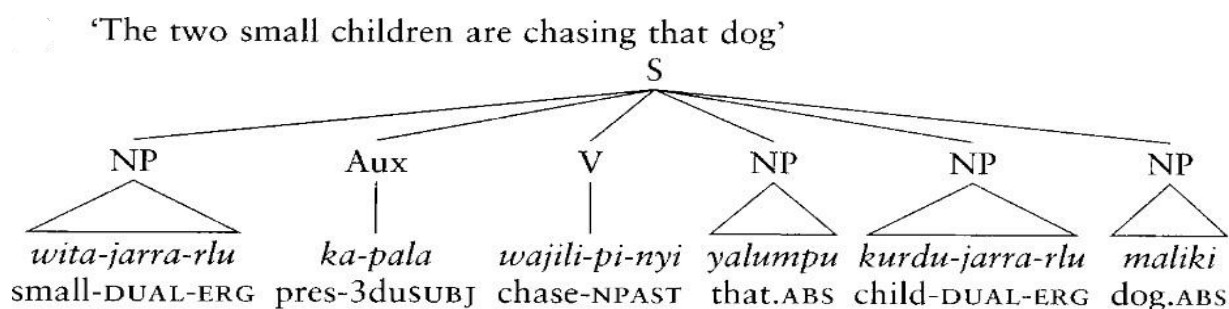
- (1) a. * *The two small are chasing that children dog.*

²⁷ No original: “The underlying thesis of generative grammar is that sentences are generated by a subconscious set of procedures (like computer programs). These procedures are part of our minds (...). The goal of syntactic theory is to model these procedures”.

- b. * *The small are dog chasing children that.*
 c. * *Chasing are the two small that dog children.*
 d. * *That are children chasing the two small dog.*

O que foi discutido para o inglês não se aplica, contudo, a outras línguas, como por exemplo, warlpiri, língua nativa australiana, não configuracional. Isso significa que quaisquer combinações dos elementos presentes na sentença da FIGURA 7 abaixo são válidas, não alterando o sentido da sentença. A única restrição é a de que o verbo auxiliar ocupe, em warlpiri, a segunda posição da frase.

FIGURA 7 - Estrutura sintagmática de uma sentença em warlpiri



Fonte: Bresnan (2001, p.6).

O que se verifica é que línguas que apresentam uma morfologia mais rica, ilustradas aqui em warlpiri pela palavra *criança*, *kurdu-jarra-rlu*, que apresenta sufixos para o dual e para o caso ergativo, podem utilizar-se com maior liberdade de fixidez ou não de estruturas sintagmáticas. Por outro lado, línguas mais pobres morfologicamente tendem a ter uma estrutura sintagmática hierárquica mais rígida, como é o caso do inglês (BRESNAN, 2001). Sendo assim, o *design* configuracional de gramáticas gerativas de vertente transformacional não parece ser apropriado a línguas não-configuracionais²⁸.

Outro caso cuja análise se diferirá na abordagem proposta pela LFG e por outros modelos gerativos será o da voz passiva. A voz passiva é tradicionalmente analisada como sendo um típico caso de derivação de uma estrutura profunda ativa. Devido ser uma teoria de vários níveis paralelos de representação, não há na LFG necessidade de transformações sintáticas, haja vista que a estrutura C, único nível de estrutura de constituintes, reflete diretamente aquilo que se observa de fato nas línguas

²⁸ Um exemplo didático de implementação de uma minigramática LFG para o latim, língua casual essencialmente não-configuracional, é proposto por Alencar (2014)

(ALENCAR, 2004, p.177). Não há transformações agindo em uma estrutura profunda para gerar uma estrutura de superfície. Daí ser a LFG uma teoria gramatical baseada em restrições (*constraint-based*), pois a gramaticalidade não depende de propriedades de derivações, mas sim da satisfação simultânea de restrições (FALK, 2001). Pela mesma razão, categorias vazias como vestígios (*traces*) e elementos pronominais vazios, comuns na TRL, não são contempladas num modelo léxico-funcional.

Como se pode depreender do próprio nome do formalismo, a LFG constitui-se de uma teoria lexicalista e funcional. O que significa, porém, essa dupla qualificação? Sobre o caráter léxico-funcional da teoria, vejamos o que assevera Dalrymple (2005):

A Gramática Léxico-Funcional é uma teoria da estrutura da língua e de como diferentes aspectos da estrutura linguística são relacionados. Como o nome sugere, a teoria é lexical: o léxico é ricamente estruturado, apresentando, em vez de transformações ou operações nas árvores de estrutura sintagmática, relações lexicais como um meio de captar as generalizações linguísticas. É também funcional (grifo nosso): funções gramaticais como sujeito e objeto são elementos primitivos da teoria, não definidos em termos da configuração da estrutura sintagmática ou dos papéis semânticos. (DALRYPLE, 2005, p. 82)²⁹

Em livro lançado na Alemanha em 2016, Schwarze e Alencar apresentam o *modus operandi* a partir do qual se pode desenvolver uma gramática em LFG/XLE em uma língua românica moderna. Nesse caso bem específico, tomaram o francês como modelo. Em seu livro intitulado *Lexikalische-funtionale Grammatik: Eine Einführung am Beispiel des Französischen mit computer-linguistischer Implementierung* (em tradução livre para o português: *Gramática Léxico- Funcional: Uma introdução a partir do francês com implementação linguístico-computacional*), encontram-se listadas as vantagens que o modelo gramatical da LFG oferece:

- É um sistema heurístico que estimula a observação exata de fatos sintáticos e morfológicos.
- É um método para a representação e testagem de declarações sobre fatos linguísticos e seu contexto.
- O modelo é explicitado de tal modo que gramáticas-LFG possam ser implementadas e testadas em computador. Isso pode acontecer, dentre outras maneiras, com a ajuda do sistema *Xerox Linguistic Environment* (XLE) que gera um *parser* a partir de uma gramática-LFG, o qual pode ser utilizado para a análise sintática automática (SCHWARZE & ALENCAR, 2016, p.1).³⁰

²⁹ No original: “Lexical Functional Grammar is a theory of the structure of language and how different aspects of linguistic structure are related. As the name implies, the theory is lexical: the lexicon is richly structured, with lexical relations rather than transformations or operations on phrase structure trees as a means of capturing linguistic generalizations. It is also functional: grammatical functions like subject and object are primitives of the theory, not defined in terms of phrase structure configuration or semantic roles.”

³⁰ No original: “– Es ist ein heuristisches Instrument, das zur genauen Beobachtung syntaktischer und morphologischer Fakten anregt. – Es ist eine Methode zur Darstellung und Überprüfung von Aussagen

Para finalizarmos esta seção, apresentamos o quadro comparativo abaixo sobre a LFG e os modelos gerativos transformacionais (GTs), cujas características enunciamos de modo resumido:

Quadro 1 - Quadro comparativo entre LFG e modelos gerativos transformacionais

LFG	GTs
Design relacional da gramática universal (BRESNAN, 2001, p.10);	Design configuracional da gramática universal;
Relações gramaticais (sujeito, objeto etc.) são primitivos sintáticos (Alencar, 2004);	Relações gramaticais (sujeito, objeto) são noções subsidiárias;
Componente sintático contraído e componente lexical expandido (maior poder explicativo, permitindo unificação de pesquisa linguística com psicolinguística (BRESNAN, KAPLAN, 1982, p. 20));	Componente sintático expandido e componente lexical contraído;
Estruturas paralelas e não-derivadas (BRESNAN, 2001);	Estruturas seriais e derivadas;
Representação monoestratal.	Representação em estrutura profunda e superficial.

Fonte: Elaborado pelo autor.

Apresentadas as características e vantagens de que esse modelo de formalismo dispõe, discutiremos, a seguir, as abordagens linguísticas de que a LFG se apropriou.

3.1 A gramática de estrutura sintagmática (PSG)

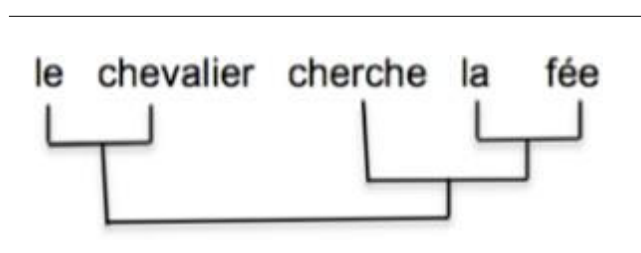
Segundo Schwarze e Alencar (2016, p.2), a gramática de estrutura sintagmática é uma concretização do pensamento de que uma frase não tem uma estrutura linear, mas sim, uma estrutura organizada hierarquicamente. Tal aceção, como salientado pelos autores, não é nova, já estando presente na própria gramática tradicional, quando esta trata de coordenação e subordinação, havendo assim, orações principais e subordinadas.

über sprachliche Fakten und ihren Zusammenhang. – Das Modell ist in dem Sinne explizit, dass LFG-Grammatiken auf dem Computer implementiert und überprüft werden können. Dies kann u.a. mit Hilfe des Systems Xerox Linguistic Environment (XLE) geschehen, das aus einer LFG-Grammatik einen Parser erzeugt, der für die automatische Syntaxanalyse verwendet werden kann”

Somente com o Estruturalismo, entretanto, é que o pensamento acerca da estrutura hierárquica da frase é desenvolvido e formalizado, sobretudo por meio da chamada análise de IC de R.S. Welss, em que IC significa *immediate constituents* (constituintes imediatos). Um exemplo de análise de IC é apresentado por Schwarze e Alencar (*op. cit.*) para a seguinte frase em francês:

- i. Le chevalier cherche la fée.³¹
 the:M;SG knight(M)[SG] seek:PRS;3ps the:F;SG fairy(F)[SG]³²
 ‘The knight seeks the fairy.’

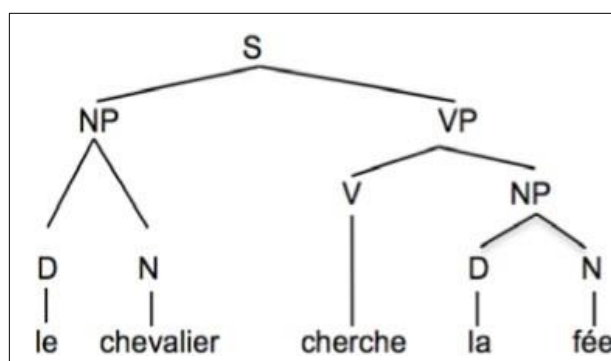
FIGURA 8 - Análise IC do exemplo (i).



Fonte: Schwarze e Alencar, 2016, p.2.

A análise de IC é, na verdade, apenas de um modelo de observação, não de uma gramática propriamente dita. Esse modelo, todavia, serviu de base para a gramática de estruturas sintagmáticas, como se pode depreender da análise da árvore sintagmática da mesma sentença (i):

FIGURA 9 - Estrutura sintagmática do exemplo (i).



Fonte: Schwarze e Alencar, 2016, p.3.

A modo de exemplo, uma gramática de estrutura de constituintes (PSG, do

³¹ Tradução: O cavaleiro procura a fada.

³² Glosa feita a partir do documento de regras de glosas interlineares de Leipzig (*The Leipzig Glossing Rules*), disponível em: <<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>>. Acesso em: 06 jun.2017.

inglês *phrase structure grammar*) de uma árvore sintagmática como a apresentada na FIGURA 9 é mostrada em seguida. Vale aqui mencionar os conceitos-chave de nó e dominância: “Cada constituinte que esteja, em uma árvore, sobre um outro ou mais constituintes forma um nó. Quando um nó A se encontra sobre um nó B, dizemos que A domina B” (*op. cit.*). Em um diagrama arbóreo fala-se também de “nó pai” e “nós filhos”.³³

- (2)
- S → NP VP
 - VP → V NP
 - NP → D N
 - V → cherche
 - N → chevalier
 - N → fée
 - D → le
 - D → la

É interessante notar que mesmo uma simples gramática como a exemplificada nesta seção já apresenta um caráter gerativo considerável, desde que as regras sintagmáticas sejam respeitadas. A gramática PSG acima é perfeitamente capaz de gerar uma sentença do tipo (ii) abaixo:

- ii. La fée cherche le chevalier.³⁴
 the:F;SG fairy(F)[SG] seek:PRS;3ps the:M;SG knight(M)[SG]
 ‘The fairy seeks the knight.’

Entretanto, se uma dada gramática contiver unicamente regras do tipo PSG como as explicitadas acima, não seria possível, inicialmente, dar conta do fenômeno de concordância presente nas línguas naturais. Como asseveram Othero e David (2011), as regras de uma PSG são, na verdade, macrorregras de boa formação gramatical de sentenças que não consideram a unificação de traços de gênero ou número, por exemplo. Adiante, apresentamos como o modelo teórico da LFG lida com os fenômenos sintáticos diversos de distribuição, dependência, concordância e regência e explicitamos também sua implementação no ambiente XLE.

3.2 A teoria de valência

³³ No original, fala-se de “nó mãe” (*Mutter*) e “nós filhas” (*Töchter*). Preferiu-se aqui utilizar as palavras masculinas, por tratar-se de uma palavra masculina em português.

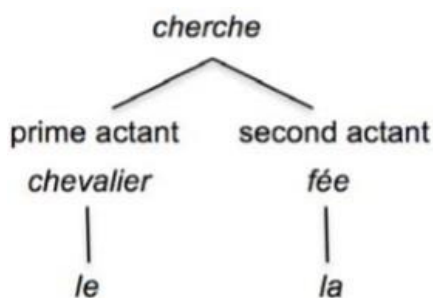
³⁴ Tradução: A fada procura o cavaleiro.

A teoria de valência representa uma explicação do termo tradicional regência (*op. cit.*, p.4) e baseia-se na ideia de que determinadas palavras abrem, na frase, espaços vazios que necessitam ser preenchidos de uma maneira definida. Tal concepção influencia e introduz na linguística moderna a chamada Gramática de Dependência (DG, *Dependency Grammar*).

Resumidamente, pode-se dizer que uma DG reconhece apenas nós lexicais, e não sintagmáticos, que estão ligados binariamente de maneira direta para formar uma árvore de dependência.

Tesnière (1959 *apud*. SCHWARZE & ALENCAR, 2016), fundador da teoria de valência, calca-a no conceito químico de valência, segundo o qual elementos químicos possuem a capacidade de formar cadeias, ligando-se a outros elementos. A partir de uma metáfora com uma peça teatral, Tesnière esclarece o conceito de valência verbal na medida em que o compara com uma cena de um drama, na qual há atores com um papel determinado pelo diretor. A figura do diretor seria ocupada pelo verbo, o qual determina que atores entrariam em cena e com quais funções. (PITTNER & BERMAN, 2007). Além dos atores, há também a presença de outros elementos que compõem a *mise-en-scène*. Tesnière distingue os elementos presentes ‘no palco frasal’ em duas categorias: os *actants* e os *circunstants* (actantes e circunstantes), que, na terminologia da LFG, são, respectivamente, funções e adjuntos. É o verbo que determina quais actantes existem; enquanto os circunstantes não são por ele determinados. Tomemos novamente o didático exemplo aqui apresentado na sentença (i) e vejamos sua representação da estrutura de valência:

FIGURA 10 - Estrutura de valência do exemplo (i)



Fonte: Schwarze e Alencar (2016, p.3).

Da FIGURA 10 pode-se depreender que sujeito e objeto são, na gramática de valência de Tesnière, actantes igualmente dependentes do verbo e, com isso, a visão tradicional de sujeito como elemento que não é exigido pelo verbo é posta em xeque.

Neste modelo representado por uma árvore dependencial que não admite nós sintagmáticas, mas sim apenas nós lexicais, o regente máximo é o verbo, ao qual encontram-se ligados os demais elementos. Na LFG, o sujeito é selecionado pelo verbo de modo análogo à seleção realizada pelo verbo para os demais complementos, o que vai de encontro a outras abordagens gerativas como a Teoria da Regência e da Ligação (TRL) (Raposo, 1992:94).

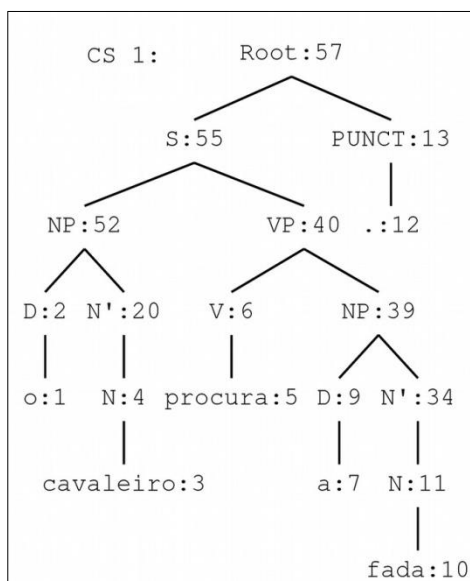
Segundo Schwarze e Alencar (2016, p.4), um modelo de gramática completo deve conter tanto as relações entre os constituintes quanto as valências dos elementos de uma sentença. Tal exigência é preenchida pela LFG, respectivamente, por meio das estruturas C e estrutura F. Desse modo, diferentemente do gerativismo chomskyano, que é monoestratal, apresentando apenas uma estrutura representacional, o esquema arbóreo, a LFG é um tipo de formalismo constituído por duas estruturas representacionais paralelas e de igual importância.

3.3 A teoria das funções gramaticais

Grosso modo, essa teoria, cujas origens remontam à gramática tradicional, organiza cada um dos elementos da frase em categorias, tais como sujeito, predicado, objeto etc. Apesar do que se observa no estruturalismo e na corrente principal da gramática gerativa, nos quais essas categorias foram consideradas superficiais, e seu uso é restrito a contextos informais, a LFG, por seu turno, utiliza-se de tais categorias para uma reconstrução teórica do conceito de valência (SCHWARZE & ALENCAR, 2016, p.5).

A fim de obter uma visão global do que foi aqui apresentado sobre a LFG, observemos a FIGURA 11 abaixo, que representa a sentença (iii) implementada por meio do sistema XLE:

FIGURA 11 - Análise da sentença (iii) por um fragmento de gramática em XLE



Fonte: Elaborado pelo autor através do XLE.

O que vemos na FIGURA 11 é a chamada estrutura C da LFG (estrutura de constituintes). “A estrutura C de uma frase nada mais é do que a estrutura que se pode gerar com um PSG” (*op. cit.*, p.16)³⁵. Carnie (2002, p.438) define a estrutura C como as árvores que são geradas, a partir das regras, na LFG. Segundo ele, a estrutura C equivaleria aproximadamente à estrutura de superfície (*S-structure*) do Programa de Princípios e Parâmetros (PP), de Chomsky, fazendo-se a ressalva de que, por não conter transformações, a LFG não apresentaria uma estrutura profunda (*D-structure*) presente no PP. Nessa estrutura, nas mais diferentes línguas, pode-se observar como palavras agrupam-se em sintagmas, e como estes últimos em sentenças

Além da estrutura de constituintes, há na LFG, a chamada estrutura funcional (estrutura F). Para Santos (2014, p.50), “a estrutura-f codifica princípios universais de construções na forma de uma matriz de atributos e valores, baseada em anotações funcionais nos nós da estrutura-c e a informação vinda do léxico”. Estas estruturas, por sua vez, expressam relações oracionais, informações sintáticas abstratas relacionadas a funções gramaticais, dependências e subcategorização. A estrutura F consiste de atributos abstratos (traços e funções), cujos valores (FALK, 2001) são descritos nas chamadas matrizes de atributos e valores. As funções podem ser classificadas em funções gramaticais e funções discursivas. Essas últimas, p.ex. Foco e Tópico, não foram por nós trabalhadas, embora seja possível sua implementação em LFG. É através do papel argumental de Foco, por exemplo, que a LFG explica,

³⁵ No original: “Die k-Struktur eines Satzes ist nichts anders als die Struktur, die man mit einer PSG erzeugen kann”.

novamente sem recorrer ao deslocamento de constituintes, os casos de frases interrogativas em que o pronome interrogativo não aparece *in situ*.

Os atributos podem ter para os valores tanto símbolos, como formas semânticas, estrutura F ou conjuntos. São as entradas lexicais que demonstram a contribuição que um item lexical traz para a estrutura F. Em termos de estrutura C/estrutura F, Carnie (2002, p.440) afirma que toda informação que termina na estrutura F inicia-se com a entrada lexical das palavras que compõem uma sentença. Cada nó da estrutura C projeta, através de um mecanismo matemático de equações funcionais, uma função na estrutura F.

Há pouco, afirmamos ser a LFG uma teoria lexicalista, uma vez que é no léxico que muito do trabalho é realizado. Por ser um modelo lexicalista, na LFG, assim como em outros modelos como na GPSG – *Generalized Phrase Structured Grammar* (Gadzar *et al.*, 1985) e na HPSG - *Head-Driven Phrase Structure Grammar* (Pollard e Sag, 1994), o léxico não só contém vasta informação sobre as palavras da língua, como também serve de critério para a boa formação de sentenças nesse modelo linguístico. No léxico podem estar contidas informações as mais diversas, tais como: matriz fonética; forma lexical; categoria sintática; estrutura argumental; restrições de seleção semântica; forma semântica; papéis temáticos; relações gramaticais entre uma função gramatical e um papel temático; traços semânticos como [+/- humano], [+/- específico], [+/- definido], [+/- animado], [+/- concreto], além de atributos para gênero, número, pessoa, voz, caso, tempo, aspecto, modo e tipo de oração etc; traços de reflexividade e ergatividade. (BRESNAN, 1982)

Vejam os exemplos de entrada lexical para forma verbal “procura” de nossa sentença (iii), com a qual estamos trabalhando. As entradas lexicais contêm, em geral, uma representação da forma do item, a categoria sintática a que ele pertence e uma lista de esquemas funcionais:

- (3) procura V (↑PRED) = ‘procurar’ <(↑SUBJ),(OBJ)>
 (↑TENSE) = present
 (↑SUBJNUM) = sng
 (↑SUBJPERS) = 3rd

Pode-se depreender de (3) que, em LFG, os itens lexicais estão abastecidos

com esquemas funcionais. A entrada lexical acima, por exemplo, nos informa que *procura* é um verbo que significa “procurar” e que toma dois argumentos obrigatórios (contidos nos parênteses angulares): um sujeito e um objeto. Todos os itens lexicais que tenham um significado (e.g., substantivos, verbos, adjetivos etc.) apresentam um traço PRED com um valor correspondente, e a valência de um dado núcleo é especificada no valor desse PRED (MÜLLER, 2016, p.215). Também nos é informado que *procura* está na forma temporal presente. Uma entrada lexical distinta, como *procurava*, teria um valor diferente para TENSE. Por fim, sabemos que o sujeito de *procura* está na terceira pessoa do singular; (\uparrow SUBJ NUM) significa dizer “o número do meu sujeito é...”. O símbolo \uparrow é uma metavariável que se refere ao nó imediatamente dominante. De modo análogo, a metavariável \downarrow refere-se ao próprio nó. Resumidamente, temos:

- 4) a. \downarrow significa este nó;
 b. \uparrow significa “meu pai” (nó imediatamente dominante)

Neste trabalho, utilizamos o *software* XLE para implementar o formalismo da LFG. Em comparação com a entrada lexical mostrada em (3), vejamos a entrada lexical correspondente de ‘procura’ na Figura abaixo, onde podem ser observadas algumas adaptações: por exemplo, em XLE, as metavariáveis \uparrow e \downarrow são representadas, respectivamente, por \wedge e $!$.

