

# Gramática e Linguagem nos Estudos Formais

Livy Real (UFPR)<sup>1</sup>

Ricardo Andrade (IME-USP)<sup>2</sup>

Resumo: Considerando a interface entre linguística e ciência da computação, faz-se necessária a discussão de seus objetos de estudo e da nomenclatura assumida por cada uma destas áreas. Para isso, abordaremos as definições de *linguagem* e *gramática* geralmente utilizadas pela ciência da computação (CC), a saber pela teoria de autômatos e linguagens formais, e pela linguística formal (LF), particularmente a de base montagoviana. Na CC, *linguagem* pode ser definida como os *strings* (palavras ou cadeias de símbolos) formados a partir dos elementos da *gramática*, que são, basicamente, regras de boa formação e léxico. Já na LF, *grosso modo*, *linguagem* é um termo mais abrangente que *gramática*, sendo a gramática parte da linguagem. Geralmente, entende-se que *gramática* é o conjunto de regras capaz de originar parte da linguagem. No entanto, sob o rótulo *linguagem*, há ainda elementos de fundamental importância para o linguista, como o léxico e a ontologia da língua. Comparando *gramática* e *linguagem* em ambas as áreas, percebe-se que enquanto na CC a linguagem é gerada pela aplicação da gramática, na LF, a gramática é um elemento a mais na linguagem, sendo esta última o rótulo pelo qual entendemos o objeto mais abrangente que temos. Considerando que as duas áreas entendem *gramática* e *linguagem* de maneira ligeiramente diferente, discutiremos (i) que tipo de objeto é criado considerando os construtos teóricos de cada área, e (ii) em que medida é a preocupação com línguas artificiais ou com línguas naturais que produziu as diferenças que encontramos nestes construtos.

## 1) Introdução

Este texto é parte de um estudo maior que visa esmiuçar semelhanças, diferenças e fronteiras entre os estudos formais na linguística e os estudos linguísticos dentro da ciência da computação. Tal estudo, entendemos, faz-se necessário, considerando-se a grande quantidade que há de trabalhos nestas áreas, ou mesmo na interface delas, que parecem fazer uso dos mesmos conceitos, ainda que sejam trabalhos em duas áreas que podem ser consideradas bastante distantes: o estudo da linguagem humana e o estudo da computação. Ao percebemos que aparentemente conceitos como *gramática*, *linguagem* e *léxico* são relevantes para ambas as áreas, ainda que estas pareçam ter preocupações diferentes - isto é, parecem ter objetos de estudo diferentes - discutimos em que medida estes conceitos são os mesmos e, caso não sejam, em que medida uma área pode emprestar certo conceito de outra sem que haja uma reformulação de sua significação.

Para iniciar esta discussão, consideraremos os conceitos de *gramática* e *linguagem* nos estudos formais da linguística e na Teoria de Autômatos e Linguagens Formais, que se insere dentro da Teoria da Computação. O que entendemos aqui por “estudos formais da linguística” são basicamente as teorias gerativista e montagoviana, que se inserem no chamado formalismo linguístico. Consideraremos apenas a visão formalista da linguística por ser esta a que mais se aproxima e a que mais interfaceia os estudos computacionais, mas de forma alguma negamos a necessidade dos estudos funcionalistas dentro da linguística. Em um estudo no qual a principal preocupação fosse comparar estes mesmos conceitos dentro da própria linguística, consideraríamos, obviamente, as concepções funcionalistas de *gramática* e *linguagem*.

Optamos tratar destes dois conceitos, *gramática* e *linguagem*, por nos parecerem muito básicos e pertinentes às duas áreas. Devemos ainda ressaltar que, para o estudo destes

---

<sup>1</sup> Programa de Pós-Graduação em Letras da Universidade Federal do Paraná/CAPES

<sup>2</sup> Programa de Pós-Graduação IME-USP/CAPES

conceitos em uma literatura em português, é necessário um certo cuidado com a tradução de determinados termos técnicos, como *language* do inglês. Adotaremos aqui *linguagem* como o termo que traduz *language* (inglês) e *langue* (do francês). Não trabalharemos casos nos quais estes dois termos claramente definem o que em português definimos como *língua*. Nestas reflexões, o único momento no qual poderíamos traduzir *language* por *língua* é na expressão *natural language*, que, em português, traduzimos normalmente por *língua natural*. Como nos pontos em que a citamos não é problemático traduzi-la por *linguagem natural*, já que não estaremos tratando especificamente de uma língua e sim da própria faculdade/habilidade da linguagem, manteremos nossa tradução de *language* por *linguagem*.

