



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

ANDRÉA FEITOSA DOS SANTOS

**UMA GRAMÁTICA LGF-XLE PARA A ANÁLISE SINTÁTICA PROFUNDA DO
PORTUGUÊS**

**FORTALEZA
2014**

ANDRÉA FEITOSA DOS SANTOS

**UMA GRAMÁTICA LFG-XLE PARA A ANÁLISE SINTÁTICA PROFUNDA DO
PORTUGUÊS**

Tese apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutora em Linguística. Área de concentração: Linguística

Orientador: Prof. Dr. Leonel F. de Alencar.
Co-orientadores: Prof. Dr. Georg Kaiser.
Profa. Dra. Miriam Butt.

FORTALEZA

2014

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca de Ciências Humanas

-
- S233g Santos, Andréa Feitosa dos.
Uma gramática LFG-XLE para a análise sintática profunda do português / Andréa Feitosa dos Santos. – 2014.
178 f. : il. color., enc. ; 30 cm.
Tese(doutorado) – Universidade Federal do Ceará, Centro de Humanidades, Departamento de Letras Vernáculas, Programa de Pós-Graduação em Linguística, Fortaleza, 2014.
Área de Concentração: Linguística.
Orientação: Prof. Dr. Leonel Figueiredo de Alencar Araripe.
Coorientação: Prof. Dr. Georg Kaiser.
Profa. Dra. Miriam Butt.
1.Língua portuguesa – Brasil – Sintaxe. 2.Língua portuguesa – Brasil – Palavras e expressões.
3.Linguística – Processamento de dados. 4.Gramática léxico-funcional. I.Título.

CDD 469.5

ANDRÉA FEITOSA DOS SANTOS

**UMA GRAMÁTICA LFG-XLE PARA A ANÁLISE SINTÁTICA PROFUNDA DO
PORTUGUÊS**

Tese apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutora em Linguística. Área de concentração: Linguística

Aprovada em: 19/12/2014.

BANCA EXAMINADORA

Prof. Dr. Leonel Figueiredo de Alencar Araripe (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dra. Vládia Celia Monteiro Pinheiro
Universidade de Fortaleza (Unifor)

Prof. Dr. Francisco A. Tavares F. Silva
Instituto Nacional de Pesquisas Espaciais (INPE)

Prof. Dra. Marias Elias Soares
Universidade Estadual do Ceará (UFC)

Profa. Dra. Márcia Teixeira Nogueira
Universidade Federal do Ceará (UFC)

Ao meu marido, Juliano, com muito amor.

AGRADECIMENTOS

Esta tese não teria sido possível sem o suporte, entusiasmo e encorajamento de muitas pessoas além de mim.

Primeiro de tudo agradeço à Universidade Federal do Ceará e ao Programa de Pós-Graduação em Linguística pelo espaço, o tempo e todo a organização envolvida para que esta pesquisa se desenvolvesse.

Quero agradecer ao Prof. Leonel F. de Alencar. Ele me encorajou a tratar de um tema intrigante e cheio de desafios. Apesar disso, tive o prazer de fazer o que fiz. Tive a permissão para trilhar um caminho que me fascina, *gramáticas*. Ele merece todo o reconhecimento por aquilo que possa ser valioso em meu trabalho, mas nenhuma responsabilidade pelas suas inadequações.

Agradeço aos meus co-orientadores Georg Kaiser e Miriam Butt, por me aceitarem como membro pesquisador em seus vibrantes ambientes de pesquisa e por me proporcionaram pensar “fora da caixa”.

Meus agradecimentos à Banca Examinadora pelo tempo dedicado à leitura desta tese, pelas considerações e pela avaliação final. A versão final desta tese, sem sombra de dúvida, ficou bem melhor graças às suas sugestões.

Uma das razões pelos ótimos momentos que tive trabalhando em linguística nos últimos anos foram os meus colegas Tiago Cunha, Gezenira Rodrigues, Elizângela Teixeira, Melanie Seiss, Sebastian Sulger, Annette Hautli e Stefano Quaglia, que gentilmente me auxiliaram e ajudaram quando deles precisei.

Outra grande razão para agradecimento cabe ainda ao Grupo de Redes de Computadores, Engenharia de Software e Sistema – GREat, pela oportunidade, a mim concedida, de colaborar com suas pesquisas e, em especial, agradeço à Lara Abreu, Mardônio França, Katiúscia Moraes, Wellington Franco, Hélio Silva e Rute Castro pelos valiosos comentários na ocasião da prévia da defesa.

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

Quero demonstrar minha mais profunda gratidão ao meu marido, Juliano, que me deu suporte constante e fez enormes concessões para que esse projeto se concretizasse. Externou minha gratidão às pessoas provavelmente mais afetadas por esta tese, Helena e Aurora, minhas filhas, que inevitavelmente passaram por momentos estressantes e turbulentos.

Por fim, agradeço à CAPES e à FUNCAP por todo o apoio financeiro que me foi concedido com a manutenção da bolsa e, particularmente, agradeço ao Serviço Alemão de Intercâmbio Acadêmico - DAAD, por ser um verdadeiro *Dad*.

A gramática de hoje não é pior que a gramática de ontem, só é mais prática, como o mundo em que vivemos.

RESUMO

A presente tese descreve a elaboração de uma gramática da frase do Português Brasileiro, desenvolvida no quadro de um modelo teórico de sofisticado formalismo computacional, a *Lexical Functional Grammar* (LFG) e implementada no sistema que constitui o estado da arte em ambiente de processamento sintático profundo no modelo gerativo da LFG, o robusto *Xerox Linguistic Environment* (XLE). A principal característica da gramática é que adota o sistema de anotação do ParGram e a metodologia convencionada por desenvolvedores de gramática XLE. No fragmento de gramática estão modelados diversificados elementos da sintaxe frasal. Em nossa gramática, foram modelados constituintes oracionais como IP e CP, elementos que encabeçam as sentenças do português. Também foram modelados determinados aspectos da subcategorização verbal e da estrutura argumental. Dos elementos verbais, nossa gramática contempla alguns casos de complexos verbais constituídos de verbos modais e verbos de controle. Os elementos nominais tratados na gramática, de modo central, foram os pronomes expletivos e reflexivos, e os casos de sintagmas nominais e determinantes com pronomes demonstrativos e interrogativos. Os demais aspectos modelados na gramática são os sintagmas preposicionados, cuja complexidade se dá na distinção entre preposições semânticas e não semânticas; os sintagmas adjetivais, cuja projeção na sentença pode ocorrer a partir de formas adjetivais atributivas, de formas ordinais ou cardinais e na forma de intensificadores; e os sintagmas adverbiais, cuja estrutura interna foi modelada levando-se em consideração tanto advérbios intransitivos quanto transitivos com complemento PP. A nossa avaliação demonstra que das 40 sentenças testadas, a nossa gramática atribui, para todas elas, análises consistentes e bem fundamentadas, ao passo que o *parser* Palavras, o atual estado da arte em processamento sintático profundo do português, atribui, a 9 sentenças, análises incorretas. Uma outra avaliação demonstra que, das 20 sentenças agramaticais testadas tanto em nossa gramática, quanto no Palavras, somente 2 receberam análises por parte de nossa gramática, enquanto o Palavras fornece análises para 19 sentenças. O trabalho tem, essencialmente, o objetivo de fazer uma descrição formal e fundamentada de um amplo leque de fenômenos do português brasileiro, mas, sobretudo, tem o objetivo de contribuir com uma gramática não trivial da frase do português no formalismo LFG-XLE, disponibilizando efetivamente um recurso gramatical do português voltado para o processamento de linguagem natural.

Palavras-chave: Gramática LFG-XLE. Análise sintática profunda. Linguística computacional. Processamento de linguagem natural. ParGram.

ABSTRACT

The present thesis describes the development of a Brazilian Portuguese sentence grammar, developed in the framework of a sophisticated computational formalism, named Lexical Functional Grammar, and implemented on a system that is state of the art in deep parsing environment in LFG generative model, the robust XLE. The main feature of the grammar is that it adopts the ParGram annotation system and the methodology agreed by XLE grammar developers. In the grammar fragment are modeled diverse elements of phrasal syntax. In our grammar were modeled constituents as IP and CP, elements that are head the sentences of the Portuguese. Also were modeled certain aspects of verbal subcategorization and argument structure. In terms of verbal elements, our grammar includes some cases of verbal complex made up of modal verbs and control verbs. The nominal elements treated in grammar, centrally, were the expletives and reflexive pronouns, and cases of nominal and determiners phrases with demonstrative pronouns and interrogative. The other aspects modeled in the grammar are PPs, whose complexity is given the distinction between semantic and nonstandard prepositions; the adjectival phrases, whose projection in the sentence can occur from attributive adjectival forms of ordinal or cardinal forms and as intensifiers; and adverbial phrases, whose internal structure was modeled taking into account both adverbs as intransitive and as transitive, with PP complement. Our evaluation shows that of the 40 tested sentences, our grammar assigns, for all of them, consistent and well-founded analysis, while the parser Palavras, the current state of the art in deep syntactic processing of Portuguese, assigns incorrect analysis for 9 sentences. Another evaluation shows that, of the 20 ungrammatical sentences tested both in our grammar, as in Palavras, only 2 received analysis by our grammar, while the Palavras provides analysis to 19 sentences. The work has essentially the goal of making a formal and grounded description in a broad range of phenomena in Brazilian Portuguese, but mainly aims to collaborate with a not trivial grammar of the sentence in the LFG-XLE formalism, effectively contributing to a grammatical resource turned to the natural language processing.

Keywords: LFG-XLE Grammar. Análise sintática profunda. ParGram. Computational linguistics. Natural Language Processing.

LISTA DE TABELAS

Tabela 1 – Estado da arte em processamento sintático profundo do português.....	21
Tabela 2 – Avaliação das sentenças gramaticais do <i>testfile</i> (Apêndice B).....	133
Tabela 3 – Avaliação das sentenças gramaticais do <i>testfile</i> (Apêndice B).....	134

LISTA DE FIGURAS

Figura 1 – Recursos e ferramentas de NLP	19
Figura 2 – Análise de uma sentença interrogativa pelo <i>parser</i> Palavras.....	22
Figura 3 – Análise de uma sentença com controle funcional pelo <i>parser</i> Palavras	23
Figura 4 – Análise de uma sentença com restrição Indicativo-subjuntivo pelo Palavras.....	24
Figura 5 – Estrutura de constituintes da sentença <i>a fada espera a rainha</i>	31
Figura 6 – Sistemas de correspondência entre estruturas	46
Figura 7 – Um exemplo de estrutura de constituintes representada no XLE	52
Figura 8 – Um exemplo de estrutura funcional representada no XLE	52
Figura 9 – Arquitetura do <i>parser</i>	54
Figura 10 – Plataforma do Aquamacs Emacs.....	55
Figura 11 – Plataforma do XLE invocada através do modo LFG do <i>Aquamacs</i>	56
Figura 12 – Informações reportadas pelo XLE após análise de uma sentença.....	57
Figura 13 – Traços do ParGram para o desenvolvimento de gramática no XLE.....	59
Figura 14 – Estrutura-c de uma oração declarativa	64
Figura 15 – Estrutura-f de uma oração declarativa.....	64
Figura 16 – Estrutura de constituintes de uma oração interrogativa deslocada.....	67
Figura 17 – Estrutura-f de uma oração interrogativa deslocada.....	67
Figura 18 – Estrutura-f de uma oração interrogativa adverbial.....	68
Figura 19 – Estrutura-c de uma sentença interrogativa adverbial	69
Figura 20 – Estrutura-c de uma sentença interrogativa <i>in situ</i>	71
Figura 21 – Estrutura-f de uma sentença interrogativa <i>in situ</i>	71
Figura 22 – Estrutura-f de uma sentença interrogativa adverbial <i>in situ</i>	72
Figura 23 – Estrutura-c de uma sentença interrogativa deslocada encaixada	73
Figura 24 – Estrutura-c de uma sentença com complemento oracional infinitivo	74
Figura 25 – Estrutura-f com complemento oracional gerundivo.....	75
Figura 26 – Estrutura-f de uma sentença com complemento oracional infinitivo.....	76
Figura 27 – Estrutura-f de uma sentença com complemento oracional gerundivo	76
Figura 28 – Estrutura-f de oração subordinada complementadora	77
Figura 29 – Estrutura-c de sentença com verbo intransitivo	79
Figura 30 – Estrutura-c para sentença com sujeito pós-verbal	81
Figura 31 – Estrutura-f de sentença com verbo bitransitivo.....	83
Figura 32 – Estrutura-f de sentença com adjunto	84

Figura 33 – Estrutura-f de sentença com Predlink	87
Figura 34 – Estrutura-f de sentença com XCOMP-PRED	89
Figura 35 – Estrutura-c de sentença com verbo auxiliar mais gerúndio	90
Figura 36 – Estrutura-c de verbo auxiliar mais infinitivo	90
Figura 37 – Estrutura-c de verbo auxiliar mais particípio	91
Figura 38 – Estrutura-f de sentença com verbo auxiliar e verbo principal.....	92
Figura 39 – Estrutura-c de uma sentença com apassivação do verbo	94
Figura 40 – Estrutura-f de sentença em forma apassivada	95
Figura 41 – Estrutura-c de uma sentença com complemento oracional aberto no infinitivo ...	97
Figura 42 – Estrutura-f de uma sentença com predicado complexo.....	99
Figura 43 – Estrutura-c de uma sentença com predicado complexo	100
Figura 44 – Estrutura-f de sentença com controle funcional de sujeito	101
Figura 45 – Estrutura-f de sentença com controle funcional de objeto	102
Figura 46 – Estrutura-c de pronome expletivo	103
Figura 47 – Estrutura-f de pronome expletivo.....	104
Figura 48 – Estrutura-f de pronome demonstrativo.....	105
Figura 49 – Estrutura-f de pronome reflexivo	106
Figura 50 – Estrutura-c de pronome reflexivo.....	107
Figura 51 – Estrutura-c de sentença interrogativa	108
Figura 52 – Estrutura-f de sentença interrogativo	109
Figura 53 – Estrutura-c com uma preposição semântica	111
Figura 54 – Estrutura-f de uma preposição semântica	112
Figura 55 – Estrutura-f de uma preposição não semântica.....	113
Figura 56 – Estrutura-c de adjetivo pré-nominal.....	115
Figura 57 – Estrutura-c de adjetivo pós-nominal	116
Figura 58 – Estrutura-f de adjetivo modificador de sintagma nominal	117
Figura 59 – Estrutura-c de sintagma nominal modificado por ordinal.....	119
Figura 60 – Estrutura-f de sintagma nominal modificado por ordinal	120
Figura 61 – Estrutura-f de sintagma nominal modificado por cardinal.....	121
Figura 62 – Estrutura-c de sintagma adjetival com modificação adverbial	122
Figura 63 – Estrutura-f de sintagma adjetival com modificação adverbial	123
Figura 64 – Estrutura-c de sintagma verbal com modificação adverbial	124
Figura 65 – Estrutura f de sintagma verbal com modificação adverbial	125
Figura 66 – Estrutura-c de advérbio transitivo	126

Figura 67 – Estrutura-f de advérbio transitivo.....	127
Figura 68 – Estrutura-c para uma construção com verbo auxiliar e verbo principal gerada pelo XLE.	131
Figura 69 – Análise do Palavras para uma construção com verbo auxiliar e verbo principal	132
Figura 70 – Estrutura-c, gerada pelo XLE, para uma sentença interrogativa.....	133
Figura 71 – Estrutura-f, gerada pelo XLE, para uma sentença interrogativa	133
Figura 72 – Análise do Palavras uma sentença interrogativa	134
Figura 73 – Análise do Palavras para uma sentença interrogativa	134
Figura 74 – Arquitetura do <i>parser</i>	140

LISTA DE ABREVIATURAS E SIGLAS

AP	Sintagma adjetival
IP	Sintagma de flexão
CP	Sintagma complementador
PP	Sintagma preposicional
DP	Sintagma determinante
NP	Sintagma nominal
AdvP	Sintagma adverbial
VP	Sintagma verbal
PossP	Sintagma possessivo
NumP	Sintagma numeral
CL	Clítico
D	Determinante
C	Complementador
S	Sentença
N	Substantivo
P	Preposição
V	Verbo
CASE	Caso
SPEC DET	Determinante especificador
SPEC NUMBER	Numeral especificador
SPEC POSS	Possessivo especificador
CONJ-FORM	Forma da conjunção
VFORM	Forma verbal
VTYPE	Tipo verbal
GEND	Gênero
NUM	Numero
PERS	Pessoa
PRED	Predicado
CLAUSE-TYPE	Tipo oracional
PRON-TYPE	Tipo pronominal

ATYPE	Tipo adjetival
PTYPE	Tipo preposicional
PSEM	Preposição semântica
PFORM	Forma preposicional
ADV-TYPE	Tipo adverbial
DET-TYPE	Tipo de determinante
NUMBER-TYPE	Tipo de numeral
NTYPE NSYN	Tipo sintático do nome
NTYPE NSEM	Característica semântica do nome
COMP-FORM	Forma do complementador
PRON-FORM	Forma pronominal
MOOD	Modo verbal
3	Terceira pessoa
loc	Locativo
expl_	Expletivo
refl	Reflexivo
demon	Demonstrativo
def	Definido
indef	Indefinido
nosem	Não semântica
sem	Semântico
acc	Acusativo
obl	Oblíquo
dat	Dativo
fem	Feminino
ind	Indicativo
inf	Infinitivo
fin	Finito
masc	Masculino
nom	Nominativo
past	Passado
pres	Presente
pl	Plural

sg	Singular
decl	Declarativo
int	Interrogativo
vpmod	Modificador verbal
adjmod	Modificador adjetival
nom	Nominalizada
SUBJ	Sujeito
OBJ	Objeto direto
OBJ-TH	Objeto secundário tema
OBL	Oblíquo
OBL-AG	Oblíquo agentivo
COMP	Complemento oracional aberto
XCOMP	Complemento oracional fechado
PREDLINK	Predicativo do sujeito
ADJUNCT	Adjunto
FOCUS	Foco
V3SG	Forma verbal na terceira pessoa do singular
V3PL	Forma verbal na terceira pessoa do plural
PRESIND	Forma verbal no presente do indicativo
PASTIND	Forma verbal no passado do indicativo
FUTIND	Forma verbal no futuro do indicativo
PRESSUBJ	Forma verbal no futuro do subjuntivo
TENSE(T)	Tempo verbal
MOOD(M)	Modo verbal
BITRANS(P)	Verbo bitransitivo
OPT-TRANS(P)	Transitividade opcional
INTRANS(P)	Verbo intransitivo
TRANS(P)	Verbo transitivo
TRANSIN(P)	Verbo transitivo indireto
BITRANS-OBL(P)	Verbo bitransitivo oblíquo
PASS(_SCHEMATA)	Esquema da passiva
OT-MARK(_mark)	Marca de otimalidade
LFG	Lexical Functional Grammar

XLE	Xerox Linguistic Environment
HPSG	Head-driven Phrase Structure Grammar
ParGram	Parallel Grammar Project
PB	Português brasileiro
PE	Português europeu

SUMÁRIO

1 INTRODUÇÃO	19
2 PRESSUPOSTOS TEÓRICOS	28
2.1 Formalismo sintático: Teoria X-barra	28
2.2 Constituintes sintagmáticos	32
2.2.1 <i>Categorias funcionais</i>.....	32
2.2.2 <i>Categorias lexicais</i>.....	37
2.3 Gramática Léxico-Funcional	44
2.3.1 <i>Algumas considerações teóricas</i>.....	45
2.3.2 <i>Níveis de representação estrutura-c e estrutura-f</i>.....	46
3 O AMBIENTE DE DESENVOLVIMENTO LFG-XLE E O PROJETO PARGRAM	51
3.1 O formalismo LFG-XLE.....	51
3.2 O projeto ParGram	57
4 PROCESSAMENTO SINTÁTICO PROFUNDO DO PB.....	62
4.1 Elementos oracionais	62
4.1.1 <i>Oração principal</i>	63
4.1.1.1 <i>Declarativa</i>	63
4.1.1.2 <i>Interrogativa</i>	65
4.1.1.2.1 <i>Deslocada</i>.....	66
4.1.1.2.2 <i>In situ</i>	70
4.1.1.2.3 <i>Encaixada</i>.....	72
4.1.2 <i>Oração subordinada</i>	73
4.1.2.1 <i>IPs como complementos verbais</i>.....	74
4.1.2.2 <i>CPs como complementos verbais</i>	77
4.2 Elementos verbais	78
4.2.1 <i>Funções gramaticais</i>.....	78
4.2.1.1 <i>Sujeito</i>	79
4.2.1.2 <i>Objeto</i>	81
4.2.1.3 <i>Objeto secundário</i>.....	82
4.2.1.4 <i>Oblíquo</i>	83
4.2.1.5 <i>Adjuntos</i>	84
4.2.1.6 <i>XCOMP e COMP</i>	84
4.2.2 <i>Predicativos</i>.....	86

4.2.2.1 PREDLINK	87
4.2.2.2 Análise do sujeito controlado – XCOMP-PRED.....	88
4.2.3 Auxiliares.....	89
4.2.4 Passivas	92
4.2.5 Predicados complexos.....	96
4.2.5.1 Verbos modais	96
4.2.5.2 Verbos de controle.....	100
4.3 Elementos nominais.....	103
4.3.1 Pronomes expletivos	103
4.3.2 Pronomes demonstrativos.....	104
4.3.3 Pronomes reflexivos.....	105
4.3.4 Pronomes interrogativos.....	107
4.4 Sintagmas preposicionados.....	109
4.4.1 Preposições semânticas	110
4.4.2 Preposições não semânticas	112
4.5 Síntagma adjetival	114
4.5.1 Adjetivos atributivos.....	114
4.5.2 Cardinais e ordinais.....	117
4.5.3 Adjetivos com intensificadores.....	121
4.6 Síntagma adverbial.....	123
4.6.1 Advérbios intransitivos	124
4.6.2 Advérbios transitivos.....	125
5 RESULTADOS OBTIDOS E DISCUSSÃO	129
5.1 Aspectos da gramática.....	129
5.2 Avaliação da gramática.....	130
5.3 Trabalhos futuros	137
5.3.1 Problemas de ambiguidade na gramática.....	138
5.3.2 Algumas considerações sobre a pipeline do XLE.....	139
6 CONSIDERAÇÕES FINAIS.....	142
REFERÊNCIAS.....	145
APÊNDICE A – UM FRAGMENTO DE GRAMÁTICA DA FRASE DO PORTUGUÊS BRASILEIRO.....	150

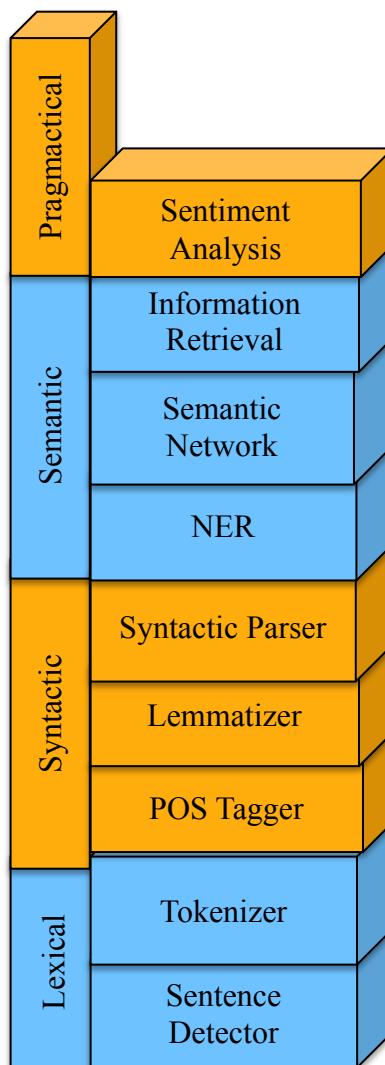
APÊNDICE B – ARQUIVO DE TESTE “TESTFILE.LFG.NEW” 173

1 INTRODUÇÃO

Muitas ferramentas para o Processamento de Línguas Naturais requerem ou se beneficiam de recursos linguísticos, tais como léxicos, gramáticas e redes semânticas (BRANCO et al. 2010; JURAFSKY & MARTIN, 2009).

Vejamos a figura abaixo:

Figura 1 – Recursos e ferramentas de NLP



Fonte: Elaborada pelo Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat)

Na figura 1, acima, temos uma arquitetura de NLP. A coluna da direita apresenta algumas camadas de ferramentas para o processamento de linguagem natural e a coluna da

esquerda apresenta algumas camadas de recursos linguísticos necessárias para a atuação das ferramentas.

Para tarefas como a análise semântica automática, uma gramática computacional é uma fonte altamente valiosa de informação. Uma vez que a análise sintática profunda por meios automáticos (*deep parsing*) é uma tarefa central no processamento de linguagem natural, neste passo da representação automática do conhecimento sentencial, quanto maior e mais qualitativa for a cobertura da gramática, melhores serão as condições de atuação dos analisadores semânticos, pois o não reconhecimento de uma estrutura durante a análise sintática pode comprometer análises posteriores.

Chamamos atenção para o fato de que gramáticas são recursos linguísticos que ficam disponíveis para a camada de ferramentas, tais como: o *syntactic parser*, o *lemmatizer* e o *POS tagger*. No caso do processamento sintático profundo, realizado por um *syntactic parser*, o *output* do analisador, ou seja, uma análise sintática profunda, fornecerá subsídios para as ferramentas da camada posterior, sendo, portanto, extremamente útil para o processamento semântico.

Nessa tese, investigamos diversificados e complexos fenômenos sintáticos da frase do Português Brasileiro (doravante PB), à luz da Gramática Léxico Funcional (LFG, *Lexical Functional Grammar*) e os implementamos no sistema XLE (*Xerox Linguistic Environment*), uma robusta ferramenta para o desenvolvimento de gramática no formalismo da LFG, cujos direitos estiveram reservados até 2001 à Xerox Corporation e de 2002 a 2011 ao Palo Alto Research Center.

Para português europeu (doravante PE), gramáticas computacionais como a LXGram (BRANCO & COSTA, 2010; SILVA et al. 2010a e 2010b), desenvolvida no âmbito do formalismo *Head-driven Phrase Structure Grammar* (HPSG) e o *parser* VISL (BICK, 2000), desenvolvido no formalismo da *Constraint Grammar* (CG), têm pavimentado o caminho para análise sintática automática de larga escala.

Nas tabela abaixo, apresentamos o estado da arte em processamento sintático profundo para o português.

Tabela 1 – Estado da arte em processamento sintático profundo do português

PARSER	QUADRO TEÓRICO	INSTITUIÇÃO	OBSERVAÇÕES
LX-Gram	HPSG	Universidade de Lisboa, Dep. de Informática, Grupo de Fala e Linguagem natural.	Precisão, larga escala, processamento sintático profundo, múltiplos propósitos do NLP.
Palavras	CG	Institute of Language and Communication, University of Southern Denmark, Projeto Linguateca	Níveis de análise: morfológico, sintático, dependência, estrutura de constituintes e papéis temáticos. Aplicações: ensino, anotação de corpus, tradução automática, QA-systems, NER.

Fontes: LXGRAM (BRANCO & COSTA, 2008)

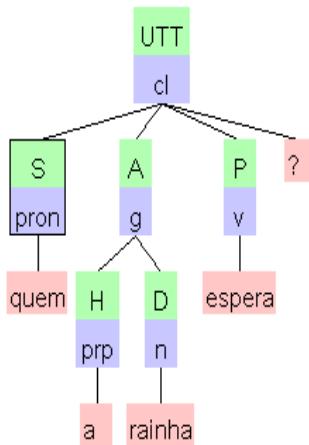
VISL (BICK, 2000)

O Palavras é, muito provavelmente, o *parser* mais utilizado em tarefas de processamento de linguagem natural. Mesmo com robustez, suas análises, no entanto, revelam algumas falhas quando as estruturas linguísticas envolvem fenômenos de maior complexidade. É o caso por exemplo das interrogativas deslocadas, das orações com controle funcional e das sentenças que sofrem apassivação.

Ao implementarmos no *parser* a sentença interrogativa *quem a rainha espera?*, cuja complexidade está em interpretar um argumento do verbo em uma posição não canônica, o Palavras não consegue analisar o pronome interrogativo *quem* como o objeto do verbo, atribuindo-lhe a função de sujeito. Confira a análise do Palavras na Figura 2, abaixo:

Figura 2 – Análise de uma sentença interrogativa pelo *parser* Palavras

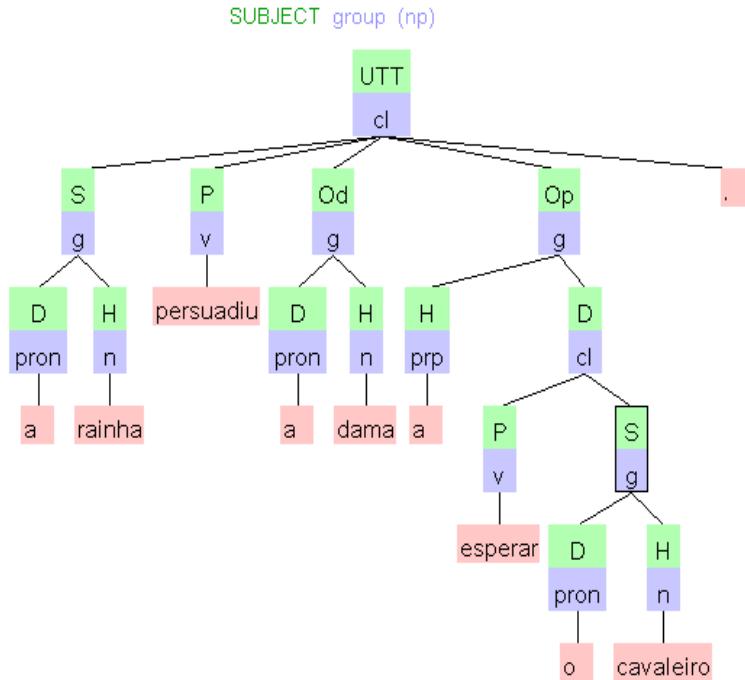
SUBJECT pronoun (indp 'quem' <interr> M/F S)



Fonte: Visual Interactive Syntax Learning (BICK, 2000), último acesso em 15.12.2014.

Também constatamos que o *parser* Palavras apresenta problemas para analisar sentenças que apresentam o fenômeno conhecido como controle funcional. Nestes casos, o sujeito de uma oração subordinada é controlado por um dos argumentos do verbo principal. Uma análise adequada para este fenômeno deve levar em consideração que o argumento externo do verbo encaixado é equivalente ao um dos argumentos do verbo principal.

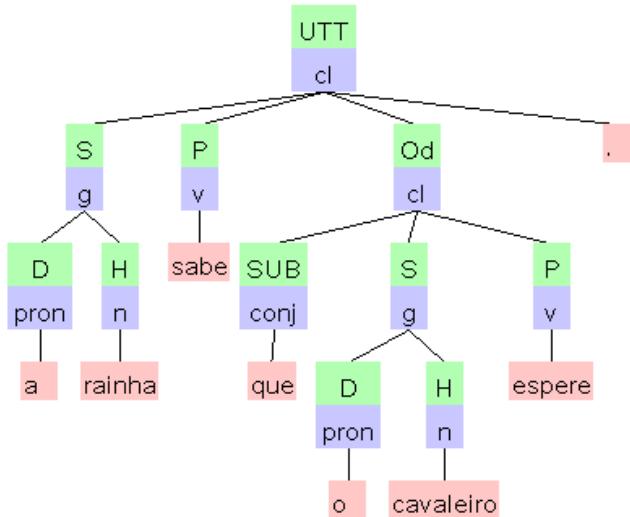
Figura 3 – Análise de uma sentença com controle funcional pelo parser Palavras



Fonte: Visual Interactive Syntax Learning (BICK, 2000), último acesso em 15.12.2014.

Observamos que o parser Palavras não captura as restrições impostas pelo modo verbal do verbo principal ao modo verbal de sentenças subordinadas. Desse modo, fornece, quando não deveria, análise para uma sentença agramatical como *a rainha sabe que o cavaleiro espere*.

Figura 4 – Análise de uma sentença com restrição Indicativo-subjuntivo pelo Palavras



Fonte: Visual Interactive Syntax Learning (BICK, 2000), acessado em 15.12.2014.

No Brasil, um exemplo de esforços aplicados na implementação de analisadores sintáticos automáticos é o *parser* Curupira (MARTINS; OTHERO, 2012). Este, por sua vez, não leva em conta as contribuições mais recentes da linguística gerativa para a descrição do português, fazendo uso das análises da gramática tradicional.

Por outro lado, até onde pudemos observar, o único esforço para elaborar gramáticas computacionais no sofisticado formalismo da LFG e implementar no XLE é demonstrado por Alencar (2004, 2013a e 2013b). Nesses trabalhos o autor faz análises léxico-funcionais de complementos verbais oracionais, modela a concordância nominal do pronome *a gente* e apresenta uma seleção de sentenças do português sintaticamente anotadas, que está disponível *online* em uma plataforma que hospeda *treebanks* para muitas línguas (ROSÉN et al., 2012).

Para o PB, para o qual esses recursos ainda são pouco disponíveis e trabalhos linguísticos teóricos no quadro da LFG ainda estão em estágios iniciais, a pesquisa precisa se concentrar em muitas questões sobre a informação sintática da frase como um todo, que devem ser codificadas com base em fundamentadas justificativas teóricas. Este é o cenário do PB, onde gramáticas computacionais são escassas e onde trabalhos em análise sintática computacional e teórica da língua são ainda limitados.

Uma das tarefas centrais no processo de desenvolvimento de uma gramática computacional é uma definição das propriedades sintáticas da frase de uma língua que serão cobertas pelo fragmento. Com base na investigação da frase do português e nas alternâncias sintáticas que nela ocorrem, nós mostramos uma enorme variedade de padrões da frase que a língua exibe, levando em consideração tanto elementos oracionais, quanto os elementos verbais, nominais, preposicionais, adjetivais e adverbiais que os constituem. Esta variedade de fenômenos é realmente bastante significativa para o processamento sintático profundo, mas variações dos padrões sintáticos não regulares ainda precisam ser modelados.

A gramática toma, como ponto de partida, tanto descrições formais do PB, quanto de outras línguas, notadamente o inglês, o francês e o alemão, cujas estruturas morfossintáticas, embora as vezes diferentes, apresentam muitas vezes consideráveis semelhanças com o português. Nesse sentido, temos que destacar quatro trabalhos. O primeiro é a implementação de um fragmento do francês no formalismo da LFG, realizada por Schwarze (1998) e a de um fragmento do inglês, realizada por Butt et al. (1999), que nos orientaram na escolha do leque de fenômenos a serem cobertos inicialmente por nossa gramática do português quando da solução a adotar no tratamento de varias questões gramaticais. O terceiro trabalho é o modo de distribuição do sistema de anotação *Parallel Grammar Project* (ParGram) (KING, 2004), adotado por Rosén et al. (2012), que nos esclareceu na disposição dos atributos e dos traços na gramática. E, por fim, o quarto trabalho é a modelação em termos da teoria X-barra, proposta por Othero (2009), de um considerável fragmento da sintaxe do português do Brasil no formalismo da CFG.

No fragmento de gramática estão modelados diversificados e relevantes elementos da sintaxe frasal. Ressaltamos que, nesta tese, assumimos que o leitor está familiarizado com conceitos elementares da abordagem teoria sintática formal. Também chamamos atenção para o fato de levantarmos, em nossa discussão, somente exemplos da variante padrão da língua portuguesa utilizada no Brasil.

O primeiro ponto da gramática a ser ressaltado são os sintagmas flexionais (IP) e os sintagmas complementadores (CP), elementos que encabeçam as sentenças do português, e, portanto, constituem o nó raiz da gramática. Neste ponto da gramática estão codificadas as principais distinções entre orações declarativas e interrogativas.

Outros fenômenos modelados bastante complexos são os elementos verbais, como subcategorização e estrutura argumental. Estes aspectos são cruciais no estabelecimento das

relações gramaticais e na formação da estrutura de predicados. Ainda sobre elementos verbais, a gramática oferece duas análises distintas para os predicativos de sujeito: a análise do sujeito controlado e a análise *predlink*, e faz uma clara distinção formal entre verbos auxiliares e verbos plenos, além disto, também apresenta uma modelação dos complexos fenômenos da inversão de voz ativa para passiva e dos verbos de controle, cruciais para uma posterior análise semântica automática, uma vez que, neste último caso, a interpretação dos argumentos do verbo subordinado dependerá da estrutura de predicados do verbo principal. Como os elementos verbais são centrais nas sentenças de uma língua, nossa gramática ainda contempla casos de predicados complexos e de verbos modais.

Os elementos nominais tratados na gramática, de modo central, foram os pronomes expletivos e reflexivos, e os casos de sintagmas nominais e determinantes com pronomes demonstrativos e interrogativos. Os demais constituintes gramaticais modelados na gramática são os sintagmas preposicionados, que exigem modelação distinta para as preposições semânticas e para as não semânticas; os sintagmas adjetivais, cuja projeção na sentença pode ocorrer a partir de formas adjetivas atributivas, de formas ordinais ou cardinais e na forma de intensificadores; e os sintagmas adverbiais, cuja estrutura interna foi modelada levando-se em consideração tanto advérbios intransitivos quanto transitivos com complemento PP. Outros fenômenos modelados na gramática, mas que não foram descritos de maneira central neste trabalho, são os casos de concordância nominal em torno do número, gênero e pessoa e os casos verbais de flexão de tempo, pessoa e número.

Não obstante abranger um variado leque de fenômenos da sintaxe frasal do português, a nossa gramática oferece, na análise sintática automática do português baseada em regras, a perspectiva da LFG, teoria gerativa que fundamenta o XLE, estado da arte em sistemas de desenvolvimento de gramáticas computacionais. Para alcançar o objetivo principal de escrever tal excerto de gramática, mesmo sendo considerada uma *toy grammar*, assim qualificada pelas próprias limitações que uma tese de doutorado oferece, esta tese tem como objetivo específico fazer uma descrição das estratégias de formação das orações em português brasileiro.

Em suma, neste trabalho temos como objetivo desenvolver uma gramática computacional da frase do português em um formalismo sofisticado como a LFG e implementada no robusto sistema XLE, utilizando, para tanto, o sistema de anotação do Projeto ParGram (KING, 2004). O trabalho partiu de análises anteriores feitas para o Português

apresentadas em Rosén et al. (2012), para o inglês apresentadas por Butt et al. (1999) e para o francês apresentadas por Schwarze (1998) e utilizou a modelação em termos de Teoria X-barra proposta por Othero (2009).

Deste modo, o desenvolvimento da gramática evidencia a importância das pesquisas desenvolvidas entre Sintaxe Formal e Processamento de Língua Natural. Congruente com o processamento de linguagem natural, observa-se a importância de o português brasileiro ser contemplado, cada vez mais, com gramáticas computacionais mais abrangentes e fundamentadas. Explica-se, desta forma, que a capacidade das aplicações de PLN de desempenharem suas tarefas depende diretamente da relação que cada recurso estabelece no processamento de léxico, gramática, rede semântica, etc.

