

UNIVERSIDADE FEDERAL DE GOIÁS  
FACULDADE DE LETRAS

ZENILDE NUNES BATISTA

ESTRUTURA LINGUÍSTICA E INFORMAÇÃO: UMA INTRODUÇÃO  
À ABORDAGEM DE ZELLIG S. HARRIS SOBRE OS FENÔMENOS  
DA LÍNGUA

Goiânia

2008

ZENILDE NUNES BATISTA

ESTRUTURA LINGUÍSTICA E INFORMAÇÃO: UMA INTRODUÇÃO  
À ABORDAGEM DE ZELIG S. HARRIS SOBRE OS FENÔMENOS  
DA LÍNGUA

Dissertação apresentada ao Curso de Mestrado em  
Estudos Lingüísticos da Faculdade de Letras da  
Universidade Federal de Goiás, para obtenção do  
título de Mestre em Lingüística.

Área de concentração: Estudos Lingüísticos

Orientador: Prof. Dr. Oto Araújo Vale

Goiânia

2008

ZENILDE NUNES BATISTA

ESTRUTURA LINGUÍSTICA E INFORMAÇÃO: UMA INTRODUÇÃO  
À ABORDAGEM DE ZELLIG S. HARRIS SOBRE OS FENÔMENOS  
DA LÍNGUA.

Dissertação defendida no Curso de Mestrado em Estudos Lingüísticos da Faculdade de Letras da Universidade Federal de Goiás, para obtenção do título de Mestre, aprovada em \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_, pela Banca Examinadora constituída pelo(a)s seguintes professore(a)s:

---

Prof. Dr. Oto Araújo Vale - UFSCar  
Presidente da Banca

---

Profa. Dra. Olga Ferreira Coelho - USP

---

Prof. Dr. Sebastião Elias Milani - UFG

---

Maria José Bocorny Finatto - UFRGS  
Suplente

Ao Altair, que sempre soube me compreender e encorajar, sempre presente, me apoiando de forma incondicional.

## AGRADECIMENTOS:

Ao Professor Dr. Oto Araújo Vale pela orientação acadêmica em todo o processo que envolveu a produção deste trabalho.

Ao Campus Catalão pela autorização e à PRPPG/UFG pela concessão do afastamento funcional para a elaboração deste texto de dissertação.

Ao Professor Dr. Sebastião Elias Milani e à Professora Dra. Joana Plaza Pinto, pelos apontamentos e sugestões quando do exame de qualificação.

À PROEC/UFG pela concessão dos dias de afastamento necessários para a adequação deste trabalho.

À cunhada e amiga Arnes, pela constante disponibilidade para consultoria em língua estrangeira.

A toda a minha família pela compreensão, especialmente meu esposo, filhas e filho.

À minha irmã Zenaide, por ser tão clara em demonstrar que sempre acredita em mim.

À amiga Iêda, que desde a graduação, sempre foi um apoio nos momentos de insegurança.

*É difícil exaurir todas as possibilidades dadas pela linguagem.*

*Gottlob Frege*

## SUMÁRIO

	RESUMO	
	ABSTRACT	
0.	INTRODUÇÃO.....	11
1.	APRESENTANDO ZELLIG SABBETAI HARRIS (1909-1992).....	14
2.	ESTRUTURALISMO AMERICANO: O LUGAR DE ZELLIG HARRIS...27	
2.1.	Origens e sequências das bases teóricas de Harris.....	31
2.1.1.	Estrutura.....	42
2.1.2.	Ontologia x Metodologia em estudo de língua.....	45
3.	ZELLIG HARRIS E A LÍNGUA.....	51
3.1.	Informação.....	54
3.1.1.	Informação sob as restrições que produzem as sentenças da língua.....	62
3.1.2.	Primeira restrição: Ordenação de Partes.....	64
3.1.3.	Segunda restrição: Probabilidade de Co-ocorrência.....	67
3.1.4.	Terceira restrição: reduções.....	69
3.2.	Metodologias de Harris.....	71
3.3.	Classes de Morfemas.....	72
3.3.1.	Abordagem Matemática da Língua .....	76
4.	O CONCEITO DE SUBLINGUAGEM EM HARRIS.....	80
4.1.	Problemática Inicial.....	80
4.2.	Definição.....	84
4.3.	Sublinguagem de uma ciência em foco.....	96
4.3.1.	Argumentos.....	99
4.3.2.	Operadores.....	99
4.3.3.	Resultados da Pesquisa em Imunologia.....	101
4.4.	Operadores Meta Científicos.....	104

4.5.	Linguagem científica.....	106
4.6.	Particularidades gramaticais relevantes.....	110
4.7	Sublinguagem de Base.....	113
4.8.	Aplicabilidade .....	114
5.	CONSIDERAÇÕES FINAIS.....	120
6.	REFERÊNCIAS BIBLIOGRÁFICAS.....	123

## **RESUMO**

As teorias de Zellig Harris no que concerne ao sistema linguístico tiveram destaque quando da virada estrutural. Mais que isso, transpuseram os limites das chamadas velhas grades e alcançaram a Lingüística hoje, como ela é praticada. Entretanto, a obra do linguista é ainda pouco discutida no meio linguístico atual. Objetivamos neste trabalho mostrar que Harris considera a língua como um sistema que está a serviço da transmissão de informação. Nesse sentido, buscamos apresentar uma introdução à teoria de informação linguística sob a ótica de Harris. Procedemos inicialmente a uma apresentação biográfica do linguista, apontando os principais rumos que sua obra tomou, desde o início de sua carreira, até a abordagem sobre informação e sublinguagem. Apresentamos também como Harris se situou nos estudos linguísticos estruturais e como a estrutura é abordada por ele, assinalando a noção de sublinguagem e a importância dessa abordagem para o embasamento teórico de diversos estudos linguísticos atuais.

**PALAVRAS-CHAVE:** estrutura, Zellig Harris, informação, restrições, sublinguagem.

**ABSTRACT**

The theories of Zellig Harris concerning to linguistic system came to evidence when structural language turn took place. More than this, they passed the boundaries of what were called ancient bars and achieved the linguistic as it is studied today. Never the less, the Harris's writings are little discussed in the actual linguistic set. Our purpose in this research is to show that Harris considers the language as a system that is at the service of transmitting information. In this sense, we intend to present an introduction to the linguistic information theory under Harris's approach. Initially, we made a biographical introduction of the linguist, showing the main courses that his writings took, since the beginning of his career, until the sublanguage and information approach. We also show how Harris placed himself in the structural linguistics studies and how the structure is approached by him, pointing out the sublanguage notion and the importance of this approach to the theoretic basis of several current linguistics studies.

**KEY WORDS:** structure, Zellig Harris, information, constraints, sublanguage.

## 0. INTRODUÇÃO

O interesse por este trabalho nasceu de um questionamento sobre o papel que teve/tem o trabalho de Zellig Sabettai Harris em relação ao que se fez em outros tempos e ao que se faz em lingüística atualmente. É interessante como a obra do linguista, uma das figuras eminentes do movimento estrutural americano, é ainda pouco discutida no meio linguístico atual. Muito do que se tem hoje em produções sobre a língua e também sobre o tratamento de linguagem, tem origem nas abordagens de Harris.

É um trabalho que traz nuances historiográficas e, dessa forma, pensamos ser necessária uma abordagem que tem como pano de fundo a história do movimento em que se localiza o autor em questão.

Harris considera a língua como estrutura que está a serviço da transmissão de informação. Entretanto, não a considera apenas enquanto estrutura: a estrutura é o recorte escolhido para suas investigações, mas o caráter social da língua é salientado pelo autor de forma recorrente. Ele percebe a língua como um sistema dinâmico, que está em constante mudança (cf. HARRIS, 1968, p. 17; 1988, p. 6). O que Harris chama de elementos não discretos da linguagem – entonação, prosódia, gestos, etc. – fazem parte de situações em que os aspectos psicológico e cultural, bem como situações interpessoais, se misturam para a realização da língua.

O relacionar a língua a formas matemáticas de representação também foi uma particularidade da obra de Harris. Em sua concepção, há propriedades comuns à língua e outros sistemas, e a representação matemática é uma delas. Essas noções se fizeram importantes para a evolução de sua abordagem de informação.

Em um primeiro momento neste trabalho, procedemos a uma apresentação biográfica de Zellig Harris, apontando os principais rumos que sua obra tomou, desde as descrições linguísticas feitas no início da carreira, até a abordagem sobre informação e sublinguagem. Para isso, além de buscarmos elementos em sua obra, procuramos também informações em escritos de alunos seus, que conviveram com o linguista e escreveram sobre sua vida e trabalho.

Em seguida, no capítulo dois, é oferecida uma visão de como Harris e suas teorias se colocam no movimento que foi chamado Estruturalismo, dada a abrangência e impacto desse movimento em diversas áreas do conhecimento, pois também entendemos que “as

idéias lingüísticas nunca se desenvolveram independentemente de outras correntes intelectuais do período”(SILVA 2006, p. 61), seja ele qual for. Na apresentação do ambiente estruturalista, embasamo-nos também na obra do historiador Françoise Dosse (1993), por ser um historiador que realizou uma busca detalhada e criteriosa, conforme explicitada em sua obra, numa descrição do estruturalismo desde seu nascimento.

Nesse processo, entendemos ser importante salientar algumas nuances diferenciadoras do desenvolvimento do Estruturalismo Europeu e Americano e dos objetos de que cada um se ocupou. Buscamos explicitar como Harris construiu suas bases teóricas, apresentando alguns dos seus precursores – com destaque para Sapir e Bloomfield - bem como de estudos contemporâneos aos seus e que se fizeram relevantes e contributivos também para o seu enfoque da língua. Nesse particular, buscamos perceber que papel a herança da Matemática e da Lógica desempenharam para o desenvolvimento de suas teorias. Discutimos ainda a noção de estrutura e como se pode percebê-la, de acordo com a concepção de diferentes autores.

No capítulo três, fazemos a exposição da concepção de estrutura da língua como Harris a apresenta. E ele o faz iniciando a partir da constituição fonêmica da língua, estendendo-se à sintaxe, à morfologia, aos enunciados, até chegar ao discurso. Trazemos para esta parte do trabalho, a concepção harrissiana de informação que o sistema lingüístico oferece à luz da gramática de operadores, criada pelo linguista em contraposição aos estudos tradicionais que consideravam até então a noção de sujeito e predicado nas construções sentenciais.

Ao apresentar sua concepção de língua, Harris parte do pressuposto desta enquanto sistema regido por mecanismos de restrição, seleção e combinação, e, dentro dessa orientação, desenvolveu sua abordagem teórica. São apresentadas as metodologias desenvolvidas por Harris, como a abordagem lingüística sob as principais restrições da língua, e como elas operam na produção da informação lingüística, bem como a relação dessa informação com a realidade. Também apresentamos o método de divisão em classes de morfemas, que foram estabelecidas por Harris para o Inglês, para uma descrição distribucional.

No capítulo quatro é apresentada a noção de sublinguagem, que permeia boa parte da obra do linguista. As sublinguagens são apresentadas como domínios específicos de realização lingüística: um fenômeno lingüístico em que o discurso de um domínio específico é construído com um léxico restrito, uma metalinguagem específica e

características gramaticais também específicas. Pode ser exemplificada pelo que se conhece comumente por textos técnicos.

Esse componente presente no sistema é apontado por Harris como um meio que, como a linguagem geral, carrega em si informação mediante restrições. Entretanto, são restrições diferenciadas em relação às do sistema maior. Há especificidades no que diz respeito às gramáticas desses subsistemas. Inclusa na abordagem de sublinguagem está a apresentação da linguagem notacional, que foi e continua sendo importante para pesquisas linguísticas, como o tratamento da linguagem natural. Nesse sentido buscamos trazer à discussão o quanto a abordagem em termo de sublinguagens se faz promissora no desenvolvimento de métodos atuais de investigação no tratamento da linguagem.

Dessa forma, buscamos recuperar algo do que o linguista descreveu como sendo elementos do sistema e que conferem informação linguística.<sup>1</sup>

---

<sup>1</sup> Para uma leitura mais facilitada e linear do trabalho como um todo, as citações em língua estrangeira - uma vez que toda a obra de Harris se encontra no Inglês - foram traduzidas por nós no corpo do texto e em seguida colocadas em notas de rodapé na língua original.

## 1. Apresentando Zellig Sabbetai Harris (1909-1992)

Zellig S. Harris foi, segundo Watt (2005), um dos poucos lingüistas que - desde que tiveram início os estudos da linguagem, no começo do séc. XIX - pode ser chamado de gênio por qualquer entendimento do assunto.

Harris Nasceu em Balta, na Ucrânia, em 1909. Foi morar com a família na Philadelphia, Estados Unidos, aos quatro anos. E o nome Zellig Sabbetai foi escolhido por ele mesmo nessa época. O nome Sabbetai identifica a família como seguidora do correspondente movimento messiânico judaico.

Sua vida acadêmica – tanto discente quanto docente – se fez na Universidade da Pennsylvania até sua aposentadoria, em 1979. Nessa Universidade fundou o 1º Departamento de Lingüística, do qual foi presidente em 1955 (cf. NEVIN, 1992; MATHEWS, 1999). Seu trabalho se tornou importante para o desenvolvimento de diversas áreas da lingüística, configurando-se através de várias nuances. O legado por ele deixado para a Lingüística foi numeroso e denso.

Apesar de fazê-lo sob protestos de seus pares, introduziu a representação algébrica e o tratamento matemático abstrato, que atualmente se tornaram eficazes no campo lingüístico. Mostrou a necessidade de se investigar os morfemas descontínuos na análise gramatical e desenvolveu maneiras para acomodá-los, incluindo os componentes de morfemas longos. Desenvolveu a análise transformacional no contexto da análise do discurso, no sentido de obter o conteúdo informacional dos textos. (NEVIN, 1992). Contribuiu de forma considerável com os estudos fonêmicos, introduzindo contemporaneamente a Bloch e Tragger, a teoria das junturas (HARRIS, 1969[1951]).

Seus primeiros estudos se relacionaram a línguas semíticas (cf. HIZ, 1994 p.1; MATTHEWS, 1999, p.1), sendo que seu primeiro trabalho conhecido foi a tese de Pós-doutorado, sobre a gramática da língua Fenícia, *A Grammar of the Phoenician Language*, (HARRIS, 1990[1936]) em que descreve características formais da língua com base em estudos não apenas lingüísticos, mas também em buscas nas quais se serviu de corpus literários e históricos, como tabletes de argila e manuscritos. Na verdade, esse foi um trabalho de certa forma histórico a serviço da descrição lingüística. O trabalho com textos semíticos possibilitou a Harris destacar-se, junto com seu professor James A. Montgomery, na aptidão em decifrar textos ugaríticos (HIZ, 1994; WATT, 2005).

Harris costumava se definir não como um linguista necessariamente, mas como um metodologista (cf. GOLDSMITH, 2008 p. 12). O termo *método* foi na Linguística de Harris deveras recorrente, pois o autor apresenta suas teorias de descrição estrutural sempre de forma que por ele são nomeadas de métodos, idéia com a qual outros estudiosos, de outras áreas também comungam. Piaget (1970, p. 111), realça que “o estruturalismo não é uma doutrina ou filosofia, mas é essencialmente um método com tudo o que esse termo implica de tecnicidade, obrigações, honestidade intelectual e progresso nas sucessivas aproximações.”

O tipo de trabalho que desenvolvia, dizia Harris, talvez não encontrasse grande interesse por parte dos linguistas, pois seu trabalho, pelas particularidades que apresentava, não era parte da Linguística como ela era institucionalmente definida. Seu interesse se voltava de forma especial para técnicas analíticas da língua nas quais aplicava pressupostos de outras ciências – do cálculo, por exemplo.

Todas as idéias de Harris eram diferentes das que estavam sendo estudadas em todos os lugares: a idéia de ‘transformação’, mais tarde modificada e tornada famosa por Noam Chomsky, a idéia de autonomia da sintaxe, e a notação matemática reduzida que fez possível representar uma gramática em poucas páginas de algo que pareciam equações<sup>2</sup> (PUTNAM, 1989, p. xi).

O linguista acreditava, entretanto, que, apesar de seus pares apresentarem alguma resistência à sua abordagem, as pessoas comuns que se interessavam pela língua, pudessem se interessar também por seu trabalho.

Tendo iniciado com as línguas semíticas, realizou também um extenso trabalho de campo com as línguas nativas dos Estados Unidos. Sapir, antes de Harris, pesquisou línguas indígenas da América do Norte, explorando as relações da língua com a antropologia. Harris foi, ainda que informalmente, pupilo de Sapir, tendo também mantido um namoro com sua filha por algum tempo. O próprio Sapir se referiu a ele como seu herdeiro intelectual (NEVIN, 2002).

---

<sup>2</sup> All of Harris’s ideas were different from those that were being studied elsewhere: the idea of a ‘transformation’, later modified and made famous by Noam Chomsky, the idea of the autonomy of syntax, and the condensed mathematical notation which made it possible to represent a grammar in a few pages of what looked like equations.

Então, a exemplo de seu mentor, Harris inventariou diversas línguas, tendo os falantes nativos como informantes, nos procedimentos que utilizava para descobrir elementos e traços formais. Dentre as muitas línguas inventariadas, - cerca de quarenta e quatro (cf. NEVIN, 1992) - pode-se citar o *swahili*, o *cherokee*, o *Navaho* e o *Yokuts*. Em relação a essas línguas, seus estudos se fizeram sentir mais no que se refere ao aspecto fonêmico-fonológico, no que obteve importantes resultados, inclusive no que concerne a elementos comparativos entre essas línguas e o Inglês (HARRIS, 1969[1951]).

Após esse período, começou a investigar diretamente problemas gerais da Linguística, num trabalho que mostrou seu primeiro fruto no *Methods in Structural Linguistics* – mais tarde apenas *Structural Linguistics* (1969[1951]), que se tornou um clássico da Linguística descritiva.

Uma particularidade importante de sua abordagem foi relativa às restrições de ocorrência de elementos linguísticos, a iniciar pelos fonemas. Dessa forma, o primeiro fonema de uma palavra pode estar perto de qualquer outro fonema. A possibilidade de ocorrência do segundo fonema é limitada pelo primeiro; a ocorrência do terceiro é limitada pelos dois primeiros. E assim por diante. Por exemplo, em Português há a ocorrência de /tr/ inicial, mas /rt/ inicial não ocorre. O fonema /t/ é limitado pela ocorrência do /r/ inicial.

Harris enfatiza que a estrutura da língua pode ser encontrada apenas a partir da não equiprobabilidade da combinação de suas partes, isso a partir da análise no nível fonemático, chegando até os níveis chamados por Benveniste (1995[1966]) de superiores, que são os níveis frasal e textual.

Essa abordagem científica da estrutura fonêmica da língua – que é o que foi feito quando da virada estrutural - se valida por estabelecer que todas as expressões nas línguas do mundo não são apenas sequências de sons insignificantes, uns após outros, mas as realizações da língua em escolhas de fonemas, num inventário complicado escolhido pela língua com ‘olho simétrico’ (cf. GOLDSMITH, 2008 p. 5 e 6).

Harris acreditava, como Sapir e Bloomfield, que o principal elemento para o estudo científico da língua é a posição relativa dos segmentos linguísticos. Entretanto, sua abordagem se diferenciou da de vários seguidores de Bloomfield no sentido de que a investigação dos fonemas e morfemas levavam a resultados que envolviam não unicamente esses elementos formais, mas envolvia a morfologia, a sintaxe e inclusive não necessariamente nessa ordem. Os procedimentos formais, para Harris não conduzem a um resultado único (HIZ, 1994).

Um pouco mais tarde, durante os anos 60, ele se dedicou a outras áreas de pesquisa: iniciou um tipo de análise linguística que hoje se ligaria ao modelo da Linguística Computacional. Isso estava se mostrando como algo possível por conta da invenção do ENIAC<sup>3</sup>, que foi desenvolvido exatamente na Universidade Pennsylvania, durante a II Guerra Mundial. Nessa atividade, desenvolveu as bases do que seria o programa piloto de primeiro analisador sintático computacional no UNIVAC<sup>4</sup> (WATT, 2005; HARRIS, 1970 [1959]). Isso pode parecer um tanto precoce, pois o próprio desenvolvimento dos computadores estava em seu momento embrionário. Entretanto, não se pode esquecer que Harris aliava o estudo tradicional da Linguística a outras ciências que de alguma forma se relacionassem a ela ou que pudessem com ela contribuir. Não foi diferente com a nascente informática.

No desenvolvimento de sua abordagem matemática do sistema linguístico, na verdade, Harris não equalizou a língua com sistemas matemáticos e a encaixou neles; o que fez foi tratar os objetos e relações da língua como relações e objetos matemáticos, procurando uma estrutura matemática que fosse suficiente para a língua (HARRIS, 1981[1978]; 1968; 1982; 1991). Segundo Watt (2005), Harris foi o primeiro a esboçar a noção de que a Linguística poderia sintetizar ou produzir as sentenças de uma dada língua em forma de algoritmo a partir de algumas regras explícitas, isto é, oferecer condições para uma explicação matemática sobre a formação de sentenças. Nesse modelo, “o tratamento matemático da língua surge de uma relação fundamental de ‘dependência sob dependência’ que divide os grupos de palavras [...] em classes de dependência [...] de operadores e seus argumentos”<sup>5</sup> (RYCKMAN, 2002, p. 21, rp).

Harris investigou alguns campos linguísticos que se fazem relevantes hoje para diversos estudos. A abordagem da sublinguagem (cf. 4.) foi uma inovação, no que se refere a investigar as peculiaridades nas gramáticas dessas sublinguagens. A exemplificação com análises da sublinguagem da ciência e de sua gramática<sup>6</sup>, o que inclui o sistema de

---

<sup>3</sup> Electrical Numerical Integrator and Calculator, computador digital eletrônico de grande escala

<sup>4</sup> UNIVersal Automatic Computer, primeiro computador fabricado e comercializado nos Estados Unidos.

<sup>5</sup> The mathematical treatment of language arises from a fundamental relation of ‘dependence on dependence’ which partitions the set of words [...] into dependence classes [...] of operators and their arguments.

<sup>6</sup> Harris se refere em sua obra a diversas sublinguagens presentes no sistema linguístico, entretanto, se atém, em suas investigações, à sublinguagem científica de pesquisa médica, mais especificamente à imunologia. Esta foi uma sublinguagem investigada por ele com o auxílio de seus alunos e de seu irmão, médico imunologista, que proporcionou a Harris o acesso a vários trabalhos da área. (cf. HARRIS 1988, p. 41 e NEVIN 1992, p. 5)

operadores gramaticais baseados na dependência entre palavras, se fez um aspecto diferenciado na obra de Harris.

Foi feita uma descrição que abrangeu especialmente uma pesquisa, abrangendo trabalhos feitos na área de imunologia, durante o período de 1935 a 1966<sup>7</sup>; essa descrição foi explicitada em publicações específicas somente já no final de sua carreira (cf. HARRIS, 1988; et al 1989; 1991). Nessa análise discursiva, além da análise sentencial voltada para a sublinguagem da pesquisa científica, é possível notar elementos da teoria transformacional de Harris<sup>8</sup>. Ele “trouxe a longa linha de pesquisa em estrutura transformacional de volta a suas origens na análise do discurso, criando o campo da gramática da sublinguagem...”<sup>9</sup> (NEVIN, 1992, p. 2). Efetivamente, Harris explicita as sublinguagens como sendo construídas sob restrições diferenciadas daquelas da linguagem. Daí, uma gramática também peculiar.

Quando um subgrupo de um sistema é fechado sob operações do sistema, o subgrupo constitui um subsistema. Se tomamos sentenças tais como são usadas na ciência, e operamos nelas com conjunções ou as transformações da língua, obtemos novamente uma sentença tal qual é usada naquela ciência<sup>10</sup> (HARRIS, 2002a, p. 218).

Por volta de 1969, Harris introduziu algumas técnicas diferenciadas de abordagem das propriedades da língua, que se fizeram relevantes para investigar a questão da informação. Nessa linha, as estruturas dos constituintes, das frases e de cadeias se fizeram importantes. Nesse novo modelo, Harris usou a relação operador-argumento, em que se atenta para o fenômeno de seleção e combinação de palavras na sentença (HIZ, 1994).

Inclusa nessa concepção, está também a dos universais linguísticos; Harris trata de um princípio universal que regula a formação de sentenças. Aqui a relação de dependência entre operadores/argumentos é fundamental para a estruturação das sentenças da língua (cf. DE BRABANTER, 2001, p. 04). “Dessa forma então, pode objetivar, a partir das

---

<sup>7</sup> Suas publicações direcionadas apenas ao tema, com resultados dos corpus analisados, foram em 1988 (Language and Information) e 1989 (The Form of Information in Science: Analysis of an Immunology Sublinguagem), mas em 1968 (Mathematical Structures of Language) e 1982 (A Grammar of English on a Mathematical Principles) e já adiantou a abordagem teórica do assunto.

<sup>8</sup> A teoria transformacional é bem detalhadamente descrita no Mathematical Structures of Language (1968) e nos papers publicados entre 1947 e 1970.

<sup>9</sup> “Harris brought the long line of research in transformational structure back to its roots in discourse analysis, creating the field of sublinguagem grammar.”

<sup>10</sup> When a subset of a system is closed under operations of the system., the subset constitutes a subsystem. If we take sentences such as are used in a science, and operate on them with the conjunctions or the transformations of the language, we obtain again a sentence such as is used in that science.

raízes “N” e “S”, uma análise desenvolvida de operador/argumento proporcionando uma intrigante reformulação de toda a tradicional [teoria de] partes do discurso e das frases, orações e sentenças em que eles ocorriam” (WATT, 2005, p. 7)<sup>11</sup>.

O tema da informação que a língua traz em si foi algo que permeou de forma recorrente a obra de Harris (1988; 1991; et al, 1989; 1968; 1970[1952]). “Em sua mais simples essência, o trabalho de Harris tem tornado possível uma consideração matematicamente construtiva da língua como um sistema auto organizável que expressa e carrega informação”<sup>12</sup> (RYCKMAN, 2002, p. 20-21). O linguista realmente se interessava em como a língua pode carregar ou transmitir informação, e seu trabalho em relação a isso foi verdadeiramente consistente durante mais de cinquenta anos (NEVIN, 1992).

Um ponto importante é que uma das conclusões da análise linguística de Harris refere-se à representação semântica que se obtém através das análises de dependência entre operador-argumento encontrada nos usos da língua. Representação essa que não retém todos os aspectos de significado. Entretanto, o mecanismo de descrição não empresta qualquer estrutura estranha àquela encontrada na língua; dessa forma, estrutura e significado se correlacionam. Sua discussão em torno da informação que se pode obter através do conteúdo do discurso – para ele e para sua abordagem, discurso é o que está explícito no texto<sup>13</sup> – se liga à teoria de como a informação se presentifica numa descrição de aspectos semânticos.

Em se tratando da gramática transformacional de Harris, há um processamento de derivação de sentenças a partir de outra sentença, uma *kernel sentence* (sentença núcleo), que chamou mais tarde de *base sentence*<sup>14</sup>. Essa teoria transformacional, que pode ser vista

---

<sup>11</sup> “In this fashion then, from just the two primitives “N” and “S”, a fully developed functor/argument analysis can aim at providing an intriguing reformulation of all the traditional parts of speech and of the phrases, clauses, and sentences they occur in.”

<sup>12</sup> “In its barest essence, Harris’s work has made possible a detailed constructive mathematical consideration of language as a self-organizing system that expresses and carries information”

<sup>13</sup> As discussões de Harris (1970[1952]) no que se refere à Análise do Discurso ao que parece, se relacionam, grosso modo, ao que se teria hoje como uma abordagem da lingüística textual, pois são as relações entre partes do texto que são analisadas, em que se percebe a importância de elementos como tópicos frasais, coesão e coerência textual, para a percepção da harmonia do texto e para a obtenção de informação no mesmo. Nesse sentido, dialoga com estudos recentes, como o de SOUZA E SILVA & KOCH (2001 p. 134), bem como de BENTES (2000, p. 247-252), que apontam o estudo sistematizado do texto como fruto da necessidade de se compreender fenômenos tais como co-referência e relações semânticas entre orações sem conectivo explícito. Essa necessidade levou alguns linguistas *contemporâneos* a tomarem como unidade básica de análise o texto, em detrimento da frase, na busca de uma gramática que verifique os mecanismos que fazem de um texto um texto, e não apenas frases que se somam.

<sup>14</sup> Desde o início da abordagem transformacional – da década de 40 - até o final da década de 70, Harris se referia às sentenças base como *kernel sentences*. A partir do trabalho de 1982, *A Grammar of English on a Mathematical Principles*, passou a se referir às *elementary sentences*, ou *base sentences*.

já no final do *Structural Linguistics* (1969[1951]), foi apropriada por Noam Chomsky e reformulada, sendo por ele relacionada a transformações dentro da teoria da gramática gerativa.

Interessa pontuar que as idéias de gramática transformacional foram objeto de discussões estimulantes entre os dois linguistas (HARRIS, 2002b), e ambos – cada um a seu modo - obtiveram um do outro contribuições para o desenvolvimento de suas teorias. Deve-se reconhecer que, apesar de Harris ter dado a largada nessa revolução transformacional que agitou o meio linguístico nos anos 50, suas teorias não tiveram o mesmo impacto que as de Chomsky. Efetivamente, Harris parece ser mais lembrado pela maioria dos linguistas apenas como o professor de Noam Chomsky. Entretanto, Huges (2001) reitera a importância, para Chomsky, da teorização pioneira de Harris, pois foi após a abordagem transformacional de Harris que aconteceu a revolução Chomskyana dessa teoria.

A gramática transformacional foi desenvolvida não por Chomsky, mas por Zellig Harris. Chomsky mais tarde escreveu que quando começou a investigar a sintaxe gerativa mais seriamente, no final dos anos 50, ele adaptaria para seu propósito um novo conceito que tinha sido desenvolvido por Zellig Harris e alguns de seus estudantes, notadamente o conceito de transformação gramatical<sup>15</sup> (HUGES, 2001, p. 40).

Goldsmith (2008) enfatiza o caráter diverso da linguística de Chomsky em relação à de Harris. Enquanto esta se pautava por um modelo científico de análise *das línguas* em estudo, a de Chomsky é uma ciência que envolve a competência dos falantes que utilizam ou criam sentenças nessas línguas (itálico nosso). É uma abordagem racionalista dentro da concepção da gramática gerativa (GOLDSMITH, 2008, p. 3). Salienta essa como sendo uma das maiores diferenças entre a concepção de análise linguística de Harris e a da Gramática Gerativa: que Harris se ocupou de analisar a totalidade dos eventos linguísticos intersubjetivos, independentemente de propriedades da mente (GOLDSMITH, 2008, p. 26). Harris não se pautava por tornar a Linguística uma extensão da psicologia, uma ciência cognitiva. Chomsky, sim, se ocupou de fazer isso. As teorias ligadas ao

---

<sup>15</sup>Transformational grammar was developed not by Chomsky but by Zellig Harris. Chomsky later wrote that when he ‘began to investigate generative syntax more seriously’ in the late 1950s, he would ‘adapt for this purpose a new concept that had been developed by Zellig Harris and some of his students, namely, the concept of grammatical transformation.’

gerativismo, segundo Ryckman (2002), se amarram à teoria formal da linguagem. Na verdade, elas surgiram da junção da psicologia cognitiva e da teoria formal da linguagem, o que propiciou para os gerativistas o estudo da linguagem e da mente.

Dessa forma, dentro da linha Chomskiana de estudos linguísticos, apesar desta se alinhar com a postura mentalista, é possível encontrar elementos que se ligam à visão empiricista da linguagem. O próprio termo que nomeia sua teoria [gerativa] foi emprestado por ele da Matemática (cf. SOUZA & SILVA e KOCH, 2001, p. 39). O gerar sentenças, sob essa perspectiva, não é mais do que a habilidade que o falante tem de selecionar e combinar elementos da língua, usando um sistema finito na produção de enunciados infinitos. Esse sistema recursivo de regras é uma idéia anteriormente sistematizada por Harris. O lingüista diz:

A computabilidade de identificação de uma subcadeia é baseada no fato de que as cadeias têm uma estrutura recursiva. Isto é necessariamente o caso: a linguagem natural é um conjunto incontável de cadeias (sentenças) de um vocabulário finito, e é produzido por um conjunto finito de processos para combinação de subcadeias (e por falantes finitos)(HARRIS, 1970[1959], p. 255)<sup>16</sup>.

Os universais linguísticos também diferem entre a teorização de Harris e a de Chomsky. Em Harris eles se colocam diretamente em uma relação de dependência entre operadores/argumentos que é fundamental para a estruturação das sentenças da língua (cf. HARRIS, 1988, p. 1-32; 1981[1976], p. 352-355; DE BRABANTER, 2001, p. 4). A diferença está em que Harris percebeu o sistema recursivo enquanto existente na estrutura lingüística, enquanto a teoria de Chomsky se projeta à criatividade do falante, sua capacidade na emissão e compreensão de frases da língua. Para Harris, a gramática transformacional representava uma ferramenta importante para analisar e trabalhar com os dados. Chomsky por sua vez, optou por colocar à parte os dados, privilegiando antes o conhecimento que permite ao usuário produzir os dados.

Em seu *Transformational Analysis*, Chomsky expressa seus agradecimentos pela estreita colaboração de Harris, salientando que as idéias subjacentes na obra eram dele. Em relação ao *Syntactic Structures* (1957), Chomsky anota no prefácio:

---

<sup>16</sup> Computability of substring identification is based on the fact that the strings have a recursive structure. This is necessarily the case: Natural language is a denumerable set of strings (sentences) of a finite vocabulary, and is produced by a finite set of processes for combining substrings (and by finite speakers).

Durante todo o período desta pesquisa eu tenho tido o benefício de conversas muito longas e frequentes com Zellig Harris. Tantas de suas idéias e sugestões estão incorporadas no texto que se segue e na pesquisa na qual ele é embasado, que eu não me atentarei em indicá-las por referência especial”<sup>17</sup> (CHOMSKY, 1976[1957], p. 6).

Harris e Chomsky faziam parte da mesma comunidade judaica, suas famílias eram próximas, mas o contato mais próximo entre os dois se deu na Pennsylvania, quando Chomsky se encontrava de certa forma desiludido com as perspectivas que tivera até então em relação aos cursos que frequentava ali. Sua esposa, Carol Chomsky, expressou que Harris exerceu talvez a principal influência no marido e que foi também responsável, de diferentes formas, pelo direcionamento posterior da vida intelectual de Chomsky. E é o Próprio Chomsky que diz: “...‘ele era uma personalidade muito cheia de energia, e era muito interessado em encorajar pessoas jovens a fazer coisas’”<sup>18</sup> (HUGES, 2001, p. 38). Na verdade, a influência de Harris sobre Chomsky é inegável, e vai desde a sua vida acadêmica até ao posicionamento político-social do linguista. Segundo o próprio Chomsky, em 1955 era ele ainda um estranho no ninho dos estudos linguísticos. Sua inserção no meio se deu graças ao interesse despertado pela abordagem linguística de Harris, bem como ao seu apoio acadêmico (JOHNSON, 1998).

Parece haver uma certa nostalgia nos escritos dos ex-alunos de Harris. De forma que tentam defender as teorias do mestre, reconhecendo ao mesmo tempo que elas foram ricocheteadas, quando de sua formulação e apresentação por ele. A abordagem linguística descritiva conforme os pressupostos harrissianos parece ter tido sua aceitação prejudicada pela atitude passiva de seus seguidores, diante do crescimento de abordagens como a cognitivista, que, como já dissemos, tem raízes na abordagem formal, mas que se posicionou pelo abandono da alternativa teórica proposta por Harris.

Em realidade, sempre há alguma distorção em relação às teorias de Z. Harris. Afirma-se, por exemplo, que os “procedimentos de descoberta”<sup>19</sup> eram uma base do trabalho dele (cf. ILARI, 2004). Nevin (1992) e Watt (2005) afirmam que os procedimentos de descoberta não eram seu objetivo, eram apenas um método auxiliar, que

<sup>17</sup> During the entire period of this research I have had the benefit of very frequent and lengthy conversations with Zellig S. Harris. So many of his ideas and suggestions are incorporated into the text below and in the research on which it is based that I will make no attempt to indicate them by special reference.

<sup>18</sup> “...he was a very powerful personality, and he was very interested in encouraging young people to do things.”

