

**MIRNA FERNANDA DE OLIVEIRA**

**ENFRENTAMENTO DO PROBLEMA DAS  
DIVERGÊNCIAS DE TRADUÇÃO POR UM SISTEMA DE  
TRADUÇÃO AUTOMÁTICA: UM EXERCÍCIO  
EXPLORATÓRIO**

**ARARAQUARA  
2006**

**MIRNA FERNANDA DE OLIVEIRA**

**ENFRENTAMENTO DO PROBLEMA DAS  
DIVERGÊNCIAS DE TRADUÇÃO POR UM SISTEMA  
DE TRADUÇÃO AUTOMÁTICA: UM EXERCÍCIO  
EXPLORATÓRIO**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Lingüística e Língua Portuguesa, da Faculdade de Ciências e Letras – UNESP – Araraquara, como exigência parcial para a obtenção do título de Doutor em Lingüística e Língua Portuguesa.

**Orientador:** Prof. Dr. Bento Carlos Dias da Silva.

**ARARAQUARA  
2006**

Oliveira, Mirna Fernanda de  
Enfrentamento do problema das divergências de tradução  
por um sistema de tradução automática: um exercício  
exploratório. – 2006  
127 f. ; 30 cm

Tese (Doutorado em Lingüística e Língua Portuguesa) –  
Universidade Estadual Paulista, Faculdade de Ciências e Letras,  
Campus de Araraquara.

Orientador: Bento Carlos Dias da Silva

1. Lingüística.. 2. Tradução automática.  
3. Processamento Automático de Línguas Naturais (PLN)  
I. Título.

**Data de aprovação: 25/04/2006**

**Componentes da Banca Examinadora**

---

Orientador: Prof. Dr. Bento Carlos Dias da Silva – FCL – UNESP – Araraquara - SP

---

Profa. Dra. Beatriz Nunes de Oliveira Longo – FCL – UNESP – Araraquara – SP

---

Profa. Dra. Dirce Charara Monteiro – FCL – UNESP – Araraquara – SP

---

Profa. Dra. Gladis Maria de Barcellos Almeida – UFSCar - São Carlos - SP

---

Prof. Dr. Heronides Maurílio de Melo Moura – UFSC – Florianópolis - SC

**Ao Prof. Dr. Bento Carlos Dias da Silva**

## AGRADECIMENTOS

A realização deste trabalho só foi possível graças à colaboração direta ou indireta de muitas pessoas. Manifesto minha gratidão a todas elas e de forma particular:

a todos os professores do Programa de Pós-Graduação em Linguística e Língua Portuguesa, pelos valiosos ensinamentos;

aos funcionários da seção de Pós-Graduação em Linguística e Língua Portuguesa, pelo carinho e auxílio;

à CAPES, pelo apoio financeiro em forma de bolsa de estudos no país e no exterior, e ao MCT/ FINEP, por apoiar projetos interdisciplinares, sem os quais esta tese não poderia ter sido desenvolvida;

à Profa. Bonnie J. Dorr, Universidade de Maryland, pelo estágio supervisionado, e aos colegas Nitin, Fazil, Christof, Okan, Tikir, Stacy, Clara, Dina, Fernando, pela convivência e trabalho conjunto;

aos funcionários da Biblioteca e do Pólo Computacional da Faculdade de Ciências e Letras, por sua colaboração;

aos colegas e professores do Núcleo Interinstitucional de Linguística Computacional (NILC-São Carlos) pelo incentivo à pesquisa em Processamento Automático das Línguas Naturais em nosso país;

aos colegas Helio, Ariani, Maria Carolina, Ana Elisa, Aline, Luciana, pela amizade, companheirismo e troca de conhecimentos, e ao CELiC, que tem dado a infraestrutura para nosso trabalho;

à Profa. Dra. Maria Helena Galvão Frem Dias da Silva, por seu carinho e incentivo;

ao Prof. Dr. Bento Carlos Dias da Silva, pela partilha de conhecimentos e paciente orientação;

à Profa. Dra. Gladis Massini-Cagliari, pelo apoio;

a toda a minha família, pelo amor incondicional;

a todos os meus amigos, sem exceção, e em especial à Soraya, Helio, Iramaia, Magdiel, Albertina, Mônica, Leonor, Delfina, pelo carinho, alegria e amizade de todas as horas;

a Deus, por me permitir a dar mais um passo em minha vida.

A todos, meus profundos agradecimentos.

*A mim, parece que fui apenas um garoto brincando na praia, me divertindo aqui e acolá ao encontrar um pedregulho mais roliço, ou uma concha mais bonita que as comuns, enquanto o grande oceano da verdade se estendia, indecifrado, diante de mim.*

*Isaac Newton*

## RESUMO

O objetivo desta tese é desenvolver um estudo lingüístico-computacional exploratório de um problema específico que deve ser enfrentado por sistemas de tradução automática: o problema das divergências de tradução quer de natureza sintática quer de natureza léxico-semântica que se verificam entre pares de sentenças de línguas naturais diferentes. Para isso, fundamenta-se na metodologia de pesquisa interdisciplinar em PLN (Processamento Automático de Línguas Naturais) de Dias-da-Silva (1996, 1998 e 2003) e na teoria lingüístico-computacional subjacente ao sistema de tradução automática UNITRAN de Dorr (1993), que, por sua vez, é subsidiado pela teoria sintática dos Princípios e Parâmetros de Chomsky (1981) e pela teoria semântica das Estruturas Conceituais de Jackendoff (1990). Como contribuição, a tese descreve a composição e o funcionamento do UNITRAN, desenhado para dar conta de parte do problema posto pelas divergências de tradução, e ilustra a possibilidade de inclusão do português nesse sistema através do exame de alguns tipos de divergências que se verificam entre frases do inglês e do português.



## ABSTRACT

This dissertation aims to develop an exploratory linguistic and computational study of an specific type of problem that must be faced by machine translation systems: the problem of translation divergences, whether syntactic or lexical-semantic ones that can be verified between distinct natural language sentence pairs. In order to achieve this aim, this work is based on the interdisciplinary research methodology of the NLP (Natural Language Processing) field developed by Dias-da-Silva (1996, 1998 & 2003) and on the linguistic computational theory behind UNITRAN, a machine translation system developed by Dorr (1993), a system that is on its turn based on Chomsky's syntactic theory of Government and Binding (1981) and Jackendoff's semantic theory of Conceptual Structures (1990). As a contribution to the field of NLP, this dissertation describes the machinery of UNITRAN, designed to deal with part of the problem of translation divergencies, and it illustrates the possibility of including Brazilian Portuguese language in the system through the investigation of certain kinds of divergencies that can be found between English and Brazilian Portuguese sentences.

## Lista de figuras

<b>Figura 1.</b> As bases de dados e os módulos de processamento de um sistema de PLN genérico. ....	28
<b>Figura 2.</b> Representação da estrutura conceitual da sentença <i>John colocou o carro na garagem</i> .....	42
<b>Figura 3.</b> Funções conceituais da TEC construídas a partir das categorias ontológicas.....	45
<b>Figura 4.</b> Representação das restrições seletivas na TEC.....	51
<b>Figura 5.</b> Estrutura conceitual do verbo <i>comprar</i> .....	54
<b>Figura 6.</b> Estrutura conceitual da sentença <i>Maria comprou o livro de João</i> .....	55
<b>Figura 7.</b> A coindexação entre as estruturas sintática e conceitual (letras latinas) e a coindexação entre posições conceituais (letras gregas). ....	56
<b>Figura 8.</b> Representação das entradas transitiva e intransitiva do verbo <i>vestir</i> . ....	58
<b>Figura 9.</b> Unificação das duas representações da Figura 8.....	58
<b>Figura 10.</b> A arquitetura do UNITRAN e o fluxo do processo da tradução através dos seus componentes e módulos. ....	61
<b>Figura 11.</b> No destaque, os módulos de <i>Análise Gramatical</i> , de <i>Síntese Gramatical</i> e de <i>Restrições da Teoria dos Princípios e Parâmetros</i> . ....	65
<b>Figura 12.</b> A interação entre os módulos computacional e lingüístico na construção da estrutura sintática das sentenças.....	69
<b>Figura 13.</b> O <i>template</i> da estrutura sintagmática canônica de línguas do tipo S-V-O, como o inglês, o espanhol e o português. ....	71
<b>Figura 14.</b> Estruturas complexas, com oração encaixada, indicada pela porção da estrutura que está inclusa no contorno pontilhado, que dá conta de estruturas com verbo no infinitivo. ....	72
<b>Figura 15.</b> Os <i>templates</i> com elementos foneticamente nulos pré-construídos pelo UNITRAN. ....	81
<b>Figura 16.</b> O lugar da interlíngua no sistema UNITRAN, prevendo-se aí a estrutura conceitual. ....	85
<b>Figura 17.</b> O componente léxico-semântico é projetado de maneira que a composição e os processos de decomposição/realização ambos se baseiam nos mesmos mapeamentos sistemáticos e parâmetros na estrutura conceitual. ....	86
<b>Figura 18.</b> Mapeamento entre as estruturas sintática e conceitual ....	87
<b>Figura 19.</b> A composição da estrutura conceitual de <i>I stabbed John</i> . ....	92
<b>Figura 20.</b> Ilustração do processo de construção da estrutura conceitual da sentença do inglês <i>I stabbed John</i> consiste de cinco entradas no procedimento de composição.....	94
<b>Figura 21.</b> Algoritmo de geração da estrutura sintática da sentença de chegada.....	98
<b>Figura 22.</b> A realização sintática da sentença <i>Eu dei punhaladas em João</i> . ....	99
<b>Figura 23</b> As entradas para os verbos <i>entrar / enter</i> ilustram o uso do parâmetro *. ....	107
<b>Figura 24</b> O parâmetro * utilizado no processo de geração de <i>João entrou na casa</i> . ....	109
<b>Figura 25</b> Ilustração do uso do parâmetro :PROMOTE.....	115
<b>Figura 26</b> Exemplo de uso do parâmetro :DEMOTE.....	116
<b>Figura 27</b> Exemplo de uso do parâmetro :CONFLATED.....	117
<b>Figura 28</b> Ilustração da aplicação do parâmetro :CONFLATED.....	118
<b>Figura 29</b> Ilustração da aplicação do parâmetro :CAT.....	120

## Lista de quadros

<b>Quadro 1.</b> Restrições da teoria de Casos.....	38
<b>Quadro 2.</b> Principais categorias ontológicas da TEC.....	43
<b>Quadro 3.</b> Funções e primitivos conceituais do UNITRAN .....	47
<b>Quadro 4.</b> A estrutura de argumentos na dimensão espacial.....	48
<b>Quadro 5.</b> Restrições da dimensão espacial .....	48
<b>Quadro 6.</b> Restrições da dimensão causal.....	48
<b>Quadro 7.</b> Restrições da dimensão campo .....	49
<b>Quadro 8.</b> Os dezenove parâmetros do módulo “Restrições da Teoria dos Princípios e Parâmetros” (dez dos quais são aplicados antes do processamento das sentenças), distribuídos em sete sub- módulos independentes .....	67
<b>Quadro 9.</b> O parâmetro Categorias Básicas .....	69
<b>Quadro 10.</b> O parâmetro Ordem dos Constituintes.....	72
<b>Quadro 11.</b> O parâmetro Especificadores de Base.....	73
<b>Quadro 12.</b> O parâmetro Adjuntos de Base .....	73
<b>Quadro 13.</b> O parâmetro Regentes Apropriados .....	75
<b>Quadro 14.</b> O parâmetro Especificadores Deslocados .....	76
<b>Quadro 15.</b> O parâmetro Os Nós de Fronteira .....	76
<b>Quadro 16.</b> O parâmetro Adjuntos Deslocados.....	76
<b>Quadro 17.</b> O parâmetro Tipo de Regência .....	78
<b>Quadro 18.</b> O parâmetro Propriedades de Caso.....	78
<b>Quadro 19.</b> O parâmetro Vestígios .....	79
<b>Quadro 20.</b> O parâmetro Pronomes Vazios.....	79
<b>Quadro 21.</b> O parâmetro Categorias Funcionais Vazias .....	80
<b>Quadro 22.</b> O parâmetro Licenciamento de pro.....	81
<b>Quadro 23.</b> O parâmetro Condições sobre Cadeias.....	81
<b>Quadro 24.</b> O parâmetro Domínio de Regência.....	82
<b>Quadro 25.</b> O parâmetro Clíticos .....	83
<b>Quadro 26.</b> O parâmetro Duplicação de Clíticos .....	83
<b>Quadro 27.</b> O parâmetro Non-drop.....	84
<b>Quadro 28.</b> As entradas lexicais construídas (RLCS) para os verbos da classe associada ao primitivo IR Poss. ....	89
<b>Quadro 29.</b> Os parâmetros do Componente Léxico-Semântico.....	95
<b>Quadro 30.</b> Divergências léxico-semânticas tratadas pelo UNITRAN e os parâmetros que as resolvem.....	106

## **Lista de abreviaturas e siglas**

ADV - advérbio

ALPAC – Automatic Language Processing Advisory Committee

C - Complementizador

CLCS – Estrutura léxico-conceitual construída

CONC - Concordância

Det - Determinante

ELS – Estrutura léxico-semântica

EXT – argumento externo

FLEX - flexão

Ing.- Inglês

FL – Forma lógica

NP – Sintagma Nominal

NUM – Número

OBJ - Objeto

P - Preposição

PLN – Processamento de Línguas Naturais

Port. - Português

RGL – Rotina Generalizada de Ligação

RLCS - Estrutura Léxico-Conceitual da Unidade Lexical

S - Sentença

SN – Sintagma Nominal

SUJ - Sujeito

SV – Sintagma Verbal

TA – Tradução Automática

TEC – Teoria de Estruturas Conceituais

TPP – Teoria de Princípios e Parâmetros

V - Verbo

## **Lista de símbolos**

$\theta$  - referente a papéis temáticos

## SUMÁRIO

INTRODUÇÃO.....	15
SEÇÃO I - OS ESTUDOS DE TRADUÇÃO E A TRADUÇÃO AUTOMÁTICA: O CONTEXTO DA PESQUISA.....	16
1.1 Os quatro momentos dos estudos de tradução.....	16
1.2 A tradução automática: um breve histórico.....	17
1.3 A TA em língua portuguesa: uma perspectiva.....	19
1.4 Tipos de sistemas.....	21
1.5 Paradigmas de TA.....	21
1.6 Os desafios da TA.....	23
1.7 A abordagem do PLN em três domínios.....	26
1.8 A base lexical e o sistema de PLN.....	27
SEÇÃO 2 – INVESTIGAÇÃO DO DOMÍNIO LINGÜÍSTICO: AS BASES TEÓRICAS DO UNITRAN.....	30
2.1 A Teoria de Princípios e Parâmetros.....	30
2.1.1 A Teoria X-barra.....	31
2.1.2 A Teoria da Regência.....	34
2.1.3 A Teoria da Fronteira.....	37
2.1.4 A Teoria dos Casos.....	37
2.1.5 A Teoria de Ligação.....	39
2.1.6 A Teoria- $\theta$ .....	40
2.1.7 A Teoria do Controle.....	41
2.2 A Teoria das Estruturas Conceituais.....	41
2.2.1 Princípios.....	41
2.2.2 Papéis temáticos.....	49
2.2.3 Restrições seletivas.....	51
2.2.4 Dois tipos de coindexação.....	53
2.2.5 Sobre a ligação sintática.....	56
2.2.6 A teoria da ligação e os elementos implícitos.....	59
SEÇÃO 3 – INVESTIGAÇÃO DO DOMÍNIO REPRESENTACIONAL I: A DESCRIÇÃO DO SISTEMA UNITRAN E UM EXERCÍCIO DE ESPECIFICAÇÃO DAS DIVERGÊNCIAS SINTÁTICAS DE TRADUÇÃO PARA O PORTUGUÊS....	60
3.1 O sistema UNITRAN: um panorama.....	60
3.2 A arquitetura e o funcionamento do UNITRAN.....	61
3.3 Os módulos de processamento do UNITRAN.....	63
3.3.1. Os Módulos de Análise e de Síntese Morfológica.....	63
3.3.2 Os Módulos de Análise e de Síntese Gramatical.....	64
3.3.3 Os Sub-módulos do Módulo de Restrições da Teoria dos Princípios e Parâmetros.....	70
3.3.3.1 O sub-módulo X'.....	70
3.3.3.5 O sub-módulo Vestígios.....	78
3.3.4 O Componente Léxico-Semântico.....	84
3.3.4.7 A geração da estrutura sintática da sentença da língua de chegada.....	96
3.3.4.8 O Cálculo da Estrutura Sintática da Sentença de Chegada.....	96
SEÇÃO 4 – INVESTIGAÇÃO DO DOMÍNIO REPRESENTACIONAL II: UM EXERCÍCIO DE ESPECIFICAÇÃO DAS DIVERGÊNCIAS LÉXICO-SEMÂNTICAS DE TRADUÇÃO PARA O PORTUGUÊS.....	101

4.1 Preliminares .....	101
4.2 Classificação das divergências de tradução léxico-semânticas .....	106
4.2.1 Divergência Formal (parâmetro *) .....	107
4.2.2 Divergência Temática (parâmetros :INT / :EXT).....	110
4.2.3 Divergência de promoção de não núcleo de L1 para núcleo de L2: (parâmetro :PROMOTE).....	114
4.2.4 Divergência de rebaixamento de núcleo de L1 para não núcleo de L2 (parâmetro :DEMOTE).....	116
4.2.5 Divergência de incorporação lexical (parâmetro :CONFLATED).....	117
4.2.6 Divergência Categorical (parâmetro :CAT) .....	119
SEÇÃO 5 - CONCLUSÕES .....	122
REFERÊNCIAS BIBLIOGRÁFICAS .....	125

## INTRODUÇÃO

O objetivo desta tese é desenvolver um estudo lingüístico-computacional exploratório de um problema específico que deve ser enfrentado por sistemas de tradução automática: o problema das divergências de tradução quer de natureza sintática quer de natureza léxico-semântica que se verificam entre pares de sentenças de línguas naturais diferentes. Para isso, fundamenta-se na metodologia de pesquisa interdisciplinar em PLN (Processamento Automático de Línguas Naturais) de Dias-da-Silva (1996, 1998 e 2003) e na teoria lingüístico-computacional subjacente ao sistema de tradução automática UNITRAN de Dorr (1993), que, por sua vez, é subsidiado pela teoria sintática dos Princípios e Parâmetros de Chomsky (1981) e pela teoria semântica das Estruturas Conceituais de Jackendoff (1990).

Na seção 1, que visa a fornecer o contexto dentro do qual as seções subseqüentes encontram sua ancoragem, apresentamos uma breve reflexão sobre o campo dos Estudos de Tradução e sobre a Tradução Automática (TA), dos sistemas de TA disponíveis no mercado e dos primeiros esforços em TA do português. Na seção 2, esta tese desenvolverá o Domínio Lingüístico da pesquisa, em que forneceremos os subsídios teóricos essenciais que sustentam as discussões posteriores sobre o sistema UNITRAN: a teoria gerativista na fase dos Princípios e Parâmetros e a teoria semântica das Estruturas Conceituais. Na seção 3, a tese abordará as questões do Domínio Representacional, em que discutiremos a composição e o funcionamento do sistema UNITRAN e uma ilustração de como o português pode ser inserido nesse sistema, com a especificação dos parâmetros que lhe são próprios. Na seção 4 a tese abordará ainda o Domínio Representacional, em que analisaremos diferentes tipos de divergências de tradução e como o UNITRAN dá conta deles, ao mesmo tempo em que discutiremos alguns tipos de divergências de tradução entre frases do inglês e do português. Finalmente, na seção 5, esta tese apresentará a conclusão, em que indicaremos as principais contribuições da pesquisa para a investigação científica interdisciplinar e, sobretudo, para a construção de sistemas de TA lingüisticamente motivados.



# SEÇÃO I - OS ESTUDOS DE TRADUÇÃO E A TRADUÇÃO AUTOMÁTICA: O CONTEXTO DA PESQUISA

## 1.1 Os quatro momentos dos estudos de tradução

Ainda que seja uma disciplina em busca de uma identidade própria (cf. ARROJO, 1998), os Estudos da Tradução, desde seus primórdios documentados na antiga Grécia até a pós-modernidade, têm acolhido em seu arcabouço uma série de propostas e metodologias de investigação que conferem à área os contornos de uma disciplina bem caracterizada, com discussões pertinentes e um grande número de trabalhos sendo produzidos, tanto do ponto de vista da teoria da tradução como do ponto de vista de inúmeras traduções de títulos de outras línguas.

De acordo com Arrojo (1998), para efeito didático, os Estudos da Tradução podem ser divididos em quatro momentos, posto que eles não são absolutamente estanques.

O primeiro momento inicia-se na Antigüidade grega, passa pela Idade Média e alcança o século XIX. Nessa extensa fase, destaca-se o conhecido princípio de Cícero, que afirma que não se deve traduzir *verbum pro verbo*, discutido no *Libellus de optimo genere oratorum*, de 46 A.C., e mais adiante, em sua *Arte Poética*. Nota-se, nesses primeiros escritos, uma reflexão sobre a teoria da tradução advinda diretamente da prática, com a adesão de autores como Lutero, du Bellay, Montaigne, Chapman, Jacques Amyot, Ben Jonson, Pope, Dryden e Rochefort (ARROJO, 1998, p. 3). Mesmo os textos teóricos, como o *De interpretatione recta*, de Leonardo Bruni, de 1420, e *De optimo genere interpretandi*, de Pierre Daniel Huet (1680), têm como característica comum seu foco empírico, conforme o *Essay on the Principles of Translation*, de Alexander Fraser Tytler (1792), e o *Ueber die verschiedenen Methoden des Ubersetzens* (1813), de Friedrich Schleiermacher (STEINER, 1975, p. 236-237).

O segundo momento, que se estende dos meados do século XIX até a década de 30, ainda se caracteriza pela reflexão entre teoria e prática, porém os trabalhos assumem caráter filosófico e abordam a tradução do ponto de vista dos estudos da linguagem e da

mente. É o que se depreende através dos trabalhos de Goethe, Schopenhauer, Matthew Arnold, Paul Valéry, Ezra Pound, I.A. Richards, Benedetto Croce, Walter Benjamin e Ortega y Gasset (STEINER, 1975, p. 237).

O terceiro momento, que abrange as décadas de 40 e 60, é marcado pela publicação do *Sous l'invocation de Saint Jérôme*, de Valery Larbaud, de 1946, no qual se incluem os primeiros trabalhos sobre a tradução automática e a aplicação de teorias lingüísticas e estatísticas desenvolvidas por formalistas russos e tchecos. Destacam-se, nesse momento, *Word and Object*, publicado por Quine (1960), e *Introdução à Teoria de Tradução*, de Andrei Fedorov, originalmente publicado em russo, em 1953. Também nascem nesse momento as primeiras associações de tradutores, refletindo a busca pela profissionalização e pela própria demarcação da área enquanto disciplina.

O quarto momento, iniciado na década de 70, ainda traz consigo a influência do período anterior, com trabalhos guiados pela lógica, pelas análises contrastiva, literária e semântica e pelos estudos comparados, além do surgimento do estudo da hermenêutica da tradução. A partir de então, as reflexões sobre a teoria e a prática da tradução transformaram-se em interface entre a filologia clássica, a literatura comparada, a psicologia, a antropologia, a sociologia, a etnolingüística e a sociolingüística (ARROJO, 1998, p. 05).

## 1.2 A tradução automática: um breve histórico

Paralelamente aos Estudos da Tradução, conforme sinalizado no terceiro período, surgiu o campo genericamente denominado Tradução Automática (doravante TA), que pode ser concebido como uma subárea do Processamento Automático das Línguas Naturais (PLN) e definido como a “tradução de um texto elaborado em uma língua natural para outro texto, em outra língua, por meio de um sistema computacional” (DORR, JORDAN & BENOIT, 1999).

Os esforços que se concentraram na elaboração de sistemas de TA partiram de pesquisadores nos Estados Unidos e na Europa e tiveram início na década de 50, estendendo-se até a década de 1960, até que, em 1966, o relatório da ALPAC (*Automatic Language Processing Advisory Committee*), comissão norte-americana de avaliação de pesquisas sobre PLN, reprovou esse tipo de empreendimento, alegando a

baixa qualidade da tecnologia desenvolvida até então e o seu alto custo, em face à disponibilidade dos inúmeros e competentes tradutores humanos.

Cumprе esclarecer que as primeiras tentativas de construção de sistemas de TA falharam essencialmente pela alta expectativa de produzir traduções perfeitas, sem a adequada fundamentação teórica. Além disso, o relatório da ALPAC acabou por provocar uma redução do investimento na pesquisa em TA nos Estados Unidos e o fomento foi direcionado para áreas correlatas em emergência, como a Linguística Computacional e a Inteligência Artificial.

Ironicamente, a pesquisa nesses dois últimos campos acabou por fornecer à TA as bases teóricas de que ela tanto precisava. Nesse ponto, ao lado da reduzida pesquisa em centros acadêmicos, destacaram-se os empreendimentos comerciais, como o desenvolvimento dos sistemas *Systran*<sup>1</sup> (EUA, final dos anos 1950) e *Logos* (EUA, 1972).

No final da década de 60, destaca-se a participação do Canadá, com suas pesquisas em TA motivadas pelo *status* anglo-francês do país.

Nas décadas de 70 e 80, surgiram dois esforços significativos que impulsionaram o desenvolvimento do campo. O primeiro foi o Projeto Eurotra, proposto pela União Européia, que objetivou construir sistemas de TA para as línguas de todos os países membros, por razões eminentemente econômicas, a fim de diminuir os custos com a tradução de documentos. O segundo, também motivado por questões econômicas, mas também políticas, foi a percepção do governo e das indústrias do Japão de que a TA do japonês, para e a partir de línguas européias, e posteriormente entre línguas das nações asiáticas, constituiria uma tecnologia linguística decisiva para o seu progresso econômico.

Nos Estados Unidos, a pesquisa e o desenvolvimento comercial de sistemas retornaram, na década de 1980, com novos investimentos governamentais para a área. Os esforços concentram-se nas universidades, com destaque para a *Carnegie Mellon University*, a *Georgetown University*, *University of Texas*, entre outras, embora os maiores interessados e fomentadores da pesquisa em TA norte-americana têm sido os militares.

---

<sup>1</sup> Trata-se de um dos sistemas de TA que chegou até nossos dias e está disponível para compra no endereço eletrônico <http://www.systransoft.com/index.html> por US\$899,00. (acesso em 20/01/2006)

Em suma, a pesquisa em TA, com seus sucessos e falhas, tem uma trajetória de mais de cinquenta anos.

### 1.3 A TA em língua portuguesa: uma perspectiva

De acordo com Hutchins (1997, p. 6), as primeiras experiências em TA voltadas para o português datam de 1948, quando da proposta de Richens, e somente publicadas em 1952, em versão revisada, por Richens & Booth (op. cit.). Os primeiros registros foram relatados em um memorando e foram testados tanto à mão quanto pelo uso de máquinas de cartões perfurados, dada a ausência, na época, de máquinas com grande poder de armazenamento e manipulação de dados.

Richens propôs que as palavras fossem reduzidas à sua raiz, separando-se suas flexões, tanto para compactar a extensão do dicionário do sistema quanto para a introdução de informação de natureza gramatical. Por exemplo, para o verbo latino *amat*, uma busca era feita pela maior raiz (*am-*) e pela flexão (*-at*), em que é colocada em correspondência com o verbo *amar* e a flexão com a terceira pessoa do singular. Assim, a anotação gramatical acabava por estender o alcance do sistema de TA palavra por palavra para um sistema capaz de fazer generalizações sobre as formas da língua. Tendo a proposta sido utilizada para várias línguas, além do português e do inglês<sup>2</sup>, o resultado obtido, conforme pode ser observado abaixo, ainda apresentava limitações:

(German) Wenn in ein\*em gröss\*eren Gebiet zwei Form\*en neben\*einander leb\*em ,ohne sich zu vermisch\*en, so gehö\*en sie verschieden\*en Form\*en\*kreis\*en an.

(English) if in a/one *d* large (more) area two form *m* beside one another live *z* without self to/too mix *z*, so belong/hear *p z z* different *m* form *m* circle *m* at

Os asteriscos no texto em alemão indicam segmentações automáticas. No texto traduzido em inglês, *d* indica uma palavra do alemão no caso dativo, *p* um verbo no passado, *m*, o número (múltiplo, plural ou dual), *z* indica não-especificidade e as barras, traduções alternativas (cf. HUTCHINS, 1997, p. 6).

Outros trabalhos pioneiros do PLN com a língua portuguesa incluem o do *Centre de Lexicologie et Stylistique des Langues Espagnoles et Portugaise* (França,

---

<sup>2</sup> Dentre as outras línguas, incluem-se: albanês, dinamarquês, holandês, finlandês, francês, húngaro, indonésio, italiano, latim, norueguês, polonês, letoniano, romeno, espanhol, sueco, turco, árabe e japonês.

década de 60), especificamente estatística léxica com vistas à análise estilística e literária, os trabalhos dos pesquisadores brasileiros Cléa Rameh e John Duncan (que elaborou um dicionário de frequência do português peninsular), nos anos 70. Destaca-se também o trabalho da Profa. Maria Tereza Camargo Biderman, também nos anos 70, quando da publicação de sua Teoria Lingüística (lingüística quantitativa e computacional), de 1978, que dá as bases para a compreensão dos estudos do PLN especialmente para o português, sempre em conjunto com teorias lingüísticas pertinentes à grande área interdisciplinar. Finalmente, destacamos o trabalho do NILC (Núcleo Interinstitucional de Lingüística Computacional), que nos anos 90 partiu para a pesquisas mais consistentes em PLN, sempre levando em conta a interdisciplinariedade, possuindo profissionais de diversas áreas do conhecimento. Nessas iniciativas encontram-se a maioria das referências a TA em língua portuguesa.

Especificamente com relação à TA para a língua portuguesa, há uma variedade de protótipos de sistemas comercializados<sup>3</sup>:

- ATS (Automatic Translation Server) <http://www.automatictrans.es/productos.htm>
- FALATUDO <http://www.falatudo.com.br/>
- Office Translator  
[http://www.digitalriver.com/dr/v2/ec\\_MAIN.Entry10?V1=641725&PN=2&SP=10023&xid=28102](http://www.digitalriver.com/dr/v2/ec_MAIN.Entry10?V1=641725&PN=2&SP=10023&xid=28102)
- Portdan –tradução para o dinamarquês  
<http://visl.hum.sdu.dk/itwebsite/port/portgram.html>
- SYSTRAN Personal  
[http://www.digitalriver.com/dr/v2/ec\\_MAIN.Entry10?V1=641878&PN=2&SP=10023&xid=28102](http://www.digitalriver.com/dr/v2/ec_MAIN.Entry10?V1=641878&PN=2&SP=10023&xid=28102)
- SYSTRAN Professional Premium  
[http://www.digitalriver.com/dr/v2/ec\\_MAIN.Entry10?V1=641777&PN=2&SP=10023&xid=28102](http://www.digitalriver.com/dr/v2/ec_MAIN.Entry10?V1=641777&PN=2&SP=10023&xid=28102)
- SYSTRAN Professional Standard  
[http://www.digitalriver.com/dr/v2/ec\\_MAIN.Entry10?V1=641803&PN=2&SP=10023&xid=28102](http://www.digitalriver.com/dr/v2/ec_MAIN.Entry10?V1=641803&PN=2&SP=10023&xid=28102)
- TraduzTudo Pro 1.2, TraduzTudo Light 1.2 <http://www.traduz.com.br/>

---

<sup>3</sup> Fonte: Disponível em [www.linguateca.pt](http://www.linguateca.pt). Consulta em: 01/07/2005.

- WebTranslator  
[http://www.digitalriver.com/dr/v2/ec\\_MAIN.Entry10?V1=641957&PN=2&SP=10023&xid=28102](http://www.digitalriver.com/dr/v2/ec_MAIN.Entry10?V1=641957&PN=2&SP=10023&xid=28102)
- WinDi (Windows Dictionaries) (de,en,es,fr,it,nl,pt)  
<http://www.windi7.com/index.html>
- Word Translator <http://www.tranexp.com/>

## 1.4 Tipos de sistemas

Os sistemas de TA distribuem-se em três arquiteturas:

- (i) **Sistemas Diretos:** trata-se de sistemas que substituem as palavras da língua de partida por palavras correspondentes da língua de chegada para gerar uma tradução. Ex: o já citado SYSTRAN (EUA, final dos anos 1950);
- (ii) **Sistemas de Transferência:** sistemas que transformam as representações sintáticas da língua de partida em representações sintáticas da língua de chegada. Essa abordagem utiliza representações que contêm primitivos (como SUJ ou OBJ), considerados primitivos funcionais em línguas diferentes. Destacam-se os seguintes sistemas: SUZY (Alemanha, década de 60), MÉTÉO (Canadá, 1960), ARIANE (GETA) (França, entre 1960 e 1970), EUROTRA (Comunidade Européia, 1978-1990), METAL (Alemanha, 1960-1989) e ROSETTA (Holanda, 1980-1988);
- (iii) **Sistemas de Interlíngua:** trata-se de sistemas que codificam a língua de partida em uma representação abstrata considerada universal, isto é, uma interlíngua, ou representação interna à máquina e independente de qualquer língua natural, e desta para a língua de chegada (AYAN, 2004). Destacam-se o DLT (Holanda, 1979-1988), e o UNITRAN (EUA, 1993).

## 1.5 Paradigmas de TA

Os três tipos básicos de arquitetura de sistemas de TA, por sua vez, permitem a classificação de diferentes paradigmas de pesquisa. O UNITRAN, o objeto de estudo

desta tese, articula-se dentro do paradigma de sistemas de TA baseados em princípios, um dos dez paradigmas sistematizados por Dorr (1993) e listados a seguir:

– TA baseada em restrições - *Constraint-based MT* –*CBMT* – Sistema de TA que permite que sejam captadas restrições lingüísticas entre as línguas envolvidas no processo de tradução, podendo essas restrições serem registradas, por exemplo, no léxico. Um exemplo desse tipo de sistema é o LFG-MT (KAPLAN & BRESNAN, 1982);

– TA baseada em diálogo - *Dialogue-based MT* –*DBMT* – sistema de TA que permite a intervenção do tradutor humano no processo de tradução, seja para inserir elementos estilísticos, seja para obter traduções de melhor qualidade. O ENtran (JOHNSON & WHITELOCK, 1987) é um exemplo desse tipo.

