

## LINGÜÍSTICA E CIÊNCIA

JOHN W. MARTIN \*

É comum as pessoas formadas em Humanidades aceitarem como artigo de fé a noção de que as "Ciências Exatas" podem englobar, dentro dum sistema coerente de princípios explicativos gerais, determinadas classes de fenômenos "objetivos" de nosso universo, quais sejam os estudados na física, na química, na matemática, e em outras áreas que tais. Mas é incomum, talvez por ser relativamente recente, a noção de que possam ser assim tratados os fenômenos da linguagem humana.<sup>1</sup> Ao contrário, estes são tidos como exclusiva-

\* O autor é natural de Winnipeg, Canadá. Doutorou-se pela Universidade de Washington (USA) em 1956, passando imediatamente depois a lecionar em diversas universidades desse país. Em 1960, trasladou-se à América Latina, onde desempenhou o magistério em Universidades do Equador, da Colômbia, do Chile, e do Brasil. Em Bogotá, foi o diretor fundador (1960-1965) do Instituto Lingüístico Colombo-Americano do Ministério de Educación Nacional (em convênio com a Universidade de Califórnia em Los Angeles) e diretor (1965-1968) do Centro de Estudios Universitarios Colombo-Americano (em convênio com onze universidades norte-americanas). No Brasil, foi professor encarregado do Programa de Pós-Graduação em Lingüística da Pontifícia Universidade Católica de São Paulo (1969-1970); especialista da Fundação Ford no Programa Unificado de Pós-Graduação em Lingüística do Museu Nacional e da Universidade Federal do Rio de Janeiro (1970-1972), e especialista para a implantação de cursos de pós-graduação no Departamento de Letras da Pontifícia Universidade Católica do Rio de Janeiro (1972), e coordenador do Programa de Pós-Graduação em Lingüística da Pontifícia Universidade Católica de Campinas, onde ocupou a presidência da Comissão de Pós-Graduação dessa entidade. Entre suas publicações se incluem: "Some Uses of the Old Spanish Post Subjunctive (with reference to the authorship of *La Celestina*)" *Romanic Philology*, XIII, 1, 1958; "Remarks on the Origin of the Portuguese Inflected Infinitive", *Word*, XVI, 3, 1960; *The Grammatical Systems of English and Spanish* (com R. J. Stockwell and J. D. Bowen), University of Chicago Press, 1965; "Observations on the Linguistic Design of Materials for the Teaching of English to Speakers of Spanish", *Actas del Segundo Simposio del Programa Interamericano de Lingüística y Enseñanza de Idiomas*, Instituto Caro y Cuervo, Bogotá, 1967; "Tense, Mood, and the Inflected Infinitive in Portuguese" in *Readings in Portuguese Linguistics* (Jürgen Schmidt-Redefeldt, org.), North-Holland Publ. Co., no prelo. Atualmente, como Professor Visitante, leciona na Universidade Federal do Paraná.

1 Por exemplo, Maurice Leroy, em seu livro *As Grandes Correntes da Lingüística Moderna* (Presses Universitaires de Bruxelles, 1967, 3.<sup>a</sup> ed. brasileira, Editora Cultrix, São Paulo, 1974) diz (p. 110): "O Language de J. Whatmough, livro vivo e original, que mistura, curiosamente, reminiscências humanísticas a uma disfarçada admiração pelas conquistas da técnica científica, como

mente "convencionais", por serem determinados e consagrados pelo "uso", ou pelas "normas de bem dizer", ou por outros fatores de igual vagueza.

Embora seja verdade que muitos fatos das diferentes línguas são, realmente, determinados por convenções, e apesar de nossa tradição gramatical obscurecer, ou ignorar de todo, a natureza de muitos outros, seria errôneo concluirmos que os fenômenos da linguagem estejam fora do alcance da ciência e, portanto, isentos de investigação racional. Na física, o fato de não sabermos, por exemplo, por que a energia toma às vezes a forma que nós percebemos como luz, ou por que a luz tem a velocidade que tem, não nos faz renunciar à investigação do significado e das conseqüências destes fenômenos. Da mesma maneira, o fato de não sabermos por que em português, a palavra 'mesa' seja *mesa*, ou por que *Eles não se dão bem* tem o significado que tem, não nos condena a uma mera catalogação dos dados superficiais dessa língua; esses dados, e outros semelhantes, apesar de ter muito de convencionais, encerram fatos que são passíveis de investigação, e esta investigação constitui a ciência lingüística. Assim, mediante o estudo dos sistemas subjacentes da língua portuguesa, podemos explicar por que não existe, nem poderia existir, nenhuma palavra da forma  $\neq$ smea,<sup>2</sup> composta dos mesmos segmentos fonológicos que *mesa*, e por que  $\neq$ *Eles dão se não bem* é necessariamente (e não só convencionalmente) uma concatenação mal formada. Isto é, ambos esses exemplos fogem a regras que são passíveis de investigação a um nível muito mais profundo e revelador do que é considerado decente em determinados círculos.<sup>3</sup>

os computadores eletrônicos,<sup>3</sup> quer mostrar, ao que parece, que a Lingüística deve ser alinhada entre as ciências da natureza; e não vemos também B. Mandelbrot, ao concluir um artigo-programa, muito bem feito aliás, acerca da possibilidade de aplicar os conceitos da teoria da comunicação ao estudo da linguagem, manifestar a esperança de ver aproximar-se o dia "em que o estudo da língua passará definitivamente a barreira fictícia que a separou, até aqui, das ciências exatas": o velho mito schleicheriano não está morto! A nota 3 do trecho citado se refere a Joshua Whatmough, *Language, A Modern Synthesis*, Londres, 1958. A nota 4 do mesmo trecho é "Benoît Mandelbrot, 'Structure formelle des textes e communication' em *Word*, t. X (1954), pp. 1-27. — E esses modos de ver fizeram escola: "Linguistics, at least in several aspects, is evolving from a humanity into a science. There is a growing recognition of languages as natural phenomena whose secrets may yield to methods of analysis and description akin to those that have proved fruitful in the physical sciences" — começo do artigo *Linguistics and Mathematics*, de Anthony G. Oettinger em *Studies Presented to Joshua Whatmough* (Haia, 1957), p. 179". Acho curioso que Oettinger, também, considere, ao que parece, que haja uma dicotomia "humanity/science", como se aquilo que é racional tivesse que ser separado daquilo que constitui "as humanidades".

2 É convencional na literatura lingüística marcar com asterisco (aqui,  $\neq$ ) concatenações mal formadas.