FIGURA 12- Entrada lexical de *procura*

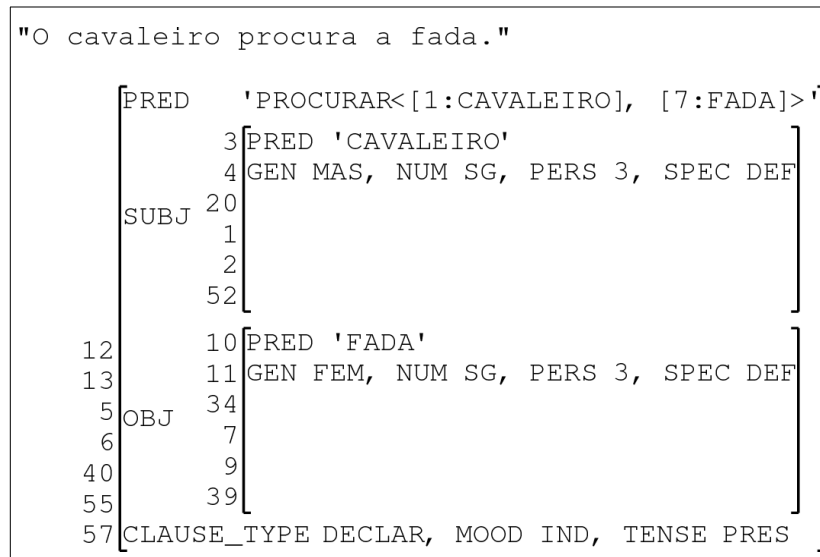
```

|
| procura   V * ( $\wedge$  PRED)='PROCURAR<( $\wedge$  SUBJ) ( $\wedge$  OBJ)>'
|                   ( $\wedge$  SUBJ PERS)=3
|                   ( $\wedge$  SUBJ NUM)=SG.
|

```

Fonte: Elaborado pelo autor a partir do XLE.

Podemos, então, apresentar também a estrutura F da sentença *O cavaleiro procura a fada* na FIGURA 13 abaixo. Nela, o traço SPEC (do inglês, *specifier*) representa o determinante do sintagma nominal *o cavaleiro*, podendo esse traço apresentar dois valores distintos, a saber, DEF para definido e INDEF para indefinido.

FIGURA 13 – Estrutura F da sentença *O cavaleiro procura a fada.*

Fonte: Elaborado pelo autor a partir do XLE.

A estrutura F da FIGURA 13 expressa os seguintes fatos:

- A sentença é constituída de um sujeito, de um verbo e de um objeto³⁶;
- O sujeito tem o significado lexical 'CAVALEIRO', é do tipo definido, e seus traços de concordância são masculino, singular e de terceira pessoa;
- O verbo tem o significado lexical de 'PROCURAR' e determina, por sua valência, que a presença de um sujeito e de um objeto é obrigatória;
- O objeto tem o significado lexical 'FADA', é definido, feminino, singular e de terceira pessoa.

Ponto importante ainda a ser mencionado é que, na FIGURA 11, encontram-se, em cada um dos nós, índices numéricos que servem para equacionar um dado nó da estrutura C a uma estrutura funcional contida no nível de representação mais abstrato, como apresentada, a estrutura F da FIGURA 13. No XLE, tais índices vão ao encontro do princípio de que há alguma estrutura funcional associada com algum nó da estrutura de constituintes (WESCOAT, 1989, p.8). Por ser um nível representacional mais abstrato, na estrutura F, diferentemente da ordem definida em que os elementos encontram-se apresentados na estrutura C, a ordem em que os atributos estão

³⁶ A função OBJ na LFG corresponde a objeto direto; Há ainda as funções de OBJ2 (dativo) e OBL (objeto preposicionado).

distribuídos não é relevante, haja vista também que tais atributos estão enumerados com os nós da estrutura C, o que garante a legibilidade da sentença.

Nas FIGURA 11 e 13, observa-se como esses índices se relacionam com a estrutura funcional do verbo *procurar* e seus argumentos. Na próxima seção, explicamos melhor como se dá a ligação entre as estruturas C e F.

3.3.1 Esquemas funcionais

A arquitetura na qual a LFG está erigida permite que diferentes níveis de informação linguística, dos mais concretos aos mais abstratos, sejam modelados por níveis representacionais distintos, os quais se encontram interligados de maneira formal por projeções funcionais. Na prática, como já mencionado, estamos falando do mecanismo de correspondência que projeto para cada nó da estrutura C uma função distinta da estrutura F.

Uma consequência dessa arquitetura de projeções é a de que os níveis representacionais restringem-se mutuamente, e, por conseguinte, uma análise só pode ser considerada bem-sucedida se as informações na estrutura F estão completas e consistentes, e, concomitantemente, se as regras de estrutura sintagmática licenciam uma dada sentença (BUTT, 2007).

Para que essa correspondência ocorra, em LFG, as regras sintagmáticas encontram-se anotadas com esquemas funcionais, os quais, para uma dada sentença, permitem a ligação entre as estruturas C e F. Os esquemas funcionais são anotações funcionais transferidas para a árvore da estrutura C, inseridas ao lado direito das categorias das regras livres de contexto. Apresentamos brevemente nesta seção o modo como se dá essa ligação baseando-nos sobretudo em Wescoat (1989).

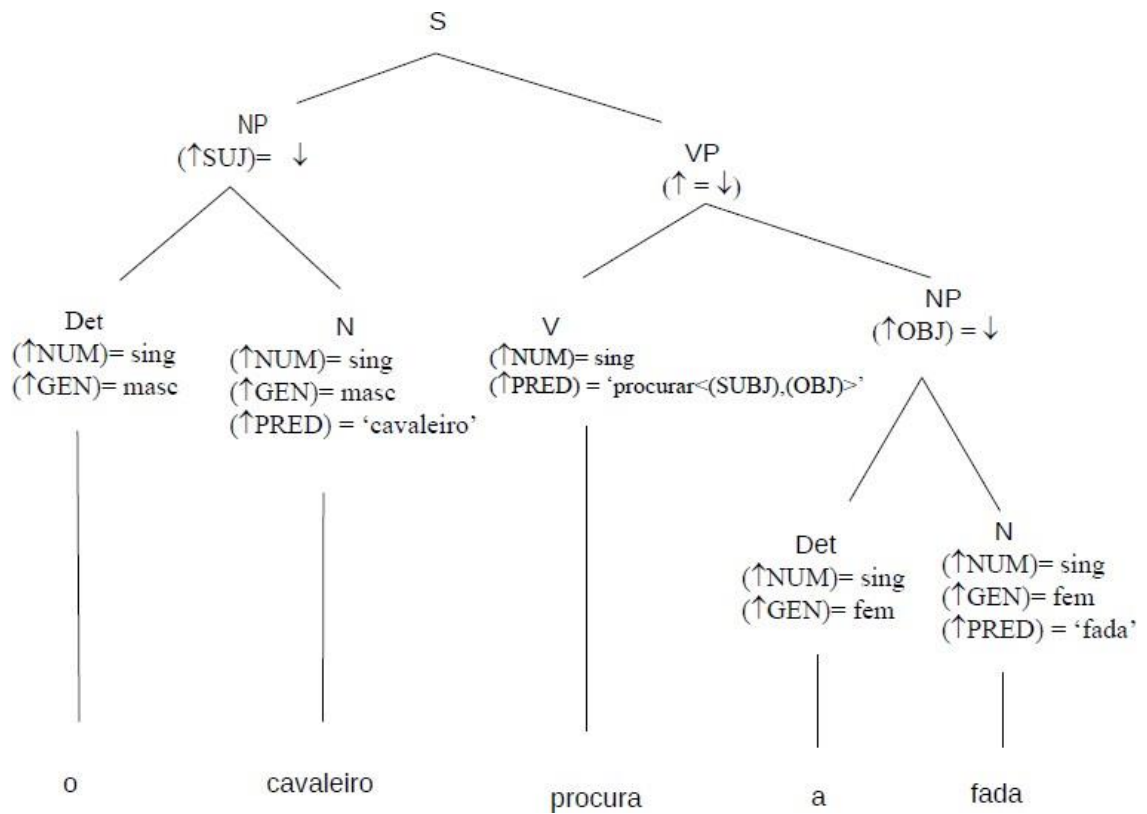
Primeiramente deve-se observar que os esquemas funcionais estão associados a cada item da estrutura C, ligação esta estabelecida tanto por meio da gramática, nas regras sintagmáticas anotadas, como por meio dos esquemas funcionais contidos no léxico. Ou seja, numa gramática LFG, há informações concernentes à estrutura F tanto no conjunto de regras que descrevem e permitem gerar sentenças gramaticais para um dado excerto de uma língua, como também no léxico de nossa gramática, cujas entradas lexicais estão enriquecidas com esquemas funcionais. À guisa de exemplo, vejamos os esquemas funcionais anotados às regras de uma minigramática, que contêm as metavariáveis já explicadas acima.

- 5) $S \rightarrow NP \quad VP$
 $(^{\wedge}S_{UJ}= !)$ $(^{\wedge}=!)$
 $NP \rightarrow (Det) \quad N$
 $VP \rightarrow \quad V \quad NP$
 $(^{\wedge}OBJ= !)$

A minigramática em (5) poderia ser utilizada para analisar frases declarativas simples de línguas como português ou inglês, e, das regras anotadas, apreende-se que uma sentença dessa gramática realiza-se por um NP seguido imediatamente de um VP, obrigatoriamente nesta ordem. Caso a ordem dos constituintes não fosse um fator determinante nessa língua, uma vírgula seria acrescida entre o NP e o VP a fim de marcar a livre ordem dos termos ($S \rightarrow NP , VP$). No que concerne ao NP, os parênteses demonstram que o Det nessa gramática é facultativo, podendo ou não ocorrer; mas, caso ocorra nesta língua, o determinante virá obrigatoriamente antes do núcleo do sintagma nominal, conclusão a que se pode chegar considerando-se novamente a vírgula inserida entre as categorias Det e N.

Utilizaremos a minigramática em (5) para analisar a frase “O cavaleira procura a fada.”. A gramática LFG de (5) gera para a sentença as duas estruturas descritas: a de constituintes e a funcional. A estrutura C é aquela apresentada na FIGURA 8. Nela estão contempladas as relações de precedência, dominância e ordem de sintagmas típicas de gramáticas livres de contexto. Também dela deriva-se a estrutura F correspondente a partir dos seguintes passos: associação dos esquemas funcionais anotados em nossa gramática (5) aos respectivos símbolos não-terminais, e, ao mesmo tempo, dos esquemas funcionais contidos no léxico aos símbolos pré-terminais.

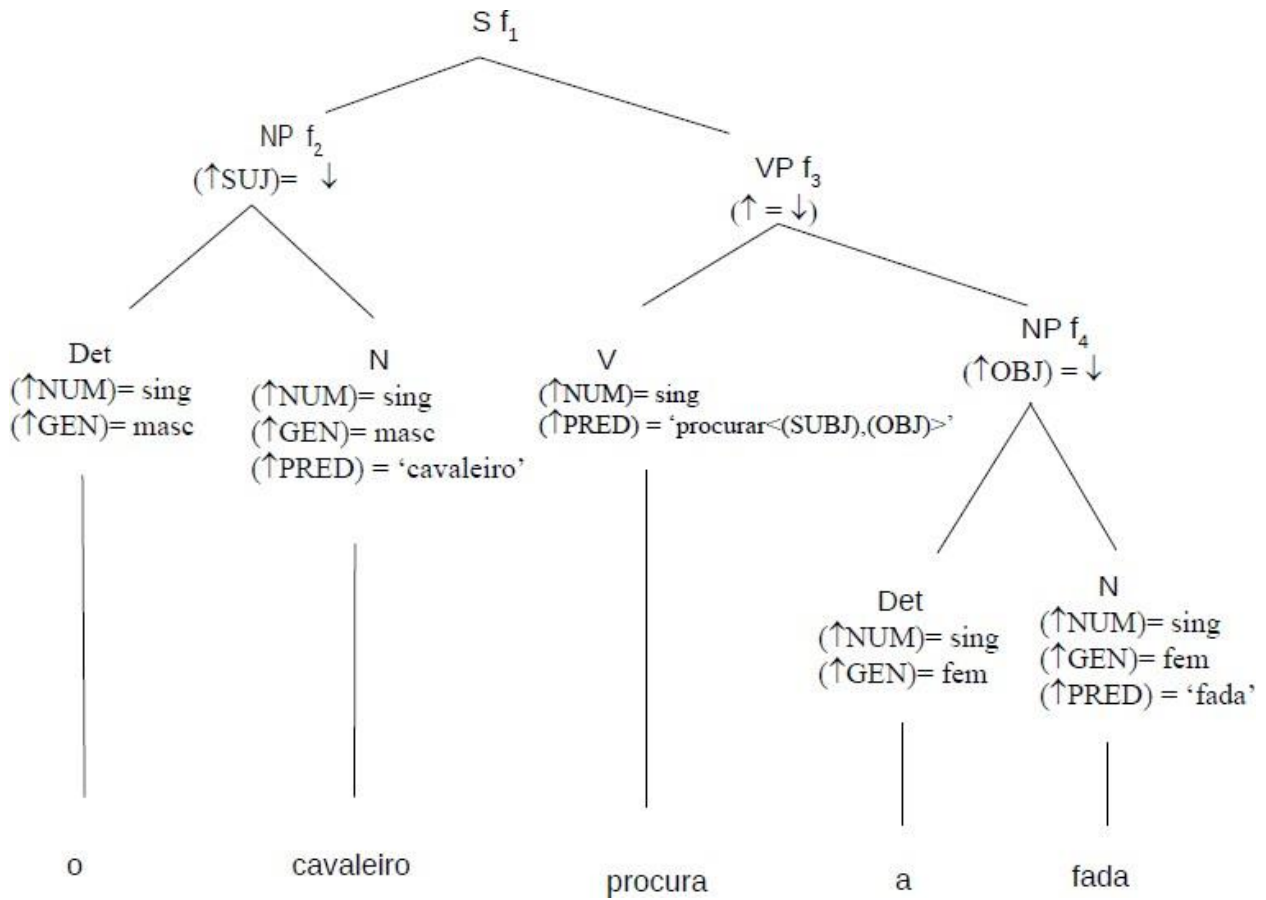
FIGURA 14 – Estrutura C anotada 1.



Fonte: Elaborada pelo autor.

A figura abaixo representa a associação de funções (f_1 , f_2 , f_3 etc) aos símbolos não- terminais, cada um desses recebendo um tipo diferente de função. A FIGURA 15 mostra um dos princípios da LFG, a saber, o de que há alguma estrutura F associada a cada nó da estrutura C (WESCOAT, 1989).

FIGURA 15 – Estrutura C anotada 2.



Fonte: Elaborada pelo autor.

Por fim, temos o chamado mecanismo de **instanciação**, quando os esquemas funcionais mostrados na figura acima são transformados em equações funcionais, “expressões específicas em uma linguagem formal usadas para falar sobre estruturas F” (WESCOAT, 1989, p.7). Nesse mecanismo, por exemplo, as metavariáveis \uparrow são substituídas pelas funções que dominam imediatamente os nós a que essas metavariáveis pertencem, ao passo que as metavariáveis \downarrow são substituídas pelas funções do próprio nó. Listamos abaixo as equações funcionais de nosso exemplo:

- 6) $(f_1 SUBJ) = f_2$
 $f_1 = f_3$
 $(f_2 NUM) = sing$
 $(f_2 GEN) = masc$
 $(f_2 PRED) = 'cavaleiro'$

$(f_3 \text{ NUM}) = \text{sing}$
 $(f_3 \text{ PRED}) = \text{'procurar}<(\uparrow \text{SUBJ})(\uparrow \text{OBJ})>$
 $(f_3 \text{ OBJ}) = f_4$
 $(f_4 \text{ NUM}) = \text{sing}$
 $(f_4 \text{ GEN}) = \text{fem}$
 $(f_4 \text{ PRED}) = \text{'fada'}$

As equações acima podem ser parafraseadas do seguinte modo: “o sujeito de f_1 é igual a f_2 ” ou “o número de f_2 é singular”, que dizer, o sujeito da sentença é a estrutura funcional associada ao sintagma nominal que está no singular. O símbolo ‘=’ expressa, do ponto de vista formal, não uma mera igualdade, mas sim um mecanismo de unificação: “a fusão de traços oriundos de diferentes fontes que se referem à mesma entidade linguística. Unificação é um formalismo que consiste em uma alternativa ao movimento como uma maneira de representar a remoção de elementos linguísticos”³⁷ (FALK, 2001, p.207).

Ao conjunto de todas as equações funcionais instanciadas dá-se o nome de descrição funcional, através da qual as estruturas funcionais são construídas (WESCOAT, 1989). Na verdade, depois do processo de instanciação, a árvore nem sequer é mais considerada no processo de construção de uma estrutura F. Toda a informação relevante encontra-se nas equações funcionais acima. É a partir das equações funcionais que se estabelecem também as condições de boa formação das sentenças a que nos referimos na seção teórica sobre a LFG. Na prática, é assim que a teoria léxico-funcional contempla questões como as de concordância de número, por exemplo, uma vez que uma dada equação funcional ‘ $(f_2 \text{ NUM}) = \text{sing}$ ’ não pode apresentar também a forma ‘ $(f_2 \text{ NUM}) = \text{plural}$ ’, pois, nesse caso, estaríamos atribuindo números diferentes para a função f_2 que está ligada ao sujeito da sentença.

Esperamos aqui ter conseguido mostrar os fundamentos do formalismo com o qual pretendemos trabalhar. Outro ponto a salientar-se diz respeito a como a LFG é capaz de lidar com o fenômeno da voz passiva, tópico da subseção seguinte.

³⁷ No original: “The merging of features from different sources which relate to the same linguistic entity. Unification is a formalism which is an alternative to movement as a way of accounting for the displacement of linguist elements” (FALK, 2001, p.207).

3.4 A análise da passiva na LFG

Nesta seção, discutiremos como o fenômeno da passiva, restringindo-nos à chamada passiva analítica, é analisado no âmbito teórico da LFG. Apresentaremos ainda a regra lexical de passivização e as estruturas C e F geradas pelo XLE para um exemplo de sentença passiva de nossa gramática a fim de discutir aspectos da análise.

Segundo Carnie (2002), o que temos na LFG é uma relação biunívoca entre as palavras e os nódulos terminais das estruturas sintáticas. Por ter uma base forte no léxico, funções gramaticais básicas da LFG nos permitem fazer construções passivas em um único passo, não havendo nesse formalismo um componente sintático para a voz passiva; ao invés disso, o que há é uma simples mudança lexical associada com morfologia passiva (CARNIE, 2002, p. 447).

Alencar (2015, p.41) faz a seguinte reflexão a respeito da voz passiva na LFG:

A característica fundamental da análise da passiva na LFG é que esse fenômeno sintático é tratado de modo puramente lexical, sem recorrer a nenhuma operação de movimento de constituintes, como o fazem outras abordagens gerativas. As diferenças estruturais entre a ativa e a passiva decorrem da vinculação de argumentos semânticos a funções sintáticas, que, na forma passiva, difere da vinculação argumental na ativa.

Vejam, então, como a chamada regra da passiva funciona em uma teoria de base lexicalista não-transformacional tal qual a LFG. Sigamos o raciocínio apresentado por Kroeger (2004, p.57), a partir dos exemplos abaixo, para descrever que tipo de regra é requerida pela gramática para derivar sentenças passivas:

- | | | | |
|-------|------------------|--------------------------|---|
| 8) a) | <i>beat</i> | < agent,
SUBJ | patient > Active
OBJ |
| b) | <i>be beaten</i> | < agent,
(OBL) | patient > Passive
SUBJ |

Para Kroeger (2004, p.57), como refletido no esquema (8a-b), “passivização é primariamente um realinhamento de relações gramaticais que sempre envolve o ato de reduzir o agente de uma forma de sujeito ativo à forma oblíqua passiva e que normalmente envolve o ato de promover o paciente da forma de objeto ativo à forma de sujeito passivo”³⁸. O processo de derivação da passiva parece, assim, bastante simples, o

³⁸ No original: “passivization is primarily a realignment of grammatical relations, always involving the demotion of the agent from active subject to passive oblique, an normally involving the promotion of the

que nos leva intuitivamente a pensar ser possível tratar a passivização “como um processo que deriva verbos passivos”³⁹ (*id.*, *ib.*) e as orações passivas como sentenças normais que contêm um tipo especial de verbo.

Esse tipo especial de verbo presente em orações passivas difere de seu correspondente ativo em dois pontos principais: i) por sua forma morfológica; ii) pelas relações gramaticais que são por ele atribuídas, ou seja, por sua subcategorização (KROEGER, 2004, p.58). Apesar do ponto (i), ou seja, da forma do verbo em si, e da presença de um verbo auxiliar (tanto em inglês quanto em português), a estrutura sintática de uma sentença passiva é bastante similar à sua correspondente ativa. No que concerne à estrutura sintagmática e ao padrão na ordem das palavras, verifica-se que não há grandes diferenças em pares de sentenças ativa e passiva. A gramática de uma língua como inglês e português, línguas configuracionais, deve, pois, incluir regras de estrutura sintagmáticas, como em (9), que permitam a geração de sentenças passivas.

$$\begin{array}{l}
 9) \quad S \rightarrow \quad NP \quad VP \\
 \qquad \qquad \qquad [SUBJ] \\
 \quad VP \rightarrow \quad V \quad (NP) \quad (PP)^{*40} \\
 \qquad \qquad \qquad \qquad \qquad [OBJ] \quad [OBL]
 \end{array}$$

(KROEGER, 2004, p.58)

Em termos de processamento, no que diz respeito à visão gerativa tradicional de ser voz passiva uma construção derivada de uma sentença ativa, pressupunha-se que uma sentença passiva seria mais dispendiosa⁴¹, uma vez que mais tempo seria requerido para o processamento das transformações. Não obstante, Müller (2016) afirma que se comparadas duas sentenças de tamanhos similares (e.g. sentenças

patient from active object to passive subject”.

³⁹ No original: “We will treat the passivization as a process that derives passive verbs”.

⁴⁰ Segundo Sag *et al.* (2003, p.563), o asterisco aqui representa a chamada estrela de Kleene (Kleene star), em homenagem ao matemático Stephen Kleene, geralmente usada na representação formal de línguas (tanto naturais como artificiais) a fim de permitir que certos padrões sejam repetidos por um número infinito de vezes (incluindo também zero repetições). A notação formal para isso é um asterisco sobrescrito. Por exemplo, uma expressão regular como ab^*c corresponde aos seguintes conjuntos de caracteres: $ac, abc, abbc, abbbc \dots$. No exemplo analisado em (9), significa que o VP pode ser formado por nenhum ou, em tese, por infinitos PPs.

⁴¹ Isso de acordo com o a Teoria Derivacional da Complexidade (DTC, *Derivational Theory of Complexity*), segundo a qual: uma sentença que requer mais transformações do que outra sentença dever ser, por isso, mais difícil de ser processada por humanos (MÜLLER, 2016, p.501).

10(a) e 10(b)), uma contendo o particípio passivo e outra um adjetivo, o custo maior de processamento esperado não se verifica.