<sup>19</sup> Na verdade, o termo “procedimentos de descoberta” não foi usado por Harris. Foi criado por Chomsky, para contrastar com *procedimentos de avaliação*, uma noção central da gramática gerativa até antes do início da década de oitenta, a partir de quando se adotou os *princípios e parâmetros* (cf. GOLDSMITH 2008, p. 9; MATHEWS 1999, p. 2; LIN 2002, p. 75)

fazia a diferença na descrição das particularidades fonêmicas. Goldsmith (2008) reitera que esses procedimentos são importantes no processo de aliar a teoria com os dados linguísticos concretos, na busca de se provar uma hipótese, ou de se construir uma.

Diferentemente do modelo racionalista, o empiricista “vê o dado que é coletado pela ampla comunidade de pesquisadores linguísticos como uma parte integral do trabalho da comunidade científica: sem dados, sem ciência.”<sup>20</sup>(GOLDSMITH, 2008, p. 10). Tendo isso em vista, o que se percebe é que Harris se interessava em determinar quais procedimentos em princípio poderiam levar a uma compreensão mais profunda de um sistema de linguagem natural (cf. GOLDSMITH, 2008 p. 19).

Harris se ocupou, desde seus primeiros trabalhos, em salientar o caráter metalinguístico da língua. Sua atenção sempre se voltou a acentuar que a língua é o único objeto de estudo existente não possuidor de uma metalinguagem externa à sua própria. Enfatiza que qualquer outra ciência tem uma metalinguagem externa à sua, que se coloca a seu serviço; por exemplo, as práticas dos domínios da Física, da Biologia, têm seu suporte na língua para expressar conteúdos e validar suas relações e conclusões. “Mas não é assim com a ciência da língua. Harris reconheceu que não há nenhum ponto de vantagem fora da língua a partir do qual descrever a língua [...]. Em contraste, a teoria gerativa postula uma metalinguagem universal, externa à língua, que é parte do aparato biológico”<sup>21</sup>(NEVIN, 1992, p. 4). Harris afirma que não se pode derivar a estrutura linguística a partir de conclusões de qualquer estudo alicerçando-se numa relação de (outra) metaciência com a Linguística (cf. HARRIS, 1988, p. 2-3; 2002b, p. 7-9; 1968).

Para além de sua vida acadêmica, destacou-se em sua trajetória política como participante ativo em movimentos libertários. Foi comprometido com a independência do estado de Israel (Watt, 2005), onde sua esposa, a física Bruria Kaufman, primeira assistente do físico Albert Einstein entre 1950 e 1955, trabalhou por alguns anos como professora do Instituto Weizmann (MATHEWS, 1999). Talvez a opção de Harris pelo Sionismo (nacionalismo judaico) tivesse um forte apelo em sua opção política, o anarquismo, que defende, entre outras posturas, uma ampla oposição a formas autoritárias de governo, e quaisquer outros tipos de ordens hierárquicas que não sejam livremente aceitas.

---

<sup>20</sup> “...sees the data which is collected by the larger community of linguistic researchers as an integral part of the work of the scientific community; no data, no science.”

<sup>21</sup> Not so for a science of language. Harris recognized that there is no vantage point outside of language from which to describe language[...]. By contrast, Generativist theory postulates a universal metalanguage, external to language, that is part of one’s biological endowment.

As aspirações de Harris se fizeram sentir também em relação à intercomunicação linguística e, principalmente, com vistas ao desenvolvimento científico. Tratando desse assunto, Harris acentuou a necessidade de uma língua franca que atingisse os usuários do mundo todo. No campo científico, eminentemente, tal língua seria profícua por oportunizar um tratamento comum de trabalhos científicos, de forma que os pesquisadores não teriam que aprender (outras) línguas estrangeiras para o trabalho com cientistas de outros países (cf. HARRIS, 1962, p. 794-795).

Como Sapir e Bloomfield, Harris também tinha interesse em problemas de comunicação internacional e propostas para uma língua auxiliar internacional. E também como eles, ele tinha um particular interesse em cooperação e comunicação internacional *para os interesses da ciência*. Esta é provavelmente uma razão por que muito de seu material para análise linguística foi extraído de jornais técnicos de várias ciências. Outra e mais imediata razão é que a sublinguagem técnica da ciência tem as restrições sintáticas e semânticas melhor definidas<sup>22</sup> (NEVIN, 1992, p. 4) Grifo nosso.

Sua constante busca por estudos linguísticos que fossem contributivos para o desenvolvimento dessa ciência, entretanto, não fizeram dele um pesquisador de horizonte único, o que pode ser observado pelas suas incursões em movimentos políticos e sociais; ademais, apesar de quase não cultivar a vida social, procurava demonstrar afeto e atenção aos que o rodeavam.

Nos escritos deixados por alguns de seus ex-alunos - Nevin (1992); Mathews (1999); Hiz, (1994) e Watt (2005) -, é unânime e constante a idéia do estilo acessível e quase paternal do professor, contrariamente à postura ditatorial que alguns atribuíam ao linguista. Goldsmith (2008, p. 1) diz: “Zellig Harris tem um pouquinho de um avô que eu nunca tive.”<sup>23</sup> Outro ponto que é consenso diz respeito ao compromisso que Harris tinha com o trabalho, compromisso esse que tentava transmitir também àqueles que trabalhavam com ele, eminentemente seus alunos, que além de alunos, eram professores iniciantes, aconselhando-os a deixar em segundo plano a vida social para priorizar o trabalho, a vida acadêmica. “Ele foi, para a sorte daqueles que o conheceram[...], uma pessoa extremamente decente, de alta integridade: ele teve absoluto e explícito descaso pelas

---

<sup>22</sup> Like Sapir and Bloomfield, Harris had an interest in problems of international communication and in proposals for an international auxiliary language. And like Sapir and Bloomfield he had in particular a long-standing interest in international cooperation and communication in science. This is probably one reason much of his material for linguistic analysis was drawn from technical journals of various sciences. Another and more immediate reason is that technical sublanguages of science have more well-defined semantic and syntactic restrictions.

<sup>23</sup> “Zellig Harris is a bit like a grandfather I never met.”

posições e status nesta sociedade, bem como por sua grande injustiça e desigualdade”<sup>24</sup>(NEVIN, 1992, p. 5).

Harris expressava se pautar sempre por demonstrar seu trabalho como sendo algo que deveria ser feito com a imparcialidade necessária para apresentar seus préstimos à sociedade, mas sem chamar atenção para si. Um episódio interessante foi quando de seu aniversário de 60 anos, em que alguns de seus alunos planejaram elaborar um *Festschrift*<sup>25</sup> em sua homenagem, o linguista respondeu da seguinte forma:

‘Tal publicação seria uma afronta profunda a mim e ao meu senso de valores. Eu tenho me conduzido ao longo da vida sob o princípio de que os cientistas podem ser pessoas que fazem o melhor trabalho que eles puderem em prol do conhecimento e de seus valores humanos. Qualquer reconhecimento especial – e inevitavelmente individual – de seu trabalho, como honras, prêmios, e *Festschriften*, é repugnante a mim e violaria o que eu sinto como retidão humana e dignidade. Portanto, eu peço a você que cancele esse projeto...[...] Eu estou certo, porém, que você me compreenderá e respeitará meus princípios ainda que eles possam parecer excessivos’<sup>26</sup> (WATT, 2005, p. 17).

A concepção do fazer ciência para o linguista se configura então como um trabalho que tem grande valor, mas que deve ser feito com modéstia e compromisso, deixando de lado os possíveis lauréis que possam ser atribuídos aos cientistas.

Nesse trabalho científico da Linguística, Harris tem como concepção a de que, a língua é um fenômeno que se apresenta eminentemente como estrutura, regida por restrições (cf. HARRIS, 2002b). Se não houvesse estrutura - e esta regida pelo condicionamento de partes – então as partes das sentenças poderiam ser colocadas aleatoriamente juntas, dessa forma seria o caos-caos.

Ao mesmo tempo em que considera a língua como estrutura organizada, a representação desse aparato, para Harris, foi tida acima de tudo, como uma instituição social, continuamente em mudança, bem como em contínua recriação em seu uso

<sup>24</sup> “... he was, to the good fortune of those who knew him, and I hope of himself, an extremely decent person of high integrity: he had utter and explicit contempt for the posturings and status in this society as well as for its vast injustice and inequality..”

<sup>25</sup> *Festschrift* é um termo emprestado do alemão que se refere a um tipo de publicação que homenageia uma personalidade destaque no meio acadêmico por ocasião da aposentadoria ou em comemoração a uma data muito especial, geralmente contendo contribuições originais de seus colegas mais próximos, bem como de primeiros estudantes de doutorado.

<sup>26</sup> ‘Such a publication would be a deep personal affront to me and to my sense of values. I have managed to live this long with the principle that scientists can be people who do the best work they can for the sake of knowledge and of its human value. Any special – and unavoidably individouous – recognition of their work, such as honors. Prizes, and *Festschriften*, is abhorrent to me, and would violate what I feel is a human right and dignity. Therefore, I ask you to Withdraw this activity...[...] I am sure, however, that you will understand me, and will respect my principles even if they may seem excessive.’

(HARRIS, 1968, p. 17; 1988 p. 92-113). “Ele discute [...] a natureza da língua como um sistema auto organizável em um contexto de uso social<sup>27</sup>” (NEVIN, 2002, p. xxx). Tendo isso em conta, não é difícil compreender que o objetivo de seus métodos não eram – e nem poderiam ser - cristalizar a língua informal em análises formais, mas verificar, para qualquer resultado dado, se esse resultado tinha uma relação válida com os dados da língua.

---

<sup>27</sup> “...he discusses [...] the nature of language as a self organizing system in a context of social use.”

## 2. Estruturalismo americano: o lugar de Zellig Harris

Seguindo em nossa proposta de retorno à obra desse lingüista americano - um dos de maior expressão - e de reencontrar ali conceitos que contribuíram para embasar vários estudos lingüísticos recentes, interessa pontuar desde já nuances diferenciadoras no movimento estruturalista da Europa e dos Estados Unidos. Conforme é de conhecimento geral, na Europa - e eminentemente na França - esse movimento teve uma grande repercussão em outras áreas de conhecimento. Entretanto, na América do Norte, pareceu circunscrever-se mais à lingüística descritiva e à antropologia, nesta última em estudos que se revestiram de natureza lingüística, uma vez que buscavam descrever estruturas das línguas.

Considerando os rumos tomados por esses estudos lingüísticos, na Europa a Linguística Estruturalista se desenvolveu influenciada de forma mais profunda pelos postulados Saussurianos - tendo como figuras de destaque a de Lévi-Strauss, Martinet, Benveniste e outros nomes de expressão - e na América, linguistas como Leonard Bloomfield, Bernard Bloch, George Tragger e Zellig Harris desenvolveram teorias de cunho estrutural abstraídas da antropologia de Boas e na Linguística de Sapir.

Zellig Harris se coloca no cenário do Estruturalismo americano demonstrando em sua obra que, para além de as línguas terem características estruturais universais (SAPIR, 1949; CHOMSKY, 1973)<sup>28</sup>, elas seriam passíveis de um tipo de representação algébrica: algo que demonstrasse uma propriedade matemática das línguas. Apesar de Louis Hjelmslev também ter focado a lingüística como uma espécie de álgebra - em que contam, para a definição de estrutura, as relações formais entre os elementos e não a materialidade dos elementos relacionados - (cf. HJELMSLEV, 1975[1961]; DOSSE, 1993 p. 93), ao que parece, essa era uma abordagem na qual a lingüística americana, com ênfase para a de Harris, se destacou, aprofundando-se nesse enfoque.

O que se pretende destacar aqui é como Harris se colocou no panorama dos estudos estruturalistas do movimento americano, quais eram os valores que se atribuía à

---

<sup>28</sup> “There are such general characteristics which apply to all languages, living or extinct, written or unwritten.” (SAPIR 1949, p. 1)

“A evidência de que atualmente dispomos dá sustento ao ponto de vista de que todas as línguas humanas compartilham de propriedades de organização e estrutura em suas camadas profundas. Essas propriedades - esses universais lingüísticos - podem com plausibilidade ser tomados como um dote mental inato, antes que como resultado de aprendizagem.” (CHOMSKY 1973, p. 42)

lingüística na época, para discutir mais adiante como/em quê a abordagem lingüística de Harris contribuiu para a evolução desses estudos.

A abordagem estruturalista então, desde os primórdios, se caracterizou pela busca das estruturas lingüísticas, tendo raízes mesmo quando dos trabalhos comparatistas. (DOSSE,1993). Desde esse período, tinha-se como essencial a necessidade da reconstituição de uma estrutura da língua. Então, mesmo antes de assim ser chamado, o estruturalismo se colocou como semente no seio dos estudos lingüísticos históricos.

Dosse (1993) discorre amplamente sobre o desenvolvimento do estruturalismo na França. E apresenta de maneira interessante, como se deram os contatos entre figuras que vieram a se tornar eminentes no estruturalismo europeu e personagens do estruturalismo americano, como Boas, Sapir e Bloomfield. Em verdade, como já dito anteriormente, a França é considerada o centro dos estudos estruturais<sup>29</sup>, entretanto, é pertinente realçar que os estudos franceses se alimentaram de fontes americanas bem antes de se projetarem para o mundo lingüístico. Ainda na primeira metade do século XX, encontram-se nos Estados Unidos linguistas que, sob a herança de Boas<sup>30</sup> e Sapir, desenvolviam pesquisa lingüística de cunho estrutural, - dentre eles Zellig S. Harris – e deram continuidade ao trabalho de descrição lingüística até então realizado.

Em se tratando da Linguística americana versus européia, é possível notar um certo ar de antagonismo entre os pares. Para Salum (1994), a Linguística européia se ressentia de que na América, não foram conferidos os devidos créditos a Saussure<sup>31</sup>, e, os americanos por sua vez, consideravam o fazer Linguística na Europa um trabalho um tanto teórico, desprovido de resultados que refletissem o manusear dos dados, de forma concreta. A proposta deste trabalho não abarca amplamente essas discussões, porém se faz interessante salientá-las porque são importantes para se compreender o direcionamento que teve a Linguística nos Estados Unidos e, conseqüentemente, na obra de Zellig Harris.

Nos Estados Unidos, as idéias estruturalistas, ainda no início do século XX, evoluíram de maneira relativamente independente da Europa, como resultado do trabalho

---

<sup>29</sup> Dosse fecha o texto de sua *História do Estruturalismo* com a seguinte frase: “É, portanto, sob a bandeira tricolor da França, e somente da França, que o estruturalismo vai expandir-se até fascinar os outros países, mas como produto específico do solo francês que se saboreia pela necessidade de exotismo.” (1993, p. 431)

<sup>30</sup> “Boas estabeleceu [...] o método de depreender da própria língua analisada os princípios gramaticais que espontaneamente resultavam da estrutura lingüística, revelada pela análise.”(CÂMARA JR., J. M. em apêndice a SAPIR, E. Linguagem 1980[1921], p. 186)

<sup>31</sup> “É bem certo que a lingüística americana moderna surgiu sem especial contribuição de Saussure; não deixa, porém, de causar espécie a onda de silêncio da quase totalidade dos linguistas norte americanos com relação ao Cours.” (SALUM, I. N. em prefácio ao Curso de Linguística Geral 1994, p. XV)

de Sapir e Boas<sup>32</sup>, conforme dito anteriormente. Boas compreendeu que uma classificação genética adequada seria impossível com o mau conhecimento dos objetos - isto é, as línguas - por classificar; dessa forma, contribuiu para a criação das bases de uma Linguística descritiva, na América do Norte. Seguindo essa mesma orientação, os estudos de Sapir também o levaram a definir a língua como forma.

Sapir realizou estudos descritivos de diversas línguas indígenas como o *Nass*, o *Chinuk*, o *Nutka*, o *Iroquês*, o *Paiúte*, dentre outras. E, tendo como referência as línguas descritas, fez um confronto entre elas - em se tratando de produtividade de combinações seletivas tanto em fonética quanto em vocábulos discriminados em classes. De acordo com o autor, "...toda língua tem um sistema fonético íntimo de padrão definido. Vemos agora que também tem o sentimento definido de constituir padrões para a formação gramatical." (SAPIR, 1980[1921], p. 55).

Características formais do Hebraico, do Chinês, do Siamês e do Latim também foram discutidas por Sapir (1980[1921]), e nessas descrições foi introduzida a noção de ambiente, que seria mais tarde aprofundada por Harris (1969[1951]).

Então, o que podemos chamar de início de uma escola linguística norte-americana se consubstancia em um primeiro momento num outro setor que não o da Linguística propriamente dita; mas as análises e descrições de línguas realizadas por Boas e Sapir se fizeram voltadas para uma abordagem linguística e estrutural, considerando que "as línguas, afinal de contas, são *estruturas* históricas excessivamente complexas" (SAPIR, 1980[1921], p. 110, grifo nosso).

Desse cenário, de forma expressiva - e já em uma abordagem essencialmente linguística - também faz parte Leonard Bloomfield como um dos pioneiros do movimento; o linguista foi o precursor do princípio da distribuição<sup>33</sup>, que ficou sendo uma das características mais salientes da técnica linguística norte-americana, e que Zellig Harris vai explorar largamente em seu trabalho de descrição (HARRIS, 1969[1951]). Sobre a distribuição praticada por Bloomfield, Câmara Jr. diz:

O termo 'morfema' já era um tanto antigo e aparece em Brugmann, por exemplo, como equivalente de forma linguística. Mas Bloomfield deu-lhe a

<sup>32</sup> "A etnologia norte-americana [...] superou rápida e definitivamente essa fase[comparatista] graças a Franz Boas, que substituiu o Major Powell no trabalho de levantamento e estudo das línguas indígenas. Devemos considerá-lo o iniciador do movimento de renovação de princípios e métodos a que se acostumou chamar a escola linguística norte-americana."(CÂMARA JR. J. M. , apendice a SAPIR, E. 1980[1921], p. 185).

<sup>33</sup> Princípio de descrição linguística em que observa-se a limitação de ocorrência de um constituinte x em determinados ambientes (fonêmicos ou frasais), e apenas neles.

significação estrita de forma indivisível, ou mínima, e pô-lo em paralelismo com “fonema” e toda uma série de novos termos com o sufixo -ema, para designar os elementos na língua últimos e distintivos. [...] Para Bloomfield, como para a escola norte-americana em geral, o conjunto total de morfemas constitui o léxico da língua (CÂMARA JR. apêndice a SAPIR, 1980[1921], p. 197)

A escola de Praga emprestou suas características gerais aos estudos morfofonêmicos norte-americanos. Mas estes adquiriram aspectos próprios, como por exemplo, a teoria distribucional, que é desenvolvida também por Zellig Harris. No *Methods in Structural Linguistics*, que teve sua primeira edição em 1951, o linguista aplicou mais ampla e coerentemente o critério da distribuição (HARRIS, 1969[1951]), para delimitar e definir os elementos vocais distintivos e para os quais Bloomfield estabelecera definitivamente, na Linguística norte-americana, a designação de “fonemas”. (BLOOMFIELD, 1978[1926]).

A teoria das junturas, desenvolvida por Harris, sob os auspícios de Bloomfield, também configura uma especificidade da Linguística norte-americana. Nas junturas, consideram-se os fatos de ligação de morfemas e se distingue entre uma integração fonética absoluta, em que foneticamente as duas formas passam a constituir uma só.

Nesse cenário, a Linguística norte-americana se constrói. Aqui, Bloomfield adaptou o pensamento de Sapir aos princípios da filosofia behaviorista, criando o que ele próprio denominou a teoria mecanicista, que faz abstração da ‘hipótese’ da mente na interpretação do fato linguístico (BLOOMFIELD, 1979[1933]).

Entre Sapir e Bloomfield há uma relação semelhante à que se depreende na ciência francesa entre Ferdinand de Saussure e Antoine Meillet. Num e noutro caso, tem-se, ao lado do pioneiro genial, o consolidador extraordinariamente dotado, que dos pontos culminantes de uma concepção vigorosa e nova, mas ainda um tanto fluido, tira um corpo de doutrina coeso que a um tempo a precisa e a restringe (CÂMARA JR. prefácio a SAPIR, 1980[1921], p. 3).

De fato, desde 1926, Bloomfield (1978[1926]) já desenvolvera seu *Conjunto de Postulados para a Ciência da Linguagem*, em que desdobra uma série de regras, abordando as noções que envolviam o comportamento dos componentes da linguagem, como os fonemas, morfemas, sememas, forma mínima, formas livre e presa, dentre outros.

A abordagem linguística em moldes matemáticos também se fez produtiva nos Estados Unidos, e teve em Bloomfield e Harris seus representantes de maior expressão. Os mesmos ‘postulados’ Bloomfieldianos demonstram o início desse pensamento matemático

na forma de tratar a linguagem. A herança de abordagem matemática, aliás, se encontra no próprio Sapir (1980[1921], p. 28-32), quando demonstra a simbolização de termos linguísticos através de fórmulas algébricas em que atribui valores a cada componente.

Zellig Harris, mais tarde, retoma e desenvolve de forma específica e minuciosa essas funções matemáticas da linguagem; esse é um aspecto do seu trabalho que iremos abordar mais adiante neste estudo.

## 2.1. Origens e seqüências das bases teóricas de Harris

O estágio atual em que se encontram os estudos lingüísticos, acreditamos, deve-se a um percurso traçado a partir da guinada estrutural que teve lugar em meados do século XX. Em verdade, seus alicerces já estavam formados no início do século, como é possível verificar pelos estudos saussurianos, que datam de 1916. A partir de então, o que passou a caracterizar o sentido de lingüística foi o status - ou pelo menos, a pretensão de status – de ciência. Autônoma e coerente. E assim, ser científico ligava-se de maneira direta ao pensamento indutivo, que permeava o fazer ciência na época.

A história como perspectiva necessária e a sucessividade como princípio de explicação, a divisão da língua em elementos isolados e a pesquisa de leis de evolução próprias a cada um deles: esses eram os caracteres dominantes da doutrina lingüística.[...] A novidade do enfoque saussuriano, que foi um dos que agiram mais profundamente, consistiu em tomar consciência de que a linguagem em si mesma não comporta nenhuma outra dimensão histórica, de que é sincronia e estrutura, e de que só funciona em virtude da sua natureza simbólica (BENVENISTE, 1995[1966], p. 05).

Dentre os lingüistas que se ocuparam dessa cientificidade da abordagem lingüística está Harris. Em seu *Structural Linguistics* (1969[1951]) coloca o método estrutural de lingüística descritiva como sendo direcionado a pessoas interessadas no caráter científico da lingüística (cf. prefácio). Junto Com outros linguistas do período, ele salientou a importância da estrutura sincrônica e do método estrutural. Enfatizou a necessidade de se organizar os dados pelo seu lugar na estrutura, de forma que essa estrutura pudesse ser descrita somente em termos de diferenças formais de suas unidades e suas relações pela observação, sem se embasar, para os estudos linguísticos, em categorias estabelecidas *a priori* (MATTHEWS, 1999).

Harris teve alguns personagens como predecessores de seus estudos, outros como contemporâneos e, certamente, outros e vários sucessores. A descrição estrutural da língua como vista em Harris relaciona-se antes de tudo ao modelo antropológico de Boas. No abandono às teorias evolucionistas, a proposta deste continha o embrião do que mais tarde seria o cerne da lingüística descritiva, o que é claramente perceptível em seu trabalho de descrição de línguas da costa do Pacífico Norte.

Nesses estudos, há a preocupação em se verificar as características estruturais das línguas e dialetos, muitos deles com o objetivo de descrever suas gramáticas (cf. BOAS, 2004[1883-1911] p. 149). Ao estabelecer a classificação dessas línguas, o antropólogo procurava perceber características comuns/diferentes às línguas/dialetos que pesquisava, tanto no aspecto fonético e morfológico, quanto no sintático.

A abordagem dos sons lingüísticos e seus alternantes já estava impressa em seu trabalho. Fazia parte de seu método a realização de estudos fonéticos, comparando como determinados sons de uma língua eram percebidos por falantes de outra língua, concluindo que cada um percebe os sons desconhecidos por meio dos sons de sua própria língua, através da analogia. A teoria de sons alternantes<sup>34</sup>, que é apropriada também por Harris (2002b; 1969[1951]), se diferenciou da de Boas, pois para este não há sons alternantes. Existem percepções alternantes do mesmo som. Analisando diferenças de audição em relação a um mesmo som, enquanto articulado por um falante de outra sociedade, Boas concluiu que essas diferenças não se deviam a causas físicas, mas à percepção diferencial do ouvinte em relação aos sons a que este estava acostumado (BOAS, 2004[1883-1911] p. 102-103).

A tarefa de Harris relativa a desenvolver a teoria de vizinhanças fonêmicas<sup>35</sup> foi antecedida por Boas, que já se ocupava de analisar esses ambientes de forma a perceber que “os sons precedentes e subsequentes, além de muitas outras circunstâncias, exercem certa influência sobre o som que pretendemos produzir.” (BOAS, 2004[1883-1911] p. 99).

<sup>34</sup> Harris segue a linha de seu professor Bloomfield, no que se refere à alternância de sons. Ele reconhece a ocorrência de fonemas que têm ocorrência em ambientes similares (como em /ekə'namiks/ ou /iykə'namiks/ para *economics*) e propõe para seu tratamento métodos de agrupamento, como o das juntas. (cf. HARRIS 1951, p. 125-149; 197-218)

<sup>35</sup> O desenvolvimento do conceito de ambiente, que num plano mais elementar, se refere aos fonemas, -e depois aos morfemas, expressões, chegando até às construções textuais - envolve o de vizinhanças, 'neighborhood' (HARRIS 1951, p. 15) que se refere à ocorrência fonêmica nos diferentes ambientes para a qual os sons antecedentes/subseqüentes são determinantes. (cf. HARRIS 1951, p. 59-68;) No Inglês, uma ocorrência de /ð/ está condicionada ao posicionamento de um segmento vocálico na casa imediatamente à direita, como em *than*. Em Português, a ocorrência de /t/ só se dá antes da vogal i, como em *tia*, enquanto que em ambientes anteriores aos outros segmentos vocálicos, só ocorre /t/. Da mesma forma, a ocorrência de morfemas está condicionada a vizinhanças morfêmicas (cf. HARRIS 1951, p. 156-196).

Ademais, a própria abordagem sobre a compreensão dos indivíduos dentro da sociedade à qual pertencem - sendo possível apenas com base nas inter-relações desses indivíduos constituintes dessa sociedade, - em si se faz um modelo para o estudo de diferentes instituições das sociedades, dentre elas, as línguas.

Harris considera, entretanto, o trabalho de Edward Sapir e de Leonard Bloomfield como principais âncoras para o desenvolvimento do seu próprio. Na introdução a uma de suas primeiras obras, confere a eles os créditos por esse mesmo trabalho.

Os procedimentos de análise discutidos aqui são o produto e desenvolvimento do trabalho de lingüistas de todo o mundo, para cujas investigações as escassas referências citadas aqui são um roteiro inadequado. *Este livro deve-se em sua maior parte, porém, ao trabalho e amizade de Edward Sapir e de Leonard Bloomfield, e particularmente ao seu último livro, Language*<sup>36</sup> (HARRIS, 1969[1951], prefácio. Grifo nosso).

Sapir realizou, como Boas, seu mentor, estudos descritivos de línguas nativas nos Estados Unidos.<sup>37</sup> De muita relevância para o desenvolvimento das teorias de Harris foram esses estudos descritivos. Dentre elas, a da definição dos sistemas fonéticos e também da constituição de padrões para a formação gramatical da língua (SAPIR,1980[1921]; HARRIS 1969[1951]; 1968;1981;1982;1988;1991). Em Sapir estão presentes muitos dos fundamentos referentes a noções que Harris desenvolveu posteriormente. Dentre elas, a de limitações em relação à composição dos constituintes da língua em seu meio de realização, a linguagem, bem como as particularidades dessas composições nas diferentes línguas. Numa comparação entre a arte e o fazer linguístico, afirma:

As possibilidades da expressão individual são infinitas, e a linguagem, em particular, é o mais fluido dos meios. *Tem de haver, contudo, certa limitação a essa liberdade, certa resistência ao meio.* Na grande arte, há a ilusão da liberdade absoluta. O constrangimento formal imposto pela matéria-prima – tinta, claro-escuro, mármore, notas de piano, ou seja o que for -, passa despercebido; é como se houvesse uma área ilimitada de separação entre a mais plena utilização da forma, por parte do artista, e o máximo de que é capaz o material utilizado . [...] A linguagem aqui é esse meio, como o mármore, o bronze ou a argila são a matéria-prima do escultor. *Como cada língua tem peculiaridades distintas, as limitações – e as possibilidades – formais inatas de*

---

<sup>36</sup> “The procedures of analysis discussed here are the product and outgrowth of the work of linguistics throughout the world, to whose investigations the meager references cited here are an inadequate guide. This book owes most, however, to the work and friendship of Edward Sapir and of Leonard Bloomfield, and particularly to the latter’s book *Language*.”

<sup>37</sup> Línguas nativas como o *nass*, o *chinuk*, *nutka*, *iroquês*, o *paiúte*, o *yana*, o *athabaskan*, o *yokut* e outras foram pesquisadas por Sapir, numa busca linguística e ao mesmo tempo, antropológica (SAPIR, 1980[1921]).

*cada literatura nunca são as mesmas que as de outra* (SAPIR, 1980[1921]. p. 175 Grifo nosso).

Salienta também as comparações entre estruturas lingüísticas enfatizando o traçado de paralelos úteis para se compreender essas diferenciações (SAPIR, 1980[1921] p. 58-67; 100-114;).

Os padrões de ocorrência de elementos fonológicos nas línguas também foram objetos de estudo para Sapir (1980[1921] p. 41-49), de forma que as descrições feitas por ele também se fizeram de muita importância para o despontar dos estudos descritivos de Harris, no que se refere a esse recorte da língua (cf. HARRIS, 1969[1951], p. 7).

Outra noção harrissiana presente já em Sapir são as representações em fórmulas no que se refere a elementos da língua (e. g. palavras, sentenças). Em exemplos de fórmulas simples para a representação algébrica de termos lingüísticos, Sapir idealiza a forma A + (b) como representativa de um elemento radical acompanhado de seu afixo (e. g. *singer*)<sup>38</sup>. Nesse caso, o elemento radical necessariamente se identifica com uma palavra. Em outros casos, em que esse constituinte é uma forma presa, passa a ter uma representação como em *hort-us*, do latim. (A) + (b); ou ainda (b) + A + (c) + (d) como fórmula representativa de palavras como *unthinkingly*, formada por radical e vários afixos.

Assim, pelo direcionamento que tomou a abordagem de Sapir, a língua foi por ele definida como forma e, naturalmente, teve em Harris um de seus herdeiros intelectuais, que deu seqüência ao desenvolvimento dessas teorizações.

Em verdade, uma seqüência de teorizações surgiram sobre a língua enquanto sistema formado por relações de dependência e também como uma combinatória fechada representável pelo sistema algébrico. Para Hjelmslev,<sup>39</sup> contemporâneo a HARRIS, a relação de interdependência se faz saliente e é fundamental para se analisar determinado objeto que se situa em um sistema, o que é aplicável aos objetos do sistema língua:

...tanto quanto suas partes, o objeto examinado só existe em virtude desses relacionamentos ou dessas dependências; a totalidade do objeto examinado é apenas a soma dessas dependências, e cada uma de suas partes define-se apenas pelos relacionamentos que existem 1) entre ela e outras partes

<sup>38</sup> Nesse caso, Sapir sugere a maiúscula A para representar o radical, e a letra minúscula entre parênteses representando o afixo, que não pode aparecer separadamente, é uma forma presa. Da mesma forma, os parênteses também indicam um radical que se faz representar por uma forma presa.

<sup>39</sup> Hjelmslev foi propositador do termo Glossemática para designar o estudo e a classificação dos glossemas, as menores unidades lingüísticas que podem servir de suporte a uma significação no estudo das formas de expressão de conteúdo no campo da semiótica.

coordenadas, 2) entre a totalidade e as partes do grau seguinte, 3) entre o conjunto dos relacionamentos e das dependências e essas partes.[...] A análise, em sua definição formal, será portanto a descrição de um objeto através das dependências homogêneas de outros objetos em relação ao primeiro e das dependências entre eles reciprocamente. Denominar-se-á classe ao objeto submetido à análise, e componentes dessa classe os objetos que são registrados por uma única análise como dependendo uns dos outros e da classe de modo homogêneo (HJELMSLEV, 1975[1971], p. 28; 34).

Aliás, essas relações de interdependência são consideradas por ele como oriundas do próprio sistema lingüístico (HJELMSLEV, 1975[1971], p. 27-45). Na sua concepção da língua como uma combinatória, ele atribuiu um valor importante a algumas propriedades formais das relações que compõem essa combinatória. As representações lingüísticas, assim, são salientadas como uma espécie de álgebra, em que as relações formais entre os constituintes são importantes para a definição de estrutura. Seria isso em nome do rigor e/ou da elegância das soluções apresentadas dessa forma? Se não em nome da elegância, certo é que o rigor se fazia necessário e presente nessa abordagem.

Em virtude de sua adequação, a teoria da linguagem realiza um trabalho empírico; em virtude de seu caráter arbitrário, realiza um trabalho de cálculo. Baseando-se em certos fatos da experiência – necessariamente limitados, embora seja útil que sejam tão variados quanto possível – o teórico empreende, num campo preciso, o cálculo de todas as possibilidades. Ele baliza arbitrariamente esse campo isolando propriedades comuns a todos os objetos a respeito dos quais se está de acordo em denominá-los de línguas, a fim de, em seguida, generalizar essas propriedades e estabelecê-las por definição. A partir desse momento ele decidiu – de um modo arbitrário mas adequado – quais são os objetos aos quais a teoria pode ser aplicada e quais aqueles aos quais ele não o pode ser. Todos os objetos assim definidos são então submetidos a um cálculo geral que prevê todos os casos concebíveis (HJELMSLEV, 1975[1971], p. 21).

A linguagem seria então para Hjelmslev possível de ser descrita por meio de procedimentos que exaurissem de forma não contraditória, seus objetos. Esse procedimento é aplicado por ele de forma mais direta à análise de textos. Nota-se uma similaridade teórica entre a sua proposta de descrição textual e a de Harris (cf. HARRIS, 1970[1952]). Nessa proposta, há um interesse pela estruturação dos constituintes do texto, que

...dizem respeito também ao sistema ou língua a partir da qual se elabora a estrutura de todos os textos de uma mesma suposta natureza, e que nos permite construir novos textos. Graças aos conhecimentos lingüísticos assim obtidos poderemos elaborar, para uma mesma língua, todos os textos concebíveis ou teoricamente possíveis.[...] A análise consiste, portanto,

efetivamente, no registro de certas dependências ou certos relacionamentos entre termos que, conforme o uso consagrado, chamaremos de partes do texto, e que existem exatamente em virtude desses relacionamentos e unicamente em virtude deles (HJELMSLEV, 1975[1971], p. 20; p.33).

No *Structural Linguistics* (1969[1951]), obra em que é dada uma especial atenção às descrições de ordem fonêmica, Harris se refere particularmente às contribuições de Bloomfield, seu professor, para o desenvolvimento desse trabalho. O princípio da classificação e da distribuição de constituintes desenvolvido por Bloomfield em seu *Conjunto de Postulados para a Ciência da Linguagem* (1978[1926]) e no *Language* (1979[1933]) evoluiu dentro da obra de Harris (1969[1951]), com o diferencial que Harris desenvolve uma didática específica, em que propõe a representação desses constituintes em fórmulas, enquanto em Bloomfield temos representações convencionais em símbolos lingüísticos aplicáveis a cada caso.

Noções importantes para a teorização de Harris, como: morfemas e suas características distribucionais (cf. HARRIS, 1969[1951]; 1990[1936]), alternância de sons, e também como as distinções fonêmicas (e.g. variantes livres) foram exploradas antes por Bloomfield, (cf. BLOOMFIELD, 1978[1926]). A abordagem bloomfieldiana sobre os constituintes imediatos foi muito importante, pois ampliou as possibilidades de análise de Harris sobre as seqüências fonêmicas (cf. HARRIS, 1969[1951], p.262-298), além de terem sido aplicadas por este também na morfologia e na sintaxe (HARRIS, 2002b, p. 03)

A metodologia de Bloomfield na descrição encontrou eco em Harris. Para este, além dos dados lingüísticos em relação aos sons, é necessário saber sobre a impressão do ouvinte, por conta da complexidade do discurso; dessa forma, sua lingüística descritiva se habilita a explorar os elementos discretos. (HARRIS, 1969[1951], p. 20)

Harris deixa entrever que não se prende à abordagem mecanicista de seu mestre, em relação ao estímulo-resposta na linguagem, mesmo porque Harris trata a maior parte do tempo da língua enquanto forma. Entretanto, o que ele nomeia ‘comportamento discursivo’ é para ele um fator importante ao se investigar traços lingüísticos. Em verdade, o comportamento discursivo seria uma ferramenta para que se pudesse perceber relações entre elementos lingüísticos e seus traços.