3 Em 1973, O Estado de São Paulo publicou uma reportagem que lembra o espanto de Leroy (nota 1): "José Gonzalo Herculano de Carvalho vai proferir no Brasil uma série de palestras sobre um novo ramo da Lingüística, que está muito em moda no mundo

As "regras" a que me referi no fim do parágrafo anterior não são regras impostas pelo uso, nem pela autoridade dos "bons autores" ou dos gramáticos escolares. Ao contrário, elas são expressões dos princípios que **subjazem** aos "diversos aspectos que se oferecem à observação direta."<sup>4</sup> São elas que nos permitem ir além da mera contemplação vácuca da superfície, para chegar a uma compreensão daquilo que os dados **significam** em termos da organização da mente humana. Elas são, enfim, as regras que são descortinadas por uma ciência que é, por excelência, "humanística".

As outras regras — aquelas que encontramos nas gramáticas normativas das escolas — são de outra índole. Elas visam, essencialmente, a **controlar** (e não a **explicar**) nossa conduta lingüística. Antes que explicativas, elas são legislativas, e vêm vinculadas a sanções determinadas por pessoas, e não pela natureza. Quando desacatamos tais regras, nos expomos à desaprovação dos defensores dum certo tipo e estilo de linguagem, mas nem por isso nossa ignorância (ou rebeldia) face a essas normas quebrará, necessariamente, os moldes do **sistema** lingüístico.

São dois, então, os tipos de regras que entram em jogo quando falamos de gramática: as que são inventadas por pessoas, e que têm o propósito de **regular** a conduta de outras pessoas, e as que **constituem** (e **definem**) a linguagem.<sup>5</sup> Quando fugimos a estas últimas, deixamos de produzir linguagem: o que produzimos nem é "certo" nem "errado": não é nada. Assim, **¶Nós vou**, embora constituído de palavras bem formadas em português, quebra certas regras constitutivas dessa língua ao nível da oração, e o resultado glo-

-----  
inteiro: 'Gramáticas Formais ou Gerativas'. Para quem não sabe o que se esconde por trás desses nomes pomposos, o assunto parece inofensivo. Entretanto, José Gonçalo denuncia esse novo campo de pesquisa lingüística como uma ameaça para a humanidade, se não for bem utilizado."

4 É comum pessoas sem preparo científico se imaginarem que "a ciência" se restringe aos "fatos" (leia-se dados) diretamente observáveis. Assim, Leroy (p. 107): "Chega-se ao ponto de pensar e de especular não mais acerca dos fatos e dos diversos aspectos que /se/ oferecem à observação direta, mas antes acerca das deduções tiradas de sua existência, das esquemas que se julgou poder estabelecer segundo sua relação recíproca no sistema. Em outras palavras, a lingüística, concebida de tal maneira, afasta-se singularmente desse fato humano por essência, vivo e multiforme, que é a linguagem, e se arrisca a tornar-se um conceptualismo formalista, acantado no campo da especulação intelectual; não foi sem razão que se pôde censurar aos estruturalistas pairarem amlúde na estratosfera sem se preocupar com as realidades concretas, que são a própria substância de nossa ciência".<sup>3</sup> A nota 3 deste trecho se refere a J. Orr em *Actes du Sixième Congrès International des Linguistes* (Paris, 1949) pp. LXIII-LXV. O presente ensaio foi inspirado pelo desejo (não muito vivo) de responder a tais confusões a respeito da natureza da atividade intelectual em geral, e da ciência e a lingüística em particular.

5 John R. Searle discute a distinção entre estes dois tipos de regra às pp. 161 e 162 de seu trabalho "Teoria da Comunicação Humana e a Filosofia da Linguagem: Alguns Comentários" (pp. 149-166 da antologia *Teoria da Comunicação Humana*, org. por Frank E. X. Dancos, Holt, Rinehart and Winston, Inc., 1967 (1.<sup>a</sup> ed. brasileira, Editora Cultrix, São Paulo, 1973).

bal não pertence a língua nenhuma. Aqui caberia, então, a admoestação "Isso não é português". Por outro lado, **Nós vai** foge somente a certas determinações reguladoras da conduta lingüística,<sup>6</sup> mas o resultado é uma oração da língua portuguesa, e quem advertisse a respeito dela "Isso não é português" simplesmente revelaria uma ignorância (ou uma cegueira) a respeito da língua falada por muitos (talvez pela maioria) dos brasileiros.

Poder-se-ia pensar que existisse algum método para isolar e identificar as regras constitutivas da linguagem e das diversas línguas que a manifestam. Com efeito, os lingüistas que trabalham com línguas indígenas desenvolveram (e continuam a refinar) métodos que facilitam o registro dos dados fornecidos pelos informantes. Mas no momento de organizar esses dados, ou de tentar aprender a língua que eles representam, o lingüista é forçado a abandonar os métodos ou técnicas de registro e criar, nem que seja somente duma maneira implícita, uma teoria que justifique a organização que ele lhes dá. E é nestas últimas atividades, quando explicitada a teoria regente, que a ciência entra em jogo. Isto é, "organizar" significa, neste contexto, desvendar os princípios sistêmicos que subjazem aos dados, ou seja, **descobrir** as regras constitutivas do sistema a que os dados pertencem. Para isto, não existem métodos.

A verdade é que não sabemos como se produzem aqueles achados que iluminam a consciência com uma luz repentina. Ignoramos de que modo se dá aquela "chance de topar, sem busca, pessoas, coisas e informações urgentemente necessárias" que João Guimarães Rosa comentou, e cujo nome ele achou arrolado em **Brewer's Dictionary of Phrase & Fable**:

SERENDIPITY: "Feliz neologismo cunhado por Horace Walpole para designar a faculdade de fazer por acosa afortunadas e inesperadas "descobertas". Numa carta a Mann (28 de janeiro de 1754) ele diz tê-lo tirado do título de um conto de fadas, **Os Três Príncipes de Serendip**, que "estavam sempre oabrando achados, por acidente ou sagacidade, de coisas que não procuravam".<sup>7</sup>

Não há, simplesmente, "métodos" para saber qual é o sistema que subjaz a determinado conjunto de dados; o que há é sagacidade e sorte. Métodos pode haver para arrumar os dados de modo

6 Ver J. Martin: "Concordância" em *Revista Brasileira de Lingüística*, II (1975), No. 1, no prelo, onde se demonstra que tais "aberrações" são produzidas por diferenças sistemáticas na aplicação das regras constitutivas do português.