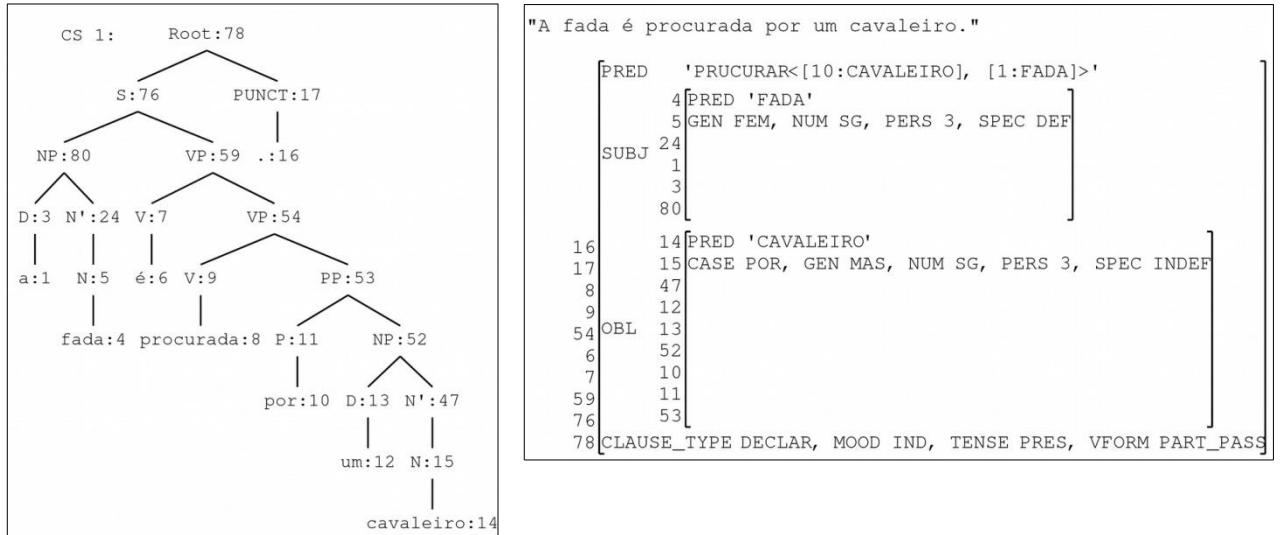
- 10) a) *The bus driver was nervous after the wreck.*
- b) *The bus driver was fired after the wreck.*

Também nesse ponto, a LFG representa uma modelo mais realista caso sejam levadas em conta pesquisas no âmbito da psicolinguística, haja vista que há uma coerência maior em uma teoria de base lexicalista, que não recorre a transformações, e teorias que se preocupam com a psicologia da linguagem.

Vejamos, então, como se configura uma análise LFG/XLE de uma sentença passiva como em (11):

- 11) A fada é procurada por um cavaleiro.

FIGURA 16 - Estrutura C e estrutura F da sentença *A fada é procurada por um cavaleiro* geradas no XLE a partir da gramática G-V



Fonte: Elaborado pelo autor através do XLE

O que se nota, comparando-se as estruturas F das FIGURAS 11 e 16, é que o valor de PRED para ambas é o mesmo. Segundo Alencar (2015a, p.40): “Isso reflete o fato de que, em termos lógico-semânticos, uma sentença na voz ativa é equivalente à sua versão passiva. O que difere, nas duas estruturas funcionais, é o mapeamento dos

argumentos semânticos sobre funções gramaticais”. Na prática, a diferença nas respectivas entradas lexicais das formas *procurar* e *procurada* dá-se apenas nas funções gramaticais que expressam os argumentos Agente e Tema.

Não obstante, uma ressalva pode ser feita, pois ainda que a grade temática seja a mesma, é interessante observar que a obrigatoriedade dos argumentos é assimétrica. O agente é opcional na sentença passiva, ao passo que seu correspondente na sentença ativa não o é. Essa opcionalidade parece refletir o princípio funcional da passiva de demissão do sujeito.

3.5 Da passiva sintética

Por se tratar de um fenômeno de complexidade considerável, restringiremos nosso objeto de estudos unicamente à chamada passiva analítica, ou seja, àquela tradicionalmente formada por um verbo auxiliar e por um particípio. Desse modo, casos como o do exemplo abaixo não serão tratados nesta pesquisa:

- 12) Alugam-se casas para turistas.

O exemplo (12) ilustra um caso da chamada passiva sintética que vem sendo, pelas teorias gerativas, sobretudo no âmbito da Teoria da Regência e Ligação - TRL (Government- Binding Theory – GB) (Chomsky, 1981), analisado de fato como voz passiva. As construções de passiva sintética conteriam um SE-passivo e teriam uma estrutura semelhante a construções com verbos inacusativos, em que o DP argumento interno recebe, segundo a Teoria de Caso, o Caso Nominativo, quando se desloca para a posição de sujeito (CAVALCANTE, 2016, p. 42).⁴² Sentenças com verbos inacusativos, como o exemplo (13a) abaixo, aceitam inclusive pósposição do sujeito, como em (13b), o que indica ser o DP um complemento do verbo:

- 13) a. A garrafa de água congelou.
b. Congelou a garrafa de água.

⁴² Sobre mais trabalhos no âmbito do gerativismo acerca da voz passiva sintética, ver, por exemplo, trabalhos de Cyrino (2007) e Mateus *et al.* (1994).

Interessante é trazer à tona o que assevera Cavalcante (*op. cit.*) quando menciona os trabalhos de Chomsky acerca de passivas sintéticas como vistas em (11). Para Chomsky, dentro do Programa Minimalista (1995), construções como estas não devem ser consideradas como voz passiva, haja vista que o DP argumento interno não é o sujeito e, tampouco, ocupa uma posição de sujeito. O DP *casas para turistas* parece funcionar muito mais como argumento interno do verbo.

Mateus *et al.* (1994), entretanto, ressaltam que tais construções com o pronome *se* compartilhariam uma propriedade definidora da passiva, a saber: “suspensão da atribuição de função semântica de AG(ente) à posição de sujeito e externalização do argumento interno do predicador” (*id.*, p. 225). Assim como na passiva perifrástica, o tema do verbo transitivo é realizado como sujeito nas construções com *se*, contudo, o que diferiria nestas das construções passivas com formação perifrástica seria o fato de o traço da passiva não decorrer de um processo de derivação morfológico, mas sim da presença do clítico *se*.

Segundo os autores, o pronome clítico absorveria o Caso acusativo e, conseqüentemente, o verbo perderia a capacidade de atribuição de Caso a seu argumento interno. Simultaneamente, pela regra *Mover α* , o clítico também suspenderia a atribuição de papel semântico de agente (AG) à posição de sujeito, que ficaria livre para ser ocupada pelas regras de deslocamento. A sentença (9) pode, assim, apresentar-se como a variante:

- 14) Casas para turistas alugam-se.⁴³

Ainda que não seja nosso objeto de estudo, achamos oportuno mencionar os problemas de análise apresentados pela chamada voz passiva sintética, havendo inclusive, dentro do próprio gerativismo, autores que negam sua real existência. Não se constitui, pois, um fato isolado a análise que propomos aqui para a passiva analítica. Por fim, a análise das passivas de *se* não será por nós implementada, lacuna com a qual pretendemos, possivelmente, trabalhar em pesquisas futuras. Entretanto, citamos aqui o trabalho de Kelling (2006), no qual a autora tece uma proposta de análise das

⁴³ Construção típica do português europeu

construções com *se* (tanto passivas, como impessoais)⁴⁴ para o espanhol usando a Teoria do Mapeamento Lexical (LMT, *Lexical Mapping Theory*), the Lexical Functional Grammar linking modul (Bresnan, 2001), a qual representa um dos desdobramentos mais modernos no âmbito teórico da LFG, mas sobre a qual não nos debruçaremos por questões de espaço e complexidade.

3.6 Evidências morfológicas e sintáticas para a classificação do PART-PASS

Quando se trabalha com implementação de gramáticas é perfeitamente legítimo construir sentenças, pois dados cotejados em *corpora* reais apresentam complexidades que vão além do fenômeno que se almeja investigar, as quais não contribuem para a verificação da existência desse fenômeno. Optou-se, assim, por trabalhar-se com um *corpus* criado, artificial, mas cuja confecção está baseada em fenômenos encontrados em *corpora* reais. No presente tópico, por exemplo, apresentamos evidências colhidas em fontes reais de que o PART-PASS em PB apresenta características típicas da classe dos adjetivos.

Na linguística gerativa, o estatuto categorial de um constituinte decorre de suas propriedades morfológicas e de sua distribuição sintática (MÜLLER, 2010). Sendo assim, procuramos, nesta subseção, mostrar como evidências morfológicas e sintáticas podem ser cotejadas de modo a reunir mais evidências em favor de nossas hipóteses. Na prática, o que buscou-se fazer foi encontrar evidências na literatura e ocorrências reais dos participios passivos em contextos morfológicos e sintáticos ocupados preferencialmente por adjetivos.

Ainda que as ideias de Perini (2010) e Alencar (2015) tenham sido as mais influenciadoras desta pesquisa, considerar a passiva uma construção adjetival não é uma novidade. O gramático Jerônimo Soares Barbosa (1875) já considerava o participio passado um adjetivo verbal com valor passivo em sentenças do tipo “eu sou amado”, semelhantes à construções com adjetivo verbal ativo: *eu sou amante* ou *eu tenho sido*

⁴⁴ Exemplos de construções de *se* passivas e impessoais tratadas por Kelling (2006) são, respectivamente, ilustradas pelos exemplos (a) e (b) seguintes:

- a) *Se firmó la paz.*
REFL sign.PAST the peace
'Assinou-se a paz'
- b) *Se impitó a todos los empleados*
REFL invite.PAST to all the employees
'Convidaram-se todos os empregados'

amado.

Mesmo não se tratando de uma análise absolutamente nova, uma abordagem interessante nos é apresentado por Alencar (2015a), quando o pesquisador faz um levantamento de um fato pouco explorado na literatura. Em seu artigo, apresentam-se evidências de morfologia derivacional que corroboram a ideia do participípio passivo pertencer à classe dos adjetivos. Os participípios das formas passivas admitem que lhes sejam acrescidos os chamados sufixos avaliativos, por exemplo, o sufixo *-inho*, típico de diminutivo. Mais exemplos cotejados por nós podem ser observados nas sentenças que seguem:

- a) “Refogue o bem o alho, jogue os ingredientes que foram *cortadinhos* juntamente com os cajús (*sic.*), refogue bem, depois jogue a água até os ingredientes ficarem cobertos e deixe cozinhar até ficarem macios (...)” (Google)
- b) “Tudo relacionado acima foi bem *explicadinho*...” (Google).
- c) “O Amigo Livro foi um tipo de Amigo Oculto e o presente deveria ser um livro sugerido pela pessoa (cada pessoa sugeria dois livros que gostaria muito de ganhar no valor de até 20,00 R\$). Tudo foi bem *organizadinho* numa lista com os nomes dos livros, nomes dos participantes e seus endereços.” (Google).

Em *Gramática Derivacional do Português*, Rodrigues *et al.* (2016, p. 360) afirmam:

O sufixo *-ísim(o, a)* é o sufixo do grau superlativo. Combina-se essencialmente com bases adjetivas, como *aflitíssimo*, *atormetadíssimo*, *caidíssimo*, *enervadíssimo*, *enormíssimo*, *feíssimo*, *lentíssimo*, *lindíssimo*, *mesmíssimo*, *ocupadíssimo*, *pequeníssimo*, *rapidíssimo*, *raríssimo*, *singularíssimo*, e também com algumas adverbiais, como *longíssimo*, *períssimo*, *pouquíssimo*, *tardíssimo*. Raramente se combina com bases nominais (cf. *coisíssima* em *coisíssima nenhuma*) e pronominais (*nadíssima*), e não se combina com bases verbais.

O PART-PASS parece aceitar, como podemos depreender dos exemplos abaixo, o acréscimo deste tipo de sufixo ao seu radical, do que se pode concluir que, aparentemente, trata-se de uma classe pertencente à mesma dos adjetivos citados acima. Eis os exemplos:

- d) “O Moda na Mala esteve presente, e inclusive abriu o desfile, que foi *aplaudidíssimo* pelo público presente.” (Google)
- e) “Em 1968 Caetano foi *vaiadíssimo* pela esquerda estudantil, na apresentação de 'É Proibido Proibir'.” (Google)

f) “Cheio de revelações e mais mistérios, o episódio foi **comentadíssimo** nas redes sociais e nós resolvemos selecionar algumas das opiniões do público.” (Google)

g) “[Eslovênia] Taí um país que não dá para entender por que não é **exploradíssimo** pelos brasileiros.” (Google).

Por fim, queremos ainda mencionar uma terceira evidência semelhante àquela apresenta por Perini (2010.). Trata-se do prefixo de negação acrescido a alguns participios. Veja-se o exemplo trazido pelo autor:

h) Eles foram incompreendidos.

Como não existe, em PB, o verbo *incompreender*, chega-se à conclusão de que a forma *incompreendido* não apresenta caráter verbal, sendo formada acrescentando-se o prefixo ‘in-’ ao participio passivo de *compreender*. O participio presente nas construções passivas teria natureza distinta ao presente em construções do passado, pois este último não admite o acréscimo do prefixo:

i) *Eu tinha incompreendido o exercício.

Comparando-se o raciocínio acima com outras línguas, tomemos o trabalho de Noma (2015), no qual a autora, ao tratar do fenômeno da voz passiva em língua alemã, argumenta que, em construções com verbo cópula (a chamada *Zustandspassiv* [passiva de estado]), há um participio de caráter diferente do participio verbal. Apresentando os exemplos (i) e (j), mostrados abaixo, afirma que o primeiro tipo de participio teria caráter adjetival, pois “o prefixo negativo *un-* não pode ser combinado com verbos, mas sim apenas com adjetivos” (NOMA, 2015, p.50)⁴⁵:

j) *Der Unfall ist noch ungeklärt.*

O caso ainda não está esclarecido.

k) *Das Paket ist zollrechtlich unabgefertigt.*

O pacote não foi expedido por razões de legislação aduaneira

⁴⁵ No original: “(...) dass das Negationspräfix *un-* nicht mit Verben, sondern nur mit Adjektiven kombiniert werden kann.“

Percebe-se no exemplo citado para o alemão, *mutatis mutandis*, um processo morfológico similar em PB em exemplos como (l) e (m), pois, como menciona Noma (*id.*), não há em alemão os verbos *unklären* e *unabfertigen*; em PB, tampouco, existem correspondentes verbais para os particípios negados (não há em PB os verbos *inacabar* e *irretocar*)⁴⁶:

- l) A obra está/ fica/permanece INACABADA.
- m) A foto permanece IRRETOCADA.
- o) *Os pedreiros inacabaram a obra.
- p) *O artista irretocou o quadro.

Situação distinta à que ocorre quando o verbo em português já possui o prefixo negativo *in-*, como em invalidar ou inabilitar:

- q) A comissão invalidou o diploma.
- r) O diploma foi invalidado pela comissão.

⁴⁶ MORAIS (1988, p.131) argumenta que embora em português não possamos afirmar que os prefixos de negação ‘i-/im-/ím’ não ocorram com verbos, tal ocorrência estaria restrita a poucos casos, cuja raridade e improdutividade não anulariam a força argumentativa a favor de que esses prefixos apresentam grande produtividade com adjetivos, o que nos leva a conclusão de serem os particípios, nos casos acima, categorizados como adjetivos.

4 METODOLOGIA

A fim de cumprimos o desiderato deste trabalho que, como já mencionado, situa-se numa interface entre a linguística gerativa e a linguística computacional, caracterizando-se, pois, como uma pesquisa de caráter interdisciplinar, lançaremos mão de metodologias de ambas as áreas com o intuito de se verificarem as hipóteses aqui levantadas.

Nossa pesquisa possui um caráter teórico-prático, uma vez que parte de abordagens teóricas acerca da classificação do particípio em construções tradicionalmente analisadas como voz passiva e, além disso, busca, através da implementação computacional em LFG/XLE, evidências que sustentem a hipótese apresentada na seção anterior. Desse modo, para que se alcance resultado satisfatório é preciso que um conjunto de procedimentos metodológicos seja seguido, executando-se cada um num dado momento específico.

4.1 Do *corpus* da pesquisa

Para o desenvolvimento de nossas gramáticas LFG-XLE, utilizamos a metodologia típica de trabalhos em LC, nos quais a construção de gramáticas formais se dá de maneira incremental, quer dizer, parte-se, inicialmente, de um recorte que contemple um leque de fenômenos, ao qual serão acrescidas melhorias e adaptações a fim de que um número cada vez maior de fenômenos possa ser analisado (ALENCAR, 2013a, p.184). É o chamado desenvolvimento em espiral ao qual voltaremos no capítulo que foca na construção de nossa gramática.

A primeira etapa desta pesquisa consistiu, então, na criação de um *corpus*, no qual esteja presente o maior número possível de variações do fenômeno objeto de estudo deste trabalho. Não é a quantidade de sentenças o mais importante para nossa pesquisa, mas sim que diversos contextos linguísticos em torno de nosso objeto sejam contemplados com pelo menos uma sentença prototípica. Citamos alguns exemplos, apenas à guisa de ilustração:

- (1) 1) A fada é esperada por um cavaleiro.
2) Ele é esperado por uma dama.
3) A fada é esperada.
4) O cavaleiro é esperado.
5) O anel é dado a uma dama por uma fada.

- 6) Os anéis são dados a uma dama.
- 7) O cavaleiro vê a floresta habitada por umas fadas.
- 8) O cavaleiro vê o castelo branco habitado por uma rainha.
- 9) O cavaleiro tem esperado.
- 10) A fada tem chegado.
- 11) As fadas têm chegado.⁴⁷

O conjunto de sentenças dado demonstra que, para a correta implementação das regras de nossa gramática, devem ser considerados exemplos em que o PART-PASS apresente concordância de gênero e número, como em (1) e (2); presença ou não do PP adjunto (sintagma preposicional que na análise tradicional funciona como agente da passiva), presente nos exemplos (5), (7) e (8). O exemplo (11), por seu turno, exemplifica, para o PB, o pretérito perfeito composto, cuja formação se faz com verbo *ter*; com este exemplo pretendemos cobrir os casos de estruturas linguísticas semelhantes às da passiva, mas que apresentam tanto um auxiliar distinto quanto um particípio sem traços de concordância de gênero e número.

Para a criação de nosso *corpus* inicial que servirá de base para a geração de regras em nossa gramática, bem como a alimentação de nosso léxico, recorreremos ao conjunto-teste em francês apresentado por Schwarze e Alencar (2016). Aqui trata-se não apenas de uma tradução entre o par de línguas português e francês, mas muito mais da seleção de que fenômenos encontram correspondência entre os dois sistemas linguísticos. No capítulo 6 de Schwarze e Alencar (2016), implementam-se também construções com o *passé composé*, tempo composto do passado em francês formado por verbo auxiliar *être* ou *avoir* e por um particípio passado do verbo principal.

Como a formação do passado em português é mais simples do que a do francês retiramos do léxico e das regras aquilo que não encontra aplicação no português. Por exemplo, os critérios de escolha do auxiliar a depender de ser o verbo principal inacusativo ou inergativo: o primeiro grupo de verbos seleciona em línguas como o francês, italiano e alemão o auxiliar *être*, *essere* e *sein*; o último, o auxiliar *avoir*, *avere* e *haben*. Ambos são verbos monoargumentais que se distinguem, entretanto, pelo fato de os inergativos selecionarem como argumento um NP/DP com papel temático de agente, ao passo que os inacusativos selecionam um argumento com papel temático de tema ou paciente.

Ainda que haja também em português a divisão dos verbos

⁴⁷ No final deste trabalho, encontram-se nosso arquivo-teste positivo <testfile.lfg> e nosso arquivo-teste negativo <negative-test.lfg> (respectivamente nos Apêndices A e B).

monoargumentais nos dois grupos supracitados, e que tais verbos apresentam características semânticas e sintáticas distintas, no que concerne à formação do passado em português, seja ele simples ou composto, não é necessário uma etiqueta acrescida aos lemas verbais que contemple essa distinção uma vez que em PB os dois grupos de verbos monoargumentais selecionam o mesmo auxiliar *ter*. Para o francês, Schwarze e Alencar (2016) utilizam etiquetas como “+UNACC”/“-UNACC” ou “+UNERG/-UNERG”.

Mesmo que não apresente correspondência com o francês na formação do passado composto, decidimos, entretanto, manter a classificação dos verbos monoargumentais em inacusativos e inergativos dentro das entradas lexicais contidas nos lemas, dado que, em trabalhos futuros, em uma versão melhorada de nossa gramática, o tratamento de outros aspectos e fenômenos do PB pode ser facilitado por essa classificação. A título de exemplo, citem-se fenômenos como i) construções com participio absoluto (as tradicionalmente chamadas orações subordinadas temporais reduzidas de participio), e ii) passivas estativas (construções classificadas como passivas, mas cujo verbo auxiliar é o *estar*), com as quais podemos trabalhar em versões posteriores de nossa gramática. Ambos os fenômenos supracitados ocorrem (cf. KATO, 2015) com AI de verbos, mas não podem ocorrer com AE; conseqüentemente, os dois fenômenos incidirão exclusivamente sobre verbos inacusativos, por estes selecionarem AI, e não sobre verbos inergativos, os quais selecionam unicamente AE. Vejam-se exemplos adaptados de Kato (*op. cit.*, p.49) que comprovam essa distinção:

- (2)
 - a. A Maria sumiu/desapareceu.
 - b. Sumida/desaparecida a Maria, a polícia foi chamada.
 - c. A Maria estava sumida/desaparecida.

- (3)
 - a. A Maria dormiu/espirrou/.
 - b. *Dormida/espirrada a Maria, todos ficaram preocupados.
 - c. *A Maria está dormida/espirrada.

Em (2a-c), temos exemplos com verbos inacusativos, em (3a-c), com verbos inergativos. Os primeiros aceitam construções com participio absoluto e passivas estativas, os segundos não são compatíveis com as mesmas.

Feita essa primeira etapa, expandimos nossa gramática a fim de que sentenças como as seguintes, listadas em (4), também pudessem ser analisadas. Nessa segunda etapa, contemplaram-se o capítulo 8 final do livro de Schwarze e Alencar

(2016) e a gramática apresentada no artigo de Alencar (2017), esta última constitui-se de versão ampliada e já melhorada do livro. Trata-se das sentenças (12-16) abaixo:

- | | | |
|-----|-----|--|
| (4) | 12) | O cavaleiro foi esperado por uma fada. |
| | 13) | O cavaleiro tem sido esperado por uma fada. |
| | 14) | O cavaleiro tem sido forçado por uma rainha a acabar a tarefa. |
| | 15) | O cavaleiro tem sido forçado a acabar a tarefa por uma rainha. |
| | 16) | A dama quer ser convidada para dançar por um cavaleiro. |

4.1.1 Da construção do conjunto-teste e das gramáticas computacionais em LFG

A construção de um *corpus* foi a primeira etapa referente à análise computacional desta pesquisa. Nosso *corpus*, a partir do qual implementaremos nossa gramática no XLE, será construído, consoante a orientação de Schwarze e Alencar (2016, p.24), de um conjunto-teste composto de duas partes:

- Um grupo de exemplos que contenham apenas sentenças bem formadas gramaticalmente em PB⁴⁸: nosso **conjunto-teste positivo**. Tais sentenças devem ser analisadas pela gramática, a qual lhes deve atribuir ainda estruturas sintagmáticas e funcionais adequadas⁴⁹;
- Um outro grupo, por seu turno, formado exclusivamente por sequências agramaticais em PB e que não devem ser geradas por nossa gramática, ou seja, nosso **conjunto-teste negativo**.