Contribuímos, portanto, com um recurso linguístico computacional que, diante das incompletudes observadas no que, atualmente, é tido como o estado da arte em gramática computacional, o *parser* Palavras, apresenta, para diversos fenômenos, análises sintáticas computacionais mais refinadas, que podem fornecer, como *input* para o processamento semântico, informações mais detalhadas para a interpretação semântica.

Esta tese está organizada da seguinte maneira. Após esta introdução, no segundo capítulo, apresentamos as questões de fundo relevantes para dar ao nosso leitor uma visão geral do módulo da gramática gerativa, conhecido como teoria X-barra, seguida de uma apresentação sobre o modelo de projeção das categorias funcionais e lexicais que adotamos nesta tese, concluindo com uma introdução sobre os níveis básicos de representação sintática assumidos pela LFG. O capítulo 3 faz uma demonstração de como a LFG, o XLE e o ParGram se integram na descrição e no desenvolvimento das gramáticas elaboradas para o processamento sintático profundo. No capítulo 4, maior e mais detalhado, discutimos as soluções formais adotadas para a modelação da gramática e apresentamos as duas estruturas de representação sintática propostas pela LFG que são geradas automaticamente pelo *parser*. Ainda neste capítulo levantamos uma discussão em torno da avaliação da gramática, tecendo algumas conclusões tanto sobre o seu poder de análise, quanto sobre suas limitações. O capítulo 5 apresenta as conclusões deste trabalho, levantando uma discussão em torno dos resultados obtidos. Por fim, o capítulo 6 apresenta as considerações finais desta tese.

2 PRESSUPOSTOS TEÓRICOS

Nesta seção apresentamos os pressupostos teóricos mais relevantes desta tese, começando com a demarcação o modelo da teoria X-barra que é assumido durante o trabalho (seção 2.1) e, em seguida, apresentando os aspectos estruturais internos de cada categoria sintagmática que adotada em nossas análises da estrutura frasal do português (seção 2.2). Finalmente, procedemos com uma discussão sobre os pressupostos básicos e os mecanismos de formalização, assumidos pela LFG, sobre os dois níveis básicos de representação sintática, a estrutura de constituintes e a estrutura funcional (seção 2.3)¹.

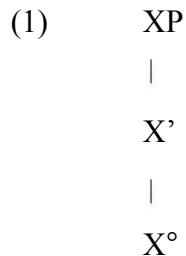
2.1 Formalismo sintático: Teoria X-barra

O objetivo da teoria X-barra é a descrição sintática dos sintagmas em variadas línguas, conforme sua natureza, sua hierarquização e suas relações internas. Seu intuito é construir hipóteses restritivas sobre a forma das regras de uma língua e sob a forma das estruturas que resultam da ação dessas regras (RAPOSO, 1992, p.159). Trata-se do módulo da Gramática Gerativa responsável pela esquematização dos sintagmas linguísticos em modo binário e pela delimitação de seus núcleos.

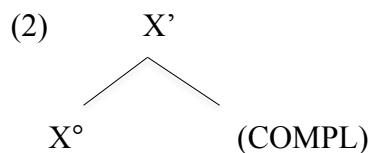
Em seu esquema, após a delimitação de seus núcleos, que podem ser tanto funcionais, como lexicais, os constituintes linguísticos são representados em três níveis.

O primeiro nível, cujo núcleo X^0 corresponde a um item lexical, possui valor correspondente a classes gramaticais como N, V, etc. Este nível pode projetar mais dois níveis de constituintes, o nível intermediário X' e o nível máximo X'' .

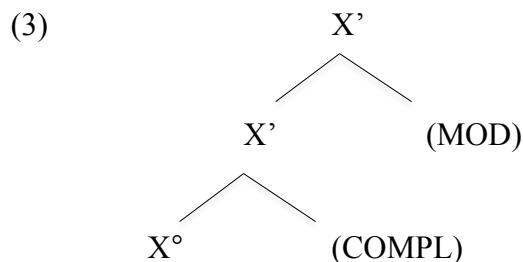
¹ Queremos ressaltar que esta tese não pretende ser uma introdução à sintaxe. Estamos assumindo que o leitor está minimamente familiarizado com conceitos elementares da teoria sintática formal.



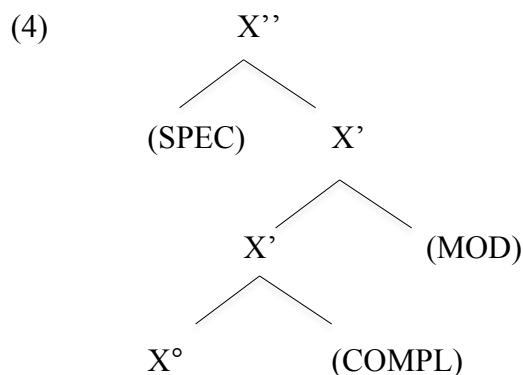
O nível X° seleciona um complemento (COMPL) e projeta o nível intermediário X' .



O nível intermediário X' , se selecionar um constituinte modificador, projetará outro nível X'' .



Ele só projetará outro nível X'' , se o constituinte selecionado for um constituinte especificador (SPEC).



O objeto da teoria X-barra são construtos linguísticos, cuja natureza é determinada pelo seu núcleo. Tais construtos representam o nível hierárquico dos seus constituintes em termos de especificação, complementação e circunstância e suas relações são estabelecidas em termos de seleção semântica e seleção sintagmática.

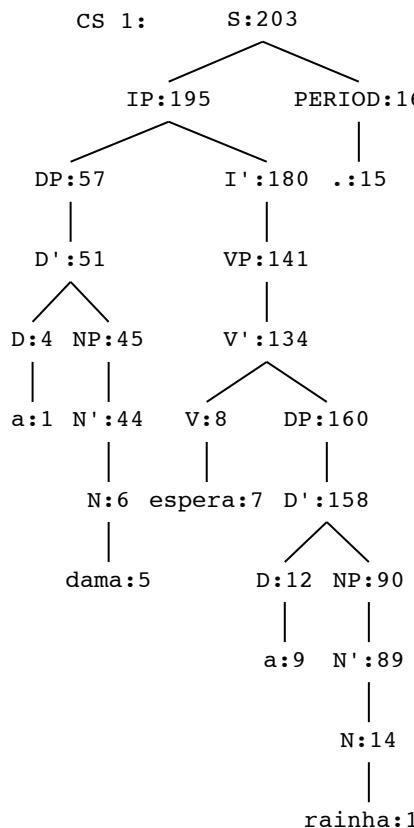
O modelo aqui adotado, por vezes, se afasta de um dos principais critérios da teoria de que todo sintagma deve ter núcleos da mesma categoria: o núcleo de um sintagma determinante (DP) é um determinante (D), o do sintagma flexionado (IP) é uma flexão (I), etc. No entanto, verificada a complexidade da implementação computacional desta propriedade, chamada de **endocentricidade**², esta noção precisou ser requalificada no tratamento dado aos elementos oracionais, os sintagmas IP e CP, visto que, na gramática, modelamos tais elementos como constituintes de S (sentença), uma categoria exocêntrica. Outro tipo de contra exemplo são os casos em que núcleos de sintagmas funcionais estão ausentes (FALK, 2001, p. 43; BRESNAN, 2001, p. 126-130).

Considere a sentença em (5):

- (5) A fada espera a rainha

² Segundo Raposo (1992, p. 163), “o princípio da endocentricidade pode ser formulado como uma condição de boa formação sobre as regras do componente categorial [...]. De acordo com este princípio (i) uma categoria sintagmática XP tem obrigatoriamente um núcleo pertencente a uma categoria lexical principal e (ii) para uma dada categoria sintagmática XP, o núcleo pertence à categoria lexical correspondente X.

Figura 5 – Estrutura de constituintes da sentença *a fada espera a rainha*



Fonte: elaborada pela autora através do XLE

Observe que, na estrutura acima³, o IP não tem o núcleo I. Isto é uma violação da endocentricidade, que não afeta o princípio da correspondência função-estrutura. Para este desvio, recorremos à argumentação apresentada em Falk (2001) e Bresnan (2001) de que a ausência de núcleo em IP é gramatical, porque o próprio especificador fornece as características que seriam fornecidas pelo núcleo.

³ As estruturas de constituintes apresentadas nesta tese, que foram geradas automaticamente pelo XLE, apresentam índices numéricos que atribuem a uma dada subárvore uma estrutura funcional da matriz total de atributos e valores. Entraremos em detalhes sobre estes índices nas seções 2.3.2 e 3.1.

2.2 Constituintes sintagmáticos

Um dos postulados da teoria X-barra é a distinção feita entre categorias funcionais e categorias lexicais. Nas próximas subseções apresentamos as projeções dessas categorias.

2.2.1 *Categorias funcionais*

As categorias funcionais são projetadas a partir de núcleos funcionais. Os núcleos funcionais são assim chamados por não selecionarem semanticamente seus complementos. Isto é, ao selecionarem seus complementos, têm em vista apenas a categoria sintagmática à qual eles devem pertencer (MIOTO, SILVA, LOPES, 2005, p.56).

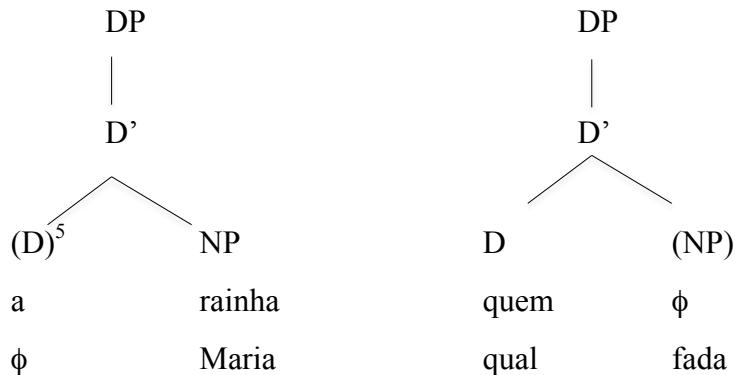
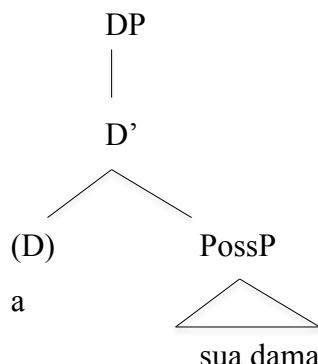
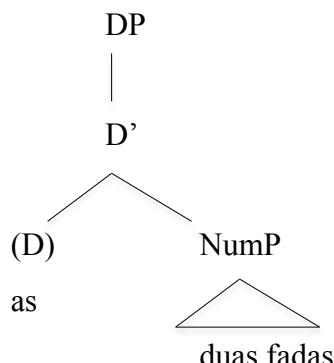
As categorias funcionais que são abordadas neste trabalho são o complementador (C), a flexão (I), o determinante (D), o possessivo (Poss) e o numeral (Num). Segundo Falk, (2001, p. 39), “estas categorias são assim chamadas porque seu propósito é fornecer traços gramaticais para seus sintagmas”.

Estas categorias funcionais são os núcleos dos sintagmas complementadores (CP), dos sintagmas flexionais (IP), dos sintagmas determinantes (DP), dos sintagmas numerais (NumP) e dos sintagmas possessivos (PossP).

As primeiras regras que apresentamos são as regras para a descrição da estrutura do DP. Em nossa gramática, assumimos que o núcleo funcional D pode selecionar os complementos categorias NP, PossP e NumP (OTHERO, 2009, 55)⁴. Vejamos, a seguir, três estruturas distintas do DP:

⁴ Para discussões aprofundadas sobre a hipótese DP, sugerimos Carnie (2002) e David (2007).

(6)

a. DP com complemento nominal**b. DP com complemento possessivo****c. DP com complemento numeral**

⁵ No formalismo CFG o uso do símbolo () serve para indicar a opcionalidade de um dado constituinte na estrutura de constituintes e o uso do símbolo | serve para indicar disjunção.

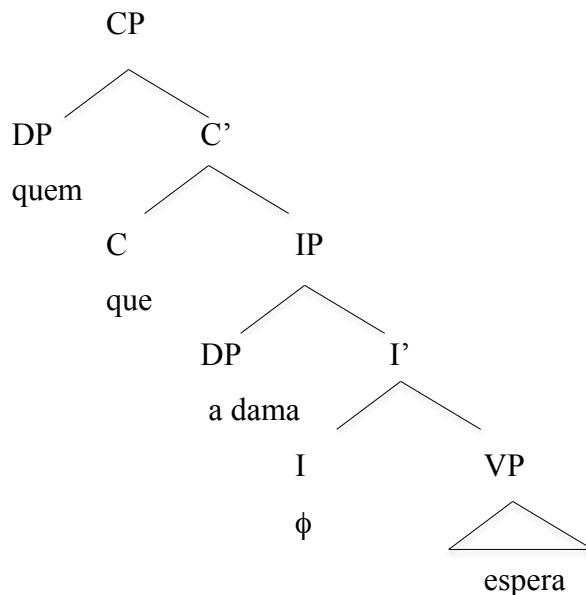
Estas três estruturas são geradas pelas seguintes regras⁶:

$$(7) \quad \begin{aligned} DP &\rightarrow D' \\ D' &\rightarrow (D) NP \mid (D) NumP \mid^7 (D) PossP \end{aligned}$$

O mesmo efeito de processamento acontece no núcleo C. Este núcleo pode selecionar tanto IPs, quanto VPs como complementos.

Vejamos as duas estruturas:

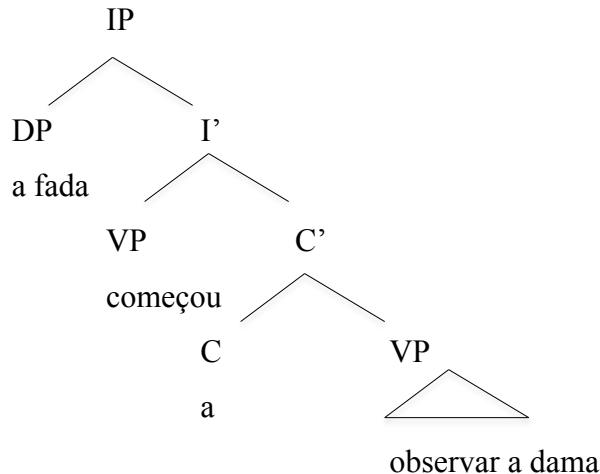
(8) **a. CP com complemento oracional**



⁶ Estas regras de reescrita categorial descrevem apenas aspectos básicos de seleção sintática. Neste formato as regras não são suficientes para fazer análises léxico-funcionais, sendo necessário, portanto, um formalismo com um grau muito maior de sofisticação para dar conta das características que assumimos estarem presentes em cada nó da árvore.

⁷ No formalismo CFG o uso do símbolo | serve para indicar disjunção.

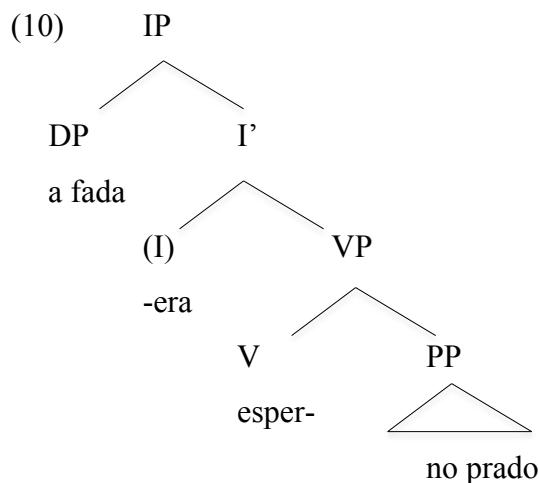
b. CP com complemento verbal



Estas duas estruturas são geradas pelas seguintes regras:

- (9) $CP \rightarrow (DP) C'$
 $C' \rightarrow C VP \mid C IP$

Adiante com a apresentação das projeções de núcleos funcionais, tratamos agora de I. O núcleo I, como todo núcleo funcional, seleciona um complemento (OTERO, 2009; SANTOS, 2009). O complemento de I será um VP. Vejamos a estrutura interna de IP.



Em português, verbos auxiliares ocupam I. Todos os outros elementos verbais, incluindo verbos sem aspecto de tempo, ocorrem dentro de VP. I não está presente quando não há auxiliar tensionado.

Para gerar a estrutura acima, a gramática precisa, basicamente, das seguintes regras:

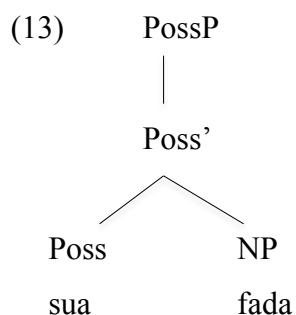
- (11) IP -> DP I'
 I' -> (I) VP

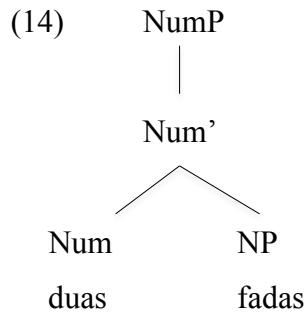
Nos casos em que o sujeito é posposto, deve haver uma regra na gramática que capture a ocorrência do DP à direita do verbo, como em (12).

- (12) IP -> I' DP

Ainda tratando das categorias funcionais, vemos nas estruturas (13) e (14), que os núcleos Poss e Num selecionam NPs. Uma boa argumentação para se propor as categorias Poss e Num pode ser vista em Othero (2009), sustentada por Perini (2000).

Para estes autores, os pronomes possessivos e os numerais não podem ser classificados como determinantes, pois não estão em distribuição complementar na estrutura dos agrupamentos nominais em português.





As estruturas arbóreas acima necessitam, para ser geradas, das seguintes regras:

(15) **a. Regras do PossP**

PossP -> Poss'

Poss' -> Poss NP

b. Regras do NumP

NumP -> Num'

Num' -> Num NP

Todas as estruturas apresentadas acima nos parecem interessantes e por isso partimos delas para seguir com a modelação da nossa gramática no capítulo quatro. A seguir apresentamos projeções de categorias lexicais.

2.2.2 *Categorias lexicais*

As categorias lexicais são projetadas a partir de núcleos lexicais. Os núcleos lexicais são assim chamados devido à capacidade que possuem de selecionarem semanticamente seus argumentos. Isto é, ao selecionarem seus complementos, têm em vista se as propriedades semânticas dos argumentos são compatíveis com as do núcleo lexical (MIOTO, SILVA, LOPES, 2005, p. 54). Portanto, espera-se que um argumento para *comer* seja algo que pode ser comido. Deste modo, uma sentença como *a criança comeu a água* não é uma sentença válida do português.

As categorias lexicais que são abordadas nesta tese são N (substantivo), V (verbo), A (adjetivo), P (preposição) e Adv (advérbios). Segundo Falk (2001, p. 34), “estas são categorias que carregam significado.” Estas categorias lexicais são os núcleos dos sintagmas nominais (NP), verbais (VP), adjetivais (AP), preposicionados (PP) e adverbiais (AdvP), respectivamente.

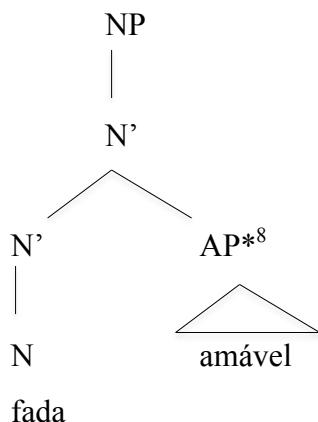
Primeiramente apresentamos a estrutura interna do NP. Em nossa gramática tratamos apenas do núcleo lexical N intransitivo, aquele que não seleciona complementos e de N modificados por APs.

Vejamos, a seguir, duas estruturas distintas do NP:

(16) **a. NP com N intransitivo**



b. NP com modificação adjetival



⁸ Este símbolo é conhecido como estrela de Kleene. É usado para expressar a capacidade de recursividade de um sintagma, i. e. quando queremos dizer que uma categoria sintagmática pode ocorrer uma ou mais vezes.

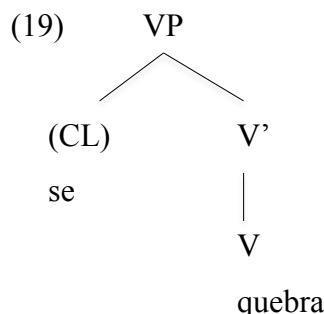
Estas duas estruturas são geradas pelas seguintes regras:

$$(17) \quad \begin{array}{l} \text{NP} \rightarrow N' \\ N' \rightarrow N \mid N' AP^* \end{array}$$

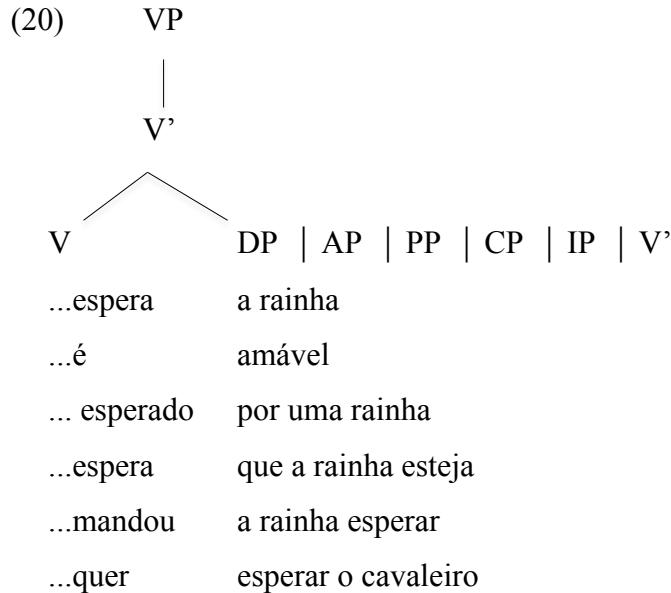
Por outro lado, os casos de modificação adjetival à esquerda são tratados com a seguinte regra adicional:

$$(18) \quad N' \rightarrow AP^* N'$$

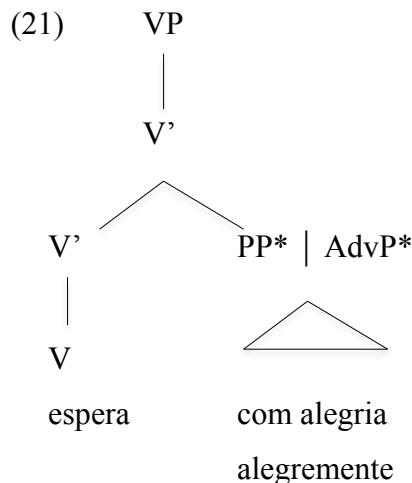
No capítulo quatro, onde trabalharemos com a modelação da gramática, partiremos das seguintes estruturas na análise do núcleo lexical V:



A estrutura acima representa estruturas com verbos reflexivos, nas quais o pronome clítico (CL) aparece como um especificador do VP.



Note que, por serem complementos, os sintagmas são selecionados por V. O contrário acontece na estrutura a seguir. Nela, o AdvP e o PP, por não serem argumentos internos do verbo, estão em condição de adjunção a V'.

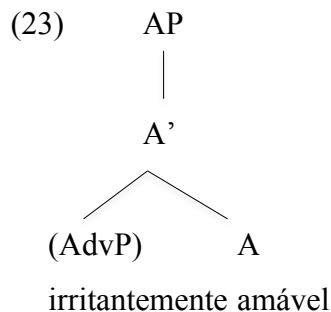


Os três exemplos acima sintetizam nove estruturas distintas que serão usadas na modelação da gramática em seção futura. Ressaltamos que este poder de síntese do formalismo se deve ao uso das expressões regulares (), | e * na representação dos constituintes.

As seguintes regras são necessárias para analisar as estruturas acima apresentadas:

- (22) $\text{VP} \rightarrow \text{CL } \text{V}' \mid \text{V}' \text{ PP}* \mid \text{V}' \text{ AdvP}*$
 $\text{V}' \rightarrow \text{V (DP)} \mid \text{V (AP)} \mid \text{V (PP)} \mid \text{V (CP)} \mid \text{V (IP)} \mid \text{V (V')}$

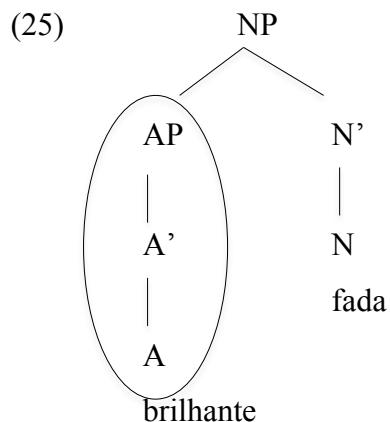
Prosseguimos com três estruturas distintas da categoria lexical A. Primeiramente, no exemplo (23), temos sintagmas adjetivais com modificadores adverbiais.



Estas estruturas são obtidas com as seguintes regras:

- (24) $\text{AP} \rightarrow \text{A}'$
 $\text{A}' \rightarrow (\text{AdvP}) \text{ A}$

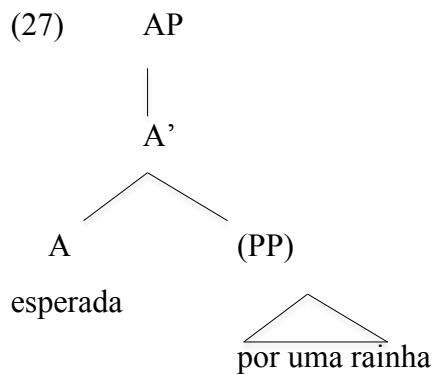
Em segundo lugar, temos uma estrutura mais simples, com apenas um adjetivo modificando um NP, como pode ser visto abaixo:



Damos, abaixo, a regra para gerar essa estrutura arbórea:

$$(26) \quad AP \rightarrow A' \\ A' \rightarrow A$$

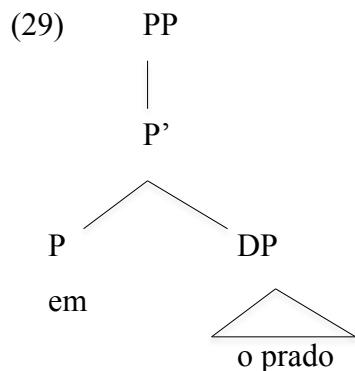
Na terceira estrutura adjetival um sintagma preposicionado pode ser um complemento de A. Vejamos:



Esta estrutura é gerada pela seguinte regra:

$$(28) \quad AP \rightarrow A' \\ A' \rightarrow A (PP)$$

Também são categorias lexicais as preposições e os advérbios. O mesmo efeito de processamento acontece com estes núcleos. O núcleo P seleciona, como complemento, um DP. Vejamos uma estrutura:

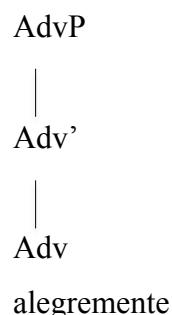


Esta estrutura é gerada pelas seguintes regras:

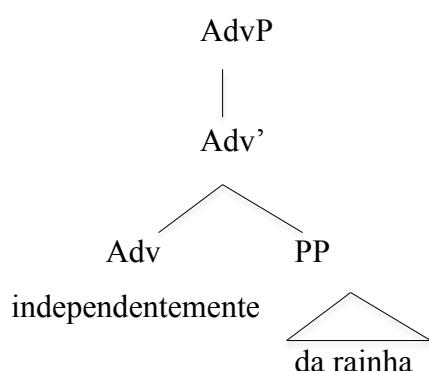
- (30) PP -> P'
 P' -> P DP

Já o núcleo lexical Adv, que tanto pode ser intransitivo, como pode selecionar um complemento PP ou ser modificado por outro AdvP, possui, pelo menos, as três estruturas abaixo:

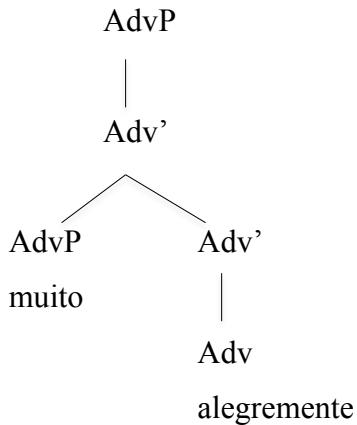
- (31) a. **AdvP com advérbio intransitivo**



- b. **AdvP com advérbio transitivo**



c. AP modificado por outro AdvP



Estas estruturas são geradas pelas seguintes regras:

- (32) $\text{AdvP} \rightarrow \text{Adv}'$
 $\text{Adv}' \rightarrow \text{AdvP } \text{Adv}' \mid \text{Adv} \mid \text{Adv PP}$

A seguir apresentamos os princípios básicos de representação da estrutura sintática na LFG e o formalismo adotado por esta teoria para a descrição gramatical.

2.3 Gramática Léxico-Funcional

A corrente gerativista que utilizamos como fundamentação teórica das nossas análises surgiu como uma proposta de teoria da gramática alternativa à gramática gerativa derivacional, porém sustentada pela hipótese da **Gramática Universal** (BRESNAN, 2001; FALK, 2001; KLENK, 2003; KROEGER, 2004). A LFG, como é conhecida, rejeita a noção derivacional fortemente presente em teorias como Regência e Ligação (Chomsky, 1982) e é orientada pela ideia de que os níveis de representação sintática ocorrem sem processos subjacentes à estrutura de constituintes. Nesta seção apresentamos, inicialmente, os principais critérios adotados pela LFG para formular a existência de estruturas de representação sintática separadas, porém relacionadas uma a outra de maneira sistemática. Particularmente, abordamos duas dessas estruturas, a estrutura de constituintes, que representa a ordem das palavras e os agrupamentos sintáticos, e a estrutura funcional, que representa funções gramaticais.

2.3.1 Algumas considerações teóricas

A razão mais importante para os gramáticos léxico-funcionais criarem uma alternativa ao gerativismo Chomyskiano reside no fato de que, para estas pessoas, as derivações são explicações paradoxais para alguns fenômenos (CARNIE, 2002). Bresnan (2001) apresenta vários argumentos para a contestação das regras de derivação. Um desses argumentos está relacionado com o fenômeno da apassivação de complementos oracionais.

Para Bresnan (2001), o fato de assumir que na formação de uma passiva o seu sujeito é o resultado do movimento do objeto da ativa falha com certos verbos.

- (33) a. [that languages are learnable] is captured _ by this theory.
- b. *This theory captures [that languages are learnable].
- c. This theory captures [the fact that langauges are learnable].

A passiva do exemplo (33a) é uma sentença gramaticalmente bem formada do inglês, que possui um sujeito oracional. Seu correspondente (33b) é uma sentença mal formada se considerarmos que o complemento do verbo *capture* deve ser um NP objeto e não um complemento oracional, como em (33c). Esse fato ilustra que o sujeito de (33a) não possui um correspondente ativo que corresponda ao NP objeto da sentença ativa em (33c), mas sim um complemento oracional. Segundo Bresnan (2001), isto demonstra que a correspondência entre estrutura e função não é perfeita.

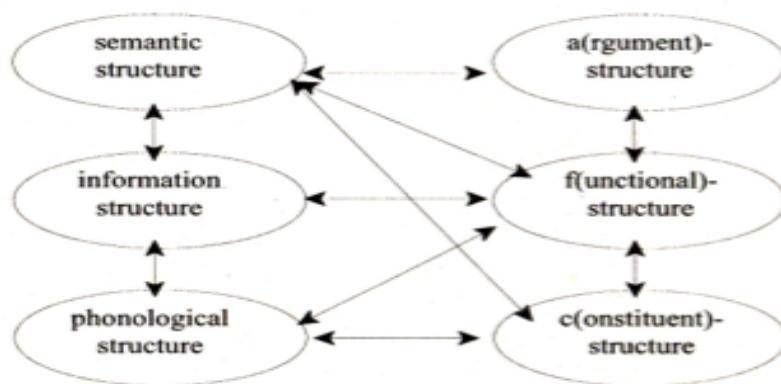
Este e outros argumentos utilizados para contestar o componente derivacional resultaram em dois postulados essenciais para a LFG. O primeiro deles é que informações funcionais podem ser adicionadas à uma estrutura de constituintes, mas não podem ser mudadas ou apagadas e segundo, a gramaticalidade não pode depender de propriedades de derivações. Bresnan e Kaplan suspeitavam que muitas informações sintáticas se estendem para além da estrutura de constituintes e por isso postularam que as funções gramaticais são primitivos linguísticos (CARNIE, 2002), presentes em todas as línguas e que não dependem necessariamente da posição dos seus elementos na estrutura de constituintes.

Essa suposição é notoriamente contrária à suposição de relações gramaticais da teoria vigente naquela época, a Teoria Padrão Estendida, para a qual tais relações são determinadas pela posição que um elemento assume em uma árvore (CARNIE, 2002). Para a Gramática Léxico-Funcional a estrutura sintática é construída fora dos seus átomos, ou seja, fora de suas palavras (FALK, 2001).

2.3.2 Níveis de representação estrutura-c e estrutura-f

A arquitetura de projeção da LFG baseia-se em níveis estruturais distintos. Cada um destes subsistemas possui seus próprios primitivos e regras de organização. Os níveis de estrutura linguística são analisados como existindo em paralelo e estão relacionados uns aos outros por funções correspondentes. O aspecto importante da pesquisa da LFG é identificar os sistemas de correspondência que ligam os níveis de representação.

Figura 6 – Sistemas de correspondência entre estruturas



Fonte: Falk (2004, p. 25)

Dois destes sistemas de correspondência estão relacionados no nível sintático. São eles: estrutura funcional (estrutura-f) e estrutura de constituintes (estrutura-c).

A estrutura-c é uma árvore de estrutura sintagmática convencional que indica o arranjo superficial e a hierarquia das palavras e sintagmas na sentença. Esta estrutura modela informação sintática relacionada a precedência, dominância e constituintes. Ela é descrita por regras de estruturas sintagmáticas e é normalmente empregada a teoria X-barra. Funções

gramaticais não são codificadas no nível de estrutura-c. Todo nó da estrutura-c está relacionado por uma função a uma estrutura-f específica, que promove uma simultânea satisfação de restrição entre estrutura-c e estrutura-f. Este mecanismo de instanciação é descrito em Bresnan (2001, p. 56-60). Em suma, a estrutura-c é a responsável pelo agrupamento das palavras em constituintes, o núcleo de um sintagma define as propriedades de um grupo inteiro, a estrutura-f é a responsável em relacionar os valores e atributos (FALK, 2001).

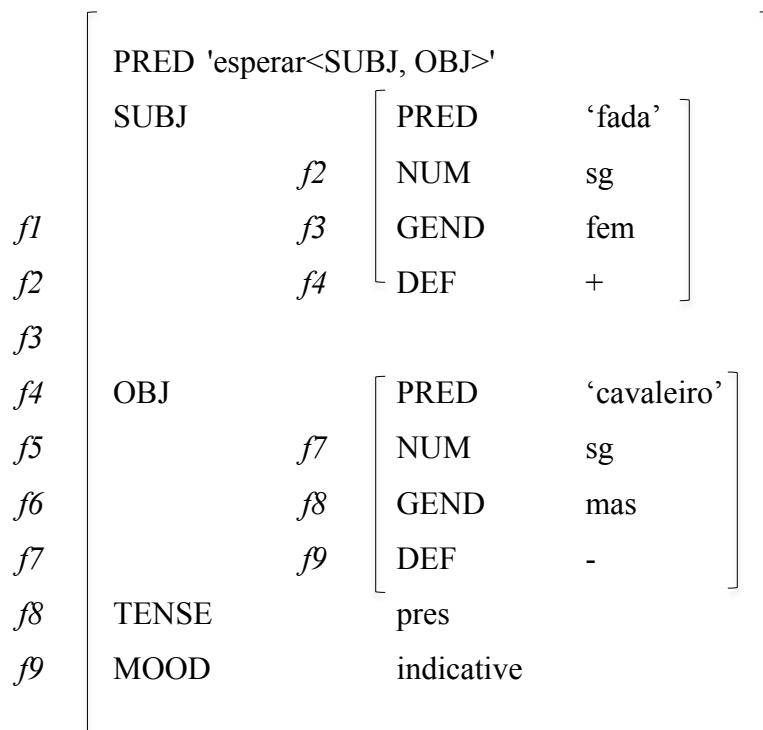
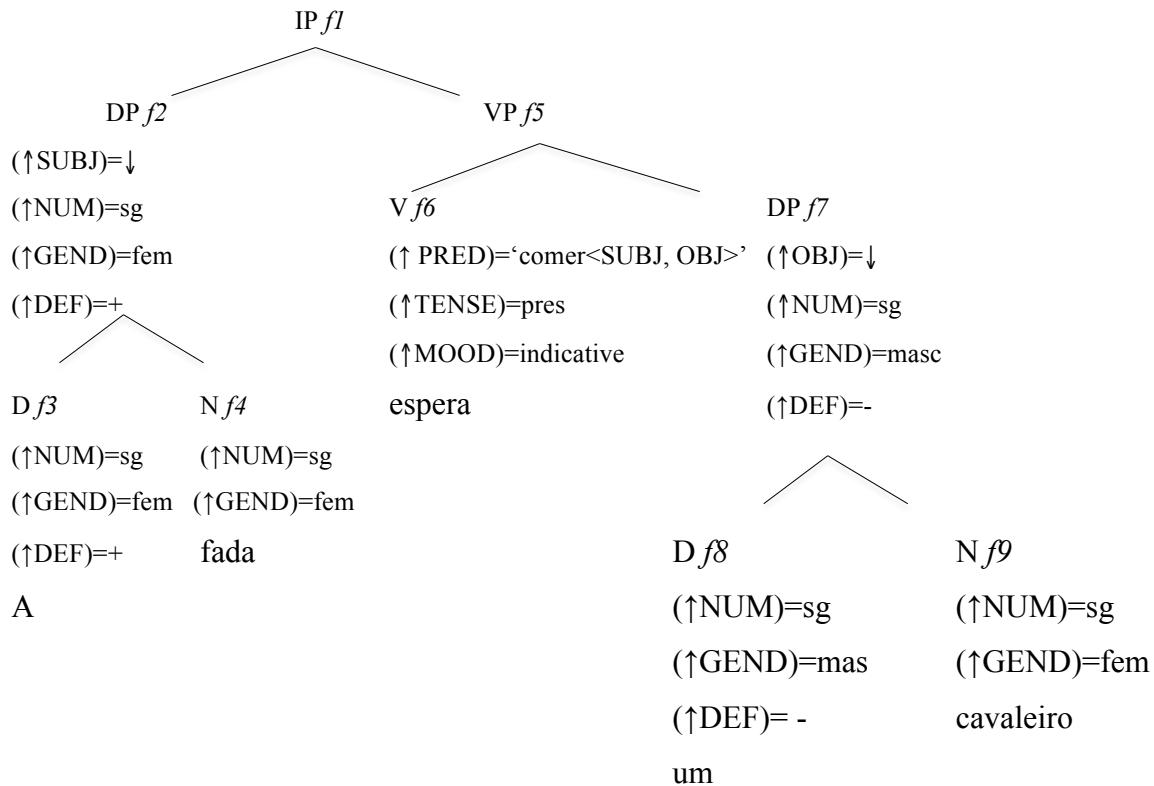
A estrutura-f codifica princípios universais de construções na forma de uma matriz de atributos e valores, baseada em anotações funcionais nos nós da estrutura-c e a informação vinda do léxico. Estas estruturas, por sua vez, expressam relações oracionais, informações sintáticas abstratas relacionadas a funções gramaticais, dependências e subcategorização. A estrutura-f consiste de atributos abstratos (traços e funções), cujos valores (FALK, 2001) são descritos em matrizes de atributos e valores. Nesta estrutura, os atributos são traços ou funções. As funções estão divididas em funções gramaticais e funções discursivas. Os atributos podem ter como valores, tanto símbolos, como formas semânticas, estrutura-f ou conjuntos. São as entradas lexicais que demonstram a contribuição que um item lexical faz para a estrutura-f.