Em investigações em lingüística descritiva, os elementos lingüísticos são associados com traços particulares do comportamento discursivo em questão, e as relações entre esses elementos são estudadas.[...] Cada elemento é

identificado com alguns traços de discurso na língua em questão<sup>40</sup> (HARRIS, 1969[1951], p. 16).

Harris confere especial crédito aos seus antecessores Sapir, Saussure<sup>41</sup> e a Bloomfield, por realizarem investigações no campo morfológico e no que se refere às relações de classe, entre morfemas e classes de morfemas. Reitera que essas investigações ficavam a desejar em relação às do campo fonológico, e poucos lingüistas se davam conta disso. (HARRIS, 1969[1951] p.197; 311). Em suma, a contribuição destes, bem como a de Boas, foram fundamentais para a formulação e prosseguimento de suas teorias relacionadas às análises fonêmica, morfológica e também sintática:

Dentro da lingüística, o sucesso da análise fonêmica de Saussure mostrou a utilidade da variação complementar e livre como uma base para definir entidades mais irrestritas, tidas como possuindo vários valores alternados que eram usualmente mais restritos. Em morfologia e sintaxe, o método ‘distribucional’ seguido por Franz Boas, e mais explicitamente por Edward Sapir e Leonard Bloomfield, analisou da mesma forma as ocorrências e combinações de elementos gramaticais em ambientes particulares de outros elementos. Eu penso, e eu fico satisfeito em pensar, que a influência intelectual e pessoal de Sapir e de Bloomfield matiza todo o trabalho esboçado abaixo<sup>42</sup> (HARRIS, 2002b, p. 1-2).

Harris também se refere a outros estudiosos em que se embasou para o desenvolvimento de suas teorias. Para aprofundar suas teorizações no que se refere à entonação no Inglês, realizou uma busca sobre diferentes estudos relativos à análise de tonicidade de segmentos discursivos, como em Rullon X. Wells, Kenneth L. Pike, Stanley Newman e outros.(cf HARRIS, 1969[1951] p.46 nota 2) O trabalho de Bernard Bloch foi importante na seqüência de estudos de Harris sobre os fonemas e suas particularidades de status e de ocorrência (HARRIS, 1969[1951], p. 65). Na verdade, a referência aos estudos de George Tragger e B. Bloch são freqüentes no *Structural Linguistics* (1969[1951]),

<sup>40</sup> “In investigations in descriptive linguistics, linguistic elements are associated with particular features of the speech behavior in question, and the relations among these elements are studied. [...] Each element is identified with some features of speech in the language in question”.

<sup>41</sup> O trabalho anterior de Saussure, foi especialmente importante no desenvolvimento das considerações de Harris sobre aspectos como: duração de fonemas, sucessão e simultaneidade como relação entre elementos lingüísticos (cf. HARRIS 1951, p. 144 nota 41)

<sup>42</sup> Within linguistics, the success of de Saussure’s phonemic analysis showed the usefulness of complementary and free variation as a basis for defining more unrestricted entities, stated as having various alternate values that were usually more restricted. In morphology and syntax, the ‘distributional’ method followed by Franz Boas, and more explicitly by Edward Sapir and Leonard Bloomfield, analyzed likewise the occurrence and combination of grammatical elements in the particular environments of other elements. I think, and I am glad to think, that the intellectual and personal influence of Sapir and of Bloomfield collors the whole of the work that is surveyed below.

inclusive no sentido de que as investigações fonêmicas dos autores supra citados foram importantes para o desenvolvimento da teoria das junturas<sup>43</sup> e dos limites morfêmicos. Para os estudos que envolviam a fonologia de línguas indígenas do Pacífico Norte, como o *Navaho* e o *Yokuts*, foram importantes as análises de Harry Hoijer (HARRIS, 1969[1951], p. 130, nota 13) e Stanley Newman. Este último foi importante para desenvolver estudos fonêmicos comparativos entre o Inglês e o *Yokuts*, inclusive para determinar fórmulas e limites - tanto fonêmicos quanto morfêmicos (HARRIS, 1969[1951], p. 151 nota 3).

Em Harris (1969[1951], p. 279; 2002b), temos ainda que os pressupostos da Lógica e da Matemática foram para ele fundamentais para formular seus métodos.

O alicerce para o trabalho veio largamente dos fundamentos da matemática e da lógica, e a análise dos formalismos; isto foi relevante para a língua porque em todos esses sistemas havia sentenças (proposições, fórmulas) com estrutura parcialmente similar (sintaxe). Mais especificamente, havia o então construtivismo corrente: nas análises pelos intuicionistas em matemática (L. E. J. Brouwer), na teoria de tipos de Russell, no trabalho de Emil Post, e no Procedimento da Turing Machine. Mais tarde, eu considerei que eu tinha suporte filosófico na abordagem construtivista (nominalista) da Estrutura da Aparência de Nelson Goodman. Retornando à primeira base, havia também o desenvolvimento de métodos recursivos por Gödel, Tarski, e outros; e na lógica havia o Cálculo de Sentenças de J. Lukasiewicz<sup>44</sup>, a Gramática Categórica de S. Lesniewski, e a sintaxe da lógica na Lógica Matemática de W. V. O. Quine de 1940.<sup>44</sup> (HARRIS, 2002b, p. 01)

O predicado logocêntrico preconizado pela lógica ligada a Russell é compreendido e abordado por Harris em termos de distribuição, ou como se queira, uma abordagem combinatória na formulação da estrutura da língua. Essa abordagem vem representar a formulação das restrições na combinação de elementos linguísticos que ocorrem no uso,

---

<sup>43</sup> A teorização sobre as junturas se originou da necessidade de se simplificar os estudos sobre fonemicização no sentido de se juntar dois grupos representantes de fonemas distintos em apenas um grupo, de forma que se façam complementares um ao outro e não mais dois sons representados de duas formas diferentes. Como na representação de minus /ay/ e slyness /Ay/ e playful /ey/ e tray-ful /Ey/. De forma que a duração diferenciada dos pares /ay/ e /Ay/ possa ser representada por uma nova e comum representação fonêmica com “um novo elemento fonêmico, chamado juntura, /-/” (HARRIS 1951, p. 79). Assim, a representação dos dois sons se faria com o símbolo /ay-/, que seria a representação fonêmica comum para os dois sons; em vez de dois: /ay/ e /Ay/, tem-se apenas /ay-/, que assinala a ocorrência tanto do segmento mais curto quanto do mais longo.

<sup>44</sup> The background for the work came largely from the foundations of mathematics and logic, and the analysis came largely from the foundations of mathematics and logic, and the analysis of formalisms; this was relevant to language because in all of these systems there were sentences (propositions, formulas) with partially similar structure (syntax). More specifically, there was the then current constructivism: in the criticisms by the intuitionists in mathematics (L. E. J. Brouwer), in Russell’s theory of types, in the work of Emil Post, and in the Turing Machine procedure. Later, I considered that I had Philosophical support in the constructivist (nominalist) approach of Nelson Goodman’s *The Structure of Appearance*. Going back to the early background, there was also the development of recursive methods by Gödel, Tarski, and others; and in logic there was J. Lukasiewicz’ *Sentential Calculus*, S. Lesniewski’s *Categorical Grammar*, and the syntax of logic in W. V. O. Quine’s *Mathematical Logic* of 1940.

tanto em se tratando da língua como um todo, como das sublinguagens. (cf. RYCKMAN, 2002, p. 22). Aqui reside, em essência, a característica matemática da língua. Essa percepção matemática já havia sido explicitada, de alguma forma, por Saussure, quando introduziu o termo ‘código’ na Linguística estrutural, quando se referiu à língua como um código utilizado em combinações de sons para expressar o pensamento. (cf. SAUSSURE, 1994[1916], p. 31)

Seria esse pendor em se representar componentes da língua, relacionando suas construções a axiomas matemáticos interessante para que a nascente Linguística se configurasse como ciência ao incluir-se de alguma forma na área das exatas? Filósofos ligados à Matemática e à Lógica, como Russell e Frege, certamente concordariam, pois temos nos dois iniciativas de associar sistemas notacionais às construções da língua. Lacey (1978) aponta que “O momento mais célebre e bem sucedido desse trabalho [o trabalho epistemológico de Russel] foi, sem dúvida, a construção da teoria aritmética a partir de elementos exclusivamente lógicos, momento expresso com perfeição nos *Principia Mathematica*.” (p. XI)

Frege (1997[1879]), por sua vez, pensou desenvolver um mecanismo que reduzisse a aritmética à lógica. Isso seria para criar uma linguagem de símbolos sem as imperfeições da linguagem natural; essa formalização da linguagem seria aplicável ao uso científico, pois nas ciências é necessário haver exatidão. O sistema notacional então proporcionaria a transmissão do que chamou de “pensamento puro”. Interessa notar que posteriormente chegou à conclusão de que esse pensamento puro se mostraria um equívoco, que “surge da imperfeição da linguagem, da qual mesmo a linguagem simbólica da análise matemática não está totalmente livre.” (FREGE, 1978, p. 76) Os lógico-filósofos, ligados que eram à matemática, buscaram encontrar nas construções da língua uma forma de exemplificar teorias que ligavam a lógica à matemática. Ao que parece, Harris fez o percurso inverso. Para sua exposição teórica da língua, buscou formas matemáticas de representação embasado naquelas dos lógicos. O pensamento de Harris se diferencia do dos lógicos na medida em que tenta aplicar essa representação na linguagem ordinária.

Nos lógicos também está a base para o rompimento com a noção de sujeito e predicado e o conseqüente desenvolvimento da noção de operador e argumento. Tanto Frege quanto Russell perceberam que, para a lógica, as categorias de sujeito e predicado eram irrelevantes. Frege declara, ao discorrer sobre o sentido e a referência que

Por isto, as palavras ‘relação entre sujeito e predicado’ designam duas relações totalmente diversas, conforme seja o sujeito um objeto, ou um conceito. Assim sendo, o melhor seria eliminar totalmente da Lógica as palavras ‘sujeito’ e ‘predicado’, posto que sempre nos levam a confundir duas relações totalmente diferentes, a de cair um objeto sob um conceito e a da subordinação de um conceito sob outro conceito (FREGE, 1978, p. 110).

De forma que os próprios problemas oriundos do conceito de sujeito e predicado fizeram com que houvesse na lógica – principalmente na de Russell e Frege - a pretensão da reformulação desse conceito para o de operadores e argumentos. (cf. LACEY, 1978).

Harris conferiu um método apropriado à análise linguística ao desenvolver essa noção. Na verdade, foi em busca de um método que propiciasse a transformação na análise de sentenças que Harris terminou por chegar a um resultado inesperado – uma metodologia para explicitar as sentenças metalingüísticas. (cf. HARRIS, 2002b, p. 3) e também para sua definição do conceito de metalinguagem como sendo parte da língua natural, o que resultou na teoria de operador- argumento (HARRIS, 2002b, p. 1-15).

No desenvolvimento da teoria de transformação Harris declara ter encontrado os pressupostos básicos também na Lógica, em Frege e Quine. No esforço para estabelecer um método transformacional que desse conta das reduções nas sentenças e também das construções parafrásticas, Harris compartilhou suas buscas e questionamentos com estudiosos como Piaget, Y.Bar-Hillel, M. P. Schützenberger, Maurice Gross e André Lentin. A Henry Hizz credita as maiores contribuições no que se refere a estabelecer as sentenças zero e também detalhes mais profundos quanto ao método transformacional. Neste particular, houve discussões importantes com Noam Chomsky, que com a abordagem gerativa da teoria transformacional, contribuiu consideravelmente (cf. HARRIS, 2002b).

Outro conceito interessante que está impresso na obra de Harris é o de informação na língua (HARRIS, 1969[1951]; 1968;1988;1991). E para ele, os constituintes lingüísticos, desde os fonemas, até a análise combinatória dos operadores e argumentos é um caminho natural para se chegar à quantidade de informação presente nas construções lingüísticas. Nesse sentido, informações relevantes sobre uma língua podem ser obtidas mais completamente e de forma interessante através de uma listagem de cada morfema, cada um com uma lista completa de propriedades – suas possibilidades de combinação com outros morfemas e suas formas nessas combinações. (HARRIS, 2002b, p. 9; 1970[1946] p. 101). A constituição de inventários extensos para demonstrar a

aplicabilidade de regras sintáticas feita por Maurice Gross (1986a; 1986b; 1986c) foi considerada produtiva por Harris (2002) para a sistematização de informação na língua.

Esses inventários foram organizados com base na teoria de estudos orientados pelos princípios do Léxico-Gramática. Tais estudos viabilizam a compreensão do fenômeno de combinação entre os componentes lingüísticos para a formação de frases simples, ou seja, frases que contêm apenas um predicado semântico. A teoria do Léxico Gramática é definida pelos mecanismos de seleção e distribuição, intermediados pela compatibilidade semântica, de maneira que os componentes de uma frase não sejam percebidos apenas por mecanismos sintáticos, mas também por regras de interpretação (GROSS, 1981).

Os estudos da área da lexicografia visam à elaboração ou descrição do Léxico gramática de uma língua dada, o que se relaciona diretamente ao estabelecimento de inventários lexicais.

Harris, por sua vez, salienta que uma lista exata de morfemas é evidentemente necessária para qualquer descrição da língua (cf. HARRIS, 2002b, p. 101). Nessa direção, o próprio Harris tratou os morfemas do Inglês (HARRIS, 1970[1946]), classificando-os no que Gross (1981) nomeou mais tarde classes de objeto, em que os diferentes tipos de morfemas e palavras são classificados conforme suas características distribucionais no estabelecimento de critérios de agrupamento em classes desses morfemas e palavras, numa investigação tipológica. De acordo com esse agrupamento, foram especificados em quais ambientes frasais as classes têm ocorrência, colocando em uma classe X, Y ou Z... todos os morfemas que têm distribuição similar, num critério que se fez essencial para a teoria de operador/argumento desenvolvida por Harris. Nesse sentido, interessa notar que a teoria de operadores e argumentos está imbricada na de informação na língua.

Pela natureza de traços estruturais que juntos criam a língua, pode-se ver um instrumento que se desenvolve naturalmente para transmitir informação. O condicionamento da entrada é uma forma adequada para a predicação que é vista[...] como um mecanismo fonte de informação para toda a língua; se uma classe de palavras Y só é utilizada desde que seja dita uma palavra de classe X, então espera-se que Y exista em sentenças apenas como algo existente em X – como uma estrutura atrelada para predicação Y de X<sup>45</sup> (HARRIS, 1981[1978], p. 410).

---

<sup>45</sup> From the nature of the separate structural features which together create language, one can see a naturally growing instrument from transmitting information. The entry requirement is an adequate form for predication, which is seen[...] to be the source information-device for all language; if a word of class Y is said only provided one has said a word of class X, then Y exists in sentences only as something said on X – as close a structure for predicating Y of X as one might expect.

Estudos atuais dão seqüência e demonstram a produtividade dessa aplicação teórica de Harris. No Brasil, há estudos descritivos para os quais a noção de operador-argumento são essenciais. Um deles é um estudo sobre expressões cristalizadas do português do Brasil (VALE, 2001). Nesse estudo as possibilidades de combinação entre um e outro é explorada com a finalidade de analisar expressões cristalizadas verbais; partindo dessa descrição, o linguista sistematiza a informação sintática e semântica contida na expressão, o que as faz serem reconhecidas pelos usuários como expressões próprias de sua língua.

Há diversos outros estudos conhecidos que têm como base a teoria de sintaxe de operadores de Harris: dentre eles, o de Gross (1975), Boons et al (1976) e o de Ranchhod, (1990), da Universidade de Lisboa. Ranchhod se vale dos princípios do léxico-gramática para um estudo aprofundado da sintaxe dos predicados nominais com verbo ‘estar’. A gramática de operadores de Harris, sendo progenitora da teoria do léxico-gramática, alicerçou de forma direta esses estudos.

### **2.1.1. Estrutura**

A par da afirmação de que o mundo dos conceitos “se reflete e sistematiza na estrutura linguística” (SAPIR, 1980[1921], p. 71), buscaremos aqui apresentar alguns conceitos da lingüística estrutural que se fazem presentes na obra de Harris e que serão mais detalhadamente discutidas em sessões posteriores a esta. Conceitos que foram também abordados em outro momento, por outros autores, mas interessa aqui observar em relação a eles a abordagem de Z. Harris e o tratamento que ele confere. A maneira como o linguista percebe o que seja a estrutura de uma dada língua, em quê há diálogo com seus contemporâneos e com a atualidade, é algo que interessa discutir e perceber neste espaço. Tal discussão certamente possibilitará compreender melhor a concepção de estrutura do autor, no desenrolar de seu trabalho.

O descritivismo lingüístico tem sido por vezes objeto de críticas provenientes de lingüistas que abordam outras correntes do pensamento. É de interesse pontuar que efetivamente, os “formalistas”, como são chamados os lingüistas que tratam da área estrutural, não se ocupam de desfiar a relação dos constituintes com seus significados, ou

entre a língua e seu meio (cf. NEVES, 1997, p. 41). Entende-se que, por ser esse um trabalho que requer um outro tipo de esforço, seria necessária uma intervenção diversa da que se propõe a análise formal. Isso, naturalmente, requereria um trabalho adicional, legado a outra(s) corrente(s), que, no caso, se consubstanciam no Funcionalismo, na Sociolingüística, Pragmática e assim por diante.

“A lingüística descritiva, como o termo tem sido usado, é um campo particular de investigação que se ocupa não de toda a atividade discursiva, mas das regularidades em certos traços do discurso”<sup>46</sup> (HARRIS, 1969[1951], p. 05). Harris define ainda o papel da lingüística descritiva em relação aos outros ramos da lingüística - e também às outras ciências - como sendo um que se faz importante no sentido de tornar claros os métodos de trabalho nesse campo, e ainda dar mais detalhes sobre outras línguas, do que a simples catalogação de sons e formas.

Um ponto que, segundo Harris, distingue a lingüística descritiva das outras áreas da lingüística é que os seus resultados se aplicam apenas a uma língua/dialeto particular (HARRIS, 1969[1951], p. 8 nota 6; SAPIR, 1980[1921], p.177), o que é diferente em outras linhas de pesquisa, (i. e. sociolingüística, funcionalismo), em que os resultados gerais se aplicam a diferentes línguas. Assim, a abordagem estrutural – como qualquer outra - representa a escolha de um ângulo pelo qual é possível individuar critérios de pertinência. De uma forma simples, para Harris, as diferentes teorias lingüísticas não têm de entrar em choque, mas serem complementares umas às outras (MATTHEWS, 1999).

Tem-se então o princípio da estrutura como objeto de estudo sendo atribuído a um pequeno grupo de lingüistas que pouco antes da década de 30, se propuseram a reagir contra a concepção exclusivamente histórica da língua, contra uma lingüística que dissociava a língua em partes isoladas e se ocupava apenas em seguir-lhes as transformações (cf. BENVENISTE, 1995[1966], p. 5).

Estrutura: derivado de *structura* em Latim, adquiriu no começo um sentido arquitetural. “... O termo assume então o sentido da descrição da maneira como as partes integrantes de um ser concreto organizam-se numa totalidade” (DOSSE, 1993, p. 15). É uma noção que se relaciona à organização no interior do sistema. O termo estrutura adquiriu o sentido de ser a forma de se estudar os fenômenos, em determinada época, não só em Linguística, mas em diversas áreas do conhecimento.

---

<sup>46</sup> “Descriptive linguistics, as the term has come to be used, is a particular field of inquiry wich deals not with the whole of speech activities, but with the regularities in certain features of speech.”

Ao estudar Zellig Sabetai Harris – lingüista estruturalista, parece ser necessária uma discussão sobre o termo estrutura, pois é ela o cerne de sua obra. A estrutura da língua seria algo ontológico ou metodológico? O linguista pesquisador procura e encontra na língua algo que organiza seu funcionamento ou estabelece um método de estudo e de busca no sistema lingüístico e procura aplicá-lo a tal sistema?

Eco (2001), através de uma série de textos analisados, chega a um modelo de estrutura como sendo em primeiro lugar, um sistema de diferenças, a que Saussure (1994[1916] certamente chamaria oposições. Outra característica apontada é a propriedade de “transponibilidade de fenômeno para fenômeno e de ordens de fenômenos para ordens de fenômenos diferentes”<sup>47</sup>(ECO, 2001, p. 258). É necessário ainda que seja um sistema regido por coesão interna, e que o acesso à observação seja possível tendo como parâmetro outro(s) sistema(s).

A partir de Saussure, tem-se a abordagem da língua como sistema organizado. Tal sistema, segundo esse ícone da lingüística moderna, se faz ao mesmo tempo uma “instituição atual” e um “produto do passado” (SAUSSURE, 1994[1916], p.16), pois, ao mesmo tempo em que é um aparato social já estabelecido, conhecido e utilizado com desenvoltura pelos falantes, é também objeto de atualizações e de evolução. É um sistema com regras próprias, leis que regem seu funcionamento, muitas vezes, difíceis de alcançar ou definir. Nesse sentido, coloca-se a importância dos estudos descritivos das estruturas das línguas. Pensamos ser necessário então procurar através de alguma discussão, caminhos para idéia de “estrutura”, dado que é um termo considerado polêmico por alguns estudiosos (ECO, 2001; BENVENISTE, 1995[1966]).

Segundo Eco (2001, p. 36), “uma estrutura é um modelo construído segundo certas operações simplificadoras que me permitem uniformar fenômenos diferentes com base num único ponto de vista.” Seria um modelo elaborado para poder nomear de maneira homogênea coisas diferentes. Na visão do autor, a partir de modelos diversos, para se estabelecer uma estrutura, há que se chegar a uma forma comum, através da abstração, de se demonstrar a disposição e/ou o funcionamento desses modelos.<sup>48</sup> Admitamos a estrutura como sendo um conjunto cujas partes estão relacionadas entre si, é um todo conexo. Assim, seria ela um objeto estruturado ou seria o conjunto de relações que estruturam o

---

<sup>47</sup> No caso da Lingüística, ocorreu a transposição do conceito de estrutura para outras áreas, como as Artes em geral, a Psicologia, Filosofia, a Antropologia e a Economia. (Dosse, 1993)

<sup>48</sup> Poderíamos a título de exemplo, construir em forma de esquema, um modelo comum representativo, ao mesmo tempo, do corpo feminino ideal e do instrumento musical violão.

objeto mas que podem dele ser abstraídas? Nesse caso, uma alternativa seria estabelecer a distinção entre a estrutura ontológica e a estrutura metodológica.

### **2.1.2. Ontologia x Metodologia em estudo de língua**

Partindo dos dois modos como Eco (2001) concebe a estrutura, a saber: 1) objeto organizado e 2) modelo generalizante, podemos apontar a estrutura em lingüística como sendo um modelo generalizante que se criou para explicar o objeto organizado língua. Os rumos que tomou o estruturalismo americano deixam entrever - como nas descrições de Bloomfield e Harris, por exemplo - a criação de metodologias próprias da língua para explicitar suas estruturas internas.

Na história do conceito de estrutura é legítimo falar de um termo que define um conjunto, as partes desse conjunto e as relações que unem essas partes “num todo formado de elementos solidários, de tal modo que cada um dependa dos demais e não possa ser o que é senão em virtude de sua relação com eles”. (ECO, 2001, p. 252). Em se tratando da Lingüística, ao que se nota, é que se trabalha com a metodologia para explicar uma estrutura latente, que se faria ontológica.

Entretanto, Eco (2001) e Dosse (1993) são claros ao dizer do vazio que se tem ao se estabelecer o método estrutural. Ou seja, na tentativa de se metodologizar uma estrutura ontológica, extingue-se o que é ontológico e em seu lugar fica só o que é metodológico, de forma que a estrutura real, se existe, a partir daí, morre. Nas palavras de Eco (2001, p. 323), “estrutura é aquilo que ainda não existe. Se existisse, se eu a tivesse individuado, teria entre as mãos apenas um momento intermédio da cadeia que me garante, abaixo daquela, uma estrutura mais elementar e oniexplicativa.” E continua, ao tratar de estabelecer uma metodologia, apontando que é necessário que

... me aproxime de um grupo de fenômenos para compreendê-los melhor mediante a predicação de homologias estruturais que me permitam estabelecer a correlação entre eles: e então a estrutura é um instrumento operacional visando ao discurso sobre o campo concreto dos fenômenos examinados (ECO, 2001, p. 322).

Mais adiante, salienta que o próprio fato de ser a estrutura um instrumento operacional, faz com que ela propicie a escolha de determinado campo de fenômenos com critérios de pertinência e pontos de vista já determinados, sendo que isso trará certos resultados. Segundo o autor, não cabe à teoria estrutural fundar critérios de avaliação pré-determinados. Dessa forma, o que ele chama de Ur-Código – que seria a Estrutura maior, estaria ausente nesse processo. E assim, afirma :

A Estrutura só se me revelará através de sua própria ausência progressiva. Tão logo houver reconhecido a ausência constitutiva do Ur-Código, terei que ter a coragem de afirmar que, enquanto ausente, a estrutura constitutiva de toda outra estrutura, não está estruturada. (ECO, 2001, p. 323)

Seguindo em seu raciocínio, lembra que, se sabemos que a estrutura é um modelo, conseqüentemente se saberá que ela não existe, ontologicamente falando.

Por outro lado, tanto Benveniste (1995[1966]) quanto Sapir (1980[1921]) e também Dosse – este através de afirmações sobre estudos lingüísticos consistentes - tornam possível a inferência de que pode existir sim, uma estrutura subjacente à língua. Ao tratar da estrutura de línguas, tanto Benveniste quanto Sapir discorrem sobre as propriedades estruturais delas de um modo amplo - tratam de combinações fonéticas, seleção, processos de composição frasal - e este último salienta, tanto em se tratando de capacidade de síntese como de classificação genética, semelhanças e diferenças possíveis de ocorrer. E trata do plano fundamental, do quadro determinado que se faz mais básico e mais profundo do que qualquer outra característica lingüística que se possa mencionar, em qualquer língua.

Vimos [...] haver umas que buscam síntese de estreita coesão ao passo que outras se contentam com a disposição mais analítica e fracionada de seus elementos, ou em uma aparecerem puras certas relações sintáticas que já em outra se combinam com certas noções que têm em si qualquer coisa de concreto, por mais abstratas que se nos apresentem elas na prática (SAPIR, 1980[1921], p. 100).

Através de discussões sobre língua e cultura, como as de Sapir(1949) e Whorf (1988[1940]) é apontado que há línguas tão sintéticas que podem expressar o conteúdo de uma oração inteira, sendo todos os componentes agregados em um único trecho, que provavelmente chamaríamos de palavra. Já outras, são de índole extremamente analítica. A idéia de estrutura lingüística em Sapir(1980[1921]) é abordada desde a ordem fonética até a disposição dos termos nas proposições frasais, a sintaxe. Interessa notar que, em relação às possibilidades de combinações, bem como a noção de ambiente fonético – o que será

aprofundado tanto na obra de Bloomfield quanto na de Harris, nesta mais do que naquela - já está expresso na obra de Sapir.

Ademais, ao fazer afirmações tais como: “toda língua tem um sistema fonético íntimo de padrão definido” e “também tem o sentimento definido de constituir padrões para a formação gramatical”(SAPIR, 1980[1921], p. 55), Sapir refere-se, pensamos, a uma estrutura ontológica presente em cada língua.

Dosse, trazendo a idéia de Louis-Jean Calvet, explicita que haveria em Saussure “a idéia da existência de uma linguagem sob a linguagem, de uma codificação consciente ou inconsciente das palavras sob as palavras, uma busca de estruturas latentes...” (DOSSE, 1993, p. 72). Também a idéia explícita em Hjelmslev (1975[1961]) é a de que, em detrimento de se analisar as realidades extralingüísticas, é preferível que o esforço do lingüista seja feito na busca de uma estrutura subjacente à ordem interna da língua.

Piaget discorre da seguinte forma, tratando da estrutura do ser em Psicologia:

Porém, se se dissocia, assim, o sujeito do ‘eu’ e do ‘vivido’, restam as suas operações, isto é, o que ele tira, por abstração reflexiva, das coordenações gerais de suas ações: *ora, essas operações são precisamente os elementos constitutivos das estruturas que ele utiliza*. Sustentar, então que o sujeito desapareceu para dar lugar ao impessoal e ao geral seria esquecer que, no plano dos conhecimentos [...], a atividade do sujeito supõe uma contínua descentralização que o liberta de seu egocentrismo intelectual espontâneo em proveito, não precisamente de um universal já pronto e exterior a ele, mas de um processo ininterrupto de coordenações e de reciprocações: ora, é esse próprio processo que é gerador das estruturas em sua construção ou reconstrução permanentes. *Em resumo, o sujeito existe porque, de maneira geral, o ‘ser’ das estruturas é sua estruturação.* [...] Não existe uma estrutura sem uma construção, ou abstrata ou genética (PIAGET, 1970, p. 114) grifo nosso.

De acordo com o princípio piagetiano exposto acima, a estrutura seria algo presente enquanto existente – independente da forma – e presente também enquanto representada, ou em outro termo, metodologizada.

No prefácio ao livro *Mathematical Structures of Languages* (1968) Harris se refere ao método apresentado da seguinte maneira: “Este livro pretende mostrar como se pode chegar a um sistema abstrato que caracteriza precisamente a linguagem natural”<sup>49</sup> (HARRIS, 1968, prefácio), ou seja, está metodologizando uma estrutura abstrata, e, não necessariamente, ausente. No *Structural Linguistics* (1969[1951]) o autor deixa claro que o método sugerido é interessante para analisar os usos que ocorrem em determinada

---

<sup>49</sup> “This book attempts to show one can arrive at na abstract system which characterizes precisely natural language.”

comunidade lingüística, identificando não apenas as ocorrências mais recorrentes, como também os “arranjos relativamente raros de elementos afirmados relativamente raros.”<sup>50</sup> (HARRIS, 1969[1951], p. 01). Em outra obra, *A Grammar of English on a Mathematical Principles* (1982), é acentuado que a análise gramatical tenta mostrar como cada sentença pode ser derivada através de operações de redução a partir de sentenças-fonte (HARRIS, 1982, p. V). Isso sugere que Harris considera a estrutura como latente, presente, pronta para ser representada – e de diferentes maneiras.<sup>51</sup> Desse modo, o linguista metodologiza estruturas consideradas presentes na língua.

Se percebermos a ontologia como o Ur-Código a que Eco (2001) se refere, sendo um código relacionado a uma “Estrutura das estruturas, que se identifica com os mecanismos Universais da mente”(p. 293), chegaremos indubitavelmente, como ele, à conclusão de que ela realmente não é representável. Aliás, Harris também discorre sobre a impossibilidade de a língua “espelhar em suas relações estruturais a relação entre os referentes” (HARRIS, 1991, p. 321). Seu conceito de estrutura se aproxima muito do que diz Eco, como sendo ela “uma rede de relações, complexo de relações, ordem que pode manter-se constante até no variar dos termos entre os quais se move” (ECO, 2001, p. 256).

Entretanto, se tomarmos ontologia como a realidade do status do sistema lingüístico, temos elementos para falar do método estrutural de Harris como uma metodologização da ontologia, pois trata-se de representar algo cuja existência pode-se alcançar, lingüisticamente falando. Se entendermos ontologia em estrutura de língua como a natureza dessa estrutura na realidade lingüística, então a metodologia de Harris explicita a ontologia estrutural da língua.

Entretanto, como esta é uma matéria que não vamos exaurir aqui - pois seria objeto para um estudo à parte - entre dogmatizar ou tornar a concepção de estrutura algo por demais difuso, apenas pretende-se buscar os caminhos para que seja possível compreender uma e outra concepção de como se (en)forma a estrutura em lingua(gem). Tomemos então para nossos objetivos a estrutura como modelo de análise, cujo objeto é, sem dúvida, a língua. A estrutura desta, concreta ou abstrata, se faz algo representável.

Tendo em vista essas considerações, é possível notar o estruturalismo em lingüística como uma corrente que observa e descreve as línguas através de critérios

---

<sup>50</sup> “...these procedures[...] provide a method for identifying all the utterances as relatively few stated arrangements of relatively few stated elements.”

<sup>51</sup> Nesse caso, tanto na forma reduzida, como na estendida, ambas são uma representação adequada da forma sentencial.

metodológicos, buscando encontrar os modelos lingüísticos. Ou seja, apesar de nomear fenômenos e de definir um conjunto e suas partes e as relações dessas partes entre si, escapa de seu objetivo o fabricar modelos de uniformização para se compreender um sistema lingüístico. Na verdade, ele se vale de modelos existentes - como os da matemática - e os adapta para explicitar o método.

A estrutura conforme se pretende focalizar aqui se validaria em uma descrição metodológica de uma realidade ontológica - pois vale-se de determinado método a partir do que é dado, é um modelo generalizante que analisa o objeto organizado língua. Optamos por entender a estrutura como instrumento operacional para explicar realidades ontológicas. Ao que parece, para Harris, a estrutura existe, mas pode ser que pareça obscura. Daí a importância da descrição. Ele não apresenta uma metodologia que pretenda dar conta do Ur-Código. Sua metodologia se direciona especificamente à ontologia do sistema enquanto sistema.

A lingüística descritiva se ocupa de metodologizar a estrutura latente nas línguas. Essa metodologização difere de lingüística para lingüística. Por exemplo, Harris em sua concepção de estrutura frasal, rompeu com a tradicional noção de Sujeito e Predicado, imprimindo um novo modelo, o de operadores e argumentos. Imersa em sua obra, está a idéia de que a estrutura é algo que está impresso nas asserções lingüísticas, e por isso mesmo, é importante que ela seja trazida à tona através de métodos de descrição, seja na ordenação fonética, na constituição de sentenças ou ainda na estruturação de textos.

Dessa maneira, a língua de forma alguma se circunscreve apenas ao esquema ou à série de esquemas que dela se abstrai: ela é apenas imobilizada em esquemas para poder estabelecer determinado mecanismo que permite a riqueza das leituras, e, portanto a contínua atribuição de sentido de que está carregada, o que conferiria seu valor como estrutura.

De acordo com Eco (2001), a estrutura realmente estará revestida de valor se se fizer um código gerador de diferentes mensagens. Aqui, o pensamento objetivo é o elo conectivo da investigação estrutural: interpreta-se a mensagem mediante um código, a língua, que pode ser transformado em outro, o esquema metodológico, e dessa forma a estrutura se faz como propriedade que define um código, que por sua vez é capaz de traduzir outros códigos.

Os estudos lingüísticos descritivos, dessa forma, se pautariam por individuar o sistema de relações de uma língua, indagando se esse sistema poderia ser comparado ao de

outra língua através de um código comum que delas desse conta, num estabelecimento de mecanismo gerativo de regras que explicasse os fenômenos de construção sentencial e discursivo em ambas.

Essa busca pela identificação de um código faz-se uma postulação teórica. E, vale salientar que esse possível ato de postulação na lingüística difere-se por ser feito *a posteriori*; somente depois da observação dos comportamentos lingüísticos concretos e da identificação das leis de dada língua é que se dá o estabelecimento das regras intrínsecas a ela. Ainda porque, para se abstrair o funcionamento de um sistema, é necessário que se tenham os dados, objeto de análise.

Esse código encontrado será então um modelo de uma série de convenções de comunicação que se postula existente como tal, para explicar o(s) possível(is) formato(s) lingüístico(s) de expressão em determinada língua. Assim,

um código é uma estrutura elaborada sob forma e modelo e postulada como regra subjacente a uma série de mensagens concretas e individuais que a ela se adequam e só em relação a ela se tornam comunicativas. Todo código pode ser comparado com outros códigos mediante a elaboração de um código comum, mais esquelético e abrangente (ECO, 2001, p. 39-40.)

Dessa forma, teria então o método estrutural uma função operacional, buscando uma generalização do discurso lingüístico no que se refere às regras de construção de um sistema. Em relação a isso, Zellig Harris resolve-se sem rodeios, ao se auto afirmar como um metodologista, sempre que chamado de lingüista (NEVIN, 1992; RYCKMAN, 2002).