A criação do *corpus* (conjuntos-teste positivo e negativo) que servirá de base para a construção de nossa gramática deu-se, inicialmente, como já mencionado, por meio da adaptação. A fim de verificar se as sentenças em PB estão bem formadas e aptas para a análise recorreremos à intuição do falante (em nosso caso, do pesquisador). Dentre os fatos que nos parecem relevantes para justificar tal escolha metodológica, mencionamos, em primeiro lugar, a praticidade e a economia de tempo empregado na abordagem de um dado fenômeno linguístico sem a necessidade de se analisar grandes e numerosos *corpora*.

A respeito da introspecção como método de análise de um fenômeno linguístico, encontramos em Perini (2006, p.39s) argumentação que parece colaborar para que consideremos plausível este recurso nos procedimentos metodológicos desta

⁴⁸ O critério para decisão sobre o *status* gramatical ou agramatical de uma sentença em PB será embasado na intuição do falante (em nosso caso, na intuição do pesquisador responsável pela pesquisa, falante nativo de PB).

⁴⁹ As estruturas constitucional e funcional serão tratadas na seção de Fundamentação Teórica.

pesquisa:

É preciso encarar uma realidade: estudar uma língua é, em última análise, estudar um fenômeno psicológico, portanto não observável diretamente. Só temos acesso a essas manifestações exteriores e aos julgamentos sobre elas, e daí temos de deduzir as características do fenômeno estudado: a língua, considerada como sistema de conhecimento armazenado na memória. Para isso é preciso lançar mão de todos os recursos disponíveis, sem esquecer a devida cautela, sabendo que lidamos com dados bastante inseguros. O *corpus*, a testagem e introspecção são recursos a ser utilizados com consciência de suas vantagens e de suas fraquezas. (PERINI, 2006, p.39s.)

Dada a ressalva feita pelo autor de que a coleta de dados em linguística não é uma tarefa trivial, e a fim de salvaguardar o caráter científico desta pesquisa, lançamos mão da introspecção do pesquisador, evitando, entretanto, para testagem da gramática casos que suscitem dúvidas.

Outro ponto importante que permeou a criação de nossa gramática foi a criação do já mencionado teste negativo, cuja existência tem por intuito treinar a gramática para identificar sentenças analisadas como agramaticais por falantes nativos. Na gramática gerativa, o teste negativo apresenta um papel importante, pois essa visa à construção de um modelo de dispositivo, formulado matematicamente, capaz de gerar todas e unicamente sentenças gramaticais de uma língua.

A importância desses dois tipos de teste é ressaltada por Butt *et al.* (1999), em seu livro *A grammar writer's cookbook*. A autora considera que para se testar uma gramática de forma completa é necessária uma variedade de suítes de testes. Dessa forma, é melhor ter suítes de testes para cada uma das construções mais importantes que a gramática aborda (aqui ilustrado pelas sentenças acima). A autora também chama a atenção que variantes agramaticais (teste negativo) devem ser igualmente incluídas de forma a assegurar que a gramática não esteja hipergerando (*overgenerating*).

4.1.2 Da criação do teste negativo

Queremos nesta subseção salientar alguns pontos acerca da criação de nosso conjunto- teste negativo. Como mencionado, as sentenças contidas neste arquivo-teste devem ser agramaticais em PB e, por conseguinte, nossa gramática LFG/XLE, se suficientemente bem construída, não lhes deve atribuir nenhuma estrutura F.

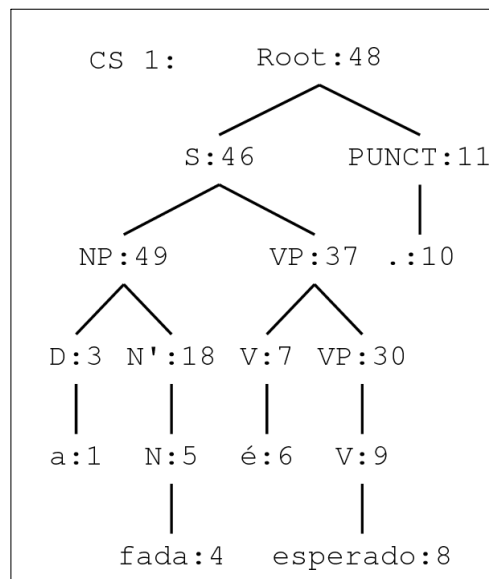
Para ilustrar, vejamos os resultados de análise da primeira sentença de nosso conjunto de teste negativo, a saber: *A fada é esperado*.

Esse exemplo, por não estar de acordo com o princípio de unificação, não pode ser analisado, haja vista que ‘fada’ tem propriedade FEMININO SINGULAR e ‘esperado’, por sua vez, MASCULINO SINGULAR. O resultado da análise em XLE, feito a partir do comando em (5), é mostrado na Figuras 17 e 18 abaixo:

(5) **% parse "A fada é esperado."**

0 solutions, 0.006 CPU seconds, 0.082MB max mem, 24 subtrees unified

FIGURA 17 - Estrutura C e estrutura F do exemplo teste negativo *A fada é esperado



Fonte: Elaborado pelo autor através do XLE.

Figura 18 – Estrutura F do exemplo teste negativo A fada é esperado

Lock	F-structure #1 (EVENTUALLY BAD)
8	PRED 'ESPERAR<NULL, [6-SUBJ]>'
9	SUBJ [GEN MAS, NUM SG, PERS 3]
30	MOOD IND, TENSE PRES, VFORM PART_PASS
6	
7	
37	

Fonte: Elaborado pelo autor através do XLE.

A análise mostra-nos que nenhuma estrutura funcional é atribuída ao exemplo (a), veja-se o resultado de “0 solutions” indicado em (5), ainda que lhe seja conferida uma estrutura de constituintes, pois a sentença é, do ponto de vista unicamente estrutural, bem formada para o PB (como se pode observar na Figura 10). Da Figura 18, percebe-se que o sujeito SUBJ de ‘ESPERAR’ tem de preencher para o

atributo GEN o valor MAS e para o atributo NUM o valor SG. Desse modo, o sistema considera a frase (a) incoerente, pois ‘fada’ tem como propriedade FEMSG.

Com base no exemplo mostrado, percebe-se que a estrutura C não é o bastante para definir categoricamente se uma sentença de uma dada língua é bem formada. Há, desse modo, a necessidade de que sejam preenchidas as condições de boa formação da estrutura F, a saber: completude, coerência e unicidade (KROEGER, 2004, p.20). Uma sentença gramaticalmente bem- formada (*well-formed*) é aquela em que todas as combinações de nós na estrutura C estão licenciadas por regras sintagmáticas bem definidas, e em que as condições de boa formação na estrutura F sejam respeitadas (*id. ib.*).

Inicialmente, pensou-se na possibilidade de se utilizar algum algoritmo que gerasse randomicamente, a partir das combinações das entradas lexicais contidas em nossa gramática, todas as possibilidades de sentenças negativas. Essa estratégia, contudo, foi rejeitada devido ao fato de almejarmos sentenças em que apenas um tipo de violação aos princípios de boa formação de sentenças estivesse presente por vez para garantir a testagem eficiente de nossa gramática. Desse modo, sentenças como *A fada são esperado* não são apropriadas aos nossos propósitos, pois tanto a concordância de número entre o VP *são* e o NP *A fada*, quanto a concordância de gênero entre *esperado* e *A fada* são violadas. Já sentenças como *A fada são esperado* prestam-se a nossos propósitos.

A estratégia utilizada, então, foi a de construirmos sentenças, partindo-se de nosso teste positivo, em que apenas um e somente um princípio de boa formação fosse violado por vez. Apresentamos abaixo nosso conjunto de teste negativo, composto, até o presente momento por 20 sentenças:

(6)

Conjunto-teste negativo: *negative-test.lfg*

SENTENCE_ID: 001

A fada é esperado.

SENTENCE_ID: 002

A fada é esperado por um cavaleiro.

SENTENCE_ID: 003

O cavaleiro é esperada por uma dama.

SENTENCE_ID: 004

A fada é espera.

SENTENCE_ID: 005

A fada quer esperada.

SENTENCE_ID: 006

A fada começa a esperada.

SENTENCE_ID: 007

A fada é esperar.

SENTENCE_ID: 008

A fada esperada.

SENTENCE_ID: 009

A fada esperado.

SENTENCE_ID: 010

A fada tem esperada.

SENTENCE_ID: 011

A fada tem espera.

SENTENCE_ID: 012

A fada tem esperar.

SENTENCE_ID: 013

A fada tem chegada.

SENTENCE_ID: 014

A fada é chegada.

SENTENCE_ID: 015

As fadas são chegadas.⁵⁰

SENTENCE_ID: 016

O cavaleiro é esperada.

SENTENCE_ID: 017

O anel é dada a uma dama por uma fada.

SENTENCE_ID: 018

Os anéis são dada a uma dama.

SENTENCE_ID: 019

⁵⁰ A sentença (15) pode ser considerada gramatical no sentido de que “as fadas são próximas, íntimas”. Mesmo assim, para o recorte com o qual almejamos trabalhar até o presente momento, resolvemos classificar como agramaticais sentenças como (15) para garantir que formações similares ao *passé composé* de verbos inacusativos em francês, analisadas em Schwarze e Alencar (2016), não fossem analisadas por nossa gramática: *Les fées sont arrivées*.

O cavaleiro vê a floresta habitado por umas fadas.

SENTENCE_ID: 020

O cavaleiro vê a floresta habitadas por umas fadas.

4.1.3 Do material utilizado

Os testes com nossos *parsers* foram realizados em uma máquina com sistema operacional Linux Ubuntu 16.04, 64-bit, processador Intel(R) Core(TM) i3-310M CPU @ 2.40 GHz, com memória de quatro *gigabytes* (4GB) e HD de um *terabytes* (1TB).

Como exposto, nossa gramática LFG foi transformada em *parser* no sistema XLE. A utilização do sistema XLE para implementação de nossa gramática não é aleatória. O sistema, além de oferecer uma interface bastante amigável na linguagem de *scripts* Tcl, apresenta uma série de vantagens que justificam seu uso, como ressaltado por Alencar (2013a, p. 36):

A facilidade de uso do XLE se torna ainda maior, por um lado, pela possibilidade de automatizar a compilação e testagem de *parsers* por meio de *scripts* em Tcl; por outro, pela possibilidade de utilizar a interface gráfica do editor de texto emacs para executar os principais comandos desse sistema, o que é especialmente interessante para linguistas sem conhecimentos de programação.

4.1.4 Do procedimento de análise

Para a verificação de nossa hipótese levantada, referente à gramática que resulta no *parsing* mais eficiente, dois fatores serão levados em conta: (a) o tempo de processamento (*parsing*) das sentenças em segundos de CPU e (b) a quantidade de *subtrees* geradas.

O tempo de processamento de cada análise é fornecido, em milissegundos, pelo XLE. O número de *subtrees* geradas fornece ao implementador da gramática uma indicação da complexidade do sistema de regras. Sentenças simples ou constituintes que apresentem um número muito alto de *subtrees* parecem fornecer um indicativo de que há algo contraditório com a implementação das regras (cf. BUTT *et al.*, 1999, p.159).

Submetemos nosso *corpus* à análise por ambas as gramáticas (G-A e G-V), repetindo a implementação, para cada uma delas, por 10 vezes a fim de verificarmos as diferenças possíveis no tempo de processamento. A quantidade de 10 repetições foi definida por ser esta a quantidade de repetições nos teste realizados no projeto PIBIC

que nos serve de parâmetro inicial desta pesquisa. A repetição é comum quando se trabalha com processamento em computação, porque o desempenho da CPU é uma variável a ser considerada. Na prática, o que se observa é sempre uma variação, ainda que infinitesimal, mesmo assim considerável, no tempo gasto para a realização de uma tarefa; mesmo mantidas as condições de processamento na utilização de um programa em um computador, a memória liberada pela CPU para o funcionamento de um dado programa varia constantemente, não sendo possível fixá-la. Recorremos, desse modo, à análise probabilística dos dados obtidos para aumentar a acurácia e o rigor na análise.

Quanto menor for o tempo de *parsing* das sentenças e a quantidade de *subtrees* geradas, mais eficiente será a gramática. Segundo BUTT *et al.* (*op. cit.*), o número de *subtrees* geradas dá uma indicação da complexidade das regras da gramática, e, em formalismo computacional, *grosso modo*, menos regras significam maior robustez.

4.2 Da criação da gramática

Uma gramática LFG é um tipo de gramática que se classifica como sendo de processamento profundo, na medida em que gera, além de informações sobre a estrutura sintagmática e dependencial de uma sentença, uma representação lógica do significado das sentenças em uma língua natural. Nas palavras de Forst (2011, p.1), dizer ser nossa gramática LFG um analisador sintático profundo equivale a dizer que ela está baseada em uma teoria linguística, em nosso caso a LFG, a qual é mais detalhada e normatizada do que as chamadas técnicas rasas (*shallow techniques*) de processamento de linguagem (*chunking*, etiquetagem morfossintática, *parser* livre de contexto).

Nossa gramática LFG é ainda monotônica, o que equivale a dizer que as entradas lexicais e regras têm permissão de unicamente adicionar ou combinar informações, mas não as deletar (FORST, 2011). Essa característica é resultado da natureza declarativa e baseada em restrições da LFG.

Nossa gramática, em seu atual estado, é resultado da adaptação para o português de uma gramática LFG-XLE a partir do modelo de implementação para o francês proposto nos oito capítulos de Schwarze e Alencar (2016) e da gramática FrGram, mais recente e de cobertura mais ampla, de Alencar (2017). Nossa tarefa inicial foi, então, a de adaptar para o português essas gramáticas, e nosso objetivo final foi ampliar a cobertura de fenômenos gramaticais do português, sobretudo no tocante às duas análises contrárias do participio nas sentenças passivas, indo além do leque de

fenômenos análogos abrangidos pelos autores para o francês.

Um exemplo de adaptação com o qual lidamos foi, por exemplo, a decisão de não simplesmente traduzir literalmente as sentenças em francês do conjunto-teste positivo contidas em Schwarze e Alencar (*op. cit.*). Observe-se a sentença (a) abaixo, cuja tradução literal para o português em (b) apresenta uma outra estrutura, a saber, a contração *no* (*em + o*):

- (7) (a) *la dame est dans la chambre.*
(b) a senhora está no quarto.

A fim de manter a mesma estrutura e simplificar as regras sintagmáticas, mantendo-as, na medida do possível, similares às da gramática LFG-XLE do francês, optamos pela adaptação em (c):

- (8) (c) a dama está sob a ponte.

Considerações mais detalhadas acerca da adaptação das regras gramaticais e das entradas lexicais da gramática do francês para a do português serão mostradas no Capítulo 5 que trata da implementação do fenômeno da passiva em uma gramática LFG-XLE. No mesmo capítulo, também apresentamos alguns pontos fundamentais na criação da segunda versão de nossa gramática a fim de podermos produzir as duas amostras a serem comparadas.

5 A IMPLEMENTAÇÃO LINGUÍSTICO-COMPUTACIONAL DA PASSIVA EM UMA GRAMÁTICA LFG/XLE

Nesta seção, almejamos apresentar nossas duas gramáticas com foco no tratamento das construções passivas em PB. Nossa metodologia foi, inicialmente, a de implementarmos a gramática que correspondesse à análise mais tradicional de classificar o PART_PASS como V. A partir dessa gramática, realizamos modificações para que fosse possível a classificação de PART_PASS como A, a fim de testarmos nossa hipótese básica.

Durante o curso de mestrado, desenvolvemos na disciplina *Sintaxe-Semântica do Português* do Programa de Pós-Graduação em Linguística da Universidade Federal Ceará uma gramática LFG-XLE para o português⁵¹. A construção da gramática deu-se de maneira incremental seguindo o modelo de implementação para o francês proposto por Schwarze e Alencar (2016), e *pari passu*, a partir das discussões feitas em sala, adaptamos os fenômenos presentes em língua portuguesa. Inicialmente, o escopo foi o de cobrir o fenômeno da passiva tal qual apresentado para o francês no capítulo sexto da referida obra.

Em seguida, continuamos a implementação e melhorias da gramática, e, seguindo o modelo dos capítulos 7 e 8 de Schwarze e Alencar (2016): a) desenvolvemos um transdutor lexical, de modo a podermos trabalhar com lemas e não mais com formas finitas que exigem entradas individuais em nosso léxico⁵²; b) acoplamos esse transdutor à nossa gramática XLE; c) ampliamos as regras para possibilitar a análise da passiva em tempos passados, como em: *a fada tem sido esperada pelo cavaleiro*; e, por fim, d) ampliamos nosso conjunto de teste-negativo, de modo a apurar ao máximo nossa gramática, evitando que sequências agramaticais ainda sejam analisadas.

No livro base utilizado para a construção de nossa gramática, ao longo de oito capítulos, Schwarze e Alencar (*op. cit.*) apresentam de maneira didática como fenômenos de uma língua românica, no caso o francês, podem ser analisados pelo formalismo da LFG e implementados em XLE. Nosso trabalho foi, então, o de adaptar

⁵¹ Expresso aqui meus agradecimentos ao colega do Programa de Pós-Graduação em Linguística da UFC, Daniel de França Brasil Soares, pelo papel fundamental na adaptação e desenvolvimento das gramáticas que utilizamos em nossas pesquisas.

⁵² Pelo paradigma de conjugação verbal do português, cada verbo, salvo exceções de verbos defectivos, apresenta, para os diferentes tempos e modos, mais de 50 formas de conjugação. O acréscimo de cada uma dessas formas, individualmente, representaria um trabalho exaustivo. O uso de ferramentas que utilizam tecnologia de estados finitos, como o XFST, da Xerox, nos possibilita um melhor manejo dessas formas.

para o português uma gramática que fosse capaz de lidar com fenômenos diversos, e nosso objetivo final consistiu em ampliar sua cobertura, indo além do escopo do livro publicado na Alemanha, cobrindo a versão mais recente da FrGramm (ALENCAR, 2017).

Uma gramática computacional volta-se para um excerto de uma língua, e, nesta pesquisa, o foco foi o tratamento da voz passiva analítica. Para isso, entretanto, fez-se necessário a implementação e adaptação de uma série de outros fenômenos linguísticos diversos, os quais estavam também presentes nas sentenças analisadas. Nossa gramática, em seu estado atual, é capaz de lidar com os seguintes fenômenos do português: (i) concordância verbal e nominal; (ii) sintaxe dos adjetivos; (iii) valência verbal; (iv) complementos oracionais com *que* e *se*; (v) sintagmas preposicionais na função de adjuntos; (vi) advérbios de gradação; e (vii) verbos de controle. Além disso, nossa gramática consegue lidar (viii) com a passiva perifrástica no tempo presente e passado, e (ix) também com o pretérito perfeito composto.

5.1 A Gramática G-V

Com um intuito meramente didático-prático, apresentaremos, de início, a gramática G-V⁵³. A partir dessa gramática teceremos considerações acerca da adaptação realizada a partir da gramática LFG do francês da autoria de Schwarze e Alencar (2016). Em seguida, mostraremos as principais modificações que foram realizadas para a criação da gramática G-A.

É fundamental levar em conta o conceito de desenvolvimento em espiral, comum em *design de software*, de que nos fala Alencar (2017). Segundo o autor, em primeiro lugar, nessa técnica, apenas uma parte do problema que se pretende solucionar é atacada. A metáfora da espiral entra em cena, e o que se tem a seguir são aprimoramentos do protótipo inicial, os quais ocorrem em etapas sucessivas, com o fito de dar conta do problema como um todo. Na elaboração de uma gramática, essa técnica “consiste em começar com a implementação de um fragmento reduzido do recorte gramatical, expandindo a cobertura desse protótipo por meio da implementação de fragmentos sucessivamente mais abrangentes” (*op. cit.*, p. 362).

Face ao método que acabamos de expor, quisemos aqui deixar claro que

⁵³ Essa gramática, código-fonte e conjuntos-teste, encontra-se disponível para *download* no seguinte sítio: <https://github.com/DanielFBrasil/lfg-portuguese-grammar>.

embora a maior parte das considerações neste capítulo dirijam-se às peculiaridades e mudanças que tivemos de acrescentar em nossa gramática G-V para que sentenças passivas pudessem ser analisadas, os demais fenômenos linguísticos envolvidos precisaram ser antes desenvolvidos. Desse modo, ainda que não as exponhamos neste texto, várias foram as decisões tomadas por nós durante a implementação da gramática para o PB, seguindo os passos dos 5 capítulos iniciais de Schwarze e Alencar (2016).

Como já descrito em nossa metodologia, o primeiro passo para a criação de uma minigramática de uma dada língua é a criação de um *corpus* em que o fenômeno alvo esteja presente. A LFG é um modelo teórico matematicamente explícito; disso resulta, segundo Alencar (2017, p. 362) “que a modelação de um fenômeno gramatical precisa se restringir a um fragmento da língua, i. e. um conjunto definido de sentenças”.

Partiremos então para a descrição referente à adaptação dos Capítulos 6 e 8, nos quais a passiva e o passado composto (*passé composé*) são implementados para o francês, e, baseando-nos no *corpus* utilizado por Schwarze e Alencar (2016), apresentamos abaixo as sentenças cujas estruturas nos serviram para a criação das regras sintagmáticas referentes à passiva e à alimentação de nosso léxico:

- 1) A fada é esperada por um cavaleiro.
- 2) Ele é esperado por uma dama.
- 3) A fada é esperada.
- 4) O cavaleiro é esperado por uma dama.
- 5) O anel é dado a uma dama por uma fada.
- 6) Os anéis são dados a uma dama.
- 7) O cavaleiro vê a floresta habitada por umas fadas.
- 8) O cavaleiro vê o castelo branco habitado por uma rainha.
- 9) O cavaleiro tem esperado a dama.
- 10) A fada tem chegado.
- 11) As fadas têm chegado.
- 12) O cavaleiro foi esperado por uma fada.
- 13) O cavaleiro tem sido esperado por uma fada.
- 14) O cavaleiro tem sido forçado por uma rainha a acabar a tarefa.

- 15) O cavaleiro tem sido forçado a acabar a tarefa por uma rainha.
- 16) A dama quer ser convidada para dançar por um cavaleiro.

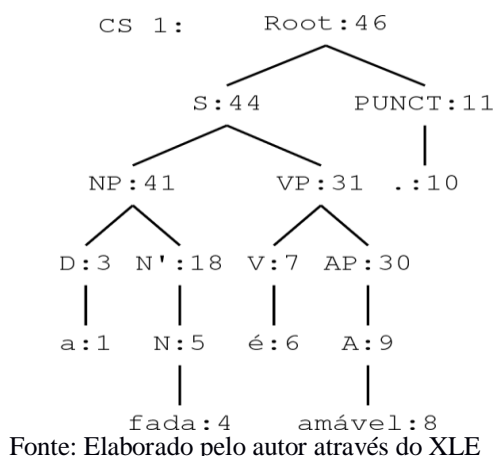
As sentenças acima contêm: i) construções tradicionalmente classificadas como passivas (1-6); ii) sentenças que apresentam um particípio que não é o verbo principal da sentença, mas sim parte de um NP modificado por uma expressão verbal (*habitada* e *habitado*, respectivamente em (7) e (8)); iii) sentenças que apresentam o pretérito perfeito composto em PB (9-11); iv) sentenças cuja passiva encontra-se no passado simples e composto (respectivamente 12 e 13); v) sentenças na voz passiva com verbos de controle que subcategorizam um XCOMP (14 e 15); e vi) sentenças que apresentam um verbo modal e outra na voz passiva (16).