As funções gramaticais são a ideia principal por trás da estrutura-f e são postuladas na LFG como funções do tipo argumentais e do tipo não argumentais (FALK, 2001). As funções do tipo argumentais, como Sujeito, Objeto, Objeto secundário (Obj2) e Oblíquo são elementos sintáticos que expressam os argumentos dos predicados. Adjuntos, Foco e Tópico também são funções gramaticais. No entanto, do tipo não argumental. As funções argumentais e a função não argumental Adjunto expressam aspectos internos da oração, já as funções não argumentais Tópico e Foco são funções gramaticais tidas como funções discursivas-funcionais por expressarem relações que são relevantes para o discurso (FALK, 2001).

Portanto, a correspondência entre a estrutura-c e a estrutura-f serve para associar constituintes sintáticos a traços e funções gramaticais. Cada um destes com tipos distintos de informação linguística, cada um com seus próprios primitivos e regras (MYCOCK, 2006).

Vejamos um exemplo de correspondência entre a estrutura-c e a estrutura-f na sentença em (34):

(34) A fada espera um cavaleiro.



Note que a estrutura-f respeita as condições de **coerência** e **completude** na formação da sentença. Diz-se que uma estrutura-f é completa quando todas as funções argumentais selecionadas pelo núcleo aparecem na estrutura. E se diz que uma estrutura funcional é coerente quando aparecem apenas aqueles argumentos selecionados pelo núcleo. Desse modo seria incompleta a estrutura-f de uma sentença como a que segue:

- (35) *O pai deu um presente.

PRED	'dar<SUBJ, OBJ, OBL>'
SUBJ	[o pai]
OBJ	[um presente]
OBL	???
TENSE	past
MOOD	indicative

A estrutura funcional acima está incompleta devido a ausência da função gramatical OBL(íquo), um dos argumentos requeridos pelo verbo.

Por sua vez, seria incoerente a estrutura funcional de uma sentença como vemos abaixo:

- (36) *O homem deu um presente ao filho ao neto.

PRED	'dar<SUBJ, OBJ, OBL>'
SUBJ	[o homem]
OBJ	[um presente]
OBL	[ao filho]
???	[ao neto]
TENSE	past
MOOD	indicative

A estrutura funcional acima está incoerente por aparecer nela um argumento que não é requerido pelo núcleo e por isso não pode ser interpretado.

Concluímos, com isso, a apresentação das questões de fundo mais relevantes desta tese. Neste capítulo nós destacamos o modelo de projeção X-barra que assumimos, chamando atenção para eventuais desvios conceituais, em seguida apresentamos aspectos estruturais internos de cada constituinte sintagmático, fazendo uma distinção entre categorias funcionais e categorias lexicais, e, finalmente, procedemos com uma discussão sobre os pressupostos básicos da LFG e as estruturas de representação sintática, estrutura-c e estrutura-f⁹.

⁹ Queremos ressaltar que esta tese não pretende ser uma introdução à sintaxe. Estamos assumindo que o leitor está minimamente familiarizado com conceitos elementares da teoria sintática formal.

3 O AMBIENTE DE DESENVOLVIMENTO LFG-XLE E O PROJETO PARGRAM

Na presente seção demonstramos o formalismo da LFG (seção 3.1), um sofisticado aparato teórico-computacional, o XLE, um robusto sistema de processamento de linguagem natural (seção 3.2) e o Projeto ParGram (seção 3.3), projeto internacional de elaboração de gramática, se integram, em um esforço colaborativo de pesquisadores no desenvolvimento de gramáticas para o processamento sintático profundo, tanto no âmbito acadêmico quanto no âmbito industrial.

3.1 O formalismo LFG-XLE

Atualmente, para o processamento sintático profundo baseado em regras, há diferentes formalismos gramaticais que podem ser destacados, como a Gramática de Unificação, a *Head-Driven Phrase Structure Grammar* (HPSG) e a *Lexical Functional Grammar* (LFG). Todos esses têm, à disposição, sistemas bastante robustos para a elaboração de gramáticas computacionais e, mais, são utilizados na implementação de *parsers* de grande envergadura para diferentes línguas (ALENCAR, 2013b; BICK, 2000; BRANCO & COSTA, 2012a, 2012b, 2012c, MÜLLER, 2010).

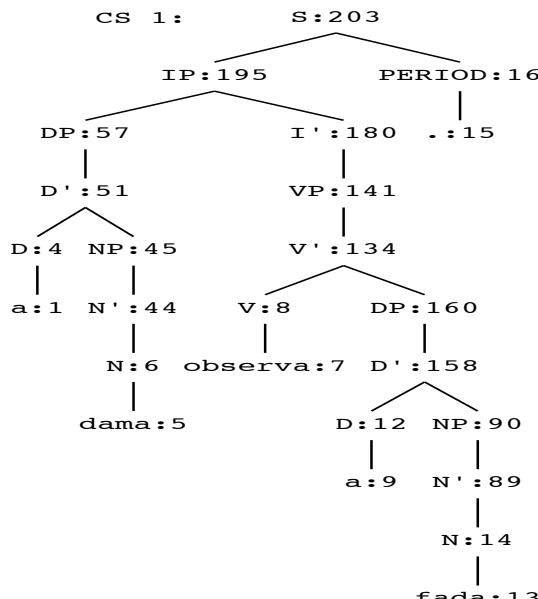
Em conformidade com a noção de níveis de representação sintática presente na LFG, a característica mais interessante de *parsers* baseados nesse formalismo é a geração de vários níveis de representação para as sentenças gramaticais, minimamente a estrutura de constituintes e a estrutura funcional, exemplificados, respectivamente na Figura 2 e na Figura 3.

Segundo Alencar (2013, p. 184),

esse segundo tipo de estrutura representa uma dimensão mais abstrata de informação gramatical, codificando a estrutura de predicado e argumentos, as relações gramaticais, bem como as propriedades morfossintáticas dos diferentes constituintes, como pessoa, número, gênero, tempo, modo, etc.

Vejamos a seguinte figura:

Figura 7 – Um exemplo de estrutura de constituintes representada no XLE



Fonte: elaborada pela autora através do XLE.

Na figura acima todos os nós da árvore possuem um índice numérico. Cada índice serve para equacionar um nó da estrutura de constituintes a uma determinada estrutura funcional. Na figura abaixo, podemos observar os índices numéricos equacionando a estrutura funcional do predicador da sentença com os seus argumentos.

Figura 8 – Um exemplo de estrutura funcional representada no XLE

"a dama observa a fada ."

```

[PRED   'observar<1:dama], [9:fada]>
  5[PRED   'dama'
  6
  44[NTYPE [NSEM [COMMON count]
SUBJ  45      [NSYN common
  1
  4[SPEC  [DET [PRED      'o'
        [DET-TYPE def]
  51
  57[CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3
  15
  13[PRED   'fada'
  16
  89[NTYPE [NSEM [COMMON count]
  7
  90      [NSYN common
OBJ   8
  9
  12[SPEC  [DET [PRED      'o'
        [DET-TYPE def]
  134
  141
  180
  158
  195
  160[CASE acc, GEND fem, HUMAN -, NUM sg, PERS 3
203[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main

```

Fonte: elaborada pela autora através do XLE.

Desta forma, o XLE, nos moldes das equações funcionais que vimos na seção 2.3.2, consegue estabelecer a interação básica, proposta pela LFG, entre estrutura de constituintes e estrutura funcional para a representação do conhecimento sintático.

Para se obter estas duas diferentes estruturas do nível sintático do processamento linguístico, a LFG possui um sofisticado formalismo para a descrição das línguas. Frente a este formalismo, o XLE propõe uma notação equivalente para os desenvolvimentos em ambientes computacionais¹⁰.

O ambiente XLE fornece a linguistas a facilidade para escrever regras sintáticas e entradas lexicais que tornem possível se obter estruturas como vimos acima. O XLE é um ambiente computacional que auxilia na escrita e na depuração de Gramáticas Léxico-Funcionais (CROUCH et al. 2011).

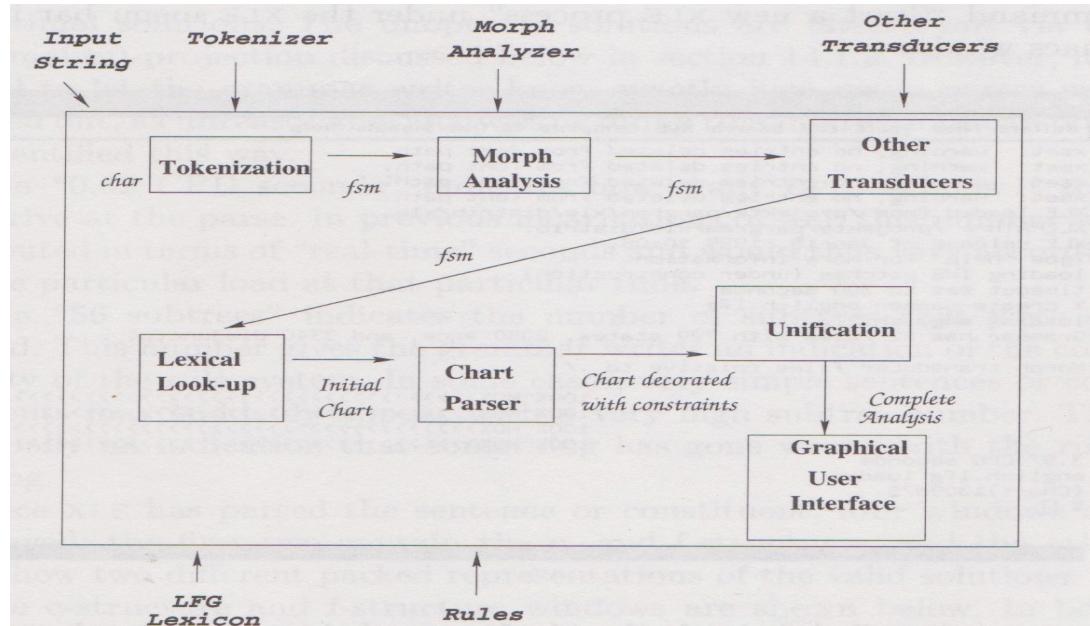
Segundo Sulger (2009, p. 8, tradução nossa)¹¹, “o XLE é uma poderosa ferramenta de processamento de língua natural que é usada para analisar e gerar texto usando gramáticas computacionais baseadas no formalismo LFG.”

A arquitetura do sistema conta com ferramentas de tokenização, análise morfológica e análise sintática, como é mostrado na figura abaixo.

¹⁰ O XLE usa apenas ascii para caracteres (de modo que a seta para cima é representado por ^ e a seta para baixo por !). A seguir apresentamos o acesso para a tabela que dá notações da LFG e seus equivalentes do XLE <http://www2.parc.com/isl/groups/nlt/xle/doc/notations.html#N0A>.

¹¹ XLE is a powerful natural language processing tool which is used to parse and generate text using computational grammars based on the LFG notation.

Figura 9 – Arquitetura do parser



Fonte: Butt et al. (1999)

O sistema, a fim de fornecer uma interface confortável e amigável para a definição e manipulação de regras linguísticas e representações, utiliza um software livremente acessível como o *Aquamacs*¹², baseado no GNU *emacs*, um componente do sistema operacional GNU.

Nesse contexto, utilizamos, na construção da nossa gramática, a metodologia convencionada para o desenvolvimento de gramáticas formais de uma língua natural, que consiste em definir, inicialmente, um recorte cobrindo um leque determinado de fenômenos, recorte esse expandido sucessivamente, resultando em versões cada vez mais amplas da gramática (ALENCAR, 2013, p. 184).

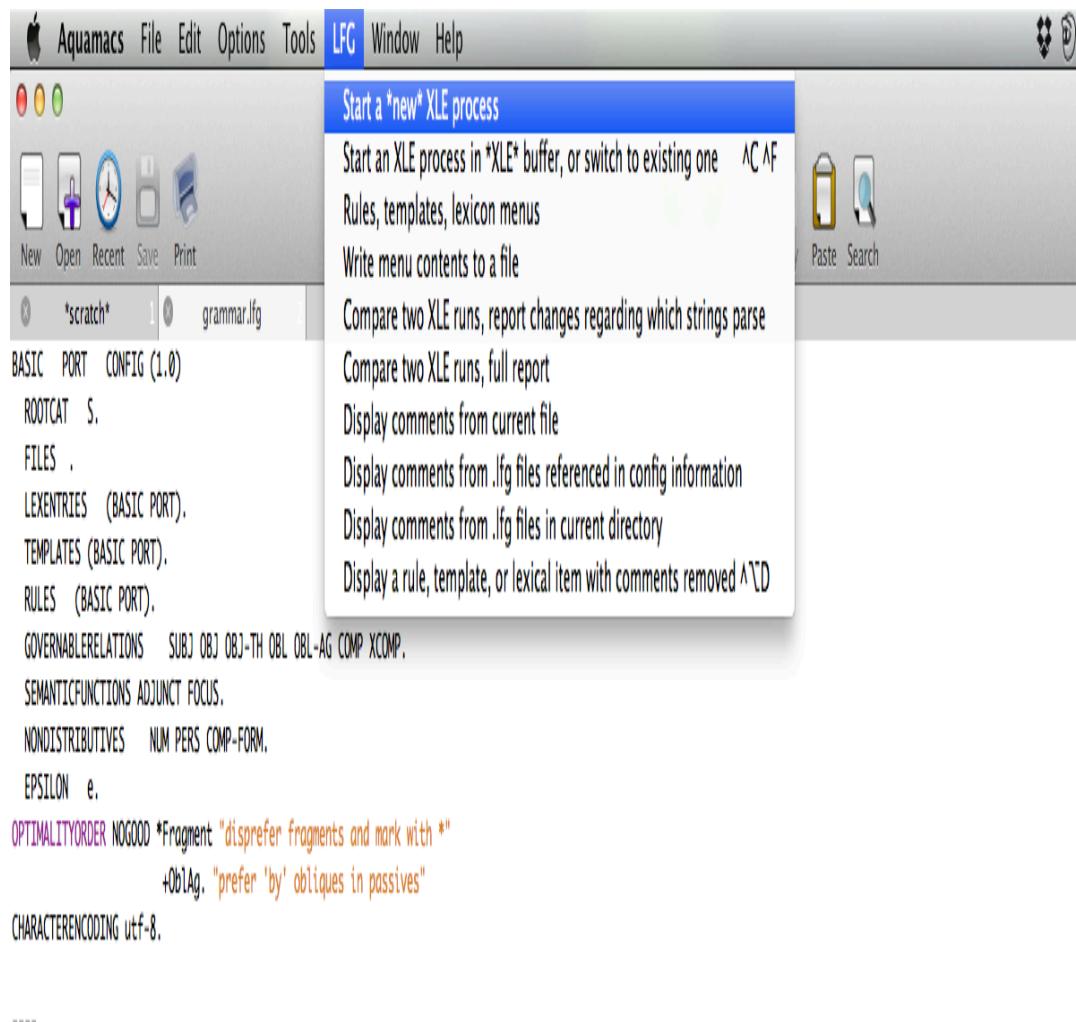
Tendo em mãos esse recorte, o próximo passo é criar um arquivo de teste com uma série de sentenças ou constituintes que pode ser alimentado para a gramática. Por outro lado, o arquivo também deve conter sentenças ou constituintes agramaticais para se ter certeza que as regras não as analisariam. No ambiente do XLE, o comando “parse-testfile testfile.lfg” resulta em uma análise gramatical de todas as sentenças do arquivo chamado, reportando o número de análises, o tempo que ele levou para ser analisado e o número de sub-árvores para a sentença

¹² <http://aquamacs.org>

(BUTT et al. 1999). Isto é necessário para saber se subsequentes mudanças na gramática afetam o comportamento de construções já tidas como certas.

Após definir uma coleção de regras sintáticas e entradas lexicais dentro dessa plataforma, chamadas ao XLE são invocadas por meio do modo LFG, desenhado por Mary Dalrymple, realizadas via comando “Start a new XLE process” na barra de menu na janela do Aquamacs Emacs (BUTT et al. 1999).

Figura 10 – Plataforma do Aquamacs Emacs



Fonte: XLE (CROUCH *et al.* 2011)

Uma vez que o XLE é chamado, ele carrega uma plataforma e então espera que o usuário faça algo. No Figura 6 abaixo, foi digitado o comando “create-parser grammar.lfg”.

Este comando permite o carregamento da gramática do português, definida no arquivo “grammar.lfg” (CROUCH et al, 2014).

Figura 11 – Plataforma do XLE invocada através do modo LFG do Aquamacs

```

Aquamacs  File Edit Options Tools XLE Complete In/Out Signals Window Help
New Open Recent Save Print *XLE*
*scratch* 1 grammar.lfg 2 *XLE*
XLE loaded from /Users/andreafeitosa/xle/bin/xle.
XLEPATH = /Users/andreafeitosa/xle.
Copyright (c) 1993-2001 by the Xerox Corporation and
Copyright (c) 2002-2011 by the Palo Alto Research Center.
All rights reserved. This software is made available AS IS,
and PARC and the Xerox Corporation make no warranty about
the software, its performance or its conformity to any specification.
XLE release of Apr 25, 2012 09:37.
Type 'help' for more information.
/Users/andreafeitosa/xlrc loaded.
loading /Users/andreafeitosa/Documents/Tese/Tese_defesa/Gramaticas/Port_declarativas/port09/grammar.lfg...
Grammar has 21 rules with 76 states, 138 arcs, and 138 disjuncts (138 DNF).
0.010 CPU seconds
/Users/andreafeitosa/Documents/Tese/Tese_defesa/Gramaticas/Port_declarativas/port09/grammar.lfg loaded
Grammar last modified on Nov 14, 2014 17:26.
/Users/andreafeitosa/Documents/Tese/Tese_defesa/Gramaticas/Port_declarativas/port09/xlrc loaded.
% create-parser grammar.lfg
loading /Users/andreafeitosa/Documents/Tese/Tese_defesa/Gramaticas/Port_declarativas/port09/grammar.lfg...
0.000 CPU seconds
/Users/andreafeitosa/Documents/Tese/Tese_defesa/Gramaticas/Port_declarativas/port09/grammar.lfg loaded
Grammar last modified on Nov 14, 2014 17:26.
(Chart)0x120d200
% parse "a fada é amável ."
parsing {a fada É amável .}
3 solutions, 0.000 CPU seconds, 0.100MB max mem, 23 subtrees unified
3
%

```

Fonte: XLE (CROUCH *et al.* 2011)

Na figura acima o que vemos é a nossa gramática carregada na plataforma do XLE, que após compilada, mostra a quantidade de regras da gramática, o número de estados, de arcos e disjunções que possui.

Na plataforma também pode-se observar que após uma análise, se obtém o número de estruturas funcionais geradas, o tempo de análise em cpu/seconds, memória e subárvores geradas, como no recorte mostrado na figura abaixo.

Figura 12 – Informações reportadas pelo XLE após análise de uma sentença

3 solutions, 0.000 CPU seconds, 0.106MB max mem, 23 subtrees unified

Neste caso, “3 solutions” indicam que a sentença possui três análises distintas. O “0.000 CPU seconds” indica o tempo de processamento para se chegar à análise. O “0.106 max mem” representa o espaço de armazenamento utilizado na memória e “23 subtrees” indica o número de sub-árvoreas que foram exploradas.

O sistema XLE está implementado em linguagem C e opera em Solaris, linus e Mac OS X, desde que se obtenha uma licença através dos seus organizadores¹³. Em particular esse sistema constitui a base para o desenvolvimento de gramáticas dentro do projeto ParGram¹⁴.

3.2 O projeto ParGram

No cenário LFG-XLE, levando-se em consideração tanto a consistência teórica da LFG, quanto o seu uso no processamento de linguagem natural, podemos dizer que, apesar de os formalismos HPSG e LFG se destacarem pelo sofisticado poder de descrição, utilizamos o formalismo da LFG, especialmente devido ao Projeto ParGram, que está voltado para a descrição de línguas tipologicamente diversas, a partir de um aparato conceitual comum (PARGRAM, 2012).

¹³ Expressamos aqui nosso mais profundo agradecimento à Professora Miriam Butt, da Universidade de Konstanz (Alemanha), por, gentilmente, ter concedido a licença com a qual instalei e passei a fazer uso dessa ferramenta.

¹⁴ Praticamente todas as estruturas-c e as estruturas-f que aparecem nesta tese foram produzidas pelo XLE.

O objetivo global do projeto é produzir colaborativamente gramáticas computacionais de grande escala, com atenção especial no paralelismo e na validade translingüística. Segundo Butt et al.(1999, p. 1), “as gramáticas são paralelas no sentido de que são guiadas por um conjunto comum de princípios linguísticos e um conjunto comumente acordado de análises gramaticais e traços.”

Em nossa modelação adotamos o sistema de anotação utilizado pelo Projeto ParGram, atendendo ao que se convencionou dentro do grupo de pesquisadores de instituições acadêmicas e industriais ao redor do mundo que adotam essa metodologia (KING, 2004).

O objetivo do projeto ParGram, colaboração internacional no desenvolvimento semântico e grammatical baseado na LFG, é produzir gramáticas de ampla cobertura para uma variedade de línguas. Essas são escritas, colaborativamente, dentro do quadro da LFG e com um conjunto de características gramaticais comumente acordados (KING, 2004).

Figura 13 – Traços do ParGram para o desenvolvimento de gramática no XLE.

```

ANIM: -> $ {+ -}.
AQUANT: -> << [ ADJUNCT PRED QUANT-TYPE DEGREE DEG-DIM ].
ATYPE: -> $ {adverbial attributive predicative}.
CASE: -> $ {acc dat erg gen inst loc nom obl}.
CLAUSE-TYPE: -> $ { adv cond decl imp int nom pol-int rel wh-int
}.
COMMON: -> $ { count gerund mass measure partitive }.
DEG-DIM: -> $ {equative neg pos}.
DEGREE: -> $ {comparative positive superlative}.
DET: -> << [ DEIXIS DET-TYPE PRED ].
DET-TYPE: -> $ {article def demon indef int rel}.
GEND: { -> $ {fem masc neut} | -> << [ FEM MASC NEUT ] }.
HUMAN: -> $ {+ -}.
GEND-SEM: -> $ {female male nonhuman}.
LOCATION-TYPE: -> $ { city country }.
MOOD: -> $ {imperative indicative subjunctive successive}.
NAME-TYPE: -> $ {first_name last_name }.
NSEM: -> << [ COMMON NUMBER-TYPE PROPER TIME ] .
NSYN: -> $ { common pronoun proper }.
NTYPE: -> << [ NSEM NSYN ] .
NUM: -> $ {pl sg}.
NUMBER: -> << [ NUMBER-TYPE PRED ADJUNCT CLASSIFIER-FORM MOD].
NUMBER-TYPE: -> $ {card fract ord percent}.
PASSIVE: -> $ {+ -}.
PERF: -> $ {+ - +_ -_}.
PERS: -> $ {1 2 3}.
PROG: -> $ {+ - +_ -_}.
PRON-TYPE: -> $ {demon expl_free inh-refl_int locative null pers
quant poss recip refl rel}.
PROPER: -> << [ PROPER-TYPE LOCATION-TYPE NAME-TYPE ].
PROPER-TYPE: -> $ {addr_form location name organization title }.
PSEM: -> $ {ag ben comit compar dir inst loc manner num part poss purp temp }.
PTYPE: -> $ {nosem sem}.
QUANT: -> << [ADJUNCT QUANT-TYPE POL PRED DEGREE DEG-DIM ].
QUANT-TYPE: -> $ {comparative equative existential gen negative superlative universal}.
SPEC: -> << [ADJUNCT AQUANT DET NUMBER POSS QUANT SPEC-TYPE].
STMT-TYPE: -> $ { decl header imp int }.
TENSE: -> $ {fut null past pres}.
TIME: -> $ { date day hour minute month season second week year}.
TNS-ASP: -> << [MOOD PERF PROG TENSE].
VTYPE: -> $ {aux copular main modal noncopular predicative raising}.

-----

```

Fonte: <http://www2.parc.com/isl/groups/nltt/xle/doc/PargramStarterGrammar/common.features.lfg>

Neste sistema de anotação, os atributos são todos modelados em letras maiúsculas, enquanto os valores aparecem todos em letras minúsculas, como exemplificado abaixo:

$$(37) \quad \begin{array}{l} \text{NUM=sg} \\ \text{NUM=pl} \end{array}$$

Note que, nas equações acima, NUM corresponde a um atributo número, enquanto o que aparece à esquerda corresponde aos valores destes atributos, no caso *plural* e *singular*.

Uma lista com todos os atributos codificados em nossa gramática é dada na lista de símbolos, no início deste trabalho.

No sistema de anotação estão convencionadas tanto funções gramaticais que são subcategorizadas pelo verbo, quanto as que não são. Nele também está convencionado um conjunto de etiquetas para os especificadores e para os elementos nominais, assim como para os traços adverbiais, adjetivais, preposicionais, verbais e oracionais.

Dentro deste sistema de anotação, é necessário sempre utilizar uma matriz de atributos e valores para codificar tanto as regras da estrutura de constituintes, quanto entradas lexicais.

$$(38) \quad \text{IP} \rightarrow \begin{array}{l} \text{DP: } (^{\wedge} \text{SUBJ})=! \\ \quad (^{!} \text{CASE})=\text{nom}; \\ \quad !^{\wedge}=!. \end{array}$$

Na regra acima, as anotações na regra indicam que a estrutura-f de IP tem um atributo SUBJ, indicado pela metavariável \wedge no nó DP, cujo valor é a estrutura-f do DP filho, indicado pela metavariável $!$ no nó DP. Por outro lado, os símbolos “,” e “.” são metavariáveis utilizadas na plataforma XLE.

Já no exemplo abaixo, temos anotações que são codificadas no léxico e servem para indicar os atributos e valores de uma determinada entrada lexical.

$$(39) \quad \text{damas N} \rightarrow \begin{array}{l} (^{\wedge} \text{PRED})='\text{dama}' \\ \quad (^{\wedge} \text{NUM})=\text{pl} \\ \quad (^{\wedge} \text{GEND})=\text{fem} \\ \quad (^{\wedge} \text{PERS})=3 \\ \quad (^{\wedge} \text{NTYPE NSEM COMMON})=\text{count} \\ \quad (^{\wedge} \text{NTYPE NSYN})=\text{common} \\ \quad (^{\wedge} \text{HUMAN})=+. \end{array}$$

A restrição $(^{\wedge} \text{PRED})='\text{dama}'$ declara que o nó terminal imediatamente dominando o símbolo terminal ‘damas’ tem uma estrutura-f, cujo valor para o atributo PRED é ‘dama’. A entrada *damas* também contém em sua matriz de atributos e valores informações sobre o

número (NUM), a pessoa (PERS), o gênero (GEND), a característica humana (HUMAN), a característica de nome (NTYPE).

Todo o inventário dos traços usados neste trabalho segue a convenção adotada por desenvolvedores de gramáticas que utilizam o XLE como plataforma de desenvolvimento de gramáticas ParGram. No entanto, para não nos fazermos exaustivos nesta seção, vamos apresentando cada atributo na medida em que forem aparecendo na gramática.

Finalmente encerramos esta seção. Em suas três partes nós demonstramos como o sofisticado aparato teórico-computacional da LFG, o sistema de processamento de linguagem natural - XLE e o projeto internacional de desenvolvimento de gramáticas, o Projeto ParGram se integram na elaboração de gramáticas, tanto em escala industrial quanto em pequena escala, para o processamento sintático profundo. A partir deste ponto, temos em mente que o leitor está situado nas três vias de desenvolvimento que norteam o desenvolvimento de nossa gramática.

4 PROCESSAMENTO SINTÁTICO PROFUNDO DO PB

Neste capítulo nós mostramos construções sintáticas da frase do português que devem ser tratadas para que se tenha um fragmento de gramática, que embora ainda pequeno, diante da possibilidade de expansão da gramática, já seja capaz de analisar um grande leque de fenômenos.

Com este intuito, nós procedemos da seguinte maneira. A seção 4.1 discute como os nós mais altos das estruturas arbóreas, os constituintes IP e CP, são modelados como elementos que constituem o nó raiz S. Na seção 4.2, modelamos diversificados e complexos fenômenos verbais como funções gramaticais, predicativos do sujeito, verbos auxiliares, passivas, predicados complexos, modais e verbos de controle. A seção 4.3 apresenta análises de elementos nominais como os pronomes expletivos, demonstrativos, reflexivos e interrogativos. Na seção 4.4, apresentamos uma distinção entre preposições semânticas e não-semânticas, crucial para o tratamento dos sintagmas preposicionados. Na seção 4.5, sobre o sintagma adjetival, tratamos a distinção entre adjetivos atributivos, ordinais, cardinais e adjetivos de intensidade. Finalmente, concluímos este capítulo, demonstrando, na seção 4.6, como advérbios intransitivos e transitivos com complemento PP estão modelados em nossa gramática¹⁵.

4.1 Elementos oracionais

Um tipo de oração pode ser reconhecido pela localização do verbo na sentença, pela morfologia verbal, pela entonação da sentença, pela presença e pela posição de palavras interrogativas, entre outros aspectos. Além disto, os argumentos de um verbo também são úteis para determinar o tipo de estrutura oracional. Além das características oracionais e argumentais, o pragmatismo oracional também desempenha papel para caracterizar os tipos de orações, como por exemplo, permitindo que certos argumentos verbais saltem para fora da oração e se coloquem a frente para topicalizar (BUTT et al. 1999).

¹⁵ Sempre que tivemos dúvidas sobre o formato da distribuição dos atributos de cada regra ou entrada lexical da nossa gramática, fizemos uso da seleção de *corpora* sintaticamente anotados, obtidos através do INESS, que os disponibiliza livremente em (<http://iness.uib.no>). Devido a isto gostaríamos de expressar nosso agradecimento a Rosen et al (2012), responsáveis pela disponibilização dos dados.

4.1.1 Oração principal

Orações matrizes aparecem independentemente. Estas podem ser declarativas e interrogativas. Em nossa gramática, modelamos a oração matriz (S) para que se expanda ou para uma oração declarativa (IP) ou para uma interrogativa (CP).

$$(40) \quad S \rightarrow \{ \{ IP;^=! \\ (! CLAUSE-TYPE)=decl; \\ (PERIOD) \\ | IP;^=! \\ (! CLAUSE-TYPE)=int; \\ (INT) \} \\ | CP;^=! \\ (! CLAUSE-TYPE)=int; \\ (INT) \}.$$

4.1.1.1 Declarativa

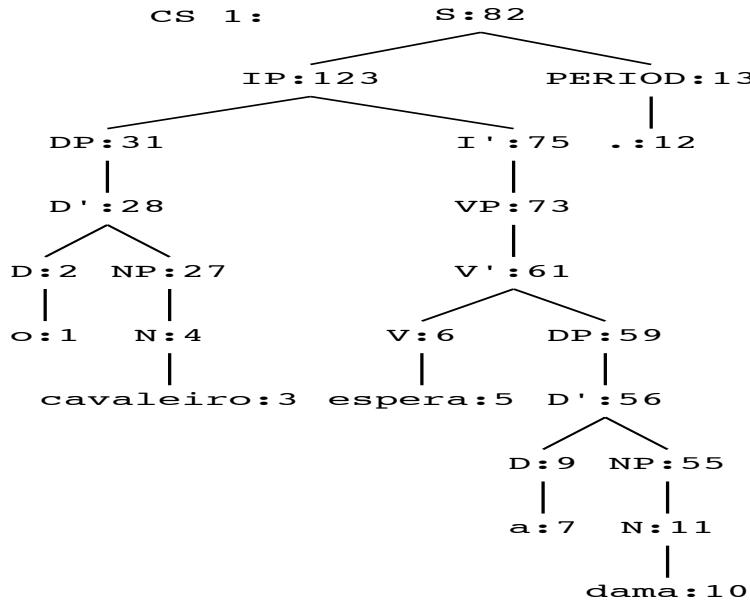
As orações declarativas são normalmente tratadas como as formas oracionais mais simples de uma língua. Nos casos de sentenças declarativas, S se expande para IP.

$$(41) \quad S \rightarrow IP;^=! \\ (! CLAUSE-TYPE)=decl; \\ (PERIOD).$$

Na estrutura-c também está codificada a forma de pontuação que distingue a declarativa da interrogativa.

$$(42) O cavaleiro espera a dama.$$

Figura 14 – Estrutura-c de uma oração declarativa

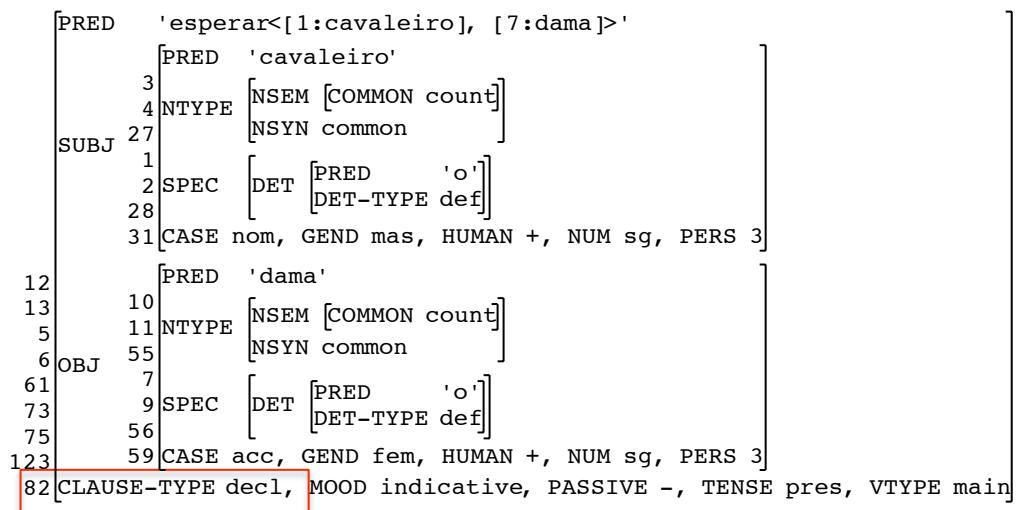


Fonte: elaborada pela autora através do XLE

A estrutura funcional de IP codifica que o tipo da oração (CLAUSE-TYPE) é declarativo (decl).

Figura 15 – Estrutura-f de uma oração declarativa

"o cavaleiro espera a dama ."



Fonte: elaborada pela autora através do XLE

Vimos, acima, a modelação da gramática para codificar as principais características das sentenças declarativas. Na sequência, trataremos da modelação das interrogativas.

4.1.1.2 Interrogativa

Em português e em outras línguas românicas, assim como em línguas germânicas inglês e alemão, há dois tipos de construções gramaticais interrogativas, as interrogativas abertas e as interrogativas fechadas (PERINI, 2010, 2009)¹⁶.

As interrogativas fechadas são comumente entendidas como perguntas que recebem respostas *sim/não*¹⁷. No português existe ainda a possibilidade de respondermos a uma interrogativa fechada utilizando o próprio verbo que foi formulado na pergunta.

- (43) a. **Declarativa**
 Carlos colocou os móveis na despensa.
 b. **Interrogativa fechada**
 Carlos colocou os móveis na despensa?

As orações interrogativas fechadas recebem variação entoacional em relação à sua equivalente declarativa. Elas são marcadas por uma entoação ascendente. Em língua escrita, a única diferença consiste no uso da interrogação¹⁸.

As interrogativas abertas são aquelas que possuem um dos elementos da classe dos interrogativos *o que, quando, como, quem, aonde*, entre outros (PERINI, 2010; MIOTO & KATO, 2005)¹⁹, que interrogam sobre um dos constituintes da oração (BOSQUE e GUTIÉRREZ-REXACH, 2009).

¹⁶ Perini (2010) apresenta ainda as noções de **interrogativa eco**, **interrogativa indireta** e interrogativas com **cadê**. Estas noções não serão relevantes para o nosso trabalho.

¹⁷ O termo original é do inglês: *yes-no questions*. Aqui usamos o termo **fechada**.

¹⁸ Este estudo provê, no entanto, somente uma cobertura parcial dos dados, ignorando aspectos da formação das interrogativas como a entoação.

¹⁹ O termo original é do inglês: *Wh-question*. Aqui usamos o termo **interrogativa-Q**.

O PB é uma língua que se caracteriza por possuir interrogativas-QU com uma tipologia tripartida. São construções do tipo elemento-QU deslocado na periferia esquerda da oração, do tipo elemento-QU *in situ* e do tipo clivada.

4.1.1.2.1 Deslocada

As interrogativas deslocadas do PB são formadas com um único elemento-QU movido para o topo da sentença. Em PB as orações permitem a cobertura de apenas uma lacuna informatacional²⁰.

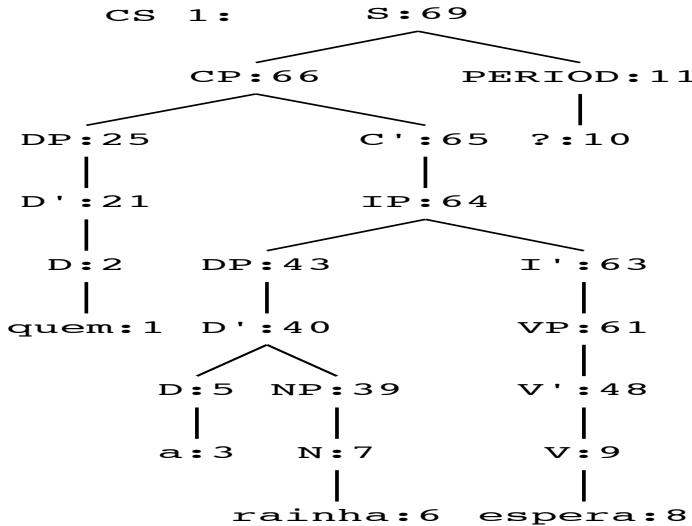
- (44) a. **Declarativa**
 A rainha espera o cavaleiro.
 b. **Deslocada simples**
 Quem a rainha espera?

A estrutura-c das interrogativas deslocadas se diferencia substancialmente da estrutura de constituintes de suas contrapartes interrogativas. Uma destas diferenças é a presença dos sintagmas interrogativos. Se S é uma oração interrogativa, então se expande para CP.

- (45) S --> CP:^{^=!}
 (! CLAUSE-TYPE)=int; "tipo de oração: interrogativa"
 (INT).