– TA baseada em casos/exemplos - *Example- (Case-/ Memory-) based MT* - *EBMT* – Sistema de TA que busca estabelecer correspondência entre padrões de sentenças previamente armazenadas para a obtenção de tradução a partir de corpora bilíngües alinhados. O Pangloss (BROWN, 1996) ilustra esse tipo.

– TA baseada em conhecimento - *Knowledge-based MT* - *KBMT* – Sistema de TA que se utiliza de conhecimentos profundos tanto lingüísticos quanto extralingüísticos, além de rotinas de inferência lógica para relacionar unidades no nível conceitual. São exemplos desse tipo os sistemas Translator e UNL.

– TA baseada em léxico - *Lexical-based MT* - *LBMT* – Sistema de TA que relaciona unidades lexicais de uma língua de partida a unidades lexicais na língua de chegada. O LTAG (que se utiliza de princípios da teoria de *Tree Adjoining Grammars*) é paradigmático.

– TA baseada em redes neurais - *Neural Net- (connectionist-) based MT* - *NBMT* – Sistema de TA que busca a utilização de uma abordagem conexionista<sup>4</sup> utilizada principalmente em analisadores gramaticais e processos de desambiguação.

---

<sup>4</sup> O conexionismo é uma das linhas de pesquisa da Inteligência Artificial que tem por objetivo investigar a possibilidade de simulação de comportamentos inteligentes através de modelos baseados na estrutura e funcionamento do cérebro humano. Fonte: [www.das.ufsc.br/gia/softcomp/node6.html](http://www.das.ufsc.br/gia/softcomp/node6.html). Acesso em 31-01-2006

– TA baseada em regras - *Rule-based MT - RBMT* – Sistema de TA baseado em regras gramaticais elaboradas para cada uma das línguas envolvidas no processo de tradução. O sistema Rosetta (1994) exemplifica esse tipo.

– TA *shake and bake - Shake and Bake MT - S&BMT* - Sistema de TA “pronto para uso” que se utilizam de regras de transferência e da correspondência entre unidades lexicais das línguas por meio de um léxico bilíngüe. O sistema busca unidades que satisfaçam a configurações sintáticas predeterminadas (*shake*) e então avalia e encaixa a que melhor satisfizer as condições da estrutura sintática (*bake*). Como exemplo, podemos citar a metodologia desse tipo de sistema proposta por Whitelock, 1992.

– TA baseada em estatística - *Statistics- (Corpus-) based MT- SBMT* – Sistema de TA que se utiliza de princípios da estatística ou probabilística na resolução do problema de TA, como a ambigüidade. O Candide (BROWN, 1990) exemplifica esse tipo.

– TA baseada em princípios - *Principle-based MT - PBMT* – Sistema de TA que tem uma estrutura que abriga princípios de teorias lingüísticas em seus níveis de processamento e de armazenamento de informações de natureza lexical, morfológica, sintática e semântica. O UNITRAN (DORR, 1992) é o exemplo paradigmático.

Na subseção seguinte, serão então descritos os princípios que lingüisticamente subsidiam o sistema UNITRAN, um sistema de interlíngua, baseado em princípios e lingüisticamente motivado.

## 1.6 Os desafios da TA

Existem três categorias principais de processos sobre os quais recaem considerações lingüísticas no processo de TA: compreensão de língua (análise), geração de língua (síntese) e a correspondência entre pares de línguas (alinhamento).

No **processo de compreensão de língua**, alimentava-se a hipótese de que, para se obter uma tradução adequada, era preciso haver um completo entendimento do texto de partida. Mais recentemente, entretanto, pesquisadores oriundos da área de tradução estatística (DORR et al, 2000) tendem a considerar suficiente uma tradução parcial, obtida através de um conjunto mínimo de conhecimento lingüístico. Ayan (2004) aponta as principais dificuldades que precisam ser equacionadas na construção de um sistema de TA:



- **A ambigüidade sintática, ou estrutural** – Ex. *Eu vi o homem no barco com o telescópio.*

Aqui, é preciso, por exemplo, determinar com quem está o telescópio: com o homem de quem o enunciador fala? Com o enunciador?

- **A ambigüidade lexical** – Ex. ing. *book* / port. *livro, reservar.*

Aqui, é preciso escolher entre as duas formas do português.

- **A ambigüidade semântica** – Ex. ing. *Every girl reads a novel/* port. *toda menina lê um romance.* Nesses casos, o sistema tem de escolher a interpretação adequada dos quantificadores, calculando seus escopos relativos (trata-se aqui da ambigüidade semântica de escopo de quantificadores): a sentença-exemplo é ambígua entre estas duas leituras: “existe um único romance que toda menina lê” (o escopo do quantificador existencial, aqui lexicalizado pela forma *um*, inclui o escopo do quantificador universal, aqui lexicalizado pela forma *every/toda*) e “cada menina lê pelo menos um romance” (aqui ocorre a situação inversa, o escopo do quantificador universal inclui o escopo do quantificador existencial).

- **A ambigüidade contextual** – Ex.

(i) Ing. *The computer outputs the data; it is fast.*

Port. *O computador imprime o dado; ele é rápido.*

(ii) Ing. *The computer outputs the data; it is stored in ascii.*

Port. *O computador imprime o dado. Ele é armazenado em ascii.*

Em negrito, os anafóricos e seus antecedentes correspondentes.

Observe-se que, no contexto de um manual de computador, determinar o antecedente apropriado de *it* é função de se calcularem, pelo menos, (a) o papel semântico que cada um dos participantes desempenha nas predicções *output/imprimir*, *store/armazenar* e *be fast/(ser) rápido* (por exemplo, no evento *imprimir* há um agente, “o que imprime”, i.e. o computador, e um paciente/efetuado, “o que é impresso”, i.e. o dado; (b) as funções sintáticas que realizam esses papéis (por exemplo, o sujeito de *armazenar* é um agente (o que armazena) e o seu objeto um paciente, “o que é armazenado”); e (c) restrições seletivas que os predicadores impõem sobre os seus participantes (por exemplo, o sujeito do predicador *(ser) rápido* exige o atributo

“velocidade”). Mas esses cálculos, eminentemente de natureza sintática e semântica, não são suficientes, dado que, o estabelecimento do vínculo anafórico entre *it/ele* e o seu antecedente (o computador, em i; os dados em ii) depende do contexto formado pela sentença anterior, sem a qual não há como calcular a entidade que é rápida (em i) ou que é armazenada (em ii); ou seja, a interpretação do *it/ele* extrapola os limites de uma sentença, exigindo trechos (um contexto maior que uma sentença) mais longos do texto para a sua interpretação.

- **Ambigüidade contextual complexa** – Ex. ing. *John hit the dog with a stick.* / port. *João bateu no cachorro com uma vara.*

Nota-se que neste caso, o contexto poderá resolver a multiplicidade de sentidos, ao permitir a recuperação de duas situações: ou João carregava uma vara para se proteger ou havia muitos cães, e dentre eles, um que carregava uma vara. Verificamos, assim, como as intenções do autor afetam a escolha das unidades lexicais e como elas são realizadas.

**No processo de geração de língua**, supõe-se que a geração de uma sentença na língua de chegada não requer uma capacidade plena de geração, ou seja, não é necessário levar em conta todo o conteúdo e a organização do texto; isso porque o texto de partida fornece grande parte dos elementos que aparecerão na superfície do texto na língua de chegada.

Dois problemas merecem destaque: o da seleção lexical apropriada na língua de chegada e o da geração do tempo verbal correspondente ao da língua de partida.

- **Seleção lexical** – Ex. ing. *wait, hope* / port. *esperar, ter esperança*; ing. *be* / port. *ser, estar*.

Neste exemplo, informações adicionais, como traços semânticos, por exemplo,  $\pm$  desejo, e  $\pm$  permanência, ajudariam a resolver o problema da escolha lexical apropriada no português.

- **Geração de tempo verbal** – Ex. ing. *Mary has lived here for many years.* port. *Maria mora aqui há muitos anos.*

Do presente perfeito do inglês para o presente do indicativo do português, a máquina precisaria da informação pragmática de que Mary morou aqui no passado e continua morando ou valer-se de uma regra de correspondência sintática que

condicionasse o emprego do presente do indicativo do português no contexto [have/has + past participle +...+ for TIME EXPRESSION]:

[...have/has + v.past participle +...+ for TIME EXPRESSION.] -> [...v.presente do indicativo +...+ há TIME EXPRESSION.]

No processo de estabelecimento de correspondências entre representações da língua de partida e da língua de chegada, a correspondência entre unidades de ambas as línguas é outra tarefa difícil de se realizar, posto que há significativas diferenças sintáticas e léxico-semânticas entre pares de sentenças de línguas diferentes. Essas diferenças, ou divergências, acabam por minar as tentativas de estabelecimento de correspondências diretas entre pares de sentenças de línguas diferentes. Entre essas divergências, podemos citar a divergência temática, a divergência de troca de núcleo, a divergência estrutural, a divergência categorial e a divergência de incorporação lexical, que serão abordadas oportunamente na seção 4 desta tese.

## **1.7 A abordagem do PLN em três domínios**

Esta tese, em particular, tem dois grandes objetivos: (i) discutir a arquitetura e funcionamento do sistema de tradução automática UNITRAN, desenvolvido por Dorr (1992), (ii) planejar a inclusão do português nesse sistema, que foi originalmente desenhado para traduzir entre o inglês, o espanhol e o alemão. Essa tarefa exploratória de inclusão do português no UNITRAN envolverá a análise do último tipo de desafio listado na subseção anterior: as “divergências de tradução”. São essas divergências a motivação primeira para a proposição do UNITRAN. Nesta tese focalizam-se, primariamente, as divergências de tradução entre o inglês e o português. Diante desses objetivos, recorreremos à abordagem de pesquisa interdisciplinar em PLN proposta por Dias-da-Silva (1996), que, dentre outras metas, visa minimizar a distância que separa os estudos lingüísticos e os estudos computacionais.

Dias-da-Silva (1996, 1998a e 1998b) delineou um modelo de pesquisa em PLN que distribui a tarefa de pesquisa em três grandes domínios: o Domínio Lingüístico, o Domínio Representacional e o Domínio Implementacional.

No Domínio Lingüístico, realizam-se as reflexões que dizem respeito a objetos lingüísticos; é neste domínio que se sistematizam os objetos lingüísticos necessários

para o desenvolvimento do sistema de PLN que se quer implementar. Nesta tese, focalizam-se os objetos que dão suporte às dimensões sintática e semântico-conceitual do sistema do UNITRAN são duas teorias: *Teoria dos Princípios e Parâmetros* (CHOMSKY, 1981) e a *Teoria das Estruturas Conceituais* (JACKENDOFF, 1990), respectivamente.

No Domínio Representacional, buscam-se os meios de representação formal para que os objetos investigados no domínio anterior possam ser, no Domínio Implementacional, codificados em um programa de computador. Nesta tese, focalizam-se as representações das divergências de tradução implementadas no UNITRAN (DORR, 1993).

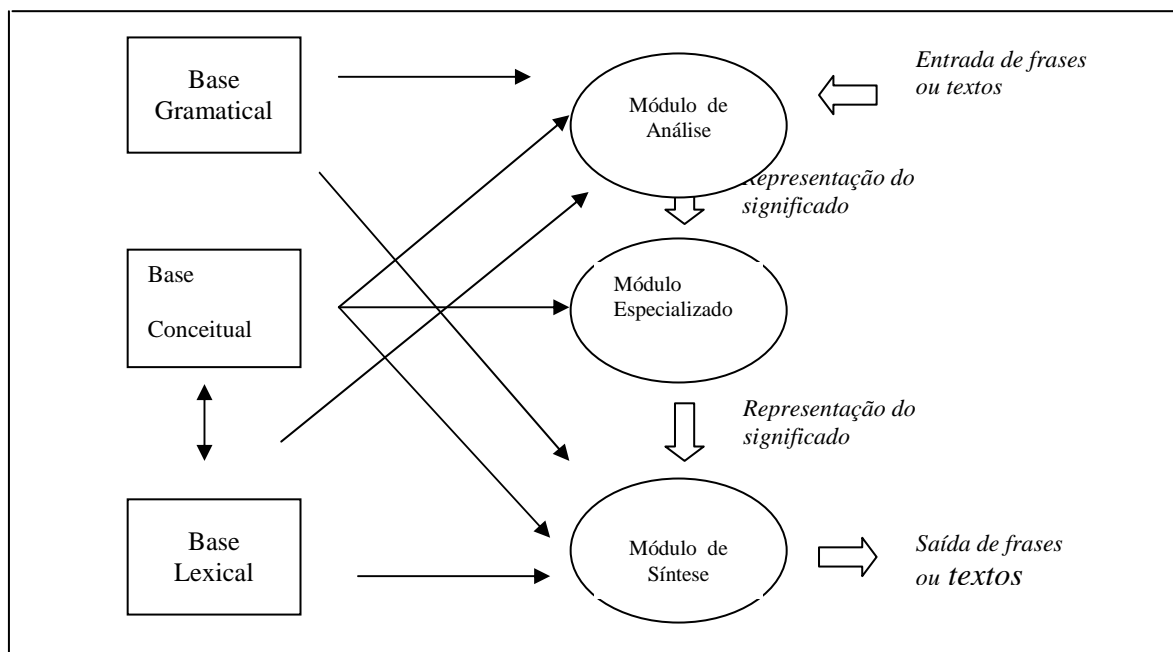
Por fim, no Domínio Implementacional, além da implementação do programa de computador que dá existência ao sistema de PLN, desenham-se também a estrutura do sistema, o fluxo de dados que por ele deve passar e os módulos de armazenamento desses dados e do seu processamento.<sup>5</sup>

## 1.8 A base lexical e o sistema de PLN

A figura 1 ilustra a arquitetura genérica de um sistema de PLN (DIAS-DASILVA, 1996, p. 232).

---

<sup>5</sup> Cabe aqui uma observação. Como não se pretende implementar um sistema de tradução automática, o que extrapolaria nossa competência de lingüista, a tese não discutirá as questões do Domínio Implementacional.



**Figura 1.** As bases de dados e os módulos de processamento de um sistema de PLN genérico.

Com exceção do **Módulo Especializado (ME)**, que deve ser projetado em função dos objetivos previstos para o sistema, os demais módulos possuem um funcionamento padrão (DIAS-DA-SILVA, 1996, p.235). Cumpre observar que no universo de discussão desta tese trata-se do módulo que realiza a tradução automática e que por isso dá nome ao sistema, isto é, o sistema de PLN alvo de nossas investigações é um sistema de TA. Desta forma, aproveitamos deste esquema as três bases: Gramatical, Lexical e Conceitual.

A **Base Lexical** é concebida como uma base de dados de natureza léxico-gramatical que contém um conjunto de unidades lexicais que, de acordo com as especificidades do sistema em desenvolvimento, poderá conter informações fonéticas, morfológicas, sintáticas, semânticas e pragmático-discursivas.

A **Base Conceitual** contém um modelo de representação de mundo, tanto físico quanto conceitual que visa descrever objetos, eventos, forças, propriedades, relações e atributos em termos de representações hierarquicamente estruturadas. A Base Conceitual contém, portanto, o conjunto de categorias semânticas básicas, isto é, primitivos semânticos que simulam a “visão de mundo” do sistema, restringindo o universo de discurso a um determinado domínio.

A Base Gramatical, por fim, contém toda a informação morfológica e sintática das línguas que alimentarão o sistema.

Nesse contexto, no sistema UNITRAN, será explorado como interação as Bases Gramatical, Lexical e Conceitual .

## SEÇÃO 2 – INVESTIGAÇÃO DO DOMÍNIO LINGÜÍSTICO: AS BASES TEÓRICAS DO UNITRAN

A fundamentação teórica do UNITRAN (DORR, 1993) é constituída pela *Teoria dos Princípios e Parâmetros* (CHOMSKY, 1981) e pela *Teoria das Estruturas Conceituais* (JACKENDOFF, 1990). A primeira subsidia a construção do seu módulo sintático e a segunda do seu módulo léxico-semântico. Esse sistema, que traduz do inglês para o alemão e o espanhol bidirecionalmente, é parametrizado nos dois níveis, a saber:

- no nível sintático: que diz respeito à estrutura da sentença;
- no nível léxico-semântico: que diz respeito às propriedades semântico-conceituais dos itens lexicais.

Como a discussão de sua arquitetura e funcionamento pressupõe a compreensão de suas bases lingüísticas, nesta seção destacam-se, de modo sucinto e simplificado, os principais constructos dessas duas teorias.

### 2.1 A Teoria de Princípios e Parâmetros

Proposta por Chomsky em 1981, a teoria de *Princípios e Parâmetros* é um desdobramento da primeira fase da teoria gerativa, iniciada com a publicação da obra *Syntactic Structures*, de 1957.

Chomsky postula a existência de princípios (características observáveis em todas as línguas) e parâmetros (que são parte da gramática universal) como componentes da faculdade da linguagem. Esses princípios, integrantes da Gramática Universal (GU) e, portanto, determinados geneticamente, segundo ele, formam módulos ou subsistemas da faculdade da linguagem. Cada um desses módulos é, então, descrito e articulado segundo cada uma das “subteorias” resumidas a seguir.

### 2.1.1 A Teoria X-barra

A Teoria X-barra permite representar as relações hierárquicas e de precedência entre os constituintes (i.e. sintagmas) de uma sentença. Nesta teoria, um ponto crucial é o fato de que um constituinte deve necessariamente construir-se a partir do seu núcleo. Assim, um SN, sintagma nominal, deve necessariamente ter como núcleo um N, uma unidade nominal. Um SV, sintagma verbal, por sua vez, só pode ter como núcleo um V, uma unidade verbal, e assim por diante, para as demais categorias sintáticas das preposições, dos adjetivos e dos advérbios (COOK, 1989, p. 103).

Dentre os princípios que compõem a faculdade da linguagem, dois estão relacionados à Teoria X-barra:

- **Princípio da Endocentricidade:** Todo sintagma é projeção do seu núcleo.
- **Princípio da Homogeneidade das Estruturas:** Todas as estruturas devem seguir o mesmo modelo estrutural.

Portanto, os princípios da teoria X-barra postulam que : (i)  $SX \rightarrow \dots X \dots$  (um sintagma SX sempre contém um núcleo do tipo X – trata-se da regra geral de projeção dos núcleos sintagmáticos); (ii)  $X'' \rightarrow$  Especificador X' (uma categoria de duas barras é formada pela projeção de uma categoria nuclear ao se combinar com um possível Especificador – trata-se da regra de projeção dos especificadores, incluindo a projeção do argumento externo do predicador, isto é, a projeção do sujeito sintático); (iii)  $X' \rightarrow$  X' adjunto (uma categoria de uma barra é formada pela projeção de uma categoria nuclear ao se combinar com possíveis Complementos – trata-se da regra de projeção sintática dos argumentos internos de um predicador, isto é, a projeção do objeto sintático); (iv)  $X' \rightarrow$  X e complemento (uma categoria de uma barra pode também ser formada pela projeção de uma categoria de nível um, isto é, uma categoria de uma barra, ao se combinar com possíveis Adjuntos -- trata-se da regra de projeção dos modificadores, isto é, a projeção dos adjuntos sintáticos).

Além desses dois princípios, outro que se descreve em termos deste módulo é o parâmetro da ordenação entre núcleo / especificador / complemento:

O núcleo            segue        o especificador  
                                  precede



O núcleo segue o complemento  
precede

Assim, do ponto de vista estrutural, algumas relações são consideradas básicas ou fundamentais e, por isso, têm consequência na construção da teoria. Observemos, então, as relações configuracionais que permitem parametrizar esse princípio: dominância, dominância imediata, paternidade, irmandade, precedência (MIOTO *et al* 1999) e c-comando (LOBATO, 1986).

## **Relações configuracionais**

### ***Dominância***

$\alpha$  domina  $\beta$  se, e somente se, existe uma seqüência conexa de um ou mais ramos entre  $\alpha$  e  $\beta$  e o percurso de  $\alpha$  até  $\beta$  através dos ramos é estritamente descendente.

### ***Dominância Imediata***

$\alpha$  domina imediatamente  $\beta$  se, e somente se,  $\alpha$  domina  $\beta$  e não existe nenhum  $\gamma$  que domina  $\beta$ .

### ***Paternidade***

$\alpha$  é pai de  $\beta$  se, e somente se,  $\alpha$  domina  $\beta$  imediatamente.

### ***Irmandade***

$\alpha$  é irmão de  $\beta$  se, e somente se,  $\alpha$  e  $\beta$  têm o mesmo pai  $\gamma$ .

### ***Precedência***

$\alpha$  precede  $\beta$  se, e somente se,  $\alpha$  está à esquerda de  $\beta$  e  $\alpha$  não domina  $\beta$  ou  $\beta$  não domina  $\alpha$ .

### ***C-comando***

$\alpha$  c-comanda  $\beta$  se, e somente se:

- (i) nem  $\alpha$  nem  $\beta$  dominam um ao outro;
- (ii) a primeira categoria ramificada que domina  $\alpha$  também domina  $\beta$ .

As escolhas para o núcleo X de uma projeção máxima são determinadas pelo parâmetro Categorias Básicas. Para as três línguas, inglês, espanhol e português, esse parâmetro tem os mesmos valores: {C, FLEX, V, N, A, P}.

Quanto à questão da ordem dos constituintes, as línguas acima são línguas em que o núcleo antecede o complemento e o especificador antecede o núcleo, daí o valor do parâmetro ser, para essas línguas, núcleo-inicial/espec-inicial com relação aos verbos, ou seja, línguas do tipo estrutural Sujeito-Verbo-Objeto (S-V-O). Observe-se, entretanto, que o português admite também inversões: “Saíram todos”, fato que deverá ser parametrizado no sistema.

Quanto aos especificadores de base, a teoria X-barra dá conta daqueles que são gerados na base. Por exemplo, o Especificador de FLEX corresponde ao sujeito da sentença. Para o inglês e o português, a posição é obrigatória, mesmo que neste último, a posição possa ser preenchida por *pro* ou PRO). Os especificadores de N são, em geral, preenchidos por determinantes, por exemplo, os artigos definidos do português e inglês e o adjetivo possessivo do inglês (*my doll, your doll, her doll, etc.*) e o N'' no caso genitivo, como em *Mary's doll (a boneca de Maria)*, em que *Mary's* é o especificador do núcleo nominal *doll*.

Quanto aos adjuntos de base, em inglês, por exemplo, um advérbio ou um sintagma preposicional podem ocorrer ao lado direito do verbo, porém no nível máximo apenas (ou seja, não entre o verbo e o objeto). Já em português, o advérbio ou o sintagma preposicional podem ocorrer à direita do verbo, ambos no nível máximo e também entre o verbo e seu objeto.

Exemplo:

- João tem comprado os remédios [ADV freqüentemente]
- João tem comprado [ADV freqüentemente] os remédios
- João tem comprado os remédios [P'' à noite]
- João tem comprado [P'' à noite] os remédios

Outra construção com adjuntos que diverge entre o inglês e o português está ligada à adjunção de sintagmas adverbiais e preposicionais com relação ao sintagma verbal. Em inglês e português, um advérbio ou um sintagma preposicional pode ocorrer ao lado direito do verbo, porém apenas no nível da projeção máxima de V (ou seja, o advérbio não pode ocorrer entre o verbo e o seu objeto). Já em espanhol, o advérbio ou o sintagma preposicional pode ocorrer à direita do verbo, tanto no nível da projeção máxima de V quanto no nível intermediário da projeção, isto é, entre o verbo e o seu objeto, conforme ilustram os exemplos 1a, 1b, 1c e 1d.

1a-

John has eaten breakfast [ADV frequently]

Juan ha comido el desayuno [ADV frecuentemente]

João tem tomado o café da manhã [ADV freqüentemente]

1b-

\*John has eaten [ADV frequently] breakfast

Juan ha comido [ADV frecuentemente] el desayuno

João tem tomado [ADV freqüentemente] o café da manhã

1c-

John has eaten breakfast [P'' in the morning]

Juan há comido el desayuno [P'' en la mañana]

João tem tomado o desjejum [P'' de manhã]

1d-

\*John has eaten [P'' in the morning] breakfast

Juan ha comido [P'' en la mañana] el desayuno

João tem tomado [P'' de manhã] o café da manhã

### 2.1.2 A Teoria da Regência

A Teoria da Regência regula as relações entre o núcleo de uma construção e as categorias dele dependentes, ou seja, refere-se à relação sintática entre um elemento “regente” e um elemento “regido”. (COOK, 1989, p.35) É a formalização da noção de complementação. Ex. *Ele ama Luiza*. O verbo *amar* governa o SN *Luiza*. A lista dos regentes inclui as categorias lexicais substantivo, verbo, adjetivo e preposição (categorias que podem ser núcleo de sintagma). A regência também garante que uma

palavra adquira um Caso apropriado (cf. subseção 2.1.4), como, no português, *para mim* e não *\*para me*. Num outro aspecto da teoria, o sujeito das línguas geralmente rege verbos, mas quando o sujeito é regido, a teoria aponta para a flexão como regente. No que se refere ao número, a teoria denomina concordância e indica se o singular ou o plural são necessários.

Por definição,  $\alpha$  rege  $\beta$  se, e somente se (SELLS, 1985, p.40).

(i)  $\alpha$  c-comanda  $\beta$  e

(ii)  $\alpha$  é uma categoria lexical ou a categoria de Flexão e

(iii) Toda projeção máxima que domina  $\beta$  também domina  $\alpha$ .

Outro aspecto diz respeito ao parâmetro *pro-drop*, referente à categoria vazia *pro*.<sup>6</sup> As línguas *pro-drop* permitem que sentenças declarativas possuam sujeito com conteúdo fonético nulo, como no português (*pro*) *Sou feliz*. Desta forma, a teoria postula que toda categoria vazia deve ser devidamente regida.

Na teoria da regência, pode-se observar que, além de SNs com conteúdo fonético, também estão os SNs não realizados foneticamente. Esse subtipo de problema é chamado de problema das categorias vazias ou vestígios. Com essa teoria, é possível determinar os vestígios, as categorias pronominais vazias *pro* (sujeito nulo) e PRO (sujeito de infinitivo não flexionado), os complementos das categorias funcionais C e FLEX, as condições de licenciamento de *pro* e as condições de regência apropriadas exercidas por categorias vazias.<sup>7</sup>

Os três parâmetros que são especificados previamente no pré-processamento das sentenças são os parâmetros Vestígios (vestígios  $v$  de elementos deslocados), Pronomes Vazios (o sujeito de infinitivo não flexionado (PRO) e o sujeito nulo (*pro*)) e Elementos Funcionais Vazios (as estruturas de complementação de C e FLEX).<sup>8</sup>

O primeiro especifica as categorias que ocorrem na posição original do elemento deslocado: por exemplo, em *Quem<sub>i</sub> ele viu v<sub>i</sub>?* o elemento *Quem* desloca-se da posição

<sup>6</sup> A Teoria dos Princípios e Parâmetros reconhece 4 tipos de categorias vazias e as classifica em função de dois traços [+ / - anafórico (a)] e [+ / - pronominal (p)]: [-a, -p]=variável (elementos Qu-deslocados). Por exemplo: *Quem<sub>i</sub> Maria viu v<sub>i</sub>?*; [+a,-p]=vestígio de N'', Por exemplo: *O vaso<sub>i</sub> quebrou v<sub>i</sub>*; [-a,+p]=*pro* (sujeito nulo); [-a,-p]=PRO (sujeito de infinitivo). O UNITRAN não diferencia os dois primeiros tipos.

<sup>7</sup> X rege apropriadamente Y se X C-rege Y e X é uma categoria lexical (N, V, A ou P) ou X é um antecedente local de Y.

<sup>8</sup> Parâmetros estabelecidos por Dorr, 1993.

de objeto do verbo para a posição do complementador da sentença, deixando o vestígio  $v_i$  com ele indexado.

O parâmetro Pronomes Vazios especifica as categorias PRO e *pro*, como em *Eu<sub>i</sub> quero PRO<sub>i</sub> ir ao cinema* e *pro tenho fome*, respectivamente. O parâmetro Elementos Funcionais Vazios especifica a estrutura de complemento dos elementos potencialmente vazios C e FLEX. O valor desses parâmetros é o mesmo para o inglês, o espanhol e o português.

Outros dois parâmetros podem ser estabelecidos de acordo com a teoria de ligação. Trata-se dos parâmetros Licenciamento de *pro* e Condições sobre Cadeias.

O parâmetro Licenciamento de *pro* especifica as restrições de licenciamento de *pro* para cada língua do sistema. As sentenças (1), (2) e (3) abaixo ilustram e explicam o papel das restrições deste sub-módulo no licenciamento do sujeito nulo *pro* para o inglês, o espanhol e o português.

1. *I saw the book / \*pro Saw the book* (CONC não é regente apropriado no inglês. *pro* não é licenciado)

2. *Yo vi el libro/ pro Vi el libro* (CONC é regente apropriado no espanhol. *pro* é licenciado)

3. *Eu vi o livro/ pro Vi o livro.* (CONC é regente apropriado no português. *pro* é licenciado)

O parâmetro Condições sobre Cadeias especifica se categorias vazias são regentes apropriados. Os exemplos 1, 2 e 3 ilustram o papel desse parâmetro na filtragem de estruturas que apresentam regência por vestígios; 3 é agramatical porque  $v_j$  não está apropriadamente regido. A regência não é apropriada porque o vestígio  $v_i$  do verbo *saw* (que foi deslocado para a esquerda de *John*) não é regente apropriado do vestígio  $v_j$ , deixado pelo deslocamento de *what* para o início da sentença. No espanhol e no português, os vestígios  $v_i$  análogos são regentes apropriados e licenciam o vestígio  $v_j$ .

1. \*What<sub>j</sub> saw<sub>i</sub> John<sub>vi</sub> v<sub>j</sub>

2. Qué<sub>j</sub> vio<sub>i</sub> Juan<sub>vi</sub> v<sub>j</sub>

3. O que<sub>j</sub> viu<sub>i</sub> João<sub>vi</sub> v<sub>j</sub>

### 2.1.3 A Teoria da Fronteira

A Teoria da Fronteira descreve e regula os deslocamentos dos constituintes de uma sentença impondo-lhes restrições. Elaborada a partir do princípio da subjacência, que tem o efeito de restringir a distância percorrida por um constituinte, postula que o movimento de um constituinte não pode atravessar mais do que um nó fronteira (RAPOSO, 1992, p.398). Para o português e o espanhol, C'' é nó fronteira, mas não FLEX''. Para o inglês, FLEX'' é nó fronteira. Assim, os nós de fronteira dão conta de divergências sintáticas entre português e espanhol, de um lado, e do espanhol e do inglês, do outro:

- (1) *\*Who<sub>i</sub> did you wonder whether<sub>i</sub> went to school?*
- (2) *Quién<sub>i</sub> contemplabas tú que<sub>vi</sub> fue a la escuela?*
- (3) *Quem<sub>i</sub> você pensa que<sub>vi</sub> foi à escola?*

### 2.1.4 A Teoria dos Casos

A Teoria dos Casos responde pela atribuição de traços de Caso sintático a SNs em posições marcadas por Caso e sobre sua realização morfológica. A atribuição de caso se dá, sob regência (com exceção de atribuição do caso genitivo, como se exemplificará logo a seguir), a todos os sintagmas nominais por meio de atribuidores de caso: o nominativo é atribuído ao sujeito da sentença pela categoria flexão com a especificação de tempo; o acusativo é atribuído pelo verbo ao seu objeto, podendo também ser atribuído, em inglês, mas não em português, pela preposição ao seu objeto; o caso oblíquo é atribuído por preposições; no inglês, o genitivo é atribuído na configuração<sub>[SN \_X]</sub>. Quanto ao Filtro de Caso, a teoria estipula que todo SN realizado foneticamente deve estar marcado por Caso. A teoria também inclui três parâmetros: (i) o parâmetro da adjacência: algumas línguas, como o inglês e o português, por exemplo, exigem que os atribuidores de caso sejam adjacentes ao SN que recebe o Caso; (ii) o

parâmetro da direção: algumas línguas, como o inglês e português, exigem que os marcadores de caso estejam à esquerda do SN que recebe o Caso; já o chinês exige que eles estejam à direita.

As principais restrições que esse sub-módulo impõe sobre as estruturas sintáticas são dadas no Quadro 1.

a) Caso Objetivo	o caso objetivo é atribuído ao objeto regido pelo verbo transitivo V
b) Caso Oblíquo	o caso oblíquo é atribuído ao objeto regido pela preposição transitiva P;
c) Caso Possessivo	o caso possessivo é atribuído ao objeto regido pelo N' transitivo;
d) Caso Nominativo	o caso nominativo é atribuído ao sujeito regido por CONC;
e) Filtro de Caso	Filtro de Caso: qualquer sentença com um N' não marcado com caso abstrato é mal formada (exceto PRO).

**Quadro 1. Restrições da Teoria dos Casos.**

Por exemplo, a determinação de que todos os elementos C-regidos que possuem uma única posição na subcategorização de um verbo serem S-regidos, dá conta da agramaticalidade de 1a, 2b e 2c.

1a. \*Lo vimos Juan.

1b. We saw Juan.

1c. Nós vimos Juan.

2a. Lo vimos a Juan.

2b. \*We saw (to) Juan.

2c. \*Nós vimos (a) Juan.

No espanhol, o clítico acusativo *lo* absorve a regência-S, isto é, é um constituinte subcategorizado pelo verbo, e *Juan* continua não regido, posto que sua posição de subcategorização foi preenchida com o clítico *lo*; portanto, *Juan* não recebe caso e a

sentença viola o Filtro de Caso. Essa sentença passa a ser gramatical com a inserção da preposição *a*, que atribui o caso objetivo a *Juan*.

### 2.1.5 A Teoria de Ligação

A Teoria da Ligação regula as condições de ligação dos elementos anafóricos, dos elementos pronominais e das expressões referenciais (COOK, 1989, p.49): (a) um elemento anafórico tem de estar necessariamente ligado num domínio local (Ex: em *Paula disse que Maria<sub>i</sub> se<sub>i</sub> machucou*, o anafórico *se* tem de estar necessariamente ligado a um antecedente no domínio local: no exemplo, o domínio local é a oração encaixada *Maria<sub>i</sub> se<sub>i</sub> machucou*; nesse domínio, *se* está necessariamente ligado ao antecedente *Maria*, o único disponível, não podendo se ligar ao elemento *Paula*, que é um antecedente potencial, mas está fora do domínio local do anafórico; a co-indexação entre anafórico e o antecedente sinaliza que as duas expressões denotam a mesma entidade, “Maria”); (b) um pronome deve estar necessariamente livre num domínio local (Ex: em *Paula<sub>j/k</sub> disse que Maria<sub>i</sub> machucou-a<sub>j</sub>*, o pronome *a* deve estar necessariamente livre no domínio local, que é a oração encaixada *Maria<sub>i</sub> machucou-a<sub>j</sub>*, podendo, entretanto, estar ligando ao antecedente *Paula*, fora do domínio local); (c) uma expressão referencial é livre em qualquer domínio. (Ex: em *Paula<sub>m</sub> disse que Maria<sub>n</sub> machucou a menina<sub>o</sub>*, a expressão referencial *a menina* deve estar necessariamente livre no domínio da sentença toda.