Para a discussão que propomos, analisaremos as definições de *gramática* e *linguagem* assumidas por J. Lyons, em *Semantics* (primeira edição em 1977) e em *Language and Linguistics* (primeira edição de 1981), e por J. Dubois *et alli*, em *Dictionnaire de Linguistique* (primeira edição de 1973), como rótulos comumente assumidos dentro da linguística, em contraposição às definições de B. Partee *et alli*, em *Mathematical Methods in Linguistics* (primeira edição de 1990), e N. Chomsky, em *Knowledge of Language* (primeira edição de 1986), como concepções largamente assumidas no que chamamos de estudos formais. Para comparar tais definições à assumida pela ciência da computação, adotaremos as definições de J. Hopcroft *et alli* (2002) e de Sipser (2006), manuais de grande aceitação por parte dos teóricos da área.

Desta forma, pretendemos, mesmo que de forma resumida, apreender o que estas áreas assumem sob estes dois rótulos tão comuns. Mais uma vez, visamos, com esta comparação, não só delimitar as possíveis diferenças que há entre estes conceitos em cada uma das subáreas, como também discutir em que medida o empréstimo destes conceitos se deu cegamente, isto é, sem considerar as diferenças entre as áreas que, em última análise, têm distintos objetos de estudo. Pretendemos ainda discutir em que medida não seria o próprio objeto de estudo que moldaria o que é entendido sob o mesmo rótulo teórico nas diferentes áreas.

## 2) Linguística

### 2.1) Dubois *et alli*

Para os autores, no renomado *Dicionário de Linguística* (1999), *linguagem*:

é a capacidade específica à espécie humana de comunicar por meio de um sistema de signos vocais (ou língua), que coloca em jogo uma técnica corporal complexa e supõe uma existência de uma função simbólica e de centro nervosos geneticamente especializados. Esse sistema de signos vocais utilizado por um grupo social (ou comunidade linguística) determinado constitui uma língua particular (...) (p.387).

É possível notar que a preocupação do texto é delimitar *linguagem* como algo intrínseco

à espécie humana, o que exclui a comunicação de animais e as chamadas linguagens formais, como a lógica e as linguagens de programação. Também é evidente, nesta definição, uma certa preocupação com a relação entre língua, linguagem e a necessidade de uma comunidade linguística para que uma língua se realize.

Já *gramática* recebe uma definição mais complexa (p.313-6). Além de possuir entradas para *gramática geral* e *gramática gerativa*, o dicionário aponta ao menos quatro conceitos que estão sob o rótulo gramática: (i) “descrição completa da língua, isto é, dos princípios de organização da língua”; (ii) “descrição dos morfemas gramaticais e lexicais, o estudo de suas formas (flexão) e de suas combinações para formar palavras (formação de palavras) ou frases (sintaxe)”; (iii) “descrição dos morfemas gramaticais (artigos, conjunções, preposições, etc), excluindo-se os morfemas lexicais (substantivos, adjetivos, verbos, advérbios de modo, etc), e a descrição de suas regras”; (iv) “para a linguística gerativa, é o modelo de competência ideal que estabelece certa relação entre o som (representação fonética) e o sentido (interpretação semântica)”.

É interessante notar a relação existente entre as três primeiras definições apontadas por Dubois *et alli*: (i) é a descrição geral da língua, (ii) é a descrição lexical e morfosintática de uma língua, e (iii) é a descrição sintática de uma língua. Por sua vez, a definição (iv) é já uma referência, ainda que bastante rasa, ao que a gramática gerativa entenderia por gramática: um modelo que gera pares de som e significado.

## 2.2) Lyons

Ainda que antiga, o *Lingua(gem) e Linguística: uma introdução* de John Lyons (1981) é ainda uma importante referência, não só histórica, para muitos cursos de graduação. Isto, acreditamos, por sua clareza e simplicidade discutindo temas, como *o que é a linguística*, que parecem óbvios, mas são de difícil definição. Por estas duas razões, sua relevância e sua simplicidade, adotamos também as definições de *linguagem* e *gramática* de seu manual como exemplo do que podemos entender sob estes rótulos “linguisticamente” falando.