### 3. Zellig Harris e a Língua

Harris se reporta à língua como algo a ser estudado em si e por si mesma, o que faz em consonância com o pensamento de Saussure (1994[1916]), inclusive no que se refere à evolução (mutabilidade) dos signos de que é composta.

...a estrutura da língua pode ser vista como uma evolução de usos em que a língua foi desenvolvida. Não há necessidade de se apelar à existência independente de alguma estrutura ou estruturalismo – cuja existência antes da língua ou independentemente da língua poderia em algum caso não ser explicada – para ver como a estrutura da língua veio a ser o que é<sup>52</sup> (HARRIS, 1981[1978], p. 411).

Ao nos atentarmos para perceber a forma de Harris tratar do objeto língua, nota-se que está permeada na obra vista, a idéia de estrutura que é metodologizada e, assim, descrita (cf. HARRIS, 1982, p. V). A partir dessa descrição, Harris afirmava ser possível estabelecer a gramática das línguas. Chegou a descrever gramáticas completas, como a da língua Hebréia (HARRIS, 1969[1951], p. 168) e Fenícia (HARRIS, 1990[1936]). Realizou estudos comparativos entre estruturas de línguas como o Navaho (cf. HARRIS, 1969[1951], p. 130; 1970[1945], p. 177-187.), o Yokuts (HARRIS, 1970[1944], p.188-208), o Hidatsa, os fonemas do Swahili (HARRIS, 1969[1951], p. 97-124), dentre outras.

Considerando a abordagem geral – conforme percebida na totalidade das obras consultadas – é saliente o pensamento de que a língua (em que se percebe um desdobramento da teoria Saussureana de língua como sistema) se faz de um sistema composto de relações cujos constituintes são alocados nos enunciados através de mecanismos internos de restrição, seleção e combinação. Isso é válido para todos os sistemas linguísticos.

A propriedade crucial da língua é que a presença de palavras em uma sentença depende de como outras palavras na sentença dependem ainda de outras palavras dela. Essa dependência é a principal restrição de igualdade de probabilidade de palavras em sentenças. [...] ...essa dependência pode ser considerada um construto de teoria sintática. [...]. Inicialmente, nós estamos falando aqui, por exemplo, de como a presença de um verbo em uma sentença

---

<sup>52</sup> ... the structure of language can be seen as an outgrowth of the uses in which language was developed. There is no need to appeal to some independently existing structure or structuralism – whose existence before language or independent of language could in any case not be explained – to see how the structure of language came to be what it is.

requer a presença nela de algum nome como seu sujeito<sup>53</sup> (HARRIS, 1991, p. 54).

Especificamente em relação ao desenvolvimento e apresentação de suas investigações linguísticas, que são feitas em uma vasta produção, a idéia de estrutura para Harris se concretiza eminentemente na forma como os componentes lingüísticos formais se ligam, se combinam para a produção de enunciados com sentido. Indo desde a abordagem fonética e suas particularidades na descrição das línguas, passando pela estrutura sintática (HARRIS, 1969[1951]) e até a teorização sobre o que ele nomeia sublinguagens (1968;1988;1991;), a nota tônica é esse recorte formal que se apresenta como um método a partir do qual os enunciados da língua podem ser obtidos a partir de algumas operações igualmente formais.

Resumidamente para Harris, a estrutura de uma língua consiste:

- 1) na construção de palavras a partir de sons, de forma que a estrutura da sintaxe não seja afetada;
- 2) na construção de sentenças a partir de palavras, por uma simples relação de ordenação de partes;
- 3) nas transformações nas formas sentenciais, essencialmente por um processo de redução<sup>54</sup> (HARRIS, 1991, p. 3).

E interessa notar que, para ele, a relação simples de seqüências de palavras na sentença, confere ao ambiente dessas mesmas seqüências um status matemático, e que pode carregar toda a informação que é transmissível pela língua. As operações feitas nessas sentenças – como as reduções – podem mudar a forma delas, mas a informação lingüística continuará a mesma. Além de considerar como as palavras se combinam para a formação de sentenças, a proposta de Harris se coloca em uma análise de como os ambientes sentenciais podem ser traduzidos em expressões algébricas. Também seria possível uma comparação da língua com outros sistemas notariais, como a música. (cf. HARRIS, 1991, prefácio; 1991, 304-318).

---

<sup>53</sup> The crucial property of language is that the presence of words in a sentence depends on how other words in the sentence depend on yet other words in it. This dependence is the essential constraint on equiprobability of words in sentences. [...] this dependence can be considered a construct of the syntactic theory. [...] To a first approximation, we are speaking here, for example, of how the presence of a verb in a sentence requires the presence there of some noun as its subject.

<sup>54</sup> 1-the construction of words from sounds, in a way that does not affect the structure of syntax;  
2-the construction of sentences from words, by a single partial-ordering relation;  
3-the transformation of sentence shapes, primarily by a process of reduction

Um traço marcante e inovador da estrutura em Harris é que as palavras e sentenças podem ser caracterizadas por uma relação definida por operadores-argumentos, com diferentes probabilidades de ocorrência para determinadas palavras como operadores, com argumentos específicos e com reduções na forma frasal para palavras que têm alta ocorrência em relação a seus operadores ou argumentos. Em verdade, esses operadores são divididos em classes de primeira e de segunda ordem, de acordo com o seu grau de dependência em relação a seus argumentos.

Nessa caracterização, Harris se afasta da então tradicional noção de estrutura sentencial, cujos componentes trazem a denominação de sujeito e predicado, noção que tem permeado os estudos gramaticais por séculos. Ele “demonstrou excepcional coragem intelectual em abandonar (ess)a noção e adotar para a descrição de sentenças o esquema geral: No V W, onde No é o sujeito gramatical, V o verbo e W a seqüência dos complementos”<sup>55</sup> (GROSS, 2002, p. 60-61).

O esquema sentencial passa então a ser estabelecido através da relação entre o operador V e seus argumentos N. Harris então vai desenvolver a partir dessa noção de estrutura frasal uma divisão das classes morfológicas do Inglês, elencando-as no que chamou de classes de morfemas. Para isso, o conceito de ambiente, bem como o de vizinhanças<sup>56</sup> se fizeram muito importantes para se estabelecer essas classes e também para explicar a relação operador-argumento. Dessa forma, a língua é descrita como sendo formada de ambientes de base e de outros, passíveis de transformação.

O tratamento estrutural que Harris confere à língua tem início com a abordagem do aspecto fonológico. Nessa abordagem, são feitas generalizações de como os fonemas se organizam nos enunciados lingüísticos. A maneira como os fonemas ocorrem ou estão distribuídos nas ocorrências lingüísticas é objeto de uma extensa exposição (HARRIS, 1969[1951]), em que são apontados procedimentos para se perceber características tais como contornos e junturas. Através da estrutura fonológica das combinações possíveis encontradas, Harris afirma ser possível fazer generalizações sobre os traços relevantes do discurso que ocorrem quando a língua é falada (HARRIS, 1969[1951], p. 152).

Ao falar sobre a estruturação dos fonemas em dada língua, em princípio, destaca uma regra básica de combinação, a mesma que se aplica aos componentes morfêmicos.

<sup>55</sup> “...Harris demonstrated exceptional intellectual courage in abandoning the notion and adopting for the description of sentences the general schema: No V W, where No is the grammatical subject, V the verb and W the sequence of the complements.”

<sup>56</sup> O conceito de ‘ambiente’ e também o de ‘vizinhanças’ foram desenvolvidos de forma mais específica no *Structural Linguistics* (1951).

Como o próprio Harris afirma, “é quase impossível a ocorrência, em qualquer língua, de todas as seqüências de todas as combinações simultâneas de todos os elementos”<sup>57</sup>(HARRIS, 1969[1951], p. 150).

Considerando essas regras de seleção e combinação para que os enunciados da língua sejam construídos de forma adequada no sistema, passamos a discutir as particularidades das restrições linguísticas propostas por Harris, salientando como essas restrições contribuem para a informação.

### 3.1. Informação

Dentro da concepção de língua adotada por Harris presentifica-se sempre a característica informacional que ela apresenta em sua estrutura. Essa propriedade se refere necessariamente à informação de acordo com a teoria matemática. “A teoria de informação linguística de Harris é alicerçada nas mesmas considerações de probabilidade relativa que subscreve a teoria da informação matemática como é ordinariamente entendida”<sup>58</sup> (NEVIN, 2002, p. xvii-xviii).

A capacidade informacional da língua - matematicamente falando - consiste na maior ou menor possibilidade de ocorrência de componentes em determinado ambiente, tanto em se tratando do nível fonemático quanto do morfemático, ou de palavras. Ainda de acordo com a teoria matemática, a estrutura se compõe também pela não equiprobabilidade de ocorrência de partes, e isso é determinante para se obter a quantidade de informação.

Um bom exemplo dessa propriedade informacional da língua pode ser encontrado na forma das modernas conversas oportunizadas pelo uso de mensageiros virtuais. As abreviações de palavras trazem em si a informação sobre o que seria possível de ocorrer naquele ambiente, e o receptor da mensagem a recebe de forma compreensível através desse canal. Mesmo sendo as palavras abreviadas, essas abreviações ocorrem de forma não aleatória. Permitem somente a inserção de elementos que estão previstos em uma relação de probabilidade. Na ocorrência

---

<sup>57</sup> “...it is almost impossible for all sequences of all simultaneous combinations of all the elements [...] to occur, in any language.”

<sup>58</sup> “Harris’s theory of linguistic information is grounded in the same considerations of relative probability that underwrite mathematical information theory as ordinarily understood.”

(1) *Vc vai fz o trabalho com o grupo?*

Não seria possível que os elementos *Vc* e *fz* fossem representados por *vaca* e *fiz*, na língua portuguesa, por exemplo. Em sua base, a teoria matemática da informação aborda sua característica mais significativa como sendo a que seleciona a mensagem de um grupo possível de mensagens, em que opera para cada seleção possível (cf. SHANNON, 1948, p. 01). No exemplo acima, as formas possíveis seriam ‘você’ e ‘fazer’.

A fórmula informacional, para Shannon, pode ser pensada e representada como uma função  $f(x, y)$ , cujo número de variáveis dependerá da natureza da mensagem, do canal, e dos demais componentes do esquema informacional. Tratando da teoria da informação, o matemático classifica os sistemas de comunicação em três principais categorias: discreta, contínua e mista – nesse caso, a língua se enquadraria neste último, que é o que nos interessa mais de perto -. Nele o sistema é ao mesmo tempo discreto, que é aquele em que tanto a mensagem quanto o sinal são uma sequência de símbolos discretos (e.g., mensagem telegráfica) e também contínuo, em que a mensagem e o sinal são tratados como funções contínuas (e.g., as mensagens via televisão).

As línguas têm uma performance essencialmente telegráfica (cf. JOOS, 1950, apud RYCKMAN, 2002, p. 28) em sua estrutura, diferindo de sistemas como o código Morse, por exemplo, em dois pontos: pelo fato de terem muitas camadas de complexidade ao invés de apenas uma, e de que em cada camada há diversas limitações sob as combinações permitidas. Desse modo, também na abordagem linguística, a língua se faz um sistema misto de informação.

Em Harris (1991; 1968; 1982, p. 5-7; 1988, p. 83), é evidenciada a relação de sua abordagem com a da Teoria Matemática da Informação; assim sendo, são tratados de maneira particular os elementos discretos da língua, ainda que o linguista se reporte aos elementos não discretos como componentes do sistema (HARRIS, 1988, p. 81; 1951, p. 20-24; 1991 p. 22; 1968 p. 6-7; 208). Harris estabelece entretanto diferenças relevantes em relação à abordagem puramente matemática e a abordagem que ele próprio faz, no estudo da informação dentro do sistema linguístico.

Na teoria matemática da Informação, desenvolvida dentro da estatística, o que foi estudado foi a quantidade de informação, como limitada pela quantidade total de redundância em um sistema. A teoria da Informação não se ocupa

compreensivamente com a relação entre as contribuições individuais para a redundância total, e não caracteriza os itens individuais de informação que estão envolvidos no acúmulo de quantidade de informação. O estudo da estrutura da língua não apenas mostra sua relação com a informação, mas também mostra uma característica estrutural de informação própria, pelo menos com respeito à informação carregada pela língua: que cada item de informação consiste não apenas de uma restrição, ou uma redundância, ou um significado isolado, mas de uma restrição agindo sobre outra <sup>59</sup>(HARRIS, 1988, p. 83).

Essa característica estrutural própria de que fala Harris, se relaciona às ditas camadas de complexidade a que Ryckman alude, que conferem à língua um status diferenciado em relação aos sistemas puramente discretos – agregando em si também características contínuas.

Da mesma forma que Shannon trata os sistemas de informação como representáveis por funções, Harris também representa a informação passada pelas sentenças da língua em forma de função, composta pelos operadores agindo sobre os argumentos. Ao tratar da informação na língua, Harris (1991, p. 345) afirma que essa informação é um efeito inerente às restrições da língua, produtoras de sentenças com sentido. Dessa forma, a proposta do linguista é a de apresentar uma teoria que procura dar conta da estrutura da língua e da informação que ela possibilita através das restrições que apresenta. Essa estrutura informacional se apresenta de forma mais saliente na relação de dependência operador/argumento e da não equiprobabilidade de ocorrência de palavras.

Em se tratando do papel das palavras na informação transmitida pela língua, é necessário refletir como elas carregam (e se carregam) em si o sentido. “É claro que a maioria das palavras têm alguma associação com o sentido independentemente de sua ocorrência em sentenças (i.e., em combinação)”<sup>60</sup> (HARRIS, 1991, p. 323). Isso quer dizer que as palavras, em estado de dicionário então, de alguma maneira, informam. De forma que elas, bem como os morfemas, têm, a priori, *algum* significado. Palavras geralmente se referem a propriedades e classificações de objetos e/ou eventos que não poderiam ser referidos de outra forma dentro da língua; mas elas não podem simplesmente ser

---

<sup>59</sup> In a mathematical Information theory, developed within statistics, what was studied was the amount of information, as limited by the total amount of redundancy in a system. Information theory did not deal comprehensively with the relation between the individual contributions to the total redundancy, and did not characterize the individual items of information that were involved in amassing the amount of information. The study of the structure of language not only shows its relation to information but also shows a structural characteristic of information itself, at least with respect to the information carried by language: that each item of information consists not only of a constraint, or a redundancy, or an isolated meaning, but of one constraint acting upon another.

<sup>60</sup> “It is clear that most words have some association with meaning independently of their occurrence in sentences”.

aprioristicamente organizadas dentro de um esquema de significado. Dessa maneira, mesmo para as muitas palavras que parecem diretamente associadas com objetos e eventos do mundo, a associação é uma relação complicada, pois as palavras não têm um único referente no mundo.

Ao se tratar das palavras inseridas em expressões, deparamos com esse problema da referência. É impossível determinar o grupo total de indivíduos/objetos que constituem a denotação de uma dada palavra. Ao que parece, as palavras não podem ser limitadas a um único referente ou a um grupo fechado de referentes. Elas não podem ser definidas pela sua extensão, ou por, conforme define Penco (2006), uma classe de objetos que se encaixam em determinado conceito.

Uma palavra pode ser usada em um sentido único e se referir a diferentes indivíduos, como *homem*, em seu sentido extensional. Bem assim duas palavras ou frases com sentidos diferentes podem se referir ao mesmo indivíduo. Assim, *Roberto Carlos* e *O Rei da música* ou *Rio de Janeiro* e *Cidade Maravilhosa* têm um único referente. Temos ainda o caso de palavras cuja referência no mundo real é nula, como os clássicos exemplos de *unicórnio* e *centauro*, em que se pode referir apenas intensionalmente.

Para Harris (1991, p. 350; 1968, p. 7), a relação não biunívoca entre língua e referentes no mundo real faz com que sejam possíveis, através dessa mesma língua, apenas generalizações, por conta da relação limitada da língua e informação com o mundo e a complexidade de seus eventos. Na verdade, as palavras são apenas associadas a aspectos do mundo percebido.

Pois se as estruturas da língua e informação são sistemas de desvios particulares de aleatoriedade, assim são os objetos e relações do mundo percebido sobre que a língua fala. Não há base aqui para qualquer afirmação geral de que a língua reflete o mundo sobre o qual ela fala, ou que a estrutura da língua e informação corresponde à estrutura do mundo<sup>61</sup> (HARRIS, 1988, p. 84).

Aqui, Harris dialoga com Sapir (1949) e Whorf (1988[1940]) na idéia de que as informações da língua não correspondem necessariamente à realidade, por não haver correspondência entre informação linguística e mundo real. “Ademais, a língua universalmente desenvolve facilidades informacionais que não surgem diretamente de

---

<sup>61</sup> For if language structure and information are systems of particular departures from randomness, so are the objects and relations of the perceived world that language talks about. There is no basis here for any general claim that language mirrors the world it talks about, or that the structure of language and information corresponds to the structure of the world.

procedimentos do mundo real”<sup>62</sup> (HARRIS, 1991, p. 347). De fato, as propriedades composicionais das palavras permitem que elas, mediante restrições e combinações dentro destas, apenas *projetem* aspectos do mundo real, mediante convenções de uma comunidade.

A língua é claramente e sobretudo, uma portadora de significado. Não porém, de todo significado. Muitas atividades humanas e estados têm significado para nós, e apenas alguns deles podem ser expressos na língua: sentimentos e sensações vagas podem ser referidas apenas indiretamente; informação não pública, tais como sensações proprioceptivas, podem ser nomeadas apenas com dificuldade; certos tipos de informação não linguística podem ser traduzidas diretamente dentro da língua (e.g. gráficos e cartazes); mas outros tipos (e.g. fotografias) podem ser representadas na língua somente vaga e seletivamente. O significado em si é um conceito de definição não clara e suas partes que podem ser expressas na língua não são um subgrupo diferente caracterizado de significado.[...] Em cada língua, nós não sabemos a priori quais aspectos específicos de significado estarão sendo referidos pelas palavras, e quanto será incluído no sentido de uma palavra. Mesmo quando os sentidos são bem definidos, não é sempre possível para palavras de uma língua espelhar em suas relações estruturais a relação entre os referentes<sup>63</sup> (HARRIS, 1991, p. 321).

Quando se parte dessas considerações de Harris, é possível afirmar que a capacidade informacional e de transmissão de significado da língua se coloca não em se tratando de representar o mundo. Em realidade, sendo a língua considerada um sistema de representação, ela não pode espelhar a realidade; quando muito pode refletir e transmitir de uma forma ordenada as percepções do mundo. Complexos são os sistemas linguísticos e a linguagem, que deles faz uso.

Sob esse ponto de vista e segundo Ryckman (2002), a língua não é, no que se refere à sua função de comunicar e transmitir informação, uma herança genética, mas é uma prática social compartilhada por excelência, e isso é um aspecto que se faz relevante se levarmos em consideração as linguagens particulares de ciências especiais - as sublinguagens das ciências.

---

<sup>62</sup> In addition, language universally develops informational facilities that do not arise directly from dealing with the real world.

<sup>63</sup> Language is clearly and above all a bearer of meaning. Not, however, of all meaning. Many human activities and states have meaning for us, and only some of this can be expressed in language: feelings and vague sensings can be referred to only indirectly; non-public informatio, such as pro-prioceptive sensations can be named only with difficulty; certain kinds of non-language information can be translated directly into language (e.g., graphs and charts); but other kinds (e.g. photographs) can be represented in language only loosely and selectively. Meaning itself is a concept of no clear definition, and the parts of it that can be expressed in language are not an otherwise characterized subset of meaning. In each language, we do not know a priori which specific aspects of meaning will be referred to by words, and how much will be included in the meaning of a single word. Even when the meanings are well-defined, it is not always possible for words of a language to mirror in their structural relations the relation among the referents.

Então, a informação linguística se enraíza acima de tudo na possibilidade de se desenvolver uma estrutura representacional discursiva, em uma relação de dependência entre partes. O resultado disso é que a estrutura linguística permite então significados como os experienciais: a imaginação, o contra-senso, os chistes, e ainda a metalinguagem.

A língua é o meio supremo de expressão humana, não de comunicar significados – pois há muitos meios não linguísticos ou quase linguísticos de fazê-lo - mas de articular, delimitar e transmitir significado, como informação de estrutura predicacional, entre um indivíduo e outro ou entre um grupo e uma ampla comunidade. Segue-se que a padronização existindo nessas práticas sociais que chamamos língua é a de informação<sup>64</sup> (RYCKMAN, 2002, p. 35)

Uma vez que o estado de dicionário das palavras permite apenas significados um tanto restritos e apresenta muitas limitações, o significado então, de uma forma mais ampla, só pode ser conferido através da propriedade que a língua tem de combinar palavras dentro de um contexto discursivo. “Os significados em relação a todo escopo linguístico não podem ser estipulados independentemente do estoque de palavras *e suas relações sentenciais*”<sup>65</sup> (HARRIS, 1988, p. 6, *italico nosso*).

Essas relações podem sofrer alterações no decorrer do tempo, com, por exemplo, as entradas na língua de diferentes expressões idiomáticas e gírias, que fazem parte das mudanças linguísticas. Assim, com mudança na dependência operador/argumento, as palavras adquirem novos e outros significados. Esses novos sentidos são prova de que eles mesmos não habitam as palavras em estado de dicionário e que o significado se constrói através das restrições/seleção linguística em contextos específicos de uso da língua.

O que é aceitavelmente plausível para a mudança é o uso da palavra, as probabilidades de combinação. A mudança é lenta na lista de palavras, ainda mais lenta em construções gramaticais e nas subclasses gramaticais de palavras, e nas distinções fonêmicas. O que não muda afinal, até onde se conhece sobre a língua, é a existência de uma ordenação de partes. Essa propriedade, que confere o status de sentença, se mostra como universal e invariável<sup>66</sup> (HARRIS, 1988, p. 95).

<sup>64</sup> Language is the paramount human means, not of communicating meaning – for there are many non-linguistic or quasi-linguistic ways of doing that – but of articulating, delimiting, and transmitting meaning, as predication-structured information, between one individual and another or between a group and a wider community. It follows that the patterning existing in those social practices that we term a language is that of information.

<sup>65</sup> “Meanings over the whole scope of language cannot be arranged independently of the stock of words and their sentential relations”.

<sup>66</sup> What is reasonable likely to change is word use, the likelihoods of combination. Change is slower in the word stock, still slower in the grammatical constructions and the grammatical subclasses of words, and in the phonemic distinctions. What doesn’t change at all, as far as we know about language, is the existence of a partial ordering. This property, which makes sentencehood, appears to be universal and invariable.

Sendo assim, a estrutura sentencial/textual se faz instrumento para a transmissão da informação linguística que se apresenta a partir das dependências contextuais do discurso. Nesse sentido, Harris afirma que “as distinções e limites de alcance do sentido são complicadas e vagas. Há de se notar que o significado é mais facilmente obtido como uma propriedade de combinações de palavras do que de palavras [...] por si mesmas”<sup>67</sup> (HARRIS, 1991, p. 325).

Se pensarmos no sentido das palavras sob uma distinção puramente semântica daparamo-nos, já de início, com o problema das relações de homonímia. Palavras que são diferentes, com sentidos diferentes, podem ter grafia idêntica, portanto, fora de um ambiente sentencial/textual, trazem problemas de sentido. O exemplo de Harris que evidencia de forma mais clara essa situação é a palavra *will* (HARRIS, 1988, p. 61) expressando *desejo* e também *futuro* são etimologicamente relacionadas, porém, levando em consideração seus sentidos, muito diversos um do outro, são consideradas duas palavras diferentes, tanto por conta de seus diferentes significados, quanto por conta de suas diferentes funções gramaticais. De forma que, para se definir o sentido de *will*, é necessário que se conheça seu ambiente de ocorrência. Da mesma forma, podemos exemplificar com a palavra *mente*, do português, que oferece a mesma problemática. Isoladamente, tem sentidos distintos e duas funções gramaticais diferentes: nome e verbo. Temos para essa palavra os seguintes significados, segundo o Dicionário Michaelis:

mente

*sf (lat mente)* **1** Faculdade de conhecer, inteligência, poder intelectual do espírito. **2** Entendimento, alma, espírito. **3** Disposição para fazer alguma coisa. **4** Idéia, resolução. **5** Concepção, imaginação. **6** Intenção, intuito, plano, tenção. organismo humano de autocondução, de aprender e de pensar.

mentir

*(lat mentiri)* *vti* e *vint* **1** Dizer mentiras, negar o que se sabe ser verdade, proferir como verdadeiro o que é falso: *Mentir aos pais é crime. Não mentiu nas suas declarações. Tudo isso declarou ele sem mentir.* *vtd* **2** Proferir mentira: *Mentiu mentira de rabo e cabeça.* *vti* e

---

<sup>67</sup> “The distinctions and boundaries of meaning-range are complicated and vague. It will be seen that meaning is more easily stated as a property of word combinations than of words [...] by themselves.”

*vint 3* Induzir em erro, ser causa de engano: "E assim não lhe minto nem o molesto" (Rodrigues Lobo). "O coração pressago nunca mente" (Luís de Camões). *vt d 4* Errar: "Meu bacamarte mentiu fogo duas vezes - disse Cabeleira" (Franklin Távora). *vti 5* Não cumprir (compromisso, dever, juramento, promessa): *Mentira ao prometido*. *vti 6* Faltar, não corresponder: *Mentiu às nossas esperanças*. *vint 7* Degenerar: *De pais sadios a prole não mente*. *vint 8* Não se concluir, não se realizar, não ter efeito, não vingar: "Os frutos mentiram" (Constâncio). *vint 9* Esmorecer, faltar (o fogo). *vint 10 Constr* Não assentar bem no lugar destinado (peça de madeira). Conjugam-se como *aderir*. *M. pela gorja*: mentir cinicamente; mentir muito.

Os significados isolados oferecidos pelo sistema são polissêmicos e não permitem que se alcance o sentido. Dessa maneira, para o ouvinte/leitor, dependerá do ambiente a associação a um sentido específico e adequado, ou seja, se, em nosso exemplo, a palavra se relaciona à mente humana - o organismo de autocondução, de aprender e de pensar - ou se a uma forma conjugada do verbo mentir.

O problema de decidir se uma dada sequência fonêmica é uma palavra com todos os seus sentidos, ou duas palavras homônimas, torna-se mais fácil pelo fato de que palavras não aparecem sozinhas na maioria das vezes. Elas aparecem em ambientes específicos [...] e podem ter sentidos diferentes sob operadores específicos ou sobre argumentos específicos<sup>68</sup> (HARRIS, 1988, p. 61-62).

Com efeito, a noção de significado como dependendo do ambiente auxilia a se pensar sobre o problema do alcance do sentido, pois os procedimentos para se determinar o sentido das palavras não repousam no significado isolado, ou seja, não são critérios de ordem semântica, no sentido mais genérico do termo. Quando se pensa no significado de uma palavra em uma ocorrência particular tendo como noção básica a de dependência de seu ambiente, a categorização de alcance de significado pode ser compreendida mediante a categorização de ambiente de palavras. Assim,

---

<sup>68</sup> The problem of deciding whether a given phoneme sequence is one word with all its meanings, or two homonymous words, is made easier by the fact that words do not for the most part appear alone. They appear in particular environments [...] and may have different meanings under particular operators or over particular arguments.

o significado das palavras são distinguidos e em parte determinados pelas palavras que ocorrem com maior probabilidade como seus argumentos ou como operadores sobre elas. O significado de uma ocorrência em particular de uma palavra é determinado pela seleção das palavras que são seus argumentos ou das palavras que são operadores sobre essa palavra nessa frase (BAPTISTA, 2005a, p. 4).

Em relação a essa dependência entre operadores e argumentos, temos em Harris (1988; 2002b; 1982; 1970[1946]; 1968) a abordagem da língua como um sistema que carrega informação mediante restrições inerentes a ele; informação que termina por se direcionar para o próprio sistema, para o que ele aceita ou não, mediante a seleção dos signos convencionados para esse mesmo sistema. Sendo essas restrições determinadas pela relação de dependência, é interessante então pensar de maneira detalhada na informação expressa pela língua através delas.

### **3.1.1. Informação sob as restrições que produzem as sentenças da língua**

Os traços informacionais que a língua carrega permitem, na teoria apresentada por Harris, que se façam generalizações sobre a relação forma-conteúdo. Nesse sentido, a informação carregada pela língua é associada com ou expressa pelas restrições de equiprobabilidade de palavras em relação a outras palavras nos usos. A informação nesse caso nem é independente dessa estrutura nem adicional a ela, mas é uma interpretação dela. A informação sob esse ponto de vista é trazida pela estrutura porque é inerente a ela. Os estudos sistematizados que se ocupam de observar a estrutura linguística sob o prisma das restrições que a produz foram propostos de forma pioneira por Harris (RYCKMAN, 2002, p. 25). O linguista considera essa característica como sendo universal às línguas. Todos os sistemas linguísticos têm suas especificidades, entretanto, todos eles são estruturados mediante restrições.

É necessário que se considere a distinção de Harris de que o padrão linguístico em si pode ser considerado como uma estrutura hierárquica de restrições sobre combinações de elementos, cujo efeito combinado é passível de acréscimo, e que, como resultado, a estrutura linguística carrega, ou melhor, expressa, informação. Isso se aplica não só ao sistema como um todo, como também a campos específicos, as chamadas sublínguas, nelas adquirindo particularidades, por conta de sua gramática, que também é específica.

“Como todo sistema sintático, o das linguagens da ciência consistem de classes e subclasses de palavras, e os tipos sentenciais que são produzidos por suas relações operador-argumento, junto com quaisquer restrições em todos eles<sup>69</sup>.” (HARRIS et al, 1989, p. 72)

As principais restrições enumeradas por Harris para a formação de sentenças adequadas na língua são três<sup>70</sup>, que, agindo umas sobre as outras, mediante uma gramática de operadores, produzem as sentenças adequadas na língua.

Para o linguista (HARRIS, 1998, p. 55), a “língua é claramente uma portadora de significado e informação”<sup>71</sup>; e esse fenômeno se dá mediante as citadas restrições que criam as sentenças. Algumas delas têm uma relação indireta com a informação, como por exemplo, as topicalizações (HARRIS, 1982, p. 109-110), em que a palavra é deslocada para o início da sentença, indicando a ênfase dada pelo falante a tal palavra/expressão. Pode-se também dizer que há significados que são transmitidos pelos componentes não discretos da língua, a saber as entonações, as pausas, expressões corporais e outros.

Sim, os componentes não discretos da língua transmitem informação, mas essa só é possível de ser expressa mediante as restrições, melhor dizendo, dentro delas. Não nos ateremos a esses componentes. Trataremos, dentro dos elementos discretos, das restrições que se colocam de forma direta na produção de informação substantiva na sentença, que, segundo Harris, são três: “Será visto que há três tipos de restrições na combinação de palavras que produzem sentenças, e que cada uma carrega um tipo de sentido, tal que o significado da sentença é determinado diretamente a partir de restrições de palavras”(HARRIS, 1988, p. 10). São elas:

- 1) a ordenação parcial de palavras, ou seja, a ordem em termos de prioridade de entrada das palavras na frase;
- 2) a desigualdade de probabilidade de palavras dentro da ordenação parcial.
- 3) as reduções sentenciais

Inclusa nessas restrições e de forma essencial, repousa a relação de dependência entre palavras, alicerçada na relação operador/argumento.

---

<sup>69</sup> Like all syntax, that of the science language consists of word classes and subclasses, and the sentence types which are produced by their operator-argument relations, together with any constraints on all these (loc. cit.)

<sup>70</sup> Harris se refere a outras restrições presentes no sistema linguístico, que não são necessariamente responsáveis pela estruturação de sentenças.

<sup>71</sup> “...language is clearly a carrier of meaning and information.”

Na presente análise uma sentença, bem como um discurso, é caracterizada(o) não diretamente como uma sequência de palavras inclusas nela(e), mas como uma representação linear de uma relação particular entre ela(e)s: a presença na sentença de certas palavras depende da presença de certas outras palavras<sup>72</sup> (HARRIS, 1982, p. 33).

Os mecanismos de combinação linguística se fazem então básicos para se estabelecerem as relações de significado das sentenças/discursos, em se tratando de todos os eventos linguísticos, inclusive os que contêm uma gramática específica, como as sublinguagens.

...a informação de um texto, grupo de discurso, ou sublinguagem, é expressa pela gramática mínima que o(a) caracteriza completamente em termos de sua recorrência de elementos e seus modos de combinação. Na falta de uma metalinguagem externa a partir da qual se derivem esses elementos e suas combinações permitidas, os elementos podem somente ser arranjados de uma forma puramente distribucional, i.e., combinatorial<sup>73</sup> (RYCKMAN, 2002, p. 30).

Passemos a tratar então desses arranjos combinatoriais, que são explicados pelas propriedades das restrições sentenciais, pois, para perceber como as sentenças carregam informação, é relevante investigar como essas restrições carregam informação (cf. HARRIS, 1988, p. 58).

### 3.1.2. Primeira restrição: Ordenação de Partes

A primeira restrição se coloca como a combinação operador-argumento na probabilidade de ocorrência de palavras em determinado ambiente. Para cada operador dado, há zero ou mais classes de palavras argumentos que podem ocorrer em combinação

---

<sup>72</sup>In the present analysis a sentence, as also a discourse, is characterized not directly as a sequence of the words in it but as a linear representation of a particular relation among them: the presence in the sentence of certain words depends upon the presence of certain other words.

<sup>73</sup> ... the information of a text, set of discourses, or sublinguagem, is expressed by the minimal grammar that completely characterizes it in terms of its recurrent elements and their modes of combination. Lacking an external metalanguage from which to derive these elements and their permitted combinations, the elements can only be set up purely distributionally, i.e., combinatorially.

com ela (e vice-versa), dado que as palavras não combinam umas com as outras de forma aleatória.

A partir de todas as restrições juntas segue-se a maior propriedade da língua: que nem todas as combinações de palavras ocorrem como sentenças da língua. Isto é uma propriedade necessária e universal. Em toda língua nem todas as sequências de fonemas, ou de morfemas (palavras), na língua ocorrem como sentenças, ou como discursos <sup>74</sup> (HARRIS, 1991, p.117).

Para a produção de sentenças, a teoria da sintaxe de operadores preconiza a presença de dois tipos de palavras:

- 1) as que aparecem quando ‘pedidas’ por outra classe de palavras, mas não exigem a presença de outras para ocorrerem – a saber, as da classe argumentos elementares (N);
- 2) as que fazem parte da sentença com a condição de que nela estejam presentes palavras (N), ou palavras que requeiram (N), os operadores.

Assim, um operador não ocorrerá na sentença a menos que uma palavra de alguma das classes de argumentos prováveis esteja presente, preenchendo o espaço ao lado. Sua entrada na frase depende desse preenchimento. Esta dependência é uma ordenação de partes (HARRIS, 1981[1978], p. 410).

A entrada do verbo *comer*, por exemplo, em uma sentença, depende preferencialmente de argumentos pertencentes a classes de palavras tais como *pessoa*, *animal*, para preenchimento da casa à esquerda, e *alimento* para preenchimento à direita. Essas têm prioridade de entrada na frase em posição argumento.

(2) *Taís comeu salada no almoço.*

(3) *Teco come a ração.*

Então essa primeira restrição consiste na propriedade pela qual uma palavra funciona em uma frase como operador sobre outras palavras, os argumentos. Ela

torna possível reunir as palavras de uma língua em classes que funcionam como operadores e como argumentos umas em relação às outras, de tal modo que no

---

<sup>74</sup>From all the constraints together there follows the major property of language: that not all combinations of words occur as sentences of a language. This is a universal and necessary property. In every language not all sequences of the phonemes, or the morphemes (words), in the language occur as sentences, or as discourses.

conjunto das frases não-reduzidas não há nenhuma palavra da classe dos operadores que apareça sem que esteja presente uma das palavras das suas classes de argumentos (BAPTISTA, 2005a, p. 15).

Nessa concepção, os operadores são classificados como pertencendo a uma primeira ( $O_n$ ), ou segunda ( $O_o$ ) ordem. Os de primeira ordem dependem da presença de palavras que não dependem de outra para entrarem na frase, o que podemos chamar de argumentos simples; estes criam junto com seus operadores as sentenças elementares. Os operadores de segunda ordem dependem de argumentos, que em verdade são operadores sobre outros argumentos. Dessa forma, nos exemplos acima, houve a entrada de operadores de primeira ordem. De forma diversa, em

(4) *Taís disse que comeria salada no almoço*

o operador *disse* ( $O_o$ ) está agindo sobre um argumento que é um outro operador, *comer*. E este, por sua vez, pede um argumento nominal, uma vez que é um operador de primeira ordem. De maneira que as construções sentenciais resultam de uma ou mais sentenças elementares, com possibilidades de operadores de primeira e de segunda ordem.