O parâmetro dessa teoria, isto é, o parâmetro Domínio de Regência, especifica o menor N’’ ou o menor FLEX’’ que contém um sujeito e o termo regente. Essa restrição prescreve que os elementos anafóricos (pronomes reflexivos e recíprocos) e os vestígios de N’’ devem estar ligados a sua categoria de regência;

- os expletivos<sup>9</sup> têm de estar livres no seu domínio de regência
- os vestígios de *qu-* devem ser ligados por um elemento externo ao seu domínio de regência.
- as expressões referenciais são livres em qualquer domínio.

---

<sup>9</sup> Os elementos expletivos são elementos com conteúdo semântico vazio. Ex. “it” em *It is raining (Está chovendo)*.



### 2.1.6 A Teoria- $\theta$

A Teoria- $\theta$  descreve a projeção da estrutura argumental das categorias lexicais; ocupa-se, assim, da atribuição de papéis temáticos ( $\theta$ ) como Agente, Paciente, etc. O Critério  $\theta$  estipula que cada SN argumento realiza um único papel temático, e cada papel temático, por sua vez, é atribuído a um único SN argumento. O papel Agente é aquele atribuído ao argumento que realiza a ação denotada pelo predicador; o papel Paciente é aquele atribuído ao argumento que é afetado pela ação denotada pelo predicador; o papel Meta é aquele atribuído ao argumento que é o destinatário da ação denotada pelo predicador (COOK, 1989, p.117).

A teoria diz respeito a três parâmetros: os Clíticos, Duplicação de Clíticos e *Non-drop*.

O parâmetro Clíticos é utilizado para especificar a categoria dos clíticos (elementos foneticamente fracos). Esse sub-módulo dá conta dos clíticos *him*, *le* e *lhe*, que ocorrem, por exemplo, em sentenças como *I gave him the book*, *Yo le di el libro*, *Eu lhe dei o livro*, respectivamente, do inglês, do espanhol e do português.

O parâmetro Duplicação de Clíticos dá conta da atribuição de papel temático a sentenças do espanhol do tipo *Lo vimos a Guille*. Nessa sentença, esse parâmetro assegura que o clítico *lo*, co-referente ao N'' *Guille* (que não tem papel temático), *lhe* transfira o seu papel temático.

Por fim, o parâmetro *Non-drop* efetua o a atribuição de elementos pleonásticos. As sentenças 1, 2 e 3 abaixo ilustram o processo, com a aplicação de duas condições: a primeira condição estipula que o parâmetro seja marcado *negativo* e o elemento seja um sujeito explícito, isto é, o pleonástico *it* do inglês; a segunda condição estipula que o parâmetro seja marcado *positivo* e a língua seja *pro-drop*, como o espanhol e o português.

1. It is raining.
2. Está lloviendo
3. Está chovendo

### 2.1.7 A Teoria do Controle

A Teoria do Controle do visa a descrever a categoria híbrida pronominal-anafórica PRO, que representa o sujeito foneticamente vazio do infinitivo em sentenças como *Maria quer PRO trabalhar meio-período*. Em princípio, PRO não é regido. Entretanto, torna-se obrigatório, por exemplo, quando tem o sujeito da oração principal como controlador (antecedente) (Ex. *João pediu PRO para sair*) ou o objeto como controlador (Ex. *João pediu para Pedro PRO sair*). O controle pode ser arbitrário quando PRO se refere a pessoas em geral, como em *Está na hora PRO de ir*. A posição vazia de sujeito de sintagmas infinitivos ou toma como antecedente um sintagma nominal (que pode ser sujeito ou objeto na oração principal), ou possui referência arbitrária, indefinida. (cf. COOK, 1989, p.162).

Na proposta do UNITRAN, a teoria de controle tem seu lugar na relação entre a interlíngua a estrutura sintática e, portanto, será retomada oportunamente na seção 3.

## 2.2 A Teoria das Estruturas Conceituais

A *Teoria das Estruturas Conceituais* (TEC) de Jackendoff (1990 e ss.), que subsidia a construção da interlíngua do sistema UNITRAN de Dorr (1992, 1993), como já se disse, fornece a metalinguagem para a representação semântica do significado sentencial independente de língua e se mostra apropriada para tratar do problema das divergências. Nesta subseção, vamos, portanto, examinar os princípios dessa teoria aplicados ao UNITRAN.

### 2.2.1 Princípios

A TEC investiga o nível conceitual da linguagem, buscando um modo de especificar a estrutura conceitual nos níveis lexical e sentencial e as relações que se estabelecem entre os planos lingüístico e conceitual. A TEC focaliza a semântica

lexical de maneira sintaticamente relevante, ao se preocupar com a construção da estrutura conceitual da sentença em função das estruturas conceituais parciais dos seus componentes. Por exemplo: a estrutura conceitual nuclear (abstraindo-se do tempo) da sentença de *João colocou o carro na garagem*, é dada na Figura 2 e explicada mais adiante.

[Evento CAUSAR  
 ([Coisa JOÃO],[Evento IR ([Coisa CARRO],[Trajetória PARA ([Lugar EM([Coisa GARAGEM])])])])])]

**Figura 2. Representação da estrutura conceitual da sentença *João colocou o carro na garagem***

A **decomposição semântica** é possível se concebermos que os significados das unidades lexicais não são atômicos. Assim, a decomposição do significado em elementos menores permite a especificação de um sistema composicional que impõe restrições à combinação desses elementos e, portanto, impõe também restrições nas configurações sentenciais.

A proposição de uma **ontologia** permite a construção de blocos de representação conceitual. Os constituintes da representação da estrutura conceitual do Quadro 3, por exemplo, correspondem, grosso modo, a categorias ontológicas do Quadro 2:

Categorias Ontológicas	Sentenças-teste para a classificação das entidades segundo as categorias
Coisa ( <i>Thing</i> )	O que você comprou? <i>(What did you buy?)</i>
Evento ( <i>Event</i> )	O que aconteceu depois? <i>(What happened next?)</i>
Estado ( <i>State</i> )	O que aconteceu foi que Max estava na África. <i>(What happened was that Max was in Africa.)</i>
Ação ( <i>Action</i> )	O que você fez? <i>(What did you do?)</i>

Lugar ( <i>Place</i> )	Onde está meu casaco?  ( <i>Where is my coat?</i> )
Trajatória ( <i>Path</i> )	Para onde eles foram?  ( <i>Where did they go?</i> )
Propriedade ( <i>Property</i> )	Como ela era?  ( <i>What was she like?</i> )
Modo ( <i>Manner</i> )	De que modo você cozinhou os ovos?  ( <i>How did you cook the eggs?</i> )
Quantidade ( <i>Amount</i> )	De que comprimento era o peixe?  ( <i>How long was the fish?</i> )

**Quadro 2. Principais categorias ontológicas da TEC.**

As estruturas conceituais sancionadas pela TEC são construídas por meio de categorias da ontologia. Cada constituinte é decomposto numa estrutura funcional do tipo [Categoria FUNÇÃO ([Categoria Argumento 1], [Categoria Argumento 2],..., [Categoria Argumento n])] . As funções codificam as relações entre os argumentos e são munidas de restrições conceituais, rotuladas pelas categorias ontológicas.

Em outras palavras, Jackendoff procura identificar funções que expliquem padrões gramaticais de combinação, usando a ontologia como um ponto de referência. A proposição das categorias ontológicas é o ponto de partida para a identificação de generalizações de como as palavras podem se combinar para formar constituintes maiores. Graças à capacidade criativa da linguagem, fica claro que não se aprende simplesmente maneiras específicas de relacionar as palavras, existem padrões de combinação regulares e produtivos que se aplicam a classes de palavras e sentenças. Ao permitir o mapeamento de palavras e sintagmas em classes mais gerais, a ontologia dá suporte para a identificação de tais padrões de combinação (VERSPOOR, 1997).

Vale ressaltar ainda que, com o crescente interesse no estudo da representação e organização lexical<sup>10</sup>, torna-se cada vez mais possível isolar os componentes de significado comuns a verbos participantes de uma mesma classe. Tais componentes podem então ser usados para determinar a representação de verbos através de línguas distintas. A abordagem da TEC, ao considerar a representação semântica como um subconjunto da estrutura conceitual, ou seja, da linguagem de representação mental. A Figura 3 mostra as principais funções conceituais da TEC (JACKENDOFF, 1990, p. 43)<sup>11</sup>.

- (a) [Lugar FUNÇÃO-LUGAR ([Coisa ])]
- (b) [Trajetória PARA ou DE ou EM\_DIREÇÃO\_A ou DISTANTE\_DE ou VIA ([Coisa ] ou [FUNÇÃO-LUGAR])]<sup>12</sup>
- (c) [Evento IR ([Coisa ], [FUNÇÃO-TRAJETÓRIA])]<sup>13</sup>
- (d) [Estado ESTENDER-SE ([Coisa ], [FUNÇÃO-TRAJETÓRIA])]
- (e) [Evento CAUSAR ([Coisa ] ou [Evento IR ([Coisa ], [FUNÇÃO-TRAJETÓRIA])] ou [Evento PERMANECER ([Coisa ], [FUNÇÃO-LUGAR])], [Evento IR ([COISA], [FUNÇÃO-TRAJETÓRIA])] ou [Evento PERMANECER ([Coisa ], [FUNÇÃO-LUGAR])]]]
- (f) [Evento PERMANECER ([Coisa ], [FUNÇÃO-LUGAR])]

---

<sup>10</sup> Grimshaw (1990), Jackendoff (1983, 1990), Levin & Rappaport (1993), Pustejovsky (1995) e Hale & Keyser (2002).

<sup>11</sup> Como se trata de uma metalinguagem formal para a representação de estruturas conceituais e, portanto, independentes de língua, optamos, por julgar mais conveniente e inteligível, por transpor as categorias ontológicas e as funções conceituais para o português.

<sup>12</sup> Esta função, a FUNÇÃO-TRAJETÓRIA, diferentemente da FUNÇÃO-LUGAR, está especificada em função dos diferentes tipos de trajetória. A descrição detalhada da FUNÇÃO-LUGAR será feita oportunamente, quando discutirmos as extensões e adaptações da TEC propostas por Dorr (1993). Cumpre destacar que tanto as categorias ontológicas quanto as funções conceituais, enquanto objetos da metalinguagem da TEC, podem ser redimensionadas, refinadas e reestruturadas em função de análises lingüísticas e conceituais mais acuradas ou em função de necessidades de aplicação a outros campos do conhecimento, como a TA.

<sup>13</sup> Sempre que for conveniente e para facilitar a leitura e sem prejuízo da compreensão do ponto em discussão, utilizaremos uma forma abreviada para as funções. Na definição desta função, usamos a forma abreviada [FUNÇÃO-TRAJETÓRIA] no lugar da especificação detalhada [Trajetória PARA ou DE ou EM\_DIREÇÃO\_A ou DISTANTE\_DE ou VIA ou EM\_TORNO\_DE ([Coisa ] ou [FUNÇÃO-LUGAR])].

- (g) [Estado ESTAR ([Coisa ], [FUNÇÃO-LUGAR])]  
 (h) [Estado ORIENTAR-SE ([Coisa ], [FUNÇÃO-TRAJETÓRIA])]

**Figura 3. Funções conceituais da TEC construídas a partir das categorias ontológicas**

As funções conceituais impõem restrições à interpretação semântica das sentenças, como pode ser observado nos exemplos a seguir. Ao examinarmos, por exemplo, os verbos de movimento de algo ao longo de uma trajetória e verbos de estado de especificação de localização espacial, percebemos uma distinção consistente entre eles. A anomalia semântica em (d) abaixo pode ser explicada em termos da incompatibilidade de duas funções conceituais: a estrutura conceitual do predicador [Estado ESTAR ([Coisa ], [FUNÇÃO-LUGAR])] e a estrutura conceitual da circunstância [Trajetória EM\_TORNO\_DE ([Coisa ])]. O predicador exige uma FUNÇÃO-LUGAR e no exemplo temos uma FUNÇÃO-TRAJETÓRIA. Não se verifica anomalia semântica em (c) porque a FUNÇÃO-LUGAR codificada por “dentro da sala” é parte da FUNÇÃO-TRAJETÓRIA (cf. Figura 5, função b).

- a) *A mosca voou ao redor da sala.* [Evento IR ([Coisa MOSCA], [FUNÇÃO-TRAJETÓRIA])]  
 b) *O livro está dentro da sala.* [Estado ESTAR ([Coisa LIVRO], [FUNÇÃO-LUGAR])]  
 c) *A mosca voou dentro da sala* [Evento IR ([Coisa ], [FUNÇÃO-TRAJETÓRIA])]  
 d) *\*O livro está ao redor da sala.*

Esses exemplos ilustram que as funções conceituais podem auxiliar a sistematização dos verbos em termos de tipos semânticos:

- e) - verbos de movimento: [Evento IR ([Coisa ], (FUNÇÃO-TRAJETÓRIA))]  
 f) - verbos de localização: [Estado ESTAR ([Coisa ], (FUNÇÃO-LUGAR))]

O evento expresso tanto pela sentença do inglês *The ball rolled toward Maria* como pela sentença do português *A bola rolou em direção à Maria* é representado pela estrutura conceitual a seguir:<sup>14</sup>

- [Evento IR ([Coisa BOLA],  
 [Trajetória EM\_DIREÇÃO\_A ([Lugar FUNÇÃO-LUGAR ([Coisa MARIA])])])]

<sup>14</sup> Nessa estrutura conceitual, tomamos BOLA e MARIA como etiquetas conceituais, não se trata de palavras do português.

Um outro ponto importante que merece destaque é a extensão das funções para domínios não-espaciais. Nos exemplos (g) e (h), podemos inferir, em (g), que a boneca é de Maria e, em (h), que a boneca é transferida para Maria e que, depois dessa transferência, Maria passa a possuí-la (trata-se, portanto, de uma analogia: o objeto percorre um caminho para chegar até Maria).

g) *A boneca pertence a Maria* [Estado ESTAR<sub>Posse</sub> ([Coisa BONECA], (FUNÇÃO-LUGAR))]

h) *Maria recebeu a boneca* [Evento IR ([Coisa BONECA], (FUNÇÃO-TRAJETÓRIA))]

Esses exemplos motivam a extensão das funções do domínio espacial para o domínio de posse, outra categoria ontológica. Assim, esse domínio de posse e os demais domínios apontados na literatura (Identidade, Tempo, Circunstância, Existência)<sup>15</sup> sugerem a utilidade das funções identificadas por Jackendoff e reflete regularidades no uso dos componentes de uma língua em domínios diferentes. Um ponto essencial é observar que a extensibilidade das funções conceituais espaciais para outros domínios demonstra que os campos semânticos têm essencialmente a mesma estrutura, e que o domínio espacial é, em geral, empregado para definir os termos a partir dos quais muitos tipos de discurso são estruturados.

Como veremos com maior detalhes a seguir, a extensão proposta por Dorr (1993) para as funções irá fornecer para a sentença *A bola rolou em direção à Maria* uma representação da estrutura conceitual com maior granularidade, conforme ilustra a estrutura conceitual abaixo:

[Evento IR ([Coisa BOLA],  
[Trajetória EM\_DIREÇÃO\_A ([Posição EM ([Coisa BOLA], [Coisa MARIA])])])]

---

<sup>15</sup> As sentenças a seguir servem de pistas para a identificação desses domínios: *Maria se tornou mãe/ O médico continuou incompetente/ Maria é dentista* (Identidade); *A reunião será das 10h às 11h/ Mantemos o jogo no domingo/ O baile é às 11 horas* (Tempo); *Maria começou a vender jóias/ Maria continua a vender jóias/ Maria vende jóias* (Circunstância); *Maria construiu uma ponte/ A ponte continua lá/ Chomsky existe* (Existência).

Assim, um evento estruturado em termo do primitivo IR associa dois argumentos, um do tipo Coisa e outro do tipo Trajetória, criando restrições léxico-semânticas combinatórias para cada tipo ontológico e ao mesmo tempo mapeando os primitivos em direção a uma correspondência direta com a estrutura sintática da sentença.

O quadro 3 abaixo lista e antecipa as funções e os primitivos conceituais que são implementos no UNITRAN.<sup>16</sup>

<b>Categorias ontológicas</b>	<b>As funções e primitivos conceituais do UNITRAN</b>
EVENTO (EVENT)	CAUSA, DEIXAR* <sup>17</sup> , IR, PERMANECER
ESTADO (STATE)	ESTAR, IR-EXTENSÃO*, ORIENTAR
POSIÇÃO (POSITION)	EM, DENTRO_DE, SOBRE
TRAJETÓRIA (PATH)	PARA, DE, EM_DIREÇÃO_A, LONGE_DE, VIA
COISA (THING)	LIVRO*, PESSOA*, REFERENTE*, FERIMENTO_COM_FACA*, FACA*, OBJETO_AFIADO*, FERIMENTO*, PÉ*, MOEDA*, TINTA*, FLUIDO*, SALA*, SUPERFÍCIE*, PAREDE*, CASA*, BOLA*, BONECA*, REUNIÃO*, SAPO*
PROPRIEDADE* (PROPERTY)	CANSADO*, FAMINTO*, SATISFEITO*, QUEBRADO*, COM_SONO*, MORTO*, ESTICADO*, FELIZ*, VERMELHO*, QUENTE*, DISTANTE*, GRANDE*, FÁCIL*, CERTO*
LOCALIZAÇÃO* (LOCATION)	AQUI*, LÁ*, ESQUERDA*, DIREITA*, PARA-CIMA*, PARA_BAIXO*
TEMPO (TIME)	HOJE*, SÁBADO*, 2:00*, 4:00*
MODO (MANNER)	DE_MODO_FORÇADO*, DE_MODO_SEMELHANTE*, BEM*, DE_MODO_RÁPIDO*, DANÇANDO*, SEMELHANTE*, DE_MODO_FELIZ*, DE_MODO_AMÁVEL*, DE_MODO_AGRADÁVEL*, DE_PRESENTE*, PARA-CIMA*, PARA_BAIXO*, PARA_DENTRO*, HABITUALMENTE*

**Quadro 3. Funções e primitivos conceituais do UNITRAN.**

<sup>16</sup> Optamos também por transpor toda a metalinguagem do UNITRAN para o português.

<sup>17</sup> O asterisco (\*) sinaliza os elementos adicionais que integram o UNITRAN.



Tais restrições ocorrem nas três dimensões léxico-semânticas descritas por Jackendoff: espacial, causal e de campo.

A estrutura de argumentos na dimensão **espacial** obedece às restrições do Quadro 4:

EVENTOS		
PRIMITIVO	ARGUMENTO 1	ARGUMENTO 2
IR	Coisa	Trajatória
PERMANECER	Coisa	Posição

**Quadro 4. A estrutura de argumentos na dimensão espacial.**

ESTADOS		
PRIMITIVO	ARGUMENTO 1	ARGUMENTO 2
ESTAR	Coisa	Posição
ORIENTAR	Coisa	Trajatória
IR-EXTENSÃO	Coisa	Trajatória

**Quadro 5. Restrições da dimensão espacial.**

A estrutura de argumentos na dimensão **causal** obedece às restrições do Quadro 6:

PRIMITIVO	ARGUMENTO 1	ARGUMENTO 2
CAUSAR	Coisa	Evento
	Evento	Estado
DEIXAR	Coisa	Evento
	Evento	Estado

**Quadro 6. Restrições da dimensão causal**

Finalmente, a estrutura de argumentos na dimensão de **campo** obedece às restrições do Quadro 7.

CAMPO	ARGUMENTO 1	ARGUMENTO 2
Localização	Coisa	Lugar
	Evento	
Posse	Coisa	Coisa

Temporal	Evento Estado	Tempo
Identificação	Coisa	Coisa Propriedade
Circunstancial	Coisa	Evento Estado
Existencial	Coisa	EXISTENCIAL

**Quadro 7. Restrições da dimensão campo**

Por exemplo, a estrutura conceitual da sentença *O sinal aponta para São Paulo* é dada pela representação a seguir, em que o primitivo ORIENTAR seleciona um argumento COISA e um argumento TRAJETÓRIA:

[Estado ORIENTAR<sub>Loc</sub>  
 ([Coisa SINAL],  
 [Trajetória EM\_DIREÇÃO\_A<sub>Loc</sub> ([Posição EM<sub>Loc</sub> ([Coisa SINAL], [Localização SÃO PAULO])])])]

### 2.2.2 Papéis temáticos

Na TEC, com se discutirá mais abaixo, os papéis temáticos são considerados posições na estrutura conceitual. Por exemplo, os papéis temáticos ATOR e PACIENTE são considerados entidades representadas em suas relações, movimentações, lugares, etc. Uma série de ações e uma série temática são introduzidas na representação dos EVENTOS. Num exemplo de Jackendoff (1990, p.130):

- a) *What Bill did to the books was to load them on the truck.*  
O que Bill fez aos livros foi carregá-los no caminhão.
- b) *? What Bill did to the truck was load the books onto it.*  
O que Bill fez ao caminhão foi carregar os livros sobre ele.
- c) *\* What Bill did to the books was load the truck with them.*  
O que Bill fez aos livros foi carregar o caminhão com eles.
- d) *What Bill did to the truck was load it with books.*  
O que Bill fez ao caminhão foi carregá-lo com livros.

Em cada caso, o evento expresso é o mesmo: os livros vão para cima do caminhão. O que muda, como se reflete na forma sintática, é a entidade que é vista como PACIENTE, a que é mais afetada pela ação de Bill. O papel PACIENTE, em a, é desempenhado por “os livros”, em b, por “o caminhão”. A razão da baixa aceitabilidade de c, por exemplo, é que há um conflito entre a entidade (X) explicitamente identificada como recebendo a ação de Bill; Bill fez alguma coisa para X, e a entidade sugerida pela sintaxe da sentença clivada é a de PACIENTE.

O exemplo sugere que a noção semântica de PACIENTE é relevante para o mapeamento entre a sintaxe e a semântica, e que essa relação temática deve ser representada para dar conta da distinção entre as sentenças paralelas acima (VERSPOOR, 1997).

Na visão de Jackendoff, como já se disse, as relações temáticas não estão associadas a posições sintáticas, mas antes correspondem a configurações particulares na estrutura conceitual (1990, p. 46). Assim, por exemplo, o papel temático Origem (*Source*), ou seja, “o objeto do qual parte o movimento”, é o argumento da FUNÇÃO-TRAJETÓRIA FROM; o papel temático Meta (*Goal*), ou seja, “o objeto em direção ao qual o movimento é direcionado”, é o argumento da FUNÇÃO-TRAJETÓRIA PARA; o papel temático Agente (*Agent*) é o primeiro argumento do FUNÇÃO-CAUSAR, e assim por diante.

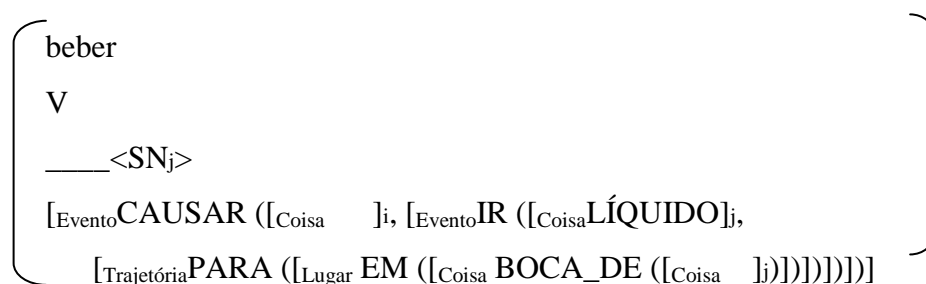
Essa estratégia de não se considerar os papéis temáticos como elementos primitivos da teoria permite tornar a teoria parcimoniosa. Por exemplo, em uma sentença como *João escalou a montanha*, a Meta do verbo *escalar* não é a montanha, mas o seu topo. Ou seja, a estrutura conceitual da sentença *João escalou a montanha* é dada por [Evento IR ([Coisa JOÃO], [Trajetória PARA (TOPO-DE ([Coisa MONTANHA]))]]. Assim, não é preciso postular esse novo papel temático.

Segundo Higginbotham (1985, p. 555, *apud* JACKENDOFF, 1990, p. 48), deve-se, no entanto, considerar as grades temáticas das unidades lexicais com um mecanismo de indexação sintática. Por exemplo, o verbo *ver* possui a grade temática <1, 2, E>, ou seja, um evento constituído de dois argumentos. Já Jackendoff (1990, p. 48) propõe apresentar o que subjaz esse e outros modelos de estrutura de argumentos. Usando as estruturas conceituais como suporte, ele propõe que se postule uma lista dos argumentos do verbo ao se extrair os índices da sua estrutura conceitual; o papel temático de cada constituinte da sentença é determinado por meio da conexão que se estabelece entre esse

constituente e a posição estrutural correspondente na estrutura conceitual. Ou seja, a estrutura de argumentos de um verbo pode ser vista como a forma abreviada da parte da estrutura conceitual que é visível para a sintaxe. As restrições referentes ao número e ao tipo de relações temáticas advêm de quaisquer restrições existentes no escopo das funções conceituais necessárias para expressar os significados de verbos e preposições. Surgem portanto três conseqüências desse ponto de vista: (i) não há papel temático para o sujeito; essa é uma relação sintática, não conceitual, e sujeitos sintáticos podem possuir uma variedade de papéis temáticos; (ii) não apenas os SNs recebem papéis temáticos. Por exemplo, Em *A luz mudou de vermelho para verde*, *verde* é Meta; (iii) não deve existir um papel temático *default* como o papel temático Objetivo (“neutro”) de Fillmore (1968).

### 2.2.3 Restrições seletivas

Na proposta de Jackendoff, as restrições seletivas são parte do significado do verbo e devem estar integradas à estrutura de argumentos. No verbo *beber*, assumindo seu significado como “fazer com que um líquido entre na boca de alguém”, é dada na Figura 4 uma possível representação:



**Figura 4. Representação das restrições seletivas na TEC**

A restrição seletiva do objeto direto (o SNj) aparece como a informação conceitual LÍQUIDO, com ele coindexada. Para que se faça uso dessa informação, não se deve simplesmente substituir a leitura do objeto direto pelo constituinte *j*; ao contrário, é necessário considerar a leitura do argumento SN de alguma maneira fundida (fusão, *merge*) com a informação semântica já contida no constituinte. Assim, propõe-se que se troque a substituição do argumento com uma regra para fusão de argumento.

Em *Pedro bebeu o vinho*, a fusão de argumentos combina a leitura de *vinho* com o constituinte [Coisa LÍQUIDO]j; o traço redundante LÍQUIDO é apagado. A leitura resultante é que em Pedro bebeu o vinho, lê-se “o líquido específico no contexto”. A fusão do argumento lida com fenômenos observáveis a respeito das restrições seletivas, incluindo o uso dessas para se ter certeza de que a maior categoria conceitual de um argumento combine com aquela estipulada pelo verbo. Portanto, as restrições seletivas dos argumentos funcionam da mesma forma na estipulação de argumentos totalmente incorporados como Tema, como no caso dos verbos (a) *amanteigar* (*Maria amanteigou o bolo*) e (b) *embolsar* (*Pedro embolsou o dinheiro*):

- a) [Evento CAUSAR ([Coisa ]i, [Events IR ([Coisa MANTEIGA],  
[Trajetória PARA ([Lugar SOBRE ([Coisa ]j))])])])]
- b) [Evento CAUSAR ([Coisa ]i, [Evento IR ([Coisa ]j,  
[Trajetória PARA ([Lugar EM ([Coisa BOLSOS])])])])]

Comparando as entradas de *beber* e *amanteigar*, por exemplo, percebe-se que não há necessidade de se postular um nível na estrutura de argumentos para codificá-los. Ambos diferem semanticamente a respeito da informação sobre TEMA e TRAJETÓRIA; eles também diferem sintaticamente em que *amanteigar* é obrigatoriamente transitivo, e *beber*, é opcionalmente transitivo. (JACKENDOFF, 1990, p. 53-54).

Assim, se o constituinte do verbo está indexado, seus traços semânticos são codificados como restrições seletivas; se o constituinte não está indexado, seus traços semânticos são codificados na estrutura conceitual e ele se torna implícito. Portanto, é possível estabelecermos como relacionar argumentos da estrutura conceitual a constituintes sintáticos: cada item lexical na sentença especifica como seus argumentos conceituais estão relacionados com posições sintáticas no sintagma de que é núcleo. A fusão de argumentos usa a informação lexical para relacionar as leituras dos complementos e sujeitos sintáticos com as posições argumentais indexadas na estrutura conceitual do núcleo sintagmático. Estabelecemos também que *papel temático* é um

termo que designa uma posição estrutural na estrutura conceitual, que consiste de um conjunto de índices que relacionam os argumentos sintáticos e conceituais. Essa integração entre estrutura de argumentos e estrutura conceitual é possível porque a estrutura de argumentos é concebida como uma representação que é parte da estrutura conceitual, e não um conjunto de diacríticos que etiquetam argumentos sintáticos. Para Jackendoff, essa é uma grande virtude da TEC. (1990, p. 55).

#### 2.2.4 Dois tipos de coindexação

Para relativizar a Teoria  $\theta$  (seção 2.1.6) e compatibilizá-la com a TEC, Jackendoff (1990, p.59-68) postula que no mapeamento entre as estruturas sintática e conceitual haja dois tipos de coindexação, como será mostrado a seguir. Esse ajuste da teoria sintática de papéis temáticos é necessária, dado que, na TEC, a um mesmo argumento sintático pode corresponder mais de uma posição na estrutura conceitual, o que significa dizer que a um mesmo SN podem estar associados mais de papel temático, contrariando, assim, a Teoria  $\theta$ . Os exemplos em que essa fato ocorre são com predicadores que denotam algum tipo de transação, como aqueles lexicalizados pelos verbos *comprar*, *vender* e *trocar*. Tomemos, então, a sentença (a) com o verbo cuja estrutura conceitual é dada na Figura 5 (o verbo *comprar*):

a) *Maria comprou o livro de João.*

O significado dessa sentença envolve, pelo menos, dois componentes conceituais, duas transferências de posse:

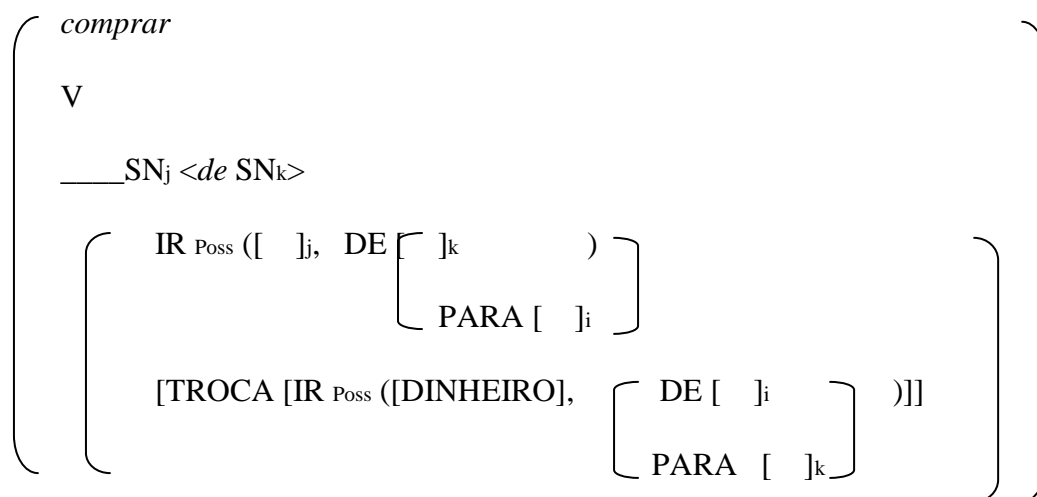
- i. a transferência de posse do livro, que passa de João para Maria (a transferência principal)
- ii. a transferência de posse do dinheiro envolvido na transação, que passa de Maria para João. (a contra-transferência)<sup>18</sup>

Essa análise mostra que os SNs *Maria* e *João* recebem dois papéis temáticos<sup>19</sup>.

---

<sup>18</sup> Embora o Tema da contra-transferência, o dinheiro, não esteja sintaticamente representado, ele está presente na estrutura conceitual como o primeiro argumento da função IR, que, por sua vez, é o argumento da função TROCA.

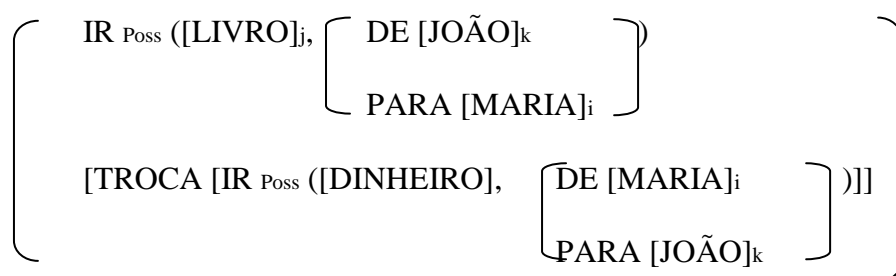
<sup>19</sup> Observe-se que não se poderia dizer que os papéis temáticos da contra-transferência são irrelevantes, dado que é a contra-transferência que permite a distinção entre *comprar* e *obter*.



**Figura 5. Estrutura conceitual do verbo *comprar*.**

Esse fato é problema para a Teoria  $\theta$ , porque, para essa teoria, os papéis temáticos são etiquetas colocadas na estrutura sintática. Já na TEC, como já foi discutido anteriormente, os papéis temáticos são definidos em termos de posições na estrutura conceitual. Em outras palavras, na TEC, duas (ou mais) posições na estrutura conceitual podem ter o mesmo índice, o que significa dizer que um mesmo constituinte da estrutura sintática pode preencher duas (ou mais) posições na estrutura conceitual.