Boa parte de seu primeiro capítulo é uma discussão sobre como pode ser definido *linguagem*. Para tal, Lyons revisita importantes teóricos - como Sapir, Bloch & Trager, Hall, Robins e Chomsky - e discute a visão de linguagem de cada um deles e suas diferentes linhas teóricas. Sua discussão é longa e interessante, partindo do estruturalismo norte-americano e chegando à gramática gerativa, podendo ser resumida no fato do conceito *linguagem* estar sempre relacionado à determinada corrente teórica, que definirá como o próprio conceito deve ser entendido. Teorias de cunho funcionalista verão na *linguagem* “o sistema pelo qual grupos sociais interagem” ou “o método de comunicação humano”, enquanto teorias de cunho formalista priorizarão a linguagem enquanto sistema estrutural.

Em relação ao termo *gramática*, Lyons o define como o elemento formal da língua que descreve sua sintaxe e morfologia em oposição a outros níveis linguísticos, como a fonologia e a semântica (1982, p.101; 1984, p.378). O que podemos notar é que enquanto Lyons apresenta uma profunda preocupação com o termo *linguagem*, *gramática* parece ser somente as regras que definem formação de palavra e de frases.

## 2.3) Linguística Formal

Nesta seção, veremos as definições de *gramática* e *linguagem* sob o rótulo que chamamos anteriormente de linguística formal, isto é, a área da linguística que vê a linguagem antes como estrutura do que como uso ou comunicação. Para tal, elegemos duas referências bibliográficas relevantes para estudiosos da área: *Mathematical Methods in Linguistics* de B. Partee *et alli* (1990) e *Knowledge of Language*, N. Chomsky (1986). Entendemos que ao esmiuçarmos as definições de Chomsky, a maior referência da gramática gerativa, e do manual de Partee *et alli*, usado largamente por teóricos formalistas não-gerativistas e gerativistas, conseguiremos desenhar um quadro razoável do que pode ser entendido como *gramática* e *linguagem* para linguistas formalistas.

### 2.3.1) Chomsky

Antes de começarmos a esmiuçar as definições dadas por Chomsky (1986) é relevante comentarmos que, diferentemente de Dubois *et alli* (1999) e Lyons (1982, 1984), o autor não pretende escrever um manual, um livro genérico a ser consultado ou para servir de introdução à linguística. Seu texto *Knowledge of Language* insere-se em um momento histórico no qual a gramática gerativa já contava com um certo prestígio, o que permite o autor ser mais direto em asserções que concernem especificamente sua teoria, isto é, Chomsky já neste livro pode contar que seus leitores já conheçam as definições de *linguagem* e *gramática* usadas normalmente na linguística, e que estes mesmos leitores já saibam que a gramática gerativa entenderá a linguagem mais como forma do que como método de comunicação e iteração. Graças a este conhecimento prévio do leitor de Chomsky aliado ao fato de ser este o texto que apresenta uma grande defesa de determinados princípios da teoria gerativa (como a delimitação de seu grande objeto de estudo: a *gramática universal*), as definições apresentadas aqui serão consideravelmente mais especializadas do que as presentes em Dubois *et alli* (1999) e Lyons (1982, 1984).

Para Chomsky, *gramática* só pode ser definido em função do que entende-se por *linguagem*: *linguagem (language)* abrange, na verdade, dois conceitos que, apesar de serem muito próximos, não são exatamente o mesmo. Para diferenciar estes conceitos, o autor distingue *I-Language* (linguagem internalizada) e *E-Language* (linguagem externalizada). A *I-Language* é o objeto de estudo da gramática gerativa: elementos mentais que caracterizam o saber uma língua (Chomsky, 1986, Cap. 1). A linguagem internalizada já não depende da comunidade linguística, do uso da língua, ou, em última instância, da fala *saussuriana*. Já a *E-language*, a linguagem externalizada, é definida como um “epifenômeno da linguagem”, isto é, a linguagem externalizada é a forma como a linguagem internalizada se realiza, mas não é a língua/linguagem em si.

Nota-se, então, que o conteúdo que autores como Dubois e Lyons entendiam estar sob o rótulo *linguagem* é dividido por Chomsky (1986) em dois e que apenas uma destas partes, a *I-language*, é objeto de seu interesse. Isto, obviamente, tem também implicações sobre o rótulo *gramática*. Enquanto os autores anteriores chamavam de gramática, *grosso modo*, as regras da língua, em *Knowledge of Language*, há pelo menos duas noções relevantes e distintas do que é *gramática* em relação ao estudo da *I-language*. A primeira diz respeito à teoria que tenta

explicar o funcionamento da I-language através de abstrações mentais que geram uma língua específica (p. 22). A segunda, já bastante mais complexa, pressupõe a ideia de que todos os seres humanos, em condições normais, possuem algo biológico que possibilita a aquisição da fala. Tal capacidade seria a *gramática universal*, um dispositivo inato no ser humano relacionado à linguagem (p.24).