O grupo iii) é importante para alimentar nossa gramática com informações que diferenciem uma construção passiva de uma construção no passado composto referido. A primeira, em uma abordagem clássica, apresenta um verbo auxiliar em sua forma finita (até o presente momento, implementamos, em nossa gramática, orações passivas apenas com o verbo *ser*) e um particípio que apresenta traços de concordância de número e gênero. No caso do passado composto em PB, o que temos é uma construção perifrástica formada pelo verbo auxiliar *ter* em sua forma finita e por um particípio verbal, que, diferentemente das construções passivas, não apresenta, em PB, traços de concordância de gênero e número, à diferença do que ocorre com a língua francesa⁵⁴.

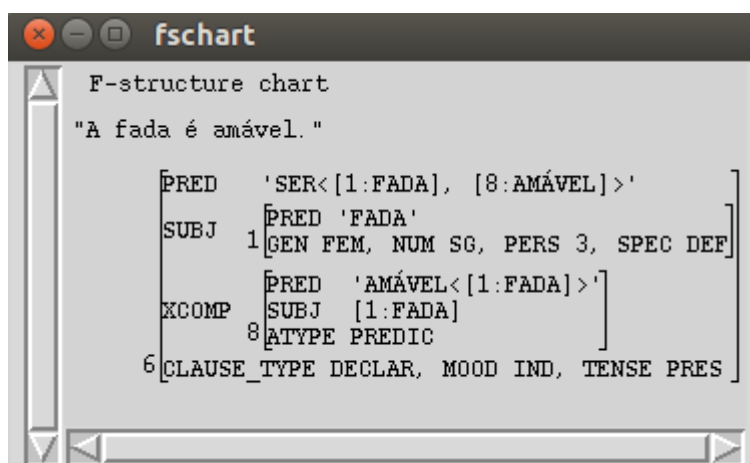
Um ponto importante da implementação em nossa gramática diz respeito à homonímia do verbo *ser* que pode apresentar-se como verbo cópula ou como verbo auxiliar. Ainda que esse verbo, nesses dois usos, apresente-se de modo idêntico na perspectiva fonética e morfológica, o mesmo não se verifica em nível sintático; conseqüentemente, teremos diferentes estruturas C e F para cada uma das funções do verbo *ser*.

À guisa de exemplo dessas diferenças, observe-se abaixo a estrutura C do verbo *ser* quando cópula na frase *A fada é amável*. Nesse caso, o V pode ter como irmão à direita um AP ou NP.

⁵⁴ Por exemplo: *Les chaises que nous avons achetées sont belles*, em que o particípio passado *achetées* apresenta marcas de concordância de gênero (feminino) e de número (plural).

FIGURA 19 – Estrutura C de sentença com verbo *ser* em função de cópula

A estrutura F da mesma sentença é a que segue:

FIGURA 20 – Estrutura F de sentença com verbo *ser* em função de cópula

Fonte: Elaborado pelo autor através do XLE

Considere-se agora o verbo *ser* como auxiliar da passiva. Em uma análise como esta, o verbo principal está na forma infinita do particípio. Segundo Schwarze e Alencar (2016, p. 152), o particípio passivo é o ponto mais alto do predicado (*das „höchste“ Prädikat*) da frase passiva, sendo ele o portador do significado lexical e também o responsável por reger as funções gramaticais. A concordância entre particípio e sujeito é assegurada por meio de equações correspondentes presentes na entrada lexical do particípio. Tomemos, como exemplo, a entrada lexical do particípio passivo *esperado* na frase *O cavaleiro é esperado*.

- (1) esperado V * (^ PRED)='ESPERAR< (^ SUBJ)>'
 (^ VFORM)= PART_PASS
 (^ SUBJ GEN)=MAS
 (^ SUBJ NUM)=SG

A entrada lexical acima dá conta da análise de uma sentença em que o agente da passiva não esteja presente, como ocorre, por exemplo, na sentença (3). Por meio da etiqueta VFORM, que, neste caso, recebe o valor de PART_PASS, a entrada lexical de *esperado* diferencia-se da entrada lexical do verbo na forma do particípio passado em uma sentença ativo, em que o valor de VFORM é distinto. Trata-se também aqui de um exemplo de homonímia que se resolve no nível lexical⁵⁵.

Analisemos agora exemplos em que o agente da passiva esteja presente. Em nossa gramática G-V, o verbo *ser* é auxiliar da passiva e, por conseguinte, não possui predicado, ou seja, não apresenta, na estrutura F do XLE, o traço PRED, sendo semanticamente vazio. Na seção sobre a voz passiva, vimos que este é um caso de um chamado verbo de alçamento, e, como tal, trata-se de um núcleo funcional e “sua contribuição para a estrutura F da sentença é meramente gramatical” (ALENCAR, 2017, p. 359), trazendo informações acerca do tempo e modo em que a sentença se encontra.

Em contrapartida, em nossa G-A, o verbo *ser* possui representação semântica PRED. Segundo Alencar (2015a, p. 40), somente verbos principais possuem esse traço, herdado a partir da estrutura funcional. Por essa razão, *ser* não pode reger um agente da passiva, expresso em PB geralmente pela preposição *por*. Nessa situação, é o próprio particípio passivo que exige a presença do agente da passiva.

Podemos, então, expandir a entrada lexical expressa em (2) para ocorrências como na sentença *O cavaleiro é esperado por uma fada*. No caso desse PP, trata-se de objeto oblíquo (OBL, em LFG), espécie de objeto que se relaciona menos diretamente com o verbo (KROEGER, 2004). Esse difere-se da função OBJ2 (objeto indireto ou secundário), na medida em que é um complemento verbal preposicionado que não aceita pronominalização por clítico dativo, ao passo que os OBJ2 podem sofrer processo de cliticização (ALENCAR, 2017, p. 361). Vejam-se os exemplos abaixo:

- (2) a. A fada dá um anel ao cavaleiro.

⁵⁵ Ver entrada lexical de *esperado* na gramática G-V no arquivo <homonyms.lfg>.

A fada dá-lhe um anel.

b. O anel foi dado à fada pelo cavaleiro.

*O anel foi dado à fada por lhe.

Em (2a), *o cavaleiro* desempenha a função de OBJ2, motivo pelo qual se deixa cliticizar pelo pronome oblíquo *lhe*. Em (2b), contudo, o mesmo sintagma nominal *o cavaleiro* não se deixa cliticizar, como se pode depreender da agramaticalidade da sentença.

A entrada lexical expandida para dar conta dos casos em que um PP exerce a função de OBL encontra-se em (3) abaixo. A primeira linha introduz o caso oblíquo regido pelo verbo *esperar*.

(3) esperado V * { (^ PRED)='ESPERAR<(^ OBL)(^ SUBJ)>'
 (^ OBL CASE)=c POR |
 (^ PRED)='ESPERAR<NULL (^ SUBJ)>'
 (^ VFORM)= PART_PASS
 (^ SUBJ GEN)=MAS
 (^ SUBJ NUM)=SG

Ainda sobre (3), no tocante à preposição *por*, trata-se de uma chamada preposição funcional, cuja função é apenas a de fornecer um caso ao argumento do verbo (cf. BUTT et al., 1999, p. 129). No XLE, esse caso é assinalado através do atributo CASE. A obrigatoriedade da preposição é determinada por meio da equação restritiva (em alemão, *fordernde Gleichung*; em inglês, *constraining equation*), caracterizada pelo símbolo '=c'. Na prática, com essa equação, exige-se que o PP que realiza o caso oblíquo tenha para o traço CASE desse complemento o valor POR; se isso não ocorrer, não será gerada nenhuma estrutura F bem formada (SCHWARZE & ALENCAR, 2016, p. 61). A entrada lexical correspondente à preposição funcional *por* é:

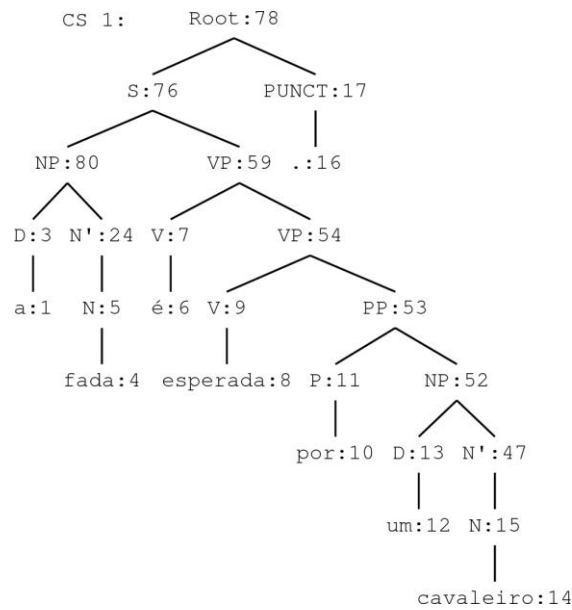
(4) por P * (^ CASE)=POR.

O que se observa também a partir da entrada lexical de *esperada* em (3) é que as possibilidades de realização sintática já estão previstas na própria entrada lexical

de um dado item através da unificação das restrições e valências. Desse modo, “o léxico não é simplesmente gerado por meio de regras de inserção lexical, (...) mas a própria sentença é produzida a partir das possibilidades combinatórias definidas no léxico” (MARTINS, 2015, p.133). É assim que o léxico ganha papel primordial num modelo lexicalista, passando inclusive a reger as próprias regras gramaticais e a controlar a boa formação de sentenças.

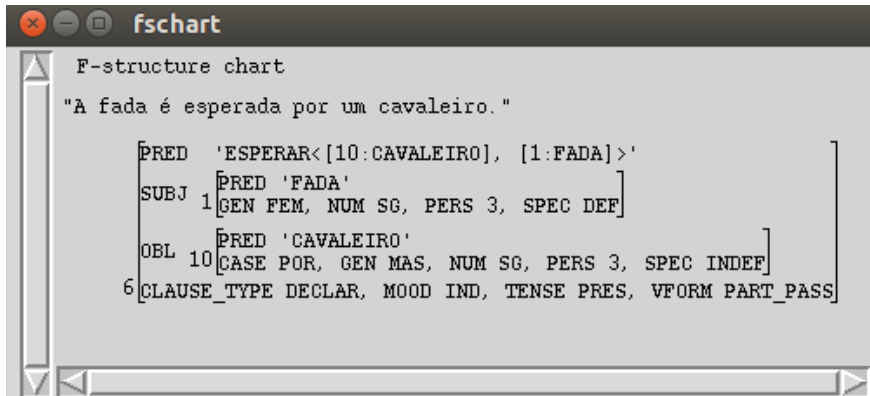
Apresentaremos agora as estruturas C e F da sentença *A fada é esperada por um cavaleiro* analisadas em nossa G-V:

FIGURA 21 – Estrutura C da sentença (1)



Fonte: Elaborado pelo autor através do XLE

FIGURA 22 – Estrutura F da sentença (1)



Fonte: Elaborado pelo autor através do XLE

É possível observar na estrutura F acima que se trata de uma sentença passiva através do VFORM PART_PASS. O traço PRED='ESPERAR<(SUBJ)(OBL)>' diferencia-se do da forma ativa, pois este último exige como argumentos um sujeito e um objeto. Além disso, o traço CASE POR informa-nos que o caso oblíquo deve ser marcado pela preposição *por*. Do ponto de vista semântico, o que temos é um argumento com função temática de agente.

Por fim, retomando a discussão sobre o verbo *ser* e seus homônimos, mostraremos abaixo a entrada lexical completa do verbo *ser* em uma de suas formas finitas:

(5) é V * { (^ PRED)='SER<(^ SUBJ)(^ XCOMP)>'
 (^ SUBJ)=(^ XCOMP SUBJ)
 @(CAT (^ XCOMP) {AP PP})
 (^ TENSE)=PRES|
 (^ AUX) = SER
 (^ PASSIVE) = +
 (^ VFORM) =c PART_PASS
 (^ TENSE)=PRES }
 @(V-AGR 3 SG)
 @IND.

Observando-se a entrada lexical da forma finita *é* em (5), temos as duas ocorrências do verbo *ser* supracitadas, respectivamente, verbo cópula e verbo auxiliar da passiva, implementadas por meio do símbolo de disjunção “|”. Para que a entrada lexical contenha o menor número de informações possíveis, utilizam-se as chaves para separar as duas últimas linhas que contêm informações comuns às duas formas verbais, ou seja, independentemente de se tratar de cópula ou auxiliar, o verbo *é* apresenta traços de concordância de terceira pessoa do singular no modo indicativo. Esses últimos traços encontram-se descritos de uma forma mais elegante através de *templates*, moldes já definidos que contêm informações repetitivas. Os *templates* utilizados em nossa gramática podem ser vistos no arquivo designado como *templates.lfg*. Na prática, ao deparar-se com o símbolo @, detecta-se automaticamente, através do XLE, que nesse ponto haverá um molde já definido, diminuindo assim a necessidade de se acrescentarem informações repetitivas em cada entrada lexical individual. O *template*

“@IND”, por exemplo, pode ser usado para todos os verbos que estejam em sua forma indicativa, e o “@(V-AGR 3 SG)”, para definir os traços de um verbo cujo sujeito é de terceira pessoa do singular. Em (6) e (7) temos o conjunto de informações resumidas por meio desses *templates*:

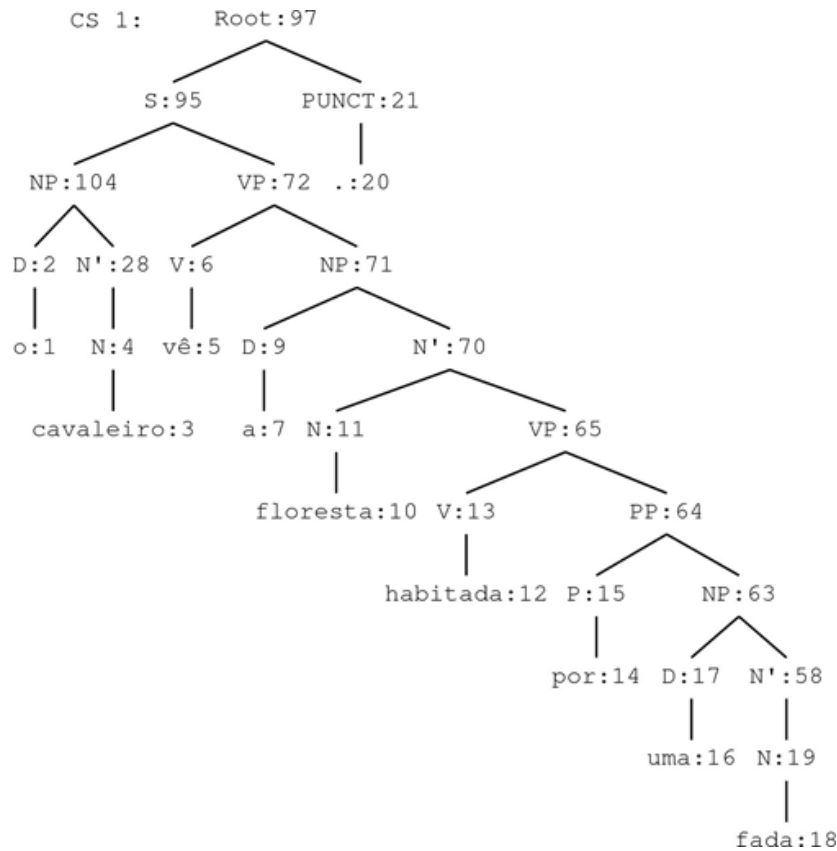
(6) IND = (^ MOOD)=IND.

(7) V-AGR(P N) = (^ SUBJ PERS) = P
(^ SUBJ NUM) = N.

Em (7) temos os parâmetros P e N, respectivamente, pessoa e número, que, em nosso exemplo, são atualizados com os valores de 3 e SG.

Partiremos agora para a descrição e análise de fenômenos presentes em sentenças como a exemplificada na sentença 7) de nosso corpus positivo: *O cavaleiro vê a floresta habitada por umas fadas*. Como já mencionado, o que há nesse tipo de sentença é um NP modificado por um VP, cujo núcleo V se encontra na forma participial. Nesse exemplo, *habitada* não é o verbo principal da sentença, mas sim parte do NP que tem como núcleo *floresta*. Da estrutura C abaixo vemos que esse VP complexo funciona como adjunto de N', uma vez que o modifica:

FIGURA 23 – Estrutura C da sentença (7)



Vejamos também a estrutura F gerada pelo XLE na FIGURA 24 abaixo. Diferentemente do exemplo (1), a sentença (7) não é uma oração passiva. Isso fica claro quando comparamos as respectivas estruturas F e notamos que o traço VFORM=PART_PASS não aparece, na sentença (7), ao final da AVM maior, como característica geral da sentença, mas sim encapsulado dentro da AVM referente ao adjunto (indicado pelo índice 12 na estrutura F da Figura 24), como característica específica deste. Isso nos mostra que um atributo pode ter valores não atômicos:

o poder descritivo das AVMs como formalismo para a descrição de estruturas linguísticas advém justamente da possibilidade de um atributo ter como valor uma outra AVM. Com isso, esse formalismo pode dar conta da recursividade das estruturas sintáticas nas línguas naturais (ALENCAR, 2017, p.358)

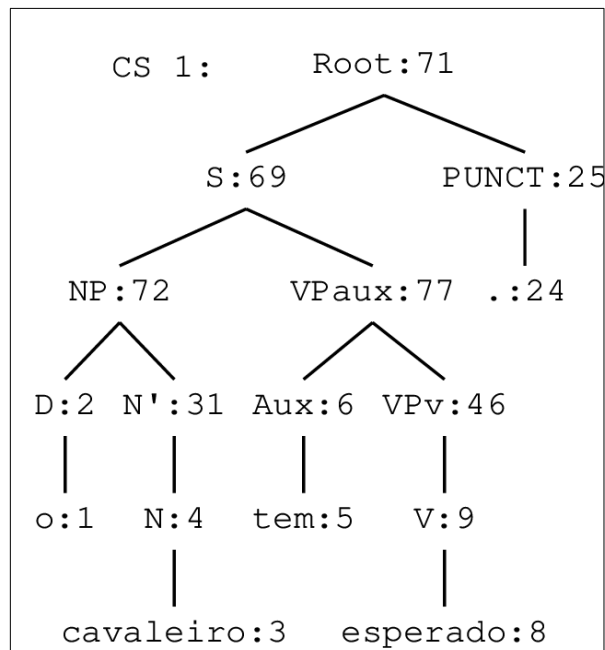
O que temos na frase (7) é uma sentença ativa que apresenta um participio funcionando como adjunto de um objeto. Essa sentença apresenta os traços CLAUSE_TYPE=DECLAR, MOOD=IND e TENSE=PRES, ou seja, é uma sentença

- (10) esperado V * { (^ PRED)='ESPERAR<(^ SUBJ)(^ OBJ)>' |
 (^ PRED)='ESPERAR<(^ SUBJ)>'}
 (^ VFORM) = **PART_PAST**
 (^ AUX) = TER}.

Os esquemas funcionais mostrados em (9) e (10) deixam claro que as funções argumentais do verbo, seja em sua forma finita ou em sua forma de particípio passado, permanecem as mesmas. A subcategorização de ambas as entradas é a mesma. Chama-se de subcategorização (cf. KROEGER, 2004, p.17) o conjunto de relações gramaticais que são explicitadas na entrada lexical de um verbo, as quais desempenham um importante papel na determinação do contexto sintático em que um dado verbo pode ocorrer. Em nossa minigramática, o verbo *esperar* apresenta duas subcategorizações: pode ser ou monovalente (intransitivo, numa análise mais tradicional; inergativo, numa análise gerativa mais moderna), apresentando como único argumento um sujeito; ou bivalente (verbo transitivo direto) com argumentos sujeito e objeto direto⁵⁷. Outrossim, um novo valor foi adicionado para o atributo VFORM: com a etiqueta PART_PAST pode-se distinguir entre o particípio passado e o particípio passivo. A última linha de (10) mostra ainda que esse tempo composto se constrói em PB selecionando o auxiliar *ter*. Como não apresenta concordância com sujeito, o PART_PAST, diferentemente do PART_PASS, não apresenta traços de gênero e número.

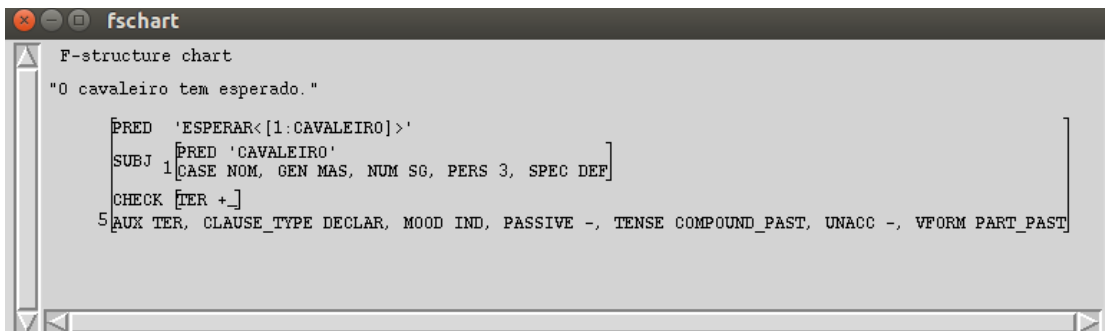
⁵⁷ Exemplos como *O cavaleiro espera por uma fada*, em que o verbo *esperar* seja transitivo indireto, ainda não foram implementados por nós.

FIGURA 25 – Estrutura C da sentença (9)



Fonte: Elaborado pelo autor através do XLE

FIGURA 26 – Estrutura F da sentença (9)



Fonte: Elaborado pelo autor através do XLE

Nossa gramática, entretanto, apresentou problemas para analisar o passado composto para verbos intransitivos como *chegar*. Acreditamos que isso se deva às implementações e regras a que nos baseamos em Schwaze e Alencar (2016) e Alencar (2017), as quais estão calcadas no francês. Nesta língua, a diferenciação entre verbos inacusativos e inergativos é imprescindível para a formação do tempo composto *passé composé*, pois daí se dá a escolha do verbo auxiliar que comporá a perífrase. Para frases como *O cavaleiro tem chegado* foram geradas duas estruturas F; a frase agramatical **A fada tem chegada* recebeu uma análise satisfatória. Essa é uma falha de nossa análise, a qual pretendemos resolver em uma versão melhorada futura.