Na Figura 13, papéis temáticos múltiplos são representados porque, na sentença (a), cada um destes SNs, *Maria* e *João* (objeto sintático da preposição *de*), está coindexado a duas posições, a saber: o SN *Maria* está coindexado simultaneamente à posição indexada pelo índice *i* da função conceitual IR (Meta do livro) e à posição *k* da função conceitual TROCA (Origem do dinheiro); o SN *João* está coindexado simultaneamente à posição indexada pelo índice *i* da função conceitual TROCA (Fonte do livro) e à posição *k* da função conceitual IR (Meta do dinheiro). A Figura 6 ilustra o resultado da operação de fusão de argumentos, necessária para construção da estrutura conceitual da sentença (a), que “preencheu”, com as duas interpretações (MARIA e JOÃO), as posições previstas na estrutura conceitual do verbo *comprar*.



**Figura 6. Estrutura conceitual da sentença *Maria comprou o livro de João.***

Essa representação, entretanto, ainda exige mais refinamento, porque em sentenças passivas como (b), por exemplo, o argumento sintático sujeito (externo ao SV) torna-se implícito. Nessa sentença, os papéis temáticos Meta da transferência principal e Origem da contra-transferência não estão sintaticamente realizados. Porém, na estrutura conceitual, as duas posições que correspondem a esses papéis precisam ser preenchidas pelo mesmo indivíduo, porque essa configuração de participantes é parte integrante da transação COMPRAR.

(b) *O livro foi comprado de Maria.*

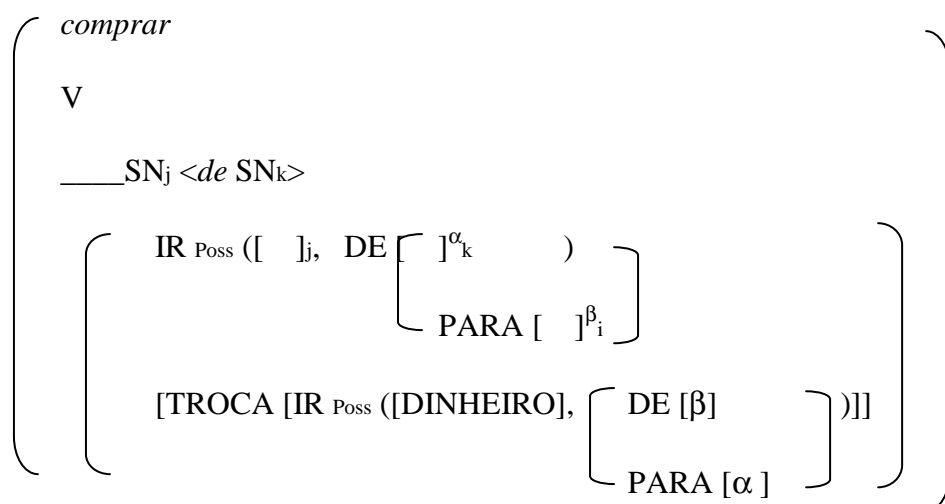
Já em sentenças como (c), em que não se explicitam a Origem da transferência principal e a Meta da contra-transferência, não se pode especificar que ambos sejam desempenhados pelo mesmo indivíduo (Maria pode muito bem ter recebido o livro de um indivíduo e ter dado o dinheiro para outro), como está especificado na Figura 6.

(c) *Maria comprou o livro.*

Para solucionar as lacunas na TEC ilustradas pelos exemplos (b) e (c), Jackendoff postula dois tipos de coindexação: (1) uma coindexação que conecta um constituinte sintático a um constituinte conceitual, os índices *i* e *k* da Figura 6); e (2), uma coindexação que estipula que dois ou mais papéis temáticos são realizados pelo mesmo indivíduo, conforme ilustra a Figura 7.<sup>20</sup>

<sup>20</sup> Observe-se que na Figura 9 não há mais duplicação de índices que especificam a correlação entre as estruturas sintática e conceitual.





**Figura 7. A coindexação entre as estruturas sintática e conceitual (letras latinas) e a coindexação entre posições conceituais (letras gregas).**

Na Figura 7, são explicitados os dois tipos de coindexação: as letras do alfabeto latino sinalizam a relação entre posições sintáticas e conceituais e as letras do alfabeto grego sinalizam a ligação entre posições na estrutura conceitual.

A Teoria  $\theta$  relativizada passa a ter a seguinte forma:

O novo critério- $\theta$

Cada índice que liga uma estrutura sintática a uma estrutura conceitual numa entrada lexical deve estar especificado uma única vez na estrutura conceitual. Todos os demais papéis temáticos que o SN por ventura desempenhar deverão estar expressos por argumentos ligados ao constituinte conceitual indexado. (JACKENDOFF, 1990, p. 64).

Assim temos o efeito formal desejado para a *Teoria de Princípios e Parâmetros*: cada argumento sintático está ligado a um único argumento conceitual.

### 2.2.5 Sobre a ligação sintática

Jackendoff (1990, p. 64) afirma que é costumeiro na literatura notar a co-referência entre SN's por meio de subscritos na estrutura sintática. Na *Teoria dos Princípios e Parâmetros* (2.1), esses subscritos são considerados parte da estrutura sintática. Como vimos, eles são marcados pela Teoria de Ligação (2.1.5), que atua na

FL das sentenças.<sup>21</sup> Entretanto, a ligação de argumentos marca, como já se observou, a correferência no interior da estrutura conceitual. A questão que se coloca é a polaridade entre ligação de constituintes sintáticos e ligação de argumentos semânticos. O mecanismo unificador dessa polaridade, para Jackendoff, é o mecanismo de ligação (*Binding*).

Em primeiro lugar, diz ele, uma correferência é uma noção fundamentalmente semântica. Qualquer que seja a notação inserida na sintaxe, a estrutura conceitual ainda precisa de algum mecanismo formal que indique que dois constituintes conceituais descrevem um mesmo indivíduo. Portanto, codificar a ligação na estrutura conceitual e não na estrutura sintática é criar um mecanismo necessário para qualquer evento.

A idéia por trás dessa abordagem é que, por exemplo, a relação entre a anáfora e seus antecedentes não deve ser tratada como coindexação na estrutura sintática, mas sim na estrutura conceitual. Observe-se, então, as duas propostas:

a) Representação da anáfora na *Teoria dos Princípios e Parâmetros*

$$\text{SN}_i \text{ liga } \left[ \begin{array}{c} \text{SN} \\ \text{anáfora} \end{array} \right]_i$$

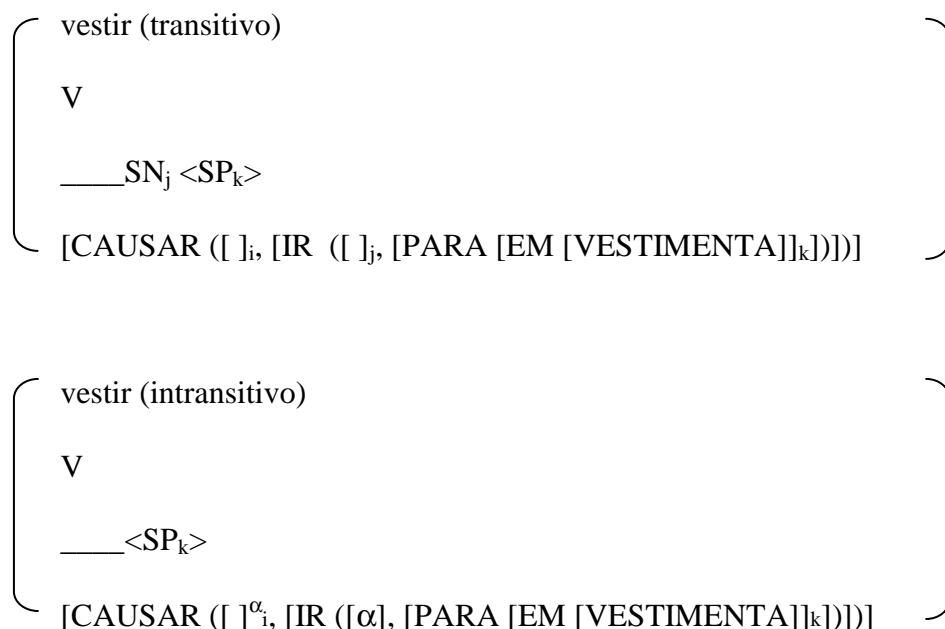
b) Representação da anáfora na TEC

$$\begin{array}{ccc} \text{SN}_i & & \left[ \begin{array}{c} \text{SN} \\ \text{anáfora } j \end{array} \right] \\ | & & | \\ \text{[Coisa } ]_i^\alpha \text{ liga} & & \text{[Coisa } \alpha]_j \end{array}$$

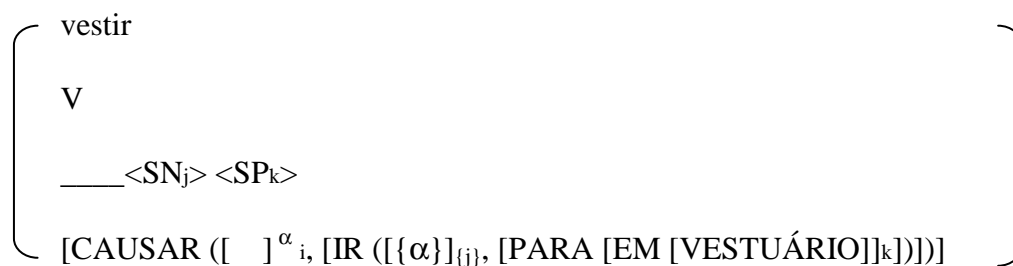
---

<sup>21</sup> Na proposta de Chomsky, considera-se que a gramática tem um nível de representação abstrato independente das estruturas profunda e de superfície, cuja função consiste em representar os aspectos do significado de uma oração que são determinados pelas suas propriedades estruturais. Este nível de representação se chama Forma Lógica (FL), e o conjunto de regras que deriva da FL chama-se componente FL (RAPOSO, 1992, p. 137).

A discussão do verbo *vestir* ilustra o problema de se estabelecer a relação entre ligação argumental e ligação sintática (adaptação da discussão de JACKENDOFF, 1990, p. 65). Na sentença *Maria vestiu (Paulo) (com um terno)*, há quatro estruturas de subcategorização do verbo. *Vestir* intransitivo não significa “vestir alguém”, mas “vestir-se”; ou seja, o indivíduo, que se veste, é um argumento ligado, não um argumento implícito. Observe-se que a forma intransitiva torna o Tema implícito e ligando ao Agente. A Figura 8 ilustra a representação das formas transitiva e intransitiva do verbo *vestir*. A Figura 9 ilustra as duas possibilidades por meio da notação de unificação de representações, em que os colchetes angulares representam constituintes opcionais e as chaves duas possibilidades de realização mutuamente excludentes: a coindexação com o objeto direto ou então a ligação com o Agente.



**Figura 8. Representação das entradas transitiva e intransitiva do verbo *vestir*.**



**Figura 9. Unificação das duas representações da Figura 8.**

## 2.2.6 A teoria da ligação e os elementos implícitos

Em um exemplo de interação da ligação com a transferência de referência, Jackendoff imagina uma situação em que estaria em um museu de cera, caminhando com o baterista/cantor Ringo Starr, e ambos deparam com a estátua de Ringo. Daí a pergunta: como dar conta de *\*Ringo cai sobre si mesmo*, pois, supostamente, Ringo tropeçou na estátua?

Jackendoff afirma que a solução depende de composição enriquecida, pois certos aspectos da ligação devem estar na estrutura conceitual e não na estrutura sintática:

*Ringo caiu sobre a estátua de si mesmo.*

[CAIR ([RINGO]<sup>α</sup>, [SOBRE [REPRESENTAÇÃO-VISUAL [α]])]

*\*A estátua de Ringo caiu sobre ele mesmo.*

[CAIR ([REPRESENTAÇÃO-VISUAL [RINGO]<sup>α</sup>], [SOBRE [α]])]

A diferença sintática decorre de que, no segundo caso, a função REPRESENTAÇÃO-VISUAL é expressa pela palavra *estátua*, enquanto na primeira ela se dá por coerção de transferência de referência (*Ringo*).

Assim, as condições sintáticas sobre a ligação são dadas em regras da forma

Se SN1 corresponde a um constituinte X<sup>α</sup> na estrutura conceitual, então [SN2, +anáfora] pode corresponder ao constituinte [α] na estrutura conceitual.

Ou seja, SN1 pode ligar SN2 somente sob as seguintes condições:

- a) condições da relação sintática de SN1 e SN2
- b) condições da relação estrutural de X<sup>α</sup> e α numa estrutura conceitual...

Em outras palavras, as condições sintáticas de ligação não estão totalmente excluídas; é que elas são parte do componente de interface e não da sintaxe em si.

## SEÇÃO 3 – INVESTIGAÇÃO DO DOMÍNIO REPRESENTACIONAL I: A DESCRIÇÃO DO SISTEMA UNITRAN E UM EXERCÍCIO DE ESPECIFICAÇÃO DAS DIVERGÊNCIAS SINTÁTICAS DE TRADUÇÃO PARA O PORTUGUÊS

### 3.1 O sistema UNITRAN: um panorama

A construção do UNITRAN (DORR, 1993), um sistema de tradução automática do tipo interlíngua, que traduz o inglês, o espanhol e o alemão entre si, com base em nas duas teorias lingüísticas esboçadas na seção anterior, parte da motivação de se modelar computacionalmente as divergências de tradução,<sup>22</sup> que são definidas como um tipo específico de problemas de natureza sintática e léxico-semântica que devem ser enfrentados por um sistema de TA e que, segundo a abordagem de Dorr, podem ser solucionados, pelo menos em parte, com a especificação de informações sintática e léxico-conceitual sintaticamente parametrizadas<sup>23</sup> no léxico do sistema. Com essa motivação, o UNITRAN demonstra que uma estrutura de representações sintática e léxico-conceituais parametrizadas no sistema pode solucionar problemas de divergências entre línguas sem recorrer a regras específicas e pontuais para cada uma delas. Em particular, o sistema demonstra que é possível a implementação da

---

<sup>22</sup> Dorr (1993) distingue sete tipos de divergências formais, que não incluem as divergências decorrentes do uso idiomático da língua e de conhecimentos aspectual, de discurso e de mundo. (DORR, 1993, *Preface*).

<sup>23</sup> É essencial ressaltar que esses parâmetros são os parâmetros do sistema UNITRAN e não os parâmetros que caracterizam a variação sintática entre as línguas, conforme a *Teoria dos Princípios e Parâmetros*. Dito de outra forma: trata-se de parâmetros implementados no sistema que são motivados por parâmetros lingüísticos. Para evitar equívocos, vamos adotar dois rótulos: o rótulo simples “parâmetro”, para designar os parâmetros lingüísticos, e o rótulo “parâmetros do sistema”, para designar os parâmetros implementados no sistema e que dão conta das divergências estruturais de tradução, quer de natureza sintática, quer de natureza léxico-conceitual.

correspondência sistemática entre a interlíngua do sistema e a estrutura sintática e léxico-semântica das línguas graças à uma base de conhecimento que codifica tipos específicos de divergências que o sistema deve manipular.

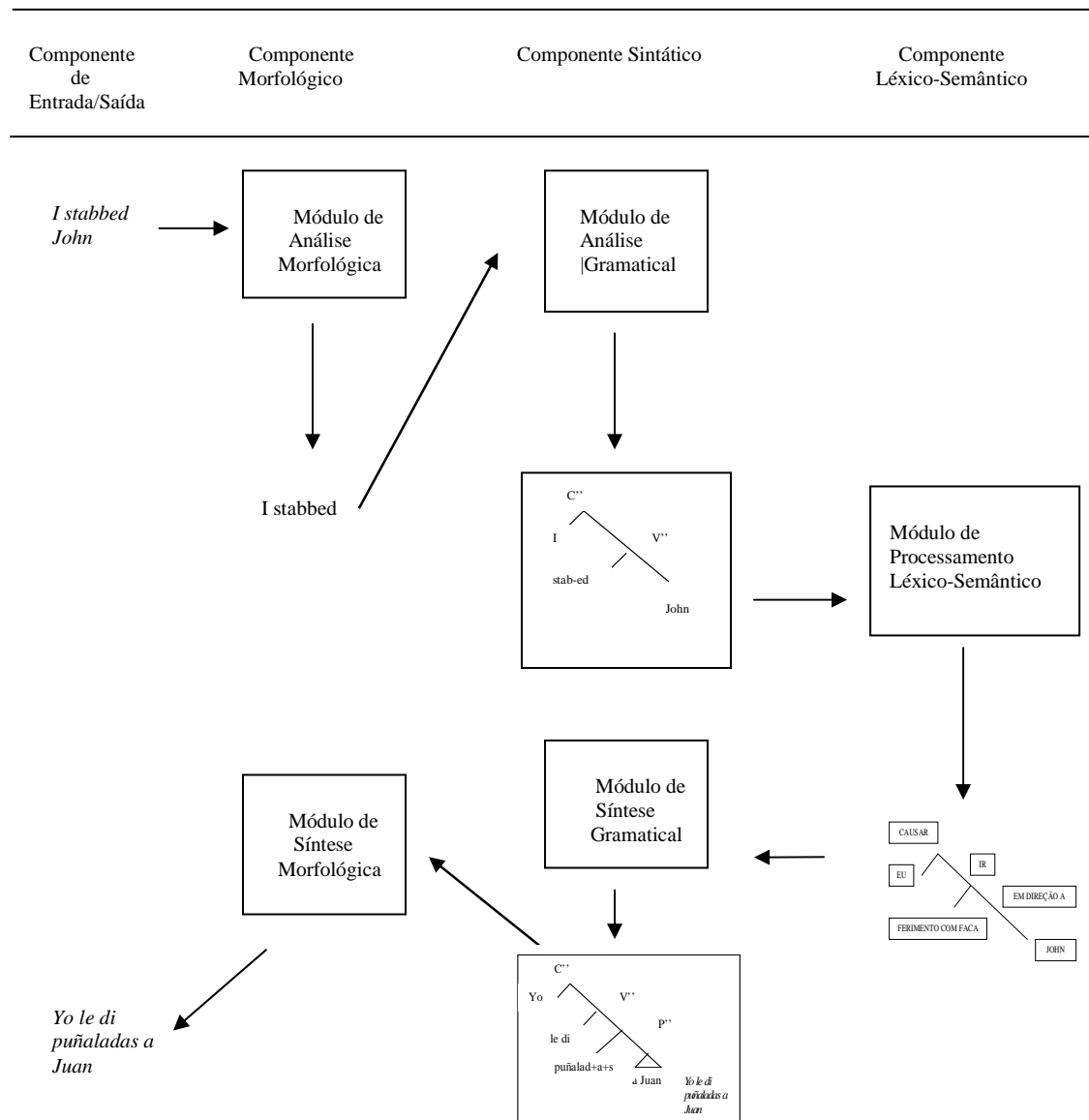
O dois módulos do sistema UNITRAN que são relevantes para a discussão nesta tese são o módulo sintático (que corresponde à Base Gramatical do sistema de PLN genérico esboçado na Seção 1) e o módulo léxico-conceitual (que funde as duas bases do sistema de PLN genérico: as Bases Lexical e Conceitual). O primeiro é construído com a implementação computacional dos constructos da *Teoria dos Princípios e Parâmetros* e parametriza a forma sintática das sentenças e dos seus constituintes; o segundo, construído com a implementação computacional dos constructos da *Teoria das Estruturas Conceituais*, parametriza a estrutura conceitual das sentenças e as propriedades léxico-conceituais dos itens lexicais a partir dos quais ela é construída.

O sistema UNITRAN é uma alternativa para sistemas de TA dos tipos direto e de transferência. O que o torna do tipo interlíngua é o fato de ele utilizar uma forma subjacente de representação da informação lingüística independente de língua e estabelecer a correspondência entre essa representação e a estrutura sintática das línguas envolvidas. Por exemplo, uma diferença gramatical entre sentenças do alemão e do inglês pode ser descrita em termos de um parâmetro sintático que descreve a ordem dos constituintes nas duas línguas. Em particular, esse parâmetro, denominado parâmetro de “núcleo de sintagma verbal”, tem o valores “*inicial*” para os verbos do inglês (Ex. *I have seen him*) e “*final*” para os verbos do alemão (Ex. *Ich habe ihn gesehen*). Assim, para o inglês, o parâmetro de núcleo de sintagma verbal colocaria, durante a geração da sentença do inglês, o SN objeto logo depois do verbo principal; para o alemão, esse parâmetro colocaria o verbo principal no final da sentença.

### **3.2 A arquitetura e o funcionamento do UNITRAN**

Baseado na composicionalidade, o UNITRAN, partindo de uma sentença da língua de partida, gera uma tradução dessa sentença na língua de chegada que não é a de substituição de palavra por palavra. A Figura 10 (adaptada de DORR, 1993, p.18), que sumariza a arquitetura do sistema e ilustra o fluxo do processo de tradução do inglês para o espanhol, destaca a interação entre a morfologia, a sintaxe, a semântica e o léxico. O esquema permite também a visualização de como o sistema seleciona no

léxico as formas correspondentes na língua de chegada (*Dar e puñaladas*) e os combina de modo a lexicalizar o mesmo conceito, que em inglês é lexicalizado pela forma verbal *to stab* (*apunhalar*), ao mesmo tempo em que constrói a estrutura sintática pertinente ao espanhol, isto é, a ordem verbo-substantivo-preposição.



**Figura 10.** A arquitetura do UNITRAN e o fluxo do processo da tradução através dos seus componentes e módulos, conforme será explicado adiante.

Como mostra a Figura 10, o sistema contém três tipos de componentes de processamento independentes (distribuídos em módulos) que permitem separar os processamentos morfológico, sintático e léxico-semântico. O *Módulo de Análise Morfológica* normaliza as formas da sentença de entrada e as passa para o *Módulo de*

*Análise Gramatical*, que as analisa e gera a estrutura sintática da sentença com base nos princípios sintáticos que são parametrizados para dar conta das idiosincrasias próprias de cada língua. O *Módulo de Processamento Léxico-Semântico* parte da representação construída pelo *Módulo de Análise Gramatical* e constrói uma representação em interlíngua, e portanto livre da sintaxe, que é então passada para como entrada para o *Módulo de Síntese Gramatical*, que gera a estrutura sintática da sentença da língua de chegada. Por fim, o *Módulo de Síntese Morfológica* gera a forma final da sentença da língua de chegada.

Note-se que, porque a variação sintática entre as línguas de partida e de chegada foi fatorada no nível do processamento léxico-semântico, o *Módulo de Processamento Léxico-Semântico* encarrega-se da tarefa de compor a representação interlingual a partir da forma sintática construída pelo *Módulo de Análise Gramatical* e passá-la para o *Módulo de Síntese Gramatical*. Assim, duas tarefas são realizadas durante o processamento léxico-semântico: (i) a composição de uma única representação interlingual a partir das formas lexicais associadas com os nós da árvore sintática; e (ii), decomposição da representação interlingual em realização na superfície.

O fato de alguns sistemas de TA operarem com base em representações não-composicionais e terem sido desenvolvidos para uma língua específica, faz com que esses sistemas não levem em conta os casos em que os argumentos possuem um *status* composicional que precisa ser levado em conta durante o processo de correspondência entre representações de línguas diferentes. Com a abordagem composicional, permite-se que os argumentos dos verbos, por exemplo, sejam considerados unidades composicionais por si próprios. Portanto, as propriedades do verbo somadas às propriedades dos seus argumentos são consideradas em cada passo do processo de TA.

O UNITRAN utiliza uma única representação subjacente para os pares de línguas envolvidas e fatora informações específicas de língua por meio de marcações de parâmetros do sistema, que são representados no léxico.

### **3.3 Os módulos de processamento do UNITRAN**

#### **3.3.1. Os Módulos de Análise e de Síntese Morfológica**

A idéia fundamental desse módulo é a reversibilidade do modelo: os mesmos léxico e gramática de uma língua são usados tanto para a análise quanto para a síntese



de uma sentença. As regras morfológicas e sintáticas são representadas por transdutores de estado finito. Uma palavra, como por exemplo, *tries*, é gerada a partir de sua raiz (*try*); a flexão (+*ies*) é obtida com base nas regras morfológicas do inglês. Além disso, o *Módulo de Análise Morfológica* retorna informações do tipo classe, número, pessoa, tempo e modo verbal, ou seja, traços. O *Módulo de Análise Morfológica* baseia-se no analisador morfológico clássico denominado *Kimmo* (1983), adaptado para o sistema com os seguintes princípios (DORR, 1993, p. 81):

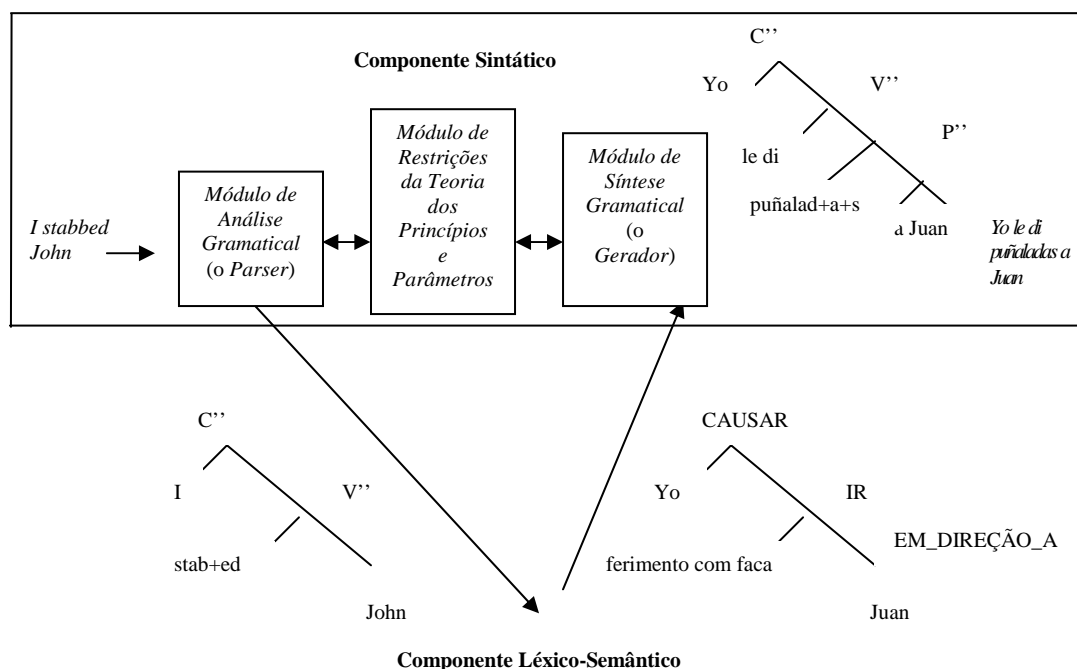
1. As formas flexionadas da mesma unidade lexical são mapeadas à sua forma canônica no dicionário do sistema (ou léxico do sistema);

2. As categorias gramaticais são representadas como parte da forma lexical.

Assim, o léxico do sistema fica adequado à utilização tanto do analisador quanto do gerador do sistema, pois consiste de raízes, uma lista de traços (eg. número, pessoa, etc.) e uma lista de *especificações de continuidade*, que consiste de uma raiz (eg. *try*) e uma *classe de continuidade* (como \*IV1), que se refere a um conjunto de sufixos (como +*s*, +*ing*, etc.) que são aplicáveis a uma raiz, isto é, afixos que “dão continuidade” à forma de base.

### 3.3.2 Os Módulos de Análise e de Síntese Gramatical

A Figura 11 esquematiza o *Componente Sintático* do UNITRAN. Nesse componente, *Módulo de Análise Gramatical* (o *parser*) constrói a estrutura sintática da sentença da língua de partida e a passa para o *Componente Léxico-Semântico*. O resultado do processamento realizado pelo *Módulo de Processamento Léxico-Semântico* é passado para o *Módulo de Síntese Gramatical* (o *gerador*), que se encarrega da construção da estrutura sintática da sentença da língua de chegada. Assim, esses dois módulos operam no sistema, respectivamente, depois da análise morfológica da sentença da língua de partida e antes da síntese morfológica da sentença da língua de chegada.



**Figura 11.** No destaque, os módulos de Análise Gramatical, de Síntese Gramatical e de Restrições da Teoria dos Princípios e Parâmetros.

Tanto o *Módulo de Análise Gramatical* quanto o *Módulo de Síntese Gramatical* interagem com o *Módulo de Restrições da Teoria dos Princípios e Parâmetros*, que fornece as restrições (isto é, os filtros próprios de cada língua) que regulam a construção das estruturas sintáticas por eles produzidas.

O *Módulo de Restrições da Teoria dos Princípios e Parâmetros*, conforme resume o Quadro 8, implementa dezenove parâmetros, distribuídos em sete submódulos, cujas restrições são aplicadas na seguinte ordem: restrições do submódulo (1) X' (X-barra), (2) Regência, (3) Fronteira, (4) Caso, (5) Vestígio, (6) Ligação e (7) Papéis Temáticos. O submódulo X' impõe restrições aos *templates* de estrutura sintática pré-compilados para cada língua, isto é, as estruturas sintáticas previamente calculadas para cada língua antes do processamento; seis das restrições dos seis submódulos restantes são também aplicadas na representação da estrutura da sentença antes do processamento e as nove restantes operam durante o processamento. Esse módulo será estudado na subseção 3.3.3.

Submódulos do Módulo da TPP	Parâmetro	Objetivo	Momento da aplicação	Exemplo <sup>24</sup>
1. X-barra	Categorias Básicas	Inventariar as categorias nucleares	Pré-processamento	<b>No inglês, no espanhol e no português, V é núcleo de V''</b>
	Ordem dos Constituintes	Determinar a ordem dos especificadores e dos complementos	Pré-processamento	<b>No inglês, no espanhol e no português, V precede objeto</b>
	Especificadores de Base	Determinar os possíveis especificadores	Pré-processamento	<b>No inglês, no espanhol e no português, N'' é especificador de FLEX''</b>
	Adjuntos de Base	Determinar os possíveis adjuntos	Pré-processamento	<b>No espanhol e no português, A é adjunto de N, podendo ser posicionado tanto à sua direita quanto à sua esquerda</b>
2. Regência	Regente Adequado	Determinar a distribuição dos núcleos lexicais em relação aos vestígios	Durante o processamento	<b>No espanhol e no português, a preposição é sempre acompanhada do seu complemento (isto é, são línguas que não admitem “preposição deslocada”)</b>
3. Fronteira	Especificadores Deslocados	Determinar os especificadores que podem ser deslocados	Pré-processamento	<b>No inglês, no espanhol e no português, ADV pode deslocar-se para a posição de especificador de FLEX<sup>25</sup></b>
	Adjuntos Deslocados	Determinar os adjuntos que podem ser deslocados	Pré-processamento	<b>No inglês, no espanhol e no português, ADV pode ligar-se por adjunção à esquerda à projeção máxima de V</b>
	Nós de Fronteira	Restringir a distância entre o vestígio e o seu antecedente	Durante o processamento	<b>No espanhol e no português, FLEX'' não é nó fronteira</b>
4. Caso	Tipo de Regência	Especificar a atribuição de caso estrutural, isto é, a atribuição de caso em função do esquema de subcategorização das unidades lexicais	Durante o processamento	<b>No espanhol, P (a) É necessária antes do objeto <i>Juan</i> em sentenças como <i>Lo vimos a Juan</i></b>

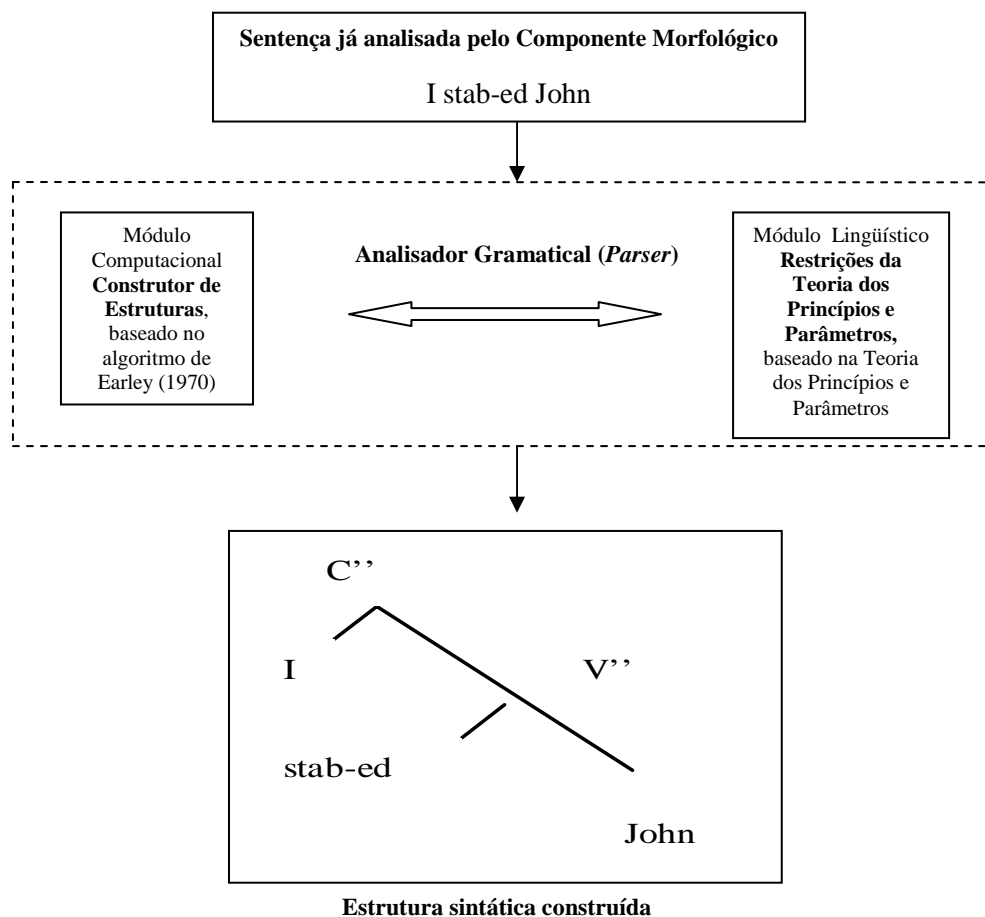
<sup>24</sup> Os exemplos ilustram o efeito da parametrização para fins específicos do UNITRAN. Não se trata de nem de uma parametrização universal e nem ela se estende para todas as construções possíveis nas línguas sobre as quais o UNITRAN opera.

<sup>25</sup> Afirmação inválida para trabalhos de Chomsky a partir de 1989.