Considerando as noções de *gramática* e *linguagem* presentes em Chomsky (1986), vemos que o objeto de estudo tratado pelo autor é distinto dos objetos tratados anteriormente e que conceitos básicos como *gramática* e *linguagem* adquiriram um estatuto bastante mais teórico.

### 2.3.2) Partee *et alli*

Como vimos, as definições apresentadas por Chomsky (1986) são de carácter bastante mais específico do que aquelas que visitamos anteriormente. Agora, consideraremos as definições bastante precisas dada por Partee *et alli* (1990), que são também bastante específicas, embora o texto se trate de um manual. Partee *et alli* (1990), em *Mathematical Methods to Linguistics*, pretendem escrever um manual de conceitos e usos básicos de termos da matemática que podem ser úteis a linguistas. Desta forma, ainda que o texto se trate de um manual, é algo consideravelmente específico, o que refletirá na natureza das definições. Tais definições serão também mais formais do que as que encontramos anteriormente, incluindo mesmo as propostas pro Chomsky (1986).

*Linguagem*, então é matematicamente definida como “um conjunto finito de palavras, morfemas ou alguma sequência finita de caracteres” (Cap. 16), ou seja, *linguagem* é todo o conjunto de elementos de uma língua. Ao definir *linguagem* como o conjunto completo de elementos de uma língua, afasta-se radicalmente das demais definições que encontramos na linguística, e, como veremos, se aproxima da definição usada na ciência da computação. Vale a pena notar que, para Partee *et alli*, nem toda sequência de caracteres faz parte da linguagem, isto é, há sequências formadas por caracteres válidos que não fazem parte da linguagem. É justamente neste ponto que o conceito de *gramática* é relevante: é a *gramática* que vai ditar quais são as sequências gramaticais e que fazem parte da linguagem e quais não são.

Assim, definem *gramática* como sendo:

um sistema dedutivo de axiomas e regras de inferência que geram as sentenças de uma linguagem e seus teoremas. Pelas definições usuais, uma gramática contém apenas um axioma, o símbolo inicial e um número finito de regras na forma:

$ip \rightarrow w,$

onde  $ip$  e  $w$  são cadeias de caracteres. (*cf.* Partee *et alli*, 1990, Sec. 16.2)

O que a definição dos autores nos diz é que uma gramática é um axioma adicionado às regras que geram uma linguagem. Vale a pena notar que embora a definição de gramática defina *sentenças* como o produto final que fará parte da linguagem, estas sentenças não são necessariamente sentenças complexas como as que vemos em sintaxe, *sentenças* (também

chamadas de *cadeias* ou *expressões*) são o resultado da aplicação da regra. Assim, se estamos tratando de uma gramática de formação de palavras, as *sentenças*, ou seja, os objetos formados pelas regras, serão palavras e os elementos formadores sons/letras/morfemas.

Na definição dos autores há também outros elementos nem sempre conhecidos dos linguistas: o *símbolo inicial* e as *cadeias*.

*Cadeias* são o resultado da aplicação das regras, em nosso exemplo acima, no qual tratamos de uma gramática de formação de palavras, as cadeias (ou sentenças) serão palavras. Em uma gramática para o nível sintático, as cadeias serão sintagmas ou frases.

O *símbolo inicial* é a variável primeira da geração de uma linguagem, em geral, S(entença).

É possível ainda notar que Partee *et alli* já na própria definição de gramática explicitam qual será o formato possível para uma regra, neste caso, a regra de reescritura, muito similar às regras de reescritura da gramática gerativa da década de oitenta.

Apresentaremos agora um pequeno exemplo de uma gramática e de uma linguagem nos moldes “matemáticos”:

Linguagem X:

Maria ama Pedro.

Pedro ama Maria.

Maria odeia Pedro.

Pedro odeia Maria.

Maria vê Pedro.

Pedro vê Maria.