7 Tutaméia (*Terceiras Estórias*), Livraria José Olympio, Editora, Rio de Janeiro, 1969, p. 157 (no "Prefácio: Sobre a Escóva e a Dúvida").

a deixar funcionar melhor essa combinação de fatores, mas é raro que um deles conduza a um "heureka", hajam vista os séculos em que os dados das línguas têm sido arrumados de diferentes maneiras pelos gramáticos, sem que ninguém tivesse topado com uma explicação do fato de a linguagem "fazer de meios finitos uso infinito"<sup>8</sup> e assim produzir uma infinidade de orações nunca antes proferidas, ou de uma criança, muito antes de saber somar seis e três, já dominar as alavancas principais da máquina de fabricar orações.

Podemos ilustrar o problema de achar os fatos que subjazem aos dados lingüísticos tomando como exemplo um grão de sal. O que aí se oferece à observação direta é uma substância cristalina. Nada há que sugira que essa substância tem uma estrutura subjacente que consiste em cloro e sódio; nada há que indique a existência de regras constitutivas da química, de acordo com as quais estes elementos se combinam numa proporção fixa para formar moléculas de cloreto de sódio que, por sua vez, se arranjam em cristais. Muito menos teríamos motivo para pensar, com base na observação direta, que o cloro e o sódio têm uma estrutura subjacente em que os formativos são cargas elétricas, estas também organizadas, de acordo com regras constitutivas da física, em átomos. Outra seria, deveras, a natureza do universo, e da história da ciência, se pudéssemos apanhar os segredos da natureza na superfície daquilo que podemos observar diretamente. As ciências com patente consagrada de "exatas" consistem, justamente no sucesso com que têm sido coroados os esforços do homem por desvendar esses segredos mediante uma penetração da superfície. E nada sabemos acerca da maneira pela qual conseguimos essa penetração.

A situação não é essencialmente diferente na lingüística, pois tanto a mente humana quanto a linguagem pela qual ela se manifesta são fenômenos naturais. E aqui também os dados que se nos oferecem "in natura" ocultam as regras constitutivas. Não é de estranhar, pois, que quem insista em trabalhar exclusivamente com os dados, sem procurar os fatos que a estes subjazem, tenha que se contentar com conhecimentos primitivos e errôneos.

Segue-se do que foi dito até aqui que existem gramáticas compostas das regras constitutivas da língua portuguesa e de todas as outras línguas e, portanto, da própria capacidade lingüística do homem. A descoberta das gramáticas que subjazem aos dados que temos à disposição é a tarefa central da lingüística de hoje, e é no

---

8 Wilhelm von Humboldt: *Über die verschiedenheit des Menschlichen Sprachbaues*, Berlin, 1836 (apud Noam Chomsky: *Aspects of the Theory of Syntax*, The M.I.T. Press, Cambridge, Massachusetts, 1965, p. v.).

desempenho dessa tarefa que esta ciência mostra mais nitidamente o que ela tem de "exata". Cabe voltarmos agora, pois, para um pequeno esboço da natureza dessas gramáticas.

As regras constitutivas das línguas humanas são, basicamente, sintáticas. Isto é, elas definem a maneira pela qual os formativos se combinam na constituição de entidades lingüísticas complexas. **Sensu lato**, pois, a sintaxe duma língua é o sistema combinatório de suas unidades simples e compostas. Na área da fonologia, as unidades simples são **traços fonéticos**, tais como "labialidade", "sonoridade", "oclusividade", "oralidade", "fricatividade", e outros.<sup>9</sup> A totalidade das regras que determinam a combinação destes traços define o inventário de **sons** que funcionam numa língua. Estes sons, por sua vez, são combinados de acordo com outras regras que definem os **morfemas**, ou seja, os itens lexicais. Exemplificando, há uma regra em português que combina os primeiros quatro traços mencionados acima, e cujo resultado é o primeiro som de **bolo**. Substituindo "sonoridade" por "surdez" nessa combinação, obtém-se o primeiro som de **pô-lo**. E substituindo "oclusividade" por "fricatividade" nesta última combinação, obtém-se o primeiro som de **folo**. O fato de não existir tal palavra em português é irrelevante ao sistema fonológico; ela poderia existir, pois as regras constitutivas da fonologia não a impedem. Por outro lado, não poderia existir uma palavra portuguesa que começasse com um som composto dos traços "labialidade", "nasalidade", e "surdez", pois as regras constitutivas dos sons desta língua não permitem que nasalidade e surdez se combinem para formar as entidades que compõem os morfemas. Tampouco poderia haver em português uma palavra  $\neq$ sflo. Isto é, qualquer falante de português pode inventar uma palavra **folo**, apontada acima, e esta será reconhecida por todos os outros falantes da mesma língua como sendo fonologicamente aceitável, embora seja provável que estes perguntem "Que é que essa palavra quer dizer?" Mas nenhum falante de português pode inventar uma palavra  $\neq$ sflo e conseguir que os outros a aceitem. Ao contrário de indagarem pelo significado dessa concatenação de sons, é mais provável que perguntem, "Que língua é essa?" ou "Você está-se resfriando?" As regras constitutivas dos morfemas do português simplesmente impedem a ge-

9 Uma boa introdução à fonologia, dum ponto de vista semelhante ao esboçado aqui, se encontra em Ronald W. Langacker: *A Linguagem e sua Estrutura: alguns conceitos lingüísticos fundamentais*, Harcourt Brace Jovanovich, Inc., 1967, 1968 (1.<sup>a</sup> ed. brasileira, Editora Vozes Limitada, Petrópolis, RJ, 1972), Capítulo 6: *Sistemas Fonológicos*. Ver, também: Suzette Haden Elgin, *Que é Lingüística?* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1973 (1.<sup>a</sup> ed. brasileira, Zahar Editores, Rio de Janeiro, 1974), pp. 15-27. A exposição mais detalhada até agora do que eu chamo aqui de "sintaxe fonológico" encontra-se em Noam Chomsky e Morris Halle: *The Sound Pattern of English*, Harper & Row, Publishers, New York, Evanston and London, 1968.

ração de tal combinação de sons, embora cada um deles pertença ao inventário dessa língua.