	Propriedades de Caso	Especificar a atribuição de casos lexicais, isto é, a atribuição inerente de caso pelo item lexical	Durante o processamento	<b>No inglês</b> , <i>believe</i> atribui caso acusativo ao sujeito ( <i>him</i> ) da oração encaixada ( <i>He was not what Shula believed him to be</i> )
5. Vestígios	Vestígios	Inventariar as categorias de vestígio	Pré-processamento	<b>No inglês, no espanhol e no português</b> , V, N'' e WH- (QU-) deslocados deixam um vestígio na posição de origem
	Pronomes Vazios	Especificar a categoria e a posição dos elementos pronominais foneticamente nulos	Pré-processamento	<b>No espanhol e no português</b> a posição de especificador de FLEX'' é um <i>pro</i>
	Elementos Funcionais Vazios	Especificar a estrutura de complemento dos elementos funcionais vazios (C e FLEX)	Pré-processamento	<b>No inglês, no espanhol e no português</b> , C seleciona FLEX'' e FLEX'' seleciona V''
	Licenciamento de <i>pro</i>	Especificar a opcionalidade do sujeito	Durante o processamento	<b>No espanhol e no português</b> , há a categoria <i>pro</i> (sujeito nulo)
	Condições sobre Cadeias	Especificar se uma categoria vazia é regente apropriado	Durante o processamento	<b>No espanhol e no português</b> , o vestígio de verbo deslocado é regente apropriado e pode reger categorias vazias
6. Ligação	Domínio de Regência	Determinar o domínio de ligação	Durante o processamento	<b>No inglês, no espanhol e no português</b> , os elementos anafóricos <i>himself</i> , <i>si</i> e <i>se</i> , respectivamente, precisam estar ligados a um antecedente em um domínio local.
7. Papéis Temáticos	Clíticos	Especificar os elementos clíticos	Pré-processamento	<b>No espanhol e no português</b> , HÁ pronomes clíticos acusativos E dativos
	Duplicação de Clíticos	Determinar a possibilidade de duplicação de clíticos	Durante o processamento	<b>No espanhol</b> , é permitida a duplicação de clíticos
	<i>Non-drop</i>	Determinar as condições de explicitação de elementos pleonásticos.	Durante o processamento	<b>No espanhol e no português</b> , há elementos pleonásticos vazios

**Quadro 8. Os dezenove parâmetros do Módulo Restrições da Teoria dos Princípios e Parâmetros (nove dos quais são aplicados antes do processamento das sentenças), distribuídos em sete sub-módulos independentes.**

Há que se destacar um ponto importante em relação ao *Componente Sintático*. As operações de construção das estruturas sintáticas do *Módulo de Análise Gramatical* e do *Módulo de Síntese Gramatical* são definidas independentemente das operações de restrição do *Módulo de Restrições da Teoria dos Princípios e Parâmetros*. Entretanto, durante a análise da sentença da língua de partida (e também durante a construção da estrutura sintática da sentença da língua de chegada), há uma interação entre cada um desses módulos e o *Módulo de Restrições da Teoria dos Princípios e Parâmetros*. A modularização das operações permite um processamento uniforme, em que ambos o *parser* e o *gerador* aplicam as mesmas restrições depois que uma forma básica da estrutura sintática é construída. As representações resultantes são subespecificadas, uma vez que não incluem informações sobre concordância, casos, papéis temáticos, estrutura de argumentos, etc. Essas estruturas básicas são então submetidas aos sub-módulos do Quadro 8, que adicionam as informações que faltavam (estrutura de argumentos, etc). Assim, o modelo parte do princípio de que uma estrutura sintática deve ser inicialmente atribuída a uma sentença, e que essa estrutura poderá ser eliminada ou modificada de acordo com os princípios dos sub-módulos. Portanto, conforme ilustra a Figura 12, a estrutura sintática final da sentença da língua de partida é resultado do processamento de dois módulos que operam em paralelo: o *Módulo Computacional Construtor de Estruturas*, baseado no algoritmo de Earley (1970), e o *Módulo Lingüístico de Restrições da Teoria dos Princípios e Parâmetros*. Em outras palavras, tanto o *parser* quanto o *gerador* do UNITRAN operam com as restrições advindas de um módulo eminentemente lingüístico.



**Figura 12.** A interação entre os módulos computacional e lingüístico na construção da estrutura sintática das sentenças.

Um passo essencial do *Componente Sintático* é verificar o esquema de subcategorização do verbo. Por exemplo, a subcategorização do verbo *colocar* do português (*Maria colocou o carro na garagem*) especifica que esse verbo seleciona dois argumentos: um deles é sintaticamente realizado por um SN (o *Tema: o carro*) e o outro por um SP (o *Locativo: em a garagem*). A checagem da subcategorização aumenta a eficiência do sistema ao permitir que o *parser* determine se uma regra específica é aplicável a determinado contexto. No exemplo do verbo *colocar*, durante o processo de busca da subcategorização para esse verbo, a regra  $V' = V N'' P''$  seria aplicada, pois passa no teste de subcategorização do verbo *colocar*).

As restrições sintáticas portanto reduzem o número de estruturas produzidas a partir da análise da sentença da língua de partida para somente uma estrutura primária; daí então, é que a estrutura é enviada para o *Módulo de Processamento Léxico-Semântico*, que dela extrai a estrutura conceitual para depois enviá-la para o gerador que, como já de observou, se encarrega de transformar essa estrutura conceitual em uma

estrutura sintática. Nesse ponto, há ainda um número de estruturas sintáticas para o gerador selecionar. As mesmas restrições (agora com os parâmetros ajustados para a língua de chegada) são aplicadas novamente, desta vez sobre estruturas sintática da língua de chegada. As possibilidades agramaticais são eliminadas. Finalmente, a estrutura na língua de chegada é gerada pelo *Módulo de Síntese Morfológica*.

### 3.3.3 Os Sub-módulos do Módulo de Restrições da Teoria dos Princípios e Parâmetros

#### 3.3.3.1 O sub-módulo X'

O sub-módulo X' do *Componente Sintático* fornece a representação dos constituintes sintáticos das sentenças e contém quatro parâmetros: Categorias Básicas, Ordem de Constituintes, Especificadores de Base e Adjuntos de Base.

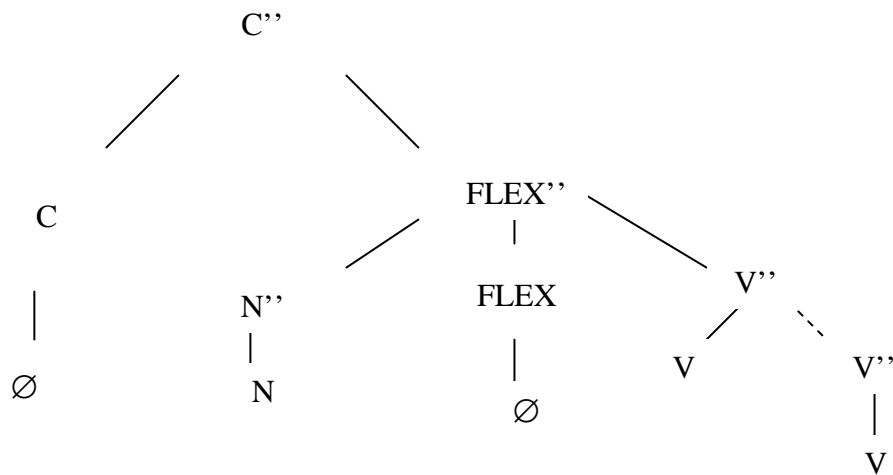
O UNITRAN implementa seis categorias sintáticas para o processamento das três línguas (cf. Quadro 9): o Complementizador (ou Complementador C), Flexão (FLEX), Verbo (V), Substantivo (N), Adjetivo (A) e Preposição (P). Dessas, C e FLEX são funcionais; as demais são lexicais, ou seja, especificadas no léxico do sistema. Por razões de implementação, C e FLEX são posições vazias (cf. Quadro 9).

<b>Categorias Sintáticas Básicas</b>
Inglês – C, FLEX, V, N, A, P
Espanhol– C, FLEX, V, N, A, P
Português – C, FLEX, V, N, A, P

**Quadro 9. O parâmetro Categorias Sintáticas Básicas.**

O parâmetro Ordem de Constituintes dá conta das distinções de ordem das palavras em uma sentença nas diferentes línguas. O estabelecimento desse parâmetro determina as posições relativas do especificador e do complemento do núcleo do sintagma. Por exemplo, em inglês, espanhol e português, o complemento do verbo (o objeto direto) está posicionado à sua direita e o especificador (o sujeito) à sua esquerda.

De modo geral, temos um *template* para a estrutura sintagmática canônica conforme ilustra as Figuras 13 e 14. Os parâmetros para as categorias C, FLEX, N, A e P são os mesmos: núcleo-inicial/espec-inicial para o inglês, espanhol e português (cf. Figura 14).



**Figura 13.** O *template* da estrutura sintagmática canônica de línguas do tipo S-V-O, como o inglês, o espanhol e o português.



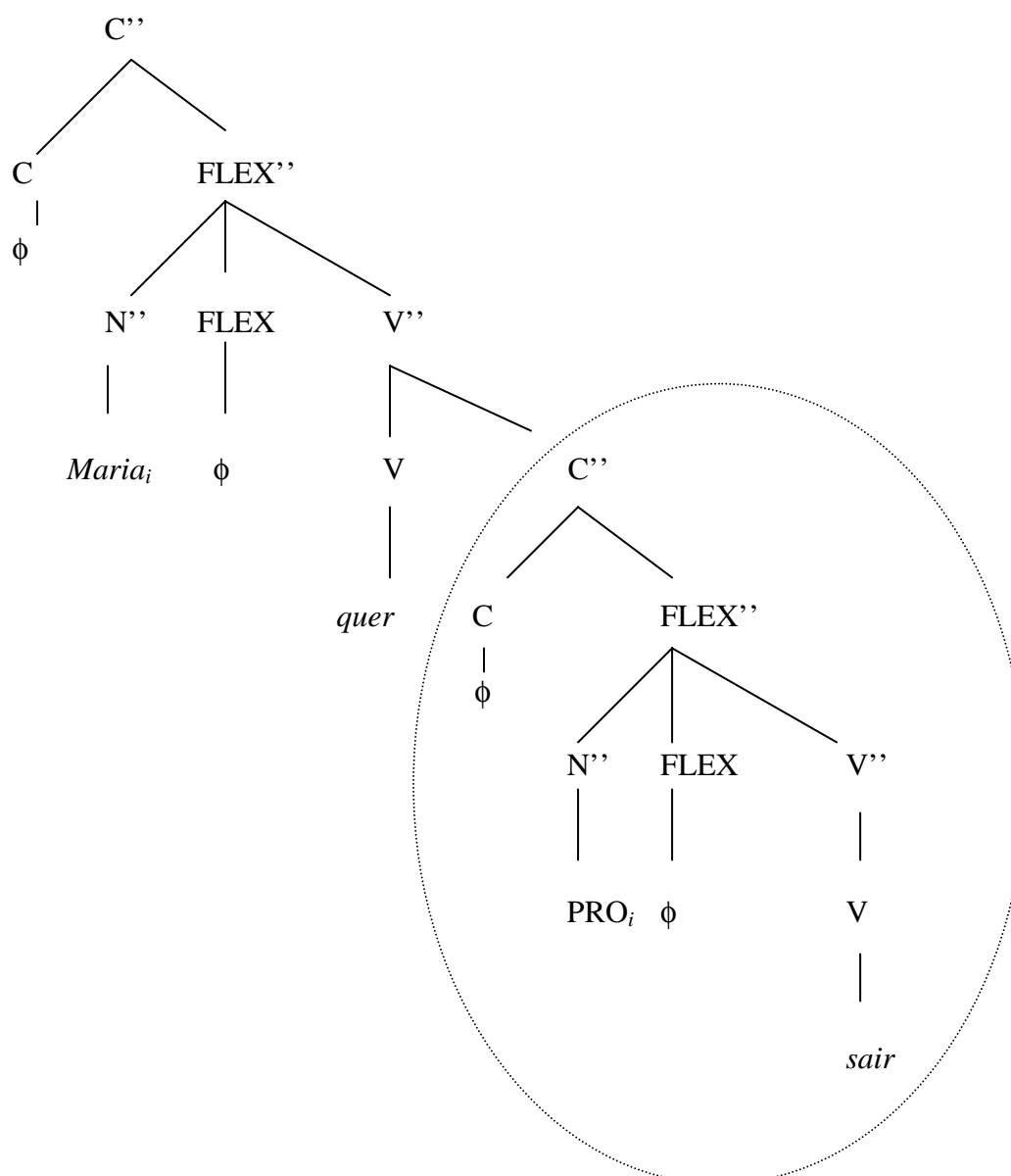


Figura 14. Estruturas complexas, com oração encaixada, indicada pela porção da estrutura que está incluída no contorno pontilhado, que dá conta de estruturas com verbo no infinitivo.

	Ordem dos Constituintes					
	C	FLEX	V	N	A	P
Inglês	1	1	2	1	2	2
Espanhol	1	1	2	1	2	2
Português	1	1	2	1	2	2

1 núcleo antes do complemento/especificador antes do núcleo

2 núcleo antes do complemento

**Quadro 10. O parâmetro Ordem dos Constituintes.**

O parâmetro Especificadores de Base determina o tipo de especificador que é permitido para cada categoria de base, e se esse especificador é obrigatório ou opcional. No UNITRAN estão implementados parâmetros para dois tipos de especificadores: os que são gerados *in situ* e os que sofrem deslocamento. Os primeiros

são especificados no sub-módulo X'; os últimos, no sub-módulo Fronteira, momento em que serão oportunamente tratados.

Tanto para o inglês quanto para o português o parâmetro Especificadores de Base é marcado com o valor opcional (cf. Quadro 11).

Especificadores de Base			
	Sujeito	determinantes/possessivas	
	FLEX	N	
Inglês	N'' 1	DET 2	N'' 2
Espanhol	N'' 1	DET 2	N'' 2
Português	N'' 1	DET 2	N'' 2

1 obrigatório / 2 opcional

**Quadro 11. O parâmetro Especificadores de Base.**

O parâmetro Adjuntos de Base especifica a posição dos adjuntos (direita, esquerda, livre). Ele diverge entre inglês e português e está ligado à adjunção de sintagmas adverbiais e preposicionais com relação ao sintagma verbal (cf. Quadro 12).

Adjuntos de Base		
	N	V
Inglês	NUM, A, C'', P'' 1, 1, 2, 2	CLITICO, ADV, P'', N'' 4, 2, 2, 2
Espanhol	NUM, C'', P'', A'' 1, 2, 2, 3	ADV, P'', CLITICO, ADV, P'', N'' 4, 4, 3, 2, 2, 2
Português	NUM, C'', P'', A 1, 2, 2, 3	ADV, P'', CLITICO, ADV, P'', N'' 4, 4, 3, 2, 2, 2

**1 – posicionamento à esquerda – nível mínimo da estrutura**

**2 – posicionamento à direita – nível máximo da estrutura**

**3 – posicionamento livre – nível mínimo da estrutura**

**4 - posicionamento à direita – nível mínimo da estrutura**

**Quadro 12. O parâmetros Adjuntos de Base**

Por exemplo, o adjetivo em inglês, em geral, precede o substantivo, mas pode ser de posicionamento livre em português e espanhol:

Inglês	The [A' large] book *the book [A' large]
Espanhol	El [A gran] libro El libro [A grande]
Português	O [A' grande] livro O livro [A' grande] <sup>26</sup>

### Os parâmetros dos demais sub-módulos

Além das restrições do sub-módulo X', as restrições dos outros sub-módulos são utilizadas na construção das estruturas sintáticas das sentenças. Embora não estejam ligadas ao sub-módulo X', são usadas concomitantemente às restrições daquele para fornecer mais informação estrutural. Os seis sub-módulos restantes, como já mencionados anteriormente, são: Vestígio, Regência, Fronteira, Caso, Ligação e Papéis Temáticos.

#### 3.3.3.2 O sub-módulo Regência

O **sub-módulo Regência** é central para a operação dos sub-módulos Fronteira, Caso, Vestígio, Ligação e Papéis Temático. O principal parâmetro do módulo Regência é o de Regente Apropriado (cf. Quadro 13), cuja especificação permite dar conta da gramaticalidade de sentenças do inglês, como (1a) que contêm “preposições deslocadas” (*preposition stranding*), isto é, preposições não acompanhadas dos seus respectivos complementos e da agramaticalidade das sentenças do espanhol (1b) e do português (1c) que não permitem esse tipo de construção.

1a [N' What store]<sub>i</sub> did John go to v<sub>i</sub>?

1b\*[N' Cuál tienda]<sub>i</sub> fue Juan a v<sub>i</sub>?

1c\*[N' Qual loja]<sub>i</sub> João foi a v<sub>i</sub>?

---

<sup>26</sup> Note-se que no português, há diferença de significado em função da posição do adjetivo, à esquerda ou à direita do nome: anteposto, “grande” significa “obra prima”; posposto, “grande” significa “extenso”.

A sentença em inglês é gramatical porque as preposições nessa língua (no exemplo a preposição *to*) são regentes apropriados do vestígio, que é um dos tipos de categorias vazias. Essa exigência decorre naturalmente do Princípio das Categorias Vazias, parte da *Teoria dos Princípios e Parâmetros*, que, como denominado na seção 2, estipula que toda categoria vazia precisa ser regida por um regente apropriado.

Regente Apropriado
Inglês: V, P
Espanhol: V, Concordância
Português: V, Concordância

**Quadro 13. O parâmetro Regentes Apropriados**

A especificação do regente apropriado também dá conta do fenômeno do sujeito-nulo. 1b é agramatical e evidencia o fato de que Concordância, diferentemente do espanhol (2) e português (3), não é um regente apropriado no inglês, não permitindo, portanto, que estruturas como [ $\phi$  *was dancing*] sejam gramaticais.

- 1a. I know that [there was dancing].
- 1b. \*I know that [ $\phi$  was dancing].
2. Yo sé que [ $\phi$  había un baile].
3. Eu sei que [ $\phi$  havia um baile].

### 3.3.3.3 O sub-módulo Fronteira

O **sub-módulo Fronteira**, com três parâmetros, regula o deslocamento de constituintes, impondo restrições à distância que separa pares de elementos correferenciais como, por exemplo, o par formado pelo vestígio e o seu antecedente: Especificadores Deslocados (cf. Quadro 14), Adjuntos Deslocados e Nós de Fronteira. Esse sub-módulo restringe a distância que separa dois elementos correferenciais à ultrapassagem de apenas um nó fronteira, notando-se que os nós são parametrizados para cada língua (cf. Quadro 15).

Especificadores Deslocados		
	FLEX	C
Inglês	N''	N'', P'', ADV
Espanhol	N''	N'', P'', ADV
Português	N''	N'', P'', ADV

**Quadro 14. O parâmetro Especificadores Deslocados.**

Nós de Fronteira
Inglês: N'', FLEX''
Espanhol: N'', C''
Português: N'', C'

**Quadro 15. O parâmetro Nós de Fronteira.**

Adjuntos Deslocados			
	C	FLEX	V
Inglês	V	ADV, P''	ADV
Espanhol	V	ADV, P''	ADV, P''
Português	V	ADV, P''	ADV, P''

**Quadro 16. O parâmetro Adjuntos Deslocados.**

Duas tarefas cabem a esse módulo: a ligação de vestígios e o deslocamento de  $\alpha$  (isto é, o deslocamento dos constituintes sintáticos da sentença), que são operações inversas. Quando a ligação de vestígios é aplicada a uma estrutura sintaticamente analisada, ela associa posições da estrutura real da sentença (estrutura da sentença que é manifesta/percebida) a posições consideradas de base (estrutura abstrata ou canônica da sentença). Já o deslocamento de  $\alpha$ , que é aplicado sobre a estrutura gerada, associa posições consideradas de base a posições na superfície.

### 3.3.3.4 O sub-módulo Caso

O **sub-módulo Caso** contém dois parâmetros que visam garantir que os N'' recebam o caso abstrato apropriado (por exemplo, nominativo, objetivo e dativo): Tipo de Regência (cf. Quadro 17) e Propriedades de Caso (cf. Quadro 18). O primeiro parâmetro regula a atribuição de caso lexical, isto é, todo item lexical rege os seus nós irmãos (regência-C) atribuindo-lhes um caso específico; o segundo regula a atribuição de caso estrutural, isto é, a atribuição de caso em função do esquema de subcategorização das unidades lexicais.

A atribuição de casos e a propriedade *s-del*<sup>27</sup> de itens lexicais são também parametrizadas no UNITRAN. As sentenças (1), (2), (3), (4), (5), (6), (7) e (8) com os verbos *believe*, *creer* e *acreditar*, respectivamente, do inglês, espanhol e português, ilustram e explicam o papel dessa parametrização na filtragem das estruturas gramaticais (cf. o sub-módulo Vestígio).

1. *John believes he is happy.* (*he* recebe caso nominativo de CONC=3<sup>a</sup>p.sing)
2. *Juan cree que pro está contento.* (*pro* é licenciado por CONC=3<sup>a</sup>p.sing)
3. *João acredita que pro está contente.* (*pro* é licenciado por CONC=3<sup>a</sup>p.sing)
4. \**John<sub>i</sub> believes PRO<sub>i</sub> to be happy.* (PRO não pode ser licenciado porque *believes* é verbo de marcação excepcional de caso, conseqüentemente, regendo PRO, mas PRO não pode ser regido, característica que define PRO)
5. *Juan<sub>i</sub> cree PRO<sub>i</sub> estar contento.* (PRO é licenciado, dado que PRO não está regido nem por *cree*, porque não se trata de verbo de marcação excepcional de caso, e nem por CONC, porque nessa sentença é vazio)
6. *João<sub>i</sub> acredita PRO<sub>i</sub> estar contente.* (PRO é licenciado, dado que PRO não está regido nem por *acredita*, porque não se trata de verbo de marcação excepcional de caso, e nem por CONC, porque nessa sentença é vazio)
7. \**Juan cree Bill comerse el desayuno.* (*cree* não é verbo de marcação excepcional de caso, logo *cree* não marca *Bill* com caso, nem *comerse*, cujo CONC é vazio. *Bill* é um N'' pleno, isto é, foneticamente realizado, não marcado por caso, situação que viola o Filtro de Caso)
8. \**João acredita Bill tomar o café da manhã.* (*acredita* não é verbo de marcação excepcional de caso, logo *acredita* não marca *Bill* com caso, nem *tomar*, cujo

---

<sup>27</sup> Verbos que possuem o traço *s-del* possuem a propriedade de atribuir caso a um N'' que é dominado pela oração que é regida pelo verbo. Trata-se de verbos de marcação excepcional de caso como, por exemplo, o verbo *believe*, na sentença, *John believes Mary to be happy*, em que *Mary* recebe papel temático de *happy* e o caso objetivo (ACUSATIVO) de *believe*. Trata-se de apagamento de C'', que, em versões anteriores da teoria gerativa, era indicado por S', isto é, a projeção de S (sentença) com o acréscimo do complementador: C''=S'=[C+S]. Daí Dorr propor o traço apagamento de S "s-del".

CONC é vazio. Bill é um N'' pleno, isto é, foneticamente realizado, não marcado por caso, situação que viola o Filtro de Caso)

Tipo de Regência				
	P	V	N	Concordância
Inglês	Regência-C	Regência-C	Regência-C	Regência-C
Espanhol	Regência-C	Regência-S	Regência-C	Regência-C
Português	Regência-C	Regência-C	Regência-S	Regência-C

**Quadro 17. O parâmetro Tipo de Regência.**

Propriedades de Caso		
	Atribuição de caso	s-del
Inglês – <i>to believe</i>	Sim	Sim
Espanhol – <i>creer</i>	Sim	Não
Português - <i>acreditar</i>	Sim	Não

**Quadro 18. O parâmetro Propriedades de Caso.**

### 3.3.3.5 O sub-módulo Vestígios

A tarefa do **sub-módulo Vestígios** é determinar os vestígios, as categorias pronominais vazias *pro* (sujeito nulo) e PRO (sujeito de infinitivo não flexionado), os complementos das categorias funcionais C e FLEX, que para o UNITRAN, projetam posições vazias, as condições de licenciamento de *pro* e as condições de regência apropriadas exercidas pelas categorias vazias. As três primeiras tarefas são executadas no pré-processamento das sentenças e as duas últimas durante o processamento. Este sub-módulo fundamenta sua operação no Princípio das Categorias Vazias: “Se um N'' foneticamente nulo é PRO, então N'' não pode ser regido; do contrário, N'' deve ser apropriadamente regido” (DORR, 1993, p.73).

Os três parâmetros que são especificados previamente no pré-processamento das sentenças são os parâmetros Vestígios (vestígios  $v$  de elementos deslocados), Pronomes Vazios (o sujeito de infinitivo não flexionado PRO e o sujeito nulo *pro* e Elementos Funcionais Vazios (as estruturas de complementação de C e FLEX).

O primeiro especifica as categorias que ocorrem na posição original do elemento deslocado: por exemplo, em *O que<sub>i</sub> ele viu v<sub>i</sub>?* o elemento *O que* desloca-se da posição de objeto do verbo para a posição do complementador da sentença, deixando o vestígio  $v_i$  com ele indexado. O parâmetro Pronomes Vazios (cf. Quadro 20) especifica as categorias PRO e *pro*, como em *Eu<sub>i</sub> quero PRO<sub>i</sub> ir ao cinema* e *pro tenho fome*, respectivamente. O parâmetro Elementos Funcionais Vazios especifica a estrutura de complemento dos elementos potencialmente vazios C e FLEX. Conforme ilustra o Quadro 20 e 21, o valor desses parâmetros é o mesmo para o inglês, o espanhol e o português. A Figura 15 ilustra os *templates* com elementos foneticamente nulos que o UNITRAN constrói para as três línguas a partir desses três parâmetros.

Vestígios
Inglês: N'', P'', ADV, V
Espanhol: N'', P'', ADV, V
Português: N'', P'', ADV, V

**Quadro 19. O parâmetro Vestígios.**

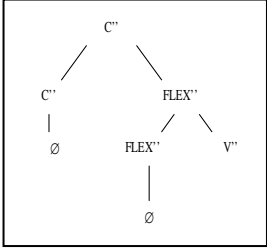
Pronomes Vazios		
	Categoria	Posição
Inglês	N''	Especificador de FLEX
Espanhol	N'', PRO/pro	Especificador de FLEX
Português	N'', PRO/ pro	Especificador de FLEX

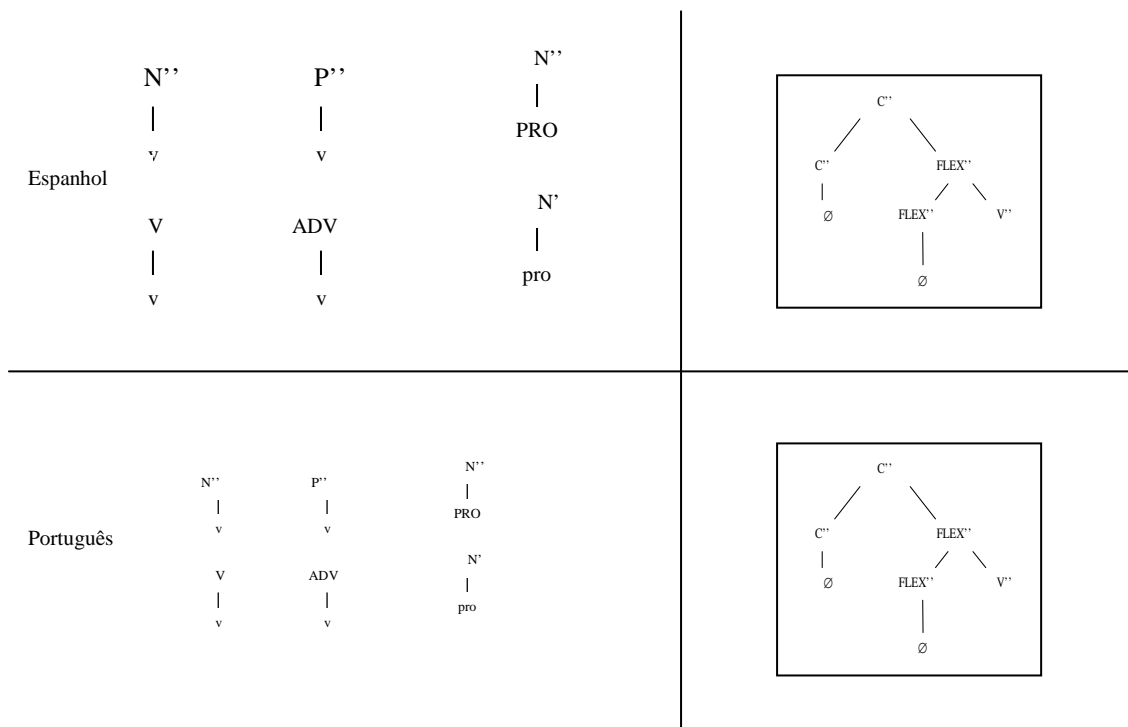
**Quadro 20. O parâmetro Pronomes Vazios.**



Elementos Funcionais Vazios		
	Categoria	Seleção de Comp.
Inglês	C	V''
	FLEX	FLEX''
Espanhol	C	V''
	FLEX	FLEX''

**Quadro 21. O parâmetro Categorias Funcionais Vazias.**

	VESTÍGIOS	PRONOMES VAZIOS	ELEMENTOS FUNCIONAIS VAZIOS	
Inglês	<p>N''</p> <p> </p> <p>v</p> <p>V</p> <p> </p> <p>v</p>	<p>P''</p> <p> </p> <p>v</p> <p>ADV</p> <p> </p> <p>v</p>	<p>N''</p> <p> </p> <p>PRO</p> <p>N'</p> <p> </p> <p>pro</p>	



**Figura 15. Os templates com elementos foneticamente nulos pré-construídos pelo UNITRAN.**

O sub-módulo **Vestígios** inclui, como já dissemos, outros dois parâmetros que se aplicam durante o processamento da sentença. Trata-se dos parâmetros **Licenciamento de *pro*** e **Condições sobre Cadeias**.

O parâmetro **Licenciamento de *pro*** (cf. Quadro 22) especifica as restrições de licenciamento de *pro* para cada língua do sistema.

Licenciamento de <i>pro</i>
Inglês: não
Espanhol: sim
Português: sim

**Quadro 22. O parâmetro Licenciamento de *pro*.**

O parâmetro **Condições sobre Cadeias** (cf. Quadro 23) especifica se o vestígio de um verbo deslocado é um regente apropriado.

Condições sobre Cadeias
Inglês: não
Espanhol: sim
Português: sim

**Quadro 23. O parâmetro Condições sobre Cadeias**

### 3.3.3.6 O sub-módulo Ligação

O **sub-módulo Ligação** opera antes do sub-módulo atribuição de Papéis Temáticos e é responsável por determinar as relações de co-referência entre N'', verificando se a categoria de regência é a oração mais próxima contendo um sujeito especificado.

Domínio de Regência	
Inglês	Tem sujeito
Espanhol	Tem sujeito
Português	Tem sujeito

**Quadro 24. O parâmetro Domínio de Regência.**

1. (i) John<sub>i</sub> likes himself<sub>i</sub>  
(ii) John<sub>i</sub> seems  $t_i$  to like Mary
2. (i) There<sub>i</sub> are men<sub>i</sub> in the room  
(ii) *pro*<sub>i</sub> me gusta Maria<sub>i</sub>.
3. (i) What<sub>i</sub> does John like <sub>$t_i$</sub>

Em (1) (ii) e (3) (i), a coindexação já foi estabelecida pelas rotinas de ligação de vestígios do sub-módulo Fronteira, e o sub-módulo Ligação deve apenas verificar se os vestígios estão ligados a seus antecedentes. Em (1) (i), (2) (i) e (2) (ii), a coindexação automática ainda não foi realizada. O sub-módulo Ligação primeiramente liga os constituintes relevantes através de coindexação; na seqüência, verifica as condições de ligação.

Uma vez que toda informação co-referencial é determinada para a árvore sintática, o sub-módulo Papéis Temáticos atribui aos N'' os papéis temáticos apropriados.

### 3.3.3.7 O sub-módulo Papéis Temáticos

O sub-módulo **Papéis Temáticos** contém três parâmetros e opera entre o *Componente Sintático* e o *Componente Léxico-Semântico* do UNITRAN: Clíticos, Duplicação de Clíticos e *Non-drop*.

Cumpra destacar que a atribuição dos papéis temáticos depois da análise gramatical leva à construção da representação do significado da sentença de entrada em termos da interlíngua do UNITRAN. A atribuição dos papéis temáticos é consolidada depois da geração da estrutura sintática, a fim de assegurar que os argumentos gerados combinem com os papéis temáticos que lhes são atribuídos pelo predicador que os seleciona. A atribuição dos papéis temáticos, seguindo a TEC, não é especificada com a aposição de símbolos como *Agente* e *Tema* aos N'', mas é automaticamente calculável a partir das posições, na Estrutura Conceitual da sentença, a que correspondem os constituintes sintáticos (cf. seção 2). Quanto ao seu grau de dependência, o sub-módulo Papéis Temáticos depende de informações contidas nas entradas lexicais dos léxicos das línguas envolvidas no processo de tradução.

O parâmetro Clíticos (cf. Quadro 25), acessado durante a precompilação, é utilizado para especificar a categoria dos clíticos (elementos foneticamente fracos).

Clíticos	
Inglês	N''
Espanhol	CL – dativo CL-acusativo
Português	CL-dativo CL - acusativo

**Quadro 25. O parâmetro Clíticos.**

O parâmetro **Duplicação de Clíticos** (cf. Quadro 26) dá conta da atribuição de papel temático a sentenças do espanhol do tipo *Lo vimos a Guille*. Nessa sentença, esse parâmetro assegura que o clítico *lo*, co-referente ao N'' *Guille* (que não tem papel temático), lhe transfira o seu papel temático.

Duplicação de Clíticos	
Inglês	Não
Espanhol	Sim
Português	Não

**Quadro 26. O parâmetro Duplicação de Clíticos**

Por fim, o parâmetro *Non-drop* (cf. Quadro 27) efetua o a atribuição de elementos pleonásticos.

<b>Non-drop</b>	
Inglês	Não
Espanhol	Sim
Português	Sim

**Quadro 27. O parâmetro *Non-drop***

Com a aplicação das restrições sintáticas implementadas nos dezenove parâmetros distribuídos nos sete sub-módulos descritos, o Componente Sintático do UNITRAN dá conta de divergências sintáticas significativas entre o inglês, o espanhol e o português e constrói tanto a estrutura sintática da sentença da língua de partida quanto a estrutura sintática da sentença da língua de chegada, a partir da representação em interlíngua (tema da próxima subseção) do significado da sentença da língua de partida.