Gramática da Linguagem X:

Caracteres:

maria, ama, pedro, odeia, vê

símbolo inicial:

S

Regras de inferência:

$S \rightarrow NP + VP$

$VP \rightarrow V + NP$

NP  $\rightarrow$  maria

NP  $\rightarrow$  pedro

V  $\rightarrow$  odeia

V  $\rightarrow$  ama

V  $\rightarrow$  vê

Como podemos ver, a gramática da Linguagem X é o que define quais as cadeias que estão ou não na Linguagem X e ao mesmo tempo a gramática da Linguagem X é construída a partir da própria Linguagem X. É notável, então, uma relação intrínseca e necessária entre *gramática* e *linguagem* dentro dos estudos formais.

### 3) Ciência da Computação

Para a Ciência da Computação, os termos *linguagem* e *gramática* são precisamente definidos e aparecem inicialmente associados à teoria de autômatos, um ramo da teoria da computação que lida com definições e propriedades de modelos matemáticos de computação, também chamados de máquinas abstratas, e que lida também com os problemas que podem ser resolvidos por tais máquinas.

Nesta seção, apresentaremos as definições de *linguagem* e *gramática* dentro da teoria de autômatos, e, para isto, apresentaremos também noções anteriores a estas, como *alfabeto* e *string*, necessárias à compreensão destes termos. Pretendemos também esclarecer a relação dos mencionados termos com a teoria de autômatos, seus objetos de estudo e suas aplicações, bem como discorrer sobre a importância desses conceitos para a ciência da computação. Utilizamos como referências para esta seção os livros de *Hopcroft et alli* (2002) e de *Sipser* (2006), considerados textos base de cursos introdutórios à teoria da computação.

#### 3.1) Alfabetos

“Um *alfabeto* é um conjunto de símbolos finito e não vazio” (cf. *Hopcroft et alli*, 2002, Sec. 1.5.1). Comumente são utilizadas as letras maiúsculas gregas  $\Sigma$  e  $\Gamma$  para denotar *alfabetos*, por exemplo:

a)  $\Sigma = \{0,1\}$

b)  $\Sigma = \{a,b,c,\dots,z\}$

#### 3.2) Strings (Cadeias)

“Um *string* (ou às vezes *palavra* ou também *cadeia*) é uma sequência finita de símbolos

escolhidos de um alfabeto” (cf. Hopcroft et alli, 2002, Sec. 1.5.2). Por exemplo, a palavra *linguagem* é uma cadeia do alfabeto  $\Sigma = \{a,b,c,\dots,z\}$ .

Se  $w$  é uma cadeia sobre um alfabeto  $\Sigma$ ,  $|w|$  é o comprimento ou número de símbolos de  $w$ . Existe uma cadeia especial chamada de cadeia vazia, comumente denotada pela letra grega  $\epsilon$ , que possui comprimento nulo.

Outro conceito relevante é o de potências de um alfabeto. Se  $\Sigma$  é um alfabeto,  $\Sigma^n$  é o conjunto de cadeias com comprimento  $n$  obtidas a partir de  $\Sigma$ . O conjunto de todas as cadeias escolhidas a partir de um alfabeto  $\Sigma$  é convencionalmente denotado por  $\Sigma^*$ .

### 3.3) Linguagens

“Uma *linguagem* é um conjunto de cadeias” (cf. Sipser, 2006, p.14). Mais formalmente, se  $\Sigma$  é um alfabeto e  $L \subseteq \Sigma^*$ , então  $L$  é uma linguagem sobre  $\Sigma$ . Uma forma comum de descrever uma linguagem é utilizando formadores de conjuntos, por exemplo, podemos descrever a linguagem  $L_{Pr}$ , uma linguagem formada apenas por números primos, da seguinte forma:  $L_{Pr} = \{w \mid w \text{ é primo}\}$ . Note que não há qualquer obrigatoriedade em utilizar uma gramática para descrever uma linguagem.

Na teoria de autômatos, existe o interesse de estudar os problemas que os computadores, ou mais precisamente os modelos matemáticos dos computadores, podem ou não resolver. O interesse pelo estudo de linguagens nasceu do fato de que alguns tipos de problemas, chamados de problemas de decisão, são equivalentes a verificar a pertinência de uma cadeia  $w$  a uma linguagem  $L$ . Essencialmente, um problema de decisão pode ser visto como uma pergunta descrita em algum sistema formal que aceita uma resposta do tipo *sim* ou *não*. Podemos formalizar o conceito de problema de decisão como um problema de pertinência a uma linguagem da seguinte forma:

Se  $\Sigma$  é um alfabeto e  $L$  é uma linguagem sobre  $\Sigma$ , então o problema  $L$  consiste em, dado um *string*  $w$  em  $\Sigma^*$ , decidir se  $w$  está ou não em  $L$  (cf. Hopcroft et alli, 2002, Sec. 1.5.4).