De maneira semelhante, e *stricto sensu*, a sintaxe determina as combinações dos "itens lexicais" duma língua.<sup>10</sup> São as regras da sintaxe entendida neste sentido que permitem a geração de **Eles não se dão bem** e que inibem a geração de ~~¶~~**Eles dão não se bem**. Estas mesmas regras permitem a geração de **Eu me cortei com esta faca**, e **Eu me ofendi com esta carta**, e são de primordial importância na explicação de como é que a primeira destas orações significa que alguém me cortou, e que esse "alguém" sou eu próprio, enquanto que a segunda, superficialmente paralela, significa que alguém me ofendeu e que esse "alguém" é outrem. Também são as regras sintáticas, neste sentido estrito, que nos permitem entender como perfeitamente normal o seguinte diálogo, o qual seria esquisito demais se tomado ao pé da letra:

- Oi, Márcia! Puxa, você está diferente!
- Oi, Lúcia! Pois é: cortei o cabelo.
- Estou vendo, e você ficou linda! Quem foi que cortou?
- O Alphonse.
- Como ele trabalha bem! Acho que vou cortar o meu com ele.

Quando digo que esse diálogo seria esquisito se tomado ao pé da letra, quero dizer que seria contraditório Márcia afirmar (com o verbo **cortei**) que foi ela que cortou seu cabelo e depois informar que quem o cortou foi Alphonse, assim como seria um contra-senso Lúcia afirmar literalmente que ela vai cortar o dela com Alphonse e não, digamos, com uma tesoura, ou coisa que o valha. Seria errôneo pensar que tais exemplos se devam ao descuido ou descaso do falante. São determinadas regras constitutivas da sintaxe que, sistematicamente, nos permitem, sob certas condições, dizer X quando o que queremos dizer é Y. Neste caso específico, pode ser demonstrada a existência duma regra que suprime o agente de **cortar** (ou o desloca, como na última linha do diálogo) e coloca em seu lu-

---

10 Por "item lexical" entende-se um formativo que veicula parte do significado duma oração. Portanto, como veremos depois, a noção "passado", por exemplo, é um "item lexical", ou morfema, do português, e estaria independentemente arrolado no dicionário que está implícito em toda teoria duma língua. Por outro lado, como também veremos mais adiante, a partícula *a* de *Maria começava a cair* não veicula parte do significado dessa oração e, portanto, não figura independentemente (pelo menos nessa função) no dicionário que está implícito na teoria da língua portuguesa. Ao contrário, essa partícula, nessa função, figuraria no dicionário como vinculada ao verbo *começar*.

gar a pessoa afetada. É isto que ocorre na acepção mais "natural" de **Quebrei uma perna**, onde pode ter sido outro que quebrou minha perna, mas ele não é identificado, e em seu lugar consto eu. É um corolário dessa regra sintática que explica, também, como entendemos neste último exemplo que a perna que "eu quebrei" é minha, e não uma perna alheia. Com efeito, a força dessas regras ocultas faz-se sentir vivamente quando tentamos interpretar **Quebrei uma perna** como se tivesse a mesma estrutura subjacente que **Matei um cachorro**. É válido e natural reagirmos a este último exemplo perguntando "Por que?" e "De quem?" mas essas perguntas não calham no caso do primeiro exemplo; produziriam surpresa, ou irritação, pois é óbvio a qualquer falante de português que o agente de **quebrei** não é, nesse exemplo, **eu**, assim como é evidente que a perna quebrada pertence ao sujeito do verbo. E se tudo isso é evidente, é por causa de todos os falantes de português serem detentores das mesmas regras que aí funcionam.

Tais deslocamentos estão no próprio âmago da língua portuguesa. Com frequência, é por causa dum deslocamento que um só conteúdo pode aparecer sob diferentes formas: **Fazer isso lhe é difícil, Isso lhes é difícil de fazer; Esta sala está cheia de fumaça. Está cheio de fumaça nesta sala; Fazer isso é nosso dever, Fazemos isso é um dever, Devemos fazer isso**. Estes exemplos estão colocados em ordem ascendente de complexidade aparente: no primeiro par, nenhum dos formativos parece ter mudado de forma (embora **lhes** tenha uma forma subjacente **eles**); no segundo par, vemos que o deslocamento que se percebe na segunda oração teve lugar antes da concordância, de modo que o adjetivo aparece sob a forma **cheio**; nos últimos três exemplos, a "primeira pessoa do plural" se apresenta em três lugares diferentes, e sob duas formas.

Esses casos de sinonímia se devem a diferentes arrumações duma única estrutura subjacente, que representa o significado comum. Mas as regras sintáticas são responsáveis, também, pela ambigüidade estrutural que vemos em **X está quente demais para comer**, onde uma só forma representa duas estruturas subjacentes. Nestes últimos casos, são nossos conhecimentos extralingüísticos que nos inclinam a favor duma determinada interpretação deste exemplo quando X é preenchido por, digamos, **esta criança**, e em favor de outra interpretação quando X é preenchido por **este bolo**, ou **este feijão**. Este é normalmente o caso quando temos que interpretar orações ambíguas: avaliamos o contexto em que são proferidas, e lhes aplicamos os conhecimentos que porventura tenhamos do mundo para tirar uma conclusão probabilística a respeito do significado que po-



dem ter. Mas tais conhecimentos extralingüísticos não têm vez no fato de reconhecermos que em **Fazer isso nos é fácil** o sujeito de **fazer** é **nós**, apesar de esse verbo aparecer na superfície sem sujeito. Poder-se-ia objetar que é o “bom senso” que nos leva a essa conclusão. Mas se fosse, porque seria que ele não nos leva à mesma conclusão em **Fazer isso nos é perigoso**, onde bem poderia ser outro o sujeito de **fazer**? O fato lingüístico é que sentimos uma diferença na estrutura interna dessas duas orações justamente porque as regras sintáticas, por um princípio geral, tratam **fácil** e **perigoso** de maneiras diferentes. É **nisso** que consiste o “bom senso” que nos guia na interpretação dessas duas orações.

As mesmas regras que estamos considerando aqui permitem a geração de **Esta bola triangular é hexagonal**, que é uma combinação “sintaticamente gramatical”, embora semanticamente anômala. O que é relevante é que qualquer falante do português poderia produzir esta oração (como eu a produzi), e todo falante do português entendê-la-ia, reconhecendo-a como bem formada ao nível da sintaxe dos morfemas e multiplamente contraditória ao nível da semântica. Com efeito, deve existir uma “sintaxe semântica”, também. Esta não impede a geração de orações como a última citada, mas as caracteriza como sendo semanticamente mal formadas. Com isto, quero dizer que a sintaxe dos itens lexicais, de que tratarei mais extensamente nos parágrafos que se sucedem abaixo, sobrepuja a qualquer sintaxe semântica que eu possa imaginar.