Quanto às relações de controle, cabe ressaltar que elas são estabelecidas no componente léxico-conceitual, como parte da representação das entradas e acessadas durante o processamento do referido componente. (DORR, 1993, P. 144).

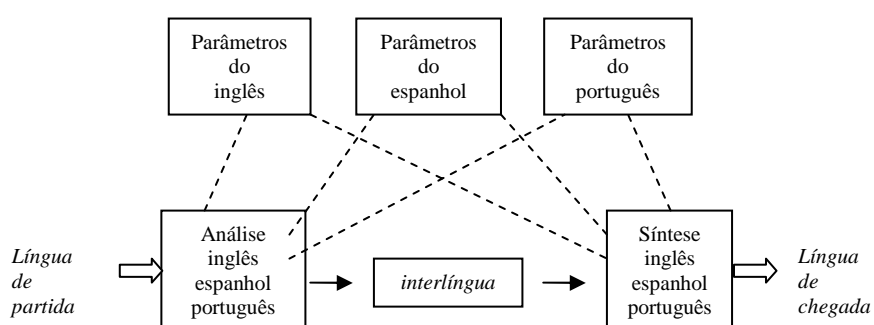
### 3.3.4 O Componente Léxico-Semântico

#### 3.3.4.1 A interlíngua do UNITRAN

A TA baseada em interlíngua pressupõe a construção de uma representação independente de língua que permita o estabelecimento de uma correspondência sistemática entre essa interlíngua e as línguas naturais. A interlíngua do UNITRAN é construída com base na *Teoria das Estruturas Conceituais* de Jackendoff (1990). A virtude dessa teoria é permitir a elaboração de uma representação relacionada à, porém independente da sintaxe e permite a codificação independente de língua, retendo um grau de estrutura suficientemente elaborado para satisfazer as exigências da tradução.

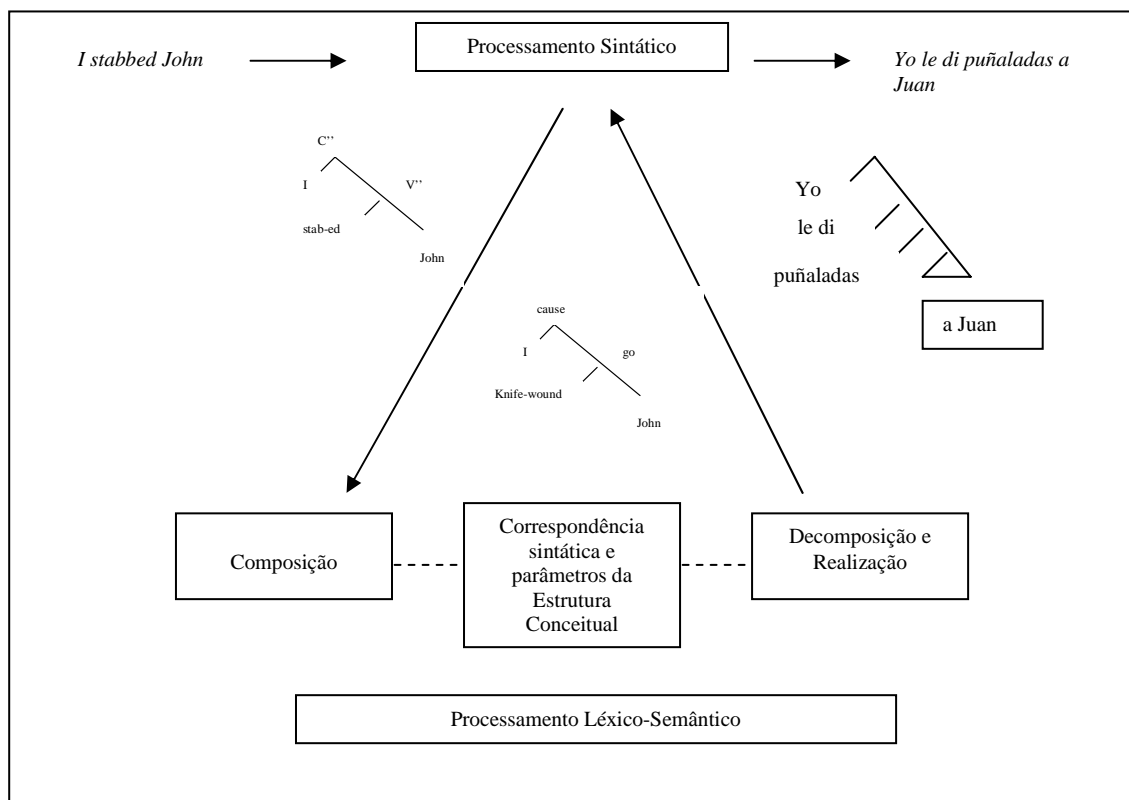
Alojada no *Módulo de Processamento Léxico-Semântico*, a interlíngua tem a funcionalidade de criar uma representação abstrata das informações lexicais e

semântico-conceituais que serve de interface entre unidades de duas ou mais línguas diferentes. Uma consequência imediata dessa propriedade é a economia de módulos, pois, para cada par de línguas envolvido no processo da TA, necessita-se apenas de um módulo de análise e de um módulo de síntese, e não um módulo para cada uma das línguas do par. Assim, a interlíngua é o denominador comum entre a língua de partida e a língua de chegada, gerando correspondências comuns entre elas. Na figura 16, Dorr (1993, p. 09) mostra o lugar da interlíngua na arquitetura do UNITRAN.



**Figura 16.** O lugar da interlíngua no sistema UNITRAN, prevendo-se aí o português.

A vantagem teórica e prática do UNITRAN sobre os outros sistemas de TA baseados em interlíngua é não multiplicar os módulos de análise e de síntese em função do número de línguas que nele possam ser incluídas, dado que, ao conformar-se com a *Teoria dos Princípios e Parâmetros*, pressupõe apenas um módulo para análise e um módulo para a síntese de sentenças, ficando as diferenças sintáticas sob a responsabilidade dos módulos específicos de cada língua, isto é, para cada língua que se queira inserir no sistema, basta a especificação dos parâmetros sintáticos e léxico-semânticos. Na Figura 17, com destaque para o *Módulo de Processamento Léxico-Semântico*, pode-se observar o módulo de Correspondência sintática e parâmetros da Estrutura Conceitual, cujos parâmetros descreveremos mais adiante.



**Figura 17.** O componente léxico-semântico é projetado de maneira que a composição e os dois processos decomposição/realização baseiam-se nos mesmos mapeamentos sistemáticos e parâmetros na estrutura conceitual. (Dorr, 1993, 164)

Um aspecto que surge na escolha da representação de qualquer sistema interlingual é a do levantamento de primitivos. Dorr (1992, p. 11) parte da seguinte generalização:

“Os primitivos devem ser definidos de forma que sua combinação capte generalizações tanto conceituais quanto sintáticas das ações de entidades que poderiam ser representados de maneira diferente.”

Por exemplo, cada ação (ex. IR) e entidade (ex. PESSOA) deve estar associada a uma representação que é conceitualmente plausível e sintaticamente ligada à estrutura sintática. Retomando o exemplo *O sinal aponta para São Paulo*, tem-se a seguinte estrutura conceitual:

[Estado ORIENTAR<sub>Loc</sub>  
 ([Coisa SINAL],

[Trajetória EM\_DIREÇÃO\_A Loc ([Posição EM Loc ([Coisa SINAL], [Localização SÃO PAULO])))]

A realização sintática dessa estrutura conceitual é dada pela representação a seguir, em que os constituintes conceituais são realizados, respectivamente, por N'', V'', P,, e N''.

[C'' [FLEX'' [N''O sinal] [V'' aponta [P'' para [N'' São Paulo]]]]]

Portanto, de acordo com a generalização inicial, as restrições dão suporte à correspondência entre as estruturas sintática e semântica ao permitir que os primitivos sejam combinados de modo que retenham informações tanto conceituais quanto sintáticas.

### 3.3.4.2 O estabelecimento da correspondência entre as estruturas sintática e conceitual

O princípio da correspondência entre as estruturas sintática e conceitual se dá da seguinte forma: a cada constituinte sintático (FLEX'', V'', N'', A'', P'') corresponde um constituinte conceitual. Por exemplo, um SN pode expressar praticamente qualquer categoria conceitual; um P'' pode expressar um Lugar, uma Trajetória, uma Propriedade; um S pode expressar um Evento ou um Estado, e assim por diante. A Figura 18 ilustra esse mapeamento.

Estrutura sintática

[Flex''[N'' João] [V'' correu [P'' para [N'' o quarto]]]]

[Evento IR ([Coisa JOÃO], [Trajetória PARA ([Lugar DENTRO\_DE<sup>28</sup> ([Coisa QUARTO])))]]

Estrutura conceitual

**Figura 18. Mapeamento entre as estruturas sintática e conceitual**

<sup>28</sup> Aqui, especificamos um dos valores da função FUNÇÃO-LUGAR, i.e. DENTRO\_DE.



A sentença corresponde à categoria Evento; o verbo corresponde à função IR; o N'' sujeito corresponde ao primeiro argumento da função conceitual IR, [Coisa JOÃO], e o P'' adjunto corresponde ao seu segundo argumento, a FUNÇÃO-TRAJETÓRIA PARA, cujo argumento é uma outra função, a FUNÇÃO\_LUGAR.

### 3.3.4.3 As entradas lexicais do UNITRAN

Cada língua processada pelo sistema requer um dicionário de entradas que contenham a especificação de estruturas conceituais. Cada entrada apresenta dois níveis de descrição: o primeiro é uma representação em estrutura conceitual independente de língua e o segundo é a especificação de parâmetros específica de língua que guia a realização sintática de uma unidade lexical e seus argumentos. Considere-se a entrada para o verbo do inglês *to stab* (*apunhalar*):

#### **to stab**

[Evento CAUSAR (Coisa \* W], [Evento IR<sub>Posse</sub>] ([Coisa FERIMENTO\_COM\_FACA], [Trajetória EM\_DIREÇÃO\_A A Posse ([Posição DENTRO\_DE<sub>Posse</sub>] ([Coisa FERIMENTO\_COM\_FACA], [Coisa \* Z])))], [COM Instr \* (Evento \*NÚCLEO\*), [Coisa OBJETO\_AFIADO U]])]

O primeiro nível de descrição especifica o significado independentemente da expressão lingüística. Em termos nocionais, essa representação pode ser assim descrita: “trata-se de um evento em que uma entidade W (Agente) faz com que Z (Paciente) sofra uma facada Y por meio de um instrumento afiado U”. O segundo nível de descrição na entrada da estrutura conceitual é dado pela notação (\*), leia-se "notação estrela", que é usada para especificar a realização sintática dos argumentos conceituais. Na seção 4 veremos como funcionam os parâmetros do sistema<sup>29</sup> na resolução das divergências de tradução.

---

<sup>29</sup> Como já advertimos, os parâmetros do sistema não são os mesmos da TPP. No sistema, são índices que apontam para operações computacionais que realizam determinada função.

### 3.3.4.4 Organização do léxico do UNITRAN

As entradas lexicais estão organizadas em classes de estruturas conceituais, montadas a partir da classificação lingüística de Levin (1985) para verbos. Essa classificação distribui os verbos em classes semânticas, como ingestão (*comer, beber, fumar, mastigar, etc.*) ou de contato/efeito (*cortar, apunhalar, esmagar, quebrar, morder, atirar, etc.*) (DORR, 1992, p. 15). Por exemplo, o verbo *to give* (da classe de mudança de posse) e *to stab* (contato/efeito) estão na mesma classe conceitual associados ao primitivo IR<sub>POSS</sub>. O quadro 28 ilustra como é uma entrada da **Estrutura Léxico-Conceitual das Unidades Lexicais do UNITRAN** (em inglês RLCS)<sup>30</sup>.

L = <i>Template</i> da função IR Poss [Evento IR Poss ([Coisa Y], [Trajetória P])]	
C = <i>Template</i> da função Causar: [Evento CAUSAR ([Coisa W], [Evento L])]	
T = <i>Template</i> da função Permitir: [Evento PERMITIR ([Coisa W], [Evento L])]	
P = <i>Template</i> de Trajetória: [Trajetória EM_DIREÇÃO_A Poss (Posição EM ([Coisa Y], [Coisa Z]))]	
Entrada	<i>ir</i> (*=[Coisa Z]; *=[Trajetória P])
Ir	<i>receber</i> (*=[Coisa: INT Y]; *=[Coisa :EXT Z])
Causar	<i>apunhalar</i> (W ≠ U, Y, Z; Y = FERIMENTO_COM_FACA; *=[Coisa W]; *=[Coisa Z]; Modificador: [COM Instr. * ([Evento *NUCLEOx ], [Coisa U INSTRUMENTO-PONTIAGUDO])])
	<i>cortar</i> (W=U, ou W≠U, Y, Z; Y = FERIMENTO_COM_FACA; *=[Coisa W]; *=[Coisa Z]; Modificador: [COM instr * ([Evento *NUCLEOx], [Coisa U INSTRUMENTO-PONTIAGUDO])])
	<i>dar</i> (W≠Y, Z; *=[CoisaW]; x = [Coisa Z]; *=[Trajetória P] ou [Coisa Z])
	<i>reaver</i> (W=Z; *=[Coisa W]; *=[Coisa Y])
	<i>obter</i> (W=Z; *=[Coisa W]; *=[Coisa Z])
Permitir	<i>aceitar</i> (W = Z; x = [Coisa W]; * = [Coisa Z])
P = <i>Template</i> de Trajetória: [Trajetória PARA-LONGE-De Poss ([Position Em ,Poss ([Coisa Y], [Coisa Z]))]	
Entrada	<i>perder</i> *=[Coisa Y]; *=[Coisa Z]
Permissivo	<i>abandonar</i> (W=Z; * = [Coisa W]; *=[Coisa Y])
	<i>render-se</i> (W=Z; *=[Coisa W]; *=[Coisa Y])

**Quadro 28. As entradas lexicais construídas (RLCS) para os verbos da classe associada ao primitivo IR poss.**<sup>31</sup>

<sup>30</sup> Observe-se o seguinte (cf. DORR, 1993, p. 132): a TEC entra no UNITRAN em DUAS dimensões, na dimensão da composição da interlíngua do sistema e na dimensão dos léxicos de cada uma das línguas. Assim, é preciso distinguir entre **Estrutura Léxico-Conceitual Composta** (em inglês abreviado por CLCS) e RLCS **Estrutura Léxico-Conceitual Raiz** (em inglês abreviado por RLCS). A primeira é construída pelo Componente Léxico-Conceitual e a segunda é especificada para cada item lexical do léxico de cada língua que integra o UNITRAN.

<sup>31</sup> As entradas lexicais são caracterizados por um *template* L na estrutura conceitual e um indicador de trajetória P. Os dois tipos de trajetória descritos aqui são EM\_DIREÇÃO\_A (ex. *ir, receber, apunhalar, cortar, dar, reaver, obter e aceitar*) e PARA-LONGE-DE Poss (ex. *perder, abandonar, render-se*).

Assim, enquanto as restrições (parâmetros) baseadas na TPP são aplicadas no Componente Sintático, as restrições (parâmetros) baseadas na TEC são aplicadas no próprio Léxico do sistema.

### 3.3.4.5 O estabelecimento da correspondência entre a interlíngua do UNITRAN e a estrutura sintática das línguas

Para relacionar as estruturas da interlíngua à língua de chegada, as entradas lexicais devem explicitar certos tipos de informação sintática e semântica específicas de cada língua. O sistema UNITRAN recebe tais informações do léxico das línguas envolvidas no processo de TA, parametrizado de forma a apontar para o sistema como construir a estrutura sintática da língua de chegada a partir de constructos da estrutura conceitual das entradas. Um exemplo de mecanismo usado nesse nível do processamento é, como já representado no Quadro 28, o marcador (\*), ou notação estrela, que marca todo argumento e modificador com realização fonética. Em outras palavras, essa notação é usada como um meio de parametrização do léxico<sup>32</sup>, uma vez que especifica as posições das estruturas conceituais que possuem realizações sintáticas correspondentes em determinada língua.

### 3.3.4.6 A geração da estrutura conceitual de uma sentença de entrada

O algoritmo para geração da estrutura conceitual depende da estrutura gerada pelo *parser* no *Componente Sintático* do sistema. Em particular, esse passo só é dado quando os papéis temáticos já tiverem sido devidamente atribuídos, uma vez que a composição da estrutura conceitual é essencialmente um processo de unificação guiados pela atribuição dos papéis temáticos. Abaixo segue o algoritmo, explicado mais adiante.

#### Composição das Estruturas Conceituais (X'')

1. Seja X = núcleo de X''
2. Seja X' = RLCS correspondente ao núcleo X.

---

<sup>32</sup> O léxico entendido como o dicionário do sistema contendo os itens lexicais das línguas, e, no caso, parametrizado com informações sintáticas e conceituais. O sistema UNITRAN se utiliza de um léxico para cada língua.

3. Seja  $W =$  argumento externo de  $X''$
4. Sejam  $Z_1, \dots, Z_n =$  argumentos internos de  $X''$
5. Sejam  $Q_1, \dots, Q_n =$  adjuntos de  $X''$
6. Para  $i \in \{W, Z_1, \dots, Z_n, Q_1, \dots, Q_n\}$ 
  - 6a. Extrair o papel temático  $i'$  associado a  $i$ . (Esta é a posição marcada com o  $(*)$  na  $X'$  que corresponde a  $i$  de acordo com a rotina de ligação).
  - 6b. Seja  $L =$  Estrutura conceitual construída
  - 6c. Unificar  $L$  com a posição  $i'$  em  $X'$ .
7. Retornar  $X'$ .

Esse algoritmo toma uma árvore sintática cujo nó do topo é  $X''$  e procura pela RLCS associada ao item lexical  $X$  (passos 1 e 2). Ele então extrai os argumentos e adjuntos de  $X''$  (passos 3 a 5) e recursivamente os passa pelo mesmo procedimento de extração dos papéis temáticos (passos 6 e 6b). A estrutura resultante é unificada com a RLCS corrente (passo 6c). Finalmente, o procedimento produz uma representação em estrutura conceitual, ou seja, uma CLCS (leia-se RLCS de  $X$  instanciada com seus argumentos) correspondendo ao núcleo do nó do topo da sentença (passo 7).

A figura 19 ilustra como se constrói a estrutura conceitual da sentença *I stabbed John*, que consiste em um processo de preenchimento de argumentos norteado pela atribuição de papéis temáticos. (cf. DORR, 1993, 149).

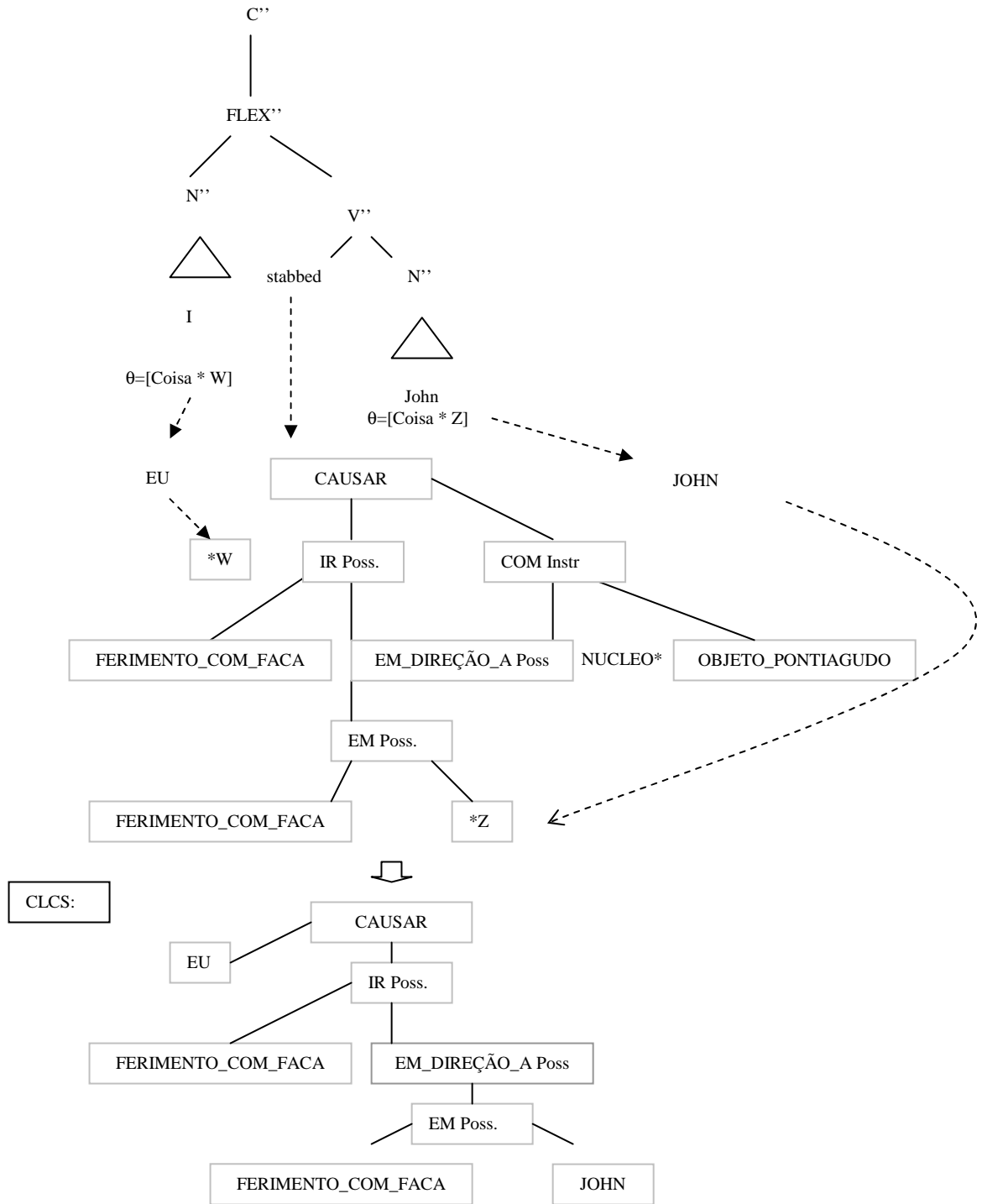


Figura 19. Composição da estrutura conceitual de *I stabbed John*.

Abaixo, o algoritmo para a geração da estrutura conceitual da sentença de partida (cf. DORR, 1993, p. 151).

1. Entrada da construção da estrutura conceitual composta (CLCS):

[C''[FLEX''[N'' I] [V'' stabbed [N'' John]]]

X =  $\emptyset$  (projeção de FLEX'')

Z1 = [FLEX'' [N'' I] [V''stabbed [N'' John]]]

2. Entrada da construção da estrutura conceitual composta (CLCS):

[FLEX'' [N'' I] [V'' stabbed [N'' John]]]

X = vazio (projeção de V'')

Z1 = [V'' stabbed [N'' John]]

3. Entrada da construção da estrutura conceitual composta (CLCS): [V'' stabbed [N'' John]]

X = stab

X' = [Evento CAUSAR [(Coisa \* W),

[Evento IR<sub>Poss</sub> ([Coisa FERIMENTO COM FACA],

[Trajetória EM DIREÇÃO A<sub>Poss</sub>

([Posição EM<sub>Poss</sub> (Coisa FERIMENTO COM FACA), [Coisa \* Z]))]],

COM<sub>Instr</sub> ([Evento \*NÚCLEO\*], [Coisa U OBJETO\_PONTIAGUDO]))]

W = [N'' I]

Z1 = [N'' John]

i' = [Coisa \* W]

4. Entrada da construção da estrutura conceitual composta (CLCS): [N'' I]

X = I

X' = [Coisa I]

4. Saída da estrutura conceitual composta: L = [Coisa I]

X' = [Evento CAUSAR ([Coisa I],

[Evento IR<sub>Poss</sub> ([Coisa FERIMENTO COM FACA], [Trajetória EM DIREÇÃO A<sub>Poss</sub>

([Posição EM<sub>Poss</sub> ([Coisa FERIMENTO COM FACA], [Coisa \*Z]))]],

[COM Instr ([Evento \*NÚCLEO\*], [Coisa U OBJETO\_PONTIAGUDO]))]

i' = [Coisa \* Z]

5. Entrada de construção da estrutura conceitual composta (CLCS): [N'' = John]

X = John

X' = [Coisa JOHN]

5. Saída da construção da estrutura conceitual composta: L = [Coisa JOHN]

$$X' = [\text{Evento CAUSAR} ([\text{Coisa I}], \\ \text{[Evento IRPoss} ([\text{Coisa FERIMENTO COM FACA}], \\ \text{[Trajetória EM DIREÇÃO A Poss} \\ \text{([Posição EM Poss} ([\text{Coisa FERIMENTO COM FACA}], [\text{Coisa JOHN}]])]), \\ \text{[COM Instr} ([\text{Evento *NÚCLEO*}], [\text{Coisa U OBJETO PONTIAGUDO}])])]$$

3. Saída da construção da estrutura conceitual composta:

$$L=X' = [\text{Evento CAUSAR} ([\text{Coisa I}], \\ \text{[Evento IRPoss} ([\text{Coisa FERIMENTO COM FACA}], \\ \text{[Trajetória EM DIREÇÃO A Poss,} \\ \text{([Posição EM Poss} ([\text{Coisa FERIMENTO COM FACA}], [\text{Coisa JOHN}])])])]$$

2. Saída da construção da estrutura conceitual composta:

$$L=X' = [\text{Evento CAUSAR} ([\text{Coisa I}], \\ \text{[Evento IRPoss} ([\text{Coisa FERIMENTO COM FACA}], \\ \text{[Trajetória EM DIREÇÃO A Poss,} \\ \text{([Posição EM Poss} ([\text{Coisa FERIMENTO COM FACA}], [\text{Coisa JOHN}])])])]$$

1. Saída da estrutura conceitual composta:

$$L=X' = [\text{Evento CAUSAR} ([\text{Coisa I}], \\ \text{[Evento IRPoss} ([\text{Coisa FERIMENTO COM FACA}], \\ \text{[Trajetória EM DIREÇÃO A Poss,} \\ \text{([Posição EM Poss} ([\text{Coisa FERIMENTO COM FACA}], [\text{Coisa JOHN}])])])]$$

**Figura 20. Ilustração do processo de construção da estrutura conceitual da sentença do inglês *I stabbed John* consiste de cinco entradas no procedimento de composição.**

Vimos no Quadro 28 que a notação \* é usada nas entradas lexicais do verbo *apunhalar*. Note-se que esse marcador faz duas distinções específicas: (i) a porção PARA<sub>POSS</sub> da entrada é marcada com (\*) para *dar* mas não para *to stab*; e (ii) o argumento FERIMENTO\_COM\_FACA não está marcado com a notação (\*) enquanto que a posição correspondente Y está marcada com (\*) para *dar*. Essas distinções paramétricas forçam a representação para a palavra *dar* ser combinada com as representações de outras palavras de tal modo que a representação composta seja equivalente à estrutura lexical especificada na entrada de *to stab*. Para o espanhol, *dar puñaladas*, a porção PARA<sub>POSS</sub> da entrada lexical de *to stab* unifica-se com a entrada lexical da palavra *a* e a posição Y é preenchida com a entrada da estrutura conceitual para *puñalada*.

A notação (\*) age com a **rotina generalizada de ligação (RGL)**<sup>33</sup>, um procedimento utilizado em conjunção com os parâmetros da estrutura conceitual que define a correspondência entre aquela estrutura e a estrutura sintática da sentença da língua de chegada. Ela consiste de quatro passos (cf. DORR, 1993, p. 139):

- a) relacionar o núcleo sintático X ao sujeito lógico
- b) relacionar a posição externa W à posição de sujeito lógico W'
- c) relacionar as posições internas Z1, ..., Zn à posição de argumento lógico Z'1, ..., Z'n
- d) relacionar as posições de adjuntos Q1, ..., Qn às posições de modificadores lógicos Q'1, ..., Q'n.

Essa rotina é usada em ambas as direções, i.e., composição da interlíngua e decomposição da interlíngua. Além dessa rotina, a interlíngua é relacionada à estrutura sintática por meio de uma função chamada **realização sintática canônica (RSC)** que, por exemplo, associa um tipo da estrutura conceitual (e.g. Coisa) com a categoria sintática típica (e.g. N'). Essa última rotina é responsável pela construção de representações usadas para geração da língua de chegada.

Os demais parâmetros do *Componente Léxico-Semântico*, além do parâmetro (\*), contém os parâmetros descritos no Quadro 29.

Parâmetros do componente léxico-semântico	Função	Restrições
*	Permite que um sujeito lógico, argumento ou modificador seja sintaticamente realizado em diferentes níveis	Associado a um argumento ou modificador nos nós irmãos ou não da estrutura conceitual
:INT/:EXT	Permite que a RGL <sup>34</sup> seja anulada ao se associar um sujeito lógico com um complemento sintático e um argumento lógico com um sujeito sintático	Associado a um argumento da estrutura conceitual no nível do nó irmão
:PROMOTE	Permite que a RGL seja anulada ao se associar um modificador lógico com um núcleo sintático	Associado a uma especificação de complemento na estrutura conceitual de um núcleo sintático
:DEMOTE	Permite que a RGL seja anulada ao se associar um núcleo lógico com um adjunto sintático	Associado a um argumento de uma estrutura conceitual de um adjunto sintático no nível do nó irmão
:CONFLATED	Permite incorporação de argumentos	Associado a um argumento ou modificador no nível do nó irmão
:CAT	Permite que uma RSC seja anulada	Associada a um argumento ou modificador ao nível do nó irmão

**Quadro 29. Os parâmetros do Componente Léxico-Semântico**<sup>35</sup>

<sup>33</sup> Na seção 4 essa rotina é denominada “Processo de especificação da estrutura conceitual canônica” (PEEC).

<sup>34</sup> Rotina generalizada de ligação.

<sup>35</sup> Esses parâmetros estão classificados em seis categorias que correspondem à tipologia de divergências léxico-semânticas de tradução que serão descritas na seção 4 deste trabalho.



### 3.3.4.7 A geração da estrutura sintática da sentença da língua de chegada

O procedimento de geração da estrutura sintática da língua de chegada utiliza-se de operações de seleção lexical e realização sintática recursivamente, começando pela entrada na unidade na **Estrutura Léxico-Conceitual Construída** (em inglês abreviado por CLCS), e gera uma estrutura seguindo os passos descritos abaixo (cf. DORR, 1993, 204).

Geração de sentenças (nó CLCS)

- 1- Selecionar um item lexical (o núcleo sintático) cuja RLCS corresponde ao nó CLCS sendo processado ao confrontá-lo com as entradas do dicionário da língua de chegada.
- 2- Construir sintaticamente uma sentença seguindo os seguintes passos:
  - 2a. Projetar uma projeção máxima ao se ativar a RSC na RLCS selecionada.
  - 2b. Gerar sintagmas satélites ao se ativar a geração de sentenças em cada nó CLCS que corresponde à posição do marcador (\*) na RLCS selecionada.
  - 2c. Anexar os núcleos sintáticos e sintagmas satélites sob a projeção máxima nas posições especificadas pela RGL.
- 3- Gerar a sentença na língua de chegada.

### 3.3.4.8 O Cálculo da Estrutura Sintática da Sentença de Chegada

A sentença de chegada *Eu dei punhaladas em João*, a partir da estrutura conceitual de *I stabbed John* é construída a partir do algoritmo da figura 21.

1. Entrada da sentença para geração  
 [Evento CAUSAR ([Coisa I],  
     [Evento IR Poss ([Coisa FERIMENTO COM FACA],  
         [Trajetória EM DIREÇÃO A Poss  
             ([Posição EM Poss ([Coisa FERIMENTO COM FACA], [Coisa JOHN]))]])]]  
 RLCS = [Evento CAUSAR ([Coisa \* W],  
     [Evento IR Poss ([Coisa \* Y],

[Trajetória \* EM DIREÇÃO A Poss

([Posição EM Poss ([Coisa Y], [Coisa Z])]) (dar)

Projeção Máxima = CSR (RLCS) -> [V''... dar...]

2. Entrada da sentença para geração: [Coisa I]

RLCS = [Coisa I] (Eu)

Projeção Máxima = RSC (RLCS) -> [N''... {Eu}...]

Sintagma projetado = [N'' {Eu}]

2. Saída gerada: [N'' {Eu}]

3. Entrada para geração: [Coisa FERIMENTO COM FACA]

RLCS = [Coisa FERIMENTO COM FACA] (punhalad)

Projeção Máxima = [N''...punhalad...]

Sintagma projetado = [N'' punhalad]

3. Saída gerada: [N'' punhalad]

4. Entrada para geração:

[Trajetória EM DIREÇÃO A Poss

([Posição EM Poss ([Coisa FERIMENTO À FACA], [Coisa JOHN])])]

RLCS = [Trajetória EM DIREÇÃO A Poss

([Posição EM Poss ([Coisa Y], [Coisa \*Z])]) (em)

Projeção Máxima = RSC (RLCS) -> [P'' ...em...]

5. Entrada para geração: [Coisa JOHN]

RLCS = [Coisa JOHN] (John)

Projeção Máxima = [N'' ...João<sub>i</sub>...]

Sintagma Projetado = [N'' João<sub>i</sub>]

5. Entrada para geração: [N'' João<sub>i</sub>]

Sintagma Projetado = [P'' em [N'' João<sub>i</sub>]]

4. Saída gerada: [C'' [FLEX'' [N'' {Eu}]]

[V'' dar [N'' punhalad]

[P'' em [N'' João<sub>i</sub>]]]

1. Saída gerada: [C'' [FLEX'' [N'' {Eu}]]

[V'' dar [N'' punhalad]

[P'' em [N'' João<sub>i</sub>]]]

Restrição de Caso:

*Eu dar punhalad em João.*

Síntese Morfológica

Eu dei punhaladas em João.