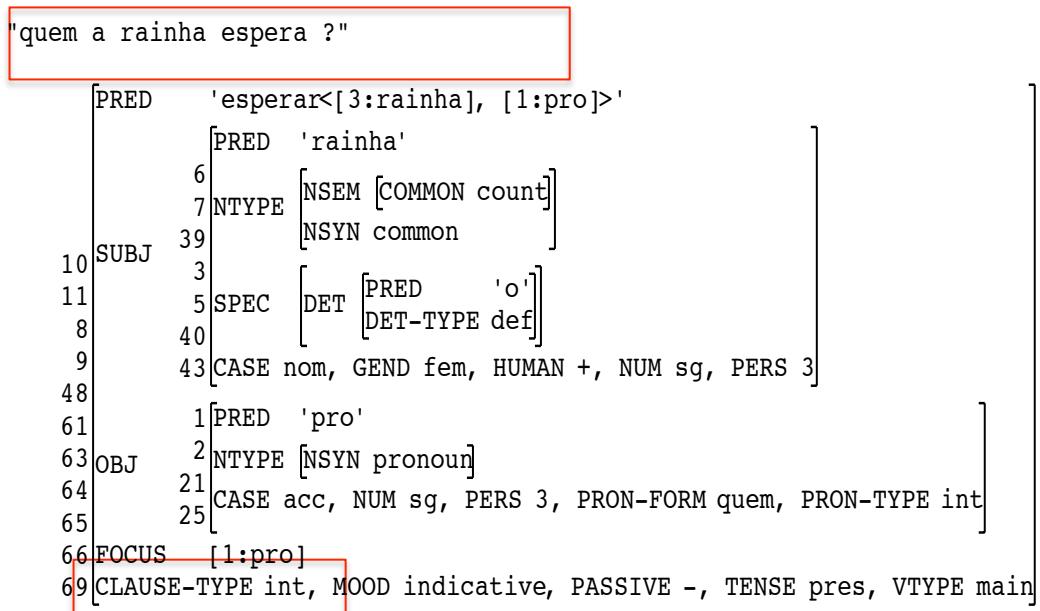
²⁰ Parece que as deslocadas simples podem ser compostas por mais de um elemento interrogativo nos casos de orações coordenadas, como por exemplo, *O que e onde o jovem estuda?*

Figura 16 – Estrutura de constituintes de uma oração interrogativa deslocada



Fonte: elaborada pela autora através do XLE

Figura 17 – Estrutura-f de uma oração interrogativa deslocada



Fonte: elaborada pela autora através do XLE

Note que PRED possui dois argumentos. Os dois estão ligados às funções gramaticais SUBJ e OBJ. Estes dois argumentos satisfazem a estrutura argumental do predicador, mesmo assim a estrutura-funcional da sentença precisa da função discursiva FOCUS para cobrir a lacuna informacional do OBJ (KING, 1995).

Sem coordenação de elementos interrogativos, em PB só é possível construir interrogativas deslocadas simples. Línguas com múltiplas deslocadas permitem a ocorrência de mais de um elemento interrogativo à esquerda da sentença (GAZDIK, 2010). Vejamos, abaixo, um exemplo de deslocada múltipla do húngaro, retirado de Mycock (2007).

(46) **a. Declarativa**

János	bemutat-t-a	Mari-t	Anná-nak
João.NOM	apresentar-ps.3sg	Maria-ACC	Ana-DAT
SUJ	VERBO	OBJ	OBL _{GOAL}
'João apresentou Maria a Ana'.			

b. Interrogativa

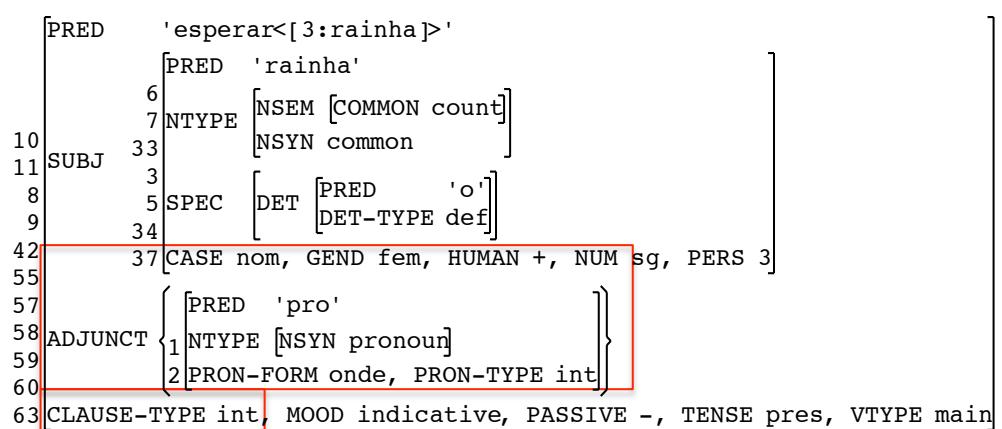
Ki	ki-t	ki-nek	mutat-t-a	be?
Quem.NOM	quem-ACC	a quem-DAT	apresentar-ps.3sg	VM
SUJ	OBJ	OBL _{GOAL}	VERBO	
'Quem apresentou quem a quem?'				

Elementos interrogativos também podem ocorrer como advérbios.

(47) Onde a rainha espera ?

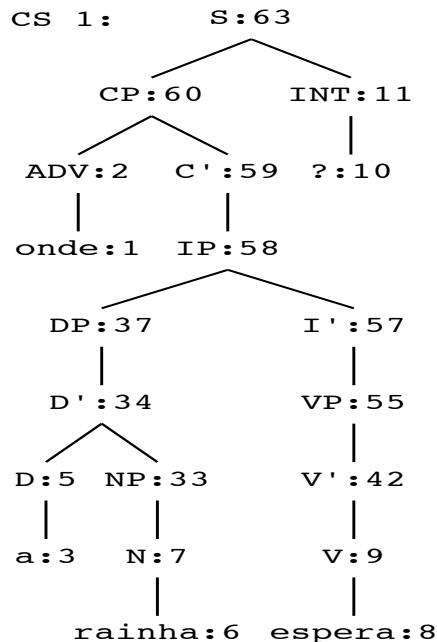
Figura 18 – Estrutura-f de uma oração interrogativa adverbial

"onde a rainha espera ?"



Fonte: elaborada pela autora através do XLE

Figura 19 – Estrutura-c de uma sentença interrogativa adverbial



Fonte: elaborada pela autora através do XLE

São mal-formadas, em português, sentenças interrogativas deslocadas múltiplas. Para serem gramaticais precisariam ser construções com coordenação.

(48) a. **Declarativa**

Quem a rainha espera no prado?

b. **Deslocada múltipla**

*Quem onde a rainha espera?

No exemplo acima vemos que em português as construções interrogativas deslocadas parecem não aceitar os casos de múltiplos elementos na periferia esquerda da sentença, assemelhando-se ao francês (GAZDIK, 2008).

(49) Onde o jovem estuda o quê²¹?

²¹ Pelo meu julgamento, „onde o jovem estuda o que?“ é pouco produtiva, embora não seja marginal.

Caso haja na sentença mais de um constituinte interrogativo e um deles esteja deslocado à esquerda, ou os outros elementos devem aparecer *in situ*, ou coordenados à esquerda

4.1.1.2.2 In situ

Enquanto, entre as línguas, as estratégias de formação das interrogativas diferem, estas estratégias podem compartilhar determinadas características.

Segundo Mycock (2005; 2004), o japonês e o cantonês são línguas *in situ*, i.e. podem ter apenas um ou mais elementos interrogativos ocupando a posição do seu equivalente declarativo. Isto significa que a ordem das palavras de uma sentença declarativa será idêntica a uma interrogativa equivalente será idêntica.

As orações *in situ* podem ser simples ou múltipla. Os morfemas interrogativos podem aparecer *in situ* em orações principais:

(50) **a. declarativa**

A rainha espera o cavaleiro.

b. in situ

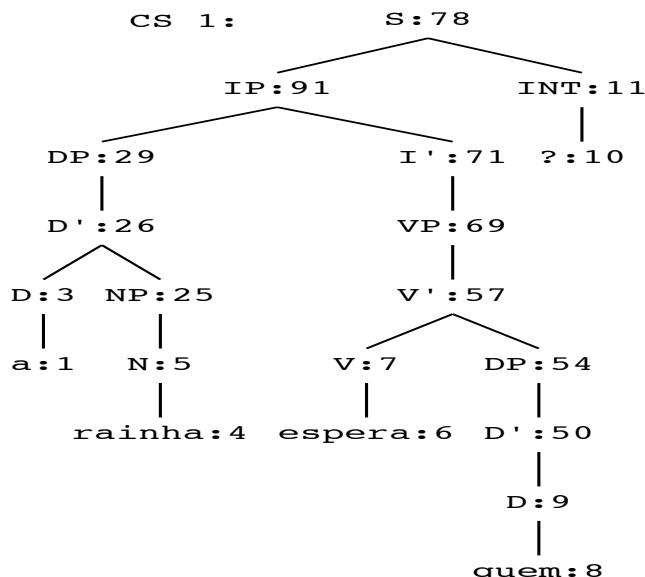
A rainha espera quem?

(51) S --> IP;^{^=!}

(! CLAUSE-TYPE)=int;

(INT)

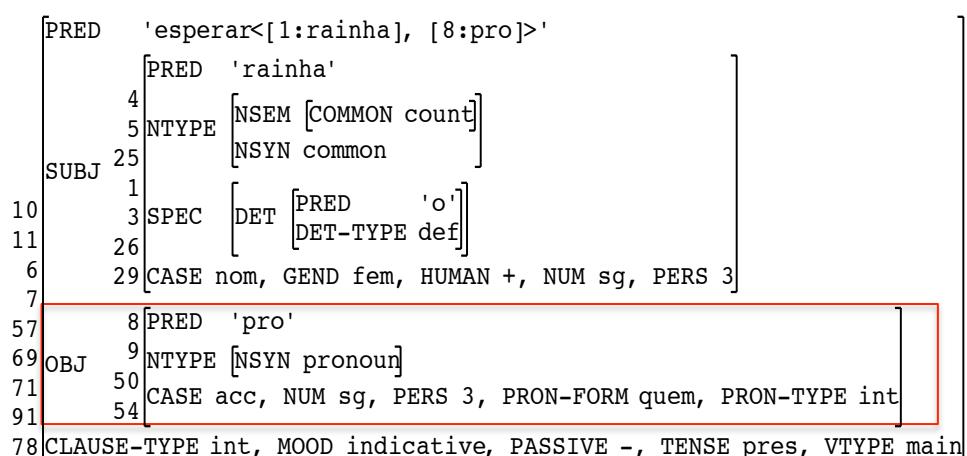
Figura 20 – Estrutura-c de uma sentença interrogativa *in situ*



Fonte: elaborada pela autora através do XLE

Figura 21 – Estrutura-f de uma sentença interrogativa *in situ*

"a rainha espera quem ?"



Fonte: elaborada pela autora através do XLE

Do mesmo modo, os pronomes interrogativos adverbiais também não sofrem restrições para ocorrerem *in situ*.

Figura 22 – Estrutura-f de uma sentença interrogativa adverbial *in situ*
 "a rainha espera onde ?"

PRED	'esperar<[1:rainha]>'	
	PRED 'rainha'	
SUBJ	4 NTYPE [NSEM [COMMON count]]	
	5 NSYN common	
10	25 1 SPEC [DET [PRED 'o']]	
11	3 DET [DET-TYPE def]	
6	26 29 CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3	
7		
38	38 ADJUNCT {PRED 'pro'	
53	53 8 NTYPE [NSYN pronoun]	
55	55 9 PRON-FORM onde, PRON-TYPE int}	
74	74 64 CLAUSE-TYPE int, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main	

Fonte: elaborada pela autora através do XLE

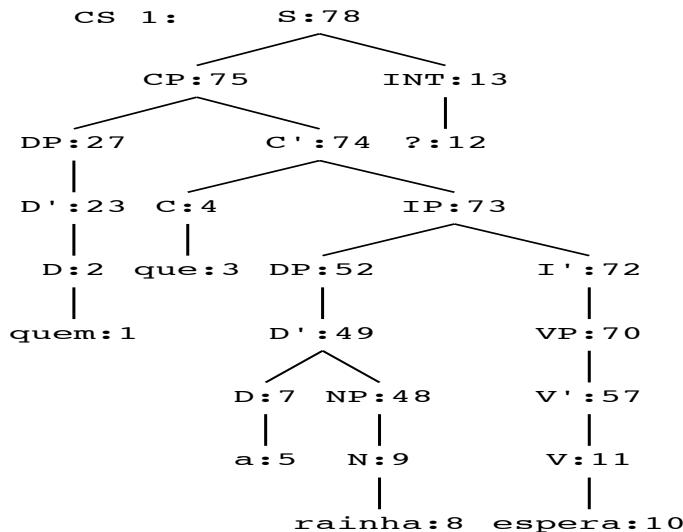
Note que a estrutura-f das duas orações vistas acima se diferenciam das interrogativas deslocadas exatamente por não terem marcação de foco sintático.

4.1.1.2.3 Encaixada

As orações interrogativas-QU deslocadas permitem construções encaixadas com os marcadores interrogativos *que*, *é que* e *que é que* (ALEXANDRE, 2006; KATO et al. 2002; KATO, 1993; MIOTO, 1997; MIOTO E KATO, 2005; NEGRÃO, 2000; OUSHIRO, 2011; SIKANSI, 1998). Vejamos um exemplo.

- (52) Quem que a rainha espera?

Figura 23 – Estrutura-c de uma sentença interrogativa deslocada encaixada



Fonte: elaborada pela autora através do XLE

Note que, na sentença interrogativa acima, temos uma construção encaixada com o marcador interrogativo *que* em posição de especificador de IP. Esta posição pode possuir, ou não, um núcleo explícito. A postulação de um CP acima de um IP na árvore garante um lugar na estrutura de constituintes para a focalização sintática do pronome interrogativo.

Em suma, nós vimos nesta seção que as interrogativas do PB constituem três estratégias distintas de formação. Elas podem ser formadas com um único elemento interrogativo deslocado na periferia esquerda da oração. As *deslocadas* ainda podem ser encaixadas com os elementos complementadores “*que*”, “é *que*”, “*que* é *que*”. E, por fim, também podem ser formadas com um ou mais elemento-QU na mesma posição de sua equivalente declarativa, chamadas *in situ*.

4.1.2 Oração subordinada

As orações subordinadas não ocorrem independentemente. Elas são subcategorizadas por um verbo e por isso se diz que não podem ocorrer sem uma oração matriz (BUTT et al, 1999). Existem dois tipos de complemento verbal oracional. Os complementos que são encabeçados por IPs e os complementos que podem ser encabeçados por CPs.

Inicialmente tratamos do IPs como complemento para, em seguida, falarmos sobre os CPs como complementos.

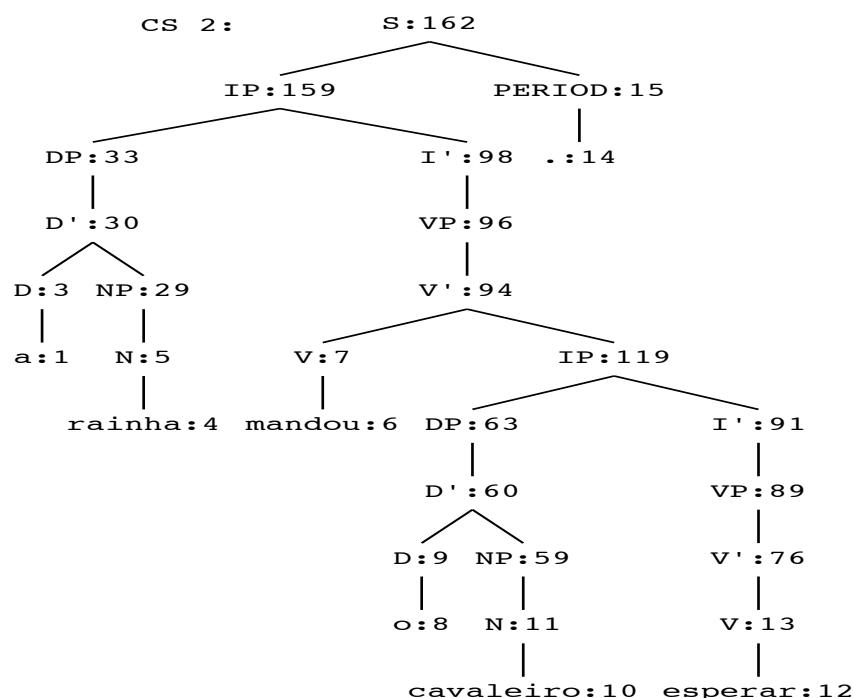
4.1.2.1 IPs como complementos verbais

Em português o verbo principal pode subcategorizar um sintagma flexionado infinitivo ou um sintagma flexionado gerundivo (OTHERO, 2009). Vejamos os seguintes exemplos, utilizados por Schwarze (1998) para o francês:

- (53) A rainha mandou o cavaleiro esperar.
- (54) A rainha viu o cavaleiro esperando.

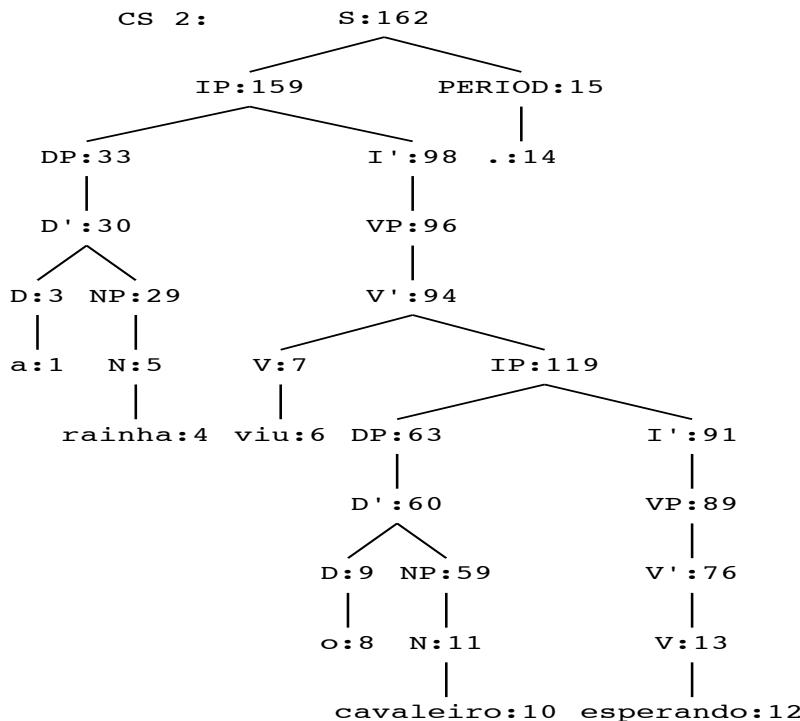
Nas estruturas de constituintes das sentenças acima, o complemento do verbo é um IP com sujeito em posição de especificador (Spec, IP).

Figura 24 – Estrutura-c de uma sentença com complemento oracional infinitivo



Fonte: elaborada pela autora através do XLE

Figura 25 – Estrutura-f com complemento oracional gerundivo



Fonte: elaborada pela autora através do XLE

Para gerar estas análises, nós elaboramos a seguinte regra:

- (55) $V' \rightarrow V;^{\wedge}!=!$
 $(IP: (^{XCOMP})=!$
 $! CLAUSE-TYPE)=nom).$

Nesta regra a anotação em IP licencia um complemento oracional aberto (XCOMP). Adotamos a análise do XCOMP para indicar que o sujeito verbal da oração subordinada é idêntico a um dos argumentos da oração matriz (BUTT et al. 1999).

As entradas lexicais *mandou* e *viu* possuem em sua estrutura-f a seguinte predicação:

- (56) $(^{PRED})='mandar<(^{SUBJ})(^{XCOMP})>'$

Note que o argumento interno da estrutura de predicados possui uma função gramatical que se unifica com o tipo de função gramatical codificado em IP.

Figura 26 – Estrutura-f de uma sentença com complemento oracional infinitivo

"a rainha mandou o cavaleiro esperar ."

PRED	'mandar<[1:rainha], [13:esperar]>'
SUBJ	5[PRED 'rainha' 6 45[NTYPE NSEM [COMMON count] 46[NSYN common 1 4[SPEC [DET [PRED 52[DET-TYPE def 58[CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3]
XCOMP	15[PRED 'esperar<[9:cavaleiro]>' 16 13[PRED 'cavaleiro' 14 92[NTYPE NSEM [COMMON count] 93[NSYN common 8 140 142 145 146 171 228 172 174[PASSIVE -, VFORM infinitive 247 251[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, PERF +, TENSE past, VTYP
	E main]

Fonte: elaborada pela autora através do XLE

Figura 27 – Estrutura-f de uma sentença com complemento oracional gerundivo

"a rainha viu o cavaleiro esperando ."

PRED	'ver<[1:rainha], [13:esperar]>'
SUBJ	5[PRED 'rainha' 6 45[NTYPE NSEM [COMMON count] 46[NSYN common 1 4[SPEC [DET [PRED 52[DET-TYPE def 58[CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3]
XCOMP	15[PRED 'esperar<[9:cavaleiro]>' 16 13[PRED 'cavaleiro' 14 92[NTYPE NSEM [COMMON count] 93[NSYN common 8 140 142 145 146 171 228 172 174[PASSIVE -, VFORM gerundive 247 251[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, PERF +, TENSE past, VTYP
	E main]

Fonte: elaborada pela autora através do XLE

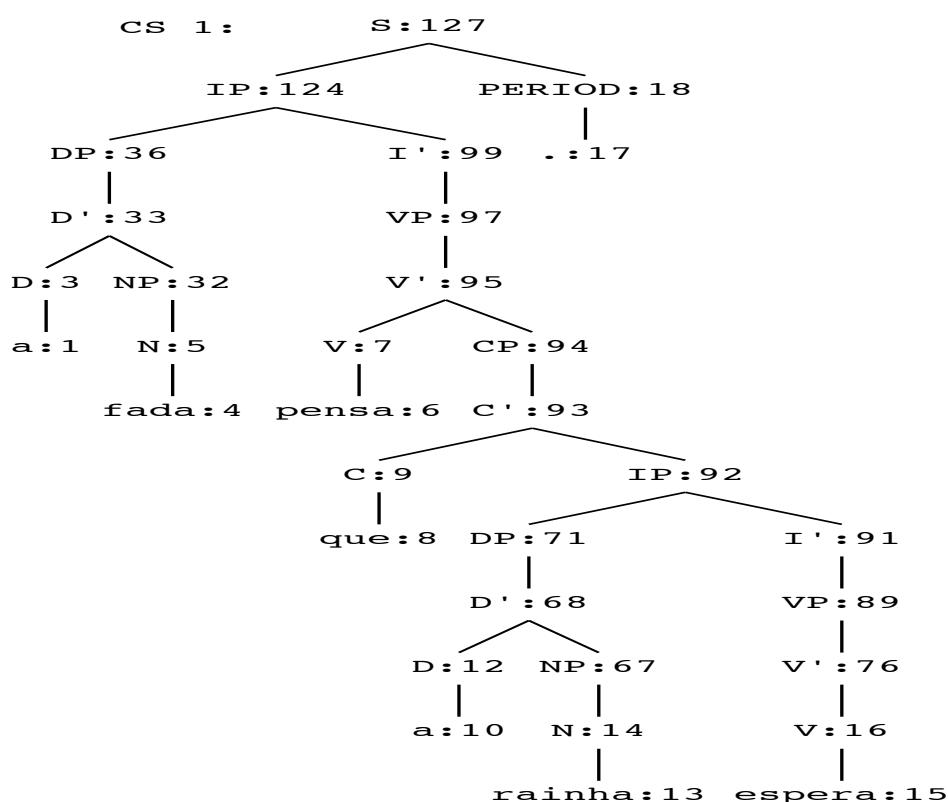
Veja que, nas análises adotadas, as sentenças possuem a mesma estrutura de constituintes e a mesma estrutura funcional. A única diferença entre os dois complementos oracionais são os atributos dos próprios verbos. A estrutura funcional de COMP apresenta esta diferença codificando em VFORM a forma verbal *infinitive* e *gerundive*.

4.1.2.2 CPs como complementos verbais

Na seção anterior vimos que um IP pode ser complemento de verbo. Ademais, também sintagmas complementadores (CPs) podem ocupar esta posição (OTERO, 2009). Vejamos o seguinte exemplo:

- (57) A fada pensa que a rainha espera.

Figura 28 – Estrutura-f de oração subordinada complementadora



Fonte: elaborada pela autora através do XLE

Na estrutura de constituinte da sentença acima, o complemento do verbo é um CP com núcleo preenchido por *que*, codificado na entrada lexical do seguinte modo:

- (58) (^ COMP-FORM)=que.

Para analisar este tipo de sentença, elaboramos as regras abaixo:

- | | | | |
|------|----|-----|--|
| (59) | V' | --> | V: \wedge =!; |
| | | | (CP: (\wedge COMP)=!
(! CLAUSE-TYPE)=decl) |

Nesta regra, a anotação em CP licencia um complemento oracional aberto (COMP), que dispõe de seu próprio sujeito e este não é idêntico a nenhum argumento da oração matriz (BUTT et al. 1999).

Em resumo, com estas regras sintagmáticas e lexicais nossa gramática analisa verbos que podem subcategorizar orações subordinadas. As orações subordinadas podem ocupar a posição de IP ou de CP na estrutura de constituintes. Nossas regras analisam os complementos oracionais infinitivos e gerundivos na posição de IP e orações complementadoras que dispõem de sujeito próprio ou possuem sujeito controlado pelo sujeito da oração principal.

4.2 Elementos verbais

Esta seção, a maior deste capítulo, envolve diversificados fenômenos relacionados com o eixo verbal de uma sentença. Nela serão abordadas, na acepção da LFG, funções gramaticais, predicativos, auxiliares, passivas, predicados complexos com verbos modais e verbos de controle.

4.2.1 Funções gramaticais

Seguindo as análises da LFG adotadas por Butt et al. (1999), nossa gramática

dispõe do seguinte inventário de funções gramaticais: SUBJ(ect), OBJ(ect), OBJ2(objeto secundário), OBL(íquo), complemento oracional fechado (COMP), complemento oracional aberto (XCOMP) e ADJ(unct).

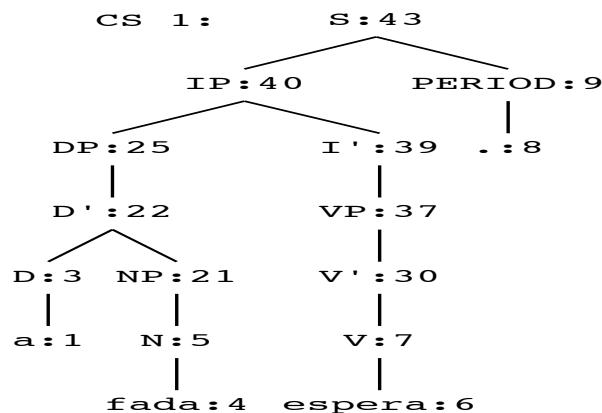
4.2.1.1 Sujeito

Em português, um SUBJ pode ser identificado em uma oração pelo papel temático associado ao DP presente na frase ou pode ter seu referente identificado pelo sufixo de pessoa-número do verbo. Para identificarmos um sujeito, este tem que ser um constituinte com potencial referencial, em termos sintáticos, um DP. A condição prévia é que a pessoa e o número do DP sujeito sejam compatíveis com a pessoa e o número indicados pelo sufixo de pessoa-número (PERINI, 2008; PERINI, 2010). Vejamos o exemplo abaixo:

- (60) A fada espera

Em nossa análise o sujeito ocupa o sintagma determinante na posição de SpecIP.

Figura 29 – Estrutura-c de sentença com verbo intransitivo



Fonte: elaborada pela autora através do XLE

A regra da gramática para esta análise é a seguinte:

- (61) IP --> DP: (^ SUBJ)=!
(! CASE)=nom; "atribui caso nominativo"

I';^=!

Nesta regra IP se expande para um DP anotado com a função de SUBJ e com a marcação de caso nominativo.

O fenômeno gramatical da concordância ligado a esta função gramatical é capturado pelos atributos de número e pessoa codificados na entrada lexical *espera*.

- (62) (^ SUBJ NUM)=sg
 (^ SUBJ PERS)=3

Como vimos, a concordância verbal é uma boa indicação na identificação do sujeito em PB: se um NP concorda com o verbo, então este NP é o sujeito.

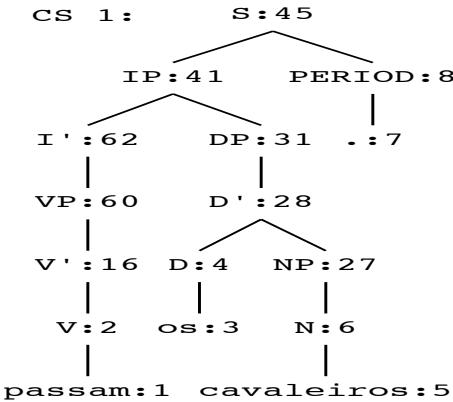
- (63) Passam os cavaleiros.

Na análise do sujeito posposto, a regra do IP incorporada na nossa gramática sofre uma inversão na ordem dos sintagmas.

- (64) IP --> I';^=!;
 (DP: (^ SUBJ)=!
 (! CASE)=nom).

Esta regra permite analisar o DP que vem após o verbo como sujeito, em uma variedade padrão do PB.

Figura 30 – Estrutura-c para sentença com sujeito pós-verbal



Fonte: elaborada pela autora através do XLE

Nas regras lexicais da gramática é inserida a entrada lexical *passam* subcategorizando um único argumento.

- (65) passam V * (^ PRED)=^P<(^ SUBJ)>

Tanto o estatuto funcional anotado no DP como os traços de concordância correspondem às exigências do verbo intransitivo *passam*.

4.2.1.2 Objeto

Tipicamente os argumentos verbais mais ligados ao verbo são tratados como OBJ. Este segundo argumento é usualmente um OBJ. Um exemplo de um verbo transitivo que subcategoriza um SUBJ e um OBJ é mostrado em (66).

- (66) A dama espera o cavaleiro

Em português, em geral, a posição do objeto é uma boa indicação desta função gramatical, pois este geralmente segue o verbo e é adjacente a ele.

- (67) (^ PRED)=^esperar<(^ SUBJ) (^ OBJ)>

4.2.1.3 Objeto secundário

Verbos bitransitivos subcategorizam três argumentos. Um exemplo de um verbo bitransitivo que subcategoriza um SUBJ, um OBJ e um OBJ2 é mostrado em (68).

- (68) A dama dá seu lenço ao cavaleiro

O típico verbo bitransitivo, tal qual o verbo *dar*, subcategorizará um sujeito, um objeto e um objeto mais distante do verbo, que chamamos de objeto secundário. Isto reflete a ideia de que a relação do verbo com seu sujeito e seus objetos é mais próxima, ou tem mais significação sintática, que a relação do verbo com outros elementos da oração (KROEGER, 2004; FALK, 2001). Normalmente NPs que podem ser substituídos por um dativo clítico são tratados como OBJ2, como ilustrado em (69b).

- (69) a. A dama dá seu lenço ao cavaleiro
 b. A dama deu-lhe seu lenço

Assim, objetos tematicamente restritos para línguas que permitem dois objetos de ocorrer ao mesmo tempo são tratados como objetos secundários, mas, ao invés de serem codificados como OBJ2, em Pargram, é codificado como OBJ-TH, mais genérico e compatível com LFG²².

Assim, introduz-se na entrada lexical de *dá* a equação $(^{\wedge} \text{PRED})='\text{dar}' < (^{\wedge} \text{SUBJ})(^{\wedge} \text{OBJ})(^{\wedge} \text{OBJ-TH})>$.

²² Thematically restricted objects for languages which allow two objects to occur at once ("we gave him a book"); these are the same thing as secondary objects (OBJ2) but in ParGram the more generic and more LFG compliant OBJ-TH is used instead (KING, 2004)

4.2.1.4 Oblíquo

Os objetos que se relacionam menos diretamente com o verbo são chamados de OBLíquo (KROEGER, 2004).

- (70) A rainha doou sua fortuna para o cavaleiro.

Para o cavaleiro é analisado como um OBL em português. OBL são geralmente PPs, se assemelham aos adjuntos pela forma e podem ser diferenciados deles somente por serem requeridos pelo predicado.

- (71) (^ PRED)=*'doar<(^ SUBJ)(^ OBJ)(^ OBL)>'*

Figura 31 – Estrutura-f de sentença com verbo bitransitivo

"a rainha doou sua fortuna para o cavaleiro ."	[PRED 'doar<[1:rainha], [8:fortuna], [12:cavaleiro]>'
	[PRED 'rainha'
SUBJ 31	4 [NTYPE [NSEM [COMMON count]]]
	5 [NSYN common]
	1 [SPEC [DET [PRED 'o']]
	3 [DET-TYPE def]]]
	32 [CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3]
	35]]
	[PRED 'fortuna'
OBJ 51	10 [NTYPE [NSEM [COMMON count]]]
	11 [NSYN common]
	8 [SPEC [POSS [PRED 'pro']
	9 [PERS 3, PRON-FORM seu, PRON-TYPE poss]]]
	52 [ANIM -, CASE acc, GEND mas, HUMAN -, NUM sg, PERS 3]
	55]]
	16 [PRED 'cavaleiro'
OBL 18	17 [NTYPE [NSEM [COMMON count]]]
	18 [NSYN common]
	14 [SPEC [DET [PRED 'o']]
	15 [DET-TYPE def]]]
	69 [CASE dat, GEND mas, HUMAN +, NUM sg, PERS 3, PFORM para, PTTYPE nosem]
	74 [12 [CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, PERF +, TENSE past, VTYPE main]
	78 [13 [CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, PERF +, TENSE past, VTYPE main]
	80 [14 [CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, PERF +, TENSE past, VTYPE main]
	81 [73 [CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, PERF +, TENSE past, VTYPE main]
	84]]

Fonte: elaborada pela autora através do XLE

4.2.1.5 *Adjuntos*

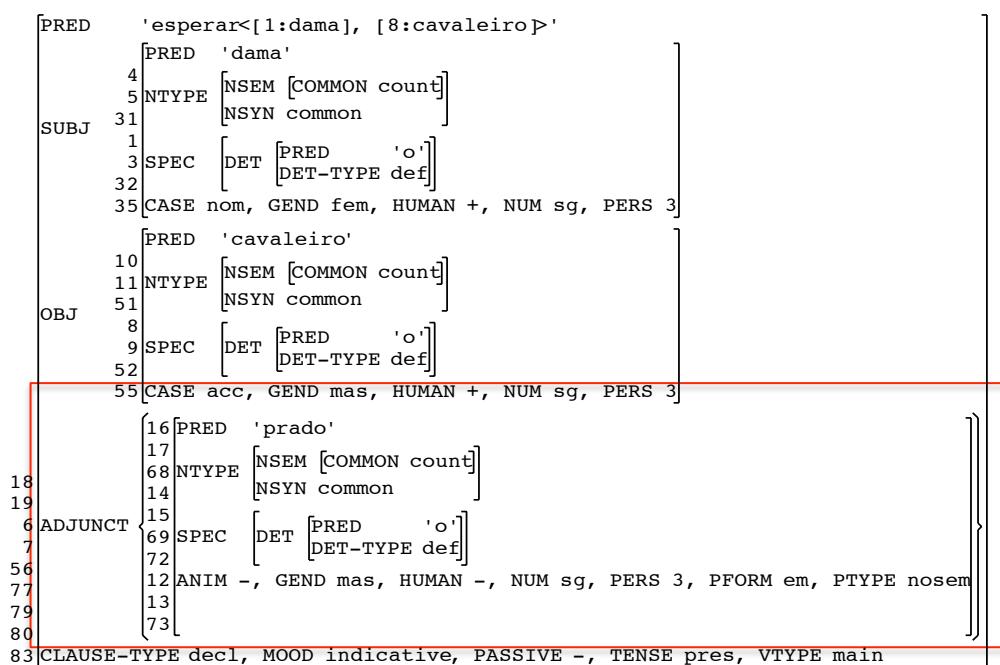
Adjuntos são funções gramaticais que não são subcategorizadas pelo verbo. Enquanto os argumentos são frequentemente obrigatórios, estes elementos são sempre opcionais, mas podem ser multiplicados livremente (BUTT et al. 1999; KROEGER, 2004).

Adjuntos são analisados como pertencendo a um grupo, que pode ocorrer com qualquer PRED. Desse modo, (40) terá a estrutura-f logo abaixo.

- (72) a dama espera o cavaleiro no prado.

Figura 32 – Estrutura-f de sentença com adjunto

"a dama espera o cavaleiro em o prado ."



Fonte: elaborada pela autora através do XLE

4.2.1.6 XCOMP e COMP

Temos visto, até aqui, argumentos verbais representados na estrutura de constituintes como NPs e PPs. Os argumentos não precisam ser, exclusivamente, sintagmas nominais ou preposicionais. Sentenças inteiras também podem servir como complementos do

verbo. Na esteira de Butt et al. (1999), adotamos a análise dos argumentos oracionais que os distinguem de duas diferentes maneiras, quais sejam, XCOMP e COMP.

Adotamos a análise do XCOMP para tratar os argumentos oracionais, cujo sujeito possui obrigatoriamente uma relação de controle funcional com algum dos argumentos externos à sua estrutura.

Na análise de COMP, o argumento oracional é um complemento que possui seu próprio sujeito. Neste casos, não há controle funcional. O verbo da oração encaixada possui seu próprio sujeito.

(73) A rainha disse que o cavaleiro espera a dama.

No exemplo acima o verbo *disse* subcategoriza dois argumentos: o SUBJ *a rainha* e o COMP *que o cavaleiro espera a dama*.

Para representar as suas propriedades argumentais, a entrada lexical *disse* possui a seguinte equação para representar:

(74) $(^{\text{PRED}})=\text{'dizer} < (^{\text{SUBJ}})(^{\text{COMP}}) >'$

Na análise do COMP, não é necessário codificar que o sujeito da subordinada deve ser controlado por um dos argumentos da oração principal.

Nos casos de COMP, um aspecto bastante interessante a ser considerado é regência exercida pelo verbo principal sobre o verbo encaixado.

(75) A dama quer que a fada espere.

Na sentença acima, o verbo principal no modo indicativo rege uma oração complementar com verbo no modo subjuntivo²³. As análises de Othero (2009) nos mostram que este fenômeno não pode ser capturado apenas com regras de constituintes.

²³ Mesmo cientes da existência de construções como „você quer que eu vou à padaria“, variante utilizada na cidade de São Paulo, neste trabalho estamos analisando apenas a variante não marcada do português.

Sendo assim, o modo do verbo encaixado é uma atribuição do verbo principal, que codificamos nas regras lexicais da nossa gramática do seguinte modo:

Já na sentença abaixo, o verbo principal no modo subjuntivo rege uma oração complementar com verbo no modo indicativo.

- (77) A dama saiba que a fada espera.

Neste exemplo, também temos um caso de argumento oracional, só que desta vez o verbo da oração subordinada está no modo indicativo. Na gramática, isto é codificado na estrutura-f do verbo principal do seguinte modo:

- (78) $(^{\text{PRED}})=$ 'acreditar $<(^{\text{SUBJ}})(^{\text{COMP}})>$ '
 $(^{\text{COMP MOOD}})=\text{indicative}$
 $(^{\text{MOOD}})=\text{subjunctive}$

Para analisar as sentenças acima não precisamos elaborar nenhuma regra sintagmática, pois como vimos, o que determina o modo do verbo encaixado é uma atribuição da própria entrada lexical. Para analisar o modo verbal das sentenças encaixadas, precisamos apenas inserir esta distinção nas entradas lexicais.