Ao tentarmos descortinar os princípios que subjazem a todos estes fenômenos, e a todos os outros que podemos perceber nas línguas, fazemos caso omisso das “regras de gramática” que aprendemos na escola. Ao contrário de aplicar aos problemas de análise essas noções de gramática escolar, tentamos descobrir os princípios gerais e específicos que realmente regem a produção de orações. Para isso, insisto de novo, temos que nos confiar à sorte, e à sagacidade.

Convém agora exemplificarmos algumas das regras que constituem (e que, constituindo, definem) a língua portuguesa. Vista somente na superfície, é “óbvio” que uma oração como.

(1) **Maria caíra**

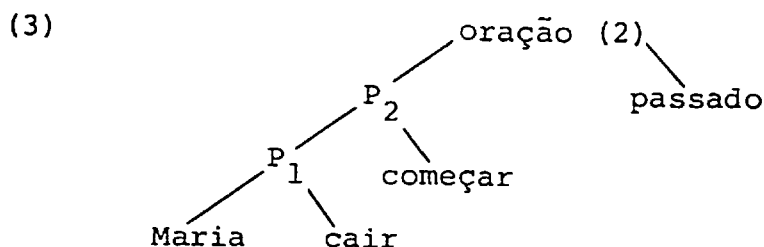
contém um verbo no “mais-que-perfeito simples do modo indicativo”. Da mesma maneira, diz-se que

(2) **Maria começava a cair**

contém um verbo no “imperfeito do modo indicativo”, e que **Maria** é o sujeito desse verbo. E, se gostarmos de taxinomias superficiais,

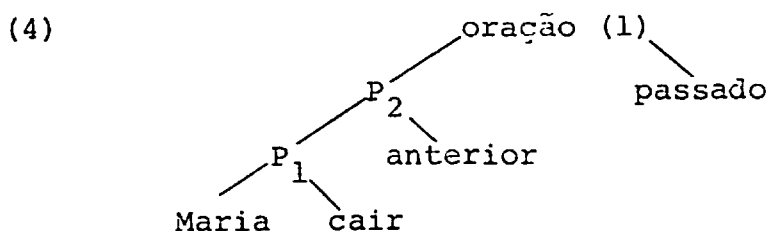
podemos até inventar de dizer que (2) (mas não (1)) contém uma "locução verbal". No espírito da gramática escolar, haveria algumas observações e classificações adicionais que poderíamos fazer, mas ao terminar a "análise", ficaríamos a perguntar, com justiça: "E daí?". A pergunta caberia justamente porque essa maneira de "fazer gramática" não explica, não transcende daquilo que comenta.

Por outro lado, se por uma estranha felicidade nos ocorresse perguntar, a respeito de (2): "Que é que queremos dizer com tal oração? — Que é que aí se afirma?", seria razoavelmente claro que essa oração afirma que **algo** começava, e que esse "algo" consiste, a algum nível, no conteúdo duma proposição da forma /Maria cair/. Isto é, existe um nível em que **começar** "modifica" /Maria cair/. E, a esse nível, o conjunto formado por /Maria cair/ e **começar** é modificado pela noção "passado". Estes fatos (observe-se que não se trata mais de dados superficiais<sup>11</sup>) podem ser expressos graficamente da seguinte maneira:



onde as unidades binárias estão constituídas de tal maneira que o elemento da direita predica ("diz algo") a respeito do que fica ao mesmo nível à esquerda.

De maneira semelhante, podemos dar a (1) a representação gráfica que se vê em



11 Dêde, neste ensaio, é aquilo que podemos apenhar "in natura", e não implica para sua apreensão a existência prévia de teoria explícita ou implícita. Fato, por outro lado, é aquilo que procuramos quando tentamos relacionar de algum modo os dados que a natureza oferece gratuitamente. Se essa distinção não é comum, nem por isso deixa ela de ser útil.

onde o que se afirma é que o conteúdo da proposição /Maria cair/ é anterior<sup>12</sup> e que a proposição //Maria cair/ anterior/ é modificada pela noção "passado".

As representações gráficas (3) e (4), evidentemente, são vinculadas a (2) e (1), respectivamente, por algum processo (ou talvez por alguns processos) de rearrumação, pois em (3) **cair** aparece à esquerda de **começar**, mas em (2) **cair** aparece à direita desse verbo. Curiosamente, entretanto, todos os outros elementos de (3) e (4) estão ordenados da mesma forma que nas correspondentes orações superficiais. Isto parece significar que se aplicarmos um único processo de rearrumação a todos os elementos de (3) e (4) podemos conseguir que **começar** e **cair** se ajeitem nos lugares certos, mas então os outros predicados estarão ordenados erradamente. Isto é, podemos postular a existência duma regra de rearrumação que, procedendo de baixo para cima, desloca um predicado inferior para a direita do predicado imediatamente superior, mas o resultado parece insatisfatório:

(5) Maria passado começar cair

pois **passado** ficou à esquerda de **começar**, enquanto que em (2) esse elemento parece estar à direita desse verbo; além disso, em (5) está faltando a partícula **a**. Mais uma vez, estamos à mercê da sorte. Se acontecer de sabermos que essa partícula, em alguns dialetos falados em Portugal, pode ser **de**, reconheceremos (talvez) ser ela desprovida de significação específica, ou própria. E se nos ocorresse substituir **começar** por **precisar** em (3), poderíamos reconhecer, também, que o aparecimento de **a** (ou **de**) em orações como (2) é uma idiossincrasia do verbo antecedente. Agora, observando que a partícula aparece somente quando **começar** vai seguido de um infinitivo, estamos em condições de concluir, razoavelmente, que ela é gerada no momento de **cair** (ou outro verbo) ser deslocado da maneira que já postulamos.

Mas isto nos deixa com o problema de "passado" estar fora de lugar em (5). Aqui, com sorte (ou sagacidade), pode acontecer de percebermos que esse conceito temporal aparece sempre sob a forma dum afixo, o que abre a possibilidade de existir uma regra<sup>13</sup> da forma

12 A noção veiculada pela palavra "anterior" é uma noção relativa. Isto é, ela representa um valor que precisa de um absoluto para se definir. No caso, "anterior" se define em termos de "passado". A representação (4), então, significa que o conteúdo da proposição /Maria cair/ é anterior com respeito ao passado.

13 Com essa maneira de me expressar, quero pôr em relevo que não "inventamos" regras; descobrimo-las. É evidente que por vezes imaginamos, erroneamente, ter descoberto uma regra, só para ter que riscá-la depois, ou para vê-la desvanecer-se e sumir à medida que a teoria global toma corpo. Ver adiante.