A ordenação parcial é uma restrição em combinação de palavras: isso quer dizer que na posição argumento perto de um dado operador a frequência (ou probabilidade) de certas palavras – que não constam na classe de argumentos para aquele operador – é zero. Cada satisfação da ordenação parcial, i. e., cada sequência de palavras na qual as palavras fonte têm suas exigências satisfeitas, é uma sentença. Consequentemente, quando um operador mais alto tem um outro operador como seu argumento [...] então o operador mais alto está agindo sobre toda a sentença, que é então um componente sentencial da sentença maior sendo criado pelo operador mais alto [...] <sup>75</sup> (HARRIS, 1988, p. 13)

É interessante notar que essa restrição é a base para as demais, por produzir as sentenças elementares da língua. Temos, dessa forma, a ordenação parcial como a primeira restrição que contribui para os significados das sentenças, na medida que o operador é dito sobre o(s) seu(s) argumentos (BAPTISTA, 2005b, p. 11).

---

<sup>75</sup> The partial order is a constraint on word combination: it says that in the argument position next to a given operator the frequency (or probability) of certain words – those not in the argument class for that operator – is zero. Each satisfaction of the partial order, i.e., each word sequence in which all the source words have their requirement satisfied, is a sentence. Hence when a higher operator has another operator as its argument [...] then the higher operator is acting on a whole sentence, which is then a component sentence of the bigger sentence being created by the higher operator[...]

É importante perceber que, do ponto de vista da ordenação parcial, a ocorrência de palavras formam um grupo de elementos arbitrários fechados sob uma relação de dependência com todas as combinações que satisfazem essa relação na formação de uma sentença. Isso se dá, naturalmente, porque os elementos da língua envolvidos, principalmente palavras, não têm propriedades inerentes a elas para serem usadas na construção de sentenças – e aí retorna-se ao pensamento inicial, as palavras em si não são portadoras de significado.

### 3.1.3. Segunda restrição: Probabilidade de Co-ocorrência

Para tratar desta restrição, Harris considera, na descrição da estrutura sentencial, a necessidade de se descrever o mecanismo que faz com que, além de uma classe particular seja selecionada para a composição da sentença, algumas combinações dentro dessa classe sejam mais prováveis que outras. Neste caso, enquanto a primeira restrição cria a estrutura sentencial, esta segunda especifica os significados das palavras na construção.

Na primeira restrição é atestada probabilidade = 0 para palavras que não pertencem à classe exigida na posição argumento. Isso quer dizer que para palavras pertencentes a tal classe, a probabilidade será sempre  $> 0$ , seja ela qual for. Aqui se apresenta a importância da segunda restrição, que demonstra o significado central de uma palavra, que é dado pela seleção de argumentos sob si mesma ou de operadores sobre si mesma. Nesta propriedade, é apontado que, para determinado operador, há palavras que têm uma probabilidade mais alta do que outras de aparecer como seu argumento (e vice-versa)<sup>76</sup>.

Retomando nossos exemplos (2) e (3), para *comer*, como visto, há grupos restritos de argumentos com probabilidade de ocorrência  $> 0$ . Dentre esses, entretanto, há os que Batista (2005b) considera de probabilidade normal, ou seja, os que produziriam as sentenças mais esperadas para esse operador na estrutura linguística. “Uma probabilidade normal de co-ocorrência corresponde a uma *distribuição normal* de um operador perante o conjunto (ordenado) de argumentos” (BATISTA, 2005b, p.17. Itálicos do autor). É, portanto, esse status de distribuição normal que confere significado ao operador.

---

<sup>76</sup> Seja em uma escala que considera probabilidade de ocorrência  $> 0$ , haverá palavras que sempre estarão em uma probabilidade = 10, por exemplo, enquanto outras estarão sempre em uma probabilidade = 1.

Entretanto, em menor ocorrência nesse tipo de construção na língua, vamos encontrar também a entrada desse verbo com argumentos tais como:

(5) *Pedro come o livro com os olhos.*

E, em ocorrências mais raras ainda teremos

(6) *O computador comeu meus arquivos.*

(7) *A escola comeu todas as minhas economias.*

De forma que o operador dado tem a propriedade de selecionar argumentos com força muito maior que 0, e estes vão conferir o significado mais recorrente do operador; mas aceita também, dentro de determinada classe, argumentos que têm ocorrência mais baixa, em uma hierarquia argumental.

Esta restrição, de probabilidades diferentes para diferentes argumentos, restringe a equiprobabilidade de palavras; ela especifica que para palavras com probabilidade  $>0$  na posição argumental de um dado operador, algumas têm [ocorrência] mais alta que a média ou uma frequência muito alta e algumas perto de zero<sup>77</sup> (HARRIS, 1988, p. 20).

Segue-se, assim, que as ocorrências da língua não acontecem de forma aleatória ou com possibilidades iguais para seus elementos formadores de sentenças, ou, como se queira, discursos. Nesse sentido, interessa realçar a propriedade da segunda restrição de fazer com que, na estrutura, seja salientado o significado específico de uma dada palavra numa frase, ao se reconhecer que, para cada palavra-argumento é mais provável aparecerem certas palavras do que outras como operadores sobre ela<sup>78</sup>.

Desde que a restrição de probabilidade seja definida na resultante de 0 versus não 0, quer dizer que não importa quão baixa é a probabilidade de uma palavra particular em relação ao operador dado, a probabilidade nunca é presumida chegar a zero;<sup>79</sup> (HARRIS, 1991, p. 115)

<sup>77</sup>This constraint, of different likelihood for different arguments, restricts the equiprobability of words; it specifies that for words with probability  $>0$  in the argument position of a given operator, some have higher than average or very high frequency and some close to zero .

<sup>78</sup> Como visto, a mesma regra vale para os operadores.

<sup>79</sup> Since the constraint of likelihood is defined on the resultant of 3.1 (0 x non zero), it follows that no matter how low is the likelihood of a particular word in respect to a given operator on it, the likelihood is presumed never to reach zero.

Dessa forma, para o par (*lia, revista*), *ler* é mais provável do que *comer*, e este mais provável do que *assistir*. Quando se diz que esta restrição é responsável pela produção de significado nas frases da língua, é porque entendemos, como Baptista (2005b) que ele é determinado pelas palavras que ocorrem com maior probabilidade como seus argumentos ou como operadores sobre elas (ocorrência normal). Daí decorre que “o significado de uma ocorrência particular de uma palavra é determinado pela selecção das palavras que são seus argumentos ou das palavras que são operadores sobre essa palavra nessa frase”(BAPTISTA, 2005b, p. 4).

### 3.1.4. Terceira restrição: reduções

A terceira das principais restrições de que Harris trata é a que se refere à propriedade de se compactar a frase, resultando em reduções na sentença. O elemento reduzido nesse caso, acrescenta pouca ou nenhuma informação, portanto, seu apagamento não prejudica a informação total da sentença. São diversos os tipos de redução que ocorrem nas sentenças da língua, mas em todos eles fica evidente que o que é redutível é o material linguístico de mais alta probabilidade: quanto maior essa probabilidade de ocorrência, maior também sua propriedade de redução; dessa forma, o ouvinte/leitor é capaz de reconstruí-lo ao compreender a sentença. Isso mostra que a propriedade de redução não é determinada simplesmente pela alta frequência, mas também pela baixa informação do material. “E uma palavra pode ser reduzida se em seu local de entrada ela tem um alto valor de expectativa: por exemplo, se ela tem uma probabilidade excepcionalmente alta de ocorrer naquela posição ou se há muitas sentenças nas quais ela poderia ocorrer naquela posição<sup>80</sup>” (HARRIS, 1982, p. 107).

Importa salientar que esta terceira restrição, assim como a segunda, é gerada a partir da primeira: não contraria a ordenação de partes, coloca-se em consonância com ela.

As reduções não alteram a presença de um componente, apenas sua configuração ou visibilidade na frase. [...] O que esta restrição reducional faz é

---

<sup>80</sup> “And a word may be reduced if at the point at which it enters it has a high expectation value: for example, if it has an exceptionally high likelihood of occurring in that position or if there are many sentences in which it could occur in that position”.

restringir certas sequências de palavras, ou mesmo de fonemas: quer dizer que, dadas combinações de palavras criadas pelas duas primeiras restrições, certas palavras aparecem como sequências fonêmicas diferentes ou não aparecem de modo algum<sup>81</sup> (HARRIS, 1988, p. 22-23).

Nesse sentido, as sentenças reduzidas também não modificam, de forma alguma, o significado em relação a sua sentença fonte<sup>82</sup>, de forma que toda informação obtida na sentença reduzida, é obtida também na sentença de base, pois toda sentença reduzida tem uma sentença de base à qual corresponde, o que é explicitado metalinguisticamente.

As diferentes línguas têm diferentes condições para a aplicação das reduções. No Inglês, por exemplo, ocorre a construção *The dog I saw was running* reduzida de *The dog that I saw was running*, enquanto no Português padrão essa redução não se coloca. Nessa modalidade, é admitida a construção *O cachorro que eu vi estava correndo*, sem possibilidade de supressão da conjunção integrante. Como é possível notar, as reduções têm participado de processos linguísticos que Harris (1988, p. 30-31) afirma ligados à gramaticalização, uma vez que tomam forma e se cristalizam no uso da língua.

Foi apresentado de forma breve como a produção de sentenças na língua não se relaciona apenas à escolha de palavras, mas é uma relação de dependência entre elas, e uma sentença em particular não apresenta apenas “um estado de sentença” mas uma relação de probabilidade de dependência entre palavras.

A essência do que se intentou apresentar aqui se projeta no sentido de demonstrar que cada restrição que contribui para a estrutura da sentença também contribui para o significado de tal sentença, incluindo os significados duplos, as chamadas ambiguidades.

Quando se parte dessa concepção de estrutura linguística e de acordo com a Teoria da Informação, é pertinente que a língua seja notada como veículo carregado de propriedades matematicamente relacionadas à informação. As restrições apresentadas por Harris, agindo umas sobre as outras na produção de sentenças, demonstram a redundância explicitada pela estrutura sintática, característica importante, segundo Shannon (1946), para o cálculo da quantidade de informação.

---

<sup>81</sup> “The reductions do not alter the presence of a word, only its shape and visibility. [...] What this reductional constraint does do is to restrict certain sequences of words, or rather of phonemes: it says that given the word combinations created by the first two constraints, certain words appear as different phoneme sequences or don't appear at all.

<sup>82</sup> a menos que resulte de uma degenerescência na redução, por serem diferentes reduções de diferentes sentenças-fonte. Ex.: *Sue pediu a Ana para cantar. Sue pediu Ana para (Sue) cantar* ou *Sue pediu que Ana cantasse?* Nesse caso, Harris salienta (1991, p. 110) que a ambiguidade de tal sentença não é caso de vaguedade, nem de uma extensão ampla de significado, mas de uma escolha entre dois ou mais sentidos específicos, ligados às sentenças fonte específicas.

Harris mostrou que (a hierarquia de) restrições em combinações de elementos linguísticos têm um efeito cumulativo de criar informação. Tomadas todas juntas, elas abrangem o que pode ser dito como estrutura informacional da sentença, discurso ou sublinguagem assim caracterizada. Fazendo assim, Harris mostrou que foi não apenas um teórico da linguagem, mas também da informação<sup>83</sup> (RYCKMAN, 2002, p. 30).

### 3.2. Metodologias de Harris

O extenso trabalho feito por Harris se apresenta através de métodos que a lingüística adotou a partir dos pressupostos teóricos da descrição. O arranjo de corpus é o elemento principal nesse tipo de trabalho. Harris imprime ao seu método descritivo um caráter comparativo entre os objetos. As estruturas – principalmente as fonêmicas e morfêmicas – são apresentadas muitas vezes em comparação com características estruturais de outras línguas. (cf. HARRIS, 1969[1951], p. 97-124; 130, 166; 1991, p. 80, 169). Essa forma de apresentar elementos e estruturas comuns a duas ou mais línguas, Harris nomeia de Transfer Grammar (HARRIS, 1970[1954], p. 139-157).

Apesar de estabelecer comparações estruturais entre línguas, Harris preocupa-se em explicitar que os dados – adquiridos em usos que têm lugar em determinada comunidade de fala (HARRIS, 1969[1951], p. 1) - se aplicam apenas à língua em questão, a não ser que se queira estudar as diferenças ou similaridades entre estruturas percebidas pelo uso de um mesmo método.

Um ponto importante a ser notado nos métodos de Harris é a consolidação de um método de investigação de estrutura de língua que percebe essa estrutura como um produto social e cultural (e.g., restrições), que condiciona e é condicionado pelo conjunto de usos da língua em um dado período. Assim, pode-se esperar importantes resultados a partir do uso de estruturas lingüísticas completas (e. g. a relação da língua com a cultura).

Em seu primeiro trabalho publicado, *A Grammar of the Phoenician Language* (1990[1936]), Harris apresenta a formação da gramática da língua Fenícia pontuando as

---

<sup>83</sup> Harris showed that the (hierarchy of) restrictions on combinations of linguistic elements have a cumulative effect of creating information. Taken as an aggregate, they comprise what can be said to be the informational structure of the sentence, discourse, or sublinguagem so characterized. In so doing, Harris showed that he was not only a theorist of language but also one of information.

diferenças dialetais das colônias (HARRIS, 1990[1936], p. 67-70; 1-10) - no que se refere à pronúncia, gramática e vocabulário - em relação à língua falada na metrópole. Na verdade “no arranjo dessa gramática, o primeiro lugar é sempre dado à língua padrão das cidades fenícias. As evoluções dialetais são notadas separadamente”<sup>84</sup> (HARRIS, 1990[1936], p. 11). Os métodos de trabalho aqui foram adotados de acordo com a realidade que se apresentava, de forma que nem sempre era possível seguir um padrão definido (cf. HARRIS, 1990[1936], prefácio). Para esse trabalho, contou com diversas fontes, diretas e indiretas<sup>85</sup>, como inscrições em tabletes, moedas e manuscritos; e através da comparação, descobriu similaridades entre a língua fenícia e a hebréia (HARRIS, 1990[1936], p. 39-64).

É acreditando nas possibilidades oferecidas pela descrição que Harris desenvolveu formas diferenciadas de fazê-la. Uma delas foi a divisão dos componentes lingüísticos em classes, que chamou de classes de morfemas, o que posteriormente denominou-se classes de objeto.

### 3.3. Classes de Morfemas

Trabalhar com operadores e argumentos propiciou a Harris obter um sistema de famílias, que foram agrupadas por critérios de combinação e seleção, através do modelo de operações matemáticas (cf. HARRIS, 1968, p. 178). Embutido nessa noção dos operadores, Harris sistematizou o que foi chamado mais tarde de classes de objeto (GROSS, 1981; LE PESANT, D. & MATHIEU-COLAS M., 1998) em que cada ‘objeto’ lingüístico é discriminado e, por suas características sintático-semânticas, incluído em determinada classe. Essa noção de classes de objeto tornou-se produtiva e embasa diversos estudos recentes que tratam dessas mesmas categorias (VALE, 2001; 2004; GROSS, 1994) utilizando-as, por exemplo, em estudos sobre Processamento de Linguagem Natural. O conceito de classes de objeto, como outros que Harris desenvolveu a partir da Lógica (cf.

---

<sup>84</sup> “In the arrangement of this grammar first place is always given to the standard language of the Phoenician cities. Dialectal developments are separately noted.”

<sup>85</sup> Como fontes diretas Harris se serviu de material arqueológico, como citados no texto. As fontes indiretas se compunham de material literário relativo à língua Fenícia provindo de outras línguas. Uma das que Harris utilizou, e que entende como tendo algumas limitações, foi o Velho Testamento. As mais produtivas, encontrou na literatura grega e latina (cf. HARRIS 1990[1936]).

FREGE, 1978, p. 108; LACEY, 1978), objetiva produzir generalizações de como as sentenças das línguas podem ser representadas, quais seqüências de morfemas ocorrem e em que ambientes. Isso vale tanto para morfemas simples quanto para seqüências de morfemas.

Segundo Harris(1970[1946]), o método é apenas uma formalização do que todo lingüista faz em seu trabalho. Na verdade, ele diz não introduzir novas operações de análise. Apenas sistematiza o que já vem sendo feito. Nenhum elemento além dos morfemas e seqüência de morfemas<sup>86</sup>, e nenhuma operação além de substituição. “Ele [o método] apenas reduz à escrita as técnicas de substituição que todo lingüista usa quando trabalha com seu material. Pode-se trabalhar mais eficientemente quando se pensa com lápis e papel”<sup>87</sup> (HARRIS, 1970[1946], p. 123).

Especificamente em relação a essas classes morfemáticas, Harris desenvolveu um método em que demonstrou como, na estrutura lingüística, é perceptível a forma de combinação dos elementos através das classes de morfemas. Na noção de classe de objeto está explícita uma classificação semântica que é feita a partir de critérios de ordem sintática. Assim, na estrutura lingüística há mecanismos que fazem com que as classes de objeto se encaixem em determinadas fórmulas, que oferecem a noção clara de limitações/liberdade de ocorrência dessas classes de objeto. Nenhuma seqüência de morfemas ocorre, segundo Harris(1970[1946]), se não se encaixar na fórmula.

Dessa maneira foram elencadas, para o Inglês, classes que envolvem os verbos(V), substantivos (N), adjetivos (A), preposições (P) e todas as demais classes de palavras presentes na língua, inclusive os morfemas livres. Apesar de se embasar na tradicional lista de elementos morfológicos, essa tipologia se difere por descrever sua distribuição, ou seja, as possibilidades de ocorrência nos ambientes frasais. Exemplificando:

N: morfemas que ocorrem antes do plural *-s* ou seus alternantes, ou depois de *the* ou de adjetivos → *hotel, butter, gain*.

Para os verbos, foram elencadas cerca de oito classes de morfemas verbais. Alguns exemplos são:

---

<sup>86</sup> No caso das seqüências de morfemas, elas não configuram uma nova classe, diferente da de morfemas simples. Essas seqüências de morfemas são agrupadas com as demais classes já estabelecidas e reduzidas ao status de morfemas simples. (cf. HARRIS, 1970[1946], p. 104)

<sup>87</sup> “It merely reduces to writing the techniques of substitution which every linguist uses as he works over his material. One works more efficiently when one thinks with pencil and paper.”

V: têm ocorrência antes da forma *-ed* que indica pretérito, ou suas formas alternantes; antes de *-ing*; depois de N acrescentado de *should, will, might, etc.* → go, take, think.

A partir dessa classe, foram elencadas várias subclasses, que apresentam outros traços e, por isso, necessitam de outro critério distribucional. Dois exemplos são:

V<sub>c</sub>: ocorrem entre N e V *-ing* → *stop, try, be.* → Max will stop walking

V<sub>d</sub>: verbos transitivos que ocorrem antes de N → *make, buy, want.* I'll buy butter

V<sub>e</sub>: verbos intransitivos que não ocorrem antes de N → *sleep.*

Um ponto que interessa notar nesse método de Harris (1970[1946], p. 109-114; 1951, p. 125-149) é a formulação de construções lingüísticas através de equações. São colocadas as possibilidades de permuta de classes de objeto através de raciocínio matemático - como em equações de representação e substituição<sup>88</sup> - em que uma classe somada a outra se iguala a uma terceira em possibilidade de ocorrência. Isso possibilita a compreensão de como diferentes classes de objeto podem ter igual distribuição em ambientes sentenciais similares. Desse modo, explicam-se equações de formação de palavras, como:

$$1) \quad A^1 - A_n = N^1$$

Em que N<sup>1</sup> = nome substantivo que se enquadra na classificação de N<sup>89</sup>

A<sup>1</sup> = adj.

-A<sub>n</sub> = sufixos, tais como -ness, -ty

<sup>88</sup> Neste caso, os morfemas passam por operações tais quais as das equações matemáticas de representação e substituição, em que a soma/ subtração entre dois termos, X e Y, resulta em um terceiro termo Z. Se temos, por exemplo, X+Y=Z, prossegue-se com a operação, com os valores atribuídos a duas das incógnitas, fazendo a substituição para encontrar valor da terceira. Essa operação, ao final, apresentará uma igualdade entre as casas esquerda e direita da equação.

<sup>89</sup> Ao representar uma classe em potência, Harris mostra que aquela classe está inscrita na classe principal de morfemas, somada a outro tipo de morfema - o que chama de "seqüência de morfemas"-, por isso a distinção.(cf. HARRIS, 1970[1946], p. 100-123)

Assim, se temos, por exemplo, para  $A^1$ , *dark*, e para  $A_n$  *-ness*, teremos à esquerda da equação, *darkness*, que é um N em Inglês, portanto, há igualdade entre os termos. Da mesma forma acontece em *below*, representada pela equação

$$2) \quad A_p - A^1 = P$$

Em que  $A_p$  = prefixo *-be* que antecede adjetivos, podendo transformá-los em preposições.

$A^1$  = classe geral de adjetivos

$P$  = preposições que têm ocorrência em determinados ambientes, dentre eles, antes de *the*<sup>90</sup>

Essas equações são importantes no sentido de mostrar por que em um mesmo ambiente podem ocorrer diferentes tipos de morfemas, como em 1:

1.a) I don't like the *darkness* here

Em que pode ocorrer outro N, uma vez que  $A^1 - A_n = N^1$ .

1.b) I don't like the *smell* here

Ou em 2

2.a) It fell *below* the dividing line

ambiente em que também pode ocorrer

2.b) It fell *at* the dividing line, sendo que  $A_p - A^1 = P$

Isso significa que diferentes classes de objeto podem ter igual distribuição em ambientes sentenciais similares.

---

<sup>90</sup> São elencadas por Harris diversas posições frasais em que essa classe de morfemas ocorre. Entretanto, como não entra no escopo deste trabalho a descrição exaustiva das classes de objeto, tomamos para exemplificação apenas as que se fazem necessárias para uma compreensão básica do método proposto.

Seguindo no método, Harris continua estabelecendo classes de equivalência também para seqüências de sentenças e de construções textuais, de forma que é possível afirmar, com o uso dessas classes, quais seqüências de morfemas ocorrem em construções do Inglês, bem como de outras línguas, o que possibilita além da descrição da sintaxe das línguas, a comparação de estruturas entre elas.

O resultado final, para cada língua que pode ser analisada desta maneira, apreende a forma de uma ou mais seqüências de classes de substituição ('construções de expressões', 'tipos de sentenças'). A fórmula nos mostra que essas são seqüências que ocorrem. A fórmula final portanto nos dá as limitações de liberdade de ocorrência de morfemas na língua (HARRIS, 1970[1946], p. 117).

### 3.3.1. Abordagem Matemática da Língua

Como já dito anteriormente, uma das particularidades da metodologia de Z. Harris é a forma matemática de abordagem lingüística, em que o estudo e representação de componentes lingüísticos é feito com o uso de linguagem e metodologia matemáticas.

Esse método se fez de tamanha importância para Harris, que foram dedicadas apenas a ele três obras completas, *A Grammar of English on a Mathematical Principles*(1982), *Mathematical Structures of Language* (1968) e *A Theory of Language and Information: a mathematical approach* (1991). Nessa abordagem, os constituintes arbitrários da língua estão dispostos em um sistema de grupos que mediante operações específicas são interpretáveis no uso.

A idéia aplicada aqui não é a da linguagem matemática dos axiomas sendo compreendida através da língua natural, como se poderia supor num primeiro momento, mas os fatos da língua natural sendo relacionados a construções matemáticas.

Isso é bem mais abrangente do que relacionar a língua a determinada matéria dentro do universo matemático. Os próprios termos lingüísticos da área da matemática são usados assim para descrever o método lingüístico. (e. g. fatoração, vizinhanças, equação). Harris procura demonstrar sua hipótese de que a língua tem uma estrutura matemática subjacente, ou seja, os axiomas podem ser utilizados para demonstrar/representar as construções lingüísticas, desde a combinação dos fonemas até a composição dos morfemas, expressões, sentenças, chegando aos textos (cf. HARRIS, 1970[1952]). Assim, é utilizado o raciocínio

matemático, que envolve leis como a de probabilidade, ou mesmo a representação notacional, a serviço da linguística.

Algumas razões são salientadas por Harris para que se estude a língua sob esse prisma, levando em consideração o que as duas áreas têm em comum: uma delas é a combinação, aplicável por exemplo, à análise da adaptação de novas palavras ao sistema. Ao entrarem no sistema lingüístico, as novas palavras precisam se adequar ao sistema já existente de combinação de palavras. Se acontece assim, as palavras já existentes têm que ser caracterizadas por suas relações de combinação entre si muito mais do que outras propriedades individuais. Assim, as novas palavras podem ter um tipo de significado ou composição fonêmica diferente das palavras já existentes, mas terá que se encaixar em um ou outro tipo de combinação de palavras que formam sentenças na língua (HARRIS, 1991, p. 145-159).

Outra característica em comum, segundo Harris, seria a possibilidade de produzir elementos em número infinito a partir de um número finito conhecido. Isso se aplicaria à limitação de experiência que os falantes de uma língua têm em relação ao conhecimento de entidades discursivas e suas combinações na língua, entretanto, podem entender um ilimitado número de sentenças, desde que não haja limitações estruturais consideráveis em sua construção. É necessário haver características construtivas ou recursivas de sentenças.

Para que seja possível esse diálogo entre o sistema lingüístico e o sistema matemático, seria então necessário haver propriedades comuns aos dois sistemas, o que tornaria relevante a relação entre eles. Harris aponta algumas das propriedades lingüísticas que se fazem presentes também na matemática, dentre elas:

- **Discrição:** refere-se aos elementos discretos da língua, os componentes morfêmicos e os elementos gramaticais em geral. “Pela análise da língua nota-se que apenas as entidades que têm regularidade gramatical são discretas. Os fonemas são elementos discretos e palavras e sentenças são formadas de seqüências específicas deles.”(HARRIS, 1991, p. 145)
- **Composição:** As distinções fonêmicas e que constituem as palavras que são utilizadas pelos falantes de uma língua são elementos que se colocam em um esquema de composição. Da mesma forma as possibilidades de ocorrência de determinados morfemas, palavras e expressões, acontecem mediante leis lingüísticas de combinação. É como se fosse uma função composta em matemática.

- Sinais arbitrários: aqui os componentes lingüísticos são vistos como elementos gramaticais que precisam ser agrupados. Isso confere ao sistema o que tem sido chamado caráter arbitrário dos sinais lingüísticos (ex: palavras diferem de uma língua para outra); segundo Harris, seria interessante para os usuários se fosse possível eliminar a necessidade de se compreender uma quantidade de elementos convencionados. Mas a quantidade fixa com distinções fonêmicas é exatamente o que é requerido para a transmissibilidade sem composições inadequadas (HARRIS, 1991, p. 148).
- Linearidade: refere-se à propriedade da língua de ordenar, linearmente, sons, seqüência de palavras e sentenças.
- Gramática Finita:

Os fonemas e o vocabulário, e os modos de combinação de palavras em sentenças, todos tem de ser usados pelos falantes de uma língua de forma que os permitam entender um ao outro. Os usuários, e suas experiências com a língua são todos finitos. Por conseguinte, é necessário ser possível caracterizar a língua por uma gramática finita: uma relação finita de elementos e também de regras de combinação com algumas regras sendo recursivamente aplicadas no fazer uma sentença<sup>91</sup> (HARRIS, 1991, p. 150).

- Por conta de propriedades como essas, a estrutura da língua seria então computável, representável e analisável através de algoritmos. É uma tarefa, segundo Harris, algo assustadora, mas possível. “As propriedades lingüísticas examinadas[...] fazem possível afirmar várias operações definíveis matematicamente e vários grupos aproximados por relações matemáticas que juntos caracterizam gramaticalmente e semanticamente quase todas as estruturas universais importantes da língua”<sup>92</sup> (HARRIS, 1991, p. 153).

Através de métodos como os apontados acima, Zellig Harris se consagrou dentro dos estudos estruturais muitas vezes chamados neo-Bloomfieldianos. Desse modo, Harris lança noções importantes através da reelaboração de conceitos que, como visto, tiveram lugar em outras abordagens, anteriores e/ou contemporâneas a ele, abordagens lingüísticas ou não. Interessa notar que essas reelaborações - quando não se fazem alicerces para - ao menos lançam luz sobre investigações de diversos campos da lingüística na atualidade.

---

<sup>91</sup> The phonemes and vocabulary, and the ways of combining words into sentences, all have to be used by the speakers of a language in a way that enables them to understand each other. The users, and their experiences with the language are all finite. Hence it must be possible to characterize a language by a finitary grammar: a finite stock of elements and also of rules of combination, with, at least some of the rules being recursively applicable in making a sentence.”

<sup>92</sup> “The properties of language surveyed [...] make it possible to state various mathematically definable operations and various sets closed under mathematically definable relations, which together characterize almost all the grammatically and semantically important universal structures of language.”

Estudos que se referem à teoria transformacional, bem como à recuperação de informação na língua encontram em pressupostos de Harris (e.g., as noções de *sublinguagem*, metalinguagem e informação) um embasamento consistente. A obra do lingüista se mostra fecunda em abordagens, por exemplo de aspectos lexicais da língua, estudos que hoje têm um aprofundamento e um direcionamento para as necessidades lingüísticas atuais. Na sessão seguinte, passaremos a apresentar a abordagem harrissiana do conceito de sublinguagem, buscando entrever sua relação e produtividade dentro dos estudos lingüísticos atuais.

## 4. O Conceito de Sublinguagem em Harris

### 4.1. Problemática Inicial

Ao iniciar a apresentação desta abordagem, deparamo-nos com uma problemática inicial e bem elementar: a definição de tradução do termo *sublanguage*. O fato de que, no Inglês, a palavra *language* se refere tanto a língua quanto a linguagem, e de que no Português, há termos distintos para as distintas concepções, faz com que haja um complicador, quando se trata de discorrer sobre o termo *language* e correlatos.

Ora, dessa forma, é relevante refletir um pouco sobre o binômio língua/linguagem antes de prosseguirmos. A definição de língua para Harris é bem abrangente. Para ele a língua é “uma atividade pública, com os elementos sonoros e elementos escritos predeterminados por convenção dentro de uma comunidade, de uma maneira que não seria necessária para expressão privada ou mesmo para comunicação interpessoal”<sup>93</sup> (HARRIS, 1988, p. 87). E continua, afirmando que

a língua é uma estrutura exigente: considerando algumas coisas como estando nela e por conseguinte, certas, e algumas coisas não estão nela e são portanto, erradas.[...] Mas é importante que se saiba que as exigências da língua, as regras da gramática, são acessíveis como o produto final de um processo de institucionalização produtiva, de algo que não é exigente e único. E, naturalmente, o processo de institucionalização em si mesmo não é de forma alguma único, sendo largamente conhecido – bem ou mal – na cultura e na organização social <sup>94</sup> (HARRIS, 1988, p. 111).

Através dessas definições, Harris reconhece o estatuto da língua como sendo o de um sistema social institucionalizado e ordenado por regras seletivas de ordem lexical e gramatical, através das quais são formados os enunciados.

---

<sup>93</sup> “... a public activity, with the sound elements and word elements preset by convention within a community, in a manner that would not have been needed for private expression or even for interpersonal communication.”

<sup>94</sup> ... language is a demanding structure: some things are regarded as being in it and therefore right, and some things are not in it and are therefore wrong. But it is important to know that the demands of language, the rules of grammar, are reachable as the end product of a process of institutionalization, from something that is not demanding and not unique. And of course, the process of institutionalization itself is by o means unique, being widely known – for better of for worse – in culture and in social organization.

Por essas definições, é possível chegar a uma compreensão de que, ao tratar de sublinguagem, o linguista se volta para o trato de fenômenos particulares dentro do sistema da língua, enquanto estrutura e também em relação à relevância de ocorrência no sistema (e.g., as sublinguagens científicas). Pois bem, esses fenômenos particulares adquirem características comuns quando investigados à luz de outros sistemas lingüísticos – os estudos de sublinguagens têm apresentado características gramaticais comuns quando feitos em diferentes línguas, - e.g., estudos feitos no Inglês e no Francês revelam uma mesma estrutura gramatical da sublinguagem da Imunologia (cf. HARRIS, 1988, p. 41; KITTREDGE & LEHRBERGER, 1982, p. 4).

Entendemos assim que a sublinguagem está em subgrupos de produções lingüísticas dentro de uma língua, e que possuem uma forma similar correspondente em outra língua. É o que ocorre, por exemplo, quando um estudioso da área de lingüística que é falante/leitor da língua inglesa, ao ter contato com textos de Lingüística no Inglês, terá acesso mais facilitado ao conteúdo, pois o texto está em uma linguagem que se constrói com muitas similaridades, tanto no Inglês quanto em Português. Harris diz ainda que o estudo de sublinguagens

repousa em uma ainda pouco explorada relação relativa a conceitos vizinhos de gíria, registro, estilo e dialeto em investigações sociais da língua. Os tipos de padrões de recorrência encontrados em vários discursos podem servir para distinguir tipos textuais ou gêneros, e.g., narrativas, observações científicas, questionários<sup>95</sup> (HARRIS *apud* GOTTFRIED, 2002, p. 89).

Aqui o linguista refere-se a ‘investigações sociais da língua’, o que nos leva a considerar que, para o Português, seria um ponto extremo nomear o fenômeno de sublíngua, porque dessa forma seriam consideradas apenas características da forma para essa nomeação. São consideradas sim as características estruturais, mas para a investigação da estrutura das sublinguagens. No que se refere a delimitá-las, entretanto, nota-se um componente que vai além de sua estrutura apenas.

---

<sup>95</sup> bear an as yet little examined relation to the neighboring concepts of argot, register, style, and dialect in social investigations of language. The types of recurrence patterns found in various discourses may serve to distinguish text-types or genres, e.g., narratives, scientific observations, questionnaires.

Tendo em vista estas considerações, a utilização da palavra *sublanguage* para se referir a domínios distintos dentro do sistema lingüístico, nos daria subsídio para talvez nos referirmos, em Português, a diferentes sublinguagens, tendo em mente que

O termo sublinguagem veio para ser usado não apenas para qualquer subgrupo marcado de sentenças que satisfazem a propriedade de fechamento, mas para aqueles grupos cujas restrições gramaticais refletem os grupos restritos de objetos e as *relações* encontradas em um dado domínio de discurso”<sup>96</sup> (KITREDGE & LEHRBERGER, 1982, p. 2 *itálico nosso*).

Assim seria: a linguagem dos bilhetes informais, a linguagem jornalística, a linguagem da pesquisa científica, e dentro desta, a linguagem científica da Química, da Física da Matemática, da Biologia, da Lingüística e assim por diante.

Com efeito, o objeto das investigações de Harris são características sistêmicas presentes na sublinguagem, (e.g., a estrutura gramatical). Se tivermos apenas isso em vista, o termo mais apropriado para o Português talvez fosse *sublíngua*, tendo claro que esta não seria uma ‘língua menor’, um apêndice do sistema maior, mas sim uma parte contida nesse sistema com características lexicais e gramaticais próprias.

Entretanto, o termo não seria suficiente para expressar essa diferença de domínios, abarcando ao mesmo tempo as características formais e também as do nível da manifestação social, que o próprio Harris aponta. Ademais, os fenômenos de ordem gramatical explicitados nessas sublinguagens, como dissemos acima, ultrapassam o domínio de uma só língua. Eles estão presentes em todas as línguas que apresentam sublinguagens específicas, eminentemente a científica. Harris (1991, p. 290; 2002a p. 219) sugere que as sublinguagens podem ser vistas como um sistema lingüístico distinto, por terem configuração lingüística similar em diferentes línguas. “As classes de palavras e símbolos de subclasses, e os tipos de sentenças, são por conseguinte não apenas uma sublinguagem de uma linguagem particular, mas um sistema lingüístico simbólico independente”<sup>97</sup> (HARRIS, 2002a, p. 219). Então, consideramos que podem ser entendidas como um tipo de linguagem, que é compartilhada, pelo menos em certos aspectos estruturais, por diferentes sistemas lingüísticos.

---

<sup>96</sup> The term sublinguagem has come to be used not just for any marked subset of sentences which satisfies the closure property, but for those sets of sentences whose lexical or grammatical restrictions reflect the restricted sets of objects and relations found in a given domain of discourse.

<sup>97</sup> “The word class and subclass symbols, and the sentence types, are therefore not just a sublinguagem of a particular language, but an independent symbolic linguistic system”.