4.2.2 Predicativos

Nesta seção trataremos de dois tipos distintos de predicativos. O caso das construções com verbos copulares e sintagma nominais e o caso dos verbos copulares com sintagmas adjetivais.

4.2.2.1 PREDLINK

Em nossa análises, consideramos como construções predicativas aquelas sentenças que possuem verbos copulares como predicadores. Estes predicadores possuem um sujeito e um outro argumento, que pode ser de várias categorias sintagmáticas (BUTT et al., 1999).

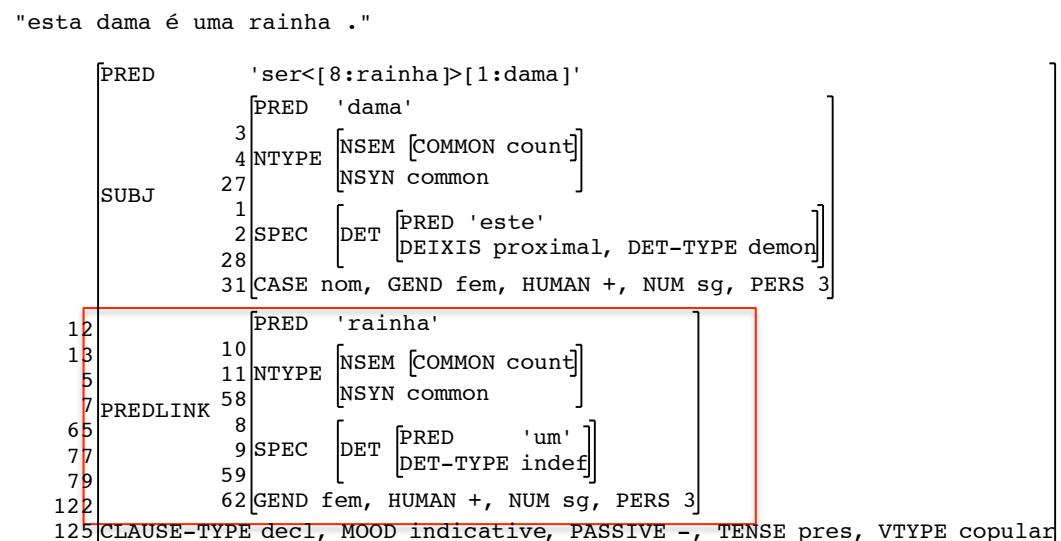
- (79) a. Esta dama é uma rainha.
 b. *Esta dama uma rainha.

O exemplo em (79a) apresenta um constituinte pós-verbal da categoria DP como elemento pós-copular. Em contraste, vemos em (79b) que em uma sentença copular com complemento nominal, o elemento copular não pode ser omitido. Este fato sugere que nomes em português não são complementos abertos e não subcategorizam um SUBJ.

Para esta sentença, adotamos a análise PREDLINK. Nesta análise, o DP complemento é um complemento fechado em posição predicativa.

- (80) (^ PRED)=*'ser<(^ PREDLINK)>(^ SUBJ)'*

Figura 33 – Estrutura-f de sentença com Predlink



Fonte: elaborada pela autora através do XLE

Observe que, nesta análise, PREDLINK faz referência ao próprio SUBJ da sentença, sendo, por isso, um predicado atributivo.

4.2.2.2 Análise do sujeito controlado – XCOMP-PRED

Nesta análise o sujeito do verbo de ligação de uma dada sentença controla o sujeito do sintagma pós-verbal. A sentença com XCOMP-PRED (ROSÉN et al., 2012) é aquela que apresenta um AP como elemento pós-copular, como mostramos na seguinte frase:

- (81) A fada é amável

Adotamos, para esta frase, a análise do complemento aberto em posição predicativa. Nesta análise, o predicado oracional copular *amável* seleciona um sujeito, que é funcionalmente controlado pelo sujeito principal da sentença.

Nesta análise, XCOMP-PRED é uma categoria aberta. É definida uma equação de controle que define a relação entre o sujeito da sentença e o sujeito de XCOMP-PRED e, por isto, permite a unificação do sujeito de XCOMP-PRED com o sujeito da sentença (SULGER, 2009; BRESNAN, 2001; BUTT et al, 1999).

XCOMP-PRED equaciona o sujeito da cópula com o sujeito do constituinte pós-cópula. Assim, introduz-se na entrada lexical de *é*, a equação $(^ \text{XCOMP-PRED} \text{SUBJ}) = (^ \text{SUBJ})$, que diz que o sujeito do complemento adjetival *amável* é igual ao sujeito de *é* (DALRYMPLE, 2004; SULGER 2009).

- (82) é $(^ \text{PRED})='\text{ser}' <(^ \text{XCOMP-PRED}) > (^ \text{SUBJ})'$
 $(^ \text{XCOMP-PRED} \text{SUBJ}) = (^ \text{SUBJ})$

Na análise do XCOMP-PRED, o adjetivo simplesmente concorda com seu próprio SUBJ, da mesma maneira que faz o verbo. No caso, o adjetivo *amável* está em posição predicativa, concordando com o sujeito da oração principal. O significado da cópula consiste, essencialmente, em atribuir ao sujeito o predicado do seu complemento.

Figura 34 – Estrutura-f de sentença com XCOMP-PRED

"a fada é amável ."

PRED	'ser<[9:amável]>[1:fada]'	
	5 PRED 'fada'	
	6	
	41 NTYPE [NSEM [COMMON count]]	
SUBJ	42 [NSYN common]	
	1	
11	4 SPEC [DET [PRED 'o']]	
12	48 [DET-TYPE def]	
7	54 CASE nom, GEND fem, HUMAN -, NUM sg, PERS 3	
8		
88	9 PRED 'amável<[1:fada]>']	
108 XCOMP-PRED	10 SUBJ [1:fada]	
126	82 ATYPE predicative, DEGREE positive	
137	86	
141 CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE copular		

Fonte: elaborada pela autora através do XLE

Isto é expresso na entrada lexical de *amável* do seguinte modo:

(83) $(^{\text{PRED}})=\text{'amável'} < (^{\text{SUBJ}})>$

Com esta equação codificamos que o sujeito da cópula também é o sujeito do elemento embutido no constituinte pós-copular.

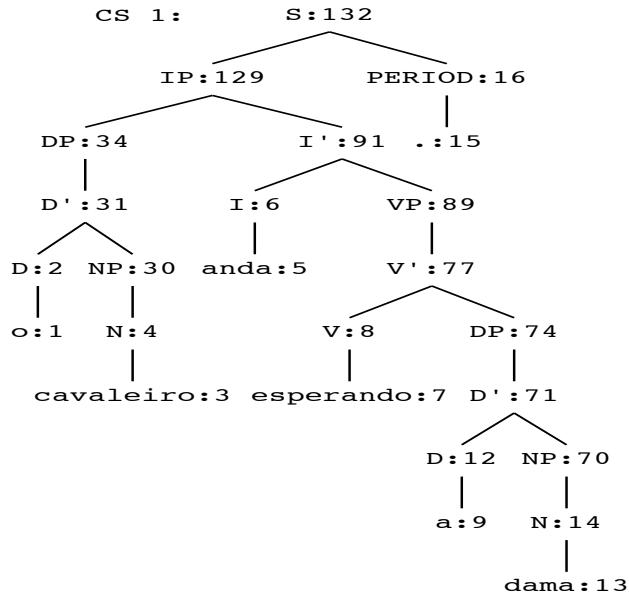
4.2.3 Auxiliares

Um importante aspecto das sentenças declarativas são as construções com verbos auxiliares. Para Perini (2010) a combinação de alguns verbos com outros verbos em forma gerundiva, infinitival ou participial não altera a sequência semântica e valencial do verbo em sua forma simples.

- (84) O cavaleiro anda esperando a dama.
(85) A dama vai esperar a rainha.
(86) A rainha havia esperado a fada.

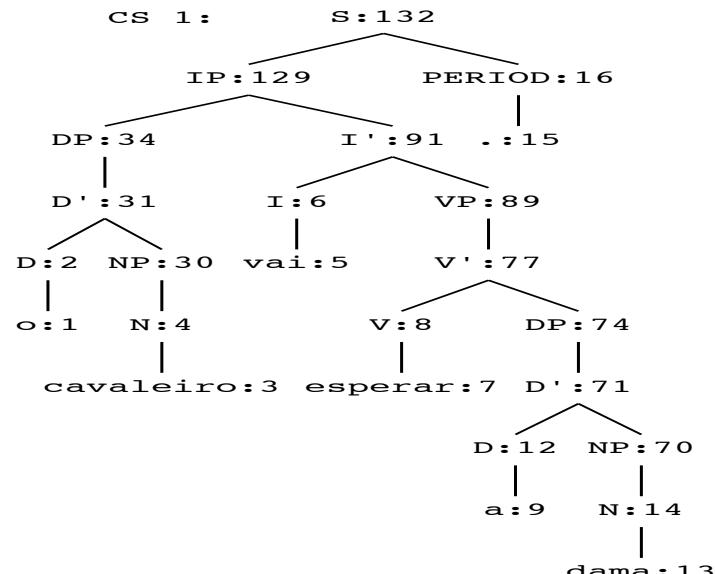
Auxiliares, em LFG, são analisados como tendo um estrutura-f rasa e nenhum predicado (BUTT et al. 1999).

Figura 35 – Estrutura-c de sentença com verbo auxiliar mais gerúndio



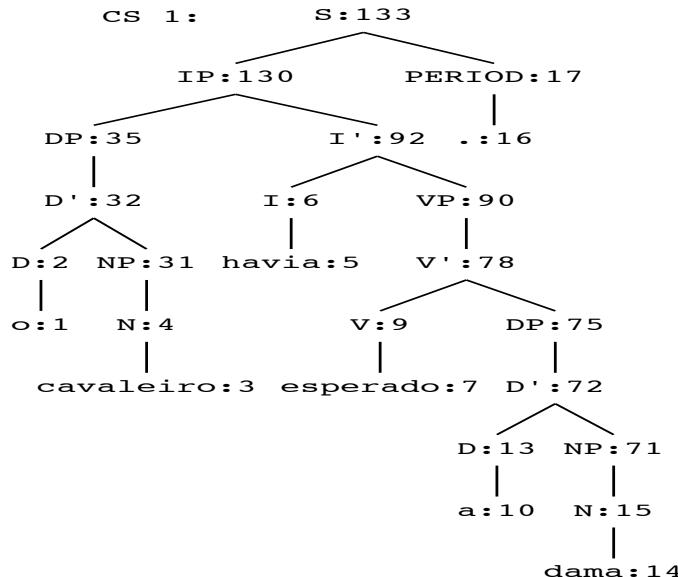
Fonte: elaborada pela autora através do XLE

Figura 36 – Estrutura-c de verbo auxiliar mais infinitivo



Fonte: elaborada pela autora através do XLE

Figura 37 – Estrutura-c de verbo auxiliar mais particípio



Fonte: elaborada pela autora através do XLE

Nestas estruturas o verbo auxiliar ocupa a posição de I e o verbo principal é o núcleo de VP. Em tais construções o verbo auxiliar atua como um elemento funcional e como tal, não apresenta alterações nem semânticas, nem valenciais (OTHERO, 2009).

Note que os verbos auxiliares das três estruturas possuem marcas de tempo diferentes. Outro aspecto importante é distinguir um verbo pleno de um verbo auxiliar. Para ambos os aspectos, a distinção é caracterizada como um traço nos itens lexicais.

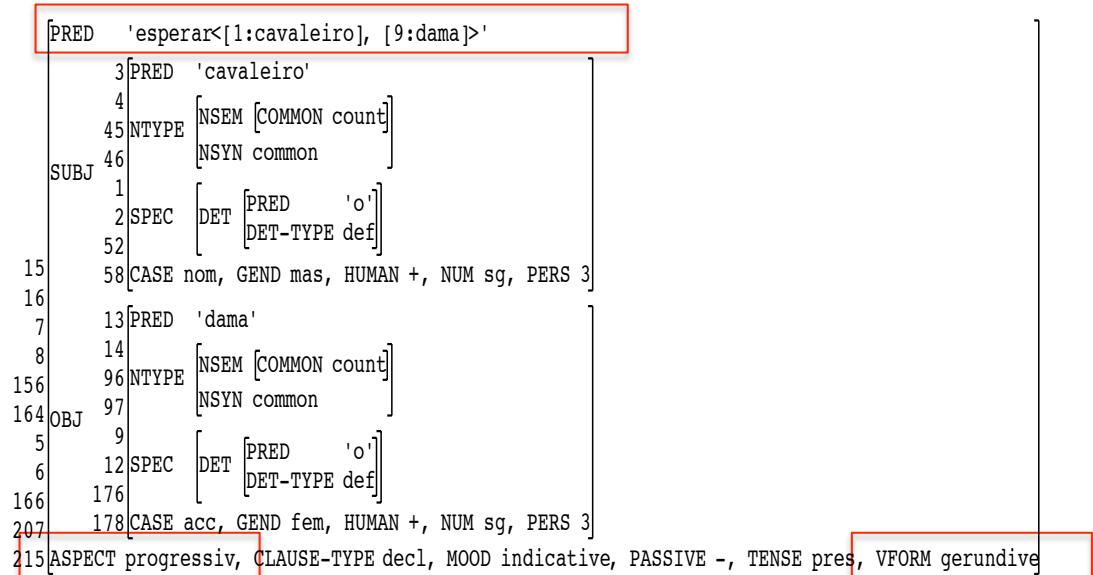
Em I codificamos a marca de tempo e em V codificamos a forma verbal. Vejamos como inserimos estas informações nas entradas lexicais da gramática:

- (87) a. anda I * (^ TENSE)=pres
 b. vai I * (^ TENSE)=fut
 c. havia I * (^ TENSE)= past

- (88) a. esperado V * (^ VFFORM)=participle
 b. esperar V * (^ VFFORM)=infinitive
 c. esperando V * (^ VFFORM)=gerundive

Figura 38 – Estrutura-f de sentença com verbo auxiliar e verbo principal

"o cavaleiro anda esperando a dama ."



Fonte: elaborada pela autora através do XLE

Observe que nesta análise *anda* colabora com o significado da sentença fornecendo o valor aspectual ASPECT *progressiv*, mas a estrutura argumental vem do predicador *esperar*.

4.2.4 Passivas

A passivação é, primariamente, um realinhamento das relações gramaticais, sempre envolvendo uma mudança do agente de sujeito ativo para um oblíquo passivo, e normalmente envolvendo a promoção do paciente de objeto ativo a sujeito passivo (KROEGER, 2004). No entanto, ela não afeta a estrutura argumental da sentença (PERINI, 2008, 2010).

Este fenômeno envolve uma mudança no mapeamento dos argumentos da sintaxe. Pode-se dizer que esta mudança é operada na estrutura argumental (FALK, 2001). Nesta perspectiva, alguns elementos podem ser mapeados da estrutura argumental para a estrutura funcional sofrendo mudança de função gramatical.

- (89) a. A fada espera um cavaleiro.
 b. O cavaleiro é esperado por uma fada.

Esta mudança é refletida na estrutura argumental da seguinte maneira:

- (90) a. Esperar <agente, paciente>
- | | | | |
|------|--|-----|--|
| | | | |
| SUBJ | | OBJ | |
- b. É esperado <agente, paciente>
- | | | | |
|-------|--|------|--|
| | | | |
| (OBL) | | SUBJ | |

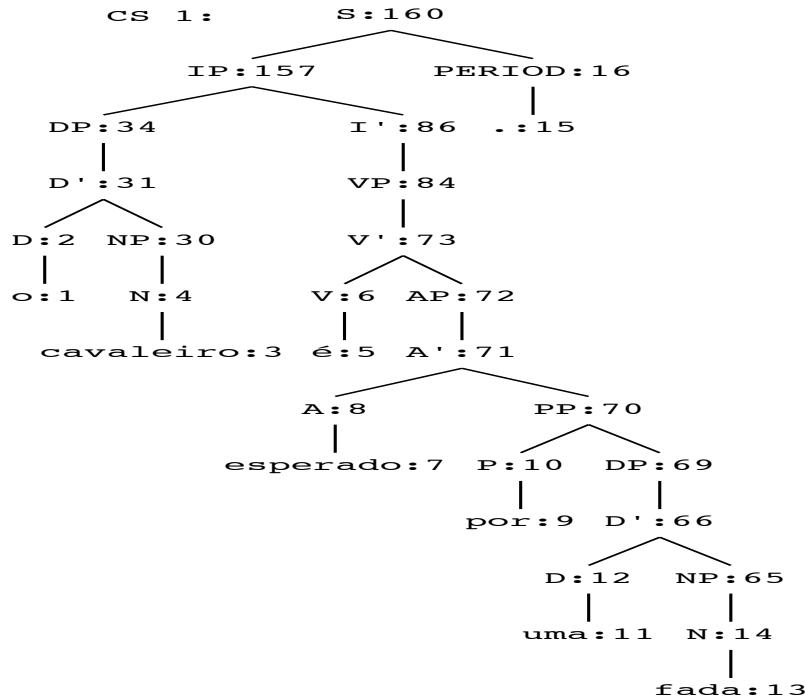
Os parênteses do OBL demonstram a possibilidade de supressão deste argumento (BUTT et al. 1999). Isto, no entanto, não significa mudança na estrutura argumental do verbo, pois cada verbo tem uma valência própria, que funciona qualquer que seja a forma assumida pelo verbo (PERINI, 2010) e por se tratar de um particípio verbal²⁴ deve ter sua estrutura argumental mantida, embora possa haver a supressão do elemento oblíquo agentivo.

Na acepção de Othero (2009), as passivas possuem estruturas semelhantes à estrutura das sentenças com verbos auxiliares, com o elemento copular ocupando a posição de núcleo do sintagma flexionado. Em sua análise, a forma participial do verbo ocupa posição de núcleo de VP. O PP agente da ação ocupa posição de adjunto, no nível intermediário da projeção verbal. Não adotamos esta análise, pois, no nosso entendimento, este reordenamento sintático da estrutura de constituintes causa uma mudança no mapeamento dos argumentos da sintaxe, afetando a estrutura argumental da sentença.

Essa abordagem, digamos tradicional, não é a adotada em nossa modelagem. Neste trabalho adotamos uma nova proposta de análise para desenvolvermos nossa gramática. Tratam-se de análises apresentadas por Rosén et al. (2012), nas quais a cópula ocupa, na estrutura de constituintes, a posição de núcleo do sintagma verbal. A voz passiva é representada por sintagma adjetival, que funciona como o predicativo da cópula.

²⁴ Assumimos a distinção feita por Perini (2010, p. 176) entre particípio nominal e particípio verbal.

Figura 39 – Estrutura-c de uma sentença com apassivação do verbo



Fonte: elaborada pela autora através do XLE

Na estrutura-c acima, o núcleo do constituinte adjetival possui uma estrutura-f responsável em fazer este mapeamento argumental na estrutura de constituintes. Para capturar a estrutura argumental do predicativo da cópula, codificamos na estrutura funcional da entrada lexical *esperado* a seguinte predicação:

- (91) @(**PASS (^ PRED)=*'P<(^ SUBJ)(^ OBJ)>'***)
 (^ PASSIVE)=+
 (^ DEGREE)=positive
 (^ ATYPE)=predicative

Em uma gramática LFG-XLE o esquema da passiva (**PASS(_SCHEMATA)**) é codificado do seguinte modo:

- (92) **PASS(_SCHEMATA)=** { _SCHEMATA (^ PASSIVE)=-
 | _SCHEMATA (^ PASSIVE)=c +
 { (^ SUBJ) --> NULL

$|(^{\wedge} \text{SUBJ}) \rightarrow (^{\wedge} \text{OBL-AG})$
 $\quad @(\text{OT-MARK OblAg})\}$
 $(^{\wedge} \text{OBJ}) \rightarrow (^{\wedge} \text{SUBJ})\}.$

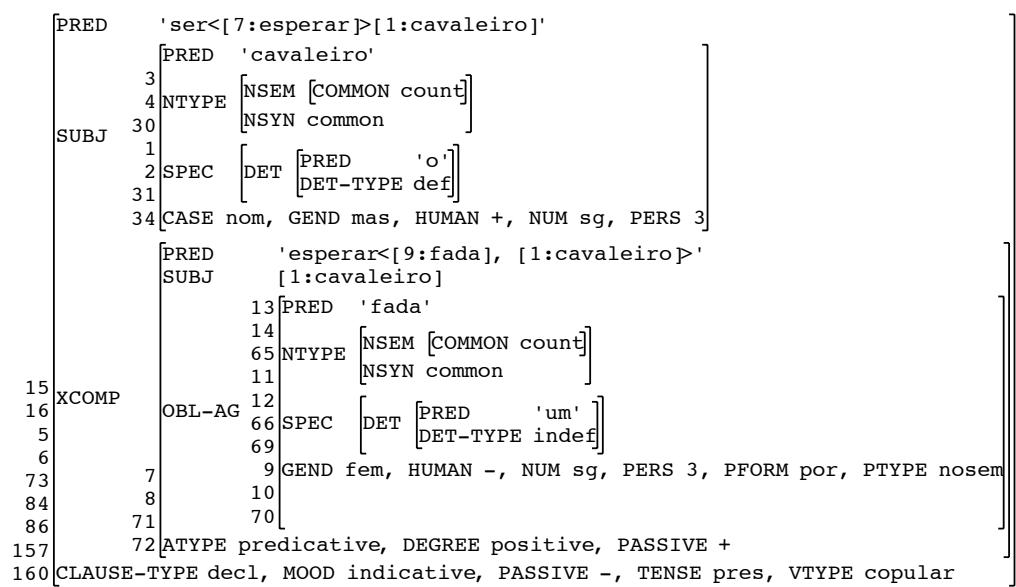
O template PASS pega um predicado como:

(93) $(^{\wedge} \text{PRED}) = 'esperar < (^{\wedge} \text{SUBJ})(^{\wedge} \text{OBJ}) >'$

e pode reescrever o sujeito ou como NULL ou como um oblíquo agente (OBL-AG). Há uma marca de optimalidade (OT mark) na disjunção que cria o OBL-AG; dada a classificação OT na BASIC PORT CONFIG (1.0), esta irá resultar em sintagmas oblíquos “por” nas passivas, sendo preferida sobre os adjuntos. A regra então reescreve o objeto como sujeito (KING, 2014).

Figura 40 – Estrutura-f de sentença em forma apassivada

"o cavaleiro é esperado por uma fada ."



Fonte: elaborada pela autora através do XLE

Observe que esta análise recupera, do constituinte adjetival, a propriedade de subcategorização do verbo *esperar*.

4.2.5 Predicados complexos

Estamos nos referindo a complexos verbais, aquelas construções em que há um agrupamento verbal, tanto aqueles em que há mais de um núcleo realizado na sentença, formando predicados complexos, quanto àquelas construções possuidoras de mais de um predicador, mas que não formam predicados complexos.

4.2.5.1 Verbos modais

O complexo verbal abaixo se constitui de um verbo modal e uma outra verbal infinitiva, formando um predicado complexo (Rech, 2011).

- (94) a. A dama quer esperar o cavaleiro.

No exemplo abaixo, os contrastes entre (95 a e b) nos mostram a necessidade de alcance do clítico para o domínio matriz.

- (95) a. A dama o quer esperar.
 b. *A dama quer o esperar.

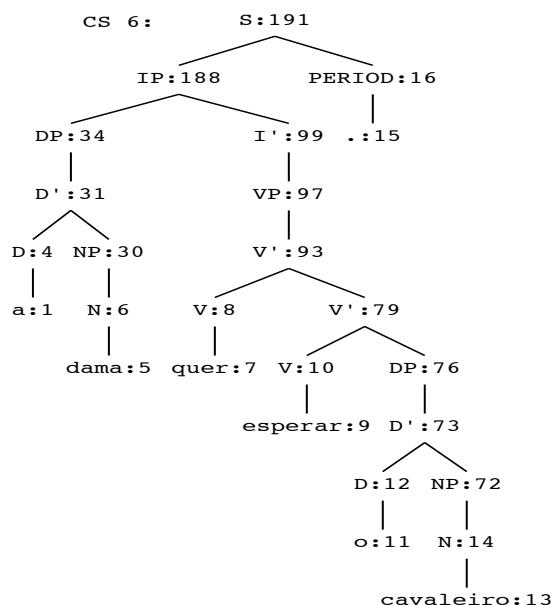
Este é, segundo Rech (2011), um dos testes de constituição de predicados complexos. Estruturalmente, na acepção de Othero (2009), o complemento verbal é um sintagma flexionado. Neste ponto nossas regras diferem, pois entendemos que o verbo principal seleciona um VP, ao invés de um IP. Em nossa gramática modelamos estes padrões verbais de forma diferente.

- (96) $V' \rightarrow V: ^{:=!};$
 $(V: (^ XCOMP)=!)$

Essa regra captura a ideia de que o verbo principal seleciona um complemento oracional aberto, cujo sujeito não se realiza dentro do seu domínio sintagmático. Há duas

razões para tratarmos o complemento do verbo como um V' e não como um IP. Vejamos a seguinte árvore:

Figura 41 – Estrutura-c de uma sentença com complemento oracional aberto no infinitivo



Fonte: elaborada pela autora através do XLE

Note que, nesta análise, o complemento do verbo é um elemento verbal sem marca de pessoa, por isso não alcança a projeção máxima. Para se completar, necessitará do especificador de IP.

Verbos modais são verbos plenos, cujas construções se caracterizam pela formação de dois sintagmas verbais. Sustentamos nossa posição nos “testes de auxiliaridade” propostos por Othero (2009). Em seus testes, o autor aplica a negação e os elementos modificadores de tempo para averiguar se os modais podem ser tratados como auxiliares. Os testes demonstram que verbos auxiliares não podem ser negados, nem modificados, exatamente por não possuírem conteúdo semântico.

Em português existem modais que se conectam a um infinitivo através de preposição. Por exemplo, *começar a*, *acabar de*, *terminar de*, *continuar a*, *parar de*.

(97) A fada {começou a/ terminou de/ continuou a/ parou de} observar a dama.

Outros verbos modalizadores, por outro lado, se conectam a um infinitivo sem o uso da preposição. Por exemplo, *dever*, *ter que* e *querer*.

- (98) A fada {deve/ tem que/ quer/ pode} encantar o cavaleiro.

Aqui, os modais são tratados como verbos plenos que, diferentemente dos auxiliares, introduzem um predicado subcategorizando um XCOMP (BUTT et al. 1999). Estes fatos podem ser facilmente levados em conta simplesmente assumindo, para formas finitas e infinitas diferentes especificações lexicais. Nós, então, teremos entradas lexicais como as seguintes:

- | | | | |
|------|---------|---|--|
| (99) | começou | V | $(^{\text{PRED}}= \text{'começar} < (^{\text{SUBJ}})(^{\text{XCOMP}}) >$ |
| | | | $(^{\text{XCOMP}} \text{ SUBJ}) = (^{\text{SUBJ}})$ |
| | | | $(^{\text{VTYPE}}) = \text{modal}$ |
| | | | $(^{\text{PASSIVE}}) = -$ |
| | | | $(^{\text{PERF}}) = -.$ |

Para uma sentença como *a fada começou a observar a dama* nós tomamos uma estrutura-c e uma estrutura-f como seguem:

Figura 42 – Estrutura-f de uma sentença com predicado complexo

"a fada começou a observar a dama ."

PRED	'começar<[1:fada], [9:observar]>'
SUBJ	5 PRED 'fada' 6 NTYPE [NSEM [COMMON count]] 51 NSYN common
	1 4 SPEC [DET [PRED 'o']] 58 64 CASE nom, GEND fem, HUMAN -, NUM sg, PERS 3
XCOMP	13 PRED 'observar<[1:fada], [15:dama]>' 14 SUBJ [1:fada] 21 19 PRED 'dama' 22 170 NTYPE [NSEM [COMMON count]] 178 113 NSYN common 7 114 8 235 OBJ 15 239 9 SPEC [DET [PRED 'o']] 183 18 186 10 201 320 82 CASE acc, GEND fem, HUMAN +, NUM sg, PERS 3 343 257 CLAUSE-TYPE nom, PASSIVE -, VFORM infinitive, XCOMP-FORM a 351 CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, PERF -, TENSE past, VTYPE modal

Fonte: elaborada pela autora através do XLE

Em termos de estrutura de constituintes, a principal diferença entre modais e auxiliares está no sintagma flexionado. No primeiro caso o núcleo do IP não é preenchido, enquanto que, no segundo, o auxiliar funciona como uma categoria funcional, preenchendo o núcleo do sintagma flexionado.

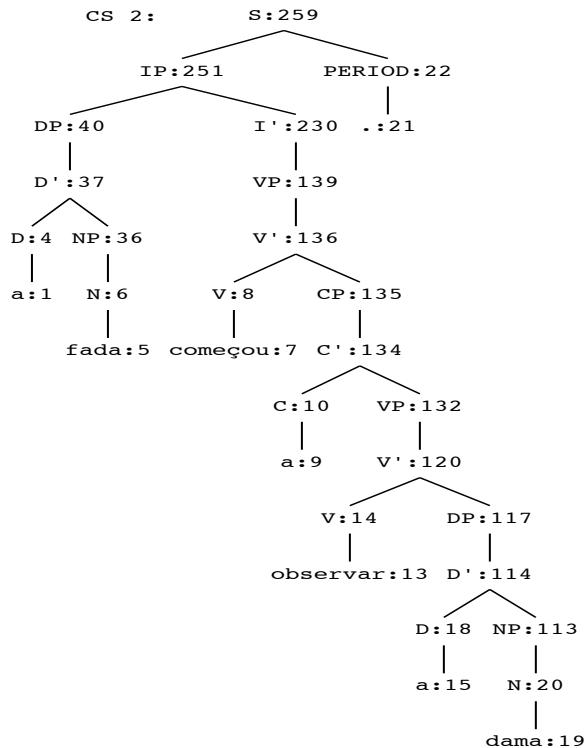
$$(101) \quad \begin{array}{lll} VP \rightarrow & V':^=!; \\ V' \rightarrow & V:^=!; \\ & (CP: (^ XCOMP)=!) \end{array}$$

Outro aspecto importante se dá no complemento de C. Em casos de modais, C só pode selecionar um sintagma verbal, uma vez que seu argumento externo está realizado fora do seu domínio, não permitindo, deste modo, a realização de um IP. Sendo assim, modelamos as regras do seguinte modo:

$$(102) \quad \begin{array}{lll} CP \rightarrow & C':^=! \\ C' \rightarrow & (C:^=!) \end{array}$$

(VP: $\wedge=!$).

Figura 43 – Estrutura-c de uma sentença com predicado complexo



Fonte: elaborada pela autora através do XLE

Note que, nesta estrutura-c, o verbo modal é encadeado por um CP e não por um PP, como se podia esperar. Nesta análise assumimos que as formas *a* e *de* não correspondem a preposições, mas sim a partículas complementadoras, pois estes se comportam exatamente como o complementador *que*.

4.2.5.2 Verbos de controle

Complexos verbais também podem se formar através de verbos de controle. Em LFG, controle funcional é realizado com um atributo XCOMP (um complemento predicativo), cujo sujeito é obrigatoriamente controlado de fora da oração. Vejamos os exemplos abaixo:

- (103) A dama prometeu observar o cavaleiro.

- (104) A rainha persuadiu a dama a esperar o cavaleiro.

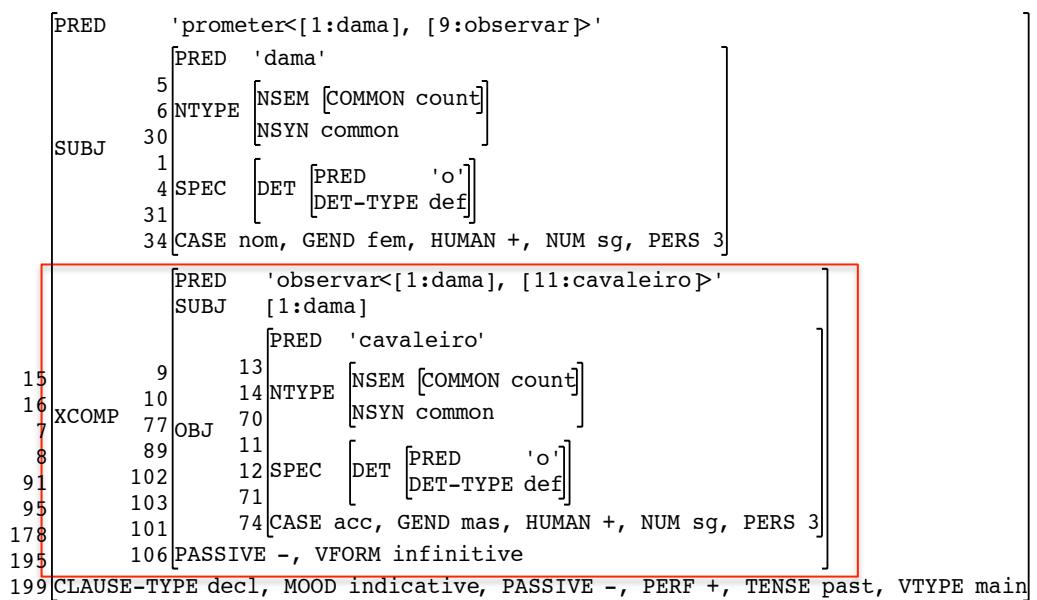
Na modelação da gramática, controle funcional é especificado lexicalmente. Em (103), o sujeito do complemento verbal é controlado pelo sujeito do verbo principal. Para capturar esta informação, codificamos na entrada lexical *prometeu* a seguinte equação:

- (105) $(^{\text{PRED}}) = \text{'prometer} < (^{\text{SUBJ}})(^{\text{XCOMP}}) >$
 $(^{\text{XCOMP}} \text{ SUBJ}) = (^{\text{SUBJ}})$

Com esta equação, expressamos que o sujeito do XCOMP é controlado pelo sujeito do verbo principal.

Figura 44 – Estrutura-f de sentença com controle funcional de sujeito

"a dama prometeu observar o cavaleiro ."



Fonte: elaborada pela autora através do XLE

Na estrutura funcional acima, está representado que, no controle funcional de sujeito, o sujeito é compartilhado pelo verbo principal e pelo complemento predicativo, i.e. uma mesma estrutura-f é o valor de dois atributos, o SUBJ de *prometer* e o SUBJ de *observar*.

No exemplo acima vimos que o sujeito da encaixada é idêntico ao sujeito da principal. Na sentença (104), vista anteriormente, acontece o contrário. O sujeito do complemento verbal é controlado pelo objeto do verbo principal. Para capturar esta informação, codificamos na entrada lexical *persuadiu* a seguinte equação:

$$(106) \quad (^{\text{PRED}}) = \begin{aligned} & \text{'persuadir<} (^{\text{SUBJ}}) (^{\text{OBJ}}) (^{\text{XCOMP}}) \text{>}' \\ & (^{\text{XCOMP}} \text{ SUBJ}) = (^{\text{OBJ}}) \end{aligned}$$

Com esta equação, codificamos lexicalmente que o objeto do verbo principal corresponde necessariamente ao sujeito do XCOMP.

Figura 45 – Estrutura-f de sentença com controle funcional de objeto

"a rainha persuadiu a dama a esperar o cavaleiro ."

	PRED	'persuadir<[1:rainha], [9:dama], [15:esperar]>'	
	5 PRED	'rainha'	
SUBJ	6 NTYPE	[NSEM [COMMON count]]	
	55 NSYN	common	
	56		
	1 SPEC	[DET [PRED 'o']]	
	4		
	62 SPEC	[DET [PRED 'o']]	
	68		
	CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3		
	13 PRED	'dama'	
	14 NTYPE	[NSEM [COMMON count]]	
OBJ	101 NSYN	common	
	102		
	9 SPEC	[DET [PRED 'o']]	
	12		
	380 SPEC	[DET [PRED 'o']]	
	382		
	CASE acc, GEND fem, HUMAN +, NUM sg, PERS 3		
	PRED	'esperar<[9:dama], [21:cavaleiro]>'	
	SUBJ	[9:dama]	
	19		
	20 23 PRED	'cavaleiro'	
	24		
	197 212 NTYPE	[NSEM [COMMON count]]	
	214 OBJ 181 NSYN	common	
XCOMP	7 182		
	298 21 SPEC	[DET [PRED 'o']]	
	220 15 22		
	16 188 SPEC	[DET [PRED 'o']]	
	222 194		
	420 216 CASE	acc, GEND mas, HUMAN +, NUM sg, PERS 3	
	451 341 PASSIVE	- , VFORM infinitive, XCOMP-FORM a	
	455 CLAUSE-TYPE	decl, MOOD indicative, PASSIVE - , PERF +, TENSE past, VTTYPE main	

Fonte: elaborada pela autora através do XLE

4.3 Elementos nominais

Nesta seção tratamos de quatro tipos distintos de elementos nominais: pronomes expletivos, demonstrativos, reflexivos e interrogativos.

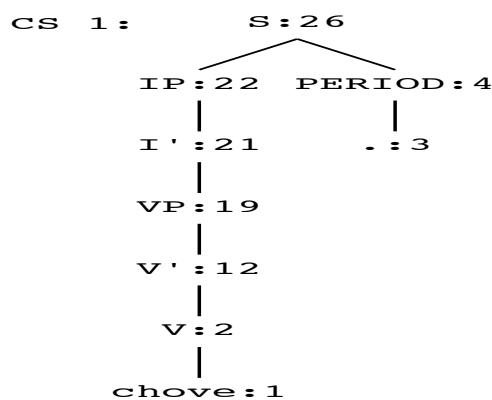
4.3.1 Pronomes expletivos

Construções com expletivos variam de língua pra língua. Em inglês há dois tipos, *it* e *there*. Em alemão há o expletivo *es*. Ao contrário destas duas línguas, o português apresenta o pronome expletivo em forma nula.

(107) Chove

Em uma sentença como esta, o verbo ocorre sem nenhum argumento. Se não há argumentos verbais presentes, a árvore da estrutura-c contém somente o verbo e não há nó correspondente para a estrutura-f do sujeito.

Figura 46 – Estrutura-c de pronome expletivo



Fonte: elaborada pela autora através do XLE

Nesta construção, o sujeito é um pronome expletivo. Ele difere de outros pronomes por não fazer referência a uma entidade atual. Por não haver a necessidade de atribuir-lhe uma resolução semântica e por serem predicionalmente vazios, expletivos são codificados sem atribuição de valor prediccional PRED (BUTT et al. 1999), da seguinte maneira:

- (108) (^ PRED)='chover<>(^ SUBJ)'
 (^ SUBJ PRON-TYPE)=expl_
 (^ SUBJ PRON-FORM)=null

Quando não há um sujeito expresso na sentença, a informação especificada pelo verbo supri o valor SUBJ para a sentença.

Já que não há sujeito explícito, todas as informações sobre o sujeito vêm de especificações no verbo.