(6)  $Af + \alpha \longrightarrow / \alpha + Af /$

onde Af significa afixo,  $\alpha$  significa qualquer elemento capaz de receber afixo e os colchetes significam que, depois da inversão efetuada pela regra, os dois elementos se aglutinam. Assim, obteríamos (em vez de (5))

(7) Maria /começar-passado/ a cair

onde aparecem todos os elementos de (2), e na ordem certa.<sup>14</sup> Vemos, também, que os processos postulados até aqui dão conta da diferença entre as respectivas formas de (4) e (1): na primeira aplicação de (6), "anterior" se inverte com *cair*, constituindo-se assim um novo domínio de (6), cuja aplicação inverte "passado" e /cair-anterior/, aglutinando-os.<sup>15</sup>

A diferença superficial entre (1) e (2) decorre, então, do fato de o predicado que modifica /Maria cair/ no primeiro destes exemplos estar preenchido por um formativo afixal, o que acarreta uma aplicação adicional de (6). Um pouco de reflexão mostrará que esta observação é suficiente para inutilizar o rótulo "locução verbal" tal como ele é empregado nas gramáticas escolares. Aqui devemos observar, também, que o elemento "anterior" de (4) pode ser realizado com um elemento não afixal, que é *ter*. Este elemento compartilha com *começar* a idiosincrasia de gerar uma partícula quando um verbo inferior é deslocado para sua direita. Neste caso, a partícula é um afixo: *-do* (que tomará diferentes formas de acordo com o elemento com que se aglutine (vide nota 14). Se tivéssemos usado *ter* em vez dum afixo em (4), o processo de deslocamento teria produzido

14 Por "ordem certa", quero dizer que uma forma como *caíra* está constituída de três camadas de significação, em que a raiz é a primeira, a segunda veicula a noção "anterior", e a terceira veicula a noção "passado". Para desfazer uma confusão persistente, quero recorrer aqui à linguagem figurada para explicar que *caíra* (ou *pusera*) se assemelha mais, neste contexto, a uma cebola do que a um trem. Isto é, considero irrelevante, ou acidental, o fato de podermos, às vezes, indicar os segmentos lineares que são acrescentados a uma raiz quando lhe são acrescentados conceitos como "anterior" ou "passado" em forma afixal. Essa segmentabilidade não é um requisito na teoria do português: *pus, punha, tive, entregue* (part. passado), etc. Houve uma época em que o desejo de segmentar as formas linguísticas levou ao dispêndio de grandes energias analíticas (ver Joaquim Mattoso Câmara Jr.: *Estrutura da Língua Portuguesa*, Editora Vozes, Petrópolis, RJ, 1970, especialmente pp. 101-114 (Capítulo XIV, "Os Padrões Especiais dos Verbos em Português"). Com isto, não quero dizer que não seja interessante estudar, como Mattoso estudou, os mecanismos morfofonêmicos que entram em jogo quando afixos são acrescentados a raízes; mas quero recalcar, isto sim, o fato de ser de todo irrelevante insistir em saber onde está, fisicamente (ou fonologicamente), o conceito "anterior" ou o conceito "passado" em *caíra* ou em qualquer outra forma verbal que os contenha. "Ordem certa", então, se refere à ordem de chegada à raiz, e não importa que esta ordem seja perceptível ou não nos segmentos fonológicos.

15 Onde se vê que /cair-anterior/ é implicitamente definido pela teoria como "elemento capaz de receber afixo".

(8) Maria passado ter —do cair

onde (6) seria aplicado duas vezes (a **passado ter** e a **-do cair**), dando

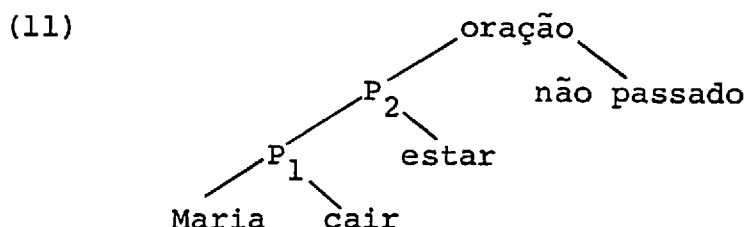
(9) Maria ter-passado cair-do

ou seja,

(10) Maria tinha caído

Ora, (10) e (1) têm exatamente o mesmo significado, e são produzidos por exatamente os mesmos processos (de deslocamento e afixação). Contudo, a gramática escolar reconhece em (10) uma locução verbal, e nega a presença da mesma em (1).

Outra incoerência da gramática escolar que se topa por vezes consiste em atribuir o conjunto **estar caindo** ao paradigma de **cair**. Para ver claramente a natureza dessa incoerência, consideremos as duas superficializações de

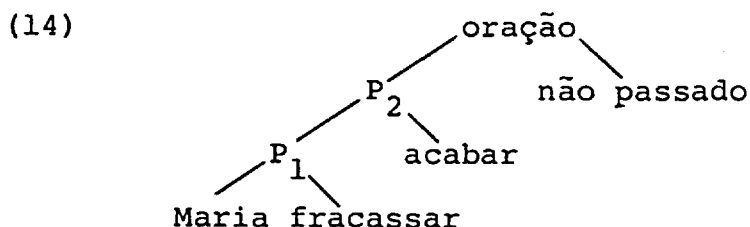


quais sejam,

(12) Maria está caindo

(13) Maria está a cair

Em (11), **estar** modifica a  $P_1$  com um significado que é, aproximadamente, "simultaneidade", ou "progressividade", ou algo semelhante. Quando o processo de deslocamento leva um predicado inferior para a direita de **estar**, este verbo gera uma partícula, a qual pode ser afixal (**-ndo**) ou não (**a**). Mas somente (12), e não (13), é atribuído (às vezes) ao paradigma de **cair**. Por outro lado, nenhum gramático cogitou jamais (que eu saiba) de considerar **acabar fracassando** como parte do paradigma de **fracassar**, mas esse conjunto tem a mesma origem estrutural que **estar fracassando**, a saber



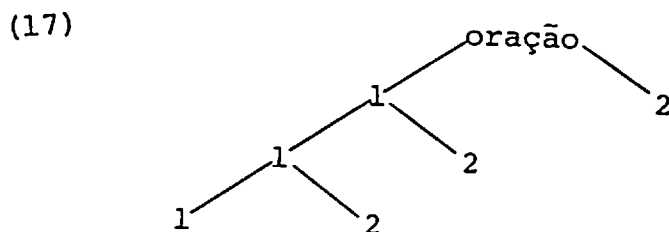
onde **acabar** significa "resultar finalmente" ou coisa que o valha. Observe-se que (14) terá duas superficializações, dependendo da partícula gerada quando do deslocamento de **fracassar** para a direita de **acabar**:

(15) Maria acaba por fracassar

(16) Maria acaba fracassando

Os dois processos esboçados acima — deslocamento para a direita, e a regra (6) — constituem uma subgramática da língua portuguesa que parece adequada para fins de descrição. Isto é ela "dá conta" da relação entre forma e significado que encontramos nas orações que foram analisadas, e parece que ela daria conta também de outras.<sup>16</sup> Mas essa subgramática deixa de explicar **por que** a relação entre forma e significado é aquela que vimos: ela **descreve** adequadamente, mas não **explica**. Isto nos leva a perguntar se formulada assim, essa subgramática é a melhor possível.