5.1.1 *Implementação da regra da passiva*

Na seção teórica deste trabalho, discutimos que na LFG a passiva é tratada como fenômeno exclusivamente lexical, não sendo necessário que se recorra a transformações sintáticas e movimentação de constituintes derivados de uma forma ativa. O que há, como apresentado, é um realinhamento dos papéis semânticos do verbo. Retomando as asseverações que foram feitas, na LFG, uma teoria lexicalista, a passivização é instanciada por meio de uma regra lexical. Apresentamos também o princípio da Regra Universal de Passivização, presente nas línguas naturais. Em PB, a instanciação da regra da passiva tem como efeito morfológico a conversão do verbo ativo em seu particípio passivo. Os particípios passivos representam, assim, um tipo de XCOMP, no caso, um complemento verbal, os quais estão sujeitos a controle funcional e que, em línguas como o português, apresentam traços de concordância de gênero e número com seu SUBJ funcional (MORAIS, 1988).

Sendo assim, as regras lexicais são de fundamental importância na análise de diáteses da LFG, uma vez que “essas regras, junto ao restante aparato formal da teoria, dispensam a postulação de transformações sintáticas. Equivalem a funções que, aplicadas sobre entradas lexicais, geram outras entradas” (ALENCAR, 2017, p.360).

É necessário, então, formalizar a regra de derivação da passiva que se aplique ao PB. Tal formalização já foi contemplada por vários autores para diferentes línguas. Bresnan (1982), no tocante ao inglês, apresenta a formalização da passivização como um processo exclusivamente lexical, independentemente de se tratar de passivas verbais ou adjetivais. Em seu texto, defende a ideia de que a construção de estruturas passivas se dá unicamente no léxico. Outros autores defendem o mesmo tipo de posicionamento, e. g. Schwarz e Alencar (2016), Alencar (2015), Kroeger (2004), Falk (2001), Morais (1988).

Baseando-nos nos trabalhos acima elencados, citamos os princípios que devem ser levados em consideração na criação da regra lexical da passiva: a) a todo particípio passivo há uma forma correspondente ativa que rege um OBJ; b) o OBJ da forma ativa é alçado à posição de SUBJ na forma passiva; c) o SUBJ da oração ativa pode ser realizado como OBL-AG na passiva (SCHWARZE & ALENCAR, 2016, p.149)⁵⁸.

⁵⁸ Embora assumamos neste trabalho ser a passiva um processo de alçamento do objeto e, ao mesmo tempo, de demissão do sujeito, no âmbito da LFG, em estudos translinguísticos (*crosslinguistics*), trata-se essencialmente de um processo de demissão do sujeito. A ocorrência de passivas impessoais (presentes

Apresentamos abaixo o esquema da regra de passivização adaptado para o português nos moldes do proposto para o francês por Schwarze e Alencar (2016, p.149):

- (11)
- | | | | |
|-------|-------------------|-----------------|-----------|
| (i) | forma ativa: | ‘ESPERAR<(^SUBJ | (^OBJ)>’ |
| (ii) | 1ª forma passiva: | ‘ESPERAR<(^OBL | (^SUBJ)>’ |
| (iii) | 2ª forma passiva: | ‘ESPERAR<NULL | (^SUBJ)>’ |
| (iv) | grade temática: | AGENTE | TEMA |

Do esquema acima, nota-se que “a passiva pode ser modelada como uma operação que incide sobre as funções gramaticais da entrada lexical da forma ativa, derivando (...) duas entradas lexicais para o particípio passivo” (ALENCAR, 2017, p.361). Do ponto de vista dos papéis semânticos, o objeto TEMA torna-se sujeito TEMA. Essa derivação é feita por meio de transformações como as que seguem:

- (12) {SUBJ → OBL | SUBJ → NULL} OBJ → SUBJ

A passivização em português apresenta aspectos importantes tanto do ponto de vista morfológico quanto do sintático. Do ponto de vista morfológico, o que a passivização faz é converter um dado verbo ativo em seu particípio passivo. Na prática, isso equivale a dizer que o verbo que sofre o processo de passivização adquirirá a forma morfológica do particípio passivo desse verbo; em português, de maneira produtiva, temos a morfologia do tipo ‘V+do’. Morais (1988), entretanto, salienta que o efeito morfológico não é só o de transformar o verbo em seu particípio passivo, mas sim em

em línguas como polonês, latim e alemão) parece corroborar essa definição mais generalizante. Vejam-se os exemplos a seguir de passivização a) de um verbo (inergativo) intransitivo em alemão *tanzen* (“dançar”) e b) de um verbo, *gedenken* (“lembrar”) que pede complemento no genitivo:

- a) *Hier wird getanzt*
Adv. V.3.p.sg PART.PASS
there is dancing here
 “Aqui se dança” (literalmente: ‘Aqui é dançado’)
- b) *Wir gedenken der Toten.* (Forma atva)
We remember the.GEN deads
 “Nós lembramos dos mortos”
- Der Toten wird gedacht.* (Forma passiva)
the.GEN deads are being remembered
 “Os mortos são lembrados”

seu particípio passivo, flexionado em gênero e número com o novo sujeito. A concordância entre particípio e sujeito será feita, como já mencionada neste trabalho, nas respectivas entradas lexicais cujos traços serão submetidos ao já referido princípio da unificação.

Sobre o aspecto sintático, observe-se a regra (12) acima, cuja primeira parte mostra que o SUBJ pode ou ser transformado em OBL, ou convertido em NULL; neste último caso, o resultado é seu apagamento. Na segunda parte da regra, o OBJ é transformado em SUBJ. É preciso observar ainda que a conversão de SUBJ em OBL, em línguas como o português, francês, inglês, espanhol, italiano etc., vem acompanhada de uma mudança na posição dos elementos na frase. Ou seja, há uma reconfiguração dos NPs presentes na sentença.

Cançado (2016) assevera que, no tocante à relação entre papéis temáticos e sintaxe, respaldando-se em vasta literatura (para citar alguns autores: FILLMORE (1968); JACKENDOFF (1972), (1983), (1990); GIVÓN(1984) etc.), em português, devido ao Princípio da Hierarquia Temática ou Princípio de Ligação entre sintaxe e semântica, “o argumento que denota o agente sempre está associado à posição de sujeito e o paciente à posição de complemento” (p.64). Isso se deve ao fato de ser o PB uma língua configuracional.

Em línguas configuracionais como as românicas citadas acima, ou seja, em línguas nas quais as funções de SUBJ e OBJ são expressas configuracionalmente de modo distinto na estrutura da sentença, reserva-se a posição de SPEC de IP, quer dizer, a de argumento externo do verbo, para o NP SUBJ. Por seu turno, a função OBJ é prototipicamente expressa como argumento interno do verbo: o NP que ocupa a função de objeto é imediatamente dominado pelo nó V. Os papéis temáticos canonicamente atribuídos às posições de SPEC de IP e complemento interno de V são, respectivamente, AGENTE e TEMA (PACIENTE). A exceção fica a cargo justamente de sentenças passivas que apresentam uma ordem não canônica, haja vista que o AGENTE não ocupa a posição SPEC de IP. Nelas, o argumento externo do verbo tem papel temático de TEMA, e o NP com papel temática de AGENTE, quando presente, é realizado como oblíquo e aparece na forma de adjunto do sintagma verbal. Por fim, pode-se afirmar que o realinhamento das funções gramaticais e dos papéis temáticos resultantes da passivização impõe uma mudança na posição dos NPs de uma sentença.

À guisa de comparação, nas línguas não-configuracionais, cite-se novamente o Wahlpiri, as quais expressam sintaticamente as funções gramaticais e,

consequentemente, cuja ordem dos sintagmas é livre, a passivização não determina uma mudança na posição dos NPs, mas sim alterações dos casos morfológicos. Resumindo, em línguas não-configuracionais, os efeitos da passivização não envolvem necessariamente mudanças estruturais. O interessante é que a LFG possibilita generalizações acerca do processo de passivização em línguas de tipologias diversas, permitindo ainda um mecanismo universal para representação das relações gramaticais que ocorrem neste processo (MORAIS, 1988).

Feitas essas observações de caráter mais teórico, discutimos os pontos principais acerca da implementação dessa regra da passiva dentro de nossa gramática, mais especificamente no XLE. Começamos destacando a importância de se trabalhar com regras que incidam sobre entradas lexicais, pois, desse modo, a implementação da regra da passiva em nossa gramática é importante para evitar que todos os participípios precisem ser acrescidos individualmente em nosso léxico. Com a regra, temos uma simplificação da codificação do léxico e, ao mesmo tempo, uma economia do espaço de armazenamento (ALENCAR, 2017, p. 354). Como, porém, essas regras apenas agem nos esquemas funcionais das entradas lexicais, não podendo manipular diretamente os lexemas, fez-se necessário um analisador morfológico, o qual foi por nós desenvolvido e, em seguida, acoplado à nossa gramática.

O uso do analisador morfológico possibilitou, dentre outras coisas, uma integração entre morfologia e léxico no tratamento da passiva, o que tornou o trabalho de alimentação do léxico mais simples, uma vez que apenas radicais como *esper-* necessitaram ser acrescidos.

Nesse ponto, cumpre assinalar que o uso de ferramentas tais como o analisador morfológico acoplado à nossa gramática, além de eficaz, parece dirimir críticas como as apontadas por Martins (2015) acerca dos formalismos de representação lexical e baseados em mecanismos de unificação de traços. Para o autor:

Embora tenham mantido o compromisso com o rigor formal, as gramáticas de unificação terminam por sobrecarregar o trabalho lexicográfico e inviabilizaram-se como alternativa robusta para o processamento automático das línguas naturais. A necessidade de incorporação, ao léxico, de todo o complexo de traços previstos pelos formalismos de unificação terminou por se revelar pouco realista, embora tenha inspirado inúmeros *parsers* de uso experimental (MARTINS, 2015, p.134)

As críticas apontadas, queremos crer, não nos parecem aplicáveis se se levam em conta novos mecanismos que podem ser incorporados às gramáticas

LFG/XLE. A parte disso, o próprio mecanismo de desenvolvimento em espiral (cf. ALENCAR (2017)) e de compartilhamento de informações e gramáticas, comum no âmbito de divulgação da LFG (vide o já referido projeto ParGram), apresentam-se como alternativa viável para a não “necessidade de incorporação” de todas as informações em nosso léxico, como se tudo tivesse de ser criado e definido desde o princípio.

Como já referido, nosso trabalho, tem por fito também servir de um modelo para possíveis outras implementações, tanto para o português como para outras línguas tipologicamente semelhantes, num movimento dialético de análise, passível de implementações contrárias ou semelhantes. Implementações já bem realizadas não precisariam assim ser novamente desenvolvidas do zero, ao mesmo tempo em que pontos falhos, alguns dos quais já apontamos, podem ser adaptados ou completamente modificados a fim de se melhorar a gramática.

O primeiro passo para o tratamento da passiva foi a expansão de nossa regra do VP, de modo que esse pudesse apresentar à sua direita um PP que exercesse a função OBL:

(13) VP --> V ({ **VP**
 NP: (^ OBJ)=!; PP: (^ OBJ2)=! (! CASE)=c A
 | NP: (^ OBJ)=!
 | **PP: (^ OBL)=!**
 | PP: (^ XCOMP)=!
 | AP: (^ XCOMP)=! (! ATYPE)=PREDIC
 | PP: (^ OBJ2)=! { (! CASE)=c PARA | (! CASE)=c A } ; CP: (^ COMP)=!
 | CP: (^ COMP)=!
 })

Em seguida, implementamos o esquema da passiva, mostrado abaixo, o qual realiza a regra lexical da passiva em nossa G-V. Esse esquema, acionado nas entradas lexicais de nossos lemas, pelo template @PASSIVE, transforma a entrada de um verbo como o do verbo transitivo *esperar*, (^PRED) = 'esperar<(^SUBJ)(^OBJ)>', transformando o OBJ em SUBJ e, ao mesmo, tempo apagando o SUBJ da forma ativa ou o realizando na forma de PP-OBL regido pela preposição *por*. Além disso, o verbo é convertido em particípio passivo 'esper-ado' por meio do analisador morfológico.

- (14) PASSIVE(_SCHEMATA) = _SCHEMATA
 (^ VFORM) = PART_PASS
 (^ OBJ --> (^ SUBJ)
 { (^ SUBJ --> NULL
 | (^ SUBJ --> (^ OBL)
 (^ OBL CASE) =c POR }.

5.2 A Gramática G-A

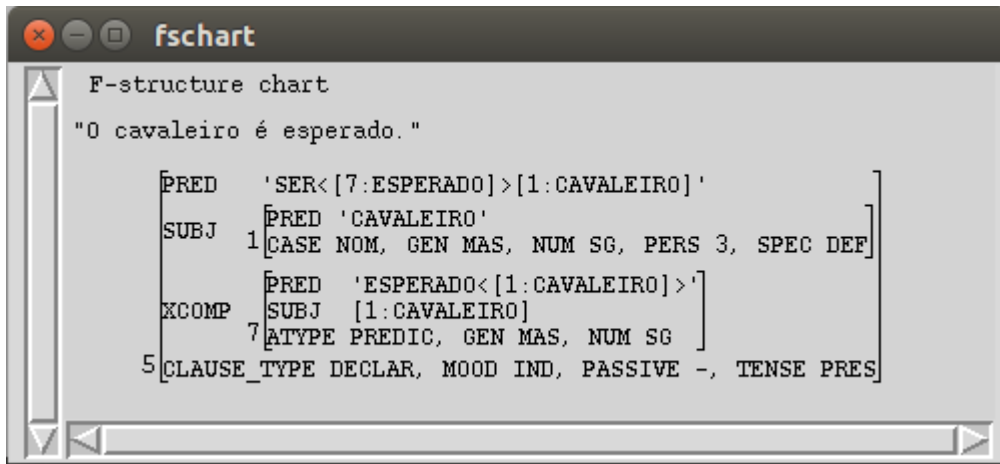
Nossa G-V é uma minigramática do PB desenvolvida em LFG/XLE com molde na gramática do francês desenvolvida por Schwarze e Alencar (2016). No tocante à passiva e ao particípio passivo, essa gramática está implementada para atribuir ao particípio a categoria V. A partir da gramática G-V, desenvolvemos a variante G-A, que atribui ao PART_PASS a classificação de A. Apresentamos nessa seção as principais modificações e adaptações que foram feitas em nossa gramática G-A.

Esse tipo de análise do PART_PASS como ADJ já foi realizado em gramáticas LFX/XLE para o polonês, a POLFIE (PATEJUK, 2013) e para o português na BrGRAM (Alencar, 2013) e na gramática de SANTOS (2014). O caráter distintivo de nosso trabalho, entretanto, é construir duas gramáticas cujas diferenças consistam apenas no tocante à forma de análise das sentenças passivas analíticas a fim de verificar qual o tipo de análise mais eficiente do ponto de vista linguístico-computacional.

Em seu artigo a respeito da gramática LFX/XLE para o polonês, Patejuk e Przepiórkowski (*id.*) trazem à baila argumentos que justificam sua análise da passiva em polonês como formação predicativa adjetival. Em sua gramática POLFIE, o verbo *być* ‘ser’ é um verbo cópula, cujo único argumento semântico é um XCOMP (Alencar, 2017, p.355).

Análise similar é implementada em nossa G-A e, dessa forma, numa sentença como (3), o sujeito *cavaleiro* representa um argumento não do verbo cópula *é*, mas sim do XCOMP *esperado*. Alencar (*id.*, *ib.*) classifica esse tipo de análise como bipredicacional: há duas predicções, a primeira delas expressa pelo verbo cópula, a segunda, pelo XCOMP. É esse o tipo de tratamento para a passiva em nossa G-A; nela o verbo cópula instancia um XCOMP, ou seja, nessa análise as construções tradicionalmente classificadas como passivas não difeririam de construções predicativas adjetivais canônicas. Vejamos a estrutura F da sentença (3) agora gerada por nossa G-A:

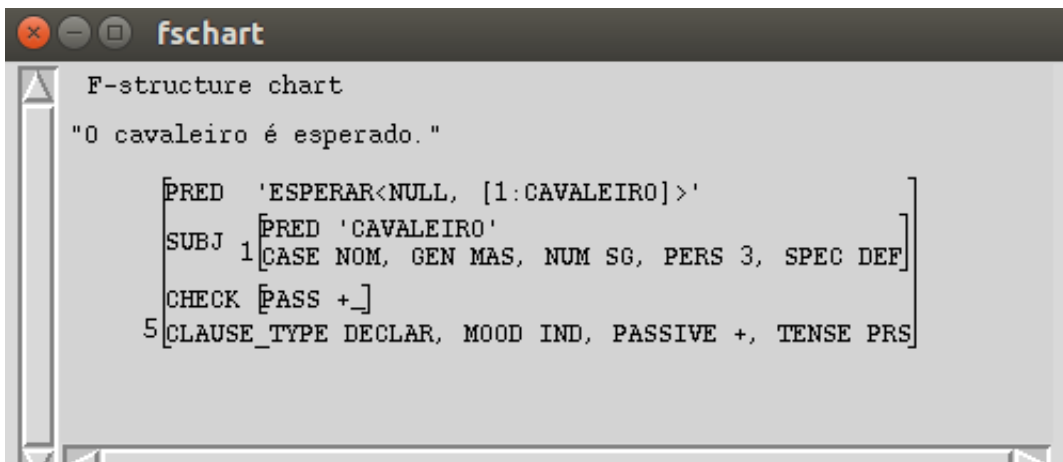
FIGURA 27 – Estrutura F da sentença (3) gerada pela G-A



Fonte: Elaborado pelo autor através do XLE.

Comparativamente, nossa abordagem mais tradicional representada pela G-V é monopredicacional, pois nela, o verbo *ser* funciona como auxiliar, não apresentando o traço semântico PRED:

FIGURA 28 – Estrutura F da sentença (3) gerada pela G-V



Fonte: Elaborado pelo autor através do XLE

No tocante às mudanças que foram feitas nas regras da gramática G-V, a Regra 7 teve de ser modificada para que a classificação do particípio nessas construções como A fosse possível. O suporte teórico do qual nos valem foi sobretudo o trabalho de Di Fellipo e Dias-Da-Silva (2005) acerca da valência adjetival e sua representação computacional. Na análise dos autores, os adjetivos são predicadores e determinam a categoria sintagmática dos constituintes que podem ou não ocorrer. Ou seja, nessa abordagem, os adjetivos, estejam em posição adnominal ou predicativa, também apresentam um quadro de subcategorização, similar àquele com que estivemos

trabalhando no caso dos verbos.

Para dar conta dessa abordagem, modificamos a regra do AP, de modo que o A possa receber à sua direita i) um PP complemento oblíquo (*o cavaleiro é esperado por uma rainha*) ; ou ii) um PP oblíquo e um PP com função de OBJ2 (*O anel é dado a fada por um cavaleiro*); iii) ou ainda um PP oblíquo e um CP com função de XCOMP (*O cavaleiro é forçado por uma rainha a acabar a tarefa*). A regra inicial, mais simples e presente em nossa G-V, apresentada em (15), foi modificada para (16) em nossa G-A:

(15) "R7" AP --> (DEG) A .

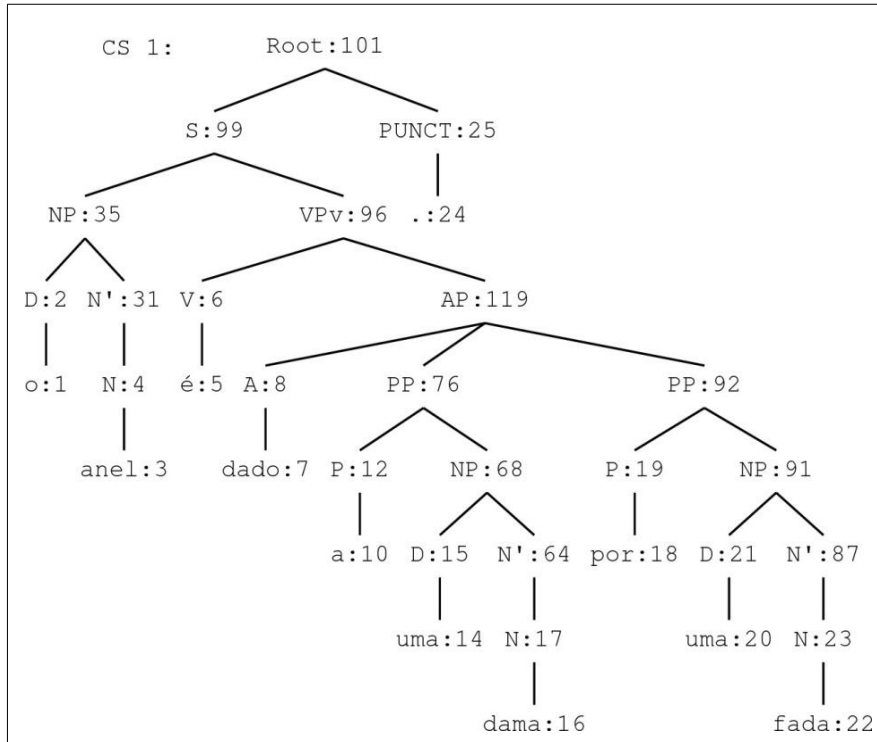
(16) R7" AP --> (DEG) A ({ **OBL-PP** | **OBL-PP, IO** | **OBL-PP, CP: (^ XCOMP)=!** })

Em (16), mostramos em negrito as modificações nas regras de nosso AP, a fim de que esse permitisse um complemento ou do tipo OBL ou do tipo OBJ2. Veja-se em (16) que o uso dos parênteses indica opcionalidade dos complementos à direita do núcleo A. Além disso, os três diferentes tipos de complementação exigidos pelo núcleo do sintagma adjetival foram acrescentados à regra por meio de disjunção lógica '|', indicando que cada uma das possibilidades pode ocorrer somente quando as outras duas não ocorram. A vírgula que separa o OBL-PP dos dois outros tipos de complemento, como já apresentado, serve para indicar que os elementos podem vir em qualquer ordem, sendo assim, tanto as frases a) e b) abaixo podem ser analisadas:

- (17) a) O cavaleiro é forçado a acabar a tarefa por uma rainha.
b) O cavaleiro é forçado por uma rainha a acabar a tarefa.