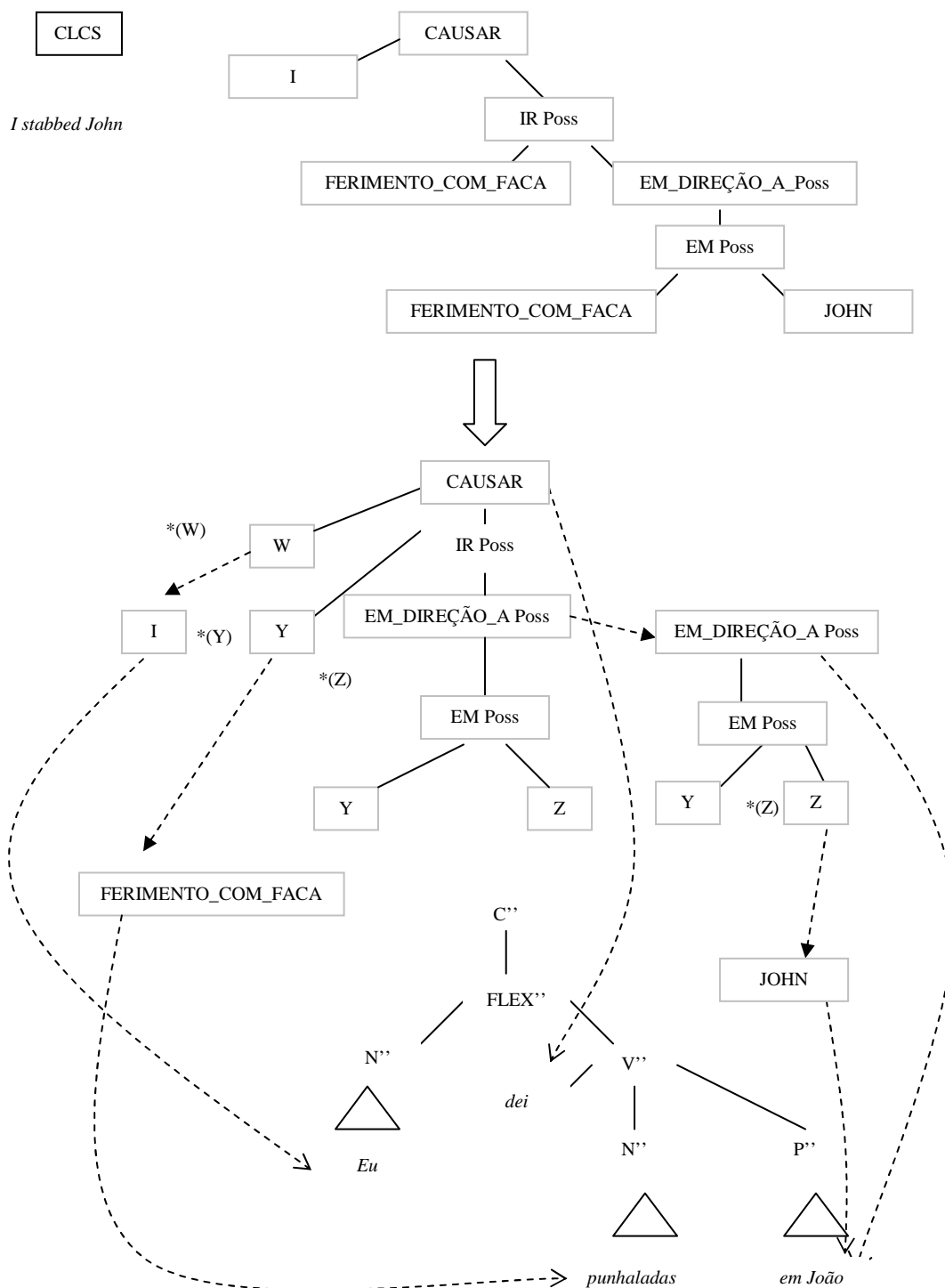
Eu dei uma punhalada em João.  
Eu dava punhaladas em João.  
Eu dava uma punhalada em João.

**Figura 21. Algoritmo de geração da estrutura sintática da sentença de chegada.<sup>36</sup>**

A Figura 22 ilustra o processo de realização de *Eu dei punhaladas em João*, a partir da estrutura conceitual de *I stabbed John* (cf. DORR, 1993, p. 227):

---

<sup>36</sup> O processo de geração da sentença *Eu dei punhaladas em João* consiste de seis entradas no procedimento de geração. As restrições de caso são então aplicadas e a sentença é sintetizada morfológicamente. Quatro possibilidades são então retornadas pelo sistema.



**Figura 22.** A realização sintática da sentença *Eu dei punhaladas em João*, que envolve a decomposição recursiva da CLCS para o verbo principal *dar* e seus três satélites *Eu*, *punhaladas*, *João*. (Cf. DORR, 1993, p. 227)

A especificação da estrutura semântica das sentenças do UNITRAN é feita pelo *Componente Léxico-Semântico*, que opera tanto na *COMPOSIÇÃO* da representação

conceitual da sentença da língua de partida, isto é, a representação do significado da sentença da língua de partida em termos da INTERLÍNGUA do UNITRAN (que é uma adaptação dos constructos da TEC), quanto na decomposição dessa representação em partes, a partir das quais o *Módulo de Síntese Gramatical* do *Componente Sintático* constrói a estrutura sintática da sentença da língua de chegada.

O *Componente Léxico-Semântico* do UNITRAN é construído segundo os constructos da TEC, isto é, é a TEC que fornece os constructos (*machinery*) e parâmetros que o UNITRAN implementa no LÉXICO do sistema para dar conta dos 6 tipos de divergência LÉXICO-SEMÂNTICA de tradução que se verificam entre as línguas naturais e que serão objeto de investigação na próxima seção. São eles: 1. divergência formal (*structural*), 2. divergência temática, 3. divergência de núcleo em L1 para não núcleo em L2 (*demotional*), 4. divergência de não núcleo em L1 para núcleo em L2 (*promotional*), 5. divergência de incorporação lexical (*conflational*), 6. divergência categorial.

## SEÇÃO 4 – INVESTIGAÇÃO DO DOMÍNIO REPRESENTACIONAL II: UM EXERCÍCIO DE ESPECIFICAÇÃO DAS DIVERGÊNCIAS LÉXICO-SEMÂNTICAS DE TRADUÇÃO PARA O PORTUGUÊS

### 4.1 Preliminares

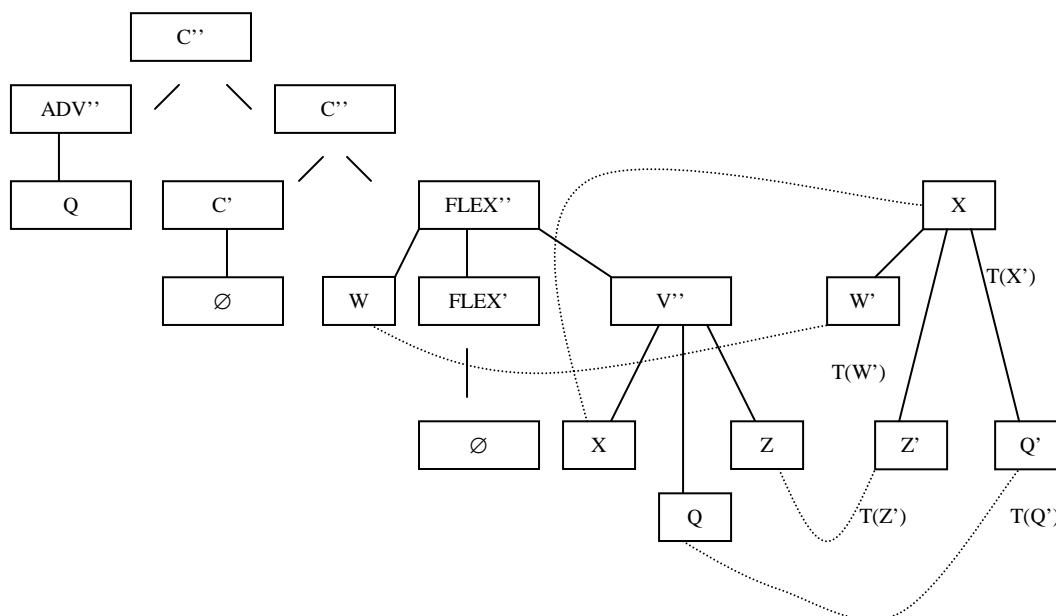
A principal inovação de Dorr (1993) no que se refere à implementação de um sistema de TA do tipo interlingual é a inclusão, no léxico de cada língua do sistema, de parâmetros que são manipulados durante o processo de construção da estrutura léxico-conceitual de sentenças da língua de partida (cf. subseção 3.3.4.6) e o processo de construção, a partir dessa estrutura léxico-conceitual, da estrutura sintática da sentença da língua de chegada (cf. subseções 3.3.4.5 e 3.3.4.7).

Esses processos, por sua vez, são realizados por meio de dois outros processos distintos, porém complementares, conforme esboçados nas subseções mencionadas no parágrafo anterior e abaixo sumarizados:

#### **PEEC – Processo de especificação da estrutura conceitual canônica**

- (a) Associar o núcleo sintático X da sentença de entrada à posição de núcleo de estrutura conceitual X’;
- (b) Associar o sujeito sintático (o argumento externo) W da sentença de entrada à posição de sujeito lógico W’ da estrutura conceitual;
- (c) Associar os objetos sintáticos (argumentos internos) Z<sub>1</sub>, ..., Z<sub>n</sub> da sentença de entrada aos argumentos lógicos da estrutura conceitual Z’<sub>1</sub>, ..., Z’<sub>n</sub>.
- (d) Associar os adjuntos sintáticos (não argumentos) Q<sub>1</sub>, ..., Q<sub>n</sub> da sentença de entrada aos adjuntos lógicos Q’<sub>1</sub>, ..., Q’<sub>n</sub> da estrutura conceitual.

Graficamente, temos a seguinte correlação:



Em que:  $T(X')$ ,  $T(W')$ ,  $T(Z')$  e  $T(Q')$  são os tipos semânticos dos componentes da estrutura conceitual.

### **PEES: Processo de Especificação da Estrutura Sintática Canônica**

Para cada tipo semântico de estrutura conceitual, gerar a categoria sintática correspondente, segundo as seguintes correlações:

EVENTO  $\Leftrightarrow V$

ESTADO  $\Leftrightarrow V$

COISA  $\Leftrightarrow N$

PROPRIEDADE  $\Leftrightarrow A$

TRAJETÓRIA  $\Leftrightarrow P$

POSIÇÃO  $\Leftrightarrow P$

LOCALIZAÇÃO  $\Leftrightarrow ADV$

TEMPO  $\Leftrightarrow ADV$

MODO  $\Leftrightarrow ADV$

INTENSIFICADOR  $\Leftrightarrow ADV$

PROPÓSITO  $\Leftrightarrow ADV$

Para facilitar a compreensão da discussão sobre as divergências de tradução tratadas pelo UNITRAN e, com ela, sinalizar sua relevância para o funcionamento do

sistema, passamos, a seguir, a resumir os elementos essenciais dos principais constructos, que em maior ou menor grau, foram abordados em seções anteriores.

- **Constructo Estrutura Léxico-Conceitual Composta**

[ T(x') X'  
     [T (w') W']  
         [T (z') Z']  
             [T (q') Q']

em que X' é o núcleo lógico, W' é o sujeito lógico, Z' é o argumento lógico, Q' o modificador lógico e T (∅) é o tipo lógico que corresponde ao primitivo ∅.

Exemplo: sentença *João foi prazerosamente para a escola.*

[Evento IR Loc  
     ([Coisa JOÃO],  
         [Trajectoria PARA Loc ([Posição EM Loc ([Coisa JOÃO],  
     [Localização ESCOLA])])] ]  
     [Modo PRAZEIROSAMENTE]

Trata-se, aqui, da estrutura conceitual composta no nível da interlíngua. No léxico do sistema, registram-se as estruturas conceituais raízes.

- **Constructo Estrutura Léxico-Conceitual Raiz**

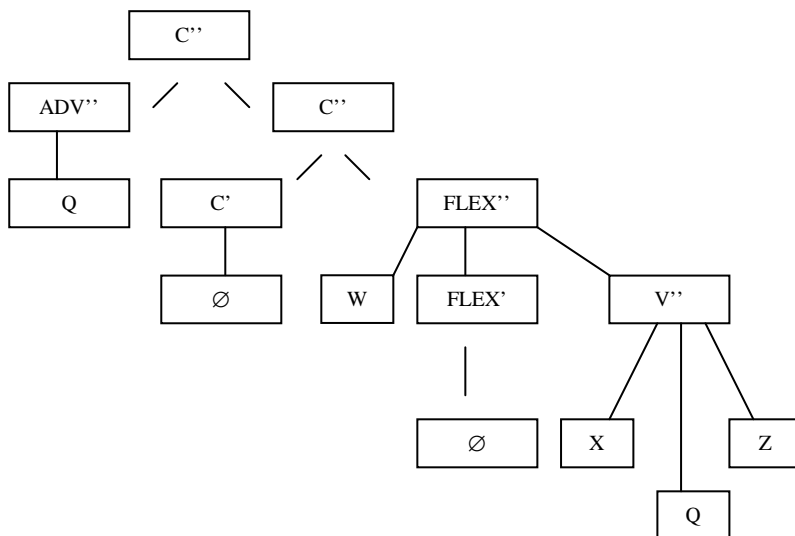
No léxico do sistema, registram-se as estruturas conceituais dos itens lexicais com as posições argumentais lógicas vazias.

Exemplo: verbo *ir*

[Evento IR Loc ([Coisa X], [Trajetória PARA Loc ([Posição EM Loc ([Coisa X], [Localização Z])])] ]

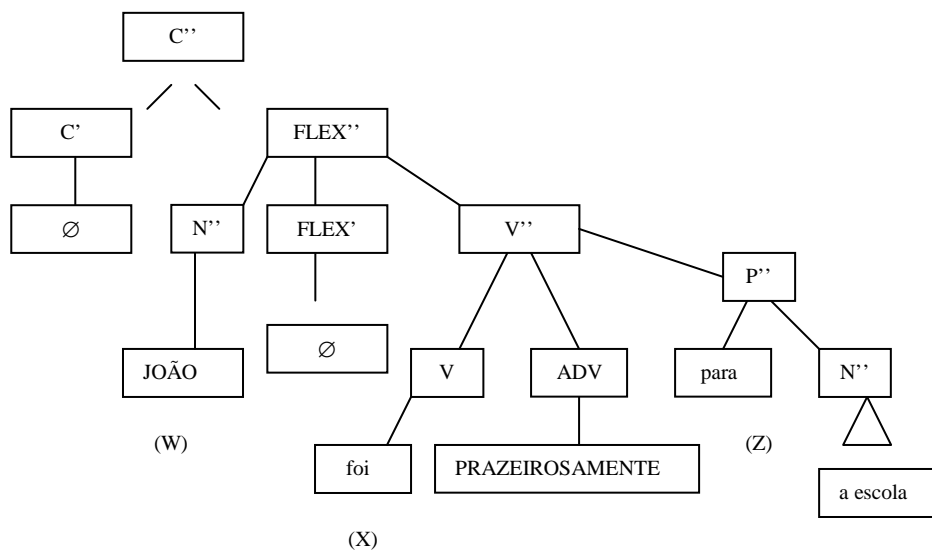
- **Constructo Estrutura Sintática**





em que X é o núcleo sintático (V, N, P, FLEX e C), W é o argumento sintático externo, Z o argumento sintático interno e Q o adjunto sintático.

Exemplo:



Feita essa sumarização de natureza didática, passamos a formalizar a noção de divergência de tradução conforme implementada no UNITRAN e ilustrá-la com

exemplos do português, em um exercício de exploração de possibilidade de inclusão dessa língua no sistema.

Intuitivamente, há uma divergência de tradução quando se verifica algum tipo de exceção nas correspondências previstas por um, por outro, ou por ambos os processos referidos anteriormente pelo PEEC (Processo de especificação da estrutura conceitual canônica). Para que se configure a divergência, a não correspondência deve ocorrer com a análise/síntese de apenas uma das línguas do par de línguas envolvido na tradução.

Entretanto, no sistema UNITRAN, não se trata de se formalizar e resolver divergências de tradução entre pares de sentenças de línguas naturais, isto é, de se formalizar e resolver as divergências de tradução que resultam da comparação entre sentenças das duas línguas que estão sob o processo de tradução.

A noção de divergência implementada no sistema não é, assim, o que se poderia caracterizar de uma divergência entre duas línguas, mas entre a língua e a interlíngua do sistema. Explicando melhor: trata-se de divergências de tradução que se verificam durante o processamento das sentenças de entrada e de saída do sistema. Em outras palavras: durante a “extração” da estrutura conceitual da sentença da língua de partida, realizada pelo PEEC e/ou durante a “construção” da estrutura sintática da sentença da língua de chegada, realizada pelo PEES.

Assim, nesta seção, discutem-se as divergências de tradução que se verificam nos mapeamentos internos ao UNITRAN.

Para tornar essa noção de divergência mais transparente para o leitor, observe-se o seguinte exemplo: a sentença do inglês *I like Mary* é convertida em interlíngua sem que qualquer mecanismo de resolução de divergências seja aplicado. Já a sentença do espanhol *Me gusta Maria* exige que se aplique um mecanismo de resolução de divergências específico para que seja sintaticamente bem formada. Assim, a resolução da divergência é alcançada com informação específica contida no léxico do espanhol e no léxico do inglês. Ou ainda, a estrutura interna das sentenças *Hans Kußt Marie gern* e *Hans kust Marie graag*<sup>37</sup>, do alemão e do holandês, respectivamente, não divergem entre si, mas divergem para fins de interlíngua do UNITRAN, porque, para o sistema, a estrutura sintática canônica correspondente à estrutura conceitual dessas sentenças exige que o núcleo conceitual seja colocado em uma posição sintaticamente periférica: no

---

<sup>37</sup> *Hans gosta de beijar Maria.*

português, o verbo *gostar* toma o evento *beijar* como complemento; já no alemão e no holandês, *gostar* é colocado em uma posição sintática inferior, isto é, a posição de adjunto (*gern* e *graag*, respectivamente), e o evento *beijar* realiza-se como o verbo principal.

## 4.2 Classificação das divergências de tradução léxico-semânticas

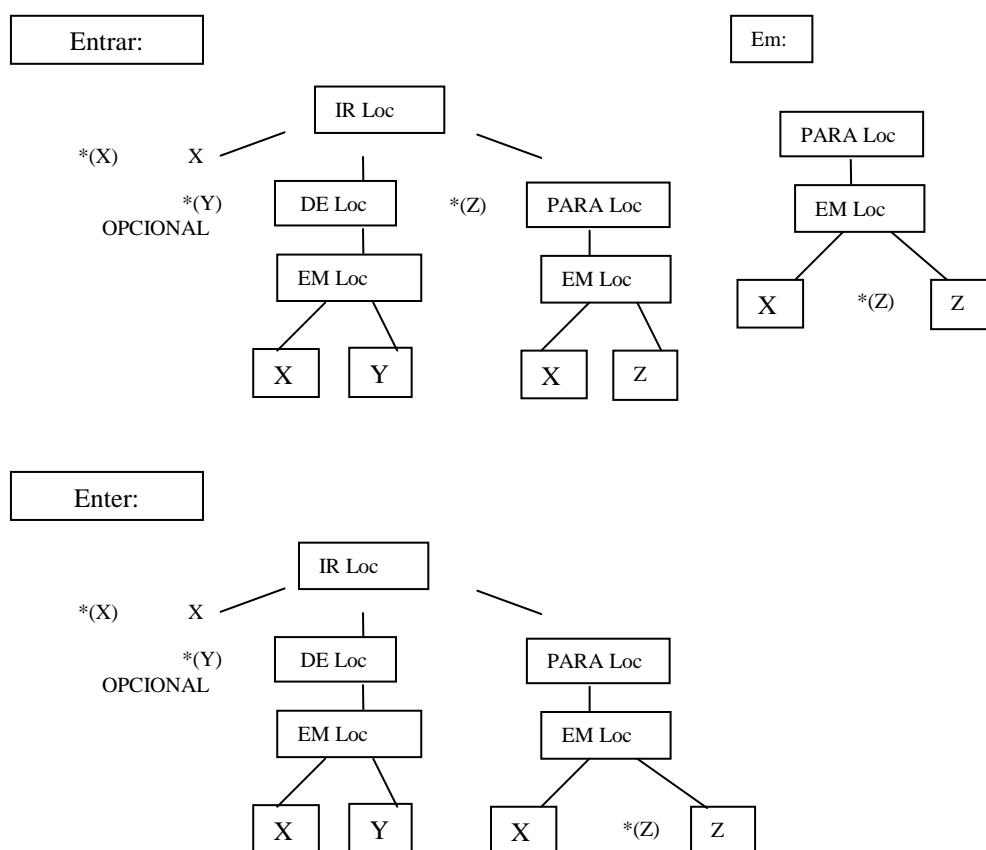
O quadro 30 apresenta uma classificação das divergências de tradução léxico-semânticas na concepção de Dorr (1993), seguida de uma descrição da tipologia em que também serão apresentadas as propostas de formalização.

**Quadro 30. Divergências léxico-semânticas tratadas pelo UNITRAN e os parâmetros que as resolvem.**

Tipo de divergência	Exemplo de tradução	Parâmetro LCS	Princípio da TEC
Formal ( <i>Structural</i> )	P – João entrou na casa. I – John entered the house.	*	Regra de correspondência
Temática ( <i>Thematic</i> )	P – Eu gosto de Maria. I – Mary pleases me.	:INT/:EXT	Regra de correspondência
De núcleo em L1 para não núcleo em L2 ( <i>Demotional</i> )	I – I like eating. A – Ich esse gern. =eu-comer-bom “Eu gosto de comer”.	:DEMOTE	Regra de correspondência
De não núcleo em L1 para núcleo em L2 ( <i>Promotional</i> )	P – Ele acabou de chegar. I – He has just arrived.	:PROMOTE	Regra de correspondência
De incorporação lexical ( <i>Conflational</i> )	I – I stabbed John. P – Eu dei punhaladas em João.	:CONFLATED	Regra de correspondência
Categorial ( <i>Categorial</i> )	P – Eu tenho fome. I – I am hungry.	:CAT	RSC

#### 4.2.1 Divergência Formal (parâmetro \*)

Esse parâmetro sinaliza as porções da estrutura conceitual que são sintaticamente realizadas. O par de frases, *John entered the house* e *João entrou na casa*, do inglês e do português respectivamente, ilustram a aplicação deste parâmetro. Trata-se de uma divergência estrutural em que o componente locativo da estrutura conceitual composta, no inglês, realiza-se na forma verbal *enter* e, no português, através da preposição *em*.



**Figura 23.** As entradas para os verbos *entrar* / *enter* ilustram o uso do parâmetro \*. Observe-se que o parâmetro OPTIONAL sinaliza que o componente de origem é parte do significado dos verbos *entrar* e *enter*.

Dadas as entradas do português (*entrar, em*) e do inglês (*enter*) a figura 23 ilustra a aplicação do parâmetro \* entre pares de sentenças do português e do inglês.

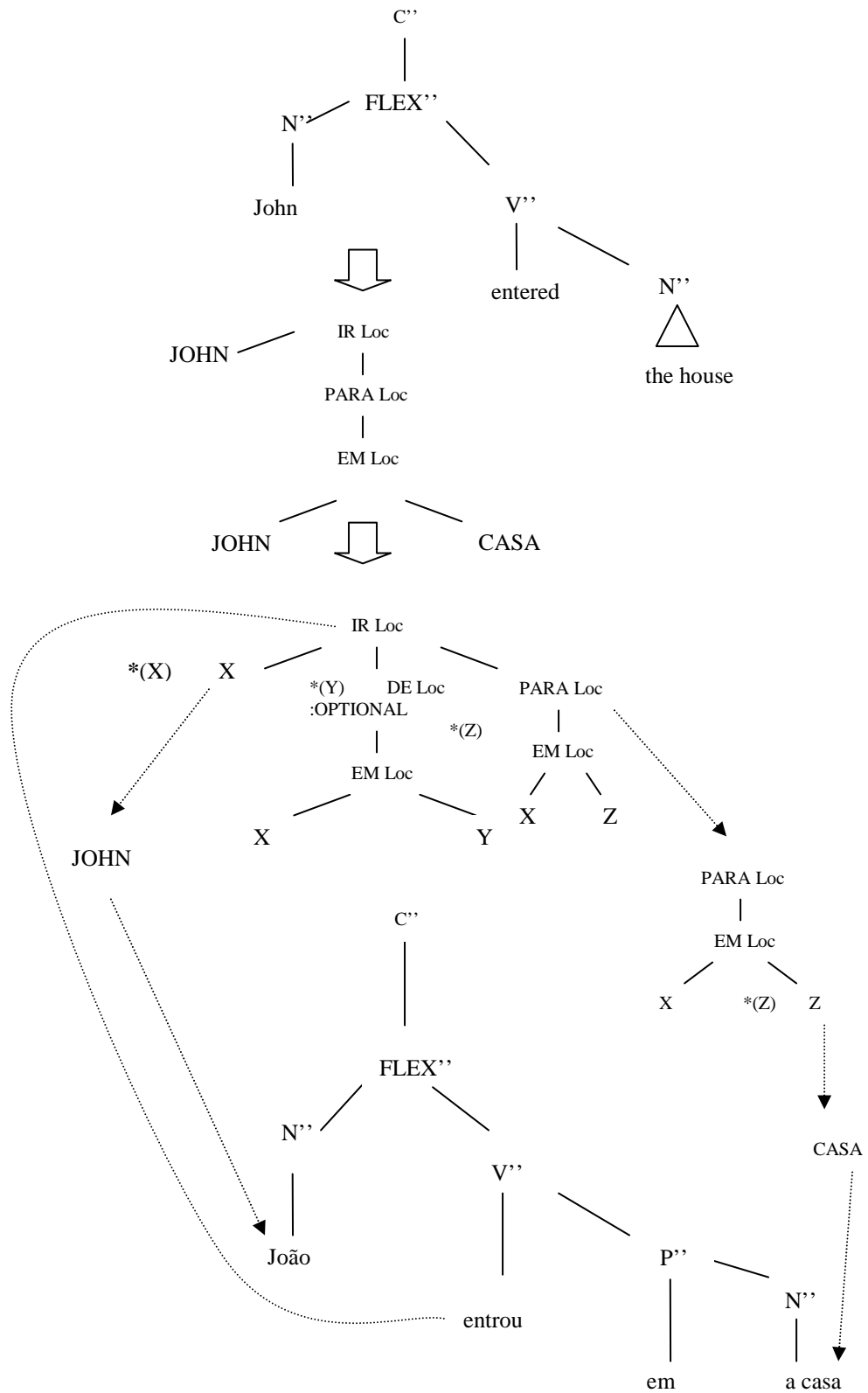


Figura 24 – O parâmetro \* utilizado no processo de geração de *João entrou na casa*.

Destacam-se os seguintes passos do processo de realização sintática da sentença em português: uma vez que a forma *entrar* é escolhida como o verbo principal da sentença, os argumentos de *entrar* são necessariamente e recursivamente realizados. O parâmetro estrela (\*) é usado para determinar o nível em que cada argumento será realizado. É importante notar que, diferentemente do verbo *to enter*, o verbo *entrar* exige a realização de *em*. Como mostra a figura 23, a palavra *em* é definida para expressar *em + a (na)*. Em seguida o segundo objeto de *entrar* é sintaticamente expresso. Esse argumento é expresso no nível da estrutura conceitual EM DIREÇÃO A Poss denotado pelo parâmetro \*(Z). Como a entrada lexical da palavra *em* corresponde à porção EM DIREÇÃO A Poss da estrutura conceitual composta, essa palavra é escolhida para ser sintaticamente realizada.

#### 4.2.2 Divergência Temática (parâmetros :INT / :EXT)

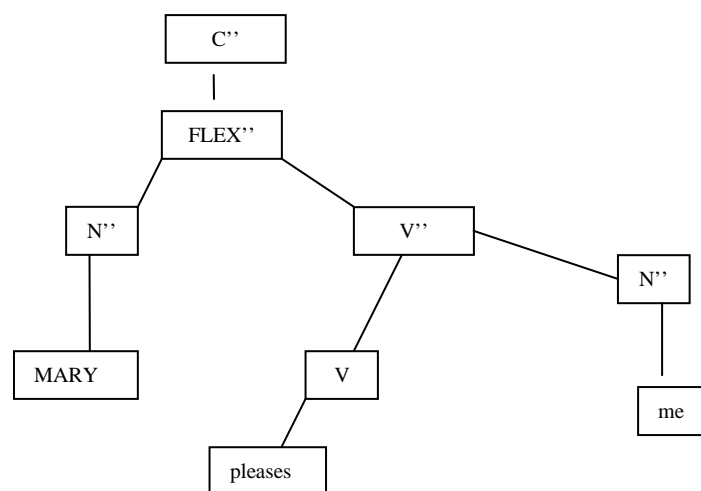
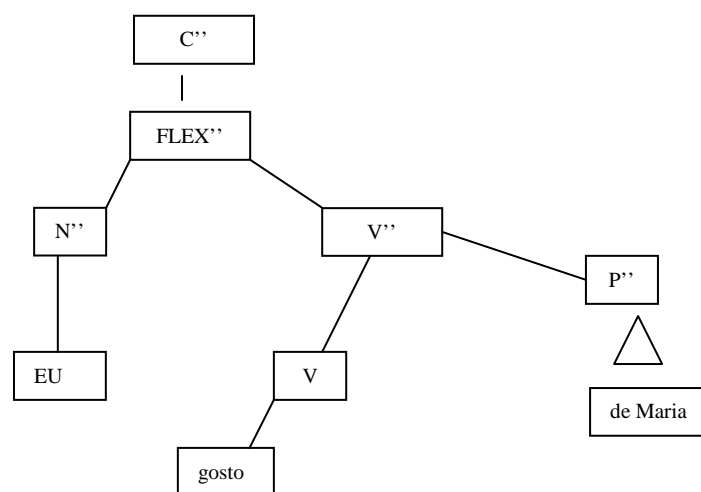
A divergência temática ocorre quando o PEEC ativa os seguintes conjuntos de relações nos passos (b) e (c) (cf. página 101).

Esse tipo de divergência é um dos três tipos que envolvem a colocação de constituintes. Os dois outros tipos, que serão discutidos na sequência da discussão da divergência temática, são o deslocamento de não núcleo em L1 para núcleo em L2 e vice-versa, o deslocamento de núcleo em L1 para não núcleo em L2. Se a divergência temática envolve o reposicionamento de elementos satélites relativos a um núcleo X, as outras duas divergências envolvem o reposicionamento de um núcleo X relativo aos seus satélites. Assim, esses três tipos de divergência dão conta dos reposicionamentos de constituintes no UNITRAN.

A divergência temática ocorre somente em situações em que há um sujeito lógico. Exemplifica esse tipo a reversão do preenchimento sintático das funções de sujeito e o objeto em pares de sentenças como:

*Eu gosto de Maria e*

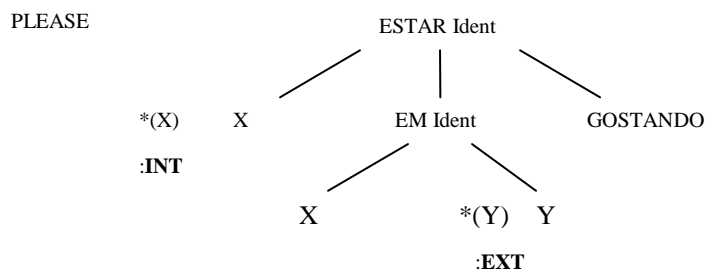
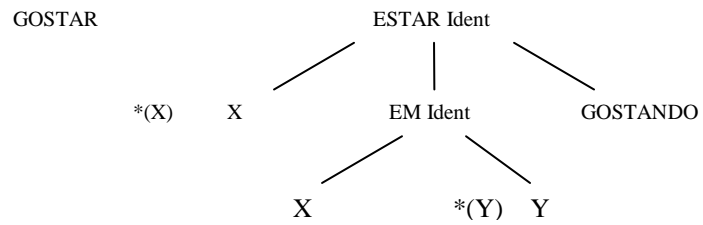
*Mary pleases me*, cujas estruturas sintáticas são dadas, respectivamente, por:

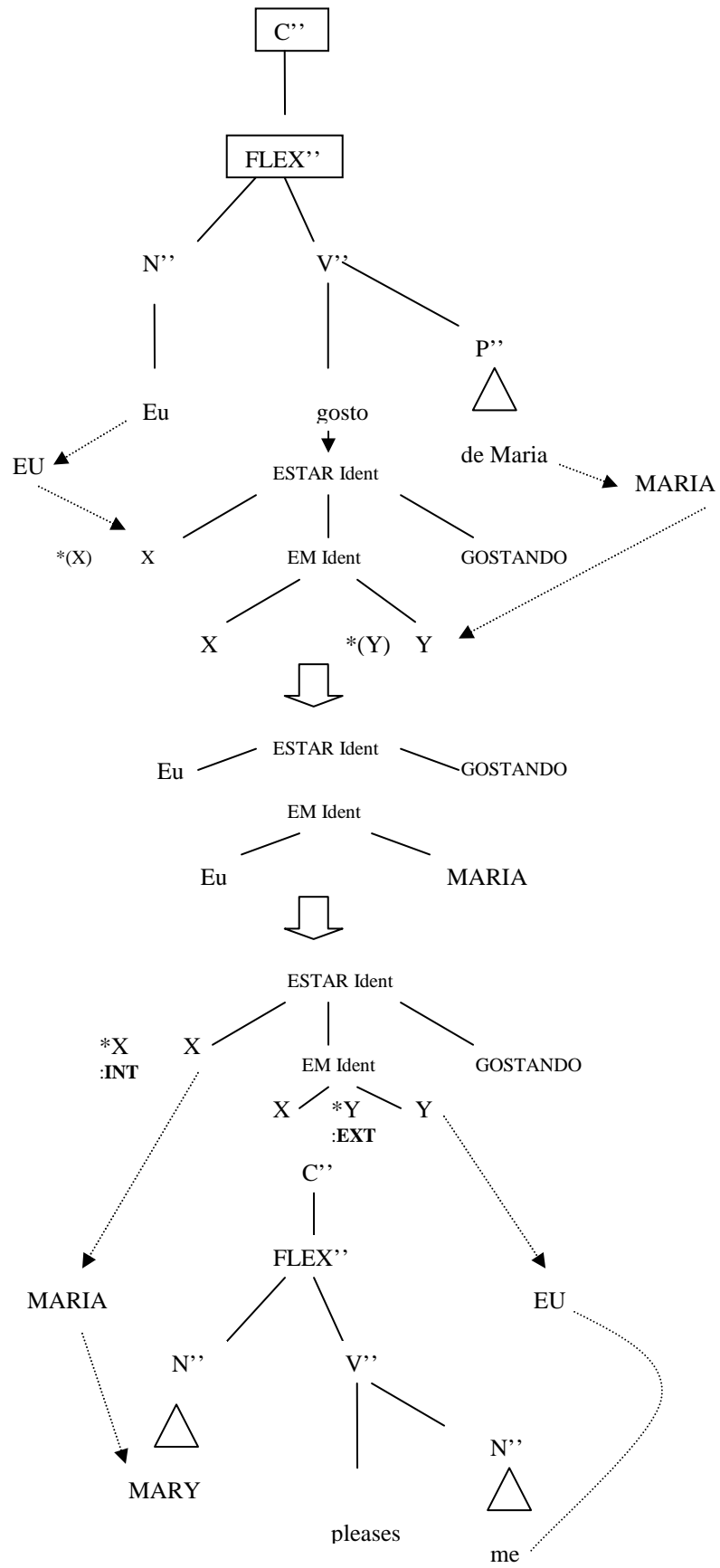


Nessas estruturas, o sujeito "eu" e o complemento "de Maria" da sentença do português passam a ser o objeto "me" e o sujeito "Mary" da sentença correspondente do



inglês, respectivamente. Essa divergência é marcada no léxico do inglês, e não no português, conforme as representações a seguir:





#### 4.2.3 Divergência de promoção de não núcleo de L1 para núcleo de L2:

(parâmetro :PROMOTE)

Esse tipo de divergência caracteriza-se pela “promoção” (isto é, colocação em uma posição sintática superior”) de um modificador lógico, como no exemplo em espanhol *Juan suele ir a casa* (*João geralmente vai para casa*), em que o modificador HABITUALMENTE é realizado pelo verbo *soler* (*tender*). A figura 24 ilustra a especificação do parâmetro para essa divergência.