Poderia nos ocorrer, diante dessa dúvida, que as orações (1) e (2) têm em comum uma estrutura ainda mais abstrata que as que vimos em (3) e (4):



16 Os processos aqui postulados e reformulados imediatamente embaixo são, com efeito, gerais, e não específicos das orações aqui escolhidas para exemplificá-los. Mas convém alertar o leitor ao fato de que diferentes itens lexicais determinam, caracteristicamente, diferentes explicações das regras. Assim, quando modificamos a proposição /Maria cair/ por poder, e

Esta representação gráfica diz que existe um conjunto de orações que têm uma organização binária. Ela diz, também, que um dos preenchimentos de 2 (o mais alto) tem que ser tal que, quando ele é englobado na estrutura, esta constitua uma oração (observe-se que os 2 inferiores não são suficientes para constituir uma oração: eles são constituintes somente de unidades denominadas "1").

Agora, voltando aos princípios que postulamos antes, podemos reformulá-los da seguinte maneira:

- (18) A língua portuguesa está constituída de tal forma que qualquer elemento rotulado 2 numa estrutura como (17) é levado para a direita de outro elemento rotulado 2 imediatamente em cima. Acabado este processo de deslocamento, aplica-se a regra (6).

Esta nova gramática tem a possível vantagem de não estar presa aos formativos específicos de (1) e (2), mas ela implica, da mesma maneira que a anterior, dois processos para relacionar (1) a (4): um destes desloca um predicado inferior para a direita dum predicado superior, e o outro aglutina afixos no lugar certo. Vistos assim, eles parecem ter pouco em comum. Mas, vistos de outra maneira, esses mesmos processos sugerem a existência duma generalização que, se devidamente captada, resultará numa importante simplificação da gramática.

Voltemos, pois, ao ataque, na esperança de serem erradas as formulações anteriores. Nosso intuito, ao fazer isto, é descobrir móveis da gramática portuguesa que conduzam a uma **explicação**, e não só a uma **descrição**, dos fatos que subjazem às orações superficiais.<sup>17</sup> Vimos acima que o predicado inferior é levado à direita do predicado imediatamente superior, e que o conjunto assim constituído é levado, numa nova aplicação do mesmo princípio, à direita do predicado superior a este. Ao chegarmos ao fim desse processo, aplicávamos, na formulação anterior, a regra (6) quantas vezes havia domínio apropriado. É neste ponto que pode ter havido um erro no raciocínio anterior, pois (6) parece, nessa formulação, dar ao predicado preenchido por "passado" ou por outro elemento afixal um tratamento excepcional. Repensando, observemos agora que ambos os

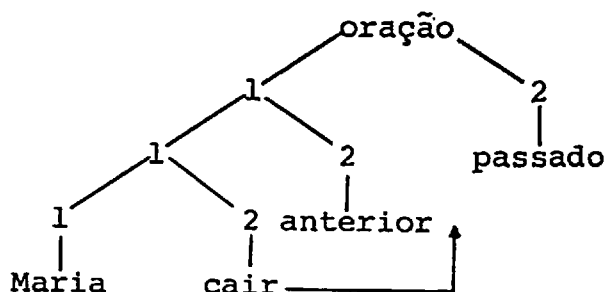
---

este conjunto por "não passado", o resultado é *Maria pode cair*, onde funcionaram os deslocamentos aqui descritos; mas quando modificamos essa mesma proposição por *possível*, vemos que este item lexical, como parte de suas características sintáticas, breca o primeiro deslocamento (de *cair*), e os resultados são *Maria cair é possível*, *Que Maria caia é possível*, *É possível que Maria caia*, *É possível Maria cair*. Ver, também, na discussão anterior, como as regras se relacionam de maneiras diversas com *fácil* e *perigoso*.

17 A discussão clássica de adequação descritiva e adequação explicativa é a de Chomsky, pp. 18-37 et passim em *Aspects of the Theory of Syntax* (ver nota 8, acima).

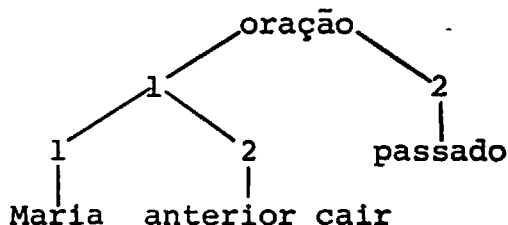
processos em causa implicam inversão de elementos. O primeiro inverte dois elementos em virtude de um destes atrair para sua direita outro inferior que é constituinte duma estrutura à esquerda, e o conjunto assim formado está sujeito à atração por um elemento ainda superior na estrutura global. Isto sugere uma hierarquia: um 2 assimila outro, e depois se submete à assimilação por um terceiro dois superior. Explorando esta noção de hierarquia, podemos postular que os elementos capazes de receber afixos são superiores a estes, de modo que essa relação hierárquica existe mesmo quando um afixo e um elemento capaz de recebê-lo se encontram no mesmo nóculo. Colocando isto em termos mais gráficos, temos uma primeira etapa de (4) em que *cair* é atraído para a direita de "anterior":

(19)



produzindo

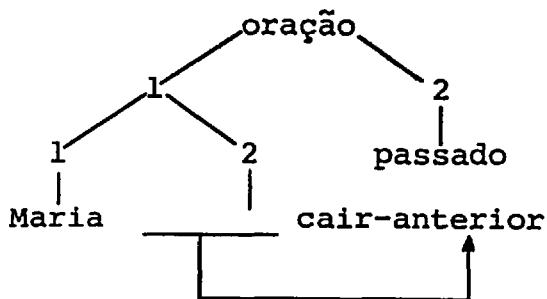
(20)



Aqui, o 2 inferior está constituído por dois elementos, e o hierarquicamente superior deles atrai o outro para sua direita. Já que o elemento assim atraído é um afixo, ele se aglutina ao elemento que o atraiu:

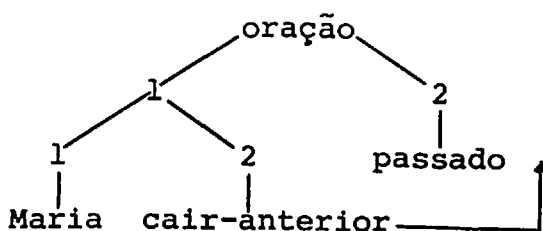


(21)



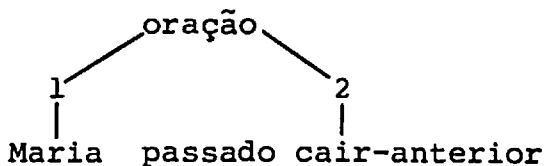
e o conjunto assim formado está sujeito a uma nova atração:

(22)



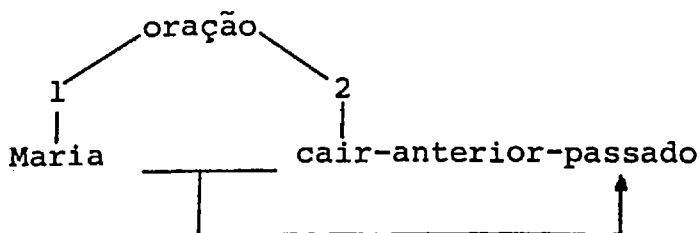
que produz

(23)



onde "passado" é hierarquicamente inferior ao novo vizinho. Portanto, a última etapa do processo é

(24)



que é o mesmo resultado produzido pela formulação anterior, mas que não implica a existência da regra (6). Ou melhor, (6) já está englobado no princípio hierárquico a que sujeitamos a operação do deslocamento.

Com isto, chegamos a nosso assunto final: se esta última formulação fôr certa, até que ponto será ela também mais explicativa? Podemos ser mais confiantes na parte da resposta que concerne ao aspecto imanente da gramática: a nova subgramática é mais simples, pois determina a ordem dos elementos constitutivos de (1) e (2) pela hierarquia dos mesmos frente a um único processo de deslocamento. No que concerne ao aspecto transcendente da gramática, a resposta, no estágio em que a ciência linguística se encontra hoje em dia, tem que ser tentativa, e terá que procurar sua ratificação em pelo menos uma outra ciência que também está ainda a tatear neste terreno: a psicologia. (Com efeito, é comum considerar-se agora que a linguística é um ramo da psicologia, e que os próximos grandes avanços nestas áreas serão feitos conjuntamente). Mas o que é sugerido pela análise anterior é que o cérebro humano está constituído de tal forma que pode converter (subconsciente e automaticamente) em "imagem espelhada" a estrutura predicativa (o conjunto dos elementos aqui rotulados "2") das orações ao mesmo tempo que destrói a estrutura proposicional, deixando como sujeito superficial unicamente um elemento rotulado "1" que não seja uma proposição. Se isto fôr verdade, o "por que" das regras constitutivas que tivemos em mira terá achado sua resposta na própria constituição do cérebro humano, e será a ela que poderemos atribuir diretamente o fato de podermos (e preferirmos) dizer, por exemplo, que Carlinhos precisa lavar a cara, quando o que queremos dizer é formalmente mais semelhante a "Carlinhos lavar a cara precisa". Muitos linguistas, enfim, acham que as regras que constituem as línguas refletem (e são decorrências diretas de) características do cérebro humano.<sup>18</sup> Espera-se que, através duma compreensão mais ampla das regras que constituem a linguagem, se possa chegar a uma compreensão mais ampla daquilo que diferencia o homem dos outros seres dotados de sistema nervoso central. Em outras palavras, a linguística está no cerne dos estudos humanísticos, onde o objeto de estudo é a mais caracterizante das atividades humanas: o pensamento e a produção de orações.

-----  
 18 O relacionamento entre o cérebro e a linguagem é o tema geral dos trabalhos que se publicam na nova revista profissional *Brain and Language*, Academic Press, Nova Iorque, nas áreas de neurologia, linguística, psiquiatria, neurofisiologia, psicologia, neuranatomia, patologias da fala, neurocirurgia, audiolgia,, psicologia fisiológico, e áreas afins.

Ao final deste ensaio (e talvez bem antes), poderia ser objetado que "Tudo isso é muita teoria para umas oraçõezinhas tão pequeninas", ou coisa semelhante. Ao que a resposta mais adequada seria, talvez, "E quanta teoria subjaz a um grão de sal?"

### Resumo

Redefine-se neste ensaio a noção de gramática, e demonstra-se que as regras, ou princípios, que constituem as línguas são tão passíveis de investigação científica quanto as que subjazem a outros fenômenos naturais. Algumas das regras constitutivas da língua portuguesa são especificadas, e é demonstrado que o conjunto de tais regras constitui uma teoria desta língua. Aventura-se a hipótese de que a forma das regras reflete características do cérebro humano.

### Summary

The notion of grammar is redefined in this essay, in which it is demonstrated that the rules, or principles, which constitute languages are as amenable to scientific inquiry as those which underlie other natural phenomena. Some of the rules which constitute the Portuguese language are specified, and it is shown that the set of such rules constitutes a theory of this language. The hypothesis is set forth that the form of the rules reflects characteristics of the human brain.

### REFERÊNCIAS BIBLIOGRÁFICAS

CHOMSKY, Noam. *Aspects of the theory of syntax*. Cambridge, Mass., M.I.T. Press, 1965. 251 p.

-----, & Morise HALLE. *The sound pattern of English*

DANCE, Frank E. X., comp. *Teoria da comunicação humana*. São Paulo Cultrix, 1973. 420 p.

ELGIN, Suzette Haden *Que é lingüística?* Rio de Janeiro, Zahar, 1974. 159 p.

GUIMARÃES ROSA, J. *Tutaméia (Terceiras estórias)*. Rio de Janeiro, J. Olympio, 1969. 201 p.

LANGACKER, Ronald W. *A linguagem e sua estrutura; alguns conceitos lingüísticos fundamentais*. Petrópolis, Vozes, 1972. 262 p.

LEROY, Maurice. *As grandes correntes da lingüística moderna*. 3. ed. São Paulo, Cultrix, 1974 193 p.

MARTIN, John W. "Concordância", *Revista Brasileira de Lingüística*, n.º 3, Petrópolis. No prelo.

MATTOSO CÂMARA JR., Joaquim. *Estrutura da língua portuguesa*, Petrópolis, Vozes, 1970. 114 p.