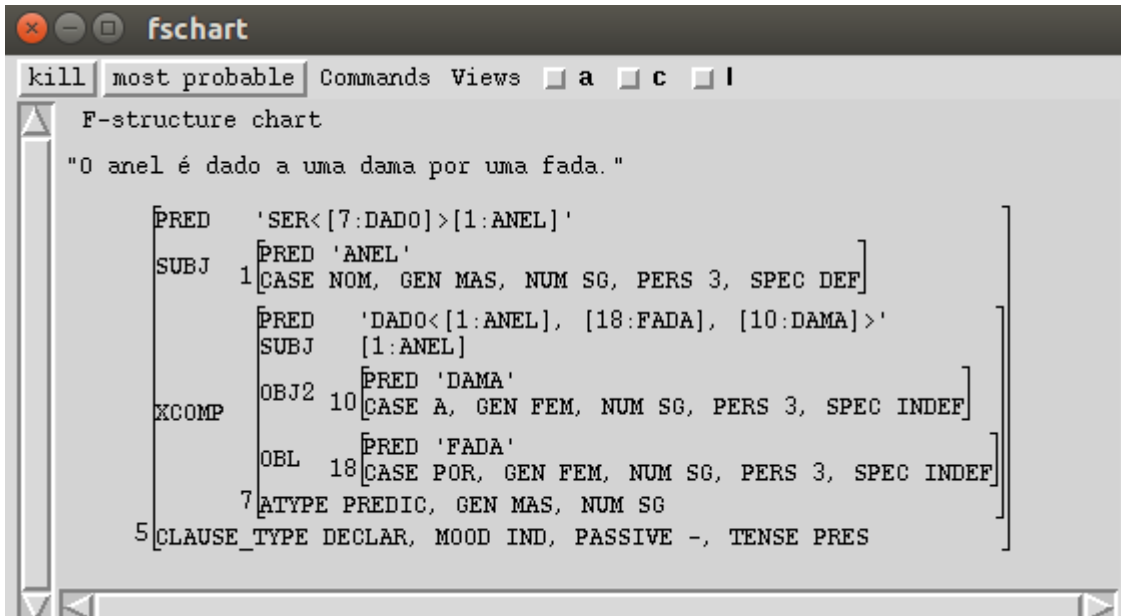
No caso da sentença (5) de nosso teste positivo, essa regra faz-se valer mais claramente, como se deixa depreender da estrutura C abaixo:

FIGURA 29 – Estrutura C da sentença (5)



Fonte: Elaborado pelo autor através do XLE

FIGURA 30 – Estrutura F da sentença (5)



Fonte: Elaborado pelo autor através do XLE

Outra modificação necessária foi no nível do léxico. Inicialmente, em uma versão mais antiga de nossa G-A, as formas similares a *dado*, por exemplo, as quais na G-V se encontravam listadas junto aos verbos, foram alocadas, no léxico da G-A, junto aos adjetivos, no masculino e feminino, no singular e plural, e se lhes foi atribuída a

dada A * { (^ PRED)='DADO<(^ SUBJ)>'
 (^ SUBJ GEN)=FEM
 (^ SUBJ NUM)=SG |
 (^ PRED)='DADO<(^ SUBJ)(^ OBJ2)>'
 (^ OBJ2 CASE)=c A
 (^ SUBJ GEN)=FEM
 (^ SUBJ NUM)=SG |
 (^ PRED)='DADO<(^ SUBJ)(^ OBJ2)(^ OBL)>'
 (^ OBL CASE)=c POR
 (^ OBJ2 CASE)=c A
 (^ SUBJ GEN)=FEM
 (^ SUBJ NUM)=SG }.

dadas A * { (^ PRED)='DADO<(^ SUBJ)>'
 (^ SUBJ GEN)=FEM
 (^ SUBJ NUM)=PL |
 (^ PRED)='DADO<(^ SUBJ)(^ OBL)>'
 (^ OBL CASE)=c A
 (^ SUBJ GEN)=FEM
 (^ SUBJ NUM)=PL |
 (^ PRED)='DADO<(^ SUBJ)(^ OBJ2)(^ OBL)>'
 (^ OBL CASE)=c POR
 (^ OBJ2 CASE)=c A
 (^ SUBJ GEN)=FEM
 (^ SUBJ NUM)=PL }.

Contudo, a solução inicial, mostrada em (18), pensada por nós para dar conta da entrada dos adjetivos nas construções passivas, não se mostrou satisfatória, tendo em vista que nossa G-A seguia analisando frases como *O anel é dado a uma fada* com o traço de *PASSIV +*, e, ao lexema *dado*, era atribuída a categoria V. Essa análise era justamente a que não esperávamos, pois, feitas as mudanças e implementações em nossa G-A, esta, se suficientemente bem implementada, não deveria admitir a presença de construções passivas, senão unicamente construções predicativas adjetivais.

O problema jazia no fato de que *dado*, diferentemente das três outras formas

(*dados, dada, dadas*), também possuía uma entrada como categoria V em nosso léxico (em negrito em (28)). A classificação como V era, e continua sendo, necessária para que construções com o passado composto (e, g. *O cavaleiro tem dado um anel à fada.*) pudessem ser analisadas. A fim de desatar esse nó górdio, tentamos diversas saídas, mas nenhuma delas mostrou-se exitosa. A solução encontrada foi, então, trabalhar as formas masculinas do particípio *dado*, quando V, e *dado*, como A, em nosso arquivo de homônimos (*homonyms.lfg*). A entrada lexical, cujas formas encontram-se já simplificadas pelo uso de *templates* em nossa versão mais moderna da G-A, pode ser vista em (19):

(19) **dado V * @(OPT-DITRANS DAR)**
 (^ VFORM) = PART_PAST
 (^ TENSE) = COMPOUND_PAST
 (^ AUX) =c **TER;**

 A * (^ PRED)='DADO< (^ SUBJ)(^ OBL)(^ OBJ2)>'
 (^ OBL CASE)=POR
 @M
 @SG.

A equação restritiva em negrito em (19) é o que garante que o compilador saiba quando se trata de um verbo ou de um adjetivo, uma vez que a primeira análise exige a presença do verbo auxiliar *ter* na formação do passado composto. Se o auxiliar não estiver presente, *dado* é analisado como adjetivo, com função de XCOMP na estrutura F, e, por conseguinte, nossa frase não mais é classificada como passiva, mas sim como predicativa adjetival.

Dentre as outras mudanças que foram feitas na G-A, está a simplificação da entrada lexical no verbo *ser*, uma vez que esse não figura mais como auxiliar da passiva. Em (20), toda a parte em negrito da gramática G-V foi removida na versão G-A:

(20) é V * (^ PRED)='SER< (^ XCOMP)>(^ SUBJ)'
 (^ SUBJ)=(^ XCOMP SUBJ)
 @(CAT (^ XCOMP) AP)

@A-FORM-V ;

Aux * { (^ **CHECK PASS**) = +_ (^ **PASSIVE**) =c +
 (^ **TENSE**)=PRS |
 (^ **VFORM**) =c PART_PAST
 (^ **CHECK SER**) = +_
 (^ **UNACC**) = +
 (^ **TENSE**)=COMPOUND_PAST }

@(V-AGR 3 SG)

@IND.

Em (20), a análise das sentenças passivas como predicativas adjetivais está assegurada pela subcategorização de *é* que exige um XCOMP como complemento, o qual deve pertencer à classe dos adjetivos (@(CAT (^ XCOMP) AP))), e cujo sujeito é o mesmo da sentença, consequência da equação de controle na segunda linha da entrada lexical: (^ SUBJ)=(^ XCOMP SUBJ).

Além disso, os demais verbos, cujos lemas encontram-se listados no arquivo *lemma.lfg*, também tiveram as entradas lexicais simplificadas, uma vez que o *template* @PASSIVE, responsável por acionar a regra lexical de passivização, não desempenha mais nenhuma função numa gramática que não considera voz passiva. Das entradas dos lemas em (21), retiramos o que se referisse à passiva:

- (21) procurar V XLE ~~@(PASSIVE~~ @(TRANS PROCURAR)).
- encontrar V XLE ~~@(PASSIVE~~ @(TRANS ENCONTRAR)).
- anunciar V XLE { ~~@(PASSIVE~~ @(OPT-DITRANS ANUNCIAR)) |
 @(COMP-V ANUNCIAR IND DECLAR) |
 @(COMP-IO-V ANUNCIAR IND DECLAR)}.
- solicitar V XLE { @(DIRECTIVE SOLICITAR OBJ2 PARA) |
 @(CTRL-V-C SOLICITAR PARA) |
~~@(PASSIVE~~ @(DITRANS SOLICITAR)) | @(COMP-

V SOLICITAR SUBJUNCT DECLAR)|
 @(COMP-IO-V SOLICITAR IND INTERR) }.

encorajar V XLE ~~@(PASSIVE~~ @(DIRECTIVE ENCORAJAR OBJA)).

comprar V XLE ~~@(PASSIVE~~ @(OPT-DITRANS COMPRAR)).

Pelo mesmo motivo de não considerar haver sentenças passivas, retiramos de nossa G- A, o *template* @PASSIVE correspondente à regra da passiva (PASS_(SCHEMATA)) que foi apresentada em 5.1.

Esperamos neste capítulo ter apresentado os principais pontos que serviram de base para a adaptação/criação de nossas gramáticas. Por fim, gostaríamos de pontuar que, como nossa gramática encontra-se disponível para *download*, verificações de nossas hipóteses, bem como análises concorrentes ou complementares podem ser implementadas e mais facilmente comparadas.

6 ANÁLISE DOS RESULTADOS

Uma gramática LFG-XLE basicamente consiste de um conjunto de regras anotadas e de um léxico (ALENCAR, 2004). Em nosso trabalho, apresentamos duas versões da gramática: a G-A e a G-V, e, de acordo com uma de nossas hipóteses, a primeira tem melhor desempenho.

Butt *et al.* (1999, p.204) afirmam que a testagem de uma gramática é um importante passo para assegurar robustez e aumento da performance da mesma. Nessa esteira, a fim de verificarmos o desempenho de nossas gramáticas, fizemos a testagem, em ambas as gramáticas, de um mesmo conjunto-teste positivo e negativo.

Com base em Butt *et al.* (1999), Schwarze e Alencar (2016) e Alencar (2017), antes da análise sintática propriamente dita, há uma série de procedimentos que necessitam ser realizados. O passo inicial é o de fragmentar um dado texto em sentenças, e estas em *tokens*. O meio pelo qual a maioria das gramáticas LFG realizada esses pré-procedimentos é acoplado à gramática um transdutor de estados finitos (BREESLEY; KARTTUNEN, 2013). Como uma gramática LFG é reversível, ou seja, capaz tanto de tanto analisar como gerar sentenças, diferentes transdutores são utilizados:

Comumente, o fragmentador de sentenças é um transdutor determinístico que simplesmente divide um determinado texto de *input* em fragmentos que terminam em sequências de caracteres localizados no final da sentença de modo não-ambíguo. Tokenizadores e analisadores morfológicos, por seu turno, normalmente são transdutores não-determinísticos que, especialmente em combinação, podem ter como output um maior número de análises possíveis. (FORST, 2011, p. 8)⁵⁹

Ainda a respeito da combinação de transdutores de estados finitos com gramáticas LFG, Forst (id., ib.) afirma que tal junção não afeta a complexidade computacional da gramática, mas, pelo contrário, ajuda na modularidade e manutenção da mesma.

6.1 Testagem

Utilizando o XLE, realizamos a análise de nosso arquivo de teste positivo

⁵⁹ Nossa tradução deste trecho original: “Typically, the sentence breaker is a deterministic transducer that simply splits a given input text into fragments ending in character sequences that are unambiguously sentence-final. Tokenizers and morphological analyzers, on the other hand, are usually non-deterministic transducers that, especially in combination, can output a large number of possible analyses.”

*testfile.lfg*⁶⁰, por meio do comando *parse-testfile <testfile.lfg>*, nas duas versões de nossa gramática. Esse comando realiza a análise gramatical (*parsing*) de todas as sentenças contidas no arquivo de extensão *.lfg* especificado e, além disso, registra o número de análises, o tempo gasto para cada uma dessas análises e o número de *subtrees* geradas para cada sentença (BUTT *et al.*, 1999, p.204). Trata-se, assim, de informações sobre performance de nosso *parser*. A quantidade de *subtrees* permite que comparações sejam feitas à medida que mudanças nas regras sintagmáticas são implementadas: é possível comparar se uma determinada mudança nas regras sintagmáticas, por exemplo, corresponde a uma mudança na análise de uma sentença, cuja análise anterior já era considerada correta, verificando-se o número de *subtrees* geradas. Além disso, mudanças relevantes na quantidade de *subtrees* podem ser um indício de que algo esteja incorreto na implementação feita.

Após o comando *parse-testfile*, o XLE gera automaticamente um novo arquivo em formato *.new* (em nosso caso: *testfile.lfg.new*). Apresentamos esse arquivo em (24) e chamamos a atenção para as informações contidas entre parênteses após cada uma das sentenças, a saber, respectivamente, o número de análises realizadas, o tempo gasto (expresso em milissegundos) e a quantidade de *subtrees*:

(1)

Teste G-V:

#⁶¹ SENTENCE_ID: 001

A fada é esperada por um cavaleiro. (1 0.008 54)

SENTENCE_ID: 002

Ele é esperado por uma dama. (1 0.006 50)

SENTENCE_ID: 003

A fada é esperada. (1 0.004 37)

SENTENCE_ID: 004

O cavaleiro é esperado por uma dama. (1 0.007 56)

SENTENCE_ID: 005

O anel é dado a uma dama por uma fada. (1 0.006 55)

SENTENCE_ID: 006

⁶⁰ O arquivo *testfile.lfg* encontra-se para consulta no **Apêndice A** deste trabalho.

⁶¹ O símbolo '#' no início da linha indica que o que vem a seguir é um comentário, algumas de cujas funções é a de tornar mais organizado e legível o código para outros leitores ou implementadores. Na prática, a linha em que se encontra a cerquilha é ignorada pelo XLE.

Os anéis são dados a uma dama. (1 0.005 40)
 # SENTENCE_ID: 007
 O cavaleiro vê a floresta habitada por umas fadas. (1 0.005 54)
 # SENTENCE_ID: 008
 O cavaleiro vê o castelo branco habitado por uma rainha. (1 0.005 57)
 # SENTENCE_ID: 009
 O cavaleiro tem esperado a dama. (1 0.004 36)
 # SENTENCE_ID: 010
 A fada tem chegado. (1 0.003 28)
 # SENTENCE_ID: 011
 Os cavaleiros têm chegado. (1 0.002 28)
 # SENTENCE_ID: 012
 O cavaleiro foi esperado por uma fada. (1 0.005 56)
 # SENTENCE_ID: 013
 O cavaleiro tem sido esperado por uma fada. (1 0.005 55)
 # SENTENCE_ID: 014
 O cavaleiro tem sido forçado por uma rainha a acabar a tarefa. (1 0.005 64)
 # SENTENCE_ID: 015
 O cavaleiro tem sido forçado a acabar a tarefa por uma rainha. (1 0.005 74)
 # SENTENCE_ID: 016
 A dama quer ser convidada para dançar por um cavaleiro. (1 0.005 76)
16 sentences, 0.080 CPU secs total, 0.008 CPU secs max (08/03/18)

Como se pode verificar, nossa gramática, na versão em que o PART-PASS é implementado como V, é capaz de analisar todas as sentenças, atribuindo a cada uma delas uma única estrutura funcional válida. Vejamos o resultado obtido por nossa gramática G-A no XLE através do mesmo comando *parse-testfile <testfile.lfg>*:

(2)

TESTE G-A:

SENTENCE_ID: 001
 A fada é esperada por um cavaleiro. (1 0.006 30)
 # SENTENCE_ID: 002
 Ele é esperado por uma dama. (1 0.005 34)

SENTENCE_ID: 003

A fada é esperada. (1 0.003 15)

SENTENCE_ID: 004

O cavaleiro é esperado por uma dama. (1 0.005 37)

SENTENCE_ID: 005

O anel é dado a uma dama por uma fada. (1 0.004 39)

SENTENCE_ID: 006

Os anéis são dados a uma dama. (1 0.005 37)

SENTENCE_ID: 007

O cavaleiro vê a floresta habitada por umas fadas. (1 0.004 38)

SENTENCE_ID: 008

O cavaleiro vê o castelo branco habitado por uma rainha. (1 0.005 41)

SENTENCE_ID: 009

O cavaleiro tem esperado a dama. (1 0.003 21)

SENTENCE_ID: 010

A fada tem chegado. (1 0.002 15)

SENTENCE_ID: 011

Os cavaleiros têm chegado. (1 0.002 15)

SENTENCE_ID: 012

O cavaleiro foi esperado por uma fada. (1 0.004 37)

SENTENCE_ID: 013

O cavaleiro tem sido esperado por uma fada. (1 0.004 40)

SENTENCE_ID: 014

O cavaleiro tem sido forçado por uma rainha a acabar a tarefa. (1 0.004 46)

SENTENCE_ID: 015

O cavaleiro tem sido forçado a acabar a tarefa por uma rainha. (1 0.004 54)

SENTENCE_ID: 016

A dama quer ser convidada para dançar por um cavaleiro. (1 0.004 40)

16 sentences, 0.066 CPU secs total, 0.006 CPU secs max (08/03/18)

As duas versões da gramática analisam nossas 16 sentenças atribuindo para

cada uma delas uma única estrutura funcional. Verifiquemos, agora, a análise de nosso conjunto-teste negativo, lembrando que nossa gramática, caso bem implementada, não poderá analisar nenhuma das sequências agramaticais presentes em nosso arquivo *negative-test.lfg*. Os resultados para as gramáticas G-V e G-A através do comando *parse-testfile <negative-test.lfg>* estão apresentados abaixo, respectivamente, em (3) e (4):

(3)

```
((1) (0 0.003 24) (4 words))
((2) (0 0.004 42) (7 words))
((3) (0 0.003 41) (7 words))
....
((18) (0 0.003 40) (7 words))
((19) (0 0.003 41) (9 words))
((20) (0 0.003 38) (9 words))
20 sentences had 0 parses
```

(4)

```
((1) (0 0.004 22) (4 words))
((2) (0 0.008 42) (7 words))
((3) (0 0.006 30) (7 words))
....
((17) (0 0.007 44) (10 words))
((18) (0 0.005 28) (7 words))
((19) (0 0.007 31) (9 words))
((20) (0 0.006 31) (9 words))
20 sentences had 0 parses
```

De (3) e (4), verifica-se que nenhuma das sentenças de nosso conjunto-teste negativo é analisada. Esse fato indica que nossas minigramáticas, dentro do recorte metodológico com o qual estamos trabalhando, não hipergeram, além de demonstrar que as restrições por nós implementadas nas regras foram suficientes para analisar o excerto com qual estamos trabalhando.

Levando-se em conta que a análise de nossas gramáticas G-A e G-V é similar no tratamento de nossos conjuntos-teste positivo e negativo, ambas analisando todas as sentenças daquele e nenhuma deste, voltamo-nos para a análise do tempo necessário pela CPU para fazê-lo. As tabelas abaixo mostram os dados a que chegamos repetindo-se a análise 10 vezes seguidas, mantendo-se, ao mesmo tempo, o uso da CPU constante, com todos os demais programas do computador desligados. O número de 10

repetições, como já mencionado, foi escolhido para repetir o procedimento utilizado no projeto PIBIC supramencionado.

Tabela 1 – Desvio padrão médio de G-V

Sentenças analisadas por G-V	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6	Teste 7	Teste 8	Teste 9	Teste 10	MÉDIA	DESVIO PADRÃO	DESVIO PADRÃO RELATIVO(%)
Frase 1	0,011	0,004	0,016	0,012	0,012	0,006	0,008	0,012	0,011	0,008	0,01	0,0035	34,96
Frase 2	0,008	0,004	0,013	0,009	0,009	0,005	0,006	0,009	0,009	0,006	0,0078	0,0026	33,54
Frase 3	0,004	0,003	0,008	0,006	0,006	0,003	0,005	0,007	0,006	0,004	0,0052	0,0017	32,43
Frase 4	0,008	0,005	0,012	0,013	0,008	0,004	0,006	0,009	0,01	0,007	0,0082	0,0029	35,34
Frase 5	0,006	0,004	0,01	0,01	0,007	0,005	0,006	0,008	0,009	0,006	0,0071	0,0021	29,28
Frase 6	0,007	0,003	0,006	0,006	0,007	0,004	0,005	0,007	0,007	0,005	0,0057	0,0014	24,88
Frase 7	0,005	0,003	0,007	0,006	0,006	0,004	0,005	0,007	0,007	0,005	0,0055	0,0014	24,62
Frase 8	0,005	0,004	0,007	0,008	0,007	0,004	0,005	0,007	0,007	0,005	0,0059	0,0014	24,56
Frase 9	0,004	0,003	0,004	0,006	0,005	0,004	0,004	0,005	0,005	0,004	0,0044	0,0008	19,17
Frase 10	0,003	0,002	0,003	0,003	0,003	0,002	0,003	0,003	0,003	0,003	0,0028	0,0004	15,06
Frase 11	0,003	0,002	0,003	0,003	0,003	0,002	0,003	0,003	0,004	0,002	0,0028	0,0006	22,59
Frase 12	0,005	0,004	0,005	0,006	0,007	0,004	0,005	0,007	0,006	0,005	0,0054	0,0011	19,91
Frase 13	0,005	0,005	0,005	0,007	0,005	0,004	0,005	0,006	0,006	0,005	0,0053	0,0008	15,53
Frase 14	0,005	0,004	0,006	0,008	0,006	0,005	0,006	0,008	0,007	0,005	0,006	0,0013	22,22
Frase 15	0,005	0,004	0,004	0,007	0,008	0,004	0,006	0,007	0,006	0,005	0,0056	0,0014	25,53
Frase 16	0,006	0,004	0,004	0,004	0,006	0,004	0,005	0,005	0,005	0,005	0,0048	0,0008	16,43
											0,00578125	0,0024	42,13

Fonte: Elaborado pelo autor.

Tabela 2 – Desvio padrão relativo médio G-A

Sentenças analisadas por (G-A)	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6	Teste 7	Teste 8	Teste 9	Teste 10	MÉDIA	DESVIO PADRÃO	DESVIO PADRÃO RELATIVO(%)
Frase 1	0,009	0,012	0,012	0,012	0,009	0,008	0,01	0,004	0,009	0,006	0,0091	0,0026	29,05
Frase 2	0,005	0,009	0,007	0,009	0,007	0,005	0,008	0,003	0,006	0,005	0,0064	0,0020	30,55
Frase 3	0,003	0,005	0,004	0,006	0,004	0,003	0,004	0,002	0,004	0,003	0,0038	0,0011	29,88
Frase 4	0,005	0,006	0,007	0,008	0,008	0,005	0,008	0,004	0,006	0,005	0,0062	0,0015	23,80
Frase 5	0,004	0,005	0,006	0,007	0,006	0,004	0,005	0,003	0,005	0,004	0,0049	0,0012	24,43
Frase 6	0,006	0,006	0,007	0,007	0,006	0,005	0,006	0,003	0,006	0,005	0,0057	0,0012	20,34
Frase 7	0,005	0,006	0,006	0,005	0,006	0,003	0,007	0,003	0,006	0,004	0,0051	0,0014	26,87
Frase 8	0,005	0,006	0,006	0,006	0,006	0,004	0,006	0,003	0,006	0,005	0,0053	0,0011	19,99
Frase 9	0,003	0,004	0,004	0,004	0,004	0,003	0,004	0,002	0,004	0,003	0,0035	0,0007	20,20
Frase 10	0,003	0,003	0,003	0,003	0,003	0,002	0,003	0,002	0,003	0,002	0,0027	0,0005	17,89
Frase 11	0,003	0,003	0,003	0,004	0,003	0,002	0,003	0,002	0,003	0,002	0,0028	0,0006	22,59
Frase 12	0,004	0,006	0,005	0,005	0,005	0,003	0,005	0,003	0,005	0,004	0,0045	0,0010	21,60
Frase 13	0,004	0,004	0,005	0,006	0,005	0,004	0,006	0,003	0,005	0,004	0,0046	0,0010	21,00
Frase 14	0,004	0,005	0,005	0,006	0,007	0,004	0,005	0,003	0,005	0,004	0,0048	0,0011	23,65
Frase 15	0,004	0,006	0,005	0,006	0,005	0,004	0,005	0,003	0,005	0,004	0,0047	0,0009	20,18
Frase 16	0,004	0,004	0,003	0,005	0,005	0,003	0,005	0,003	0,004	0,004	0,004	0,0008	20,41
											0,00488125	0,0019	39,56

Fonte: Elaborado pelo autor.