### 3.4) Gramática

Para a teoria da computação, uma *gramática* é um modelo ou uma técnica utilizada para gerar uma *linguagem*. Em particular, estuda-se a gramática livre de contexto (*context-free grammars*), também chamada de gramática de tipo 2, considerando-se a hierarquia de Chomsky (cf. Chomsky, 1959), responsável por gerar uma classe de linguagens chamada de linguagem livre de contexto.

De acordo com Sipser (2006, p.102), as gramáticas livres de contexto foram inicialmente utilizadas no estudo de linguagens humanas como uma forma de compreender o relacionamento entre nomes, verbos e preposições, bem como seus correspondentes sintagmáticos.

Uma gramática livre de contexto é formalmente definida (cf. Sipser, 2006, p.102) como

uma 4-tupla<sup>3</sup>  $(V, T, R, S)$  onde:

- a)  $V$  é um conjunto finito chamado de *as variáveis*;
- b)  $T$  é um conjunto finito, disjunto de  $V$ , chamado de *os terminais*;
- c)  $R$  é um conjunto finito de *regras* (também chamadas de *produções*), onde cada regra consiste em uma variável e uma cadeia de variáveis e terminais;
- d)  $S$  é a variável inicial, ou símbolo de início.

Um exemplo de gramática livre de contexto é (cf. Sipser, 2006, p.103):

$G = (\{S\}, \{a, b\}, R, S)$ , onde o conjunto de regras  $R$  é definido como:

$$S \rightarrow aSb$$

$$S \rightarrow SS$$

$$S \rightarrow \varepsilon.$$

Esta gramática gera cadeias como “ab”, “aabb”, “aababb”. Uma forma de entender a linguagem gerada por esta gramática é entender “a” como um parêntese esquerdo “(“ e “b” como um parêntese direito “)”. Vista dessa forma,  $L(G)$  é a linguagem de todas as cadeias de parênteses adequadamente aninhados.

Um outro conceito, relacionado ao conceito de gramáticas, são os reconhecedores gramaticais, que são máquinas formais capazes de reconhecer se uma dada cadeia  $w$  pertence ou não a uma linguagem  $L$ . Por exemplo, um reconhecedor para uma linguagem livre de contexto é um autômato de pilha. Não prolongaremos tal discussão a respeito de reconhecedores de linguagens por se tratar de um aspecto mais computacional e fugir do escopo deste texto, mais informações sobre o tema podem ser encontradas nos livros de Hopcroft et alli (2002) e de Sipser (2006).

Uma importante aplicação de gramáticas livres de contexto dentro da ciência da computação é a especificação e compilação (tradução de um programa escrito em uma linguagem de programação para uma outra linguagem, chamada de linguagem objeto) de linguagens de programação, embora pesquisadores da área de inteligência artificial (e da área de linguística computacional) também utilizem gramáticas livres de contexto, bem como gramáticas sensíveis ao contexto, para modelar a linguagem natural a partir de uma perspectiva computacional. Na construção de compiladores, muitas vezes, é necessário extrair aspectos semânticos do programa para, por exemplo, poder realizar a tradução para uma linguagem de máquina, que de fato dirá ao computador como realizar uma determinada tarefa.

Knuth em *Semantics of Context-Free Languages* (1968) examina uma forma de se fazer isso através do uso de atributos associados às produções de uma gramática livre de contexto, construindo uma gramática chamada de *gramática de atributos*. Apesar de existirem outras técnicas formais para capturar informação semântica de um programa, a gramática de atributos tem a vantagem de ser independente de qualquer forma de análise sintática.

Assim, apesar de gramáticas serem vistas apenas como uma forma de gerar uma linguagem, existem extensões formais capazes de agregar algum conhecimento semântico às

<sup>3</sup> Isto é, uma lista finita de quatro elementos em uma ordem previamente estabelecida (cf. Spiser, 2006, p.06).

mesmas.