Figura 47 – Estrutura-f de pronome expletivo

```
"chove ."
3 [PRED 'chover<>[1-SUBJ]'
4 [SUBJ [NUM sg, PERS 3, PRON-FORM null, PRON-TYPE expl_]
1 [CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTTYPE main
2
12
19
21
22
26]
```

Fonte: elaborada pela autora através do XLE

4.3.2 Pronomes demonstrativos

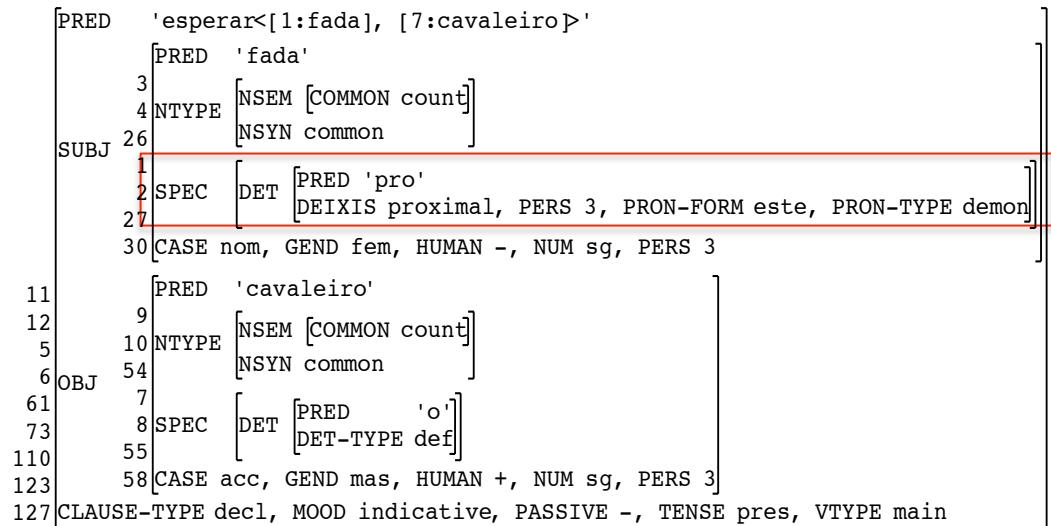
Os pronomes demonstrativos ocorrem como núcleo do sintagma determinante e admitem, outros termos dentro do DP, como sintagmas nominais.

- (109) esta fada espera o cavaleiro.

Sua estrutura funcional codifica informações, tais como, gênero, número e pessoa.

Figura 48 – Estrutura-f de pronome demonstrativo

"esta fada espera o cavaleiro ."



Fonte: elaborada pela autora através do XLE

Note que os demonstrativos também são designados de PRON-TYPE *demon* e têm um atributo DEIXIS, cujo valor é *proximal*. Esta característica é codificada para demonstrar que o pronome aparece como determinante, ao contrário das formas que funcionam como pronomes pessoais.

4.3.3 Pronomes reflexivos

Os reflexivos são usados para expressar que o objeto de uma oração é entendido como referencialmente idêntico ao sujeito da mesma oração²⁵ (PERINI, 2010, p. 245).

- (110) A rainha se maquia.

²⁵ Conforme Perini (2010) os reflexivos também podem expressar que o sujeito e o objeto de uma sentença são partes simétricas de um evento ou estado. Dada a complexidade envolvida neste fenômeno, o seu tratamento ficará para trabalhos posteriores.

No exemplo (110) temos um caso em que o objeto é expresso pela forma reflexiva. Tais pronomes são analisados como tendo um valor PRED de pronome, indicando que estes são anáforas aguardando resolução dentro do componente semântico (BUTT et al., 1999, p. 76). Estes reflexivos são argumentos verbais que, para fornecer o máximo de informações para tal componente, devem ser modelados com o máximo de informações, a estrutura-f abaixo.

Figura 49 – Estrutura-f de pronome reflexivo

	"a rainha se maquia ."	
	PRED	'maquiar<1:rainha], [7-OBJ:pro>'
	5	PRED 'rainha'
	6	NTYPE [NSEM [COMMON count]]
11	41	NSYN common
	12	SUBJ
	9	1
	10	4 SPEC [DET [PRED 'o']]
	63	48
	7	54 CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3
	8	PRED 'pro'
	85	OBJ NTYPE [NSYN pronoun]
	94	[CASE acc, GEND mas, NUM sg, PERS 3, PRON-TYPE refl]
101	105	CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main

Fonte: elaborada pela autora através do XLE

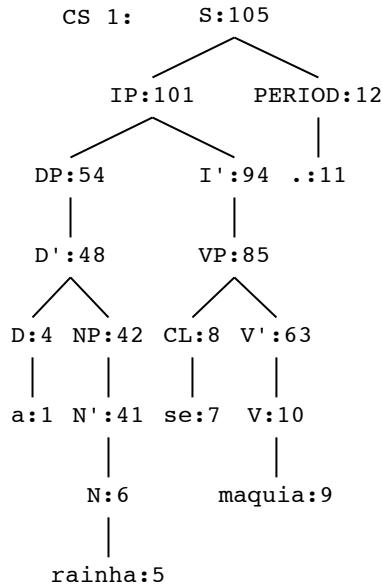
Note que a forma de superfície do pronome é codificado no valor PRON-TYPE, utilizado para distinguir entre vários tipos de pronomes. Também são codificados caso, gênero, número e pessoa, características necessárias para a avaliação semântica do pronome.

Pronomes reflexivos não podem aparecer com determinantes ou com modificação pronominal, por isto são instanciados, na estrutura de constituintes, diretamente sobre CL. Para capturar esta forma de distribuição destes reflexivos, codificamos a seguinte regra:

$$(111) \quad VP \rightarrow (CL: ^{=}!) \\ V'^{=}!;$$

Com esta regra, a gramática licencia a estrutura-c que mostramos abaixo, na qual na condição de clítico, ou seja, como pronome oblíquo, em regra, ocorre antes do verbo (PERINI, 2010, p. 119).

Figura 50 – Estrutura-c de pronome reflexivo

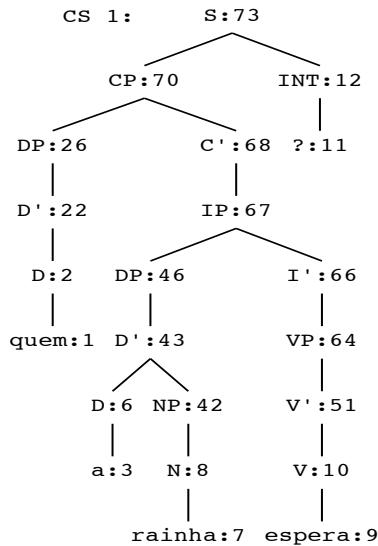


Fonte: elaborada pela autora através do XLE

4.3.4 Pronomes interrogativos

Vimos na seção anterior que pronomes reflexivos são instanciados na estrutura de constituintes como clíticos. Já os interrogativos são instanciados como núcleos de sintagmas determinantes, conforme mostra a Figura 50 .

Figura 51 – Estrutura-c de sentença interrogativa



Fonte: elaborada pela autora através do XLE

A principal motivação para modelarmos os interrogativos como núcleos funcionais é a condição de alguns destes pronomes de selecionarem complementos, como, por exemplo, os pronomes *qual* e *que*.

(112) Quais/que damas esperam?

Um aspecto dos DPs interrogativos é que estes, quando deslocados na periferia esquerda da sentença, são instanciados por um CP e recebem, além das características de OBJ e CASE próprias do argumento verbal que representam, uma marcação de foco sintático e uma restrição exclusiva para o tipo de pronome (PRON-TYPE) interrogativo.

(113) CP --> DP: (^ OBJ)=!
 (! CASE)=acc
 (^ FOCUS)=!
 (! PRON-TYPE)=c int)
 C':^=!

Outro aspecto dos pronomes interrogativos é que possuem características nominais, e possuem, modeladas na estrutura-f, a distinção sintática (NTYPE) que os diferencia dos nomes e o PRON-TYPE que representam.

Figura 52 – Estrutura-f de sentença interrogativo

"quem a rainha espera ?"

	PRED	'esperar<[3:rainha], [1:pro]>'
		[PRED 'rainha'
	7	
	8	8 NTYPE [NSEM [COMMON count]
		42 NSYN common
11	SUBJ	3
12		12 SPEC [DET [PRED 'o']]
9		6 DET [DET-TYPE def]
10		43 10 CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3
51		46
64		1 64 FOCUS [PRED 'pro']
66	OBJ	2 66 FOCUS [NTYPE [NSYN pronoun]]
67		22 67 NUM sg, PERS 3, PRON-FORM quem, PRON-TYPE int
68		26 68
70	FOCUS	[1:pro]
73		73 CLAUSE-TYPE int, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main

Fonte: elaborada pela autora através do XLE

De acordo com o PARGRAM o pronome interrogativo possui um NTYPE com uma característica sintática (NSYN) pronominal e um PRON-TYPE atribuído de valor interrogativo. Ademais, o pronome também é especificado pela forma (PRON-FORM), pelo número e pessoa (BUTT et al.; KING, 2006).

4.4 Sintagmas preposicionados

Na modelação dos sintagmas preposicionados, a principal dificuldade a ser implementada está relacionada com a distinção entre preposições semânticas e não-semânticas.

A preposição semântica possui um predicado, expressa bastante conteúdo semântico e normalmente dá origem a um PP, como em *embaixo da mesa*. Por outro lado, as preposições não-semânticas, normalmente, representam, apenas, uma exigência do verbo e não trazem ou trazem pouca contribuição semântica para o significado do PP que se origina após o verbo, como em *a assistir a um filme*. Nossa tratamento destes fenômenos é apresentado nas seções seguintes.

4.4.1 Preposições semânticas

As preposições semânticas, como são chamadas por Butt et al. (1999, p. 125), se referem àquelas preposições que atribuem papel temático ao sintagma que ajudam a formar (PERINI, 2010, p. 312).

Segundo Butt et al. (1999, p. 125), “o tipo mais comum de PP envolve uma preposição que tem um claro conteúdo semântico em si mesmo, tais como os locativos *em*, *sobre*, *sob*, etc., o instrumental *com*, e os direcionais *para dentro de*, *para cima de*, etc”²⁶.

- (114) Em o prado

Em nossa gramática, este tipo de preposição é dotada de um valor PRED e subcategoriza um quadro que indica que a preposição exige um objeto, que normalmente é um NP. Para diferenciar esta preposição das preposições não semânticas, deve ser codificada na estrutura-f com o atributo PTYPE semântico.

- (115) (^ PRED) = 'em<(^ OBJ)>'
 (^ PTYPE)=sem
 (^ PSEM)=loc

No exemplo (116) temos um sintagma preposicionado que funciona como um adjunto do verbo.

- (116) A fada espera a dama no prado.

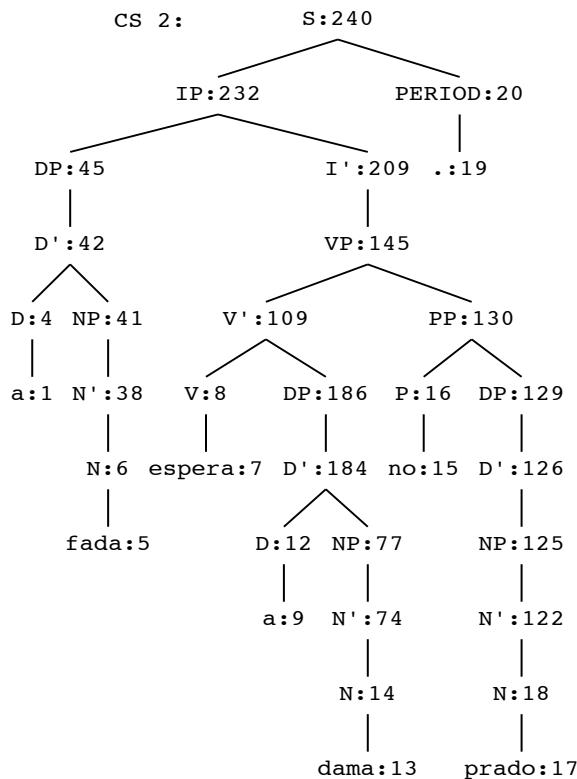
Para codificarmos esta condição, modelamos em nossa gramática a seguinte regra:

²⁶ *The most common type of PP involves a preposition which has a clear semantic content of its own, such as the locatives on, in, under, etc., the instrumental with, and the directional into, onto, etc.* (BUTT et al., 1999, 125).

- (117) VP --> V';^=!;
 PP*: ! \$ (^ ADJUNCT)
 (! CASE)=obl
 (! PTYPE)=c sem

Observe que, na árvore abaixo, este sintagma é instanciado na estrutura-c em um nível intermediário.

Figura 53 – Estrutura-c com uma preposição semântica

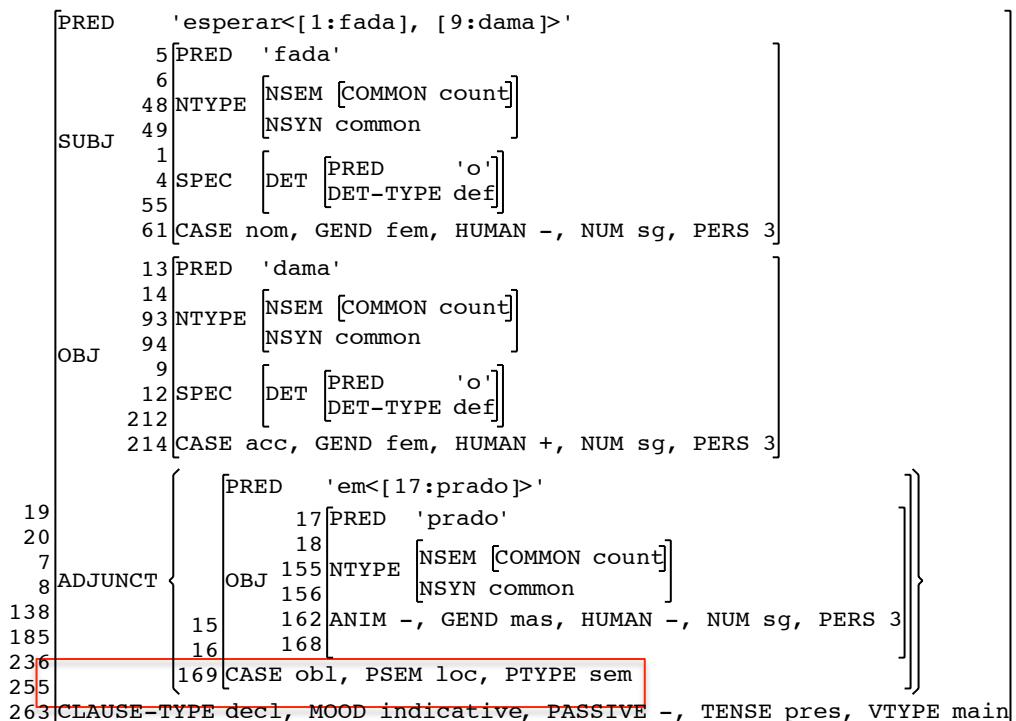


Fonte: elaborada pela autora através do XLE

Não é apenas na estrutura-c que está expressa a condição de adjunto do PP. Esta condição também está configurada na estrutura-f da sentença pela função semântica *locativa* que exerce sobre o verbo.

Figura 54 – Estrutura-f de uma preposição semântica

"a fada espera a dama no prado ."



Fonte: elaborada pela autora através do XLE

Observe que, na estrutura-f acima, o adjunto, não sendo um argumento do verbo não constitui um valor de PRED e, portanto, é representado fora da sua estrutura argumental.

4.4.2 Preposições não semânticas

Segundo Butt et al. (1999, p. 127), “em certas construções uma preposição em particular é exigida pelo verbo em um de seus sentidos particulares. Nestes casos, a preposição não faz, ou faz muito pouca, contribuição semântica por si própria”.

- (118) a. A dama espera um cavaleiro no prado.
 b. A dama espera por um cavaleiro no prado.

Nas duas construções acima, tanto o NP *o cavaleiro* como o PP *por um cavaleiro* são complementos do verbo. Em contraste com o PP *no prado*, que claramente indica um

adjunto locativo do verbo, o uso do PP complemento é tratado, neste caso em particular, como argumento de *espera*.

Como o PP complemento requer uma preposição não semântica, a estratégia para diferenciar os dois tipos de sintagmas é a codificação da preposição não semântica com um PTYPE *nosem* (BUTT et al., 1999, p. 127-128). Preposições que ocorrem nestas condições são analisadas como não tendo um PRED. Elas não subcategorizam um objeto, o que significa que o PP argumento *por um cavaleiro* é tratado como objeto do verbo (BUTT et al. 1999). Uma amostra de estrutura-f para esta construção é apresentada abaixo, com destaque para as atribuições do OBJ.

Figura 55 – Estrutura-f de uma preposição não semântica

"a dama espera por um cavaleiro no prado ."

PRED	'esperar<[1:dama], [9:cavaleiro]>'	
	5[PRED 'dama'	
	6[NTYPE [NSEM [COMMON count]	
	41[NSYN common]	
SUBJ	1[SPEC [DET [PRED 'o']	
	4[DET-TYPE def]	
	42[CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3]	
	45[
	13[PRED 'cavaleiro'	
	14[NTYPE [NSEM [COMMON count]	
	75[NSYN common]	
OBJ	11[SPEC [DET [PRED 'um']]	
	12[DET-TYPE indef]	
	79[CASE acc, GEND mas, HUMAN +, NUM sg, PERS 3, PFORM por, PTYPE nosem]	
	82[
	9[
	10[
	83[
	19[PRED 'em<[17:prado]>'	
	20[17[PRED 'prado'	
	7[18[NTYPE [NSEM [COMMON count]	
ADJUNCT	8[107[NSYN common]]	
	84[110[ANIM -, GEND mas, HUMAN -, NUM sg, PERS 3]	
	131[111[
	185[114[
	204[115[CASE obl, PSEM loc, PTYPE sem]	
	208[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main]	

Fonte: elaborada pela autora através do XLE

Segundo Butt et al. (1999, p. 129), a preposição serve meramente para designar caso ao argumento. Como vimos no exemplo acima, em PB *esperar* que normalmente é um

verbo transitivo, as vezes pode ser um verbo que subcategoriza, ao invés de um DP, um PP encabeçado pela preposição *por*.

4.5 Sintagma adjetival

Adjetivos são caracterizados como modificadores nominais. Em PB, pronomes adjetivos aparecem de maneiras variadas. Neste trabalho apresentamos três destas maneiras: adjetivos atributivos, cardinais e ordinais e adjetivos com modificadores. Chamamos atenção para o fato de que os ordinais e cardinais, embora sejam modelados em nossa gramática como NumPs e não como APs, na abordagem tradicional da gramática, ficaram situados neste capítulo.

4.5.1 Adjetivos atributivos

Os adjetivos prototípicos para o PB são os adjetivos atributivos pós-nominais (119), embora também sejam usuais em posição pré-nominal (120)²⁷.

- (119) A fada cintilante observa a dama.
- (120) A amável dama espera o cavaleiro.

Em PB os adjetivos flexionam para concordar com o núcleo nominal que eles modificam em gênero e número (MIOTO et al., 2005, p. 104) e são codificados tendo um ATYPE *atributivo*.

Em PB os adjetivos podem ser predicativos (seção 4.2.2.2) e atributivos. Nos casos dos atributivos, a exigência deste tipo de adjetivo é forçada via equação funcional na estrutura-

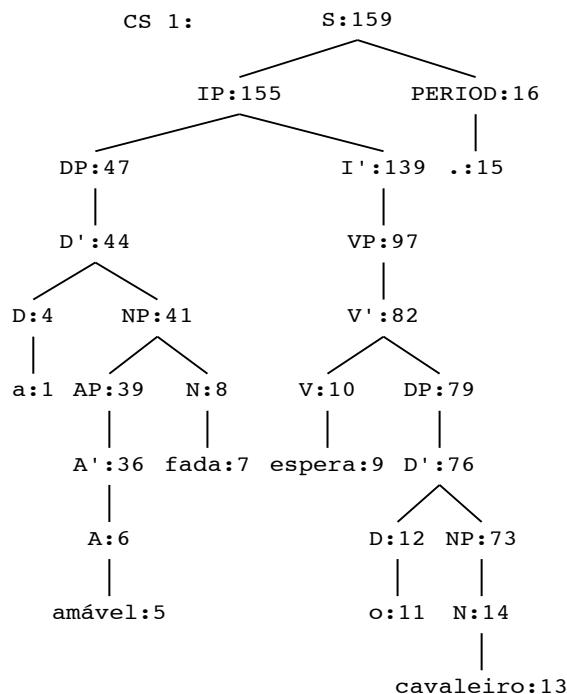
²⁷ O constituinte AP aponta uma série de complicações, como os casos de separação dos adjetivos atributivos por vírgulas ou coordenação. No entanto, tais aspectos vão além dos limites deste trabalho.

c se um atributo é chamado. Também deve estar codificado na estrutura-c que a estrutura-f do sujeito do AP corresponde à estrutura-f do nó superior, o nó NP.

Elaboramos a seguinte regra para analisar a estrutura do AP pré-nominal.

- (121) NP --> AP*: ! \$ (^ ADJUNCT)
 (! SUBJ)=^
 (! ATYPE)=c attributive;
 N:^=!.

Figura 56 – Estrutura-c de adjetivo pré-nominal



Fonte: elaborada pela autora através do XLE

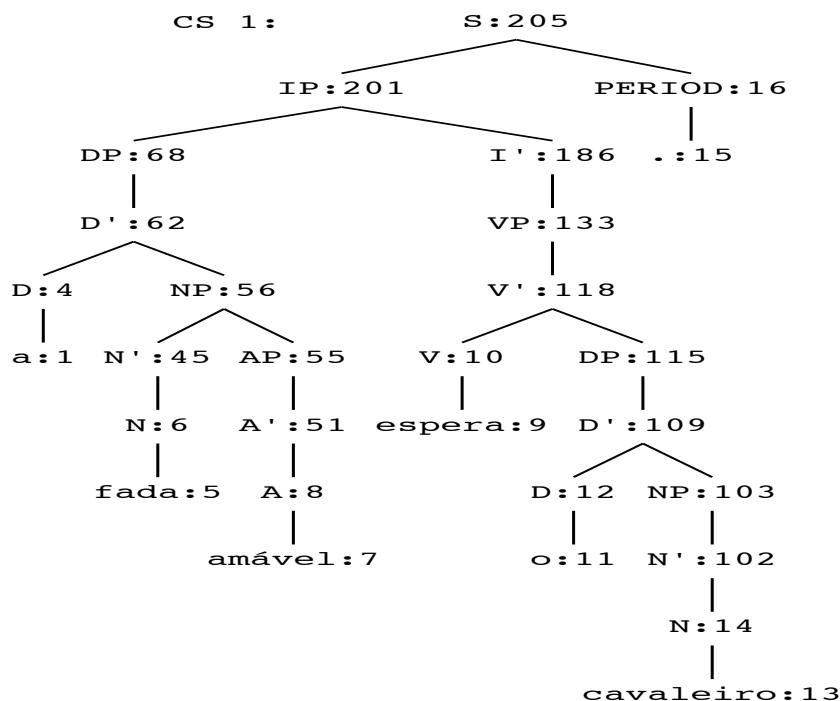
Adjetivos pós-nominais em português são equivalentes à sua contraparte pré-nominal. Assim como os pré-nominais, os adjetivos pós-nominais também concordam em gênero e número com o elemento modificado.

Para analisar o adjetivo em posição pós-nominal a gramática precisa ter uma regra que traga o AP adjunto à direita do núcleo nominal, como em (122).

- (122) NP --> N:^=!;

AP*: ! \$ (^ ADJUNCT)
 (! SUBJ)=[^]
 (! ATYPE)=c attributive

Figura 57 – Estrutura-c de adjetivo pós-nominal



Fonte: elaborada pela autora através do XLE

As duas regras, acima apresentadas, descrevem um NP com modificação adjetival. Estas regras, essencialmente, olham para a estrutura-f do AP e exigem que o SUBJ do ADJUNCT seja fornecido pelo NP. Como o NP não possui SUBJ, é necessário codificar no léxico o PRED do SUBJ do ADJUNCT e do NP (ALENCAR, 2013).

- (123) amável A * (^ PRED)='amável<(^ SUBJ)>'
 (^ SUBJ NUM)=sg
 (^ SUBJ GEND)=fem
 (^ ATYPE)=attributive
 (^ DEGREE)=positive.

Uma estrutura-f bem formada é dada abaixo.

Figura 58 – Estrutura-f de adjetivo modificador de sintagma nominal

"a amável fada espera o cavaleiro ."

PRED	'esperar<[1:fada], [11:cavaleiro]>'	
	PRED 'fada'	
ADJUNCT	{ 5[PRED 'amável<[1:fada]>' 6[SUBJ [1:fada] 36[ATYPE attributive, DEGREE positive 39]	}
SUBJ	7[NSEM [COMMON count] 8[NSYN common	
	41	
	1[PRED 'o']	
4[SPEC	[DET [PRED 44[DET-TYPE def]]	
	47[CASE nom, GEND fem, HUMAN -, NUM sg, PERS 3	
15[PRED 'cavaleiro'		
16[13[NSEM [COMMON count]		
9[NSYN common		
10[OBJ 73		
82[11[PRED 'o']		
97[12[SPEC [DET [PRED 76[DET-TYPE def]]		
139[79[CASE acc, GEND mas, HUMAN +, NUM sg, PERS 3		
155[159[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main		

Fonte: elaborada pela autora através do XLE

4.5.2 *Cardinais e ordinais*

Ordinais como em (124) se comportam como um simples adjetivo atributivo no que diz respeito a flexão.

- (124) A primeira dama

No entanto, eles têm uma distribuição diferente na estrutura-c e são assinalados constituintes sintáticos NumP (BUTT et al., 1999, p. 108). Observe os exemplos abaixo:

- (125) a. A nossa primeira dama
b. *A primeira nossa dama

Como se observa no contraste acima, os dois elementos podem co-ocorrer dentro do agrupamento nominal, no entanto, há uma ordem linear que deve ser respeitada. O contraste, portanto, evidencia que tais elementos não são da mesma natureza, pois não estão em distribuição complementar.

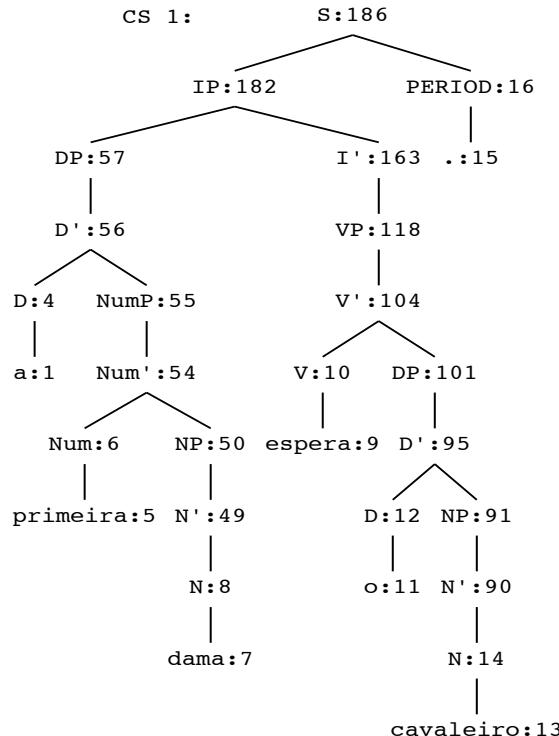
Desse modo, estamos encarando o NumP como um especificador do NP e não como seu adjunto. Vejamos abaixo as regras da gramática:

- (126) DP --> D': $\wedge=!$.
 D' --> (D: $\wedge=!$)
 (NumP: $\wedge=!$).
 NumP --> Num': $\wedge=!$.
 Num' --> Num: $\wedge=!$
 (NP: $\wedge=!$).
 NP --> N': $\wedge=!$.
 N' --> N: $\wedge=!$.

Isto se reflete na estrutura-c como um núcleo funcional que seleciona um complemento NP²⁸.

²⁸ Othero (2009) também apresenta análise do núcleo funcional Num que seleciona um complemento PP. Neste trabalho, por limitação de tempo, não trazemos esta análise.

Figura 59 – Estrutura-c de sintagma nominal modificado por ordinal



Fonte: elaborada pela autora através do XLE

Somente em termos de estrutura-f os ordinais são reconhecidos como adjetivos, pelo atributo ATYPE *ordinal*.

(127) NUMBER * (^ SPEC NUMBER PRED)=*'primeiro'*
 (^ SPEC NUMBER NUM)=sg
 (^ SPEC NUMBER GEND)=fem
 (^ SPEC NUMBER NUMBER-TYPE)=ordinal.

Uma estrutura-f válida é mostrada abaixo.

Figura 60 – Estrutura-f de sintagma nominal modificado por ordinal

"a primeira dama espera o cavaleiro . "

PRED	'esperar<[1:dama], [11:cavaleiro]>'	
7	PRED 'dama'	
8	NTYPE [NSEM [COMMON count]]	
45	NSYN common	
SUBJ		
5		
6	DET [PRED 'o']	
50	DET-TYPE def	
SPEC		
1	NUMBER [PRED 'primeiro']	
4	GEND fem, NUM sg, NUMBER-TYPE ordinal	
51	CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3	
52		
15		
16	13 PRED 'cavaleiro'	
9	14 NTYPE [NSEM [COMMON count]]	
83	NSYN common	
10	OBJ	
95	84	
11	15 PRED 'o'	
109	12 SPEC [DET [PRED 'o']]	
152	88	
169	169 CASE acc, GEND mas, HUMAN +, NUM sg, PERS 3	
173	173 CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main	

Fonte: elaborada pela autora através do XLE

Cardinais, por outro lado, não flexionam, mas exigem que o nome esteja no plural (a menos que o cardinal seja *um*), e exibem um padrão sintático ligeiramente diferente na estrutura-c, se compararmos o contraste em (127) (BUTT et al., 1999, p. 109; OTHERO, 2009).

- (127) a. As duas fadas esperam a rainha.
 b. *as fadas duas esperam a rainha.

Assim como os ordinais, os cardinais são introduzidos por uma regra NumP na estrutura-c e são diferenciados na estrutura-f pelo atributo NUMBER-TYPE *cardinal*, como vemos abaixo.

Figura 61 – Estrutura-f de sintagma nominal modificado por cardinal

"as duas damas esperam a rainha ."

```

PRED  'esperar<1:dama], [9:rainha]>
      5[PRED  'dama'
      6
      49[NTYPE [NSEM [COMMON count]
      50[NSYN common
      3
      4[SUBJ [DET [PRED      'o'
      54[SPEC [DET-TYPE def
      55[NUMBER [PRED 'dois'
      1[NUMBER [GEND fem, NUM sg, NUMBER-TYPE cardinal
      2[CASE nom, GEND fem, HUMAN +, NUM pl, PERS 3
      56
      57

      15[PRED  'rainha'
      16
      89[NTYPE [NSEM [COMMON count]
      7[NSYN common
      8[OBJ [DET [PRED      'o'
      90[SPEC [DET-TYPE def
      130[9
      137[12[SPEC [DET [PRED      'o'
      174[152[CASE acc, GEND fem, HUMAN +, NUM sg, PERS 3
      193[154[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main
      201[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main
  
```

Fonte: elaborada pela autora através do XLE

Observe, na parte em destaque, que NUMBER e NUMBER-TYPE são codificados como traços especificados SPEC do sujeito.

4.5.3 Adjetivos com intensificadores

Os adjetivos também aceitam algum tipo de determinação que se traduz nos intensificadores como *bastante*, *muito*, *pouco*, *mais* etc (MIOTO et al., 2005, p. 104).

(128) A mais bela dama espera o cavaleiro.

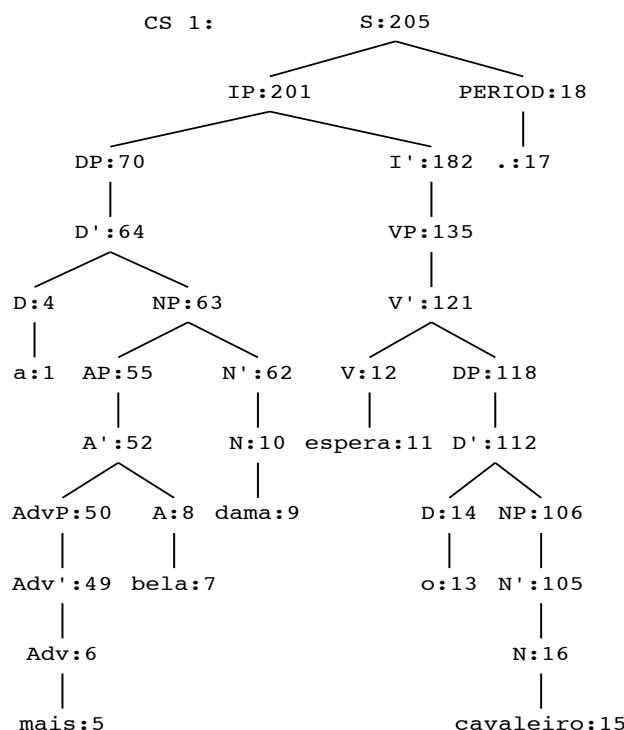
Estamos encarando o advérbio *mais* com um adjunto do AP e não com seu especificador. Advérbios que aparecem dentro das regras do AP devem ser do tipo que podem modificar adjetivos e outros advérbios.

As regras a seguir descrevem um AP com modificação adverbial.

- (129) AP --> A':^=!.
A' --> (AdvP:^=!)
A:^=!;
AdvP --> Adv':^=!.
Adv' --> Adv:^=!.

Na estrutura-c abaixo vemos um AP com modificação adverbial à esquerda do núcleo.

Figura 62 – Estrutura-c de sintagma adjetival com modificação adverbial

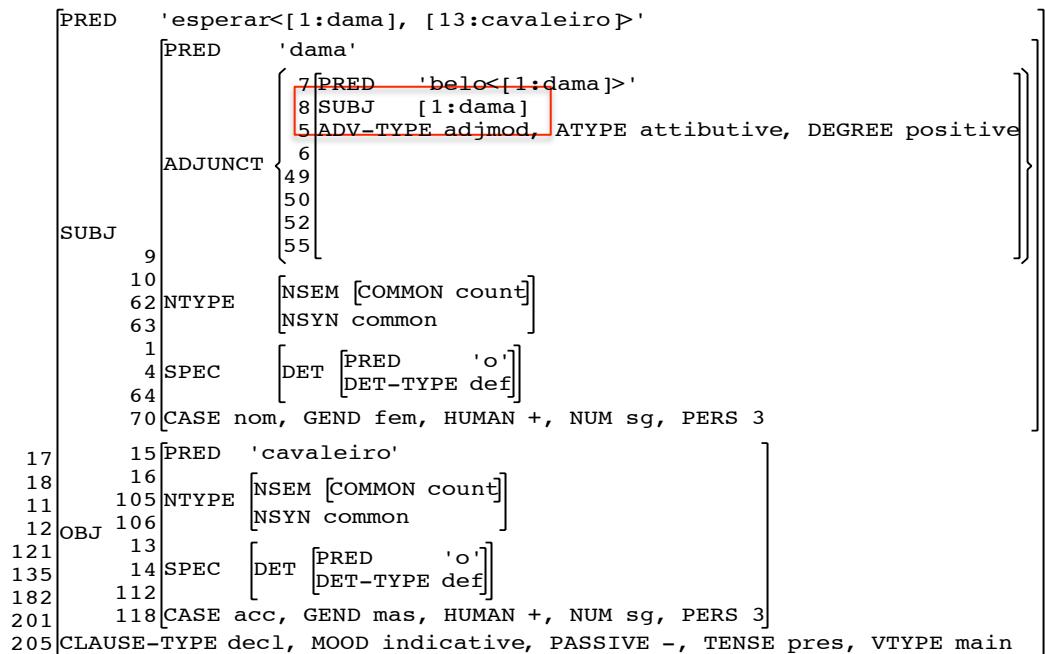


Fonte: elaborada pela autora através do XLE

O léxico, por sua vez, designa o tipo de advérbio que é chamado pelo AP com o atributo ADV-TYPE *adjmod*.

Figura 63 – Estrutura-f de sintagma adjetival com modificação adverbial

"a mais bela dama espera o cavaleiro ."



Fonte: elaborada pela autora através do XLE

Nesta seção apresentamos a modelação de adjetivos atributivos, cardinais, ordinais e adjetivos com intensificadores. Por fim, a próxima e última seção, traz uma discussão sobre dois tipos de advérbios, a saber, intransitivos e transitivos com complemento PP.

4.6 Sintagma adverbial

Segundo Butt et al. (1999), “advérbios têm algumas propriedades que caracterizam a classe como um todo. Ao contrário dos adjetivos e verbos, advérbios não flexionam. Eles servem para modificar tanto uma oração como um constituinte.”

Na seção anterior, para tratarmos de modificação adjetival, falamos sobre advérbios como modificadores de APs. Nesta seção demonstramos como nossa gramática contempla, além do referido advérbio modificador de AP, advérbios intransitivos e advérbios transitivos que selecionam um complemento PP.

4.6.1 Advérbios intransitivos

Os advérbios intransitivos são aqueles que aparecem na sentença sem qualquer modificador. Vejamos, abaixo, um exemplo:

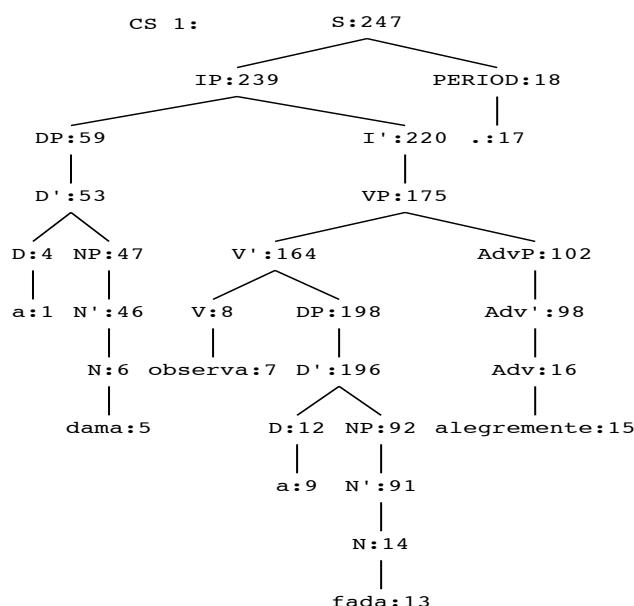
- (130) A dama observa a fada alegremente.

Entendemos que, no exemplo acima, temos um caso de modificação do constituinte verbal. Desse modo, o constituinte adverbial é chamado para a estrutura-c por uma regra de adjunção do nível intermediário V'.

Este tipo de estrutura motiva a elaboração da seguinte regra para a gramática:

- (131) VP \rightarrow V'^{!;}
 AdvP*! \$ (^ ADJUNCT)
 (! ADV-TYPE)=c vpmod.
 AdvP \rightarrow Adv'^{!;}
 Adv' \rightarrow Adv^{!;}

Figura 64 – Estrutura-c de sintagma verbal com modificação adverbial



Fonte: elaborada pela autora através do XLE

Note que duas notações importantes estão presentes nas regras em (131). Primeiro, ao ser chamado pelo VP, o AdvP adjunto deve possuir uma equação de restrição para os tipos adverbiais modificadores do VP. Segundo, o núcleo do AdvP é um advérbio intransitivo.