No tocante à complexidade temporal de nossas gramáticas, os resultados foram os seguintes:

- A gramática G-V apresentou um tempo médio de processamento de 0,00578125s ((DPR⁶² ±42,13%);
- A gramática G-A apresentou um tempo médio de processamento de 0,00488125s (DPR ±39,56%).

Para controle estatístico dos testes, foi realizada a Análise de Variância (ANOVA) de dois fatores, apropriado para verificar se existe diferença significativa entre duas amostras, em nosso, caso o processamento por G-A e G-V. A Tabela 3 abaixo mostra o resultado do teste:

Tabela 3 – Análise de variância com fator duplo.

ANOVA						
<i>Fonte da variação</i>	<i>SQ</i>	<i>GI</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Amostra	6,48E-05	1	6,48E-05	27,9964	2,41E-07	3,87395
Colunas	0,000839	15	5,59E-05	24,1595	1,78E-42	1,701228
Interações	3,07E-05	15	2,05E-06	0,884248	0,582414	1,701228
Dentro	0,000667	288	2,31E-06			
Total	0,001601	319				

Fonte: elaborada pelo autor.

A partir dos dados da Tabela 3, pode-se inferir que os valores de probabilidade (p) foram bem menores que o valor crítico convencionado de 0,05 ($valor-P = 2,41 \times 10^{-7}$). Desse modo, deve-se aceitar que há diferença estatisticamente significativa entre as análises das sentenças pela gramática G-A em relação à G-V. Chega-se à mesma conclusão quando se comparam os valores de F e F críticos para as amostras: uma vez que $F > F$ crítico ($F = 27,996 > F$ crítico = 3,8739), a hipótese nula, a qual afirma que não há diferença nas médias de processamento das duas amostras, deve ser descartada, e, conseqüentemente, deve-se aceitar a hipótese de há diferença significativa na média de processamento.

A partir dos resultados e da análise estatística, a gramática G-A é, portanto, a que apresenta o menor tempo de processamento, sugerindo que a abordagem que trata o PART-PASS como adjetivo é mais adequada, por ser menos complexa computacionalmente.

⁶² Desvio padrão relativo médio.

Além disso, pode-se também comparar a quantidade de subtrees geradas pelas duas gramáticas para as sentenças de nosso teste positivo:

Tabela 4 – Quantidade de *subtrees*

Frases	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GA	30	34	15	37	39	37	38	41	21	15	15	37	40	46	54	40
GV	54	50	37	56	55	40	54	57	36	28	28	56	55	64	74	76

Fonte: Elaborada pelo autor

Vê-se claramente, pelos dados da Tabela 4, que a gramática G-A é a que gera a menor quantidade de *subtrees* para as mesmas sentenças. Como já mencionado, esse resultado indica que se trata de uma gramática de menor complexidade no que se refere à implementação das regras sintagmáticas, ou seja, uma gramática de maior robustez, visto que realiza as mesmas análises com um *design* mais elegante, recorrendo a menos regras. Outra leitura possível é a de que pode haver algo de equivocado com a implementação de nossas regras na G-V, uma vez que sentenças simples (e. g., sentença 3: *A fada é esperada*) apresentam um número relativamente alto de *subtrees* em comparação com a quantidade gerada pela G-A. Os valores da Tabela 4 podem, assim, servir de parâmetro para uma análise crítica de nossas gramáticas com o intuito de que melhorias possam ser feitas.

Outro aspecto a ser considerado em uma fase posterior diz respeito à complexidade das duas abordagens na dimensão espacial. O XLE gera informações acerca de complexidade espacial de nossa gramática. Essas informações são geradas automaticamente pelo XLE quando carregamos nossa arquivo de extensão *.lfg*. A esse respeito, as informações de que dispomos são as seguintes: A gramática G-V tem 10 regras sintagmáticas, 37 estados, 56 arcos e 64 disjuntos. Nossa G-A, por seu turno, tem 10 regras com 39 estados, 61 arcos e 73 disjuntos. Infelizmente, não foi possível, por questões de tempo, a realização de uma análise mais aprofundada desses valores. Tais dados são importantes e devem ser também levados em conta em trabalhos futuros, a fim de se verificar se também na dimensão espacial, a gramática G-A mostra-se mais eficiente computacionalmente do que a G-V.

7 CONCLUSÃO

A presente dissertação teve como objetivo geral investigar, à luz dos princípios da Gramática Léxico-Funcional e da Linguística Computacional (LC), qual a classificação mais apropriada para o particípio passivo (PART-PASS) em construções tradicionalmente analisadas como passivas analíticas. A escolha deste objetivo geral deu-se tanto do ponto de vista da linguística teórica, a partir das hipóteses defendidas por Perini (2010), como do ponto de vista da linguística aplicada, em particular, na área de LC, haja vista que não encontramos na literatura resposta à questão levantada de qual a classificação mais condizente para o particípio: verbo ou adjetivo,

Para tanto, apresentamos dois objetivos específicos: a) testar a hipótese de que o PART-PASS é um elemento de categoria A (adjetivo); b) testar a hipótese de que a gramática em que o PART-PASS é implementado como A, ou seja, a gramática G-A, é mais eficiente do que a G-V, aquela em que o PART-PASS é implementado como V (verbo).

Nossa hipótese básica foi a de que a classificação mais consistente (do ponto de vista linguístico-gerativo) e mais eficiente (do ponto de vista computacional) para o PART-PASS é a de pertencente à categoria A, ou seja, que as construções de voz passiva analítica em PB são, na verdade, formações predicativas adjetivais. Nessa esteira, a gramática G-A seria a que apresentaria o design mais eficiente.

Desenvolvemos nos moldes de Schwarze e Alencar (2016) e Alencar (2017) uma gramática LFG-XLE capaz de analisar, dentre outros fenômenos, construções com PART-PASS tanto no presente como no passado: a gramática G-V que contempla a análise mais canônica das construções passivas. Em seguida, adaptamos esta gramática para que o PART-PASS fosse analisado como A, ou seja, uma gramática (G-A) em que as sentenças passivas seriam, na verdade, analisadas como predicativas adjetivais. A única distinção entre as duas variantes da gramática dá-se unicamente no tocante à análise do PART-PASS.

Para testagem da eficiência das gramáticas, criamos um *corpus* positivo e um negativo. Nosso *corpus* positivo é formado de 16 sentenças, cada uma delas contendo aspectos importantes a serem considerados no tratamento da passiva. O procedimento de teste foi o de analisar, em cada uma das duas gramáticas, no XLE, nosso conjunto de teste positivo. A análise do arquivo de teste positivo foi repetida 10 vezes em para cada gramática, os resultados do tempo de processamento foram, então,

submetidos a análise estatística com o teste ANOVA com fator de variância dupla. Os resultados demonstram que a análise pela G-A mostrou-se a de melhor desempenho, por menos complexa e dispendiosa computacionalmente. Tanto o tempo médio de processamento quanto a quantidade de *subtrees* foram menores na G-A do que na G-V. Os resultados a que chegamos corroboram nossa hipótese básica.

Acreditamos ainda que este trabalho pode servir de parâmetro para que outras abordagens do mesmo fenômeno ou de outros fenômenos linguísticos, tanto do PB como de outras línguas, possam ser comparados dentro do âmbito da LC. E esperamos ter contribuído para a maior discussão, no Brasil, no âmbito da LFG, sobre a voz passiva.

REFERÊNCIAS

- ALENCAR, L. F. de. Uma implementação computacional de construções verbais perifrásticas em francês. **Revista Alfa**, São Paulo, v.61, n.2, p.351-380, 2017.
- _____. A passiva em português como construção predicativa adjetival: evidência morfológica e implementação computacional em LFG/XLE. **Estudos da Língua(gem)** (Online), v. 13, p. 35-57, 2015a.
- _____. Técnicas em softwares livres na linguística de corpus. 2015. 5f. Projeto (PIBIC). \ Universidade Federal do Ceará, Fortaleza. 2015b.
- _____. Modelação computacional de padrões variáveis de concordância em português. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 21, p. 43-110, 2013a.
- _____. BrGram: uma gramática computacional de um fragmento do português brasileiro no formalismo da LFG. In: **BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY 2013 STIL**, 9., 2013. Fortaleza. Proceedings. Fortaleza: Sociedade Brasileira de Computação, 2013b. p. 183-188.
- _____. Línguas formais, gramáticas e autômatos no processamento automático das palavras. In: ALENCAR, L. F. de; OTHERO, G.A. (Org.). **Abordagens computacionais da teoria da gramática**. 1 ed. Campinas: Mercado de Letras, 2012, p. 13-75.
- _____. Modelação computacional de padrões variáveis de concordância em português. **Revista de Estudos da Linguagem**, Belo Horizonte, n. 21, vol. 1, p. 43- 110, 2013.
- _____. Resenha de “Teoria X-barras: descrição do português e aplicação computacional”, de Gabriel de Ávila Othero. **Revista Virtual de Estudos da Linguagem – ReVEL**. Vol. 6, n. 10, março de 2008. [www.revel.inf.br].
- _____. Linguagem e inteligência artificial. In: MATTES, Marlene (Org.). **Linguagem. As expressões do múltiplo**. Fortaleza: Premium, 2006.
- _____. Complementos verbais oracionais - uma análise léxico-funcional. **Revista Língua(gem)**, Santa Maria (RS), v. 1, n.1, p. 173-218, 2004.
- ALMEIDA, Napoleão Mendes de. **Gramática metódica da língua portuguesa**. 29ª ed. São Paulo: Saraiva, 1980.
- ARNAULD, Antoine; LANCELOT, Claude. **Gramática de Port-Royal**. Trad. de Bruno Fregni Bassetto, Henrique Graciano Murachco. 2ª ed. São Paulo: Martins Fontes, 2001.
- BAGNO, Marcos. **Gramática pedagógica do português brasileiro**. São Paulo: Parábola, 2011. BECHARA, Evanildo. **Moderna gramática portuguesa**. Rio de Janeiro: Lucerna, 2005.

- BEESELEY, K. R. and Karttunen, L. **Finite State Morphology**. Stanford: CSLI, 2003.
- BORER, H.; WEXLER, K. **The maturation of syntax**. In: ROEPER, T.; WILLIAMS, E. (eds.) *Parameter Setting*, Massachusetts: D. Reidel, 1987.
- BRESNAN, Joan, *et al.* **Lexical-functional syntax**. Oxford: Wiley-Blackwell. 2nd ed., 2016.
- BRESNAN, Joan. **Lexical-functional syntax**. Oxford: Blackwell, 2001.
- BUTT, Miriam *et al.* **A grammar writer's cookbook**. Stanford: CSLI publications, 1999.
- CASTILHO, Ataliba Teixeira de. **Nova Gramática do Português Brasileiro**. São Paulo: Editora Contexto, 2010, 768p.
- CÂMARA JR., Joaquim Mattoso. **Estrutura da língua portuguesa**. Petrópolis: Vozes, 2004.
- CÂMARA JR., Joaquim Mattoso. **História e estrutura da língua portuguesa**. Rio de Janeiro: Padrão, 1979.
- CANÇADO, Márcia; AMARAL, Luana. **Introdução à semântica lexical: papéis temáticos, aspecto lexical e decomposição de predicados**. Petrópolis: Vozes, 2016.
- CARNIE, Andrew. **Syntax: a generative introduction**. Oxford: Blackwell, 2002.
- CARRASCO, Félix. Sobre el formante de la voz passiva en español. In: **Revista española de lingüística**. La Rioja: Universidad de La Rioja, 1973, p. 331-341.
- CAVALCANTE, Silvia Regina de Oliveira. Mudanças e estabilidade em construções “passivas” na história do português. **Revista Linguística (Online)**, v.32, p.41-60, 2016.
- CEGALLA, Domingos Paschoal. **Novíssima gramática da língua portuguesa**. 40ª ed. São Paulo: Ed. Nacional. 1967.
- CHOMSKY, Noam. **Linguística cartesiana – um capítulo da história do pensamento racionalista**. Petrópolis/São Paulo: Editora Vozes/Edusp, 1972.
- CHOMSKY, N. (1981). **Lectures on Government and Binding**. Dordrecht: Foris.
- COSERIU, Eugenio. **Sprachwissenschaft**. Vol. 12, caderno 1. Heilderberg: Carl Winter Universitätsverlag, 1987.
- CROUCH, D. et al. XLE Documentation. **Palo Alto**: Palo Alto Research Center, 2011. Disponível em: <http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html>. Último acesso: 15.06.2017.
- CUNHA, Celso; CINTRA, Luis F. Lindley. **Nova gramática do português contemporâneo**. 2ª. ed. Rio de Janeiro: Nova Fronteira, 1985a.

CUNHA, Celso; CINTRA, Luis F. **Nova gramática do português contemporâneo**. 11ª. ed. Lisboa: Edições João de Sá da Costa, 1985b.

CYRINO, Sonia. Construções com SE e promoção de argumento no português brasileiro: uma investigação diacrônica. **Revista da ABRALIN**, 2, p.85-116, 2007.

DALRYMPLE, M. Lexical-Functional Grammar. In: **Encyclopedia of Language & Linguistics**. 2. Elsevier, 2005.

D'AVILLA, Nerci Terezinha. **Voz verbal**: uma abordagem enunciativa. 2006. Tese (doutorado em estudos da linguagem). Universidade Federal do Rio Grande do Sul.

DI FELIPPO, A; DIAS-DA-SILVA, B. C. Os adjetivos valenciais do português e sua representação linguístico-computacional. **Revista do GEL**, v. 02. p. 55-81. Araraquara, São Paulo, Brasil, 2005.

GRUNE, D.; JACOBS, C.J.H. Parsing Techniques: A Practical Guide (Monographs in Computer Science). 2nd. ed. Nova Iorque, Springer, 2010.

FALK, Yehuda N. **Lexical-Functional Grammar**: an introduction to parallel constraint-based syntax. Stanford: CSLI Publications, 2001.

FORST, M. **Computational Aspects of Lexical Functional Grammar**. Language and Linguistics Compass, 5: 1-18, 2011. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-818X.2010.00261.x>>. Acesso em julho de 2018.

HAEGEMAN, L. **Introduction to Government and Binding Theory**. 2nd ed. New Jersey: Wiley- Blackwell, 1994.

INDURKHYA, Nitin; DAMERAU, Fred (Orgs.). **Handbook of Natural Language Processing**. 2ª ed. Cambridge: Series Editors, 2010.

KAPLAN, Ronald M.; BRESNAN, Joan. **The Mental Representation of Grammatical Relations** ed. Joan Bresnan: Cambridge, 1982.

KAPLAN, R. M.; BRESNAN, J. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In: DALRYMPLE, M. *et al.* (eds.). **Formal Issues in Lexical Functional Grammar**. Stanford (CA): Stanford University, 1995.

KATO & NASCIMENTO, Mary e Milton (orgs.). **Gramática do português culto falado no Brasil**: A construção da sentença. São Paulo: Contexto, 2015.

KELLING, C. Spanish se-constructions: the passive and the impersonal construction. In: BUTT, M.; HOLLOWAY, T. K. (eds.). **Proceedings of the LFG06 Conference (Konstanz)**. Stanford (CA): Stanford University: CSLI Publications, 275-288.

KROEGER, Paul R. **Analyzing Syntax**: A Lexical-functional Approach. Cambridge: Cambridge University Press, 2004.

LEMNITZER, L.; ZINSMEISTER, H. **Korpuslinguistik**: eine Einführung. Tübingen: Narr, 2006.

LIMA, Carlos Henrique da Rocha. **Gramática Normativa da Língua Portuguesa**. 33^a ed. Rio de Janeiro: José Olympio, 1996.

LINGUINHO, Marcus Vinicius da Silva. **Verbos auxiliares e a sintaxe dos domínios não-finitos**. Tese de Doutorado. Faculdade de Filosofia Letras e Ciências Humanas. Universidade de São Paulo. 2011.

LUFT, Celso Pedro. **Moderna gramática brasileira**. 2^a. ed., ver. e atual. São Paulo: Globo, 2002.

MATEUS, Maria Helena Mira *et al.* **Gramática da Língua Portuguesa**. 4^a. ed., Lisboa: Caminho. 1994.

MIOTO, Carlos; FIGUEIREDO, Maria Cristina; LOPES, Ruth. **Novo manual de sintaxe**. Florianópolis: Insular, 2005.

MORAIS, M. A. T. **A passivização no português**: uma abordagem léxico-funcional. Dissertação (Mestrado em Linguística). Universidade Federal do Paraná. Curitiba, p.181. 1988.

MÜLLER, Stefan. **Gramatical theory: From transformational grammar to constraint-based approaches**. (Textbooks in Language Sciences 1). Berlin: Language Science Press, 2016.

MÜLLER, S. **Grammatiktheorie**. Tübingen: Stauffenburg, 2010.

NEVES, Maria Helena de Moura. **Gramática de usos do português**. São Paulo: Ed. UNESP, 2000.

OTHERO, G. A. ; DAVID, K. A. . Agrupamentos nominais no português brasileiro: uma análise formal e sua aplicação computacional. In: OTHERO, G. A.; ALENCAR, L.. (Org.). **Abordagens Computacionais da Teoria da Gramática**. 1ed.Campinas: Mercado de Letras, 2012, v. , p. 77-98.

OTHERO, Gabriel. **Teoria X-barra**: descrição do português e aplicação computacional. São Paulo: Contexto, 2006.

OTHERO, Gabriel; MENUZZI, Sérgio. **Linguística Computacional**. São Paulo: Parábola, 2005.

PARDO, T. A. S. **Estudo comparativo dos formalismos gramaticais DCG e LFG**. São Carlos, 1999 (Relatório Técnico).

PARGRAM / ParSem: **An international collaboration on LFG-based grammar and semantics development**. [S.l.]: [s.n.], 2012. Disponível em: <<http://pargram.b.uib.no/>>. Acesso em: 11 jul.2016.

PATEJUK, A. **POLFIE – an LFG grammar of Polish**. 2013. Disponível em: <<http://zil.ipipan.waw.pl/LFG>>. Acesso em: 02.nov.2017.

PATEJUK, Agnieszka; PRZEPIÓRKOWSKI, Adam. In favour of the raising analysis of passivisation. In: **Lexical Functional Grammar Conference**, 2014. Ann Arbor. Abstracts. Disponível em: <<http://lfg-conference.org/program/>>. Acesso em: 06 jul.2016.

PERINI, Mário Alberto. **Gramática descritiva do português brasileiro**. Petrópolis, RJ: Vozes, 2016.

_____. **Gramática do Português Brasileiro**. São Paulo: Parábola, 2010.

_____. **Princípios de linguística descritiva: introdução ao pensamento gramatical**. São Paulo: Parábola Editorial, 2006.

PITTNER, K; BERNAN, J. **Deutsche Syntax**. 2ed., Tübingen: Narr Studienbücher, 2007.

RAPOSO, Eduardo Paiva. **Teoria da gramática: a faculdade da linguagem**. Lisboa: Caminho, 1992.

SAG, Ivan A.; WASOW, Thomas; BENDER, Emily M.. **Syntactic Theory: A Formal Introduction**. Stanford, CA: CSLI Publications, 2003.

SANTOS, Andréa Feitosa dos. **Uma Gramática LFG-XLE para o Processamento Sintático Profundo da Frase do Português Brasileiro**. 2014. Tese (Doutorado em Linguística) - Universidade Federal do Ceará.

SCHWARZE, C.; ALENCAR, Leonel F. de. **Lexikalisch-funktionale Grammatik. Eine Einführung am Beispiel des Französischen mit computerlinguistischer Implementierung**, vol 1, 1ª ed. Tübingen: Stauffenburg, 2016.

SIMIONI, Leonor. The relationship among the subject positions, focus, and agreement in passive constructions in Brazilian Portuguese. **Estudos da Língua(gem)** 8, p.173-187,2010. Disponível em: <<http://estudosdalinguagem.org/index.php/estudosdalinguagem/article/viewFile/192/273>>. Acesso em: 25 jun.2016.

TARALLO, Fernando; KATO, Mary. A. **Harmonia trans-sistêmica: variação intra- e inter- lingüística**. Preedição 5. Campinas, SP: Unicamp, 1989.

TEIXEIRA, M. T.; OTHERO, G. A. **Aquisição de sentenças passivas: uma retrospectiva teórico- experimental**. Fórum Linguístico, v. 15, p. 3241-3258, 2018.

WANDRUSZKA, Mario. **Sprachen vergleichbar und unvergleichlich**. Munique: R. Piper & Co. Verlag, 1969.

WESCOAT, Michael. **Practical Introduction for working with the formalism of lexical functional grammar**. Xerox Palo Alto Research Center, 1989.

APÊNDICE A - Arquivo de teste positivo *testfile.lfg*

SENTENCE_ID: 001

A fada é esperada por um cavaleiro. # SENTENCE_ID: 002

Ele é esperado por uma dama. # SENTENCE_ID: 003

A fada é esperada.

SENTENCE_ID: 004

O cavaleiro é esperado. # SENTENCE_ID: 005

O anel é dado a uma dama por uma fada. # SENTENCE_ID: 006

Os anéis são dados a uma dama. # SENTENCE_ID: 007

O cavaleiro vê a floresta habitada por umas fadas. # SENTENCE_ID: 008

O cavaleiro vê o castelo branco habitado por uma rainha. # SENTENCE_ID: 009

O cavaleiro tem esperado. # SENTENCE_ID: 010

A fada tem chegado.

SENTENCE_ID: 011

As fadas têm chegado. # SENTENCE_ID: 012

O cavaleiro foi esperado por uma fada.

SENTENCE_ID: 013

O cavaleiro tem sido esperado por uma fada.

SENTENCE_ID: 014

O cavaleiro tem sido forçado por uma rainha a acabar a tarefa.

SENTENCE_ID: 015

O cavaleiro tem sido forçado a acabar a tarefa por uma rainha.

SENTENCE_ID: 016

A dama quer ser convidada para dançar por um cavaleiro.

APÊNDICE B - Arquivo de teste negativo *negative-test.lfg*

SENTENCE_ID: 001

A fada é esperado.

SENTENCE_ID: 002

A fada é esperado por um cavaleiro.

SENTENCE_ID: 003

O cavaleiro é esperada por uma dama.

SENTENCE_ID: 004

A fada é espera.

SENTENCE_ID: 005

A fada quer esperada.

SENTENCE_ID: 006

A fada começa a esperada.

SENTENCE_ID: 007

A fada é esperar.

SENTENCE_ID: 008

A fada esperada.

SENTENCE_ID: 009

A fada esperado.

SENTENCE_ID: 010

A fada tem espera

SENTENCE_ID: 011

A fada tem espera.

SENTENCE_ID: 012

A fada tem esperar.

*# SENTENCE_ID: 013
A fada tem chegada.*

*# SENTENCE_ID: 014
A fada é chegada.*

*# SENTENCE_ID: 015
As fadas são chegadas.*

*# SENTENCE_ID: 016
O cavaleiro é esperada.*

*# SENTENCE_ID: 017
O anel é dada a uma dama por uma fada.*

*# SENTENCE_ID: 018
Os anéis são dada a uma dama.*

*# SENTENCE_ID: 019
O cavaleiro vê a floresta habitado por umas fadas.*

*# SENTENCE_ID: 020
O cavaleiro vê a floresta habitadas por umas fadas.*