### 3.5) Reconhecedores Gramaticais

Quando se deseja identificar se uma determinada cadeia  $w$  faz parte de uma linguagem  $L$ , os cientistas da computação utilizam ferramentas formais conhecidas como *reconhecedores*. Um reconhecedor pode ser visto como um programa que recebe uma gramática  $G$  e uma cadeia  $w$  e verifica se  $w$  pode ser gerada por  $G$ . Resumidamente podemos estabelecer a seguinte relação entre reconhecedores e tipos de gramáticas:

Gramática	Reconhecedor
Com estrutura de frase	Máquina de Turing
Sensíveis ao contexto	Máquina de Turing com memória limitada
Livres de contexto	Autômatos a pilha
Regulares	Autômatos finitos

Tabela 1

*Grosso modo*, o que fazem os reconhecedores é identificar se determinada cadeia é gramatical.

## 4) Comparação

Considerando, então, a teoria da computação, é possível notar que o termo *linguagem* é anterior à *gramática*, isto é, o interesse do pesquisador recai sobre a linguagem que se pretende examinar e a gramática é a ferramenta usada para isto. Na linguística, nem sempre é assim: se pensarmos em uma linha gerativista da linguística formal, podemos perceber uma certa inversão de prioridades: para os gerativistas, o que se pretende examinar é a própria gramática, e a linguagem, ou ainda a *E-language*, isto é, o epifenômeno da linguagem internalizada, é a ferramenta que temos para se atingir a gramática, inacessível consciente e diretamente tanto ao falante quanto ao linguista.

Ainda em relação aos gerativistas, é interessante notar que para outros linguistas formalistas e para cientistas da computação, uma gramática não é a única forma de gerar uma linguagem e sim uma possibilidade, possivelmente entre muitas outras. Tal diferença é proveniente do objeto de estudo das áreas: adeptos da gramática gerativista estão interessados, sobretudo, em descobrir a gramática universal, o componente biológico e inato da linguagem, e cientistas da computação e outros linguistas formalistas estão interessados em gramáticas possíveis, como proposto por Quine (1972). Para Quine, se duas gramáticas geram a mesma linguagem, elas são equivalentes e não há porque assumir uma e não outra como correta. No

entanto, para gerativistas, haveria razões biológicas para se assumir uma gramática e não outra. Neste sentido, teoria da computação e a parte não-gerativista da linguística se aproximam, cf. Chomsky (1986, p. 20), na medida em que encontrar uma e somente uma gramática para dada linguagem não faz sentido e que qualquer gramática que gere/descreva adequadamente uma linguagem pode ser adotada.

É em relação ao proposto por Partee *et alli* (1990) que as definições da teoria de autômatos mais se assemelham: entendem a *linguagem* como um conjunto finito de objetos e *gramática* como um sistema inferencial com um axioma que gera determinada linguagem. O que é interessante notar é que, ainda que o objeto matemático *gramática* da ciência da computação e de linguistas formalistas seja muito próximo, a função de uma *gramática* dentro do sistema muda a depender da área de investigação: para linguistas, a gramática nos diria que sequências são ou não gramaticais, enquanto na teoria de autômatos, este é papel dos reconhecedores gramaticais. A *gramática per se* na ciência da computação, gera a linguagem, mas se há a necessidade de testar se determinada sequência está ou não dentro desta linguagem, estudiosos da computação utilizam reconhecedores de linguagem<sup>4</sup>. *Grosso modo*, para linguistas a gramática serviria para gerar a linguagem, mas também para avaliá-la, e para cientistas da computação, não.

Considerando agora as definições da linguística geral, de Dubois *et alli* (1999) e Lyons (1982, 1984), é também perceptível a distância existente entre o que a “linguística geral” e a teoria da computação entendem serem gramáticas ou linguagens. Dificilmente, a ciência da computação se interessa pela linguagem natural; inversamente, o mesmo acontece para os linguistas, que dificilmente se interessam por outros objetos que não a linguagem natural. Isto é dizer que os objetos de estudo, as perguntas que os teóricos de diferentes áreas se colocam são de natureza distinta e, obviamente, isto gerará conceitos distintos, ainda que sob o mesmo rótulo. Desta forma, seria difícil achar em um âmbito não-linguista uma preocupação social ou funcional com a linguagem, por exemplo.<sup>5</sup>

Vale comentar também a relação necessária entre linguagem e gramática nos diversos casos que visitamos. Ainda que estejamos tratando de objetos diferentes sob o mesmo rótulo, sempre haverá uma relação necessária entre o que se entende por *gramática* e o que se entende por *linguagem*. Aparentemente, apenas a gramática gerativa<sup>6</sup>, dentro as áreas que visitamos, se interessa primordialmente pelo que chamam de *gramática*, enquanto estudiosos da ciência da computação vêm uma gramática como uma ferramenta para categorizar uma dada linguagem e outros linguistas entendem a gramática como um conjunto de regras que formalizam a linguagem, mas não vêm necessariamente este conjunto de regras como seu objeto em si.