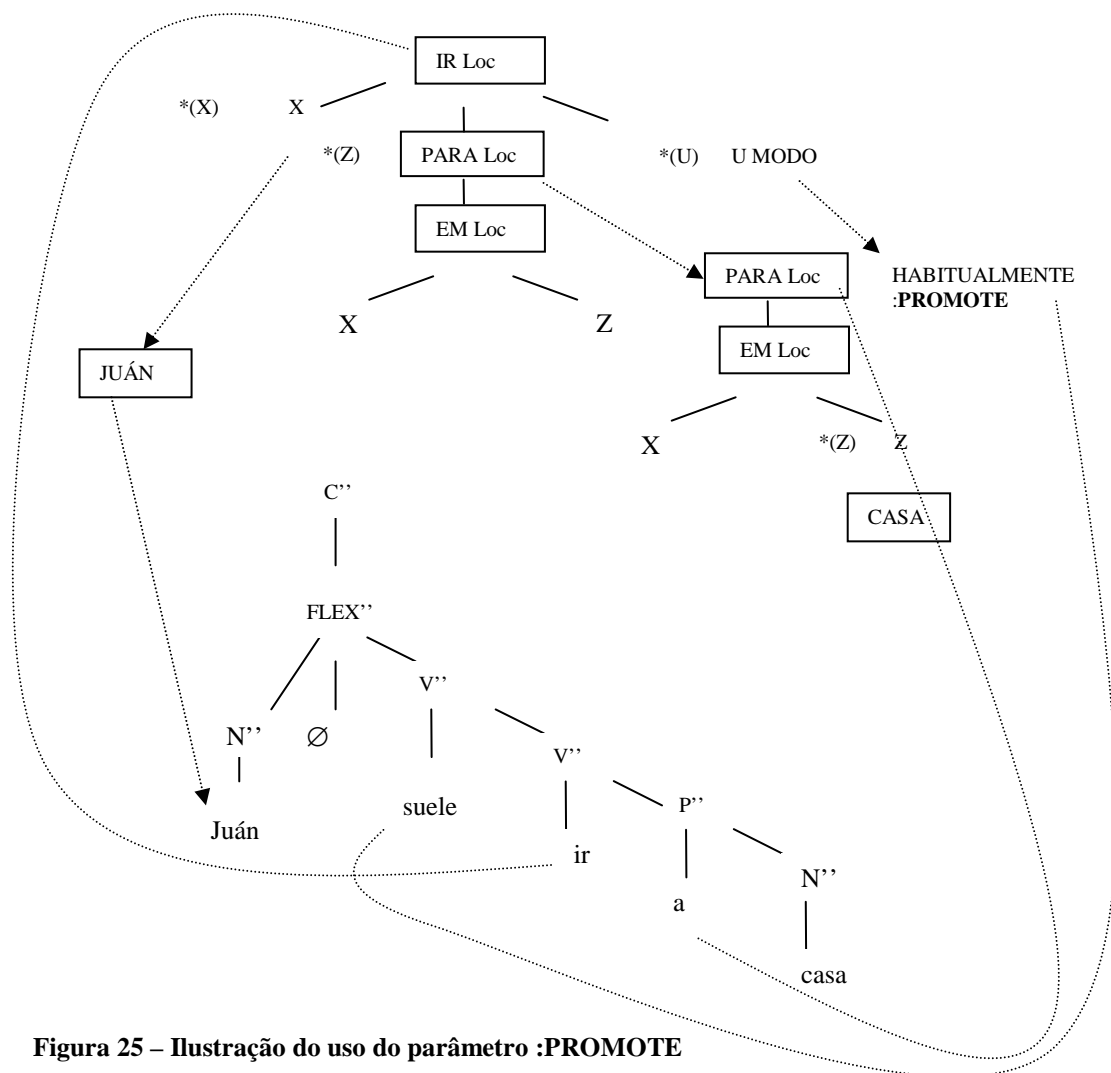


Figura 25 – Ilustração do uso do parâmetro **:PROMOTE**

A estrutura conceitual da entrada lexical do verbo *soler* do espanhol contém o parâmetro **:PROMOTE**, que faz com que o sistema dê prioridade para a realização sintática desse constituinte, que realiza-se como um núcleo verbal. A sentença correspondente do português *João geralmente vai para casa* é gerada pelo sistema com o componente de modo colocado na posição canônica de adjunto dado que a estrutura conceitual da entrada lexical de *geralmente* não é marcada com esse parâmetro.

#### 4.2.4 Divergência de rebaixamento de núcleo de L1 para não núcleo de L2 (parâmetro :DEMOTE)

Este tipo de divergência caracteriza-se pelo rebaixamento (isto é, pela colocação em uma posição sintática inferior) de um núcleo lógico como no exemplo do alemão *Ich esse gern* (*Eu gosto de comer*) em que, o núcleo lógico ESTAR Circ. é sintaticamente realizado por um advérbio (*gern*).

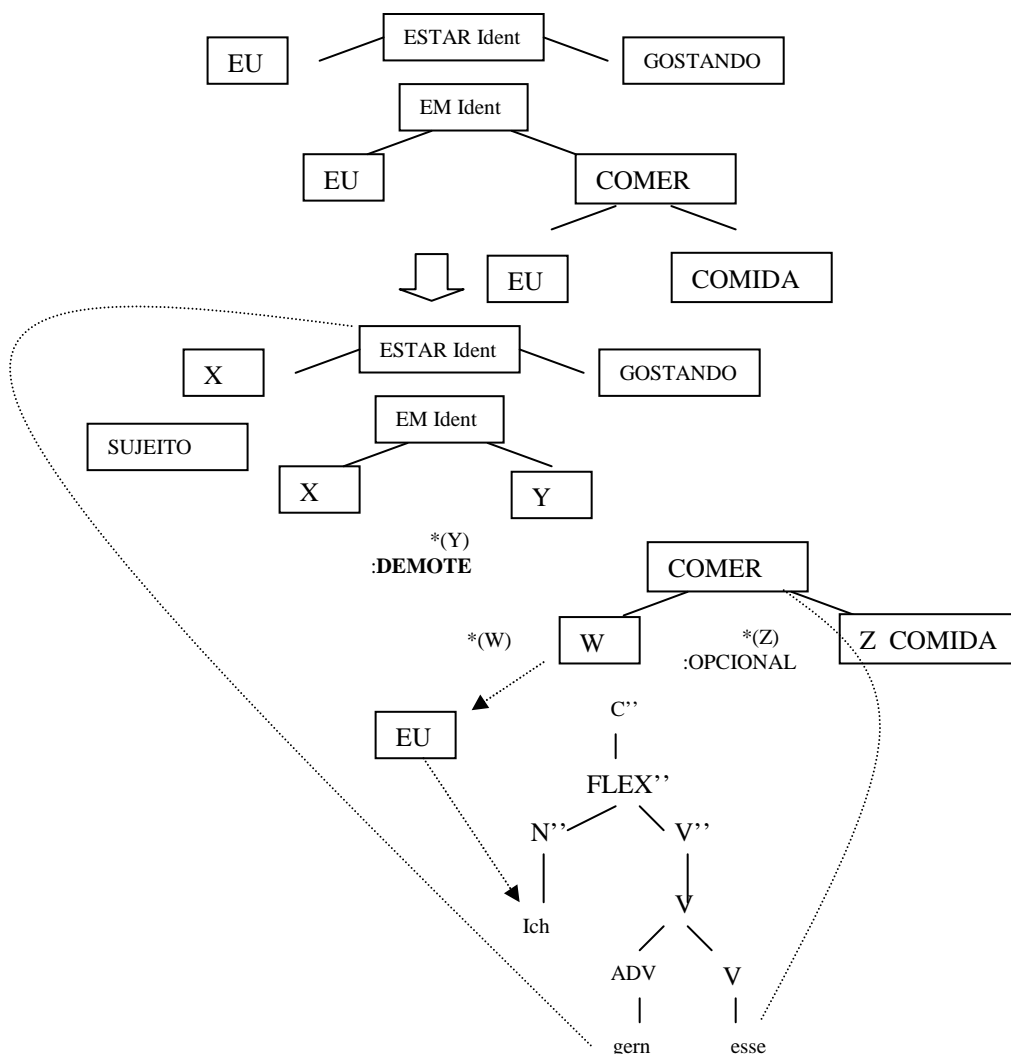


Figura 26. Exemplo de uso do parâmetro **:DEMOTE**.

Como o componente conceitual ESTAR circ. da estrutura conceitual composta corresponde à estrutura conceitual raiz de *gern*, essa palavra é escolhida como parte da realização sintática da sentença. Entretanto, o argumento de evento de *gern* está

associado ao marcador :DEMOTE; assim, o sistema determina que o componente conceitual COMER seja realizado em primeiro lugar. Dado que o componente conceitual raiz da palavra *essen* corresponde ao evento COMER, essa palavra é escolhida e projetada na estrutura sintática como o verbo principal.

#### 4.2.5 Divergência de incorporação lexical (parâmetro :CONFLATED)

O parâmetro :CONFLATED é usado para indicar que um determinado argumento não precisa ser realizado na estrutura sintática. É esse parâmetro que dá conta da divergência de frases como *Eu dei punhaladas em João* e *I stabbed John*. Na sentença em inglês o argumento FERIMENTO\_COM\_FACA é incorporado no verbo. A construção com o verbo *stab* não realiza esse argumento. Diferentemente do inglês, no português a construção *dar punhaladas em* realiza sintaticamente esse argumento por meio da palavra *punhalada*.

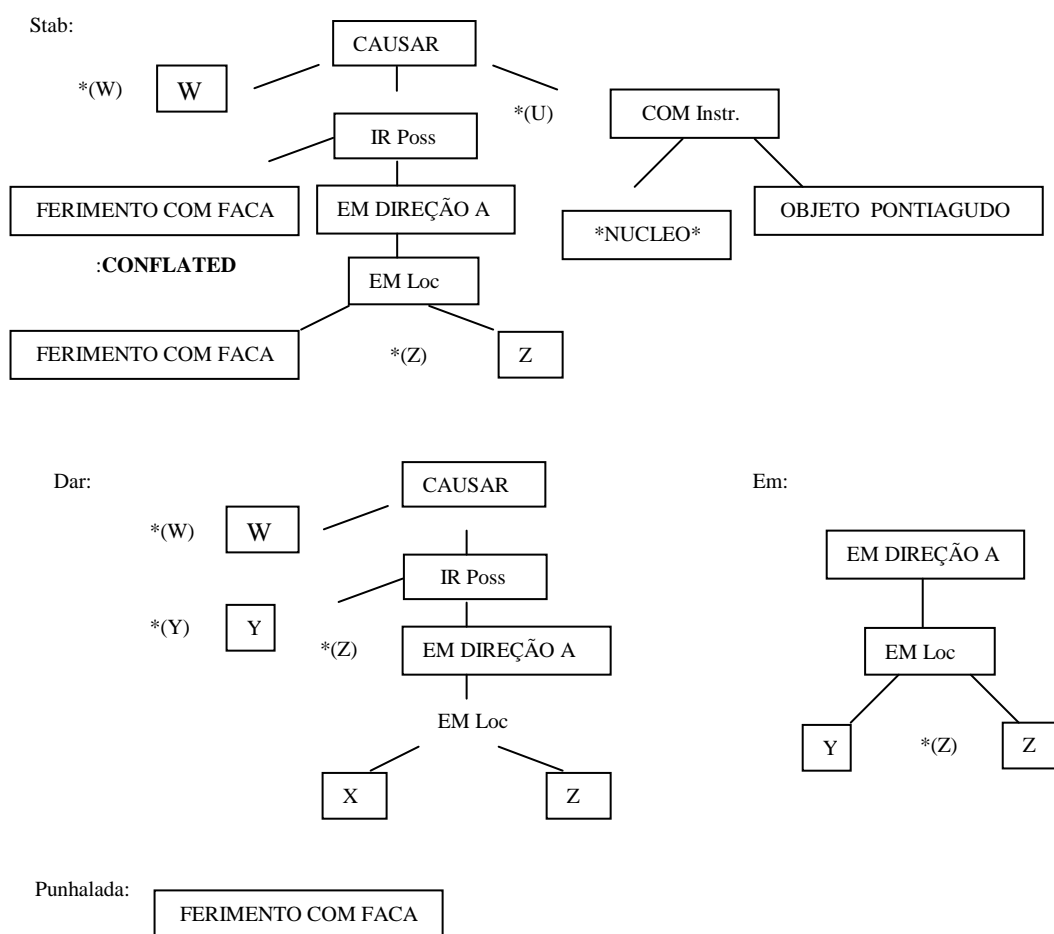


Figura 27. Exemplo de uso do parâmetro :CONFLATED.

Dadas as entradas das sentenças do inglês (*stab*) e do português (*dar, punhalada, em*) a figura 27 ilustra a aplicação do parâmetro :CONFLATED.

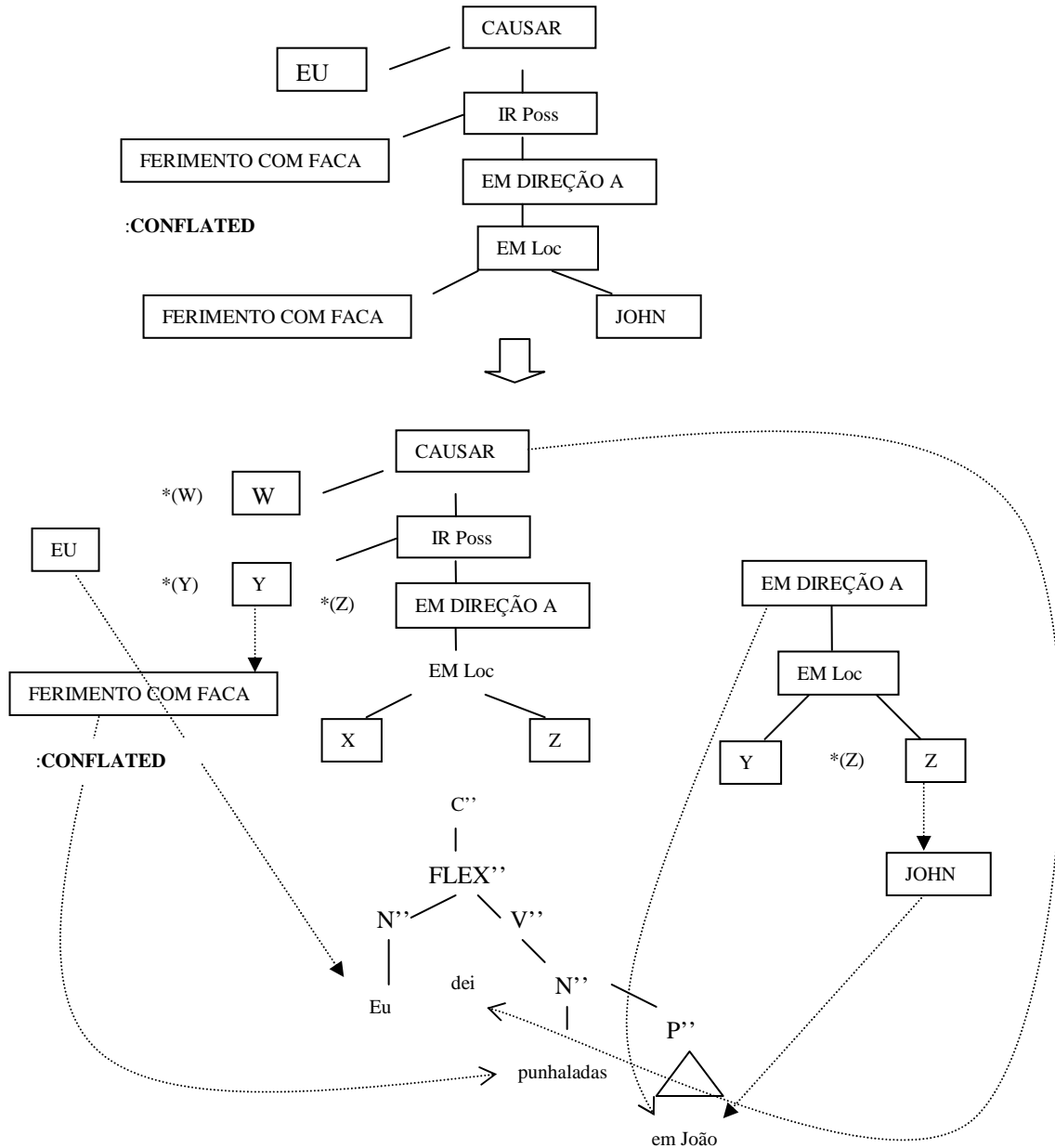


Figura 28 - Ilustração da aplicação do parâmetro :CONFLATED.

Destacam-se os seguintes passos do processo de realização sintática da sentença em português: uma vez que a forma *dar* é escolhida como o verbo principal da sentença, os argumentos de *dar* são necessariamente e recursivamente realizados. O parâmetro

estrela (\*) é usado para determinar o nível em que cada argumento será realizado. É importante notar que, diferentemente do verbo *to stab*, o verbo *dar* exige que o argumento Y seja expresso, dado que ele não contém o parâmetro :CONFLATED. Dado que a estrutura conceitual composta contém o argumento FERIMENTO\_COM\_FACA esse argumento será realizado como valor semântico da variável Y da estrutura conceitual da entrada lexical correspondente do português. Como mostra a figura 28, a palavra *punhalada* é definida para expressar o componente FERIMENTO\_COM\_FACA, e, portanto, essa é a palavra escolhida para ser o objeto direto do verbo principal. Em seguida o segundo objeto de *dar* é sintaticamente expresso. Esse argumento é expresso no nível da estrutura conceitual EM DIREÇÃO A Poss denotado pelo parâmetro \*(Z). Como a entrada lexical da palavra *em* corresponde à porção EM DIREÇÃO A Poss da estrutura conceitual composta, essa palavra é escolhida para ser sintaticamente realizada.

#### 4.2.6 Divergência Categorial (parâmetro :CAT)

O parâmetro :CAT fornece a categoria sintática para um componente conceitual da estrutura conceitual raiz das entradas lexicais. É importante notar que o processo PEEC canônica faz a correspondência entre um tipo semântico da estrutura conceitual e uma categoria sintática da estrutura da sentença. Quando essa correspondência é violada por alguma propriedade lexical, o marcador :CAT sinaliza o fato. As frases *I am hungry* do inglês e *Eu tenho fome* do português exemplificam o emprego do parâmetro, conforme ilustra a figura 29.



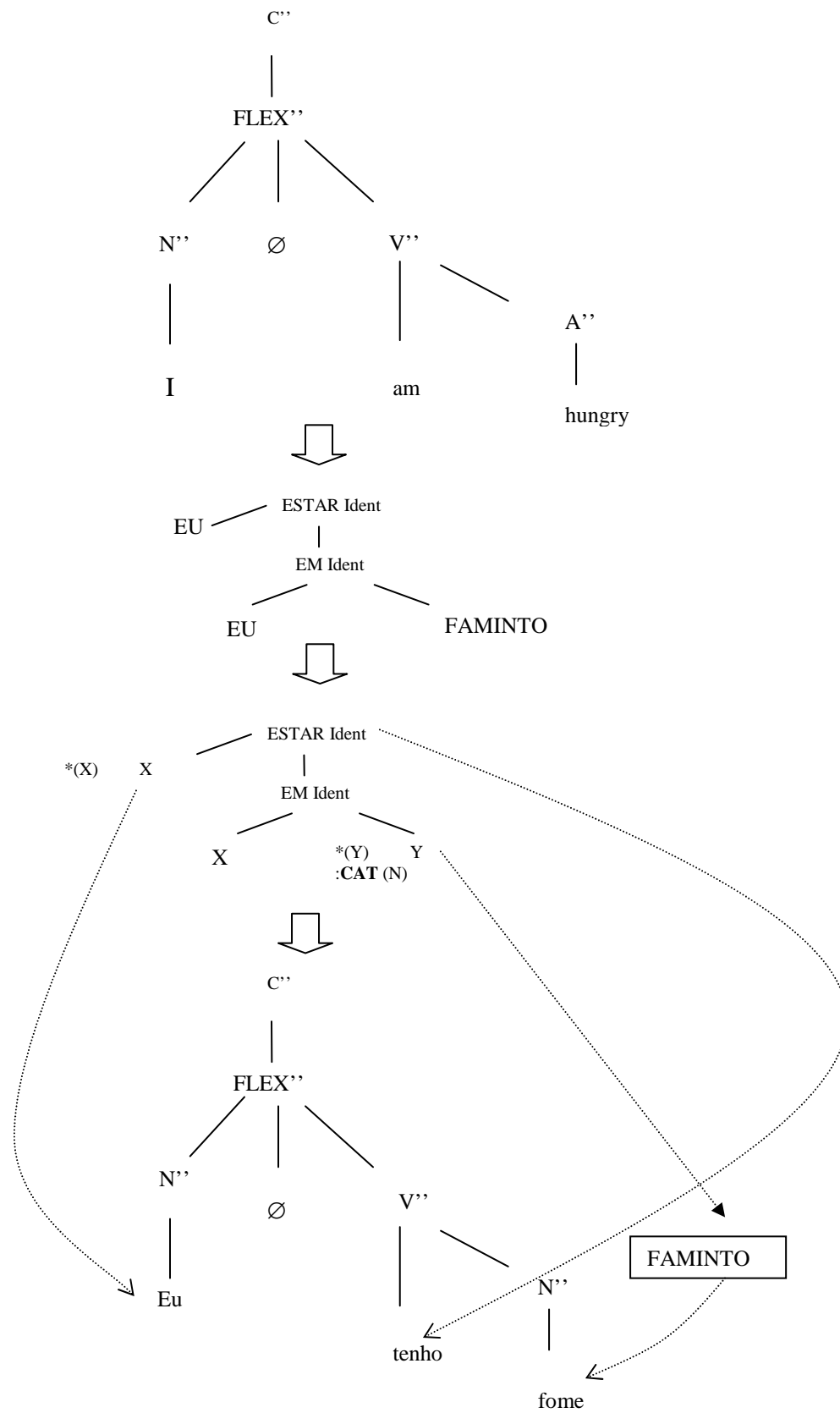


Figura 29 – Ilustração da aplicação do parâmetro `:CAT`.

Cumprе notar que nessa divergência ocorre o que se poderia denominar “divergência lexical”, já que houve uma escolha lexical distinta nas duas línguas: numa se escolhe BE, em outra se escolhe o TER.

Esta seção apresentou uma classificação sistemática dos tipos de divergência léxico-semântica de tradução e mostrou como essa classificação é formalmente definida e sistematicamente empregada para resolver as discrepâncias de realização estrutural entre as sentenças das línguas e a representação de interlíngua interna ao sistema. Dado que a especificação desses parâmetros nos léxicos das línguas individuais dá conta da variedade de realização sintática das unidades lexicais quando se compõem para integrar estruturas oracionais, a especificação de regras de transferência torna-se absolutamente dispensável, conferindo assim maior simplicidade e modularidade ao sistema. Essas duas propriedades são essenciais para a inclusão de novas línguas.

Por fim, cumprе ressaltar que a parametrização dessas divergências léxico-semânticas juntamente com a parametrização das divergências sintáticas descritas na seção 3 constituem a principal inovação do sistema de TA proposto por Dorr (1993) ao conferir aos sistemas de TA a necessária motivação lingüística que torna pesquisas nesse domínio do PLN relevantes quer para os Estudos Lingüísticos quer para os Estudos da Tradução.

## SEÇÃO 5 - CONCLUSÕES

Como pudemos verificar ao longo da tese, a abordagem utilizada na construção do sistema UNITRAN possui dois grandes méritos: (i) o fato de compatibilizar duas teorias lingüísticas aparentemente díspares como embasamento teórico (a *Teoria dos Princípios e Parâmetros* e a *Teoria das Estruturas Conceituais*), no sentido de tornar o sistema lingüisticamente motivado e (ii) o fato de seu *design* permitir a inclusão de outras línguas, como fizemos com o português, com uma adaptação de parâmetros específicos de nossa língua.

A motivação dessa tese é a mesma que deu origem ao próprio UNITRAN: o estudo de divergências de tradução. Além de serem um problema para a análise/geração de sentenças bem formadas das línguas, constituem também um desafio para os próprios estudos lingüísticos.

A solução para o problema das divergências de tradução apontada pela proposta para o UNITRAN possibilitou investigações em diferentes domínios do conhecimento. O estudo das teorias de *Princípios e Parâmetros* e das *Estruturas Conceituais*, possibilitou o reconhecimento dessas teorias como fundamento para sistemas de TA. O exame dessas teorias levou à considerações de caráter representacional e ao levantamento de aspectos da língua portuguesa relevantes para proposição de uma representação para alguns tipos importantes de divergências de tradução.

Ao observarmos a tese de uma perspectiva maior, percebemos sua inserção na proposta de metodologia de pesquisa em PLN de DIAS-DA-SILVA (1996), com investigações nos Domínios Lingüístico e Representacional. A metodologia é clara e aplicável à construção de sistemas de TA lingüisticamente motivados e ao tratamento do problema posto pelas divergências de tradução.

Acreditamos ter contribuído para o campo da TA em português, explorando uma metodologia original que permite a discussão de questões relevantes para os estudos lingüísticos.

Em termos mais específicos, consideramos ter contribuído para:

-o desenvolvimento de um *design* uniforme que se utiliza de módulos de processamento sintático e semântico para a língua portuguesa;

-o uso de uma só representação formal subjacente independente de língua que permite a acomodação dos dados da língua portuguesa;

-o uso de uma correspondência sistemática entre as estruturas sintática e léxico-semântica dos pares de sentenças do inglês e do português, via exame das divergências de tradução;

-a classificação das divergências de tradução como uma classe de problemas relevantes tanto para os Estudos Lingüísticos e para a TA;

-a formalização da noção de divergência e a explicitação dos mecanismos de correspondência sintática e léxico-semântica que se estabelece entre as línguas e a interlíngua do sistema de TA;

-o uso da composicionalidade para derivar a interlíngua de maneira a acomodar uma lista aberta de diferentes tipos de divergências de tradução, posto que não se trata de uma tipologia fechada.

Quanto à parametrização das divergências, considera-se tal abordagem desejável sob cinco perspectivas. Em primeiro lugar, ela permite que conhecimentos específicos de língua sejam representados independentemente dos princípios sintáticos e da representação conceitual geral prevista pelo sistema. Em segundo lugar, ela permite que a descrição de gramáticas de línguas naturais sejam simplificadas pela especificação dos

parâmetros. Em terceiro lugar, a especificação de parâmetros permite que o sistema de TA seja facilmente modificado para acomodar novas línguas, através da seleção dos parâmetros necessários e do fornecimento do léxico apropriado à tarefa da TA. Em quarto lugar, pode-se contar com um sistema modularizado, cujos componentes de análise e geração possam ser utilizados pelas línguas envolvidas na tradução. Finalmente, em quinto lugar, a abordagem do UNITRAN permite a elaboração de uma metodologia de trabalho em PLN mais constricta sob o ponto de vista do processamento: o ajuste de parâmetros somado aos princípios lingüísticos envolvidos na resolução dos problemas das divergências age como filtro para bloquear sentenças traduzidas mal formadas.

Não podemos afirmar que a investigação aqui realizada é suficiente para resolver o problema das divergências de tradução no âmbito da TA, posto que se trata de uma classe de problemas em aberto, que a pesquisa lingüística pode ainda aperfeiçoar, mas, com certeza, constitui um passo importante no intuito de se jogar uma luz sobre os problemas e abrir perspectivas para futuras investigações.

## REFERÊNCIAS BIBLIOGRÁFICAS

ARROJO, R. Os “estudos da tradução” como área de pesquisa independente: dilemas e ilusões de uma disciplina em (des)construção. In: **DELTA – Documentação de Estudos em Lingüística Teórica e Aplicada**. Campinas, n. 14, v.2, 1988.

AYAN, N. F. **Injecting linguistic information to improve word alignments for statistical MT systems**. Research Proposal. University of Maryland: UMIACS, 2004.

CHOMSKY, N. **Some concepts and consequences of the theory of government and binding**. 5. ed. The MIT Press: 1990.

COOK, V. J. **Chomsky’s Universal Grammar**. An introduction. Oxford & Cambridge (MA): Blackwell, 1989.

DIAS-DA-SILVA, B.C. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais**. Araraquara, 1996. 272p. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara.

\_\_\_\_\_. Os Domínios Lingüístico e Tecnológico do Estudo do Processamento Automático das Línguas Naturais. **Estudos Lingüísticos**, Assis, vol. 26, p.612-617, 1998a.

\_\_\_\_\_. Bridging the gap between linguistic theory and natural language processing. **16th International Congress of Linguists**. Proceedings... Pergamon: Oxford/Elsevier, Paper 0425, 1998b.

DIAS-DA-SILVA, B. C. ; MORAES, Helio Roberto de . A construção de thesaurus eletrônico para o português do Brasil. **Alfa - Revista de Lingüística**, São Paulo, v. 47, p. 101-115, 2003.

DORR, B. J. The use of lexical semantics in interlingual machine translation. **Journal of Machine Translation**, **7:3**, pp. 135-192, 1992.

\_\_\_\_\_. Interlingual Machine Translation: a parameterized approach. **Artificial Intelligence** **63** (1993a), 429-492.

\_\_\_\_\_. **Machine Translation: a view from the lexicon**. Cambridge (MA): The MIT Press, 1993b.

DORR, B.; JORDAN, P. W.; BENOIT, J. W. **A survey of current paradigms in machine translation**. EUA: Lamp Technical Report Series, UMIACS, 1998.

FILLMORE, C. J. “The Case for Case”. In: BACH, E., HARMS, R. T. (eds.). **Universals in Linguistic Theory**. Holt, Rinehart and Winston, 1968. p. 1-88.

GRIMSHAW, J. **Argument Structure**. Cambridge (MA): MIT Press, 1990.

HABASH, N., DORR, B. A Categorical Variation Database for English. In: NAACL/HLT 2003, **Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference**, p. 96-102, Edmonton, Canada, May-June 2003.

HAEGEMAN, L. **Introduction to Government and Binding Theory**. 2.ed. Oxford & Cambridge (MA): Blackwell, 1996.

HATIM, B.; MASON, I. **Discourse and the translator**. London/New York: Longman, 1990.

HUTCHINS, J.; SOMERS, H. L. **An introduction to Machine Translation**. London: Academic Press, 1992.

HUTCHINS, J. From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology. In: **Machine Translation**, v. 12, n. 3, 1997, pp. 195-252. (Disponível em <http://ourworld.compuserve.com/homepages/wjhutchins/> . Acesso em 28/07/2005.

\_\_\_\_\_. First steps on mechanical translation. In:

Disponível em <http://ourworld.compuserve.com/homepages/wjhutchins/> 1997. Acesso em 28/07/2005.

JACKENDOFF, R. **Consciousness and the computational mind**. Cambridge (Mass): The MIT Press, 1987.

\_\_\_\_\_. **Semantic Structures**. Cambridge/Mass.: The MIT Press, 1990.

\_\_\_\_\_. **The architecture of the language faculty**. Cambridge (Mass.): The MIT Press, 1997.

\_\_\_\_\_. **Foundations of language**. Brain, meaning, grammar, evolution. OUP: 2002.

LOBATO, L. **Sintaxe gerativa do português: da teoria padrão à teoria da regência e ligação**. Belo Horizonte: Vigília, 1986.

MAY, R. **Logical Form: Its structure and derivation**. Cambridge (Mass): The MIT Press, 1985.

MIOTO, C.; SILVA, M. C. F. ; LOPES, R. E. V. **Manual de Sintaxe**. 1. ed. Florianópolis: Insular, 1999.

MOUNIN, G. **Os problemas teóricos da tradução**. São Paulo: Cultrix, 1975.

NEVES, M.H.M. **Gramática de usos do português**. São Paulo: Ed. UNESP, 2000.

OLIVEIRA, M. F. **A teoria do léxico gerativo e evidência de corpus: explorações**. Florianópolis: UFSC, 2002. Mimeo. (Monografia).

- PUSTEJOVSKY, J. **The Generative Lexicon**. Cambridge (Mass.)/London: The MIT Press, 1995.
- RAPOSO, E. P. **Teoria da Gramática**. A Faculdade da Linguagem. Lisboa: Caminho, 1992.
- SADOCK, J. **Autolexical syntax**. Chicago: The University of Chicago Press, 1991.
- SELLS, P. **Lectures on contemporary syntactic theories**. Stanford: CSLI, 1985.
- SHIEBER, S. & SCHABES, Y. Generation and synchronous tree adjoining grammars. **Journal of Computational Intelligence** 7, 220-228. (1992).
- STEINER, G. **After Babel — Aspects of Language and Translation**. 2. ed. Londres, Oxford e Nova York: Oxford University Press, 1992.
- VERSPOOR, C. M. **Contextually-Dependent Lexical Semantics**. The University of Edimburg, 1997. Disponível em [http://www.cogsci.ed.ac.uk/~kversp/ftp\\_html/thesis.html](http://www.cogsci.ed.ac.uk/~kversp/ftp_html/thesis.html). Acesso em: 15 jul. 1999.
- WHITELOCK, P. Shake-and-Bake translation. In: **Proceedings of Fourteenth International Conference on Computational Linguistics**, Nantes, France, 1992. p. 784-791.