Por conta dessas intersecções de fórmulas estruturais em comum das sublinguagens científicas, acreditamos realmente que o termo *sublanguage* aproxima-se mais de sublinguagem do que de sublíngua, pois este último teria traços particulares – que as sublinguagens têm, sem dúvida – mas que não se estenderiam para além da fronteira de uma língua. “Ademais, a sublinguagem para uma dada ciência em uma língua é similar à sublinguagem para aquela ciência em outra língua, então seu status de sublíngua do Inglês, Francês ou outra, não é sua principal caracterização” (HARRIS, 1989, p. 80).

Outra característica da sublinguagem é que a distingue da linguagem como um todo resulta do fato de ela não ter uma metalinguagem interna, já que a sublinguagem é constituída por operações externas que a delimitam; Essas operações são a sua metalinguagem e fazem parte da linguagem geral. Por sua vez, a linguagem no seu todo não tem uma metalinguagem externa (cf. HARRIS, 1988, p. 38, 103; 1991, p. 272). Nesse sentido, a metalinguagem de uma sublinguagem é uma parte da linguagem geral que serve para determinar essa dada sublinguagem<sup>98</sup>.

Entendemos, através da descrição de língua do próprio Harris, que a sublinguagem, para ser uma sublíngua, teria, da mesma forma que a língua, e por extensão, sua própria metalinguagem – a ‘submetalinguagem’. Uma vez, porém, que sua metalinguagem é emprestada do sistema lingüístico maior, pensamos ser bem pouco provável que *sublanguage* refira-se, strictu sensu, a sublíngua.

Se tivermos em mente a discussão de Saussure (1994[1916], p. 142-145) sobre língua, também sobre as relações sintagmáticas e associativas nela, no que ele nomeia ‘unidade lingüística’, não será difícil chegar à conclusão de que, ao considerar os dois componentes da língua como estando juntos, teremos apenas linguagem, sem a necessidade de dissociar estrutura de expressão lingüística. Assim, entendemos que, para efeito de terminologia, é desnecessário e, diria, desinteressante dissociar língua e linguagem, porque para os fins deste estudo, observamos que a primeira está a serviço da segunda.

Dessa forma, optamos por utilizar o termo sublinguagem; e importa deixar claro que não deixam de estar imbricadas aqui tanto a concepção de língua quanto a de linguagem, pois estamos considerando as sublinguagens como formas de expressão que têm lugar nos ditos domínios específicos dentro do sistema lingüístico, como modos particulares de realização da língua dentro do sistema lingüístico maior. Em outras

---

<sup>98</sup> Jorge Baptista, em comunicação pessoal.

palavras, é uma manifestação da linguagem dentro do universo de uma língua natural. Esse é um traço da abordagem linguística de Harris, o sentido de *language* – e, conseqüentemente, *sublanguage* - não se restringir apenas à língua natural<sup>99</sup>.

Tendo em vista o que foi posto, deste ponto em diante, e para a teorização de Harris, utilizaremos o termo *sublinguagem*, o que assumimos como um signo arbitrário nesta proposta, até por entender que ele não é excludente. Ao contrário, abrange fenômenos que se referem à estrutura e da mesma maneira, às formas de realização dessa estrutura em determinados grupos; então refere-se a algo que seriam *sublínguas* e a suas distintas linguagens que se cristalizam através de signos convencionados dentro do sistema principal. Assim, entendemos como Gottfried (2002, p. 89), que “sublinguagens podem ser consideradas uma consequência de hábitos compartilhados de usos de palavras formados em atividades especiais de vários subgrupos, parte da ‘divisão do trabalho sociolinguístico’, obtido em muitas comunidades de discurso”<sup>100</sup>.

#### 4.2. Definição

Segundo Kittredge (1982), quando uma língua natural é usada em um ambiente restrito, pode-se referir a tal ambiente como a uma *sublinguagem*. Complementando essa conceituação, aponta outros fatores presentes em tal ambiente e que o caracterizam como *sublinguagem*, quais sejam:

- Domínios restritos de referência: neles o grupo de objetos e relações a que as expressões linguísticas se referem é relativamente limitado.
- Orientação e propósitos restritos: a relação entre os participantes são de um tipo particular e o propósito da troca é orientado através de certos objetivos.
- A comunidade de participantes compartilha conhecimentos especializados.

É uma definição que também está presente em Harris, em sua abordagem de *sublinguagem*. Em relação a esse conceito e à guisa de defini-lo, ele diz que “as distintas situações sociais ou ocupacionais são na maioria dos casos reconhecíveis como tratando de

---

<sup>99</sup> Jorge Baptista (Comunicação pessoal) nos alertou sobre essa abrangência de conceitos.

<sup>100</sup> “Sublinguagens may be considered a consequence of the shared habits of word usage formed in the special activities of various subgroups, part of the ‘sociolinguistic division of labor’ obtaining in many speech communities” (GOTTFRIED 2002, p. 89)

objetos distintos, mais caracteristicamente nas várias subciências, mas também em situações como regulamentos, artigos de opinião, ou até mesmo [mensagem] heráldica”<sup>101</sup> (HARRIS, 1991, p. 272).

Como teoria, esse conceito foi desenvolvido algum tempo depois da produção de seus primeiros escritos, que se direcionavam, à época, mais à descrição dos componentes lingüísticos relativos à língua como um todo, que ele nomeia linguagem geral. Esta expressão, aliás, é a que é usada para contrapor o ambiente de sublinguagem à linguagem como realização do todo lingüístico. Em todos os textos de Harris que se referem a esse assunto, são impressos conceitos relativos à sublinguagem realçando-a em relação à linguagem geral.

Em relação à abordagem inicial do assunto pelo autor, observamos que em seu clássico – e tido por muitos como seu primeiro trabalho - a obra *Structural Linguistics* (1951), o assunto não é investigado ainda. Nessa obra, ao falar dos diferentes estilos de discurso, o autor se limita a discutir sobre o que mais tarde seria um campo para o desenvolvimento de sua teoria de sublinguagem.

Pode ser possível mostrar, em muitas línguas, que há diferenças em estilo ou moda de discurso em relação às quais os usos comuns e mesmo discursos são consistentes. Esses estilos podem ser relativos a várias situações culturais e interpessoais.[...]; podemos considerar estilos os que marcam falantes particulares ou grupos de falantes socialmente diferenciados (e.g., estilo das garotas adolescentes), estilos que marcam tipos particulares de relação interpessoal<sup>102</sup> (HARRIS, 1969[1951], p. 10).

O tema da sublinguagem só vem à tona posteriormente, no *Dyscourse Analisys* (1970[1952]), em que o conceito já pode ser percebido – embora o autor não lance mão ainda dessa categorização – quando Harris apresenta uma proposta de análise textual, através da qual mostra a produtividade de um método criado por ele, no qual se faz possível reconhecer, estruturalmente, o tipo de texto que é escrito. Dentro da proposta de investigação das sublinguagens, vale lembrar que a análise descritiva de Harris se faz tendo

---

<sup>101</sup> “The distinct social or occupational situations are in most cases recognizable as dealing with distinct subject matters, most characteristically in the various sub-sciences, but also in such situations as law cases, opinion surveys, or even heraldry.”

<sup>102</sup> It may be possible to show, in many languages, that there are differences in style or fashion of speech, in respect to which whole utterances or even discourses are consistent. These styles may be related to various cultural and interpersonal situations[...], we could. consider styles which mark particular speakers or socially differentiated groups of speakers (e.g. adolescent girls’ style), styles which mark particular types of interpersonal relation.

como material somente corpus de textos escritos. Igualmente, quando o autor se refere a ‘discurso’, está considerando o componente texto (HARRIS, 1970[1952]).

Na obra *Mathematical Structures of Language* (HARRIS, 1968, p. 152-156) o linguista discute como a gramática descritiva das sentenças de uma língua pode se revestir de nuances diferenciadas quando utilizada a serviço de ambientes textuais em áreas específicas, trazendo já - mesmo que em seu status ainda em formação - o conceito de sublinguagem.

Em uma obra posterior, *A Grammar of English on Mathematical Principles* (HARRIS, 1982), o conceito de sublinguagem é apresentado direcionado especificamente para a sublinguagem da própria língua, que é chamada de Sublinguagem de Base. Aqui as considerações se estendem às particularidades da gramática da língua Inglesa.

Já no *Language and Information* (HARRIS, 1988), são apresentadas, em linhas gerais, as características das sublinguagens, de forma que, além da definição do termo, e de apresentar a sublinguagem da língua, nessa obra o autor aprofunda mais a discussão sobre esse componente presente nos diferentes domínios lingüísticos. São apresentadas diferentes formas de sublinguagem (e.g. a sublinguagem de domínios específicos dentro da língua; a linguagem de notação, específica da representação científica, que se faz também uma sublinguagem). Aqui o autor traz a resenha de uma investigação feita por um grupo de lingüistas, com a colaboração de pesquisadores da área de imunologia, assunto sobre o qual falaremos mais adiante. Essa investigação é também apresentada em obras posteriores: *A Theory of Language and Information: a mathematical approach* (HARRIS, 1991), em que Harris trata, entre outros assuntos, também do tema sublinguagem, relacionando-a a sistemas notacionais, como a música e as representações dos axiomas matemáticos.

As obras *The form of Information in Science: analysis of an Immunology Sublanguage* (HARRIS et. al. 1989) e *The Structure of Science Information* (HARRIS, 2002) tratam exclusivamente da noção de sublinguagem. O estudo de Harris da sublinguagem, eminentemente da científica, sugere uma abordagem da estrutura sintática da ciência baseada na sintaxe das sublinguagens usadas para se produzir determinado tipo de texto. Essas sublinguagens são formadas tendo, para expressar seus conceitos e relações, uma metalinguagem específica dentro da língua (cf. HARRIS, 1988, p. 38).

Por exemplo, os conceitos da Química e da Biologia têm uma forma própria de se colocarem, com expressões específicas das referidas áreas. Mas ao mesmo tempo em que

têm essa linguagem própria, não há uma metalinguagem interna através da qual possam se constituir, seu meio de transmissão lingüística está fora desses domínios. É uma metalinguagem particular, mas para que aconteça é necessária uma outra, que só é possibilitada por conta da característica da língua de se referir a qualquer entidade, inclusive a ela própria (HARRIS, 2002b, p. 8-10; 1988, p. 35). E, nesse caso, a metalinguagem tida como para falar de diferentes domínios é um subgrupo de sentenças específico, constituindo assim uma sublinguagem (HARRIS, 2002b, p. 10).

A metalinguagem de cada sublinguagem, sua gramática, não seria um subgrupo dela mesma, como acontece na *linguagem geral*. Ela seria expressa por um grupo de sentenças da língua (Português, Inglês etc.), que por sua vez seria externo à sublinguagem que descreve. A partir dessa realidade, Harris afirma:

Em geral, a investigação de um campo, e a definição de suas entidades, é feita em uma metalinguagem para esse campo, numa linguagem de capacidade informacional mais ampla que a do campo dado. [...]  
Ter uma metalinguagem externa faz uma diferença. Isso significa que definir os grupos de palavras e a estrutura da sublinguagem não é restrito a uma regularidade combinatorial interna da forma que temos visto para a linguagem toda. Qualquer grupo de palavras ou seqüências de palavras pode ser definido(a) para a sublinguagem na metalinguagem daquela sublinguagem<sup>103</sup> (HARRIS, 1988, p. 2; 38).

Tendo em vista a definição feita do que é, para Harris, sublinguagem, a questão que se coloca para nós seria: como pode a análise lingüística formal lançar luz sobre uma área específica de realização da língua, em textos escritos, que se destacam por suas particularidades lexicais<sup>104</sup> – e, conseqüentemente, sintáticas e semânticas – auxiliando a compreensão de como são estruturados os textos nesses domínios? Acreditamos que até o final deste capítulo esta questão será esclarecida, em uma apresentação que tem como base a relação que as sublinguagens apresentam, antes de tudo, com a estrutura da língua como um todo, o que é perceptível na forma como Harris embasa sua investigação, a saber, na teoria sintática de operadores e argumentos (HARRIS, 1988, p. 33).

Dentro do sistema lingüístico que confere a certo tipo de linguagem o status de sublinguagem é o fato de se compor de um subgrupo de sentenças da língua que são

---

<sup>103</sup> Having an external metalanguage makes a difference. It means that defining the word sets and structure of the sublinguagem is not restricted to internal combinatorial regularity in the way we have seen for the whole. Any set of words or word sequences can be defined for the sublinguagem in the metalanguage of that sublinguagem.

<sup>104</sup> Kittredge (1982, p. 80), enfatiza que “O traço mais óbvio de uma sublinguagem é seu léxico especializado”.

fechadas sob determinadas operações definidas. “Um subgrupo de sentenças de uma língua constitui uma sublinguagem dessa língua se ela for fechada sob algumas operações da língua”<sup>105</sup> (HARRIS, 1988, p. 34). Nesse sentido, Gottfried (2002) enfatiza que as especificações gramaticais das sublinguagens requerem afirmações explícitas das operações sob as quais o subgrupo de sentenças é fechado. Sendo assim, torna-se possível perceber as similaridades e diferenças gramaticais entre sublinguagem e linguagem geral.

A estrutura que é fixa na gramática<sup>106</sup> das sublinguagens então é adequada a elas e diferente daquela da linguagem, que tem um vocabulário aberto e restrições mais amplas. Nesse sentido, há subgrupos de sentenças de textos que satisfazem algumas condições gramaticais que não são satisfeitas pela linguagem em geral (cf. HARRIS, 1991, p. 279). Isso pode ser exemplificado pela própria metalinguagem do ensino gramatical, ou, como se queira, do ensino de estrutura de línguas. Aliás, Harris salienta dentre os variados tipos de sublinguagem, uma das mais importantes é a metalinguagem própria da língua:

Uma outra sublinguagem estruturalmente distinta é a metalinguagem. Desde que palavras podem se referir a palavras não menos que a outras coisas, podemos investigar todas aquelas sentenças de uma língua que se referem a palavras da mesma língua. O grupo de todas tais sentenças em uma língua é a metalinguagem daquela língua: ela identifica os elementos da língua e suas combinações e relações, e os significados de todos eles<sup>107</sup> (HARRIS, 1988, p. 35).

A linguagem gramatical de uma língua restringe-se a falar de estruturas sintáticas específicas daquela língua. As sentenças, assim, dizem algo sobre sentenças da língua, ou suas funções, ou suas partes. Essas sentenças da sublinguagem gramatical contêm palavras que nomeiam palavras, ou falam sobre elas. Então, consideramos aqui o fato de que essa metalinguagem não apenas está na linguagem, mas ela constitui uma sublinguagem dessa linguagem (cf. HARRIS, 1991, p. 34).

O fato de que a metalinguagem de uma língua natural constitui uma sublinguagem não entra em choque com o fato de que suas sentenças são ao mesmo tempo sentenças da linguagem natural, satisfazendo sua gramática. A

<sup>105</sup> “A subset of the sentences of a language constitutes a sublinguagem of that language if it closed under some operations of the language..”

<sup>106</sup> Para a definição de gramática, adotamos aqui além da de Harris, a de TRASK apud SILVA (2006), que a apresenta como um sistema através do qual as palavras e morfemas de uma língua estão organizados em unidades amplas, particularmente em sentenças.

<sup>107</sup> Another important structurally distinguished sublinguagem is the metalanguage. Since words can refer to words no less than to other things, we can investigate all those sentences of a language that refer to words of the same language. The set of all such sentences in a language is the metalanguage of that language: it identifies the elements of the language and their combinations and relations, and the meanings of all these.

metalinguagem segue descrevendo a estrutura operador/argumento que cria sentenças apenas afirmando-as em sentenças que elas mesmas têm aquela estrutura [...] Em contrapartida, a metalinguagem então satisfaz para caracterizar (ou gerar, se se desejar assim) as sentenças da língua<sup>108</sup> (HARRIS, 1991, p. 275).

Então, “se essas palavras particulares são tomadas como a extensão de novas subclasses  $N_{meta}$  de  $N_n$ , obtém-se a gramática como uma sublinguagem que consiste de todas as sentenças que contém  $N_{meta}$ .” (HARRIS, 1968, p. 152). Tomemos a sentença abaixo:

(8) *A frase ‘estou faminta’ é uma sentença nominal.*

Aqui, o  $N_{meta}$  *A frase estou faminta*, ele próprio contém uma frase em si, e ao mesmo tempo está se automeando ao explicitar que é uma sentença nominal. Assim, há para a descrição de sentenças da língua, um subgrupo de operadores e argumentos que são específicos desse subgrupo e que se configuram em termos técnicos para tratar da linguagem. Isso difere da concepção semântica de Lyons (1980[1977], p. 19) sobre uma forma “técnica” especialmente construída para se tratar da linguagem. Dessa forma, pode-se dizer que a sublinguagem do domínio que produz a metalinguagem das sublinguagens (e. g., a sublinguagem gramatical) também é caracterizada por uma propriedade que difere da linguagem geral.

Também nas sentenças que são características da linguagem de algumas ciências, ao descreverem alguns fenômenos, nota-se que são sentenças gramaticalmente adequadas em se tratando da sintaxe da língua, entretando, para aquela sublinguagem em particular, as restrições aplicáveis na formulação da frase são inaceitáveis. Nas sentenças da linguagem bioquímica abaixo:

(9) *Polypeptides were washed in HCL*

(10) *HCL was washed in polypeptides*

---

<sup>108</sup> The fact that the metalanguage of a natural language constitutes a sublinguagem does not conflict with the fact that its sentences are at the same time sentences of a natural language, satisfying its grammar. The metalanguage succeeds in describing the operator-argument structure that creates sentences only stating it in sentences that themselves have that structure. [...] In turn, the metalanguage then suffices to characterize (or generate, if we will) the sentences of the language”.

A sentença (10) não é uma sentença aceitável na linguagem bioquímica. É necessário que se aplique uma propriedade seletiva dentro de um grupo fechado de componentes lexicais – aliás, esta é uma das mais salientes características das sublinguagens: particularidades de ordem lexical, bem como sintática - que irão satisfazer as restrições específicas de formação sentencial. No caso da bioquímica, há subclasses definidas de argumentos, que nomeiam as substâncias, soluções, reagentes e outros componentes relativos à área; também operadores que se adequam à descrição dos fenômenos, reações, experiências. É certo que esse tipo de sentença pode ser encontrado na linguagem fora do domínio da bioquímica, entretanto, as outras sentenças da linguagem natural não exigem necessariamente as mesmas seqüências de palavras particulares.

Da mesma forma, há algumas regras gramaticais da linguagem que não são satisfeitas, ou não se aplicam em alguns casos na sublinguagem (cf. HARRIS, 1991, p. 272). Mas, uma vez que a sublinguagem é um domínio dentro da linguagem, como é possível que haja desencontros nos sistemas gramaticais? Em verdade, não são desencontros, mas particularidades. Harris diz que “uma vez que a sublinguagem precisa satisfazer as regras da língua, esse apagamento é possível somente se as regras forem satisfeitas [ao menos] vagamente na sublinguagem”(HARRIS, 1968, p. 154).

Uma sublinguagem pode acontecer não para conter as condições para uma dada restrição. O subgrupo de sentenças que forma a sublinguagem precisa satisfazer a gramática da linguagem natural, mas ela pode fazê-lo vagamente no caso de algumas das restrições, que são em consequência disso omitidas na gramática da sublinguagem (HARRIS, 1991, p. 273).<sup>109</sup>

Por exemplo, algumas sentenças bem formadas na linguagem não aparecem na sublinguagem. Na verdade, essas regras, nesse ambiente específico, encontram-se subentendidas, e por conta do contexto, não se fazem vagas como seria se as construções estivessem sendo usadas na linguagem “e quando se coloca em prática usos de tipos distintos de tipos de situações sociais ou ocupacionais [...] eles [os usos] produzem sublinguagens com classes de palavras tipos de sentenças específicos” (HARRIS, 1991, p. 272)<sup>110</sup>. Alguns tipos de redução, por exemplo, são verdadeiramente vagos, quando apresentados em um contexto de *linguagem geral*. Entretanto, nas restrições das

<sup>109</sup> A sublinguagem may happen not to contain the conditions for a given constraint. The subset of sentences which forms a sublinguagem must satisfy the whole-language grammar, but it can do so vacuously in the case of some of the constraints, which are thereupon omitted in the sublinguagem grammar”.

<sup>110</sup> And when we carried out on the utterances of distinct social or occupational types of situation[...] they yield sublinguagements with specific word classes and sentence type.

sublinguagens, essas reduções expressam clareza, da mesma forma que expressariam estando as sentenças construídas em sua forma completa.

Uma sublinguagem pode diferir da linguagem por omitir algumas propriedades gramaticais da língua ou por permitir reduções diferentes. As restrições sintáticas e reduções do domínio específico não são necessariamente as mesmas daquelas para o Inglês em geral<sup>111</sup> (FRIEDMAN et. al., 2002, p. 224).

Em domínios como o da clínica, por exemplo, em seções que tratam de queixas dos pacientes (relatórios), foi observado que as sentenças podem consistir de frases nominais apenas. Assim,

(11) *Dor*

(12) *Febre*

(13) *Indisposição*

são frases reduzidas e têm um operador (que se infere sendo, por exemplo *apresentou-se no paciente*) já esperado nesse contexto, e, dessa forma, pode ser suprimido. Na verdade, em domínios específicos, em que o contexto é conhecido, a informação contextual é sempre conhecida. De acordo com Harris, essa informação que é recorrente no contexto, é dispensável e pode ser omitida.

Quando se investiga o material da sublinguagem sob os mesmos critérios de investigação da linguagem, tais como ordenação de partes, encontra-se realmente uma gramática diferenciada em alguns aspectos, com restrições sentenciais diferentes. Dessa forma, um mesmo método aplicado a dois tipos de dados diferentes, (e.g. duas sublinguagens) produzem gramáticas com diferenças entre si. Dessa maneira, a estrutura gramatical da sublinguagem, além de satisfazer suas regras próprias, satisfaz também a gramática da língua como um todo, dentro de seu escopo lingüístico. “Será visto que as estruturas daquelas sublinguagens têm importantes novas propriedades delas mesmas, e também lançam luz aprofundada na estrutura da linguagem como um todo” (HARRIS, 1991, p. 272).<sup>112</sup>

---

<sup>111</sup> A sublinguagem may differ from the language geral by omitting some grammatical properties of the language or by allowing different reductions. The domain-specific syntactic constraints and reductions are not necessarily the same as those of general English.

<sup>112</sup> It will be seen that the structures of these sublinguagens have important new properties of their own, and also throw further light on the structure of language as a whole.

Harris (1968, p. 155; 1988, p. 33-56) aponta para as gramáticas das sublinguagens, um status diferenciado daquele de estarem simplesmente inclusas na da língua. Suas gramáticas, para além de se inscreverem, se interligam à da língua, uma vez que há uma relação de complementaridade.

Embora as sentenças da sublinguagem sejam um subgrupo de sentenças dela, digo, do Inglês, a gramática da sublinguagem não é uma subgramática do Inglês. A sublinguagem tem restrições importantes que não estão na língua: as subclasses particulares de palavras, e os tipos particulares de sentenças feitas por elas. E a língua tem restrições importantes que não são seguidas na sublinguagem. Naturalmente, desde que sentenças da sublinguagem são também sentenças da língua, elas não podem violar as restrições da língua, mas elas podem evitar as condições que requerem aquelas restrições<sup>113</sup> (HARRIS, 1988, p. 39).

Um exemplo interessante para se perceber essa diferença gramatical da sublinguagem em relação à linguagem toda, é a relação entre forma ativa e passiva das sentenças. É possível encontrar na análise feita pelo grupo de pesquisadores do corpus investigado - o da imunologia – que essa perspectivização confere um status diferenciado ao status que confere a um texto da linguagem comum. Em ocorrências como:

(14) Léa apareceu lá em casa ontem

(15) Léa foi encontrada no bar

Os operadores *apareceu* e *foi encontrada* representam uma mudança de perspectiva (CAMACHO, 1999), suas nuances semânticas são diferenciadas. A nuance semântica em (14) é conferida por um predicador de movimento. Em (15), o predicador é estativo. Diversamente, em artigos de imunologia, em que há ocorrências do tipo

(16) Anticorpos *apareceram* em células plasmáticas

(17) Anticorpos *foram encontrados* em células plasmáticas

---

<sup>113</sup> Though the sentences of the sublinguagem are a subset of the sentences of, say, English, the grammar of the sublinguagem is not a subgrammar of English. The sublinguagem has important constraints which are not in the language: the particular word subclasses, and the particular sentence types made by these. And the language has important constraints which are not in the language: the particular word subclasses, and the particular sentence types made by these. And the language has important constraints which are not followed in the sublinguagem. Of course, since the sentences of the suglanguage are also sentences of the language, they cannot violate the constraints of the language, but they can avoid the conditions that require those constraints.

Os operadores *apareceram* e *foram encontrados* apresentam uma só perspectiva, a de os anticorpos terem se desenvolvido nas células após determinado período que envolvia a pesquisa. Assim, pode-se dizer que desempenham a função de sinônimos puros, o que dificilmente se encontraria na linguagem (cf. HARRIS, 1988, p. 39; 1991, p. 282 nota 4). Essa maior possibilidade de ocorrência de sinônimos perfeitos ocorre porque as sublinguagens contam com um número mais restrito de elementos para a expressão de suas relações do que a linguagem. Diferenças desse tipo conferem à sublinguagem particularidades relevantes em traços gramaticais quando comparadas à linguagem.

Para essa análise dos componentes lingüísticos das sublinguagens em relação à *linguagem geral*, Harris considera que é essencial a abordagem não em termos de classes fixas de palavras, como substantivos, adjetivos ou verbos. O que se faz relevante para produzir resultados é a abordagem sob o prisma da gramática de operadores, mesmo porque alguns elementos de áreas específicas ou classes deles fogem da classificação algo ingênua de palavras.

Ao se analisar um dado assunto-objeto, encontra-se bem mais que uma relação geral de dependência, consegue-se notar grupos específicos de argumentos que ocorrem somente em ambientes particulares de operadores.

Ou seja, obtemos vários tipos de sentenças, cada qual uma combinação particular de subclasses de palavras. Esta é a propriedade característica das sublinguagens de um assunto objeto; isto se sustenta seja as sentenças estando em discursos ou ocorrendo separadamente. É mais nítido na ciência, especialmente nas ciências mais ‘sólidas’, quando palavras e sentenças tiradas de fora da ciência são irrelevantes e inexistentes<sup>114</sup> (HARRIS, 1991, p. 278).

E isso certamente possibilita postular uma gramática diferenciada para esse assunto-objeto. Nos exemplos anteriores (9) e (10) são sentenças gramaticais para a linguagem geral, mas, na linguagem laboratorial o operador *is washed* exige como argumentos um grupo específico de palavras diretamente relacionadas ao campo; nesse caso, algo dentro do grupo de ácidos, soluções ou similares.

Relatórios e discussões em um aspecto bem estruturado do mundo mostram limitações em uso de palavras que são suficientes para constituir restrições em

---

<sup>114</sup> That is, we obtain various sentence types, each a particular combination of word subclasses. This is the characteristic property of subject matter sublinguagens; it holds whether the sentences are in discourses or occur singly. It is sharpest in science, especially in the ‘hard’ sciences, where words and sentences drawn from outside the science are both irrelevant and absent.

ocorrência de palavras; isto é especialmente evidente quando o que dito é limitado ao que parece relevante para aquele campo<sup>115</sup> (HARRIS, 1988, p. 37).

Essas exigências ocorrem porque trata-se de um uso limitado e de ocorrência restrita de palavras, e também ao fato da sublinguagem ser expressa em uma metalinguagem direcionada somente a ela e seus conceitos; isso traz alguns problemas de ordem de compreensão, de forma que somos levados a entender as sublinguagens, e não apenas as das ciências, mas principalmente elas, como um ambiente lingüístico hermético, e, nesse sentido, é o próprio Harris (cf. 1988, p 34; 1991 pp. 272, 274, 281; 1968, p. 152) que se refere às operações originadoras de uma sublinguagem específica como ‘fechadas’.

Ao passo que o fechamento das sublinguagens gramaticais [...] emergiram de restrições gramaticais específicas, o fechamento aqui emerge de que limitações do uso de palavra se devem a limitações dos assuntos objeto, as quais são fortes o suficiente para constituir restrições em seleção de palavras quando se está falando dentro de um assunto objeto<sup>116</sup> (HARRIS, 1991, p. 278).

A partir dessas considerações temos que a sublinguagem tem duas características elementares que a diferem da linguagem em geral: a metalinguagem específica e sua estrutura gramatical.

Como foi explicitado no início do texto, Harris escolheu se debruçar preferencialmente sobre a linguagem da ciência para tratar do objeto sublinguagem inscrito nela. Nessa sublinguagem da ciência estão inclusos alguns componentes sobre os quais se faz indispensável anotar, por serem parte mesmo do componente sublingüístico. Dessa forma, são abordadas a metalinguagem da linguagem, também a linguagem científica, que consiste de fórmulas apresentadas através de símbolos. Harris faz uma distinção entre a linguagem científica e sublinguagem da ciência, mas todos esses componentes, inclusive as fórmulas, igualmente se relacionam à estrutura da língua e também ao conteúdo informacional das diferentes linguagens científicas existentes.

À parte de puramente características estruturais dessas sublinguagens, há uma propriedade importante em relação à informação; [...] quando combinações de

<sup>115</sup>“ Reports and discussions in well structures aspects of the world show limitations on word use that are sharp enough to constitute constraints on word occurrence; this is especially evident when what is said is limited to what seems relevant to the field.”

<sup>116</sup> Whereas the closure of the grammatical sublinguagens above arose from specific grammatical constraints, the closure here arises from the limitations of word use due to the limitations of the subject matter, which are sharp enough to constitute constraints on word selection when one is speaking within the subject matter.

palavras de uma língua são escritas mais eficientemente, obtemos uma forte correlação entre diferenças em estrutura e diferenças em informação. Esta correlação é mais forte ainda nas sublinguagens<sup>117</sup> (HARRIS, 1988, p. 39-40).

Além disso, algo do que Harris chama de “incremento” na estrutura, possibilita um incremento também na capacidade informacional:

...podemos questionar como palavras carregam significado e como sentenças carregam informação. Consideraremos então exemplos de como um incremento na estrutura produz um incremento na capacidade informacional, i.e., em que tipo de informação a estrutura pode carregar<sup>118</sup> (HARRIS, 1988, p. 58).

Dessa forma, as diferenças nas restrições das sublinguagens refletem uma informação que se desenvolve a partir dessas restrições específicas. O lingüista apresenta então maior interesse em analisar a linguagem científica, sendo a mais saliente propriedade dessa análise a de transmitir a informação numa dada ciência, em detrimento do estudo voltado apenas para demonstrar sua sistematicidade formal.

Quando o assunto objeto em questão é um subcampo da ciência, nós obtemos um sistema lingüístico complexo que pode apresentar precisamente a informação da[quela]ciência,[...]na sublinguagem da ciência nós encontramos maneiras distintas de indicar a estrutura da informação, e os desarranjos na informação, e mudanças nela<sup>119</sup> (HARRIS, 1988, p. 40-41).

Na verdade, ao se fazer uma leitura de Harris, nota-se um imbricamento que envolve a sintaxe de operadores, informação, metalinguagem e sublinguagem. Na abordagem de sublinguagem da ciência não é diferente, de forma que a informação também é explicitada através das operações de restrição e seleção que envolvem os operadores e argumentos da construção sentencial, obedecendo a determinado sistema de regras próprios da sublinguagem.

---

<sup>117</sup> Aside from the purely structural characteristics of these sublinguages, there is an important property in respect to information;[...] when the word combinations of a language are described most efficiently, we obtain a strong correlation between differences in structure and differences in information.

<sup>118</sup> ...we can ask how words carry meaning and how sentences carry information. We will then consider examples of how an increment in structure yields an increment in informational capacity, i.e., in what information the structure can carry.

<sup>119</sup> When the subject matter in question is a subfield of science, we obtain a complex linguistic system that can present precisely the information of the science [...] In the science sublinguage we find distinguished ways of indicating the structure of the information, and the disagreements in information, and changes in it.

### 4.3. Sublinguagem de uma ciência em foco

O estudo das sublinguagens especializadas - em procedimento diverso do que faz a sociolinguística e a Linguística funcional, ao estudarem as mudanças nas sublinguagens no que tange aos usos de falares dos diversos grupos pertencentes a um sistema linguístico - investiga como grupos fixos de formas de discurso e escrita se caracterizam no desenvolvimento de uma disciplina em uma ciência particular (HIZ, 1982, p. 206). Nessa concepção, “o que é especial para as ciências é que certas subclasses de palavras (e frases) co-ocorrem de forma regular para produzir certos tipos especificáveis de sentenças (como combinações sentenciais de classes de palavras)”<sup>120</sup> (HARRIS, 2002b, p. 10).

A linguagem geral tem as restrições para a ordenação parcial, que rege a relação operador-argumento e que cria e transmite a informação da língua. Da mesma forma, uma sublinguagem da ciência traz consigo restrições que lhe são próprias, pois é um subambiente dentro do universo linguístico. Então, “o grupo fechado de classes de palavras em uma ciência, e o grupo fechado de tipos sentenciais e fórmulas individuais extraídas delas, carregam tipos únicos de informação, a informação da ciência dada.<sup>121</sup>” (HARRIS et al, 1989, p. 80)

A sublinguagem da ciência é efetivamente, para Harris, um domínio da língua que se faz realmente interessante como objeto de investigação, dadas as particularidades nele apresentadas.

Entre possíveis sublinguagens objeto algumas se colocam como, de longe, muito mais importantes. São as sublinguagens da ciência, cada uma consistindo das sentenças e discursos de praticantes de uma sub-ciência, resultado do curso de uma pesquisa e discussão. Elas são importantes porque seu objeto é importante, e porque elas lançam luz nas estruturas de sua informação<sup>122</sup> (HARRIS, 1991, p. 283).

<sup>120</sup> “What is special to sciences is that certain subclasses of words (and phrases) co occur in a regular way to make certain specifiable sentence-types (as sentential combinations of word subclasses)”.

<sup>121</sup> ...the closed set of word-classes in a science, and the closed set of sentence-types and individual formulas formed out of these, carry unique kinds of information, the information of the given science.

<sup>122</sup> Among possible subject matter sublinguagens certain ones stand out as by far the most important. These are the science sublinguagens, each consisting of the sentences and discourses of the practitioners of a sub-science, issued in the course of research and discussion. They are important because their subject is important, and because they throw light on the structure of its information .

Esses traços próprios, bem como as semelhanças/diferenças gramaticais entre os diversos campos afins (e.g. imunologia, farmacologia) são de interesse para se obter uma estrutura gramatical da(s) ciência(s); a diferenciação entre as estruturas dessas sublinguagens em relação à linguagem, para o lingüista, é um ponto saliente e importante de se observar.

Pode mesmo ser que esses escritos científicos provem ter estruturas mais complicadas, como conjunções ternárias, i.e., pares de conjunções que requerem tipos particulares de similaridade e diferenças entre as três sentenças envolvidas. Em todo caso, ela[a sublinguagem] tem estruturas de discurso diferentes dos textos coloquiais ou literários. Tais regras seriam estabelecidas se mostrarmos que apenas certas seqüências sentenciais são encontradas nesses escritos, mais restritos que na linguagem como um todo<sup>123</sup> (HARRIS, 1968, p. 154).

Para apresentar as considerações e as conclusões em relação a essa teoria, Harris optou, para explicitar o estatuto da sublinguagem, por trabalhar com a sublinguagem da Imunologia. Esse estudo envolveu o tratamento de textos de artigos sobre a área da ciência imunológica, com o propósito de perceber se seria possível representar através de procedimentos formais a informação dos textos de maneira prática e ordenada.

O período coberto foi de 1935-1966, quando o campo era muito menor e mais incipiente que é agora, e quando havia um problema central de determinar qual célula era produtora de anticorpos. Havia também uma controvérsia sobre se era o linfócito ou a célula plasmática, ambos do sistema linfático. Depois foi mostrado, pelo microscópio eletrônico e outros métodos, que ambos os tipos de célula produziam anticorpos, a controvérsia foi resolvida pelo entendimento de que os dois nomes de células indicavam diferentes estágios de desenvolvimento da mesma linha celular<sup>124</sup> (HARRIS, 1988, p. 41).