Uma estrutura-f para a sentença em (130) é mostrada abaixo:

Figura 65 – Estrutura f de sintagma verbal com modificação adverbial

"a dama observa a fada alegramente ."

```

PRED   'observar<1:dam>, [9:fada]>'  

      5 PRED   'dama'  

      6 NTYPE  [NSEM [COMMON count]  

      46 NSYN common  

SUBJ  

      13 PRED   'fada'  

      14 NTYPE  [NSEM [COMMON count]  

      91 NSYN common  

      92  

OBJ  

      9 PRED   'o'  

      12 DET    [PRED   'o'  

      17 SPEC   [DET    [DET-TYPE def]  

      18 196  

      7 CASE   nom, GEND fem, HUMAN +, NUM sg, PERS 3  

      198  

      8 PRED   'alegremente'  

      164 ADV-TYPE vpmmod  

      175 ADJUNCT { 15 PRED   'alegremente'  

      220 16 ADV-TYPE vpmmod  

      239 98  

      102 }  

      247 CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main

```

Fonte: elaborada pela autora através do XLE

A matriz destacada acima mostra o ADJUNCT como modificador de toda a sentença.

4.6.2 Advérbios transitivos

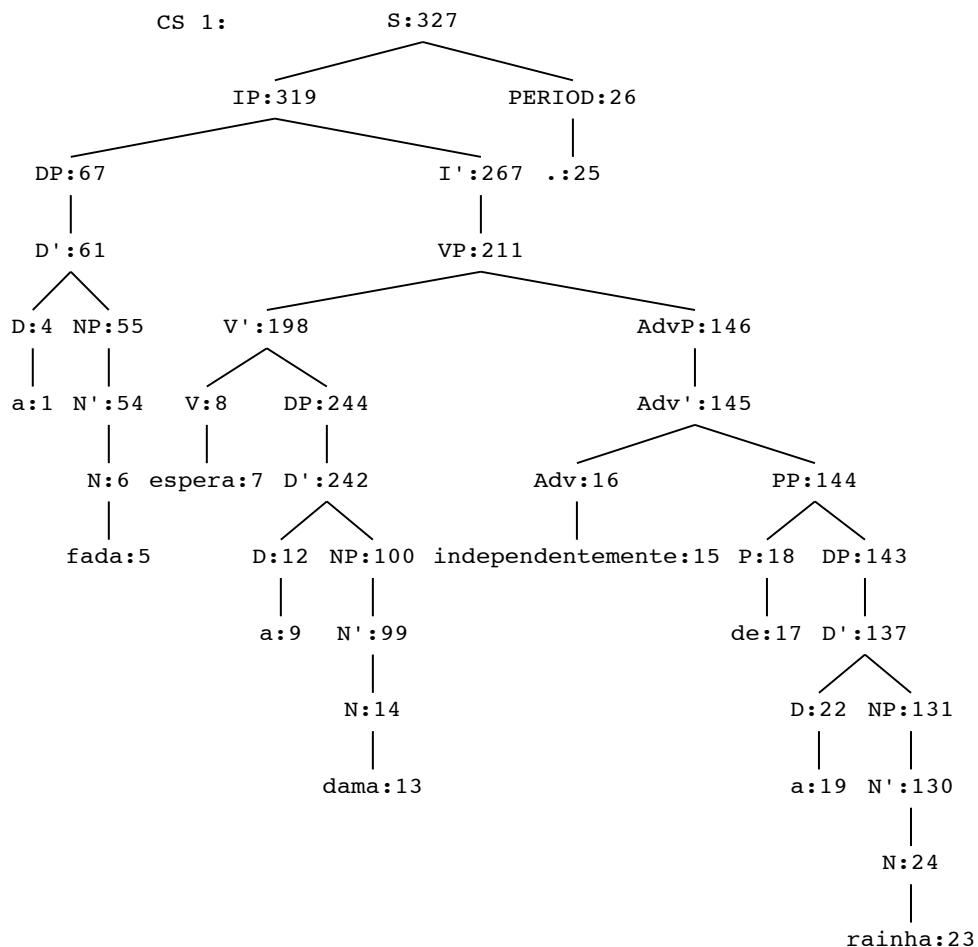
Por outro lado, advérbios podem selecionar complementos. Quando isto acontece, diz-se que são transitivos.

- (132) A fada espera a dama independentemente da rainha.

Para capturar essa propriedade dos advérbios transitivos, a gramática deve possuir a seguinte regra:

- (133) $\text{AdvP} \rightarrow \text{Adv}'^{\wedge=!}.$
 $\text{Adv}' \rightarrow \text{Adv}^{\wedge=!};$
 $(\text{PP}^{\wedge=!}).$

Figura 66 – Estrutura-c de advérbio transitivo



Fonte: elaborada pela autora através do XLE

Para restringir os advérbios que podem aparecer na regra do VP, estes precisam ser de um tipo que possa modificar constituintes verbais, i.e. seu ADV-TYPE precisa ser *vmod*. No léxico isto é codificado do seguinte modo:

(134) Adv * (^ ADV-TYPE)=vmod.

Uma estrutura-f bem formada para a sentença em (132) é dada abaixo:

Figura 67 – Estrutura-f de advérbio transitivo

"a fada espera a dama independentemente de a rainha ."

PRED	'esperar<[1:fada], [9:dama]>'	
	5[PRED 'fada'	
	6[NTYPE [NSEM [COMMON count]	
SUBJ	54[NSEM [COMMON count]	
	55[NSYN common	
	1[SPEC [DET [PRED 'o']]	
	4[SPEC [DET [DET-TYPE def]	
	61[CASE nom, GEND fem, HUMAN -, NUM sg, PERS 3]	
	67[CASE nom, GEND fem, HUMAN -, NUM sg, PERS 3]	
	13[PRED 'dama'	
	14[NTYPE [NSEM [COMMON count]	
OBJ	99[NSEM [COMMON count]	
	100[NSYN common	
	9[SPEC [DET [PRED 'o']]	
	12[SPEC [DET [DET-TYPE def]	
	242[CASE acc, GEND fem, HUMAN +, NUM sg, PERS 3]	
	244[CASE acc, GEND fem, HUMAN +, NUM sg, PERS 3]	
ADJUNCT	23[PRED 'rainha'	
	24[NTYPE [NSEM [COMMON count]	
	130[NSEM [COMMON count]	
	131[NSYN common	
	19[SPEC [DET [PRED 'o']]	
	22[SPEC [DET [DET-TYPE def]	
	137[CASE TYPE vpmod, GEND fem, HUMAN +, NUM sg, PERS 3, PFORM de, PTYPENosem]	
	25	
	26	
	7	
	8	
	198	
	211	
	267	
	319	
	327[CLAUSE-TYPE decl, MOOD indicative, PASSIVE -, TENSE pres, VTYPE main	

Fonte: elaborada pela autora através do XLE

Nossa gramática, no entanto, precisaria ser muito mais complexa para dar conta de todas as particularidades dos advérbios, pois, como chama atenção Butt et al. (1999, p. 134),

“a posição do advérbio é notoriamente difícil de determinar, em parte por que a semântica de um dado advérbio pode mudar em certos contextos, permitindo-o aparecer em uma posição diferente da usual.”

A dificuldade no tratamento dos advérbios também é compartilhada por Othero (2009, p. 111), para quem “as evidências empíricas parecem apontar para o fato de que há mais do que uma classe de palavras no que tradicionalmente se classifica sob o rótulo *advérbio*. ”

Deste modo, devido às limitações próprias de uma tese de doutorado, damos por encerrado o tratamento dos advérbios.

5 RESULTADOS OBTIDOS E DISCUSSÃO

Nesta seção, nós, inicialmente, fazemos um levantamento geral dos aspectos modelados na gramática. Em seguida demonstramos a nossa avaliação da gramática, através da comparação das análises da nossa gramática com as análises do Palavras para as 40 sentenças do nosso arquivo de teste. Concluímos o capítulo levantando algumas considerações sobre a necessidade e os direcionamentos para trabalhos futuros de aperfeiçoamento da gramática.

5.1 Aspectos da gramática

Ao longo deste capítulo nós mostramos como construções sintáticas da frase do português podem ser implementadas em um recurso computacional, levando em conta exigências linguísticas teóricas, mas certificando-se de que o recurso é apropriadamente codificado para propósitos linguístico-computacionais.

O capítulo mostrou que a estrutura frasal do português, mesmo modelada na condição de um fragmento de gramática, apresentou um número heterogêneo de regras de estrutura sintagmática e regras lexicais. Visando contribuir para o enriquecimento de recursos linguísticos para o PB, implementamos um *parser* para o português a partir de um fragmento que abrange a estrutura interna da frase em diversificados e relevantes fenômenos.

O primeiro ponto da gramática a ser ressaltado são os constituintes oracionais IP e CP, elementos que encabeçam as sentenças do português, e, portanto, constituem o nó raiz da gramática. Neste ponto da gramática estão codificadas as principais distinções entre orações declarativas e interrogativas.

Os elementos verbais modelados envolvem subcategorização e estrutura argumental. Estes aspectos são cruciais no estabelecimento das relações gramaticais e na formação da estrutura de predicados. A gramática oferece duas análises distintas para os predicativos de sujeito, a análise do sujeito controlado e a análise *predlink* e faz uma distinção entre verbos auxiliares e verbos plenos, além disso, também apresenta uma modelação da inversão de voz ativa para passiva e dos verbos de controle, cruciais para uma posterior análise semântica automática, uma vez que, neste último caso, a estrutura argumental do verbo

subordinado é recuperada por um dos argumentos do verbo principal. Os demais elementos verbais modelados são alguns predicados complexos e alguns verbos modais.

O resultado é uma gramática com um amplo conjunto de fenômenos verbais, capaz de dar conta da análise de aspectos sintáticos cruciais para uma posterior interpretação semântica, como os verbos de controle, e que engloba aspectos bastante complexos de outras naturezas, como a predicação adjetival e a apassivação.

Os elementos nominais da gramática, de modo central, são os pronomes expletivos e reflexivos, e os casos de sintagmas nominais e determinantes com pronomes demonstrativos e interrogativos. Os demais constituintes gramaticais modelados são os sintagmas preposicionados, cuja complexidade se dá na distinção entre preposições semânticas e não semânticas; os sintagmas adjetivais, cuja projeção na sentença pode ocorrer a partir de formas adjetivas atributivas, de formas ordinais ou cardinais e na forma de intensificadores; e os sintagmas adverbiais, cuja estrutura interna foi modelada levando-se em consideração tanto advérbios intransitivos quanto transitivos com complemento PP.

Um ponto a ser ressaltado é que também modelamos aspectos sintático-discursivos para explicar a focalização de pronomes interrogativos. Outros fenômenos modelados, que não estão descritos de maneira central neste trabalho são a concordância nominal em torno do número, gênero e pessoa e flexão verbal de tempo, pessoa e número.

5.2 Avaliação da gramática

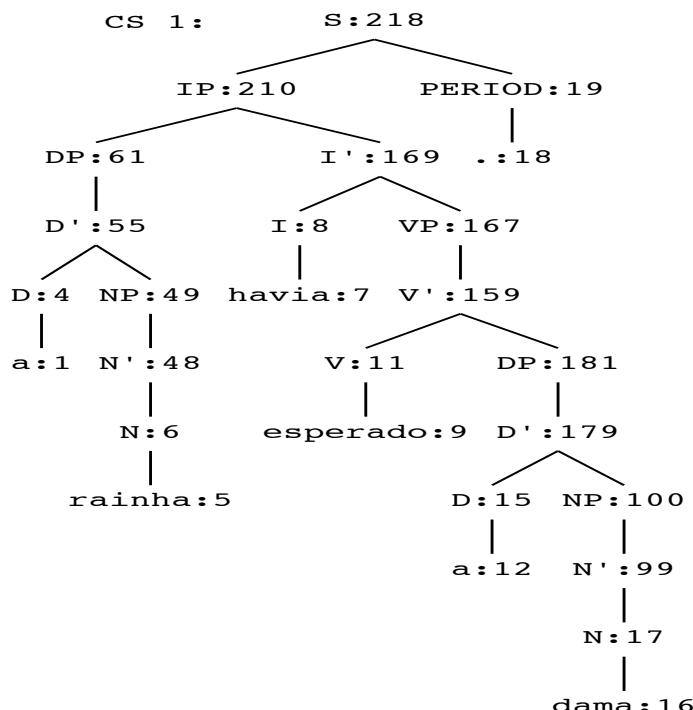
Para avaliarmos nossa gramática, adotamos uma metodologia baseada em dois parâmetros de avaliação: (i) análises corretas; e (ii) análises incorretas.

Primeiramente, consideramos como corretas aquelas análises que, independentemente da abordagem teórica e da terminologia adotada, não ferem as relações hierárquicas e de parentescos das estrutura e não violam as relações argumentais de um predicador. Vejamos o seguinte exemplo:

(135) A rainha havia esperado a dama.

Para a nossa gramática, *havia esperado* é um sintagma flexionado IP com o núcleo I preenchido, pois as marcas de tempo e pessoa são nele representados e não em um morfema adjungido à forma verbal lexical. Confira na Figura 69 abaixo.

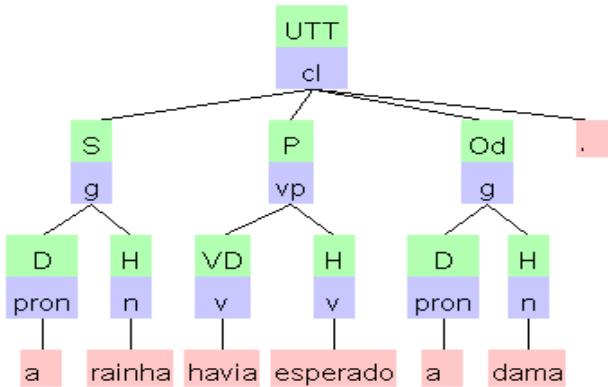
Figura 68 – Estrutura-c para uma construção com verbo auxiliar e verbo principal gerada pelo XLE.



Fonte: elaborada pela autora através do XLE

Para o *parser* Palavras o contexto verbal não se constitui de um sintagma flexionado, pois esta não é uma abordagem teórica da *Constraint Grammar* para a composição das locuções verbais. Confira, na figura 68 abaixo, a análise do Palavra para a sentença em (135).

Figura 69 – Análise do Palavras para uma construção com verbo auxiliar e verbo principal



Fonte: Visual Interactive Syntax Learning (BICK, 2000), último acesso em 15.12.2014

Entretanto isto não significa que a análise não seja adequada, pois não há razões para dizer que *havia* necessita necessariamente ser tratado como uma flexão I. Entendemos, desta forma que ambas as análises são adequadas do ponto de vista estrutural, independentemente da abordagem teórica e terminológica adotadas.

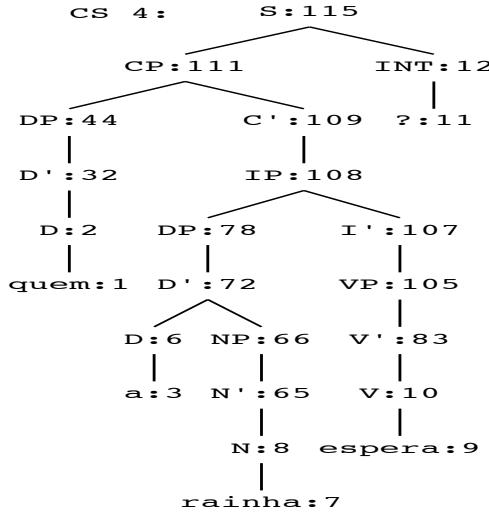
Em segundo lugar, consideramos como incorretas aquelas sentenças que, independentemente de ponto de vista teórica e dos termos utilizados para fazer referências aos objetos linguísticos, ferem condições estruturais ou argumentais internas das sentenças. Vejamos o seguinte exemplo:

(136) quem a rainha espera?

Nesta sentença, temos o fenômeno do deslocamento de pronome interrogativo para a periferia esquerda da sentença. Este deslocamento, no entanto, não nos impossibilita de interpretar a estrutura argumental do predicador e de reconhecer que a posição do NP *a rainha* esta reservada ao sujeito.

A nossa gramática é capaz de reconhecer que *quem* é um dos argumentos de *espera*, e que mesmo no início da sentença em forma de pronome interrogativo, continua sendo o objeto do verbo. Veja abaixo a análise apresentada por nossa gramática:

Figura 70 – Estrutura-c, gerada pelo XLE, para uma sentença interrogativa



Fonte: elaborada pela autora através do XLE

Figura 71 – Estrutura-f, gerada pelo XLE, para uma sentença interrogativa

"queim a rainha espera ?"

	PRED	'esperar<[7:rainha], [1:pro]>'
11	7	PRED 'rainha'
12	8	
9	65	NSEM [COMMON count]
10	66	NSYN common
83	127	CASE nom, GEND fem, HUMAN +, NUM sg, PERS 3
105	129	
107		
131	1	PRED 'pro'
3	2	NTYPE [NSYN pronoun]
4	32	CASE acc, NUM sg, PERS 3, PRON-FORM quem, PRON-TYPE int
109	44	
111	FOCUS [1:pro]	
115	CLAUSE-TYPE int, MOOD indicative, PASSIVE -, TENSE pres, VFORM finite, VTYPE main, XCOMP-FORM a	

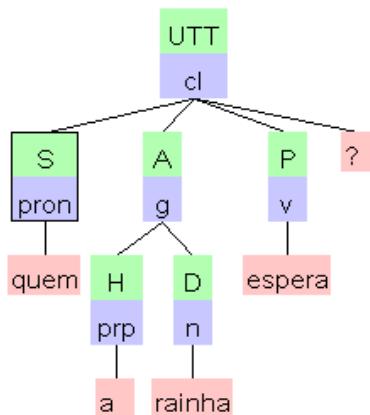
Fonte: elaborada pela autora através do XLE

Observe que, em termos de estrutura de constituintes, o objeto do verbo, representado pela forma pronominal *queim* ocupa a posição de especificador de CP. Deste modo, a posição de especificador de IP, que é própria dos sujeitos, fica reservada ao NP *a rainha*.

O parser Palavras, por sua vez, equivoca-se na análise da sentença *quem a rainha espera*. Vejamos, abaixo, a análise apresentada pelo Palavras:

Figura 72 – Análise do Palavras uma sentença interrogativa

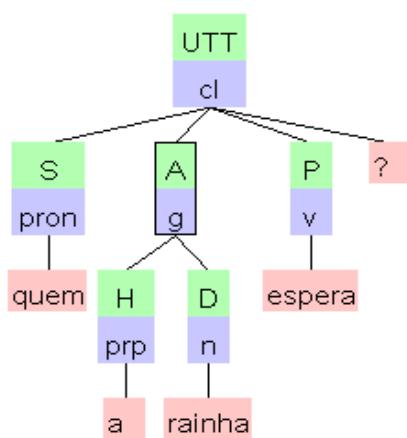
SUBJECT pronoun (indp 'quem' <interr> M/F S)



Fonte: Visual Interactive Syntax Learning (BICK, 2000), último acesso em 15.12.2014

Figura 73 – Análise do Palavras para uma sentença interrogativa

ADVERBIAL group (pp)



fonte: Visual Interactive Syntax Learning (BICK, 2000), último acesso em 15.12.2014

As figuras acima representam uma estrutura de constituintes de uma oração interrogativa deslocada. Observe que, na figura 72 à esquerda, *quem* está destacado como um pronome (pron) em posição de sujeito (S), mas já vimos que *quem* é, na verdade, o objeto do verbo.

Além disto, a figura 73 representa *a rainha* como um sintagma preposicionado (pp) exercendo uma função adverbial. Essa análise é bastante equivocada, pois, como vimos, *a rainha* cumpri o papel de sujeito da estrutura argumental de *espera*. Não há razão para que uma análise apresente uma inversão drástica de funções gramaticais para uma sentença interrogativa.

Avaliamos, com base nos parâmetros citados, todas as sentenças gramaticais do arquivo de teste *testfile.lfg* (Apêndice B) no *parser Palavras*. A Tabela 2, abaixo, nos mostra o número de erros e acertos do *parser*.

Tabela 2 – Avaliação das sentenças gramaticais do *testfile* (Apêndice B)

	Análises corretas	Análises incorretas
Santos (2014)	40	0
<i>Parser Palavras</i>	31	9

A nossa gramática apresenta 40 análises corretas, se pensarmos que estão em conformidade com a orientação gerativista.

Por outro lado, das 40 análises para as sentenças gramaticais do nosso *testfile.lfg* (Apêndice B), o *parser Palavras* apresentou 31 análises corretas e 9 análises incorretas. Incorretas, porque, independentemente da orientação teórica, seja funcionalista, gerativista ou estruturalista, apresentam erros estruturais ou gramaticais já consolidados entre as teorias sintáticas.

Isto significa que aproximadamente 80% das sentenças implementadas no *Palavras* possuem análises bem fundamentadas, em contraposição aos 20% das sentenças analisadas que possuem problemas teóricos que comprometeriam análises semânticas posteriores.

Além de avaliar o percentual de acerto das análises de uma gramática computacional para sentenças tidas como gramaticais, é necessário, também, avaliar se sentenças agramaticais são ou não analisadas por uma gramática de reescrita.

Para tanto, partimos da consideração da hipótese mentalista da linguagem, de que o conhecimento sintático que o falante de uma língua possui baseia-se, entre outras coisas, em um conjunto finito de regras gramaticais, capaz de analisar sentenças de uma determinada língua, se e somente se, as sentenças forem sentenças gramaticais desta mesma língua.

Deste modo, criamos, no mesmo arquivo das sentenças gramaticais, uma seção somente para sentenças agramaticais do português. Todas as sentenças agramaticais do nosso arquivo de teste (Apêndice B), além de terem sido analisadas pela nossa gramática, foram também analisadas pelo Palavras. Nesta avaliação contamos também com o julgamento de gramaticalidade de um avaliador especialista.

Confira, abaixo, os dados da Tabela 3.

Tabela 3 – Avaliação das sentenças gramaticais do *testfile* (Apêndice B)

	Sentenças gramaticais	Sentenças agramaticais
Especialista	0	20
Santos (2014)	18	2
<i>Parser</i> Palavras	1	19

Como vimos um pouco mais acima, uma sentença agramatical não deve receber nenhuma análise por parte das regras de uma gramática. Das 20 sentenças agramaticais do nosso *testfile.lfg* (Apêndice B), todas as 20 foram consideradas agramaticais por um especialista. Nossa gramática entende que 18 sentenças não devem receber análises, enquanto duas sentenças recebem. Por outro lado, o *parser* Palavras considera que 19 sentenças são gramaticais, enquanto que apenas 1 é realmente tida como agramatical e, portanto não recebe nenhuma análise.

A avaliação nos mostra que nossa gramática apresenta análises fundamentadas na vertente gerativista, conhecida como LFG, para as 40 sentenças dos *testfile.lfg*. Estas análises, embora possam levantar discussões com outras abordagens teóricas, são amplamente aceitas por estudiosos de orientação gerativa. Do modo como foi formalizada, a nossa gramática é capaz de fornecer análises sintáticas profundas bem fundamentadas para uma etapa posterior de interpretação semântica.

Notamos que as análises da nossa gramática se destacam em relação às análises do Palavras principalmente por causa das sentenças interrogativas, dos verbos de controle, das estruturas com valência três e passivas.

As sentenças interrogativas possuem, no caso das deslocadas, um reordenamento dos argumentos verbais, sem que haja modificações nas funções gramaticais. E,

principalmente, possuem um pronome interrogativo que deixa em aberto o conteúdo do argumento. É exatamente esta lacuna que motiva o uso das construções interrogativas.

Algo semelhante acontece com as orações com verbos de controle. Em construções desse tipo, o falante de uma língua é capaz de recuperar a informação argumental do verbo de uma estrutura encaixada, a partir de um dos argumentos do verbo da oração matriz. Desse modo, é importante para a interpretação da sentença que as análises automáticas forneçam, para um eventual processador semântico, esta mesma possibilidade que o falante tem de recuperar a estrutura argumentativa do verbo encaixado.

Nos casos de estruturas com três argumentos verbais nominais como em a rainha doou sua fortuna para o cavaleiro a distinção que a nossa análise faz para os dois objetos do verbo é muito importante para a compreensão do objeto mais diretamente ligado ao verbo, no caso sua fortuna.

As estruturas com apassivação, como sabemos, sofrem um reordenamento do constituintes da sentenças e também um remapeamento das funções gramaticais. Nestes casos, a inversão de funções gramaticais não implica um remapeamento dos papéis temáticos dos constituintes. Deste modo, capturando a inversão das funções sintáticas e mantendo a estrutura dos papéis temáticos, as análises fornecem dados mais precisos para tarefas de processamento semântico.

Por fim, com base na implementação de sentenças agramaticais, pode-se observar que as regras da nossa gramática, de um modo geral, conseguem capturar muito bem as propriedades internas das sentenças gramaticais e restringem, na maioria dos casos, sua atuação apenas às sentenças gramaticais. Isto se deve, principalmente, ao sofisticado formalismo da LFG, que, com o intuito de analisar os mais diversificados fenômenos das línguas, consegue, através da unificação de traços, criar restrições necessárias para a diminuição das análises e da hipergeração.

5.3 Trabalhos futuros

Nesta seção apresentamos algumas limitações do nosso trabalho e estabelecemos alguns direcionamentos para eventuais trabalhos futuros que possam ser desenvolvidos a partir do atual estágio da nossa gramática. Primeiramente, abordamos o problema de ambiguidade da nossa gramática e, em seguida, discutimos alguns ponto ligados à *pipeline* do XLE.

5.3.1 Problemas de ambiguidade na gramática

Nesta seção levantamos uma discussão em torno da problemática da ambiguidade da nossa gramática. A partir da avaliação da gramática pudemos constatar que uma das suas principais limitações está na capacidade de fazer fornecer o mínimo possível de análises, deixando um número maior de análises apenas para os casos reais de ambiguidade estrutural ou lexical.

Como podemos observar no Apêndice B, o arquivo *testfile.lfg* possui um conjunto de sentenças gramaticais e agramaticais do PB. O arquivo, após ser rodado por completo no sistema, reporta, em um novo arquivo, o número de estruturas funcionais, o tempo de análise e as sub-árvores unificadas da sentença, respectivamente.

Tais informações podem ser vistas entre parênteses ao lado de cada sentença do novo arquivo gerado, o *testfile.lfg.new*. Vejamos um exemplo retirado do *testfile.lfg.new* (APÊNDICE B):

(135) quem que a rainha espera ? (3 0.000 27)

Há três grupos de números nos parênteses acima, que poderíamos representar desta maneira ((3) (0.000) (27)) . (3) reporta o número de análises *f-structures*, (0.000) é o tempo de análise em CPU/seconds e (27) reporta o número de sub-árvores.

O teste revela que a gramática ainda apresenta certas ambiguidades, mesmo sendo claro que algumas das sentenças ambíguas possuem apenas uma leitura. Por exemplo:

(136) o cavaleiro quer esperar a dama . (72 0.000 66)

Esta sentença deveria ter apenas uma única estrutura-f, mas a gramática atribui a esse exemplo 72 análises. Essa ambiguidade excessiva é um defeito da gramática em seu estado atual, problema para ser corrigido em trabalhos futuros.

O conjunto das sentenças agramaticais apresenta, até o momento, apenas duas sentenças mal formadas do português que não deveriam ter análises. A gramática deveria ter excluído as seguintes sentenças:

(137) **a dama espera para o cavaleiro*. (9 0.000 35)

Nesta sentença deveria haver algum tipo de restrição implementada na regra dos complementos preposicionados de V.

(138) **fada chove*. (3 0.000 17)

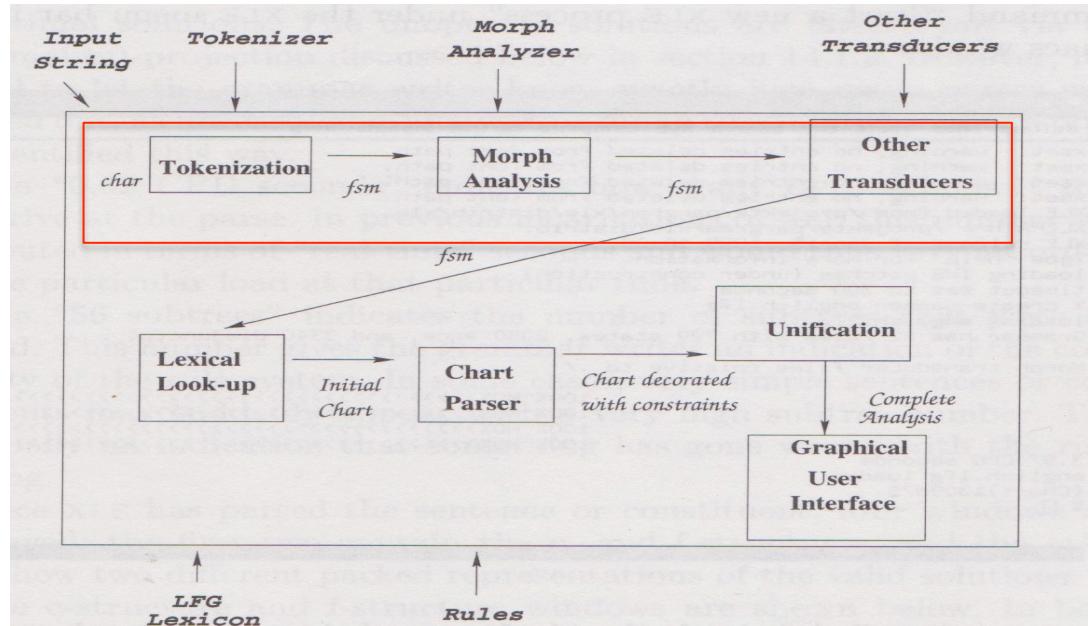
A sentença agramatical do exemplo acima precisa ser trabalhada no predicado da entrada lexical *chove*.

Além do problema da ambiguidade, apresentado pela nossa gramática, apresentamos na próxima seção alguns componentes da *pipeline* do XLE que podem, futuramente, ser introduzidos na gramática.

5.3.2 Algumas considerações sobre a pipeline do XLE

Além do problema da ambiguidade, apresentado pela nossa gramática, queremos chamar atenção para a figura abaixo, a qual apresenta a *pipeline* do XLE. Utilizamos esta figura para ilustrar quais aspectos da *pipeline* a nossa gramática não abarca.

Figura 74 – Arquitetura do *parser*



Fonte: Butt et al. 1999

Observe que, na *pipeline*, o processamento de língua natural se inicia com a tokenização, passando por uma análise morfológica e por um transdutor de estados finitos.

Do modo como foi desenvolvida, nossa gramática não contempla estas possibilidades. Diante desta impossibilidade, portanto, sinais de pontuação estão separados da última palavra da sentença por um espaço (ver APÊNDICE B), para que o *tokenizer* possa fazer a distinção entre esses elementos, caso contrário reconheceria o sinal de pontuação como uma parte da palavra.

Devemos ressaltar que o XLE oferece aos desenvolvedores de gramáticas que utilizam este sistema recursos que integram Tecnologia de Estados Finitos com o processamento sintático profundo (KAPLAN et al. 2004), como se pode observar na arquitetura do XLE apresentada na seção 3.2.

Isto significa que a morfologia de estados finitos pode ser integrada na gramática, permitindo, por exemplo, que haja uma única entrada lexical para a raiz de uma palavra, evitando ter que listar todas as suas formas de superfície. Fazendo-se a integração morfológica, só há necessidade de modelar no léxico raízes e lemas.

Estas são apenas sugestões para um eventual trabalho de continuidade da elaboração da gramática. Tais limitações não tornam a gramática menos importante no cenário do processamento sintático profundo do PB.

6 CONSIDERAÇÕES FINAIS

Nesta tese tivemos como objetivo apresentar uma gramática computacional da frase do português. O trabalho partiu de análises anteriores feitas para o português, para o inglês e para o francês e utilizou uma modelação nos termos da Teoria X-barra, contribuindo, desse modo, com uma gramática não trivial da frase do português, disponibilizando um recurso gramatical do português voltado para o processamento de linguagem natural.

Além de abranger um variado leque de fenômenos da sintaxe frasal do português, a nossa gramática oferece a perspectiva da LFG, teoria gerativa utilizada pelo XLE, sistema robusto e com uma amigável interface para o usuário, representando o estado da arte no que diz respeito aos aspectos de engenharia de desenvolvimento de gramáticas computacionais. Para alcançar o objetivo principal de escrever tal excerto de gramática, esta tese teve como objetivo específico fazer uma descrição das estratégias de formação das orações em português brasileiro.

A investigação apresentada destinou-se a aprofundar, teórica e empiricamente, a problemática da implementação computacional como fundamental no desenvolvimento de recursos linguísticos para o processamento de linguagem natural. Nos cinco capítulos, cada um com sua especificidade, foram salientados, no capítulo II, os princípios e os aspectos teóricos da relação existente entre teoria da gramática e formalismo. No capítulo III, foi apresentado o aparato utilizado pelos usuários do XLE, no âmbito do Projeto Pargram. Em seguida, no capítulo IV foram analisadas e implementadas as correspondências entre as estruturas de constituintes e as estruturas funcionais de um conjunto de 40 sentenças. O capítulo V traz uma avaliação das análises da nossa gramática em comparação com as análises das mesmas sentenças, feitas pelo *parser* Palavras. Na avaliação, também foi considerada a capacidade de ambos os analisadores reconhecerem sentenças agramaticais, não gerando-lhes análises.

Uma das vantagens das descrições no formalismo da LFG é a existência do XLE, que permite compilar um *parser* para uma dada gramática elaborada no formalismo. Por um lado, com base no XLE, dezenas de gramáticas no formalismo LFG vêm sendo desenvolvidas, no âmbito do Projeto Pargram, para línguas tipologicamente bastante distintas, como o inglês antigo, o urdu, o turco, o polonês, entre outras. Por outro lado, as gramáticas podem ser, e algumas já são, utilizadas por companhias que investem em tecnologia da informação, mais

especificamente em aplicações de processamento de linguagem natural de grande importância, nas áreas de extração e recuperação de informação, sistemas de perguntas e respostas, etc.

Em verdade, a necessidade crescente de recursos linguísticos computacionais, tais como: léxicos, gramáticas e redes semânticas, suscitou algumas dúvidas a respeito dos atuais recursos para o português e motivou um estudo aprofundado que revelou um número ainda pequeno de gramáticas computacionais e de um projeto contínuo de desenvolvimento de gramáticas para o PB, que garantissem as condições equitativas do processamento sintático profundo para outras línguas.

Neste momento, torna-se útil ressaltar que esta tese, no sentido de efetuar um contributo para a elaboração de um fragmento de gramática computacional, parte do conceito de regras de estrutura sintagmática na visão de uma teoria linguística não transformacional, ultrapassando a ótica gerativo-transformacional, tradicionalmente aplicado para a descrição sintática do português. Para tanto, torna-se essencial ressaltar que, apesar de darem uma perspectiva global e integrada do trabalho realizado, não podem, todavia, ser assumidas como soluções práticas absolutas para o fim do processamento sintático profundo da língua portuguesa.

Com base nesse contexto, a gramática, concebida e elaborada levando-se em consideração a descrição formalizada do português sob uma perspectiva gerativa, evidencia a importância das pesquisas desenvolvidas entre Sintaxe Formal e Processamento de Língua Natural. Congruente com o processamento de linguagem natural, observa-se a importância de o português brasileiro ser contemplado, cada vez mais, com gramáticas computacionais mais abrangentes e fundamentadas. Explica-se, desta forma, que a capacidade das aplicações de PLN de desempenharem suas tarefas depende diretamente da relação que cada recurso estabelece no processamento de léxico, gramática, rede semântica, etc.

A avaliação da gramática demonstrou que este recurso fornece mais de uma análise, mesmo em casos, em que as sentenças não possuem nem ambiguidade estrutural e nem lexical. Isto mostra que a gramática é ainda bastante ambígua, passível, portanto, de maior refinamento de suas regras.

Enfim, contribuímos com um recurso linguístico computacional que, diante das incompletudes observadas no que, atualmente, é tido como o estado da arte em gramática computacional, o *parser* Palavras, acrescenta, para diversos fenômenos, análises sintáticas

computacionais mais refinadas, que podem fornecer, como *input* para o processamento semântico, informações mais detalhadas para a interpretação semântica.

Em suma, é esta a proposta de análise e implementação de uma gramática computacional, produzida no interior de um contexto particular de uma tese de doutorado. O conceito de gramática computacional para o processamento de linguagem natural aqui defendida, tem por intenção despertar uma consciência de pesquisa acadêmica e industrial, com base no entendimento essencial da teoria e da aplicação, vislumbrando um caminho mais amplo para o desenvolvimento de recursos linguísticos.

REFERÊNCIAS

- ALENCAR, Leonel Figueiredo. Complementos verbais oracionais em português: uma análise léxico-funcional. **Revista Lingua(gem)/Instituto Latino-Americano de Pesquisas Científicas**, Santa Maria, v. 1, n.1, p. 173-218, 2004.
- _____. Modelação computacional de padrões variáveis de concordância em português. **Relin**. Belo Horizonte, v. 21, n.1, p.43-110, jan./jun. 2013a.
- _____. BrGram: uma gramática computacional de um fragmento do português brasileiro no formalismo da LFG. **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**, Fortaleza, p. 183–188, out. 2013b.
- _____. A passiva em português como construção predicativa adjetival: evidência morfológica e implementação computacional em LFG/XLE. **Estudos da Língua(gem)**, Vitória da Conquista. No prelo.
- ALEXANDRE, Nélia. Estruturas em movimento: alguns tópicos sobre as construções-Q e de clivagem. **Letras de Hoje**, Porto Alegre, v. 41, n.1, 99-119, 2006.
- BICK, E. **The parsing system “Palavras”**: automatic grammatical analysis of Portuguese in a Constraint Grammar framework. 2000. 505 f. Tese (Doutorado em Linguística.) – Departamento de Linguística, Universidade de Århus, Århus, 2000.
Disponível em: <<http://visl.sdu.dk/~eckhard/pdf/PLP20-amilo.ps.pdf>> Acesso em: 11 nov. 2013.
- BOSQUE, Ignacio, GUTIÉRREZ-REXACH, Javier. **Fundamentos de Sintaxis Formal**. Madrid: Ediciones Akal, 2009.
- BRANCO, António et al. Developing a Deep Linguistic Databank Supporting a Collection of Treebanks: the CINTIL DeepGramBank. **Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)**, 2010.
- _____, COSTA, Francisco, A Computational Grammar for Deep Linguistic Processing of Portuguese: LXGram, version A.4.1, **Relatório técnico**. Departamento de Informática, Universidade de Lisboa, 2008. Último acesso em: 30.03.2015
Disponível em : <<http://nlxgroup.di.fc.ul.pt/lxgram/>>
- _____, COSTA, Francisco. **HPSG**: Arquitetura. In: ALENCAR, Leonel Figueiredo, OTHERO, Gabriel de Ávila (Orgs). **Abordagens computacionais da teoria da gramática**. Campinas: Mercado de Letras, 2012a.

_____, COSTA, Francisco. Representação sintáctica em HPSG. In: ALENCAR, Leonel Figueiredo, OTHERO, Gabriel de Ávila (Orgs). **Abordagens computacionais da teoria da gramática**. Campinas: Mercado de Letras, 2012b.