---

<sup>4</sup> Para esta tarefa, é possível também utilizar um analisador gramatical (*parser*), que é um programa capaz de reconhecer a gramaticalidade de uma sequência através de sua análise sintática.

<sup>5</sup> Não estamos afirmando que não existam estudos na área da ciência da computação que considerem a linguagem natural e tampouco assumindo que todo linguista trabalha com linguagem natural, mas sim, assumindo que tais estudos são encontrados mais raramente entre as duas áreas.

<sup>6</sup> Pretendemos, em trabalho futuro, aprofundar a discussão sobre o objeto de estudo dos linguistas formalistas não-chomskyanos. Ao nosso ver, tais estudiosos têm uma postura que se afasta da visão chomskyana de *gramática* e *linguagem* e se aproxima das outras áreas que visitamos. No entanto, para que esta afirmação seja realmente contundente, precisaríamos de uma análise mais cuidadosa.

## 5) Conclusão

Apresentamos, ao longo destas páginas, diferentes definições dos conceitos *gramática* e *linguagem* para a ciência da computação e para distintas áreas da linguística, tendo a linguística formal, chomskyana e não-chomskyana, como foco. Utilizamos, para isto, referências bibliográficas de grande relevância para as áreas e esmiuçamos as definições apresentadas por cada um dos autores. Por fim, comparamos pontos de contato entre os conceitos utilizados pelas áreas – como, por exemplo, a relação intrínseca existente entre o que se entende por *gramática* e o que se entende por *linguagem* – e ressaltamos algumas diferenças entre as áreas: para que serve uma gramática, a possibilidade de termos mais de uma gramática gerando a mesma linguagem, a precedência teórica do conceito *gramática* sobre *linguagem* ou vice-versa.

Considerando todos os pontos citados na seção acima, parece-nos possível concluir não só que cada área reinventou os conceitos *gramática* e *linguagem* à medida que os incorporou, mas também que estas reinvenções são de caráter teórico, isto é, aconteceram considerando os objetos teóricos de cada uma das áreas. A ciência da computação, por exemplo, apropriou-se destes conceitos surgidos na linguística, mas os reformulou de maneira que a preocupação com a linguagem humana perdesse a relevância e que fosse a formalização matemática o carro chefe destes conceitos. Já a gramática gerativa reformulou estes conceitos, adaptando-os às suas concepções biológicas e inatistas, o que gerou objetos sob o rótulo *gramática* e *linguagem* bastante diferentes do que vemos na linguística como um todo.

Como dissemos, este trabalho é uma pequena parte de um estudo maior que tem como objetivo examinar semelhanças, diferenças e fronteiras entre os estudos formais na linguística e os estudos linguísticos dentro da ciência da computação. Partimos destes conceitos por entendermos que são muito básicos a ambas as áreas e já pudemos perceber que, quando há diferenças – e quase sempre há – estas diferenças são fruto do fato de que as duas áreas têm objetos de estudos diferentes, que necessitam olhares e recortes de natureza distinta, o que força determinada nomenclatura nomear conceitos similares porém distintos.

## Referências

- PARTEE, B.; ter MEULEN, A.; WALL, R. *Mathematical Methods in Linguistics*. Kluwer, 1990.
- CHOMSKY, N. *Knowledge of Language*. Praeger, 1986.
- CHOMSKY, N. On certain formal properties of grammars in *Information and Control*, n.2, v. 2, 1959.
- DUBOIS, J. et alli. *Dicionário de Linguística*. Cultrix, 1999.
- HOPCROFT, J.; ULLMAN, J.; MOTWANI, R. *Introdução à Teoria de Autômatos, Linguagens e Computação*. Rio de Janeiro: Campus, 2002.
- KNUTH, D. Semantics of Context-Free Languages in *Mathematical Systems Theory*, 1968, v. 2, n. 2, p.127-145.

LYONS, J. *Lingua(gem) e Linguística*. Zahar, 1982 .

LYONS, J. *Semántica Lingüística: Una Introducción*. Cambridge, 1997. (1ed. 1984)

QUINE, W. *Methodological reflections on current linguistic theory*. Reidel, 1972.

SIPSER, M. *Introduction to the Theory of Computation*. Thomsom Course Technology, 2006.