Através desse trabalho também foi observado se seria possível localizar na estrutura sentencial e se caracterizar estruturalmente a mudança na informação através dos anos pelos mesmos procedimentos formais de análise. Harris (1968, p. 17; 1988 p. 9, 62, 92-

---

<sup>123</sup> It may even be that scientific writing will prove to have more complicated structures, such as ternary conjunctions, i.e., conjunction pairs requiring particular kinds of similarity and different discourse structures than colloquial or literary texts. Such rules would be established if we show that only certain sentence sequences are found in this writing, more restricted than in the language as a whole.

<sup>124</sup> The period covered was c. 1935-1966, when the field was far smaller and more inspectable than it is now, and when it had a central research problem of determining which cell was the producer of antibodies. There was also a controversy about whether it was the lymphocyte or the plasma cell, both of the lymphatic system. After it was shown, by electron microscopy and other methods, that both cell types produced antibodies, the controversy was resolved by the understanding that the two cell names indicated different stages of development of the same cell line.

113; 1991, p. 297) reconhece as mudanças que envolvem o fenômeno lingüístico através do tempo como sendo importantes para mudanças que se refletem também em suas estruturas. As mudanças relacionadas ao signo lingüístico (cf. SAUSSURE, 1994[1916], p. 89-93) se fez um elemento importante na colocação e resolução dessa problemática, pois o deslocamento de significado dos nomes celulares ocupou posição central na aludida controvérsia. Procurava-se nela entrever se as transformações na língua afetavam também a estrutura representativa da sublinguagem.

Foi analisado como termos específicos da área de pesquisa combinavam entre si, como é que se dava nessa sublinguagem o movimento das restrições originadoras de frases bem formadas na língua.

O objetivo do método é estabelecer classes de objeto relevantes no domínio, e classes de relações em que os objetos participam. A técnica agrupa diferentes argumentos de sentenças (sujeitos ou objetos gramaticais) em uma classe de acordo com sua ocorrência nos textos com o mesmo operador (verbo principal, adjetivo ou preposição). Os operadores são agrupados em classes de acordo com sua ocorrência com a mesma classe de argumentos. Quando a análise é feita em uma amostra de suficiente extensão, as classes de argumentos correspondem aos objetos do domínio e as classes de operadores, às relações do domínio <sup>125</sup> (JOHNSON, 1989, p. 191).

Assim, as palavras foram divididas em classes, para registro de ocorrência. Classes como de nomes, verbos e conjunções, foram estabelecidas, com os respectivos tipos variantes<sup>126</sup>, para que se chegasse, através da mesma análise estrutural de classes propostas por Harris (1970[1946]) para os componentes que produzem eventos na linguagem comum, ao(s) modelo(s) de sentenças possíveis dentro da referida sublinguagem.

As principais classes encontradas dentro da sublinguagem da imunologia, com suas respectivas combinações/probabilidade de ocorrência conforme as restrições impostas pelo ‘subsistema’, foram:

---

<sup>125</sup> The purpose of the method is to establish classes of objects relevant in the domain, and classes of relations in which the objects participate. The technique groups different arguments of sentences (grammatical subjects or objects) into a class according to their occurrence in the texts with the same operator (main verb, adjective, or preposition). Operators are grouped into classes according to their occurring with the same classes of arguments. When the analysis is carried out on a sample of sufficient size, argument classes are found to correspond to domain objects, and operator classes to domain relations.

<sup>126</sup> As formas variantes de determinada classe são marcadas, a saber, as palavras com modificadores trazem-nos na forma de sobrescritos, como na fórmula  $GU^{ft}$  (com *ft* representando *from...to*), e as que representam membros de diferentes subclasses, com subscritos. Por exemplo, tem-se *V* para verbos, *V<sub>i</sub>* para *appears in*, *V<sub>p</sub>* para *produced by*.

### 4.3.1 Argumentos

**G** para *antigen*, que poderiam também representar sinônimos ou outros nomes de antígenos.

**A** para *antibody*, podendo o símbolo também representar palavras ou lexias sinônimas para os anticorpos.

**B** representando corpo, partes de corpo (*stomach, footpad*) ou mesmo nomes de animais, como *rabbit*, e outros.

**T** para *tissue*. O símbolo representa uma classe ampla de palavras para se referir a diferentes tecidos e órgãos. As diferenciações são representadas por subscritos, como em **T<sub>1</sub>** para *lymph*, **T<sub>s</sub>** para *spleen*, etc. Foi observado que tanto o símbolo padrão, como o que é acompanhado de subscritos, ocorrem em ambientes similares, exceto sob a ocorrência de particularidades, como os modificadores.

**C** *cell*, para representar nomes de células, também com as respectivas formas variantes: **C<sub>y</sub>** *lymphocyte*, **C<sub>b</sub>** *blast cells*, **C<sub>z</sub>** *plasma cells*.

### 4.3.2. Operadores

De acordo como ocorriam ligando tais e quais argumentos, os operadores foram classificados em:

**J** para *inject*, que também representa verbos sinônimos, como *immunize* (para esta sublinguagem) verbo que se observou ocorrendo entre **G** e **B**:

(18) *Antigen is injected into a body part.*

**V** sendo verbos que ocorrem entre **A** e **C** ou **A** e **T**, com as diferentes classes verbais indicadas pelos subscritos, como **V<sub>i</sub>** para verbo + in (*appear in, found in*), **V<sub>p</sub>** para *produced in*, **V<sub>s</sub>** para *secreted by*:

(19) *Antibody is secreted by the cell.*

**W** representando uma classe ampla de verbos e adjetivos que ocorrem depois de **T** ou **C**, que são propriedades de tecidos ou células:

(20) *Tissue is inflamed*

(21) *Cells proliferate*

**U** para representar uma classe de verbos que ocorrem entre (**G**)antigen e (**T**)tissue ou mesmo cell:

(22) *Antigen moves to tissue*

**Y** como uma última classe de verbos, que ocorrem entre *cell* e *cell*, com variantes em  $Y_c$  (*changes into*) e  $Y_u$  (*contaminated with*):

(23) *Plasma cell is similar to lymphatic cell.*

A partir dessa classificação, foram observadas as cadeias de ocorrência que tinham lugar no corpus da sublinguagem. As fórmulas prototípicas encontradas foram:

**GJB** → *antigen is injected into a body part ou an animal*

**GU<sup>ft</sup>TT** → *antigen moves from some tissue to some tissue*

**TW** e **CW** → *A tissue (cell) has some property or undergoes a change*

**AVC** → *antibody appears in a cell ou is produced by, is secreted from*

**CYC** → *some cell is similar to another cell*

**CY<sub>c</sub>C** → *A cell develops into another cell*

**GJB:<sup>t</sup>AVC** → *Antigen is injected into a body part; three days later antibody appears in cells.*

**AVT:** → *Antibody is found in lymphocytes*

Dentro desta penúltima fórmula, ocorre sempre a conjunção *thereafter* - e sinônimas a ela - ligando as duas sentenças (*three days later, five hours*); isso leva a crer que dentro das características gramaticais que revestem as sublinguagens, há também em

relação às conjunções que ligam as sentenças uma determinação de grau de ocorrência mais alto ou mais baixo, que é interna ao domínio (cf. HARRIS, 1968, p. 155).

### 4.3.3. Resultados da Pesquisa em Imunologia

Estabelecer os tipos de estrutura predominante na sublinguagem oportunizou que se pudesse explicitar diversas particularidades inerentes a elas. Foi visto que as fórmulas a que se chegou são importantes para se obter informação tanto generalizada quanto específica em relação ao domínio investigado. As seqüências lineares dos símbolos apresentam as classes e suas disposições na cadeia sentencial, bem como os sobrescritos e subscritos mostram as informações específicas, que indicam os modificadores nas sentenças. Dessa forma, de posse dos símbolos correspondentes às classes e à ocorrência dessas classes em cadeia, é possível, pelas fórmulas, depurar como estão organizados os textos científicos, os tipos de estrutura de suas sentenças, os ambientes em que determinados operadores e argumentos ocorrem, e, por conseguinte, a estruturação do texto (HARRIS, 1970[1952]).

Um dos problemas propostos quando do início da investigação no corpus da Imunologia foi o de se buscar as mudanças na informação textual ocorridas em um intervalo de tempo. No caso da pesquisa em questão, foram utilizados artigos de 1935 a 1966 (HARRIS et al, 1989), portanto, com uma interposição de um período de 31 anos entre os primeiros artigos e os últimos. As fórmulas sentenciais demonstraram que houve diferenciação de tipos de componentes que entraram nas sentenças com o passar do tempo. Devido ao desenvolvimento da ciência em questão, as mudanças lexicais se fizeram perceber através das fórmulas.

Podemos localizar a mudança através do tempo: Primeiro [a seqüência] AVT é trocada por AVC. Nos primeiros artigos (1935) encontramos T (e apenas no final, brevemente, C), porque os tipos celulares não eram então prontamente distinguidos nos tecidos. Mais tarde, um novo tipo de sentença, CYC aparece, quando mais tipos celulares são distinguidos e suas similaridades notadas, e quando a proliferação de nomes de células é controlada por dizer que nelas alguns nomes identificam a mesma célula. Pudemos localizar ambigüidade, como quando a proliferação de nomes das células não é embasada por diferentes propriedades (na classe W) relatado diferentemente pelos nomes celulares, com

a ambigüidade sendo finalmente reconhecida por sentenças CYC afirmando que esses eram nomes para a mesma célula<sup>127</sup> (HARRIS, 1988, p. 46-47).

Esse surgimento de uma nova classe para denominar células, mostra, também na sublinguagem – e com consequências importantes – a evolução que ocorre a todo tempo no sistema linguístico. Uma vez que o léxico da língua passa por mudanças, esse movimento se reflete nos diferentes subsistemas, novas subclasses e tipos de sentenças podem entrar na sublinguagem a qualquer tempo. “A relação da estrutura dessa sublinguagem com a informação da ciência pode ser vista de várias formas. Primeiro pode-se ver como a conhecida mudança da informação através do tempo é vista na mudança de subclasses de palavras e de tipos de sentenças em sucessivos artigos desse período”<sup>128</sup> (HARRIS, 1991, p. 286).

O sistema das fórmulas se mostrou eficiente para localizar possíveis contradições no discurso, pela diferenciação dos símbolos. Em uma série de artigos ocorria a fórmula  $AV_pC_y$  (*Antibody is produced by lymphocytes* ou *lymphocytes produce antibody*); havia a ocorrência de  $AV_pC_z$  em outra série (*Antibody is produced by plasma cells*) e ainda em outra,  $AV_{\sim p}C_y$  (*Lymphocytes do not produce antibody*). Ora, se um texto diz que os anticorpos são produzidos por linfócitos e outro diz que os linfócitos não os produzem, isso é, dentro da sublinguagem uma contradição comprometedora da credibilidade da pesquisa científica.

Através de uma fórmula sentencial também foi possível localizar a explicação encontrada nos últimos artigos para a contradição presente nos textos do corpus.

O artigo 3 argumentava que as células plasmáticas do sistema linfático, e não os linfócitos, eram os produtores de anticorpos. O artigo 4 apresentava resultados na mesma direção, junto com a evidência de que as células plasmáticas são o estágio final de um desenvolvimento de células reticulares que vão completar o estágio dos linfoblastos, mas, foi afirmado, não completam o estágio de linfócitos.[...] O artigo 10, no curso de um problema de pesquisa diferente, mostrou que pequenos linfócitos não são uma célula final mas um estágio avançado de desenvolvimento, e que elas são essenciais para a produção inicial

<sup>127</sup> We can locate change over time: First AVT is replaced by AVC. In the earliest paper (1935) we find T (and only at the end, briefly, C), because cell types were not then readily distinguished in the tissues. Later, a new sentence type, CYC, enters, when more cell types are distinguished and their similarities noted, and when the proliferation of cell names is controlled by say in that some different names identify the same cell. We can locate unclarity, as when the proliferation of cell names is not supported by different properties (In the W class) reported for the differently named cells, with the unclarity being finally recognized by CYC sentences stating that these are names for the same cell.

<sup>128</sup> The relation of this sublinguagem structure to the information of the science can be seen in various ways. We first how the known change of information through time is seen in the change of word subclasses and sentence types in the successive articles of this period.

de anticorpos. Os artigos 11 e 13 mostram pequenos linfócitos individuais, bem como células plasmáticas, produzindo anticorpos, e argumenta que células plasmáticas são provenientes de linfócitos. [...] As mudanças nos tipos sentenciais refletiram mesmo mais de perto as mudanças na informação<sup>129</sup> (HARRIS et al, 1989, p. 64-65).

Nos artigos finais do período em que foi feita a pesquisa, ocorreu uma estrutura nova,  $C_y Y_c C_z$ , (*Lymphocytes develop into plasma cells*), em que os argumentos  $C_z$  e  $C_y$  não haviam sido até então combinados com o operador  $Y_c$ .  $C_y$  (linfócitos) e  $C_z$  (células plasmáticas), sendo os dois argumentos que estavam no centro da investigação, foram apresentadas agora ambas como produtoras de anticorpos. Foi visto então que as diferenças em classes de palavras e em fórmulas de sentenças aparecem onde há diferenças na informação.

O deslocamento gradual na observação experimental foi logo visível em detalhes das fórmulas. De alguma forma transformações gramaticais podem em alguns casos mostrar a direção de uma mudança no comportamento gramatical, então as mudanças nos ambientes de operador-argumento de uma dada palavra pode em alguns casos mostrar a direção de uma mudança na compreensão<sup>130</sup> (HARRIS, 1991, p. 296).

Estava terminada a controvérsia. Em verdade, os linfócitos e as células plasmáticas indicavam apenas diferentes estágios de uma mesma linha de células, sendo, portanto, as duas, produtoras de anticorpos.

Foi possível notar que as fórmulas sentenciais deixam entrever uma evolução no discurso, pois como já salientava Saussure (1994[1916], p. 86), “um dado estado de língua é sempre produto de fatores históricos” e, aqui o status da sublinguagem também reflete fatores históricos, como evolução científica, e em decorrência dela, mudança lexical. As fórmulas se mostram eficazes tanto para resumir a informação quanto para criticá-la, pois

---

<sup>129</sup> Paper 3 argued that the plasma cells of the lymphatic system, and not the lymphocytes, were the antibody producers. Paper 4 presented results in the same direction, together with evidence that the plasma cells are the end stage of a development of reticulum cells which go through lymphoblast stages, but, it was claimed, not through lymphocyte stages. [...] Paper 10, in the course of a different research problem, showed that small lymphocytes are not an end cell but develop further, and that they are essential to initial antibody production. Papers 11 and 13 show individual small lymphocytes, as well as plasma cells, producing antibody, and argue that plasma cells are descended from lymphocytes. [...] The changes in sentence-types reflected even more closely the changes in information.

<sup>130</sup> The gradual shift in experimental observation was visible early in details of the formulas. Somewhat grammatical transformations can in some cases show the direction of an ongoing grammatical change, so the changes in operator-argument environments of a given word can in some cases show the direction of an ongoing change in understanding.

por conta das mudanças na representação formulaica, notam-se os acréscimos, percebe-se um refinamento das construções, mostrando a opacidade e também as contradições nos textos.

#### 4.4. Operadores Meta Científicos

Outro tipo de particularidade é a distinção entre operadores metacientíficos e operadores próprios da sublinguagem, ambos recorrentes nela (cf. HARRIS, 1988, p. 45; 1991, p. 289; GOTTFRIED, 2002, p. 97). A sublinguagem da ciência é construída com base em uma metalinguagem voltada para, além dessa ciência em particular, também para a linguagem científica como um todo. Então, o que ocorre é que os trabalhos científicos têm, em geral, um formato peculiar, com dizeres igualmente peculiares. As diferentes seções são explicitadas com dizeres, operadores, adequados a elas. Disso resulta que, dentro da sublinguagem, há uma subdivisão de tipos de sentenças nas diferentes seções dos textos.

Nesses casos, cada classe de segmentos tem seu próprio vocabulário interrelacionado, comum a suas ocorrências em todas as ciências e não específicos ao material da imunologia; sua relação com o material científico específico nas sentenças é o mesmo que é em outras ciências. Pode-se portanto dizer que cada sublinguagem da ciência, de certa forma e em diferentes maneiras, baseia-se nesses sistemas sublinguísticos gerais em construir suas sentenças<sup>131</sup> (HARRIS, 1991, p. 290).

Então, operadores como “era esperado que” ou “tem mostrado que” são tidos como material meta científico, que permeia a sublinguagem e dos quais as sentenças dessa sublinguagem dependem para representar a informação do texto de forma precisa. Hiz (1982) acentua que, efetivamente, esses operadores se encontram em uso metalinguístico para falar sobre a linguagem, e não estão em uso próprio da sublinguagem. (cf. HIZ, 1982, p. 209). Dessa forma, as construções, *antibody is found in lymphocytes* e *it was found that antibody is in lymphocytes*, seriam representadas, a primeira, através de fórmula simples da

---

<sup>131</sup>In these cases, each class of segments has its own interrelated vocabulary, common to its occurrence in all sciences and not specific to the immunology material; its relation to the specific science material in the sentences is much the same as it is on other sciences, one could therefore say that each science sublinguagem, in somewhat different ways, draws upon these more general sublinguagem systems in constructing its sentences.

sublinguagem (AVC); e a segunda, por uma formulação que destacaria a sentença de seu operador meta científico ( $O_{meta}AVC$ ).

Gottfried (2002) enfatiza que as sessões de textos que contêm os operadores metacientíficos, ficariam de fora, em princípio, dos tipos sentenciais específicos da sublinguagem. Essas seções caracterizam uma relação epistêmica do investigador para com o resultado, que é descrito pelas sentenças da linguagem científica (cf. HARRIS, 2002a p. 218); os operadores meta científicos dessas sentenças apenas ligam sentenças da linguagem de ciência dentro do texto, não estabelecendo necessariamente nenhuma relação sentencial entre os membros de classe de palavras. Na frase

(24) *The authors have demonstrated that [antibody is in lymphnodes]*

a sentença *The authors have demonstrated* se faz uma sentença meta científica que se refere a *antibody is in lymphnodes*.

No exemplo, a sentença que se encontra dentro dos colchetes, nessa sublinguagem pode ocorrer sozinha; já a sentença que se posiciona fora dos colchetes não poderia aparecer sozinha, configurando como pertencente à sublinguagem. Mas apesar das sentenças da linguagem da ciência serem independentes na sublinguagem, elas têm dependências informacionais em relação aos operadores meta.

Em relação a esses operadores metacientíficos, no estudo de Harris, foi verificado que os tipos sentenciais diferiam nas sessões dos textos específicos de imunologia, de maneira que na maioria dos artigos foram encontradas diferenças em tipos de sentenças entre as sessões dos Procedimentos, Resultados, e Discussões. Essas nuances demonstravam distinções gramaticais, possíveis de se obter através das fórmulas:

Se olharmos para a divisão típica de artigos experimentais em seções de Procedimentos, Resultados e Discussões, encontraremos que eles não são apenas um assunto de tradição, mas são gramaticalmente distinguíveis. Na seção de Procedimentos as fórmulas da imunologia identificam antígenos particulares, o número e condições de injeção, a localização precisa da injeção, qual animal, e assim por diante. A seção de Resultados tem tipos de sentenças diferentes, por exemplo, o GJB:AVC para *antigen is injected,thereafter antibody appears in cell*. Mas há certas palavras que são compartilhadas pelas duas seções, e o que é mais importante que o mero compartilhar de palavras é o fato de que as sentenças de Resultado são vistas como sendo parcialmente dependentes das sentenças de Procedimento. Certas sentenças de resultado não puderam aparecer se as sentenças de procedimento fossem diferentes. Isto é realmente surpreendente, mas o ponto é que isso pode ser mostrado através de

fórmula estrutural, de uma maneira precisa<sup>132</sup>(HARRIS, 1988, p. 53). grifos do autor.

Verificou-se que as sessões de Discussão geralmente não apresentavam novos tipos de fórmulas sentenciais. Elas se traduziam nas fórmulas de Resultados, geralmente modificadas por classificadores, com seqüências e conjunções particulares. Dessa forma, as estruturas de construção demonstraram a sistematização do texto, sua natureza conclusiva e/ou assertiva sobre um objeto pesquisado. Harris aplicou os resultados de fórmulas sentenciais à possibilidade de analisar a sintaxe de conclusões de trabalhos da pesquisa atual, para analisar como sentenças de resultados se fazem pertinentes se associadas às de Conclusão nos relatórios científicos.

#### 4.5. Linguagem científica

Na discussão da noção de sublinguagem são associadas a ela outras noções que são igualmente importantes, noções essas que se fazem relevantes para compreender a abordagem harrissiana de língua e suas manifestações. Para a investigação de uma sublinguagem importa compreender que ela, além de estar permeada pela gramática de operadores (HARRIS, 1988, p. 49; 1991, p. 291) – que, como dissemos, lhe é peculiar -, necessita de uma metalinguagem externa para se constituir e também é passível de se traduzir por um sistema de representações. Atenhamo-nos a este último.

Os ditos sistemas de representação foram, em momentos anteriores a Harris, tidos como possíveis transmissores do pensamento puro. Para Culioli (1995, p. 21-30), essa tentativa de transmissão se apresenta como um problema, porque não se tem acesso direto às representações mentais.

---

<sup>132</sup> If we look at the typical division of experimental articles into Procedures, Results, and Discussion sections, we find that these are not just a matter of tradition, but are grammatically distinguishable. In the Procedures section, the immunology formulas identify particular antigens, the number and conditions of injection, the precise location of injection, what animal, and so on. The Results section has different sentence types, for example the GJB:AVC for antigen is injected, thereafter antibody appears in cell. But there are certain words that are shared by the two sections, and what is more important than the mere sharing of words is the fact that the Results sentences are seen to be partially dependent on the Procedures sentences. Certain Results sentences could not appear if the Procedures sentences were different. This is hardly surprising, but the point is that it can be shown in the formula structure, and in a precise way.

Efetivamente, a criação da linguagem notacional está na base da lógica moderna. “O que inaugurou a lógica moderna foi a invenção de Frege da notação de quantificadores exposta no § 11 do *Begriffsschrift*; e foi sua construção resultante do que conhecemos como predicado da lógica que fez possível um tratamento satisfatório da generalidade múltipla.”<sup>133</sup> (BEANEY 1997[1879], p. 47). Frege, (1997[1879]), no seu *Begriffsschrift*, pretendeu elaborar um sistema de notação que, por ser assim, se faria mais claro que a linguagem natural, sem os problemas de ambigüidade, vagueza e mal entendidos que a permeiam. Esse sistema seria aplicado de maneira produtiva na linguagem científica, em que haveria necessidade, mais do que na linguagem comum, de precisão. Assim, segundo Frege (1997[1879], p. 49), seu *Begriffsschrift* seria um auxílio para propósitos científicos, e não merecia ser condenado por não servir para outros propósitos. Entretanto, não foi possível para Frege colocar em prática seu projeto tal qual o idealizara, porque recebeu muitas críticas relativas a sua pretensão de estabelecer uma linguagem que transmitisse o que ele chamou de “pensamento puro”, aplicável à linguagem científica.

Harris, que se declara um herdeiro da Lógica filosófica e da Matemática (cf. HARRIS, 2002b, p. 1-13; 1991, p. 284), retoma essa idéia da representação notacional, tanto na linguagem comum quanto nas sublinguagens (HARRIS, 1991, p. 303, 1970[1959], p. 253-277). Nessa empreitada de fazer possível a representação da linguagem através de símbolos, o linguista não tem a pretensão de uma pureza de representação. Na realidade, ele coloca em discussão os critérios da lógica extensional para se chegar ao significado das palavras (HARRIS, 1988, p. 64).

Além disso, dado o que conhecemos sobre o estatuto de ‘verdade’ em lógica e sobre a descrição alternativa da estrutura, uma teoria não necessita ser pensada como apresentando uma verdade final, mas apenas como organizando os resultados de certos métodos de análise, ‘verdadeiros’ tanto quanto possível<sup>134</sup> (HARRIS, 2002b, p. 9).

No que se refere às sublinguagens, Harris apresenta as linguagens científicas (notacionais), que se constroem a partir de notações especificamente elaboradas para traduzir o conteúdo formal dessas ciências. A utilização dessa linguagem científica tem,

---

<sup>133</sup> What inaugurated modern logic was Frege’s invention of quantifier notation first explained in § 11 do of the *Begriffsschrift*; and it was his resulting construction of what we know as predicate logic that made possible a satisfying treatment of multiple generality.

<sup>134</sup> Furthermore, given what we know about the status of ‘truth’ in logic and about the alternative description of structure, a theory should not be thought of as presenting the final truth, but only as organizing the results of certain methods of analysis, ‘true’ as far as it goes.

para Harris, algumas vantagens em relação ao uso da linguagem comum, como por exemplo, evitar redundâncias, e o uso desnecessário de sinônimos (HARRIS, 1991, p. 291; 1988, p. 48).

Os símbolos oportunizam discretizar os *status* de diferentes classes e subclasses de argumentos e também as relações prepositivas que ligam os operadores aos seus argumentos. Com a linguagem dos símbolos também “é possível omitir exigências gramaticais da língua, que são irrelevantes para a ciência particular...”<sup>135</sup> (HARRIS, 1988, p. 49).

Ao se tratar da linguagem simbólica, há uma distinção entre esta e a sublinguagem da ciência (cf. HARRIS, 1991, p. 290-291; 1988, p. 48-54). A sublinguagem da ciência corresponde a um domínio, e aquela corresponde a formas de representação artificialmente construídas (e.g., a classificação dos operadores e argumentos em tipos traduzidos em letras e sinais), ou seja, são sistemas de fórmulas de informação compostos por símbolos.

As linguagens da ciência são uma representação mais precisa da informação da ciência do que o são as sublinguagens. Nas sublinguagens, que embasam o vocabulário e a gramática da linguagem geral mas a usa sob restrições da ciência, a homonímia é inevitável. [...]... a sinonímia (em relação à subciência) é comum na sublinguagem: e. g., *anticorpos*, *aglutinina*, *globulina gama*, são todos membros indeferenciados de A em relação ao assunto nesses artigos. Porém, nas linguagens da ciência, apenas A é um elemento de vocabulário<sup>136</sup> (HARRIS, 1991, p. 294). Itálicos do autor.

Nessa distinção repousa o fato de que as sublinguagens, mesmo tendo um domínio comum, são expressas em códigos diferentes, em diferentes sistemas lingüísticos.

Entretanto, dentro de uma ciência, pode-se afirmar a existência de correlação estrutural e semântica entre as fórmulas que ocorrem em textos de mesmo material lingüístico, ou seja, fórmulas de língua da ciência; as análises feitas de artigos escritos em diferentes línguas têm mostrado a mesma representação formulaica; isso se deve ao fato de

---

<sup>135</sup> “...it is possible to omit grammatical requirements of the linguagem geral that are irrelevant to the particular science,...”

<sup>136</sup> The science languages are a more precise representation of science information than are the sublinguagens. In the sublinguagens, which draw upon linguagem geral vocabulary and grammar but use it under science constraints, homonymy is inescapable. [...]... synonymy (in respect to the subsience) is common in the sublinguagem: e.g. *antibody*, *agglutinin*, *gamma globulin*, are all undifferentiated members of A in respect to the issue in these articles. However, in the science language, only A itself is a vocabulary element.

que há, entre diversas línguas, uma correspondência de estrutura sintática, possibilitando o que Harris chamou de Transfer Grammar:<sup>137</sup>

também no estudo de imunologia[...], a análise de artigos escritos em Francês produziram substancialmente a mesma representação formulaica como no caso dos artigos em Inglês. Igualmente, é relevante que muitos artigos escritos em uma língua sejam lidos e discutidos por cientistas que falam outras línguas, e que a pressão similar exercida nos escritores de artigos para aderir a formulações precisas e padronizadas podem vir tanto de colegas estrangeiros quanto de colegas que falam a mesma língua (HARRIS 1991, p. 290 nota 9)<sup>138</sup>.

Através dessa constatação, podemos entender, como Harris, que as linguagens científicas constituem um novo tipo de sistema; aqui coloca-se a noção de que é possível investigar a linguagem de textos de uma área específica sob a orientação de que inerente a eles há um sistema lingüístico de sintaxe distinta, que possibilita a investigação formal dessas mesmas sublinguagens. “Essa visão é embasada pelo fato de que quando em uma dada ciência analisamos artigos escritos em línguas naturais diferentes, obtemos essencialmente a mesma estrutura de sublinguagem” (HARRIS, 1991, p. 290). O referido sistema lingüístico distinto se torna possível porque os sistemas de notação têm à sua disposição, da mesma forma que as sublinguagens, uma metalinguagem externa a eles, a da língua natural.

A linguagem da ciência se apresenta assim por um corpo de fórmulas canônicas, representando as relações de sintaxe da ciência. Sua gramática aponta os símbolos de classes (as letras maiúsculas), os membros de classes (subscritos), os modificadores (sobrescritos), com as restrições expressas neles, e com suas combinações, que constituem tipos de sentenças (cf. HARRIS, 1988, p. 51).

---

<sup>137</sup> Fenômeno de correspondência entre as estruturas de línguas, que possibilita a transposição entre as gramáticas de duas línguas dadas. Nele, as estruturas sintáticas da primeira língua são traduzidas pelas equivalentes na segunda (cf. SALKOF 2002, p. 167-178).

<sup>138</sup> Thus in the immunology study [...], analysis of articles written in French yielded substantially the same formulaic representation as in the case of English articles. It is also relevant that many articles written in one language are read and discussed by scientists who speak other languages, and that the peer pressure which is exerted on article writers to adhere to precise and standard formulations can come from foreign colleagues as well as from same language colleagues.

#### 4.6. Particularidades gramaticais relevantes

As combinações entre os componentes (e.g., operador e argumento) criam os significados gramaticais (cf. HARRIS, 1988, p. 64-65). Nesse princípio repousa a compreensão sobre as particularidades das gramáticas das sublinguagens.

Na linguagem coloquial temos classes de palavras tais como nomes e verbos e adjetivos, com sentenças sendo formadas por sequências particulares dessas palavras. Essas sentenças apresentam afirmações sobre o mundo, algo que, por exemplo, a notação musical não pode fazer não importa que valores sejam dados a seus símbolos. As afirmações podem ser verdadeiras ou falsas ou sem sentido, e pode ser sobre qualquer coisa. No caso dos textos da ciência encontramos as mesmas estruturas básicas, exceto que em cada subciência há subgrupos particulares de nomes que ocorrem com subgrupos particulares de verbos ou outras palavras...<sup>139</sup>(HARRIS, 2002a, p. 215)

Tendo em mente essa concepção, e uma vez que a ordenação de partes é um traço gramatical – uma das restrições que formam a gramática da língua - as sublinguagens são portadoras de restrições diferenciadas e importantes, por conta das peculiaridades em seu sistema de ordenação de partes. Sendo elas estruturadas através de operadores e argumentos que são específicos daquele domínio e que, por ser assim, criam significados também específicos, pois estão sob um mecanismo diferenciado do da língua no que se refere à ordenação de partes, então, sua gramática é diferenciada em relação à da língua.

Para compreendermos esse fenômeno é suficiente que pensemos nos operadores e argumentos do domínio da bioquímica, por exemplo, em duas situações: em ambientes da *linguagem geral* e em seu próprio domínio. Em um texto que não se refere necessariamente ao domínio, os operadores estão restritos a classes específicas de argumentos, porém, a escolha lexical pode se dar sobre uma lista imensamente maior. Assim, para a sublinguagem da bioquímica, a frase

(25) *As proteínas foram tratadas com ácido*

---

<sup>139</sup> In colloquial language we have word-classes such as nouns and verbs and adjectives, with sentences being formed by particular sequences of these words. These sentences present statements about the world, something which, for example, musical notation could not do no matter what values were given to its symbols. The statements may be true or false or non-sensical, and may be about anything. In the case of science writing we find the same basic structure, except that for each subscience there are particular subsets of nouns that occur with particular subsets of verbs or other words...

se apresenta como uma sentença estruturada dentro da regularidade gramatical dessa sublinguagem, tendo como argumentos nomes específicos do domínio, que vão preencher adequadamente as casas à esquerda e à direita do operador *foram tratados*. Tal não ocorre em

(26) *As proteínas foram tratadas com humor*

pois o argumento *com humor* se apresenta como um nonsense em relação à sublinguagem. Entretanto, em um texto que não fosse específico, ela se faria igualmente satisfatória e gramaticalmente adequada. Isso se explica pelo fato de que o significado das palavras são modificados pela relação de probabilidade entre seus operadores e argumentos, que em alguns casos diferem em relação a eles (e. g., nos ambientes de sublinguagens). Dessa forma, podemos mostrar que o significado de uma sentença ou discurso se dá pelo significado de suas palavras sob as restrições. As restrições das sublinguagens para a formação de sentenças produzem diferenças na gramática de seus operadores. Isso se dá porque

quando as sentenças da sublinguagem não são consideradas separadamente, como um universo separado de dados, mas junto com outras sentenças arbitrárias da língua, as palavras  $O_{sol}$  não são encontradas para serem restritas aos pares  $N_{mol}N_{sol}$  nos dados da língua e não constituem classes separadas. Também a sublinguagem tem restrições que não se aplicam à língua<sup>140</sup> (HARRIS, 1991, p. 279).

Dentro dessa concepção, interessa salientar sobre a sublinguagem na citada relação operador/argumento, a probabilidade zero, mesmo dentro de classes pertencentes à sublinguagem. Visto que essas classes são construídas com a ocorrência de seus membros em relações particulares de operador argumento na formação de sentenças, há nesse sentido diferenças em relação à linguagem no que se refere a estabelecer grupos de possíveis tipos sentenciais.

Em se tratando da linguagem natural, as palavras também têm um grupo de operadores/argumentos específicos em relação aos quais elas têm probabilidade de

---

<sup>140</sup> when the sentences of the sublinguagem are considered not separately, as a separate universe of data, but together with arbitrary other sentences of the language, the  $O_{sol}$  words are not found to be restricted to ordered  $N_{mol}N_{sol}$  pairs in the whole-language data, and do not constitute separate classes. Thus the sublinguagem has constraints which do not apply in the language geral.

ocorrência, caracterizando assim seu significado. Em relação a cada operador muitos argumentos diferentes têm diferentes probabilidades de ocorrência, sendo mais alta ou mais baixa. A ocorrência de um argumento depende de estar ele em um grupo (classe) de ocorrências possíveis em relação ao operador; então, operadores como *dormir*, *cair*, *surgir* - que são verbos indicadores de processo e exigem um argumento que possa carregar o traço *afetado* (cf. CAMPOS et al, 1996) - requerem uma palavra que, por sua vez, não requer necessariamente outra e é considerada como tendo probabilidade positiva, não importa quão pequena seja essa probabilidade. Assim, a probabilidade de ocorrência do argumento vai depender da sua relevância semântica em relação ao operador, sendo possível que ocorram argumentos pertencentes à classe *animado* ou *não animado* para o operador *dormir*, por exemplo.

Para isso, a noção de probabilidade de co-ocorrência se coloca como importante. É certo que determinados argumentos apresentam um grau maior de probabilidade de ocorrência em relação à combinação com o operador. Dessa forma, o verbo *dormir* - indicador de experiência - teria, prototipicamente, a princípio, um argumento humano com traço semântico experimentador; entretanto, o verbo admite também argumentos não animados, em construções metafóricas. Harris apresenta a probabilidade *mais alta* e *mais baixa* de tipos de argumento que seriam mais e/ou menos prováveis de ocorrer junto aos operadores, considerando a função semântica dessa combinação.

Na linguagem da ciência não parece ser assim, ao menos no que se refere aos operadores de primeira ordem (cf. HARRIS, 2002a, p. 217). Assim sendo, em uma linguagem da ciência muitas palavras dentro de uma classe de dependências têm probabilidade zero.

Dessa forma, os argumentos de  $Y_c$  (*develops into*), por exemplo, seriam duas palavras de nível zero - duas palavras que não requerem outra, e.g., *cell*, mas não todas as palavras dessa classe.

(27) *Antibody develops into cell*

Muitas palavras de nível zero dentro da sublinguagem, tais como animais ou parte do corpo (B) têm probabilidade zero de posição argumento de  $Y_c$ . Não ocorre, por exemplo,

(28) *The liver develops into a splen*

mesmo sendo esses argumentos representativos de partes do organismo - o que os faria parte do grupo de combinações possíveis - eles não ocorrem, sua probabilidade é zero. Apenas nomes de células e, ocasionalmente, outras palavras ocorrem aqui como argumento de  $Y_c$ .