_____, COSTA, Francisco. Representação do significado em HPSG. In: ALENCAR, Leonel Figueiredo, OTHERO, Gabriel de Ávila (Orgs). **Abordagens computacionais da teoria da gramática**. Campinas: Mercado de Letras, 2012c.

BRESNAN, Joan. **Lexical-Functional Syntax**. Oxford: Blackwell Publishers, 2001.

BUTT, Miriam J., KING, Tracy Holloway. Structural Topic and Focus Movement. In: **Proceedings of the First LFG Conference**, XRCE, Grenoble. CSLI Publications, 2006. Ultimo acesso: 02.10.2012 disponível em <<http://csli-publications.stanford.edu/LFG/>>

BUTT, Miriam, KING, Tracy Holloway, NIÑO, María-Eugenio, SEGOND, Frédérique. **A grammar writer's cookbook**. Stanford: CSLI publications, 1999.

CARNIE, Andrew. **Syntax: a generative introduction**. Oxford: Blackwell Publishers, 2002.

CHOMSKY, Noam. **Lectures on Government and Biding**. 2. ed. Dordrecht: Foris Publications, 1982.

CLARK, Stephen. Statistical Parsing. In: CLARK, Alexander, FOX, Chris, LAPPIN, Shalom. **The Handbook of Computational Linguistics and Natural Language Processing**. Oxford: Wiley-Blackwell Publishing, 2013, p. 333-363.

CROUCH, Dick, DALRYMPLE, Mary, KAPLAN, Ron, KING, Tracy, MAXWELL, John, NEWMAN, Paula. **XLE Documentation**. Palo Alto: Palo Alto Research Center, 2011. Disponível em <http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html>. Acesso em: 17 nov. 2014.

COSTA, Francisco; BRANCO, António. LXGram: A Deep Linguistic Processing Grammar for Portuguese. In: PARDO, T. A. S. et al. (Eds.). **Proceedings of the International Conference on Computational Processing of the Portuguese Language**, Porto Alegre, n. 9, p. 86-89, 2010.

DALRYMPLE, Mary, DYVIK, Helge, KING, Tracy. Copular complements: closed or open? **Proceedings of the Ifg04 conference**, Christchurch, New Zealand, jul. 2004.

DAVID, Karine Alves. **Sintaxe das expressões nominais no português do Brasil: uma abordagem computacional**. 2007. 117f. Dissertação (Mestrado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, 2007.

FALK, Yehuda. **Lexical-functional grammar**: an introduction to parallel constraint-based syntax. Stanford: CSLI Publications, 2001.

GAZDIK, Anna. French Interrogatives in an OT-LFG Analysis: An LFG account. **Proceedings of the LFG08 Conference**. Stanford: CSLI Publications, 2008.

GAZDIK, Anna. Multiple questions in French and in Hungarian: An LFG account. **Proceedings of the LFG10 Conference.** Stanford: CSLI Publications, 2010.

JURAFSKY, Daniel, MARTIN, James H. **Speech and Language Processing:** An introduction to natural language processing, computational linguistics, and speech recognition. 2. ed. New Jersey: Pearson Education, 2009.

KATO, Mary Aizawa. Word order change: the case of Brazilian Portuguese WH-questions. **XI International Congress of Historical Linguistics.** Los Angeles, UCLA, 1993.

_____, BRAGA, Maria Luiza, CORRÊA, Vilma Reche, ROSSI, Aparecida Lopes, SIKANSI, Nilmara Soares. As construções-Q no português brasileiro falado: perguntas, clivadas e relativas. In: KOCH, Ingodore G. Villaça (org.). **Gramática do Português Falado.** Desenvolvimentos. v. 6. Campinas: Editora da Unicamp, 2002.

KAPLAN, Ronald M. et al. Integrating finite-state technology with deep LFG grammars. **Combining Shallow and Deep Processing for NLP/ESSLLI,** Nancy, ago. 2004.

KING, Tracy Holloway. **Starting a ParGram Grammar.** 2004. Disponível em: <<http://www2.parc.com/isl/groups/nlt/xle/doc/PargramStarterGrammar/common.features.lfg>>. Acesso em: 17 nov. 2014.

_____, Tracy Holloway. **Configuring topic and focus in Russian.** Stanford: CSLI Publications, 1995.

KLENK, Ursula. **Generative Syntax.** Tübingen: Gunter Narr Verlag, 2003.

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

MARTINS, R.; OTHERO, G. A. Parsing do português. In: ALENCAR, L. F. de; OTHERO, G. A. (Org.). **Abordagens computacionais da teoria da gramática.** 1. ed. Campinas: Mercado de Letras, 2012, p. 99-126.

MIOTO, Carlos. WH que / WH é que. **XXVI Anais do Grupo de Estudos Linguísticos do Estado de São Paulo,** 1997.

_____, KATO, Mary. As interrogativas Q do português europeu e do português brasileiro atuais. **Revista da Abralin**, vol. 4, n. 1 e 2, p. 171-196, 2005.

_____, SILVA, Maria Cristina Figueiredo; LOPES, Ruth. **Novo Manual de Sintaxe.** Florianopolis: Insular, 2005.

MYCOCK, Louise. The Wh-expletive construction. **Proceedings of the LFG04 Conference.** Stanford: CSLI Publications: 2004.

_____. Wh-in-situ in constituents questions. **Proceedings of the LFG05 Conference.** Stanford: CSLI Publications: 2005.

_____. **The typology of Constituent Questions:** a Lexical-Functional Grammar analysis of ‘wh’- questions. 2009. 419 f. Tese (Doutorado em Linguística) – Faculdade de Humanidades, University of Manchester, Manchester, 2006.

_____. Constituent Question Formation and Focus: a new typological perspective. **Transactions of the Philological Society**, v. 105:2, p. 192-251, 2007.

MÜLLER, Stefan. **Grammatiktheorie**. Tübingen: Stauffenburg Verlag, 2013.

NEGRÃO, Esmeralda. Wh-extractions and relative clauses in Brazilian Portuguese. **D.E.L.T.A**, vol. 16, n. Especial, p. 141-163, 2000.

OTHERO, Gabriel de Ávila. **A gramática da frase em português**. Algumas reflexões para a formalização da estrutura frasal em português. EdiPUCRS: Porto Alegre, 2009.

OUCHIRO, Lívia. **Uma análise variacionista para as Interrogativas-Q**. 2011. Dissertação (Mestrado em Linguística) – Faculdade de Filosofia, Letras e Ciências, Universidade de São Paulo, 2011.

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: 17 nov. 2014.

PERINI, Mario A. **Estudos de gramática descritiva**: as valências verbais. São Paulo: Parábola, 2009.

_____, Mário A. **Gramática descritiva do português**. 4º. ed. São Paulo: Ática, 2000.

_____, Mario A. **Gramática do português brasileiro**. São Paulo: Parábola, 2010.

SULGER, Sebastian. **Irish Clefting: the LFG perspective**. 2009. 101 f. Dissertação (mestrado em linguística), Universidade de Konstanz, Konstanz, 2009.

RAPOSO, Eduardo. **Teoria da gramática. A faculdade da linguagem**. Lisboa: Editorial, Caminho, 1992.

RECH, Núbia Ferreira. Verbos de reestruturação no português brasileiro. **Revista do Gel**, São Paulo, v.8, n.1, p. 165 – 182, 2011.

ROSÉN, Victoria, SMEDT, Koenraad De, MEURER, Paul and Helge Dyvik. An open infrastructure for advanced treebanking. In HAJIĆ, Jan, SMEDT, Koenraad De, TADIĆ, Marko, BRANCO, António. (Orgs.) **META-RESEARCH Workshop on Advanced Treebanking at LREC2012**, Istanbul, p. 22–29, mai. 2012.

SANTOS, Andréa Feitosa. **Movimento do verbo e categorias vazias em i e v em um fragmento de gramática computacional do português.** 2009. 95f. Dissertação (Mestrado em Linguística) – Centro de Humanidades, Universidade Federal do Ceará, 2009.

SILVA, João et al. Out-of-the-box Robust Parsing of Portuguese. In: T.A.S. Pardo et al. (Eds.): **PROPOR 2010**, LNAI 6001, Porto Alegre, p. 75-85, 2010a.

SILVA, João et al. Top-Performing Robust Constituency Parsing of Portuguese: Freely Available in as Many Ways as you Can Get it. **Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)**, Malta, 2010b.

SCHWARZE, Christoph. **Lexikalisch-funktionale Grammatik.** Eine Einführung in 10 Lektionen mit französischen Beispielen. Fachgruppe Sprachwissenschaft der Universität Konstanz, Arbeitspapier Nr. 76, 1998.

SIKANSI, Nilmara Soares. Interrogativas Q- do português brasileiro moderno: quadro geral. **Cadernos de Estudos da Linguagem**, v. 34, p. 119-129, Jan./Jun. 1998.

ZWEIGENBAUM, Pierre. Attributive adjectives, adjuncts and cyclic f-structures in Lexical-Functional Grammar. DIAM. Département Intelligence Artificielle et Médecine. **Relatório Interno**, Paris, 2011.

APÊNDICE A – UM FRAGMENTO DE GRAMÁTICA DA FRASE DO PORTUGUÊS BRASILEIRO

Durante todo esta tese nós vimos elaborando um fragmento de gramática do português brasileiro. Neste apêndice nós coletamos a versão final de todas as regras e entradas lexicais que nós desenvolvemos.

BASIC PORT CONFIG (1.0)

ROOTCAT S.

LEXENTRIES (BASIC PORT).

TEMPLATES (BASIC PORT).

RULES (BASIC PORT).

GOVERNABLERELATIONS SUBJ OBJ OBJ-TH OBL OBL-AG COMP XCOMP.

SEMANTICFUNCTIONS ADJUNCT FOCUS.

NONDISTRIBUTIVES NUM PERS COMP-FORM.

EPSILON e.

OPTIMALITYORDER NOGOOD *Fragment "disprefer fragments and mark with *"

+OblAg. "prefer 'by' obliques in passives"

CHARACTERENCODING utf-8.

BASIC PORT RULES (1.0)

S --> {{IP:^=!

(! CLAUSE-TYPE)=decl; "usado pela oração principal para indicar o tipo de oração: declarativa"

(PERIOD)

|IP:^=!

(! CLAUSE-TYPE)=int; "tipo de oração: interrogativa"

(INT)}

|CP:^=!

(! CLAUSE-TYPE)=int;

(INT)}.

IP --> {DP: (^ SUBJ)=!
 (! CASE)=nom; "atribui caso nominativo"
 I';^=!
 |I';^=!;
 (DP: (^ SUBJ)=!
 (! CASE)=nom)}.

I' --> (I:^=!)

VP:^=!.

VP --> (CL: ^=!)

V';^=!;

PP*: ! \$ (^ ADJUNCT)
 (! CASE)=obl "usado para função gramatical que não é argumento
 subcategorizado"
 (! PTYPE)=c sem;
 AdvP*: ! \$ (^ ADJUNCT)
 (! ADV-TYPE)=c vpmmod;
 Adv*: ! \$ (^ ADJUNCT).

V' --> V:^=!;

(V': (^ XCOMP)=!)

(AP: (^ XCOMP)=!) "complemento aberto: o sujeito é proveniente de fora do predicado"
 (DP: (^ PREDLINK)=!
 (^ VTYPY)=c copular) "um complemento fechado em posição predicativa: o sujeito
 do predicado é o sujeito do verbo principal"
 (DP: (^ OBJ)=!
 (! CASE)=acc) "atribui caso acusativo"
 (PP: (^ OBJ)=!
 (! CASE)=acc

(! PFORM)=por)
 (PP: (^ OBL)=!
 (! CASE)=dat) "atribui caso dativo"
 (PP: (^ OBJ-TH)=!
 (! CASE)=obl) "atribui caso oblíquo"
 (PP: (^ OBL-AG)=!
 (! PFORM)=c por)"exemplo de uma equação de restrição: requere um valor com =c"
 (CP: (^ COMP)=!
 (! CLAUSE-TYPE)=decl) "complemento oracional fechado. tipo de oração:
 delcarativa"
 (CP: (^ XCOMP)=!) "complemento oracional aberto"
 (IP: (^ COMP)=!
 (! CLAUSE-TYPE)=nom) "complemento oracional fechado. tipo de oração:
 nominalizada".

CP --> { C':^=!
 |(DP: (^ OBJ)=!
 (! CASE)=acc
 (^ FOCUS)=!
 (! PRON-TYPE)=c int)
 C':^=!}
 Adv*: ! \$ (^ ADJUNCT).

C' --> (C:^=!)
 (IP:^=!)
 (VP:^=!)..

DP --> D':^=!.
 D' --> (D:^=!)
 (NP: ^=!)
 (PossP: ^=!)

(NumP; $\wedge=!$).

NP --> { {AP*: ! \$ (\wedge ADJUNCT)
 (! SUBJ)= \wedge
 (! ATYPE)=c attributive;
 N': $\wedge=!$
 |N': $\wedge=!$;
 AP*: ! \$ (\wedge ADJUNCT)
 (! SUBJ)= \wedge
 (! ATYPE)=c attributive}
 |N': $\wedge=!$ }.

N' --> N: $\wedge=!$.

PossP --> Poss': $\wedge=!$.

Poss' --> Poss: $\wedge=!$;
 (NP; $\wedge=!$).

NumP --> Num': $\wedge=!$.

Num' --> Num: $\wedge=!$;
 (NP; $\wedge=!$).

AP --> A': $\wedge=!$.

A' --> (AdvP: $\wedge=!$)
 A: $\wedge=!$;
 (PP: (\wedge OBL-AG)=!
 (! PFORM)=c por).

AdvP --> Adv': $\wedge=!$.

Adv' --> {AdvP:^=!;

 Adv':^=!

 |Adv:^=!

 |Adv:^=!;

 (PP:^=!)}..

PP --> P':^=!..

P' --> {P:^=! "núcleo do PP"

 (! PTYPE)=c sem;

 DP:(^ OBJ)=!

 |P:^=! "núcleo do PP"

 (! PTYPE)=c nosem;

 DP:^=!}..

BASIC PORT TEMPLATES (1.0)

V3SG = (^ SUBJ NUM)=sg

(^ SUBJ PERS)=3. "forma verbal na terceira pessoa do singular"

V3PL = (^ SUBJ NUM)=pl

(^ SUBJ PERS)=3. "forma verbal na terceira pessoa do plural"

PRESIND = @(TENSE pres) "tempo do verbo: presente"

@(MOOD indicative)."modo do verbo/oração: indicativo"

PASTIND = @(TENSE past) "tempo do verbo: passado"

@(MOOD indicative).

FUTIND = @(TENSE past) "tempo do verbo: futuro"
 @(MOOD indicative).

PRESSSUBJ = @(TENSE pres)
 @(MOOD subjunctive)."modo do verbo/oração: subjuntivo"

TENSE(T) = (^ TENSE) = T.

MOOD(M) = (^ MOOD) = M.

BITRANS(P) = (^ PRED)= $'P<(^ \text{SUBJ})(^ \text{OBJ})(^ \text{OBJ-TH})>'$. "OBL: sintagmas preposicionados normalmente subcategorizado pelo verbo" "DITRANS: verbo ditransitivo"

OPT-TRANS(P) = { @(TRANS P)
 | @((INTRANS P))."template para verbos que podem ser trans or intrans"

INTRANS(P) = (^ PRED)= $'P<(^ \text{SUBJ})>'$. "INTRANS: verbo intransitivo"

TRANS(P) = @(PASS (^ PRED)= $'P<(^ \text{SUBJ})(^ \text{OBJ})>'$). "TRANS: verbo transitivo / TRANS Template que permite regra lexical passiva"

TRANSIN(P) = @(PASS (^ PRED)= $'P<(^ \text{SUBJ})(^ \text{OBL})>'$). "TRANSIN: verbo transitivo indireto"

BITRANS-OBL(P) = (^ PRED)= $'P<(^ \text{SUBJ})(^ \text{OBJ})(^ \text{OBL})>'$. "BITRANS: verbo bitransitivo"

PASS(_SCHEMATA) = { "versão ativa" _SCHEMATA (^ PASSIVE)=
 | "versão passiva" _SCHEMATA (^ PASSIVE)=c +
 {(^ SUBJ) --> NULL "apaga o sujeito"
 |(^ SUBJ) --> (^ OBL-AG) "transforma o sujeito em sintagma oblíquo 'por'"
 |@((OT-MARK OblAg)}

$(^{\wedge} \text{OBJ}) \rightarrow (^{\wedge} \text{SUBJ})$ "transforma o objeto em sujeito" }.

OT-MARK(_mark) = _mark \$ o::*."assigns an OT mark, from common templates"

BASIC PORT LEXICON (1. 0)

"determinantes aqui"

a D * $(^{\wedge} \text{SPEC DET PRED})='o'$ "codifica determinantes"
 $(^{\wedge} \text{SPEC DET DET-TYPE})=\text{def}$
 $(^{\wedge} \text{NUM})=\text{sg}$
 $(^{\wedge} \text{GEND})=\text{fem};$ "tipo de determinante do: definido"
 P * $(^{\wedge} \text{PFORM})='a'$ "preposição que não tem um PRED"
 $(^{\wedge} \text{PTYPE})=\text{nosem}$ "a preposição encabeça um sintagma preposicionado exigido pelo verbo";
 C * $(^{\wedge} \text{XCOMP-FORM})='a'.$

as D * $(^{\wedge} \text{SPEC DET PRED})='o'$
 $(^{\wedge} \text{SPEC DET DET-TYPE})=\text{def}$
 $(^{\wedge} \text{NUM})=\text{pl}$
 $(^{\wedge} \text{GEND})=\text{fem}.$

o D * $(^{\wedge} \text{SPEC DET PRED})='o'$
 $(^{\wedge} \text{SPEC DET DET-TYPE})=\text{def}$
 $(^{\wedge} \text{NUM})=\text{sg}$
 $(^{\wedge} \text{GEND})=\text{mas}.$

os D * $(^{\wedge} \text{SPEC DET PRED})='o'$
 $(^{\wedge} \text{SPEC DET DET-TYPE})=\text{def}$

(^ NUM)=pl
 (^ GEND)=mas.

uma D * (^ SPEC DET PRED)='um'
 (^ SPEC DET DET-TYPE)=indef "tipo de determinante: indefinido"
 (^ NUM)=sg
 (^ GEND)=fem.

um D * (^ SPEC DET PRED)='um'
 (^ SPEC DET DET-TYPE)=indef "tipo de determinante: indefinido"
 (^ NUM)=sg
 (^ GEND)=mas.

"adverbs here"

mais Adv * (^ ADV-TYPE)=adjmod. "tipo de advérbio: modificador de adjetivo"

independentemente Adv * (^ ADV-TYPE)=vpmod.
 "tipo de advérbio: modificador de oração"

alegremente Adv * (^ PRED)='alegremente'
 (^ ADV-TYPE)=vpmod.

muito Adv * (^ PRED)='muito'
 (^ ADV-TYPE)=advmod.

"numerals here"

primeira Num * (^ SPEC NUMBER PRED)='primeiro' "números modificadores de nome"
 (^ SPEC NUMBER NUM)=sg
 (^ SPEC NUMBER GEND)=fem
 (^ SPEC NUMBER NUMBER-TYPE)=ordinal. "tipo do numeral"

duas Num * (^ SPEC NUMBER PRED)='dois' "números modificadores de nome"

(^ SPEC NUMBER NUM)=sg

(^ SPEC NUMBER GEND)=fem

(^ SPEC NUMBER NUMBER-TYPE)=cardinal. "tipo do numeral"

"pronomes aqui"

esta D * (^ SPEC DET PRED)='pro'

(^ SPEC DET PRON-TYPE)=demon "tipo de determinante: demonstrativo"

(^ SPEC DET DEIXIS)=proximal "demonstrativo que codifica/funciona como dêixis: proximal"

(^ SPEC DET PRON-FORM)=este

(^ PERS)= 3 "pessoa do demonstrativo: terceira"

(^ NUM)=sg

(^ GEND)=fem.

seu D * (^ SPEC POSS PRED)='pro' "codifica o possuidor do NP"

(^ SPEC POSS PRON-TYPE)=poss "tipo de pronome: possessivo"

(^ SPEC POSS PRON-FORM)=seu "forma do pronome possessivo: indefinido"

(^ SPEC POSS PERS)= 3 "pessoa do possessivo: terceira"

(^ NUM)=sg

(^ GEND)=mas.

sua D * (^ SPEC POSS PRED)='pro'

(^ SPEC POSS PRON-TYPE)=poss

(^ SPEC POSS PRON-FORM)=seu

(^ SPEC POSS PERS)= 3

(^ NUM)=sg

(^ GEND)=fem.

nossa D * (^ SPEC POSS PRED)='pro'
 (^ SPEC POSS PRON-TYPE)=poss
 (^ SPEC POSS PRON-FORM)=nosso
 (^ SPEC POSS PERS)= 3
 (^ NUM)=pl
 (^ GEND)=fem.

quem D * (^ PRED)='pro'
 (^ NTYPE NSYN)=pronoun "tipo sintático do nome: pronome"
 (^ PRON-TYPE)=int "tipo de pronome: interrogativo"
 (^ PRON-FORM)=quem
 (^ NUM)=sg
 (^ PERS)=3.

quais D * (^ PRED)='pro'
 (^ NTYPE NSYN)=pronoun "tipo sintático do nome: pronome"
 (^ PRON-TYPE)=int "tipo de pronome: interrogativo"
 (^ PRON-FORM)=qual
 (^ NUM)=sg
 (^ PERS)=3.

onde Adv * (^ PRED)='pro'
 (^ NTYPE NSYN)=pronoun "tipo sintático do nome: pronome"
 (^ PRON-TYPE)=int "tipo de pronome: interrogativo"
 (^ PRON-FORM)=onde.

se CL * (^ OBJ PRED)='pro'
 (^ OBJ NTYPE NSYN)=pronoun
 (^ OBJ PRON-TYPE)=refl
 (^ OBJ PERS)=3
 (^ OBJ CASE)=acc
 (^ OBJ NUM)=sg

(^ OBJ GEND)=mas.

"nouns here"

fada N * (^ PRED)='fada'

(^ NUM)=sg "número do nome: singular"

(^ GEND)=fem "gênero sintático do nome: feminino"

(^ PERS)=3

(^ NTYPE NSEM COMMON)=count "característica semântica do nome: comum (contável)"

(^ NTYPE NSYN)=common "tipo sintático do nome: (comum)"

(^ HUMAN)=-. "designação do nome: não-humano"

fadas N * (^ PRED)='fada'

(^ NUM)=pl "número do nome: singular"

(^ GEND)=fem "gênero sintático do nome: feminino"

(^ PERS)=3

(^ NTYPE NSEM COMMON)=count "característica semântica do nome: comum (contável)"

(^ NTYPE NSYN)=common "tipo sintático do nome: (comum)"

(^ HUMAN)=+. "designação do nome: não-humano"

dama N * (^ PRED)='dama'

(^ NUM)=sg

(^ GEND)=fem

(^ PERS)=3

(^ NTYPE NSEM COMMON)=count

(^ NTYPE NSYN)=common

(^ HUMAN)=+. "designação do nome: humano"

damas N * (^ PRED)='dama'

(^ NUM)=pl

([^] GEND)=fem
 ([^] PERS)=3
 ([^] NTYPE NSEM COMMON)=count
 ([^] NTYPE NSYN)=common
 ([^] HUMAN)=+.

rainha N * ([^] PRED)='rainha'
 ([^] NUM)=sg
 ([^] GEND)=fem
 ([^] PERS)=3
 ([^] NTYPE NSEM COMMON)=count
 ([^] NTYPE NSYN)=common
 ([^] HUMAN)=+.

lenço N * ([^] PRED)='lenço'
 ([^] NUM)=sg
 ([^] GEND)=mas "gênero sintático do nome: masculino"
 ([^] PERS)=3
 ([^] NTYPE NSEM COMMON)=count
 ([^] NTYPE NSYN)=common
 ([^] HUMAN)=-
 ([^] ANIM)=-."designação do nome: inanimado"

fortuna N * ([^] PRED)='fortuna'
 ([^] NUM)=sg
 ([^] GEND)=fem
 ([^] PERS)=3
 ([^] NTYPE NSEM COMMON)=count
 ([^] NTYPE NSYN)=common
 ([^] HUMAN)=-
 ([^] ANIM)=-.

prado N * (^ PRED)='prado'
 (^ NUM)=sg
 (^ GEND)=mas
 (^ PERS)=3
 (^ NTYPE NSEM COMMON)=count
 (^ NTYPE NSYN)=common
 (^ HUMAN)=
 (^ ANIM)=-.

conselho N * (^ PRED)='conselho'
 (^ NUM)=sg
 (^ GEND)=mas
 (^ PERS)=3
 (^ NTYPE NSEM COMMON)=count
 (^ NTYPE NSYN)=common
 (^ HUMAN)=
 (^ ANIM)=-.

cavaleiro N * (^ PRED)='cavaleiro'
 (^ NUM)=sg
 (^ GEND)=mas
 (^ PERS)=3
 (^ NTYPE NSEM COMMON)=count
 (^ NTYPE NSYN)=common
 (^ HUMAN)=+.

cavaleiros N * (^ PRED)='cavaleiro'
 (^ NUM)=pl
 (^ GEND)=mas
 (^ PERS)=3
 (^ NTYPE NSEM COMMON)=count
 (^ NTYPE NSYN)=common

(^ HUMAN)=+.

vaso N * (^ PRED)='vaso'

(^ NUM)=sg
 (^ GEND)=mas
 (^ PERS)=3
 (^ NTYPE NSEM COMMON)=count
 (^ NTYPE NSYN)=common
 (^ HUMAN)=-.

"adjectives here"

amável A * (^ SUBJ PRED)='amável<(^ SUBJ)>'

(^ SUBJ NUM)=sg
 (^ SUBJ GEND)=fem
 (^ ATYPE)=predicative;
 A * (^ PRED)='amável<(^ SUBJ)>'
 (^ SUBJ NUM)=sg
 (^ SUBJ GEND)=fem
 (^ ATYPE)=attibutive
 (^ DEGREE)=positive.

amáveis A * (^ SUBJ PRED)='amável<(^ SUBJ)>'

(^ SUBJ NUM)=pl
 (^ SUBJ GEND)=fem
 (^ ATYPE)=predicative;
 A * (^ PRED)='amável<(^ SUBJ)>'
 (^ SUBJ NUM)=pl
 (^ SUBJ GEND)=fem
 (^ ATYPE)=attibutive
 (^ DEGREE)=positive.

cintilante A * (^ PRED)='cintilante<(^ SUBJ)>'

(^ SUBJ NUM)=sg
 (^ SUBJ GEND)=fem
 (^ ATYPE)=attributive
 (^ DEGREE)=positive.

bela A * (^ PRED)='belo<(^ SUBJ)>'

(^ SUBJ NUM)=sg
 (^ SUBJ GEND)=fem
 (^ ATYPE)=attributive
 (^ DEGREE)=positive.

"conjunctions here"

que C * (^ COMP-FORM)=que.

"verbs here"

é V * {(^ PRED)='ser<(^ XCOMP)>(^ SUBJ)'
 (^ XCOMP SUBJ)=(^ SUBJ) "predicativo aberto"
 (^ XCOMP VFORM)=particip
 |(^ PRED)='ser<(^ PREDLINK)>(^ SUBJ)' "predicativo fechado"}
 @V3SG
 @PRESIND
 (^ VTYPY)=copular "tipo do verbo que encabeça a oração: cópula"
 (^ PASSIVE)=- "o verbo não está na forma passiva".

anda I * @V3SG

@PRESIND.

vai I * @V3SG

@FUTIND.

havia I * @V3SG

@PASTIND.

dá V * @(BITRANS dar)

@V3SG

@PRESIND

(^ VTYPE)=main "tipo do verbo que encabeça a oração: principal"

(^ PASSIVE)=-.

quer V * { (^ PRED)='querer<(^ SUBJ)(^ COMP)>'

(^ COMP MOOD)=subjunctive

|(^ PRED)='querer<(^ SUBJ)(^ XCOMP)>'

(^ XCOMP SUBJ)=(^ SUBJ)

(^ XCOMP VFORM)=infinitive}

@V3SG

@PRESIND

(^ VTYPE)=main

(^ PASSIVE)=-.

viu V * (^ PRED)='ver<(^ SUBJ)(^ COMP)>'

@V3SG

@PASTIND

(^ VTYPE)=main

(^ PASSIVE)=-

(^ PERF)=+.

começou V * (^ PRED)='começar<(^ SUBJ)(^ XCOMP)>'

(^ XCOMP SUBJ)=(^ SUBJ) "complemento aberto"

@V3SG

@PASTIND

(^ VTYPE)=modal

(^ PASSIVE)=-

(^ PERF)=-.

prometeu V * (^ PRED)='prometer<(^ SUBJ)(^ XCOMP)>'

(^ SUBJ)=(^ XCOMP SUBJ) "complemento oracional aberto: controle verbal"

@V3SG

@PASTIND

(^ VTYPE)=main

(^ PASSIVE)=-

(^ PERF)=+.

persuadiu V * (^ PRED)='persuadir<(^ SUBJ)(^ OBJ)(^ XCOMP)>'

(^ XCOMP SUBJ)=(^ OBJ) "complemento oracional aberto: controle verbal"

@V3SG

@PASTIND

(^ VTYPE)=main

(^ PASSIVE)=-

(^ PERF)=+.

disse V * (^ PRED)='dizer<(^ SUBJ)(^ COMP)>'

@V3SG

@PASTIND

(^ VTYPE)=main

(^ PASSIVE)=-

(^ PERF)=+.

queira V * (^ PRED)='quer<(^ SUBJ)(^ COMP)>'

(^ COMP MOOD)=subjunctive

@V3SG

@PRESSUBJ

(^ VTYPE)=main
 (^ PASSIVE)=-.

espera V * {@(TRANSIN esperar)
 |@(OPT-TRANS esperar)}
 @V3SG
 @PRESIND
 (^ VFORM)=finite
 (^ VTYPE)=main
 (^ PASSIVE)=-.

pergunta V * (^ PRED)='acreditar<(^ SUBJ)(^ COMP)>'
 (^ COMP MOOD)=indicative
 @V3SG
 @PRESIND
 (^ VTYPE)=main
 (^ PASSIVE)=-.

tem V * (^ PRED)='ter<(^ SUBJ)(^ XCOMP)>'
 (^ XCOMP SUBJ)=(^ SUBJ) "complemento aberto"
 @V3SG
 @PASTIND
 (^ VTYPE)=modal
 (^ PASSIVE)=
 (^ PERF)=-.

esperou V * @OPT-TRANS esperar
 @V3SG
 @PASTIND
 (^ VTYPE)=main
 (^ PASSIVE)=-.

esperar V * {@(OPT-TRANS esperar)
 |@(OPT-TRANS esperar)
 (^ XCOMP SUBJ)=(^ SUBJ)}
 (^ PASSIVE)=-
 (^ VFORM)=infinitive.

observar V * @OPT-TRANS observar
 (^ PASSIVE)=-
 (^ VFORM)=infinitive.

observa V * @OPT-TRANS observar
 @V3SG
 @PRESIND
 (^ PASSIVE)=-
 (^ VTYPY)=main.

esperando V * @OPT-TRANS esperar
 (^ PASSIVE)=-
 (^ VFORM)=gerundive.

esperado V * @OPT-TRANS esperar
 (^ PASSIVE)=-
 (^ VFORM)=particip;
 A * @OPT-TRANS esperar
 (^ PASSIVE)=+
 (^ DEGREE)=positive
 (^ ATYPE)=predicative.

espere V * @OPT-TRANS esperar
 @V3SG
 @PRESSUBJ
 (^ PASSIVE)=-.

esperem V * @OPT-TRANS esperar)

@V3PL
 @PRESSUBJ
 (^ VTYPE)=main
 (^ PASSIVE)=-.

esperam V * @OPT-TRANS esperar)

@V3PL
 @PRESIND
 (^ VTYPE)=main
 (^ PASSIVE)=-.

pensa V * (^ PRED)='pensar<(^ SUBJ)(^ COMP)>'

(^ COMP MOOD)=indicative
 @V3SG
 @PRESIND
 (^ VTYPE)=main
 (^ PASSIVE)=-.

acredita V * (^ PRED)='acreditar<(^ SUBJ)(^ COMP)>'

(^ COMP MOOD)=indicative
 @V3SG
 @PRESIND
 (^ VTYPE)=main
 (^ PASSIVE)=-.

acredite V * (^ PRED)='acreditar<(^ SUBJ)(^ COMP)>'

(^ COMP MOOD)=indicative
 @V3SG
 @PRESSUBJ
 (^ VTYPE)=main

(^ PASSIVE)=-.

doou V * @ (BITRANS-OBL doar)

@V3SG
 @PASTIND
 (^ VTYPE)=main
 (^ PASSIVE)=-
 (^ PERF)=+.

dormiu V * {@ (INTRANS dormir)}

|(^ PRED)='querer<(^ SUBJ)(^ XCOMP)>'
 (^ XCOMP SUBJ)=(^ SUBJ)}
 @V3SG
 @PASTIND
 (^ VTYPE)=main
 (^ PASSIVE)=-
 (^ PERF)=+.

mandou V * (^ PRED)='mandar<(^ SUBJ)(^ COMP)>'

@V3SG
 @PASTIND
 (^ VTYPE)=main
 (^ PASSIVE)=-
 (^ PERF)=+.

passam V * {@ (INTRANS passar)}

@V3PL
 @PRESIND
 (^ VTYPE)=main
 (^ PASSIVE)=-.

chove V * (^ PRED)='chover<>(^ SUBJ)'

([^] SUBJ PRON-TYPE)=expl_
 ([^] SUBJ PRON-FORM)=null
 ([^] VTYPE)=main
 ([^] PASSIVE)=-
 @V3SG
 @PRESIND.

saiba V * ([^] PRED)='saber<([^] SUBJ)([^] COMP)>'
 ([^] COMP MOOD)=indicative
 @V3SG
 @PRESSUBJ
 ([^] VTYPE)=main
 ([^] PASSIVE)=-.

sabe V * ([^] PRED)='saber<([^] SUBJ)([^] COMP)>'
 ([^] COMP MOOD)=indicative
 @V3SG
 @PRESIND
 ([^] VTYPE)=main
 ([^] PASSIVE)=-.

exige V * ([^] PRED)='exigir<([^] SUBJ)([^] COMP)>'
 ([^] COMP MOOD)=subjunctive
 @V3SG
 @PRESIND
 ([^] VTYPE)=main
 ([^] PASSIVE)=-.

quebra V * @(TRANS quebrar)
 ([^] VTYPE)=main
 ([^] PASSIVE)=-
 @V3SG

@PRESIND.

"prepositions here"

no P * (^ PRED)=*'em*<(^ OBJ)>
 (^ PTYPE)=sem
 (^ PSEM)=loc.

por P * (^ PFORM)=por
 (^ PTYPE)=nosem.

para P * (^ PFORM)=de
 (^ PTYPE)=nosem.

de P * (^ PFORM)=de
 (^ PTYPE)=nosem.

"punctuations here"

. PERIOD * (^ CLAUSE-TYPE)=decl.

? INT * (^ CLAUSE-TYPE)=int.

APÊNDICE B – ARQUIVO DE TESTE “testfile.lfg.new”

GRAMMATICAL SENTENCES

o cavaleiro espera a dama . (18 0.010 39)

#

quem a rainha espera ? (6 0.000 33)

#

a rainha espera quem ? (18 0.000 54)

#

a rainha espera onde ? (27 0.010 64)

#

quem que a rainha espera ? (3 0.000 27)

#

a rainha mandou o cavaleiro esperar . (27 0.000 42)

#

a rainha viu o cavaleiro esperando . (27 0.010 42)

#

o cavaleiro quer esperar a dama . (72 0.000 66)

#

a fada pensa que a rainha espera . (18 0.000 48)

#

a fada espera . (3 0.000 19)

#

passam os cavaleiros . (3 0.000 34)

#

a rainha doou sua fortuna para o cavaleiro . (27 0.000 46)

#

a dama espera o cavaleiro no prado . (27 0.010 45)

#

a rainha disse que o cavaleiro espera a dama . (108 0.000 68)

#

a dama quer que a fada espere . (18 0.000 50)

#

a dama saiba que a fada espera . (18 0.010 48)

#

esta dama é uma rainha . (9 0.000 30)

#

a fada é amável . (3 0.000 23)

#

o cavaleiro anda esperando a dama . (9 0.010 39)

#

a dama vai esperar a rainha . (9 0.000 39)

#

a rainha havia esperado a fada . (9 0.000 39)

#

o cavaleiro é esperado por uma fada . (9+45 0.010 73)

#

a fada começou a observar a dama . (36 0.000 70)

#

a dama prometeu observar o cavaleiro . (45 0.000 59)

#

a rainha persuadiu a dama a esperar o cavaleiro . (108 0.000 83)

#

chove . (1 0.000 8)

#

esta fada espera o cavaleiro . (9 0.000 30)

#

a fada quebra o vaso . (9 0.010 30)

#

o vaso se quebra . (3 0.000 21)

#

a fada espera por o cavaleiro . (18 0.000 36)

#

a dama espera por um cavaleiro no prado . (54 0.000 54)

#

a dama espera um cavaleiro no prado . (27 0.010 45)

#

a fada cintilante observa a dama . (3 0.000 38)

#

a amável dama espera o cavaleiro . (3 0.000 31)

#

a primeira dama espera o cavaleiro . (9 0.010 34)

#

as duas fadas esperam a rainha . (9 0.000 41)

#

a mais bela dama espera o cavaleiro . (6 0.000 37)

#

a dama observa a fada alegremente . (27 0.000 45)

#

a fada espera a dama independentemente de rainha . (54 0.010 56)

#

a fada observa a rainha muito alegremente . (9 0.000 51)

NON-GRAMMATICAL SENTENCES

#

quem onde a rainha espera ? (0 0.000 0)

#

a fada é amáveis . (0 0.000 17)

#

a dama espera para o cavaleiro . (9 0.000 35)

#

fada chove . (3 0.000 17)

#

a dama acredita que a fada espere . (0 0.000 42)

#

a dama acredite que a fada espere . (0 0.010 42)

#

a rainha sabe que o cavaleiro espere . (0 0.000 42)

#

a rainha saiba que o cavaleiro espere . (0 0.010 42)

#

a rainha quer que a fada espera . (0 0.000 43)

#

a rainha queira que a fada espera . (0 0.010 42)

#

a fada exige que a rainha espera . (0 0.000 42)

#

a fada exige espera o cavaleiro . (0 0.000 52)

#

a fada quer espera o cavaleiro . (0 0.010 54)

#

o cavaleiro é esperando por uma fada . (0 0.000 54)

#

o vaso quebra . (0 0.000 19)

#

o fada cintilante observa a dama . (0 0.000 10)

#

a primeira nossa dama espera o cavaleiro . (0 0.010 0)

#

a dama observa a alegmente fada . (0 0.000 0)

#

a fada espera a dama de a rainha independentemente . (0 0.000 39)

#

a bela mais dama espera o cavaleiro . (0 0.000 0)

62 sentences, 0.160 CPU secs total, 0.010 CPU secs max (11/19/14)