#### 4.7. Sublinguagem de Base

Uma das sublinguagens de que Harris trata é a sublinguagem do domínio de descrição da própria língua: a linguagem específica que é utilizada para se construir e expressar as relações presentes em um sistema lingüístico. Nas obras *A Grammar of English on a Mathematical Principles* (1982), *A Theory of Language and Information: a mathematical approach* (1991) e *Language and Information* (1988), Harris procede à descrição de como os componentes da língua se acomodam sob as restrições fundamentais, o que ele nomeia de Sublinguagem de Base (e.g., a gramatical). Compreende-se que é denominada assim porque trata de discussões sobre a estrutura da própria linguagem. É através dos elementos formais que estão presentes no domínio dessa sublinguagem que as demais são construídas. Nesse caso, temos elementos para inferir que esta seria a única sublinguagem que se constrói com elementos meta presentes em seu próprio domínio.

Nas obras citadas, são descritas características particulares da gramática da língua inglesa, e, ao tratar dessa sublinguagem de base, isso é feito com vistas a ser uma representação linear de relações particulares entre os componentes da sentença e do discurso (cf. HARRIS, 1982, p. 33) do Inglês. O autor se ocupa de descrever as restrições criadas pela gramática do Inglês e da representação das relações particulares que se dão na construção da sentença e do discurso.

Essa descrição das restrições que criam as sentenças da língua é feita através de uma sublinguagem do domínio lingüístico, com operadores e argumentos específicos - para explicar, por exemplo, os mecanismos de exigência de argumentos, a linearidade da sentença, os indicadores de operadores, indicadores de argumentos e outros mecanismos

intrínsecos ao sistema. Dessa forma, são abordadas as relações sintáticas que regem a ordenação de palavras na frase.

Para essa discussão, são descritos os tipos de operadores que ocorrem na língua (1ª e 2ª ordem), os tipos de argumentos e a forma como os operadores agem sobre os argumentos e também são descritas as restrições que criam as subclasses gramaticais e sua entrada na frase. Nessa relação, ocorre a não equiprobabilidade de ocorrência de palavras; e para descrever esses conceitos e relações, a sublinguagem de base, ou seja, a da língua, se apresenta.

#### 4.8. Aplicabilidade

A representação notacional para Harris, apesar dos problemas que possam vir desse uso, é eficaz, como dissemos anteriormente, para o reconhecimento de estruturas sentenciais e textuais - de forma que possam ser aplicadas em métodos para sumarização de textos – e, principalmente para o processamento em computador (HARRIS, 1970[1959], p. 253-255; 2002a, p. 218).

Uma vez que uma representação formulaica foi desenvolvida para uma sublinguagem, o processamento em computador pode encontrar exemplos de fórmulas individuais (com subclassificadores e modificadores) que representam qualquer informação que esteja sendo buscada – isto por conta da localização intrínseca de informação na estrutura das fórmulas <sup>141</sup>(HARRIS, 2002a, p. 221).

Harris anteviu a aplicabilidade maior de sua teoria de sublinguagem em métodos de processamento de textos de áreas específicas. E, nesse sentido, reitera a sua eficácia para a área da pesquisa. Nesse caso, uma aplicação seria verificar as possibilidades de obtenção de modelos notacionais, para as linguagens da ciência, através de redução simplificada de usos linguísticos atuais (cf. HARRIS, 2002a, p. 220).

A força da abordagem da sublinguagem repousa em embasar a representação do conhecimento na análise de textos atuais. A importância dessa abordagem para a Linguística computacional é que a fase inicial da análise de

---

<sup>141</sup> Once a formulaic representation is developed for a sublinguagem, computer processing could find instances of the individual formulas(with subclassifiers and modifiers) that represent whatever information is being sought – this because of the intrinsic locating of information in the structure of the formulas.

sublinguagem estabelece uma relação direta entre formas de sentenças de superfície e a representação semântica<sup>142</sup> (JOHNSON, 1989, p. 190).

O linguista propõe a teoria apresentada, para se colocar a serviço de processar a linguagem em domínios textuais, especialmente porque as linguagens técnicas (sublinguagens) têm uma estrutura e regularidade possíveis de serem observadas e adequadas para a linguagem computacional (cf. HARRIS, 2002a, p. 215; FRIEDMAN et al, p. 223).

Uma vez que as fórmulas da sublinguagem têm sido estabelecidas, programas de computador podem carregar vários tipos de estoque de informação estruturada e processar, incluir, sumarizar e comparar as questões, especialmente com relação à redundância e fraseologias alternativas, com um instrumento ou vários<sup>143</sup> (HARRIS, 2002a, p. 220).

Essa noção de sintaxe diferenciada para diferentes domínios tem sido deveras produtiva na pesquisa atual. Assume-se que um computador não pode simular uma linguagem natural satisfatoriamente se não compreender o assunto que está em discussão, sendo necessário para isso que se ofereça ao programa um modelo detalhado do domínio específico do discurso com o qual se está trabalhando (cf. DIAS DA SILVA et. al., 2007 p. 13). Para isso, em um primeiro momento, é necessário um inventário lexical relativo ao domínio, seguido da sistematização de regras sintáticas próprias para que o texto possa ser processado a contento.

Quando um analisador, gerador, ou sistema de tradução incorpora tal descrição linguística, torna-se não somente mais eficiente, mas também capaz de discriminar entre sentenças [...] que são apropriadas ao domínio e aquelas que são gramaticais, mas impróprias<sup>144</sup> (KITTTREDGE, 1982. p. 79).

Outro tipo de aplicação seria a que Friedman et al (2002) faz na sublinguagem clínica, que relaciona a estrutura de informação de uma ciência para além de apenas se

---

<sup>142</sup> The strenght of the sublinguagem approach lies in basing the knowledge representation on the analysis of actual texts. The significance of this approach to computational linguistics is that the initial phase of sublinguagem analysis establishes a direct relationship between surface sentence forms and the semantic representation (formulas).

<sup>143</sup> Once the sublinguagem formulas have been established, computer programs can carry out various kinds of structured information storage and processing, including summarizing and comparing the questions, especially with regard to redundancy and alternative wordings, within one instrument or several (loc. cit).

<sup>144</sup>When a parser, generator, or translation system incorporates such a precise linguistic description, it becomes not only more efficient bus also capable of discriminating between sentences [...] that are appropriate to the domain and those that are grammatical but inappropriate.

caracterizar o campo, verificando também se é possível estabelecer uma estrutura da ciência em questão.

Pode-se também investigar similaridades e diferenças entre as fórmulas encontradas para sublinguagens afins. Assim, é possível “ver quão distintas são as coisas sobre que elas falam ou a forma como falam sobre essas coisas; ou seja, para ver até que ponto elas se tornam subciências distintas”<sup>145</sup> (HARRIS, 2002a, p. 220).

Esse trabalho de descrição das sublinguagens é importante também no que se refere a complementar uma descrição da língua como um todo, pois, segundo Kittredge & Lehrberger (1982, p. 2), “uma descrição apropriada de uma língua natural envolve a descrição de suas sublinguagens e as relações entre elas”.

Sendo a teoria de sublinguagem produtiva para se estabelecer as gramáticas de diferentes domínios, ela possibilita a descrição dessas sublinguagens vizinhas, com proveito no que se refere a particularidades ou à intersecção de suas configurações sintáticas.

No trabalho de Friedman et al (2002), por exemplo, em que se buscou verificar particularidades gramaticais das sublinguagens das áreas clínica e biomolecular, com vistas à obtenção e estruturação da informação nas áreas investigadas, ficou evidente que o domínio clínico é de natureza eminentemente descritiva, sendo os nomes predominantes. Por outro lado, a linguagem do domínio biomolecular se apresentou com a dominância de relações de ordem verbal. Isso possibilitou a construção de manuais de padrões de ocorrência nas duas áreas em questão, o que foi útil para possibilitar a leitura de textos da área via programa de computador.

Para que tais programas sejam eficientes, são elaborados dicionários específicos para as sublinguagens, bem como as também específicas regras gramaticais que possibilitarão a leitura adequada. (KITTRIDGE, 1982).

Ainda que não se reconheça oficialmente em citações de pesquisas, que não haja referências a seu nome nos atuais textos que tratam da automatização de textos, a origem para os estudos em análise linguística computacional está em bases desenvolvidas por Harris (cf. 1970[1959]). Nesse artigo, é oferecido algo como um plano-piloto de programa analisador sintático computacional. Nessa proposta, as unidades básicas de cadeias frasais são relacionadas à análise sintática automática, produtiva para procedimentos, como por

---

<sup>145</sup> “... see how distinct are the things they talk about or the way they talk about them; that is, so see to what extent they have become separate subsience”.

exemplo, a sumarização de textos e a tradução automática, assunto-alvo de intensas pesquisas atualmente (STRUBE DE LIMA et al, 2007).

Atualmente, o acesso à informação digital é um dos principais objetivos de pesquisadores envolvidos em pesquisas que visam a ampliar a disponibilidade de informação em diferentes formatos e mídias. E acrescente-se, em diferentes áreas, por isso, de acordo com a concepção de Harris, torna-se necessário trabalhar com uma gramática própria do domínio; ocorrendo o contrário, é possível que os programas descartem “frases que reportam a entidades fora do domínio do discurso estabelecido” (DIAS DA SILVA et. al, 2007, p. 11).

Essa recuperação de informação se torna interessante e necessária em se tratando de documentos e textos disponíveis na web – sublinhe-se, disponíveis em grande quantidade.

A tecnologia empregada varia de simples máquinas de estados finitos a sistemas mais sofisticados, que levam em conta algum tratamento linguístico da consulta e também características do perfil do usuário. O enriquecimento desse processo pode-se dar pela consideração de informação linguística e contextual, em ordem crescente de complexidade, nos níveis morfológico (a formação das palavras da consulta e do texto pode guiar, de alguma forma, a busca), morfossintático (as classes das palavras – nome, verbo, etc. – são indicadores de relevância), sintático (a interrelação entre os elementos de um texto – nome-adjetivo, sujeito-verbo, verbo-objeto, etc. adiciona informação importante), semântico (em nível lexical – o sentido daquela ocorrência – ou em nível sentencial – o significado de uma sentença, ambos contribuem para o sucesso de uma busca) e pragmático-discursivo (pela consideração de questões culturais, por exemplo) (STRUBE DE LIMA, 2007, p. 4).

Em verdade, esses trabalhos envolvendo PLN<sup>146</sup> se fazem mais numerosos, consistentes e necessários a cada dia, agregando as áreas de estudos linguísticos e de tecnologia da informação. Neles, são estabelecidas as classes de palavras e agrupadas por critérios de ordem semântica. Na verdade, há diversos tipos de gramáticas em diferentes formalismos para representação computacional. O objetivo é representar de forma prática as discussões específicas em diferentes ambientes textuais, que são expressos em forma de gráficos conceptuais ou listas adaptadas para o trabalho em computador, através de “construção de interfaces gráficas, isto é, programas que transformam a informação em

---

<sup>146</sup> Processamento de Linguagem Natural

objetos gráficos, facilitando sobremaneira a comunicação entre o usuário e o computador” (DIAS DA SILVA et al, 2007, p. 7).

Com o crescimento da disponibilidade de informações textuais *on line*, é crescente também a demanda por técnicas de PLN, em diversos domínios; isso demonstra a produtividade das pesquisas envolvendo o PLN, e, conseqüentemente, das teorias que a embasam.

A teoria de sublinguagem apresentada se coloca na base dos estudos sobre PLN; Harris trabalhou com a hipótese de que o conteúdo informacional e estrutural delineado pela gramática da sublinguagem poderia ser usado por um processador de linguagem para capturar e codificar informações salientes e relações em textos. Nesse sentido, essa teoria é importante para a Linguística Computacional porque

...é assumido que qualquer sistema aplicado será orientado através de variedades particulares da linguagem natural associadas com um domínio de conhecimento específico. Isso surge do fato atual amplamente aceito de que tais sistemas requerem sem dúvida restrições justas, essencialmente semânticas, para obter uma análise correta, e que tais restrições possam no momento serem afirmadas somente para as sublinguagens, não para a linguagem como um todo. Embora um sistema prático possa bem ter componentes que são designados para acomodar a linguagem toda, ele precisa também antecipar as propriedades particulares sintáticas, lexicais, semânticas e discursivas da sublinguagem na qual ele opera <sup>147</sup>(KITREDGE, 1982, p.. 79).

Essa teoria se fez crucial para o desenvolvimento de trabalhos como o de Kittredge & Lehrberger (1982) de Sagger (apud Friedman et. al. 2002), o já citado de Friedman, et. al (2002), em que, nas palavras dos autores: “os princípios de Harris são aplicáveis a muito de nosso trabalho em processamento de linguagem biomédica”<sup>148</sup> (p. 223).

Uma vez que o método de PLN se propõe no auxílio à busca de informação textual, em diversas áreas, é interessante que se sublinhe que através dele se transpõe os portais da pesquisa puramente linguística. Efetivamente, nesse sentido, tem-se investigações de cunho linguístico que se aplicam a diversas outras áreas do conhecimento. “O tratamento computacional da informação em língua natural, a modelagem matemática da língua e a

---

<sup>147</sup>For each of these areas, it is assumed that any applied system will be oriented toward the particular variety of natural language associated with a single knowledge domain. This follows from the now widely accepted fact that such systems require rather tight, primarily semantic, constraints to obtain a correct analysis, and that such constraints can at present be stated only for sublanguages, not for the language as a whole. Although a practical system may well have components that are designed to accommodate the whole language, it must also anticipate the particular syntactic, lexical, semantic, and discourse properties of the sublanguage in which it will operate.

<sup>148</sup> “... the Harris’s principles are applicable to much of our work in biomedical language processing”.

aquisição do conhecimento lingüístico podem representar inestimável auxílio em termos do acesso universal à informação,...”(STRUBE DE LIMA et al, 2007, p. 12).

Procuramos aqui apresentar a aplicabilidade, na pesquisa atual, da teoria de sublinguagem como proposta por Harris. A sintaxe peculiar, que pode ser observada nos comumente chamados textos técnicos, tanto produz textos por vezes assaz herméticos ao olhar de um leitor não inteirado da referida área (e.g., informática, química), como também produz textos de fácil compreensão para o falante da língua (e.g., gastronomia), mas que necessitam de uma gramática específica para tradução. Como se pode verificar, são noções úteis para estudos lingüísticos formais em que se estabelecem inventários lexicais de domínios específicos. Mas são igualmente promissoras para outros campos, como o de desenvolvimento da tecnologia da informação - que se reconhece - é um campo em que ainda há muitos desafios a serem superados.

## 5. CONSIDERAÇÕES FINAIS

Ao iniciar o estudo da obra de Zellig Harris e sua abordagem dos fenômenos da língua, sabíamos que estávamos diante de alguns desafios. Procuramos superá-los seguindo os passos e métodos da Historiografia Linguística. Mostramos de início nesta pesquisa a figura do linguista Zellig S. Harris, com detalhes biográficos, apresentando também sua trajetória acadêmica como linguista/metodologista.

Apresentamos através desta pesquisa um retorno ao trabalho que Harris realizou em relação à Linguística. Pudemos refletir com mais vagar sobre o tipo de trabalho produzido por ele, bem como sobre as fontes teóricas que lhe deram embasamento para desenvolver as teorias formuladas.

Mostramos no primeiro capítulo um pouco de sua origem e sua trajetória como pesquisador da língua e da linguagem. Para que isso fosse possível, buscamos informações em escritos de ex-alunos seus, colaboradores, bem como de documentos que trouxessem informações a seu respeito. É bem verdade que esses documentos são escassos, e neles, a informação sobre a vida pessoal do linguista, bem resumidas. Pode ser que tenha ficado uma lacuna quanto a esse particular, de modo que fica aberto o espaço para uma construção biográfica mais detalhada nesse sentido. Por outro lado, apresentamos uma resenha da sua carreira como linguista, com base nas obras de maior expressão publicadas por ele.

Foi discutido o movimento estruturalista como sendo um terreno fértil em que germinaram as idéias de Harris sobre a abordagem de língua e linguagem. Pudemos observar que sua teorização se robustece e adquire especificidades dentro do modelo estrutural. Isso foi enriquecedor para a Língua, mesmo porque é interessante que a produção do conhecimento linguístico possa ser construída de diferentes modos, de forma que lance luz de diferentes maneiras sobre esse conhecimento.

Como herdeiro de Sapir e de Bloomfield, vimos que Harris se ocupou da descrição linguística que também foi realizada por aqueles, entretanto, como foi visto, Harris agregou à sua prática de linguista ingredientes da Matemática e da Lógica, que fizeram toda a diferença em suas discussões sobre língua. E conferiu a elas um toque de originalidade (e, deve-se reconhecer, de hermetismo, sobretudo para aqueles que chegam à linguística pelas vias dos cursos de Letras). Entretanto, pelo que foi apresentado nos

capítulos três e quatro, esse contato com os pressupostos da Lógica e da Matemática foi essencial para o desenvolvimento da sua teoria de informação na língua enquanto sistema. A relação que estabeleceu entre os diferentes sistemas de representação foi fundamental para que seu método se tornasse linguisticamente válido e inovador. Por exemplo, sua gramática de operadores dialoga diretamente com a teoria matemática de probabilidades e combinação.

Por conta dessa metodologia diferenciada, reconhecemos Harris como um linguista/metodologista à frente de seu tempo. E o dizemos por termos observado suas descrições de como percebe a informação linguística.

Como vimos no capítulo quatro, a abordagem de sublinguagem hoje abre um campo fértil para a linha de estudos lingüísticos que visa a automação de processamento textual. Investigações que se desenvolvem sobre o processamento de linguagem natural têm suas bases tanto na concepção de sublinguagem quanto na da gramática de operadores de Harris.

Em programas de tradução automática, por exemplo, ao levar em conta que as regras de coocorrência são diferentes nas sublinguagens em relação às regras sintáticas da língua geral, o programa poderá fazer um processamento voltado para o que a gramática daquela sublinguagem admite, e devolver assim uma tradução mais apropriada. Para isso, nos processadores de textos é necessário incluir regras especiais para as possíveis formações sentenciais específicas.

Então, suas gramáticas precisam estar aptas a desconsiderar algumas regras de produção que existem na língua e não na sublinguagem e também utilizar regras sintáticas que se aplicam só a elas e não são reconhecidas pela linguagem natural. Assim, a análise lingüística formal lança luz sobre áreas específicas de realização da língua, em textos escritos, revestidos de particularidades lexicais, sintáticas e semânticas, auxiliando na compreensão de como são estruturados os textos nesses domínios.

Dessa maneira, as fórmulas que são necessárias para a elaboração de programas de tradução automática de línguas necessitam representar a estrutura gramatical das duas línguas envolvidas no processo.

Em verdade, como foi possível perceber pelas idéias harrissianas trazidas a este estudo, em todos os eventos envolvendo a língua(gem), a gramática de operadores se faz presente, até mesmo quando se trata dos elementos não discretos, porque, como expressa o

próprio Harris (1988), esses elementos informam, mas estão a serviço de uma informação que é estruturada pelas restrições que criam as sentenças. Em suas palavras:

Há também significados expressos em associação regular com o discurso, mas externos a ele, tal como os gestos. Qualquer informação substantiva (objetiva) carregada por todos eles pode também ser expressa na língua propriamente, mas somente através das duas restrições [...] que criam sentenças<sup>149</sup> (HARRIS, 1988, p. 59).

Dessa forma, há espaço dentro da obra de Harris para pesquisa voltada para a análise dos componentes linguísticos em termos de restrições, pesquisas que abarquem não só o processamento de linguagem, mas outras áreas da Linguística. Acreditamos que a investigação em termos de classes de palavras e sua distribuição em termos de restrições linguísticas pode ser promissora na investigação de um sistema linguístico que esteja em processo de aprendizado.

A forma como a lingua(gem) é abordada por Zellig Harris abre, enfim, diversas possibilidades para a pesquisa linguística. Desde o estudo dos fonemas, passando pela Sintaxe, a Filosofia da Linguagem, chegando até as interfaces da Linguística com a Informática, são campos em que as discussões de Harris se aplicam com pertinência e mostram que o seu fazer linguístico não se prendeu às origens, ao contrário, eles são base para um movimento dinâmico em que se busque embasar pesquisas atuais em teorias clássicas.

---

<sup>149</sup> There are also meanings expressed in irregular association with speech, but external to it, such as gesture. Any substantive (objective) information carried by all of these can be also expressed in language proper, but only via the[...] two constraints that create sentences.(op. cit)

## 6. Referências Bibliográficas

- BAPTISTA, J. Uma teoria em termos de restrições. Materiais de apoio para a disciplina de Sintaxe. Faro: UALG, 2005a.
- BAPTISTA, J. Mini-curso de Sintaxe. Materiais de apoio para o curso organizado no L2F [Módulo 1 pdf], INESC-ID Lisboa. Faro/Lisboa:UALG/L2F 2005b.
- BEANEY, M. The Frege reader. Oxford: Blackwell, 1997[1879]. p. 47-78.
- BENVENISTE, E. Problemas de Lingüística Geral I; trad de Maria da Gloria Novak e Maria Luiza Néri; 4ª ed. SP, Pontes: Editora Universidade Estadual de Campinas 1995[1966].
- BLOOMFIELD, L. Um conjunto de postulados para a ciência da linguagem. In: DASCAL, M. (org.) Fundamentos Metodológicos da lingüística. São Paulo: Global, 1978[1926]. p. 45-60.
- BLOOMFIELD, L. Language. Boston: George Aleen & Unwin Ltd., 1979[1933].
- BOAS, F. Textos Kathlamet. In: STOCKING JR., G. W.(org.). A formação da Antropologia americana: antologia. Trad. Rosaura Maria C. L. Eichenberg. Rio de Janeiro, Ed. UFRJ, 2004[1883-1911]. p.148-152.
- BOAS, F. Sobre sons alternantes. In: STOCKING JR., G. W.(org.). A formação da Antropologia americana: antologia. Trad. Rosaura Maria C. L. Eichenberg. Rio de Janeiro, Ed. UFRJ, 2004[1883-1911]. p. 98-104.
- BOONS, J.; GUILLET, A.; LECLÈRE, C. La structure des phrases simples en français: constructions intransitives. Genebra: Droz, 1976.
- DE BRABANTER, P. Zellig Harris's theory of syntax and linguistic reflexivity. Ijn\_00000415 version 1. IJN\_Institut Jean Nicod 2001. Disponível em <[http://jeannicod.ccsd.cnrs.fr/ijn\\_00000415/en/](http://jeannicod.ccsd.cnrs.fr/ijn_00000415/en/)>.
- CAMACHO, R. G. Estrutura argumental e funções semânticas. Alfa, v. 43, p 145-170. São Paulo: UNESP, 1999.
- CÂMARA JR. Prefácio e Apêndice a SAPIR, E. A Linguagem: Introdução ao estudo da fala. Trad. Joaquim M. Câmara Jr. São Paulo: Cultrix, 1980[1921]. p. 1-8; 183-203.
- CAMPOS, O. G. L. A. S.; RODRIGUES, A. C. S.; GALEMBECK, P. T. A flexão modo-temporal no português culto do Brasil: formas de Pretérito Perfeito e Imperfeito do Indicativo. In: Gramática do Português Falado v. IV. Campinas, Editora da Unicamp, 1996. p. 35-74.

CHOMSKY, N. Prefácio a CHOMSKY, N. *Syntactic Structures*. Paris: Mouton, 1976[1957]. p. 5-7.

CHOMSKY, N. A linguagem e a mente. In: COELHO, M. LEMLE, M. LEITE, Y. (orgs.) *Novas perspectivas lingüísticas*. Trad. Miriam Lemle. Petrópolis, Rj: Vozes Ltda. 1973 p. 28-42.

CULIOLI, A. Defining the Territory. In: CULIOLI, A. *Cognition and Representation*. Amsterdam: John Benjamins Publishing Company, 1995. p. 12-31.

DIAS DA SILVA, B. C.; RINO, L. H. M.; SPECIA, L.; NUNES, M. G. V.; OLIVEIRA JR., O. N. de; MARTINS, R. T.; PARDO, T. A. S. Introdução ao Processamento das Línguas Naturais e Algumas Aplicações. Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional. USP/UFSCar/UNESP, 2007. p. 1-59.

DOOSSE, F. História do Estruturalismo, v. 1: o campo do signo. 1945-1966. Trad. Álvaro Cabral. São Paulo: Ensaio, 1993.

ECO, U. A estrutura ausente. São Paulo: Perspectiva, 2001.

FREGE, G. Begriffsschrift: a formula language of pure thought modelled on that of arithmetic. In: BEANEY, Michael. *The Frege reader*. Oxford: Blackwell, 1997[1879]. p. 47-78.

FREGE, G. Sobre o sentido e a referência. In: *Lógica e Filosofia da Linguagem*. São Paulo: Cultrix, 1978. p. 60-86.

FRIEDMAN, C.; KRA, P.; RZHETSKY, A. Two biomedical sublanguages: a description based on the theories of Zellig Harris. In: *Journal of Biomedical Informatics* v. 35 p. 222-235. Agosto, 2002.

GOLDSMITH, J. Review of The Legacy of Zellig Harris: Language and Information into the 21<sup>st</sup>. century. Vol 1. Disponível em [http://hum.uchicago.edu/~jagoldsm//Papers/Nevin\\_ZelligHarris\\_Goldsmith\\_review.pdf](http://hum.uchicago.edu/~jagoldsm//Papers/Nevin_ZelligHarris_Goldsmith_review.pdf).> Acessado em 20/03/2008.

GOTTFRIED, M. Grammatical specification of scientific sublanguages *In* NEVIN, B. (org.). *The Legacy of Zellig Harris V 1*, John Benjamins Publishing Company. Philadelphia, 2002.

GROSS, M. *Méthodes en syntaxe*. Paris: Hermann, 1975.

GROSS, M. Les Bases Empiriques de la notion de predicat semantique, *Langages*, n. 63 1981, p. 7-52.

GROSS, M. *Grammaire transformationelle du français 1 : syntaxe du verbe*. Paris: Cantilène, 1986a.

GROSS, M. Grammaire transformationnelle du français 2: syntaxe du nom. Paris: Cantilène, 1986b.

GROSS, M. Grammaire transformationnelle du français 3: syntaxe de l'adverbe. Paris: ASSTRIL, 1986c.

GROSS, G. Sélection et sémantique: classes d'objets, compléments appropriés, compléments analysables. In: Langages vol. 28, n° 115. Paris: 1994, p. 31-46.

GROSS, M. Consequences of the metalanguage being included in the language. In: NEVIN, B.(org.) The Legacy of Zellig Harris. Vol. I. Philadelphia: John Benjamins Publishing Company. 2002, p. 57-67.

HARRIS, Z. S. A Grammar of the Phoenician Language. Ph.D. dissertation. New Haven: American Oriental Society, 1990[1936].

HARRIS, Z. S. Yokuts Structure And Newman's Grammar. In: Papers in Structural and Transformational Linguistics. New York: D. Reidel Publishing Company, 1970[1944]. p. 188-208.

HARRIS, Z. S. Navaho Phonology and Hoijer's analysis. In: Papers in Structural and Transformational Linguistics. New York: D. Reidel Publishing Company, 1970[1945] p. 177-187.

HARRIS, Z. S. From morpheme to utterance. In: Papers in Structural and Transformational Linguistics. Holanda, D. Reidel Publishing Company, 1970[1946]. p. 100-125.

HARRIS, Z. S. Structural Linguistics. Chicago, The University of Chicago Press, eight impression, 1969[1951].

HARRIS, Z. S. Discourse analysis. In: Papers in Structural and Transformational Linguistics. Holanda, D. Reidel Publishing Company, 1970[1952] p. 312-348.

HARRIS, Z. S. Transfer Grammar. In: Papers in Structural and Transformational Linguistics. Holanda, D. Reidel Publishing Company, 1970[1954] p. 139-157.

HARRIS, Z. S. Computable Syntactic Analysis: the 1959 Computer Science-Analyzer. In: Papers in Structural and Transformational Linguistics. Holanda, D. Reidel Publishing Company, 1970[1959] p. 252-277.

HARRIS, Z. S. A Language for international Cooperation. In: Preventing World War III: Some Proposals (ed. By Q. Wright et al.), Simon & Scuster, New York, 1962.

HARRIS, Z. S. Mathematical Structures of Language. Pennsylvania: Interscience publishers, 1968.

HARRIS, Z. S. A Theory of Language Structure. In: Papers on Syntax. Dordrecht: D. Reidel Publishing Company, 1981[1976] p. 352-376.

HARRIS, Z. S. Grammar on Mathematical Principles. In: Journal of Linguistics. In Papers on Syntax. Boston, D. Reidel Publishing Company, 1981[1978].

HARRIS, Z. S. A Grammar of English on a Mathematical Principles. New York: John Wiley & Sons, 1982.

HARRIS, Z. S. Language and Information. New York: Columbia University Press, 1988.

HARRIS, Z. S.; GOTTFRIED, M.; RYCKMAN, T.; MATTICK JR. P.; DALADIER, A.; HARRIS, T. N.; HARRIS, S. The form of Information in Science: analysis of an immunology sublanguage. Dordrecht: Kluwer Academic, 1989.

HARRIS, Z. S. A Theory of Language and Information: a mathematical approach New York: Oxford University Press, 1991.

HARRIS, Z. S. The Structure of Science Information. In: Journal of Biomedical Informatics, Volume 35. August, 2002a, p. 215-221.

HARRIS, Z. S. The background of Transformational and Metalanguage analysis. In: NEVIN, Bruce (org). The Legacy of Zellig Harris. Vol. I. John Benjamins Publishing Company. Philadelphia, 2002b p. 1-13.

HIZ, H. Specialized Languages of Biology, Medicine and science and connections between them. In: KITTREDGE, R.; LEHRBERGER J., (ed.). Sublanguage: Studies of Language in Restricted Domains. Berlin, 1982 p. 206-212

HIZ, H. Zellig S. Harris. In: Proceedings of the American Philosophical Society, v. 138, n° 4. 1994 p. 1-6.

HJELMSLEV, L. Prolegômenos a uma teoria da linguagem. Trad. J. Teixeira Coelho Neto. São Paulo, Perspectiva, 1975[1961].

HOUAISS, A. (ed.) Webster's Dicionário Inglês-Português. Rio de Janeiro: Record, 1982.

HUGHES, S. "Speech," The Pennsylvania Gazette, Volume 99 July/August 2001, pages 38-45. Disponível em: <<http://www.upenn.edu/gazette/0701/hughes.html>>

ILARI, R. O estruturalismo lingüístico: alguns caminhos. In: MUSSALIM, Fernanda; BENTES, Anna Christina (Orgs.). Introdução à lingüística: fundamentos epistemológicos. São Paulo: Cortez, 2004 p. 53-92.

JOHNSON, D. (reviewer). of: BARSKY, R. F. Noam Chomsky: a Life of Dissent. In: Journal of the History of the Behavioral Sciences: v. 34(2), p. 173-180. John Wiley & Sons, 1998[1997].

JOHNSON, S. B. (reviewer). HIZ, H.; GOTTFRIED, M.; RYCKMAN, T.; MATTICK Jr., P. DALADIER, A.; HARRIS, T. N.; HARRIS, Z. S. The form of Information in Science: Analysis of an Immunology Sublanguage. In: Computational linguistics, v. 15 n° 3, September 1989 p. 190-192.

KITTREDGE, R.; LEHRBERGER J., (ed.). Sublanguage: studies of language in restricted semantic domains. Berlin: De Gruyter, 1982. Disponível em:

<http://books.google.com/books?id=jaLsYuzT3uwC&pg=PA1&dq=%22mathematical+structures+of+language%22&hl=pt-BR&sig=3mwG1KPUeOeXZmf7ifwSycy1TU#PPA1,M1>

KITTREDGE, R. Sublanguages. In: American Journal of Computational Linguistics v. 8 n° 2. Montreal, Université de Montreal, abril-june 1982 p. 79-84.

LACEY, H. M. Prefácio a: RUSSEL, B. Ensaio escolhidos. Trad. Pablo Ruben Mariconda. São Paulo: Abril Cultural, 1978 (coleção *Os pensadores*) p. VI-XX.

LE PESANT, D. & MATHIEU-COLAS M.. Introduction aux classes d'objets. Langages 131 : 6-33. Paris , Larousse, 1998.

LYONS, J. Semântica. Lisboa: Editorial Presença/Martins Fontes, 1980[1977] p. 11-25.

MATTHEWS, P. Zellig Sabbetai Harris. In: Language 75.1, 1999. p. 112-119.

MICHAELIS ON LINE. Disponível em:

<http://michaelis.uol.com.br/moderno/portugues/index.php?lingua=portugues-portugues&palavra=mentir> acessado em 15/05/08

NEVES, M. H. de M. A gramática funcional. São Paulo: Martins Fontes, 1997.

NEVIN, B. Zellig S. Harris: An appreciation. In: California Linguistic Notes 1992. p. 60-64.

NEVIN, B. The Legacy of Zellig Harris (org.). Vol. I: Philosophy of Science, Syntax and Semantics. John Benjamins Publishing Company. Philadelphia, 2002 p. i-xxxiv.

PENCO, C. Introdução à filosofia da linguagem. Trad. Ephraim F. Alves. Petrópolis: Editora Vozes, 2006.

PIAGET, J. O Estruturalismo. Trad. Moacir R. de Amorim. São Paulo: Difusão Européia do Livro, 1970.

PUTNAM, H. Preface. In: GOTTFRIED, M.; RYCKMAN, T.; MATTICK JR. P.; DALADIER, A.; HARRIS, T. N.; HARRIS, S. The form of Information in Science: analysis of an immunology sublanguage. Dordrecht: Kluwer Academic, 1989. p. xi-xiii.

RANCHHOD, E. M. Sintaxe dos predicados Nominiais com estar. Instituto nacional de investigação Científica – Centro de lingüística da Universidade de Lisboa. Lisboa, 1990.

RYCKMAN, T. A. Method and theory in Harris's Grammar of Information. In: NEVIN, Bruce E. The Legacy of Zellig Harris. Vol. I. John Benjamins Publishing Company. Philadelphia, 2002 p. 19-37.

SALKOFF, M. Some New results on Transfer Grammar. In: NEVIN, Bruce E. *The Legacy of Zellig Harris*. Vol. I. John Benjamins Publishing Company. Philadelphia, 2002. p. 167-178.

SALUM, I. N. Prefácio a SAUSSURE, F. *Curso de Linguística Geral 1994[1916]*, p. xv.

SAPIR, E. Language. In: MANDELBAUN, D. G. (Ed.) *Culture, Language and Personality*. Berkeley: University of Califórnia Press, 1949, p. 1-44.

SAPIR, E. *A Linguagem: Introdução ao estudo da fala*. Trad. Joaquim M. Câmara Jr. São Paulo: Cultrix, 1980[1921].

SAUSSURE, F. *Curso de Linguística Geral*. São Paulo: Cultrix, 1994[1916].

SHANNON C. E. A Mathematical Theory of Communication. In *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July-October, 1948. Disponível em: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>

SILVA, M. Princípios metodológicos e fundamentação teórica da gramaticografia: por uma história cultural da gramática portuguesa. In: *Revista da ABRALIN*, v. 5 n. 1 e 2 p. 68-81. São Paulo, IP PUC, dez. 2006.

SOUZA e SILVA, M. C. P.; KOCH, I. G. V. *Lingüística Aplicada ao Português: Sintaxe-10ª ed.* São Paulo: Cortez, 2001.

STRUBE DE LIMA, V. L. ; NUNES, M. G. V. ; VIEIRA, R. . Desafios do Processamento de Línguas Naturais. In: 34o. Seminário Integrado de Software e Hardware, 2007, Rio de Janeiro. *Anais do XXVI Congresso da Sociedade Brasileira de Computação*. Rio de Janeiro, 2007. v. 1. p. 202-216.

VALE, O. A. *Expressões Cristalizadas do Português do Brasil: uma proposta de tipologia*. Tese (doutorado) – UNESP, Araraquara, 2001.

VALE, O. A. Substantivos portugueses não flexionam em gênero. In: *Palavra* nº 12. Rio de Janeiro, Ed. Galo Branco, 2004. p. 193-209.

WATT, W.C. . Zellig Sabbetai Harris: A biographical Memoir. In: *Biographical Memoirs*, v. 87.p. 1-24. Washington, D.C.: The National Academies Press, 2005.

WHORF, B. L. Science and linguistics In: SMOLINSKY, F. (ed.). *Landmarks of American language and linguistics*. Washington, D. C.: Bureau of Educational and Cultural Affairs, 1988[1940]. p. 31-38.